RESAMPLING: CONSISTENCY OF SUBSTITUTION ESTIMATORS[1]

By Hein Putter and Willem R. van Zwet

University of Leiden and University of North Carolina, Chapel Hill

On the basis of N i.i.d. random variables with a common unknown distribution P we wish to estimate a functional $\tau_N(P)$. An obvious and very general approach to this problem is to find an estimator $\hat{P}_N$ of P first, and then construct a so-called substitution estimator $\tau_N(\hat{P}_N)$ of $\tau_N(P)$. In this paper we investigate how to choose the estimator $\hat{P}_N$ so that the substitution estimator $\tau_N(\hat{P}_N)$ will be consistent.

Although our setup covers a broad class of estimation problems, the main substitution estimator we have in mind is a general version of the bootstrap where resampling is done from an estimated distribution $\hat{P}_N$. We do not focus in advance on a particular estimator $\hat{P}_N$, such as, for example, the empirical distribution, but try to indicate which resampling distribution should be used in a particular situation. The conclusion that we draw from the results and the examples in this paper is that the bootstrap is an exceptionally flexible method which comes into its own when full use is made of its flexibility. However, the choice of a good bootstrap method in a particular case requires rather precise information about the structure of the problem at hand. Unfortunately, this may not always be available.

1. Substitution estimators.

Let $(\mathscr{X}, \mathscr{A})$ be a measurable space and let $\mathscr{P}$ be a collection of probability measures on $(\mathscr{X}, \mathscr{A})$. Let Π be a topology on $\mathscr{P}$, so that $(\mathscr{P}, \Pi)$ is a topological space. Finally, let $X_1, X_2, \ldots$ denote a sequence of i.i.d. random variables with values in $\mathscr{X}$ and (unknown) common distribution $P \in \mathscr{P}$.

For $N = 1, 2, \ldots$, we consider a map $\tau_N \colon (\mathscr{P}, \Pi) \to (\mathscr{R}, r)$, where $(\mathscr{R}, r)$ is a metric space. Both spaces $(\mathscr{P}, \Pi)$ and $(\mathscr{R}, r)$ are equipped with the σ-algebra of Borel sets $\mathscr{B}(\mathscr{P}, \Pi)$ and $\mathscr{B}(\mathscr{R}, r)$, which are generated by the open sets in $(\mathscr{P}, \Pi)$ and $(\mathscr{R}, r)$, respectively. Probability distributions on these spaces are probability measures on the Borel sets and are induced by measurable maps from $(\mathscr{X}^\infty, \mathscr{A}^\infty, P^\infty)$ to $(\mathscr{P}, \mathscr{B}(\mathscr{P}, \Pi))$ or $(\mathscr{R}, \mathscr{B}(\mathscr{R}, r))$. We assume throughout that each τ_N is measurable.

Having observed the i.i.d. sample $X_1, \ldots, X_N$ with common distribution $P \in \mathscr{P}$, our aim is to estimate the somewhat abstract $\mathscr{R}$-valued "parameter"

$\tau_N(P)$. For a measurable map $t_N: \mathscr{X}^N \to \mathscr{R}$, let $T_N = t_N(X_1, \ldots, X_N)$ be an estimator of $\tau_N(P)$ based on $X_1, \ldots, X_N$. We shall say that T_N is a *consistent estimator* of $\tau_N(P)$ for $P \in \mathscr{P}$ if

$$(1.1) \qquad r(T_N, \tau_N(P)) \to_P 0 \quad \text{for every } P \in \mathscr{P},$$

where $\to_P$ indicates convergence in probability under P as $N \to \infty$. The more formally inclined reader should view this expression as shorthand for the correct but laborious statement that the sequence $\{T_N\}_{N=1}^{\infty}$ is a consistent sequence of estimators of the sequence $\{\tau_N(P)\}_{N=1}^{\infty}$. If we wish to stress the role of the metric r in (1.1), we call T_N *r-consistent*.

In the absence of any special structural properties of $\tau_N(P)$, a popular estimator of $\tau_N(P)$ is the *substitution estimator* $\tau_N(\hat{P}_N)$. (This is commonly called a "plug-in estimator," but this expression is of the same sad grammatical level as "see-through clothes.") It is obtained by first estimating P by $\hat{P}_N = p_N(X_1, \ldots, X_N)$ for a measurable map $p_N: \mathscr{X}^N \to \mathscr{P}$ and then substituting this estimator in τ_N. We shall call the estimator $\hat{P}_N$ consistent with respect to the topology $\boldsymbol{\Pi}$ ($\boldsymbol{\Pi}$-*consistent*) if for every $P \in \mathscr{P}$ and every neighborhood U of P,

$$(1.2) \qquad P^N(\hat{P}_N \in U) \to 1 \quad \text{as } N \to \infty.$$

In the particular applications we have in mind, the topology $\boldsymbol{\Pi}$ on $\mathscr{P}$ will often be metrized by a metric p, so that the topological space $(\mathscr{P}, \boldsymbol{\Pi})$ is a metric space $(\mathscr{P}, p)$. Consistency of $\hat{P}_N$ will then be p-consistency, defined by

$$(1.3) \qquad p(\hat{P}_N, P) \to_P 0 \quad \text{for every } P \in \mathscr{P}.$$

We shall study the consistency of $\tau_N(\hat{P}_N)$ as an estimator of $\tau_N(P)$, assuming that $\hat{P}_N$ is a consistent estimator of P.

The metric p in (1.3) will often be the Hellinger metric H on $\mathscr{P}$. Recall that for $P, Q \in \mathscr{P}$ with densities f and g with respect to a common σ-finite measure μ on $(\mathscr{X}, \mathscr{A})$, the Hellinger distance H of P and Q is defined by

$$(1.4) \qquad H(P, Q) = \left\{ \int_{\mathscr{X}} (f^{1/2} - g^{1/2})^2 \, d\mu \right\}^{1/2}.$$

Note that this definition does not depend on the choice of the dominating measure μ and that H is indeed a metric on $\mathscr{P}$. If $p = H$, (1.3) becomes $H(\hat{P}_N, P) \to_P 0$ for every $P \in \mathscr{P}$ and we say that $\hat{P}_N$ is *Hellinger-consistent*. We call $\hat{P}_N$ a $\sqrt{N}$-*Hellinger-consistent estimator* of P when

$$(1.5) \qquad H(\hat{P}_N, P) = \mathscr{O}_P(N^{-1/2}) \quad \text{for every } P \in \mathscr{P},$$

which means that for every $P \in \mathscr{P}$ and $\varepsilon > 0$ there exists a $C > 0$ such that

$$P^N(H(\hat{P}_N, P) \geq C N^{-1/2}) \leq \varepsilon \quad \text{for all } N.$$

Many results in asymptotic statistics do not hold for all underlying distributions $P \in \mathscr{P}$, but only for $P \in \mathscr{P} \setminus D$, where the exceptional set D is in some sense small compared to $\mathscr{P}$. For a finite dimensional parametric family $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, we may identify $\mathscr{P}$ with Θ, and the exceptional subset of Θ will typically be small in the sense that it has Lebesgue measure zero. On the more general spaces of distributions $\mathscr{P}$ that we consider in this paper, there is no obvious analogue of Lebesgue measure for which "small" sets can naturally be described as sets of measure zero. Moreover, our formulation of the consistency problem as well as our proofs of the results are largely topological rather than measure theoretic. It is therefore hardly surprising that the exceptional set D in our results will be small in a topological sense: D will be a *set of the first category* in $(\mathscr{P}, \mathbf{\Pi})$. We recall that a set of the first category is a countable union of nowhere dense sets, and that a set is *nowhere dense* in $(\mathscr{P}, \mathbf{\Pi})$ if its closure does not contain an open set in $(\mathscr{P}, \mathbf{\Pi})$.

We begin our study of the consistency of substitution estimators with an elementary observation. Suppose that the sequence $\tau_N \colon (\mathscr{P}, \mathbf{\Pi}) \to (\mathscr{R}, r)$ is *equicontinuous* on $\mathscr{P}$, that is, for every $P \in \mathscr{P}$ and $\varepsilon > 0$ there exists a neighborhood U_ε of P such that $r(\tau_N(P), \tau_N(Q)) < \varepsilon$ for all $Q \in U_\varepsilon$ and $N = 1, 2, \ldots$. Then consistency of $\hat{P}_N$ clearly implies consistency of $\tau_N(\hat{P}_N)$, since for every $P \in \mathscr{P}$ and $\varepsilon > 0$,

$$(1.6) \qquad P^N \left(r \left(\tau_N(\hat{P}_N), \tau_N(P) \right) \geq \varepsilon \right) \leq P^N \left(\hat{P}_N \notin U_\varepsilon \right) \to 0$$

as $N \to \infty$. Trivial though this observation may be, we shall dignify it by including it among the four theorems in this section.

THEOREM 1.1. *Suppose the following statements hold*:

(i) *The sequence of maps* $\tau_N \colon (\mathscr{P}, \mathbf{\Pi}) \to (\mathscr{R}, r)$ *is equicontinuous on* $\mathscr{P}$.
(ii) *There exists an estimator* $\hat{P}_N = p_N(X_1, \ldots, X_N)$ *of P with values in* $\mathscr{P}$, *which is* $\mathbf{\Pi}$*-consistent for* $P \in \mathscr{P}$.

