UNIFORMITY IN P
OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by

Anne Sheehy and Jon A. Wellner

Technical Report No. 254
November 1988

Prepared under the Auspices
of
National Science Foundation
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
UNIFORMITY IN $P$
OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by

Anne Sheehy and Jon A. Wellner

Stanford University and University of Washington

August, 1986
Revision 1: February, 1987
Revision 2: September, 1988

AMS 1980 Subject Classifications: Primary 60F05, 62G05, 62G20;
Secondary 60B12, 60F17, 60G44, 62G30, 46G05, 46G10.

Key Words and Phrases: bootstrap, central limit theorem, delta method, empirical process, regular estimators, Skorokhod - Dudley - Wichura theorem, strong law of large numbers, uniformity, uniform integrability, weak approximation

1 Research supported by: National Science Foundation Grant DMS-8708083

2 Research supported by: National Science Foundation Grant DMS-8400893 and the John Simon Guggenheim Foundation
UNIFORMITY IN P
OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by

Anne Sheehy and Jon A. Wellner
Stanford University and University of Washington

ABSTRACT

The uniformity of convergence in the true underlying P is studied for general empirical measures and processes indexed by a collection of functions F. We first give sufficient conditions for uniformity in P ∈ P of the convergence in Pollard's (1982) Glivenko-Cantelli theorem and Donsker (or central limit) theorem. In fact, we strengthen the latter to a uniform in P weak approximation in the spirit of Dudley and Philipp (1983).

We use these uniform in P limit theorems and weak approximations to study the bootstrap for general empirical measures, and show that for indexing collections F satisfying a Pollard-Kolcinskii entropy condition, the bootstrap enjoys a strong regularity property. These theorems are compared to a recent bootstrap theorem (for a fixed P) of Giné and Zinn (1988). We also show that any P₀-Donsker class F with P₀-square integrable envelope F is automatically a \{Pₙ\}-uniform Donsker class for n⁻¹/² Hellinger tangent (or "contiguous") sequence \{Pₙ\} if limsupₙ Pₙ(F²) < ∞.

These results are then applied to study the delta method for compact (or Hadamard) differentiable functions v of a general empirical measure IPₙ. We extend the results of Reeds (1976) and Gill (1988), and study several closely related regularity and bootstrap versions of the basic delta method for such functions.

The delta method theorems are illustrated by examples which include length biased sampling, quantiles and the quantile process, mean residual life and a bivariate version thereof, percentile residual life, cumulative hazard functions in one and two dimensions, and estimation of a probability measure P subject to constraints.
UNIFORMITY IN $P$ OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by Anne Sheehy and Jon A. Wellner

Outline

0. Introduction
1. Uniform limit theorems
   a. Uniform in $P$ strong laws of large numbers and Glivenko-Cantelli theorems
   b. Uniform in $P$ Donsker theorems or weak convergence
   c. Regularity of the empirical measure as an estimator of $P$
   d. Convergence of the "bootstrapped empirical process"
2. The Skorokhod-Dudley$^2$-Wichura theorem
3. The delta method
   a. The delta method via compact differentiation
   b. Representing the derivatives
4. Examples: the delta method in action
5. Proofs for section 1

References
0. Introduction.


Most of the theory developed so far (with the notable exception of Le Cam (1983), (1986)) has focussed on the setting of iid observations with \( P \), the underlying distribution thereof, fixed. For applications in statistics, however, it is frequently useful (or even necessary!) to know what happens when \( P \) varies with \( n \) or perhaps even over a fixed set \( P \). The present paper is aimed at two sets of questions in this direction:

I. (Uniformity). How uniform (in \( P \in P \)) are the recently developed limit theorems for general empirical distributions \( IP_n \) and processes?

II. (The delta method). For a (nonlinear) function \( v \) of \( P \) and of \( IP_n \), is there a workable generalization of the classical delta - method, and how is it affected when \( P \) varies with \( n \)?

The first question is addressed, and given a partial answer, in section 1. We show that for classes of functions \( F \) with envelope function \( F \) satisfying a Pollard - Kolčinskii type entropy condition, the convergence in Pollard’s (1982a) version of the classical Glivenko - Cantelli theorem is uniform in \( P \in P \) if \( F \) is uniformly integrable over \( P \), and the convergence in Pollard’s (1982a) Donsker (or central limit) theorem is uniform in \( P \in P \) if \( F \) is uniformly square - integrable over \( P \). We also consider sequences \( \{P_n\} \) converging to a fixed \( P_0 \) in the Hellinger metric, and "contiguous" sequences \( \{P_n\} \) converging to a fixed \( P_0 \) at rate \( n^{-1/2} \) with tangent \( h \in L_2(P_0) \). For the latter type of sequences, we show that the Donsker - property (CLT) is preserved: if \( F \) is a \( P_0 \) - Donsker class with envelope function \( F \) and \( \{P_n\} \) is a \( n^{-1/2} \) - Hellinger tangent sequence (with \( \limsup_n \rightarrow \infty P_n(F^2) < \infty \)), then \( F \) is a "\( \{P_n\} \) - uniform Donsker class" as defined in section 1. This is somewhat comparable to (although perhaps a bit simpler than) a recent result of Giné and Zinn (1988) who show that "the bootstrap works" for general empirical measures (at a fixed \( P_0 \)) if and only \( F \) is a Donsker class with square - integrable envelope function \( F \)
(under mild measurability hypotheses). We go on to show that the bootstrap is (Hellinger regular; i.e. that it "works" uniformly in \( \{P_n\} \) converging to \( P_0 \) in Hellinger distance for classes \( \mathcal{F} \) satisfying a Pollard - Kolčinskii entropy condition.

The results in section 1, are closely related to the theorems of Pollard (1982a) in the case of theorems 1.1 - 1.3, and van der Vaart (1988) in the case of theorem 1.4, but have many precursors and predecessors in the statistical literature: uniformity of convergence in a parameter or df or measure \( P \) has been an issue in the work of many authors, including Chung (1951), Dvoretzky, Kiefer, and Wolfowitz (1956), Ibragimov and Has’minskii (1981), Kiefer and Wolfowitz (1959), Lai (1978), Le Cam (1956), and Wald (1943). For more on the relationship of our results to those of Le Cam (1983), see the remarks in section 1. Dudley’s (1987) work on "universal Donsker classes" (classes \( \mathcal{F} \) for which the CLT holds for every \( P \)) are also related, but aimed in a slightly different direction.

The proofs of the theorems in section 1 are given in section 5.

The second question is addressed in section 3, with preparation and background work in section 2. A variety of examples are given in section 4.

Section 2 introduces and briefly describes Dudley’s (1985) version of the well known Skorokhod - Dudley - Wichura theorem. This "fourth generation theorem" (which we call the \( S - D^2 - W \) theorem for short), is used there to develop several corollaries in preparation for the delta method results in section 3.

Our generalization of the classical delta method to the setting of empirical measures is developed and explored in section 3. We follow the direction of Reeds (1976) and Gill (1988) by introducing a version of compact (or Hadamard) differentiation which seems well suited for applications in statistics. Because Reeds’ (1976) work is not widely available, we highly recommend Gill (1988) as an exposition and nice extension of Reeds’ theory. In particular, we extend Gill’s notion of "differentiability tangentially to a subspace", and show how various notions of continuity of this type of differentiation theory have corresponding consequences for regularity of estimators and performance of the bootstrap.

Our basic delta method result has the following flavor. Suppose that:

(i) \( \nu \) is (compactly -) differentiable at \( P \) with derivative \( \dot{\nu} \), and

(ii) \( \mathbb{X}_n = \sqrt{n} (P_n - P) \Rightarrow \mathbb{X}_0 \) as \( n \to \infty \).

Then

(1) \( \sqrt{n} (\nu(P_n) - \nu(P)) \Rightarrow \dot{\nu}(\mathbb{X}_0) \) as \( n \to \infty \).

A nice pair of results in section 3 (theorems 3.4 and 3.5) are to the following effect:
A. (Local regularity result). Suppose the delta method "works" at a fixed $P$; i.e. (1) holds. If $\{P_n\}$ is an $n^{-1/2}$ -- Hellinger tangent sequence (i.e. "contiguous" sequence), then (1) continues to hold under (essentially) no further assumptions (with sampling under $P_n$) if $\nu(P)$ on the left side is replaced by $\nu(P_n)$.

B. (Bootstrap result). Suppose the delta method "works" at a fixed $P$; i.e. (1) holds. Then the bootstrap (for estimating the distribution of the left side in (1)) is weakly (i.e. in probability) consistent.

Virtually all our results in section 3 have been inspired by Gill (1988).

The examples in section 4, especially the quantile and quantile process examples 3 and 3.A, clarify what is involved in proving a.s. consistency of the bootstrap for nonlinear functions $\nu$ of $P$. They also illustrate the wide applicability of the basic delta method idea in the setting of general empirical distributions. We do not however, believe that this method will yield the most "highly tuned" or refined results in individual cases, especially results involving the interaction of the functional and the probabilistic structure of the process. But, as illustrated by Reeds’ (1976) and Gill (1988) the choice of metrics and norms for the various spaces involved is sufficiently rich to come very close to optimal theorems in many cases. In particular, we give some (limited) attention in section 4 to comparison of conclusions via our methods versus conclusions obtainable via the finely tuned constructions of Csörgő, Csörgő, Horváth, and Mason (1986). We do not advocate any one method to the exclusion of others. We do feel that the delta-method perspective is frequently illuminating and useful, and often provides a good starting point for further exploration. The theorems of section 3 seem to confirm this.
1. Uniform limit theorems

We now present our uniform in \( P \in \mathcal{P} \) limit theorems.

Uniform in \( P \) strong laws of large numbers and Glivenko - Cantelli theorems.

Our first goal is a strengthening of Pollard's (1982) Glivenko - Cantelli theorem for the empirical measure indexed by functions which parallels Chung's (1951) strengthened version of the classical strong law of large numbers (SLLN). To emphasize the parallels, we first briefly recall Chung's theorem. The strong law of large numbers asserts that if \( X, X_1, \ldots, X_n, \ldots \) are iid real-valued random variables with distribution \( P \) and \( E_P|X| < \infty \), then

\[
1 \quad \overline{X}_n \to_{a.s.} \mu = E_P(X) \quad \text{as} \quad n \to \infty.
\]

Equivalently, for every \( \varepsilon > 0 \)

\[
2 \quad \text{Pr}_P \left\{ \max_{m \geq n} |\overline{X}_m - \mu| > \varepsilon \right\} \to 0 \quad \text{as} \quad n \to \infty.
\]

The uniform in \( P \in \mathcal{P} \) version of (2) is the following:

**Theorem 1.0.** (Chung, 1951). Suppose that \( X_1, \ldots, X_n, \ldots \) are iid \( P \in \mathcal{P} \) and that \( \mathcal{P} \) satisfies the uniform integrability condition

\[
3 \quad \sup_{P \in \mathcal{P}} E_P|X|1_{|X| \geq \lambda} \to 0 \quad \text{as} \quad \lambda \to \infty.
\]

Then \( \overline{X}_n \to_{a.s.} \mu \) as \( n \to \infty \) uniformly in \( P \in \mathcal{P} \); i.e.

\[
4 \quad \sup_{P \in \mathcal{P}} \text{Pr}_P \left\{ \max_{m \geq n} |\overline{X}_m - \mu| \geq \varepsilon \right\} \to 0 \quad \text{as} \quad n \to \infty.
\]

Uniform integrability conditions like (3) of Chung's theorem 1.0 will recur repeatedly in the theorems to follow. Thus it will be useful to recall the following characterization of uniformly integrable families.

**Theorem.** (Vallée - Poussin). A family of \( L_1 \) rv's \( \{X_t : t \in T\} \) is uniformly integrable if and only if there exists a convex function \( G \) on \([0, \infty)\) with \( G(0) = 0 \), \( G(x)/x \to \infty \) as \( x \to \infty \), and

\[
\sup_{t \in T} E G(|X_t|) < \infty.
\]

For a proof, see Meyer (1966). For example, (3) holds if for some \( \delta > 0 \) we have \( E_P|X|^{1+\delta} \leq M < \infty \) for all \( P \in \mathcal{P} \).

Pollard (1982) (see also Dudley (1984) page 108) proved a Glivenko - Cantelli theorem for the empirical distribution indexed by functions which generalizes both (1) and (2) and the classical Glivenko - Cantelli theorem for the empirical distribution function (see e.g. Shorack and Wellner (1986) pages 95 and 106 for the latter). We will
provide a uniform version of Pollard's theorem which strengthens it in exactly the same way that Chung's theorem 0 strengthens the SLLN. It will also play a key role in proving uniform versions of Pollard's (1982,1984) weak convergence theorems for the empirical process.

To state our uniform version of Pollard's theorem, we first need to define the Kolčinskii - Pollard combinatorial entropy for a class of functions. Let \( P \) be a collection of probability measures on the measurable space \((A, \mathcal{A})\), and let \( F \) be a collection of \( \mathcal{A} \) - measurable functions defined on \( A \) with envelope function \( F \); i.e. \(|f| \leq F\) for each \( f \in F \). We will usually assume that \( F \cup \{F\} \subset L_r(P) \) for some \( r > 0 \). Suppose that \( X_1, \ldots, X_n, \ldots \) are iid \((A\text{-valued})\) random variables with distribution \( P \in P \); thus \( X(j) \equiv X_j, j = 1, 2, \ldots \) are the coordinates for a countable product \((A^\infty,\mathcal{A}^\infty, P^\infty)\) of copies of \((A,\mathcal{A}, P)\). For definiteness we take (as in Dudley (1984), page 11), the underlying probability space(s)

\[(\Omega, \Sigma, Pr_p) = (A^\infty, \mathcal{A}^\infty, P^\infty) \times ([0,1], \mathcal{B}, \lambda)\]

where \( \mathcal{B} \) denotes the Borel subsets of \([0,1]\) and \( \lambda \) is Lebesgue measure.

We let \( P_n \) denote the empirical measure of the first \( n \) \( X \)'s:

\[P_n(B) \equiv \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i) \quad \text{for } B \in \mathcal{A}.
\]

For any \( f \in L_1(P) \), we will use the notation

\[P(f) \equiv \int f \, dP = \mathbb{E}_P f(X)\]

and similarly for \( P_n \).

**Definition 1.1.** (Kolčinskii, 1981; Pollard, 1982). Let \( r > 0 \). Suppose that \( F \) is an envelope function for \( F \). For any measure \( Q \) on \((A,\mathcal{A})\) with \( Q(F^r) < \infty \), and \( \delta > 0 \) let

\[N_F^{(r)}(\delta,F,Q) \equiv \sup \left\{ m : \text{for some } f_1, \ldots, f_m \in F \text{ and all} \right\}
\]

\[i \neq j, \int |f_i - f_j|^r \, dQ > \delta' \int F^r \, dQ \].

Let \( \Gamma \) be the set of all measures on \((A,\mathcal{A})\) with finite support; i.e. \( Q = k^{-1} \sum_{i=1}^{k} \delta_{x(i)} \) for some points \( x(1), \ldots, x(k) \in A \) not necessarily distinct. Then let

\[N_F^{(r)}(\delta,F) \equiv \sup_{Q \in \Gamma} N_F^{(r)}(\delta,F,Q), \quad \text{and} \]

\[H_F^{(r)}(\delta,F) \equiv \log N_F^{(r)}(\delta,F) \quad \ldots \]

Throughout most of the following we assume that \( F \) is a permissible class of functions in the sense of Pollard (1984), page 196; this assumption is needed to ensure
measurability of suprema such as \( D_n \) below. A closely related alternative assumption is Dudlay's notion of an image admissible Suslin class \( F \); see Dudley (1984), section 10.2; or, more simply, that \( F \) satisfies a separability assumption as in Pollard (1982a) or Gaenssler (1984). We have chosen not to work with the most general measurability assumptions (Gin'ev and Zinn (1984) or Talagrand (1987)) since our main emphasis in this section is on the uniformity in \( \mathcal{P} \) of the limit theorems.

**Theorem 1.1.** Suppose \( F \) is permissible and that:

(i) \( N_K^{(1)}(\delta, F_K) < \infty \) for every \( \delta > 0 \) and \( K > 0 \) where \( F_K = \{ f 1_{F \leq K}: f \in F \} \).

(ii) \( \sup_{P \in \mathcal{P}} E_P F 1_{F \geq \lambda} \to 0 \) as \( \lambda \to \infty \).

Then

\[
D_n = \sup \{ \| (P_n - P)(f) \|: f \in F \}
= \| P_n - P \|_F \to_{a.s.} 0 \quad \text{uniformly in } P \in \mathcal{P}
\]

as \( n \to \infty \); that is, for every \( \varepsilon > 0 \)

(6) \[ \sup_{P \in \mathcal{P}} P_{P} \left\{ \max_{m \geq n} D_m \geq \varepsilon \right\} \to 0 \quad \text{as } n \to \infty. \]

Note that (ii) holds if for some \( \delta > 0 \) we have \( E_P F^{1+\delta} \leq M < \infty \) for all \( P \in \mathcal{P} \).

Chung's (1951) theorem 1.0 is used heavily in the proof of theorem 1.1, so it not surprising that it is still contained therein; by taking \( F = \{ \text{identity} \} \), theorem 1.1 reduces to theorem 1.0. Theorem 1.1 also contains the (uniform in all \( P \) ) Glivenko-Cantelli theorem due to Vapnik and Chervonenkis (1971) and Steele (1978).

**Corollary 1.1.** (Vapnik - Chervonenkis, 1971; Steele, 1978). Suppose that \( C \) is a permissible Vapnik-Chervonenkis class of subsets of \( A \) and set

\[
D_n(C) \equiv \| P_n - P \|_F \quad \text{with } F = \{ 1_C: C \in C \}.
\]

Then \( D_n(C) \to_{a.s.} 0 \) as \( n \to \infty \) uniformly in \( P \in M = \{ \text{all } P \} \):

(7) \[ \sup_{P \in M} P_P \left\{ \max_{m \geq n} D_m(C) \geq \varepsilon \right\} \to 0 \quad \text{as } n \to \infty. \]

**Remark 1.1.** The measurability conditions of Vapnik and Chervonenkis (1971) are correct and apparently weaker than the permissibility condition used in our version of corollary 1.1. The conditions of Steele (1978) are still weaker, but are not sufficient. See Dudley (1984) section 11.2 for further discussion.
As preparation for the proofs of uniform weak convergence theorems for the empirical process, we now introduce several useful pseudometrics on \( L_2(P) \) and on \( P \). For \( f, g \in L_2(P) \) let

\[
\tau^2_P(f, g) \equiv \left[ \int (f - g)^2 \, dP - \left( \int (f - g) \, dP \right)^2 \right] = \text{Var}_P(f(X) - g(X)).
\]

**Definition 1.2.** For \( P, Q \in \mathcal{P} \) define the pseudo-metrics \( \rho_F \) and \( \rho_F^0 \) by

\[
\rho_F(P, Q) \equiv \sup_{f, g \in \mathcal{F} \cup \{1\}} |(P - Q)(f - g)|,
\]

\[
\rho_F^0(P, Q) \equiv \sup_{f, g \in \mathcal{F}} |\text{Var}_P(f - g) - \text{Var}_Q(f - g)|.
\]

The pseudo metric \( \rho_F \) relates \( L_2(Q) \) norms to \( L_2(P) \) norms, and \( \rho_F^0 \) relates \( \tau_Q \) norms to \( \tau_P \) norms, for \( Q, P \in \mathcal{P} \): if \( f, g \in \mathcal{F} \),

\[
\|f - g\|_{L_2(Q)} - \rho_F(Q, P) \leq \|f - g\|_{L_2(P)} \leq \|f - g\|_{L_2(Q)} + \rho_F(Q, P)
\]

(8)

\[
\tau^2_Q(f, g) - \rho_F^0(P, Q) \leq \tau^2_P(f, g) \leq \tau^2_Q(f, g) + \rho_F^0(Q, P).
\]

The following corollary of theorem 1.1 will play an important role in the proofs of uniform weak convergence theorems for the empirical process.

**Corollary 1.2.** Suppose \( F \) is permissible and that:

(i) \( N^{(2)}_{\delta}(\delta, F) < \infty \) for every \( \delta > 0 \).

(ii) \( \sup_{P \in \mathcal{P}} E_P F^2 1_{\{F \geq \lambda\}} \to 0 \) as \( \lambda \to \infty \).

Then

\[
\hat{D}_n \equiv \sup\{|(P_n - P)(fg)|: f, g \in \mathcal{F}, 1\} \to_{a.s.} 0,
\]

\[
D^#_n \equiv \sup\{|(P_n - P)(f - g)|^2: f, g \in \mathcal{F} \} \to_{a.s.} 0, \quad \text{and}
\]

(10) \( \rho_F^0(P_n, P) \to_{a.s.} 0 \)

uniformly in \( P \in \mathcal{P} \) as \( n \to \infty \).
Uniform in $P$ Donsker theorems or weak convergence

Let $P$ be a collection of probability measures on the space $(A, \mathcal{A})$, and let $F$ be a collection of $\mathcal{A}$-measurable functions defined on $A$ with envelope function $F$; i.e., $|f| \leq F$ for each $f \in F$. Assume that $F \cup \{F\} \subset L_2(P)$ for each $P \in P$. Let $\{P_n\}$ be a sequence in $P$. Suppose that $X_{n1}, \ldots, X_{nn}$ are row independent, i.i.d. $P_n$, $(A$-valued) random variables. Assume that the triangular array is defined on a common probability space $(\Omega, \Sigma, Pr)$. For definiteness, in this setting we take

$$(\Omega, \Sigma, Pr) = (A^1, \mathcal{A}^1, P_1) \times \cdots \times (A^n, \mathcal{A}^n, P_n) \times \cdots \times ([0,1], \mathcal{B}, \lambda).$$

Let the empirical measure $P_n$ be defined in the usual way, in terms of the $n$ random variables in the $n$th row of the array, by

$$P_n(B) = n^{-1} \sum_{i=1}^n 1_B(X_{ni}) \text{ for } B \in \mathcal{A}.$$

and

$$\Xi_n = \sqrt{n} \left(P_n - P_n\right).$$

Thus

$$\Xi_n(f) = \int f \, d\Xi_n = \sqrt{n} \int f \, d(P_n - P_n), \text{ for } f \in F.$$

If

$$P_n(fg) - P_n(f)P_n(g) \to P_0(fg) - P_0(f)P_0(g) \text{ for all } f, g \in F,$$

then the finite dimensional distributions of the processes $\Xi_n$ converge to those of the mean zero Gaussian process $\Xi_0$ with covariance

$$E\Xi_0(f)\Xi_0(g) = \int fg \, dP_0 - \int f \, dP_0 \int g \, dP_0 \text{ for } f, g \in F.$$

Let $L_{\infty}(F)$ be the space of all bounded real functions on $F$, and define $C(F, P)$ to be the set of all functions $x(\cdot)$ in $L_{\infty}(F)$ that are uniformly continuous with respect to the $\tau_P$ seminorm. As usual, $L_{\infty}(F)$ is equipped with the uniform norm $\|x\| = \sup_{f \in F} |x(f)|$. We call a process $\Xi_0$ with covariance given by (13) and all sample paths in $C(F, P_0)$ a $P_0$-Brownian bridge.

Here is the definition of weak convergence which we will use; it is a slight variation on the definition of Hoffmann-Jørgensen (1984); see e.g. Dudley (1985). Let $B^k$ denote the collection of Borel sets in $R^k$.

If $(\Omega, \Sigma, P)$ is a probability space and $f$ is a real-valued function on $\Omega$, let

$$E^*_p f \equiv \inf \{ E_p h : f \leq h \text{ and } h \text{ is measurable} \},$$

$$E^*_p f \equiv \sup \{ E_p h : f \geq h \text{ and } h \text{ is measurable} \}.$$
\[ E_p f = E_p f^* \quad \text{and} \quad E_p f = E_p f_* \]

where \( f^* , f_* \) are real-valued, \( \Sigma \) - measurable functions defined by
\[
\begin{align*}
  f^* & \equiv \text{ess.inf}\{h \geq f : h \text{ is measurable}\} \geq f \\
  f_* & \equiv \text{ess.sup}\{h \leq f : h \text{ is measurable}\} \leq f.
\end{align*}
\]

**Definition 1.3.** Suppose that \((S,d)\) is a metric space, \(\{(X_n,\mathcal{A},Q_n)\}_{n \geq 0}\) is a sequence of probability spaces, and \(\mathbb{X}_n : X_n \to S\). Let \(C_b(S)\) be the collection of bounded, uniformly continuous functions \(h\) from \(S\) to \(R\). Then we say that \(\mathbb{X}_n\) \(\text{converges weakly to a random element } \mathbb{X}_0 \in (S,B_{\text{borel}})\), and write \(\mathbb{X}_n \Rightarrow \mathbb{X}_0\), if for every \(h \in C_b(S)\),
\[
(14) \quad E^* h(\mathbb{X}_n) \to E h(\mathbb{X}_0), \quad \text{as } n \to \infty.
\]

Since \(\mathbb{X}_0\) is a random element in \((S,B_{\text{borel}})\) and \(h\) is continuous, \(E h(\mathbb{X}_0)\) is well-defined; this entails that \(\mathbb{X}_0\) is concentrated on a separable \(S_0 \subset S\) (with the possible exception of set-theoretic pathological cases); i.e. \(Q_0(\mathbb{X}_0 \in S_0) = 1\). See the discussion in Dudley (1985), pages 148 - 149.

For our applications we will take \((S,d) = (L_\infty(F),\|\cdot\|_F)\), or \((S,d) = (B,\|\cdot\|_B)\) for a Banach space \(B\).

To state our results, it will be helpful to first extend the terminology of Dudley (1984).

**Definition 1.4.** A class \(F \subset L_2(A,\mathcal{A},P)\) for all \(P \in P\) will be called a \(P\)-**uniform** \(G_P\) BUC class if and only if the \(P\)-Brownian bridge processes \(\mathbb{X}(f)(\omega)\), \(f \in F\), \(\omega \in \Omega\), can be chosen so that: for every \(\varepsilon > 0\) there exist \(\lambda = \lambda(\varepsilon) > 0\) (large) and \(\delta = \delta(\varepsilon)\) (small) such that
\[
(15) \quad \sup_{P \in P} \Pr_P\{\|\mathbb{X}\|_F > \lambda\} < \varepsilon;
\]
and
\[
(16) \quad \sup_{P \in P} \Pr_P\{\sup_{[\delta]} |\mathbb{X}(f - g)| > \varepsilon\} < \varepsilon
\]

where \([\delta]_P = \{(f,g) : f, g \in F \text{ and } \tau_P(f,g) < \delta\}\).

Note that if \(P\) contains just a single measure \(P\), then \(F\) is a \(P\)-**uniform** \(G_P\) BUC class implies that \(F\) is a \(G_P\) BUC class in the sense of Dudley (1984).

