

Optimal Confidence Sets, Bioequivalence, and the Limaçon of Pascal

Lawrence D. Brown¹

George Casella²

J. T. Gene Hwang¹

Cornell University, Ithaca, N. Y. 14853

BU-1205-M

April 1993

Key words and phrases: decision theory, Bayes estimation, frequentist estimation, hypothesis testing, uniformly most powerful, uniformly most accurate

AMS 1990 Subject Classification: Primary 62F25, 62C99; Secondary: 62F03

¹ Research supported by National Science Foundation Grant No. DMS9107842.

² Research supported by National Science Foundation Grant No. DMS9100839 and National Security Agency Grant No. 90F-073.

Abstract

We begin with a decision-theoretic investigation into confidence sets that minimize expected volume at a given parameter value. Such sets are constructed by inverting a family of uniformly most powerful tests, hence also enjoy the optimality property of being uniformly most accurate. In addition, these sets possess Bayesian optimal volume properties, and represent the first case (to our knowledge) of a frequentist $1-\alpha$ confidence set that possesses a Bayesian optimality property.

The hypothesis testing problem that generates these sets is similar to that encountered in bioequivalence testing. Our sets are optimal for testing bioequivalence in certain settings, and in the case of the normal distribution, the optimal sets are curves known as the limaçon of Pascal. We illustrate the use of these curves with an example.

1. Introduction

The construction of good set estimates of a parameter, both frequentist and Bayesian, has long been a goal of statisticians. The formalization of “good” set estimates is usually in terms of some measure of the size of the set, often taken to be the volume of the set. Alternatively, the size of the set can be measured by its probability of false coverage. Thus, if $C(x)$ is a set estimate of a parameter θ , $P_\theta(\theta \in C(X))$ is the probability of true coverage, while $P_\theta(\theta' \in C(X))$, $\theta \neq \theta'$, is the probability of false coverage.

The false coverage of $C(x)$ can be related to its volume through the Ghosh-Pratt identity (Ghosh 1961; Pratt 1961),

$$E_\theta \text{vol}(C(X)) = \int P_\theta(\theta' \in C(X)) d\theta', \quad (1.1)$$

but this has rarely been used in establishing volume optimality. An exception is the work of Cohen and Strawderman (1973).

Equation (1.1) illustrates that possession of an optimal expected volume is a somewhat stronger property than possessing optimal false coverage probabilities, since expected volume can be regarded as a sum over all false coverages. Since admissibility with respect to expected volume implies admissibility with respect to false coverage probability, a procedure with optimal expected volume will have attractive false coverage properties. However, the converse is not true. As domination of false coverage probabilities ties directly into testing theory, where much is known about optimality, we find many cases where set estimates with optimal false coverage properties do not have optimal volume properties. For example, the usual multivariate normal confidence set cannot be uniformly dominated in false coverage, but it can be dominated in volume (see Casella and Hwang 1983 or Shinozaki (1989).

There is, however, an instance in which false coverage and volume are equivalent. That is when there is interest in producing a procedure that is optimal at some point in the parameter space. Thus, if there is interest in minimizing volume at a parameter value $\theta = \theta^*$, then this can be accomplished by minimizing all of the false coverages at $\theta = \theta^*$. Doing so brings the construction of optimal volume confidence sets back into a Neyman-Pearson testing set-up.

At first it may seem surprising that one can construct a confidence set that has optimal size at $\theta = \theta^*$ while maintaining a nominal coverage probability for all parameter values. However, this problem is a version of what was solved by Sterne (1954) in the binomial case (see also Crow 1956). For $X \sim \text{binomial}(n, p)$, Sterne proposed to construct a confidence set for p by inverting acceptance regions composed of the fewest X values necessary to have a rejection region with prespecified size α . He noted that such a set minimized the sum of the $n+1$ lengths. It turns out that such a construction, which is a Neyman-Pearson-type construction, will yield sets of minimum volume at $\theta = \theta^*$.

Interestingly, there is another aspect to the construction outlined here. We will see that the process of minimizing the expected volume at a value $\theta=\theta^*$ can also be used to minimize a Bayesian expected volume, that is, an expected volume integrated over a prior distribution. Thus, our construction gives a frequentist confidence set (one that maintains a nominal coverage probability) that optimizes a Bayesian measure of volume.

In Section 2 we formalize the decision-theoretic problem, and establish an optimality theorem in the frequentist setting. We also consider the normal case in detail, where the limaçon of Pascal appears. Section 3 connects these results to the Bayesian formulation, and shows how to construct optimal frequentist/Bayes intervals. Section 4 addresses the unknown variance case, and we see that the known variance optimality results can be generalized to this case. Lastly, Section 5 discusses the connections to the problem of bioequivalence testing. We also present an example showing how bioequivalence confidence sets based on the limaçon can provide sharper inferences.

2. The Frequentist Interpretation

2.1 A General Formulation

Let X have continuous density (for convenience) $f(\cdot|\theta)$ with respect to Lebesgue measure. Given that we observe $X = x$, we set up a confidence set for θ . This set, $C(x)$, may be a randomized set, and has inclusion probability

$$P(\theta \in C(x) | x) = \varphi(\theta | x). \quad (2.1)$$

That is, $\varphi(\theta | x)$ is the probability of including the value θ in the set when x is observed. For nonrandomized sets $\varphi(\theta | x) = I(\theta \in C(x))$, the indicator function of the set $C(x)$. (Randomized rules are only included for completeness of the theory, as they should never be recommended in practical applications.)

The volume of the set $C(x)$, $\text{vol}(C(x))$, with respect to Lebesgue measure, is given by

$$\text{vol}(C(x)) = \int_{\Theta} \varphi(t | x) dt, \quad (2.2)$$

with expected volume

$$E_{\theta} \text{vol}(C(X)) = \int_{\mathfrak{S}} \text{vol}(C(x)) f(x | \theta) dx. \quad (2.3)$$

In addition to calculating (2.3) as a measure of size, it is usual to calculate the frequentist coverage probability of the set $C(x)$, that is

$$P_{\theta}(\theta \in C(X)) = \int_{\mathfrak{S}} \varphi(\theta | x) f(x | \theta) dx. \quad (2.4)$$

A standard frequentist requirement is to have this coverage probability greater than some nominal level, say $1-\alpha$, for all values of θ . Subject to that constraint, we seek to minimize the expected volume of $C(x)$ at a selected value of θ . Without loss of generality we take $\theta=0$ and, to avoid trivial pathologies, $0 < \alpha < 1$. Thus, the problem of interest becomes:

$$\begin{aligned} &\text{Over all confidence sets } C(x), \text{ minimize } E_0 \text{vol}(C(X)) \text{ subject to} \\ &P_0(\theta \in C(X)) \geq 1-\alpha \text{ for all } \theta. \end{aligned} \quad (2.5)$$

Before stating and proving a formal theorem, note that the specifying of $\theta=0$ in the volume requirement, which puts a particular importance on this value, makes one think of a hypothesis testing formulation. However, the formulation of a hypothesis test that is equivalent to (2.5) is not entirely straightforward, for the specified value $\theta=0$ is not part of the null hypothesis, but of the alternative hypothesis.