Then $\tau_N(\hat{P}_N)$ *is an r-consistent estimator of* $\tau_N(P)$; *thus*,

$$(1.7) \qquad r \left(\tau_N(\hat{P}_N), \tau_N(P) \right) \to_P 0 \quad \text{for every } P \in \mathscr{P}.$$

We can push this argument a little bit further by assuming that $\mathscr{P} = \bigcup_{i \in I} \mathscr{P}_i$ for an arbitrary index set I and disjoint measurable $\mathscr{P}_i$, and that the assumptions of Theorem 1.1 hold on each $\mathscr{P}_i$ separately. If $\mathbf{\Pi}_i = \{U \cap \mathscr{P}_i : U \in \mathbf{\Pi}\}$ denotes the relative topology on $\mathscr{P}_i$, we have the following corollary:

COROLLARY 1.1. *Suppose that* $\mathscr{P} = \bigcup_{i \in I} \mathscr{P}_i$ *and that the following statements hold*:

(i) *For each* $i \in I$, *the sequence of maps* $\tau_N \colon (\mathscr{P}_i, \mathbf{\Pi}_i) \to (\mathscr{R}, r)$ *is equicontinuous on* $\mathscr{P}_i$.

(ii) *There exists a* $\boldsymbol{\Pi}$*-consistent estimator* $\hat{P}_N = p_N(X_1, \ldots, X_N)$ *of* $P \in \mathscr{P}$ *with the additional property that for each* $i \in I$, $P^N(\hat{P}_N \in \mathscr{P}_i) \to 1$ *for every* $P \in \mathscr{P}_i$.

Then $\tau_N(\hat{P}_N)$ *is an r-consistent estimator of* $\tau_N(P)$.

This result also follows directly from Theorem 1.1 by replacing the topology $\boldsymbol{\Pi}$ by the smallest topology containing $\boldsymbol{\Pi}_i$ for all $i \in I$. This has the effect of isolating the $\mathscr{P}_i$ from one another by making each $\mathscr{P}_i$ both open and closed. Note that assumption (ii) of Corollary 1.1 implies that $\hat{P}_N$ can serve as a test statistic for testing the hypothesis $P = P_i$ versus $P = P_j$, whenever $P_i \in \mathscr{P}_i$, $P_j \in \mathscr{P}_j$ and $i \neq j$. This test is asymptotically perfect in the sense that both error probabilities tend to zero as $N \to \infty$.

It is clear that the equicontinuity assumption for τ_N cannot be weakened much further unless one is willing to impose even more severe restrictions on the estimator $\hat{P}_N$. However, the equicontinuity of τ_N does merit further attention. It is often the case that τ_N converges pointwise to a function τ: $(\mathscr{P}, \boldsymbol{\Pi}) \to (\mathscr{R}, r)$, that is,

$$(1.8) \qquad r\big(\tau_N(P), \tau(P)\big) \to 0 \quad \text{for every } P \in \mathscr{P}.$$

We shall show that in this case continuity of each τ_N ensures equicontinuity of τ_N outside of a set of the first category. As a result we have the following theorem:

THEOREM 1.2.   *Suppose the following statements hold*:

(i) *For every N, the map* $\tau_N \colon (\mathscr{P}, \boldsymbol{\Pi}) \to (\mathscr{R}, r)$ *is continuous.*
(ii) *For every* $P \in \mathscr{P}$, $\tau_N(P)$ *converges to a limit* $\tau(P)$ *in* $(\mathscr{R}, r)$.
(iii) *There exists an estimator* $\hat{P}_N = p_N(X_1, \ldots, X_N)$ *of P with values in* $\mathscr{P}$, *which is* $\boldsymbol{\Pi}$*-consistent for* $P \in \mathscr{P}$.

Then there exists a set D of the first category in $(\mathscr{P}, \boldsymbol{\Pi})$ *such that the sequence* τ_N *is equicontinuous at every point* $P \in \mathscr{P} \setminus D$, *and hence*

$$(1.9) \qquad r\big(\tau_N(\hat{P}_N), \tau(P)\big) \to_P 0 \quad \text{for every } P \in \mathscr{P} \setminus D.$$

Since $\tau_N \to \tau$ in Theorem 1.2, we have replaced $\tau_N(P)$ by $\tau(P)$ in (1.9): consistent estimation of $\tau_N(P)$ and $\tau(P)$ amounts to the same thing in this case. We do insist, however, that the substitution estimator be of the form $\tau_N(\hat{P}_N)$, rather than $\tau(\hat{P}_N)$. This is because in applications such as the bootstrap one often has no way of knowing the functional form of τ. Nevertheless, the reader should note that under the assumptions of Theorem 1.2, $\tau(\hat{P}_N)$ is indeed a consistent estimator of $\tau(P)$ for $P \in \mathscr{P} \setminus D$, because equicontinuity of τ_N implies continuity of τ.

Let us briefly discuss the results stated so far. Theorem 1.1 makes it clear that, as long as we make no assumptions about the speed of convergence of $\hat{P}_N$ to P, the equicontinuity of the sequence τ_N is the key to consistency of

$\tau_N(\hat{P}_N)$. Of course one can reduce the severity of the equicontinuity assumption somewhat by placing restrictions in probability on the possible values of $\hat{P}_N$. Corollary 1.1 is an example of this. If $\tau_N \to \tau$, Theorem 1.2 provides a worst case scenario: the substitution estimator can only fail to be consistent on a set D of exceptional points, which is at most a set of the first category. Without further investigation, however, such a statement is of only limited practical value. After all, the true underlying distribution P may be one of the exceptional points. Also the convergence may of course be slow near these points. Thus Theorem 1.2 merely indicates the structure of the consistency problem rather than providing a complete solution. In any particular case one will have to investigate whether such exceptional points actually exist, and if so, where they are located. It often turns out that with a judicious choice of the topology Π and the estimator $\hat{P}_N$, there are no exceptional points and the substitution estimator will be consistent for all $P \in \mathcal{P}$.

This last remark may need further clarification. In applications, the choice of the metric r on $\mathcal{R}$ will usually be determined in advance by the type of consistency that one would like the substitution estimator to possess. On the other hand, the choice of the topology Π on $\mathcal{P}$, or of the metric p inducing it, is completely open to us. If Π is a coarse topology, it will be relatively easy to find a consistent estimator $\hat{P}_N$ of P, but relatively many sequences of maps τ_N will possess only limited continuity properties and the set of exceptional P for which $\tau_N(\hat{P}_N)$ is not consistent will be relatively large. Conversely, if Π is a fine topology, there will be few, if any, consistent $\hat{P}_N$, but having found one, it will produce substitution estimators $\tau_N(\hat{P}_N)$ which are consistent for relatively many sequences τ_N, except on relatively small sets of exceptional P. If the sequence τ_N is given in a particular application, the trick will be to find a topology which is fine enough to provide τ_N with sufficient continuity properties, yet coarse enough to admit a consistent estimate $\hat{P}_N$ of P. In Section 3 we illustrate this search for an appropriate topology and for an estimator $\hat{P}_N$ which is consistent in this topology by a number of examples.

Another point worth noting concerns our interpretation of a set D of the first category as a "small" set. In a certain sense, this is indeed correct if $(\mathcal{P}, \Pi)$ is topologically complete. In this case the category theorem asserts that $\mathcal{P} \setminus D$ is at least dense in $\mathcal{P}$ [cf. Dudley (1989), page 44]. In more general cases, however, D may be quite large. In fact, the entire space $\mathcal{P}$ may be of the first category in $(\mathcal{P}, \Pi)$ and we may have $D = \mathcal{P}$, so that Theorem 1.2 is vacuous. We discuss an example of this phenomenon in Section 3. Fortunately it turns out that the pathological character of this example is due to an unfortunate choice of the topology Π. A different choice of topology leads to an estimator $\hat{P}_N$ for which the substitution estimator $\tau_N(\hat{P}_N)$ is consistent for all $P \in \mathcal{P}$.

In the preceding paragraphs we have stressed the constructive aspects of our results so far by explaining how these results may be used to arrive at an estimator $\hat{P}_N$ which makes the substitution estimator $\tau_N(\hat{P}_N)$ consistent. However, one may also approach the consistency problem from a different angle and investigate the existence of a consistent substitution estimator

without worrying about its construction. For a result of this type, a logical assumption is the existence of a consistent estimator $T_N = t_N(X_1, \ldots, X_N)$ of $\tau_N(P)$. If no such estimator exists, there is no hope of finding a consistent substitution estimator.

THEOREM 1.3.   *Suppose the following statements hold*:
   (i) *For every N, the map $\tau_N: (\mathscr{P}, \mathbf{\Pi}) \to (\mathscr{R}, r)$ is measurable.*
   (ii) *The metric space $(\mathscr{R}, r)$ is separable.*
   (iii) *There exist an r-consistent estimator $T_N = t_N(X_1, \ldots, X_N)$ of $\tau_N(P)$ for $P \in \mathscr{P}$.*

Then there exists an estimator $\hat{P}_N = p_N(X_1, \ldots, X_N)$ with values in $\mathscr{P}$ such that $\tau_N(\hat{P}_N)$ is an r-consistent estimator of $\tau_N(P)$ for every $P \in \mathscr{P}$.

