Ann. Inst. Statist. Math.
Vol. 49, No. 1, 1–24 (1997)

DIAGNOSING BOOTSTRAP SUCCESS*

RUDOLF BERAN

Department of Statistics, University of California, Berkeley, CA 94720-3860, U.S.A.

(Received August 8, 1995; revised February 13, 1996)

Abstract. We show that convergence of intuitive bootstrap distributions to the correct limit distribution is equivalent to a local asymptotic equivariance property of estimators and to an asymptotic independence property in the bootstrap world. The first equivalence implies that bootstrap convergence fails at superefficiency points in the parameter space. However, superefficiency is only a sufficient condition for bootstrap failure. The second equivalence suggests graphical diagnostics for detecting whether or not the intuitive bootstrap is trustworthy in a given data analysis. Both criteria for bootstrap convergence are related to Hájek's (1970, *Zeit. Wahrscheinlichkeitsth.*, **14**, 323–330) formulation of the convolution theorem and to Basu's (1955, *Sankhyā*, **15**, 377–380) theorem on the independence of an ancillary statistic and a complete sufficient statistic.

Key words and phrases: Bootstrap convergence, local asymptotic equivariance, local asymptotic sufficiency, asymptotic independence, superefficiency points, convolution theorem.

1. Introduction

Behind the bootstrap lies a grand dream: freeing statistical inference from ingenious analytical approximations to sampling distributions that must be devised case-by-case. Earlier this century, the two-sample Behrens-Fisher problem forced frequentist statistical theory to derive approximations based on asymptotic expansions. Welch's (1937) clever solution to the problem relied on his discovery that the first two terms in the Edgeworth expansion for the Behrens-Fisher statistic coincide with those for a certain t-distribution, whose degrees-of-freedom depend upon the estimable ratio of the two variances. Using this t-distribution to derive critical values eased the construction of tables and finessed the poor behavior of Edgeworth approximations in their tails.

The normal-model parametric bootstrap solution to the Behrens-Fisher problem has the same good second-order asymptotics as Welch's solution (Beran (1988)). Unlike the analytical approach, the bootstrap solution is highly intuitive and may be carried out easily on a computer. Moreover, it extends readily to

* This research was supported by National Science Foundation grant DMS 9224868.

several samples, to non-normal models, and to other roots that mimic the structure of the t-statistic asymptotically. Hall (1992) and Mammen (1992) described systematically such asymptotic theory for the bootstrap. Provable success in such settings has encouraged the use of intuitive bootstrap inference in data analyses that may involve smoothing, or model-selection, or some form of shrinkage (cf. Efron and Tibshirani (1993), Hjorth (1994)). However, theoretical support for natural bootstrap procedures in these more complex contexts is often lacking. Of course, it is precisely in sophisticated data analyses that we would most like the bootstrap to work reliably and automatically.

1.1 *Correct convergence of bootstrap estimators*

Suppose that the sample X_n consists of n random elements $(X_{n,1}, X_{n,2}, \ldots, X_{n,n})$ whose joint distribution is $P_{\theta,n}$. The unknown parameter θ lies in a space Θ that need not be finite-dimensional. Of interest is the distribution $H_n(\theta)$ of a root $R_n(X_n, \theta)$ under this model. By *root*, we mean a function of the sample and of the parameter that takes values in a complete separable metric space. Roots include pseudo-pivots for confidence sets (possibly simultaneous), loss functions, or test statistics. Let $\hat{\theta}_n = \hat{\theta}_n(X_n)$ denote an estimator of θ. The corresponding plug-in or bootstrap estimator of $H_n(\theta)$ is then $H_n(\hat{\theta}_n)$. As defined, this bootstrap distribution is a random probability measure. The interpretation of $H_n(\hat{\theta}_n)$ as a conditional distribution leads to useful Monte Carlo approximations.

PROPOSITION 1.1. *Suppose that Θ is metric, $\theta_0 \in \Theta$, and*

a) *for every sequence $\{\theta_n \in \Theta : n \geq 1\}$ that converges to θ_0, $H_n(\theta_n) \Rightarrow H(\theta_0)$;*

b) *$\hat{\theta}_n \to \theta_0$ in $P_{\theta_0,n}$-probability as $n \to \infty$.*

Then $H_n(\hat{\theta}_n) \Rightarrow H(\theta_0)$ in $P_{\theta_0,n}$-probability.

This simple result, a version of Theorem 1 in Beran (1984), provides a template for arguing that the limit in $P_{\theta_0,n}$-probability of the bootstrap distributions $\{H_n(\hat{\theta}_n)\}$ coincides with the limit of the sampling distributions $\{H_n(\theta_0)\}$. Such reasoning, or the related equicontinuity consideration in Bickel and Freedman (1981), has been used frequently in the bootstrap literature.

Most important is the choice of the metric on the parameter space Θ. Condition a) is easier to satisfy if the metric is strong, so that the class of convergent sequences $\{\theta_n\}$ is relatively small. Condition b) is easier to meet if the metric is weak, so that the class of consistent estimators is relatively large. Thus, proving bootstrap convergence by means of Proposition 1.1 usually requires a clever metrization of the parameter space and a carefully crafted estimator $\hat{\theta}_n$. The details depend strongly on the mathematical structure of the problem. Because of this, Proposition 1.1 does not support the idea that intuitive bootstrap estimates will "usually" converge.

Another aspect of bootstrap convergence was pointed out by Putter (1994) in his doctoral thesis. Recall that a set in a topological space is nowhere dense if its closure has empty interior. A set is said to be of (Baire) category I if it is a countable union of nowhere dense sets. A set is of category II if it is not of category I.

PROPOSITION 1.2. *Suppose that Θ is complete metric and*

a) $H_n(\theta) \Rightarrow H(\theta)$ *as* $n \to \infty$, *for every* $\theta \in \Theta$;

b) $H_n(\theta)$ *is continuous in* θ, *in the topology of weak convergence, for every* $n \geq 1$;

c) $\hat{\theta}_n \to \theta$ *in* $P_{\theta,n}$*-probability, for every* $\theta \in \Theta$.

Then, there exists a set E *of category I such that* $H_n(\hat{\theta}_n) \Rightarrow H(\theta)$ *in* $P_{\theta,n}$*-probability, for every* $\theta \in \Theta - E$. *Moreover, the set* $\Theta - E$ *is of category II.*

Proposition 1.2, a variant of Theorem 2.2.2 in Putter (1994), ultimately stems from mathematical results on category and on interchanging the order of two limits. We return to this point in Section 4. Topologically speaking, the set of θ at which the bootstrap distributions in Proposition 1.2 converge correctly is "large"; and the set of points where convergence is not assured is "small". In general, sets of category I are not easily visualized. For example, when Θ is a subset of a Euclidean space, a set of category I can have large Lebesgue measure. If we impose additional assumptions on the model in this finite dimensional case, then the exceptional set E in Proposition 1.2 does have Lebesgue measure zero (cf. Section 3 and Theorem 2.6.1 in Putter (1994)). While encouraging, such results do not identify, in a useful way, those values of θ at which convergence of the bootstrap distributions fails.

1.2 *Superefficiency and bootstrap failure*

At first glance, Proposition 1.2 suggests that the failure of bootstrap distributions to converge correctly is an unusual event, of no great practical importance. The following examples cast doubt on this impression.

Example 1. Suppose that the $\{X_{n,i}\}$ are iid random variables, each distributed according to $N(\theta, 1)$. Here the parameter dimension $k = 1$. Let $\bar{X}_n$ be the sample mean and let $T_{n,H}$ be the Hodges estimator of θ, given by

$$(1.1) \qquad T_{n,H} = \begin{cases} b\bar{X}_n & \text{if} \quad |\bar{X}_n| \leq n^{-1/4} \\ \bar{X}_n & \text{otherwise} \end{cases},$$

where $b^2 < 1$ (LeCam (1953)). Observe that, when b is zero, $T_{n,H}$ is a model selection estimator that chooses between fitting the $N(0,1)$ and $N(\theta,1)$ models on the basis of the data.

The distribution $H_n(\theta)$ of the root $n^{1/2}(T_{n,H} - \theta)$ converges weakly to $N(0,1)$ when $\theta \neq 0$ and to $N(0, b^2)$ when $\theta = 0$. Let $\bar{X}_n$ denote the sample mean. The bootstrap distribution $H_n(\bar{X}_n)$ converges weakly in probability to $N(0,1)$, the correct limit, when $\theta \neq 0$. However, if $\theta = 0$, then $H_n(\bar{X}_n)$ converges *in distribution*, as a random element of the space of all probability measures on the real line metrized by weak convergence, to the random probability measure $N((b-1)Z, b^2)$. Here Z has a standard normal distribution. This failure in bootstrap convergence at the single point $\theta = 0$ is an instance of Theorem 2.3 and illustrates Proposition 1.2.

The asymptotic risk of the Hodges estimator is

$$\lim_{n\to\infty} nE_\theta(T_{n,H}-\theta)^2 = \begin{cases} b^2 & \text{if } \theta = 0 \\ 1 & \text{if } \theta \neq 0 \end{cases} \tag{1.2}$$

while the Fisher information bound is 1. The origin is thus a point of superefficiency for the Hodges estimator. For fixed n, the risk of $T_{n,H}$ is less than 1 in a neighborhood of the origin, then rises steeply above one, and subsequently drops slowly towards 1 as $|\theta|$ tends to infinity (cf. Lehmann (1983), Chapter 6). The neighborhood of improved risk narrows as n increases, so that the asymptotic picture is (1.2). At finite n, the Hodges estimator has larger risk than the sample mean for most values of θ. Such poor behavior in risk near any point of superefficiency is characteristic of one-dimensional estimators (LeCam (1953), Hájek (1972)).

It is noteworthy that the failure in convergence of the bootstrap distribution $H_n(\bar{X}_n)$ occurs at $\theta = 0$, the value at which the Hodges estimator improves upon the sample mean.

Example 2. For higher dimensional estimators, superefficiency at a point need not entail poor risk nearby. Let I_k denote the $k \times k$ identity matrix. Suppose that the $\{X_{n,i}\}$ are iid random k-vectors, each distributed according to $N_k(\theta, I_k)$, where $\theta \in R^k$. This is a simple model for n repeated observations on a discrete time series measured at k time points. The goal is to estimate the unknown signal θ. For $k \geq 4$, consider the Stein estimator that shrinks each component of the sample mean $\bar{X}_n$ towards the average of all nk numbers in the sample. Let e denote the vector in R^k whose components each equal 1. For every vector x, let $m(x)$ denote the average of the components of x. Define

$$T_{n,S} = m(\bar{X}_n)e + [1-(k-3)/(n|\bar{X}_n - m(\bar{X}_n)e|^2)](\bar{X}_n - m(\bar{X}_n)e). \tag{1.3}$$

This estimator is a sharp early example of what are now called regularization methods for signal recovery (cf. Titterington (1991)).

