The Annals of Statistics
1998, Vol. 26, No. 5, 1687–1718

THE PROBLEM OF REGIONS

BY BRADLEY EFRON[1] AND ROBERT TIBSHIRANI

Stanford University and University of Toronto

In the problem of regions, we wish to know which one of a discrete set of possibilities applies to a continuous parameter vector. This problem arises in the following way: we compute a descriptive statistic from a set of data, notice an interesting feature and wish to assign a confidence level to that feature. For example, we compute a density estimate and notice that the estimate is bimodal. What confidence can we assign to bimodality? A natural way to measure confidence is via the bootstrap: we compute our descriptive statistic on a large number of bootstrap data sets and record the proportion of times that the feature appears. This seems like a plausible measure of confidence for the feature. The paper studies the construction of such confidence values and examines to what extent they approximate frequentist p-values and Bayesian a posteriori probabilities. We derive more accurate confidence levels using both frequentist and objective Bayesian approaches. The methods are illustrated with a number of examples, including polynomial model selection and estimating the number of modes of a density.

1. Introduction. The title of this paper refers to a class of problems that combine elements of both hypothesis testing and estimation. Figure 1 illustrates an example we are going to discuss concerning the choice of a polynomial regression model. We have observed a K-dimensional multivariate normal vector y having unknown expectation vector μ and covariance matrix the identity,

$$y \sim N_K(\mu, I). \tag{1.1}$$

$K = 2$ in the illustration. The space of possible μ vectors is partitioned into regions called $\mathscr{R}_{\text{con}}, \mathscr{R}_{\text{lin}}, \mathscr{R}_{\text{quad}}$, and the maximum likelihood estimate (MLE) $\widehat{\mu} = y$ is observed to lie in $\mathscr{R}_{\text{quad}}$. How confident should we be that μ itself lies in $\mathscr{R}_{\text{quad}}$?

The meaning of "confidence" here will be approached in two complementary ways: in the classical frequentist manner of p-values and confidence levels, and also from a Bayesian viewpoint involving objective (uninformative) priors. We will develop a bootstrap methodology that gives reasonably good answers from both points of view.

A first, possibly naive, answer to the question posed in Figure 1 is obtained by resampling from model (1.1) with $\widehat{\mu}$ replacing the unknown vector μ, say

$$y^* \sim N_K(\widehat{\mu}, I). \tag{1.2}$$

Received February 1997; revised June 1998.

[1] Supported by NSF Grant DMS-95-04379 and NIH Grant CA59039.

AMS 1991 *subject classifications*. Primary 62G10; secondary 62G09.

Key words and phrases. Discrete estimation problems, objective Bayes methods, bootstrap reweighting, metric-free methods.

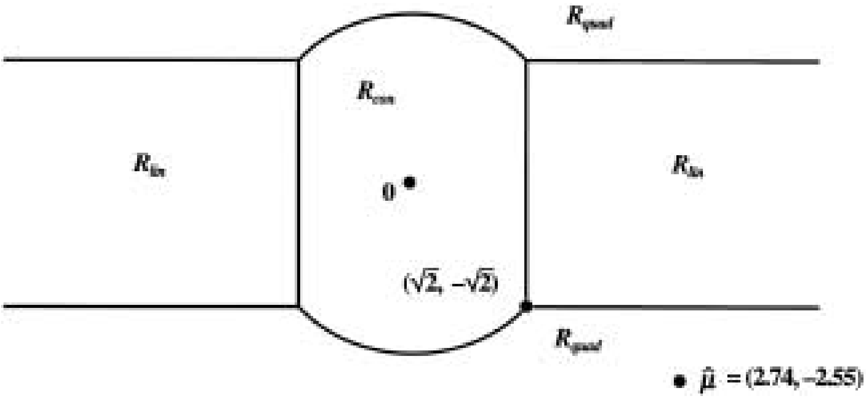

FIG. 1. *An example of the problem of regions: a normally distributed vector* $y = \widehat{u}$, *with covariance I, is observed to lie in the region* $\mathscr{R}_{\text{quad}}$. *With what confidence can we say that the true expectation vector* μ *lies in* $\mathscr{R}_{\text{quad}}$? *This example, which concerns the choice of a polynomial regression model using the* C_p *criterion, is discussed in Section* 5.

One thousand "parametric bootstrap" vectors y^* obtained independently according to (1.2) included 18 in $\mathscr{R}_{\text{con}}$ and 117 in $\mathscr{R}_{\text{lin}}$, with the remaining 865 falling into $\mathscr{R}_{\text{quad}}$. It seems plausible to say that there is a 1.8% chance that $\mu \in \mathscr{R}_{\text{con}}$ and an 11.7% chance that $\mu \in \mathscr{R}_{\text{lin}}$, leaving 86.5% confidence that $\mu \in \mathscr{R}_{\text{quad}}$. We will call this a first-order bootstrap analysis.

The first-order answer turns out not to be a bad first guess. However, more elaborate resampling procedures are necessary if we want "confidence" to better agree with either its frequentist or Bayesian meanings. The shapes of the regional boundaries play an important role, leading to second-order corrections of the first-order analysis. Sections 2 and 3 show that the fact that the boundary between $\mathscr{R}_{\text{con}}$ and $\mathscr{R}_{\text{quad}}$ curves *away* from $\widehat{\mu}$ should increase our belief that $\mu \in \mathscr{R}_{\text{con}}$ and decrease our confidence that $\mu \in \mathscr{R}_{\text{quad}}$. The bootstrap methods presented in Sections 2 through 6 make these corrections automatically, without requiring detailed geometrical knowledge of the shapes or distances of the boundaries. This is crucial for practical applications, where the situation can easily be much more complicated than Figure 1.

The problem of regions arises when we wish to know which one of a discrete set of possibilities applies to a continuous parameter vector. Figure 1 actually concerns a regression situation where we are trying to choose the degree of a polynomial model using the C_p criterion. $\mathscr{R}_{\text{con}}$, $\mathscr{R}_{\text{lin}}$ or $\mathscr{R}_{\text{quad}}$ are the regions where a constant, linear or quadratic model would be preferred (see Section 5). In other words, we are trying to assign a measure of statistical uncertainty to the data-based choice of "quadratic" as the best-fitting polynomial regression model. Similar questions include:

1. How many modes does a density function have? (See Section 7.)
2. How many terms should be included in a principal components or factor analysis?

3. The bioequivalence problem: does a newly produced drug have efficacy between 80% and 125% of its predecessor?
4. Ranking and selection: how confident should we be that an apparently best treatment is actually best?
5. Simultaneous significance testing: given K observed treatment effects, with what confidence can we say that some treatments are better or worse than others?

The normal model (1.1) is convenient for exposition, and we will follow it in the first several sections of this paper, but it does not play a critical role in our theory. Section 6 generalizes (1.1) to all multiparameter exponential families. A particularly important case is the multinomial family, which has a central role in nonparametric applications of our theory.

The first-order bootstrap analysis above was introduced by Felsenstein (1985) as a method of assigning confidence values to phylogenetic trees constructed on the basis of genetic sequence data. Felsenstein's method, which is an application of nonparametric bootstrapping, is discussed from the point of view of this paper in Efron, Halloran and Holmes (1996), where the regions problem was introduced. The regions problem becomes much more intricate in the trees context, having enormous numbers of regions separated by complicated and hard-to-locate high-dimensional boundaries. Our bootstrap-based methods are designed to function in these difficult situations. A key feature is that they do not depend on metric properties such as "distances between trees," and so can be applied automatically without requiring a detailed analysis of specific problems.

There are three main themes in this paper: frequentist confidence levels, objective Bayesian posterior probabilities and bootstrap methods. Sections 2, 3 and 4 carry these themes through for the normal model (1.1). The frequentist methods work well for two regions but can give unreasonable answers for more than two. The Bayesian approach is simple and works for any number of regions, but is strongly dependent on the choice of prior. The objective Bayes approach, implemented through bootstrap sampling, combines the objectivity of the frequentist method with the conceptual simplicity of the Bayesian framework. Section 5 applies this theory to the C_p problem of Figure 1. The theory is extended to general exponential family probability models in Section 6, and applied in Section 7 to estimating the number of modes of a density function. We conclude with a few remarks in Section 8. A longer version of this paper, Efron and Tibshirani (1996), is available as a hardcopy technical report or on the web site http://stat.stanford.edu/tibs/research.html.

2. Frequentist confidence levels.

Familiar versions of what we have called the problem of regions show up in standard hypothesis testing situations. These situations have well-accepted frequentist solutions that we will want our methods to agree with when they apply. This section considers the regions problem for the normal model $y \sim N_K(\mu, I)$, when there are only two regions and they are separated by a smooth boundary. Standard frequentist confidence levels are available here and we will use them to suggest improvements to the first-order bootstrap method of the Introduction.

Figure 2 shows a schematic diagram of the situation, as well as summarizing some of this section's notation. The two regions, $\mathscr{R}_0$ and $\mathscr{R}_1$, are separated by a smooth boundary $\mathscr{B}$. We observe the data vector $y = \widehat{\mu}$ to lie in $\mathscr{R}_0$, at distance x_0 from the nearest point on $\mathscr{B}$, and wonder how confident we should be that μ itself lies in $\mathscr{R}_0$.

The first-order bootstrap answer, which we will call the *confidence value* and denote by $\tilde{\alpha}$, is

$$\tilde{\alpha} = \text{prob}\{y^* \in \mathscr{R}_0\} \quad \text{where } y^* \sim N_K(\widehat{\mu}, I). \tag{2.1}$$

However, $\tilde{\alpha}$ will not match the usual frequentist *confidence level*, denoted $\widehat{\alpha}$, unless the boundary $\mathscr{B}$ is flat. We will show that $\tilde{\alpha}$ exceeds $\widehat{\alpha}$ if $\mathscr{B}$ curves away from $\widehat{\mu}$ as in Figure 2, and is less than $\widehat{\alpha}$ if $\mathscr{B}$ curves toward $\widehat{\mu}$.

What is the confidence level $\widehat{\alpha}$? In the usual frequentist formulation, it is the probability of being closer than y to the boundary, minimized over the choice of μ in $\mathscr{R}_1$. In other words, $1 - \widehat{\alpha}$ is the p-value for testing the null hypothesis that $\mu \in \mathscr{R}_1$. "Attained confidence level" would be a more accurate but cumbersome terminology for $\widehat{\alpha}$. If, in fact, $\mu \in \mathscr{R}_1$, then $\widehat{\alpha}$ will exceed .90 no more than 10% of time, .95 no more than 5% of time, etc., which is its usual frequentist interpretation.

EXAMPLE 1. Suppose that

$$\mathscr{R}_1 = \{\mu \colon \|\mu\| \le \theta_1\}, \qquad \mathscr{R}_0 = \{\mu \colon \|\mu\| > \theta_1\}, \tag{2.2}$$

so that the two regions are separated by a spherical boundary in K dimensions, and that the data vector $y \sim N_K(\mu, I)$ falls into $\mathscr{R}_0$, say

$$\|y\| = r > \theta_1. \tag{2.3}$$

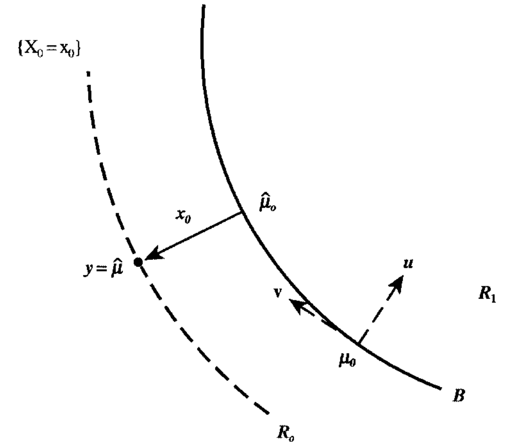

FIG. 2. *Two regions, $\mathscr{R}_0$ and $\mathscr{R}_1$, separated by a smooth boundary $\mathscr{B}$; data vector $y \sim N_K(\mu, I)$ is in $\mathscr{R}_0$; nearest point to $y = \widehat{\mu}$ not in $\mathscr{R}_0$ is $\widehat{\mu}_0$; signed distance from $\widehat{\mu}_0$ to y is x_0. How confident should we be that $\mu \in \mathscr{R}_0$?*

Then according to (2.1),

$$\tilde{\alpha} = \text{prob}\{\chi_K^2(r^2) \geq \theta_1^2\}, \tag{2.4}$$

the notation indicating a noncentral chi-squared random variable with K degrees of freedom and noncentrality parameter r^2. The confidence level for $\mu \in \mathscr{R}_0$, or more precisely one minus the p-value against the hypothesis that $\mu \in \mathscr{R}_1$, is

$$\widehat{\alpha} = \text{prob}\{\chi_K^2(\theta_1^2) < r^2\}. \tag{2.5}$$

NOTE. Here we are using a one-sided p-value and behaving as if $\mathscr{R}_1$ were preselected to be the null hypothesis region. It is easy to calculate two-sided p-values [from (2.12), e.g.], but these become problematical in multiregion problems, nor do they agree with the Bayesian considerations of Sections 3 and 4.

For the case where

$$r = 7, \quad \theta_1 = 5, \quad K = 4, \tag{2.6}$$

we compute $\tilde{\alpha} = .9880$ and $\widehat{\alpha} = .9596$, so in this situation the confidence value is much bigger than the confidence level. This is shown more dramatically in terms of the "non-confidences"

$$\tilde{\beta} \equiv 1 - \tilde{\alpha} = .0120, \qquad \widehat{\beta} \equiv 1 - \widehat{\alpha} = .0404. \tag{2.7}$$

The non-confidence level $\widehat{\beta}$, which is the usual p-value, is more than three times bigger than the non-confidence value $\tilde{\beta}$.

