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

by

ANNE SHEEHY
JON A. WELLNER

TECHNICAL REPORT No. 87
August 1986
(Revised February, 1987)

Department of Statistics, GN-22
University of Washington
Seattle, Washington 98195 USA
UNIFORMITY IN P
OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by

Anne Sheehy and Jon A. Wellner

University of Washington

August, 1986
Revised February, 1987

AMS 1980 Subject Classifications: Primary 62G05, 60F05;
Secondary 62G30, 60G44

Key Words and Phrases: empirical process, central limit theorem, regular estimators,
strong law of large numbers, uniformity, uniform integrability, bootstrap.

Research supported by: National Science Foundation Grant DMS-8400893

February 19, 1987
UNIFORMITY IN $P$
OF SOME LIMIT THEOREMS
FOR EMPIRICAL MEASURES AND PROCESSES

by

Anne Sheehy and Jon A. Wellner
University of Washington

ABSTRACT

Let $P$ be a collection of probability measures on the measurable space $(A,\mathcal{A})$. Suppose that $P_n$ is the empirical measure for $X_1, \cdots, X_n$ iid $P \in P$. Let $F$ be a class of measurable functions with envelope $F$; thus $|f| \leq F$ for all $f \in F$. Then $X_n = \sqrt{n} (P_n - P)$ is the empirical process on $F$. It is shown that uniform integrability of $F$ over $P$ and a certain combinatorial entropy condition on $F$ imply that

$$\sup_{F} |P_n(f) - P(f)| \to 0 \text{ a.s. uniformly in } P \in P.$$ 

If $F$ is sparse and $F^2$ is uniformly integrable over $P$, then we establish a uniform asymptotic equicontinuity property of the empirical process $X_n$ uniformly in $P \in P$. This result is used to prove weak convergence of the empirical processes $X_n$ when the $X_i$’s are iid $P$, where $P$ converges in a certain sense to a fixed $P_0$. Applications include: (1) a proof of the convergence of the "bootstrapped empirical process"; (2) a method of establishing the "regularity" of estimators; and, (3) (local asymptotic) power approximations for tests.
0. Introduction.

Limit theorems (usually as the sample size $n \to \infty$) are a staple of mathematical statistics; they provide assurance of the consistency of estimators and useful approximations to the distributions of estimates and power of tests. Typically the limit theorems are established (under some hypotheses) at a fixed probability measure $P$ in a collection of probability measures (model) $\mathcal{P}$, or at a fixed parameter value $\theta$ in the parameter set $\Theta$ indexing the model. However, $P$ (or $\theta$) is usually unknown to the statistician, and hence it is desirable to have the limit theorems or approximations hold as uniformly in $P$ (or $\theta$) as possible. If the convergence is not, in some sense, uniform, then an adequate sample size cannot be specified even in principle since $P$ (or $\theta$) is unknown.

Attention to uniformity in $P$ (or $\theta$) of the convergence in limit theorems in statistics has been spotty. We now briefly review some of the literature (known to us) in which attention has been paid to this type of uniformity of convergence.

Uniformity was clearly of considerable concern to Wald (1943) in his study of the large sample optimality of tests. Apparently Wald's interest in the uniformity of convergence prompted Chung (1951) to prove his uniform in $P \in \mathcal{P}$ strong law of large numbers under the assumption that $|X|$ is uniformly integrable over $P$; see Chung (1951) page 341 and our section 1 below for Chung's theorem. Le Cam (1953), pages 307 - 308 gave a careful treatment of the uniform (in $\theta \in K$ compact $\subset \Theta$) convergence to normality of the distributions of maximum likelihood estimates. [It is clear that Le Cam had studied Wald (1943) carefully.] Le Cam used the fact that uniform convergence in the ordinary CLT for sums follows from the assumption that $|X|^2$ is uniformly integrable over $P$ by application of a refinement of the Berry - Esseen theorem; see e.g. Petrov (1975) pages 118 - 120. Another example of careful attention to uniformity in $P$ is the classic paper of Chernoff and Savage (1958) on the asymptotic normality of rank statistics under alternatives. In related work, Lai (1978) proved uniform invariance principles for partial sum processes, and used them to analyze sequential tests based on rank statistics. A useful consequence of the uniform asymptotic normality proved by these authors is the behavior of the power of the rank tests under local alternatives and derivations of Pitman efficiency of the rank tests. Pollard (1980) outlines an approach to proving the needed uniformity for a class of tests based on minimum distance estimation; see his section 6, pages 62 ff. His methods are closely related to ours; see section 2 below. Ibragimov and Has'minskii (1981) give a useful summary of uniform limit theorems; see their Appendix I pages 363 ff.