When the components of θ are not all equal, the distribution $H_n(\theta)$ of the root $n^{1/2}(T_{n,S}-\theta)$ converges weakly to $N(0, I_k)$. The bootstrap distribution $H_n(\bar{X}_n)$ converges weakly in probability to the same limit. To describe what happens when the components of θ are equal, let Z_k be a random vector with standard normal distribution on R^k, let $Z_k(h) = Z_k + h$, and define the family of distributions

$$\pi(h) = \mathcal{L}\{Z_k - (k-3)[Z_k(h) - m(Z_k(h))e]/|Z_k(h) - m(Z_k(h))e|^2]\}, \quad h \in R^k. \tag{1.4}$$

When the components of θ are equal, $H_n(\theta)$ equals $\pi(0)$ for every n and so converges weakly to $\pi(0)$. However, the bootstrap distribution $H_n(\bar{X}_n)$ converges in distribution, as a random element of the space of all probability measures on the real line metrized by weak convergence, to the random probability measure $\pi(Z_k)$. This result also exemplifies Theorem 2.3. The values of θ at which bootstrap convergence fails form a one-dimensional subspace of R^k, a set of category I and of Lebesgue measure zero.

For $k \geq 4$, the asymptotic risk $\lim_{n\to\infty} nE_\theta|T_{n,S} - \theta|^2$ equals the information bound k whenever the components of θ are not all equal and is strictly smaller than k when they are equal. The points of superefficiency of this Stein estimator thus coincide with the subspace where the bootstrap distribution $H_n(\bar{X}_n)$ fails to converge correctly. Unlike Example 1, the estimator $T_{n,S}$ strictly dominates $\bar{X}_n$ over the entire parameter space, substantially so when k is much larger than n (cf. Lehmann (1983), p. 305). The superefficiency points detected by asymptotics in n while k is held fixed are only a ghost of the risk function at finite n and k.

The possibility of superefficiency is at heart of modern estimation theory. Sophisticated signal estimators or model selection estimators implicitly create points of superefficiency, though often without articulating this as goal; and they do so because, as in Example 2, superefficiency points can reduce risk over the entire parameter space when k is not too small. That points of superefficiency in a Euclidean parameter space form a Lebesgue null set (cf. LeCam (1953) and Section 3) does not make them unimportant to the task of improved estimation. Unfortunately, as we will prove in this paper, superefficiency is a sufficient condition for failure of intuitive bootstrap distributions to converge correctly. In this sense, intuitive bootstrap methods break down precisely when we need them most.

1.3 *Bootstrap diagnostics*

To guard against possibly misleading bootstrap distributions in data analyses, it is desirable to characterize theoretically those situations where intuitive bootstrapping fails to converge correctly. Equally important are direct diagnostic methods for detecting bootstrap failure in a specific data analysis. Because θ is unknown, we cannot say with certainty that the bootstrap fails in a particular data analysis. However, as in traditional regression diagnostics, it is reasonable to seek data-based indicators of possible bootstrap failure.

The main results of this paper are both structural and diagnostic. First, we show that, in locally asymptotically normal parametric models, correct convergence of intuitive bootstrap distributions is equivalent to a local asymptotic equivariance property of estimators. This result defines what we mean by the phrase "intuitive bootstrap" and implies that superefficiency is a sufficient (but not necessary) condition for failure of the intuitive bootstrap. Second, we prove that correct convergence of the intuitive bootstrap is equivalent to an asymptotic independence property in the bootstrap world. Because this asymptotic independence property takes place in the world of the data, it suggests diagnostic plots for detecting bootstrap failure. Third, we extend both results to nonparametric bootstrapping when the parent distribution has finite support, but the cardinality and values of the support points are unknown. The theory of the paper is deeply related to Hájek's (1970) formulation of the convolution theorem as well as to Basu's (1955) theorem on the independence of a complete sufficient statistic and an ancillary statistic.

2. Bootstrap convergence

We will characterize successful bootstrap convergence in a setting more structured than that of the Introduction. The assumptions to be made imply results substantially more detailed than Propositions 1.1 and 1.2. It is useful to treat parametric and nonparametric bootstrapping separately.

2.1 *Parametric bootstrap*

We begin with assumptions on the model. Suppose that the sample X_n consists of n iid random elements, each having distribution P_{θ_0}. The parameter space Θ is an open subset of R^k. The distribution $P_{\theta_0,n}$ of X_n is now the n-fold product of P_{θ_0}. For every $h \in R^k$ such that $\theta_0 + n^{-1/2}h \in \Theta$, let $P^c_{\theta_0+n^{-1/2}h,n}$ denote the absolutely continuous part of $P_{\theta_0+n^{-1/2}h,n}$ with respect to $P_{\theta_0,n}$.

DEFINITION 2.1. Let $L_n(h,\theta_0)$ denote the log-likelihood ratio of $P^c_{\theta_0+n^{-1/2}h,n}$ with respect to $P_{\theta_0,n}$. The model $\{P_{\theta,n} : \theta \in \Theta\}$ is locally asymptotically normal (LAN) at θ_0 if there exist a random $k \times 1$ vector $Y_n(\theta_0)$ and a nonsingular $k \times k$ matrix $I(\theta_0)$ such that, under $P_{\theta_0,n}$,

$$L_n(h_n,\theta_0) = h'Y_n(\theta_0) - 2^{-1}h'I(\theta_0)h + o_p(1) \tag{2.1}$$

for every $h \in R^k$ and for every sequence $\{h_n \in R^k\}$ converging to h; and $\mathcal{L}[Y_n(\theta_0) \mid P_{\theta_0,n}] \Rightarrow N(0, I(\theta_0))$.

This definition, which implicitly determines the Fisher information matrix $I(\theta_0)$, was introduced by LeCam (1960). The local uniformity is important for our bootstrap discussion. When (2.1) holds at every $\theta_0 \in \Theta$, we can modify the score function $Y_n(\cdot)$ so as to achieve both (2.1) and

$$Y_n(\theta_0 + n^{-1/2}h_n) = Y_n(\theta_0) - I(\theta_0)h + o_p(1) \tag{2.2}$$

under $P_{\theta_0,n}$ (see LeCam (1969), p. 68). We will assume (2.2) hereafter. The LAN property is possessed by many classical models, including smoothly parameterized exponential families. Hájek and Šidák (1967) and Hájek (1972) provided convenient sufficient conditions for LAN. For an LAN model, the log-likelihood ration behaves asymptotically like the log-likelihood ratio of $N(h, I^{-1}(\theta_0))$ with respect to $N(0, I^{-1}(\theta_0))$. We expect that the limit likelihood ratio can be used as a good approximation to the actual likelihood ratio—an idea that the preceding references support.

Let $\tau = \tau(\theta)$ be a parametric function that takes values in R^m, where $m \leq k$. Let $T_n = T_n(X_n)$ be any estimator of τ. We begin with the root $R_n(X_n,\theta) = n^{1/2}(T_n - \tau)$, whose distribution under $P_{\theta,n}$ is $H_n(\theta)$. We will characterize the set $\mathcal{C}$ of θ values such that both $H_n(\theta) \Rightarrow H(\theta)$ and $H_n(\hat{\theta}_n) \Rightarrow H(\theta)$ in $P_{\theta,n}$-*probability in other words, the values of* θ *at which the bootstrap distribution* converges correctly.

Motivating this study are the following statistical considerations. Loss functions and some roots used to construct confidence sets have the form $g[n^{1/2}(T_n -$

$\tau(\theta))]$, where g is real-valued. Let $H_{n,g}(\theta) = \mathcal{L}\{g[n^{1/2}(T_n - \tau(\theta))] \mid P_{\theta,n}\}$ and let $H_g(\theta) = \mathcal{L}\{g(R)\}$, where R has distribution $H(\theta)$. Assume that the discontinuity points of g form a null set with respect to $H(\theta)$. Then, for every $\theta \in \mathcal{C}$, both $H_{n,g}(\theta) \Rightarrow H_g(\theta)$ and $H_{n,g}(\hat{\theta}_n) \Rightarrow H_g(\theta)$ in $P_{\theta,n}$-probability. The set $\mathcal{C}$ thus describes points where the bootstrap distribution of the root $g[n^{1/2}(T_n - \tau(\theta))]$ converges correctly. For an analysis of studentized roots, which are not covered by this discussion, see Subsection 2.5.

The next definition isolates an approximate equivariance property possessed by many classical estimator sequences $\{T_n : n > 1\}$ in LAN models.

DEFINITION 2.2. Suppose that $H_n(\theta_0) \Rightarrow H(\theta_0)$ as $n \to \infty$. The estimators $\{T_n\}$ of $\tau(\theta)$ are locally asymptotically equivariant (LAE) at θ_0 if, for every $h \in R^k$ and every sequence $\{h_n \in R^k\}$ converging to h,

$$H_n(\theta_0 + n^{-1/2}h_n) \Rightarrow H(\theta_0). \tag{2.3}$$

This definition technically strengthens Hájek's (1970) concept of *regularity*, which asserts only that $H_n(\theta_0 + n^{-1/2}h) \Rightarrow H(\theta_0)$ for every $h \in R^k$. Suppose that τ is Fréchet differentiable at θ_0 and that the derivative $\nabla\tau(\theta_0)$, which is the $m \times k$ matrix $\{\partial\tau_i(\theta_0)/\partial\theta_{0,j}\}$, has full rank m. Let

$$\Sigma_\tau(\theta_0) = \nabla\tau(\theta_0)I^{-1}(\theta_0)\nabla'\tau(\theta_0), \tag{2.4}$$

where $'$ denotes matrix transpose. An important consequence of regularity, and so of LAE, is the convolution theorem: there exists a distribution $D(\theta_0)$ such that $H(\theta_0) = D(\theta_0) * N(0, \Sigma_\tau(\theta_0))$. Proved by Hájek (1970) and, in less generality, by Inagaki (1970), the convolution theorem is closely tied to the question of successful bootstrap convergence. The next theorem supports this assertion.