Example 1 is misleadingly simple because the minimum distance x_0 in Figure 2 is a pivotal quantity, having the same distribution for all μ on the boundary $\mathscr{B}$. This allows us to consider only one distribution in the computation (2.5) of $\widehat{\alpha}$. The discussion below shows that, in general, x_0 acts as a very good *approximate* pivotal quantity, for any smooth boundary $\mathscr{B}$, and that we can use x_0 to obtain a high-order approximation to $\widehat{\alpha}$. It will also lead to a bootstrap method for converting an $\tilde{\alpha}$ value into a good approximation for $\widehat{\alpha}$, based on the following result.

THEOREM. *To second order of accuracy*,

$$\widehat{\alpha} \doteq \Phi\{\Phi^{-1}(\tilde{\alpha}) - 2\widehat{z}_0\}, \tag{2.8}$$

where Φ *is the standard normal cumulative distribution function* (cdf) *and*

$$\widehat{z}_0 = \Phi^{-1}\{\text{prob}(y_0^* \in \mathscr{R}_0)\}, \qquad y_0^* \sim N_K(\widehat{\mu}_0, I). \tag{2.9}$$

"Second-order accuracy" means that (2.8) errs by order $O_p(n^{-1})$ in repeated sampling situations, in the sense defined below. The theorem allows us to compute a second-order accurate frequentist confidence level $\widehat{\alpha}$ solely by bootstrap calculations: first, we calculate $\tilde{\alpha}$ using the first-order bootstrap method (2.1); then, $\widehat{z}_0$ is obtained from the second kind of bootstrapping in (2.9); finally, we get $\widehat{\alpha}$ from (2.8). All of these calculations require only the knowledge of whether or not the resampled point falls into $\mathscr{R}_0$, which makes the algorithm applicable even to very complicated problems.

We will now verify the theorem, and in fact give better results that are accurate to a third order of approximation. (Remark A of Section 8 describes a full third-order bootstrap algorithm for $\widehat{\alpha}$.) This material is more technical than the other sections and may be omitted at first reading.

Looking at Figure 2 again, let μ_0 be a point on $\mathscr{B}$, and let $\mathscr{T}_{\mu_0}$ be the $(K-1)$-dimensional tangent plane to $\mathscr{B}$ through μ_0. The figure shows a vector v in $\mathscr{T}_{\mu_0}$ and also a distance u measured orthogonally to $\mathscr{T}_{\mu_0}$, taken positive in the direction pointing away from $\mathscr{R}_0$ and into $\mathscr{R}_1$. We assume that there is a $(K-1)$ by $(K-1)$ symmetric matrix $\mathbf{d}(\mu_0)$, not necessarily positive definite, such that the Taylor series describing the boundary $\mathscr{B}$ for points near μ_0 begins

$$u = v'\mathbf{d}(\mu_0)v. \tag{2.10}$$

The matrix $\mathbf{d}(\mu_0)$ measures the curvature of $\mathscr{B}$ at μ_0 with $\mathbf{d}(\mu_0) = \mathbf{0}$, corresponding to a flat spot. By $\mathscr{B}$ being smoothly defined we mean that $\mathbf{d}(\mu_0)$ exists and is continuously differentiable. These definitions are spelled out more carefully in Efron (1985).

The *signed distance* X_0 is defined to be the distance from point y to the nearest point $\widehat{\mu}_0$ on $\mathscr{B}$, taken positive if $y \in \mathscr{R}_0$ and negative if $y \in \mathscr{R}_1$. Let

$$d_1(\mu_0) = \operatorname{trace}(\mathbf{d}(\mu_0)), \qquad d_2(\mu_0) = \operatorname{trace}(\mathbf{d}(\mu_0)^2). \tag{2.11}$$

Theorem 1 of Efron (1985) states that if $y \sim N_K(\mu_0, I)$ for $\mu_0 \in \mathscr{B}$, then X_0 is approximately normal,

$$X_0 \sim N\big(d_1(\mu_0), \{1 - d_2(\mu_0)\}^2\big). \tag{2.12}$$

Approximation (2.12) is highly accurate in the following asymptotic sense: suppose that Figure 2 applies, but that instead of $y \sim N_K(\mu, I)$, we observe a sequence of situations indexed by n, with

$$y \sim N_K(\mu, I/n) \tag{2.13}$$

at the nth stage. This would be the case if $y = \sum_{i=1}^{n} y_i/n$ with the y_i independently distributed as $N_K(\mu, I)$. Multiplying y by $\sqrt{n}$ restores the covariance matrix in (2.13) to I, but the magnified version of the boundary $\mathscr{B}$ has curvature matrix *divided* by $\sqrt{n}$. In what follows, we assume that this rescaling has been done and that $\mathbf{d}(\mu_0)$ is of order $O(n^{-1/2})$. This implies orders

$$d_1(\mu_0) = O(n^{-1/2}), \qquad d_2(\mu_0) = O(n^{-1}), \tag{2.14}$$

for the terms in (2.12). Theorem 1 of Efron (1985) shows that (2.12) is *third-order accurate*. That is, (2.12) is correct to order $O(n^{-1})$, only erring by $O(n^{-3/2})$. In particular, the skewness and kurtosis of X_0 are both $O(n^{-3/2})$.

Now let $\widehat{\mathbf{d}} = \mathbf{d}(\widehat{\mu}_0)$ be the curvature matrix of $\mathscr{B}$ at $\mu = \widehat{\mu}_0$, the closest point to $y = \widehat{\mu}$, and let

$$\widehat{d}_1 = \operatorname{trace}(\widehat{\mathbf{d}}), \qquad \widehat{d}_2 = \operatorname{trace}(\widehat{\mathbf{d}}^2). \tag{2.15}$$

A third-order normal pivotal quantity can be derived from (2.12),

$$Z \equiv \frac{X_0 - \widehat{d}_1}{1 - \widehat{d}_2} \sim N(0,1), \tag{2.16}$$

by which we mean that if $y \sim N_K(\mu_0, I)$ for $\mu_0 \in \mathscr{B}$, then $\text{prob}\{Z < z\} = \Phi(z) + O(n^{-3/2})$ for any fixed z. This is demonstrated in the Appendix of Efron (1985).

Because Z is a third-order pivotal, it can be used to construct a highly accurate frequentist confidence level for $\{\mu \in \mathscr{R}_0\}$, namely,

$$\widehat{\alpha} = \Phi\left(\frac{x_0 - \widehat{d}_1}{1 - \widehat{d}_2}\right); \tag{2.17}$$

$\widehat{\alpha}$ is the approximate probability that Z is less than its observed value $z = (x_0 - \widehat{d}_1)/(1 - \widehat{d}_2)$, if the true expectation vector lies on the boundary $\mathscr{B}$. It is shown in the Appendix of Efron and Tibshirani (1996) that $\widehat{\alpha}$ also equals to third order the probability that X_0 is less than its observed value x_0.

We can also calculate a third-order approximation for the bootstrap confidence value

$$\tilde{\alpha} = \text{prob}_{\widehat{\mu}}\{y^* \in \mathscr{R}_0\}, \tag{2.18}$$

where $y^* \sim N_K(\widehat{\mu}, I)$ as in (1.2). The Appendix shows that

$$\tilde{\alpha} = \Phi\left(\frac{x_0 + \widehat{d}_1}{1 + \widehat{d}_2}\right) + O_p(n^{-3/2}), \tag{2.19}$$

O_p indicating stochastic order of magnitude as usual.

Comparing (2.17) with (2.19) and using (2.14) leads to a simple relationship between $\tilde{\alpha}$ and $\widehat{\alpha}$.

LEMMA 1. *Define*

$$\tilde{Z} = \Phi^{-1}(\tilde{\alpha}), \qquad \widehat{Z} = \Phi^{-1}(\widehat{\alpha}). \tag{2.20}$$

Then the third-order relationship between $\widehat{Z}$ and $\tilde{Z}$ is

$$\widehat{Z} = (1 + 2\widehat{d}_2)(\tilde{Z} - 2\widehat{d}_1) + O_p(n^{-3/2}). \tag{2.21}$$

A less precise second-order relationship is

$$\widehat{Z} = (\tilde{Z} - 2\widehat{d}_1) + O_p(n^{-1}). \tag{2.22}$$

The improved bootstrap algorithms of this paper work by doing the first-order bootstrap calculation (2.1) and then improving $\tilde{\alpha}$ by means of the second-order correction (2.22). In order to do so, we need a bootstrap method for approximating $\widehat{d}_1$. This is provided by the following result, verified in the Appendix.

LEMMA 2. *Define $\widehat{z}_0$ as in* (2.9). *Then*

$$\widehat{z}_0 = \frac{\widehat{d}_1}{1+\widehat{d}_2} + O_p(n^{-3/2}) = \widehat{d}_1 + O_p(n^{-1}), \tag{2.23}$$

so $\widehat{z}_0$ is a second-order approximation to $\widehat{d}_1$.

Important: the point $\widehat{\mu}_0$ does not have to be determined very accurately; (2.9) gives second-order accuracy for any $\widehat{\mu}_0$ within $O_p(1)$ of the true nearest point. This is important in the examples of Sections 5 and 7, where the "nearest points" will be found by bootstrap calculations.

Combining the two lemmas gives

$$\widehat{Z} = \tilde{Z} - 2\widehat{z}_0 + O_p(n^{-1}), \tag{2.24}$$

which verifies the theorem (2.8). The quantity $\widehat{z}_0$ is $O_p(n^{-1/2})$, so typically $\widehat{Z} - \tilde{Z}$ will be $O_p(n^{-1/2})$.

As a check on (2.8), consider Example 1 where $\mathscr{R}_1 = \{\mu\colon \|\mu\| \le \theta_1 = 5\}$ and we have observed a four-dimensional vector y with $\|y\| = r = 7$. In this case, we can do the bootstrap computations theoretically, without Monte Carlo, as in (2.4): $\tilde{\alpha} = \text{prob}\{\chi_4^2(49) > 25\} = .988$, $\tilde{Z} = \Phi^{-1}(\tilde{\alpha}) = 2.26$. Definition (2.9) gives

$$\widehat{z}_0 = \Phi^{-1}\{\text{prob}(\chi_4^2(25) > 25)\} = \Phi^{-1}(.619) = .302, \tag{2.25}$$

so (2.8) results in

$$\widehat{\alpha} \doteq \Phi(2.26 - 2(.302)) = .9508. \tag{2.26}$$

This compares with the first-order approximation $\tilde{\alpha} = .9880$ and with the exact value $\widehat{\alpha} = 1 - \widehat{\beta} = .9596$ in (2.7). In this example, the boundary $\mathscr{B}$ curves away from $\widehat{\mu}$, as in Figure 2. This makes $\widehat{z}_0 > 0$ and $\widehat{\alpha} < \tilde{\alpha}$. The opposite happens if $\mathscr{B}$ curves *toward* $\widehat{\mu}$, in which case $\widehat{\alpha}$ would exceed $\tilde{\alpha}$.

If we want a still better approximation, we can use the third-order formula (2.21). In this case, the curvature matrix is easy to calculate theoretically, $\widehat{\mathbf{d}} = I/(2\theta_1)$ for any μ_0 with $\|\mu_0\| = \theta_1$, so

$$\widehat{d}_1 = (K-1)/2\theta_1 = .300, \qquad \widehat{d}_2 = (K-1)/4\theta_1^2 = .030. \tag{2.27}$$

Then (2.21) gives

$$\widehat{Z} \doteq 1.756, \qquad \widehat{\alpha} \doteq .9604, \tag{2.28}$$

an order of magnitude more accurate approximation than (2.26) to $\widehat{\alpha} = .9596$.

Most problems are too complicated to allow these kinds of theoretical calculations. The crucial quantities $\tilde{\alpha}$ and $\widehat{z}_0$ in (2.8) have to be estimated from bootstrap simulations, requiring on the order of 1000 bootstrap replications of y^* in (2.1), and another 1000 of y_0^* in (2.9). Section 5 carries out a second-order bootstrap analysis for the C_p example of Figure 1, illustrating some of the practical pitfalls of the theory.

In some situations, it is possible to directly compute $\widehat{d}_1$, and use (2.22) instead of (2.24) to approximate $\widehat{\alpha}$. This approach is particularly simple and accurate if the boundary $\mathscr{B}$ in Figure 2 is the level set of a smoothly defined real-valued parameter $\theta = t(\mu)$, say $\mathscr{B} = \{\mu\colon t(\mu) = \theta_1\}$. Define

$$\dot{t}_0 = \dot{t}(\widehat{\mu}_0), \qquad \ddot{t}_0 = \ddot{t}(\widehat{\mu}_0), \tag{2.29}$$

where $\dot{t}(\mu)$ is the gradient vector $(\partial t/\partial \mu_i)$ and $\ddot{t}(\mu)$ is the second derivative matrix $(\partial^2 t/\partial \mu_i \, \partial \mu_j)$ assumed to exist continuously as a function of μ. Then it can be shown that

$$\widehat{d}_1 = \frac{1}{2}\left[\frac{\dot{t}_0' \ddot{t}_0 \dot{t}_0}{\|\dot{t}_0\|^3} - \frac{\operatorname{trace}(\ddot{t}_0)}{\|\dot{t}_0\|}\right] \doteq \widehat{z}_0. \tag{2.30}$$

This formula assumes that $t(\mu)$ is increasing as we go from $\mathscr{R}_0$ to $\mathscr{R}_1$ in Figure 2. In Example 1, we could take $t(\mu) = -\|\mu\|$ and $\theta_1 = -5$, in order to get the sign right.

