Optimal Confidence Sets, Bioequivalence, and the Lamacon of Pascal
Author(s): Lawrence D. Brown, George Casella, J. T. Gene Hwang
Source: *Journal of the American Statistical Association*, Vol. 90, No. 431 (Sep., 1995), pp. 880-889
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/2291322
Accessed: 25/03/2010 16:01

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=astata.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association.*

http://www.jstor.org

Optimal Confidence Sets, Bioequivalence, and the Limaçon of Pascal

Lawrence D. BROWN, George CASELLA, and J. T. Gene HWANG*

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, and hence they 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; in the case of the normal distribution, the optimal set is a curve known as the limaçon of Pascal. We illustrate the use of these curves with a biopharmaceutical example.

KEY WORDS: Bayes estimation; Decision theory; Frequentist estimation; Hypothesis testing; Uniformly most accurate; Uniformly most powerful.

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, then $P_\theta(\theta \in C(X))$ is the probability of true coverage, whereas $P_\theta(\theta' \in C(X))$, $\theta \neq \theta'$, 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),

$$E_\theta \text{vol}(C(X)) = \int P_\theta(\theta' \in C(X))\, d\theta', \qquad (1)$$

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

Equation (1) illustrates that possession of an optimal expected volume is a somewhat stronger property than possessing optimal false coverage probabilities, because expected volume can be regarded as a sum over all false coverages. Because 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. But 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 setup.

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. But this problem is a version of what was solved by Sterne (1954) in the binomial case (see also Crow 1956). For $X \sim$ 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 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. In Section 3 we connect these results to the Bayesian formulation and show how to construct optimal frequentist/Bayes intervals. We address the unknown variance case in Section 4, where we see that the known variance optimality results can be generalized to this case. In Section 5 we discuss the connections to the problem of bioequivalence testing and also present an example of inferences from

* Lawrence D. Brown is Meiers Busch Professor of Statistics, Department of Statistics, Wharton School of Business, University of Pennsylvania, Philadelphia, PA 19104. George Casella is Professor, Biometrics Unit, and J. T. Gene Hwang is Professor, Department of Mathematics, Cornell University, Ithaca, NY 14853. Brown and Hwang's research was supported by National Science Foundation Grant DMS9107842. Casella's research was supported by National Science Foundation Grant DMS9100839 and National Security Agency Grant 90F-073. The authors thank Jason Hsu for many interesting discussions about different aspects of this problem, and Weizhen Wang for numerical calculations leading to Figure 4, and Stephen Stigler for historical help. They also thank the editor, associate editor, and three referees for their helpful comments.

© 1995 American Statistical Association

bioequivalence confidence sets based on the limaçon. Finally, we provide some observations and conclusions in Section 6.

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

$$P(\theta \in C(x)\,|\,x) = \varphi(\theta\,|\,x); \tag{2}$$

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 included only for completeness of the theory, as they should never be recommended in practical applications.) Note also that θ need not be a scalar. Indeed, the real usefulness of the procedures developed here are for the vector parameter case, as is illustrated later.

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

$$\text{vol}(C(x)) = \int_\Theta \varphi(t\,|\,x)\,dt, \tag{3}$$

with expected volume

$$E_\theta \text{vol}(C(X)) = \int_{\mathcal{X}} \text{vol}(C(x)) f(x\,|\,\theta)\,dx. \tag{4}$$

Besides calculating (4) as a measure of size, it is usual to calculate the frequentist coverage probability of the set $C(x)$; that is,

$$P_\theta(\theta \in C(X)) = \int_{\mathcal{X}} \varphi(\theta\,|\,x) f(x\,|\,\theta)\,dx. \tag{5}$$

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

over all confidence sets $C(x)$, *minimize* $E_0 \text{vol}(C(X))$

$$\textit{subject to } P_\theta(\theta \in C(X)) \geq 1 - \alpha \quad \forall\, \theta. \tag{6}$$

Before stating and proving a formal theorem, note that specifying $\theta = 0$ in the volume requirement, which puts a particular importance on this value, makes one think of a hypothesis testing formulation. But the formulation of a hypothesis test that is equivalent to (6) is not entirely straightforward, for the specified value $\theta = 0$ will become part of the alternative hypothesis, rather than the null hypothesis. (This is because calculation of the expected volume at $\theta = 0$ is equivalent to a power calculation.)

Consider testing

$$H_0\colon \theta = \theta_0 \quad \text{versus} \quad H_1\colon \theta = 0 \tag{7}$$

