The Annals of Statistics
1989, Vol. 17, No. 1, 125–140

A STOCHASTIC MINIMUM DISTANCE TEST FOR MULTIVARIATE PARAMETRIC MODELS[1]

By R. Beran and P. W. Millar

University of California, Berkeley

Stochastic procedures are randomized statistical procedures which are functions of the observed sample and of one or more artificially constructed auxiliary samples. As the size of the auxiliary samples increases, a stochastic procedure becomes nearly nonrandomized. The stochastic test of this paper arises as a numerically feasible approximation to a natural minimum distance goodness-of-fit test for multivariate parametric models. The distance being minimized here is the half-space metric for probabilities on a Euclidean space. It is shown that the various approximations used in constructing the stochastic test and its critical values do not detract from its first-order asymptotic performance.

1. Introduction. The practical constraint of computational simplicity strongly shaped early statistical methods, such as Pearson's chi-squared test and the analysis of variance. The chi-squared statistic and the F-statistic are algebraically straightforward. The usual asymptotic null distribution of each statistic can be tabulated conveniently because it does not depend on unknown nuisance parameters whose values must be estimated from the data. The present availability of inexpensive high-speed computing has greatly widened the definition of a practical statistical procedure. In particular, bootstrap methods and random search ideas have recently solved several statistical problems which are difficult or intractable for purely analytical approaches.

Characteristic of the new results is their reliance on relatively abstract triangular array asymptotics combined with intensive computing. Here are several examples:

(i) Affinely invariant confidence sets for an unknown multivariate distribution [Beran and Millar (1986)]. The bootstrap aspect of this solution has been extended recently by Gaenssler (1986), Romano (1988) and by Sheehy and Wellner (1986). The random search aspect was foreshadowed in Pyke's (1984) discussion of tests for a simple multivariate hypothesis.

(ii) Bootstrap and random search implementations of minimum Kolmogorov distance estimates and tests [Beran (1986) and Beran and Millar (1987)]. This approach resolves the problem of calculating such minimum distance estimates and the problem of obtaining asymptotically valid critical values for such tests [see Durbin (1973) and Pollard (1980) for discussions of the difficulties in a purely analytical approach].

Received August 1987; revised April 1988.

[1]Research supported by National Science Foundation Grant DMS-87-01426.

AMS 1980 *subject classifications*. Primary 62E20; secondary 62H15.

Key words and phrases. Stochastic procedure, goodness-of-fit test, minimum distance test, bootstrap, stochastic search.

(iii) Iterated bootstrap methods for making second and higher order refinements to critical values for tests and confidence sets [Hall (1986) and Beran (1987)]. Semianalytical approximations to such procedures have been discussed by Abramovitch and Singh (1985) and by Efron (1987).

Underlying these examples is the concept of stochastic procedure, defined by Beran and Millar (1987). A stochastic procedure is a randomized estimate, test or confidence set with two properties:

(a) It is a function of the original sample and of one or more artificially constructed auxiliary samples.

(b) It becomes nearly nonrandomized when the auxiliary samples are increased in size.

An early example is a test for a simple hypothesis whose critical value is obtained by Monte Carlo approximation of the appropriate null distribution [cf. Dwass (1957)]. The three examples cites above are more elaborate instances of stochastic procedures. In general, stochastic procedures are randomized procedures, in the sense of decision theory, which arise as useful approximations to numerically intractable statistical procedures.

This paper studies minimum distance goodness-of-fit tests for multivariate parametric models. Let Θ be an open subset of R^d and let $\{P_\theta: \theta \in \Theta\}$ be a parametric family of probabilities on R^q. Let $\mathbf{x}_n = (x_1, \dots, x_n)$ be a sample of independent identically distributed R^q-valued random variables whose common distribution P is unknown. The null hypothesis to be tested asserts that P is some P_{θ_0}, where θ_0 is unknown. Let $\hat{P}_n = \hat{P}_n(\mathbf{x}_n, \cdot)$ denote the empirical measure which assigns probability n^{-1} to each of the observations in the sample. Define the minimum distance statistic for the model $\{P_\theta\}$ to be

$$T_n = \inf_{\theta} \sup_{A} n^{1/2} |\hat{P}_n(A) - P_\theta(A)|, \tag{1.1}$$

where the infimum is over Θ and the supremum is over all half-spaces of R^q.

When $q = 1$, the statistic T_n reduces to the minimum Kolmogorov distance statistic on the real line. For general q, the statistic T_n has two attractive properties: It is invariant under affine transformations of the data whenever the family $\{P_\theta: \theta \in \Theta\}$ is so invariant; and it makes good sense as a test statistic whether the distribution P of the data is discrete, absolutely continuous or singular with respect to Lebesgue measure.

The asymptotic null distribution of T_n can be characterized abstractly by using asymptotic theory for the empirical process on a Vapnik–Červonenkis class and a standard analysis of minimum distance statistics (see Section 2.1). In general, the cdf of this limit law is intractable and does not yield usable critical values for testing purposes. Moreover, the test statistic T_n itself is difficult to evaluate when the dimension of parameter space or sample space exceeds 1.

This paper introduces a stochastic goodness-of-fit which approximates the minimum distance test just described in a natural way. The construction of the new test in Section 2.2 involves stochastic approximations to both the supremum and infimum in (1.1) and a suitable bootstrap algorithm for the critical value.

The stochastic test is numerically feasible, yet has the same first-order asymptotics as the motivating minimum distance test under the null hypothesis and under contiguous alternatives. The main results are described in Section 2.3. The theoretical success of the stochastic test as an approximation to the minimum distance test depends crucially upon the finite-dimensionality of both parameter space and sample space.

2. The tests and their asymptotics. To provide necessary background for the stochastic goodness-of-fit test, this section first reviews the asymptotic theory, under the null hypothesis, of the test based on the minimum distance statistic T_n. The definition and asymptotic theory of the stochastic test follows in Sections 2.2 and 2.3.

2.1. *Asymptotics of the minimum distance test.* Let 1.1 and $\langle \cdot , \cdot \rangle$ denote, respectively, Euclidean norm and inner product in R^q. Let $S_q = \{s \in R^q: |s| = 1\}$ be the unit sphere in R^q. Any half-space A of R^q can be parametrized as

$$A(s, t) = \{x \in R^q: \langle s, x \rangle \leq t\}, \tag{2.1}$$

where $(s, t) \in S_q \times R$. Let L_∞ be the set of all bounded measurable functions on $S_q \times R$, metrized by the supremum norm $\|\cdot\|$. The σ-algebra in L_∞ is that generated by open balls. Any probability P on R^q can be regarded as an element of L_∞ by identifying P with the function which maps $(s, t) \in S_q \times R$ into $P[A(s, t)]$. With this 1–1 identification, the minimum distance statistic T_n defined in (1.1) can be written as

$$T_n = \inf_\theta n^{1/2} \|\hat{P}_n - P_\theta\|. \tag{2.2}$$

The measurability of expressions such as (2.2) follows from two facts: The supremum defining the norm $\|\cdot\|$ can be replaced by a supremum over a countable number of half-spaces [cf. Section 2 of Beran and Millar (1986)] and P_θ will be assumed $\|\cdot\|$-continuous as a function of θ.