Consider testing

$$H_0: \theta = \theta_0 \text{ vs } H_1: \theta = 0 \quad (2.6)$$

where $X \sim f(\cdot | \theta)$. The most powerful (Neyman-Pearson) size α test is given by a rejection rule satisfying

$$\psi_{\theta_0} = \begin{cases} 1 & \text{if } f(x|0) > k(\theta_0)f(x|\theta_0) \\ 0 & \text{otherwise} \end{cases},$$

for which $E_{\theta_0}(\psi_{\theta_0}(X)) = \alpha$. As in (2.1), define confidence sets C^* with

$$\phi^*(\theta|x) = 1 - \psi_{\theta}(x). \quad (2.7)$$

It is this confidence set that solves the problem in (2.5). Note that in the usual non-randomized case the confidence set is

$$C^*(x) = \left\{ \theta : f(x|\theta) > f(x|0) / k(\theta) \right\}. \quad (2.8)$$

Theorem 2.1: Let $X \sim f(\cdot | \theta)$, and let $\phi^*(\theta|x)$ be given by (2.7). The confidence set $C^*(x)$ minimizes the expected volume at $\theta = 0$ among all $1-\alpha$ confidence sets.

Proof: The proof is based on the Ghosh-Pratt identity (Ghosh 1961; Pratt 1961), and can be found in Pratt's paper (along with the one-dimensional normal example). We have for any confidence set $C(x)$

$$E_0 \text{vol}(C(X)) = \int_{\mathfrak{X}} \text{vol}(C^*(x)) f(x|0) dx = \int_{\mathfrak{X}} \int_{\Theta} \varphi(\theta|x) d\theta f(x|0) dx,$$

where $\varphi(\theta|x)$ is the probability that θ is included in $C(x)$ when x is observed. Interchanging the order of integration gives

$$E_0 \text{vol}(C(X)) = \int_{\Theta} \int_{\mathfrak{X}} \varphi(\theta|x) f(x|0) dx d\theta = \int_{\Theta} P_0(\theta \in C(X)) d\theta. \quad (2.9)$$

The integrand in (2.9) is the probability of false coverage, which is minimized, subject to (2.5), by the uniformly most accurate set $C^*(x)$, and in turn produces the minimum expected volume. \square

It should be noted that any weighted volume measure can be used and Theorem 2.1 would remain valid. That is, if we measure the size of a set $C(x)$ by

$$E_{\theta} \text{size}(C(X)) = \int_{\mathfrak{X}} \left[\int_{\Theta} \varphi(\theta|x) \nu(\theta) d\theta \right] dx, \quad (2.10)$$

where $\nu(\cdot) > 0$ is some weight function, then $C^*(x)$ of (2.8) minimizes (2.10) over all $1-\alpha$ confidence sets.

2.2 The Normal Case

To better understand the behavior of $C^*(x)$, we look at it more closely in the normal case. If X has a p -variate normal distribution, $X \sim N_p(\theta, I)$, then for the hypothesis test (2.6) we would reject

H_0 if

$$\frac{e^{-\frac{1}{2}|x|^2}}{e^{-\frac{1}{2}|x-\theta_0|^2}} > k(\theta_0) \Leftrightarrow x'\theta_0 < k^*(\theta_0),$$

where $k^*(\theta_0)$ is chosen to give the test size α . Thus, the confidence set is $C^*(x) = \{\theta : x'\theta \geq k^*(\theta)\}$. To evaluate the form of $k^*(\theta)$, we use the fact that $W = \theta'X/|\theta| \sim n(|\theta|, 1)$, regardless of the dimension of X . Then, for $\theta \neq 0$ the coverage probability is

$$\begin{aligned} P_\theta(\theta \in C^*(X)) &= P_\theta(X'\theta \geq k^*(\theta)) \\ &= P(|\theta|W \geq k^*(\theta)) \\ &= \Phi(a) \end{aligned}$$

for $k^*(\theta) = |\theta|(|\theta| - a)$, where $\Phi(\cdot)$ is the standard normal cdf. Choosing $a = \Phi^{-1}(1 - \alpha)$ yields a $1 - \alpha$ confidence interval.

To better understand the shape of $C^*(x)$, write $x'\theta = |x| |\theta| \cos \beta$, where $\cos \beta = x'\theta/|x| |\theta|$, and β is the angle between x and θ . Then, in the normal case the optimal confidence set is

$$C^*(x) = \{\theta : |\theta| \leq a + |x| \cos \beta\}. \quad (2.10)$$

The boundary of this set is the main lobe of a curve known as the limaçon of Pascal, a curve that is often used in calculus courses to illustrate polar coordinate techniques. (The limaçon was actually studied by Etienne Pascal, the father of the famous Blaise Pascal, see Archibald 1900). The limaçon is shown in Figure 1, and the confidence set (2.10) is graphed in Figure 3 for various values of x when $p = 2$. It is interesting to note that when $x = 0$ the set is a sphere, but as x moves away from zero there is a distinct nonconvexity to the set. (The limaçon is actually a generalization of the cardioid, a "heart-shaped" polar curve.) As x tends toward infinity the limaçon becomes more like a sphere. In higher dimensions, the limaçon shape is retained. If we graph a higher dimensional limaçon by identifying the x -axis with the data x , and using β as the angle between θ and x , equation (2.10) will resemble the limaçon in Figure 1. The remainder of the set is then generated by rotating the limaçon about the x -axis.

Notice that $C^*(x)$ has coverage probability $\Phi(a)$ regardless of the dimension of the problem. In one dimension some further simplifications can be made. Here we have $W = |X| \text{sgn}(\theta X) \sim n(|\theta|, 1)$, and

$$C^*(x) = \{\theta : |x| \text{sgn}(\theta x) \geq |\theta| - a\} = \{\theta : \min(0, x - a) \leq \theta \leq \max(0, x + a)\}. \quad (2.11)$$

The 95% confidence interval has $a = 1.645$, and is equal to $x \pm 1.645$ for small $|x|$. The usual two-sided 95% confidence interval is $x \pm 1.96$. Thus, $C^*(x)$ is narrower than the usual interval for small

values of x , but wider for larger values. $C^*(x)$ does have an interesting property, however, in that it falls on one side of zero for smaller values of $|x|$ than the usual interval, thus allowing an experimenter more power in making formal conclusions about the sign of θ . Figure 2 compares $C^*(x)$ with the usual one-dimensional interval.