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

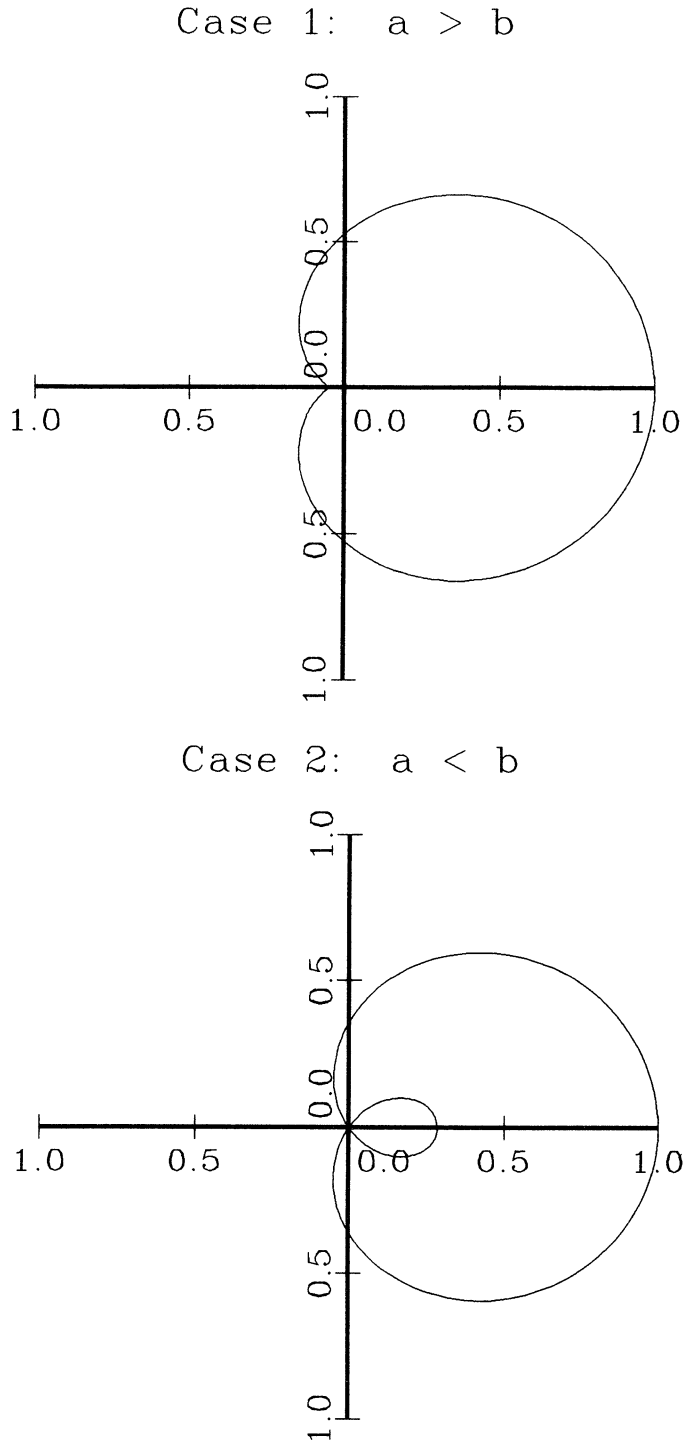

Figure 1. The Limaçon of Pascal $r = a + b \cos \beta$. If $a < b$, the limaçon has an inner loop, which does not occur in the confidence set. The confidence set $\{\theta : |\theta| \leq z_\alpha + |x| \cos \beta\}$ is in effect a "positive-part" limaçon.

$$\psi_{\theta_0} = 1 \quad \text{if } f(x\,|\,0) > k(\theta_0) f(x\,|\,\theta_0),$$

$$= 0 \quad \text{otherwise,}$$

for which $E_{\theta_0}(\psi_{\theta_0}(X)) = \alpha$. Using inclusion probabilities as specified in (2), define a confidence set C^* with

$$\phi^*(\theta\,|\,x) = 1 - \psi_\theta(x). \tag{8}$$

It is this confidence set that solves the problem in (6). Note that in the usual nonrandomized case, the confidence set is

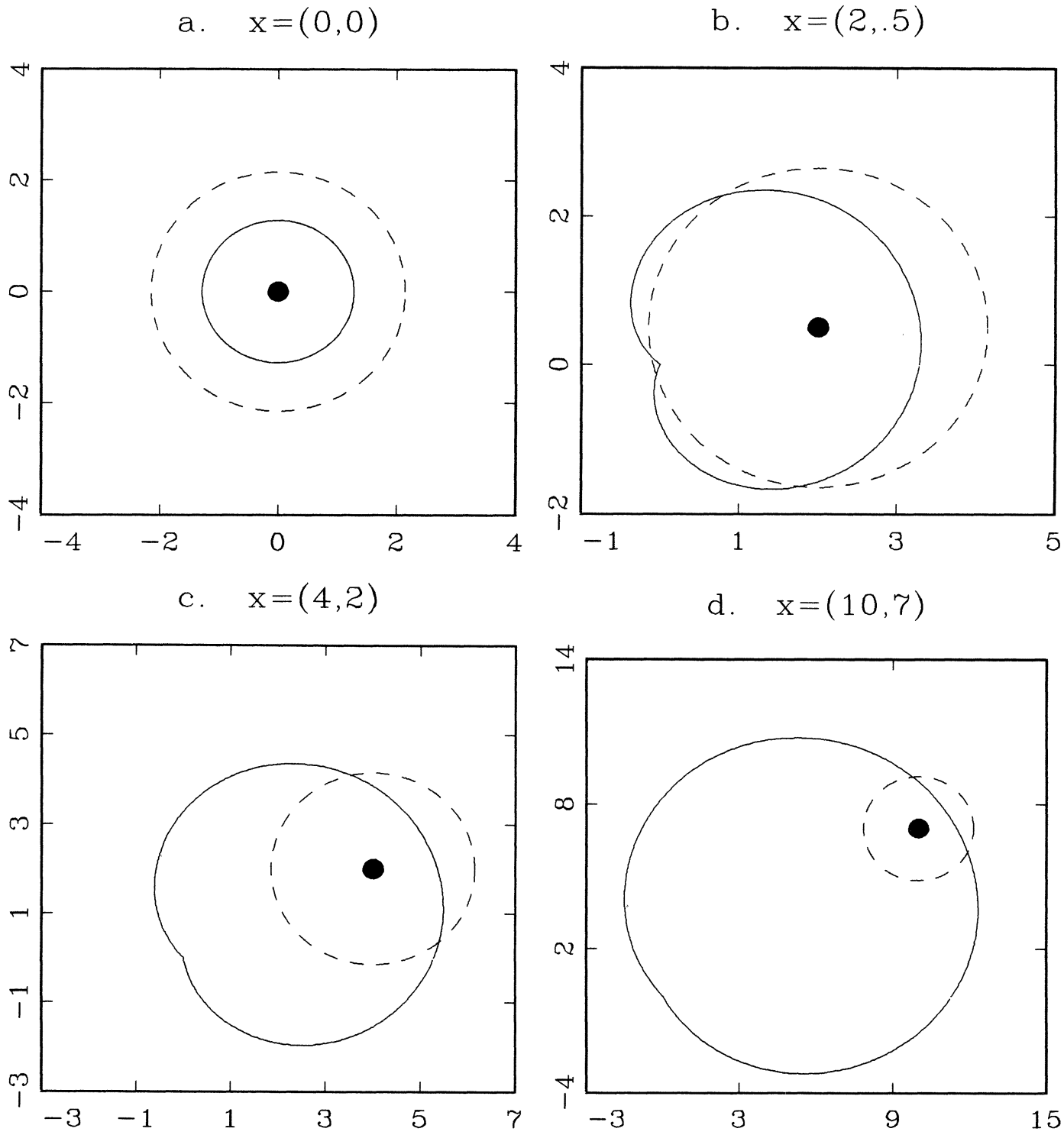

Figure 2. Two-Dimensional Limaçon (Solid Lines) and Usual Confidence Sphere (Dashed Lines) for Four Different Data Points: (a) x = (0, 0); (b) x = (2, .5); (c) x = (4, 2); (d) x = (10, 7). The confidence sphere has a constant radius, while the limaçon enlarges as the data move away from the origin. (Note that the four graphs have different scales to accommodate this.)