The null hypothesis to be tested asserts that the actual distribution P of each observation is some P_{θ_0}, where θ_0 is an unknown element of Θ, which is an open subset of R^d. The asymptotic null distribution of T_n can be characterized as follows: Let $W_{\theta_0} = \{W_{\theta_0}(s, t): (s, t) \in S_q \times R\}$ be a Gaussian process with sample paths in L_∞, mean 0 and covariance function

$$\operatorname{Cov}\left[W_{\theta_0}(s, t), W_{q_0}(s', t')\right] = P_{\theta_0}(A \cap A') - P_{\theta_0}(A) P_{\theta_0}(A'), \tag{2.3}$$

where A, A' stand for $A(s, t)$, $A(s', t')$, respectively. Suppose the parametric model satisfies the following conditions, for every $\theta_0 \in \Theta$:

ASSUMPTION A1. Identifiability. For every positive c,

$$\inf\left\{\|P_\theta - P_{\theta_0}\|: |\theta - \theta_0| > c\right\} > 0. \tag{2.4}$$

ASSUMPTION A2. Norm differentiability. There exists a bounded vector function m_{θ_0}, mapping $S_q \times R$ into R^d, such that

$$\|P_\theta - P_{\theta_0} - \langle m_{\theta_0}, \theta - \theta_0 \rangle\| = o(|\theta - \theta_0|). \tag{2.5}$$

ASSUMPTION A3. Nonsingularity. There exists a positive constant C such that

$$\|\langle m_{\theta_0}, u \rangle\| \geq C|u| \tag{2.6}$$

for every vector $u \in R^d$.

ASSUMPTION A4. Quarter-space continuity. For every pair of half-spaces A and A' of R^q,

$$\lim_{\theta \to \theta_0} |P_\theta(A \cap A') - P_{\theta_0}(A \cap A')| = 0. \tag{2.7}$$

PROPOSITION 2.1. *Let* $\{\theta_n \in \Theta\}$ *be any sequence such that* $\{n^{1/2}(\theta_n - \theta_0)\}$ *is bounded. Suppose Assumptions* A1–A4 *hold. Then*

$$\mathbf{L}(T_n | P_{\theta_n}^n) \Rightarrow H(\theta_0), \tag{2.8}$$

where

$$H(\theta_0) = \mathbf{L}\Big(\inf_u \|W_{\theta_0} - \langle m_{\theta_0}, u \rangle\|\Big). \tag{2.9}$$

PROOF. Indeed, under Assumption A4,

$$\mathbf{L}\Big[n^{1/2}\big(\hat{P}_n - P_{\theta_n}\big) | P_{\theta_n}^n\Big] \Rightarrow \mathbf{L}(W_{\theta_0}) \tag{2.10}$$

as random elements of L_∞. This triangular array central limit theorem for the empirical process on half-spaces follows from Le Cam (1983); see Section 4 of Beran and Millar (1986) for further details. Consequently, by the analysis in Pollard (1980) of minimum distance test statistics and by (2.10),

$$\begin{aligned} T_n &= \inf_\theta n^{1/2} \|\hat{P}_n - P_{\theta_0} + \langle m_{\theta_0}, \theta - \theta_0 \rangle\| + o_p(1) \\ &= \inf_\theta \|n^{1/2}\big(\hat{P}_n - P_{\theta_n}\big) + n^{1/2} \langle m_{\theta_0}, \theta - \theta_n \rangle\| + o_p(1) \\ &\Rightarrow H(\theta_0) \end{aligned} \tag{2.11}$$

under $\{P_{\theta_n}^n\}$. □

In general, Proposition 2.1 does not directly yield usable critical values for a goodness-of-fit test based on T_n. Let $\hat{\theta}_n = \hat{\theta}_n(\mathbf{x}_n)$ be a consistent estimate of θ_0. The corresponding estimated asymptotic null distribution $H(\hat{\theta}_n)$ is rarely tractable, either analytically or numerically.

A more promising bootstrap approach to the critical values of T_n runs as follows. Let

$$H_n(\theta) = \mathbf{L}(T_n | P_\theta^n). \tag{2.12}$$

The bootstrap estimate for the null distribution of T_n is defined to be $H_n(\hat{\theta}_n)$. The corresponding bootstrap test based on T_n is

$$\phi_n = \begin{cases} 1, & \text{if } T_n > c_n(\alpha), \\ 0, & \text{otherwise,} \end{cases} \tag{2.13}$$

where $c_n(\alpha)$ is the largest $(1 - \alpha)$th quantile of $H_n(\hat{\theta}_n)$. Let ρ be any metric which metrizes weak convergence on the real line. Introduce the following additional assumption on the estimates $\{\hat{\theta}_n\}$:

ASSUMPTION A5. Root-n consistency. If $\{\theta_n \in \Theta\}$ is any sequence such that $\{n^{1/2}(\theta_n - \theta_0)\}$ is bounded, then $\{\mathbf{L}[n^{1/2}(\hat{\theta}_n - \theta_n)|P^n_{\theta_n}]\}$ is tight.

PROPOSITION 2.2. *Fix α. Suppose the assumptions of Proposition* 2.1 *hold together with* A5. *Then*

$$\rho\left[H_n(\hat{\theta}_n), H(\theta_0)\right] \to 0 \quad \textit{in } P^n_{\theta_n}\textit{-probability.} \tag{2.14}$$

Suppose, in addition, that the $(1 - \alpha)$*th quantiles of* $H(\theta_0)$ *are all continuity points of its cdf. Then*

$$\lim_{n\to\infty} E\left(\phi_n|P^n_{\theta_n}\right) = \alpha. \tag{2.15}$$

PROOF. Indeed, conclusion (2.14) is immediate from Proposition 2.1, in view of (2.12) and A5. Let $c_L(\alpha)$ and $c_U(\alpha)$ denote the smallest and largest $(1 - \alpha)$th quantiles of $H(\theta_0)$, defined formally as in Section 2 of Beran (1986). Let F denote the right-continuous cdf of $H(\theta_0)$. By a variant of Theorem 2.1 in the paper just cited,

$$\begin{aligned} 1 - F\left[c_U(\alpha)\right] &\le \liminf_{n\to\infty} E\left(\phi_n|P^n_{\theta_n}\right) \\ &\le \limsup_{n\to\infty} E\left(\phi_n|P^n_{\theta_n}\right) \le 1 - F\left[c_L(\alpha)-\right]. \end{aligned} \tag{2.16}$$

When the choice of α is restricted, as in the second part of Proposition 2.2, then (2.15) ensues.