As mentioned before, the one-dimensional $C^*(x)$ was first derived by Pratt (1961), who also discussed the connection with the Sterne-Crow intervals for a binomial success probability (Crow 1956). However, the interval (2.11) has another history, emerging from the work of Hsu (1981, 1984). He derived the interval in the context of a multiple decision problem, where one is interested in confidence intervals for the distance from the best mean. A discussion of $C^*(x)$, and demonstration of some properties, can be found in Exercise 9.31 of Casella and Berger, 1990.

The expected volume, at $\theta = 0$, can also be evaluated for $C^*(x)$. Since $C^*(x) = \{\theta : |\theta| \leq w + a\}$, where $W \sim n(|\theta|, 1)$, we have

$$\begin{aligned} E_0 \text{vol}(C^*(X)) &= \int_{\Theta} \left[\int_{\mathfrak{E}} I(\theta \in C^*(x)) f(x|0) dx \right] d\theta \\ &= \int_{\Theta} \left[\int_{\left\{x: \frac{x'\theta}{|\theta|} \geq |\theta| - a\right\}} f(x|0) dx \right] d\theta \\ &= \int_{\Theta} \Phi(a - |\theta|) d\theta, \end{aligned} \tag{2.12}$$

where the last equality follows from the fact that for $X \sim N(0, I)$, $X'\theta/|\theta| \sim N(0, 1)$ for any nonzero θ . If we then apply a polar transformation we have

$$\begin{aligned} E_0 \text{vol}(C^*(X)) &= \frac{\pi^{p/2}}{\Gamma(p/2+1)} \int_0^\infty r^{p-1} \Phi(a-r) dr \\ &= \frac{\pi^{p/2}}{\Gamma(p/2+1)} \int_{-\infty}^a \frac{(a-t)^p}{p} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt. \end{aligned} \tag{2.13}$$

Note that $\pi^{p/2} / \Gamma(p/2 + 1)$ is the volume of a p -sphere of radius 1, so the p^{th} root of the integral in (2.13) is, effectively, the radius of the set. For $p = 1$ we can write

$$E_0 \text{vol}(C^*(X)) = 2 \left[a\Phi(a) + \frac{e^{-a^2/2}}{\sqrt{2\pi}} \right],$$

but for other values of p the integral is harder to evaluate. Table 1 gives some values of the p^{th} root of $E_0 \text{vol}(C^*(x))$ and, for comparison, the corresponding values for $C_0(x)$, the usual confidence sphere.

Table 1
 Effective volume (p^{th} root of expected volume) of $C^*(x)$ and $C_0(x)$,
 the usual confidence sphere

<u>p</u>	<u>1-α</u>	<u>$[E_0 \text{ vol } C^*(X)]^{1/p}$</u>	<u>$[E_0 \text{ vol } (C_0(X))]^{1/p}$</u>	<u>Ratio</u>
1	.90	2.66	3.29	0.809
1	.95	3.33	3.92	0.849
1	.99	4.66	5.15	0.905
3	.90	2.03	4.03	0.504
3	.95	2.36	4.50	0.524
3	.99	3.01	5.45	0.552
10	.90	2.35	4.39	0.535
10	.95	2.58	4.69	0.550
10	.99	3.04	5.29	0.575

Of course, for $\theta \neq 0$, $E_\theta \text{ vol}(C^*(X))$ will grow larger than $E_\theta \text{ vol}(C_0(X))$ (which is constant in θ), the discrepancy increasing as $|\theta|$ increases. We illustrate this with Figure 3, which compares realized values of the two sets for a variety of x values. Note that the different graphs have different scales, and the value of x is the center of the sphere.

3. The Bayes/Frequentist Interpretation

Interestingly, the same mathematical technique that produces the $1-\alpha$ confidence set of minimum expected volume at a particular θ also minimizes the expected Bayesian volume using a prior for θ . If $X \sim f(\cdot | \theta)$ where θ has a prior distribution π , then the expected Bayesian volume of a set $C(x)$ is

$$E_{\pi} \text{vol}(C(X)) = \int_{\Theta} \left[\int_{\mathfrak{S}} \text{vol}(C(x)) f(x | \theta) dx \right] \pi(\theta) d\theta. \quad (3.1)$$

We now seek to minimize (3.1), a Bayesian measure, among all sets $C(x)$ that satisfy the frequentist coverage probability constraint, $P_{\theta}(\theta \in C(x)) \geq 1-\alpha$.

In Section 2, the minimizing set was constructed from testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = 0$ or, equivalently, $H_0 : X \sim f(x | \theta_0)$ vs. $H_1 : X \sim f(x | 0)$. In the Bayesian formulation, the minimizing set is constructed from the test

$$H_0 : X \sim f(x | \theta_0) \quad \text{vs} \quad H_1 : X \sim m_{\pi}(x),$$

where $m_{\pi}(x) = \int_{\Theta} f(x | \theta) \pi(\theta) d\theta$. The confidence set is, therefore, given by

$$C_{\pi}^*(x) = \left\{ \theta : f(x | \theta) \geq m_{\pi}(x) / k(\theta) \right\}, \quad (3.2)$$

where $k(\theta)$ is chosen so that $P_{\theta}(\theta \in C_{\pi}^*(x)) = 1 - \alpha$. We have the following theorem

Theorem 3.1: Let $X \sim f(x | \theta)$, $\theta \sim \pi(\theta)$, and $C_{\pi}^*(x)$ be given by (3.2). The confidence set $C_{\pi}^*(x)$ minimizes the expected Bayesian volume (3.1) among all $1-\alpha$ confidence sets.

Proof: For any confidence set $C(x)$ we have

$$E_{\pi} \text{vol}(C(X)) = \int_{\Theta} \left[\int_{\mathfrak{S}} \text{vol}(C(x)) f(x | \theta) dx \right] \pi(\theta) d\theta = \int_{\mathfrak{S}} \text{vol}(C(x)) m_{\pi}(x) dx.$$

Now proceed as in the proof of Theorem 2.1, with $m_{\pi}(x)$ in place of $f(x | 0)$. □

For illustration, consider again the normal case $X \sim N(\theta, 1)$ and $\theta \sim N(0, \tau^2 1)$. The marginal distribution of X is $N(0, (\tau^2+1)1)$, and the confidence set is

$$C_{\pi}^*(x) = \left\{ \theta : \left| \theta - \frac{\tau^2}{\tau^2+1} x \right|^2 \leq k^*(\theta) \right\} \quad (3.3)$$

where $(\tau^2+1 / \tau^2)^2 k^*(\theta)$ is the upper α critical point of a noncentral chi-squared distribution with noncentrality parameter $|\theta|^2 / \tau^2$. It can be shown that as $\tau^2 \rightarrow 0$ this set reduces to $C^*(x)$ of the previous section and, as $\tau^2 \rightarrow \infty$, this set approaches the usual sphere $C_0(x)$.