$$C^*(x) = \{\theta : f(x|\theta) > f(x|0)/k(\theta)\}. \quad (9)$$

Theorem 2.1. Let $X \sim f(\cdot\,|\theta)$, and let $\phi^*(\theta|x)$ be given by (8). 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 $C(x)$,

$$E_0\text{vol}(C(X)) = \int_{\mathcal{X}} \text{vol}(C(x)) f(x|0)\, dx$$

$$= \int_{\mathcal{X}} \int_{\Theta} \varphi(\theta|x)\, d\theta f(x|0)\, dx,$$

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

$$E_0\text{vol}(C(X)) = \int_\Theta \int_{\mathcal{X}} \varphi(\theta|x) f(x|0)\, dx\, d\theta$$
$$= \int_\Theta P_0(\theta \in C(X))\, d\theta. \quad (10)$$

The integrand in (10) is the probability of false coverage, which is minimized, subject to (6), by the uniformly most accurate set $C^*(x)$, which in turn produces the minimum expected volume.

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 expected size of a set $C(x)$ by

$$E_0[\text{size}(C(X))] = \int_{\mathcal{X}} \left[\int_\Theta \varphi(\theta|x)\nu(\theta)\, d\theta \right] f(x|0)\, dx, \quad (11)$$

where $\nu(\cdot) > 0$ is some weight function, then $C^*(x)$ of (8) minimizes (11) 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 (7) we would reject H_0 if

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

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

$$P_\theta(\theta \in C^*(X)) = P_\theta(X'\theta \geq k^*(\theta))$$
$$= P(|\theta| W \geq k^*(\theta)) = \Phi(a)$$

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.

We next look more closely at the shape of $C^*(x)$ and 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

$$C^*(x) = \{\theta : |\theta| \leq a + |x| \cos\beta\}. \quad (12)$$

The boundary of this set is the main lobe of a curve known as the limaçon of Pascal, a curve often used in mathematics 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 (p. 881), and the confidence set (12) is graphed in Figure 2 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 cardiod, a "heart-shaped" polar curve.) As x tends toward infinity, the limaçon again 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, then the boundary of the set (12) will resemble the limaçon in Figure 1a. 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

$$C^*(x) = \{\theta : |x|\,\text{sgn}(\theta x) \geq |\theta| - a\}$$
$$= \{\theta : \min(0, x - a) \leq \theta \leq \max(0, x + a)\}. \quad (13)$$

The 90% confidence interval has $a = 1.28$ and is equal to $x \pm 1.28$ for small $|x|$. The usual two-sided 90% confidence interval is $x \pm 1.645$. Thus $C^*(x)$ is narrower than the usual interval for small values of x, but wider for larger values. Figure 3 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). But the interval (13) has other histories. It is strongly connected with the application of bioequivalence testing, which will be discussed in Section 5. In a different context, the interval emerged in the work of Hsu (1981, 1984), who derived the interval in the context of a multiple decision problem, where one is interested in confidence in-

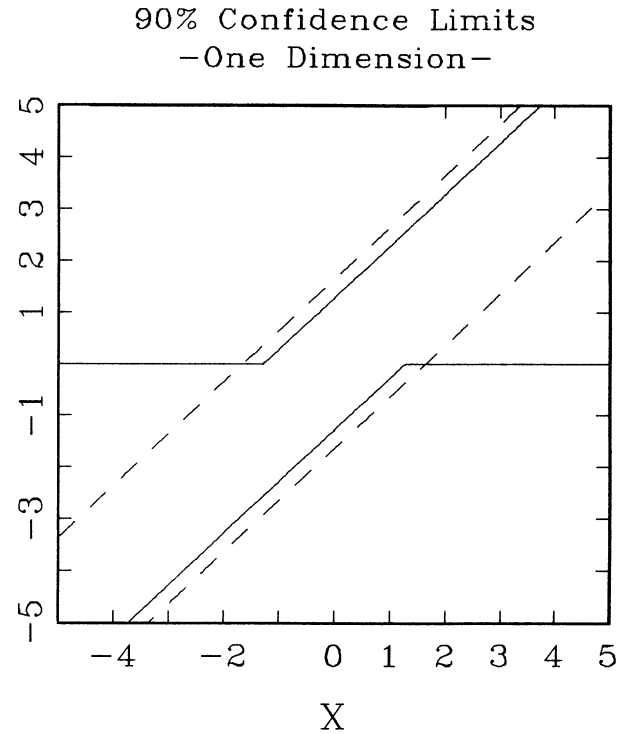

Figure 3. Comparison of C(x) (Solid lines) of (13) With the Usual 90% Confidence Interval (Dashed Lines).*

tervals for the distance from the best mean. (A short discussion of $C^*(x)$ and some of its properties can be found in exercise 9.31 of Casella and Berger [1990].)