   In the proof of Theorem 1.3 we construct the estimator $\hat{P}_N$ explicitly on the basis of T_N. Hence, if T_N is not known to us, we cannot construct $\hat{P}_N$, and if it is known, it may not make much sense to construct $\hat{P}_N$ and $\tau_N(\hat{P}_N)$ since we already have a consistent estimator T_N of $\tau_N(P)$. Thus Theorem 1.3 should indeed be viewed purely as an existence statement to the effect that anything that can be estimated consistently at all, can be estimated consistently by a substitution estimator. The problem is of course to find an appropriate $\hat{P}_N$.
   The final result of this section allows us to construct a substitution estimator in some cases where T_N is not known, but its existence is. We consider the case where $(\mathscr{P}, \mathbf{\Pi})$ is a metric space $(\mathscr{P}, H)$, the τ_N are assumed to be continuous but not necessarily convergent and $\hat{P}_N$ is $\sqrt{N}$-Hellinger-consistent. If $\tau_N(P)$ can be estimated consistently at all, we show that $r(\tau_N(P_N), \tau_N(P)) \to 0$ for every sequence P_N with $H(P_N, P) = \mathscr{O}(N^{-1/2})$ and for all P outside of a set of the first category. Substituting $\hat{P}_N$ for P_N we find the following theorem:

THEOREM 1.4.   *Let the topology $\mathbf{\Pi}$ be metrized by the Hellinger metric H and suppose the following statements hold*:
   (i) *For every N, the map $\tau_N: (\mathscr{P}, \mathbf{\Pi}) \to (\mathscr{R}, r)$ is continuous.*
   (ii) *There exists an r-consistent estimator $T_N = t_N(X_1, \ldots, X_N)$ of $\tau_N(P)$ for $P \in \mathscr{P}$.*
   (iii) *There exists an estimator $\hat{P}_N = p_N(X_1, \ldots, X_N)$ with values in $\mathscr{P}$, which is $\sqrt{N}$-Hellinger-consistent for $P \in \mathscr{P}$.*

Then there exists a set D of the first category in $(\mathscr{P}, H)$ such that $\tau_N(\hat{P}_N)$ is an r-consistent estimator of $\tau_N(P)$ for $P \in \mathscr{P} \setminus D$, that is,

$$(1.10) \qquad r\big(\tau_N(\hat{P}_N), \tau_N(P)\big) \to_P 0 \quad \text{for every } P \in \mathscr{P} \setminus D.$$

   Because the requirement that $\hat{P}_N$ is $\sqrt{N}$-Hellinger-consistent may be somewhat unexpected, we shall show by means of a counterexample (Example 3.4 in Section 3) that this assumption is really needed. Ordinary Hellinger consistency is not sufficient. It is clear from the work of Le Cam (1973, 1986)

and Birgé (1983, 1986) that $\sqrt{N}$-Hellinger-consistent estimators will generally exist for finite dimensional families $\mathscr{P}$ where dimension is defined in terms of metric entropy. Typical examples of such families are parametric families $\mathscr{P} = \{P_\theta\colon \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, provided that Hellinger distance in $\mathscr{P}$ and Euclidean distance in Θ are compatible in some sense. For these families Theorem 1.4 enables us to find a substitution estimator that will work for "most" P if anything does. For infinite dimensional families $\mathscr{P}$, $\sqrt{N}$-Hellinger-consistent estimators of P will generally not exist, and without further assumptions on τ_N, one will generally not be able to construct a satisfactory estimator of $\tau_N(P)$ either.

We also note that for parametric families $\mathscr{P} = \{P_\theta\colon \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, it is possible to prove results similar to Theorem 1.4, where the exceptional set equals $D = \{P_\theta\colon \theta \in \Theta_0\}$ and Θ_0 has Lebesgue measure 0 [cf. Putter (1994)].

The two main results in this section are concerned with the interplay between conditions on τ_N and conditions on $\hat{P}_N$, needed to obtain reasonable substitution estimators $\tau_N(\hat{P}_N)$ of $\tau_N(P)$. Theorem 1.2 discusses what is needed for τ_N under the weakest possible condition (consistency) on $\hat{P}_N$. Theorem 1.4, on the other hand, operates under the weakest possible condition [estimability of $\tau_N(P)$] on τ_N.

In the remainder of the paper we proceed as follows. In Section 2 we apply the results of this section to the bootstrap and discuss the significance of our results in this context. Section 3 provides a number of examples that clarify the relationship between our results and standard bootstrap theory. Proofs of Theorems 1.2, 1.4 and 1.3 are given in Sections 4, 5 and 6, respectively.

2. The bootstrap. In the setup of the previous section, consider a sequence of random variables $Y_N = y_N(X_1, \ldots, X_N; P)$, where y_N is a measurable map from $\mathscr{X}^N \times \mathscr{P}$ to a *separable* metric space $(\mathscr{S}, s)$. Let $\mathscr{R}$ be the space of all probability distributions on $(\mathscr{S}, s)$ equipped with a metric r, which metrizes weak convergence. An obvious choice for r is Prohorov's metric ϱ. For distributions $R_1, R_2 \in \mathscr{R}$ this is defined by

$$(2.1) \quad \varrho(R_1, R_2) = \inf\{\varepsilon > 0\colon R_1(A) \leq R_2(A^\varepsilon) + \varepsilon, \text{ for all } A \in \mathscr{B}(\mathscr{S}, s)\},$$

where A^ε is an ε-neighborhood of A. Since $(\mathscr{S}, s)$ is separable, ϱ does indeed metrize weak convergence of probability measures in $\mathscr{R}$ [cf. Dudley (1989), Section 11.3], but of course other choices of r are also possible. Note that the separability of $(\mathscr{S}, s)$ also implies that $(\mathscr{R}, r)$ is separable [cf. Billingsley (1968), page 239]. Our aim is to estimate the law $\tau_N(P)$ of Y_N under P. Obviously $\tau_N(P) \in \mathscr{R}$.

As before, let us estimate P by $\hat{P}_N = p_N(X_1, \ldots, X_N)$ for a measurable map $p_N\colon \mathscr{X}^N \to \mathscr{P}$. With $\hat{P}_N$ as the resampling distribution, the bootstrap estimator of $\tau_N(P)$ is simply a substitution estimator $\tau_N(\hat{P}_N)$. To see this, note that if the resampling distribution is $\hat{P}_N$, the bootstrap estimates the distribution of Y_N by that of

$$Y_N^* = y_N(X_1^*, \ldots, X_N^*; \hat{P}_N),$$

where $X_1^*, \ldots, X_N^*$ are i.i.d. with distribution $\hat{P}_N$. However, this is just a description of $\tau_N(\hat{P}_N)$. The bootstrap estimate $\tau_N(\hat{P}_N)$ can be computed either analytically or by Monte Carlo simulation, but we shall not be concerned with that question here.

Note that the resampling distribution $\hat{P}_N$ is not necessarily the empirical distribution of $X_1, \ldots, X_N$, as is customary for the nonparametric bootstrap. In fact, our requirement that $\hat{P}_N$ takes its values in $\mathscr{P}$ prohibits this in many cases. If, for example, $\mathscr{P}$ is a parametric family $\{P_\theta : \theta \in \Theta\}$, our estimate $\hat{P}_N = P_{\hat{\theta}_N}$ will typically be based on an estimate $\hat{\theta}_N$ of the parameter θ and our bootstrap procedure will be the so-called *parametric bootstrap*. As we have indicated in Section 1, the purpose of this paper is to emphasize the importance of a judicious choice of the resampling distribution $\hat{P}_N$ so as to satisfy the requirements of our theorems. All but one of the examples in Section 3 will concern cases where the nonparametric bootstrap fails, but a proper choice of $\hat{P}_N$ will make the bootstrap work. On the one hand, this illustrates the great flexibility of the bootstrap method. On the other hand, it also shows that precise information about the behavior of the distribution $\tau_N(P)$ of Y_N as a function of P is needed to arrive at the correct resampling distribution $\hat{P}_N$. Unfortunately, such information may often not be available.

With the present choice of $(\mathscr{R}, r)$ and τ_N, Theorems 1.1–1.4 and Corollary 1.1 become results on the consistency of the bootstrap. For the sake of brevity, we shall not reformulate these results in this particular context. All the reader has to remember is that $\tau_N(P)$ is now the distribution of the random variable $Y_N = y_N(X_1, \ldots, X_N; P)$ taking values in a separable metric space, r is a metric metrizing weak convergence of probability distributions on this space and $\tau_N(\hat{P}_N)$ is the bootstrap estimate of $\tau_N(P)$ with $\hat{P}_N$ as the resampling distribution. As we pointed out above, assumption (ii) of Theorem 1.3 is automatically satisfied.

We begin by noting that the equicontinuity condition of Theorem 1.1 has been used to prove consistency of the bootstrap estimator ever since the beginning of research on bootstrap asymptotics [cf., e.g., Bickel and Freedman (1981) and Beran (1984)].

In Theorem 1.2, the choice of the metric r is irrelevant as long as it metrizes weak convergence of probability distributions on the separable metric space $(\mathscr{S}, s)$. Assumption (ii) of Theorem 1.2 now means that the distributions $\tau_N(P)$ of Y_N converge weakly to a limit distribution $\tau(P)$. The conclusion of Theorem 1.2 is that for $P \in \mathscr{P} \setminus D$, the bootstrap estimator $\tau_N(\hat{P}_N)$ converges weakly to $\tau(P)$ in probability.

The situation is more complicated in the remaining results where we do not require convergence of $\tau_N(P)$. The conclusions of these theorems refer to sequences $\tau_N(\hat{P}_N)$ and $\tau_N(P)$ of distributions for which the distance tends to zero in r-metric. This is not a matter which depends only on the topology of weak convergence which is induced by r. For different choices of the metric r metrizing weak convergence, $r(P_N, Q_N) \to 0$ may mean the same or different things [cf. Dudley (1989), Theorem 11.7.1 and problem 8 on page 313]. This problem disappears if the sequence $\tau_N(P)$ is uniformly tight.