Formula (2.29) is the *ABC method* of calculating the "bias-correction constant" $\widehat{z}_0$ in the normal family $y \sim N_K(\mu, I)$, as described in DiCiccio & Efron (1992), Section 2, except that here the calculation is done at the boundary point $\widehat{\mu}_0$ rather than at $\widehat{\mu} = y$. The first and second derivatives of $t(\mu_0)$ are computed in a numerically efficient way that requires only $2K + 2$ recomputations of $t(\mu)$.

There are two drawbacks to the ABC approach. It fails for situations where the boundary $\mathscr{B}$ consists of piecewise flat facets, as for the multimodality example of Section 7. Second, it requires $\mathscr{B}$ to be described as a level set of a known function $t(\mu)$. This is the kind of mathematical specification we are trying to avoid with our metric-free methods. Efron and Tibshirani (1996), Remark K suggests a way of directly calculating $\widehat{d}_1$ without the specification of $t(\mu)$.

The frequentist paradigm is of great help in understanding and improving upon the first-order bootstrap method. By itself, though, it is not flexible enough to encompass situations like that in Figure 1, let alone the phylogenetic trees of Efron, Halloran and Holmes (1996). Section 3 develops a Bayesian justification for our second-order bootstrap methods that is less precise than the frequentist justification but covers a wider range of situations.

3. Bayesian measures of confidence.

The Bayesian solution to the problem of regions is straightforward: we begin with a prior density $h(\mu)$ for the unknown parameter vector, and take as our measure of confidence in $\mathscr{R}_0$ (the region containing $y = \widehat{\mu}$) the a posteriori probability that $\mu \in \mathscr{R}_0$, say

$$\bar{\alpha} = \operatorname{prob}\{\mu \in \mathscr{R}_0 \mid y\}. \tag{3.1}$$

The difficulty of course lies in the choice of the prior $h(\mu)$, especially for complicated situations such as those of the examples in Sections 5 and 7. This section and the next one relate our bootstrap methods to *objective Bayes* solutions, showing that the bootstrap approximates $\bar{\alpha}$ in (3.1) for choices of $h(\mu)$ that can reasonably be termed "uninformative." The discussion is in terms of the normal model $y \sim N_K(\mu, I)$, with more general results appearing in Section 6.

Suppose first that the prior density $h(\mu)$ for the unknown mean vector μ in (1.1) is perfectly flat,

$$h(\mu) \equiv 1. \tag{3.2}$$

Then the a posteriori density for μ having observed $\widehat{\mu} = y$ is

$$\mu \mid \widehat{\mu} \sim N_K(\widehat{\mu}, I); \tag{3.3}$$

exactly the same as the parametric bootstrap distribution of y^* in (1.2), so the first-order bootstrap method of Section 1 provides an implementation of the flat-prior Bayes analysis. In other words, $\tilde{\alpha}$, (2.1), is the same as $\bar{\alpha}$.

More generally, suppose that our problem involves $J+1$ regions $\mathscr{R}_0, \mathscr{R}_1, \dots, \mathscr{R}_J$ that partition Euclidean K-space $\mathscr{R}^K$,

$$\mathscr{R}^K = \bigcup_{j=0}^{J} \mathscr{R}_j. \tag{3.4}$$

Then the flat-prior Bayes a posteriori probability that $\mu \in \mathscr{R}_j$, say

$$\bar{\beta}_j = \text{prob}\{\mu \in \mathscr{R}_j \mid y\}, \tag{3.5}$$

equals what we called the first-order bootstrap non-confidence value $\tilde{\beta}_j = \text{prob}_{\widehat{\mu}}\{y^* \in \mathscr{R}_j\}$. The bootstrap probabilities .018 for $\mathscr{R}_{\text{con}}$, .117 for $\mathscr{R}_{\text{lin}}$ and .865 for $\mathscr{R}_{\text{quad}}$ reported for Figure 1 are, except for simulation error, the a posteriori probabilities starting from prior (3.2).

This can be taken as support for the first-order bootstrap method, or at least as an argument that it cannot be systematically biased in some way. [Critics of Felsenstein's method had suggested it was biased downward, producing too-small confidence values; see Efron, Halloran and Holmes (1996).] The trouble is that the choice of the flat prior (3.2) is rather arbitrary, and can give answers that disagree with p-values and confidence levels. In Example 1, for instance, case (2.6), $\bar{\beta}_1$ equals .0120 as in (2.7), compared to the p-value $\widehat{\beta} = .0404$.

Welch and Peers (1963) showed how to choose prior densities $h(\mu)$ that better match frequentist confidence levels, providing a more sophisticated theory of what constitutes an uninformative prior density. For example, the prior

$$h^{wp}(\mu) = 1/\|\mu\|^{K-1} \tag{3.6}$$

gives good agreement with frequentist confidence levels for the parameter $\theta = \|\mu\|$ relating to Example 1; see Tibshirani (1989). [Equation (3.6) is the uniform distribution in polar coordinates on $\mathscr{R}^k$.] With this prior, the Bayesian non-confidence value (3.5) becomes $\bar{\beta}_1 = .0408$, nearly matching $\widehat{\beta} = .0404$.

In complicated situations such as those considered in Sections 5 and 7, it becomes impossible to write down the Welch–Peers density. The next section shows that, in a sense, the second-order bootstrap method (2.8) automatically carries out the Welch–Peers calculations, so that the second-order bootstrap result can be thought of as a Bayesian analysis starting from an uninformative prior density.

Why should we be interested in Bayesian solutions to the regions problem? The next example shows that in situations even slightly more complicated than Figure 2, the frequentist p-value interpretation of confidence levels may not make much sense. Remark E of Section 8 gives another such example.

EXAMPLE 2. Suppose we augment (2.2) in Example 1 with a third region $\mathscr{R}_2 = \{\mu\colon \|\mu\| \geq \theta_2\}$, $\theta_2 > \theta_1$, so that $\mathscr{R}_0$ is reduced to the spherical shell between $\mathscr{R}_1$ and $\mathscr{R}_2$,

$$\mathscr{R}_0 = \{\mu\colon \theta_1 < \|\mu\| < \theta_2\}, \tag{3.7}$$

and consider the case

$$r = 7, \qquad \theta_1 = 5, \qquad \theta_2 = 9.5, \qquad K = 4. \tag{3.8}$$

This is the same situation as in (2.6) except that $\mathscr{R}_0$ has been reduced by the subtraction of the outer region $\mathscr{R}_2$. Now how confident should we be that $\mu \in \mathscr{R}_0$?

Let x_0 be the distance from y to the nearest μ vector not in $\mathscr{R}_0$, $x_0 = 2$ in this case. In the normal family $y \sim N_K(\mu, I)$, x_0^2 equals Wilks' likelihood ratio statistic for testing the null hypothesis that μ does not lie in $\mathscr{R}_0$,

$$x_0^2 = -2\log\left[\left\{\sup_{\mathscr{R}_0^c} f_\mu(y)\right\}\Big/\left\{\sup_{\mathscr{R}_0} f_\mu(y)\right\}\right].$$

As before, we take the confidence level for $\{\mu \in \mathscr{R}_0\}$ to be the probability that x_0 is less than its observed value, minimized over μ in the complement of $\mathscr{R}_0$,

$$\widehat{\alpha} = \inf_{\mu\in\mathscr{R}_0^c}\{\text{prob}_\mu(x_0 \leq 2)\}. \tag{3.9}$$

A standard calculation shows that

$$1 - \widehat{\alpha} = \widehat{\beta} = \text{prob}\{7^2 < \chi_4^2(5^2) < 7.5^2\} = .0283 \tag{3.10}$$

for case (3.8). Comparing this with (2.7) shows that reducing $\mathscr{R}_0$ by the subtraction of $\mathscr{R}_2$ has increased our confidence that $\mu \in \mathscr{R}_0$ from .9596 to .9717!

This kind of paradox cannot occur with Bayesian methods. In the spherical shell example, the Welch–Peers prior (3.6) gives Bayesian posterior probabilities (3.5)

$$\bar{\beta}_1 = .0408, \qquad \bar{\beta}_2 = .0036. \tag{3.11}$$

These compare almost perfectly with the frequentist levels for $\mathscr{R}_1$ and $\mathscr{R}_2$ tested *separately* against $\mathscr{R}_0$,

$$\begin{aligned}\widehat{\beta}_1 &= \text{prob}\{\chi_4^2(5^2) > 7^2\} = .0404,\\ \widehat{\beta}_2 &= \text{prob}\{\chi_4^2(9.5^2) < 7^2\} = .0035.\end{aligned} \tag{3.12}$$

Looking at (3.11) and (3.12), a reasonable Bayesian or frequentist assessment of confidence for $\{\mu \in \mathscr{R}_0\}$ in situation (3.8) is

$$\widehat{\alpha} = 1 - \widehat{\beta}_1 - \widehat{\beta}_2 \doteq 1 - \bar{\beta}_1 - \bar{\beta}_2 = .956. \tag{3.13}$$

This is the recipe we will follow in the more complicated example of Section 5: we will use second-order bootstrap analyses to estimate non-confidences $\widehat{\beta}_j$ for each alternative region $\mathscr{R}_j$, and then take $\widehat{\alpha} = 1 - \sum_{j=1}^{J} \widehat{\beta}_j$. The theory presented next provides a Bayesian justification for this recipe.

4. Bootstrap reweighting. Confidence values produced by first-order bootstrap resampling, as in (2.1), are the same as Bayes a posteriori probabilities starting from a flat prior $h(\mu) = 1$. This section shows that using the second-order bootstrap (2.8) amounts to doing a Bayesian analysis starting from a Welch–Peers uninformative prior, for instance $h(\mu) = 1/\|\mu\|^{K-1}$ in Examples 1 and 2. Moreover, the second-order answers are obtained by *reweighting* the first-order resamples according to a simple importance sampling scheme. We continue to work within the normal model $y \sim N_K(\mu, I)$, more general results appearing in Section 6. Formally speaking, our statements of accuracy apply to the asymptotics of (2.13)–(2.14), but in any given situation the notional sample size "n" is fixed and we must, as always, hope that the asymptotic calculations are a good guide to practice.

Suppose that the boundary $\mathscr{B}$ in Figure 2 is the level set of a smoothly defined function $\theta = t(\mu)$, say

$$\mathscr{B}(\theta_1) = \{\mu\colon t(\mu) = \theta_1\}, \tag{4.1}$$

with $\mathscr{R}_0(\theta_1) = \{\mu\colon t(\mu) < \theta_1\}$ and $\mathscr{R}_1(\theta_1) = \{\mu\colon t(\mu) \geq \theta_1\}$. Now we will consider what happens as the value of θ_1 changes.

For a vector μ [an arbitrary point in $\mathscr{R}^K$ not necessarily the true expectation vector μ in (1.1)] and with $y = \widehat{\mu}$ fixed as its observed value, let $W(\mu)$ be the derivative of the confidence level with respect to the confidence value,

$$W(\mu) = d\widehat{\alpha}_\theta / d\tilde{\alpha}_\theta, \tag{4.2}$$

defined as follows: μ determines $\theta = t(\mu)$, $\mathscr{B}(\theta)$, $\mathscr{R}_0(\theta)$ and $\mathscr{R}_1(\theta)$, and then $\tilde{\alpha}_\theta = \text{prob}_{\widehat{\mu}}\{y^* \in \mathscr{R}_0(\theta)\}$; it also determines the nearest point $\widehat{\mu}_0(\theta)$ in Figure 2, and $\widehat{z}_{0\theta} = \Phi^{-1}\{\text{prob}_{\widehat{\mu}_0}(y_0^* \in \mathscr{R}_0(\theta))\}$, (2.9); these give $\widehat{\alpha}_\theta = \Phi(\tilde{Z}_\theta - 2\widehat{z}_{0\theta})$, (2.8), where $\tilde{Z}_\theta = \Phi^{-1}(\tilde{\alpha}_\theta)$ as in (2.20). Finally, $W(\mu) = [d\widehat{\alpha}_\theta/d\theta]/[d\tilde{\alpha}_\theta/d\theta]$. In what follows, we will assume that $\widehat{\alpha}_\theta$ goes from 0 to 1 as θ goes from $-\infty$ to ∞. Differentiating (2.8) gives a simple expression for $W(\mu)$,

$$W(\mu) = \exp\left(\tfrac{1}{2}\left(\tilde{Z}_\theta^2 - \widehat{Z}_\theta^2\right)\right) \quad \text{where } \widehat{Z}_\theta = \tilde{Z}_\theta - 2\widehat{z}_{0\theta}. \tag{4.3}$$

Let $f_\mu(y)$ be the normal density corresponding to (1.1),

$$f_\mu(y) = \frac{1}{(2\pi)^{K/2}} \exp\left(-\frac{1}{2}\|y - \mu\|^2\right), \tag{4.4}$$

and take $y^* \sim N_K(\widehat{\mu}, I)$ as in (1.2). Then the non-confidence value $\tilde{\beta}_{\theta_1} = 1-\tilde{\alpha}_{\theta_1}$ equals $\int_{t(y^*)\geq\theta_1} f_{\widehat{\mu}}(y^*)\,dy^*$. Similarly, we have the following.