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

$$\begin{aligned} E_0\text{vol}(C^*(X)) &= \int_\Theta \left[\int_{\mathcal{X}} I(\theta \in C^*(x)) f(x|0)\, dx \right] d\theta \\ &= \int_\Theta \left[\int_{\{x : x'\theta/|\theta| \ge |\theta| - a\}} f(x|0)\, dx \right] d\theta \\ &= \int_\Theta \Phi(a - |\theta|)\, d\theta, \end{aligned} \tag{14}$$

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 then have

$$\begin{aligned} E_0\text{vol}(C^*(X)) &= \frac{\pi^{p/2}}{\Gamma(p/2 + 1)} \int_0^\infty r^{p-1} \Phi(a - r)\, dr \\ &= \frac{\pi^{p/2}}{\Gamma(p/2 + 1)} \int_{-\infty}^a \frac{(a - t)^p}{p} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt. \end{aligned} \tag{15}$$

Note that $\pi^{p/2}/\Gamma(p/2 + 1)$ is the volume of a p-sphere of radius 1, so the pth root of the integral in (15) 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 pth root of $E_0\text{vol}(C^*(x))$ and, for comparison, the corresponding values for $C_0(x)$, the usual confidence sphere.

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. This is illustrated in Figure 2, which compares realized values of the two sets for a variety of x values. Note that the different graphs have different scales, with the value of x at 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

$$E_\pi\text{vol}(C(X)) = \int_\Theta \left[\int_{\mathcal{X}} \text{vol}(C(x)) f(x|\theta)\, dx \right] \pi(\theta)\, d\theta. \tag{16}$$

We now seek to minimize (16), a Bayesian measure, among all sets $C(x)$ that satisfy the frequentist coverage probability constraint, $P_\theta(\theta \in C(X)) \ge 1 - \alpha$.

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

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

where $m_\pi(x) = \int_\Theta f(x|\theta)\pi(\theta)\, d\theta$. Thus the confidence set is given by

$$C^*_\pi(x) = \{\theta : f(x|\theta) \ge m_\pi(x)/k(\theta)\}, \tag{17}$$

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 (17). The confidence set $C^*_\pi(x)$ minimizes the expected Bayesian volume (16) among all $1 - \alpha$ confidence sets.

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

$$\begin{aligned} E_\pi\text{vol}(C(X)) &= \int_\Theta \left[\int_{\mathcal{X}} \text{vol}(C(x)) f(x|\theta)\, dx \right] \pi(\theta)\, d\theta \\ &= \int_{\mathcal{X}} \text{vol}(C(x)) m_\pi(x)\, dx. \end{aligned}$$

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 $X \sim N(\theta, I)$ and $\theta \sim N(0, \tau^2 I)$. The marginal distribution of X is $N(0, (\tau^2 + 1)I)$, and the confidence set is

$$C^*_\pi(x) = \left\{ \theta : \left| \theta - \frac{\tau^2}{\tau^2 + 1} x \right|^2 \le k^*(\theta) \right\}, \tag{18}$$

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 $\tau^2 \to 0$, this set reduces to $C^*(x)$ of the previous section and, as $\tau^2 \to \infty$, this set approaches the usual sphere $C_0(x)$.

Table 1. Effective Volume (pth Root of Expected Volume) of $C^*(x)$ and $C_0(x)$, the Usual Confidence Sphere

p	$1 - \alpha$	$[E_0\text{vol}\, C^*(X)]^{1/p}$	$[E_0\text{vol}\,(C_0(X))]^{1/p}$	Ratio
1	.90	2.66	3.29	.809
1	.95	3.33	3.92	.849
1	.99	4.66	5.15	.905
3	.90	2.03	4.03	.504
3	.95	2.36	4.50	.524
3	.99	3.01	5.45	.552
10	.90	2.35	4.39	.535
10	.95	2.58	4.69	.550
10	.99	3.04	5.29	.575

4. GENERALIZATIONS TO THE CASE OF UNKNOWN VARIANCE

The set $C^*(x)$ of (9) 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 $X_1, \ldots, X_n$ from $n(\theta, \sigma^2)$ with both parameters unknown, there are two ways of generalizing the procedure of Section 2. The first, and perhaps the more obvious, is to test the hypotheses

$$H_0: \theta = \theta_0, \quad \sigma > 0 \quad \text{versus} \quad H_1: \theta = 0, \quad \sigma > 0. \tag{19}$$

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

$$C_t(\bar{x}, s) = \left\{\theta : \min\left(0, \bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right) \le \theta \le \max\left(0, \bar{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right)\right\}, \tag{20}$$

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 does not enjoy the same optimality properties as the interval (9). This interval was also considered by Bofinger (1992) and by Hsu, Hwang, Liu, and Ruberg (1994), although they did not investigate its optimality. We detail its exact optimality later.

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

$$\varphi(\theta \mid x_1, \ldots, x_n) = \varphi\left(\theta \left| \frac{\bar{x} - \theta}{s}\right.\right). \tag{21}$$

Note that the intervals of (20) have this form.

A second, perhaps less obvious way of generalizing Section 2.2 is to modify the hypotheses of (7) by dividing by σ to obtain

$$H_0: \frac{\theta}{\sigma} = \eta_0 \quad \text{versus} \quad H_1: \frac{\theta}{\sigma} = 0, \tag{22}$$

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 (22) are also quite interesting, because the "signal-to-noise ratio" $\eta = \theta/\sigma$ is often of interest. We will see that (22) leads to a confidence interval for η that is different from any confidence sets for θ; in particular, different from (20).

For the hypotheses (22), a reasonable invariant procedure will be of the form

$$\varphi(\eta \mid x_1, \ldots, x_n) = \varphi(\eta \mid \bar{x}/s), \tag{23}$$

with corresponding confidence intervals given by

$$C_t^*(\bar{x}/s) = \{\eta : f(\bar{x}/s \mid \eta) \ge f(\bar{x}/s \mid 0)/k(\eta)\}, \tag{24}$$

where $f(\bar{x}/s \mid \eta)$ 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 (22) 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.