In the context of the bootstrap, the assumption that the distributions $\tau_N(P)$ of Y_N converge weakly to a limit distribution $\tau(P)$ is important for another reason as well. It allows us to use the $(M\text{--}N)$-bootstrap $\tau_M(\hat{P}_N)$, where $M = M_N$ tends to infinity with N, but at a slower rate $M_N = \mathcal{o}(N)$. Since consistent estimation of $\tau_M(P)$, $\tau_N(P)$ or $\tau(P)$ amounts to the same thing in this case, the $(M\text{--}N)$-bootstrap may be viewed as an attempt to estimate $\tau_M(P)$ with the advantage of having at our disposal a resampling distribution $\hat{P}_N$ which is much closer to the underlying P than $\hat{P}_M$. As a result the $(M\text{--}N)$-bootstrap is consistent much more generally than the traditional $(N\text{--}N)$-bootstrap [cf. Politis and Romano (1994)]. Viewed in this light, we may weaken condition (iii) of Theorem 1.4 to Hellinger consistency at an arbitrarily slower rate $H(\hat{P}_N, P) = \mathcal{O}_P(a_N)$ with $a_N \to 0$, but $Na_N^2 \to \infty$, provided that we replace the bootstrap $\tau_N(\hat{P}_N)$ by the $(M\text{--}N)$-bootstrap $\tau_M(\hat{P}_N)$ with $M = a_N^{-2}$ and that $\tau_N(P) \to \tau(P)$. However, with these modifications, Theorem 1.4 is simply contained in Theorem 1.2 and it follows that we have nothing new to say about this method of improving the bootstrap by employing a smaller resample size. We are solely concerned with an appropriate choice of the resampling distribution $\hat{P}_N$.

The assumption (i) in Theorems 1.2 and 1.4 that τ_N is continuous for each N will generally not cause any problems. For most reasonable choices of the topology $\boldsymbol{\Pi}$ and the metric r, the distribution τ_N of Y_N for a fixed value of N would be continuous if Y_N did not depend on P. The direct dependence of Y_N on P is not likely to make matters worse, and the assumption that τ_N is continuous for every fixed N will be satisfied in all reasonable cases. It is the equicontinuity that may be lacking for certain P.

It was mentioned in Section 1 that for a parametric model $\mathscr{P} = \{P_\theta \colon \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, a $\sqrt{N}$-Hellinger-consistent parametric estimator $\hat{P}_N = P_{\hat{\theta}_N}$ will typically exist, provided that Hellinger distance in $\mathscr{P}$ and Euclidean distance in Θ are compatible in an appropriate sense. In this case, Theorem 1.4 asserts that for continuous τ_N, the parametric bootstrap with resampling distribution $P_{\hat{\theta}_N}$ will work for "most" P if anything does, even if $\tau_N(P)$ does not converge to a limit distribution $\tau(P)$.

As we pointed out at the end of Section 1, Theorems 1.2 and 1.4 deal with two extreme cases with minimal conditions on $\hat{P}_N$ and τ_N, respectively. In the context of the bootstrap this distinction attains an added significance. Before applying the bootstrap one should answer two questions:

1. What should one bootstrap?
2. How should one bootstrap?

The first of these questions refers in particular to choosing the proper dependence of Y_N—and hence of $\tau_N(P)$—on P. As a general rule one should do this in such a way that τ_N depends on P as little as possible. Theorem 1.2 suggests that whenever possible one should normalize Y_N so that its distribution $\tau_N(P)$ tends to a limit distribution $\tau(P)$. Any consistent choice of a resampling distribution $\hat{P}_N$ will then produce a bootstrap $\tau_N(\hat{P}_N)$ that works outside a set D of the first category. To get rid of this set D, one may search

for a topology $\boldsymbol{\Pi}$ on $\mathscr{P}$ which is fine enough to make τ_N equicontinuous on $\mathscr{P}$, and then for a $\boldsymbol{\Pi}$-consistent estimator $\hat{P}_N$. Unfortunately, this step will often be impossible for lack of the necessary knowledge of τ_N.

The second question refers to the choice of the resampling distribution $\hat{P}_N$. Theorem 1.4 asserts that even if we do not know how to normalize Y_N properly, the parametric bootstrap will generally still work on $\mathscr{P} \setminus D$. In nonparametric models, however, we had better make sure that τ_N converges.

In the extensive literature on the bootstrap it is usually shown that the bootstrap is strongly consistent, in the sense that $r(\tau_N(\hat{P}_N), \tau_N(P)) \to 0$ P-almost surely. For perfectly good reasons, strong consistency has not played an important role in the development of statistics so far, and hence we have been content to formulate our results in terms of ordinary (weak) consistency $r(\tau_N(\hat{P}_N), \tau_N(P)) \to_P 0$.

3. Examples. In this section we shall give some examples that illustrate the importance of choosing an appropriate resampling distribution $\hat{P}_N$ in applying the bootstrap. The first example exhibits a function $\tau_N(P)$ that is continuous in P with respect to Hellinger distance for every fixed N, but where the pointwise limit $\tau(P) = \lim_{N \to \infty} \tau_N(P)$ has a single discontinuity at a point P_0. It is shown that a parametric bootstrap fails in the point of discontinuity of τ. With a suitable metric that isolates that point, the equicontinuity is recaptured.

EXAMPLE 3.1. Let $\mathscr{P} = \{P_\alpha : 0 \le \alpha < 1/2\}$, where P_α is the probability distribution on $\mathbb{R}$ with distribution function F_α, defined for $0 < \alpha < 1/2$ by

$$(3.1) \qquad F_\alpha(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 1 - (1 + \alpha x)^{-1/\alpha}, & \text{if } x > 0, \end{cases}$$

and for $\alpha = 0$ by

$$(3.2) \qquad F_0(x) = \lim_{\alpha \to 0} F_\alpha(x) = \begin{cases} 0, & \text{if } x \le 0, \\ 1 - e^{-x}, & \text{if } x > 0. \end{cases}$$

Let $\mathscr{P}$ be equipped with Hellinger distance H and let the metric r on $\mathscr{R}$ be Lévy's metric. The Hellinger distance on $\mathscr{P}$ is related to Euclidean distance on the parameter space $[0, 1/2)$ by the relation

$$(3.3) \quad H(P_\alpha, P_\beta) = \frac{|\alpha - \beta|}{\sqrt{2(1 + \alpha)(1 + 2\alpha)}} + \mathscr{o}(|\alpha - \beta|) \quad \text{for } \alpha, \beta \in (0, 1/2)$$

and

$$(3.4) \qquad H(P_0, P_\alpha) = \frac{\alpha}{\sqrt{2}} + \mathscr{o}(\alpha).$$

We are interested in the distribution of the random variables

$$(3.5) \qquad Y_N = N^{-\alpha}(M_N - \log N) \quad \text{for } 0 \le \alpha < 1/2,$$

where M_N stands for $\max_{i=1,\ldots,N} X_i$. Note that Y_N depends on the underlying distribution through $N^{-\alpha}$.

Let $\tau_N(P_\alpha)$ be the law of Y_N when $X_1, \ldots, X_N$ are i.i.d. with distribution P_α and let $G_{N,\alpha}$ denote the distribution function of $\tau_N(P_\alpha)$. Then for $0 < \alpha < 1/2$,

$$G_{N,\alpha}(x) = P(N^{-\alpha}(M_N - \log N) \leq x) = [F_\alpha(N^\alpha x + \log N)]^N$$
$$= [1 - (1 + \alpha N^\alpha x + \alpha \log N)^{-1/\alpha}]^N, \qquad x > -N^{-\alpha} \log N,$$

and for $\alpha = 0$,

$$G_{N,0}(x) = P(M_N - \log N \leq x) = (1 - \exp(-x - \log N))^N, \quad x > -\log N.$$

For fixed N, it is easily seen that τ_N is continuous at P_α for $\alpha > 0$. Furthermore, $G_{N,\alpha}(x) \to G_{N,0}(x)$ as $\alpha \to 0$ for every $x > -\log N$ and hence $r(\tau_N(P_\alpha), \tau_N(P_0)) \to 0$ as $\alpha \to 0$, so that $\tau_N(P)$ is continuous at P_0. Hence, for each N, τ_N is clearly continuous on $\mathscr{P}$.

Now let N tend to infinity and let $\tau(P_\alpha)$ be the pointwise limit of $\tau_N(P_\alpha)$. The distribution function of $\tau(P_\alpha)$ will be denoted by G_α. Then for $0 < \alpha < 1/2$,

$$G_\alpha(x) = \lim_{N \to \infty} G_{N,\alpha}(x) = \exp(-(\alpha x)^{-1/\alpha}), \qquad x > 0,$$

and for $\alpha = 0$,

$$G_0(x) = \lim_{N \to \infty} G_{N,0}(x) = \exp(-e^{-x}), \qquad x \in \mathbb{R}.$$

We find that $\tau(P)$ is not continuous at P_0, since $\lim_{\alpha \to 0} G_\alpha(x) = 0$ for all x.

Application of Theorem 1.2 yields the existence of a set D of the first category in $(\mathscr{P}, H)$ such that the sequence τ_N is equicontinuous at P for all $P \in \mathscr{P} \setminus D$. Consequently if $\hat{P}_N$ is a Hellinger-consistent sequence of estimators of P, the bootstrap with resampling distribution $\hat{P}_N$ is consistent for all $P \in \mathscr{P} \setminus D$. Note that P_0 belongs to the exceptional set D since the limit τ is not continuous at P_0. A closer analysis reveals that τ_N is equicontinuous at P_α for $\alpha > 0$.

