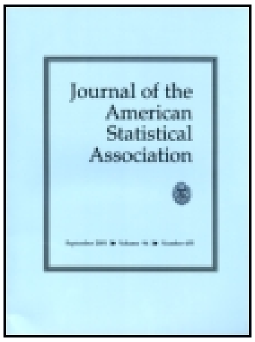

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Balanced Simultaneous Confidence Sets

Rudolf Beran

To cite this article: Rudolf Beran (1988) Balanced Simultaneous Confidence Sets, Journal of the American Statistical Association, 83:403, 679-686

To link to this article: https://doi.org/10.1080/01621459.1988.10478648

Balanced Simultaneous Confidence Sets

RUDOLF BERAN*

Suppose that $T(\theta) = \{T_u(\theta)\}$ is a family of parametric functions indexed by the variable u. For each u, an approximate confidence set $C_{n,u}$ for $T_u(\theta)$ may be obtained by referring a function of $T_u(\theta)$ and the sample to an estimated quantile $d_{n,u}$ of that function's sampling distribution. This approach is sometimes called the pivotal method for constructing a confidence set, even when it is not based on a true pivot. A simultaneous confidence set C_n for $T(\theta)$ is then obtained by simultaneously asserting the individual confidence sets $\{C_{n,u}\}$. The problem is to choose the critical values for the $\{C_{n,u}\}$ in such a way that the overall coverage probability of C_n is correct and the coverage probabilities of the individual confidence sets $\{C_{n,u}\}$ are equal. The second property is termed balance. It means that the simultaneous confidence set C_n treats each constituent confidence statement $C_{n,u}$ fairly. Aside from a few special cases, the problem just described is too difficult for analytical approaches, whether exact or asymptotic in nature. A new bootstrap method described in this article, however, yields a practical solution to the problem that is asymptotically valid under general conditions. The new method recovers, as special cases, the Tukey and Scheffé simultaneous confidence intervals for the normal linear model. In addition, it handles distributionally harder problems, such as simultaneous confidence cones for the eigenvectors of an unknown covariance matrix or confidence bands for a linear predictor, in either parametric or nonparametric settings.

KEY WORDS: Balance; Bootstrap estimate; Prepivoting; Stochastic procedure.

1. INTRODUCTION

Tukey (1953) and Scheffé (1953) developed simultaneous confidence intervals for linear functions of regression parameters in the normal linear model. The requirement of tractable distribution theory strongly shaped and limited these classical confidence procedures. To what extent does the availability of bootstrap estimates for unknown distributions enlarge the possibilities for simultaneous inference? This article argues that the answer is *substantially*.

I consider simultaneous confidence statements in the following setting. Let x_n be a sample of size n whose distribution $P_{\theta,n}$ belongs to a specified parametric family of distributions. The parameter value θ is unknown but is assumed to lie in a parameter space Θ, which can be either finite or infinite dimensional. Let T be a parametric function, defined on Θ, which has *components* labeled by an index set U. In other words, $T(\theta) = \{T_u(\theta) : u \in U\}$, where the parametric function $T_u(\theta)$ is the uth component. The index set U can be finite (see Example 2) or infinite [Ex. 1(b)].

I wish to construct a simultaneous confidence set for $T(\theta)$, that is, for the components $\{T_u(\theta)\}$ of $T(\theta)$. This simultaneous confidence set is to have overall coverage probability $1 - \alpha$ for $T(\theta)$. Moreover, it is to be *balanced* in the sense that the coverage probability for the confidence statement concerning the component $T_u(\theta)$ remains unchanged as u varies. This second requirement ensures that the simultaneous confidence set treats each constituent confidence statement fairly.

Example 1: Linear Combinations of Means. Suppose that $x_n = (X_1, \ldots, X_n)$ is a sample of iid random r-dimensional vectors, each having unknown distribution P. Assume that the mean vector μ_P and covariance matrix Σ_P of this distribution both exist. In this example, $\theta = P$, the parameter space is infinite dimensional, and $P_{\theta,n}$ is the joint distribution of x_n under the model just described.

Let U be any fixed set of nonnull $r \times 1$ vectors. Define

$$T_u(P) = u'\mu_P, \qquad u \in U. \tag{1.1}$$

Then $T(P) = \{T_u(P)\}$ is the set of all such linear combinations of means generated as u ranges over U. For instance: (a) U consists of all pairwise contrasts in r-dimensional space; that is, u belongs to U iff one component of u is 1, another component of u is -1, and the remaining $r - 2$ components are 0. (b) U is any nonnull subspace of r-dimensional space.

The index set U is finite in (a) but infinite in (b). In this setting, neither the Tukey nor the Scheffé methods yield balanced simultaneous confidence intervals for the linear combinations of means $\{T_u(P)\}$. Difficulties arise because the components of the sample mean vector may be heteroscedastic, dependent, and not normally distributed.

Example 2: Confidence Cones for Eigenvectors. Suppose that $x_n = (X_1, \ldots, X_n)$ is a sample of iid random r-dimensional column vectors, each having an $N(\mu, \Sigma)$ distribution. In this case, the unknown parameter θ is the pair (μ, Σ), the parameter space is finite-dimensional, and $P_{\theta,n}$ is the joint normal distribution of the sample. Suppose that the eigenvalues of Σ are distinct and strictly positive. Define $T_u(\theta)$ to be the uth eigenvector of Σ, corresponding to the uth largest eigenvalue and expressed as a column vector of unit length whose direction is determined by a preset rule. The set of all k orthonormal eigenvectors of Σ is then $T(\theta)$. The problem is to construct balanced simultaneous confidence cones for the eigenvectors of Σ.

This example also has a nonparametric formulation, in which the sample vectors are distributed as in Example 1 and the covariance matrix Σ_P has distinct nonzero eigenvalues. The complex asymptotic theory for sample eigen-

* Rudolf Beran is Professor, Department of Statistics, University of California, Berkeley, CA 94720. This research was supported by National Science Foundation Grant MCS84-03239.

Journal of the American Statistical Association
September 1988, Vol. 83, No. 403, Theory and Methods

vectors, even under the normal model, precludes analytical construction of the desired balanced simultaneous confidence cones.

In the general setting, consider first the construction of a confidence set $C_{n,u}$ for the component parametric function $T_u(\theta)$. The pivotal method runs as follows. Let $R_{n,u}(T_u(\theta)) = R_{n,u}(x_n, T_u(\theta))$ be a function of the sample and of $T_u(\theta)$. Let $d_{n,u}$ denote a consistent estimate for the $(1 - \beta)$th quantile of the distribution of $R_{n,u}$ under the model. Let $\mathbf{T}_u$ denote the set of all possible values of $T_u(\theta)$ as θ ranges over the parameter space. Then

$$C_{n,u} = \{t_u \in \mathbf{T}_u : R_{n,u}(t_u) \le d_{n,u}\} \tag{1.2}$$

is a confidence set with asymptotic coverage probability $1 - \beta$ for $T_u(\theta)$. When $R_{n,u}$ is not a true pivot, in the sense of classical statistics, we will call it the *root* of the confidence set $C_{n,u}$.