LEMMA 3. *For a given value θ_1, the non-confidence level $\widehat{\beta}_{\theta_1} = 1 - \widehat{\alpha}_{\theta_1}$ is*

$$\widehat{\beta}_{\theta_1} = \int_{t(y^*)>\theta_1} f_{\widehat{\mu}}(y^*) W(y^*)\,dy^*. \tag{4.5}$$

PROOF. Let $\widehat{g}(\theta)$ be the bootstrap density

$$\widehat{g}(\theta) = \frac{d}{d\theta}\widehat{G}(\theta), \quad \text{where } \widehat{G}(\theta) = \text{prob}_{\widehat{\mu}}\{\widehat{\theta}^* < \theta\} = \tilde{\alpha}_\theta, \tag{4.6}$$

$\widehat{\theta}^* \equiv t(y^*)$. Integrating over the level sets $t(\mu)$ on the right-hand side of (4.5) gives

$$\int_{t(y^*)>\theta_1} f_{\widehat{\mu}}(y^*) W(y^*)\,dy^* = \int_{\theta_1}^{\infty} \widehat{g}(\theta)\frac{d\widehat{\alpha}_\theta}{d\tilde{\alpha}_\theta}\,d\theta. \tag{4.7}$$

But $\widehat{g}(\theta) = d\tilde{\alpha}_\theta/d\theta$, so

$$\int_{t(y^*)>\theta_1} f_{\widehat{\mu}}(y^*) W(y^*)\,dy^* = \int_{\theta_1}^{\infty} d\widehat{\alpha}_\theta = 1 - \widehat{\alpha}_{\theta_1} = \widehat{\beta}_{\theta_1}. \qquad \square \tag{4.8}$$

Lemma 3 leads to a direct resampling connection between confidence values and confidence levels. Suppose that there are only two regions $\mathscr{R}_0$ and $\mathscr{R}_1$ as in Figure 2, that the boundary $\mathscr{B}$ is the level set $\{t(\mu) = \theta_1\}$ and that $\widehat{\theta} = t(\widehat{\mu})$ is less than θ_1 so $\widehat{\mu} \in \mathscr{R}_0$. We generate $\mathscr{B}$ bootstrap samples from $f_{\widehat{\mu}}$, say

$$y^{*(1)}, y^{*(2)}, \dots, y^{*(B)} \sim f_{\widehat{\mu}}, \tag{4.9}$$

and approximate the first-order bootstrap non-confidence value $\tilde{\beta} = 1 - \tilde{\alpha}$ by

$$\tilde{\beta} \doteq \sum_{t(y^{*(b)})>\theta_1} 1/B. \tag{4.10}$$

Letting $W^{(b)} = W(y^{*(b)})$, notice that the second-order non-confidence level $\widehat{\beta} = 1 - \widehat{\alpha}$ is approximated by

$$\widehat{\beta} \doteq \sum_{t(y^{*(b)})>\theta_1} W^{(b)}/B, \tag{4.11}$$

since (4.11) is the usual Monte Carlo estimate of the integral in (4.5). The weighting factor $W(\mu) = \exp(\tilde{Z}_\theta^2/2 - \widehat{Z}_\theta^2/2)$ is just the ratio of the approximate normal densities for $\widehat{Z}_\theta$ and $\tilde{Z}_\theta$, so we see that $W(\mu)$ acts as an importance sampling factor, converting random variables drawn from $\widehat{\mu}$ into ones drawn from $\widehat{\mu}_0$.

Lemma 3 also has a Bayesian interpretation. A Welch–Peers objective prior density $h^{wp}(\mu)$ such as (3.6) is defined by the fact that it produces Bayesian a posteriori probabilities agreeing to second order with frequentist confidence limits. According to Bayes rule, this means that

$$\widehat{\beta}_{\theta_1} \doteq \int_{t(\mu)>\theta_1} f_\mu(\widehat{\mu}) \frac{h^{wp}(\mu)}{f(\widehat{\mu})}\, d\mu, \tag{4.12}$$

where $f(\widehat{\mu})$ is the marginal density $\int f_\mu(\widehat{\mu}) h^{wp}(\mu)\, d\mu$. See Tibshirani (1989).

We can rewrite (4.5) as

$$\widehat{\beta}_{\theta_1} = \int_{t(\mu)>\theta_1} f_\mu(\widehat{\mu}) W(\mu)\, d\mu. \tag{4.13}$$

A comparison of (4.12) with (4.13) shows that $W(\mu)$ *is* a Welch–Peers prior density, scaled so that it integrates to 1 with respect to the likelihood function $f_\mu(\widehat{\mu})$,

$$W(\mu) = \frac{h^{wp}(\mu)}{f(\widehat{\mu})}, \qquad \int_{\mathscr{R}^K} f_\mu(\widehat{\mu}) W(\mu)\, d\mu = 1, \tag{4.14}$$

the latter following from (4.13) with $\theta_1 \to -\infty$. Remark B of Section 8 describes a more direct connection between a second-order bootstrap analysis and the Welch–Peers theory of uninformative priors.

All of this has the following interpretation: reweighting the first-order bootstrap samples according to $W(y^*)$ as in (4.11) converts $\tilde{\beta}$ into $\widehat{\beta}$; and from a Bayesian viewpoint it converts the flat-prior a posteriori probability for $\mathscr{R}_1$, or for $\mathscr{R}_0$, into the appropriate Welch–Peers a posteriori probability.

In practice, the reweighting does not have to be done since the conversion formula (2.8) gives $\widehat{\alpha}$ or $\widehat{\beta}$ directly once $\widehat{z}_0$ has been calculated. However, (4.11) and its interpretations are helpful for thinking about complicated regional problems such as those in Sections 5 and 7.

Figure 3 is a schematic diagram of a hypothetical problem involving four regions. Resampling from $\widehat{\mu}$ in $\mathscr{R}_0$ has given some y^*'s in $\mathscr{R}_1$, $\mathscr{R}_2$ and $\mathscr{R}_3$. Their proportions are the first-order non-confidence values $\tilde{\beta}_j$. Reweighting these proportions according to (4.11) converts each $\tilde{\beta}_j$ into a non-confidence level $\widehat{\beta}_j$. In Figure 3, $\widehat{\beta}_1$ would be less than $\tilde{\beta}_1$ since the $\mathscr{R}_0/\mathscr{R}_1$ boundary curves toward $\widehat{\mu}_1$, with the opposite being true for $\mathscr{R}_2$ and $\mathscr{R}_3$. From a Bayesian point of view, we can imagine starting with a flat prior on μ, and bending it to accommodate the different boundary situations. Except in simple situations such as Figure 2, it will be difficult to fully describe the bent prior, but it will behave like the appropriate Welch–Peers prior density near each boundary.

5. The C_p example.

Figure 1 provides a realistically complicated example of a regions problem. It refers to a C_p model-selection procedure for the data set of 201 points shown in the left panel of Figure 4. Each point represents a participant in the Control group of the Minnesota arm of the LRC-CPRT, a large-scale investigation of the drug cholostyramine [Efron and Feldman (1991)]. The two measurements for each man are his percentage compliance with the intended dose of the drug (actually a placebo) and his decrease in

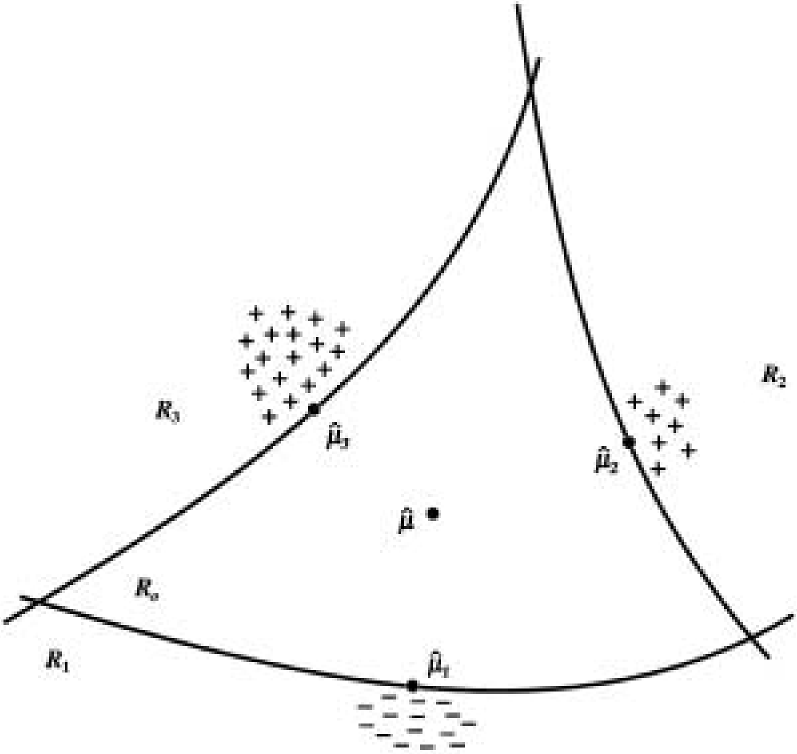

FIG. 3. *Schematic diagram of a four-regions problem,* $\widehat{\mu} \in \mathscr{R}_0$; *first-order bootstrap sampling* $y^* \sim f_{\widehat{\mu}}$ *has yielded some points not in* $\mathscr{R}_0$; *these are shown as minuses in* $\mathscr{R}_1$ *and plusses in* $\mathscr{R}_2$ *and* $\mathscr{R}_3$. *The plusses have weights greater than* 1, *the minuses weights less than* 1, *in accordance with the weights assigned by Welch–Peers composite prior density for* μ. (*Points* y^* *in* $\mathscr{R}_0$ *not shown.*)

total blood cholesterol level over the course of the experiment, this being the response variable of interest.

We consider models that predict cholesterol decrease as a polynomial function of compliance. The right panel of Figure 4 traces the C_p estimate of prediction error, defined below, as a function of the model's degree j, for j going from 0 to 7. There is a sharp minimum at $j = 2$, strongly suggesting a quadratic model. How confident should we be that the quadratic model is actually best?

The cholesterol decrease values were standardized to have approximate variance 1, by division by an estimate of their measurement error $\widehat{\sigma} = 15.09$ based on the residuals from the seventh-degree polynomial model. Then an appropriate rotation of coordinates made the successive coordinates of the response vector $y = \widehat{\mu}$ correspond to the innovations of successive polynomial models. The first eight coordinates of $\widehat{\mu} = (\widehat{\mu}_0, \widehat{\mu}_1, \widehat{\mu}_2, \ldots)$ were

$$\widehat{\mu} = (7.20, 2.74, -2.55, -0.32, -0.09, 0.78, -0.82, 0.53, \ldots). \tag{5.1}$$

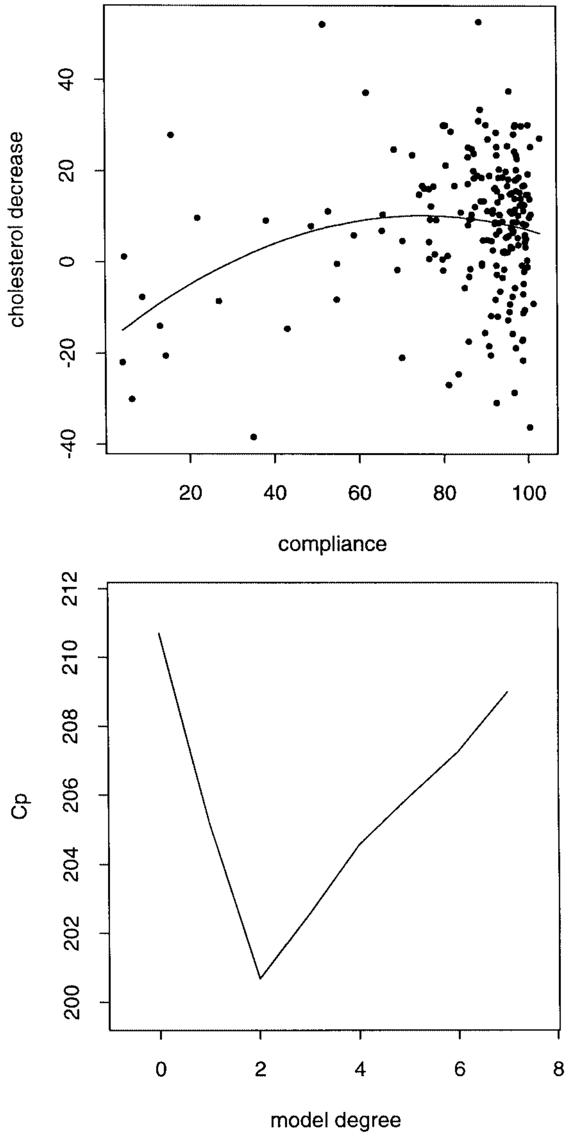

FIG. 4. *Cholesterol decrease versus compliance,* 201 *men in the Minnesota Control group, LRC-CPRT study. Right panel: C_p statistic for model that predicts cholesterol decrease as a jth-degree polynomial in compliance, $j = 0, 1, \ldots, 7$. The C_p statistic is minimized for the quadratic model, solid curve in left panel.*

Thus going from a linear to a quadratic model reduced the residual sum of squares by 2.55^2. We will take

$$y = \widehat{\mu} \sim N_{201}(\mu, I) \tag{5.2}$$

in what follows, assuming normality and ignoring the fact that the standardizing constant $\widehat{\sigma} = 15.09$ was actually an estimate.