It appears therefore that P_0 is the only trouble spot in the model $\mathscr{P}$, so the problem can be resolved by choosing a metric on $\mathscr{P}$ that isolates P_0. Take, for instance,

$$(3.6) \qquad \pi(P, Q) = \begin{cases} H(P, Q), & \text{if } P, Q \neq P_0, \\ \sqrt{2}, & \text{if } P = P_0 \neq Q \text{ or } P \neq P_0 = Q. \end{cases}$$

Clearly π defines a metric on $\mathscr{P}$ and the sequence τ_N is trivially equicontinuous with respect to π at P_0, and hence on $\mathscr{P}$. A π-consistent estimator of P will have to satisfy $P_0^N(\hat{P}_N = P_0) \to 1$ and for all $P \in \mathscr{P} \setminus \{P_0\}$ both $P^N(\hat{P}_N = P_0) \to 0$ and $H(\hat{P}_N, P) \to_P 0$. If we set $\hat{P}_N = P_{\hat{\alpha}_N}$, then this implies that $\hat{\alpha}_N$ has to be a consistent estimate of α satisfying

$$P_0^N(\hat{\alpha}_N = 0) \to 1 \quad \text{and} \quad P_\alpha^N(\hat{\alpha}_N = 0) \to 0 \quad \text{for } 0 < \alpha < 1/2.$$

It is indeed possible to detect the isolated point P_0 in $\mathscr{P}$ with probability tending to 1 by choosing, for instance,

$$(3.7) \qquad \hat{\alpha}_N = \begin{cases} 0, & \text{if } M_N \leq 2 \log N, \\ \left[1 - \left(\overline{X}_N \right)^{-1} \right]^+, & \text{otherwise.} \end{cases}$$

Since τ_N is equicontinuous with respect to π and $\pi(P_{\hat{\alpha}_N}, P) \to_P 0$, for every $P \in \mathscr{P}$, the bootstrap with resampling distribution $P_{\hat{\alpha}_N}$ is consistent for every P in $\mathscr{P}$. Note that this result may also be obtained directly by applying Corollary 1.1 combined with (3.7), but it seemed instructive to exhibit a metric π that separates $\{P_0\}$ and $\mathscr{P} \setminus \{P_0\}$ explicitly.

EXAMPLE 3.2.   If the class of all possible distributions $\mathscr{P}$ is complete, the exceptional set of the first category, appearing in the result of Theorems 1.2 and 1.4, is small in the sense that its complement is dense in $\mathscr{P}$. If $\mathscr{P}$ is not complete, however, these sets can be quite large. In this example we discuss a particular statistical model $\mathscr{P}$ equipped with Hellinger metric H, such that $\mathscr{P}$ is of the first category in $(\mathscr{P}, H)$. This model is not an artificial construct, but it is the natural model for a statistical situation of interest.

Let $\mathscr{P}$ be the class of probability distributions P on $(0, \infty)$ with distribution functions F satisfying

$$(3.8) \qquad \lim_{x \searrow 0} \frac{F(x)}{x} = a(P) \in (0, \infty).$$

Let $X_1, X_2, \ldots$ be i.i.d. random variables taking values in $(0, \infty)$ with unknown common distribution P in $\mathscr{P}$ and distribution function F. Consider the random variable

$$(3.9) \qquad Y_N = N \min\{X_1, \ldots, X_N\}$$

and let $\tau_N(P)$ be the distribution of Y_N under P. Note that $\mathscr{P}$ is precisely the class of underlying distributions P for which $\tau_N(P)$ converges to a nondegenerate limit $\tau(P)$, which is an exponential distribution with parameter $a(P)$. Let $\mathscr{P}$ be equipped with Hellinger distance H. Then assumptions (i) and (ii) of Theorem 1.2 are satisfied. It is shown in Putter and van Zwet (1994) that application of Theorem 1.2 does not yield any positive information in the sense that the exceptional set D appearing in the conclusion of the theorem equals the entire space $\mathscr{P}$. The aforementioned paper also contains a direct proof that $\mathscr{P}$ is indeed a set of the first category in $(\mathscr{P}, H)$. Luckily, all this trouble is caused only by a wrong choice of the metric p. If we define

$$(3.10) \qquad \pi(P, Q) = \sup_{x > 0} \frac{|F(x) - G(x)|}{x},$$

where F and G denote the distribution functions corresponding to P and Q, then it is shown in Putter and van Zwet (1994) that the sequence τ_N is equicontinuous with respect to π for all $P \in \mathscr{P}$. A π-consistent estimator $\hat{P}_N$

of P is also provided. Its distribution function $\hat{F}_N$ is given by

$$(3.11) \qquad \hat{F}_N(x) = \begin{cases} \dfrac{F_N(\xi_N)}{\xi_N}\, x, & \text{if } 0 < x < \xi_N, \\ F_N(x), & \text{if } x \geq \xi_N, \end{cases}$$

where F_N denotes the empirical distribution function and ξ_N is a sequence of positive numbers converging to zero with $N\xi_N \to \infty$. It follows that the bootstrap with $\hat{P}_N$ as resampling distribution works for all P in $\mathscr{P}$.

In practice, this example occurs in a slightly modified form. Instead of $X_1, \ldots, X_N$, one observes $Z_1 = \theta + X_1, \ldots, Z_N = \theta + X_N$ for the purpose of estimating the parameter $\theta \in \mathbb{R}$ which is the lower endpoint of the support of the distribution of the Z_i. When using $\min(Z_1, \ldots, Z_N)$ as an estimator of θ, one is indeed interested in the distribution of $Y_N = N(\min(Z_1, \ldots, Z_N) - \theta) = N\min(X_1, \ldots, X_N)$. Obviously, $\hat{P}_N$ as defined by (3.11) cannot be used for the resampling distribution of $X_1^*, \ldots, X_N^*$, since the empirical distribution function F_N of $X_1, \ldots, X_N$ is now unknown. However, a slight modification will work. If G_N denotes the empirical distribution function of $Z_1, \ldots, Z_N$, one can estimate the distribution P of X_1 by a distribution $\bar{P}_N$ with distribution function.

$$\bar{F}_N(x) = \begin{cases} \dfrac{G_N(\min(Z_1, \ldots, Z_N) + \xi_N)}{\xi_N}\, x, & \text{if } 0 < x < \xi_N, \\ G_N(\min(Z_1, \ldots, Z_N) + x), & \text{if } x \geq \xi_N, \end{cases}$$

where $\xi_N \to 0$ and $N\xi_N \to \infty$. It is easy to see that $\pi(\bar{P}_N, \hat{P}_N) \to_P 0$, so that $\pi(\bar{P}_N, P) \to_P 0$ and the resampling distribution $\bar{P}_N$ will produce a consistent bootstrap for all $P \in \mathscr{P}$.

EXAMPLE 3.3 (Superefficiency). Another example, related to Example 3.1, is provided by Beran (1982). Consider an i.i.d. sequence $X_1, \ldots, X_N$ with a common normal distribution P_θ with unknown mean θ and unit variance. The Hodges estimator of θ is given by

$$(3.12) \qquad T_N = \begin{cases} \bar{X}_N, & \text{if } |\bar{X}_N| > N^{-1/4}, \\ b\bar{X}_N, & \text{if } |\bar{X}_N| \leq N^{-1/4}, \end{cases}$$

where $\bar{X}_N = (1/N)\sum_{i=1}^N X_i$ and $b \in (0,1)$. We wish to find a bootstrap estimate for the distribution $\tau_N(P_\theta)$ of

$$Y_N = N^{1/2}(T_N - \theta)$$

under P_θ. We equip the class $\mathscr{P} = \{P_\theta : \theta \in \mathbb{R}\}$ with the Euclidean metric d in the parameter space, that is, $d(P_\theta, P_{\theta'}) = |\theta - \theta'|$. Thus, in effect we are identifying $\mathscr{P}$ and its parameter space $\mathbb{R}$. Let $(\mathscr{R}, l)$ denote the class of all distributions on $\mathbb{R}$ equipped with the Lévy metric l. Consider τ_N as a map from $(\mathscr{P}, d)$ to $(\mathscr{R}, l)$.

If we denote the distribution function corresponding to $\tau_N(P_\theta)$ by $G_{N,\theta}$, we find

$$G_{N,\theta}(x) = \begin{cases} \Phi(x), & \text{if } N^{-1/4}|x + \theta N^{1/2}| \geq 1, \\ \Phi\left(\dfrac{x + (1 - b)\theta N^{1/2}}{b}\right), & \text{if } N^{-1/4}|x + \theta N^{1/2}| \leq b, \\ \Phi(-N^{1/4} - \theta N^{1/2}), & \text{if } -1 < N^{-1/4}(x + \theta N^{1/2}) < -b, \\ \Phi(N^{1/4} - \theta N^{1/2}), & \text{if } b < N^{-1/4}(x + \theta N^{1/2}) < 1, \end{cases}$$

where Φ denotes the standard normal distribution function. It follows that the pointwise limit G_θ of $G_{N,\theta}$ is given by

$$(3.13) \qquad G_\theta(x) = \begin{cases} \Phi(x/b), & \text{if } \theta = 0, \\ \Phi(x), & \text{otherwise.} \end{cases}$$

This implies that the limit $\tau(P_\theta)$ of $\tau_N(P_\theta)$ is a normal distribution with variance b^2 if $\theta = 0$ and unity otherwise. Since $0 < b < 1$, the Hodges estimator is superefficient at $\theta = 0$.

Obviously, τ_N is continuous at every P_θ and since $\tau_N \to \tau$, Theorem 1.2 applies, and hence the sequence τ_N is equicontinuous on $\mathscr{P} \setminus D$, where D is of the first category. As τ has a discontinuity at P_0, this distribution clearly belongs to D. Observing that $G_{N,\theta}(x) = G_{N,\theta'}(x) = \Phi(x)$ if both $|x + \theta N^{1/2}| \geq N^{1/4}$ and $|x + \theta' N^{1/2}| \geq N^{1/4}$, we see that $l(\tau_N(P_\theta), \tau_N(P_{\theta'}))$ can be made arbitrarily small for θ' in a small neighborhood of a fixed $\theta \neq 0$ and large N, so that τ_N is equicontinuous at every P_θ with $\theta \neq 0$. Hence, $D = \{P_0\}$, the single point of discontinuity of the limit distribution τ.