In another, slightly weaker guise, uniformity of convergence in $P$ (or $\theta$) has entered the statistics literature in the form of the regular estimates introduced by Hájek (1970), (1972) as a part of his approach to asymptotic efficiency in estimation. Hájek's notion of a regular estimator $\hat{\theta}_n$ entails (in the context of a parametric model) convergence in law of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ under $P_{\theta_0}$ to a limit distribution $L_0$ not depending...
on \( h \) whenever \( \theta_n = \theta + n^{-1/2}h \). Thus regularity of an estimator is a weaker kind of uniformity of convergence which is implied by uniform convergence of its distributions uniformly in \( \theta \) in compact subsets \( K \) of \( \Theta \). Extended versions of Hájek’s notion of regularity have been used in a wide range of nonparametric and semiparametric models; see e.g. Beran (1977), Wellner (1982), Begun, Hall, Huang, and Wellner (1983), and Millar (1985). One of the main motivations for the present work was to provide sufficient conditions for the regularity of estimators in nonparametric problems such as those of Millar (1985). Further treatment and elaboration on these themes will be given in the monograph by Bickel, Klaassen, Ritov, and Wellner (1987); applications to constrained estimation problems will be contained in the forthcoming University of Washington thesis of Sheehy (1987).

Uniformity of convergence has been a continuing theme in the study of the limit theory of empirical distributions and processes as well. For the empirical df of (real-valued) random variables (observations in \( \mathbb{R}^d \)), the fact that the distribution of many statistics of interest does not depend on the true distribution \( P \) (or \( F \)) of the data is simply a consequence of the (inverse) probability integral transformation; see e.g. Shorack and Wellner (1986) chapters 1 and 3. This fact was exploited by Dvoretzky, Kiefer, and Wolfowitz (1956) in their study of the asymptotic optimality of the empirical distribution function as an estimator of the true distribution function. Their exponential bound for the distribution of the supremum distance between the empirical distribution and true distribution is uniform in \( P \) (or \( F \)), and this remains true of the exponential bounds for the corresponding suprema in the case of observations in \( \mathbb{R}^d \) due to Kiefer (1961), and for the case of the empirical distribution indexed by a Vapnik-Chervonenkis class of subsets of a general sample space \( A \) due to Alexander (1984). However, the distribution of the empirical process and its supremum does depend on \( P \) (or \( F \)) when the sample space is \( \mathbb{R}^d \), and hence Kiefer and Wolfowitz (1959) had to invest considerable effort to establish the uniformity (in \( P \) or \( F \)) of convergence of the empirical process to the limit Gaussian process in the course of their proof of the asymptotic optimality of the empirical distribution function (as an estimator of \( F \) on \( \mathbb{R}^d \)). Both section 12 of Dudley (1984) and Dudley (1986) are primarily concerned with the closely related notion of convergence of, and bounds for, the empirical process for every \( P \) [Dudley calls the former the "universal Donsker" property].