**Definition 1.5.** A class \(F \subset L_2(A,\mathcal{A},P)\) for all \(P \in P\) will be called a \(P\)-**uniform functional Donsker class** if and only if it is a \(P\)-uniform \(G_P\) BUC class and there are iid \(P\)-uniform \(G_P\) BUC processes \(Z_i(f,\omega)\), \(i \geq 1\), \(f \in F\), \(\omega \in \Omega\) defined on the spaces \((\Omega,\Sigma,Pr_P) : P \in P\) such that for every \(\varepsilon > 0\)
\[
(17) \quad \sup_{P} P_{r}^{*} \left\{ n^{-1/2} \max_{m \leq n} \sup_{f \in F} \left| \sum_{i=1}^{m} f(X_i) - P(f) - Z_i(f) \right| > \varepsilon \right\} \to 0
\]
as \( n \to \infty \). If we let \( E_i = \delta_{X_i} - P \), \( i \geq 1 \), so that \( n^{-1/2} \sum_{i=1}^{n} E_i = \mathbb{X}_n \), then (17) can be rewritten as
\[
(18) \quad \sup_{P} P_{r}^{*} \left\{ n^{-1/2} \max_{m \leq n} \left\| \sum_{i=1}^{m} (E_i - Z_i) \right\|_{F} > \varepsilon \right\} \to 0
\]
as \( n \to \infty \). Here \( P_{r}^{*} \) denotes outer probability; \( P_{r}^{*}(C) \equiv \inf\{ P_{r}(B) : B \supset C, B \in \Sigma \} \).

For a sequence \( \{P_n\} \) with \( F \subseteq L_2(A,\mu,P_n) \) for every \( n = 1, 2, \ldots \), we say that \( F \) is a \( \{P_n\} \)-uniform functional Donsker class if (17) holds without the supremum over \( P \) and \( P \) replaced by \( P_n \) throughout: i.e.
\[
(19) \quad P_{r}^{*} \left\{ n^{-1/2} \max_{m \leq n} \sup_{f \in F} \left| \sum_{i=1}^{m} f(X_i) - P_n(f) - Z_i(f) \right| > \varepsilon \right\} \to 0
\]
as \( n \to \infty \); here \( ZZ_i, \ i = 1, 2, \ldots \) are iid \( P_n \)-Brownian bridges for each \( n \geq 1 \).

The following theorem gives conditions for the crucial uniform in \( .P \in P \) tightness needed to prove (17) or (19).

**Theorem 1.2.** Suppose \( F \) is permissible and that:

(i) \( \sum_{j=1}^{\infty} 2^{-j} [H_{P}^{(2)}(2^{-j},F)]^{1/2} < \infty \) or equivalently \( \int_{0}^{1} [H_{P}^{(2)}(x,F)]^{1/2} dx < \infty \).

(ii) \( \sup_{P} E_{P} F_{21[F \geq \lambda]} \to 0 \) as \( \lambda \to \infty \).

Then for every \( \varepsilon > 0 \) there are \( \delta = \delta(\varepsilon) > 0 \) and \( \delta' = \delta'(\varepsilon) > 0 \) for which
\[
(20) \quad \limsup_{n \to \infty} \sup_{P} \left\{ \sup_{[\delta]_{P}} P_{r} \{ |X_n(f - g)| > \varepsilon \} < \varepsilon \right\}
\]
and
\[
(20') \quad \limsup_{n \to \infty} \sup_{P} \left\{ \sup_{[\delta']_{P}} P_{r} \{ |X_n(f - g)| > \varepsilon \} < \varepsilon \right\}
\]
where \( [\delta]_{P} = \{ (f,g) : f, g \in F \text{ and } \|f - g\|_{L_2(P)} < \delta \} \) and \( [\delta']_{P} = \{ (f,g) : f, g \in F \text{ and } \tau_{P}(f,g) < \delta' \} \).

Theorem 1.2 gives sufficient conditions for a uniform in \( P \in P \) tightness condition on the process \( \mathbb{X}_n \). It can be viewed as the most difficult part of proving the following "\( P \)-uniform weak approximation theorem" for the entire process \( \{\mathbb{X}_m\}_{m \leq n} \).
Theorem 1.3. Suppose that $F \subseteq L_2(A, \mathcal{A}, P)$ for all $P \in \mathcal{P}$ is permissible with envelope function $F$, and that hypotheses (i) and (ii) of theorem 1.2 hold. Then $F$ is a $\mathcal{P}$-uniform functional Donsker class.

This theorem generalizes the $\mathcal{P}$-uniform weak approximation theorem for sums of mean zero uniformly square-integrable (real-valued) random variables due to Lai (1978) in much the same way that our theorem 1.1 generalizes Chung’s theorem 1.0. (Lai’s (1978) theorem treats the interpolated partial sum process and shows that it can be approximated (uniformly in $P \in \mathcal{P}$) in probability by a Brownian motion; our approximation here is in terms of the corresponding partial sums as in Philipp (1980) and Dudley and Philipp (1983); see also Dudley (1984), page 13.)

Although we do not know a complete $\mathcal{P}$-uniform analogue of Dudley’s (1984) theorem 4.1.1 characterizing functional Donsker classes for a fixed $P$, the following proposition is a partial result in the "only if" direction.

Proposition 1.1. If $F$ is a $\mathcal{P}$-uniform functional Donsker class, then for every $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon)$ and $N = N(\varepsilon)$ such that, for $n \geq N$

$$
\sup_{P \in \mathcal{P}} P_{PrP}^* \left\{ \sup_{[0], r} |X_n(f - g)| > \varepsilon \right\} < \varepsilon;
$$

i.e. the conclusion (20) of theorem 1.2 holds with $P_{PrP}$ replaced by $P_{PrP}^*$.

Theorem 1.3 easily implies that (18) holds without the maximum on $m \leq n$, and hence yields the following corollary. In fact, we will not give a complete proof of theorem 1.3 here, but only prove the corollary.

Corollary 1.3. Suppose that $F$ is permissible and satisfies hypotheses (i) and (ii) of theorem 1.2. Then, on the probability spaces of theorem 1.3, for any $\varepsilon > 0$

$$
\sup_{P \in \mathcal{P}} P_{PrP} \left\{ ||X_n - X(n)||_F > \varepsilon \right\} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty
$$

where $X(n) = n^{-1/2} \sum_{i=1}^{n} Z_i$ is a sequence of $P$-Brownian bridges.

Now consider corollary 1.3 for a sequence $\{P_n\}$ of probability measures for which the envelope $F$ of $F$ is $\{P_n\}$-uniformly square integrable. The conclusion (22) gives a weak approximation of the empirical process $X_n$ by a sequence $X(n)$ of $P_n$-Brownian bridges (even if the sequence $\{P_n\}$ does not converge). If we add the convergence hypothesis $\rho_F(P_n, P_0) \rightarrow 0$, then we can replace the $P_n$-Brownian bridges by a sequence $\{X_0(n)\}$ of $P_0$-Brownian bridges to obtain the following useful further corollary.
Corollary 1.4. Suppose $F$ is permissible and that:

(i) $F$ is totally bounded in the $\tau_{P_0}$ seminorm.

(ii) $\limsup_{n \to \infty} P_n(F^2 1_{\{F \geq \lambda\}}) \to 0$ as $\lambda \to \infty$.

(iii) $\rho_F(P_n, P_0) \to 0$ as $n \to \infty$.

(iv) For each $\epsilon > 0$, $\eta > 0$ there exists a $\delta > 0$ for which

$$\limsup_{n \to \infty} Pr_{P_n^*} \{ \sup_{[\delta]_n} |X_n(f - g)| > \eta \} < \epsilon$$

where $[\delta]_n = \{ (f, g) : f, g \in F$ and $\tau_{P_n^*}(f, g) < \delta \}$.

Then $\{X_n\}$ and a sequence $\{X_0^{(n)}\}$ of $P_0$-Brownian bridges can be defined on $(\Omega, \Sigma, Pr)$ so that: for every $\epsilon > 0$

$$Pr\{ \|X_n - X_0^{(n)}\|_F > \epsilon \} \to 0 \quad \text{as} \quad n \to \infty.$$  (23)

In particular, $X_n \Rightarrow X_0$ in $L_\infty(F)$.

Here are several further corollaries of corollaries 1.3 and 1.4.

Corollary 1.5. Suppose the hypotheses of corollary 1.3 hold. Then for any bounded, uniformly continuous function $g$ on $L_\infty(F)$,

$$\sup_{P \in P} |E_{P_n^*} g(X_n) - E_P g(X)| \to 0 \quad \text{as} \quad n \to \infty.$$  

Here $X \equiv X^{(n)}$ is a $P$-Brownian bridge process on $F$.

A further corollary is obtained by specializing corollary 1.4 to Vapnik-Chervonenkis classes of sets:

Corollary 1.6. Suppose that $C$ is a permissible Vapnik-Chervonenkis class of subsets of $A$, let $F \equiv \{ 1_C : C \in C \cup \{ A \} \}$, and suppose that $\{P_n\}$ satisfies

$$\sup_{B, C \in C \cup \{ A \}} |P_n(B \cap C) - P_0(B \cap C)| \to 0 \quad \text{as} \quad n \to \infty.$$  (24)

Then $X_n \Rightarrow X_0$ in $L_\infty(F)$.

Regularity of the empirical measure as an estimator of $P$

Corollary 1.4 easily implies that $P_n$ is a "regular estimator" of $P$ in the sense of the following definition. We let $H(P, Q)$ denote the Hellinger metric between two probability measures $P$ and $Q$:

$$H^2(P, Q) = \int \left( \frac{dP}{d\mu} \right)^{1/2} - \left( \frac{dQ}{d\mu} \right)^{1/2} \frac{d\mu}{d\mu}$$

for $\mu$ dominating both $P$ and $Q$. 

Definition 1.6. Suppose that $X_{n1}, \ldots, X_{nn}$ are iid $P_n$ for each $n \geq 1$. Then 
\[ \{\hat{\mathcal{F}}_n \} \equiv \{\hat{\mathcal{F}}_n(f) : f \in \mathcal{F} \} \]

is a Hellinger - regular estimator of \[ \{P\} \equiv \{P(f) : f \in \mathcal{F} \} \text{ at } P_0 \text{ if } \hat{\mathcal{X}}_n \equiv \sqrt{n} (\hat{\mathcal{F}}_n - P_n) \Rightarrow \hat{\mathcal{X}} \text{ in } L_\infty(F) \text{ for every sequence } \{P_n\} \subset P \text{ with } H(P_n, P_0) \to 0 \text{ where the (distribution of the)} \]

limit process $\hat{\mathcal{X}}$ depends on $P_0$ but not on the sequence $\{P_n\}$.

If this holds for every $P_0 \in P$, we say that $\{\hat{\nu}_n\}$ is Hellinger - regular on $P$.

If $\hat{\mathcal{X}}_n \Rightarrow \hat{\mathcal{X}}$ for every sequence $\{P_n\}$ satisfying
\[ \int \{ \sqrt{n} [(dP_n)^{1/2} - (dP_0)^{1/2}] - \frac{1}{2} h (dP_0)^{1/2} \}^2 \to 0 \]

for some $h \in L_2^0(P_0) \equiv \{ h \in L_2(P_0) : P_0(h) = 0 \}$ where the (distribution of the) limit process depends on $P_0$ but not the sequence $\{P_n\}$ (and hence, in particular, not on $h$), then we say that $\hat{\mathcal{X}}_n$ is $n^{-1/2}$-Hellinger - regular at $P_0$.

Here $\hat{\mathcal{F}}_n$ denotes an arbitrary estimator of $P$, not necessarily the empirical measure $P_n$. For example, $\hat{\mathcal{F}}_n$ could be a smoothed version of $P_n$.

Corollary 1.7. Suppose that $F$ is permissible and that $F$ and $P$ satisfy (i) and (ii) of theorem 1.2. Then $H(P_n, P_0) \to 0$ implies that $\rho_F(P_n, P) \to 0$, and it follows that $\mathcal{F}_n$ is a Hellinger - regular estimator of $\{P\}$ on $P$.

We now give a final stability or uniform - Donsker theorem for contiguous sequences $\{P_n\}$. This theorem applies to any $P_0$-Donsker class $\mathcal{F}$ with (uniformly) square integrable envelope function $F$, not just those classes $\mathcal{F}$ described by the Pollard type entropy condition (i) of theorem 1.2. We will use this theorem in section 2.

Theorem 1.4. Suppose that:

(i) $\{P_n\}$ is a sequence (in $P$) satisfying (25) for some $h \in L_2^0(P_0)$.

(ii) $F$ is a $P_0$-Donsker class with square integrable envelope function $F$ satisfying
\[ \limsup_{n \to \infty} P_n(P^2) < \infty. \]

Then $x_n \equiv \sqrt{n} (P_n - P_0)$ satisfies
\[ \|x_n - x_0\|_P \to 0 \quad \text{as } n \to \infty \]

where $x_0 \in C(F, P_0)$ is defined by
\[ x_0(f) = \int f \, h \, dP_0 = P_0(h \, f), \]

and $F$ is a $\{P_n\}$-uniform Donsker class: with
\[ \mathcal{X}_n \equiv \sqrt{n} (\mathcal{F}_n - P_n), \quad n \geq 1, \]
there is a probability space \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})\) with \{\(X_n\)\} and a sequence of \(P_0 - \text{Brownian bridge processes} \ X_0^{(n)}\) with sample paths in \(C(F, P_0)\) defined thereon satisfying, for every \(\varepsilon > 0\)

\[
Pr_{P_0}^* \{|\|X_n - X_0^{(n)}\|_F > \varepsilon\} \to 0 \quad \text{as} \quad n \to \infty;
\]

i.e. \(IP_n\) is a \(n^{-1/2} - \text{Hellinger - regular estimator of} \ {P}\) on \(P\). Moreover, \(\bar{X}_n^0 = \sqrt{n} (IP_n - P_0)\) satisfies

\[
|\|X_n^0 - (X_0^{(n)} + x_0)\|_F^* \to p. \quad 0.
\]

Convergence of the "bootstrapped empirical process"

The preceding theorems imply convergence of the "bootstrapped empirical process". Suppose that \(X_1, X_2, \cdots,\) are iid \(P_0\) on \((A, \mathcal{A})\), and let

\[
P_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}
\]

be the empirical measure of the first \(n\) of the \(X_i\)'s. Suppose further that \(X_1^*, \cdots, X_m^*\) are iid \(P_n\), let

\[
P_m^* = m^{-1} \sum_{i=1}^{m} \delta_{X_i^*}
\]

and set

\[
\bar{X}_m^* = \sqrt{m} (P_m^* - P_n).
\]

Thus \(X_1^*, \cdots, X_m^*\) is the "bootstrap sample", \(P_m^*\) is the "bootstrap empirical measure", and \(\bar{X}_m^*\) is the "bootstrap empirical process". If we assume \(P_0 F^2 < \infty\) and \(F\) is sparse with envelope \(F\), then hypotheses (i) and (ii) of theorem 1.2 are satisfied a.s. \(P_0\) for \(P = \{IP_n\}\). Thus (iv) of corollary 1.4 holds a.s., and it therefore follows that corollary 1.4 may be applied to deduce weak convergence (a.s. conditional on \(X_1, X_2, \cdots,\)) of \(\bar{X}_m^*\) whenever \(n\) and \(m = m(n) \to \infty\). The following theorems generalize and extend some results of Bickel and Freedman (1981), Shorack (1982) (see Shorack and Wellner (1986) chapter 23), Gaenssler (1986), Beran and Millar (1986), and Beran, Le Cam, and Millar (1987).

**Theorem 1.5.** Suppose \(F\) is permissible and that:

(i) \[
\sum_{j=1}^{\infty} 2^{-j} [H^{(2)}_F (2^{-j}, F)]^{1/2} < \infty \quad \text{or equivalently} \quad \int_0^{1} [H^{(2)}_F (x, F)]^{1/2} dx < \infty.
\]

(ii) \[
\int F^2 dP_0 < \infty.
\]

If \(m \wedge n \to \infty\), then for almost all sample sequences \(X_1, X_2, \cdots\), conditional on \(X_n = (X_1, \cdots, X_n)\),
\[(32) \quad \mathbb{X}_m^\# \Rightarrow \mathbb{X}_0^\# \equiv \mathbb{X}_0 \]
in \(L_\infty(F)\): i.e. for \(h \in C_b(L_\infty(F))\)
\[
E^*[h(\mathbb{X}_m^\#)|X_n] \to_a.s. E[h(\mathbb{X}_0^\#)] = E[h(\mathbb{X}_0)]
\]
as \(m \land n \to \infty\).

It is also possible to use our results to deduce a stability or regularity property of the bootstrap. Now suppose that \(X_{n1}, \ldots, X_{nm}\) are iid \(P_n\) where \(\rho_F(P_n, P_0) \to 0\) (and that the whole triangular array is defined on a common probability space as before). Then the resulting "bootstrap process" still converges:

**Theorem 1.6.** Suppose that \(F\) is permissible and that (i) - (iii) of corollary 1.4 hold. If \(m \land n \to \infty\), then conditional on \(X_n = (X_{n1}, \ldots, X_{nm})\), the processes \(\mathbb{X}_m^\# \Rightarrow \mathbb{X}_0^\# \equiv \mathbb{X}_0\) a.s. in \(L_\infty(F)\).

We conclude with the statement of a recent result of Ginë and Zinn (1988), and a corollary thereof. The following Ginë - Zinn bootstrap theorem is compared with our theorems 1.5 and 1.6 in remark 7 below. We use it in section 2.

**Theorem 1.7.** (Ginë and Zinn, 1988). Suppose that \(F\) is permissible. Then the following statements are equivalent:

(i) \(F\) is a \(P_0\)-Donsker class with \(P_0(F^2) < \infty\).

(ii) For a.e. sample sequence \(X_1, X_2, \ldots, \mathbb{X}_n^\# \Rightarrow \mathbb{X}_0^\# \equiv \mathbb{X}_0\) in \(L_\infty(F)\).

In fact, the measurability hypothesis on \(F\) used by Ginë and Zinn is weaker than permissible, but we have stated it here in terms of a permissible class \(F\) for simplicity.

**Remarks and discussion:**

2. Corollary 1.3 can be viewed as extending some of the results of Le Cam (1983; 1986). While Le Cam allows independent but non-identically distributed \(X_{ni}\)'s, his treatment is primarily for the case of bounded functions \(f \in F = F_n\) which may also depend on \(n\). [But note that Le Cam (1986) indicates that extensions of his results to unbounded functions are possible; see p. 549.] Our results treat only the iid case (with \(P_{ni} = P_n\) changing from row to row), but allow unbounded classes \(F\).

3. Since the collection of half-spaces in \(R^d\) is a Vapnik - Chervonenkis (1971) class (see also Dudley (1979)), corollary 1.4 contains proposition 4.1 of Beran and Millar as a special case. Their proof relies on the results of Le Cam (1983). Of course many other classes \(C\) are Vapnik - Chervonenkis classes; see e.g. Dudley (1984) or Shorack and Wellner (1986) chapter 26.
4. The notion of Hellinger-regularity in definition 1.6 and proved in corollary 1.6 is considerably stronger (i.e. involves more regularity) than the $n^{-1/2}$-Hellinger regularity of definition 1.6 and theorem 1.4. The latter is essentially the type of (local) regularity required by Beran (1977), Begun et al. (1983), and Millar (1985) following Hajek (1970), (1972) as a hypothesis for their convolution theorems for (locally) regular estimators, while the former is more closely related to the uniformity of convergence established by Kiefer and Wolfowitz (1959). We feel that even the stronger Hellinger-regularity of definition 1.6 is of interest and important in it's own right.

5. Dudley (1987) has shown that Pollard's entropy condition, hypothesis (i) of theorem 1.2 is sufficient for $\mathbf{F}$ to be a "universal Donker class": $\mathbf{X}_n \Rightarrow \mathbf{X}$ for every (fixed) $P$ on $(\mathcal{A}, \mathcal{P})$. Because Dudley (1986) insists on convergence for every $P$, his collections of functions $\mathbf{F}$ are forced to be bounded. Theorem 1.2 shows that Pollard's entropy condition, in combination with the uniform integrability hypothesis (ii) in fact yields uniform asymptotic equicontinuity, the crucial ingredient for weak convergence, uniformly in $P \in \mathbf{P}$.

6. While preparing the second revision of this paper (submitted first in August 1986) we learned of the work of Massart (1986). Massart also uses Pollard's (1982) notion of entropy. Theorem 5.10, page 411, of Massart (1986) gives rates of convergence for corollary 1.3 under the stronger assumptions $E F^{2+\delta} < \infty$ with $\delta > 0$ and

$$e^{(2)}_F(F) = \inf \{ s : \lim_{u \to 0} u^s H^{(2)}_F(u, F) < \infty \} < 2,$$

or

$$d^{(2)}_F(F) = \inf \{ s : \lim_{u \to 0} u^s N^{(2)}_F(u, F) < \infty \} = 2d < \infty.$$

Both of these hypotheses imply our condition (i). Massart does not emphasize or make explicit the uniformity in $P$ (which is, however, true) in his theorem. Both our argument and Massart's rely on the bounds of Yurinskii (1977), but with different truncation arguments.

7. While preparing the second revision of this paper, we also received a copy of Gin'e and Zinn (1988); their result for the bootstrap of a general empirical measure is stated above as theorem 1.7. We conjecture that their theorem remains valid for general $m$ if only $m \wedge n \to \infty$. Thus Gin'e and Zinn's theorem implies that "the bootstrap works" at $P_0$ for any $P_0-Donsker$ class with square-integrable envelope function. This is consistent with our theorem 1.5 which shows the "if" part of this assertion for a particular type of collection $\mathbf{F}$. Thus (if their theorem is valid for general $m$ as conjectured above) our theorem 1.5 is strictly contained in the Gin'e - Zinn theorem. Note that no uniform integrability assumption is needed for theorem 1.5. Our theorem 1.6 goes beyond the results of Gin'e and Zinn, and shows that, for the particular Donsker classes described by Pollard's entropy, the bootstrap estimators of the distribution of $\mathbf{X}_n$ have a strong Hellinger-regularity property, and hence are even more regular than the
regularity hypothesized by Beran (1982), (1984) in his studies of the asymptotic optimality of the bootstrap estimators. We return to this issue and consequences of the results of this section for estimation of function \( v \) of \( P \) in section 3.

8. If \( P \) is "sufficiently large" (e.g. if \( F \) is bounded and \( P = M \equiv \) all probability measures on \( (A, A_\varepsilon) \)), then we conjecture that the conclusion of theorem 1.3 implies condition (i) of theorem 1.2; i.e. that sparseness of the collection \( F \) in the sense of Pollard (1982) is necessary as well as sufficient. In any case, it is apparently still an open problem to characterize \( P \) – uniform Donsker classes \( F \).
2. The Skorokhod - Dudley$^2$ - Wichura theorem

To prepare the way for our delta - method results in section 3, we now give a brief account of the fourth generation Skorokhod - Dudley - Wichura theorem due to Dudley (1985). This theorem is geared to weak convergence in the sense of definition 1.3, thereby avoiding issues involving choice of a sigma-field for the nonseparable space $L_{\infty}(F)$ and measurability of the empirical processes.

We use Dudley's theorem to give "almost surely convergent constructions" for the empirical processes studied in section 1. These corollaries will be exploited in section 3 to study the delta method.

Readers interested primarily in only the basic delta method theorem 3.1 in section 3, rather than its regularity and bootstrap extensions, need to read only as far as corollary 2.1.

Preparation; perfect functions

As in Dudley and Philipp (1983) and Dudley (1984), chapter 3, for not necessarily measurable real - valued functions $g_n$ we say $g_n \to 0$ in probability, and write $g_n \to_p 0$, if $Pr^*(|g_n| > \varepsilon) \to 0$ as $n \to \infty$ for every $\varepsilon > 0$. By Dudley and Philipp (1983), lemma 2.5, $g_n \to_p 0$ if and only if $g_n^* \to_p 0$. We will also make frequent use of Dudley's (1985) proposition 1.1 giving basic equivalences for almost sure convergence ($\to_{a.s.}$) of nonmeasurable real - valued functions.

To state Dudley's theorem, we first need the notion of a perfect function.

**Definition 2.1** Let $(X, \mathcal{A}, P)$ be a probability space, let $(Y, \mathcal{B})$ be a measurable space, and let $\phi : X \to Y$ be measurable. Let $Q$ be the restriction of $P \circ \phi^{-1}$ to $\mathcal{B}$. Then $\phi$ is perfect if and only if for any bounded real - valued function $h$ on $Y$ $Q^*(h) = P^*(h \circ \phi)$.

As shown by Dudley (1985), theorem 2.1, page 146, this is equivalent to $(h \circ \phi)^* = h^* \circ \phi$ a.s. $P$ and to $Q^* = P^* \circ \phi^{-1}$.

Here is a key property of perfect functions that makes them useful and important: suppose that $Y : (\Omega, \Sigma, Q) \to (S, d)$ where $(S,d)$ is a metric space. Suppose there exists a perfect function $\phi : (\tilde{\Omega}, \tilde{\Sigma}, \tilde{Q}) \to (\Omega, \Sigma, Q)$ with $Q = \tilde{Q} \circ \phi^{-1}$. Define $\tilde{Y} : (\tilde{\Omega}, \tilde{\Sigma}) \to (S, d)$ by $\tilde{Y} = Y \circ \phi$. Then for any set $A \subset S$ $Q^*(\tilde{Y} \in A) = Q^*(Y \in A)$;

(1) $Q^*(\tilde{Y} \in A) = Q^*(Y \in A)$;

this follows from the definition since, with $h(\omega) \equiv 1_{[Y(\omega) \in A]}$ for $\omega \in \Omega$, $Q^*(Y \in A) = \tilde{Q}^*(Y \circ \phi \in A) = \tilde{Q}^*(\tilde{Y} \in A)$.

In this sense we can say that "$\tilde{Y} =_d Y$".
The S-D^2-W theorem

Here, then, is the fourth-generation Skorokhod theorem.

**Theorem 2.1.** (Skorokhod - Dudley^2 - Wichura). Let \((S,d)\) be any metric space, \((X_n,A_n,Q_n)\) any probability spaces, and \(Y_n\) a function from \(X_n\) into \(S\) for each \(n = 0, 1, \cdots\). Suppose that \(Y_0\) has separable range \(S_0 \subset S\) and is measurable. Then \(Y_n \Rightarrow Y_0\) (in the sense of definition 1.3) if and only if there exists a probability space \((\bar{\Omega}, \bar{\Sigma}, Q)\) and perfect measurable functions \(\phi_n\) from \((\bar{\Omega}, \bar{\Sigma})\) to \((X_n, A_n)\) for each \(n = 0, 1, \cdots\) such that:

(i) \(Q \circ \phi_n^{-1} = Q_n\) on \(A_n\) for each \(n = 0, 1, \cdots\),

(ii) \(\bar{Y}_n \equiv Y_n \circ \phi_n \to Y_0 \circ \phi_0 \equiv \bar{Y}_0\) almost uniformly; i.e.

\[
d(\bar{Y}_n, \bar{Y}_0)^* \to_{a.s.} 0 \quad \text{as} \quad n \to \infty
\]

For a complete discussion of theorem 2.1 and the notion of perfect functions, see Dudley (1985). For another discussion and exposition (with the same insistence on uniformly continuous functions \(h\) as in our definition 1.3), see Pollard (1988).