Of course this does not imply that

$$(3.14) \qquad \liminf_N l\left(\tau_N(P_0), \tau_N(P_{\theta_N})\right) > 0$$

for sequences $\theta_N \to 0$. However, as Beran (1982) points out, (3.14) does hold for sequences θ_N converging at rate $N^{-1/2}$. To see this, notice that for such sequences

$$G_{N,\theta_N}(x) = \Phi\left(\frac{x + (1 - b)\theta_N N^{1/2}}{b}\right) + o(1).$$

In fact, for every $\varepsilon > 0$, (3.14) holds uniformly for $|\theta_N| \geq \varepsilon N^{-1/2}$. This shows that for $\hat{\theta}_N = \bar{X}_N$, for instance, the parametric bootstrap with resampling distribution $P_{\hat{\theta}_N}$ will work for $\theta \neq 0$, but fails for $\theta = 0$.

In Example 3.1 we have shown how to deal with a situation like this. By applying Corollary 1.1, we find that all we have to do to make the parametric bootstrap work is to modify the estimator $\hat{\theta}_N = \bar{X}_N$ to ensure that

$$P_0^N(\hat{\theta}_N = 0) \to 1, \qquad P_\theta^N(\hat{\theta}_N = 0) \to 0 \quad \text{for } \theta \neq 0.$$

Choosing

$$\hat{\theta}_N = \begin{cases} \overline{X}_N, & \text{if } |\overline{X}_N| > N^{-1/4}, \\ 0, & \text{if } |\overline{X}_N| \le N^{-1/4}, \end{cases}$$

which is the Hodges estimator for $b = 0$, will accomplish this and the corresponding parametric bootstrap $\tau_N(P_{\hat{\theta}_N})$ will work for all θ.

The reason we discuss the Hodges estimator T_N is that it is perhaps the best known example of an estimator which is superefficient for a single parameter value $\theta = 0$. Le Cam (1953) has pointed out that one can modify T_N in an obvious way to construct an estimator which is superefficient for all θ belonging to a countable closed set in $\mathbb{R}$. Moreover, Le Cam showed that an estimator of θ can only be superefficient on a set of the first category in $\mathbb{R}$ equipped with the Euclidean metric. Since superefficiency can only occur at points where τ_N is not equicontinuous, this may be viewed as a consequence of Theorem 1.2.

EXAMPLE 3.4. Our last example concerns the question of existence of a consistent bootstrap estimator. It may also clarify why a $\sqrt{N}$-Hellinger-consistent estimator $\hat{P}_N$ is needed in Theorem 1.4.

Let $X_1, X_2, \ldots$ be i.i.d. random variables with a common normal distribution with expectation $\theta \in \mathbb{R}$ and variance 1, which we shall indicate as P_θ or $\mathcal{N}(\theta, 1)$. Define $\overline{X}_N = (1/N)\Sigma_{i=1}^N X_i$ and

$$Y_N = N^{1/2}(\overline{X}_N - a_N \theta).$$

We distinguish three different cases.

Case (i): $a_N \equiv 1$. The distribution $\tau_N(P_\theta)$ of Y_N is $\mathcal{N}(0, 1)$ independent of θ, which can obviously be estimated consistently for any metric r on $\mathscr{R}$. Also the sequence τ_N is equicontinuous for any topology $\boldsymbol{\Pi}$ on $\mathscr{P} = \{P_\theta: \theta \in \mathbb{R}\}$ and any metric r on $\mathscr{R}$. As we can choose any $\boldsymbol{\Pi}$ and r, Theorem 1.1 ensures that the bootstrap $\tau_N(P_{\hat{\theta}_N})$ equals the true distribution $\tau_N(P_\theta)$ for any "estimator" $\hat{\theta}_N$ of θ, consistent or not.

Case (ii): $a_N \equiv 0$. Now $\tau_N(P_\theta)$ is $\mathcal{N}(N^{1/2}\theta, 1)$, which cannot be estimated consistently in Prohorov or Lévy metric. The reason for this is that any estimator of θ has an error which is at least of order $N^{-1/2}$ in probability and as a result $N^{1/2}\theta$ cannot be estimated consistently. It follows that there is no consistent bootstrap estimator of $\tau_N(P_\theta)$ either.

Case (iii): $a_N = 1 - \varepsilon_N$, $\varepsilon_N \searrow 0$. Now $\tau_N(P_\theta)$ is $\mathcal{N}(\varepsilon_N N^{1/2}\theta, 1)$ which can be estimated consistently by $\mathcal{N}(\varepsilon_N N^{1/2}\overline{X}_N, 1)$ in Prohorov or Lévy metric. If $\hat{\theta}_N$ is an estimator of θ, the bootstrap $\tau_N(P_{\hat{\theta}_N})$ will work if and only if

$$\varepsilon_N N^{1/2}(\hat{\theta}_N - \theta) \to_{P_\theta} 0$$

for every $\theta \in \mathbb{R}$. This is true for every sequence $\varepsilon_N \searrow 0$ if and only if $\hat{\theta}_N - \theta = \mathscr{O}_{P_\theta}(N^{-1/2})$ or $H(P_{\hat{\theta}_N}, P_\theta) = \mathscr{O}_{P_\theta}(N^{-1/2})$. It follows that the assumption in Theorem 1.4 that $\hat{P}_N$ is $\sqrt{N}$-Hellinger-consistent cannot be relaxed. The bootstrap $\tau_N(P_{\overline{X}_N})$, which incidentally is the same as $\mathcal{N}(\varepsilon_N N^{1/2}\overline{X}_N, 1)$, is obviously consistent.

4. Proof of Theorem 1.2. In this section $(\mathscr{P}, \mathit{\Pi})$ will be a topological space, $(\mathscr{R}, r)$ a metric space and τ_N a sequence of continuous maps from $(\mathscr{P}, \mathit{\Pi})$ to $(\mathscr{R}, r)$, converging to a limit $\tau \colon (\mathscr{P}, \mathit{\Pi}) \to (\mathscr{R}, r)$.

DEFINITION. τ_N is *locally uniformly convergent* at P_0 if for every $\varepsilon > 0$ there exists a neighborhood U_ε of P_0 and a number N_ε such that $r(\tau_N(P), \tau(P)) \le \varepsilon$ for all $N \ge N_\varepsilon$ and for all $P \in U_\varepsilon$.

DEFINITION. τ_N is *equicontinuous* at P_0 if for every $\varepsilon > 0$ there exists a neighborhood U_ε of P_0 such that $P \in U_\varepsilon$ implies

$$\sup_N r(\tau_N(P), \tau_N(P_0)) \le \varepsilon.$$

For the sequence of continuous maps τ_N, define

(4.1) $E_1 = \{P \in \mathscr{P} \colon \tau_N \text{ is locally uniformly convergent at } P\}$,

(4.2) $E_2 = \{P \in \mathscr{P} \colon \tau_N \text{ is equicontinuous at } P\}$.

The following lemma asserts that E_1 and E_2 are equal.

LEMMA 4.1.   *Suppose that τ_N is continuous for $N = 1, 2, \ldots$ and that τ_N converges pointwise to a limit τ. Then τ_N is equicontinuous at $P_0 \in \mathscr{P}$ if and only if τ_N is locally uniformly convergent at P_0.*

PROOF.   Suppose that τ_N is locally uniformly convergent at P_0 and fix $\varepsilon > 0$. Then there exists a neighborhood U_ε of P_0 and an integer N_ε such that

$$r(\tau_N(P), \tau(P)) \le \varepsilon \quad \text{for } P \in U_\varepsilon \text{ and } N \ge N_\varepsilon.$$

Hence, for $P \in U_\varepsilon$ and $N \ge N_\varepsilon$,

$$r(\tau_N(P), \tau_N(P_0)) \le r(\tau_N(P), \tau_{N_\varepsilon}(P)) + r(\tau_N(P_0), \tau_{N_\varepsilon}(P_0))$$
$$+ r(\tau_{N\varepsilon}(P), \tau_{N_\varepsilon}(P_0))$$
$$\le 4\varepsilon + r(\tau_{N_\varepsilon}(P), \tau_{N_\varepsilon}(P_0)).$$

Since τ_N is continuous for every N, there exists a neighborhood U_ε' of P_0 such that for $P \in U_\varepsilon'$, $r(\tau_N(P), \tau_N(P_0)) \le \varepsilon$ for $N \le N_\varepsilon$, so

$$r(\tau_N(P), \tau_N(P_0)) \le 5\varepsilon$$

for $P \in U_\varepsilon \cap U_\varepsilon'$ and all N. Hence τ_N is equicontinuous at P_0.