Formula (2.15) says that $E(\phi_n|P^n_\theta)$—the probability that ϕ_n rejects when the null hypothesis is true—converges to α uniformly over balls of the form $\{\theta \in \Theta: |\theta - \theta_0| \le n^{-1/2}c\}$. In general, we cannot expect uniform convergence over fixed compact subsets of Θ unless we are able to strengthen the assumptions on the parametric model and on the estimates $\{\hat{\theta}_n\}$. □

In principle, the bootstrap critical value $c_n(\alpha)$ in (2.13) can be approximated by Monte Carlo methods. However, the supremum over all half-spaces in the definition (1.1) of T_n is hard to evaluate when the dimension q exceeds 1. In addition, the infimum over all Θ in (1.1) is difficult to find when the dimension d is not small. The next section introduces stochastic approximations to the supremum and infimum which reduce the computational burden substantially.

2.2. *The stochastic test.* The motivating idea is to use three stochastic approximations in calculating T_n:

(i) Approximate the supremum in (1.1) by the supremum over a set of randomly chosen half-spaces.

(ii) Approximate the infimum over θ in (1.1) by the minimum over a finite number of randomly chosen values of θ.

(iii) For each value of θ chosen in step (ii), approximate P_θ in (1.1) by the empirical measure of a sample of independent observations from that distribution P_θ.

The third step serves to approximate $P_\theta(A)$ for an arbitrary half-space A of R^q. Of course, step (iii) can be omitted when $\{P_\theta: \theta \in \Theta\}$ is a multivariate normal family of distributions, since an analytical calculation is then available for $P_\theta(A)$.

Let $\mathbf{s}_n = (s_1, \dots, s_{j_n})$ be a sample of j_n iid random unit vectors, each uniformly distributed over S_q. For every function f in L_∞, define the *stochastic norm* $\|\cdot\|_n$ of f to be

$$\|f\|_n = \max_{1 \le j \le j_n} \sup_t |f(s_j, t)|. \tag{2.17}$$

Let $\hat{\Theta}_n$ be a *local* random search sample of size k_n in Θ, constructed as follows. Suppose $\hat{\theta}_n = \hat{\theta}_n(\mathbf{x}_n)$ is a root-n consistent estimate of θ. Conditionally on the original sample $\mathbf{x}_n$, draw k_n independent bootstrap samples $\mathbf{x}_1^*, \dots, \mathbf{x}_{k_n}^*$, each of size n from the fitted model $P_{\hat{\theta}_n}$. Let $\theta_{n,k}^* = \hat{\theta}_n(\mathbf{x}_k^*)$ be the value of the estimate of θ recalculated from the kth bootstrap sample $\mathbf{x}_k^*$ and set

$$\hat{\Theta}_n = \left(\hat{\theta}_n, \theta_{n,1}^*, \dots, \theta_{n,k_n}^*\right). \tag{2.18}$$

The adjective local used above to describe $\hat{\Theta}_n$ refers to the following property: Under $P_{\theta_n}^n$, the expected number of elements in $\hat{\Theta}_n$ which fall within a ball of radius $O(n^{-1/2})$ about θ_n tends to infinity whenever $\lim_{n\to\infty} k_n = \infty$, regardless of the dimension d of Θ. This concentration of $\hat{\Theta}_n$ on balls of radius $O(n^{-1/2})$ about θ_n is important for computational efficiency because the infimum over θ in (2.2) is typically achieved within such a shrinking ball. For further discussion, see Section 1 of Beran and Millar (1987).

For each $\theta \in \hat{\Theta}_n$, let $\hat{P}_{\theta,n}$ be the empirical distribution of i_n conditionally independent (given $\mathbf{x}_n$ and $\hat{\Theta}_n$) identically distributed random variables drawn from P_θ. By construction the variables $(\mathbf{x}_n, \hat{\Theta}_n, \{\hat{P}_{\theta,n}: \theta \in \hat{\Theta}_n\})$ and the variables $\mathbf{s}_n$ are independent.

The proposed stochastic approximation to the minimum distance statistic T_n is

$$\hat{T}_n = \max\left\{n^{1/2}\|\hat{P}_n - \hat{P}_{\theta,n}\|_n : \theta \in \hat{\Theta}_n\right\}. \tag{2.19}$$

The evaluation of $\hat{T}_n$ is straightforward, though computer-intensive, and is usually much simpler than the evaluation of T_n. The null distribution of $\hat{T}_n$ is determined by the joint distribution, under the null hypothesis, of the original sample $\mathbf{x}_n$ and of the auxiliary random variables $\mathbf{s}_n$, $\hat{\Theta}_n$ and $\{\hat{P}_{\theta,n}: \theta \in \hat{\Theta}_n\}$.

Nevertheless, under conditions to be described in the next section, the asymptotic null distribution of $\hat{T}_n$ coincides with that of T_n. This result is a pleasant surprise which depends crucially upon the finite-dimensionality of both the parameter space Θ and of the sample space.

The entire construction of $\hat{T}_n$ can be bootstrapped, using the fitted null hypothesis model, to yield asymptotically valid critical values for a goodness-of-fit test. This direct approach, discussed more fully in Section 3.1, requires extensive computing. The following *conditional* bootstrap algorithm is considerably more practical and still gives asymptotically valid critical values.

Given the original sample $\mathbf{x}_n$, draw m_n new conditionally independent bootstrap samples $\mathbf{y}_1^*, \dots, \mathbf{y}_{m_n}^*$, each of size n, from the fitted model $P_{\hat{\theta}_n}$. By construction, the $\{\mathbf{y}_m^*\}$ and $\mathbf{s}_n$ are independent and the $\{\mathbf{y}_m^*\}$ and $(\hat{\Theta}_n, \{\hat{P}_{\theta, n}: \theta \in \hat{\Theta}_n\})$ are conditionally independent, given $\mathbf{x}_n$. Keeping the auxiliary variables in $\mathbf{s}_n$, $\hat{\Theta}_n$ and $\{\hat{P}_{\theta, n}: \theta \in \hat{\Theta}_n\}$ *unchanged*, compute

$$T_m^* = \min\{n^{1/2}\|\hat{P}_n(\mathbf{y}_m^*) - \hat{P}_{\theta, n}\|_n: \theta \in \hat{\Theta}_n\} \tag{2.20}$$

for $1 \le m \le m_n$. The statistic T_m^* is simply a recalculation of $\hat{T}_n$, in which the empirical distribution $\hat{P}_m(\mathbf{y}_m^*)$ of the mth bootstrap sample replaces the empirical distribution $\hat{P}_n$ of the original sample. Nothing else is bootstrapped.