Example 1 (continued). Let $\overline{X}_n$ and S_n denote, respectively, the sample mean and sample covariance matrix. Recall that $T_u(P)$ is the linear combination of means $u'\mu_P$. Define

$$R_{n,u}(T_u(P)) = n^{1/2}|u'(\overline{X}_n - \mu_P)|/(u'S_nu)^{1/2} \tag{1.3}$$

for every value of u in U. Substituted into (1.2), this root generates a confidence interval for the linear combination $u'\mu_P$.

Example 2 (continued). Let $c_{n,u}$ denote the uth eigenvector of the sample covariance matrix. Define

$$R_{n,u}(T_u(P)) = n[1 - |c'_{n,u}T_u(\theta)|], \qquad 1 \le u \le r, \tag{1.4}$$

where $T_u(\theta)$ is now the uth eigenvector of Σ. Substituted into (1.2), this root generates a confidence cone for the uth eigenvector.

By simultaneously asserting the confidence sets $\{C_{n,u}\}$, we obtain a simultaneous confidence set C_n for the family of parametric functions $T(\theta) = \{T_u(\theta)\}$. More fully, let $\mathbf{T}$ denote the set of all possible values of $T(\theta)$ as θ varies over the parameter space. As in Example 2, constraining relationships among its components may restrict the range of $T(\theta)$. Every point t in this range can be written in component form $t = \{t_u : u \in U\}$, where t_u lies in the range of $T_u(\theta)$. In this notation,

$$C_n = \{t \in \mathbf{T} : t_u \in C_{n,u}\}. \tag{1.5}$$

Alternatively, because of (1.2),

$$C_n = \{t \in \mathbf{T} : R_{n,u}(t_u) \le d_{n,u}\}. \tag{1.6}$$

In Example 1, C_n becomes a set of simultaneous confidence intervals. In Example 2, C_n is a set of simultaneous confidence cones for the eigenvectors of interest.

How should we choose the critical values $\{d_{n,u}\}$ in (1.6) and (1.2) so that the coverage probability of C_n for $T(\theta)$ is $1 - \alpha$ and the component coverage probability of $C_{n,u}$ for $T_u(\theta)$ does not vary with u (balance)? Classical answers to this question, such as the methods of Tukey and of Scheffé, work only when the roots $\{R_{n,u}\}$ are true pivots, whose distribution does not depend on θ. This point is developed further in Section 3. The main result of this article, described in Section 2, is a bootstrap construction for the critical values that is asymptotically valid in general and exact in the classical situations.

2. THE BOOTSTRAP APPROACH

2.1 Choosing Critical Values

Let $H_{n,u}(\cdot, \theta)$ be the left-continuous cdf of the root $R_{n,u}$. Let $H_n(\cdot, \theta)$ be the left-continuous cdf of the transformed root $\sup\{H_{n,u}(R_{n,u}, \theta) : u \in U\}$. Suppose that $\hat{\theta}_n$ is a consistent estimate of the parameter θ. The natural, plug-in estimates of $H_{n,u}$ and H_n are, respectively, $\hat{H}_{n,u} = H_{n,u}(\cdot, \hat{\theta}_n)$ and $\hat{H}_n = H_n(\cdot, \hat{\theta}_n)$. Efron (1979) called such plug-in cdf estimates bootstrap estimates.

Let $\hat{H}_{n,u}^{-1}(t)$ and $\hat{H}_n^{-1}(t)$ denote the largest tth quantiles of $\hat{H}_{n,u}$ and $\hat{H}_n$, respectively. The proposed bootstrap version of simultaneous confidence set C_n is obtained by taking the critical values in (1.6) to be

$$d_{n,u} = \hat{H}_{n,u}^{-1}[\hat{H}_n^{-1}(1 - \alpha)]. \tag{2.1}$$

The calculation of $d_{n,u}$ involves two steps: (a) Find the largest $(1 - \alpha)$th quantile of $\hat{H}_n$ and call it c_n. (b) Find the largest c_nth quantile of $\hat{H}_{n,u}$. This is the critical value in (2.1). Note that it depends heavily on the assumed model $\{P_{\theta,n} : \theta \in \Theta\}$.

Under assumptions described in Section 4, the simultaneous bootstrap confidence set C_n has asymptotic coverage probability $1 - \alpha$ for $T(\theta)$ and is asymptotically balanced (Theorem 4.1). Moreover, a consistent estimate of the componentwise coverage probability that $C_{n,u}$ contains $T_u(\theta)$ is $\hat{H}_n^{-1}(1 - \alpha)$. Underlying these technical results is another representation for C_n,

$$C_n = \{t \in \mathbf{T} : \hat{H}_n[\sup_u \hat{H}_{n,u}\{R_{n,u}(t_u)\}] \le 1 - \alpha\}, \tag{2.2}$$

which is equivalent to (1.6) when the critical values are given by (2.1).

The key idea visible in (2.2) is the concept of prepivoting—the transformation of a root by its estimated cdf. Prepivoting takes the root $R_{n,u}$ into $\hat{H}_{n,u}(R_{n,u})$, whose asymptotic distribution is usually uniform (0, 1) for every choice of u. It is this property that ensures the asymptotic balance of the simultaneous confidence set C_n. Similarly, prepivoting takes the root $\sup_u \hat{H}_{n,u}(R_{n,u})$ into $\hat{H}_n[\sup_u \hat{H}_{n,u}(R_{n,u})]$, whose asymptotic distribution is again typically uniform (0, 1). This property ensures that the asymptotic coverage probability of C_n for $T(\theta)$ is $1 - \alpha$. An earlier, more delicate use of prepivoting—to reduce the coverage probability error of a confidence set—is discussed in Beran (1987).

2.2 Approximating C_n for Finite U

Except in special cases, such as those described in Section 3.1, closed-form expressions for $\hat{H}_{n,u}$ and $\hat{H}_n$ are not available. In the important case in which U is a finite set, the following Monte Carlo algorithm generates an effective stochastic approximation to the critical values $\{d_{n,u}\}$ and hence to the simultaneous confidence set C_n.