4. Generalizations to the Case of Unknown Variance

The set $C^*(x)$ of (2.8) is optimal in cases where there are no nuisance parameters, and the normal examples of Section 2.2 all reflect this. Of course, the more practical problems usually involve nuisance parameters, and we now consider that case. We restrict our discussion to the normal distribution with unknown mean and variance.

With a sample X_1, \dots, X_n from $n(\theta, \sigma^2)$ with both parameters unknown, there are two ways of generalizing the procedure of Section 2. The first, which is perhaps the more obvious way, is to test the hypotheses

$$H_0 : \theta = \theta_0, \sigma > 0 \quad \text{vs} \quad H_1 : \theta = 0, \sigma > 0. \quad (4.1)$$

Using a standard Student's t test, this leads to intervals of the form

$$C_t(\bar{x}, s) = \left\{ \theta : \min\left(0, \bar{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}}\right) \leq \theta \leq \max\left(0, \bar{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}}\right) \right\}, \quad (4.2)$$

where $t_{\alpha, n-1}$ is the upper α cutoff from Student's t distribution with $n-1$ degrees of freedom. It is straightforward to verify that $C_t(\bar{x}, s)$ is a $1-\alpha$ confidence interval, although it doesn't enjoy the same optimality properties as the interval (2.8). This interval was also considered by Hsu, Hwang, Liu, and Ruberg (1993), although they did not investigate its optimality. We detail its exact optimality below.

For this problem it is natural to consider only confidence sets related to the usual (scale-invariant) t-tests of H_0 . This means that the inclusion probabilities of the confidence set must be of the form

$$\varphi(\theta | x_1, \dots, x_n) = \varphi\left(\theta \mid \frac{\bar{x} - \theta}{s}\right). \quad (4.3)$$

Note that the intervals of (4.2) have this form.

A second, perhaps less obvious way of generalizing Section 2.2 is to modify the hypotheses of (2.6) by dividing by σ to obtain

$$H_0 : \frac{\theta}{\sigma} = \eta_0 \quad \text{vs.} \quad H_1 : \frac{\theta}{\sigma} = 0, \quad (4.4)$$

where η_0 is a fixed constant. On defining $\eta = \theta/\sigma$, we see that we are reduced to considering a one-parameter problem. In practical terms, the hypotheses (4.4) are also quite interesting, because the "signal-to-noise ratio" $\eta = \theta/\sigma$ is often of interest. We will see that (4.4) leads to a confidence interval for η that is different from any confidence sets for θ , in particular (4.2).

For the hypotheses (4.4) a reasonable invariant procedure will be of the form

$$\varphi(\eta | x_1, \dots, x_n) = \varphi(\eta | \bar{x}/s), \quad (4.5)$$

with corresponding confidence intervals given by

$$C_t^*(\bar{x}, s) = \left\{ \eta : f(\bar{x}/s | \eta) \geq f(\bar{x}/s | 0)/k(\eta) \right\}, \quad (4.6)$$

where $f(\bar{x}/s | \eta)$ is the noncentral t distribution with noncentrality parameter η and $n-1$ degrees of freedom. The function $k(\eta)$ is chosen so that C_t^* is a $1-\alpha$ confidence set, that is, so that the

corresponding tests in (4.4) have level α . Figure 4 shows a plot of these intervals, along with the normal (σ known) intervals. It is interesting to note that the resulting boundaries are curved, in contrast to the straight line boundaries in the known σ case.

Construction of the intervals given by (4.6) is actually quite straightforward, exploiting monotonicity properties of both the density and distribution function of the noncentral t . Note first that since $f(t|\eta)$ has monotone likelihood ratio, the acceptance region of the test (4.4), that is $H_0:\eta$ vs. $H_1:\eta = 0$, is given by

$$A_\eta(t) = \begin{cases} \{t:t \geq k_1(\eta)\} & \text{if } \eta > 0 \\ \{t:t \leq k_2(\eta)\} & \text{if } \eta < 0 \end{cases}.$$

Now the distribution function, $F(t|\eta)$, is decreasing in η , so an α -level test is constructed by solving

$$F(t|k_1^{-1}(t)) = \alpha \quad \text{and} \quad F(t|k_2^{-1}(t)) = 1 - \alpha,$$

and setting $\eta_U(t) = k_1^{-1}(t)$ and $\eta_L(t) = k_2^{-1}(t)$ yields

$$C_t^*(\bar{x}/s) = \left\{ \eta: \min(0, \eta_L(\bar{x}/s)) \leq \eta \leq \max(0, \eta_U(\bar{x}/s)) \right\}.$$

Note that this construction holds, in general, as long as the density satisfies suitable monotonicity conditions. It also follows that the confidence interval $(\eta_L(t), \eta_U(t))$ is a $1 - 2\alpha$ interval.

For both the set-ups leading to (4.1) and (4.4) we can establish optimality properties of the resulting confidence sets.

Theorem 4.1: Among all $1 - \alpha$ confidence sets for θ of the form (4.3), the intervals (4.2) minimize $E_{0,\sigma}(\text{vol}(C(\bar{X}, S)))$ for every $\sigma^2 > 0$.

Proof: As in the proof of Theorem 2.1, use the Ghosh-Pratt identity to write

$$E_{0,\sigma}(\text{vol}(C(\bar{X}, S))) = \int_{-\infty}^{\infty} P_{0,\sigma}(\theta \in C(\bar{X}, S)) d\theta.$$

Among all $1 - \alpha$ confidence sets of the form (4.4) the integrand is minimized for each θ by the C_t of (4.2), because these intervals correspond to most powerful level α invariant tests of the hypotheses (4.1). □

Theorem 4.2: Among all invariant $1 - \alpha$ confidence sets for $\eta = \theta/\sigma$, the intervals (4.6) minimize $E_{0,\sigma}(\text{vol}(C(\bar{X}, S)))$.

Proof: This follows from Theorem 2.1 upon taking into account the monotone likelihood ratio property of the noncentral t distribution. □

Without the restriction (4.3) the confidence intervals (4.2) are not optimal. In fact, they are not even admissible, since it is possible to construct $1 - \alpha$ confidence intervals with smaller expected length for every $\theta = 0, \sigma > 0$. This is possible because the results of Brown and Sackrowitz (1984) enables construction of level α tests, $\varphi_{BS}(\theta|\bar{X}, S)$ of (4.1), whose power strictly dominates that of the one-

sided t-tests leading to (4.2) Consequently, this family of critical functions leads to intervals which dominate those of (4.2), as can be seen by applying the Ghosh-Pratt identity as in the proof of Theorem 4.1.