**Corollaries**

For our first application of theorem 2.1, we take \((S,d) = (L_\infty(F), \|\cdot\|_F)\), \(Y_n = X_n\), based on \(X\)’s iid \(P_0\) for \(n = 0, 1, \cdots\), so that \(Y_0 = X_0\) has range \(S_0 = C(F, P_0) = \Omega_0\) which is separable under our assumptions on \(F\). Let \(\Sigma_0\) denote the Borel sigma-field of \(C(F, P_0) = \Omega_0\). Thus a first consequence of the \(S-D^2-W\) theorem is (this is, in fact, an immediate consequence of Dudley’s (1985) theorem 5.2):

**Corollary 2.1.** (Fixed \(P_0\)). Suppose that \(F\) is a \(P_0\)-Donsker class, and let \((\Omega, \Sigma, Pr_{P_0})\) be the probability space of theorem 4.1.1, Dudley (1984). Then on some probability space \((\bar{\Omega}, \bar{\Sigma}, Q)\) there is a sequence of perfect functions

\[
\phi_n : (\bar{\Omega}, \bar{\Sigma}, Q) \to (\Omega, \Sigma, Pr_{P_0}), \quad n \geq 1,
\]

\[
\phi_0 : (\bar{\Omega}, \bar{\Sigma}, Q) \to (\Omega_0, \Sigma_0, L(X_0)),
\]

so that \(Q \circ \phi_n^{-1} = Pr_{P_0}(A^n, A^n), \quad n = 1, 2, \cdots\), \(Q \circ \phi_0^{-1} = L(X_0) = L(G_{P_0})\) on \((\Omega_0, \Sigma_0)\), and, with

\[
\bar{X}_n \equiv X_n \circ \phi_n, \quad n = 0, 1, \cdots
\]

(so that \(Q^*(\bar{X}_n \in A) = Pr_{P_0}^*(X_n \in A), \quad n = 0, 1, \cdots\) for any set \(A \in L_\infty(F)\))

\[
\|\bar{X}_n - \bar{X}_0\|_F^* \to_{a.s.} 0 \quad \text{as} \quad n \to \infty.
\]
Proof. This follows immediately from Dudley's (1984) theorem 4.1.1 and theorem 2.1. □

For the remaining corollaries we introduce and use the following notation: if \( \theta_i : (\Omega_i, \Sigma_i, P_i) \rightarrow (\mathcal{S}, d) \), \( i = 1, 2 \), satisfy

\[
P_1^*(\theta_1 \in A) = P_2^*(\theta_2 \in A) \quad \text{for any set } A \in \mathcal{S},
\]
then we write \( \theta_1 \overset{d}{=}_d \theta_2 \).

For our second application, we consider the empirical process of \( X \)'s iid \( P_n \) with \( \rho_F(P_n, P_0) \rightarrow 0 \) as in corollary 1.4.

Corollary 2.2. (General sequence \( \{P_n\} \)). Suppose that the hypotheses (and hence the conclusion) of corollary 1.4 hold, and let \( (\Omega, \Sigma, Pr) \) denote the probability space on which the weak approximation of corollary 1.4 holds. Then on some probability space \( (\Omega, \Sigma, Q) \) there is a sequence of perfect functions \( \phi_n : (\Omega, \Sigma, Q) \rightarrow (\Omega, \Sigma, Pr_{P_n}) \), \( n = 0, 1, \cdots \), so that \( \overline{X}_n \overset{d}{=} \overline{X}_n \circ \phi_n \overset{d}{=} \overline{X}_n \), \( n = 0, 1, \cdots \), and

\[
(4) \quad \|\overline{X}_n - \overline{X}_0\|^*_F \rightarrow a.s. 0 \quad \text{as } n \rightarrow \infty.
\]

Proof. This follows immediately from corollary 1.4 and theorem 2.1. □

For our third application, consider the empirical process of \( X \)'s iid \( P_n \) as in theorem 1.4.

Corollary 2.3. (Contiguous sequence \( \{P_n\} \)). Suppose that \( F \) and \( \{P_n\} \) satisfy the hypotheses of theorem 1.4, and let \( (\Omega, \Sigma, Pr) \) denote the probability space constructed there. Then on some probability space \( (\Omega, \Sigma, Q) \) there is a sequence of perfect functions \( \phi_n : (\Omega, \Sigma, Q) \rightarrow (\Omega, \Sigma, Pr) \), \( n = 0, 1, \cdots \), so that

\[
\overline{X}_n \overset{d}{=} \overline{X}_n \circ \phi_n \overset{d}{=} \overline{X}_n , \quad n = 0, 1, \cdots ,
\]
and, with

\[
\overline{X}_n^0 \overset{d}{=} \overline{X}_n \circ \phi_n + x_n
\]
\[
= (\overline{X}_n + x_n) \circ \phi_n = \overline{X}_n^0 \circ \phi_n , \quad n = 1, 2, \cdots ,
\]
where \( x_n = \sqrt{n} (P_n - P_0) \) as in (1.27), we have,

\[
(5) \quad \|x_n - x_0\|_F \rightarrow 0 \quad \text{as } n \rightarrow \infty,
\]

\[
(6) \quad \|\overline{X}_n - \overline{X}_0\|_F^* \rightarrow a.s. 0 \quad \text{as } n \rightarrow \infty.
\]

and

\[
(7) \quad \|\overline{X}_n^0 - (\overline{X}_0 + x_0)\|_F^* \rightarrow a.s. 0 \quad \text{as } n \rightarrow \infty.
\]

Proof. Immediate from theorem 1.4 and theorem 2.1. □
Now for the bootstrap corollaries.

**Corollary 2.4.** (Single bootstrap construction). Suppose that $F$ is a $P_0$-functional Donsker class with $P_0(F^2) < \infty$ as in theorem 1.7. Then, for any fixed $\omega \in C$ where $C \in \Sigma$ is the set with $Pr_{P_0}(C) = 1$ guaranteed by theorem 1.7, on some probability space $(\tilde{\Omega}^\#$, $\tilde{\Sigma}^\#$, $\tilde{Q}^\#$) there is a sequence of perfect functions $\phi_n^\# = \phi_n^\# : (\tilde{\Omega}^\#, \tilde{\Sigma}^\#, \tilde{Q}^\#) \to (\Omega^\#, \Sigma^\#, Pr_{P_0}(\omega))$, $n = 0, 1, \cdots$, so that, conditional on $X_1, \cdots, X_n$, \cdots,

$$\mathfrak{X}_n^\# \equiv \mathfrak{X}_n^\# \circ \phi_n^\# \overset{d}{=} \mathfrak{X}_n^\#,$$ \hspace{1cm} n = 1, 2, \cdots,

$$\mathfrak{X}_0^\# \overset{d}{=} \mathfrak{X}_0^\#,$$

and

$$||\mathfrak{X}_n^\# - \mathfrak{X}_0^\#||_F^* \to_{a.s.} 0 \quad \text{as} \quad n \to \infty.$$  

**Proof.** This follows from the Gin'e - Zinn theorem 1.7 and theorem 2.1. $\square$

For our weak consistency theorem for the bootstrap in the next section, the following more elaborate "double construction" will be used.

**Corollary 2.5.** (Double bootstrap construction). Suppose that $F$ is a $P_0$-Donsker class with $P_0(F^2) < \infty$ as in theorem 1.7. Then on some probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{Q})$ there is a sequence of perfect functions $\phi_n : (\tilde{\Omega}, \tilde{\Sigma}, \tilde{Q}) \to (\Omega, \Sigma, Pr_{P_0})$, $n = 0, 1, 2, \cdots$, so that $\mathfrak{X}_n \equiv \mathfrak{X}_n \circ \phi_n \overset{d}{=} \mathfrak{X}_n$, $n = 0, 1, \cdots$, $IP_n \equiv IP_n \circ \phi_n \overset{d}{=} IP_n$ for $n = 1, 2, \cdots$, and, for fixed $\bar{\omega} \in \tilde{\Omega}$, $X_1^\#, \cdots, X_n^\#$ iid $IP_n(\omega) = IP_n \circ \phi_n(\bar{\omega})$ on $(A, A)$ with empirical measure $IP_n^\#$ and empirical process

$$\mathfrak{X}_n^\# \equiv \sqrt{n} \left( IP^\#_n - IP^\#_n(\bar{\omega}) \right),$$

we have both

$$||\mathfrak{X}_n - \mathfrak{X}_0||_F^* \to_{a.s.} 0$$

and

$$\overset{a.s.}{\tilde{Q}^\#} \Rightarrow \mathfrak{X}_n^\# \overset{d}{=} \mathfrak{X}_0^\# \quad \text{as} \quad n \to \infty.$$  

Let $\tilde{C} \in \tilde{\Sigma}$ be the set with $\tilde{Q}(\tilde{C}) = 1$ such that both (9) and (10) hold. For any fixed $\bar{\omega} \in \tilde{C}$, on some probability space $(\tilde{\Omega}^\#, \tilde{\Sigma}^\#, \tilde{Q}^\#)$ there is a sequence of perfect functions $\phi_n^\# = \phi_n^\# : (\tilde{\Omega}^\#, \tilde{\Sigma}^\#, \tilde{Q}^\#) \to (\Omega^\#, \Sigma^\#, Pr_{P_0}(\bar{\omega}))$ for $n = 0, 1, \cdots$, so that $\mathfrak{X}_n^\# \equiv \mathfrak{X}_n^\# \circ \phi_n^\# \overset{d}{=} \mathfrak{X}_n^\#$,

$$||\mathfrak{X}_n^\# - \mathfrak{X}_0^\#||_F^* \to 0 \quad \text{a.s.} \quad \tilde{Q}^\# \quad \text{as} \quad n \to \infty.$$  

Let $\tilde{C}^\# \subset \tilde{\Sigma}^\#$ be the set with $\tilde{Q}^\#(\tilde{C}^\#) = 1$ so that (11) holds. Then, finally, with $\mathfrak{X}_0 \equiv \sqrt{n} \left( IP^\#_n - P_0 \right)$,
for \( \tilde{\omega} \in \tilde{C} \) and \( \tilde{\omega}^\# \in \tilde{C}^\# \),
\[
\|\tilde{X}_n^{\#0} - (\tilde{X}_0^\# + \tilde{X}_0(\tilde{\omega}))\|_F(\tilde{\omega}^\#) \to 0
\]
as \( n \to \infty \).

**Proof.** This requires a little care in order to arrange that both (9) and (10) hold. First note that since \( F \) is a \( P_0 \)-Donsker class with \( P_0(F^2) < \infty \), then it follows from the argument of Giné and Zinn (1988), page 8, that
\[
(a) \quad Y_{n1} = \max_{m \geq n} \rho_F(P_m, P_0) \to_p 0 \quad \text{as} \quad n \to \infty ,
\]
\[
(b) \quad Y_{n2} = \max_{m \geq n} \max_{1 \leq i \leq m} F(X_i)/m^{1/2} \to_p 0 \quad \text{as} \quad n \to \infty ,
\]
and, for each \( \omega \in \Omega \) and \( \varepsilon > 0 \) there is a \( \delta = \delta(\varepsilon, \omega) \) such that with
\[
(c) \quad Z_n(\omega) = Pr^*_{\varepsilon} \{ \sup_{[\delta], r, n} |X_n^\#(f - g)| > \varepsilon |X_1, \cdots, X_n| \}(\omega),
\]
\[
(d) \quad Y_{n3} = \max_{m \geq n} Z_m \to_p 0 .
\]

Hence, by Dudley (1985) corollary 3.4, \( Y_n = (Y_{n1}, Y_{n2}, Y_{n3}) \Rightarrow 0 \in R^3 \) as \( n \to \infty \) in the sense of definition 1.3. Thus, with \( (S, d) \) of theorem 2.1 taken to be 
\[
(L_\omega(F) \times R^3, \|\|_F \vee |\|) ,
\]
it follows that
\[
(e) \quad (X_n, Y_n) \Rightarrow (X_0, 0) \quad \text{as} \quad n \to \infty .
\]

Now theorem 2.1 yields perfect maps \( \phi_n : (\hat{\Omega}, \hat{\Sigma}, \hat{\mathcal{Q}}) \to (\Omega, \Sigma, Pr_{P_0}), \quad n \geq 1 , \)
\( \phi_0 : (\hat{\Omega}, \hat{\mathcal{X}}, \hat{\mathcal{Q}}) \to (\Omega_0, \Sigma_0, L(X_0)) \) such that, with \( (X_n \circ \phi_n, Y_n \circ \phi_n) \equiv (\tilde{X}_n, \tilde{Y}_n) ,
\]
\[
(f) \quad (\tilde{X}_n, \tilde{Y}_n) \to_{a.s.} (\tilde{X}_0, 0) \quad \text{as} \quad n \to \infty .
\]
Thus (9) holds, and by the argument of Giné and Zinn (1988), \( \tilde{Y}_n \to_{a.s.} 0 \) implies that
(10) holds.

For fixed \( \tilde{\omega} \in \tilde{C} \), (11) follows from a second application of theorem 2.1 in the same way as corollary 2.4. Finally, for fixed \( \tilde{\omega} \in \tilde{C} \), \( \tilde{\omega}^\# \in \tilde{C}^\# \),
\[
\|\tilde{X}_n^{\#0} - (\tilde{X}_0^\# + \tilde{X}_0(\tilde{\omega}))\|_F(\tilde{\omega}^\#)
\]
\[
\leq \|\tilde{X}_n^\# - \tilde{X}_0^\#\|_F(\tilde{\omega}^\#) + \|\tilde{X}_n(\tilde{\omega}) - \tilde{X}_0(\tilde{\omega})\|_F
\]
\[
\leq \|\tilde{X}_n - \tilde{X}_0\|_F(\tilde{\omega}^\#) + \|\tilde{X}_n - \tilde{X}_0\|_F(\tilde{\omega})
\]
\[
\to 0 + 0 = 0 \quad \text{as} \quad n \to \infty
\]
by (9) and (11). \( \Box \)
3. The delta method

We often want to consider estimation of some other function (or parameter) $v(P)$ instead of $P$ itself. Here $v: P \to B$, and we assume that $B$ is a Banach space. The estimator of choice (based on $X_1, \ldots, X_n$ iid $P$) is often just $v(P_n)$. Then, if $v$ is differentiable with derivative $\dot{v}$ in an appropriate sense, a generalization of the classical "delta method" should yield

$$\sqrt{n} \left\{ v(P_n) - v(P) \right\} = \dot{v}(\sqrt{n} (P_n - P)) = \dot{v}(\mathbb{X}) \Rightarrow \dot{v}(\mathbb{X}_0)$$

by weak convergence of $\mathbb{X}_n$ to $\mathbb{X}_0$.

Our goals in this section are: to make (1) precise; and to give extensions of (the precise version of) (1) which:

(i) yield regularity (and $n^{-1/2}$ regularity) of such estimators $v(P_n)$; i.e. allow replacement of $P$ by $P_n$ in (1) and $X_{n1}, \ldots, X_{nn}$ iid $P_n$.

(ii) imply strong (and weak) consistency of the "bootstrap" (i.e. allow replacement of $P$ by $IP_n$ and $IP_n$ by $IP_n^*$ in (1), with $\text{m} \wedge n \to \infty$).

Our approach is based upon and extends the "delta method" results of Reeds (1976) and Gill (1988). It involves differentiability hypotheses related to compact (= Hadamard) differentiation as in Reed's (1976), and the notion of compact differentiation tangentially to a subspace introduced by Gill (1988). Both Reeds (1976) and Gill (1988) consider primarily the case of real-valued random variables and empirical df's thereof as estimates of the corresponding population df, rather than more general random variables and empirical measures as we do here. Gill (1988) makes good use of the Skorokhod - Dudley - Wichura theorem (see e.g. Shorack and Wellner 1986, chapter 2) in his setting. We generalize Gill's approach and results to functions of empirical measures more generally, viewed as elements of $L_\infty(F)$, and use Dudley's (1985) $S - D^2 - W$ theorem 2.1. Our formulation removes measurability conditions and puts the theorems in a simple form.

A merit of this approach is that it yields easily stated "clean" theorems which clearly separate the stochastic aspects of the problem (convergence of $IP_n$ to $P$ and convergence, or weak approximability, of the process $\mathbb{X}_n$) from the analytical aspects of the problem (differentiability of the map $v$). Furthermore, the conclusion is simple in spirit and closely tied to the basic "delta method idea". It is especially valuable if the same (differentiable) function $v$ is to be used in a variety of different sampling situations; i.e. with different estimators of $P$ rather than just the empirical measure $IP_n$ based on iid $X$'s : the derivative $\dot{v}$ really needs to be calculated just once! This will be illustrated in section 4 with several examples.

An alternative approach, via asymptotic linearity, is also possible, but does not seem to give as much insight, so we will not pursue it here.
Readers primarily interested in the basic method result itself, and not the regularity and bootstrap extensions thereof, need read only as far as theorem 3.1.

The delta method via compact, or Hadamard, differentiation

We first develop an appropriate definition of a differentiable function \( v \) of \( P \in P \subseteq M \equiv \{ \text{all probability measures on } (A,\mathcal{A}) \} \). We assume that \( P \) is large enough to contain the empirical measures \( IP_n \) (or at least that \( Pr_p(IP_n \in P) = 1 \) for \( n \) sufficiently large and \( P \in P \)).

Suppose that \( P \in P \) and that \( F \subseteq L_2(P) \) for each \( P \in P \) is permissible with envelope function \( F \) satisfying \( \sup_{P \in P} P(F) < \infty \); more will often be required of \( F \) in our theorems.

Suppose that \( v: P \to B \) where \( B \) is a Banach space. We regard \( P \) as a subset of \( L_\infty(F) \), which is possible by our current assumption on \( F \). We do not impose any measurability requirement on the function \( v \).

Here is the variant of compact differentiability which we will use. We fix \( P_0 \in P \) and a pseudo-metric \( d \) on \( P \).

**Definition 3.1.** Let \( P(P_0,d) \) be a collection of sequences \( \{P_n\} \subseteq P \) with \( d(P_n,P_0) \to 0 \) as \( n \to \infty \). Then \( v: P \subseteq L_\infty(F) \to B \) is **differentiable tangentially to** \( C(F,P_0) \) at \( P_0 \) relative to \( P(P_0,d) \) if there is a continuous linear function \( \dot{v}(P_0) \equiv \dot{v}_0: C(F,P_0) \to B \) such that: for any \( \{P_n\} \subseteq P(P_0,d) \), any sequence \( \{D_n\} \) with \( \|D_n - D_0\|_F \to 0 \) where \( D_0 \in C(F,P_0) \), and any sequence of real numbers \( \varepsilon_n \to 0 \) satisfying \( Q_n = P_n + \varepsilon_n D_n \in P \),

\[
(2) \quad \frac{v(P_n + \varepsilon_n D_n) - v(P_n)}{\varepsilon_n} \to \dot{v}_0(D_0) \quad \text{in } B \quad \text{as } n \to \infty.
\]

If \( \dot{v}_0 \) has an extension (again denoted by \( \dot{v}_0 \)) to a continuous linear function on all of \( L_\infty(F) \), then (2) implies that

\[
(3) \quad \left\| \frac{v(P_n + \varepsilon_n D_n) - v(P_n)}{\varepsilon_n} - \dot{v}_0(D_n) \right\|_B \to 0 \quad \text{as } n \to \infty,
\]

and, of course,

\[
(4) \quad \|\dot{v}_0(D_n) - \dot{v}_0(D_0)\|_F \to 0 \quad \text{as } n \to \infty
\]

by continuity of \( \dot{v}_0 \).

If (2) holds for just one sequence \( \varepsilon_n \to 0 \), then we say that \( v \) is \( \{\varepsilon_n\} \)-differentiable tangentially to \( C(F,P_0) \) at \( P_0 \) relative to \( P(P_0,d) \). In all of our applications here \( \varepsilon_n = n^{-1/2} \) is the key sequence, and it suffices that \( v \) be \( \{n^{-1/2}\} \)-differentiable tangentially to \( C(F,P_0) \) at \( P_0 \) relative to \( P(P_0,d) \). For many functions \( v \), this distinction is not important, and the differentiability holds for any \( \varepsilon_n \to 0 \). For others it is crucial -- often in connection with a description of the
collection \( P(P_0, d) \); e.g. see our treatment of quantiles in section 4.

If \( \nu \) is differentiable tangentially to \( C(F, P_0) \) – relative to \( P_0 \) – tangentially differentiable at \( P_0 \) and \( d(P_n, P_0) \to 0 \), then we say that \( \nu \) is \( d \)-continuously \( C(F, P_0) \) – tangentially differentiable at \( P_0 \). If \( P(P_0, d) \) consists of only the trivial (constant) sequence \( \{ P_0 \} \), then we say that \( \nu \) is \( C(F, P_0) \) – tangentially differentiable at \( P_0 \). In view of the results of section 1, we usually choose \( d \) to be either \( \rho_F \) or the Hellinger metric \( H \).

For a more leisurely discussion and motivation of this type of differentiability, see the illuminating exposition of Gill (1988), especially his theorems 3 and 4 and lemma 1.

First the basic result which makes (1) precise in the case of a fixed law \( P_0 \): assume that that \( X_1, \ldots, X_n \) are iid \( P_0 \in P \) with corresponding empirical measure \( \hat{P}_n \) and empirical process

\[
\mathbb{X}_n = \sqrt{n} (\hat{P}_n - P_0). 
\]

Theorem 3.1. (Basic \( \delta \)-method). Suppose that:

(i) \( \nu: P \to B \) is \( C(F, P_0) \) – tangentially differentiable at \( P_0 \) with derivative \( \nu_0: C(F, P_0) \to B \).

(ii) \( F \) is a \( \nu \)- (functional) Donsker class: \( \mathbb{X}_n = \sqrt{n} (\hat{P}_n - P) \Rightarrow \mathbb{X}_0 \) in \( \text{L}_{\infty}(F) \) in the sense of definition 2.1.

Then, in the probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) of corollary 2.1,

\[
||\sqrt{n} (\nu(\hat{P}_n) - \nu(P_0)) - \nu_0(\mathbb{X}_0)||_B^* \to_{a.s.} 0 \quad \text{as} \quad n \to \infty.
\]

Proof. Let \( \mathbb{X}_n \hat{=} \) be defined as in corollary 2.1, and note that

(a) \( \hat{P}_n = P_0 + n^{-1/2} \mathbb{X}_n \).

By (2.3) of corollary 2.1, for \( \tilde{\omega} \in C \) with \( Q(C) = 1 \),

(b) \( ||\mathbb{X}_n(\tilde{\omega}) - \mathbb{X}_0(\tilde{\omega})||_F \leq ||\mathbb{X}_n - \mathbb{X}_0||_F(\tilde{\omega}) \to 0 \) as \( n \to \infty \). Hence for \( \tilde{\omega} \in C \),

\[
||\sqrt{n} (\nu(\hat{P}_n(\tilde{\omega})) - \nu(P_0)) - \nu_0(\mathbb{X}_0(\tilde{\omega}))||_B \\
= ||\sqrt{n} (\nu(P_0 + n^{-1/2} \mathbb{X}_n(\tilde{\omega})) - \nu(P_0)) - \nu_0(\mathbb{X}_0(\tilde{\omega}))||_B \\
\to 0
\]

by (b) and definition 2.1. Hence, with \( Y_n = \sqrt{n} (\nu(\hat{P}_n) - \nu(P_0)) \) and \( Y_0 = \nu(\mathbb{X}_0) \),

\[
[||Y_n - Y_0||_B \to 0] \supset C
\]

or \( [||Y_n - Y_0||_B \not\to 0] \subset C^c \), which implies that

\[
Q^*(||Y_n - Y_0||_B \not\to 0) \leq Q(C^c) = 0.
\]
Hence (6) holds by proposition 1.1 of Dudley (1985).

This is closely related to the final conclusion of Gill's (1988) theorem 3 with his $X_n = \text{our } P_n$ and his $E \subset B_1$ in his lemma our $P \subset L_\infty(F)$.

Note that, as usual with the delta-method, the iid structure which we have used to guarantee the weak convergence result in the hypothesis (ii) of theorem 3.1 is not needed at all; the only thing used in the proof is that $P_n = P_0 + n^{-1/2}X_n$ where $X_n$ converges weakly. Hence the theorem can be applied far beyond the iid setting made explicit above. We will give examples of this in section 4.

Now for a theorem of this type yielding regularity of the estimator sequence $v(P_n)$. We first generalize the regularity definition 1.6; in keeping with the spirit of this section, we use a weak approximation formulation of regularity which avoids measurability hypotheses.

**Definition 3.2. (Regular estimators).** Let $P(P_0, d)$ be a collection of sequences \{P_n\} $\subset P$ with $d(P_n, P_0) \to 0$ as $n \to \infty$. Suppose that $v : P \to B$ is a $B$-valued parameter, and that $\hat{v}_n$ is an "estimator" of $v(P)$: i.e. $\hat{v}_n$ is a not necessarily measurable function from $\Omega_n \equiv A^n$ to $B$. Then \{\hat{v}_n\} is a (weakly approximable) regular "estimator" of $v(P)$ at $P_0$ relative to $P(P_0, d)$ if $\mathbb{Z}_n \equiv \sqrt{n} (\hat{v}_n - v(P_n))$ and $\mathbb{Z}_0 \in B_0$ taking values in a separable subspace $B_0 \subset B$ (with its Borel sigma field) can be constructed on a common probability space so that $\|\mathbb{Z}_n - \mathbb{Z}_0\|_B \to_\rho 0$ for every sequence \{P_n\} $\subset P(P_0, d)$ where the (distribution of the) limit process $\mathbb{Z}_0$ depends on $P_0$ but not on the sequence \{P_n\}; i.e. there exists some probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{Q})$ with processes $\{\bar{\mathbb{Z}}_n\}_{n \geq 0}$ defined thereon so that $\bar{\mathbb{Z}}_n \overset{d}{=} \mathbb{Z}_n$ for $n = 0, 1, \ldots$ and $\|\bar{\mathbb{Z}}_n - \bar{\mathbb{Z}}_0\|_B \to_\rho 0$.

When \{\hat{v}_n\} is a regular estimator of $v(P)$ at $P_0$ relative to the collection $P(P_0, H)$ of all sequences \{P_n\} $\subset P$ with $H(P_n, P_0) \to 0$, we say that $\hat{v}_n$ is a (weakly approximable) Hellinger-regular "estimator" of $v(P)$ at $P_0$.

If such a construction and conclusion holds for any sequence \{P_n\} satisfying (1.25), we say that $\hat{v}_n$ is a (weakly approximable) $n^{-1/2}$-Hellinger-regular "estimator" of $v(P)$ at $P_0$.

As in section 1, weakly approximable Hellinger regularity is a stronger notion of regularity than the (weakly approximable) $n^{-1/2}$-Hellinger regularity (or local regularity) often hypothesized in efficiency studies; recall remark 1.4.