The empirical distribution $\hat{H}_n$ of the values $\{T_m^*: 1 \le m \le m_n\}$ is the proposed *conditional* bootstrap estimate of the null distribution of $\hat{T}_n$. The corresponding stochastic goodness-of-fit test is

$$\hat{\phi}_n = \begin{cases} 1, & \text{if } \hat{T}_n > \hat{c}_n(\alpha), \\ 0, & \text{otherwise,} \end{cases} \tag{2.21}$$

where $\hat{c}_n(\alpha)$ is the largest $(1 - \alpha)$th quantile of $\hat{H}_n$. The next section studies the asymptotic behavior of $\hat{\phi}_n$ under the null hypothesis.

2.3. *Asymptotics of the stochastic test.* This theme will be developed under assumptions which differ somewhat from those in Section 2.1. When $\mathbf{x}_n$ has distribution $P_{\theta_n}^n$ let Q_n denote the joint distribution of $\mathbf{x}_n$ and of the auxiliary variables $\mathbf{s}_n$, $\hat{\Theta}_n$ and $\{\hat{P}_{\theta, n}: \theta \in \hat{\Theta}_n\}$. Recall the definition (2.17) of the stochastic norm $\|\cdot\|_n$.

Assumption B1. Stochastic identifiability. For every positive c and ε, there exists positive δ such that

$$\liminf_{n \to \infty} Q_n\left[\inf\{\|P_\theta - P_{\theta_0}\|_n: |\theta - \theta_0| > c\} \ge \delta\right] \geqq 1 - \varepsilon. \tag{2.22}$$

Assumption B2. Stochastic norm differentiability. There exists a bounded continuous vector function m_{θ_0}, mapping $S_q \times R$ into R^d, such that the following assertion holds: For every positive ε, there exists positive δ such that

$$\lim_{n \to \infty} Q_n\left[\sup_{|\theta - \theta_0| \le \delta}\{\|P_\theta - P_{\theta_0} - \langle m_{\theta_0}, \theta - \theta_0\rangle\|_n / |\theta - \theta_0|\} \le \varepsilon\right] = 1. \tag{2.23}$$

ASSUMPTION B3. Stochastic nonsingularity. There exists a sequence of positive random variables $\{C_n\}$ such that $\{\mathbf{L}(C_n^{-1}|Q_n)\}$ is tight and

$$\|\langle m_{\theta_0}, u\rangle\|_n \geq C_n|u| \tag{2.24}$$

for every vector $u \in R^d$.

ASSUMPTION B5. Regularity. If $\{\theta_n \in \Theta\}$ is any sequence such that $\{n^{1/2}(\theta_n - \theta_0)\}$ is bounded then $\{\mathbf{L}[n^{1/2}(\hat{\theta}_n - \theta_n)|P_{\theta_n}^n]\}$ converges weakly to a distribution μ_{θ_0} which does not depend on the sequence $\{\theta_n\}$. Moreover, $\mu_{\theta_0}(O) > 0$ for every open set O in R^d.

Note that Assumptions B1 and B2 are weaker than A1 and A2 respectively, apart from the continuity in B2, while B3 and B5 are stronger than A3 and A5 respectively. Other aspects of the assumptions are discussed in Section 3.3. Propositions 2.3 and 2.4, which follow, are analogs for the stochastic test of the earlier Propositions 2.1 and 2.2.

PROPOSITION 2.3. *Let $\{\theta_n \in \Theta\}$ be any sequence such that $\{n^{1/2}(\theta_n - \theta_0)\}$ is bounded. Suppose Assumptions* B1–B3 *and* A4 *and* B5 *hold and*

$$\begin{aligned} &\lim_{n\to\infty} i_n/[n \log(k_n)] = \infty, \\ &\lim_{n\to\infty} j_n = \lim_{n\to\infty} k_n = \infty. \end{aligned} \tag{2.25}$$

Then

$$\mathbf{L}(\hat{T}_n|Q_n) \Rightarrow H(\theta_0), \tag{2.26}$$

where $H(\theta_0)$ is defined by (2.9).

The proof of this result is deferred to Section 4. Note that the rate of convergence condition (2.25) on i_n, j_n and k_n does not depend on either the dimension q of the sample space or on the dimension d of the parameter space. In particular, the sizes of the auxiliary samples needed to construct $\hat{T}_n$ do not explode as q or d increases. This property of $\hat{T}_n$—of great practical importance—is remarkable. It is ultimately a consequence of two facts: the careful design of the random searches over S_q and over Θ and the finiteness of both q and d. Nonstochastic grid searches over S_q and Θ require search sample sizes which grow exponentially in q and d to obtain a result like Proposition 2.3. When q and d are both infinite, Proposition 2.3 can fail (work in progress by the authors).

When $\mathbf{x}_n$ has distribution $P_{\theta_n}^n$, let Q_n' denote the joint distribution of $\mathbf{x}_n$, of $\mathbf{s}_n, \hat{\Theta}_n$ and $\{\hat{P}_{\theta,n}: \theta \in \hat{\Theta}_n\}$ and of the bootstrap samples $\{\mathbf{y}_m^*: 1 \leq m \leq m_n\}$.

PROPOSITION 2.4. *Fix α. Suppose the assumptions of Proposition* 2.3 *hold and*

$$\lim_{n\to\infty} m_n = \infty. \tag{2.27}$$

Then

$$\rho\left[\hat{H}_n, H(\theta_0)\right] \to 0 \quad \textit{in } Q_n'\textit{-probability}. \tag{2.28}$$

Suppose, in addition, that the $(1-\alpha)$*th quantiles of* $H(\theta_0)$ *are all continuity points of its cdf. Then*

$$\lim_{n\to\infty} E\left(\hat{\phi}_n | Q_n'\right) = \alpha. \tag{2.29}$$

Thus, the probability that the stochastic minimum distance test $\hat{\phi}_n$ rejects when the null hypothesis is true converges to α, uniformly over balls of the form $\{\theta \in \Theta: |\theta - \theta_0| \le n^{-1/2}c\}$. It is straightforward to extend Propositions 2.2 and 2.4 to contiguous local alternatives. Suppose that the sequence of distributions for $\mathbf{x}_n$ does not belong to the parametric model $\{P_\theta^n: \theta \in \Theta\}$ but is contiguous to $\{P_{\theta_0}^n\}$. Then, the asymptotic powers of ϕ_n and $\hat{\theta}_n$ under these alternatives coincide. This conclusion follows readily from the proofs of Propositions 2.2 and 2.4, the latter to be found in Section 4.

3. Extensions. This section sketches several possible extensions of the statistical methods and theory presented in Section 2. Proofs are omitted since they resemble those in Section 4 or are straightforward.