1. Given the data x_n, draw j_n conditionally independent bootstrap samples $(x^*_{n,1}, \ldots, x^*_{n,j_n})$, each of size n, from the fitted model $P_{\hat\theta_n,n}$. In practice, j_n should be at least 1,000 if the overall coverage probability $1 - \alpha$ is to be .90 or .95. For every u in the index set U, form $H'_{n,u}$, the left-continuous cdf of the bootstrapped roots $\{R^*_{n,u,j} = R_{n,u}(x^*_{n,j}, T_u(\hat\theta_n)) : 1 \le j \le j_n\}$. This cdf approximates $\hat H_{n,u}$.

2. For every value of j, let

$$s_{n,j} = \sup_u H'_{n,u}[R^*_{n,u,j}] = j_n^{-1} \sup_u [\operatorname{rank}(R^*_{n,u,j}) - 1], \quad (2.3)$$

the rank being calculated among the j_n possible values of its argument. Form H'_n, the left-continuous empirical cdf of the $\{s_{n,j}\}$. This cdf approximates $\hat H_n$.

3. Find c'_n, the largest $(1 - \alpha)$th quantile of H'_n. Then find $d'_{n,u}$, the largest c'_nth quantile of $H'_{n,u}$. The latter approximates the bootstrap critical value $d_{n,u}$ defined in (2.1).

Define the simultaneous confidence set C'_n by analogy with (1.6):

$$C'_n = \{t \in \mathbf{T} : R_{n,u}(t_u) \le d'_{n,u}\}. \quad (2.4)$$

This new confidence set is a widely applicable practical approximation to the theoretical bootstrap confidence set C_n defined in Section 2.1. Under conditions described in Section 4, the simultaneous confidence set C'_n also has asymptotic coverage probability $1 - \alpha$ for $T(\theta)$ and is asymptotically balanced, provided j_n tends to infinity with n (Theorem 4.2).

We observe that confidence set C'_n is a randomized procedure, in the sense of decision theory, since it depends on the artificial bootstrap samples $y_n = (x^*_{n,1}, \ldots, x^*_{n,j_n})$ as well as on the original data x_n. The distribution and coverage probabilities of C'_n are thus determined by the joint distribution of (x_n, y_n). The effects of the randomization on C'_n can be made very small, however, simply by drawing sufficiently many bootstrap samples. Confidence set C'_n exemplifies an emerging class of statistical methods, called stochastic procedures, which make controlled use of randomization to resolve computational problems (compare Beran and Millar 1987).

Example 2 (continued). Mardia, Kent, and Bibby (1979) reported test scores for $n = 88$ college students, each of whom took two closed-book and three open-book tests ($r = 5$). Which averages of the test scores best discriminate among students? How trustworthy are the averages suggested by the estimated principal components? We can address these questions by constructing simultaneous confidence cones for the eigenvectors of the covariance matrix Σ_p, using the previously described nonparametric model for the data. Less formal bootstrap analyses of the data have appeared in Diaconis and Efron (1983) and Beran and Srivastava (1985).

The sample eigenvectors, in order of decreasing sample eigenvalues, are as follows:

$$\begin{aligned} c_{n,1} &= (.505, .368, .346, .451, .535)' \\ c_{n,2} &= (.749, .207, -.076, -.301, -.548)' \\ c_{n,3} &= (-.300, .416, .145, .597, -.600)' \\ c_{n,4} &= (-.296, .783, .003, -.518, .176)' \\ c_{n,5} &= (.079, .189, -.924, .256, .151)'. \end{aligned} \quad (2.5)$$

The asymptotically balanced simultaneous confidence cones C'_n, based on the roots $\{R_{n,u} : 1 \le u \le 5\}$ of (1.4), are defined by (2.4). Here $\theta = P$; $U = \{1, 2, \ldots, 5\}$; $\hat\theta_n$ is the empirical measure of the sample, $\hat P_n$; and $\mathbf{T}$ is the set of all orthonormal bases for five-dimensional space. The bootstrap samples are drawn from $\hat P_n$. Consider as well another simultaneous confidence set for the eigenvectors, obtained by bootstrapping nonparametrically the root $\sup_u R_{n,u}$. This alternative approach yields the simultaneous confidence cones

$$D'_n = \{t \in \mathbf{T} : R_{n,u}(t_u) \le e'_n, 1 \le u \le 5\}, \quad (2.6)$$

whose common critical value e'_n is determined as follows: Let K'_n be the left-continuous empirical cdf of the values $\{\sup_u R_{n,u}(x^*_{n,j}, T_u(\hat\theta_n)) : 1 \le j \le j_n\}$. Define e'_n to be the largest $(1 - \alpha)$th quantile of K'_n.

Both simultaneous confidence cones C'_n and D'_n have the same form, although the critical values $\{d'_{n,u}\}$ and e'_n determining the confidence cones for the individual eigenvectors differ. Both simultaneous confidence sets have asymptotic coverage probability $1 - \alpha$ for the set of eigenvectors $\{T_u(\theta) : 1 \le u \le 5\}$, provided that $\lim_{n\to\infty} j_n = \infty$ and the regularity conditions of Section 4 are met; however, C'_n is asymptotically balanced and D'_n is not.

The difference is dramatically visible in Table 1, which exhibits C'_n and D'_n for the test-score data, using $j_n = 1{,}000$ nonparametric bootstrap samples and setting intended overall coverage probability $1 - \alpha$ to be .95. The coverage probability of each confidence cone in C'_n is estimated consistently by the largest $(1 - \alpha)$th quantile of H'_n. On the other hand, the coverage probability of the uth confidence cone in D'_n is estimated consistently by $H'_{n,u}(e'_n)$. Rows 2 and 4 of Table 1 report the values of these coverage

Table 1. Comparison on the Test-Score Data, at Asymptotic Level .95, of Two Families of Simultaneous Confidence Cones for the Eigenvectors of the Unknown Covariance Matrix

	Eigenvector number u				
	1	2	3	4	5
C'_n					
Critical value $d'_{n,u}$	1.88	20.8	82.3	81.3	5.45
Estimated marginal coverage probability	.988	.988	.988	.988	.988
D'_n					
Critical value e'_n	69.5	69.5	69.5	69.5	69.5
Estimated coverage probability	1.000	.998	.955	.955	1.000

probabilities. Simultaneous confidence set D_n' is useless in analyzing the test-score data because it generates excessively wide confidence cones for eigenvectors 1, 2, and 5. Balancing the simultaneous confidence set by construction of C_n' makes the individual confidence cones equally trustworthy.

The estimated marginal coverage probability of each confidence cone in C_n' is slightly less than .99. This is not surprising, since the Bonferroni inequality ensures that five confidence cones, each having marginal coverage probability .99, will have simultaneous coverage probability of at least .95.