For each θ, σ the $(1-\alpha)$ confidence intervals constructed above have coverage exceeding $1-\alpha$, since the tests of Brown and Sackrowitz have size smaller than α . But $\inf \left\{ P_{\theta, \sigma}(\theta \in C_{BS}(\bar{X}, S)) \mid \theta, \sigma \right\} = 1-\alpha$. We do not know whether the intervals C_{BS} are admissible. Brown and Sackrowitz show their tests are admissible for testing $H_0: \theta = \theta_0, \sigma > 0$ vs. either $H_1: \theta > 0$ if $\theta_0 < 0$ or $H_1: \theta < 0$ if $\theta_0 > 0$; however, they do not prove admissibility for H_0, H_1 of (4.1) as would be needed to establish admissibility of C_{BS} .

By contrast, the intervals in Theorem 4.2 are admissible for μ/σ . This is because the tests involved are UMP invariant and hence are Bayes among invariant tests. The invariance group here is the group of scale transformations and, applying the results of Brown and Fox (1974), a procedure that is Bayes among invariant procedures is admissible. Admissibility of the tests implies admissibility of the confidence sets.

Generalizations to higher dimensions should also be of interest. Presumably, the limaçon does not appear here because the relevant distributions are multivariate t rather than multivariate normal as in Section 3.

5. Connections with Bioequivalence

The intervals discussed here, particularly the form (4.2), have seen practical use in bioequivalence testing. This brings us to the interesting connection between bioequivalence and the limaçon of Pascal.

The problem of declaring bioequivalence is typically that of deciding whether the difference of two population means, $\mu_1 - \mu_2$, is close to zero. Typically, μ_1 and μ_2 represent mean responses of two different types of treatments or drugs (treatments vs. control, oral vs. injection, brand name vs. generic) and the interest (usually of pharmaceutical companies) is to demonstrate that the effects are equivalent, yielding bioequivalent formulations of the treatment. At present, a typical statistical approach is to test

$$H_0: |\mu_1 - \mu_2| > \delta \quad \text{vs.} \quad H_1: |\mu_1 - \mu_2| < \delta, \quad (5.1)$$

where δ is a specified threshold. Rejection of such a test leads to the declaration of bioequivalence. Note that what is typically the "null hypothesis" is placed in the alternative, as this is the research hypothesis of interest. One of the first researchers to formulate the bioequivalence problem in this was Anderson and Hauck (1983). (See also Hauck and Anderson 1984, 1992. The latter is a review paper.)

A technique of carrying out a test of (5.1), as required by the FDA (FDA, 1992), is to perform two one-sided tests, as described by Schuirmann (1987), for example. This procedure establishes bioequivalence of μ_1 and μ_2 , at level α , if both of the following two one-sided test of $\theta = \mu_1 - \mu_2$ reject the null hypothesis at level α :

$$\begin{aligned} \text{i) } H_0: \theta \leq -\delta \quad \text{vs.} \quad H_1: \theta > -\delta \\ \text{ii) } H_0: \theta \geq \delta \quad \text{vs.} \quad H_1: \theta < \delta. \end{aligned} \quad (5.2)$$

It is interesting to note that (5.2) is a case of an intersection-union test, as developed by Berger (1982). As the overall hypothesis of interest, that $-\delta \leq \theta \leq \delta$ is an intersection of the two other hypotheses, individual α -level tests lead to an overall α -level test for (5.1). Additionally, referring to the discussion following (4.6), it turns out that Schuirmann's procedure leads to basing the conclusion on the non-truncated confidence interval, which seems to have confidence $1 - 2\alpha$, the wrong level. The α -level is actually correct, however, the intervals are wider than necessary.

Although interest has usually centered around the testing problem in (5.1), there is also considerable interest (and benefit) in constructing confidence intervals for $\mu_2 - \mu_1$. Given a $1 - \alpha$ confidence interval $C(x)$ for $\mu_2 - \mu_1$, an α -level test of (5.1) can be conducted by rejecting H_0 whenever $C(x) \subset (-\delta, \delta)$, as proposed by Westlake (1972, 1976). (Notice that this use of confidence sets as tests is reversed from the usual use. The values of $\mu_2 - \mu_1$ in the interval $(-\delta, \delta)$ are those for which bioequivalence will be declared, which is the alternative hypothesis.)

The confidence set approach to bioequivalence testing has an added benefit in that the constant δ does not have to be prespecified. Indeed, for a given $1-\alpha$ bioequivalence confidence set $C(x)$, and any set Δ , we could conclude $H_1: \mu_1 - \mu_2 \in \Delta$ if $C(x) \subset \Delta$. It then follows that $C(x)$ is the smallest set of parameter values $\mu_1 - \mu_2$ for which the data will reject H_0 (and hence conclude bioequivalence).

Recent research has continued to take the confidence interval approach to bioequivalence. The work of Westlake (1976) was apparently prompted by the mismatch in error probabilities of Schuirmann's procedure (α for the test and 2α for the interval). Although Westlake's $1-\alpha$ interval does not correspond to Schuirmann's α -level test, it led to the further research of Bofinger (1985) and Hsu *et al.* (1993), who proposed the interval (4.2). This $1-\alpha$ interval, which is shorter than Westlake's (1976) interval, corresponds to the level α two-one-sided tests procedure. Generalizations of this problem, including a nonparametric approach, are discussed in Hsu *et al.* (1993).

There are other formulations of the bioequivalence hypothesis, formulations that lead to alternate tests and confidence intervals. For example, one could specify δ of (5.1) in terms of the variance σ ; that is, two procedures are declared bioequivalent if their mean difference is no more than a specified proportion of their variance. Of course, this formulation leads directly to the hypotheses specified in (4.4), and to the noncentral t-based intervals of (4.6). Although this formulation of the bioequivalence problem has been used, the optimal procedure of (4.6) has not been employed.

Lastly, the bioequivalence problem can be a multivariate one, where the full advantage of the limaçon can be enjoyed. To test the bioequivalence of p formulations, one might specify a set Δ (possibly a hyperrectangle) in which the differences must lie. By constructing a limaçon confidence set, overall bioequivalence can be examined. Bioequivalence would be concluded if the limaçon confidence set falls entirely within Δ .