3.1. *The unconditional bootstrap.* The triangular array aspect of Proposition 2.3 justifies a more elaborate bootstrap algorithm for approximating the null distribution of the stochastic test statistic $\hat{T}_n$. Unlike the conditional bootstrap algorithm described in Section 2.2, the alternative algorithm bootstraps the entire construction of $\hat{T}_n$, as follows:

Given the original sample $\mathbf{x}_n$, draw m_n conditionally independent bootstrap samples $\mathbf{y}_1^*, \dots, \mathbf{y}_{m_n}^*$, each of size n, from the fitted model $P_{\hat{\theta}_n}$. Treating the mth bootstrap sample $\mathbf{y}_m^*$ as the original data set, recalculate the stochastic test statistic, as described in Section 2.2, to obtain the value $T_{m,1}^*$. For each m, this process involves several operations: drawing a new search sample $\mathbf{s}_n$ of size j_n from the uniform distribution on S_q, constructing a new search sample $\hat{\Theta}_n$ by parametrically bootstrapping $\theta_n(\mathbf{y}_m^*)$ k_n times and recalculating a new empirical approximation $\hat{P}_{\theta,n}$ to P_θ for each θ in $\hat{\Theta}_n$.

Let $\hat{H}_{n,1}$ denote the joint empirical distribution of the $\{T_{m,1}^*: 1 \le m \le m_n\}$. Let $Q_{n,1}'$ denote the distribution, when $\mathbf{L}(\mathbf{x}_n) = P_{\theta_n}^n$, of $\mathbf{x}_n$, and of all the auxiliary samples generated by the unconditional bootstrap algorithm just described. It follows from Proposition 2.3, Assumption B5 on $\hat{\theta}_n$ and Theorem 2.1 of Beran and Millar (1987) that

$$\rho\left[\hat{H}_{n,1}, H(\theta_0)\right] \to 0 \quad \text{in } Q_{n,1}'\text{-probability}. \tag{3.1}$$

Let $\hat{\phi}_{n,1}$ denote the stochastic test which rejects the null hypothesis whenever $\hat{T}_n$ exceeds the largest $(1-\alpha)$th quantile of the unconditional bootstrap distribution

$\hat{H}_{n,1}$. Then

$$\lim_{n\to\infty} E(\hat{\phi}_{n,1}|Q'_{n,1}) = \alpha, \tag{3.2}$$

provided α is chosen as in Proposition 2.4.

The unconditional bootstrap test $\hat{\phi}_{n,1}$ requires considerably more computing than does the conditional bootstrap test $\hat{\phi}_n$ of Section 2.2. It is not known whether $\hat{\phi}_{n,1}$ has any theoretical superiority over $\hat{\phi}_n$.

3.2. *Other goodness-of-fit tests.* An alternative to the minimum distance statistic T_n defined in (2.2) is the simpler statistic

$$R_n = n^{1/2}\|\hat{P}_n - P_{\hat{\theta}_n}\|, \tag{3.3}$$

which compares the fitted parametric model $P_{\hat{\theta}_n}$ with the empirical distribution of the sample. Suppose Assumptions A2, A4 and B5 hold and the joint distributions $\{\mathbf{L}[(W_n, n^{1/2}(\hat{\theta}_n - \theta_n))|P^n_{\theta_n}]\}$ converge weakly to $\mathbf{L}[(W_{\theta_0}, Y_{\theta_0})]$, where $W_n = n^{1/2}(\hat{P}_n - P_{\theta_n})$ and $\mathbf{L}(Y_{\theta_0}) = \mu_{\theta_0}$, as in B5. Then

$$\begin{aligned}\mathbf{L}(R_n|P_{\theta_n}) &\Rightarrow \mathbf{L}(\|W_{\theta_0} - \langle m_{\theta_0}, Y_{\theta_0}\rangle\|)\\ &= K(\theta_0), \quad \text{say.}\end{aligned} \tag{3.4}$$

A more computable stochastic approximation to R_n is the statistic

$$\hat{R}_n = n^{1/2}\|\hat{P}_n - \hat{P}_{\hat{\theta}_n, n}\|_n, \tag{3.5}$$

where the stochastic norm $\|\cdot\|_n$ and the empirical measure $\hat{P}_{\hat{\theta}_n, n}$ are defined as in Section 2.1. Note that $\hat{R}_n$ is obtained formally from $\hat{T}_n$ by setting $\hat{\Theta}_n = \{\hat{\theta}_n\}$. In this special case, Q_n reduces to the joint distribution of $\mathbf{x}_n$, $\mathbf{s}_n$ and $\hat{P}_{\hat{\theta}_n, n}$ when the distribution of $\mathbf{x}_n$ is $P^n_{\theta_n}$. Suppose

$$\begin{aligned}&\lim_{n\to\infty} i_n/n = \infty,\\ &\lim_{n\to\infty} j_n = \infty.\end{aligned} \tag{3.6}$$

Under assumptions B2, A4, B5 and the joint weak convergence assumption on $\{W_n, n^{1/2}(\hat{\theta}_n - \theta_n)\}$ stated in the previous paragraph,

$$\mathbf{L}(\hat{R}_n|Q_n) \Rightarrow K(\theta_0). \tag{3.7}$$

The method of Section 2.2, with $\hat{\Theta}_n = \{\hat{\theta}_n\}$, yields asymptotically valid conditional bootstrap critical values for $\hat{R}_n$.

Intermediate between $\hat{R}_n$ and $\hat{T}_n$, in computational complexity, are test statistics of the same form as $\hat{T}_n$ in which the cardinality k_0 of the search sample $\hat{\Theta}_n$ is held fixed as n increases. The statistic $\hat{R}_n$ corresponds to the special case $k_0 = 1$. Under assumptions nearly identical to those of the preceding paragraph,

$$\mathbf{L}(\hat{T}_n|Q_n) \Rightarrow L(\theta_0), \tag{3.8}$$

where

$$L(\theta_0) = \mathbf{L}\left[\min\left\{\|W_{\theta_0} - \langle m_{\theta_0}, Y_{\theta_0}\rangle\|, \min_{1\le k\le k_0}\|W_{\theta_0} - \langle m_{\theta_0}, Y_{\theta_0} + Y_{\theta_0,k}\rangle\|\right\}\right] \tag{3.9}$$

and Y_{θ_0}, $\{Y_{\theta_0,k}\}$ are iid with distribution μ_{θ_0} of Assumption B5. The power of the various tests described in this paper needs further study.

3.3. *Comments on the assumptions.* In what follows, $P_\theta = P_\theta(s, t)$ is the probability assigned to the half-space $A(s, t)$. This function P_θ is viewed as an element of L_∞.

Sufficient conditions for A2. Suppose that for every $(s, t) \in S_q \times R$ and for every parameter value θ in a neighborhood of θ_0, P_θ has gradient vector $m_\theta = m_\theta(s, t)$ whose components are continuous in θ. Suppose, in addition, that these components $\{m_{\theta,j}: 1 \le j \le d\}$ are norm-continuous at θ_0 in the sense

$$\lim_{\theta\to\theta_0}\|m_{\theta,j} - m_{\theta_0,j}\| = 0, \qquad 1 \le j \le d, \tag{3.10}$$

and that the $\{m_{\theta_0,j}\}$ are elements of L_∞. Then A2 holds, by an argument resting on the fundamental theorem of calculus.