Now assume that $X_{n1}, \ldots, X_{nn}$ are iid $P_n \in P$ with $H(P_n, P_0) \to 0$, with corresponding empirical measure $P_n$ and empirical process

\begin{equation}
\mathbb{X}_n = \sqrt{n} (\bar{P}_n - P_n)
\end{equation}

as in theorem 1.2.
Theorem 3.2. (Hellinger regularity of the $\delta$-method). Suppose that $\{P_n\}$ satisfies $H(P_n, P_0) \to 0$ and:

(i) $\nu : P \to B$ is $H$-continuously, $C(F, P_0)$ - tangentially differentiable at $P_0$ with derivative $\dot{\nu}_0 : C(F, P_0) \to B$.

(ii) $F$ is a $\{P_n\}$ - uniform (functional) Donsker class with $\{P_n\}$ - uniformly square integrable envelope function $F$.

Then, on the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, Q)$ of corollary 2.2

$||\sqrt{n} \left(\nu(P_n) - \nu(P)\right) - \dot{\nu}_0(\bar{X}_0)||_B^* \to a.s. 0$ as $n \to \infty$.

In particular, if $F$ is a $P_0$ - uniform (functional) Donsker class with $P_0$ - uniformly square envelope function $F$ and (i) holds at every $P_0 \in P_0 \subset P$, then $\{\nu(P_n)\}$ is a Hellinger - regular estimator of $\nu(P)$ on $P_0$.

Proof. Exactly like the proof of theorem 3.1, but with $C(F, P_0)$ - tangential differentiability strengthened to $H$ - continuously, $C(F, P_0)$ - tangential differentiability, and corollary 2.1 replaced by corollary 2.2. Note that $\{P_n\}$ - uniform square integrability of $F$ implies that $\rho_F(P_n, P_0) \to 0$ as $n \to \infty$ by corollary 1.7.

Note that (ii) holds under the hypotheses of our corollary 1.4. Also note that if we replace hypothesis (i) by

(i') $\nu : P \to B$ is differentiable tangentially to $C(F, P_0)$ relative to $P(P_0, H)$

for some particular collection $P(P_0, H)$, then since (ii) holds for any sequence $\{P_n\} \in P(P_0, H)$ under the hypotheses of our theorem 1.3, the amount of regularity available in the conclusion (8) is essentially governed by the collection $P(P_0, H)$ in the differentiability condition (i').

To complete this first collection of results, we give a theorem which guarantees consistency of the bootstrap for estimation of the distribution of measurable functions of $\sqrt{n} (\nu(P_n) - \nu(P))$. We use the notation of theorem 1.7: $X_1, X_2, \cdots$ are assumed to be iid $P_0$ with empirical measures $\{IP_n\}$; then for a fixed $n \geq 1$, $X_1^#, \cdots, X_n^#$ are iid $IP_n(\omega) \equiv IP_n^{\omega}$, with empirical measure $IP_n^\omega$, and

$\bar{X}_n^# \equiv \sqrt{n} (IP_n^# - P_n^{\omega})$

is the bootstrap empirical process.

Theorem 3.3. (Strong consistency of the bootstrap). Suppose that:

(i) $\nu : P \to B$ is $\rho_F$ - continuously, $C(F, P_0)$ - tangentially differentiable at $P_0$ with derivative $\dot{\nu}_0 : C(F, P_0) \to B$.
(ii) \( F \) is a \( P_0 - \) (functional) Donsker class with square integrable envelope function \( F \).

Then, for the same fixed \( \omega \in C \) as in corollary 2.4, on the probability space \((\hat{\Omega}, \hat{\Sigma}, \hat{\mathcal{Q}})\) of corollary 2.4,

\[
\|\sqrt{n} \big( \nu(P_n^{\omega}) - \nu(P_0^{\omega}) \big) - \nu_0(\hat{\mathcal{X}}_0) \|_B \to_{a.s.} 0 \quad \text{as} \quad n \to \infty
\]

where \( \nu(\hat{\mathcal{X}}_0) \equiv \nu_0(\hat{\mathcal{X}}_0) \equiv \nu_0(\mathcal{X}_0) \).

**Proof.** Exactly like the proof of theorem 3.1, but with \( C(F, P_0) - \) tangential differentiability strengthened to \( \rho_F - \) continuously, \( C(F, P_0) - \) tangential differentiability, and corollary 2.1 replaced by corollary 2.4. \( \square \)

This extends theorem 4 of Gill (1988) and lemma 8.10 of Bickel and Freedman (1981). It can be refined still further: instead of requiring that \( \nu \) be \( \rho_F - \) continuously, \( C(F, P_0) - \) tangentially differentiable at \( P_0 \), we can in fact just require \( \nu \) to be \( \{n^{-1/2}\} - \) differentiable tangentially to \( C(F, P_0) \) at \( P_0 \) relative to \( P(P_0, \rho_F) \) where \( P(P_0, \rho_F) \) is any collection satisfying

\[
Pr_{P_0}(\{P_n\} \in P(P_0, \rho_F)) = 1.
\]

In other words, using properties of \( IP_n \) which hold a.s., such as oscillation moduli, tail behavior, law of the iterated logarithm, etc. (see e.g. Shorack and Wellner (1986), chapters 13 - 17 for results of this type in the case of one-dimensional empirical processes, and Alexander (1987) for related results in more general situations), the continuity requirement in the differentiability condition in theorem 3.3 can be reduced quite substantially. The following refinement of theorem 3.3 will be useful in treating quantile and quantile processes in the following section. It extends Gill's (1988) theorem 4 in the spirit of Beran (1984).

**Theorem 3.3.A.** (Strong consistency of the bootstrap). Suppose that:

(i) \( \nu : P \to B \) is \( \{n^{-1/2}\} - \) differentiable tangentially to \( C(F, P_0) \) at \( P_0 \) relative to \( P(P_0, \rho_F) \) with derivative \( \nu_0 : C(F, P_0) \to B \) where

\[
Pr_{P_0}(\{IP_n\} \in P(P_0, \rho_F)) = 1.
\]

(ii) \( F \) is a \( P_0 - \) (functional) Donsker class with square integrable envelope function \( F \).

Then, for the same fixed \( \omega \in C \) as in corollary 2.4, on the probability space \((\hat{\Omega}, \hat{\Sigma}, \hat{\mathcal{Q}})\) of corollary 2.4,

\[
\|\sqrt{n} \big( \nu(P_n^{\omega}) - \nu(P_0^{\omega}) \big) - \nu_0(\hat{\mathcal{X}}_0) \|_B \to_{a.s.} 0 \quad \text{as} \quad n \to \infty
\]

where \( \nu(\hat{\mathcal{X}}_0) \equiv \nu_0(\hat{\mathcal{X}}_0) \equiv \nu_0(\mathcal{X}_0) \).
Theorems 3.2 and 3.3 require not just \( C(F, P_0) \) - tangential differentiability at \( P_0 \), but some continuity in the differentiation: \( H \) - continuity for theorem 3.2, and \( \rho_p \) - continuity for theorem 3.3. The pay-off is a very strong kind of regularity in the conclusion of theorem 3.2, and almost sure consistency of the bootstrap in the conclusion of theorems 3.3 and 3.3.A. The following two theorems parallel theorems 3.2 and 3.3. They show that a weaker differentiability assumption -- namely just \( C(F, P_0) \) - tangential differentiability as in theorem 2.1 -- yields weaker conclusions: a parallel result to theorem 3.2 is that for \( C(F, P_0) \) - tangentially differentiable \( \nu \) and (contiguous) sequences \( P_n \) satisfying (1.23) (so \( H(P_n, P_0) = O(n^{-1/2}) \)), we automatically have \( n^{-1/2} \) - Hellinger regularity of \( \nu(P_n) \) without any additional continuity assumption on the differentiability of \( \nu \). Similarly, a parallel result to theorem 3.3 is that for \( C(F, P_0) \) - tangentially differentiable \( \nu \) and \( F \) a \( P_0 \) - Donsker class, it automatically follows that the bootstrap is weakly - consistent without any further continuity assumptions on the differentiability of \( \nu \). In view of theorem 1.4 and the Gin'e - Zinn bootstrap theorem 1.7 respectively, which guarantee that the needed empirical process results holds without any further hypotheses than those needed for a fixed \( P_0 \), the following theorems assert that \( n^{-1/2} \) - Hellinger regularity and weak consistency of the bootstrap hold automatically for \( C(F, P_0) \) - tangentially differentiable functions \( \nu \).

While it is nice to know that these weaker properties come for free, it is not clear to us that the weaker properties -- namely \( n^{-1/2} \) - Hellinger regularity and weak consistency of the bootstrap -- are the relevant properties for many applications of interest. It is clear that the stronger properties -- Hellinger regularity and strong consistency of the bootstrap -- are of interest for many applications; our results clarify the price to be paid in order that they hold. The analogies and parallels between theorems 1.4 and 1.7 and theorems 3.4 and 3.5 also seem instructive.

**Theorem 3.4.** (\( n^{-1/2} \) - Hellinger regularity of the \( \delta \) - method).

Suppose that:

(i) \( \nu: P \rightarrow B \) is \( C(F, P_0) \) - tangentially differentiable at \( P_0 \) with derivative \( \nu_0: C(F, P_0) \rightarrow B \).

(ii) The hypotheses (and hence also the conclusions) of theorem 1.5 hold.

Then, on the probability space \( (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{Q}) \) of corollary 2.3,

\[
\left\| \sqrt{n} (\nu(P_n) - \nu(P_0)) - \nu_0(\tilde{X}_0) \right\|_B^* \rightarrow_{a.s.} 0 \quad \text{as} \quad n \rightarrow \infty.
\]

In particular, if \( F \) is a \( P_0 \) - (functional) Donsker class with envelope function \( F \) satisfying \( \sup_{P \in \mathcal{P}} P(F^2) < \infty \), then \( \{\nu(P_n)\} \) is a weakly - approximable \( n^{-1/2} \) - Hellinger - regular estimator of \( \nu(P) \) at \( P_0 \).
Proof. The key observation is that both (2.5) and (2.7) hold. Therefore, by linearity of \( \nu_0 \),

\[
\sqrt{n} \left( \nu(P_n) - \nu(P_0) \right) - \nu_0(\mathbb{F}_0)
\]

\[
= \sqrt{n} \left( \nu(P_0 + n^{-1/2}\mathbb{F}_n) - \nu(P_0) \right) - \nu_0(\mathbb{F}_0 + x_0)
\]

\[
- \{\sqrt{n} \left( \nu(P_0 + n^{-1/2}x_n) - \nu(P_0) \right) - \nu_0(x_0)\}
\]

and hence

\[
\left\| \sqrt{n} \left( \nu(P_n) - \nu(P_0) \right) - \nu_0(\mathbb{F}_0) \right\|^*_{B}
\]

\[
\leq \left\| \sqrt{n} \left( \nu(P_0 + n^{-1/2}\mathbb{F}_n) - \nu(P_0) \right) - \nu_0(\mathbb{F}_0 + x_0) \right\|^*_{B}
\]

\[
+ \left\| \sqrt{n} \left( \nu(P_0 + n^{-1/2}x_n) - \nu(P_0) \right) - \nu_0(x_0) \right\|_{B}
\]

\[\rightarrow_{a.s.} 0 + 0 = 0\]

by (2.5), (2.7), and definition 2.1 just as in the proof of theorem 3.1. \( \square \)

Now for the bootstrap parallel of theorem 3.4; it both generalizes (and is inspired by) theorem 5 of Gill (1988).

**Theorem 3.5.** (Weak consistency of the bootstrap).

Suppose that:

(i) \( \nu: \mathbb{F} \rightarrow \mathbb{B} \) is \( C(F, P_0) \) – tangentially differentiable at \( P_0 \) with derivative \( \nu_0: C(F, P_0) \rightarrow \mathbb{B} \).

(ii) The hypotheses (and hence also the conclusions) of theorem 1.7 (Giné - Zinn) hold; i.e. \( F \) is a \( P_0 \) – Donsker class with \( P_0(F^2) < \infty \).

Then, with

\[
Y_n^\# \equiv \sqrt{n} \left( \nu(P_n^\#) - \nu(P_n) \right), \quad n = 1, 2, \cdots
\]

and

\[
Y_0^\# \equiv \nu_0(\mathbb{F}_0^\#) \equiv \nu(\mathbb{F}_0) \equiv Y_0,
\]

for any \( h \in C_b(B) \),

\[
E^* \{ h(Y_n^\# | X_1, \cdots, X_n \} \rightarrow_{p} E h(Y_0^\#) = E h(Y_0)
\]

as \( n \rightarrow \infty \); i.e. the bootstrap is weakly consistent.

**Proof.** This is like the proof of theorem 3.4, but with corollary 2.3 replaced by corollary 2.5. The key observation is now that both (2.9) and (2.11) hold. By (ii), the hypotheses and conclusions of corollary 2.5 hold. Let \( \tilde{C} \) and \( \tilde{C}^\# \) denote the sets with \( \tilde{Q}(\tilde{C}) = 1 \) and \( \tilde{Q}^\#(\tilde{C}^\#) = 1 \) defined there, and consider fixed \( \tilde{\omega} \in \tilde{C} \), \( \tilde{\omega}^\# \in \tilde{C}^\# \). Then
(a) \[ \sqrt{n} \left( \nu(A_n^\#(\tilde{\omega}^\#)) - \nu(A_n(\tilde{\omega})) \right) - \nu_0(A_0^\#(\tilde{\omega}^\#)) \]
\[ = \sqrt{n} \left( \nu(P_0 + n^{-1/2} A_{n,0}^\#(\tilde{\omega}^\#)) - \nu(P_0) \right) \]
\[ - \nu_0(A_0^\#(\tilde{\omega}^\#) + A_0(\tilde{\omega})) \]
\[ - \{ \sqrt{n} \left( \nu(P_0 + n^{-1/2} A_{n}(\tilde{\omega})) - \nu(P_0) \right) - \nu(A_0(\tilde{\omega})) \} \]
so that

(b) \[ \| \sqrt{n} \left( \nu(A_n^\#(\tilde{\omega}^\#)) - \nu(A_n(\tilde{\omega})) \right) - \nu_0(A_0^\#(\tilde{\omega}^\#)) \|_B \]
\[ \leq \| \sqrt{n} \left( \nu(P_0 + n^{-1/2} A_{n,0}^\#(\tilde{\omega}^\#)) - \nu(P_0) \right) - \]
\[ \nu_0(A_0^\#(\tilde{\omega}^\#) + A_0(\tilde{\omega})) \|_B \]
\[ + \| \sqrt{n} \left( \nu(P_0 + n^{-1/2} A_n(\tilde{\omega})) - \nu(P_0) \right) - \nu(A_0(\tilde{\omega})) \|_B \]
\[ \rightarrow 0 + 0 = 0 \quad \text{as} \quad n \rightarrow \infty \]
by (2.12), (2.9), and (i). Thus, as in the proof of theorem 3.1, with \( A_n \equiv X_n \circ \phi_n \), \( A_n^\# \equiv A_{n,0}^\#, \) for any fixed \( \tilde{\omega} \in \bar{C} \),

(c) \[ \| A_n^\# - A_0^\# \|_B \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \quad \text{a.s.} \quad \bar{Q}^\#. \]

By Dudley (1985) corollary 3.4 and (2.1), this yields, for any \( h \in C_b(B) \) and the same fixed \( \tilde{\omega} \in \bar{C} \),

(d) \[ E^* \left\{ h(T_n^\#|X_1, \cdots, X_n \right\}(\phi_n(\tilde{\omega})) \rightarrow E \left\{ h(T_0^\# \right\} \quad \text{as} \quad n \rightarrow \infty; \]
i.e., a.s. \( \bar{Q} \)

(e) \[ E^* \left\{ h(T_n^\#|X_1, \cdots, X_n \right\}(\phi_n) \rightarrow E \left\{ h(T_0^\# \right\}. \]

But (e) implies that (14) holds by Dudley (1985) theorem 3.5. \( \square \)

We could go on to give a theorem yielding regularity of the bootstrap estimator of the distribution of \( \sqrt{n} \left( \nu(A_n) - \nu(P) \right) \) using theorem 1.6, but we forego this extension here.

Representing the derivative; covariance formulas

The map \( \nu_0 \) induces a Gaussian distribution on the Banach space \( B \). To calculate covariances of this distribution, the following proposition -- an extension of the classical Riesz representation theorem -- is useful.

Proposition 3.1. Suppose that \( F \) is totally bounded in the \( \tau_{P_0} \) - pseudo-metric, and \( \nu_0: C(F, P_0) ightarrow B \) is continuous and linear. Then there exists a weak* - countably additive, \( B^{**} \) - valued measure \( \mu \) defined on the Borel sets of \( \bar{F} \), the (compact) completion of \( F \) under \( \tau_{P_0} \), such that:
(i) $\mu(\cdot)b^*$ is a regular, countably additive Borel measure for each $b^* \in B^*$; 
(ii) the map $b^* \rightarrow \mu(\cdot)b^*$ of $B^*$ into $C(\overline{F}, P_0)^*$ is weak* - to weak* continuous; 
(iii) $b^* \dot{v}_0(x) = \int_F x d(b^* \mu)$ for each $b^* \in B^*$ and $x \in C(\overline{F}, P_0)$; 
(iv) $\|\dot{v}_0\| = \|\mu\|_1(\overline{F})$; 
(v) if $x \in C(\overline{F}, P_0)$ is linear, then 

$$b^* \dot{v}_0(x) = x(\dot{v}_{b^*})$$

where

$$\dot{v}_{b^*} \equiv \int_F f d(b^* \mu)(f) \in L_2(P_0).$$

**Proof.** Since $F$ is totally bounded, its completion $\overline{F}$ in $L_2(P_0)$ with respect to $\tau_{P_0}$ is compact. Moreover, any $x \in C(F, P_0)$ has a unique extension to a continuous function $\overline{x}$ on $\overline{F}$, and the correspondence $x \rightarrow \overline{x}$ is a norm isomorphism between $C(F, P_0)$ and $C(\overline{F}, P_0)$. Since $\overline{F}$ is compact and Hausdorff (here we identify $L_2(P_0)$ equivalence classes), (i) - (iv) follow from Diestel and Uhl (1977), theorem 1, page 152.

Part (v) follows immediately from (iii) and linearity of $x$. \(\square\)

In general, the representing measure $\mu$ of $\dot{v}_0$ is not norm - countably additive nor $B$ - valued. But according to the Bartle - Dunford - Schwartz theorem, $\dot{v}_0$ is weakly compact if and only if its representing measure $\mu$ is norm - countably additive if and only if $\mu$ takes all of its values in $B \subset B^{**}$; see Diestel and Uhl (1977), page 153.

**Corollary 3.1.** If $F$ is a $G_P$ BUC class and $\dot{v}_0 : C(F, P_0) \rightarrow B$ is bounded and linear, then, with $\overline{X}_0$ a $P_0$ - Brownian bridge, for $b^*, b_1^*, b_2^* \in B^*$,

$$b^* \dot{v}_0(\overline{X}_0) = \overline{X}_0(\dot{v}_{b^*})$$

where $\dot{v}_{b^*}$ is given by (16), and

$$\text{Cov}[b_1^* \dot{v}_0(\overline{X}_0), b_2^* \dot{v}_0(\overline{X}_0)] = \int \dot{v}_{b_1^*} \dot{v}_{b_2^*} dP_0 - \int \dot{v}_{b_2^*} dP_0 \int \dot{v}_{b_2^*} dP_0.$$

**Proof.** This follows from proposition 3.1 and linearity of $\overline{X}_0$. \(\square\)

Before leaving this subsection, we note that $C(F, P_0)$ - tangential differentiability of $\nu : P \rightarrow B$ in the sense of definition 3.1 implies pathwise norm - differentiability often used in efficiency studies; see e.g. van der Vaart (1988) or Bickel, Klaassen, Ritov, and Wellner (1989) chapter 5.
Remarks and discussion

1. Since $H$-convergence often implies $\rho_F$-convergence of $P_n$ to $P_0$, more continuity is required of the differentiability hypothesized for convergence of the bootstrap in theorem 3.3 than is required for the regularity theorem 3.2. It would be interesting to have examples of functions $\nu$ for which Hellinger-regularity is true, but the convergence of the bootstrap fails.

2. Lohse (1987) also uses a (different) weakening of compact differentiability and properties of the empirical df of $R^k$ valued random variables (the classical case of $A = R^k$ and $F = \{\text{indicators of lower left orthants}\}$) to establish validity of the bootstrap in the sense of weak consistency for functions $\nu = (\text{Lohse's } T)$ taking values in $R^p$. Our theorems 3.3 and 3.7 strengthen his in several respects: $R^p$ is replaced by a general Banach space $B$ or $L_\alpha(T)$, df's $F$ of $R^k$ random variables are replaced by general $P$ on a measurable space $(A,A')$, and measurability conditions are removed. It appears that Lohse (1987) relies on hypothesized behavior of the bootstrap empirical process, and does not prove the important hypothesis A.3 in his theorem 3.3, page 358 except in special cases.

3. van der Vaart (1988) gives results concerning measurability of derivative maps in cases of interest; see his lemmas 4.3 and 4.4 page 93. He also proves efficiency of Hadamard differentiable functions of efficient estimators, which entails a local regularity (or $n^{-1/2}$-Hellinger regularity) of convergence weaker than our Hellinger regularity (recall remark 1.4) via contiguity methods; see his theorems 4.10 and 4.11 page 101 and 103. Our theorem 3.4 is comparable to his theorem 4.10 and 4.11, but with measurability hypotheses removed.

4. While revision #2 of this manuscript was underway, we received a copy of Pons and Turuckheim (1988). They state a special case of our corollary 1.4 and a delta method result in the spirit of our theorems 2.2 and 2.3. They apply their results to develop a bootstrap test of independence for bivariate censored data.

5. Our definition 3.1 does not require that $\dot{\nu}_0$ exist as a continuous map from $L_\alpha(F)$ to $B$, and in this respect we differ slightly from Gill (1988). When $B = R$, a continuous extension always exists (by the Hahn-Banach theorem), but for general $B$ a continuous extension may not exist. In fact, in all our examples $\nu_0$ is well-defined and continuous on $L_\alpha(F)$. Then, assuming $IP_n \in P$ a.s. and taking $\varepsilon_n = n^{-1/2}$, it follows from (3) and corollary 2.1 that

$$\|\sqrt{n} (\nu(\widetilde{P}_n) - \nu(P_0)) - \dot{\nu}_0(\widetilde{X}_n)\|_B^* \rightarrow a.s. 0$$

where, by linearity of $\dot{\nu}_0$,

$$\dot{\nu}_0(\widetilde{X}_n) = n^{-1/2} \sum_{i=1}^n \dot{\nu}(X_i)$$

with $\dot{\nu} : A \rightarrow B$ defined by $\dot{\nu}(X) \equiv \dot{\nu}_0(\delta_X - P)$. Of course (19) is an "asymptotic
linearity" form of the delta method which leads to consideration of the empirical process \( \mathcal{X}_n \) indexed by \( F = \{ b^* \mathcal{V} : b^* \in C^* \subset B^* \} \). Asymptotic linearity statements of the form (19) have been used by many authors, including Bahadur and Kiefer in the case of quantiles and the quantile process in our examples 3 and 3.A, to reduce a complicated nonlinear process involving \( \mathcal{V} \) to a simpler linear one involving the empirical process \( \mathcal{X}_n \). We have established theorems parallel to the present theorems 3.1 - 3.5 with the differentiability hypotheses in the form of definition 3.1 replaced by asymptotic linearity hypotheses in the form of (19) (in probability on the original probability space) where the functions \( \mathcal{V} \) may depend on \( P_0 \) or \( P_n \). These theorems also seem to be useful and important, but we will not present them here.
4. Examples -- the delta method in action

Now we give several examples to illustrate the theorems of section 3.

The theorems of section 3 yield many known (first order) asymptotic results as special cases, and perhaps yield some clarification and unification of earlier bootstrap results. For example, see our treatment and discussion of the bootstrap for sample quantiles in example 2.

A notable class of exceptions are the results of Csörgő, Csörgő, and Horváth (1986) and Csörgő and Mason (1988) concerning asymptotic theory and weak consistency of the bootstrap for mean residual life, total time on test, and the Lorenz curve. These authors use the very finely tuned weak approximation methods of Csörgő, Csörgő, Horváth, and Mason (1986) to prove some highly refined limit theorems, and the bootstrap analogues thereof, for these natural estimators. Their methods are geared toward and apparently limited to functions $\nu$ of the one-dimensional empirical df $IP_n(x) \equiv IP_n I_{(x \leq x)}$ of $X_i$ 's iid in $R$. Our theorems in section 3 can be extended to include their results by introducing a generalization of definition 3.1 involving a weight function $\psi : L^2_\infty (F) \to B$ as follows: Suppose that a multiplication is defined in $B$ so that $b_1 \cdot b_2 \in B$ is well-defined. (For $B = L^\infty (T)$, as we have in mind, this is true.) Say that $\nu$ is $\psi$-weightedly differentiable tangentially to $C(F, P_0)$ at $P_0$ relative to the collection $P(P_0, d)$ if

$$\frac{\nu(Q_n) - \nu(P_n)}{\varepsilon_n} \psi(Q_n) \to \nu_0 (D_0) \psi(P_0) \quad in \ B$$

$n \to \infty$ whenever $D_n = (Q_n - P_n)/\varepsilon_n \to D_0$ in $L^\infty_\infty (F)$ with $D_0 \in C(F, P_0)$, $\{Q_n\} \in P$, and $\{P_n\} \in P(P_0, d)$. (This type of differentiability is apparently related to Serfling's (1980), page 221, notion of a "quasi-differential".) We will not pursue this extension here, but simply note that even our rather crude methods suffice to give results that are stronger in a different direction than those of these authors; by establishing the existence of a derivative of a function $\nu$ such as mean residual life or a hazard function, we have established an analytic fact that can be used in different probabilistic settings. All that is really needed is the convergence of the appropriate process for estimation of $P$.

Example 1. (The mean squared). Suppose that $(A, \mathcal{A}) = (R, \mathcal{B})$, $B = R$, and

$$\nu(P) = (\int x \, dP(x))^2 \equiv \mu^2(P),$$

the square of the mean $\mu$ of $P$. We take $F = \{f_0(x) = x \equiv$ the identity function on $R \}$, and suppose that

$$P = \{P \text{ on } R : \int x^2 \, dP(x) < \infty\}.$$ 

Then $\nu$ is trivially $p_F$-continuously, $C(F, P_0)$--tangentially differentiable at any
$P_0 \in \mathcal{P}$ with derivative $\dot{v}_0$ given by

$$\dot{v}_0(D_0) = 2\mu(P_0)D_0(f_0) = 2\mu(P_0)\int x dD_0(x).$$

Thus all of theorems 3.1 - 3.5 apply. Of course

$$\dot{v}_0(\mathcal{X}_0) = 2\mu(P_0)\mathcal{X}_0(f_0) \equiv 2\mu(P_0)N(0, \sigma^2) = N(0, 4\mu^2(P_0)\sigma^2)$$

where $\sigma^2 \equiv Var(X)$. 