Let $\{T_{n,E}\}$ be any sequence of estimators such that, under $P_{\theta_0,n}$,

$$T_{n,E} = \tau(\theta_0) + n^{-1/2}\nabla\tau(\theta_0)I^{-1}(\theta_0)Y_n(\theta_0) + o_p(n^{-1/2}). \tag{2.5}$$

Suppose that $\{h_n \in R^k\}$ is any sequence converging to h. By (2.2), differentiability of τ, and contiguity (LeCam (1960)), the limiting distribution of $n^{1/2}(T_{n,E} - \tau(\theta_0 + n^{-1/2}h_n))$ under $P_{\theta_0+n^{-1/2}h_n,n}$ is $N(0, \Sigma_\tau(\theta_0))$. Thus, the estimators $\{T_{n,E}\}$ are LAE at θ_0 and have the least dispersed limit distribution that the convolution theorem permits. For this reason, estimators that satisfy (2.5) are called asymptotically *efficient*. We remark that maximum likelihood estimators in smooth exponential families and one-step MLE's in LAN models are typical constructions of $T_{n,E}$ (cf. LeCam (1956, 1969)).

Recall that, in complex variables, a set $U \subset R^k$ is called a *uniqueness set* if any analytic function defined on an open connected subset of C^k that contains U is uniquely determined by its values on U. For example, U could be R^k, or a k-dimensional box in R^k with edges parallel to the coordinate axes, or a dense subset of these.

In what follows, let $\hat{\theta}_n$ be the estimator of θ that is used to construct the bootstrap distribution $H_n(\hat{\theta}_n)$ for $n^{1/2}(T_n - \tau(\theta))$. Let $J_n(\theta) = \mathcal{L}[n^{1/2}(\hat{\theta}_n - \theta) \mid P_{\theta,n}]$ and let

$$K_n(\theta) = \mathcal{L}[n^{1/2}(T_n - T_{n,E}), Y_n(\theta) \mid P_{\theta,n}]. \tag{2.6}$$

THEOREM 2.1. *Suppose that the model $\{P_{\theta,n} : \theta \in \Theta\}$ is LAN at θ_0 and τ is Fréchet differentiable at θ_0 with derivative $\nabla\tau(\theta_0)$ of full rank. Suppose that $H_n(\theta_0) \Rightarrow H(\theta_0)$, that $J_n(\theta_0) \Rightarrow J(\theta_0)$, and that the support of $J(\theta_0)$ contains a uniqueness set. Then the following statements are equivalent:*

a) *$H_n(\hat{\theta}_n) \Rightarrow H(\theta_0)$ in $P_{\theta_0,n}$-probability.*

b) *$K_n(\hat{\theta}_n) \Rightarrow D(\theta_0) \times N(0, I(\theta_0))$ in $P_{\theta_0,n}$-probability for some distribution $D(\theta_0)$ such that $H(\theta_0) = D(\theta_0) * N(0, \Sigma_\tau(\theta_0))$.*

c) *$\{T_n\}$ is LAE at θ_0 with limit distribution $H(\theta_0)$.*

Remarks. For standard estimators $\hat{\theta}_n$, the support of $J(\theta_0)$ is the whole of R^k. Theorem 2.1 thus defines and addresses what we have called the "intuitive bootstrap". The interpretation of part b) is discussed after Theorem 3.1. Its use in diagnosing bootstrap failure is taken up in Subsection 2.3.

Part c) of Theorem 2.1 implies part a) regardless of the support of $J(\theta_0)$ (cf. the proof in Section 4). Thus, the parametric bootstrap distribution of an asymptotically efficient estimator—that is, any estimator satisfying (2.5)—always converges correctly in probability when $J_n(\theta_0) \to J(\theta_0)$. It is parametric bootstrap distributions of estimators superefficient at θ_0 that fail to converge correctly there unless the support of $J(\theta_0)$ is the origin. This point will be developed in Section 3 and Subsection 2.4.

Let $\{\theta_n \in \Theta\}$ be any sequence such that $n^{1/2}(\theta_n - \theta_0)$ converges to a finite limit. Theorem 2.1 remains valid if $P_{\theta_0,n}$ is replaced by $P_{\theta_n,n}$ in parts a) and b). This is so because $\{P_{\theta_n,n}\}$ and $\{P_{\theta_0,n}\}$ are contiguous.

Example 2. (continued) The normal location model in this example is LAN, with $Y_n(\theta_0) = n^{1/2}(\bar{X}_n - \theta_0)$ and $I(\theta_0) = I_k$; and the best LAE estimator $T_{n,E}$ is $\bar{X}_n$. We may therefore apply Theorem 2.1 to the Stein estimator (1.3). Let $\theta_n = \theta_0 + n^{-1/2}h_n$, where $\{h_n \in R^k\}$ is any sequence converging to h. When the components of θ_0 are not all equal, then $H_n(\theta_n)$ converges weakly to $N(0, I_k)$, whatever the choice of h. However, when the components of θ_0 are equal, the weak limit of $H_n(\theta_n)$ is the distribution $\pi(h)$ defined in (1.4). Consequently, the Stein estimator is not LAE when the components of θ_0 are all equal.

Consider the bootstrap distribution $H_n(\bar{X}_n)$ for the Stein estimator. The weak limit of $J_n(\theta_0) = \mathcal{L}[n^{1/2}(\bar{X}_n - \theta_0) \mid P_{\theta_0,n}]$ is standard normal on R^k and so has full support. Parts a) and c) of the Theorem 2.1 thus confirm what we observed in the Introduction: the bootstrap distribution $H_n(\bar{X}_n)$ cannot converge correctly when the components of θ_0 are all equal.

By reasoning similar to the first paragraph above, we find that, when the components of θ_0 are not equal, part b) of the theorem holds with $D(\theta_0)$ being the point mass at the origin. However, when the components of θ_0 are all equal, then $K_n(\bar{X}_n)$ converges in distribution to the random probability measure $\lambda(Z_k)$, where, in the notation of (1.4),

$$\lambda(h) = \mathcal{L}\{-(k-3)[Z_k(h) - m(Z_k(h))e]/|Z_k(h) - m(Z_k(h))e|^2, Z_k\}. \tag{2.7}$$

Thus, the asymptotic independence in part b) breaks down at the points where the intuitive bootstrap fails to converge correctly.

2.2 *Nonparametric bootstrap*

The method used to prove Theorem 2.1—specifically the step from (4.5) to (4.6)—does not work when the parameter θ takes values in a Banach space. For this technical reason, we will study the nonparametric bootstrap not over the class of all distributions on a Euclidean space but over a rich subset of this to which our approach applies. Let Π denote the set of all distributions that have finite support on a given Euclidean space $\mathcal{E}$. Because neither the finite support set nor its cardinality are specified, Π is dense, under weak convergence, in the set of all distributions on $\mathcal{E}$. Suppose that the observations $\{X_{n,i}\}$ in the sample are iid, each distributed according to an unknown distribution $P \in \Pi$. Let $s = \{s_1, s_2, \ldots, s_d\}$ denote the support points of P, in lexicographical order; let $\theta = (\theta_1, \theta_2, \ldots, \theta_{d-1})$ be the first $d-1$ of the probabilities $\theta_j = P(\{s_j\})$; and let $\delta(s_j)$ denote the probability measure that puts unit mass at s_j. Then

$$P = \sum_{j=1}^{d} \theta_j \delta(s_j) = P(\theta, s, d). \tag{2.8}$$

We assume that the representation (2.8) is minimal in the sense that each θ_j lies strictly between 0 and 1. Because P is unknown, so are the values of θ, s, and d. The class of distributions Π reflects the common experience that measurements are made to a finite number of digits over a finite range, with the accuracy and range not known a priori.

For $1 \le j \le d$, let $\hat{\theta}_{n,j}$ be the proportion of observations $\{X_{n,i}\}$ that equal θ_j and let $\hat{\theta}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2}, \ldots, \hat{\theta}_{n,d-1})$. The empirical distribution of the sample is then $\hat{P}_n = P(\hat{\theta}_n, s, d)$. An important property of the empirical distribution $\hat{P}_n$ is that its support lies within the support of P.

For fixed d and s, let $L(s,d)$ denote the set of all signed measures $\nu(a,s,d) = \sum_{j=1}^{d} a_j \delta(s_j)$, where $a = (a_1, a_2, \ldots, a_{d-1})$ ranges over R^{d-1} and $a_d = -\sum_{j=1}^{d-1} a_j$. Metrize $L(s,d)$ by Euclidean distance on the argument a. Let $\text{diag}(\theta)$ denote the diagonal matrix formed by arranging the $d-1$ components of θ down the diagonal and let $W(P)$ be a random vector in R^{d-1} whose distribution is $N(0, \text{diag}(\theta) - \theta\theta')$. Under P^n, the random vectors $\{n^{1/2}(\hat{\theta}_n - \theta)\}$ converge weakly to $W(P)$. Correspondingly, the empirical processes $\{n^{1/2}(\hat{P}_n - P)\}$ converge weakly, as random elements of $L(s,d)$, to the Gaussian process $\nu(W(P), s, d)$.

Let T be a functional that maps Π into R^m and let $T_n = T_n(X_n)$ be an estimator of $T(P)$. The *nonparametric bootstrap* estimator of the sampling distribution $H_n(P) = \mathcal{L}[n^{1/2}(T_n - T(P)) \mid P^n]$ is $H_n(\hat{P}_n)$. In this nonparametric setting, the plug-in estimator $T(\hat{P}_n)$ for $T(P)$ plays the role of the best LAE estimator introduced in (2.5). In place of (2.6), define

$$K_n(P) = \mathcal{L}[n^{1/2}(T_n - T(\hat{P}_n)), n^{1/2}(\hat{P}_n - P) \mid P^n]. \tag{2.9}$$

In Theorem 2.3 below, we relate correct weak convergence of the nonparametric bootstrap distribution $H_n(\hat{P}_n)$ to the asymptotic structure of $K_n(\hat{P}_n)$. The Fréchet differentiability and LAE conditions of Subsection 2.1 are replaced as follows.

DEFINITION 2.3. The functional T is said to be partially Fréchet differentiable at $P_0 = P(\theta_0, s_0, d_0)$ if $T[P(\theta, s_0, d_0)]$ is Fréchet differentiable in θ at θ_0 with $m \times (d_0 - 1)$ derivative $\nabla T(P_0) = \{\partial T_i[P(\theta_0, s_0, d_0)]/\partial \theta_{0,j}\}$.