Consider the implications of C_n' for the test-score data. From (1.4), (2.4), (2.5), and Table 1, it is evident that the constituent confidence cone for the first eigenvector $T_1(\theta)$ contains the vector $5^{1/2}$ $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})'$, which corresponds to the average test score for each student. On general grounds, a natural candidate for $T_2(\theta)$ is the vector $(\frac{6}{5})^{1/2}$ $(\frac{1}{2}, \frac{1}{2}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3})'$; the corresponding principal component is the difference between the average closed-book score and the average open-book score for each student. In fact, this vector lies well within the confidence cone for $T_2(\theta)$. The following conclusions can be drawn: The average score is a rough summary of a student's performance in the five tests. Finer distinctions among students can be made by calculating the difference between the average closed-book score and the average open-book score for each student.

2.3 Approximating C_n for Infinite U

When the number of elements in U is very large or infinite, it may be hard to compute the supremum over U in step 2 of the algorithm for confidence set C_n' (see Sec. 2.2). A stochastic approximation to the supremum is often available. Suppose that μ is a probability distribution on U that has full support. Let $(u_1, \ldots, u_{k_n})$ be iid random elements of U, each distributed according to μ. The $\{u_k\}$ are also independent of the data x_n and of the bootstrap samples $\{x_{n,j}^*\}$ described in Section 2.2. Replace steps 2 and 3 in that section by the following steps:

2$'$. For every value of j, let

$$s_{n,j}' = j_n^{-1} \max_{1\le k\le k_n} [\operatorname{rank}(R_{n,u_k,j}^*) - 1]. \tag{2.7}$$

Form H_n'', the left-continuous empirical cdf of the $\{s_{n,j}'\}$. This cdf is a further approximation to $\hat{H}_n$.

3$'$. Find the critical values $d_{n,u}''$ as in step 3, using the cdf H_n'' in place of the cdf H_n'.

The associated further approximation to confidence set C_n is then

$$C_n'' = \{t \in \mathbf{T} : R_{n,u}(t_u) \le d_{n,u}''\}. \tag{2.8}$$

Note that the distribution of C_n'' depends on the joint distribution of the original sample x_n, the bootstrap samples $\{x_{n,j}^* : 1 \le j \le j_n\}$, and the search sample $\{u_k : 1 \le k \le k_n\}$. Under circumstances described in Section 4, if both j_n and k_n tend to infinity with n, then confidence set C_n'' is also asymptotically balanced and has asymptotic coverage probability $1 - \alpha$ for $T(\theta)$ (Theorem 4.2). Maximizing in (2.7) over a randomly chosen set of $\{u_k\}$ rather than over a deterministic grid thus has the following advantage: The rate at which k_n, the number of grid points, must tend to infinity to achieve the desired asymptotic behavior for C_n'' does *not* depend on the dimension of U.

2.4 Further Examples

The examples described in this section are intended to illustrate further the scope of simultaneous confidence set C_n and its stochastic approximations C_n' and C_n''. In all cases, the asymptotic theory of Section 4 is applicable. Many other examples can be treated similarly.

Example 3: Confidence Intervals for Eigenvalues. In the setting of Example 2, let $T_u(\theta)$ denote the uth largest eigenvalue of the covariance matrix $\Sigma(\theta)$. The eigenvalues are assumed to be distinct and strictly positive. The parameter θ is either (μ, Σ) in the $N(\mu, \Sigma)$ model for the data or P in the nonparametric formulation. The function $T(\theta)$ is the $r \times 1$ vector of the ordered eigenvalues; its range is $\mathbf{T} = \{(z_1, \ldots, z_r) \in R^r : z_1 > \cdots > z_r > 0\}$.

Let $\mathbf{l}_{n,1} \ge \cdots \ge \mathbf{l}_{n,r}$ denote the ordered eigenvalues of the sample covariance matrix S_n. Since the eigenvalues $\{T_u(\theta)\}$ are scale parameters, a natural family of roots on which to base simultaneous confidence intervals is

$$R_{n,u}(T(\theta)) = n^{1/2}|\log[\mathbf{l}_{n,u}/T_u(\theta)]|, \qquad 1 \le u \le r. \tag{2.9}$$

The estimate of θ is $(\overline{X}_n, S_n)$ in the normal formulation or the empirical measure of the sample, $\hat{P}_n$, in the nonparametric version.

Example 4: Confidence Band for a Linear Predictor. Suppose that x is a random $r \times 1$ vector whose distribution P has finite mean vector μ_P and nonsingular covariance matrix Σ_P. Partition $x' = (y \mid v')$ so that y is a random scalar and v is a random $(r - 1) \times 1$ vector. Correspondingly, partition

$$\mu_P = \begin{pmatrix} m_1(\mu_P) \\ m_2(\mu_P) \end{pmatrix}, \qquad \Sigma_P = \begin{pmatrix} \sigma_{P,11} & \sigma_{P,12} \\ \sigma_{P,21} & \Sigma_{P,22} \end{pmatrix}, \tag{2.10}$$

so that $m_1(\mu_P)$ and $\sigma_{P,11}$ are scalar. The best (minimum mean squared error) linear predictor of y given v is

$$m_1(\mu_P) + b(\Sigma_P)(v - m_2(\mu_P)), \tag{2.11}$$

where $b(\Sigma_P) = \sigma_{P,12}\Sigma_{P,22}^{-1}$. The problem is to construct a balanced confidence band for the values of this linear predictor as v varies over a designated subset V of R^{r-1}. Usually, V is a bounded set.

Define the $r \times 1$ vectors

$$g(\mu_P, \Sigma_P) = \begin{pmatrix} m_1(\mu_P) - b(\Sigma_P)m_2(\mu_P) \\ b(\Sigma_P) \end{pmatrix}, \qquad u = \begin{pmatrix} 1 \\ v \end{pmatrix}. \tag{2.12}$$

The range of u is $U = \{u : u' = (1 \mid v') \text{ and } v \in V\}$. The best linear predictor (2.18) can be rewritten as

$$T_u(P) = u'g(\mu_P, \Sigma_P), \qquad u \in U. \tag{2.13}$$

Suppose that $x_n = (X_1, \ldots, X_n)$ are n iid random vectors, each distributed according to P. Let $\overline{X}_n$ and S_n denote the sample mean vector and the sample covariance matrix, respectively. The least squares estimate of $T_u(P)$ is then $u'g(\overline{X}_n, S_n)$. A family of roots that generates a simultaneous confidence band for the linear predictor function $T(P) = \{T_u(P) : u \in U\}$ is

$$R_{n,u}(T_u(P)) = n^{1/2}|u'g(\overline{X}_n, S_n) - T_u(P)|. \quad (2.14)$$

The estimate of P is $\hat{P}_n$, the empirical measure of the sample x_n.