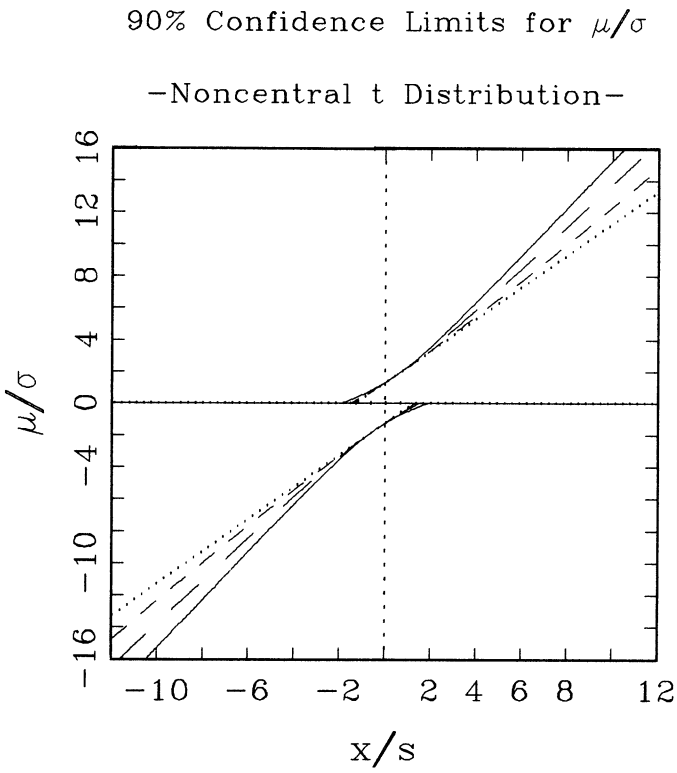

Figure 4. Comparison of C_t^* of (24) With the Normal (Known σ) Interval. The dotted lines are the normal interval, and the noncentral t interval is shown for 2 df (solid lines), 5 df (long dashed lines), and 20 df (short dashed lines).

Construction of the intervals given by (24) is actually quite straightforward, exploiting monotonicity properties of both the density and distribution function of the noncentral t. Note first that because $f(t \mid \eta)$ has monotone likelihood ratio, the acceptance region of the test (22) (i.e., H_0: η versus H_1: $\eta = 0$) is given by

$$A_\eta(t) = \{t : t \ge k_1(\eta)\} \quad \text{if } \eta > 0,$$
$$= \{t : t \le k_2(\eta)\} \quad \text{if } \eta < 0,$$

where k_1 and k_2 are increasing functions to be determined. Because the distribution function, $F(t \mid \eta)$, is decreasing in η, an α-level test is constructed by solving

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

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

$$C_t^*(\bar{x}/s) = \{\eta : \min(0, \eta_L(\bar{x}/s)) \le \eta \le \max(0, \eta_U(\bar{x}/s))\}. \tag{25}$$

Note that this construction holds in general as long as the density satisfies suitable monotonicity conditions. Of course, $C_t^*(\bar{x}/s)$ is a $1 - \alpha$ confidence interval, and it also follows that the confidence interval $(\eta_L(t), \eta_U(t))$ is a $1 - 2\alpha$ interval.

For both the setups leading to (19) and (22), we can establish optimality properties of the resulting confidence sets.

Theorem 4.1. Among all $1 - \alpha$ confidence sets for θ of the form (21), the intervals (20) 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 (21), the integrand is minimized for each θ by the C_t of (20), because these intervals correspond to most powerful level α invariant tests of the hypotheses (19).

Theorem 4.2. Among all invariant $1 - \alpha$ confidence sets for $\eta = \theta/\sigma$, the intervals (25) 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 (21), the confidence intervals (20) are not optimal. In fact, they are not even admissible, because 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) enable construction of level-α tests, $\varphi_{\text{BS}}(\theta | \bar{X}, S)$ of (19), whose power strictly dominates that of the one-sided t tests leading to (20). Consequently, this family of critical functions leads to intervals C_{BS} that dominate those of (20), 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 C_{BS} have coverage probability exceeding $1 - \alpha$, because the tests of Brown and Sackrowitz have size smaller than α. But $\inf\{P_{\theta,\sigma}(\theta \in C_{\text{BS}}(\bar{X}, S)) | \theta, \sigma\} = 1 - \alpha$. We do not know whether the intervals C_{BS} are admissible. Brown and Sackrowitz showed that their tests are admissible for testing H_0: $\theta = \theta_0$, $\sigma > 0$ versus either H_1: $\theta > 0$ if $\theta_0 < 0$ or H_1: $\theta < 0$ if $\theta_0 > 0$, but they did not prove admissibility for H_0 and H_1 of (19), as would be needed to establish admissibility of C_{BS}. But we conjecture that the intervals in Theorem 4.2 are admissible for μ/σ. This is because the tests involved are uniformly most powerful invariant and are admissible by the results of Brown and Fox (1974).

Generalizations to higher dimensions should also be of interest. Presumably, the limaçon will 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 (20), have seen practical use in bioequivalence testing. This brings us to the interesting connection between bioequivalence and the limaçon of Pascal.

5.1 Approaches to Declaring Bioequivalence

The problem of bioequivalence is typically that of deciding if the difference of two parameters, $\mu_1 - \mu_2$, is close to zero. Typically, these parameters represent different types of treatments or drugs (e.g., treatments versus control, oral versus injection, brand name versus generic). The parameters measured are often pharmacokinetic parameters, such as the area under the blood concentration–time curve (AUC) and the maximum concentration ($C_{\max}$) within a specified time period after taking a drug. The interest (usually of pharmaceutical companies) is to demonstrate that the effects are equivalent, yielding marketable bioequivalent drugs.