DEFINITION 2.4. Suppose that $H_n(P_0) \Rightarrow H(P_0)$ as $n \to \infty$. The estimators $\{T_n\}$ of $T(P)$ are partially LAE at $P_0 = P(\theta_0, s_0, d_0)$ if, for every $h \in R^{d-1}$ and every sequence $\{h_n \in R^{d-1}\}$ converging to h,

$$H_n[P(\theta_0 + n^{-1/2}h_n, s_0, d_0)] \Rightarrow H(P_0). \tag{2.10}$$

Let

$$\Sigma_T(P_0) = \nabla T(P_0)[\mathrm{diag}(\theta_0) - \theta_0\theta_0']\nabla' T(P_0). \tag{2.11}$$

When T is partially differentiable at P_0, the distribution of $n^{1/2}[T(\hat{P}_n) - T(P_0)]$ converges weakly under P_0^n to $N(0, \Sigma_T(P_0))$.

THEOREM 2.2. *Suppose that the functional $T(P)$ is partially Fréchet differentiable at $P_0 = P(\theta_0, s_0, d_0)$ with derivative $\nabla T(P_0)$ of full rank and that $H_n(P_0) \Rightarrow H(P_0)$. Then the following statements are equivalent:*

a) *$H_n(\hat{P}_n) \Rightarrow H(P_0)$ in P_0^n-probability.*

b) *$K_n(\hat{P}_n) \Rightarrow D(P_0) \times \mathcal{L}[\nu(W(P_0), s_0, d_0)]$ in P_0^n-probability for some distribution $D(P_0)$ such that $H(P_0) = D(P_0) * N(0, \Sigma_T(P_0))$.*

c) *$\{T_n\}$ is partially LAE at P_0 with limit distribution $H(P_0)$.*

Example 3. Suppose that the support points of each distribution $P \in \Pi$ are vectors in R^k. In the notation of (2.8), the mean of P is

$$\mu(P) = \sum_{j=1}^{d} \theta_j s_j. \tag{2.12}$$

Evidently, $\mu(P)$ is partially differentiable at every $P_0 \in \Pi$. Let $\sigma^2(P)$ denote the variance of P. By the triangular array central limit theorem, the plug-in estimator $\bar{X}_n = \mu(\hat{P}_n)$ of $\mu(P)$ is partially LAE at every $P_0 \in \Pi$, the limit distribution being $N(0, \sigma^2(P_0)I_k)$. Part c) of Theorem 2.2 implies that the nonparametric bootstrap distribution for $\mathcal{L}[n^{1/2}(\bar{X}_n - \mu(P_0)) \mid P_0^n]$ converges weakly in probability to the correct limit for every $P_0 \in \Pi$.

In the notation of Example 2, consider the *scaled* Stein estimator for $\mu(P)$,

$$T_{n,SS} = m(\bar{X}_n)e + [1 - (k-3)s_n^2/(n|\bar{X}_n - m(\bar{X}_n)e|^2)](\bar{X}_n - m(\bar{X}_n)e), \tag{2.13}$$

where $s_n^2 = n\sigma^2(\hat{P}_n)/(n-1)$ is the usual unbiased estimator of $\sigma^2(P)$. When $P_0 = P(\theta_0, s_0, d_0)$ is such that the components of $\mu(P_0)$ are not all equal, then $H_n(P_0) = \mathcal{L}[n^{1/2}(T_{n,SS} - \mu(P_0)) \mid P_0^n]$ is partially LAE at P_0 with limit distribution $N(0, \sigma^2(P_0)I_k)$.

When the components of $\mu(P_0)$ are equal, the situation is more complex. Let V_k be a $N(0, \sigma^2(P_0)I_k)$ random vector. For every $h \in R^{d_0-1}$, let $h_{d_0} = -\sum_{j=1}^{d_0-1} h_j$. Define the shifted random vector

$$V_k(h) = V_k + \sum_{j=1}^{d_0} s_{0,j}h_j \tag{2.14}$$

and the distribution

$$\nu(P_0, h) = \mathcal{L}\{V_k - (k-3)\sigma^2(P_0)[V_k(h) - m(V_k(h))e]/|V_k(h) - m(V_k(h))e|^2\}. \tag{2.15}$$

If the components of $\mu(P_0)$ are all equal, then $H_n[P_0(\theta_0 + n^{-1/2}h_n, s_0, d_0)]$ converges weakly to $\nu(P_0, h)$ for every sequence $\{h_n \in R^{d_0-1}\}$ that converges to h. Consequently, the scaled Stein estimators $\{T_{n,SS}\}$ are not partially LAE at such P_0.

By Theorem 2.2, the nonparametric bootstrap distribution $H_n(\hat{P}_n)$ for the scaled Stein estimator converges correctly in probability if and only if the components of the mean $\mu(P_0)$ are not all equal. The set of distributions $P_0 \in \Pi$ where bootstrap convergence fails is indexed by those values of θ_0 such that $\mu[P(\theta_0, s_0, d_0)]$ has equal components. This is a hyperplane in R^{d_0-1} of Lebesgue measure zero.

2.3 *Graphical diagnostics*

Is bootstrap inference credible in a given data analysis? A reasonable answer would consist in diagnostic plots philosophically akin to regression diagnostics. Theorem 2.1 suggests how to do this for parametric bootstraps. Suppose that X_n^* is a bootstrap sample of size n, constructed so that the conditional distribution of X_n^*, given X_n, is $P_{\hat{\theta}_n,n}$. Let $T_n^* = T_n(X_n^*)$, let $Y_n^* = Y_n(X_n^*, \hat{\theta}_n)$, and let $T_{n,E}^* = T_{n,E}(X_n^*)$. In Theorem 2.1, correct convergence of the parametric bootstrap distributions for $n^{1/2}(T_n - \theta_0)$ cannot occur unless $n^{1/2}(T_n^* - T_{n,E}^*)$ and Y_n^* are asymptotically independent, given X_n. Graphical checks for this approximate independence serve as the basic diagnostic for bootstrap success.

The following algorithm carries out the idea:

1) Given the original sample X_n, construct B conditionally independent bootstrap samples $X_n^{*1}, X_n^{*2}, \ldots, X_n^{*B}$, each of which has conditional distribution $P_{\hat{\theta}_n,n}$.

2) Compute $R_n^{*j} = n^{1/2}[T_n(X_n^{*j}) - T_{n,E}(X_n^{*j})]$ and $Y_n^{*j} = Y_n(X_n^{*j}, \hat{\theta}_n)$ for $1 \leq j \leq B$.

3) Select two real-valued test functions f and g that are continuous over R^m and R^k respectively. Plot the points $\{(f(R_n^{*j}), g(Y_n^{*j})) : 1 \leq j \leq B\}$ in R^2.

4) Use this scatterplot and summaries such as robust correlations to assess whether approximate independence of the $\{R_n^{*j}\}$ and the $\{Y_n^{*j}\}$ breaks down. If so, mistrust the bootstrap distribution $H_n(\hat{\theta}_n)$ computed from this data set.

Theorem 2.1 provides a rationale for this algorithm, provided the test functions f and g are chosen without reference to the sample X_n. The critical assumption on the support of $J(\theta_0)$ is met by typical estimators $\{\hat{\theta}_n\}$.

Adjusting this algorithm to fit the nonparametric bootstrap consists in minor changes:

1′) Given the original sample X_n, construct B conditionally independent bootstrap samples $X_n^{*1}, X_n^{*2}, \ldots, X_n^{*B}$, each of which has conditional distribution $\hat{P}_n^n$.

2′) Let $\hat{P}_n^{*j}$ denote the empirical distribution of bootstrap sample X_n^{*j}. Compute $R_n^{*j} = n^{1/2}[T_n(X_n^{*j}) - T_{n,E}(X_n^{*j})]$ and $Y_n^{*j} = n^{1/2}(\hat{P}_n^{*j} - \hat{P}_n)$ for $1 \leq j \leq B$.

3′) Select two real-valued test functions f and g that are continuous over R^m and $L(s,d)$ respectively. Plot the points $\{(f(R_n^{*j}), g(Y_n^{*j})) : 1 \leq j \leq B\}$ in R^2.

4′) Assess as before.

Theorem 2.2 supplies a rationale for this algorithm.

Example 4. In the nonparametric model of Example 3, with $k = 1$, consider the Hodges estimator $T_{n,H}$ in (1.1), with $b = 0$, as an estimator of the mean $\mu(P)$. To diagnose empirically whether the nonparametric bootstrap distribution for $H_n(P) = \mathcal{L}[n^{1/2}(T_{n,H} - \mu(P)) \mid P^n]$ is reliable, we take, in the preceding algorithm,

$$f(R_n^{*j}) = n^{1/2}(T_{n,H}(X_n^{*j}) - \bar{X}_n^{*j}) \qquad g(Y_n^{*j}) = n^{1/2}(\bar{X}_n^{*j} - \bar{X}_n). \tag{2.16}$$

Here f is the identity function while $g(\nu) = \int x d\nu(x)$ for every signed measure $\nu \in L(s,d)$.

Figure 1 displays scatterplots produced from two samples X_n by step 3′) of the diagnostic algorithm. The first sample was an artificial $N(0,1)$ sample of size $n = 30$, for which $\mu(P) = 0$. Note that this sample comes from a distribution in Π. The second sample was the first sample shifted by 1 so that $\mu(P) = 1$. Both diagnostic plots exhibit points from $B = 50$ bootstrap samples. In the plot for the first sample, 80% of the points lie on a line with slope -1, indicating strong dependence and failure of the nonparametric bootstrap. In the plot for the second sample, 94% of the points have abscissa 0, indicating near independence and success of the nonparametric bootstrap. As $n \to \infty$, the proportion of plotted points not on the respective lines tends to zero. This theoretical result follows readily from the definition of the Hodges estimator and (2.16). Note that sample correlation is an untrustworthy measure of dependence in Fig. 1 because of the outlying points. Theorem 2.2 only supports the use of dependence measures that are continuous with respect to weak convergence of distributions.