Sufficient condition for A3. Suppose the components $\{m_{\theta_0,j}\}$ are linearly independent functions of (s, t). Then A3 holds. Indeed,

$$\begin{aligned}\inf\left\{|\langle m_{\theta_0}, u\rangle/|u|: u \in R^d\right\} &= \inf\left\{|\langle m_{\theta_0}, u\rangle|: u \in R^d, |u| = 1\right\}\\ &= c > 0,\end{aligned} \tag{3.11}$$

because the infimum is attained on the compact unit ball in R^d.

Alternatives to B1 *and* B2. For proving Proposition 2.3, the extrema over θ in Assumptions B1 and B2 can be further restricted to values of $\theta \in \hat{\Theta}_n$. This follows by inspection of the proof in Section 4. Alternatively, B1 can be replaced by a rate condition on k_n,

$$\lim_{n\to\infty} k_n P^n_{\theta_n}\left(|\hat{\theta}_n - \theta_n| > c\right) = 0 \tag{3.12}$$

for every positive c whenever $\{n^{1/2}(\theta_n - \theta_0)\}$ is bounded. In typical examples, the condition

$$\lim_{n\to\infty}\left[\log(k_n)/n\right] = 0 \tag{3.13}$$

implies (3.12), by a large deviations argument.

Indeed, to establish (4.7) in the proof of Proposition 2.3, it suffices to verify that, under (3.12), the probability of there being at least one $\theta \in \hat{\Theta}_n$ such that $|\theta - \theta_0| > c$ tends to zero as k_n increases. By the definition of $\hat{\Theta}_n$,

$$\begin{aligned}&Q_n\left[\max\{|\theta - \theta_0| > c\}: \theta \in \hat{\Theta}_n\right]\\ &\quad\le Q_n\left[\max_{1\le k\le k_n}|\theta^*_{n,k} - \hat{\theta}_n| > c/2\right] + P^n_{\theta_n}\left[|\hat{\theta}_n - \theta_n| > c/2\right].\end{aligned} \tag{3.14}$$

In view of B5, it is enough to show that the first term on the right side of (3.14) tends to 0 as n increases. Evidently,

$$
\begin{aligned}
Q_n\left[\max_{1\le k\le k_n}|\theta^*_{n,k}-\hat{\theta}_n|>c/2|\mathbf{x}_n\right] &= P^n_{\hat{\theta}_n}\left[\max_{1\le k\le k_n}|\theta^*_{n,k}-\hat{\theta}_n|>c/2|\mathbf{x}_n\right] \\
&= 1-\left[1-\hat{P}^n_{\hat{\theta}_n}\{|\theta^*_{n,1}-\hat{\theta}_n|>c/2|\mathbf{x}_n\}\right]^{k_n} \qquad (3.15)\\
&\le \left\{k_n P^n_{\hat{\theta}_n}\left[|\theta^*_{n,1}-\hat{\theta}_n|>c/2|\mathbf{x}_n\right]\right\},
\end{aligned}
$$

which tends to 0 in probability under B5 and (3.12). Hence the unconditional probability also converges to 0.

The support of μ_{θ_0} *in* B5. Under Hájek's (1970) regularity conditions on the model $\{P_\theta: \theta\in\Theta\}$, the limit distribution μ_{θ_0} in B5 can be written as the convolution of a proper normal distribution with another, possibly degenerate, distribution. In these circumstances, μ_{θ_0} has full support in R^d, as required by the last part of B5.

4. Proofs. This section proves Propositions 2.3 and 2.4. The argument extends the methods of Wolfowitz (1957), Pollard (1980) and others. Complications arise because both the norm $\|\cdot\|_n$ and the search set $\hat{\Theta}_n$ are random and because $\hat{\Theta}_n$ is not an open set.

PROOF OF PROPOSITION 2.3. By Alexander's (1984) inequality for the empirical process on a Vapnik–Červonenkis class, there exist positive constants C_1, C_2 such that

$$P^n_\theta\left[i_n^{1/2}\|\hat{P}_{\theta,n}-P_\theta\|\ge\lambda\right]\le C_1\exp\left(-C_2\lambda^2\right) \qquad (4.1)$$

for every $\theta\in\Theta$. Consequently, for every positive λ,

$$
\begin{aligned}
&Q_n\left[\sup\{n^{1/2}\|\hat{P}_{\theta,n}-P_\theta\|: \theta\in\hat{\Theta}_n\}>\lambda\right] \\
&\qquad\le (k_n+1)C_1\exp\left[-C_2(i_n/n)\lambda^2\right]\to 0,
\end{aligned} \qquad (4.2)
$$

because of condition (2.25). Thus

$$\inf\{n^{1/2}\|\hat{P}_n-\hat{P}_{\theta,n}\|_n: \theta\in\hat{\Theta}_n\} = \inf\{n^{1/2}\|\hat{P}_n-P_\theta\|_n: \theta\in\hat{\Theta}_n\}+o_{Q_n}(1). \qquad (4.3)$$

Let $V_n = n^{1/2}(\hat{P}_n-P_{\theta_0})$. The distributions $\{\mathbf{L}(\|V_n\|_n|Q_n)\}$ are tight because of Assumption B2, the conditions on $\{\theta_n\}$ and the weak convergence of $\{n^{1/2}(\hat{P}_n-P_{\theta_n})\}$ under Q_n to the Gaussian process W_{θ_0}. See Proposition 1 in Beran and Millar (1986) for the last point.

Let $\hat{\Theta}_{n,c}=\{\theta\in\hat{\Theta}_n: |\theta-\theta_0|\le c\}$, where c is positive and finite. Since

$$
\begin{aligned}
&\inf\left\{\|\hat{P}_n-P_\theta\|_n: \theta\in\hat{\Theta}_n-\hat{\Theta}_{n,c}\right\} \\
&\qquad\ge \inf\left\{\|P_\theta-P_{\theta_0}\|_n: \theta\in\hat{\Theta}_n-\hat{\Theta}_{n,c}\right\}-n^{-1/2}\|V_n\|_n,
\end{aligned} \qquad (4.4)
$$

it follows from B1 and the preceding paragraph that the left side of (4.4) is bounded away from 0 in Q_n-probability. On the other hand, since

$$\lim_{n\to\infty} Q_n\left[\inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in \hat{\Theta}_{n,c}\} \le \|\hat{P}_n - P_{\hat{\theta}_n}\|_n\right] = 1, \tag{4.5}$$

it follows from B2 and B5 that

$$\inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in \hat{\Theta}_{n,c}\} = o_{Q_n}(1). \tag{4.6}$$

Thus, for every positive c,

$$\lim_{n\to\infty} Q_n\left[\inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in \hat{\Theta}_n\} = \inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in \hat{\Theta}_{n,c}\}\right] = 1. \tag{4.7}$$