We illustrate this last situation with data from Ruberg and Stegeman (1991), on equivalence of batch degradation slopes. (We make a simplifying (but seemingly reasonable) assumption about the batch variation. For an alternative analysis, based on multiple comparison techniques, see Ruberg and Hsu (1992).) Each of $p=6$ batch slopes, β_1, \dots, β_6 , is used to estimate the shelf-life of a product. Here we look at the bioequivalence of the differences

$$\theta_i = \beta_i - \beta_1, \quad i = 2, \dots, p.$$

The data are assumed to be independent normal with known variance; that is, we observe $\hat{\beta}_i \sim n(\mu, \sigma^2)$. This results in observed $\hat{\theta}_i = \hat{\beta}_i - \hat{\beta}_1$ that are correlated, $\hat{\theta} \sim N(\theta, \Sigma)$ where, in this case, $\Sigma = \sigma^2(I + J)$, J being a vector of ones. In general, for $\hat{\theta} \sim N(\theta, \Sigma)$, the $1-\alpha$ limaçon confidence set is

$$\left\{ \theta : \left(\theta' \Sigma^{-1} \theta \right)^{\frac{1}{2}} \leq z_\alpha + \frac{\hat{\theta}' \Sigma^{-1} \theta}{\left(\theta' \Sigma^{-1} \theta \right)^{\frac{1}{2}}} \right\}, \quad (5.2)$$

where z_α is the upper α cutoff point from a univariate standard normal distribution. This set reduces

to the form of (2.10) with the transformation $\eta = \Sigma^{\frac{1}{2}}\theta$, $\hat{\eta} = \Sigma^{\frac{1}{2}}\hat{\theta}$. The set (5.2) is somewhat of an elliptical limaçon, and is shown in Figure 5. We show two-dimensional projections of the five-dimensional set, which gives the smallest Δ for which bioequivalence will be declared.

Acknowledgment: We thank Jason Hsu for many interesting discussions about different aspects of this problem, and Weizhen Wang for numerical calculations leading the Figure 4.

References

- Anderson, S. and Hauck, W.W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Commun. Statist., Theor. Meth.* **12**, 2663-2692.
- Archibald (1900). Le limaçon de Pascal. *L'Intermédiaire des Mathématiciens*, 106-107.
- Berger, R.L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295-300.
- Bofinger, E. (1985). Expanded confidence intervals, one-sided tests, and equivalence testing. *J. Biopharm. Statistics* **2(2)**, 181-188.
- Brown, L.D. and Sackrowitz, H. (1984). An alternative to Student's t test for problems with indifference zones. *Ann. Statist* **12**, 451-469.
- Brown, L.D. and Fox, M. (1974). Admissibility in statistical problems involving a location or scale parameter. *Ann. Statist* **2**, 807-814.
- Casella, G. and Berger, R.L. (1990). *Statistical Inference*. Pacific Grove, California: Wadsworth and Brooks/Cole.
- Casella, G. and Hwang, J.T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**, 688-698.
- Cohen, A. and Strawderman, W.E. (1973). Admissibility implications for different criteria in confidence estimation. *Ann. Statist.* **1**, 363-366.
- Crow, E.L. (1956). Confidence intervals for a proportion. *Biometrika* **43**, 423-425.
- FDA (1992). Draft recommendation on statistical procedures for bioequivalence studies using the standard two-treatment crossover design. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD.
- Ghosh, J.K. (1961). On the relation among shortest confidence intervals of different types. *Calcutta Statistical Association Bulletin*, 147-152.
- Hauck, W.W. and Anderson, S. (1984). A new statistical procedure for testing equivalence in two-group comparative bioavailability trials. *J. Pharmacokinetic Biopharm.* **12**, 83-91.
- Hauck, W.W. and Anderson, S. (1992). Types of bioequivalence and related statistical considerations. *Int'l. J. Clin. Pharm. Ther. Toxic.* **30**, 181-187.
- Hsu, J.C. (1981). Simultaneous confidence intervals for all distances from the best. *Ann. Statist.* **9**, 1026-1034.
- Hsu, J.C. (1984). Constrained two-sided simultaneous confidence intervals for multiple comparisons with the best. *Ann. Statist.* **12**, 1136-1144.
- Hsu, J.C., Hwang, J.T.G., Liu, H.K., and Ruberg, S.J. (1992). Equivalence confidence intervals in bioequivalence trials. To appear in *Biometrika*.
- Pratt, J.W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56**, 541-567.

- Ruberg, S.J. and Hsu, J.C. (1992). Multiple comparison procedures for pooling batches in stability studies. *Technometrics* **34**, 465-472.
- Ruberg, S.J. and Stegeman, J.W. (1991). Pooling data for stability studies: Testing the equality of batch degradation slopes. *Biometrics* **47**, 1059-1069.
- Schuirmann, D.J. (1987). A comparison of the two one-sided test procedures and the power approach for assessing the equivalence of average bioavailability. *J. Pharm. Biopharm.* **15**, 657-680.
- Shinozaki, N. (1989). Improved confidence sets for the mean of a multivariate normal distribution. *Ann. Inst. Statist. Math.* **41**, 331-346.
- Sterne, T.E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41**, 275-278.
- Westlake, W.J. (1972). Use of confidence intervals in analysis of comparative bioavailability trials. *J. Pharm. Sci.* **61**, 1340-1341.
- Westlake, W.J. (1976). Symmetrical confidence intervals for bioequivalence trials. *Biometrics* **32**, 741-744.

Figure 1. The limaçon of Pascal $r = b + a \cos \beta$. If $b < a$ the limaçon has an inner loop, which does not occur in the confidence set. The confidence set $\{\theta: |\theta| \leq Z_\alpha + |x| \cos \beta\}$ is, in effect, a "positive-part" limaçon.