An initial approach taken to this problem, sometimes called the power approach, was to test H_0: $\mu_1 - \mu_2 = 0$ versus H_1: $\mu_1 - \mu_2 \neq 0$ at level $\alpha = .05$ and declare bioequivalence if the test cannot be rejected with the estimated power at least .8. This testing approach was criticized by many statisticians, including Westlake (1972) and Metzler (1974). In a number of papers, Westlake (1972, 1974, 1975, 1976, 1979) and Metzler (1974) then proposed to construct a confidence interval for $\mu_1 - \mu_2$. The confidence interval would be used to conclude bioequivalence in the following way. Given a prespecified tolerance Δ, usually set by a regulatory agency, bioequivalence would be declared if the confidence interval were completely contained in $[-\Delta, \Delta]$. To use this approach, usually called the confidence approach, it it important to decide on an appropriate confidence interval. Metzler (1974) proposed using the usual $1 - \alpha$ Student's t interval, whereas Westlake (1972, 1976) constructed a different $1 - \alpha$ interval, one that is symmetric about zero (unlike the usual t interval which is symmetric about the unbiased point estimator of $\mu_1 - \mu_2$). In justifying his interval, Westlake (1981) explained that using the "conventional $1 - \alpha$ confidence interval with $\alpha = .05$ is unduly conservative since the probability that the interval falls within the $\pm\Delta$ limits when the difference in means is Δ can be shown to be $< \alpha/2$ or .025."

In the setting of hypothesis testing, Westlake's comment translates to saying that using the usual $1 - \alpha$ t interval leads to a test of

$$H_0: |\mu_1 - \mu_2| \geq \Delta \quad \text{versus} \quad H_1: |\mu_1 - \mu_2| < \Delta, \tag{26}$$

with type I error probability at $|\mu_1 - \mu_2| = \Delta$ no greater than $\alpha/2$. (In fact, it is true that the size of the test is exactly $\alpha/2$.) But even with these possible interpretational difficulties, the hypothesis testing formulation of bioequivalence became increasingly popular. Anderson and Hauck (1983) were one of the first researchers to formulate the bioequivalence problem as a hypothesis test like (26), where rejection of H_0 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. (See also Hauck and Anderson 1984, 1992 [the latter is a review paper].)

Schuirmann (1987) proposed an alternative technique for carrying out a test of (26), called the two one-sided tests procedure. 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:

$$\text{(a)} \quad H_0: \theta \leq -\Delta \quad \text{versus} \quad H_1: \theta > -\Delta$$

and

$$\text{(b)} \quad H_0: \theta \geq \Delta \quad \text{versus} \quad H_1: \theta < \Delta. \tag{27}$$

It is interesting to note that (27) 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 alternative hypotheses, individual α-level tests lead to an overall α-level test for (26). This approach is currently

recommended by the U.S. Food and Drug Administration (1992).

Relating to the mismatch of error probabilities mentioned before (26), it has been observed that there is a similarity between the α level two one-sided tests procedure and a $1 - 2\alpha$ Student's t confidence interval. This similarity is somewhat of a fiction, based more on an algebraic coincidence rather than a statistical equivalence. Both tests in (27) will reject H_0, and bioequivalence will be concluded, if and only if the $1 - 2\alpha$ Student's t confidence interval is contained in $[-\Delta, \Delta]$. This is because the middle of the limaçon interval is a $1 - 2\alpha$ interval (see Fig. 3), but the limaçon interval widens away from this middle region. In fact, formally inverting the family of tests (indexed by Δ) in (27) will yield confidence intervals like (13) or (20), the one-dimensional limaçon.

More recent research on this topic has centered on constructing confidence intervals for $\mu_2 - \mu_1$, and has paid less attention to the testing problem in (26). One possible reason for this is that the confidence set approach has the added benefit that the constant Δ does not have to be prespecified. Indeed, for a given $1 - \alpha$ confidence set $C(x)$, and any set Γ, we could conclude that H_1: $\mu_1 - \mu_2 \in \Gamma$ if $C(x) \subset \Gamma$. 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).

Liu (1990) has produced a $1 - \alpha$ symmetric confidence interval that is shorter than that of Westlake (1976) and is the smallest interval symmetric about zero that contains (20). Other improved intervals have been constructed by Bofinger (1992) and by Hsu et al. (1994). Interestingly, with the exception of Westlake's interval, all of these other intervals, when used as hypothesis tests, lead to Schuirmann's α level test, further showing that there is no mismatch of error probabilities. Further generalizations of this problem, including a nonparametric approach, were discussed by Hsu et al. (1994).

There are other formulations of the bioequivalence hypothesis that lead to alternate tests and confidence intervals. For example, one could specify Δ of (26) 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 standard deviation. Of course, this formulation leads directly to the hypotheses specified in (22), and to the noncentral t-based intervals of (24). Although this formulation of the bioequivalence problem has been used, the optimal procedure of (24) has not been used.

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

5.2 An Example

We now illustrate the use of the limaçon with data from Sheen, Kim, Petillo, and Serajuddin (1991) on the bioavailability of drug as a function of different delivery systems; that is, the question of interest is whether different delivery systems are bioequivalent. The particular drug investigated, a 5-lipoxygenase inhibitor that is orally active against hypersensitivity diseases, is a "sparingly water-soluble" drug. Although Sheen et al. (1991) looked at a number of parameters, for illustration here we will concentrate on only two questions. The drug is available in both tablet and capsule form, and for each formulation we will investigate the bioequivalence of dosages received with or without food. Eight healthy males were used in the study, with appropriate washout periods between treatments. For each formulation of the drug (tablet or capsule) the response differences (with food–without food) were measured for the responses AUC and $C_{\max}$. Thus for each formulation, we construct a two-dimensional limaçon with parameters

$$\theta_1 = \mu_1 - \mu_2 = \text{AUC difference}$$

and

$$\theta_2 = \tau_1 - \tau_2 = C_{\max} \text{ difference.}$$

The data are assumed to be bivariate normal with known variance; that is, we observe $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, where $\hat{\theta} \sim N(\theta, \Sigma)$. (We could drop the assumption of known variance by replacing the normal and chi-squared cutoff points with t and F cutoff points.). In general, for $\hat{\theta} \sim N(\theta, \Sigma)$, the $1 - \alpha$ limaçon confidence set is

$$\left\{\theta : (\theta'\Sigma^{-1}\theta)^{1/2} \le z_\alpha + \frac{\hat{\theta}'\Sigma^{-1}\theta}{(\theta'\Sigma^{-1}\theta)^{1/2}}\right\}, \tag{28}$$