Conversely, suppose that τ_N is equicontinuous at P_0. Fix $\varepsilon > 0$. Then there exists a neighborhood U_ε of P_0 such that $r(\tau_N(P), \tau_N(P_0)) \le \varepsilon$ for $P \in U_\varepsilon$ and all N. Since $\tau_N \to \tau$, this implies that $r(\tau(P), \tau(P_0)) \le \varepsilon$ for $P \in U_\varepsilon$. Hence, for $P \in U_\varepsilon$ and all N,

$$r(\tau_N(P), \tau(P)) \le r(\tau_N(P), \tau_N(P_0)) + r(\tau(P), \tau(P_0)) + r(\tau_N(P_0), \tau(P_0))$$
$$\le 2\varepsilon + r(\tau_N(P_0), \tau(P_0)).$$

As $\tau_N \to \tau$, there exists N_ε such that for all $P \in U_\varepsilon$ and $N \ge N_\varepsilon$,

$$r(\tau_N(P), \tau(P)) \le 3\varepsilon,$$

so τ_N is locally uniformly convergent at P_0. $\square$

LEMMA 4.2. *Suppose that τ_N is continuous for each N and that τ_N converges pointwise to a limit τ. Let E_1 and E_2 be defined as in* (4.1) *and* (4.2). *Then E_1^c (and hence also E_2^c) is a set of the first category in $(\mathscr{P}, \mathit{\Pi})$.*

PROOF. The sequence of maps τ_N is locally uniformly convergent at P_0 iff for every m there exists $M = M(m)$ such that P_0 is an interior point of

$$G_{mM} = \left\{ P : r\big(\tau_N(P), \tau(P)\big) \leq \frac{1}{2m} \text{ for all } N \geq M \right\},$$

and hence an interior point of the larger set

$$F_{mM} = \big\{ P : r\big(\tau_N(P), \tau_{N'}(P)\big) \leq 1/m \text{ for all } N, N' \geq M \big\}.$$

As τ_N is continuous, F_{mM} is the intersection of closed sets and is therefore closed. Since $\tau_N \to \tau$, clearly $\bigcup_{M=1}^{\infty} F_{mM} = \mathscr{P}$ for every m.

Let $\mathring{F}_{mM}$ denote the interior of F_{mM}. The sequence τ_N is not locally uniformly convergent at P iff there exists m such that $P \notin \bigcup_M \mathring{F}_{mM}$, that is, iff for some m,

$$P \in \bigcup_{M=1}^{\infty} F_{mM} \setminus \bigcup_{M'=1}^{\infty} \mathring{F}_{mM'}.$$

Hence

$$E_1^c = \bigcup_{m=1}^{\infty} \left(\bigcup_{M=1}^{\infty} F_{mM} \setminus \bigcup_{M'=1}^{\infty} \mathring{F}_{mM'} \right) \subset \bigcup_{m=1}^{\infty} \bigcup_{M=1}^{\infty} \big(F_{mM} \setminus \mathring{F}_{mM} \big).$$

Since F_{mM} is closed, $F_{mM} \setminus \mathring{F}_{mM}$ is a closed set with empty interior and therefore nowhere dense. It follows that the set $\bigcup_m \bigcup_M (F_{mM} \setminus \mathring{F}_{mM})$ is a countable union of nowhere dense sets, and hence of the first category and a fortiori so is E_1^c. The lemma is proved. □

PROOF OF THEOREM 1.2. When τ_N is equicontinuous at P and $\hat{P}_N$ is a consistent estimator of P, it is clear from the argument leading to Theorem 1.1 that $r(\tau_N(\hat{P}_N), \tau_N(P)) \to_P 0$. It follows from Lemmas 4.1 and 4.2 that there exists a set D of the first category in $\mathscr{P}$ such that τ_N is equicontinuous at all $P \in \mathscr{P} \setminus D$. The theorem is proved. □

5. Proof of Theorem 1.4. In this section we shall assume that $\mathscr{P}$ is a metric space, equipped with Hellinger distance H and that $\tau_N \colon (\mathscr{P}, H) \to (\mathscr{R}, r)$ is continuous for each N. We shall omit the assumption that τ_N converges pointwise. For the proof of Theorem 1.4 we follow an entirely different path. Let us first collect some results about Hellinger distance. Suppose that for each $N = 1, 2, \ldots$, $(\mathscr{X}_N, \mathscr{A}_N)$ is a measurable space and Q_{1N} and Q_{2N} are probability measures on $(\mathscr{X}_N, \mathscr{A}_N)$ with densities q_{1N} and q_{2N} with respect to a σ-finite dominating measure μ_N. Let us define an *asymptotically perfect test* for distinguishing between $\{Q_{1N}\}$ and $\{Q_{2N}\}$ as a sequence of tests for Q_{1N} against Q_{2N} for which the probabilities of errors of both type I

and type II tend to zero as N tends to infinity. Existence of such an asymptotically perfect test between $\{Q_{1N}\}$ and $\{Q_{2N}\}$ is related to the Hellinger distance between Q_{1N} and Q_{2N}. In particular, an asymptotically perfect test cannot exist if $\limsup_{N \to \infty} H^2(Q_{1N}, Q_{2N}) < 2$.

LEMMA 5.1. *Suppose that* $\limsup_{N \to \infty} H^2(Q_{1N}, Q_{2N}) < 2$. *Then for any sequence* $A_N \in \mathscr{A}_N$,

$$\liminf_{N \to \infty} (Q_{1N}(A_N) + Q_{2N}(A_N^c)) > 0.$$

PROOF.   We can write

$$(5.1) \quad H^2(Q_{1N}, Q_{2N}) = \int (q_{1N}^{1/2} - q_{2N}^{1/2})^2 \, d\mu_N = 2 - 2\int (q_{1N} q_{2N})^{1/2} \, d\mu_N,$$

and hence $\liminf \int (q_{1N} q_{2N})^{1/2} \, d\mu_N > 0$. The Cauchy–Schwarz inequality ensures that for every N and for any set $A_N \in \mathscr{A}_N$,

$$\int (q_{1N} q_{2N})^{1/2} \, d\mu_N = \int_{A_N} (q_{1N} q_{2N})^{1/2} \, d\mu_N + \int_{A_N^c} (q_{1N} q_{2N})^{1/2} \, d\mu_N$$

$$\leq \left(\int_{A_N} q_{1N} \, d\mu_N \int_{A_N} q_{2N} \, d\mu_N \right)^{1/2}$$

$$+ \left(\int_{A_N^c} q_{1N} \, d\mu_N \int_{A_N^c} q_{2N} \, d\mu_N \right)^{1/2}$$

$$\leq (Q_{1N}(A_N))^{1/2} + (Q_{2N}(A_N^c))^{1/2}.$$

The lemma follows. □

Let P and P_N be probability measures on a measurable space $(\mathscr{X}, \mathscr{A})$ with densities p and p_N with respect to a σ-finite measure μ on $(\mathscr{X}, \mathscr{A})$. On the product measurable space $(\mathscr{X}^N, \mathscr{A}^N)$ we define the product measures $Q_{1N} = P^N$ and $Q_{2N} = P_N^N$ with densities $\prod_{i=1}^{N} p(x_i)$ and $\prod_{i=1}^{N} p_N(x_i)$ with respect to $\mu_N = \mu^N$. By (5.1) we have

$$1 - \tfrac{1}{2} H^2(Q_{1N}, Q_{2N}) = \int \cdots \int \prod_{i=1}^{N} \{p(x_i) p_N(x_i)\}^{1/2} \, d\mu_N$$

$$(5.2) \qquad\qquad = \left[\int \{p(x) p_N(x)\}^{1/2} \, d\mu \right]^N$$

$$= \left[1 - \tfrac{1}{2} H^2(P, P_N) \right]^N.$$

It follows that $H(P, P_N) = \mathscr{O}(N^{-1/2})$ implies that $\limsup_N H^2(Q_{1N}, Q_{2N}) < 2$, and Lemma 5.1 yields the following corollary:

COROLLARY 5.1. *Suppose* $H(P, P_N) = \mathscr{O}(N^{-1/2})$. *Then for any sequence* $A_N \in \mathscr{A}^N$,

$$\liminf_{N \to \infty} (P^N(A_N) + P_N^N(A_N^c)) > 0.$$

LEMMA 5.2. *Suppose that* $\tau_N\colon (\mathscr{P}, H) \to (\mathscr{R}, r)$ *is continuous for every* N *and that* $T_N = t_N(X_1, \ldots, X_N)$ *is a consistent estimator of* $\tau_N(P)$, *that is,*

$$r(T_N, \tau_N(P)) \to_P 0 \quad \text{for every } P \in \mathscr{P}.$$

Then there exists a set D *of the first category in* $(\mathscr{P}, H)$ *such that for every* $P_0 \in \mathscr{P} \setminus D$, *every* $\varepsilon > 0$ *and every sequence* $\delta_N \searrow 0$,

$$(5.3) \qquad \lim_{N \to \infty} \sup_{\{P\colon H(P, P_0) \le \delta_N\}} P^N\big(r(T_N, \tau_N(P)) \ge \varepsilon\big) = 0.$$

PROOF. Fix an integer $k > 0$ and define $\psi_N^{(k)}\colon (\mathscr{P}, H) \to \mathbb{R}$ with Euclidean distance by

$$\psi_N^{(k)}(P) = P^N\big(r(T_N, \tau_N(P)) \ge 1/k\big), \qquad N = 1, 2, \ldots.$$

Clearly, $\psi_N^{(k)}(P) \to 0$ as $N \to \infty$ for every $P \in \mathscr{P}$. Since we would like to apply Lemma 4.2 to $\psi_N^{(k)}$, we would also need continuity and hence we modify $\psi_N^{(k)}$ slightly: choose $\delta_N \searrow 0$ and define

$$\tilde{\psi}_N^{(k)}(P) = \frac{1}{\delta_N} \int_{1/k}^{1/k + \delta_N} P^N\big(r(T_N, \tau_N(P)) \ge u\big) \, du.$$

We have

$$(5.4) \qquad 0 \le \tilde{\psi}_N^{(k)}(P) \le \psi_N^{(k)}(P) \le \tilde{\psi}_N^{(k+1)}(P),$$

the last inequality for $N \ge N_0 = N_0(k)$, such that $\delta_{N_0} \le 1/(k(k+1))$. Therefore, $\tilde{\psi}_N^{(k)} \to 0$ on $\mathscr{P}$, but $\tilde{\psi}_N^{(k)}$ is also continuous on $\mathscr{P}$. To see this, note that for any $P_1, P_2 \in \mathscr{P}$ and $A \in \mathscr{A}^N$, $|P_1^N(A) - P_2^N(A)| \le H(P_1^N, P_2^N)$, and for fixed N, we can make this arbitrarily small by taking $H(P_1, P_2)$ small [cf. (5.2)]. Hence, for every fixed N,

$$\left| \tilde{\psi}_N^{(k)}(P_1) - \frac{1}{\delta_N} \int_{1/k}^{1/k + \delta_N} P_2^N\big(r(T_N, \tau_N(P_1)) \ge u\big) \, du \right|$$

can be made as small as we wish to taking $H(P_1, P_2)$ small. Since τ_N is continuous and the integral defining $\tilde{\psi}_N^{(k)}$ depends continuously on the upper and lower bound of the range of integration, the same is true for

$$\left| \frac{1}{\delta_N} \int_{1/k}^{1/k + \delta_N} P_2^N\big(r(T_N, \tau_N(P_1)) \ge u\big) \, du - \tilde{\psi}_N^{(k)}(P_2) \right|,$$

which proves the continuity of each of the functions $\tilde{\psi}_N^{(k)}$.