**Example 1A.** (Smooth functions of moments). Suppose that $(\mathcal{A}, \mathcal{M})$ is general, $B = \mathbb{R}^m$, $F = \{f_1, \cdots, f_k\}$, and $g : \mathbb{R}^k \to \mathbb{R}^m$ is a $C^1$ function (i.e. each component of $g$ is continuously differentiable or $C^1$), at least in a neighborhood of $P(f) = (P(f_1), \cdots, P(f_k))$. Write $\nabla g$ for the $m \times k$ matrix of derivatives. Consider estimation of

$$v(P) = g(P(f_1), \cdots, P(f_k)) = g(P(\_\_)).$$

Then $F$ has envelope function $F = \max_{1 \leq j \leq k} |f_j|$, and we take

$$P_0 = \{P \in \mathcal{M} : P(F^2) < \infty \text{ and } g \in C^1 \text{ at } P(\_\_)\}.$$

Then, by essentially the same argument as in example 1, $v$ is $\rho_F$-continuously, $C(F, P_0)$-tangentially differentiable at any $P_0 \in P_0$ with derivative

$$\dot{v}_0(D_0) = \nabla g(P_0(\_\_)) \cdot D_0(\_\_)$$

where $\nabla g$ is $m \times k$

$$= D_0(\nabla g(P_0(\_\_)) \cdot \_\_).$$

Thus all of theorems 3.1 - 3.5 apply. Of course

$$\dot{v}_0(\mathcal{X}_0) = \mathcal{X}_0(\nabla g(P_0(\_\_)) \cdot \_\_)$$

$$\equiv \nabla g(P_0(\_\_)) \cdot N_k(0, \Sigma_0) \equiv N_m(0, \nabla g \Sigma_0(\nabla g)^T)$$

where

$$\Sigma_0 = P_0(\_\_^T) - P_0(\_\_)P_0(\_\_)^T$$

is the covariance matrix of $\mathcal{X}_0(\_\_)$. Further subcases of this example include many classical statistics, including variances, covariances, correlation, coefficient of variation, and higher central and ordinary moments.

**Example 2.** (Length biased sampling). Let $(\mathcal{A}, \mathcal{M}) = (\mathbb{R}^+, \mathcal{B}^+)$, and suppose that $X_1, \cdots, X_n$ are iid $P$ on $R^+ \equiv [0, \infty)$ where the df corresponding to $P$ is given by

(a) $P(x) = \frac{1}{\mu(Q)} \int_{[0,x]} y \, dQ(y) = \frac{\int_{[0,x]} y \, dQ(y)}{\int_{[0,\infty)} y \, dQ(y)}$;

here $Q$ is a measure (or corresponding df) on $R^+$ with
\[(b) \quad 0 < \mu(Q) \equiv \int_0^\infty y \, dQ(y) < \infty.\]

\(P\) is the length-biased distribution corresponding to \(Q\). Note that we can easily express the df \(Q \equiv v\) in terms of \(P\) by

\[(c) \quad v(P)(x) = Q(x; P) = \frac{\int_{[0,x]} y^{-1} \, dP(y)}{\int_{[0,\infty]} y^{-1} \, dP(y)}, \quad x \geq 0.\]

Here we are interested in estimating \(v(P) \equiv Q(\cdot, P)\) based on observations from \(P\).
In view of (c), a natural estimator of \(Q\) based on a sample from \(P\) is

\[(d) \quad v(P_n)(x) \equiv Q_n(x) = \frac{\int_{[0,x]} y^{-1} \, dP_n(y)}{\int_{[0,\infty]} y^{-1} \, dP_n(y)} = \hat{\mu}_n \int_{[0,x]} y^{-1} \, dP_n(y)\]

with

\[\hat{\mu}_n^{-1} \equiv \int_{[0,\infty]} y^{-1} \, dP_n(y) = \frac{1}{n} \sum_{i=1}^n X_i^{-1};\]

note that \(P(0) = 0\) even if \(Q(\{0\}) > 0\), so all the \(X_i\)'s are strictly positive with probability one.

Here we take \(B = L_\infty(R^+)\) with the uniform norm, and let

\[P \equiv \{P\ \text{of the form (a) where } Q \text{ satisfies (b) and } \int_0^\infty y^{-2} \, dP(y) = \int_0^\infty y^{-1} \, dQ(y) < \infty \}.\]

Let \(F\) be the collection

\[F \equiv \{y \to \frac{1}{y} 1_{[0,x]}(y) : 0 \leq x < \infty \},\]

with envelope function \(F(y) = 1/y\). Note that \(F\) is a \(P\)-uniform Donsker class for any collection \(P\) for which \(F\) is uniformly square integrable over the class, since it is of the form fixed square integrable times indicators of a VC collection; see e.g. Pollard (1982), theorem 9.

Then it is easily verified that \(v\) is \(P\)-continuously, \(C(F, P_0)\)-tangentially differentiable at \(P_0 \in P\) with derivative \(v_0 : C(F, P_0) \to B = L_\infty(R^+)\) given by

\[(e) \quad v_0(D_0) = D_0(v_x)\]

where

\[v_x(y) = \mu(Q_0) y^{-1} \{1_{[0,x]}(y) - v(P_0)\}\]

\[= \mu(Q_0) y^{-1} \{1_{[0,x]}(y) - Q_0(x)\}.\]

Thus all the theorems 3.1 - 3.5 apply to this example. In particular, (3.6) holds where, by (e),

\[\dot{v}_0(\mathbb{X}_0)(x) = \mathbb{X}_0(\dot{v}_x) \equiv v'_0(x) \quad \text{for } x \in R^+\]
has covariance

\[
\text{Cov}[\mathcal{Y}_0(x), \mathcal{Y}_0(y)] = \mu^2(\mathcal{Q}_0) \int_0^\infty z^{-2} (1_{[0,x]}(z) - \mathcal{Q}_0(x))(1_{[0,y]}(z) - \mathcal{Q}_0(y)) dP_0(z) = \mu(\mathcal{Q}_0) \int_0^\infty z^{-1} (1_{[0,x]}(z) - \mathcal{Q}_0(x))(1_{[0,y]}(z) - \mathcal{Q}_0(y)) d \mathcal{Q}_0(z).
\]

The above treatment of this special case of the general biased sampling model of Vardi (1983), (1985) can easily be extended, using arguments similar to those of Gill, Vardi, and Wellner (1988), to show Vardi’s (1985) nonparametric maximum likelihood estimates are Hellinger - regular and that the bootstrap is a.s. consistent.

**Example 3.** (The $t$ - th quantile). Suppose that $(\mathcal{A}, \mathcal{G}) = (R, \mathcal{B})$, $\mathcal{B} = R$, $\mathcal{F} = \{ 1_{(-\infty, x)} : x \in R \}$. In this, and the following one-dimensional examples in this section, we let $P = \mathcal{M}$ and write

\[
P(x) = P(X \leq x) = P 1_{[x, \infty)}, \quad x \in R
\]

for the df associated with any $P$, and

\[
P^{-1}(u) = \inf \{ x : P(x) \geq u \}, \quad 0 \leq u \leq 1,
\]

for the quantile function. Let $p(x) = P'(x)$ (if it exists) be the density of $P$ at $x$. For fixed $0 < t < 1$, consider estimation of

(a) $v(P) \equiv P^{-1}(t)$, \quad for $P \in \mathcal{P}_0$

where

(b) $\mathcal{P}_0 = \{ P \in \mathcal{P} : p = P' \text{ exists at } v(P) \text{ and } p(P^{-1}(t)) > 0 \}$.

Then, as shown in Gill (1988), $v$ is $C(\mathcal{F}, \mathcal{P}_0)$ - tangentially differentiable at each $P_0 \in \mathcal{P}_0$ with

(c) $v_0(D_0) = -\frac{D_0(1_{(-\infty, v(P_0))})}{P_0(v(P_0))}$.

Since $\mathcal{F}$ is a $\mathcal{M}$ - uniform Donsker class, hypotheses (i) and (ii) of theorems 3.1, 3.4, and 3.5 are satisfied, and $v(\mathcal{IP}_n) = \mathcal{IP}_n^{-1}(t)$ is an $n^{-1/2}$ – Hellinger - regular estimator of the $t$ – th quantile $v(P_0)$ at $P_0 \in \mathcal{P}_0$, and the bootstrap is weakly consistent by theorem 3.5. Weak consistency of the bootstrap for the $t$ – th quantile has also been established by Beran (1984) and Gill (1988). Of course it follows easily from (c) that

(d) $v_0(\mathcal{X}_0) = -\frac{1}{p_0(P_0^{-1}(t))} \mathcal{X}_0(1_{(-\infty, v(P_0))}) \equiv N(0, -\frac{t(1-t)}{p_0^2(P_0^{-1}(t))})$.

Bickel and Freedman (1981) proved strong consistency of the bootstrap for the $t$ – th quantile (assuming only slightly more, namely existence and continuity of the derivative $p$ in a neighborhood of $v(P)$), so ordinary tangential differentiability at a
point and theorem 3.5 do not yield the Bickel - Freedman result. Moreover, it is easily seen that \( v(P) \) is not \( \rho_F \) - continuously, \( C(F, P_0) \) - tangentially differentiable at any \( P_0 \in P_0 \), so theorem 3.3 does not apply.

But, it is not hard to see that \( v(P) \) is \( \{ n^{-1/2} \} \) - differentiable tangentially to \( C(F, P_0) \) at \( P_0 \) relative to the collection of sequences \( P(P_0, \rho_F) \) satisfying \( osc_{C_n}(a_n) \to 0 \) whenever \( a_n \to 0 \) where \( C_n = \sqrt{n} (P_n - P_0) \) and, with \( v_0 = v(P_0) \),

\[
osc_{C_n}(a) = \sup_{x, y \in [v_0 - \varepsilon, v_0 + \varepsilon], |y - x| \leq a} |C_n(y) - C_n(x)|.
\]

Since \( P_0 \) is continuous in a neighborhood of \( v_0 \), so is the limit process \( X_0(x) = X_0(1[\cdot \leq x]) \), \( x \in R \), and it can therefore be shown that \( IP_n \) is in \( P(P_0, \rho_F) \) with probability 1. (This follows from the fact that the oscillation modulus \( \omega_n(a_n) \) of the simple uniform empirical process \( U_n \) converges to 0 a.s. whenever \( a_n \to 0 \).) In fact, this is essentially what Bickel and Freedman (1981) use in the course of the proof of their theorem 5.1. (Bickel and Freedman (1981), page 1207, refer to theorem 3 of Komlós, Major, and Tusnády (1975), when, it seems to us, they should refer to theorem 4 of that paper so that the construction has the correct joint in \( n \) distributions. Then the oscillation modulus of the process \( K(n, \cdot)^{1/n} \) does indeed converge to 0 a.s. as shown in theorem 14.3.1 of Shorack and Wellner (1986). In any case, the Bickel and Freedman proof is basically correct.) Hence our theorem 3.3.A does apply, and yields the same conclusion as proposition 5.1, Bickel and Freedman (1981) : the bootstrap is a.s. consistent for the \( t \) - th quantile under essentially the weakest possible conditions.

Bickel and Freedman (1981) describe their theorem 5.1 proof as being "ad hoc", and this is also appropriate in describing any application of our theorem 3.3.A because of the necessity of choosing the collection of sequences \( P(P_0, \rho_F) \). Thus this example clarifies theorems 3.3.A and 3.5, and shows the limitations of the present approach to asymptotic theory via differentiation theory as in section 3: here is the type of case in which proof of the a.s. consistency of the bootstrap via a differentiability argument essentially breaks down, and discovery of the right collection \( P(P_0, \rho_F) \) apparently requires us to "repeat the original argument" as advocated by Csörgő and Mason (1988). This is also in line with the comments by Bickel and Freedman (1981) page 1201. Nonetheless, theorem 3.5 demonstrates that \( C(F, P_0) \) - tangential differentiability at a point, as is true for our present \( v \), suffices for the weak consistency of the bootstrap: the differentiation theory has taken us a long way, even if it hasn't told the full story.

Singh (1981) uses the Bahadur representation of a sample quantile to show that the rate of this convergence is \( n^{-1/4} (loglog n)^{1/2} \) if, in addition, \( |p''((P^{-1}(r)))| < \infty \).

Example 3.A. (The quantile process). Now suppose the set-up of example 2 with \( F = \{ 1_{(-\infty, x]} : x \in R \} \), but suppose that
(a) \( P_0 = \{ P \in M : p = P' \text{ exists and is continuous and positive on } R \} \).

Let \( 0 < a < b < 1 \) and take \( B = L_\infty(T) \) with \( T = [a, b] \). Consider \( \nu : P \to B \) defined by

\[
\nu(P)(t) = P^{-1}(t), \quad a \leq t \leq b.
\]

Essentially in the same way as shown by Gill (1988) (he worked in \( D[a, b] \) rather than our \( L_\infty(T) = L_\infty([a, b]) \)), \( \nu \) is \( C(F, P_0) - \) tangentially differentiable at each \( P_0 \in P_0 \) with derivative \( \nu_0 \) given by

\[
\nu_0(D_0)(t) = \frac{D_0(1_{a, b}(\nu(P_0)(t)))}{P_0(P_0^{-1}(t))}, \quad a \leq t \leq b.
\]

Since \( F \) is an \( M \)-uniform Donsker class, theorems 3.1, 3.4, and 3.5 apply and show that \( \nu(P_n) \) is a \( n^{-1/2} \)-Hellinger regular estimator of \( \nu(P_0) \) and that the bootstrap is weakly consistent. This is as in Gill (1988), proposition 1, and Csörgő and Mason (1988), example 2, and is strictly weaker than Bickel and Freedman (1981), theorem 5.1.

Just as in example 3, \( \nu(P) \) is not \( \rho_F \)-continuously, \( C(F, P_0) - \) tangentially differentiable, so theorem 3.3 does not apply. But, also as in example 2, and essentially as argued by Bickel and Freedman (1981), \( \nu(P) \) is \( \{ n^{-1/2} \} - \) differentiable tangential to \( C(F, P_0) \) relative to the collection \( P(P_0, \rho_F) \) of sequences \( \{ P_n \} \) satisfying \( osc_{C_n}(d_n) \to 0 \) whenever \( d_n \to 0 \) where \( C_n \equiv \sqrt{n}(P_n - P_0) \) and now

\[
osc_{C_n}(d) \equiv \sup_{x, y \in [\nu(a) - \epsilon, \nu(b) + \epsilon], |x - y| \leq d} |C_n(y) - C_n(x)|.
\]

Thus by theorem 3.3.A the bootstrap is a.s. consistent for estimation of the distribution of \( \sqrt{n} (\nu(P_n) - \nu(P)) \).

For extensions of the basic limit theorem to (0,1) via strong approximation methods, see Csörgő and Révész (1978) (or SW, chapter 18); the asymptotic linearity form of our result under the assumption of \( \|\rho'(P^{-1})\|_a < \infty \) is the famous Bahadur-Kiefer theorem (see e.g. Shorack and Wellner (1986), page 586).

Example 4. (Mean residual life). Suppose again that \( (A, \sigma) = (R^+, E^+) \), \( B = L_\infty(T) \) with \( T = [0, M] \) for some \( 0 < M < \infty \), and let

(a) \( P = \{ P \text{ on } R^+: \int x^2 dP(x) < \infty, \overline{P}(M) \equiv P(X > M) > 0 \} \).

Consider estimation of \( \nu(P) \) defined by

(b) \( \nu(P)(x) \equiv E(X - x | X > x) = \frac{\int_{(x, \infty)} (y - x) dP(y)}{\overline{P}(x)} \), \( x \in T \).

For this example we let

(c) \( F \equiv \{ y \to 1_{(x, \infty)}(y), y \to (y - x)^+ : 0 \leq x \leq M \} \).
with envelope function $F(y) = y$. Note that $F \subset F_1 - F_2$ where $F_1 \equiv \{ y \ 1_{(x,\infty)}(y) : 0 \leq x \leq M \}$ and $F_2 \equiv \{ x \ 1_{(x,\infty)}(y) : 0 \leq x \leq M \}$ are both sparse classes by Pollard (1982) theorem 9 and theorem 10(iv), respectively. Hence $F \subset F_1 - F_2$ is sparse by Pollard (1982) theorem 10(i).

Then $v$ is $\rho_F$-continuously, $C(F, P_0)$-tangentially differentiable at any $P_0 \in P$ with derivative $\nu_0$ given by

(d) \[ \nu_0(D_0)(x) \equiv D_0(\nu_x), \quad x \in T \]

where

(e) \[ \nu_x(y) \equiv \frac{1}{P_0(x)} \{ (y-x)^+ - \nu(P_0)(x) \ 1_{(x,\infty)}(y) \}. \]

Hence all of theorems 3.1 to 3.5 apply to this example with $B$ chosen as above. In particular, (3.6) holds where, by (d),

\[ \nu_0(\mathbf{x}_0)(x) = \mathbf{x}_0(\nu_x) \equiv \mathbf{X}_0(x) \quad \text{for} \quad x \in T. \]

As noted by Hall and Wellner (1979), the limit process has covariance function

(g) \[ Cov[\nu_x(y), \nu_y(x)] = \frac{\sigma^2(0)}{P_0(x) P_0(y)} \tilde{K}(x \vee y) \]

where $\tilde{K}(x) \equiv \tilde{F}_0(x) \sigma^2(x) / \sigma^2(0)$ with $\sigma^2(x) \equiv Var(X - x | X > x)$ is a survival function on $R^+$ (i.e. $K \equiv 1 - \tilde{K}$ is a df), and hence $\nu_0 \equiv \frac{\sigma}{P_0} B_0(\tilde{K})$ where $B_0$ is standard brownian motion on $[0, \infty)$; see Shorack and Wellner (1986), pp. 775 ff., or Csörgő, Csörgő, and Horváth (1986).

Since the supremum in (f) is over $0 \leq x \leq M$, this is not as fine a result as that obtained by Csörgő, Csörgő, and Horváth (1986), or by Csörgő and Mason (1988) for the bootstrapped version using the construction developed by Csörgő, Csörgő, Horváth, and Mason (1986). It does, however give the basic behavior on $[0, M]$ with $\tilde{F}(M) > 0$; it does unify the various regularity and bootstrap versions of the theorem via compact differentiation theory; and the differentiation result can now be applied repeatedly in the sense above, for estimators of $P$ derived from sampling schemes other than iid.

Note that the finest versions of the Csörgő, Csörgő, and Horváth (1986) and Csörgő and Mason (1988) results (e.g. CCH (1986), theorem 4.1, page 39, or CM (1988), example 4) involve a sort of weighted differentiation theory as mentioned at the beginning of this section with $\psi(P) \equiv P(1_{(\cdot, \omega)})$.

Example 4.A. (Bivariate mean residual life). Suppose that $(A, A) = (R^{+2}, B_0^{+2})$, let $\tilde{P}(x) = P(X > x) \equiv P(X_1 > x_1, X_2 > x_2)$, and let

(a) \[ P \equiv \{ P \text{ on } R^{+2} : \int |x|^2 \ dP(x) < \infty, \ \tilde{F}(M) > 0 \} \]
where $M = (M, M)$ for some $0 < M < \infty$. Consider estimation of

(b) $\nu(P)(\bar{x}) = E(X - \bar{x} | X > \bar{x}) = \frac{\iint_{(0, \infty)} (y - \bar{x})^+ dP(y)}{P(\bar{x})}$

$= \gamma(P)(\bar{x}) - \bar{x}$

where

(c) $\gamma(P)(\bar{x}) = \frac{EX1_{(\bar{x}, \infty)}(X)}{P(\bar{x})}$

and where $\bar{x}^+ = (x_1, x_2)1_{(0, \infty)}(x_1)1_{(0, \infty)}(x_2)$. Here $\nu$ takes values in $B = L_\infty(T) \times L_\infty(T)$ where $T = [\Omega, M] \subset R^{+2}$, and we can take the class $F$ to be

(d) $F = \{ 1_{(\bar{x}, \infty)}(y), (y_1 - x_1)1_{(\bar{x}, \infty)}(y), (y_2 - x_2)1_{(\bar{x}, \infty)}(y) : \bar{x} \in T \}$.

Note that $F$ has envelope function $F(y) = (y_1 + y_2) \nu 1$, and is a $P_0$-Donsker class. Then $\nu$ is $p_0$-continuously, $C(F, P_0)$-tangentially differentiable at any $P_0 \in P$ with derivative $\nu_0$ given by

$\nu_0(D_0) = D_0(\nu_\bar{x}), \quad \bar{x} \in T$

where

$\nu_\bar{x}(y) = \frac{1}{P_0(\bar{x})} \left\{ (y - \bar{x})^+ - \nu(P_0)(\bar{x}) 1_{(\bar{x}, \infty)}(y) \right\}$.

$= \frac{1}{P_0(\bar{x})} (y - \gamma(P_0)(\bar{x})) 1_{(\bar{x}, \infty)}(y)$.

Thus all of theorems 3.1 - 3.5 apply; in particular (3.6) holds with

(e) $Y_0(\bar{x}) = \nu_0(\bar{x} 0) = \bar{x} 0(\nu_\bar{x})$

which has covariance

(f) $\text{Cov} \{ Y_0(\bar{x}), Y_0(y) \}$

$= \frac{1}{P_0(\bar{x})} \frac{1}{P_0(y)} E (X - \gamma_0(\bar{x}))(X - \gamma_0(y)) 1_{(\bar{x} \nu y, \infty)}$

where $\gamma_0 = \gamma(P_0)$.

Note that the univariate (marginal) mean residual life functions for both $X_1$ and $X_2$ are contained in $\nu$ since

$\nu(P)(x_1, 0) = (E(X_1 - x_1 | X_1 > x_1), E X_2)$

and similarly for $\nu(P)(0, x_2)$. By (f) the covariance between their estimators (at $x_1$ and $x_2$ respectively) is given by
\[
\frac{1}{\overline{F}_0(x_1,0)\overline{F}_0(0,x_2)} E(X_1 - \gamma_0(x_1,0))(X_2 - \gamma_0(0,x_2)) 1_{\langle x_1,x_2,\infty \rangle} \\
= \frac{\overline{F}_0(x_1,x_2)}{\overline{F}_0(x_1,0)\overline{F}_0(0,x_2)} \text{Cov}_0[X_1,X_2|X > x].
\]

Example 5. (Percentile residual life). Suppose that \((A,A) = (R^+,B^+ )\) as in example 5, but now let \(T = [a,b] \times [0,M]\) where \(0 < a < b < 1\) and \(0 < M < \infty\). Consider estimation of \(\nu : P \rightarrow L_\nu(T)\) defined by

\[
\nu(P)(t,x) \equiv P^{-1}(1 - (1-t)(1-P(x))) - x
\]

where we now take \(P\) and \(F\) as in example 3.A. This function \(\nu\) is the \(t\)-th quantile analogue of the mean residual life function in example 5; e.g. for \(t = 1/2\) it is the median residual life function. We consider it here as yielding a function of both \(t\) and \(x\); note that for \(x = 0\) and \(t\) fixed it reduces to example 3, and to example 3.A when considered as a function of \(t\).

As in examples 3 and 3.A, \(\nu\) is \(C(F,P_0)\) - tangentially differentiable at any \(P_0 \in P\) with derivative \(\nu_0\) given by

\[
\nu_0(D_0)(t,x) = (1-t)D_0(1_{(-\infty,x)} - D_0(1_{(-\infty,\nu_0(t,x)+x)}))
\]

\[
\text{p}_0(\nu_0(t,x)+x)
\]

where we write \(\nu_0 \equiv \nu(P_0)\). Since \(F\) is an \(M\)-uniform Donsker class, theorems 3.1, 3.4, and 3.5 hold. Just as in examples 3 and 3.A, although \(\nu\) is not \(\rho_F\) - continuously, \(C(F,P_0)\) - tangentially differentiable, it is \(\{n^{-1/2}\}\) - differentiable tangentially to \(C(F,P_0)\) at \(P_0 \in P\) relative to \(P(P_0,\rho_F)\) as defined in example 3.A (with a slight modification of the definition of \(osc_{C_\nu}\)), and hence theorem 3.3.A does apply: the bootstrap is a.s. consistent.

The limit process for estimation of \(\nu\) is

\[
\nu_0(\mathcal{X}_0)(t,x) = \frac{(1-t)\mathcal{X}_0(1_{(-\infty,x)}) - \mathcal{X}_0(1_{(-\infty,\nu_0(t,x)+x)})}{p_0(\nu_0(t,x)+x)}
\]

\[
\equiv \frac{(1-t)U(P_0(x)) - U(1-(1-t)(1-P_0(x)))}{p_0(\nu_0(t,x)+x)}
\]

\[
\equiv \frac{\mathcal{Z}(t,P_0(x))}{p_0(\nu_0(t,x)+x)}
\]

where \(\mathcal{U}\) is a \((\text{Uniform}(0,1)\) - )Brownian bridge process on \([0,1]\) and

\[
\mathcal{Z}(t,u) \equiv (1-t)U(u) - U(1-(1-t)(1-u))
\]

satisfies, for each fixed \(0 \leq u < 1\), \(\mathcal{Z}(\cdot,u)/\sqrt{1-u} \equiv N\), a Brownian bridge
process on \([0,1]\); see e.g. Shorack and Wellner (1986), exercise 2.2.12.

Note that our treatment here of the bootstrap for estimation of \(v\) (essentially using the Bickel-Freedman argument in combination with our theorem 3.3.A) seems to yield a stronger result than that obtained in Csörgő and Csörgő (1987) and example 7 of Csörgő and Mason (1988).

There is an obvious percentile residual life analogue of example 4.A., but we will not pursue it here.

**Example 6.** (Cumulative hazard function). Suppose that \((A, B) = (R^+, E^+)\), let \(P\) denote the df corresponding to \(P\), \(B = L_\infty(T)\) with \(T = [0, M]\), and
\[
P \equiv \{ P \text{ on } R^+: \bar{F}(M) = P(X > M) > 0 \}.
\]

Consider estimation of
\[
v(P)(x) = \int_{[0,x]} \frac{1}{1 - P(y^-)} dP(y) = \Lambda(x; P), \quad x \in T,
\]
the cumulative hazard function corresponding to \(P\). Here we take
\[
F \equiv \{ y \rightarrow 1_{[0,x]}(y): 0 \leq x < \infty \}.
\]
Then \(v\) is \(\rho_F\)-continuously, \(C(F, P_0)\)-tangentially differentiable at any \(P_0 \in P\) with derivative
\[
\dot{v}_0(D_0)(x) = \int_{[0,x]} \frac{1}{P_0(y^-)} d\{ D_0(y) - \int_{[0,y]} \bar{D}_0(z) d\Lambda_0(z) \}
\]
\[
= D_0(\dot{v}_x)
\]
where
\[
\dot{v}_x(y) = \frac{1_{[0,x]}(y)}{P_0(y^-)} - C(x \wedge y),
\]
\[
C(x) = \int_{[0,x]} \frac{1}{P_0(y^-)} d\Lambda_0(y),
\]
\[
D_0(y) = D_0(1_{[0,y]}), \quad \text{and}
\]
\[
\bar{D}_0(y) = D_0(1_{[y,\infty)}) = -D_0(y^-).
\]

Our motivation for writing \(\dot{v}_0\) in the form (d) comes from martingale theory; see e.g. Shorack and Wellner (1986), equation (6.1.4), page 265. Hence all of theorems 3.1 - 3.5 apply to this example with the above choice of \(B\). In particular, by theorem 3.1, (3.6) holds where, by (d),
\[
Y_0(x) = v_0(\bar{X}_0)(x)
\]
\[
= \int_{[0,x]} \frac{1}{P_0(y^-)} d\{ \bar{X}_0(y) - \int_{[0,y]} \bar{ar{X}}_0(z) d\Lambda_0(z) \}
\]
\[ \int_{[0,x]} \frac{1}{\bar{P}_0(y-)} \, dM_0(y) \, . \]

It is well-known that the process \( M_0 \) is a Gaussian martingale with predictable variation process
\[ \int_{[0,T]} (1 - \Delta A_0) \, dP_0, \]
and by martingale theory \( X_0 = B_0(C_\Delta) \) where \( B_0 \) is standard Brownian motion on \( \mathbb{R}^+ \) and
\[ C_\Delta(x) = \int_{[0,x]} \frac{(1 - \Delta A_0)}{\bar{P}_0(y-)} \, dA_0 ; \]
see e.g. Shorack and Wellner (1986), chapters 6 and 7. This example can be extended and refined by taking \( B = L_\infty(T) \) and \( \|b\|_B \) defined by
\[ \|b\|_B = \sup_{0 \leq x < \infty} |b(x)| \, \bar{P}_0(x) , \]
but we will not pursue this refinement here.