where z_α is the upper α cutoff point from a univariate standard normal distribution. This set reduces to the form of (12) with the transformation $\eta = \Sigma^{-1/2}\theta$, $\hat{\eta} = \Sigma^{-1/2}\hat{\theta}$. The set (28) is somewhat of an elliptical limaçon and is shown in Figure 5 for both the tablet and capsule formulations. In both cases the limaçon confidence set is completely contained within the usual one. This would allow the experimenter to conclude bioequivalence against smaller values Δ, and in this sense the limaçon provides a sharper inference for bioequivalence. Of course, we again note that this reduction in volume is obtained only when the "exceptional point" (here (0, 0)) is well supported by the data, and the reduction is obtained at the expense of an increase in volume in other portions of the parameter space. In the study by Sheen et al. (1991), however, the data are quite supportive of bioequivalence. In fact, the capsule data are so supportive that the limaçon in Figure 5a is nearly an ellipse, just as the limaçon in Figure 2a is a circle.

6. CONCLUSIONS

We hope that this article reflects some of the surprise and delight that we experienced when discovering these many connections between decision-theoretic mathematical statistics, Bayes and frequency inference, limaçons, and bioequivalence. None of these connections were conjectured at the beginning of this research, but as we followed the statistical and mathematical path, the connections revealed themselves. It is perhaps most gratifying that what started

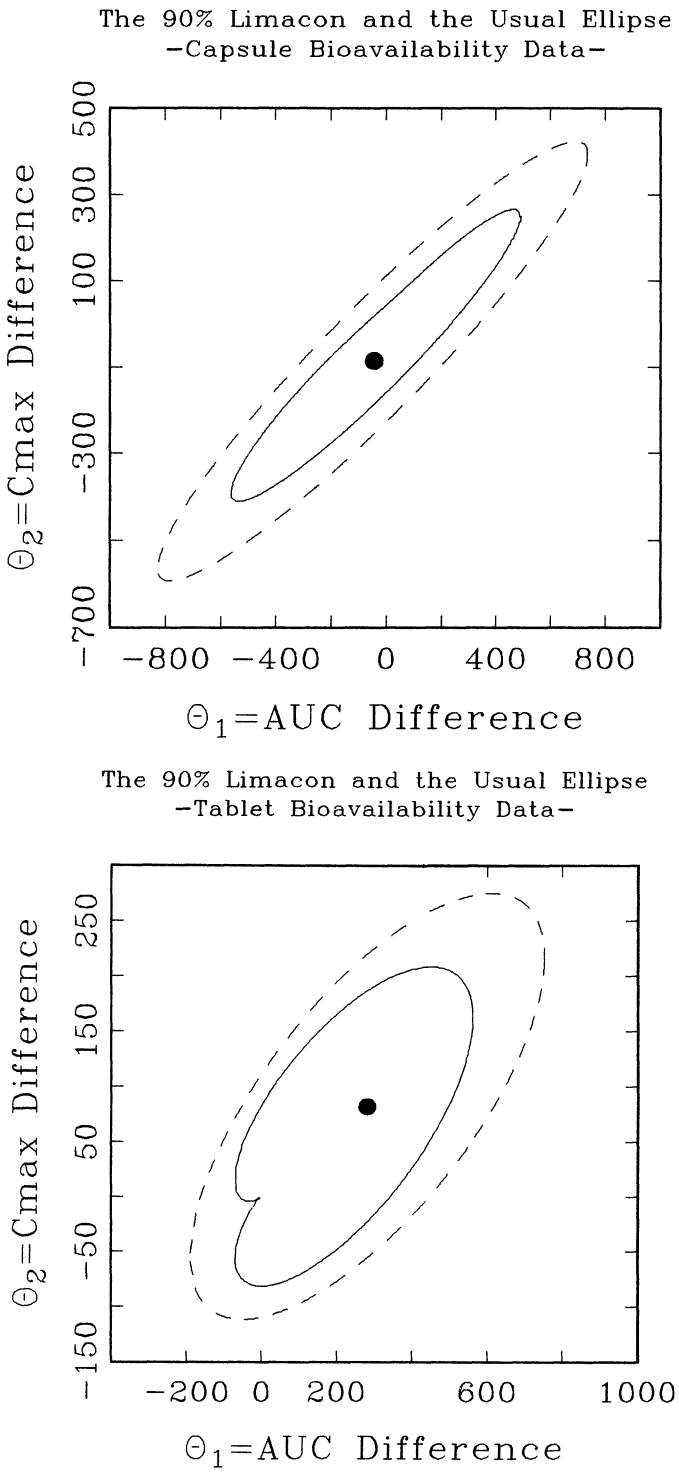

Figure 5. Two-Dimensional "Elliptical" Limaçons (Dashed Line, Usual Ellipse; Solid Line, Limaçon) for the Capsule (a) and Tablet (b) Formulations in the Sheen et al. (1991) Bioavailability Study.

as purely an exercise in decision-theoretic optimality has led to some interesting and useful rethinking of an important problem in applied statistics.

Part of what this article illustrates is that the decision-theoretic formulation of the confidence set problem can lead to optimal solutions that perhaps would not have been seen otherwise. The fact that a curve like the limaçon of Pascal appears as the optimal shape of a confidence set shows how a somewhat different formulation of the confidence set problem can result in alternate shapes of confidence sets. In particular, the case of the normal distribution, where ellipses have ruled, has succumbed to an alternate form. It would be interesting to see what shapes are optimal for other distributions.

We have also seen the the decision theory can span both Bayes and frequency theory, and may provide a solution that is acceptable under both types of inference. This dual optimality also reflects the power of the Ghosh–Pratt identity, which establishes connections between classical properties of confidence sets, such as uniformly most accurate (UMA), with admissibility under a certain loss structure.