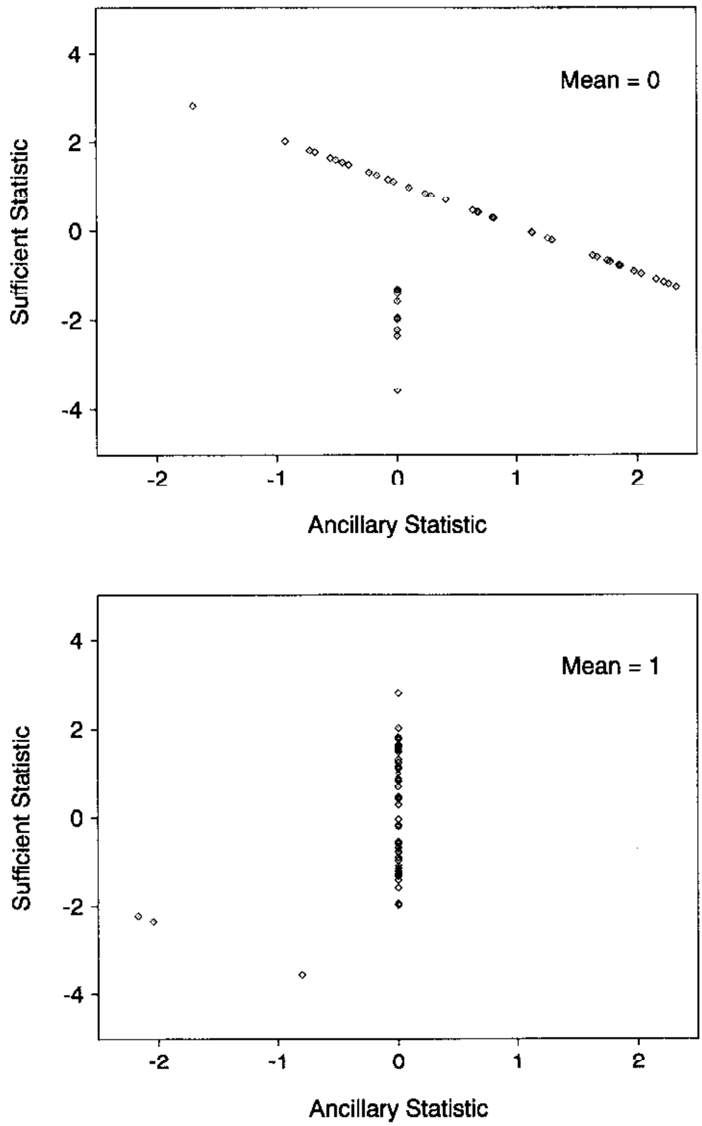

Fig. 1. Diagnostic scatterplots for Example 4 at sample size $n = 30$. When $\mu(P) = 0$, 80% of the $B = 50$ points lie on a line of slope -1, indicating strong dependence and failure of the nonparametric bootstrap. When $\mu(P) = 1$, 94% of the points lie on the vertical line with abscissa 0, indicating near independence and success of the nonparametric bootstrap.

2.4 *Repairing bootstrap failure*

To better understand the role of the support and LAE conditions in Theorem 2.1, we consider the following relaxation of LAE.

DEFINITION 2.5. The estimators $\{T_n\}$ are said to be locally asymptotically weakly convergent (LAWC) at θ_0 if there exists a family of distributions $\{\pi(\theta_0, h) : h \in R^k\}$ such that

$$H_n(\theta_0 + n^{-1/2} h_n) \Rightarrow \pi(\theta_0, h) \tag{2.17}$$

for every $h \in R^k$ and every sequence $\{h_n \in R^k\}$ converging to h.

For instance, both the Stein and Hodges estimators of Examples 1 and 2 are LAWC at every point in the parameter space. Evidently, an estimator that is LAWC at θ_0 is LAE at θ_0 if and only if $\pi(\theta_0, h) = \pi(\theta_0, 0)$ for every $h \in R^k$.

THEOREM 2.3. *Suppose that the estimators $\{T_n\}$ are LAWC at θ_0 and the $\{\hat{\theta}_n\}$ are such that $J_n(\theta_0) \Rightarrow J(\theta_0)$. Let V be a random vector whose distribution is $J(\theta_0)$. Then, under $P_{\theta_0,n}$, the bootstrap distribution $H_n(\hat{\theta}_n)$ converges in distribution, as a random probability measure, to the random probability measure $\pi(\theta_0, V)$.*

Thus, the bootstrap distribution of an LAWC estimator converges correctly if and only if

$$\pi(\theta_0, V) = H(\theta_0) = \pi(\theta_0, 0) \quad w.p.1. \tag{2.18}$$

The following Corollary develops this observation into two methods for repairing the intuitive bootstrap at non-LAE points.

COROLLARY 2.1. *Suppose that the conditions for Theorem 2.3 hold. The following assertions hold under $P_{\theta_0,n}$:*

a) *Let $\{\hat{\theta}_n\}$ be such that the support of $J(\theta_0)$ is the origin whenever θ_0 is a point where $\{T_n\}$ is not LAE. Then the bootstrap distribution $H_n(\hat{\theta}_n)$ converges weakly in probability to the correct limit distribution $H(\theta_0)$.*

b) *Let $\{m_n : n \geq 1\}$ be any sequence of positive integers such that $m_n \to \infty$ and $m_n/n \to 0$ as $n \to \infty$. Then the sub-sample bootstrap distribution $H_{m_n}(\hat{\theta}_n)$ converges weakly in probability to the correct limit distribution $H(\theta_0)$.*

Wu (1986), Mammen (1992), Putter (1994) and Beran (1995) exhibited several ways to modify $\hat{\theta}_n$ so as to remedy bootstrap failure at certain points in the parameter space while preserving correct convergence elsewhere. Similar to part a) of Corollary 2.1 is Putter's work. The alternative cure b) by modifying bootstrap sample size is also effective over the whole parameter space. Introduced by Bretagnolle (1983), this sample size device was used by Beran and Srivastava (1985), Swanepoel (1986), and Athreya (1987). Künsch (1989), Franke and Härdle (1992), and Politis and Romano (1992) developed related methods for bootstrapping stationary time series.

Some disclaimers are in order. A practical drawback to strategy a) is the prior mathematical analysis required to determine the non-LAE points of T_n and to construct suitable $\hat{\theta}_n$. Conceivably, a more sophisticated adaptive construction of $\hat{\theta}_n$ might overcome this obstacle. A drawback to strategy b) is that $H_{m_n}(\bar{X}_n)$ is highly inefficient as an estimator of $H(\theta_0)$ whenever θ_0 is an LAE point (cf. Beran (1982)). Bickel *et al.* (1997) suggest efficiency improvements to strategy b). Typically, the pointwise bootstrap convergence achieved by either strategy is not locally uniform in θ at non-LAE points.

Example 2. (continued) To apply part a) of the Corollary to the Stein estimator, let

$$\hat{\theta}_n = \begin{cases} m(\bar{X}_n)e & \text{if } \ |\bar{X}_n - m(\bar{X}_n)e| \le n^{-1/4} \\ \bar{X}_n & \text{otherwise} \end{cases} . \tag{2.19}$$

Evidently, $P_{\theta_0,n}[\hat{\theta}_n = m(\bar{X}_n)e]$ tends to 1 if the components of θ_0 are all equal and to 0 otherwise. Hence, the limit distribution $J(\theta_0)$ is a point mass at the origin if the components of θ_0 are all equal and is $N(0, I_k)$ otherwise. The Stein estimator $T_{n,S}$ is LAWC when the components of θ_0 are all equal, with $\pi(\theta_0, h)$ given by (1.4); and it is LAE at all other values of θ_0, with $\pi(\theta_0, h) = N(0, I_k)$ for every h. By Theorem 2.1 at LAE points and by part a) of Theorem 2.3 elsewhere, the bootstrap distribution $H_n(\hat{\theta}_n)$ converges to $H(\theta_0)$ in $P_{\theta_0,n}$-probability for *every* choice of θ_0.

Applying Part b) of the Corollary to the Stein estimator, with $\hat{\theta}_n = \bar{X}_n$, is straightforward.

2.5 *Studentized roots*

Suppose that S_n estimates consistently the Euclidean-valued parametric function $\sigma(\theta)$. Let $G_n(\theta)$ denote the distribution under $P_{\theta,n}$ of the pair $(n^{1/2}(T_n - \tau), S_n)$. Let $G(\theta) = \mathcal{L}\{R, \sigma(\theta)\}$, where R has distribution $H(\theta)$. In this subsection, we will characterize the set $\mathcal{D}$ of θ values such that both $G_n(\theta) \Rightarrow G(\theta)$ and $G_n(\hat{\theta}_n) \Rightarrow G(\theta)$ in $P_{\theta,n}$-probability—that is, the values of θ at which the latter bootstrap distribution converges correctly.

Motivating this analysis is the observation that studentized roots used to construct confidence sets have the form $f[n^{1/2}(T_n - \tau(\theta)), S_n]$, where f is real-valued or possibly vector-valued and S_n estimates the symmetric square root $\sigma(\theta)$ of the asymptotic covariance of $n^{1/2}(T_n - \tau(\theta))$. Such roots follow the example of Hotelling's T^2-statistic. Let $G_{n,f}(\theta) = \mathcal{L}\{f[n^{1/2}(T_n - \tau(\theta)), S_n] \mid P_{\theta,n}\}$ and let $G_f(\theta) = \mathcal{L}\{f(R, \sigma(\theta))\}$. Assume that the discontinuity points of f form a null set with respect to $G(\theta)$. Then, for every $\theta \in \mathcal{D}$, both $G_{n,f}(\theta) \Rightarrow G_f(\theta)$ and $G_{n,f}(\hat{\theta}_n) \Rightarrow G_f(\theta)$ in $P_{\theta,n}$-probability. The set $\mathcal{D}$ thus describes points where the bootstrap distribution of a studentized root converges correctly.

Let $F_n(\theta) = \mathcal{L}(S_n \mid P_{\theta,n})$ and let $\delta(\sigma(\theta))$ denote the point mass at $\sigma(\theta)$. Because a constant and a random element are independent, the bootstrap convergence $G_n(\hat{\theta}_n) \Rightarrow G(\theta)$ in $P_{\theta,n}$-probability is equivalent to the pair of convergences $H_n(\hat{\theta}_n) \Rightarrow H(\theta)$ and $F_n(\hat{\theta}_n) \Rightarrow \delta(\sigma(\theta))$ in $P_{\theta,n}$-probability. Equivalent conditions for the second of these are as follows.

THEOREM 2.4. *Suppose that the model $\{P_{\theta,n} : \theta \in \Theta\}$ is LAN at θ_0 and that $J_n(\theta_0) \rightarrow J(\theta_0)$. Then the following statements are equivalent:*

a) *$F_n(\hat{\theta}_n) \Rightarrow \delta(\sigma(\theta_0))$ in $P_{\theta_0,n}$-probability.*

b) *$S_n \rightarrow \sigma(\theta_0)$ in $P_{\theta_0,n}$-probability.*

c) *$S_n \rightarrow \sigma(\theta_0)$ in $P_{\theta_0+n^{-1/2}h_n,n}$-probability, for every $h \in R^k$ and for every sequence $\{h_n\}$ converging to h.*

This theorem has several implications. In view of Theorem 2.1, the bootstrap distribution $G_n(\hat{\theta}_n)$ converges correctly under $P_{\theta_0,n}$ if and only if the estimators $\{T_n\}$ are LAE at θ_0 and b) holds there. Second, the obvious diagnostic for b)—checking that recomputations of S_n from B conditionally independent bootstrap samples are similar—has the theoretical support of a) and may be added to the diagnostic steps of Subsection 2.3. Third, as in Subsection 2.2, it is possible to extend Theorem 2.4 and its applications to nonparametric bootstrapping.