In this paper we focus on "uniformity in \( P \)" of some limit theorems for empirical measures and processes. We first prove a "uniform in \( P \)" version of a Glivenko-Cantelli theorem for empirical measures due to Pollard (1982). Chung’s (1951) uniform strong law of large numbers is a special case of our strengthened version of Pollard’s theorem. We then give a "uniform in \( P \)" version of a part of a limit theorem for the empirical process indexed by a subset \( F \) of \( L_2(P) \) which is also due to Pollard (1982); our theorem gives very natural sufficient conditions for uniform asymptotic equicontinuity of the empirical process -- uniformly in \( P \in \mathcal{P} \). We then use this theorem to deduce several weak convergence results for the empirical process of
observations \( X_1, \ldots, X_n \) iid \( P_n \). A corollary of our theorems is a set of sufficient conditions for regularity of \( P_n \) as an estimator of \( P \).

1. Uniform limit theorems

We now present our uniform in \( P \in \mathbb{P} \) 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 explain 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

\[
\bar{X}_n \to_a.s. \mu = E_P(X) \quad \text{as} \; n \to \infty.
\]

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

\[
Pr_P \left\{ \max_{m \geq n} |\bar{X}_m - \mu| > \varepsilon \right\} \to 0 \quad \text{as} \; n \to \infty.
\]

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

**Theorem 0.** (Chung, 1951). Suppose that \( X, X_1, \ldots, X_n, \ldots \) are iid \( P \in \mathbb{P} \) and that \( P \) satisfies the uniform integrability condition

\[
\lim_{\lambda \to \infty} \sup_{P \in \mathbb{P}} E_P|X| 1_{|X| \geq \lambda} = 0.
\]

Then \( \bar{X}_n \to_a.s. \mu \) as \( n \to \infty \) uniformly in \( P \in \mathbb{P} \); i.e.

\[
\sup_{P \in \mathbb{P}} Pr_P \left\{ \max_{m \geq n} |\bar{X}_m - \mu| \geq \varepsilon \right\} \to 0 \quad \text{as} \; n \to \infty.
\]

Uniform integrability conditions like (3) of Chung’s theorem 0 will recur repeatedly in the theorems to follow. Thus it will be useful to recall for the reader the following useful characterization of uniformly integrable families.

**Theorem.** (Vallée - Poussin). A family of \( L_1 \) rv’s \( \{X_i : i \in T \} \) 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

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

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 \( P \in \mathbb{P} \).

February 19, 1987
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 (1981,1984) weak convergence theorems for the empirical process.

To state our uniform version of Pollard’s theorem, we first need Pollard’s combinatorial entropy for a class of functions. Let \( P \) be a collection of probability measures on the measurable space \( (\mathcal{A}, \mathcal{F}) \), and let \( \mathcal{F} \) be a collection of \( \mathcal{A} \)-measurable functions defined on \( \mathcal{A} \) with envelope function \( F \); i.e. \( |f| \leq F \) for each \( f \in \mathcal{F} \). We will usually assume that \( \mathcal{F} \cup \{F\} \subset L_r(P) \) for some \( r > 0 \). Suppose that \( X_1, \cdots, X_n, \cdots \) are iid \( (\mathcal{A}\text{-valued}) \) random variables with distribution \( P \in \mathcal{P} \), and let \( P_n \) denote the empirical measure of the first \( n \) \( X \)'s:

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

**Definition 1.** (Pollard, 1982). Let \( r > 0 \). Suppose that \( F \) is an envelope function for \( \mathcal{F} \). For each finite subset \( S \) of \( \mathcal{A} \) and each \( \delta > 0 \) let \( N^{r,F}_{\delta}(S, \mathcal{F}) \) be the smallest value of \( m \) for which there exists \( f_1, \cdots, f_m \in \mathcal{F} \) such that

\[
\min\left\{ \sum_{x \in S} |f(x) - f_i(x)|^r \leq \delta^r \sum_{x \in S} F(x) \right\}
\]

for every \( f \in \mathcal{F} \). Then let

\[
N^{r,F}_{\delta}(\delta, \mathcal{F}) = \sup_{S} N^{r,F}_{\delta}(S, \mathcal{F}), \quad \text{and} \quad H^{r,F}_{\delta}(\delta, \mathcal{F}) = \log N^{r,F}_{\delta}(\delta, \mathcal{F})
\]

where the supremum is over all finite subsets \( S \) of \( \mathcal{A} \).