**Example 6.A.** (A bivariate hazard function). Suppose that \( (A,P) = (\mathbb{R}^2, \mathbb{B}^2) \), let \( \bar{P}(\bar{x}) \equiv P(X_1 > x_1, X_2 > x_2) \) and \( \bar{P}(-\bar{x}) \equiv P(X_1 \geq x_1, X_2 \geq x_2) \). We take \( B = L_\infty(T) \) with \( T = [0,M] \), and
\[ P = \{ P \text{ on } \mathbb{R}^2 : P(X \geq M) = \bar{P}(M-) > 0 \} . \]

Consider estimation of

(a) \[ \nu(P)(\bar{z}) = \int_{[0,\bar{z}]} \frac{1}{\bar{P}_0(y-)} \, dP(y) = \Lambda(\bar{z} ; P) \quad \text{for } x \in T ; \]

\( \nu \) is one possible generalization of the cumulative hazard function in example 8 to the bivariate situation. Here we take

(b) \[ F = \{ 1_{[0,\bar{z}]}(y) : 0 \leq \bar{z} < \infty \} . \]

Then \( \rho_F \) is continuously, \( C(F, P_0) \) - tangentially differentiable at any \( P_0 \in P \) with derivative \( \nu \) given by

(c) \[ \dot{\nu}_0(D_0)(\bar{z}) = \int_{[0,\bar{z}]} \frac{1}{\bar{P}_0(y-)} \, d\{ D_0(y) - \int_{[0,\bar{z}]} \bar{D}_0(\bar{z}) dA_0(\bar{z}) \} \]
\[ = D_0(\nu_\bar{z}) \]

where

(e) \[ \dot{\nu}_\bar{z}(y) = \frac{1_{[0,\bar{z}]}(y)}{\bar{P}_0(y-)} - C(\bar{z} \Lambda y) \]

(f) \[ C(\bar{z}) = \int_{[0,\bar{z}]} \frac{1}{\bar{P}_0(-)} \, dA_0 , \]
(g) \( D_0(y) = D_0(1_{[0,y]}) \), and \( \overline{D_0(y)} = D_0(1_{[y,\infty)}) = -D_0(y^-) \) almost exactly as in the one-dimensional case. Since \( F \) is a \( M \)-uniform Donsker class, all of theorems 3.1 - 3.5 apply, and we conclude that \( \nu(P_n) \) is a Hellinger regular estimator of \( \nu(P_0) \) and that the bootstrap is a.s. consistent at any \( P_0 \in \mathcal{P} \).

Here in two dimensions the martingale motivation is less compelling since the corresponding process \( M_0 \), and hence also \( \nu_0(X_0) \), is only a weak martingale; see e.g. Pons (1986). This example has been treated in more detail, with the additional complication of (bivariate) censoring, by Pons and Turckheim (1988) using methods similar to ours.

**Example 7.** (Cumulative hazard function with censored data). Suppose that \( A = R^+ \times \{0,1\} \) with the natural product sigma-field, and that \( P \) is the distribution of \( X = (Z, \delta) = (Y \Lambda C, 1_{[Y \leq C]}) \) where \( Y \equiv F \) and \( C \equiv G \) are independent. We set

(a) \[ H_u(z) = P(Z \leq z, \delta = 1) = \int_{[0,z]} (1 - G_-) \, dF \]
and

(b) \[ \overline{H}(z) = P(Z \geq z) = (1 - F_-)(1 - G_-)(z) \]

Let \( 0 < M < \infty \). We take

(c) \[ P = \{ P \text{ on } A : \overline{H}(M) > 0 \} \]

and consider estimation of \( \nu: P \to B \equiv L_{\omega}(T), T = [0,M], \) defined by

(d) \[ \nu(P)(x) = \int_{[0,x]} \frac{1}{H} \, dH_u = \int_{[0,x]} \frac{dF}{1 - F_-} \equiv \Lambda(x; F) \]

for \( x \in T \). In this case we take

(e) \[ F = \{ (z, \delta) \to \delta 1_{[0,x]}(z), 1_{[x, \infty)}(z) : x \geq 0 \} \]

This collection \( F \) has envelope function \( F(x) = 1 \) and is sparse; i.e. hypothesis (i) of theorem 1.3 holds.

Much as in the uncensored case, \( \nu \) is \( \rho_F \)-continuously, \( C(F, P_0) \)-tangentially differentially at any \( P_0 \in \mathcal{P} \) with derivative

(f) \[ \dot{\nu}_0(D_0(y)) = \int_{[0,y]} \frac{1}{H(z)} \, d\{ D_u(z) - \int_{[0,z]} \overline{D}(t) \, d\Lambda_0(t) \} \]

where

\[ \dot{\nu}_y(z, \delta) = \frac{\delta 1_{[0,y]}(z)}{H(z)} - C(y \Lambda z), \]
\[ C(y) \equiv \int_{[0,y]} \frac{1}{H_0} d\Lambda_0, \]
\[ D_u(y) \equiv D_0(\delta 1_{[0,y]}(z)), \]
and
\[ \overline{D}(y) \equiv D_0(1_{[y,\infty)}). \]

Hence all of theorems 3.1 - 3.5 apply; in particular (3.6) holds with
\[ \dot{\nu}_0(\mathbf{X}_0)(y) = \int_{[0,y]} \frac{1}{H(z)} d(\mathbf{X}_u(z) - \int_{[0,z]} \mathbf{X}(t) d\Lambda_0(t)) \]
\[ = \int_{[0,y]} \frac{1}{H(z)} dM_u(z). \]

It is well-known that the process \( M_u \) is a Gaussian martingale with predictable variation process
\[ \int_{[0,\cdot]} (1 - \Delta \Lambda_0) \, dH_u, \]
and by martingale theory \( Z_0 \equiv \dot{\nu}_0(\mathbf{X}_0) \equiv \mathcal{B}_0(C_\Delta) \) where
\[ C_\Delta(y) \equiv \int_{[0,y]} \frac{1}{H^2_0} (1 - \Delta \Lambda_0) \, dH_u, \]
\[ = \int_{[0,y]} \frac{1}{H_0} (1 - \Delta \Lambda_0) \, d\Lambda_0; \]

see e.g. Shorack and Wellner (1986), chapter 7, or Gill (1988). The observation that this function \( \nu \) is \( C(\mathbf{F},P_0) - \) tangentially differentiable is due to Gill (1988); the slight strengthening to \( \rho_F - \) continuously, \( C(\mathbf{F},P_0) - \) tangential differentiability seems to be new.

**Example 7.A.** (The product limit estimator). Here we continue with the set-up of example 9, but consider estimation of \( \gamma : P \to L_\infty(T) \) defined by
\[ \gamma(P)(x) = \prod_{y \leq x} \{ 1 - d \nu(P)(y) \} ^{\psi} \nu(P)(y) \quad \text{for} \ x \in T \]

where \( \nu \) is defined by 9(d). Of course the "product integral" \( \psi \) in (a) yields the survival function \( 1 - F \); see Gill and Johansen (1987) theorem 17 for this fact and an excellent exposition of product integrals and their properties. (See Shorack and Wellner (1986), pages 301 and 897 for related material.) It follows from theorem 14 of Gill and Johansen (1987) that \( \gamma \) is \( C(\mathbf{F},P_0) - \) tangentially differentiable at any \( P_0 \in P \) with derivative \( \gamma_0 \) given by
\[ \gamma_0(D_0)(y) = - \overline{F}(y) \int_{[0,y]} \frac{\overline{F}}{F} \frac{1}{H} d\{ D_u(z) - \int_{[0,z]} \overline{D}(t) \, d\Lambda_0(t) \} \]
where

\[ \gamma_y(z, \delta) = -\bar{F}(y) \left\{ \frac{\bar{F}_-(z)}{F(z)} \frac{\delta 1_{[0,1]}(z)}{H(z)} - \bar{C}(y \wedge z) \right\}, \]

with

\[ \bar{C}(y) = \int_{[0,1]} \frac{\bar{F}_-}{F} \frac{1}{H} d\Lambda. \]

Thus theorems 3.1, 3.4, and 3.5 apply to this example. In particular, (3.6), (3.11), and (3.14) hold (with \( \nu \) replaced by \( \gamma \)) and

\[ Y_0(y) = \gamma_0(\Xi_0)(y) = \bar{F}(y) \int_{[0,1]} \frac{\bar{F}_-}{F} \frac{1}{H} dM_u \]

with \( M_u \) as in example 9; again it is martingale theory which motivates us to write the limit process in this form. The limit process \( \bar{M} \bar{F} \) is a Gaussian martingale, and it follows easily from standard martingale calculations that

\[ \text{Cov}[Y_0(x), Y_0(y)] = \bar{F}(x) \bar{F}(y) \int_{[0,1]} \frac{1}{H(1 - \Delta \Lambda)} d\Lambda. \]

These results are not the strongest possible. The basic limit theorem of Breslow and Crowley (1974) (for fixed \( P_0 \) and \( B = L_{\infty}(\{0,1\}) \)) has been strengthened to the whole real line by Gill (1983) by use of martingale methods. The almost sure approximation approach was taken by Burke, Csörgő, and Horváth (1981). The bootstrap has been shown to be a.s. consistent, for intervals \([0,M]\) as considered here, by Akritas (1986) and Lo and Singh (1986). Akritas uses martingale methods, while Lo and Singh give asymptotic linearity forms of the theorems with rates of convergence (under continuity assumptions on \( F \) and \( G \)). The \( n^{-1/2} \)-Hellinger regularity result given here is apparently new, but is not surprising.

Although we have not attempted this yet, it would be interesting to show that \( \gamma \) is \( \rho_F \)-continuously, \( C(F,P_0) \)-tangentially differentiable as a function to \( B = L_{\infty}(T) \); or just \( C(F,P_0) \)-tangentially differentiable as a function to \( B = L_{\infty}(R^+) \) with one of the weighted metrics discussed in Gill (1983) or Shorack and Wellner (1986), chapter 7; or perhaps both. The first differentiability result would yield Akritas (1986) as a corollary.

Example 8. (Constrained estimation of a probability measure). Suppose now that \((A, \mathcal{A}) = (R^d, \mathcal{B}^d)\), let \( T : R^d \to R^b \) be a given (vector of) measurable functions, and let \( C \) be any (permissible) Vapnik - Chervonenkis collection of sets in \( R^d \). Consider estimation of \( P_0 \) (as a function from \( C \) to \( R \)) when it is known that
\[ E_{p_0} T \equiv P_0 T = 0 \; ; \; \text{i.e.} \]

(a) \[ P_0 \in \mathcal{P}_0 \equiv \{ P \; \text{on} \; R^d : E_p T = 0 \} . \]

Here we take \( v \) to be a function defined implicitly as follows: for any \( \delta > 0 \) let

(b) \[ \mathcal{P} = \mathcal{P}_\delta \equiv \{ P : E_p \exp(\theta T) < \infty \; \text{for} \; |\theta| < \delta ; \]

\[ 0 \in \text{int supp} \; P \; T^{-1} ; \; \text{Var}_p T > 0 \} . \]

For \( P \in \mathcal{P} \), define a measure \( v(P) = Q_{\text{min}} \) on \( R^d \) by

(c) \[ K(v(P), P) = \inf_{Q: E_Q T = 0} K(Q, P) , \]

where \( K(Q, P) \) is the Kullback-Leibler divergence between the measures \( Q \) and \( P \); i.e.

(d) \[ K(Q, P) = \begin{cases} \int \log \frac{dQ}{dP} dQ & \text{if} \; Q << P , \\ \infty & \text{otherwise} . \end{cases} \]

Results of Csiszar (1975) imply that (for \( P \in \mathcal{P}_\delta \) for some \( \delta > 0 \)) \( v(P) \) is unique, well-defined, and that

(d) \[ \frac{dQ_{\text{min}}}{dP} = \frac{d v(P)}{dP} = \frac{\exp(\theta T)}{\int \exp(\theta T) dP} = c(\theta, P) \exp(\theta T) , \]

where \( \theta \) is chosen so that \( v(P) = Q_{\text{min}} \) satisfies the constraint: i.e.

(e) \[ Q_{\text{min}}^T = c(\theta, P) \int T \exp(\theta T) dP = 0 . \]

This procedure yields a function \( v : P \rightarrow L_{\infty}(C) = B \).

It is known (e.g. Haberman (1984)), that if \( P_0 \in \mathcal{P}_\delta \) for some \( \delta > 0 \), then \( P_n \in \mathcal{P}_\delta \) a.s. for \( n \) sufficiently large. Therefore a natural estimator of \( P_0 \) is simply

(f) \[ \mathcal{Q}_n = v(P_n) , \]

which is called the minimum Kullback-Leibler divergence estimator of \( P_0 \). Note that this estimator satisfies the constraint \( \mathcal{Q}_n^T = 0 \) by construction.

For this example we take \( F \) to be the collection of functions

(g) \[ F = F_{\delta / 2} \equiv \{ T_i \exp(\theta T) , \exp(\theta T) , 1_C \exp(\theta T) , \]

\[ |\theta| < \delta / 2 , \; C \in C , \; i = 1 , \cdots , b \} . \]

An envelope function \( F \) for this collection is

(h) \[ F(y) = F_\delta(y) = (1 \gamma |T(y)|) \exp(\delta |T(y)|) . \]

With this choice of envelope function \( F \), \( F \) is sparse (i.e. (i) of theorem 1.2 holds);
see lemma A.2.2.2, Sheehy (1987). Thus $\mathbf{F}$ is a $P_0$–Donsker class if $P_0(F^2) < \infty$, and $\mathbf{F}$ is a $\{P_n\}$–Donsker class if $F$ is $\{P_n\}$–uniformly square integrable.

Suppose that $P_0 \in P_0 \cap P_\delta$ and that $P_0(F^2_\delta) < \infty$ for some $\delta > 0$, and write $E_0, Var_0, Cov_0$ for expectations, variances, and covariances computed under $P_0$. Then, by an argument along the lines of Sheehy (1987), (1988), $\nu$ is $\rho_F$–continuously, $C(F, P_0)$–tangentially differentiable at $P_0$ with derivative $\nu_0$ given by

(i) \[ \nu_0(D_0)(C) = D_0(\nu_C) \]

where

(j) \[ \nu_C(y) \equiv (1_C(y) - P_0(C)) - Cov_0(1_C, T)'(Var_0 T)^{-1} T(y) \]

for $C \in \mathcal{C}$. Hence all of theorems 3.1 - 3.5 apply. In particular, (3.6) holds where the process $\mathbf{Y}_0$ is given by

(k) \[ \mathbf{Y}_0(C) = \nu_0(\mathbf{X}_0) = \mathbf{X}_0(\nu_C) \quad \text{for} \quad C \in \mathcal{C} \]

\[ = \mathbf{X}_0(1_C) - Cov_0(1_C, T)'(Var_0 T)^{-1} \mathbf{X}_0(T) \]

with covariance

(l) \[ Cov[\mathbf{Y}_0(C), \mathbf{Y}_0(D)] = P_0(C \cap D) - P_0(C)P_0(D) \]

\[ - Cov_0(1_C, T)'(Var_0 T)^{-1} Cov_0(1_D, T) \]

for $C, D \in \mathcal{C}$; note that the asymptotic variance of the estimator $\nu(IP_n)(C)$ of $P_0(C)$ for any fixed $C \in \mathcal{C}$ is smaller than that of the empirical measure by the amount

\[ Cov_0(1_C, T)'(Var_0 T)^{-1} Cov_0(1_C, T). \]

Sheehy (1988) shows that $\nu(IP_n)$ is an asymptotically efficient estimator of $P_0 \in P_0 \cap P_\delta$, and studies $\nu(IP_n)$ for $P_0 \notin P_0$. 
5. Proofs for Section 1

We begin with some preliminary results.

Proposition 5.1. Suppose that

(i) \( P \) is totally bounded in the \( \rho_F \) pseudo-metric.
(ii) \( H_F^{(2)}(\delta, F) < \infty \) for every \( \delta > 0 \).
(iii) \( F \in L_2(P) \) for each \( P \in P \).

Then \( F \) is totally bounded in the \( \tau_F \) pseudo-metric uniformly in \( P \in P \); i.e. for every \( \varepsilon > 0 \) there exist \( f_1, \cdots, f_k \) such that \( \min_{1 \leq i \leq k} \tau_F(f, f_i) < \varepsilon \) for all \( P \in P \).

Proof. Pollard (1982) shows that \( F \) is totally bounded in \( L_2(P) \) for each \( P \in P \); i.e. given any \( \varepsilon > 0 \) there exists a finite collection of functions \( F_{\varepsilon, P} \) so that for any \( f \in F \)

\[
\min_{f' \in F(\varepsilon, P)} \tau_F(f', f) \leq \min_{f' \in F(\varepsilon, P)} ||f - f'||_{L_2(P)} < \varepsilon.
\]

This in turn implies that given any finite collection \( \{P_1, \cdots, P_m\} \subset P \), there exists a finite collection \( F(\varepsilon) \cup \bigcup_{i=1}^m F(\varepsilon, P_i) \) will do, so that for any \( f \in F \)

\[
\min_{f' \in F(\varepsilon)} \tau_F(f', f) < \varepsilon, \quad \text{for} \quad i = 1, \cdots, m.
\]

Now cover \( P \) by \( \rho_F \) - balls of radius \( \varepsilon^2 \) with centers \( \{P_1, \cdots, P_m\} \) and let \( F(\varepsilon) \) be the finite collection of functions defined above. If \( P \in P \), then \( \rho_F(P, P_i) < \varepsilon^2 \) for some \( i = 1, \cdots, m \) and if \( f \in F \), \( \tau_F(f, f') < \varepsilon \) for some \( f' \in F(\varepsilon) \), so that

\[
\tau_F^2(f, f') \leq \tau_F^2(f, f^*) + \varepsilon^2 \leq 2 \varepsilon^2.
\]

Hence

\[
\min_{f' \in F(\varepsilon)} \tau_F(f, f') \leq \sqrt{2} \varepsilon \quad \text{for all} \quad P \in P;
\]

i.e. \( F \) is totally bounded in the \( \tau_F \) pseudometric uniformly in \( P \in P \). \( \square \)

Proposition 5.2. Suppose that \( F \) is permissible with \( N_F^{(2)}(\delta, F) < \infty \) for all \( \delta > 0 \), and suppose that \( P \in F^2 \) < \( \infty \) for all \( P \in P \). Then

\[
\sup_{P \in P} N_F^{(2)}(\delta, F, P) \leq N_F^{(2)}(\delta/2, F).
\]

Proof. Pollard's (1982) proof of his corollary 13, page 244 works. Note that uniform square integrability of \( F \) over \( P \) is not needed for this conclusion. \( \square \)

Our proofs of theorem 1 and its corollaries depend on the following lemmas.
Lemma 5.1. If $N_F^{(d)}(\delta, F) < \infty$ and $F_K \equiv \{ f 1_{[F \leq K]} : f \in F \}$, then $N_K^{(d)}(\delta, F_K) \leq N_F^{(d)}(\delta, F) < \infty$.

Proof. Since $N_F^{(d)}(\delta, F) < \infty$, given a set $S$ we can find $\{ f_1, \cdots, f_m \} \subset F$ so that for each $f \in F$ there is an $i$ with

$$\sum_{x \in S} |f(x)1_{[F(x) \leq K]} - f_i(x)1_{[F(x) \leq K]}|^d$$

$$= \sum_{x \in S \cap [F(x) \leq K]} |f(x) - f_i(x)|^d$$

$$\leq \delta^d \sum_{x \in S \cap [F(x) \leq K]} F(x)^d$$

$$\leq \delta^d \sum_{x \in S \cap [F(x) \leq K]} K^d$$

$$\leq \delta^d \sum_{x \in S} K^d$$

where

$$m \leq N_K^{(d)}(\delta, S \cap \{ F(x) \leq K \}, F_K) \leq N_F^{(d)}(\delta, F) < \infty. \quad \Box$$

Lemma 5.2. Suppose that $F$ has envelope function $K$. If $N_K^{(2)}(\delta, F) < \infty$ for all $\delta > 0$ and $F^* \equiv \{ fg : f, g \in F \}$, then $N_K^{(1)}(\delta, F^*) < \infty$ for all $\delta > 0$.

Proof. We show, in fact, that

$$N_K^{(1)}(2\delta, F^*) \leq \left[ \frac{N_K^{(2)}(\delta, F) + 1}{2} \right] = \left[ \frac{N_K^{(2)}(\delta, F)}{2} \right] + N_K^{(2)}(\delta, F).$$

Let $\{ f_1, \cdots, f_m \} \equiv F_\delta$ be chosen so that given any $f \in F$ there exists $f' \in F_\delta$ so that

(a) $\sum_{x \in S} |f(x) - f'(x)|^2 \leq \delta^2 \sum_{x \in S} K^2$

where $m \leq N_K^{(2)}(\delta, F)$. Then for any $f, g \in F$ with $f_i$ and $g_i$ chosen so that (a) is true (if $f = g$ we can choose $f_i = g_i$), we have, with $n = \#(S)$,

$$\frac{1}{n} \sum_{x \in S} |fg - f_i g_i|$$

$$\leq \frac{1}{n} \sum_{x \in S} |f||g - g_i| + \frac{1}{n} \sum_{x \in S} |g_i||f - f_i|$$

$$\leq K \{ \frac{1}{n} \sum_{x \in S} |g - g_i| + \frac{1}{n} \sum_{x \in S} |f - f_i| \}$$

$$\leq K \{ (\frac{1}{n} \sum_{x \in S} |g - g_i|^2)^{1/2} + (\frac{1}{n} \sum_{x \in S} |f - f_i|^2)^{1/2} \}$$

$$\leq K \{ 2 \delta K \} = 2 \delta K^2. \quad \Box$$
Lemma 5.3. Suppose that $F$ is permissible and has envelope function $K$, a constant, and that $N_K^{(1)}(\delta, F) < \infty$ for all $\delta > 0$. Then $D_n \equiv \|P_n - P\|_F \to a.s. 0$ as $n \to \infty$ uniformly in $P \in \mathfrak{P}$ where $\|P_n - P\|_F \equiv \sup_{f \in F} |(P_n - P)(f)|$:

$$
(1) \quad \sup_{P \in \mathfrak{P}} Pr_P \left\{ \max_{m \geq n} \|P_m - P\|_F > \varepsilon \right\} \to 0 \quad \text{as} \quad n \to \infty
$$

for every $\varepsilon > 0$.

Proof. Let $\varepsilon > 0$. We will prove (1) by showing that we can choose $n(\varepsilon)$ so large that

$$
(a) \quad \sup_{P \in \mathfrak{P}} Pr_P \left\{ \max_{m \geq n} \|P_m - P\|_F > \varepsilon \right\} < \varepsilon \quad \text{for} \quad n \geq n(\varepsilon).
$$

In fact, a choice of $n(\varepsilon)$ that works is

$$
(b) \quad n(\varepsilon) \geq \max \left\{ \frac{8K^2}{\varepsilon^2}, \frac{256K^2}{\varepsilon^2}, H_K^{(1)}(\varepsilon/8K, F), n(\varepsilon, K) \right\}
$$

where $n(\varepsilon, K)$ is so large that

$$
(c) \quad \sum_{m=n(\varepsilon, K)}^\infty \exp \left( -\frac{m \varepsilon^2}{256K^2} \right) < \varepsilon.
$$

The proof uses the symmetrized empirical measure

$$
P_n^0(A) = \frac{1}{n} \sum_{i=1}^n \sigma_i 1_A(X_i)
$$

where $\sigma_1, \sigma_2, \cdots$ are iid Rademacher rvs's (so that $P(\sigma_i = \pm 1) = 1/2$). By Pollard's (1984, page 15, equation (11)) symmetrization lemma it follows that

$$
(d) \quad \sup_{P \in \mathfrak{P}} Pr_P \{ \|P_n - P\|_F > \varepsilon \}
$$

$$
\leq 4 \sup_{P \in \mathfrak{P}} Pr_P \{ \|P_n^0\|_F > \varepsilon/4 \}
$$

$$
= 4 \sup_{P \in \mathfrak{P}} E_P [Pr_P \{ \|P_n^0\|_F > \varepsilon/4 \}|X_n]
$$

for $n \geq 8K^2/\varepsilon^2$; this depends on

$$
\sup_{P \in \mathfrak{P}} \sup_{f \in F} Pr_P \{ |(P_n - P)(f)| > \varepsilon/2 \}
$$

$$
\leq \sup_{P \in \mathfrak{P}} \sup_{f \in F} \frac{4E_P f^2}{n \varepsilon^2}
$$

$$
\leq \frac{4K^2}{n \varepsilon^2} \leq \frac{1}{2} \quad \text{for} \quad n \geq \frac{8K^2}{\varepsilon^2}.$$
to obtain the factor of 4 on the right side. Given \( X_n \), choose functions \( g_1, \ldots, g_M \), \( M = N^{(1)}_K(\varepsilon/8K, F) \) so that
\[
\min_j \mathcal{P}_n |f - g_j| \leq \frac{\varepsilon}{8} \quad \text{for each } f \in F.
\]
Write \( f^* \) for a \( g_j \) at which the minimum is achieved. For any function \( g \)
\[
\mathcal{P}^0_n|g| = n^{-1} \sum_{i=1}^n \sigma_i g(X_i) \leq n^{-1} \sum_{i=1}^n |g(X_i)| = \mathcal{P}_n|g|.
\]
Choose \( g = f - f^* \) for each \( f \) in turn to obtain, by Hoeffding's inequality at the next to last step (Hoeffding (1963) theorem 2; see e.g. Shorack and Wellner (1986) inequality A.4.6),
\[
\text{Prob}\{ \sup_{f} |\mathcal{P}^0_n f| > \frac{\varepsilon}{4} |X_n| \} \leq \text{Prob}\{ \sup_{f} ||\mathcal{P}^0_n f^*| + \mathcal{P}_n|f - f^*|| > \frac{\varepsilon}{4} |X_n| \} \leq \text{Prob}\{ \max_j |\mathcal{P}^0_n g_j| > \frac{\varepsilon}{8} |X_n| \}
\]
because \( \mathcal{P}_n|f - f^*| \leq \varepsilon/8 \)
\[
\leq N^{(1)}_K(\varepsilon/8K, F) \max_j \text{Prob}\{ |\mathcal{P}^0_n g_j| > \varepsilon/8 |X_n| \}.
\]
\[
\leq N^{(1)}_K(\varepsilon/8K, F) \max_j \text{Prob}\{ |\sum_{i=1}^n \sigma_i g_j(X_i)| > n \varepsilon/8 |X_n| \}
\]
\[
\leq 2 \exp[H^{(1)}_K(\frac{\varepsilon}{8K}, F) - 2(n \varepsilon/8)^2/\sum_{i=1}^n (2 g_j(X_i))^2]
\]
by Hoeffding's inequality
\[
\leq 2 \exp[-n \varepsilon^2/256K^2] \quad \text{using } |g_j| \leq K \text{ and (b)}.
\]
Combining (d) and (e) yields, for \( n \geq n(\varepsilon) \),
\[
\sup_{P \in P} \text{Pr}_P\{ \max_{m \geq n} |\mathcal{P}_m - P|_F > \varepsilon \}
\]
\[
\leq 4 \sum_{m=n}^\infty \sup_{P \in P} \text{Pr}_P\{ |\mathcal{P}_m^0|_F > \frac{\varepsilon}{4} \}
\]
\[
\leq 8 \sum_{m=n}^\infty \exp(-\frac{n \varepsilon^2}{256K^2}) < \varepsilon,
\]
and hence (a) holds. \( \square \)
With these three lemmas as preparation we can now prove theorem 1.1.