3. EXTENSIONS

3.1 Relation to the Tukey–Scheffé Approach

The Scheffé (1953) and Tukey (1953) simultaneous confidence intervals for certain linear functions of regression parameters in the normal linear model can be rederived as special cases of confidence set C_n.

Scheffé's Method for Linear Functions. Let x_n have an $N(C\beta, \sigma^2 I)$ distribution, where β is r-dimensional, the regression matrix C has rank r, and I is the identity matrix. The unknown parameter $\theta = (\beta, \sigma^2)$ is estimated by $(\hat{\beta}_n, \hat{\sigma}_n^2)$ from the usual least squares theory. Suppose that U is a non-empty subspace of dimension q in r-dimensional space. The problem is to construct balanced simultaneous confidence intervals for the linear combinations $T_u(\theta) = u'\beta$ that are generated as u ranges over the subspace U.

As the roots for C_n, take

$$R_{n,u}(T_u(\theta)) = |u'(\hat{\beta}_n - \beta)|/\hat{\sigma}_{n,u}, \quad (3.1)$$

where $\hat{\sigma}_{n,u}^2 = u'(C'C)^{-1}u\hat{\sigma}_n^2$. Then $H_{n,u}(\cdot, \theta)$ is the cdf of a t distribution with $n - r$ degrees of function, say Ψ_{n-r}. Consequently, $H_n(\cdot, \theta)$ is the cdf of $\Psi_{n-r}(\sup_u |R_{n,u}|)$. By a well-known argument, $H_n(\cdot, \theta)$ is the cdf of $\Psi_{n-r}[(qW)^{1/2}]$, where W has an F distribution with q and $n - r$ df (Miller 1966, chap. 2, sec. 2). Both $R_{n,u}$ and $\sup_u \hat{H}_{n,u}(R_{n,u})$ are true pivots here. Moreover, C_n reduces here to Scheffé's simultaneous confidence intervals

$$C_n = \{u'\beta : |u'(\hat{\beta}_n - \beta)|^2 \le q\sigma_n^2 F_{q,n-r}(1 - \alpha) \text{ for every } u \in U\}, \quad (3.2)$$

where $F_{q,n-r}(1 - \alpha)$ is the $(1 - \alpha)$th quantile of the F distribution with q and $n - r$ df.

Tukey's Method for Pairwise Contrasts. Specialize the foregoing normal linear model to the one-way layout. We observe r independent samples of size m; the jth sample is drawn from an $N(\beta_j, \sigma^2)$ distribution; $\beta = (\beta_1, \ldots, \beta_r)'$; and the overall sample size is $n = mr$. The unknown parameter $\theta = (\beta, \sigma^2)$ is estimated by $(\hat{\beta}_n, \hat{\sigma}_n^2)$ from the usual least squares theory. Suppose that U is the set of all pairwise contrasts in r-dimensional space, already defined in Example 1 of Section 1. The problem is to construct simultaneous confidence intervals for the pairwise differences of means $T_u(\theta) = u'\beta$ obtained as u ranges over this choice of U.

Define the roots $\{R_{n,u}\}$ by

$$R_{n,u}(T_u(\theta)) = m^{1/2}|u'(\hat{\beta}_n - \beta)|/(2^{1/2}\hat{\sigma}_n). \quad (3.3)$$

Then $H_{n,u}(\cdot, \theta)$ is again the t-distribution cdf Ψ_{n-r}; however, $H_n(\cdot, \theta)$ is the cdf of $\Psi_{n-r}(2^{-1/2}R)$, where R has the studentized range distribution with parameters r and $n - r$ (Miller 1966, chap. 2, sec. 1). Now C_n reduces to Tukey's (1953) simultaneous confidence intervals for the pairwise differences in means based on the studentized range distribution cited previously.

3.2 Errors in Coverage Probability

The simultaneous confidence set C_n defined in (1.5) and (1.2) can exhibit two types of coverage probability error, as follows:

Type 1. The difference between the overall coverage probability $P_{\theta,n}[C_n \ni T(\theta)]$ and $1 - \alpha$ is nonzero.

Type 2. Inequalities occur among the marginal coverage probabilities $\{P_{\theta,n}[C_{n,u} \ni T_u(\theta)]\}$ as u varies. In other words, C_n lacks balance.

An ideal situation arises when the roots $\{R_{n,u}\}$ and $\sup_u \hat{H}_{n,u}(R_{n,u})$ are true pivots with continuous distributions and the critical values of C_n are constructed by the bootstrap method (2.1). Then, C_n has no coverage probability errors of either Type 1 or Type 2. The Tukey and Scheffé methods, in the context of Section 3.1, illustrate this rare, highly favorable case. More generally, Theorem 4.1 gives theoretical conditions under which the coverage probability errors of C_n are asymptotically negligible, as sample size n increases. Several factors determine the rate at which these errors converge to 0: the distribution of the sample, the choice of the roots $\{R_{n,u}\}$, the method used to construct the critical values of C_n (if not this article's bootstrap method), and the true value of θ.

Several authors have studied coverage probability errors of bootstrap and competing confidence sets in cases in which asymptotic expansions are available (Abramovitch and Singh 1985; Beran 1987; Efron 1987; Hall 1986). A key finding in this: Properly constructed bootstrap critical values for the pivotal method are *at least* as good, in terms of asymptotic order of coverage probability error, as critical values for the pivotal method obtained by any other asymptotic method. In the case of C_n, this conclusion refers to the overall coverage probability error. The rate at which the lack of balance in C_n diminishes as n increases is unknown.

4. ASYMPTOTIC COVERAGE PROBABILITIES

This section presents two theorems. The first theorem gives the asymptotic coverage probability of confidence set C_n and its component confidence sets $\{C_{n,u}\}$. The second theorem establishes analogous results for the stochastic approximations C_n' and C_n''. The notation is that of Section 2. Proofs of the theorems follow in Section 5. The theorem assumptions are satisfied by Examples 1–4 of Sections 1 and 2. Details are available from me.

4.1 Confidence Set C_n

Suppose that Θ is an open subset of a metric space with metric d and U is a metric space with metric m. The choice of d and m will depend on the application of the theory. Let $C(U)$ denote the set of all continuous bounded functions on U, made metric by supremum norm. Assume that $C(U)$ is topologically complete and the processes $R_n(x_n, \theta) = \{R_{n,u}(x_n, T_u(\theta)) : u \in U\}$ $(n \geq 1)$ have their sample paths in $C(U)$. Boundedness of the sample paths can always be ensured by making a continuous, bounded, and strictly monotone transformation of the $\{R_{n,u}\}$. This transformation does not affect confidence set C_n, defined by (1.6) and (2.1).

Introduce the following assumptions for every $\theta \in \Theta$:

Assumption 1. If $\lim_{n\to\infty} d(\theta_n, \theta) = 0$, the estimates $\{\hat{\theta}_n\}$ converge to θ in $P_{\theta_n,n}$ probability.