The best-fitting quadratic regression curve is shown in the left panel of Figure 4. The choice of a quadratic model was based on the C_p criterion. The C_p value for the jth polynomial model is

$$\widehat{C}_j = \sum_{j+1}^{201} \widehat{\mu}_i^2 + 2(j+1), \tag{5.3}$$

this being an unbiased estimate of the model's prediction error [Mallows (1973)]. If we are only willing to consider models up to degree J, including the constant model with $j = 0$, then the C_p criteria partition $\mathscr{R}^{201}$ into $J+1$ regions, $\mathscr{R}_j$ comprising those $\widehat{\mu}$ where model j is preferred,

$$\mathscr{R}_j = \Big\{\widehat{\mu}\colon \widehat{C}_j = \min_{0 \le i \le J} \widehat{C}_i\Big\}. \tag{5.4}$$

For the cholesterol data, $\widehat{\mu} \in \mathscr{R}_2 \equiv \mathscr{R}_{\text{quad}}$, the quadratic preference region, for all choices of $J \ge 2$. The question of interest can be phrased as: how confident can we be that $\mu \in \mathscr{R}_2$?

NOTE. Even though it is notationally convenient to describe the regions in terms of $\widehat{\mu}$, it is really the region of the vector μ we are interested in. Remark C of Section 8 suggests a different regional set-up that might be preferable here.

Vectors on the boundary between $\mathscr{R}_j$ and $\mathscr{R}_k$, $j < k$, must satisfy

$$\big\{\widehat{C}_j = \widehat{C}_k\big\} \quad \text{or} \quad \Bigg\{\sum_{i=j+1}^{k} \widehat{\mu}_i^2 = 2(k-j)\Bigg\}. \tag{5.5}$$

Figure 1 shows the situation if we only consider models up to degree 2, in which case the three possible boundaries are determined by $(\widehat{\mu}_1, \widehat{\mu}_2) = (2.70, -2.55)$. According to (5.5), the boundary between $\mathscr{R}_j$ and $\mathscr{R}_k$ is a portion of a $(k-j)$-dimensional sphere of radius $[2(k-j)]^{1/2}$. The ABC formula (2.30) can be calculated theoretically in this case. For the portion of the jth region's boundary with $\mathscr{R}_2$, we have

$$\widehat{z}_{0j} = \pm\frac{|j-2|-1}{\sqrt{8|j-2|}} \quad \begin{cases} + & \text{for } j < 2, \\ - & \text{for } j > 2. \end{cases} \tag{5.6}$$

Figure 1's geometry is reflected in (5.6), with $\widehat{z}_{01} = 0$ for the flat $\mathscr{R}_{\text{lin}}/\mathscr{R}_{\text{quad}}$ boundary, while $\widehat{z}_{00} > 0$ for the $\mathscr{R}_{\text{con}}/\mathscr{R}_{\text{quad}}$ boundary, which curves away from $\widehat{\mu}$.

Table 1 presents a regional analysis of the confidence level for $\mu \in \mathscr{R}_{\text{quad}}$. We consider four other regions, $\mathscr{R}_0$ constant, $\mathscr{R}_1$ linear, $\mathscr{R}_3$ cubic, $\mathscr{R}_4$ quartic.

TABLE 1
*Regional analysis for the cholesterol data C_p problem**

Region: **Model:**	$\mathscr{R}_0$ **constant**	$\mathscr{R}_1$ **linear**	$\mathscr{R}_3$ **cubic**	$\mathscr{R}_4$ **quartic**	$\mathscr{R}_2$ **quadratic**
$\tilde{\beta}_j$	.019	.096	.142	.076	$\tilde{\alpha} = .667$
$\hat{z}_{0j}$	.250	.000	.000	-.250	
$\hat{\beta}_j$	.057	.096	.142	.027	$\hat{\alpha} = .678$

*Five polynomial models considered: constant, linear, quadratic, cubic, quartic; $\widehat{\mu} \in \mathscr{R}_{\text{quad}}$. Value $\tilde{\beta}_j$ is proportion of 1000 bootstrap replications falling into jth region; adjusted levels $\widehat{\beta}_j = 1 - \widehat{\alpha}_j$ from (5.8) with $\widehat{z}_0$ given by ABC formula (5.6). Sum of $\widehat{\beta}_j$ is 0.328, leaving confidence level 0.678 for $\mu \in \mathscr{R}_{\text{quad}}$.

$B = 1000$ bootstrap replications of $y^* \sim N_K(\widehat{\mu}, I)$, perhaps ten times more than necessary for a first-order analysis, gave empirical probabilities

$$\tilde{\beta}_j = \#\{y^{*(b)} \in \mathscr{R}_j; b = 1, 2, \ldots, B\}/B, \quad j = 0, 1, 3, 4, \tag{5.7}$$

these being the non-confidence values pertaining to the four alternative regions. The first two of them differ from the values reported for $\mathscr{R}_{\text{con}}$ and $\mathscr{R}_{\text{lin}}$ in Section 1 because of simulation error, but also because this analysis includes $\mathscr{R}_{\text{cub}}$ and $\mathscr{R}_{\text{quart}}$.

The conversion formula (2.8) can be expressed as an adjustment of the flat-prior non-confidence values $\tilde{\beta}_j$ to non-confidence levels $\widehat{\beta}_j = 1 - \widehat{\alpha}_j$,

$$\widehat{\beta}_j = \Phi[\Phi^{-1}(\tilde{\beta}_j) + 2\widehat{z}_{0j}]. \tag{5.8}$$

This is the same as (2.8) when there are only two regions, but it is more convenient to work with in multiregional situations.

The adjusted values $\widehat{\beta}_j$ appear in the bottom line of Table 1. The adjustments are quite substantial, multiplying $\widehat{\beta}_0$ by 3 and dividing $\widehat{\beta}_4$ by nearly as much. The objective Bayesian interpretation of these results is as follows: $\mu \in \mathscr{R}_2$ with a posteriori probability

$$\widehat{\alpha} = 1 - (.057 + .096 + .142 + .027) = .678, \tag{5.9}$$

the most likely alternative region being $\mathscr{R}_3$ with probability .142, etc. Notice that $\widehat{\alpha}$ is nearly the same as $\tilde{\alpha}$ despite big changes for regions $\mathscr{R}_0$ and $\mathscr{R}_4$.

The bootstrap formula (2.9) for estimating $\widehat{z}_0$ did not work well in the C_p example. Unlike the spherical shell situation of Example 2, this situation locates $\widehat{\mu}$ near three-way and higher-way regional boundaries, for example the point $(\sqrt{2}, -\sqrt{2})$ in Figure 1. A typical boundary point $\widehat{\mu}_0$ on the boundary between the constant and quadratic regions, selected as in Section 7, had $y_0^* \sim N_K(\widehat{\mu}_0, I)$ falling into $\mathscr{R}_j$ with these probabilities:

$$\mathscr{R}_0\text{: } 27\% \qquad \mathscr{R}_1\text{: } 13\% \quad \mathscr{R}_2\text{: } 29\% \quad \mathscr{R}_3\text{: } 12\% \quad \mathscr{R}_4\text{: } 19\%. \tag{5.10}$$

This kind of global "leakage" undercuts the local asymptotics behind Lemma 2. The problem did not occur in the other examples of this paper. It can be avoided here by following the tactic used in Efron, Halloran and Holmes (1996) for the tree example, where the alternative regions were considered one at a time instead of simultaneously. For example, we can consider just two regions, whether or not $\widehat{C}_0$ exceeds $\widehat{C}_2$, in which case $\widehat{z}_0$ obtained from (2.9) nearly equals the ABC value .250.

6. More general probability families.

So far we have considered only the normal family $y \sim N_K(\mu, I)$. A similar theory goes through almost as easily for multiparameter exponential families. The key result for the normal family was (2.8), which converts first-level bootstrap confidence values $\tilde{\alpha}$ into second-order frequentist confidence levels $\widehat{\alpha}$. The analogue of (2.8) for general exponential families appears under the names "BC_a" and "ABC" in Efron (1987) and DiCiccio and Efron (1992). Here is a brief review of what we will call the ABC conversion theory.

We now suppose that the observed data vector y has its density function $f_\mu(y)$ in a K-parameter exponential family of densities:

$$f_\mu(y) = e^{\eta y - \psi(\eta)}, \tag{6.1}$$

where y is the K-dimensional sufficient statistic, μ is the expectation parameter vector $\mu = E_\mu\{y\}$ and η is the natural or canonical parameter vector, a one-to-one function of μ; $\psi(\eta)$ is a normalizing function designed to make $f_\mu(y)$ integrate to 1 with respect to some common carrier density for the family. The second derivative matrix of $\psi(\eta)$ with respect to η gives the covariance matrix of y_0 at the corresponding value of μ,

$$\Sigma(\mu) = \operatorname{cov}_\mu(y). \tag{6.2}$$

Again we suppose that there are just two regions $\mathscr{R}_0$ and $\mathscr{R}_1$ in K-dimensional space separated by a boundary $\mathscr{B}$ defined by a smooth real-valued parameter $\theta = t(\mu)$, say $\mathscr{B} = \{\mu\colon t(\mu) = \theta_1\}$, with $\mathscr{R}_0 = \{t(\mu) < \theta_1\}$ and $\mathscr{R}_1 = \{t(\mu) \geq \theta_1\}$. As in Figure 2, the maximum likelihood estimate $\widehat{\mu} = y$ determines a "closest" point $\widehat{\mu}_0$ on $\mathscr{B}$, which we take to be the restricted MLE of μ given $t(\mu) = \theta_1$. The MLE of θ is $\widehat{\theta} = t(\widehat{\mu})$. For convenience, assume $\widehat{\theta} < \theta_1$ so $\widehat{\mu} \in \mathscr{R}_0$ as in Figure 2. The first-order bootstrap confidence value for $\{\mu \in \mathscr{R}_0\}$ is still given by (2.1),

$$\tilde{\alpha} = \operatorname{prob}_{\widehat{\mu}}\{y^* \in \mathscr{R}_0\} = \operatorname{prob}_{\widehat{\mu}}\{t(y^*) < \theta_1\}, \tag{6.3}$$

the notation now indicating that

$$y^* \sim f_{\widehat{\mu}}. \tag{6.4}$$

The frequentist confidence level $\widehat{\alpha}$ can be taken to be

$$\widehat{\alpha} = \operatorname{prob}_{\widehat{\mu}_0}\bigl\{t(y_0^*) > \widehat{\theta}\bigr\}, \tag{6.5}$$

where y_0^* is drawn from the f_μ density having $\mu = \widehat{\mu}_0$.

Corresponding to (2.8), there is a second-order accurate formula for converting $\tilde{\alpha}$ to $\widehat{\alpha}$. Letting $\tilde{Z} = \Phi^{-1}(\tilde{\alpha})$, the *ABC conversion formula* is

$$\widehat{Z} = \frac{\tilde{Z} - \widehat{z}_0}{1 + \widehat{a}(\tilde{Z} - \widehat{z}_0)} - \widehat{z}_0, \qquad \widehat{\alpha} = \Phi(\widehat{Z}). \tag{6.6}$$

Here $\widehat{z}_0$ is the *bias-correction* quantity appearing in (2.9),

$$\widehat{z}_0 = \Phi^{-1}\{\text{prob}(y_0^* \in \mathscr{R}_0)\}, \qquad y_0^* \sim f_{\widehat{\mu}_0}, \tag{6.7}$$

while $\widehat{a}$, the *acceleration*, is another $O_p(n^{-(1/2)})$ quantity, calculated from (6.9) below. Formula (6.6), like (2.22), errs by order $O_p(n^{-1})$. In the normal family (1.1), $\widehat{a} = 0$, so (6.6) is the same as (2.22), but $\widehat{a}$ can make a substantial difference in other families.

To compute $\widehat{a}$, we first find $\widehat{\eta}_0$, the natural parameter vector corresponding to the restricted MLE $\widehat{\mu}_0$ on $\mathscr{B}$, and $\dot{t}_0 = \dot{t}(\widehat{\mu}_0) = (\partial t/\partial \mu_i)_{\widehat{\mu}_0}$, which determines the delta-method estimate of variance for $\widehat{\theta}$ (evaluated at $\mu = \widehat{\mu}_0$),

$$\widehat{\sigma}_0^2 = \dot{t}_0' \widehat{\Sigma}_0 \dot{t}_0 \quad (\Sigma_0 \equiv \Sigma(\widehat{\mu}_0)). \tag{6.8}$$

Then the acceleration $\widehat{a}$ is computed from

$$\widehat{a} = \frac{\partial^2}{\partial \varepsilon}\big[\dot{t}_0' \mu(\widehat{\eta}_0 + \varepsilon \dot{t}_0)\big]_{\varepsilon = 0} / (6\widehat{\sigma}_0^3), \tag{6.9}$$

where $\mu(\eta)$ indicates the vector μ expressed as a function of η. This is formula (2.9) of DiCiccio and Efron (1992), carried out at $\widehat{\mu}_0$ instead of $\widehat{\mu}$. Formula (6.7) is numerically evaluated by substituting small values of ε into the bracketed term. Section 7 discusses the calculation of $\widehat{a}$ in the case where we can only tell whether or not a given vector $y^* \in \mathscr{R}_0$.