For the positive random variables $\{C_n\}$ in Assumption B3, let

$$N_n = \left\{\theta \in \hat{\Theta}_n : \|P_\theta - P_{\theta_0} - \langle m_{\theta_0}, \theta - \theta_0\rangle\|_n \le 2^{-1}C_n|\theta - \theta_0|\right\}. \tag{4.8}$$

For every positive δ, let

$$\begin{aligned} r_n(\delta) &= \sup\{\|P_\theta - P_{\theta_0} - \langle m_{\theta_0}, \theta - \theta_0\rangle\|_n / |\theta - \theta_0| : |\theta - \theta_0| \le \delta\}, \\ S(\delta) &= \{\theta \in \Theta : |\theta - \theta_0| \le \delta\}. \end{aligned} \tag{4.9}$$

Because of B2, for every positive ε there exists positive δ such that

$$\begin{aligned} Q_n\left[N_n \supset S(\delta) \cap \hat{\Theta}_n\right] &\ge Q_n\left[r_n(\delta) \le 2^{-1}C_n\right] \\ &\ge Q_n\left[r_n(\delta) \le \varepsilon, \varepsilon \le 2^{-1}C_n\right] \\ &= Q_n\left[C_n^{-1} \le (2\varepsilon)^{-1}\right] + o(1). \end{aligned} \tag{4.10}$$

Hence, for every positive γ, there exists positive δ such that

$$Q_n\left[N_n \supset S(\delta) \cap \hat{\Theta}_n\right] \ge 1 - \gamma \tag{4.11}$$

for every n sufficiently large. It follows from (4.11) and (4.7) that

$$\begin{aligned} &\lim_{n\to\infty} Q_n\left[\inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in \hat{\Theta}_n\}\right. \\ &\qquad \left.= \inf\{\|\hat{P}_n - P_\theta\|_n : \theta \in N_n\}\right] = 1. \end{aligned} \tag{4.12}$$

Let $A_n = n^{1/2}\|P_{\hat{\theta}_n} - P_{\theta_0}\|$ and let

$$d_n = \max\left\{C_n^{-1}(4\|V_n\|_n + 2A_n), n^{1/2}|\hat{\theta}_n - \theta_0|\right\}. \tag{4.13}$$

In view of B2, B3 and B5 and the properties of V_n, $\{\mathbf{L}(d_n|Q_n)\}$ is tight. If $\theta \in N_n$,

$$\begin{aligned} \|\hat{P}_n - P_\theta\|_n &\ge \|\langle m_{\theta_0}, \theta - \theta_0\rangle\|_n - 2^{-1}C_n|\theta - \theta_0| - \|\hat{P}_n - P_{\theta_0}\|_n \\ &\ge 2^{-1}C_n|\theta - \theta_0| - n^{-1/2}\|V_n\|_n, \end{aligned} \tag{4.14}$$

the second inequality following from B3. Thus, from (4.14) and the definition (4.13),

$$\|\hat{P}_n - P_\theta\|_n \ge \|\hat{P}_n - P_{\hat{\theta}_n}\|_n, \tag{4.15}$$

whenever $\theta \in N_n \cap S(n^{-1/2}d_n)$. Note that, by (4.11),

$$\lim_{n\to\infty} Q_n\big[N_n \supset S(n^{-1/2}d_n) \cap \hat{\Theta}_n\big] = 1. \tag{4.16}$$

In view of (4.15) and (4.13),

$$\begin{aligned}\inf\{\|\hat{P}_n - P_\theta\|_n: \theta \in N_n\} &= \inf\{\|\hat{P}_n - P_\theta\|_n: \theta \in N_n \cap S(n^{-1/2}d_n)\}\\ &\geq \inf\{\|\hat{P}_n - P_\theta\|_n: \theta \in S(n^{-1/2}d_n) \cap \hat{\Theta}_n\}.\end{aligned} \tag{4.17}$$

From this and (4.16) it follows that

$$\begin{aligned}\lim_{n\to\infty} Q_n\big[&\inf\{\|\hat{P}_n - P_\theta\|_n: \theta \in N_n\}\\ &= \inf\{\|\hat{P}_n - P_\theta\|_n: \theta \in S(n^{-1/2}d_n) \cap \hat{\Theta}_n\}\big] = 1.\end{aligned} \tag{4.18}$$

Combining (4.12) with (4.18) yields

$$\begin{aligned}&\inf\{n^{1/2}\|\hat{P}_n - P_\theta\|_n: \theta \in \hat{\Theta}_n\}\\ &\quad= \inf\{n^{1/2}\|\hat{P}_n - P_\theta\|_n: \theta \in S(n^{-1/2}d_n) \cap \hat{\Theta}_n\} + o_{Q_n}(1)\\ &\quad= \inf\{\|V_n - n^{1/2}\langle m_{\theta_0}, \theta - \theta_0\rangle\|_n: \theta \in S(n^{-1/2}d_n) \cap \hat{\Theta}_n\} + o_{Q_n}(1),\end{aligned} \tag{4.19}$$

the second line using B2.

Next, observe that if $\theta \in \bar{S}(n^{-1/2}d_n) \cap \hat{\Theta}_n$, then in view of B3, B2 and (4.13),

$$\begin{aligned}&\|V_n - n^{1/2}\langle m_{\theta_0}, \theta - \theta_0\rangle\|_n\\ &\quad\geq n^{1/2}\|\langle m_{\theta_0}, \theta - \theta_0\rangle\|_n - \|V_n\|_n\\ &\quad\geq C_n n^{1/2}|\theta - \theta_0| - \|V_n\|_n\\ &\quad> 3\|V_n\|_n + 2A_n\\ &\quad\geq \|V_n - n^{1/2}(P_{\hat{\theta}_n} - P_{\theta_0})\|_n\\ &\quad= \|V_n - n^{1/2}\langle m_{\theta_0}, \hat{\theta}_n - \theta_0\rangle\|_n + o_{Q_n}(1)\\ &\quad\geq \inf\{\|V_n - n^{1/2}\langle m_{\theta_0}, \theta - \theta_0\rangle: \theta \in S(n^{-1/2}d_n) \cap \hat{\Theta}_n\} + o_{Q_n}(1).\end{aligned} \tag{4.20}$$

Hence, from (4.3), (4.19) and (4.20)

$$\begin{aligned}\hat{T}_n &= \inf\{\|V_n - n^{1/2}\langle m_{\theta_0}, \theta - \theta_0\rangle\|_n: \theta \in \hat{\Theta}_n\} + o_{Q_n}(1)\\ &= \inf\{\|W_n - n^{1/2}\langle m_{\theta_0}, \theta - \theta_n\rangle\|_n: \theta \in \hat{\Theta}_n\} + o_{Q_n}(1),\end{aligned} \tag{4.21}$$

where $W_n = n^{1/2}(\hat{P}_n - P_{\theta_n})$.