Application of Lemma 4.2 yields the existence of a set $D^{(k)}$ of the first category in $(\mathscr{P}, H)$ such that $\{\tilde{\psi}_N^{(k)}\}$ is locally uniformly convergent at P_0 for every $P_0 \in \mathscr{P} \setminus D^{(k)}$, so a fortiori

$$(5.5) \qquad \sup_{\{P\colon H(P, P_0) \le \delta_N\}} \tilde{\psi}_N^{(k)}(P) \to 0$$

for every $P_0 \in \mathscr{P} \setminus D^{(k)}$ and every sequence $\delta_N \searrow 0$. Taking $D = \bigcup_{k=1}^\infty D^{(k)}$ and noting that D is also of the first category, we find that (5.5) holds for all $k = 1, 2, \ldots$, provided that $P_0 \in \mathscr{P} \setminus D$. Because of (5.4), this implies the same for $\psi_N^{(k)}$ itself, and hence we obtain (5.3) and the lemma. □

LEMMA 5.3.   *Suppose that $\tau_N \colon (\mathscr{P}, H) \to (\mathscr{R}, r)$ is continuous for every N and that $T_N = t_N(X_1, \ldots, X_N)$ is a consistent estimator of $\tau_N(P)$. Then there exists a set D of the first category in $(\mathscr{P}, H)$ such that for every $P_0 \in \mathscr{P} \setminus D$ and every $C > 0$,*

$$(5.6) \qquad \sup_{\{P \colon H(P, P_0) \le CN^{-1/2}\}} r\big(\tau_N(P), \tau_N(P_0)\big) \to 0.$$

PROOF.   According to Lemma 5.2 we can choose a set D of the first category in $(\mathscr{P}, H)$ such that for every $P_0 \in \mathscr{P} \setminus D$, every $C > 0$ and every $\varepsilon > 0$,

$$(5.7) \qquad \sup_{\{P \colon H(P, P_0) \le CN^{-1/2}\}} P^N\big(r(T_N, \tau_N(P)) \ge \varepsilon\big) \to 0.$$

Fix $P_0 \in \mathscr{P} \setminus D$ and take any sequence $P_N \in \mathscr{P}$ with $H(P_N, P_0) \le CN^{-1/2}$. For $N = 1, 2, \ldots,$ define

$$A_N = \big\{r(T_N, \tau_N(P_0)) \ge r(T_N, \tau_N(P_N))\big\} \in \mathscr{A}^N.$$

On A_N we have

$$r(T_N, \tau_N(P_0)) \ge 1/2\{r(T_N, \tau_N(P_N)) + r(T_N, \tau_N(P_0))\}$$
$$\ge 1/2 r\big(\tau_N(P_N), \tau_N(P_0)\big),$$

and similarly on A_N^c,

$$r(T_N, \tau_N(P_N)) > 1/2 r\big(\tau_N(P_N), \tau_N(P_0)\big).$$

It follows that

$$\big[P_0^N(A_N) + P_N^N(A_N^c)\big] \le 2\mathbf{I}_{(0,2\varepsilon)}\big(r(\tau_N(P_N), \tau_N(P_0))\big)$$
$$+ 2 \sup_{\{P \colon H(P, P_0) \le CN^{-1/2}\}} P^N\big(r(T_N, \tau_N(P)) \ge \varepsilon\big).$$

Because $H(P_N, P_0) \le CN^{-1/2}$, we can combine Corollary 5.1 and (5.7) to conclude that

$$\limsup_N r\big(\tau_N(P_N), \tau_N(P_0)\big) \le 2\varepsilon.$$

Since ε is an arbitrary positive number and P_N is an arbitrary sequence with $H(P_N, P_0) \le CN^{-1/2}$ the proof is complete. $\square$

PROOF OF THEOREM 1.4.   Take D as in Lemma 5.3, fix $P \in \mathscr{P} \setminus D$ and take $\varepsilon > 0$. By assumption (iii), we can find $C > 0$ such that for every N,

$$P^N\big(H(\hat{P}_N, P) \le CN^{-1/2}\big) \ge 1 - \varepsilon.$$

Application of Lemma 5.3 yields that for every positive δ and for every $P \in \mathscr{P} \setminus D$,

$$\limsup_N P^N\big(r\big(\tau_N(\hat{P}_N), \tau_N(P)\big) > \delta\big) \le \varepsilon,$$

and as ε is positive but otherwise arbitrary, this proves the theorem. $\square$

6. Proof of Theorem 1.3. In this section we only assume measurability of the maps τ_N, separability of $(\mathscr{R}, r)$ and the existence of an r-consistent estimator T_N of $\tau_N(P)$. To construct an r-consistent substitution estimator $\tau_N(\hat{P}_N)$, we begin by choosing $\hat{P}_N = p_N(X_1, \ldots, X_N)$ to be an approximate minimum distance estimator, that is, an estimator satisfying

$$(6.1) \qquad r\left(T_N, \tau_N(\hat{P}_N)\right) \le \inf_{P \in \mathscr{P}} r(T_N, \tau_N(P)) + \varepsilon_N$$

for some sequence $\varepsilon_N \searrow 0$. If this can be done in such a way that $p_N \colon \mathscr{X}^N \to \mathscr{P}$ is measurable so that $\hat{P}_N$ is a proper estimator, then the consistency of $\tau_N(\hat{P}_N)$ will follow, since for every $P_0 \in \mathscr{P}$,

$$\begin{aligned} r\left(\tau_N(\hat{P}_N), \tau_N(P_0)\right) &\le r\left(T_N, \tau_N(\hat{P}_N)\right) + r(T_N, \tau_N(P_0)) \\ &\le \inf_{P \in \mathscr{P}} r(T_N, \tau_N(P)) + \varepsilon_N + r(T_N, \tau_N(P_0)) \\ &\le 2r(T_N, \tau_N(P_0)) + \varepsilon_N \to_{P_0} 0 \end{aligned}$$

because of the consistency of T_N.

It remains to be shown that (6.1) can be satisfied for a measurable p_N. Define $\mathscr{R}^0 = \{\tau_N(P) \colon P \in \mathscr{P}\}$, let $\bar{\mathscr{R}}^0$ denote the closure of $\mathscr{R}^0$ in $(\mathscr{R}, r)$ and let T_N^0 denote the projection of $T_N \in \mathscr{R}$ on $\bar{\mathscr{R}}^0$. Inequality (6.1) asserts that $\hat{P}_N$ must be chosen in such a way that $\tau_N(\hat{P}_N)$ lies in a ball with center T_N and radius $r(T_N, T_N^0) + \varepsilon_N$. One easily convinces oneself that this implies that for any ball $B \subset \mathscr{R}$ with radius $\varepsilon_N/3$, we can choose a fixed $\hat{P}_N$ satisfying (6.1) for all $T_N \in B$. As $(\mathscr{R}, r)$ is separable, we can cover $\mathscr{R}$ with a countable number of balls B_k with radius $\varepsilon_N/3$. For every m, we can therefore define $\hat{P}_N$ as a fixed point of $\mathscr{P}$ for all T_N in the measurable set $B_m \cap \bigcap_{k=1}^{m-1} B_k^c$, so that $\hat{P}_N$ is an elementary measurable function of T_N. As T_N is a measurable function of $X_1, \ldots, X_N$, the proof is complete. $\square$

REFERENCES

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

BERAN, R. J. (1984). Bootstrap methods in statistics. *Jber. Deutsch. Math.-Verein.* **86** 14–30.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BILLINGSLEY, P. (1986). *Convergence of Probability Measures*. Wiley, New York.

BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237.

BIRGÉ, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291.

DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth, Belmont, CA.

LE CAM, L. M. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *University of California Publications in Statistics* **1** 277–330.

LE CAM, L. M. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

LE CAM, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LE CAM, L. M. and YANG, G. L. (1990). *Asymptotics in Statistics*: *Some Basic Concepts*. Springer, New York.

POLITIS, D. and ROMANO, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050.

PUTTER, H. (1994). Consistency of resampling methods. Ph.D. thesis, Univ. Leiden.

PUTTER, H. and VAN ZWET, W. R. (1996). On a set of the first category. In *Festschrift for Lucien Le Cam*. Springer, New York. To appear.

VAN ZWET, W. R. (1996). Resampling: the jackknife and the naïve bootstrap. Unpublished manuscript.

DEPARTMENT OF MATHEMATICS
FREE UNIVERSITY OF AMSTERDAM
1081 DE BOELELAAN
1081 HV AMSTERDAM
THE NETHERLANDS

DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
UNIVERSITY OF LEIDEN
P.O. BOX 9512
2300 RA LEIDEN
THE NETHERLANDS