A second-order accurate bootstrap algorithm for approximating the confidence level $\widehat{\alpha}$ proceeds as follows: first-order bootstrap sampling gives a Monte Carlo estimate of $\tilde{\alpha}$ as in (6.3); second-level bootstrap sampling, $y_0^* \sim f_{\widehat{\mu}_0}$, gives a Monte Carlo estimate of $\widehat{z}_0$ as in (6.7); the method described in Section 7 gives $\widehat{a}$; and finally the ABC conversion formula (6.6) gives $\widehat{\alpha}$. The multimodality example of Section 7 carries through this algorithm for a case where the regions have complicated multifaceted linear boundaries.

Here is a much simpler case where we can check the algorithm's accuracy. CD4 counts were measured for 20 HIV-positive subjects before and after administration of an experimental antiviral drug: $x_i = (B_i, A_i)$ for $i = 1, 2, \ldots, 20$. Assuming a bivariate normal family, with unknown mean vector and covariance matrix [*not* (1.1)], the $K = 5$-dimensional sufficient statistic $y = (\bar{B}, \bar{A}, \bar{B}^2, \overline{BA}, \bar{A}^2)$ was

$$y = \widehat{\mu} = (3.29, 4.09, 11.44, 14.10, 18.03), \tag{6.10}$$

giving sample correlation coefficient $\widehat{\theta} = .723$. [The data set appears as Table 1 of DiCiccio and Efron (1992).] What confidence should we attach to the region $\mathscr{R}_0 = \{\theta > .5\}$, the set of bivariate normal distributions with correlation exceeding 0.50? The exact one-sided p-value of $\widehat{\theta} = 0.723$ under the null hypothesis that $\theta = 0.50$ is

$$\widehat{\beta} = .072 \tag{6.11}$$

in a bivariate normal family.

Four thousand first-order bootstrap vectors y were sampled from the bivariate normal distribution determined by (6.10), of which 219 had sample correlation coefficient $\widehat{\theta}^* \leq .50$, giving

$$\tilde{\beta} = .0548 = 219/4000. \tag{6.12}$$

Several boundary vectors $\widehat{\mu}_0$ were located according to the method described in Section 7, all of which had $\widehat{a} = .000$ and $\widehat{z}_0 = .0559$ using (6.9) and (6.13) below. The conversion formula (6.6) then gave $\widehat{\beta} = .068$, agreeing with the exact level (6.11), within the simulation error.

There is an analytic formula like (2.30) for approximating $\widehat{z}_0$ without having to do the Monte Carlo simulations of (6.7):

$$\widehat{z}_0 \doteq \widehat{a} + \tfrac{1}{2}\big[\widehat{\delta}'\ddot{t}_0\widehat{\delta}/\widehat{\sigma}^3 - \mathrm{tr}\big(\ddot{t}_0\widehat{\Sigma}_0\big)/\widehat{\sigma}\big] \quad \big(\widehat{\delta} = \widehat{\Sigma}_0\dot{t}_0\big). \tag{6.13}$$

Formulas (6.9) and (6.13) are easy to evaluate numerically as shown in DiCiccio and Efron (1992), but they only apply to situations where the boundary is the level set of a known function $\theta = t(\mu)$. Section 7 discusses situations where this is not the case.

These results extend the frequentist theory of Section 2 to multiparameter exponential families. It is also easy to extend the Bayes/bootstrap theory of Sections 3 and 4. Differentiating (6.6) gives

$$W(\mu) = \exp\bigg(\frac{1}{2}\big(\tilde{Z}_0^2 - \widehat{Z}_0^2\big)\bigg)\bigg(\frac{\widehat{Z}_0 + \widehat{z}_0}{\tilde{Z}_0 - \widehat{z}_0}\bigg)^2, \tag{6.14}$$

in place of (4.3). Lemma 3 and its interpretations go through with just this change, as shown in Efron and Tibshirani (1996), Section 7.

The *Poisson family* is a particularly useful exponential family (6.1): we observe K independent Poisson variates $(y_1, y_2, \ldots, y_K) \equiv y$ with expectations $(\mu_1, \mu_2, \ldots, \mu_K) \equiv \mu$,

$$y \sim Po_K(\mu). \tag{6.15}$$

Then $\eta = \log(\mu) = (\log(\mu_1), \log(\mu_2), \ldots, \log(\mu_K))$, $\psi(\eta) = \Sigma e^{\eta_k}$, $\Sigma(\mu) = \mathrm{diag}(\mu)$ the $K \times K$ diagonal matrix with diagonal elements $\mu_1, \mu_2, \ldots, \mu_K$ and $\widehat{\sigma}^2 = \sum_{k=1}^K y_k \dot{t}_{0k}^2$. The numerator in (6.9) for $\widehat{a}$ is $(\partial^2/\partial\varepsilon^2)[\Sigma \dot{t}_{0k} \exp\{\widehat{\eta}_{0k} + \varepsilon \dot{t}_{0k}\}]_0$.

Another useful case is the *multinomial family*, where we observe a vector of proportions p from n independent draws on K categories,

$$p \sim \mathrm{Mult}_K(n, \pi)/n. \tag{6.16}$$

Here $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$, the vector of true probabilities, plays the role of μ, with the simplex of possible π vectors being partitioned into regions $\mathscr{R}_j$.

It turns out that the Poisson family also applies to the multinomial. Instead of (6.16), we can assume that the count vector $y = np$ is in the Poisson family (6.15), with regions $\mathscr{R}'_j$ determined by

$$\mathscr{R}'_j = \{\mu = c\pi \text{ for } \pi \in \mathscr{R}_j,\ c > 0\}. \tag{6.17}$$

Then the second-level bootstrap algorithm will give the same results in the Poisson and multinomial families, and so will (6.9) and (6.13); see DiCiccio and Efron (1992). If the regions $\mathscr{R}_j$ are determined by a real-valued statistic $\widehat{\theta} = t(p)$, then we need to take $\widehat{\theta} = t(y/\Sigma y_k)$ in the Poisson formulation.

Nonparametric applications of the bootstrap use the multinomial model with $K = n$, the number of independent observations and $\widehat{\pi} = (1, 1, \dots, 1)/n$; see Efron (1987), Section 7. The multimodality example of Section 8 will be carried out nonparametrically and will make use of the Poisson–multinomial relationship to simplify some of the calculations.

7. Multimodality example. This section discusses a nonparametric example of the problem of regions, one where it is difficult to describe the regional boundaries in terms of a real-valued parameter $\theta = t(\mu)$. The process of executing Section 6's two-level bootstrap analysis illustrates some of its practical and interpretational difficulties, as well as its power.

Figure 5 shows a histogram of the data, the thicknesses in millimeters of $n = 485$ stamps issued in 1872, the Hidalgo issue of Mexico, and also a Gaussian kernel estimate of the thickness density. The kernel estimate has two modes, but how much confidence should we have that the true density function is bimodal? This is a question of philatelic interest since it is suspected that the issue might be a mixture, printed on more than one type of paper [Efron and Tibshirani (1993), Section 16.5; Izenman and Sommer (1988)].

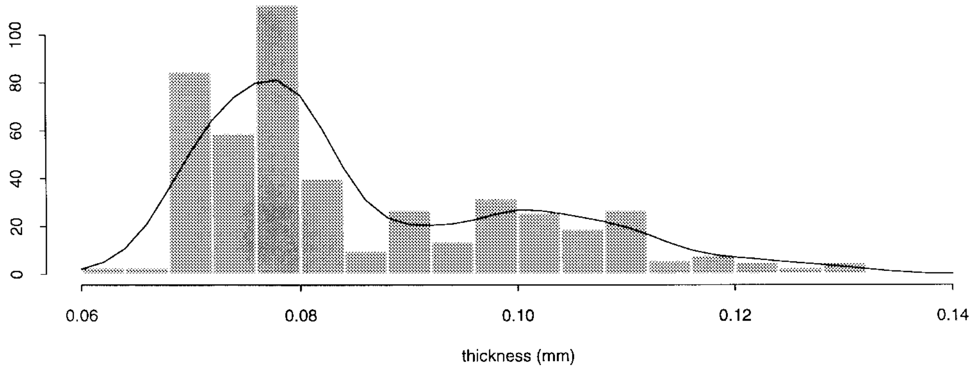

FIG. 5. *The stamp data*; *thicknesses of* 485 *stamps issued in* 1872; *solid curve a Gaussian kernel density estimate, is bimodal, suggesting that the stamps were issued on two different papers. How confident can we be that the true density is bimodal?*

The density estimate in Figure 5 is

$$\widehat{d}(x) = \frac{1}{nh}\sum_{i=1}^{n}\varphi\left(\frac{x - x_i}{h}\right) \quad (h = .0039), \tag{7.1}$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is the stamp data and $\phi(t)$ is the normal kernel $\exp(-t^2/2)\sqrt{2\pi}$. The kernel width $h = .0039$ was chosen as best by a 10-fold cross-validation procedure. In what follows, h will be fixed at this value, though in principle it would be no more difficult to recompute h by cross-validation for each bootstrap sample.

Bootstrap samples $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ were selected in the usual nonparametric way as a random sample of size n drawn with replacement from $x_1, x_2, \dots, x_n$. Each $\mathbf{x}^*$ gave a bootstrap density estimate $\widehat{d}^*$,

$$\widehat{d}(x)^* = \frac{1}{nh}\sum_{i=1}^{n}\varphi\left(\frac{x - x_i^*}{h}\right), \tag{7.2}$$

having *modality* $\widehat{M}^*$ defined as follows: let

$$\widehat{d}_k^* = \widehat{d}(s_k)^* \quad \text{for } s_k = .06 + .002k \quad (k = 0, 1, 2, \dots, 40), \tag{7.3}$$

and define s_k as a mode of $\widehat{d}^*$ if

$$\widehat{d}_k^* > \widehat{d}_{k-1}^* \quad \text{and} \quad d_k^* > d_{k+1}^* \tag{7.4}$$

(with $\widehat{d}_{-1}^*$ and $\widehat{d}_{41}^* = 0$). Then $\widehat{M}^*$ is the number of such modes for $\widehat{d}^*$, with $\widehat{M} = 2$ for the original estimate (7.1). We wish to make a statement of confidence for M, the modality as defined in (7.3)–(7.4), for the true density d. It should be noted that the choice $h = .0039$ is crucial in defining $\widehat{d}$'s modality [though the mesh width .002 in (7.3) is not]. For example, $h = .0039/2$ gives $\widehat{M} = 7$. This kind of definitional dependence is important of course, but it has no special relevance to the regions problem. The analysis which follows could be carried out in the same way for any choice of h.

One thousand nonparametric bootstrap samples gave the modalities shown in Table 2. We see, perhaps surprisingly, that there is very little chance of unimodality, the only substantial alternative being trimodality. With this in mind, we consider only two regions in what follows:

$$\begin{aligned} \mathscr{R}_0 &= \{d \text{ has 2 or fewer modes}\} \quad \text{versus} \\ \mathscr{R}_1 &= \{d \text{ has three or more modes}\}, \end{aligned} \tag{7.5}$$

TABLE 2

Modalities $\widehat{M}^*$ of 1000 nonparametric bootstrap density estimates for the stamp data

Modality	**1**	**2**	**3**	**4**
Number	4	850	144	2

called "bimodal" versus "trimodal" for simplicity. Table 2 gives first-order confidence value $\tilde{\alpha} = .854$ for $\mathscr{R}_0$, or non-confidence $\tilde{\beta} = 1 - \tilde{\alpha} = .146$.

A nonparametric bootstrap sample $\mathbf{x}^*$ can be represented by $\mathbf{p}^* = (p_1^*, p_2^*, \dots, p_n^*)$, where p_i^* is the proportion of bootstrap values x_j^* equalling the ith original value x_i,

$$p_i^* = \#\{x_j^* = x_i\}/n \quad \text{for } i = 1, 2, \dots, n. \tag{7.6}$$

The vector $\mathbf{p}^*$ has a scaled multinomial distribution on n categories,

$$\mathbf{p}^* \sim \text{Mult}_n(n, \widehat{\pi})/n, \qquad \widehat{\pi} = (1, 1, \dots, 1)/n; \tag{7.7}$$

$\mathbf{p}^*$ and $\widehat{\pi}$ play the roles of y^* and $\widehat{\mu}$ in the normal model (1.2). We can consider a nonparametric regional problem in terms of the multinomial family (6.16) with $K = n$, where the MLE of π is $\widehat{\pi} = (1, 1, \dots, 1)/n$; see Efron (1987), Section 8. The multinomial is an exponential family, which allows us to use the theory of Section 6 for converting $\tilde{\alpha}$ into a second-order accurate confidence level $\widehat{\alpha}$.

The triangle in Figure 6 represents the simplex $\mathscr{S}_n$ of n-dimensional probability vectors,

$$\mathscr{S}_n = \left\{\mathbf{p}\colon p_i \geq 0 \text{ and } \sum_{i=1}^{n} p_i = 1\right\}. \tag{7.8}$$

Each probability vector $\mathbf{p}$ determines a density

$$\widehat{d}_{\mathbf{p}}(x) = \frac{1}{h} \sum_{i=1}^{n} p_i \varphi\left(\frac{x - x_i}{h}\right), \tag{7.9}$$

this formula agreeing with $\widehat{d}^*$ in (7.2) for a bootstrap probability vector $\mathbf{p}^*$. Figure 6 partitions $\mathscr{S}_n$ into regions $\mathscr{R}_0$ and $\mathscr{R}_1$ corresponding to definition (7.5). The MLE $\widehat{\pi}$ is in $\mathscr{R}_0$, while 146 of the 1000 bootstrap vectors $\mathbf{p}^*$ fell into $\mathscr{R}_1$.