3. Local asymptotic equivariance and superefficiency

The results in this section are needed to prove Theorem 2.1 and to clarify the structure of the exceptional set E where bootstrap failure occurs.

DEFINITION 3.1. Suppose that $H_n(\theta_0) \Rightarrow H(\theta_0)$ as $n \rightarrow \infty$. The estimators $\{T_n\}$ are essentially locally asymptotically equivariant (ELAE) at θ_0 if there exists a uniqueness set $U \subset R^k$ and, for every $h \in U$, a sequence $\{h_n \in R^k\}$ converging to h such that

$$H_n(\theta_0 + n^{-1/2}h_n) \Rightarrow H(\theta_0). \tag{3.1}$$

The ELAE property is formally weaker than LAE in two respects: the values of h are restricted to a uniqueness set U that can be a proper subset of R^k; and (3.1) is required to hold for only one sequence $\{h_n\}$ converging to h. Convergence in probability of the bootstrap distribution $H_n(\hat{\theta}_n)$ to $H(\theta_0)$ proves to be a stochastic form of ELAE, where $\{h_n\}$ is a realization of $\{n^{1/2}(\hat{\theta}_n - \theta_0)\}$ and U is the support of $J(\theta_0)$.

Clearly LAE implies Hájek (1970) regularity, which in turn implies ELAE. The first two parts of the next theorem establish the equivalence of these concepts for models that are LAN. The third part contains Pfanzagl's ((1994), Theorem 8.4.1) formulation of the convolution theorem as the special case $h_n = 0$.

THEOREM 3.1. *Suppose that the model $\{P_{\theta,n} : \theta \in \Theta\}$ is LAN at θ_0 and τ is Fréchet differentiable at θ_0 with derivative $\nabla\tau(\theta_0)$ of full rank. Suppose that $H_n(\theta_0) \Rightarrow H(\theta_0)$. Then the following statements are equivalent:*

a) *$\{T_n\}$ is ELAE at θ_0 with limit distribution $H(\theta_0)$.*

b) *$\{T_n\}$ is LAE at θ_0 with limit distribution $H(\theta_0)$.*

c) *For every $h \in R^k$ and every sequence $\{h_n \in R^k\}$ converging to h,*

$$K_n(\theta_0 + n^{-1/2}h_n) \Rightarrow D(\theta_0) \times N(0, I(\theta_0)) \tag{3.2}$$

for some distribution $D(\theta_0)$ such that $H(\theta_0) = D(\theta_0) * N(0, \Sigma_\tau(\theta_0))$.

Part c) implies that $n^{1/2}(T_n - T_{n,E})$ is locally asymptotically invariant in the approximating local normal model at θ_0. At the same time, the statistic $Y_n(\theta)$ is locally asymptotically sufficient and complete in the approximating local normal model (cf. LeCam (1956)). The local asymptotic independence asserted in part c) recalls Basu's (1955) theorem on the independence of an ancillary statistic and a complete sufficient statistic. Part c) of Theorem 2.1 then carries Basu's theorem into the bootstrap world associated with $H_n(\hat{\theta}_n)$.

Theorem 3.1 is also linked to the result of Lehmann and Scheffé that a uniformly minimum variance unbiased estimator is uncorrelated with any unbiased estimator of zero (Lehmann (1983), Theorem 1.1). On the one hand, $T_n - T_{n,E}$ is an LAE estimator of zero; $T_{n,E}$ is a best LAE estimator for $\tau(\theta)$ in view of Hájek's convolution theorem; and the LAE property may be interpreted as local asymptotic unbiasedness. On the other hand, part c) of Theorem 3.1 implies local asymptotic independence of $n^{1/2}(T_n - T_{n,E})$ and $n^{1/2}(T_{n,E} - \tau(\theta))$; and so entails local asymptotic correlation zero.

On close reading, remarks by LeCam ((1973), pp. 169 and 176) imply an almost everywhere form of the convolution theorem. Results of this type were proved by Jeganathan (1981), Droste and Wefelmeyer (1984), and Pfanzagl ((1994), Theorem 8.4.14). The next theorem reaches a stronger conclusion: LAE holds almost everywhere.

THEOREM 3.2. *Suppose that the model $\{P_{\theta,n} : \theta \in \Theta\}$ is LAN at every $\theta \in \Theta$ and that $H_n(\theta) \Rightarrow H(\theta)$ for every $\theta \in \Theta$. Then there exists a Lebesgue null set $E \subset \Theta$ such that, for every $h \in R^k$ and every sequence $\{h_n\}$ converging to h,*

$$K_n(\theta + n^{-1/2}h_n) \Rightarrow D(\theta) \times N(0, I(\theta)) \quad \text{for every} \quad \theta \in \Theta - E. \tag{3.3}$$

Hence the estimators $\{T_n\}$ are LAE at every $\theta \in \Theta - E$ and $H(\theta) = D(\theta) * N(0, \Sigma_\tau(\theta))$ for every $\theta \in \Theta - E$.

Theorems 2.1 and 3.2 establish that intuitive bootstrap distributions fail to converge correctly at precisely those points in Θ where $\{T_n\}$ is not LAE; and that this exceptional set has Lebesgue measure zero. In Examples 1 and 2, the points of bootstrap failure were also superefficiency points. A connection between superefficiency and non-LAE exists more generally. Let w be a symmetric, subconvex, continuous, non-negative function on R^k. We say that $\{T_n\}$ is *superefficient* at θ_0 for loss function w if

$$\limsup_{n\to\infty} \mathrm{E}_{\theta_0} w[n^{1/2}(T_n - \tau(\theta_0))] < \mathrm{E}w[\Sigma_\tau^{-1/2}(\theta_0)Z_k], \tag{3.4}$$

where Z_k has a standard normal distribution on R^k.

If $\{T_n\}$ is LAE at θ_0, the limiting distribution $H(\theta_0)$ has the convolution structure in part c) of Theorem 3.1. Fatou's lemma and Anderson's lemma (cf. Ibragimov and Has'minskii (1981), p. 157) then imply that

$$\limsup_{n\to\infty} \mathrm{E}_{\theta_0} w[n^{1/2}(T_n - \tau(\theta_0))] \geq \mathrm{E}w[\Sigma_\tau^{-1/2}(\theta_0)Z_k]. \tag{3.5}$$

Thus, a point of superefficiency must be a non-LAE point. It is useful to note that discontinuity of $H(\theta)$ or of $\Sigma_\tau(\theta)$ at θ_0 is a necessary condition for breakdown of convolution structure at θ_0 (see the Addendum to Theorem 8.4.14 in Pfanzagl (1994)); and consequently is a necessary condition for superefficiency at θ_0. Examples 1 and 2 illustrate discontinuity of the limit distribution $H(\theta)$ at the superefficiency points. The following example shows, however, that the set of non-LAE points can be strictly larger than the set of superefficiency points.

Example 5. In the setting of Example 1, we construct a modified Hodges estimator as follows. Split the sample into two subsamples of sizes $n_1 = [n/2]$ and $n_2 = n - n_1$. Let $\bar{X}_{n,1}$ denote the average of the first n_1 components of the sample X_n and let $\bar{X}_{n,2}$ denote the average of the remaining n_2 components. Define

$$T_{n,MH} = \begin{cases} (\bar{X}_{n,1} - \bar{X}_{n,2})/2 & \text{if} \quad |\bar{X}_n| \le n^{-1/4} \\ \bar{X}_n & \text{otherwise} \end{cases} . \tag{3.6}$$

For every sequence $\{h_n \in R^k\}$ that converges to h, the distributions $H_n(\theta_0 + n^{-1/2}h_n) = \mathcal{L}[n^{1/2}(T_{n,MH} - (\theta_0 + n^{-1/2}h_n)) \mid P_{\theta_0+n^{-1/2}h_n,n}]$ converge weakly to $N(0,1)$ if $\theta_0 \neq 0$ and to $N(-h,1)$ if $\theta_0 = 0$. Consequently, $\{T_{n,MH}\}$ is not LAE at 0. Nor is it superefficient there, because the pointwise limit $H(\theta)$ is $N(0,1)$ for every θ.

We may summarize the preceding discussion quite simply. Under the LAN and support conditions of Theorem 2.1, LAE at θ_0 is a necessary and sufficient condition for correct convergence of intuitive bootstrap distributions at θ_0. This condition remains critical in the more general discussion of Subsection 2.5. The set of non-LAE points, or of bootstrap failure, has Lebesgue measure zero. Superefficiency at θ_0 is a sufficient condition for bootstrap failure there, but it is not necessary.

4. Proofs

The logical sequence is to prove Section 3 and then Section 2. We end with a short proof of Proposition 1.2 in the Introduction.

Proof of Theorem 3.1. For every $h \in R^k$ and every sequence $\{h_n \in R^k\}$ converging to h, write θ_n as shorthand for $\theta_0 + n^{-1/2}h_n$. The LAN property implies that $\{P_{\theta_n,n}\}$ and $\{P_{\theta_0,n}\}$ are contiguous. By contiguity and (2.2),

$$Y_n(\theta_n) = Y_n(\theta_0) - I(\theta_0)h + o_p(1) \tag{4.1}$$

under $P_{\theta_n,n}$. From this, property (2.5), contiguity, and differentiability of τ,

$$n^{1/2}(T_{n,E} - \tau(\theta_n)) = \nabla\tau(\theta_0)I^{-1}(\theta_0)Y_n(\theta_n) + o_p(1) \tag{4.2}$$

under $P_{\theta_n,n}$. Observe that $n^{1/2}(T_n - \tau(\theta_n))$ is the sum of $n^{1/2}(T_n - T_{n,E})$ and $n^{1/2}(T_{n,E} - \tau(\theta_n))$. Part c) of the theorem and (4.2) thus imply part b), the limit distribution $H(\theta_0)$ having the convolution structure asserted in c). Since b) obviously implies a), it remains only to show that a) implies c).