Let

$$G_n(u) = \|W_n - \langle m_{\theta_0}, u\rangle\|_n \tag{4.22}$$

and let $\hat{v}_n$ be the empirical measure of the recentered search sample $\{n^{1/2}(\theta - \theta_n): \theta \in \hat{\Theta}_n\}$. Evidently, from (4.21),

$$\hat{T}_n = \operatorname*{ess\,inf}_{\hat{v}_n} G_n(u) + o_{Q_n}(1). \tag{4.23}$$

Since $\hat{v}_n$ is the empirical measure of the $\{n^{1/2}(\theta^*_{n,j} - \hat{\theta}_n) + n^{1/2}(\hat{\theta}_n - \theta_n): 1 \le j \le j_n\}$ and the value $n^{1/2}(\hat{\theta}_n - \theta)$, it follows from B5 and Theorem 2.1 of Beran and Millar (1987) that $\rho(\hat{v}_n, v) \to 0$ in Q_n-probability, where v is the distribution $\mu_{\theta_0} * \mu_{\theta_0}$. Moreover, $\{W_n\}$ converges weakly under $\{Q\}$ to W_{θ_0}, as random elements of L_∞.

Let γ be the uniform distribution on S_q and let

$$G(u) = \operatorname*{ess\,sup}_{\lambda} \sup_{t} |W_{\theta_0} - \langle m_{\theta_0}, u\rangle|.$$

By Wichura (1970), there exist versions of $\{W_n, \hat{v}_n\}$ and of W_{θ_0} such that $\|W_n - W_{\theta_0}\| \to 0$ a.s. and $\hat{v}_n \to v$ a.s. For these versions, the corresponding versions of $\{G_n(u)\}$ and $\{G(u)\}$ converge uniformly on compact balls in R^d. Moreover, the distribution ν has full support in R^d, because of B5. Thus, by Lemma 4.1 of Beran and Millar (1987), it follows that

$$\operatorname*{ess\,inf}_{\hat{v}_n} G_n(u) \to \inf_u G(u) \quad \text{a.s.,} \tag{4.24}$$

for the special versions.

Observe that

$$\mathbf{L}\Big[\inf_u G(u)\Big] = \mathbf{L}\Big[\inf_u \|W_{\theta_0} - \langle m_{\theta_0}, u\rangle\|\Big], \tag{4.25}$$

because of (2.14) and (2.15) in Beran and Millar (1986) and the assumptions on m_{θ_0} in B2. Proposition 2.3 follows from (4.23), (4.24) and (4.25). □

Proof of Proposition 2.4. Let $\{v_n\}$ be probabilities on R^d which converge weakly to a probability v, which has full support on R^d. Let $\{w_n \in L_\infty\}$ converge in norm to $w \in L_\infty$. Let $\mathbf{z}_n = (z_1, \dots, z_{j_n})$ be the first j_n elements of a sequence $\{z_j \in S_q\}$.

Define

$$\begin{aligned}\Lambda_n(w_n, v_n, \mathbf{z}_n) = \operatorname*{ess\,inf}_{v_n} \sup_{1\le j\le j_n} \sup_t |w_n(z_j, t) \\ - n^{1/2}(P_{\theta_n + n^{-1/2}u} - P_{\theta_n})(z_j, t)|.\end{aligned} \tag{4.26}$$

The essential infimum is taken over $u \in R^d$. Let

$$\xi_n(\theta_n, v_n, \mathbf{z}_n) = \mathbf{L}\big[\Lambda_n(W_n, v_n, \mathbf{z}_n) | P^n_{\theta_n}\big], \tag{4.27}$$

where W_n is the empirical process $n^{1/2}(\hat{P}_n - P_{\theta_n})$ defined in Section 2.

Slightly modified, the argument for Proposition 2.3 establishes

$$\begin{aligned}\Lambda_n(w_n, v_n, s_n) \to_{Q_n} \operatorname*{ess\,inf}_{v} \operatorname*{ess\,sup}_{\lambda} \sup_t |w - \langle m_{\theta_0}, u\rangle| \\ = \Lambda(w, v), \quad \text{say,}\end{aligned} \tag{4.28}$$

where λ is the uniform distribution on S_q. By the discussion at the end of the proof for Proposition 2.3, $\mathbf{L}[\Lambda(W_{\theta_0}, v)]$ coincides with $H(\theta_0)$. It follows from (4.28) and the independence of $\{W_n\}$, $\{\mathbf{s}_n\}$, by Lemma 2.2(ii) of Beran and Millar

(1987), that

$$\rho\left[\xi_n(\theta_n, v_n, \mathbf{s}_n), H(\theta_0)\right] \to 0 \quad \text{in } Q_n\text{-probability.} \tag{4.29}$$

Let $\{\hat{v}_n\}$, v be as in the proof of Proposition 2.3. It follows from (4.29) and the independence of $\{(\hat{\theta}_n, \hat{v}_n)\}, \{\mathbf{s}_n\}$, by Lemma 2.2(i) of Beran and Millar (1987), that

$$\rho\left[\xi_n\left(\hat{\theta}_n, \hat{v}_n, \mathbf{s}_n\right), H(\theta_0)\right] \to 0 \quad \text{in } Q_n\text{-probability.} \tag{4.30}$$

The conditional bootstrap distribution $\hat{H}_n$ is the empirical distribution of a random sample of size m_n drawn from the random measure $\xi_n(\hat{\theta}_n, \hat{v}_n, \mathbf{s}_n)$. Equation (2.28) is immediate from (4.30) and Theorem 2.1 of Beran and Millar (1987). Proposition 2.4 follows from Proposition 2.3 and (2.28). □

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.

BERAN, R. (1986). Simulated power functions. *Ann. Statist.* **14** 151–173.

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

BERAN, R. and MILLAR, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14** 431–443.

BERAN, R. and MILLAR, P. W. (1987). Stochastic estimation and testing. *Ann. Statist.* **15** 1131–1154.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

DWASS, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Statist.* **28** 181–187.

EFRON, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82** 171–185.

GAENSSLER, P. (1986). Bootstrapping empirical measures indexed by Vapnik–Chervonenkis classes of sets. *Probab. Theory Math. Statist.* **1** 467–481.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–330.

HALL, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14** 1431–1452.

LE CAM, L. (1983). A remark on empirical measures. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 305–327. Wadsworth, Belmont, Calif.

POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70.

PYKE, R. (1984). Comment on "Some limit theorems for empirical processes," by E. Giné and J. Zinn. *Ann. Probab.* **12** 996–997.

ROMANO, J. R. (1988). A bootstrap revival of some nonparametric tests. *J. Amer. Statist. Assoc.* **83** 698–708.

SHEEHY, A. and WELLNER, J. A. (1986). Uniformity in P of some limit theorems for empirical measures and processes. Unpublished.

WICHURA, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. *Ann. Math. Statist.* **41** 284–291.

WOLFOWITZ, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28** 75–88.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720