Figure 6 shows one of the 146 $\mathbf{p}^*$ vectors in $\mathscr{R}_1$. Also shown is the boundary point

$$\widehat{\pi}_0 = w\mathbf{p}^* + (1 - w)\pi, \tag{7.10}$$

found by the binary search algorithm in Efron, Halloran and Holmes (1996): starting with $w_1 = .5$, we check whether or not $w_1\mathbf{p}^* + (1 - w_1)\widehat{\pi}$ is in $\mathscr{R}_0$. If it is, then $w_2 = .25$; if not, then $w_2 = .75$, etc. Twenty steps of the binary search determines w in (7.10) within $1/2^{20}$. Notice that *the search process is metric-free* in the sense that the only computations involve whether or not a vector $\mathbf{p} \in \mathscr{R}_0$, in this case whether or not $\widehat{d}(\mathbf{p})$ has no more than two modes.

Twenty-five of the 146 $\mathbf{p}^*$ vectors in $\mathscr{R}_1$ were selected at random for use in the second level of bootstrap computations, say $p^{*(j)}$ for $j = 1, 2, \dots, 25$, each yielding a boundary vector $\widehat{\pi}_0^{(j)}$ by the binary search process. Each $\widehat{\pi}_0^{(j)}$ then gave 400 independently generated second-level bootstrap samples,

$$\mathbf{p}_0^{*(jk)} \sim \text{Mult}_n(n, \widehat{\pi}_0^{(j)})/n \qquad (k = 1, 2, \dots, 400; n = 485), \tag{7.11}$$

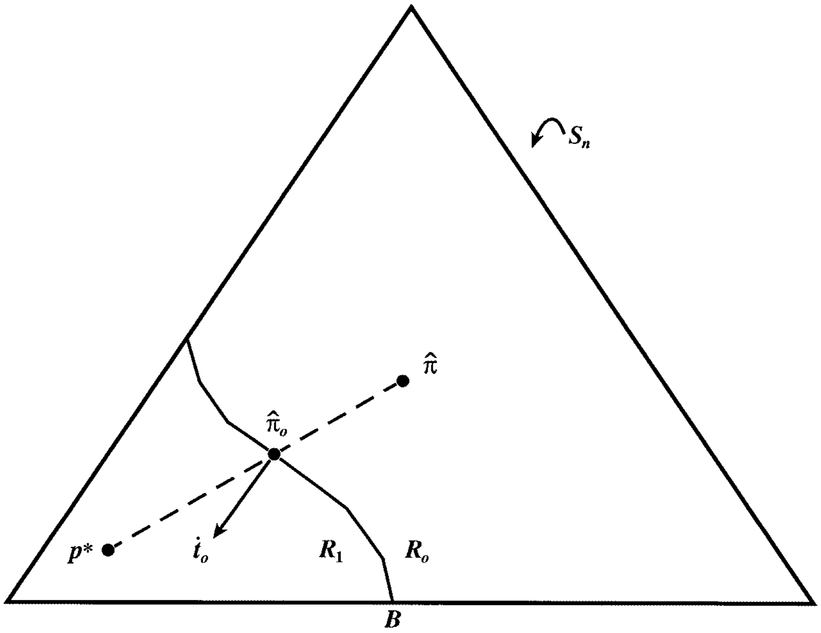

FIG. 6. *Multinomial representation of the multimodality problem*; $\mathscr{S}_n$ *is the probability simplex in n dimensions*; *MLE* $\hat{\pi} = (1, 1, \dots, 1)/n$ *is in* $\mathscr{R}_0$, *the set of probability vectors giving bimodality*; 14.6% *of the bootstrap vectors* $\mathbf{p}^*$ *fall into* $\mathscr{R}_1$, *the region of trimodality*; *each such* $\mathbf{p}^*$ *gives a boundary vector* $\hat{\pi}_0$ *according to the binary search procedure* (7.10).

and an estimate of the bias-correction $\hat{z}_0$,

$$\hat{z}_0^{(j)} = \Phi^{-1}\left\{\frac{\#(\mathbf{p}_0^{(jk)} \in \mathscr{R}_0)}{400}\right\}. \tag{7.12}$$

The average of the 25 $\hat{z}_0^{(j)}$ values was -0.152. This says that, in a global sense, the boundary between $\mathscr{R}_0$ and $\mathscr{R}_1$ curves *toward* $\hat{\pi}$, so that $\hat{\alpha}$, the confidence level for bimodality, should be bigger than the confidence value $\tilde{\alpha} = 1 - \tilde{\beta} = .854$. The average value of the accelerations $\hat{a}^{(j)}$, computed as at (7.13) below, was only .006. Using $\hat{z}_0 = -.152$ and $\hat{a} = .006$ in the conversion formula (6.6) gives $\widehat{\beta} = .089$, as shown in the top line of Table 3, so $\hat{\alpha} = 1 - \widehat{\beta} = .911$.

In simple situations such as that of Figure 2, it can be demonstrated that averaging over the choice of boundary points, 25 of them in this case, produces second-order accurate confidence levels. There is also a rough Bayesian justification; see Remark E of Section 8. The main advantage of this tactic is that it avoids having to calculate a "nearest point" $\hat{\mu}_0$ or $\hat{\pi}_0$. It also provides a more global picture of the boundary geometry, as further analysis of the multimodality problem showed.

TABLE 3
Confidence levels for bimodality, the stamp data[*]

Region	$\tilde{\boldsymbol\beta}$	$\widehat{\boldsymbol z}_{\mathbf 0}$	$\widehat{\boldsymbol a}$	$\widehat{\boldsymbol\beta}$	$\widehat{\boldsymbol\alpha}$
Combined trimodal $\mathscr{R}_1$	.146	−.152	.006	.089	.911
Third mode far right	.111	−.264	.008	.042	.878
Third mode at .108	.035	.202	.002	.080	

[*] Top line assumes a single alternative region $\mathscr{R}_1$ as in (7.5); bottom lines partition $\mathscr{R}_1$ into alternative regions as in text. Both analyses yield confidence levels for bimodality bigger than the first-order confidence value $\tilde{\alpha} = 1 - \tilde{\beta} = .854$.

As **p** passes from $\mathscr{R}_0$ to $\mathscr{R}_1$ through boundary point $\widehat{\pi}_0$ in Figure 6, a third mode emerges on the density estimate $\widehat{d_{\mathbf{p}}}(x)$, say at location $s^{(j)}$, for point $\widehat{\pi}_0^{(j)}$. Among the 25 cases, 19 had $s^{(j)} \geq .120$, at the far right end of the x-axis in Figure 5, and for those 19 cases the $\widehat{z}_0^{(j)}$ values were all negative, averaging -0.264. The remaining six cases behaved differently, having third mode at .108, just to the right of the actual second mode in Figure 5, with these $\widehat{z}_0^{(j)}$ averaging $+0.202$.

In other words, there were two alternative regions for bimodality, one with the boundary curving toward $\widehat{\pi}$ and the other with the boundary curving away (as in Example 2 of Section 3 or in the schematic diagram of Figure 3). The first-order confidence value $\tilde{\alpha} = \text{prob}_{\widehat{\pi}}\{\mathbf{p}^* \in \mathscr{R}_0\}$ depends only on $\mathscr{R}_0$ and not on how we partition the complement of $\mathscr{R}_0$. The second-order answer $\widehat{\alpha}$ does depend on the partition, which may be naturally defined as in the C_p example or emerge from the analysis as here. In this case, the bottom lines of Table 3 show that partitioning $\mathscr{R}_0$ reduces $\widehat{\alpha}$ from .911 to .878.

Without trying to put too fine a point on our methodology, we can state the following conclusions: the first-order analysis gives substantial confidence for bimodality, $\tilde{\alpha} = .85$, with trimodality and not unimodality being the only viable alternative. The second-order analysis suggests that, if anything, $\tilde{\alpha}$ is too low, a better confidence level being $\widehat{\alpha}$ in the range .88 to .91.

The values of $\widehat{\alpha}$ in Table 3 are based on (6.9), which has a simple form in the nonparametric situation,

$$\widehat{a} = \sum_{i=1}^{n} \widehat{\pi}_{0i} \dot{t}_{0i}^{3} \Bigg/ \left[6\sqrt{n} \left(\sum_{i=1}^{n} \widehat{\pi}_{0i} \dot{t}_{0i}^{2} \right)^{3/2} \right], \tag{7.13}$$

$\dot{t}_0$ being an orthogonal vector to $\mathscr{B}$ at boundary point $\widehat{\pi}_0$, pointing outwards from $\mathscr{R}_0$ as in Figure 6. Formula (7.13) comes from the multinomial representation (7.7) and the Poisson–multinomial connection mentioned after (6.16). Notice that multiplying $\dot{t}_0$ by any positive constant does not change $\widehat{a}$. Efron, Halloran and Holmes (1996) used $\dot{t}_0 = \widehat{\pi}_0 - \widehat{\pi}$ in (7.13), also replacing $\widehat{\pi}_{0i}$ with $1/n$, which is easy but not very accurate.

In this case, it is easy to compute $\dot{t}_0$. Using the notation of (7.3)–(7.4), suppose that the density (7.9) corresponding to $\widehat{\pi}_0$ has its third mode emerging at s_k as we go from $\mathscr{R}_0$ to $\mathscr{R}_1$, with

$$\widehat{d}_k = \widehat{d}_{k-1}, \qquad \widehat{d}_k > \widehat{d}_{k+1}. \tag{7.14}$$

(This was the case for all 25 of the selected boundary points $\widehat{\pi}_0^{(j)}$.) Let $\mathbf{o}_k$ be the vector with ith element

$$\mathbf{o}_{ki} = \varphi\left(\frac{s_k - x_i}{h}\right) - \varphi\left(\frac{s_k - 1 - x_i}{h}\right), \qquad i = 1, 2, \dots, n. \tag{7.15}$$

It follows from definition (7.9) that $\mathbf{o}_k$ is orthogonal to $\mathscr{B}$ at $\widehat{\pi}_0$, pointing out of $\mathscr{R}_0$, and so $\dot{t}_0$ in (7.13) can be set equal to $\mathbf{o}_k$ for the estimation of $\widehat{a}$.

The boundary $\mathscr{B}$ in Figure 8 consists of flat facets determined by $\mathbf{o}_k'\mathbf{p} = 0$. The local curvature of $\mathscr{B}$ is zero except where the facets join, but the kind of global curvature captured by (7.12) is quite substantial. Efron and Tibshirani (1996), Remark C presents a simple example where this kind of global curvature gives accurate quantitative predictions. Remark F of Section 8 gives a metric-free algorithm for computing $\dot{t}_0$ in the absence of a theoretical expression like (7.15).

8. Remarks.

We conclude with some remarks.

REMARK A. A third-order bootstrap analysis is possible, based on the idea of calibration. Working within the normal model $y \sim N_K(\mu_1 I)$, we first sample $y_0^* \sim N_K(\widehat{\mu}_0, I)$ as in (2.9) and then do a second tier of resampling,

$$y_0^{**} \sim N_K(y_0^*, I). \tag{8.1}$$

The second-tier quantity

$$\tilde{\alpha}^* = \text{prob}\{y_0^{**} \in \mathscr{R}_0 \mid y_0^*\} \tag{8.2}$$

is seen to be a bootstrap replication of $\tilde{\alpha}$, (2.1).

The *calibrated* value of $\tilde{\alpha}$ is

$$\tilde{\alpha}_{\text{cal}} = \text{prob}_{\widehat{\mu}_0}\{\tilde{\alpha}^* < \tilde{\alpha}\}; \tag{8.3}$$

see Hall (1992), Beran (1987) and Loh (1987). The Appendix of Efron and Tibshirani (1996) shows that, in the context of Section 2, $\tilde{\alpha}_{\text{cal}}$ is a third-order accurate estimate of $\widehat{\alpha}$.

The calibrated confidence level $\tilde{\alpha}_{\text{cal}}$ is metric-free, third-order accurate and applies automatically without special programming to any situation. The problem with using it is the enormous number of bootstrap replications required. First, we need on the order of 1000 y_0^* vectors, and for each of these, a comparable number of second-tier replications y_0^{**}. This requires checking $\{y_0^{**} \in \mathscr{R}_0\}$ perhaps 1,000,000 times in all. Also, because $\tilde{\alpha}_{\text{cal}}$ is defined in terms of frequentist level sets, it can sometimes produce paradoxes of the type encountered in Example 2.

REMARK B. In the case where the boundary $\mathscr{B}$ is the level set of a parameter $\theta = t(\mu)$, the curvature of $\mathscr{B}$ determines the Welch–Peers uninformative prior density $h^{wp}(\mu)$ in the normal family (1.1). Using the notation applying to (2.29)–(2.30), we have

$$\frac{\partial}{\partial x} \log h^{wp}\big(\widehat{\mu}_0 + x\dot{t}_0 / \parallel \dot{t}_0 \parallel\big)\big|_{x=0} = 2\widehat{d_1}, \tag{8.4}$$

this being true because the authors have shown that (8.4) is equivalent to Stein's condition for $h^{wp}(\mu)$; see Tibshirani (1989). Note: (8.4) assumes that $\dot{t}_0$ points away from $\mathscr{R}_0$ into $\mathscr{R}_1$, and that the positive direction for measuring $\widehat{d_1}$, called u in Figure 2, agrees with $\dot{t}_0$.