First step. Suppose not. By the hypothesized weak convergence, LAN, and tightness, there exists a subsequence M such that the weak convergence in c) fails on M while

$$(4.3)\qquad \mathcal{L}[n^{1/2}(T_n - \tau(\theta_0)), Y_n(\theta_0) \mid P_{\theta_0,n}] \Rightarrow \mathcal{L}(R, Y) \quad \text{for} \quad n \in M$$

with $\mathcal{L}(R) = H(\theta_0)$ and $\mathcal{L}(Y) = N(0, I(\theta_0))$. Fix $h \in R^k$ and let $\{h_n\}$ be the distinguished sequence whose existence is assured by ELAE. By contiguity, differentiability of τ, and Corollary 2.1 of LeCam (1960), the characteristic function of $H_n(\theta_n)$ satisfies

$$(4.4)\qquad \begin{aligned} &\lim_{n\in M} \mathrm{E}_{\theta_n} \exp\{iu'n^{1/2}[T_n - \tau(\theta_n)]\} \\ &\quad = \mathrm{E}\{\exp[iu'R - iu'\nabla\tau(\theta_0)h] \exp[h'Y - 2^{-1}h'I(\theta_0)h]\}. \end{aligned}$$

By ELAE, the left side of (4.4) also equals $\mathrm{E}\exp(iu'R)$. Thus

$$(4.5)\qquad \mathrm{E}\exp(iu'R) = \mathrm{E}\{\exp[iu'R - iu'\nabla\tau(\theta_0)h] \exp[h'Y - 2^{-1}h'I(\theta_0)h]\}$$

for every h in the uniqueness set U. Since the right side of (4.5) is analytic in h while the left side does not depend on h, equation (4.5) holds for every $h \in C^k$.

Setting $h = i[t - I^{-1}(\theta_0)\nabla'\tau(\theta_0)u]$ in (4.5) yields, after simplification,

$$(4.6)\quad \mathrm{E}\exp(iu'R) = \mathrm{E}\exp(iu'W + it'Y)\exp[2^{-1}t'I(\theta_0)t]\exp[-2^{-1}u'\Sigma_\tau(\theta_0)u],$$

where $W = R - \nabla\tau(\theta_0)I^{-1}(\theta_0)Y$ and $\Sigma_\tau(\theta_0)$ is defined in (2.4). Consequently,

$$(4.7)\quad \mathrm{E}\exp(iu'W + it'Y) = \exp[2^{-1}u'\Sigma_\tau(\theta_0)u]\mathrm{E}\exp(iu'R)\exp[-2^{-1}t'I(\theta_0)t],$$

which establishes independence of W and Y.

Second step. Let

$$(4.8)\qquad W_n = n^{1/2}(T_n - T_{n,E}) = n^{1/2}[T_n - \tau(\theta_0)] - n^{1/2}[T_{n,E} - \tau(\theta_0)].$$

From (2.5) and (4.3), $\mathcal{L}[W_n, Y_n(\theta_0) \mid P_{\theta_0,n}] \Rightarrow \mathcal{L}(W, Y)$ on the subsequence M. Let $\{h_n \in R^k\}$ now be *any* sequence that converges to h. By contiguity, Corollary 2.1 of LeCam (1960), and the independence of (W, Y),

$$(4.9)\qquad \begin{aligned} &\lim_{n\in M} \mathrm{E}_{\theta_n} \exp[iu'W_n + it'Y_n(\theta_0)] \\ &\quad = \mathrm{E}\{\exp(iu'W + it'Y)\exp[h'Y - 2^{-1}h'I(\theta_0)h]\} \\ &\quad = \mathrm{E}\exp(iu'W)\mathrm{E}\exp[(it + h)'Y]\exp[-2^{-1}h'I(\theta_0)h]. \end{aligned}$$

Because the distribution of Y is $N(0, I(\theta_0))$, the transform $\mathrm{E}\exp(s'Y) = \exp[2^{-1}s'I(\theta_0)s]$ for every $s \in C^k$. Thus, (4.9) simplifies to

$$(4.10)\qquad \begin{aligned} &\lim_{n\in M} \mathrm{E}_{\theta_n} \exp[iu'W_n + it'Y_n(\theta_0)] \\ &\quad = \mathrm{E}\exp(iu'W)\mathrm{E}\exp[it'I(\theta_0)h - 2^{-1}t'I(\theta_0)t]. \end{aligned}$$

If we let $D(\theta_0) = \mathcal{L}(W)$, then (4.10) is equivalent to the weak convergence

$$(4.11) \qquad \mathcal{L}[W_n, Y_n(\theta_0) \mid P_{\theta_n,n}] \Rightarrow D(\theta_0) \times N(I(\theta_0)h, I^{-1}(\theta_0)) \quad \text{for} \quad n \in M.$$

From this and (4.1), it follows that $K_n(\theta_n) \Rightarrow D(\theta_0) \times N(0, I^{-1}(\theta_0))$ for $n \in M$, contradicting our initial supposition. Hence, a) implies the weak convergence in c). The convolution structure for $H(\theta_0)$ follows by specializing that weak convergence to $h_n = 0$, then using (4.8) and (2.5).

LEMMA 4.1. *Let* $\{f_n : n \geq 1\}$ *and* f *be Lebesgue measurable real-valued functions on* R^k *such that*

$$(4.12) \qquad \lim_{n\to\infty} f_n(x) = f(x) \quad \text{a.e. Lebesgue.}$$

Then, for every sequence $\{y_n \in R^k\}$ *converging to zero, there exists a subsequence* M *such that*

$$(4.13) \qquad \lim_{n\in M} f_n(x + y_n) = f(x) \quad \text{a.e. Lebesgue.}$$

PROOF OF LEMMA 4.1. For this generalization of Bahadur's lemma, see Droste and Wefelmeyer ((1984), pp. 140–141) or Pfanzagl ((1994), pp. 285–286).

PROOF OF THEOREM 3.2. Let $\phi_n(u, \theta)$ and $\phi(u, \theta)$ denote, respectively, the characteristic functions of $H_n(\theta)$ and $H(\theta)$. Let D be a countable dense subset of R^k. Fix $(h, u) \in D^2$. By the weak convergence hypothesized in Theorem 3.2 and by Lemma 4.1 with $f_n = \phi_n(u, \cdot)$, $x = \theta$, and $y_n = n^{-1/2}h$, there exists a subsequence $M(h, u)$ and a Lebesgue null set $E(h, u)$ such that

$$(4.14) \qquad \lim_{n\in M(h,u)} \phi_n(u, \theta + n^{-1/2}h) = \phi(u, \theta) \quad \text{for every} \quad \theta \in \Theta - E(h, u).$$

Let

$$(4.15) \qquad E = \bigcup_{(h,u)\in D^2} E(h, u).$$

By reasoning akin to the first step in the proof of Theorem 3.1, equation (4.5) holds for every $(h, u) \in D^2$ and for every $\theta \in \Theta - E$. Because D is a uniqueness set for analytic functions and because a characteristic function is determined by its values on the dense subset $D \subset R^k$, equation (4.5) continues to hold for every $h \in C^k$, every $u \in R^k$, and every $\theta \in \Theta - E$. The remaining argument for the a.e. convergence (3.3) and for the a.e. convolution structure of $H(\theta)$ parallels the proof of Theorem 3.1 after (4.5).

PROOF OF THEOREM 2.1. a) implies c). The weak convergence of $\{J_n(\theta_0)\}$ and the convergence in probability of $\{H_n(\hat{\theta}_n)\}$ imply that $\mathcal{L}[n^{1/2}(\hat{\theta}_n - \theta_0),$

$H_n(\hat{\theta}_n)] \Rightarrow J(\theta_0) \times \delta(H(\theta_0))$, where $\delta(H(\theta_0))$ denotes the point mass at $H(\theta_0)$ and the topology is the product of Euclidean convergence and weak convergence. Consider the probability space $(\Omega, \mathcal{B}, \mu)$, where Ω is the unit interval equipped with Borel sets $\mathcal{B}$ and μ is Lebesgue measure. By a standard Skorokhod construction, which inherits w.p.1 the relation $H_n(\hat{\theta}_n) = H_n(\theta_0 + n^{1/2}(\hat{\theta}_n - \theta_0))$, there exist a null set $N \subset \Omega$ and random vectors $\{V_n(\omega)\}$ and $V(\omega)$ defined on Ω such that $\mathcal{L}(V_n) = J_n(\theta_0)$, $\mathcal{L}(V) = J(\theta_0)$, and

$$\lim_{n\to\infty}(V_n(\omega), H_n(\theta_0 + n^{-1/2}V_n(\omega))) = (V(\omega), H(\theta_0)) \quad \forall \omega \in \Omega - N. \tag{4.16}$$

The set $\{V(\omega) : \omega \in \Omega - N\}$ is dense in the support of $J(\theta_0)$, and so itself constitutes a uniqueness set for analytic functions defined on C^k. For each $\omega \in \Omega - N$, put $h = V(\omega)$ and $h_n = V_n(\omega)$. By (4.16), $h_n \to h$ and $H_n(\theta_n) \Rightarrow H(\theta_0)$. Thus, a) implies that the estimators $\{T_n\}$ are ELAE. The desired LAE property follows from Theorem 3.1.

c) implies b). In (3.2) of Theorem 3.1, set $h_n = V_n(\omega)$ and $h = V(\omega)$ to deduce b) from the LAE property.

b) implies a). The argument draws on the derivation of c) from a). As there, the assumed convergence in probability of $K_n(\hat{\theta}_n)$ implies that, for every h in a dense subset of the support of $J(\theta_0)$, there exists a sequence $\{h_n \in R^k\}$ converging to h such that $K_n(\theta_n) \Rightarrow D(\theta_0) \times N(0, I(\theta_0))$. From this and (4.2), we see that $H_n(\theta_n) \Rightarrow H(\theta_0) = D(\theta_0) * N(0, \Sigma_\tau(\theta_0))$ and the estimators $\{T_n\}$ are ELAE. By Theorem 3.1 they are LAE. In the definition (2.3) of LAE, set $h_n = V_n(\omega)$ and $h = V(\omega)$ to deduce a).

Proof of Theorem 2.2. This argument consists mostly in applying Theorem 3.1 to the parametric model $P^n(\theta, s_0, d_0)$ and to the parametric function $\tau(\theta) = T[P(\theta, s_0, d_0)]$. Here θ ranges over those vectors in R^{d_0-1} whose components are strictly between 0 and 1 and sum to strictly less than 1. Under this model, the density of the sample X_n with respect to counting measure on the support points s_0 is

$$\prod_{j=1}^{d_0} \theta_j^{n\hat{\theta}_{n,j}}. \tag{4.17}$$

Let e denote the vector in R^{d_0-1} whose components each equal 1. By Taylor expansion, the parametric model $P^n(\theta, s_0, d_0)$ is LAN at θ_0 with

$$\begin{aligned} I(\theta_0) &= [\mathrm{diag}(\theta_0)]^{-1} + (1 - e'\theta_0)^{-1}ee' \\ Y_n(\theta_0) &= I(\theta_0)n^{1/2}(\hat{\theta}_n - \theta_0). \end{aligned} \tag{4.18}$$

A standard matrix identity yields $I^{-1}(\theta_0) = \mathrm{diag}(\theta_0) - \theta_0\theta_0'$.