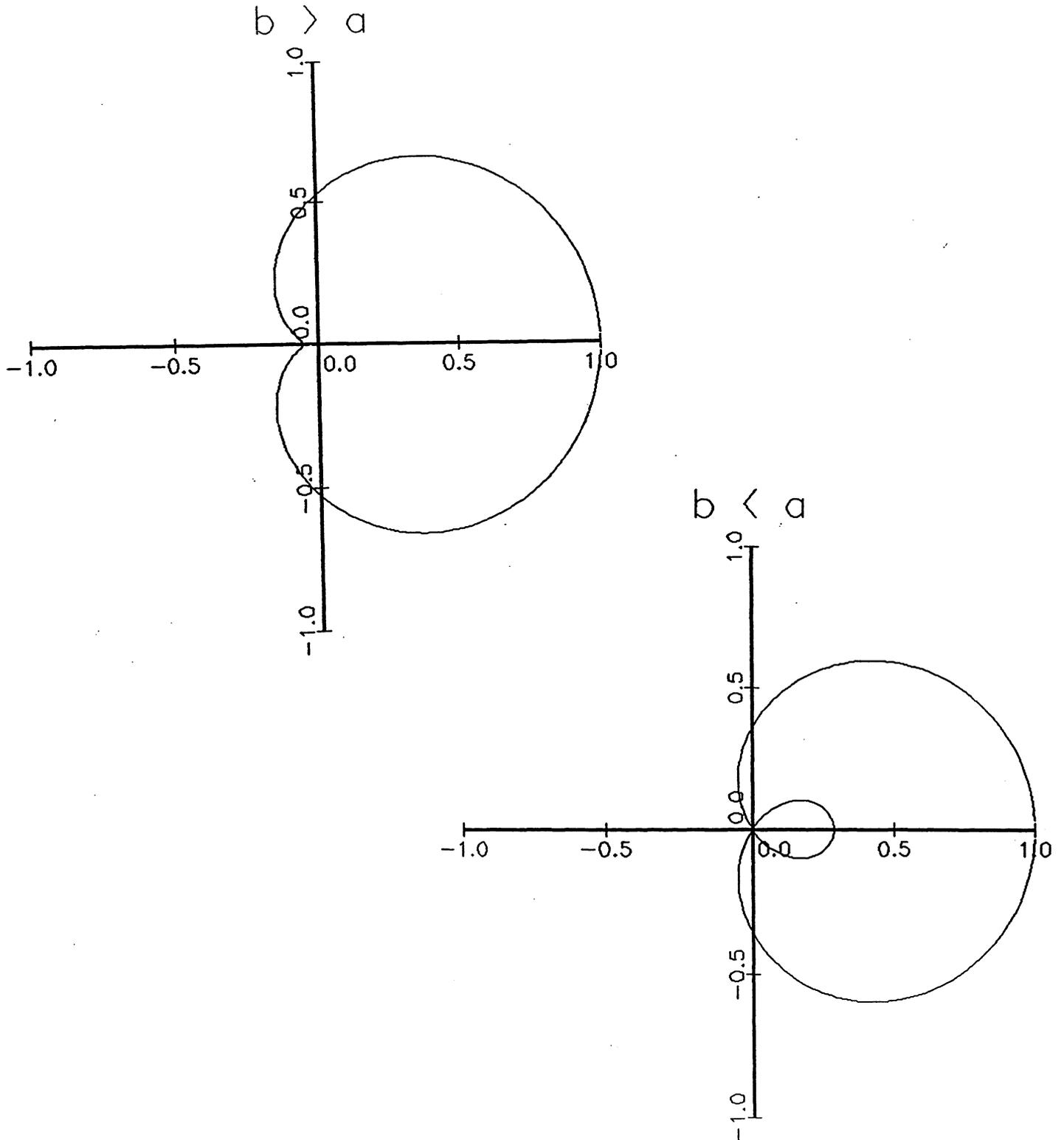


Figure 2. Comparison of $C^*(x)$ (solid lines) of (2.10) with the usual 90% confidence interval (dashed lines).

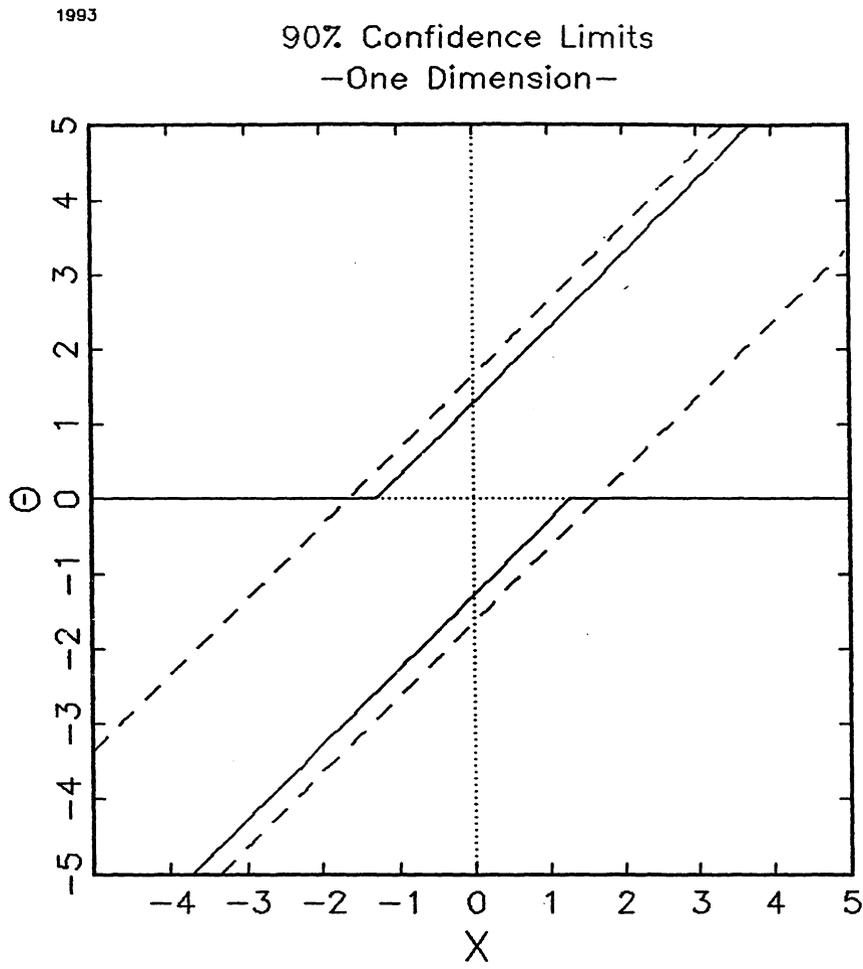
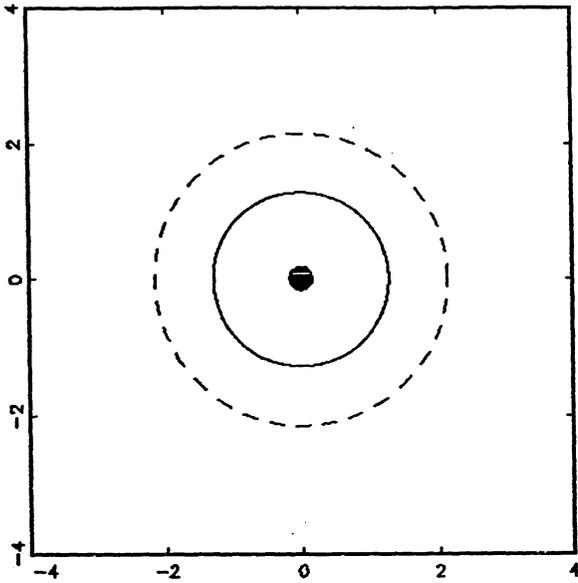
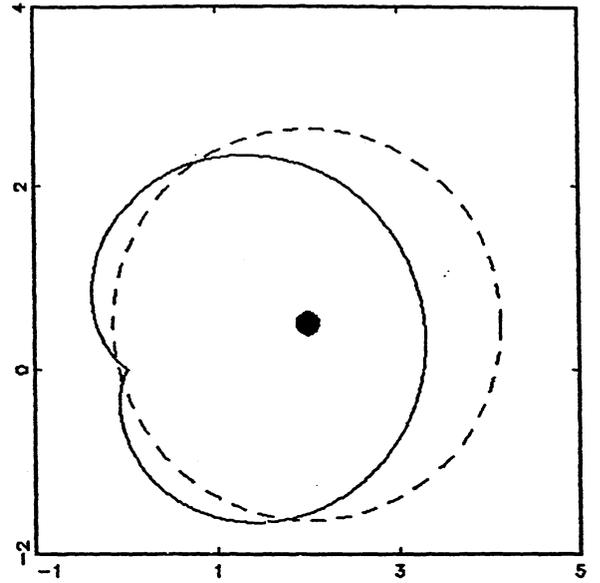


Figure 3. Two-dimensional limaçon (solid lines) and usual confidence sphere (dashed lines) for four different data points. The confidence sphere has a constant radius, while the limaçon enlarges as the data move away from the origin.