Assumption 2. There exists a process $R(\theta) = \{R_u(\theta) : u \in U\}$ with sample paths in $C(U)$ such that if $\lim_{n\to\infty} d(\theta_n, \theta) = 0$, then $\mathbf{L}[R_n(x_n, \theta_n) \mid P_{\theta_n,n}]$ converges weakly in $C(U)$ to $\mathbf{L}[R(\theta)]$. The support of $\mathbf{L}[R(\theta)]$ is separable.

Assumption 3. Let $H_u(\cdot, \theta)$ be the cdf of $R_u(\theta)$ $(u \in U)$. The family of cdf's $\{H_u(\cdot, \theta) : u \in U\}$ is equicontinuous in the omitted argument: For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_{u\in U} |H_u(x, \theta) - H_u(y, \theta)| < \varepsilon$ whenever $|x - y| < \delta$.

Assumption 4. Let $H(\cdot, \theta)$ be the cdf of $\sup_{u\in U} H_u[R_u(\theta), \theta]$. The cdf $H(x, \theta)$ is continuous and strictly monotone in x.

Theorem 4.1. Suppose that Assumptions 1–4 hold, $\alpha \in (0, 1)$, and $\lim_{n\to\infty} \theta_n = \theta$. Then

$$\lim_{n\to\infty} P_{\theta_n,n}[T(\theta_n) \in C_n] = 1 - \alpha \tag{4.1}$$

and

$$\lim_{n\to\infty} \sup_{u\in U} |P_{\theta_n,n}[T_u(\theta_n) \in C_{n,u}] - H^{-1}(1 - \alpha, \theta)| = 0. \tag{4.2}$$

Moreover, $\hat{H}_n^{-1}(1 - \alpha)$ converges in $P_{\theta_n,n}$ probability to $H^{-1}(1 - \alpha, \theta)$.

The theorem gives circumstances under which the following hold: The simultaneous confidence set C_n for $T(\theta)$ has asymptotic coverage probability $1 - \alpha$, uniformly over compact subsets of Θ; the component confidence set $C_{n,u}$ for $T_u(\theta)$ has asymptotic coverage probability $H^{-1}(1 - \alpha, \theta)$, uniformly over U and over compact subsets of Θ; and the natural estimate $\hat{H}_n^{-1}(1 - \alpha)$ for the coverage probability of $C_{n,u}$ is consistent, uniformly over compacts in Θ. The uniformity of these convergences enhances their trustworthiness as approximations to the large-sample behavior of C_n, $C_{n,u}$, and $\hat{H}_n^{-1}(1 - \alpha)$.

4.2 Confidence Sets C_n' and C_n''

Analogs of Theorem 4.1 for these two stochastic approximations to C_n are the topics of this section. Confidence set C_n', defined in (2.4), is a function of the original data x_n and the j_n bootstrap samples $y_n = (x_{n,1}^*, \ldots, x_{n,j_n}^*)$. Given x_n, the $\{x_{n,j}^*\}$ are independent, each distributed according to $P_{\hat{\theta}_n,n}$. Let $Q_{\theta,n}$ denote the joint distribution of (x_n, y_n) when x_n has distribution $P_{\theta,n}$. Formally,

$$Q_{\theta,n}(A) = \int_A P^{j_n}_{\hat{\theta}_n(x),n}(dy) P_{\theta,n}(dx) \tag{4.3}$$

for every measurable subset A in the range of (x_n, y_n).

Let μ be a probability measure on U, endowed with its Borel σ algebra, and let $v_n = (u_1, \ldots, u_{k_n})$ be a sample of k_n iid random elements of U, each distributed according to μ. Construct v_n and (x_n, y_n) to be independent. Let $Q_{\theta,n} \times \mu^{k_n}$ denote the joint distribution of (x_n, y_n, v_n) when x_n has distribution $P_{\theta,n}$. This distribution determines the behavior of confidence set C_n'', defined in (2.8).

Theorem 4.2. Suppose that Assumptions 1–4 hold, $\alpha \in (0, 1)$, $\lim_{n\to\infty} \theta_n = \theta$, and $\lim_{n\to\infty} j_n = \infty$. Then

$$\lim_{n\to\infty} Q_{\theta_n,n}[T(\theta_n) \in C_n'] = 1 - \alpha \tag{4.4}$$

and

$$\lim_{n\to\infty} \sup_{u\in U} |Q_{\theta_n,n}[T_u(\theta_n) \in C_{n,u}'] - H^{-1}(1 - \alpha, \theta)| = 0. \tag{4.5}$$

Moreover, $(H_n')^{-1}(1 - \alpha)$ converges in $Q_{\theta_n,n}$ probability to $H^{-1}(1 - \alpha, \theta)$.

Suppose further that $\lim_{n\to\infty} k_n = \infty$ and μ gives positive probability to every open subset of U. Then the foregoing conclusions remain valid if C_n', $C_{n,u}'$, H_n', and $Q_{\theta_n,n}$ are simultaneously replaced by C_n'', $C_{n,u}''$, H_n'', and $Q_{\theta_n,n} \times \mu^{k_n}$, respectively.

This result, the interpretation of which strictly parallels that given previously for Theorem 4.1, justifies the use of confidence set C_n' or C_n'' as a computational approximation to C_n. The analysis in Theorem 4.2 takes into account the finiteness of j_n (the number of bootstrap samples) and k_n (the size of the search sample over U). Note that the rate at which j_n or k_n must tend to infinity does not depend on the dimensions of Θ or U.

5. THEOREM PROOFS

The proofs for Theorems 4.1 and 4.2 rest on a more abstract proposition. As in Section 4, let U be a metric space and let $C(U)$ be the set of all continuous bounded functions on U, made metric by supremum norm $\|\cdot\|$ and endowed with its Borel σ algebra. Assume that $C(U)$ is topologically complete. For every probability measure ν on $C(U)$, define the cdf's

$$F_u(x, \nu) = \nu\{z \in C(U) : z(u) < x\}, \tag{5.1}$$

for every $u \in U$, and

$$F(x, \nu) = \nu\{z \in C(U) : \sup_{u\in U} F_u[z(u), \nu] < x\}. \tag{5.2}$$

In addition, for every probability measure μ on U, endowed with its Borel σ algebra, define the cdf

$$G(x, \nu, \mu) = \nu\{z \in C(U) : \operatorname{ess\,sup}_\mu F_u[z(u), \nu] < x\}. \tag{5.3}$$

Here the ess sup notation means essential supremum over U with respect to the measure μ.