REMARK C. In the C_p example, we might set

$$\widehat{D}_j = \sum_{j+1}^{201} \widehat{\mu}_i^2 + (j+1), \tag{8.5}$$

and then define the regions $\mathscr{R}_j$ by replacing $\widehat{C}_j$ with $\widehat{D}_j$ in (5.3). This makes sense because

$$D_j = \sum_{j+1}^{201} \mu_i^2 + (j+1) = E_\mu\{\parallel \mu - \widehat{\mu}(j) \parallel^2\}, \tag{8.6}$$

where $\widehat{\mu}(j)$ is the MLE of μ assuming a polynomial model of degree j, so $\mathscr{R}_j$ now becomes the region of μ vectors with $\widehat{\mu}(j)$ having smallest expected squared error.

Table 4 is the equivalent of Table 1; $\widehat{\mu}$ still falls into $\mathscr{R}_2$, but now there is increased probability of $\mathscr{R}_3$, $\mathscr{R}_4$, and decreased probability of $\mathscr{R}_0$, $\mathscr{R}_1$. The ABC values $\widehat{z}_{0j}$ are $\sqrt{2}$ times those in Table 1.

REMARK D. The ABC confidence interval theory of DiCiccio and Efron (1992) evaluates $\widehat{a}$ and $\widehat{z}_0$ at $\widehat{\mu}$. In Section 2, where we are using hypothesis tests rather than confidence intervals, $\widehat{a}$ and $\widehat{z}_0$ are evaluated at boundary points $\widehat{\mu}_0$. Asymptotically, this causes only a third-order difference that does

TABLE 4
Equivalent of Table 1 *if regions defined by* (8.5) *instead of* (5.3)*

Region	$\mathscr{R}_0$	$\mathscr{R}_1$	$\mathscr{R}_3$	$\mathscr{R}_4$	$\mathscr{R}_2$
Model	**constant**	**linear**	**cubic**	**quartic**	**quadratic**
$\tilde{\beta}_j$	.003	.033	.229	.240	$\tilde{\alpha} = .495$
$\hat{z}_{0j}$	.354	.000	.000	-.354	
$\hat{\beta}_j$	.018	.033	.229	.079	$\hat{\alpha} = .641$

*Now the cubic and quartic models have greater probabilities, the constant and linear models less; confidence level $\widehat{\alpha} = .640$ for $\mu \in \mathscr{R}_{\text{quad}}$ is not much different than before, though $\tilde{\alpha}$ is much smaller.

not affect the second-order accuracy of our methods. It would be easier to evaluate $\widehat{a}$ and $\widehat{z}_0$ at $\widehat{\mu}$, but less appropriate to the objective Bayes theory of Section 4, and impossible in situations where the boundary is not defined by a known function $\theta = t(\mu)$.

The ABC theory is motivated in Efron (1987), Section 2, by assuming that a monotone transformation $\phi = m(\theta)$, $\phi = m(\theta)$ produces a normal translation family, possibly with bias and changing variance,

$$\widehat{\phi} \sim N(\phi - z_0\sigma_\phi, \sigma_\phi^2) \quad \text{where } \sigma_\phi = 1 + a\phi. \tag{8.7}$$

In this case, the ABC conversion formula (6.6) is exact, with $\widehat{z}_0 = z_0$ and $\widehat{a} = a$ evaluated at any value of μ.

REMARK E. Frequentist p-values can give misleading results even in problems having only two regions. Consider again the spherical shell example (2.2), but now let

$$\mathscr{R}_0 = \{\mu\colon \|\mu\| \le \theta_1\}, \qquad \mathscr{R}_1 = \{\mu\colon \|\mu\| > \theta_1\}. \tag{8.8}$$

The parameter θ is $\|\mu\|$. Take $\theta_1 = 1.5$, $K = 2$ and suppose we observe $y = \widehat{\mu} = (\sqrt{2}/4,\ \sqrt{2}/4)$ so that $\widehat{\theta} = 0.5$. The first-order confidence value is

$$\tilde{\alpha} = \text{prob}_{\hat{\mu}}(\widehat{\theta}^* < 1.5) = 0.631. \tag{8.9}$$

Now the closest point to $\widehat{\mu}$ on the boundary between $\mathscr{R}_0$ and $\mathscr{R}_1$ is $\widehat{\mu}_0 = (3\sqrt{2}/4, 3\sqrt{2}/4)$, and the pure frequentist confidence value is $\text{prob}_{\hat{\mu}_0}(\widehat{\theta}^* > 0.5) = 0.959$. This value is close to 1 because the set $\{\|\mu\| < .5\}$ is so small that the tail event $\{\widehat{\theta}^* < 0.5\}$ is not assigned much probability. But $\widehat{\mu}$ is not that unlikely when sampling from μ_0. The Bayes a posteriori probability for $\mathscr{R}_0$ is a more reasonable 0.842 under the Weld–Peers prior $1/\|\mu\|$.

REMARK F. The $\widehat{z}_0$ values in Table 5 were obtained by averaging the individual $\widehat{z}_0^{(j)}$ values over the relevant boundary points. There is a rough Bayesian argument in favor of this tactic. Going back to the normal family $y \sim N_K(\mu, I)$, suppose μ_0 is a point on the boundary $\mathscr{B}$ in Figure 2. The attained significance level at μ_0, using X_0 as the test statistic, is

$$\widehat{\alpha}(\mu_0) = \text{prob}_{\mu 0}\{X_0 < x_0\} \doteq \Phi\big(x_0 - \widehat{z}_0(\mu_0)\big), \tag{8.10}$$

with $\widehat{z}_0(\mu_0) = \Phi^{-1}\{\text{prob}_{\mu_0}(y_0^* \in \mathscr{R}_0)\}$ as in (2.15), (2.21). If instead of a single point μ_0, we have a Bayesian distribution $\xi(\mu)$ on $\mathscr{B}$, then

$$\begin{aligned}\widehat{\alpha}(\xi) &\doteq \int_{\mathscr{B}} \Phi(x_0 - \widehat{z}_0(\mu_0))\xi(\mu_0)\,d\mu_0 \\ &\doteq \Phi(x_0) - \varphi(x_0)\int_{\mathscr{B}} \widehat{z}_0(\mu_0)\xi(\mu_0)\,d\mu_0.\end{aligned} \tag{8.11}$$

In other words, we average the $\widehat{z}_0$ values as in Table 5. [Rather than, say, averaging $\Phi(\widehat{z}_0)$.] Of course, $\xi(\mu_0)$ is unavailable, but the empirical distribution of the boundary points obtained from the first-level bootstraps gives a reasonable estimate of ξ, thought of as the posterior distribution on $\mathscr{B}$ having observed y.

REMARK G. In most cases, we would not have a formula such as (7.15) to furnish $\dot{t}_0$ for the computation of $\widehat{a}$. Here is a metric-free algorithm for finding $\dot{t}_0$:

1. Select an orthogonal basis $\{u_1, u_2, \ldots, u_{n-1}\}$ for the $(n-1)$-dimensional subspace orthogonal to $\widehat{\pi}_0 - \widehat{\pi}$.
2. Let $v_i = \widehat{\pi}_0 + \varepsilon u_i$ for $i = 1, 2, \ldots, n-1$, with ε a small value such as .001.
3. Use a one-dimensional search algorithm like that at (7.10) to find $v_i^0 = wv_i + (1-w)\widehat{\pi}$ on the boundary $\mathscr{B}$, for $i = 1, 2, \ldots, n-1$.
4. Finally, take $\dot{t}_0$ to be the vector orthogonal to $\mathrm{v}_1^0, \mathrm{v}_2^0, \ldots, \mathrm{v}_{n-1}^0$, and having positive inner product with $\widehat{\pi}_0 - \widehat{\pi}$.

APPENDIX

The results of Section 2 are based on the Appendix in Efron (1985), and in particular on this lemma: write $z \sim N_K(0, I)$ as $(z_1, z_{(2)})$, where $z_{(2)} = (z_{(2)}, z_{(3)}, \ldots, z_{(K)})'$ and let

$$Q(z) = \left[1 + A(z_{(2)})\right] z_1 + B(z_{(2)}), \tag{A.1}$$

$A(z_{(2)})$ and $B(z_{(2)})$ being $O_p(n^{-1/2})$. Then the first four cumulants of Q are

$$Q \sim \left\{E(B), \operatorname{var}(B) + E(1+A)^2, 6(\operatorname{cov}(A, B)), 12(\operatorname{var}(A))\right\} + O(n^{-3/2}). \tag{A.2}$$

In addition, it was shown that our Figure 2 can always be transformed so that

$$\widehat{\mu}_0 = 0, \qquad \widehat{\mu} = (-x_0, 0)', \tag{A.3}$$

and where the boundary $\mathscr{B}$ is approximated near 0 by

$$\mathscr{B} = \left\{z \colon z_1 = z'_{(2)} \widehat{\mathbf{d}} z_{(2)}\right\}. \tag{A.4}$$

Here $\widehat{\mathbf{d}} = \mathbf{d}(\widehat{\mu}_0) = \mathbf{d}(0)$ as in (2.15). Using approximation (A.4) causes errors only at the third level $O_p(n^{-3/2})$. Note: the sign convention in (A.4) is opposite to that in Efron (1985), in order to accommodate the boundary-oriented theory of this paper.

PROOF OF (2.19) AND (2.23). Define

$$Q(z) = z_1 - z'_{(2)} \widehat{\mathbf{d}} z_{(2)}, \tag{A.5}$$

with $\widehat{\mathbf{d}}$ considered as fixed at its observed value. Applying (A.1), (A.2) with $A = 0$, $B = -z'_{(2)} \widehat{\mathbf{d}} z_{(2)}$ gives

$$Q(z) \sim \left\{-\widehat{d}_1, 1 + 2\widehat{d}_2, 0, 0\right\} + O_p(n^{-3/2}), \tag{A.6}$$

using definitions (2.15). In other words, $Q(z)$ is approximately normal,

$$Q(z) \sim N\big(-\widehat{d}_1, (1+\widehat{d}_2)^2\big) + O_p(n^{-3/2}). \tag{A.7}$$

Result (2.19) concerns first-order bootstrap samples $y^* \sim N_K(\widehat{\mu}, I)$. Using (A.4)–(A.5) gives

$$\tilde{\alpha} = \mathrm{prob}_{\widehat{\mu}_0}\{y^* \in \mathscr{R}_0\} \doteq \mathrm{prob}_{\widehat{\mu}_0}\big\{Q(y^*) < 0\big\}. \tag{A.8}$$

Notice that

$$z = y^* + (x_0, 0) \sim N_K(0, I), \tag{A.9}$$

(A.3), so we have

$$\begin{aligned} \tilde{\alpha} &\doteq \mathrm{prob}_{\widehat{\mu}_0}\big\{y_1^* - y_{(2)}^* \widehat{\mathbf{d}} y_{(2)}^* < 0\big\} \\ &= \mathrm{prob}\big\{z_1 - x_0 - z'_{(2)} \widehat{\mathbf{d}} z_{(2)} < 0\big\} \\ &= \mathrm{prob}\{Q(z) < x_0\} \\ &\doteq \Phi\left(\frac{x_0 + \widehat{d}_1}{1 + \widehat{d}_2}\right) \end{aligned} \tag{A.10}$$

by (A.7), which verifies (2.19).

Lemma 2 involves second-order bootstrap samples $y_0^* \sim N_K(\widehat{\mu}_0, I)$. Since $\widehat{\mu}_0 = 0$ (A.3), we can take $z = y_0^* \sim N_K(0, I)$, apply (A.7) and get

$$\Phi(\widehat{z}_0) = \mathrm{prob}_{\widehat{\mu}0}\{y_0^* \in \mathscr{R}_0\} \doteq \mathrm{prob}\{Q(z) < 0) \doteq \Phi\left(\frac{\widehat{d}_1}{1 + \widehat{d}_2}\right). \tag{A.11}$$

REFERENCES

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.

DICICCIO, T. and EFRON, B. (1992). More accurate confidence limits in exponential families. *Biometrika* **79** 231–245.

EFRON, B. (1982). The jackknife, the bootstrap and other resampling plans. SIAM, Philadelphia.

EFRON, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72** 45–58.

EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82** 171–200.

EFRON, B. and FELDMAN, D. (1991). Compliance as an explanatory variable in clinical trials, *J. Amer. Statist. Assoc.* **86** 9–26.

EFRON, B., HALLORAN, E. and HOLMES, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Nat. Acad. Sci. U.S.A.* **93** 13429–13434.

EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.

EFRON, B. and TIBSHIRANI, R. (1996). The problem of regions. Stanford Technical Report 192. Available at ftp://utstat.toronto.edu/pub/tibs/regions.ps.

FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 783–791.

HALL, P. (1992). *The Bootstrap and Edgeworld Expansion*. Springer, New York.

IZENMAN, A. and SOMMER, L. (1988). Philatelic mixtures and multimodal densities. *J. Amer. Statist. Assoc.* **83** 941–953.

LOH, W. Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82** 152–162.

MALLOWS, C. (1973). Some comments on cp. *Technometrics* **15** 661–675.

RUBIN, D. (1981). The bayesian bootstrap. *Ann. Statist.* **9** 130–134.

TIBSHIRANI, R. (1989). Non-informative priors for one parameter of many. *Biometrika* **76** 604–608.

WELCH, B. and PEERS, H. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **25** 318–329.

WENG, C. S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17** 705–710.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANDFORD, CALIFORNIA 94305
E-MAIL: brad@stat.stanford.edu

DEPARTMENT OF PUBLIC HEALTH SCIENCES
AND DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
TORONTO M5S 1A8
CANADA
E-MAIL: tibs@utstat.toronto.edu