**Proof of theorem 1.1.** In view of (ii) we can choose $K$ so large that

(a) \[ \sup_{P \in P} E_P F 1_{[F > K]} < \frac{\varepsilon}{4}. \]

Then, since

\[ \|P_m - P\|_F \leq \|P_m - P\|_{F_K} + \sup_{f \in F} |(P_m - P)f 1_{[F > K]}| \]
\[ \leq \|P_m - P\|_{F_K} + P_m F 1_{[F > K]} + P F 1_{[F > K]} \]
\[ \leq \|P_m - P\|_{F_K} + |(P_m - P)F 1_{[F > K]}| + 2P F 1_{[F > K]} \]

(b) \[ \leq \|P_m - P\|_{F_K} + |(P_m - P)F 1_{[F > K]}| + \frac{\varepsilon}{2} \quad \text{by (a)}, \]

it follows that

(c) \[ \sup_{P \in P} P_{rP} \left\{ \max_{m \geq n} \|P_m - P\|_F > \varepsilon \right\} \]
\[ \leq \sup_{P \in P} P_{rP} \left\{ \max_{m \geq n} \|P_m - P\|_{F_K} > \frac{\varepsilon}{4} \right\} \]
\[ + \sup_{P \in P} P_{rP} \left\{ \max_{m \geq n} |(P_m - P)F 1_{[F > K]}| > \frac{\varepsilon}{4} \right\} \]
\[ \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \quad \text{for } n \geq \text{some } N(\varepsilon) \]

by lemma 3 (since $N_{K}^{(1)}(\delta, F_K) < \infty$ by lemma 1) and theorem 0 respectively. \qed

**Proof of corollary 1.1.** Since $C$ is a Vapnik - Chervonenkis class of sets $F = \{ 1_C : C \in C \}$ satisfies (i) of theorem 1 with constant envelope function $F \equiv K = 1$; see e.g. Pollard (1982, 1984) or Dudley (1984, 1986). Condition (ii) holds trivially. Thus (1.7) follows from theorem 1.1. \qed

**Proof of corollary 1.2.** Now $N_{K}^{(2)}(\delta, F) < \infty$ for every $\delta > 0$ implies that $N_{K}^{(2)}(\delta, F_K) < \infty$ for every $\delta > 0$ by lemma 3.1. By lemma 3.2 this implies that $N_{K}^{(1)}(\delta, \tilde{F}_K) < \infty$ for all $\delta > 0$ where $\tilde{F}_K \equiv \{ f g 1_{[F^2 \leq K]} : f, g \in F \}$. Hence the conditions of theorem 1 are satisfied with
\[ \tilde{F} \equiv \{ fg : f, g \in F \} \]
replacing \( F \). Thus (1.8) holds. Since \( \rho_{\tilde{F}}(IP_n, P) \leq D_n^\# \leq 4 \tilde{D}_n \), the remaining two assertions follow from (1.8).

Proof of theorem 1.2. Our proof of theorem 1.2 is a modification of Pollard's (1982) proof of his theorem 7. The basic techniques are not new, but in order to show clearly how (i) and (ii) yield (20), we give the proof in detail.

First note that (ii) implies that there is a constant \( M < \infty \) so that
\[
(\text{a}) \quad \sup_{P \in \mathcal{P}} \|F\|_{L^2(P)}^2 = \sup_{P \in \mathcal{P}} E_P F^2 \leq M < \infty.
\]
This constant enters repeatedly in the remainder of the proof.

Now let \( \delta_j \equiv 2^{-j} \) for \( j \geq 1 \), and set \( H_j \equiv H_{\tilde{F}}^{(2)}(2^{-j}, F) \) so that \( \sum_{j=1}^\infty \delta_j H_j^{1/2} < \infty \) by condition (i). Select a sequence of positive numbers \( \{\eta_j\} \) for which
\[
(\text{b}) \quad \sum_{j=1}^\infty \eta_j < \infty,
\]
\[
(\text{c}) \quad \eta_j \geq (144 M \delta_j^2 H_j)^{1/2}, \quad \text{(so that} \quad H_j \leq \frac{\eta_j^2}{144 M \delta_j^2})\,.
\]
\[
(\text{d}) \quad \sum_{j=1}^\infty \exp\left(-\frac{\eta_j^2}{72 \delta_j^2 M}\right) < \infty.
\]
This is possible because of the growth condition (i) on \( H_{\tilde{F}}^{(2)}(\bullet, F) \). For example, \( \eta_j \equiv \max\{j \delta_j, (144 \|F\|_{L^2} \delta_j^2 H_j)^{1/2}\} \) works.

We now give our choices of \( \delta = \delta(\varepsilon) > 0 \) and \( n = n(\varepsilon, \delta) \) (not dependent on \( P \in \mathcal{P} \)) which yield (20). Choose an integer \( r = r(\varepsilon) \) so large that, with \( \eta = \varepsilon/8 \),
\[
(\text{e}) \quad 2 \sum_{j=r+1}^\infty \exp\left(-\frac{\eta_j^2}{72 M \delta_j^2}\right) < \frac{\varepsilon}{16}, \quad \text{(by (d))}
\]
\[
(\text{f}) \quad \sum_{j=r+1}^\infty \eta_j < \frac{\varepsilon}{16}, \quad \text{(by (b))}
\]
\[
(\text{g}) \quad 2 \exp\left(-\frac{\eta^2}{72 \delta_r^2 M}\right) < \frac{\varepsilon}{16}, \quad \text{(since} \quad \delta_r \to 0)
\]
\[
(\text{h}) \quad \eta^2 \geq 144 M H_r \delta_r^2 \quad \text{(by condition (i))}
\]
all hold. Now choose \( \delta \equiv \delta(\varepsilon) > 0 \) so that
\[
(\text{i}) \quad \delta \leq \min\{ \sqrt{\frac{2}{3}} \delta_r(\varepsilon) M^{1/2}, \frac{\varepsilon}{\sqrt{2}} \}.
\]
With this \( \delta = \delta(\varepsilon) \) we choose \( n = n(\varepsilon, \delta) \) so large that

\[
\sup_{P \in \mathcal{P}} Pr_{P} \{ \rho_{F}(\mathcal{P}, P) > \delta^2/2 \} \leq \frac{\varepsilon}{16}
\]

and

\[
\sup_{P \in \mathcal{P}} Pr_{P} \{ \|F\|_{n}^2 > \|F\|_{L^2(P)}^2 + M \} < \frac{\varepsilon}{16}
\]

where \( \|F\|_{n}^2 = \int F^2 d\mathcal{P} \). Such a choice is possible in view of corollary 1.2 and theorem 1.0 respectively.

To prove (20) we will show that this choice of \( \delta \) and \( n(\varepsilon, \delta) \) imply that

\[
\sup_{P \in \mathcal{P}} Pr_{P} \{ \sup_{[\delta]_{P}} |X_n(f - g)| > \varepsilon \} \leq \varepsilon \quad \text{for all} \quad n \geq n(\varepsilon, \delta).
\]

Write \( X_n = \{X_1, \cdots, X_n\} \).

By lemma II.8, page 14, of Pollard (1984) [or see e.g. inequality A.14.4, page 882, Shorack and Wellner (1986)] it follows that

\[
\sup_{P \in \mathcal{P}} Pr_{P} \{ \sup_{[\delta]_{P}} |X_n(f - g)| > \varepsilon \} \leq 4 \sup_{P \in \mathcal{P}} Pr_{P} \{ \sup_{[\delta]_{P}} \mathcal{P}^0_n(f - g) > \frac{\varepsilon}{4\sqrt{n}} \}
\]

for all \( n \geq 1 \) where \( \mathcal{P}^0_n \) is the symmetrized empirical measure as in lemma 5.3. The validity of (m) depends on

\[
\sup_{P \in \mathcal{P}} \sup_{[\delta]_{P}} Pr_{P} \{ |\sqrt{n}(\mathcal{P}_n - P)(f - g)| > \varepsilon \}
\]

\[
\leq \sup_{P \in \mathcal{P}} \sup_{[\delta]_{P}} \frac{Var_{P}(f - g)}{\varepsilon^2}
\]

\[
\leq \frac{\delta^2}{\varepsilon^2} \leq \frac{1}{2}
\]

by the choice (i) of \( \delta \).

We use (j) to replace \([\delta]_{P}\) on the right side in (m) by \( [[3^{1/2} \delta^{1/2}]_{n}] \) where

\[
[[\delta]]_{n} = \{ (f, g) : f, g \in \mathcal{F} \text{ and } \|f - g\|_{n} < \delta \}, \text{ and } \|f\|_{n}^2 = \int f^2 d\mathcal{P}_n.
\]

Let \( D_n^# = \rho_{F}(\mathcal{P}_n, P) \) be as defined in (1.9), and let \( B_n^c \) denote the event \( [[\|F\|_{n}^2 > \|F\|_{L^2(P)}^2 + M ]] \) in (k). It follows from our choice of \( n \) and (j) and (k) that the right side of (m) is bounded by

\[
4 \sup_{P \in \mathcal{P}} Pr_{P} \{ \sup_{[\delta]_{P}} |\mathcal{P}^0_n(f - g)| > \frac{\varepsilon}{4\sqrt{n}}, D_n^# > \frac{\delta^2}{2} \}
\]

\[
+ 4 \sup_{P \in \mathcal{P}} Pr_{P} \{ \sup_{[\delta]_{P}} |\mathcal{P}^0_n(f - g)| > \frac{\varepsilon}{4\sqrt{n}}, D_n^# \leq \frac{\delta^2}{2} \}
\]
\[ \leq \frac{e}{4} + 4 \sup_{P \in P} \{ \sup_{[\log \delta^2]} \frac{\mathbb{P}^0_n(f - g)}{\mathbb{P}^0_n(f - g)} \mathbb{P}_n \{ X_n \} \} \]

(n) \[ \leq \frac{e}{2} + 4 \sup_{P \in P} E \left[ \sup_{[\log \delta^2]} \frac{\mathbb{P}^0_n(f - g)}{\mathbb{P}^0_n(f - g)} \mathbb{P}_n \{ X_n \} \} 1_{B_n} \right] . \]

In view of (n), the desired inequality (l) will hold if we show that

(o) \[ \text{Prob}\{ \sup_{[\log \delta^2]} \frac{\mathbb{P}^0_n(f - g)}{\mathbb{P}^0_n(f - g)} \mathbb{P}_n \{ X_n \} \} < \frac{e}{8} \quad \text{on} \quad B_n . \]

Choose finite subclasses \( F(1), F(2), \ldots \) of \( F \) such that

(p) \[ \min_{\phi \in F(i)} \| f - \phi \|_n \leq \delta_i \| F \|_n \quad \text{for each fixed} \quad f \in F . \]

By definition 1.1, \( F(i) \) need contain at most \( \exp(H) \) functions (recall that \( \delta_i = 2^{-i} \) and \( H_i = H_i^2(2, i, F) \)). For a given \( f \in F \), denote by \( f_i \) a function \( \phi \) in \( F(i) \) for which the left-hand side of (p) achieves its minimum. Note that \( \| f - f_i \|_n \to 0 \) as \( i \to \infty \). Thus, for any fixed \( r \)

(q) \[ f - f_r = \sum_{j=r+1}^\infty (f_j - f_{j-1}) \]

pointwise on \( X_n \).

The proof of (o) breaks into two parts. The first is to show that for our choice of \( r \geq r(e) \) we have

(r) \[ \text{Prob}\{ \sup_{F} \frac{\mathbb{P}^0_n(f - f_r)}{\mathbb{P}^0_n(f - f_r)} \mathbb{P}_n \{ X_n \} \} < \frac{e}{16} \]
on

\[ B_n = \text{\{ } \| F \|_n^2 \leq \| F \|_{L^2(P)}^2 + M \text{ \}} . \]

The second part is to show that for our choice of \( r \) we have

(s) \[ \text{Prob}\{ \sup_{f, g \in [\log \delta^2]} \frac{\mathbb{P}^0_n(f_r - g_r)}{\mathbb{P}^0_n(f_r - g_r)} \mathbb{P}_n \{ X_n \} \} < \frac{e}{16} \]
on \( B_n \). Since

(t) \[ \sup_{f, g \in [\log \delta^2]} \frac{\mathbb{P}^0_n(f - g)}{\mathbb{P}^0_n(f - g)} \]

\[ \leq 2 \sup_{F} \frac{\mathbb{P}^0_n(f - f_r)}{\mathbb{P}^0_n(f - f_r)} + \sup_{f, g \in [\log \delta^2]} \frac{\mathbb{P}^0_n(f_r - g_r)}{\mathbb{P}^0_n(f_r - g_r)} \]

the inequality (o) follows from (r) - (t).

To prove (r), use (f) and (q) to bound the left side of (r) by

\[ \text{Prob}\{ \sup_{F} \frac{\mathbb{P}^0_n(f - f_r)}{\mathbb{P}^0_n(f - f_r)} \mathbb{P}_n \{ X_n \} \} > \frac{1}{\sqrt{n}} \sum_{j=r+1}^\infty \eta_j |X_n| \]
\[
\begin{align*}
\leq & \sum_{j=r+1}^{\infty} \operatorname{Prob}\{ \sup_{F} |F_n^0(f_j - f_{j-1})| > \frac{\eta_j}{\sqrt{n}} |X_n| \} \\
(u) & \leq \sum_{j=r+1}^{\infty} |F_j||F_{j-1}| \sup_{F} \operatorname{Prob}\{ |F_n^0(f_j - f_{j-1})| > \frac{\eta_j}{\sqrt{n}} |X_n| \} .
\end{align*}
\]

where \(|F_j| = \exp(H_j)|. Consider one of these last conditional probabilities, noting that

\[
F_n^0(f_j - f_{j-1}) = \frac{1}{n} \sum_{i=1}^{n} \sigma_i(f_j - f_{j-1})(X_i) = \frac{1}{n} \sum_{i=1}^{n} \sigma_i h_i .
\]

Thus by theorem 2 of Hoeffding (1963) [see e.g. inequality A.4.6 of Shorack and Wellner (1986)],

\[
(v) \quad \operatorname{Prob}\{ |n^{-1/2} \sum_{i=1}^{n} \sigma_i h_i| > \eta_j |X_n| \} \leq 2 \exp(- \frac{2n \eta_j^2}{4 \sum h_i^2})
\]

where

\[
\sum_{i=1}^{n} h_i^2 = n \|f_j - f_{j-1}\|_n^2
\]

\[
\leq n (\|f - f_j\|_n + \|f - f_{j-1}\|_n)^2
\]

\[
\leq n \|F\|_n^2 (\delta_j + \delta_{j-1})^2
\]

\[
\leq 9n \delta_j^2 [\|F\|_{L_2(P)}^2 + M ] \quad \text{on} \quad B_n
\]

\[
\leq 18n \delta_j^2 M \quad \text{by (a)} .
\]

Therefore on \(B_n\) the sum in (u) is less than

\[
\sum_{j=r+1}^{\infty} \exp(2H_j)2 \exp(- \frac{\eta_j^2}{36 \delta_j^2 M})
\]

\[
\leq 2 \sum_{j=r+1}^{\infty} \exp(- \frac{\eta_j^2}{72 \delta_j^2 M}) \quad \text{by (c)}
\]

\[
< \frac{\varepsilon}{16} \quad \text{by (e)} ;
\]

hence (r) holds.

To prove (s), note that on the event \(B_n \cap \{|f - g|_n < \sqrt{3/2} \delta \} \) we have

\[
\|f_r - g_r\|_n \leq \|f_r - f\|_n + \|f - g\|_n + \|g - g_r\|_n
\]
\[
\leq (\frac{3}{2})^{1/2} \delta + 2 \delta_r \|F\|_n \\
\leq (\frac{3}{2})^{1/2} \delta + 2 \delta_r \left[ \|F\|_{L_2(P)}^2 + M \right]^{1/2} \quad \text{on } B_n \\
(w) \quad < (1 + 2\sqrt{2}) \delta_r M^{1/2} < 3\sqrt{2} \delta_r M^{1/2}
\]

by (a) and our choice of \( \delta \) in (i). Recall that \( \eta \equiv \epsilon/8 \) as in (e) - (h). Use of Hoeffding's inequality again allows us to bound the left side of (s) on \( B_n \) by

\[
|F(r)|^2 \sup_{\|f_r - g_r\|_n} 2 \exp\left(-\frac{\eta^2}{2 \|f_r - g_r\|_n^2}\right) \\
\leq 2 \exp(2H_r - \frac{\eta^2}{36 \delta_r^2 M}) \quad \text{by (w)} \\
\leq 2 \exp\left(-\frac{\eta^2}{72 \delta_r^2 M}\right) \quad \text{by (h)} \\
\leq \frac{\epsilon}{16} \quad \text{by (g)};
\]

Hence (s) holds.

We now show that (20') holds. Define

\[ F^M = \{ f - c : f \in F, \ |c| < M \} . \]

Then, under the hypotheses of the theorem, (i) and (ii) are true with \( F \) replaced by \( F^M \) and \( F \) replaced by \( F + M \). To see this, note that \( F_1^M = \{ c : |c| \leq M \} \) satisfies

\[ \int_0^1 H^R_M(x, F_1^M)^{1/2} \, dx \leq \int_0^1 (\log(1/x))^{1/2} \, dx < \infty, \]

and then apply theorem 10 of Pollard (1982).

Now define

\[ F^P = \{ f - Pf : f \in F \} , \]

\[ [\delta]^P = \{ f , g \in F^P : \|f - g\|_{L_2(P)} < \delta \} , \]

and

\[ [\delta]^M = \{ f , g \in F^M : \|f - g\|_{L_2(P)} < \delta \} . \]

Clearly, for every \( P \in \mathcal{P} \)

\[ F^P \subset F^M \quad \text{and} \quad [\delta]^P \subset [\delta]^M . \]

Also

\[
(x) \quad \sup_{[\delta]^P} |X_n(f - g)| = \sup_{[\delta]^P} |X_n(f - Pf - (g - Pg))| = \sup_{[\delta]^P} |X_n(f - g)| \\
(y) \quad \leq \sup_{[\delta]^M} |X_n(f - g)| .
\]
Since (i) and (ii) imply (20), for every $\epsilon > 0$ there is a $\delta' \equiv \delta'(\epsilon) > 0$ such that
\[
\epsilon > \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} P_{R_P} \{ \sup_{[\delta']} |X_n(f - g)| > \epsilon \}
\geq \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} P_{R_P} \{ \sup_{[\delta']} |X_n(f - g)| > \epsilon \} \quad \text{by (y)}
= \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} P_{R_P} \{ \sup_{[\delta']} |X_n(f - g)| > \epsilon \} \quad \text{by (x)},
\]
and this completes the proof. \(\square\)

Our proof of theorem 1.3 is rather long, but follows closely the "if" part of Dudley's (1984) proof of his theorem 4.1.1, pages 27 - 31. It will be given in complete detail elsewhere. The main additional technical tools needed are:

(i). Uniform in $P \in \mathcal{P}$ bounds on the Prohorov distance between the finite-dimensional laws of the process $X_n$ and the corresponding $P$ - Brownian bridge $X \equiv G_P$. These are obtained from theorem 5.1 below which is derived from the bounds of Yurinskii (1977).

(ii). A uniform in $P \in \mathcal{P}$ weak approximation for sums of mean-zero uniformly square integrable random vectors based on the bounds of (i). This result adds uniformity in $P$ to the weak approximation result of Philipp (1979) (see Dudley (1984), theorem 1.1.3, page 7), and extends (without interpolation to continuous time) the uniform in $P \in \mathcal{P}$ weak approximation result of Lai (1978) from $R^1$ to $R^k$.

Theorem 5.1. For $P \in \mathcal{P}$ let $X_1, \cdots, X_n$ be iid $P$ in $R^k$ with $E_P X_i = 0$ and $\Sigma_P = E_P(X_i X_i^T)$. Suppose that $|X|$ is uniformly square integrable over $P \in \mathcal{P}$ (where $| \cdot |$ denotes the usual Euclidean norm):

\[
(1) \quad \sup_{P \in \mathcal{P}} E_P |X|^2 1_{|X| \geq \lambda} \to 0 \quad \text{as} \quad \lambda \to \infty.
\]

Let $F_n$ denote the law of $n^{-1/2} S_n = n^{-1/2} \sum_{i=1}^n X_i$ on $R^k$ and let $F_{N(0, \Sigma_P)}$ denote the $N_k(0, \Sigma_P)$ law on $R^k$. Then for every $\epsilon > 0$ the Prohorov distance $\pi(F_n, F_{N(0, \Sigma_P)})$ is bounded by

\[
(2) \quad \pi(F_n, F_{N(0, \Sigma_P)}) \leq 2 e^{-2} E_P |X|^2 1_{|X| \geq e^{\sqrt{n}}} + 2^{2/3} (E_P |X|^2 1_{|X| \geq e^{\sqrt{n}}})^{1/3} + C (k M)^{1/4} e^{1/4} (1 + |\log(2^3 e M/k)|^{1/2})
\]

where $C$ is an absolute constant and $\sup_{P \in \mathcal{P}} E_P |X|^2 \leq M$ (which follows for some constant $M$ from (1)). Hence

\[
(3) \quad \sup_{P \in \mathcal{P}} \pi(F_n, F_{N(0, \Sigma_P)}) \to 0 \quad \text{as} \quad n \to \infty.
\]
Theorem 5.1 is the basic result needed to prove the following "uniform in $P$ weak approximation" for sums of independent random variables in $R^k$.

Theorem 5.2. Let $P$ be a collection of laws on $R^k$ with mean zero and covariance matrices $\Sigma_P = E_P(XX^T)$ and suppose that (1) holds. Then there exists a family of probability spaces $\{(\Omega, \mathcal{F}, P): P \in P\}$ with random variables $\{X_i\}_{i \geq 1}$ and $\{Z_i\}_{i \geq 1}$ defined thereon (for each $P \in P$), such that: $X_i$ are iid $P$, $Z_i$ are iid $N_k(0, \Sigma_P)$, and

$$(4) \quad \sup_{P \in P} P \left( n^{-1/2} \max_{m \leq n} \left| \sum_{i=1}^m X_i - \sum_{i=1}^m Z_i \right| \geq \varepsilon \right) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Proof of theorem 5.1. Let $\varepsilon > 0$. Define the truncated variables $X_{in} \equiv X_i 1_{\{|X_i| \leq \varepsilon \sqrt{n}\}}$, and the corresponding centered variables $X_{in}^c \equiv X_{in} - E(X_{in})$, $i = 1, \cdots, n$. If $F_n^\#$ is the law of $n^{-1/2} \sum_1^n X_{in}$ and $F_n^\ast$ is the law of $n^{-1/2} S_n^\ast \equiv \sum_1^n X_{in}^c$, it is easy to show that

(a) $\quad \pi(F_n^\#, F_n^\ast) \leq \varepsilon^{-2} E_P |X|^2 1_{\{|X| \geq \varepsilon \sqrt{n}\}}$

and that

(b) $\quad \pi(F_n^\#, F_n^\ast) \leq \varepsilon^{-2} E_P |X|^2 1_{\{|X| \geq \varepsilon \sqrt{n}\}}$.

Furthermore

$$E_P |X_{in}^c|^3 \leq 23 E_P |X_{in}|^3 \leq 23 \varepsilon \sqrt{n} E_P |X|^2 \leq 23 \varepsilon \sqrt{n} M$$

since (1) implies that

$$\sup_{P \in P} E_P |X|^2 \leq M$$

for some absolute constant $M$. Thus by Yurinskii's (1977) theorem 1 (see also Dehling (1983), theorem B, page 395),

(c) $\quad \pi(F_n^\ast, F_{N(0, \Sigma_n)}) \leq C k^{1/4} (23 \varepsilon \sqrt{n} M)^{1/4} n^{-1/8} (1 + \frac{\log(23 \varepsilon \sqrt{n} M)}{\sqrt{n} k})$

$$\quad \quad \quad = C k^{1/4} (23 \varepsilon M)^{1/4} (1 + \frac{\log(23 \varepsilon M)}{k})^{1/2}.$$

where $\Sigma_n \equiv E(XX^T)$. Finally,

$$E_P |X_i - X_{in}^c|^2 \leq 2^2 E_P |X|^2 1_{\{|X| \geq \varepsilon \sqrt{n}\}}$$

so that

(d) $\quad \pi(F_{N(0, \Sigma_n)}, F_{N(0, \Sigma_n)}) \leq 2^{23} (E_P |X|^2 1_{\{|X| \geq \varepsilon \sqrt{n}\}})^{1/3}$

by Dehling (1983), lemma 2.1, page 402.

Combining (a) - (d) yields (2), and (3) follows from (2). Alternatively, (3) can be proved using a proof by contradiction from (1) and the classical Lindeberg - Feller
central limit theorem.

Proof of theorem 5.2. This proof follows along the lines of the proof of theorem 1 in Philipp (1980) [Corr. (1986)]; the needed uniformity in \( P \) is based fundamentally on the uniformity in \( P \in P \) in theorem 5.1.