For every $n \geq 1$, let $\hat{\mu}_n$ and $\hat{\nu}_n$ be random probability measures whose possible values are probability measures on U and $C(U)$, respectively, and let Y_n be a process whose sample paths belong to $C(U)$. The probability space on which $\hat{\mu}_n$, $\hat{\nu}_n$, and Y_n are defined is endowed with the probability measure Q_n. Let ρ denote Prohorov metric. Introduce the following assumptions:

Assumption 5. There exists a probability measure ν on $C(U)$ such that $\lim_{n\to\infty} \rho[\mathbf{L}(Y_n \mid Q_n), \nu] = 0$ and $\rho(\hat{\nu}_n, \nu)$ converges to 0 in Q_n probability.

Assumption 6. The family of cdf's $\{F_u(\cdot, \nu) : u \in U\}$ is equicontinuous in the omitted argument. The cdf $F(\cdot, \nu)$ is continuous and strictly monotone in the omitted argument.

Assumption 7. There exists a probability measure μ that gives positive mass to every open subset of U and is such that $\rho(\hat{\mu}_n, \mu)$ converges to 0 in Q_n probability.

Proposition 5.1. Suppose that Assumptions 5 and 6 hold. Then (a) $\mathbf{L}[F\{\sup_{u\in U} F_u[Y_n(u), \hat{\nu}_n], \hat{\nu}_n\} \mid Q_n]$ converges weakly to a uniform distribution on (0, 1); (b) for every sequence $\{u_n \in U\}$, $\mathbf{L}[F\{F_{u_n}[Y_n(u_n), \hat{\nu}_n], \hat{\nu}_n\} \mid Q_n]$ converges weakly to the distribution on (0, 1) with cdf $F^{-1}(\cdot, \nu)$; and (c) for every $t \in (0, 1)$, $F^{-1}(t, \hat{\nu}_n)$ converges to $F^{-1}(t, \nu)$ in Q_n probability. If Assumption 7 also holds, $F(\cdot, \hat{\nu}_n)$ can be replaced by $G(\cdot, \hat{\nu}_n, \hat{\mu}_n)$ in these conclusions.

Proof. If Assumptions 5 and 6 hold, then

$$\sup_{u\in U} \|F_u(\cdot, \hat{\nu}_n) - F_u(\cdot, \nu)\| \to 0 \quad \text{in } Q_n \text{ probability} \tag{5.4}$$

and

$$\|F(\cdot, \hat{\nu}_n) - F(\cdot, \nu)\| \to 0 \quad \text{in } Q_n \text{ probability}. \tag{5.5}$$

If Assumption 7 also holds, then

$$\|G(\cdot, \hat{\nu}_n, \hat{\mu}_n) - F(\cdot, \nu)\| \to 0 \quad \text{in } Q_n \text{ probability}. \tag{5.6}$$

Indeed, let $\{\nu_n : n \geq 1\}$ be any sequence of probability measures on $C(U)$ such that $\lim_{n\to\infty} \rho(\nu_n, \nu) = 0$. By the Skorokhod–Dudley–Wichura theorem (Wichura 1970) there exist on some probability space processes $\{Z_n : n \geq 1\}$ and Z, with sample paths in $C(U)$, such that $\mathbf{L}(Z_n) = \nu_n$, $\mathbf{L}(Z) = \nu$, and

$$\lim_{n\to\infty} \sup_{u\in U} |Z_n(u) - Z(u)| = 0 \tag{5.7}$$

with probability 1. This implies that

$$\lim_{n\to\infty} \sup_{u\in U} \beta[F_u(\cdot, \nu_n), F_u(\cdot, \nu)] = 0, \tag{5.8}$$

where β is the bounded Lipschitz metric for distributions on the real line.

Let $\{u_n \in U\}$ be an arbitrary sequence. In (5.8), the bounded Lipschitz metric can be replaced by the Lévy metric. Consequently, for every $\varepsilon > 0$, there exists $n_0(\varepsilon)$ such that for every $n \geq n_0(\varepsilon)$ and for every real x

$$F_{u_n}(x - \varepsilon, \nu) - \varepsilon \leq F_{u_n}(x, \nu_n) \leq F_{u_n}(x + \varepsilon, \nu) + \varepsilon. \tag{5.9}$$

On the other hand, by Assumption 6, for every $\gamma > 0$

$$\sup_x |F_{u_n}(x \pm \varepsilon, \nu) - F_{u_n}(x, \nu)| < \gamma, \tag{5.10}$$

provided that ε is chosen sufficiently small. These two inequalities imply that

$$\lim_{n\to\infty} \sup_{u\in U} \|F_u(\cdot, \nu_n) - F_u(\cdot, \nu)\| = 0 \tag{5.11}$$

and hence (5.4) is true, in view of Assumption 5.

In particular, (5.11) entails that

$$\lim_{n\to\infty} \sup_{u\in U} |F_u[Z_n(u), \nu_n] - F_u[Z_n(u), \nu]| = 0 \tag{5.12}$$

with probability 1. From Assumption 6 and (5.7),

$$\lim_{n\to\infty} \sup_{u\in U} |F_u[Z_n(u), \nu] - F_u[Z(u), \nu]| = 0 \tag{5.13}$$

with probability 1. Combining (5.12) with (5.13) yields

$$\lim_{n\to\infty} \sup_{u\in U} F_u[Z_n(u), \nu_n] = \sup_{u\in U} F_u[Z(u), \nu] \tag{5.14}$$

with probability 1. Hence

$$\lim_{n\to\infty} \|F(\cdot, \nu_n) - F(\cdot, \nu)\| = 0, \tag{5.15}$$

in view of Assumption 6; consequently, (5.5) is true because of Assumption 5.

Let $\{\mu_n : n \geq 1\}$ be any sequence of probability measures on U such that $\lim_{n\to\infty} \rho(\mu_n, \mu) = 0$. From (5.12) and (5.13), it follows that

$$\lim_{n\to\infty} |\text{ess sup}_{\mu_n} F_u[Z_n(u), \nu_n] - \text{ess sup}_{\mu_n} F_u[Z(u), \nu]| = 0 \tag{5.16}$$

with probability 1. Since Z has continuous sample paths, so does the process $\{F_u[Z(u), \nu] : u \in U\}$, in view of Assumption 6. Since, in addition, μ has full support,

$$\lim_{n\to\infty} \text{ess sup}_{\mu_n} F_u[Z(u), \nu] = \sup_{u\in U} F_u[Z(u), \nu] \tag{5.17}$$

with probability 1. (Compare Lemma 3.1 in Beran and Millar 1987.) Combining (5.16) with (5.17) yields the conclusion

$$\lim_{n\to\infty} \|G(\cdot, \nu_n, \mu_n) - F(\cdot, \nu)\| = 0 \tag{5.18}$$

whenever $\max\{\rho(\mu_n, \mu), \rho(\nu_n, \nu)\}$ tends to 0 as n increases. Since $\max\{\rho(\hat{\mu}_n, \mu), \rho(\hat{\nu}_n, \nu)\}$ converges to 0 in Q_n probability, Assertion (5.6) follows from (5.18).