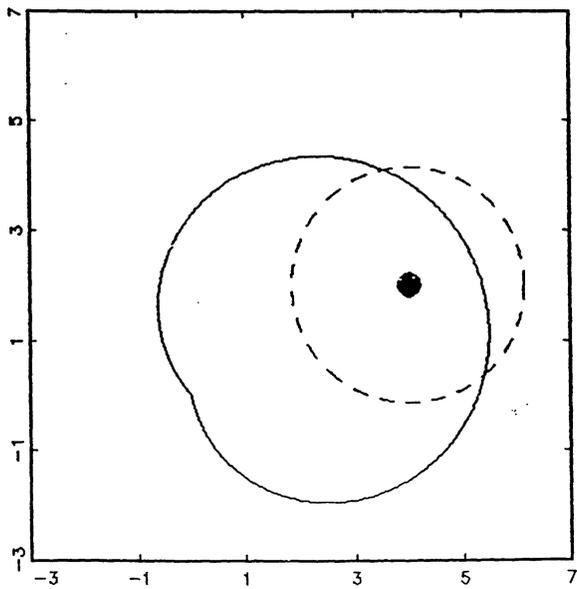
$x=(0,0)$



$x=(2,.5)$



$x=(4,2)$



$x=(10,7)$

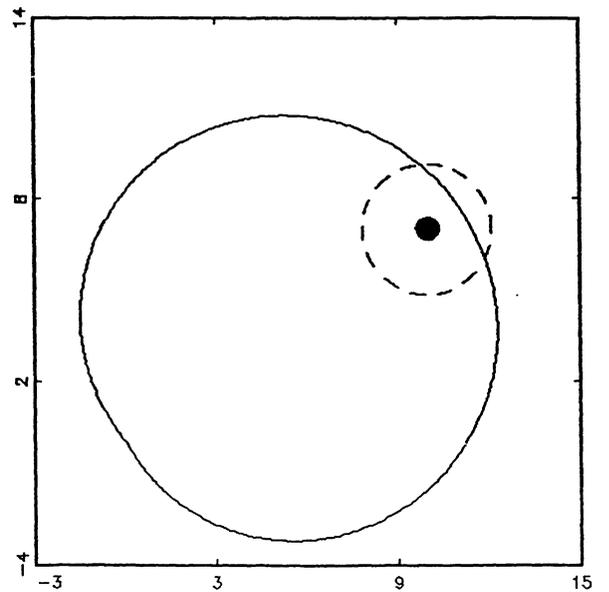


Figure 4. Comparison of C_t^* of (4.6) with the normal (known σ) interval. The dotted lines are the normal interval, and the noncentral t interval is shown for a 2 df (solid lines), 5 df (short dashed lines), and 20 df (long dashed lines).

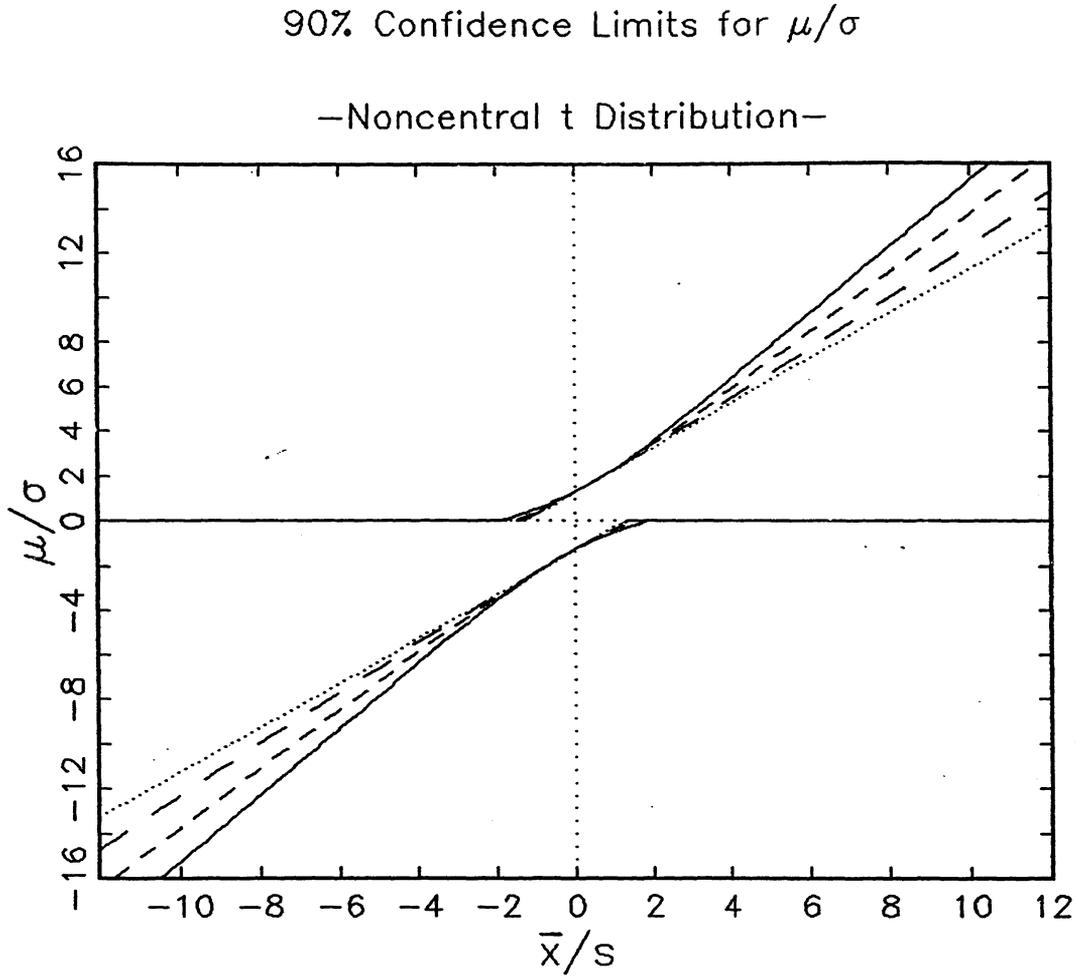


Figure 5. Two-dimensional projections of the five-dimensional “elliptical limaçon” and usual confidence ellipse for the batch degradation data. Note that both $\hat{\theta}_1$ and $\hat{\theta}_2$ were close to zero, but $\hat{\theta}_4$ was farther from zero, which is reflected in the size and placement of the respective limaçons.