As was noted in Subsection 2.2, the random vectors $\{n^{1/2}(\hat{\theta}_n - \theta_0)\}$ converge weakly under $P(\theta_0, s_0, d_0)$ to $W(P_0)$, whose distribution is $J(\theta_0) = N(0, I^{-1}(\theta_0))$. This distribution has full support on R^{d_0-1}.

A sequence of estimators $\{T_n\}$ is partially LAE for $\tau(\theta) = T[P(\theta, s_0, d_0)]$ at θ_0 if and only if it is LAE for τ at θ_0. The plug-in estimator $T(\hat{P}_n) = \tau(\hat{\theta}_n)$ plays the role of $T_{n,E}$. Indeed, (4.18) and the assumed differentiability of τ (i.e. the partial differentiability of T) entail

$$\begin{aligned}(4.19) \qquad \tau(\hat{\theta}_n) &= \tau(\theta_0) + \nabla\tau(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(n^{-1/2}) \\ &= \tau(\theta_0) + n^{-1/2}\nabla\tau(\theta_0)I^{-1}(\theta_0)Y_n(\theta_0) + o_p(n^{-1/2})\end{aligned}$$

under $P^n(\theta_0, s_0, d_0)$, as required in (2.5). The covariance matrix $\Sigma_\tau(\theta_0)$ defined in (2.4) coincides here with the covariance matrix $\Sigma_T(P_0)$ of (2.11).

With these identifications, Theorem 2.1 is applicable. The distributions $H_n(\hat{\theta}_n)$ and $H_n(\hat{P}_n)$ coincide. The distribution $K_n(\hat{\theta}_n)$ in Theorem 2.1 here becomes

$$(4.20) \qquad K_n(\theta) = \mathcal{L}[n^{1/2}(T_n - \tau(\hat{\theta}_n)), I(\theta)n^{1/2}(\hat{\theta}_n - \theta) \mid P^n(\theta, s_0, d_0)].$$

It does *not* coincide with $K_n(\hat{P}_n)$. However, the one-to-one correspondence between $I(\theta)n^{1/2}(\hat{\theta}_n - \theta)$ and $n^{1/2}[\hat{P}_n - P(\theta, s_0, d_0)]$ justifies replacing $K_n(\hat{\theta}_n)$ by $K_n(\hat{P}_n)$ in stating Theorem 2.2.

PROOF OF THEOREM 2.3. Let $\{V_n(\omega)\}$ and $V(\omega)$ be the Skorokhod random vectors constructed for the proof of Theorem 2.1. The LAWC condition (2.17) then yields

$$(4.21) \qquad H_n(\theta_0 + n^{-1/2}V_n(\omega)) \Rightarrow \pi(\theta_0, V(\omega)) \quad \forall\omega \in \Omega.$$

The theorem follows.

PROOF OF COROLLARY 2.1. Part a) is immediate from Theorem 2.3 because $H(\theta_0) = \pi(\theta_0, 0)$. To prove part b), observe that, in the notation of the previous proof, $H_{m_n}(\hat{\theta}_n)$ has the same distribution as $H_{m_n}(\theta_0 + m_n^{-1/2}U_n)$, where $U_n = (m_n/n)^{1/2}V_n$. Since $U_n(\omega) \to 0$ for every $\omega \in \Omega$ and $m_n \to \infty$, LAWC implies

$$(4.22) \qquad H_{m_n}(\theta_0 + m_n^{-1/2}U_n) \to \pi(\theta_0, 0) = H(\theta_0) \quad \forall\omega \in \Omega.$$

Assertion b) follows.

PROOF OF THEOREM 2.4. Statements b) and c) are equivalent because $\{P_{\theta_0,n}\}$ and $\{P_{\theta_0+n^{1/2}h_n,n}\}$ are contiguous under the LAN assumption.

a) implies b). As in the first part of the proof of Theorem 2.1, construct Skorokhod versions $\{V_n(\omega)\}$ and $V(\omega)$ such that $\mathcal{L}(V_n) = J_n(\theta_0)$, $\mathcal{L}(V) = J_n(\theta_0)$, and

$$(4.23) \qquad \lim_{n\to\infty}(V_n(\omega), F_n(\theta_0 + n^{-1/2}V_n(\omega)) = (V(\omega), \delta(\sigma(\theta_0))) \quad \forall\omega \in \Omega - N,$$

where N is a null set. Fix $\omega \in \Omega - N$ and put $h = V(\omega)$, $h_n = V_n(\omega)$. By (4.23), $h_n \to h$ and $F_n(\theta_n) \Rightarrow \delta(\sigma(\theta_0))$. Part b) now follows because $\{P_{\theta_0+n^{-1/2}h_n,n}\}$ and $\{P_{\theta_0,n}\}$ are contiguous.

b) implies a). Since b) is equivalent to c), set $h_n = V_n(\omega)$ and $h = V(\omega)$ in c) to deduce a).

PROOF OF PROPOSITION 1.2. The exercises and theorems cited below are all in Chapter 1 of Munroe (1953). Let g be any bounded, continuous function on the range space of the root R_n. Because of ex. o on p. 70 and the Baire category theorem, hypothesis a) implies that there exists a set E of category I such that

$$\int g dH_n(\cdot,\theta) \to \int g dH(\cdot,\theta) \quad \text{subuniformly on} \quad \Theta - E \tag{4.24}$$

as $n \to \infty$. Fix $\theta_0 \in \Theta - E$. By hypothesis b),

$$\lim_{\theta \to \theta_0} \int g dH_n(\cdot,\theta) = \int g dH_n(\cdot,\theta_0), \qquad n \geq 1. \tag{4.25}$$

By the subuniform version of Theorem 5.4 (ex. j on p. 44) and the two preceding limits,

$$\begin{aligned} \lim_{n\to\infty,\theta\to\theta_0} \int g dH_n(\cdot,\theta) &= \lim_{n\to\infty} \lim_{\theta\to\theta_0} \int g dH_n(\cdot,\theta) \\ &= \lim_{n\to\infty} \int g dH_n(\cdot,\theta_0) = \int g dH(\cdot,\theta_0). \end{aligned} \tag{4.26}$$

Let $\{\theta_n \in \Theta\}$ be any sequence converging to $\theta_0 \in \Theta - E$. Convergence (4.26) implies that $H_n(\theta_n) \Rightarrow H(\theta_0)$. The first assertion of Proposition 1.2 follows from this and hypothesis c). The second assertion is the Baire category theorem.

REFERENCES

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case, *Ann. Statist.*, **15**, 724–731.

Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhyā*, **15**, 377 380.

Beran, R. (1982). Estimated sampling distributions: the bootstrap and competitors, *Ann. Statist.*, **10**, 212–225.

Beran, R. (1984). Bootstrap methods in statistics, *Jahresber. Deutsch. Math.-Verein.*, **86**, 212–225.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements, *J. Amer. Statist. Assoc.*, **83**, 687–697.

Beran, R. (1995). Stein confidence sets and the bootstrap, *Statist. Sinica*, **5**, 109–127.

Beran, R. and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix, *Ann. Statist.*, **13**, 95–115 (Correction: *Ann. Statist.*, **15**, 470–471).

Bickel, P. J. and Freedman, D. F. (1981). Some asymptotic theory for the bootstrap, *Ann. Statist.*, **9**, 1196–1217.

Bickel, P. J., Götze, F. and von Zwet, W. R. (1997). Bootstrapping fewer than n observations: gains, losses, and remedies for losses, *Statist. Sinica*, **7** (in press).

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles, *Ann. Inst. Henri Poincaré.*, **X9**, 281–296.

Droste, W. and Wefelmeyer, W. (1984). On Hájek's convolution theorem, *Statist. Decisions*, **2**, 131–144.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.

Franke, J. and Härdle, W. (1992). On bootstrapping kernel spectral estimates, *Ann. Statist.*, **20**, 121–145.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates, *Zeit. Wahrscheinlichkeitsth.*, **14**, 323–330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation, *Proc. Sixth Berkeley Symp. on Math. Statist. Prob.* (eds. L. M. LeCam, J. Neyman, and E. M. Scott), Vol. **1**, 175–194, Univ. California Press, Berkeley.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Hjorth, J. S. U. (1994). *Computer Intensive Statistical Methods*, Chapman and Hall, London.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*, Springer, New York.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property, *Ann. Inst. Statist. Math.*, **22**, 1–13.

Jeganathan, P. (1981). On a decomposition of the limit distribution of a sequence of estimators, *Sankhyā Ser. A*, **43** (1), 26–36.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, **17**, 1217–1241.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, *Univ. Calif. Pub. Statist.*, **1**, 277–330.

LeCam, L. (1956). On the asymptotic theory of estimation and testing hypotheses, *Proc. Third Berkeley Symp. on Math. Statist. Prob.* (ed. J. Neyman), Vol. **1**, 129–156, Univ. California Press, Berkeley.

LeCam, L. (1960). Locally asymptotically normal families of distributions, *Univ. Calif. Pub. Statist.* **3**, 27–98.

LeCam, L. (1969). *Théorie Asymptotique de la Décision Statistique*, Univ. Montreal Press.

LeCam, L. (1973). Sur les contraintes imposées par les passages à limite usuels en statististique, *Bull. Internat. Statist. Inst.*, **45**, 169–180.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Mammen, E. (1992). *When Does Bootstrap Work? Lecture Notes in Statist.*, **77**, Springer, New York.

Munroe, M. E. (1953). *Introduction to Measure and Integration*, Addison-Wesley, Reading, Massachusetts.

Pfanzagl, J. (1994). *Parametric Statistical Theory*, Walter de Gruyter, Berlin.

Politis, D. N. and Romano, J. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, *Ann. Statist.*, **20**, 1985–2007.

Putter, H. (1994). *Consistency of Resampling Methods*, Ph.D. dissertation at Leiden University, The Netherlands.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Comm. Statist. Theory Methods*, **15**, 3193–3203.

Titterington, D. M. (1991). Choosing the regularization parameter in image restoration, *Spatial Statistics and Imaging* (ed. A. Possolo), IMS Lecture Notes-Monograph Series, **20**, 392–402.

Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal, *Biometrika*, **29**, 350–362.

Wu, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion), *Ann. Statist.*, **14**, 1261–1295.