Next, we verify conclusions (a), (b), and (c). Because of (5.4),

$$\sup_{u\in U} |F_u[Y_n(u), \hat{\nu}_n] - F_u[Y_n(u), \nu]| \to 0 \tag{5.19}$$

in Q_n probability. On the other hand, Assumptions 5 and 6 imply that

$$\mathbf{L}\{\sup_{u\in U} F_u[Y_n(u), \nu] \mid Q_n\} \Rightarrow \mathbf{L}\{\sup_{u\in U} F_u[Y(u), \nu]\}, \tag{5.20}$$

where $\{Y(u) : u \in U\}$ is a process with distribution ν on $C(U)$. Combining (5.19) with (5.20) establishes

$$\mathbf{L}\{\sup_{u\in U} F_u[Y_n(u), \hat{\nu}_n] \mid Q_n\} \Rightarrow \mathbf{L}\{\sup_{u\in U} F_u[Y(u), \nu]\}. \tag{5.21}$$

Because of (5.5),

$$F\{\sup_{u\in U} F_u[Y_n(u), \hat{v}_n], \hat{v}_n\} - F\{\sup_{u\in U} F_u[Y_n(u), \hat{v}_n], v\} \to 0 \quad (5.22)$$

in Q_n probability. From this, (5.21), and Assumption 6 follows that

$$\mathbf{L}[F\{\sup_{u\in U} F_u[Y_n(u), \hat{v}_n], \hat{v}_n\} \mid Q_n] \Rightarrow \mathbf{L}[F\{\sup_{u\in U} F_u[Y(u), v], v\}]. \quad (5.23)$$

Since the distribution on the right side of (5.23) is uniform on (0, 1), conclusion (a) follows.

Let $\{u_n \in U\}$ be an arbitrary sequence. Assumptions 5 and 6 imply that

$$\lim_{n\to\infty} \beta[\mathbf{L}\{F_{u_n}[Y_n(u_n), v] \mid Q_n\}, \mathbf{L}\{F_{u_n}[Y(u_n), v]\}] = 0, \quad (5.24)$$

where β is bounded Lipschitz distance. Since $F_{u_n}[Y(u_n), v]$ is uniformly distributed on (0, 1), it follows from (5.24) and (5.19) that $\mathbf{L}\{F_{u_n}[Y_n(u_n), \hat{v}_n] \mid Q_n\}$ converges weakly to a uniform distribution on (0, 1). This convergence, (5.5), and Assumption 6 imply conclusion (b).

Conclusion (c) is immediate from (5.5) and Assumption 6.

If Assumption 7 holds, the use of (5.6) in place of (5.5) throughout the foregoing argument justifies the final assertion of the proposition.

Proof of Theorem 4.1. Refer to the notation of Section 4. Let $D_n(\theta) = \mathbf{L}[R_n(x_n, \theta) \mid P_{\theta,n}]$. Suppose that $\lim_{n\to\infty} d(\theta_n, \theta) = 0$. Make the following identification: $Q_n = P_{\theta_n,n}$, $\hat{v}_n = D_n(\hat{\theta}_n)$, $v = \mathbf{L}[R(\theta)]$, and $Y_n = R_n(x_n, \theta_n)$. Then Assumptions 5 and 6 hold, $F_u(\cdot, \hat{v}_n) = \hat{H}_{n,u}$, and $F(\cdot, \hat{v}_n) = \hat{H}_n$. Conclusions (a), (b), and (c) of Proposition 5.1 thus imply (4.1), (4.2), and the final assertion of Theorem 4.1, respectively.

Proof of Theorem 4.2. First part: Make the following changes in the proof for Theorem 4.1. Let $Q_n = Q_{\theta_n,n}$, and let $\hat{v}_n$ be the empirical measure on $C(U)$ of the $\{R_n(x^*_{n,j}, \hat{\theta}_n) : 1 \le j \le j_n\}$. Since $\hat{v}_n$ is the empirical measure of a sample of size j_n drawn from $D_n(\hat{\theta}_n)$, it follows by theorem 2 of Beran, LeCam, and Millar (1987) that $\rho(\hat{v}_n, v) \to 0$ in Q_n probability. Note that, for this choice of $\hat{v}_n$, $F_u(\cdot, \hat{v}_n) = H'_{n,u}$ and $F(\cdot, \hat{v}_n) = H'_n$.

Second part: Let $Q_n = Q_{\theta_n,n} \times \mu^{k_n}$. Observe that Assumption 7 holds and that $G(\cdot, \hat{v}_n, \hat{\mu}_n) = H''_n$; then refer to the last assertion of Proposition 5.1.

[*Received July 1986. Revised August 1987.*]

REFERENCES

Abramovitch, L., and Singh, K. (1985), "Edgeworth Corrected Pivotal Statistics and the Bootstrap," *The Annals of Statistics,* 13, 116–132.

Beran, R. (1987), "Prepivoting to Reduce Level Error of Confidence Sets," *Biometrika,* 74, 457–468.

Beran, R., Le Cam, L., and Millar, P. W. (1987), "Convergence of Stochastic Empirical Measures," *Journal of Multivariate Analysis,* 23, 159–168.

Beran, R., and Millar, P. W. (1987), "Stochastic Estimation and Testing," *The Annals of Statistics,* 15, 1131–1154.

Beran, R., and Srivastava, M. S. (1985), "Boostrap Tests and Confidence Regions for Functions of a Covariance Matrix," *The Annals of Statistics,* 13, 95–115.

Diaconis, P., and Efron, B. (1983), "Computer-Intensive Methods in Statistics," *Scientific American,* 248, 116–130.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics,* 7, 1–26.

——— (1987), "Better Bootstrap Confidence Intervals" (with discussion), *Journal of the American Statistical Association,* 82, 171–200.

Hall, P. (1986), "On the Bootstrap and Confidence Intervals," *The Annals of Statistics,* 14, 1431–1452.

Mardia, K. U., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis,* New York: Academic Press.

Miller, R. (1966), *Simultaneous Statistical Inference,* New York: McGraw-Hill.

Scheffé, H. (1953), "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika,* 40, 87–104.

Tukey, J. W. (1953), "The Problem of Multiple Comparisons," unpublished manuscript.

Wichura, M. J. (1970), "On the Construction of Uniformly Convergent Random Variables With Given Weakly Convergent Image Laws," *Annals of Mathematical Statistics,* 41, 284–291.