Proof of corollary 1.3. We first prove that the \( P - \) Brownian bridges, \( P \in P \), satisfy (15) and (16) by arguing as in Dudley (1984), page 28. By theorem 1.2, (20) holds. Then for any finite subset \( G \) of \( F \), it follows from the assumption (ii) and theorem 5.1 that we can replace \( \mathbb{X}_n(f - g) \) by \( \mathbb{X}(f - g) \) with \( f, g \in G \) in (20) if the \( \varepsilon \) inside the probability is replaced by \( \varepsilon/2 \) and the \( \varepsilon \) outside the probability is replaced by \( 2\varepsilon \): thus

\[
\sup_{P \in P} \Pr_P \left\{ \sup_{[\delta; f, g \in G]} |\mathbb{X}(f - g)| > \varepsilon/2 \right\} < 2\varepsilon.
\]

Since \( \delta \) does not depend on \( G \) or \( P \), we can let \( G \) increase up to a countable dense set \( H \) in \( F \) for \( L_2(P) \), and thereby (using \( 2\varepsilon > \varepsilon/2 \)) obtain

\[
\sup_{P \in P} \Pr_P \left\{ \sup_{[\delta; f, g \in H]} |\mathbb{X}(f - g)| > 2\varepsilon \right\} < 2\varepsilon.
\]

Thus \( \mathbb{X} \equiv \mathbb{X}_P \) defined for \( f \in F \) by

\[
\lim \{ \mathbb{X}_P(h) : h \in H, \tau_P(h, f) \to 0 \}
\]

exists a.s. (uniformly in \( P \in P \)) and defines a (family) of version(s) of \( \mathbb{X} \) satisfying (16).

To prove that (15) holds, let \( F(\delta) \) be a \( \delta - \) net for \( F \) in \( L_2(P) \) with \( k \equiv \#(F(\delta)) \) independent of \( P \) in view of definition 1.1 and proposition 3.2. Let \( \Pi_\delta \) denote a projection onto \( F(\delta) \). Then, since

\[
||\mathbb{X}||_F \leq ||\mathbb{X} - \mathbb{X}_P||_F + ||\mathbb{X}||_{F(\delta)}
\]

(c)

\[
\leq \sup_{[\delta]} |\mathbb{X}(f - g)| + ||\mathbb{X}||_{F(\delta)},
\]

it follows, for \( 0 < \varepsilon/2 < \lambda \), that

\[
\sup_{P \in P} \Pr_P \left( ||\mathbb{X}||_F \geq \lambda \right) \leq \sup_{P \in P} \Pr_P \left\{ \sup_{[\delta]} |\mathbb{X}(f - g)| \geq \varepsilon/2 \right\}
\]

\[
+ \sup_{P \in P} \Pr_P \left\{ ||\mathbb{X}||_{F(\delta)} \geq \lambda - \varepsilon/2 \right\}
\]

\[
\leq \frac{\varepsilon}{2} + \frac{k M}{(\lambda - (\varepsilon/2))^2}
\]

by (16) for \( \delta \leq \delta(\varepsilon) \).
for \( \lambda \geq \varepsilon/2 + (2k M/e)^{1/2} \).

Now we can prove corollary 1.3 itself. For \( k = 1, 2, \ldots \) take \( \varepsilon = 2^{-k} \) in

theorem 1.3; we thereby obtain \( \delta = \delta_k > 0 \) and \( N = N_k \) so that (1.14) holds with

\( \varepsilon = 2^{-k} \) and \( \delta = \delta_k \) for \( n \geq N_k \). For \( k = 1, 2, \ldots \) let \( F_{k,P} \) be a finite

subset of \( F \) such that

\[
\sup_{f \in F} \inf_{g \in F_{k,P}} \{ \tau_p(f, g) : g \in F_{k,P} \} < \delta_k.
\]

Let \( T_{k,P} \) denote the finite-dimensional space of all real functions on \( F_{k,P} \) also with

supremum norm \( \| \cdot \|_{k,P} \). Let \( F_{k,P} = \{ g_1, P, g_2, P, \ldots, g_m(k,P), P \} \). By proposition

5.2 we know that \( m(k, P) \leq m(k) < \infty \). For each \( f \in F \) and \( P \in P \), let \( \hat{f}_{k,P} = g_j, P \) for the least \( j \) such that \( \tau_p(f, g_j, P) < \delta_k \). For any \( \phi \in L_{\infty}(F) \) let \( \phi_{k,p}(f) = \phi(f_{k,P}), f \in F \). Then \( \phi_{k,P} \in L_{\infty}(F) \).

Write \( E_{j, P} = \delta_{X(j)} - P \in L_{\infty}(F), j \geq 1 \). Let \( \Lambda_{k,P}(\phi) = \phi_{k,P} \),

\( E_{k,j,P} = E_{j, P} - \Lambda_{k,P} E_{j, P} \).

The union of the finite-dimensional ranges of the \( \Lambda_{k,P}, k = 1, 2, \ldots \) is

included in a complete separable subspace \( T_P \) of \( L_{\infty}(F) \) with \( C(F, P) \in T_P \), for each \( P \in P \). Note that \( \| \phi_{k,P} \| \leq \| \phi \| \) for all \( K, P \) and all \( \phi \in L_{\infty}(F) \). Then by

(20) we have for \( n \geq N_k \)

(a) \[
\sup_{P \in P} P r_p \{ n^{-1/2} \sum_{j=1}^{n} E_{k,j,P} \| > 2^{-k} \} \leq 2^{-k}.
\]

Let \( P_k \) be the law on \( T_{k,P} \) of \( f \to f(X_1) - P f, f \in F_{k,P} \). Then by theorem

5.1, Strassen's (1965) theorem, and the fact that \( \#(F_{k,P}) \leq m(k) \) for all \( P \in P \),

\( k \geq 1 \), we conclude that there exist random variables \( V_{k,j,P} \) iid \( P_k \), \( W_{k,j,P} \) iid with

a Gaussian law \( Q_{k,P} \) all defined on some common probability space, and some

\( n_k = n(k) \geq N_k, k \geq 1 \) with

(b) \[
\sup_{P \in P} P r_p \{ n^{-1/2} \sum_{j=1}^{n} V_{k,j,P} - W_{k,j,P} \| > 2^{-k} \} \leq 2^{-k}
\]

for all \( n \geq n_k \). Also, \( Q_{k,P} \) must be the law of the restriction of \( \mathbb{X} = G_P \) to

\( F_{k,P} \). We assume that the sequences \( \{ (V_{k,j,P}, W_{k,j,P}) \}_{j=1}^{\infty} \) are independent of each

other for different \( k \). We also take \( 1 \equiv n_0 < n_1 < \cdots \).

For each \( k = 0, 1, \ldots \) if \( n_k \leq j < n_{k+1} \), write \( k = k(j) \) and set

\( V_{j,P} = V_{k(j), P}, W_{j,P} = W_{k(j), P}, \) and \( F_{(j), P} = F_{k(j), P} \). Arguing as in Dudley

(1984), pages 29 - 30, we show that we can define independent Gaussian processes

\( Z_{j,P} \) distributed \( G_P \), on \( (\Omega, \Sigma, P_r) \) so that

\[
L((E_{j,F_{(j), P}}), Z_{j,P}|F_{(j), P}, Z_{j,P}) = L(V_{j,P}, W_{j,P}, Z_{j,P})
\]
on
\[ \prod_{j=1}^{\infty} T(j)_P \times \prod_{j=1}^{\infty} T(j)_P \times \prod_{j=1}^{\infty} U_j_P \equiv X \times Y \times Z \]

where the $U_j_P$'s are copies of $C(F, P_0)$ and $W_j_P = ZZ_j_P|_{F_0,P}$.

For any $k \geq 1$

\[ n^{-1/2} \| \sum_{j=1}^{n} E_j_P - Z_j_P \| \]

\[ = n^{-1/2} \| \sum_{j=1}^{n} E_j_P - \Lambda_k_P E_j_P + \Lambda_k_P E_j_P - \Lambda_k_P Z_j_P \]

\[ + \Lambda_k_P Z_j_P - Z_j_P \| \]

\[ \leq n^{-1/2} \| \sum_{j=1}^{n} E_j_P - \Lambda_k_P E_j_P \|
\]

\[ + n^{-1/2} \| \sum_{j=1}^{n} \Lambda_k_P E_j_P - \Lambda_k_P Z_j_P \|
\]

\[ + n^{-1/2} \| \sum_{j=1}^{n} \Lambda_k_P Z_j_P - Z_j_P \|. \]

So, for any $\varepsilon > 0$

\[ Pr_p \{ n^{-1/2} \| \sum_{j=1}^{n} E_j_P - Z_j_P \| > \varepsilon \} \]

(c)

\[ \leq Pr_p \{ n^{-1/2} \| \sum_{j=1}^{n} E_j_P - \Lambda_k_P E_j_P \| > \varepsilon/3 \}
\]

\[ + Pr_p \{ n^{-1/2} \| \sum_{j=1}^{n} (\Lambda_k_P E_j_P - \Lambda_k_P Z_j_P) \| > \varepsilon/3 \}
\]

\[ + Pr_p \{ n^{-1/2} \| \sum_{j=1}^{n} \Lambda_k_P Z_j_P - Z_j_P \| > \varepsilon/3 \}. \]

The first term in (c) can be made less than $\varepsilon/3$ by choosing $k$ large enough so that $2^{-k} < \varepsilon/3$ and $n \geq N_k$ where $N_k$ was defined in (a). Since, for each $P \in P$, $n^{-1/2} \sum_{j=1}^{n} ZZ_j_P$ is distributed as a $P$-Brownian bridge $X = G_p$, and so is equal in law to a version of $X = G_p$ satisfying (16), by choosing $k$ large enough we can make the third term in (c) less than $\varepsilon/3$ also.

Finally, since the second term in (c) is equal to

\[ Pr_p \{ n^{-1/2} \| \sum_{j=1}^{n} V_{kj,P} - W_{kj,P} \| > \varepsilon/3 \}, \]
from (b) we conclude that for \( n \geq n_k \) (where \( n_k \) was defined in (b)), the third term is less than \( \varepsilon/3 \), and we are done. \( \square \)

**Proof of proposition 1.1.** Let \( F \) be a \( P \)-uniform functional Donsker class. Let \( E_i \) and \( Z_i \) be as in definition 1.5. Given \( \varepsilon > 0 \), let \( N(\varepsilon) \) be so large that

(a) \[ \sup_{P \in \mathcal{P}} \mathbb{P}_P \left\{ n^{-1/2} \left \| \sum_{i=1}^n (E_i - Z_i) \right \|_F > \varepsilon/3 \right\} < \varepsilon/2. \]

Since \( F \) is a \( P \)-uniform \( G_P \) BUC class, there is a \( \delta = \delta(\varepsilon) \) such that, with \( \mathcal{X} = G_P \) for each \( P \in \mathcal{P} \),

(b) \[ \sup_{P \in \mathcal{P}} \mathbb{P}_P \left\{ \sup_{[\delta]} (\mathcal{X}(f - g)) > \varepsilon/3 \right\} < \varepsilon/2. \]

By assumption each \( (Z_1 + \cdots + Z_n)/n^{1/2} \) is such a version of \( \mathcal{X} = G_P \). But

(c) \[ \sup_{[\delta]} \left| \mathcal{X}_n(f - g) \right| \leq 2 \sup_{f \in F} \left| \mathcal{X}_n(f) - n^{-1/2} \sum_{i=1}^n Z_i(f) \right| \]

\[ + \sup_{[\delta]} \left| n^{-1/2} \sum_{i=1}^n Z_i(f - g) \right|, \]

and hence (21) follows from (a), (b), and (c). \( \square \)

**Proof of corollary 1.4.**

This can be proved in several ways. The following proof is along the lines of our proof of corollary 1.3, but using theorem 7 of Dehling (1983) to control \( \pi(F_n(0, \Sigma(P_n)), F_n(0, \Sigma(P_0))) \). Let \( \Sigma_n = \Sigma(P_n) \) and \( \Sigma_0 = \Sigma(P_0) \) denote the \((k \times k)\) covariance matrices of the \( P_n \)- and \( P_0 \)-Brownian bridges, respectively, restricted to a \( \delta \)-net \( F(\delta) \) for \( F \) with \( k \equiv \#(F(\delta)) \); and let \( \|A\|_1 \equiv \text{trace class norm of } A \). Then

\[ t_n = \|\Sigma_n - \Sigma_0\|_1 \leq k^2 \sup_{f, g \in F(\delta)} |\text{Cov}_{P_n}(f, g) - \text{Cov}_{P_0}(f, g)| \]

\[ \leq k^2 \sup_{f, g \in F} |\text{Cov}_{P_n}(f, g) - \text{Cov}_{P_0}(f, g)| \]

\[ \leq k^2 \left( \frac{9}{2} + 8 M^{1/2} \rho_F(P_n, P_0) \right) \]

where \( \sup_{n \geq 1} \mathbb{P}_n(F^2) \leq M \)

(a) \[ \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{by hypothesis (iii)} \].

Thus by theorem 7 of Dehling (1983) page 400,

\[ \pi_{kn} = \pi(F_n(0, \Sigma_n), F_n(0, \Sigma_0)) \]

\[ \leq C t_n^{1/3} k^{1/6} (1 + |\log(k/t_n)|^{1/2}) \]
\[ \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \quad \text{by (a)}. \]

Also note that under the hypotheses of corollary 1.4 we have

(b) \[ \limsup_{n \to \infty} \sup_{[\delta]} Pr_{\mathcal{P}_n} \left\{ |X_n(f - g)| > \varepsilon \right\} = 0. \]

Then the rest of the proof is the same as that of corollary 1.3. \qed

**Proof of corollary 1.5.** We use the \( X_n \)'s and \( X^{(n)} \)'s as constructed in corollary 1.3. Since \( g \) is uniformly continuous on \( L_\infty(\mathcal{F}) \), given any \( \varepsilon > 0 \) there is a \( \delta > 0 \) so that \( \phi_1, \phi_2 \in L_\infty(\mathcal{F}) \) and

\[ ||\phi_1 - \phi_2||_F \leq \delta \]

imply

\[ |g(\phi_1) - g(\phi_2)| \leq \varepsilon. \]

Suppose, with loss of generality, that \( |g| \leq M \). Then

(a) \[ g(X_n) \leq (g(X^{(n)}) + \varepsilon)1[||X_n - X^{(n)}||_F^* \leq \delta] \]

\[ + M1[||X_n - X^{(n)}||_F^* > \delta] \]

and since the RHS of (a) is measurable,

(b) \[ E_{\mathcal{P}} g(X_n) \leq E_{\mathcal{P}} g(X^{(n)})1[||X_n - X^{(n)}||_F^* \leq \delta] + \varepsilon \]

\[ + M Pr_{\mathcal{P}} \{ ||X_n - X^{(n)}||_F^* > \delta \}. \]

From (b) we obtain

(c) \[ E_{\mathcal{P}}^* g(X_n) \leq E_{\mathcal{P}} g(X^{(n)}) + M Pr_{\mathcal{P}} \{ ||X_n - X^{(n)}||_F^* > \delta \} + \varepsilon. \]

Applying (1.22) to (c) yields: for any \( \varepsilon > 0 \), for \( n \geq n(\varepsilon) \) sufficiently large,

(d) \[ \sup_{P \in \mathcal{P}} \{ E_{\mathcal{P}}^* g(X_n) - E_{\mathcal{P}} g(X^{(n)}) \} \leq 2\varepsilon. \]

We now show that given any \( \varepsilon > 0 \), for \( n \geq n(\varepsilon) \) sufficiently large

(e) \[ \sup_{P \in \mathcal{P}} \left\{ E_{P} g(X_0) - E_{P}^* g(X_n) \right\} \leq \varepsilon. \]

First note that

\[ E_{\mathcal{P}}^* g(X_n) \leq E_{\mathcal{P}} g(X_n) \]

so that

(f) \[ E_{\mathcal{P}} g(X_0) - E_{\mathcal{P}}^* g(X_n) \]

\[ \leq E_{\mathcal{P}} g(X_0) - E_{\mathcal{P}} g(X_n) \]

\[ = E_{\mathcal{P}} g(X_0) - E_{\mathcal{P}}^* g(X_n). \]
From (d), for any bounded uniformly continuous function $g$ on $L_\infty(F)$, for $n \geq n(\varepsilon)$,

(h) \[ \sup_{P \in \mathcal{P}} E_P^*(-g(X_n)) - E_P(-g(X_0)) \leq \varepsilon. \]

Moreover,

(i) \[ (-g(X_n))^* = -g^*(X_n). \]

Therefore

\[ \sup_{P \in \mathcal{P}} \{ E_P g(X_0) - E_P^* g(X_n) \} \]
\[ \leq \sup_{P \in \mathcal{P}} \{ E_P g(X_0) - E_P^* g(X_0) \} \quad \text{by (f)} \]
\[ = \sup_{P \in \mathcal{P}} \{ -E_P g^*(X_n) + E_P g(X_0) \} \quad \text{by (g)} \]
\[ = \sup_{P \in \mathcal{P}} \{ E_P^*(-g(X_n)) - E_P(-g(X_0)) \} \quad \text{by (i)} \]
\[ \leq \varepsilon \quad \text{for } n \geq n(\varepsilon) \quad \text{by (h)}, \]

thereby proving (e) and the corollary. □

Proof of corollary 1.6. It suffices to verify the hypotheses of corollary 1.4. Hypothesis (i) follows from theorem 9 of Pollard (1982); or see Dudley (1984), (1986). Hypothesis (ii) holds trivially since $F \equiv 1$ is bounded in this case, and (iii) holds by hypothesis. Hence corollary 1.4 applies. □

Proof of corollary 1.7. By corollary 1.4 it suffices to show that $H(P_n, P_0) \to 0$ implies $\rho_F(P_n, P_0) \to 0$ for $\{P_n\} \subset \mathcal{P}$. But, by Cauchy - Schwarz, for $f, g \in F$, with $\mu_n \equiv P_n + P_0$, $p_n \equiv dP_n/d\mu_n$, $s_n \equiv \sqrt{p_n}$, $P_0 \equiv dP_0/d\mu_n$, $s_0 \equiv \sqrt{P_0}$, it follows that

\[ \left\{ \int f g d(P_n - P_0) \right\}^2 \leq \left\{ \int f \varnothing [s_n - s_0][s_n + s_0] d\mu_n \right\}^2 \]

(a) \[ \leq \int f^2 [s_n - s_0]^2 d\mu_n \int g^2 [s_n + s_0]^2 d\mu_n \]

where

\[ \int g^2 [s_n + s_0]^2 d\mu_n \leq 2 \int g^2 d(P_n + P_0) \]

(b) \[ \leq 2 \int F^2 d(P_n + P_0) \leq \text{some } M < \infty \]

by the uniform integrability hypothesis (ii), and, for $\lambda > 0$,

\[ \int f^2 [s_n - s_0]^2 d\mu_n \]
\[ = \int f^2 1_{[|f| > \lambda]} [s_n - s_0]^2 d\mu_n + \int f^2 1_{[|f| \leq \lambda]} [s_n - s_0]^2 d\mu_n \]

\[ \leq \int f^2 1_{[|f| > \lambda]} [s_n - s_0]^2 d\mu_n + \int f^2 1_{[|f| \leq \lambda]} [s_n - s_0]^2 d\mu_n \]

by the uniform integrability hypothesis (ii), and, for $\lambda > 0$,
\[
\leq 2 \int f^2 1_{|f| > \lambda} [s_n^2 + s_0^2] d\mu_n + \lambda^2 H^2(P_n, P_0)
\]
\[
\leq 2 \{ E_{P_n} F^2 1_{[F > \lambda]} + E_{P_0} F^2 1_{[F > \lambda]} \} + \lambda^2 H^2(P_n, P_0).
\]

Hence, combining (a), (b), and (c) yields
\[
\rho^2(P_n, P_0) \leq 4M \left\{ 2 \{ E_{P_n} F^2 1_{[F > \lambda]} + E_{P_0} F^2 1_{[F > \lambda]} \} + \lambda^2 H^2(P_n, P_0) \right\}
\]
\[
\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty
\]
in view of uniform integrability and \( H(P_n, P_0) \rightarrow 0 \) by choosing \( \lambda \) sufficiently large to make the first terms small and then letting \( n \rightarrow \infty \). \( \square \)

**Proof of theorem 1.4.** We first prove (1.27). This goes along the lines of lemma 5.21 of van der Vaart (1988); we give the details for completeness. Define \( \mu_n, p_n, p_0, s_n, \) and \( s_0 \) as in the proof of corollary 1.6. Then for \( f \in F \) and \( \{ P_n \} \) satisfying (1.25)
\[
x_n(f) - x_0(f) = \int f \left\{ \sqrt{n} (s_n - s_0) - \frac{1}{2} h s_0 \right\} (s_n + s_0) d\mu_n
\]
\[
+ \frac{1}{2} \int f h (s_n - s_0) s_0 d\mu_n.
\]
(a)
\[
= A_n(f) + B_n(f)
\]
where
(b)
\[
|A_n(f)| \leq \{ 2 \int f^2 dP_n + 2 \int f^2 dP_0 \}^{1/2}
\]
\[
\cdot \left\{ \left[ \sqrt{n} (s_n - s_0) - \frac{1}{2} h s_0 \right]^2 d\mu_n \right\}^{1/2}
\]
\[
\leq \{ 2 P_n(F^2) + 2 P_0(F^2) \}^{1/2}
\]
\[
\cdot \left\{ \left[ \sqrt{n} (dP_n^{1/2} - dP_0^{1/2}) - \frac{1}{2} h dP_0^{1/2} \right]^2 \right\}^{1/2}
\]
(c)
\[
\rightarrow 0 \quad \text{uniformly in} \quad f \in F \quad \text{as} \quad n \rightarrow \infty.
\]

by (1.25) and (1.26). Furthermore,
(d)
\[
|2B_n(f)| \leq \int_{|f| \leq n^{1/4}} f h s_0 (s_n - s_0) d\mu_n
\]
\[
+ \int_{|f| > n^{1/4}} h s_0 f (s_n - s_0) d\mu_n
\]
\[
\leq \{ n^{1/2} h^2 dP_0 \int (s_n - s_0)^2 d\mu_n \}^{1/2}
\]
\[
+ \{ P_0(h^2 1_{[F > n^{1/4}]})(P_n(F^2) + P_0(F^2)) \}^{1/2}
\]
(e) \[ 0 + 0 = 0 \] uniformly in \( f \in F \) as \( n \to \infty \)

since \( P_0(h^2) < \infty \) and by (1.23) and (1.26). Combining (a) - (e) yields (1.27).

Now we prove that \( F \) is a \( \{P_n\} \) - uniform Donsker class. Since \( F \) is a \( P_0 \)-Donsker class, it follows from Dudley (1984) theorem 4.1.1 that, for every \( \varepsilon > 0 \) there is a \( \delta = \delta(\varepsilon) > 0 \) such that

\[
(f) \quad Pr_{P_0}^* \{A_n(\varepsilon)\} = Pr_{P_0}^* \{\sup_{[\delta]_F} |\mathbb{X}_n^0(f - g)| > \varepsilon \} = Pr_{P_0}^* \{A_n^*(\varepsilon)\} \to 0 \quad \text{as} \quad n \to \infty.
\]

But (1.25) implies that \( \{Pr_{P_0}^* \} \) is contiguous to \( \{Pr_{P_0} \} \) by Le Cam's first lemma: (1.25) implies that

\[
\Lambda_n = \log \prod_{i=1}^{n} \frac{P_n(X_i)}{P_0} \quad \text{satisfies}
\]

\[
(g) \quad \Lambda_n - \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) - \frac{1}{2} P_0(h^2) \right) \equiv o_{P_0}(1),
\]

and hence that \( \Lambda_n \to_d N(-\sigma^2/2, \sigma^2) \) under \( P_0 \). But then (f) implies that, for every \( \varepsilon > 0 \) there is a \( \delta(\varepsilon) \) (the same \( \delta \) as in (f)) so that

\[
(h) \quad Pr_{P_n}^* \{A_n(\varepsilon)\} = Pr_{P_n}^* \{\sup_{[\delta]_F} |\mathbb{X}_n^0(f - g)| > \varepsilon \} = Pr_{P_n}^* \{A_n^*\} \to 0 \quad \text{as} \quad n \to \infty.
\]

by the definition of contiguity; see e.g. Shorack and Wellner (1986), pages 152 - 160.

Now

\[
(i) \quad \mathbb{X}_n(f - g) = \mathbb{X}_n^0(f - g) - x_n(f - g),
\]

so

\[
(j) \quad \sup_{[\delta]_F} |\mathbb{X}_n(f - g)| \leq \sup_{[\delta]_F} |\mathbb{X}_n^0(f - g)| + \sup_{[\delta]_F} |x_n(f - g)|,
\]

where

\[
\sup_{[\delta]_F} |x_n(f - g)| \leq \sup_{[\delta]_F} |x_n(f - g) - x_0(f - g)| + \sup_{[\delta]_F} |x_0(f - g)|
\]

\[
\leq 2 \|x_n - x_0\|_F + \sup_{[\delta]_F} |P_0(f - g)h|\]

\[
\leq 2 \|x_n - x_0\|_F + \{\delta^2 P_0(h^2)\}^{1/2}
\]
if \( \delta < \varepsilon/(4\|h\|_{L_2(P_0)}) \) and \( n \) is so large that \( \|x_n - x_0\|_F < \varepsilon/8 \). Combining (h), (j), and (k) yields: for every \( \varepsilon > 0 \) there is a \( \delta = \delta(\varepsilon) \) (chosen to satisfy (f) and (k)) so that

\[
\text{Pr}_{P_n}^* \{ \sup_{[\delta]_n} |X_n(f - g)| > \varepsilon/2 \} \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\]

Since \( \rho_F(P_n, P_0) \rightarrow 0 \), it follows from (1.10) via the same argument as in the proof of theorem 1.2 that (l) implies: for every \( \varepsilon > 0 \) there is a \( \delta = \delta(\varepsilon) \) such that

\[
\text{Pr}_{P_n}^* \{ \sup_{[\delta]_n} |X_n(f - g)| > \varepsilon \} \rightarrow 0 \quad \text{as } n \rightarrow \infty.
\]

Asymptotic normality of the finite dimensional distributions of the process \( X_n \) follows from Le Cam’s third lemma and (1.27).

The remainder of the proof is like the proof of corollary 1.4 using (m) and the asymptotic normality of the finite-dimensional distributions; however, since permissibility of \( F \) is not assumed, the argument uses outer measures and measurable covering functions as in Dudley (1984).

Proofs of theorems 1.5 and 1.6. In both cases it suffices to check that

\[
\rho_F(P_n, P_0) \rightarrow_{a.s.} 0 \quad \text{as } n \rightarrow \infty.
\]

In the case of theorem 1.5, this follows easily from corollary 1.2 applied to \( P = \{P_0\} \). In the case of theorem 1.6 we have

\[
\rho_F(P_n, P_0) \leq \rho_F(P_n, P_n) + \rho_F(P_n, P_0) \rightarrow_{a.s.} 0 + 0 = 0
\]

by (9) of corollary 1.2 and by hypothesis. The conclusion then follows from corollary 1.3.

Acknowledgments: We owe thanks to: a referee for suggestions which lead to a clearer presentation of the proof of theorem 1.2; Lucien LeCam for encouraging us to formulate our results in terms of theorem 1.3 and corollary 1.3; Richard Gill for stimulating discussions concerning Hadamard derivatives and the delta method; Rudy Beran for several helpful suggestions and questions about the bootstrap; and Peter Gaenssler for his interest and encouragement.
REFERENCES