Of course, there is a limitation to the formulation presented here, in that a specified "exceptional" point must be identified. In the bioequivalence problem the value zero is a clear point of interest, but in other problems such a point may not be obvious. Of course, any particular value may be substituted for zero, and in other cases such values may be obvious, but a value must be chosen. (A situation that is reminiscent of Stein-type point estimation). The resulting confidence sets are, by construction, exact $1 - \alpha$ confidence sets, but they can be uselessly large when the data do not support the specified point. For example, in the bioequivalence problem the limaçon is not appropriate for detecting when two treatments are different. This limitation can be anticipated from the decision-theoretic formulation, in that the limaçon delivers optimality at zero by sacrificing performance in other regions of the parameter space.

Interestingly, by using a hierarchical model and a "pseudo-empirical Bayes" construction, the limitations of the previous paragraph can be avoided. Tseng (1994) has applied Theorem 2.1 in this way, and has constructed exact $1 - \alpha$ confidence sets for a multivariate normal mean that are uniformly smaller than the usual set. These new sets have a complicated analytical form but can be shown to be convex in many cases and can be displayed graphically. Further work in this direction may lead to confidence sets that are optimal in a subspace rather than merely near a point, reminiscent of Stein-type estimators that shrink toward subsets of the parameter space.

[Received August 1993. Revised August 1994.]

REFERENCES

Anderson, S., and Hauck, W. W. (1983), "A New Procedure for Testing Equivalence in Comparative Bioavailability and Other Clinical Trials," *Communications in Statistics, Part A—Theory and Methods,* 12, 2663–2692.

Archibald (1900), "Le Limaçon de Pascal," *L'Intermediaire des Mathématiciens,* 106–107.

Berger, R. L. (1982), "Multiparameter Hypothesis Testing and Acceptance Sampling," *Technometrics,* 24, 295–300.

Bofinger, E. (1992), "Expanded Confidence Intervals, One-Sided Tests, and Equivalence Testing," *Journal of Biopharmacological Statistics,* 2(2), 181–188.

Brown, L. D., and Fox, M. (1974), "Admissibility in Statistical Problems Involving a Location or Scale Parameter," *The Annals of Statistics,* 2, 807–814.

Brown, L. D., and Sackrowitz, H. (1984), "An Alternative to Student's *t* Test for Problems With Indifference Zones," *The Annals of Statistics,* 12, 451–469.

Casella, G., and Berger, R. L. (1990), *Statistical Inference,* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Casella, G., and Hwang, J. T. (1983), "Empirical Bays Confidence Sets for the Mean of a Multivariate Normal Distribution," *Journal of the American Statistical Association,* 78, 688–698.

Cohen, A., and Strawderman, W. E. (1973), "Admissibility Implications for Different Criteria in Confidence Estimation," *The Annals of Statistics,* 1, 363–366.

Crow, E. L. (1956), "Confidence Intervals for a Proportion," *Biometrika,* 43, 423–425.

U.S. Food and Drug Administration (1992), *Draft Recommendation on Statistical Procedures for Bioequivalence Studies Using the Standard Two-Treatment Crossover Design,* Rockville, MD: Center for Drug Evaluation and Research, Food and Drug Administration.

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," *Journal of Pharmacokinetic Biopharmacology,* 12, 83–91.

——— (1992), "Types of Bioequivalence and Related Statistical Considerations," *International Journal of Clinical Pharmacology and Therapeutic Toxicology,* 30, 181–187.

Hsu, J. C. (1981), "Simultaneous Confidence Intervals for All Distances From the Best," *The Annals of Statistics,* 9, 1026–1034.

——— (1984), "Constrained Two-Sided Simultaneous Confidence Intervals for Multiple Comparisons With the Best," *The Annals of Statistics,* 12, 1136–1144.

Hsu, J. C., Hwang, J. T. G., Liu, H. K., and Ruberg, S. J. (1994), "Equivalence Confidence Intervals in Bioequivalence Trials," *Biometrika,* 81, 103–114.

Liu, H. K. (1990), "Confidence Intervals in Bioequivalence," in *Proceedings of the Biopharmacology Section,* American Statistical Association, pp. 51–54.

Metzler, C. M. (1974), "Bioavailability—A Problem in Equivalence," *Biometrics,* 30, 309–317.

Pratt, J. W. (1961), "Length of Confidence Intervals," *Journal of the American Statistical Association,* 56, 541–567.

Schuirmann, D. J. (1987), "A Comparison of the Two One-Sided Test Procedures and the Power Approach for Assessing the Equivalence of Average Bioavailability," *Journal of Pharmacology and Biopharmacology,* 15, 657–680.

Sheen, P-C., Kim, S-I., Petillo, J. J., and Serajuddin, A. T. M. (1991), "Bioavailability of a Poorly Water-Soluable Drug From Tablet and Solid Dispersion in Humans," *Journal of Pharmacological Science,* 80, 712–714.

Shinozaki, N. (1989), "Improved Confidence Sets for the Mean of a Multivariate Normal Distribution," *Annals of the Institute of Statistical Mathematics,* 41, 331–346.

Sterne, T. E. (1954), "Some Remarks on Confidence or Fiducial Limits," *Biometrika,* 41, 275–278.

Tseng, Y. L. (1994), "Good Exact Confidence Sets for a Multivariate Normal Mean," Ph.D. thesis, Cornell University, Statistics Center.

Westlake, W. J. (1972), "Use of Confidence Intervals in Analysis of Comparative Bioavailability Trials," *Journal of Pharmacological Science,* 61, 1340–1341.

——— (1974), "The Use of Balanced Incomplete Block Designs in Comparative Bioavailability Studies," *Biometrics,* 32, 741–744.

——— (1975), "Statistical Problems in Interpreting Comparative Bioavailability Trials," *International Journal of Chemical Pharmacology,* 11, 342–345.

——— (1976), "Symmetrical Confidence Intervals for Bioequivalence Trials," *Biometrics,* 32, 741–744.

——— (1979), "Statistical Aspects of Comparative Bioavailbaility Trials," *Biometrics,* 35, 273–280.

——— (1981), "Bioequivalence Testing—A Need to Rethink (Reader Reaction Response)," *Biometrics,* 37, 580–594.