Throughout the following we assume that \( \mathcal{F} \) is a permissible class of functions in the sense of Pollard (1984), page 196; this assumption is needed to ensure measurability of supra as \( D_n \) below. (An alternative approach involving outer measures and measurable cover functions as in Dudley and Philipp (1983) and Dudley (1984) could also be used.)

**Theorem 1.** Suppose \( \mathcal{F} \) is permissible and that:

(i) \( N^{(0,F)}_{K}(\delta, \mathcal{F}_K) < \infty \) for every \( \delta > 0 \) and \( K > 0 \) where \( \mathcal{F}_K = \{f 1_{[F \leq K]} : f \in \mathcal{F} \} \).

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

Then

\[
D_n = \sup \{ |(P_n - P)(f)| : f \in \mathcal{F} \}
\]

\[
= \|P_n - P\|_F \to_{a.s.} 0 \quad \text{uniformly in } P \in \mathcal{P} \text{ as } n \to \infty;
\]

February 19, 1987
that is, for every \( \varepsilon > 0 \)
\[
\sup_{P \in \mathcal{P}} \Pr_{P} \left\{ \max_{m \geq n} D_m \geq \varepsilon \right\} \to 0 \quad \text{as } n \to \infty.
\]

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

Chung's (1951) theorem 0 is used heavily in the proof of theorem 1, so it not surprising that it is still contained therein; by taking \( F = \{ \text{id} \} \), theorem 1 reduces to theorem 0. Theorem 1 also contains the (uniform in all \( P \) ) Glivenko - Cantelli theorem due to Vapnik and Chervonenkis (1971) and Steele (1978).

**Corollary 1.** (Vapnik - Chervonenkis, 1971; Steele, 1978). Suppose that \( C \) is a permissible Vapnik - Chervonenkis class of subsets of \( A \) and set
\[
D_n(C) = \| P_n - P \|_F \quad \text{with } F = \{ 1_C : C \in C \}.
\]
Then \( D_n(C) \to_{a.s.} 0 \) as \( n \to \infty \) uniformly in \( P \in M = \{ \text{all } P \} \):
\[
\sup_{P \in M} \Pr_{P} \left\{ \max_{m \geq n} D_m(C) \geq \varepsilon \right\} \to 0 \quad \text{as } n \to \infty.
\]

The following corollary of theorem 1 will play an important role in the proofs of uniform weak convergence theorems for the empirical process.

**Corollary 2.** Suppose \( F \) is permissible and that:
(i) \( N_F^0(\delta,F) < \infty \) 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
\[
D_n^* = \sup \{ \| (P_n - P)(f - g) \| : f, g \in F \} \to_{a.s.} 0 \quad \text{uniformly in } P \in \mathcal{P} \text{ as } n \to \infty
\]
and
\[
D_n^{**} = \sup \{ \| (P_n - P)(f - g)^2 \| : f, g \in F \} \to_{a.s.} 0 \quad \text{uniformly in } P \in \mathcal{P} \text{ as } n \to \infty.
\]

Corollary 2 will be used to infer uniform total boundedness of a collection of functions \( F \).

**Definition 2.** For \( P, Q \in \mathcal{P} \) define the pseudo-metric \( \rho \) by
\[
\rho(P,Q) = \sup_{f, g \in \mathcal{F} \cup \{1\}} \| P - Q (f - g)^2 \|.
\]

The pseudo-metric \( \rho \) can be thought of as similar to the dual-bounded Lipschitz metric; see e.g. Dudley (1976) for this metric.
Uniform in $P$ Donsker theorems or weak convergence

Let $P$ be a collection of probability measures on the space $(A,A)$, and let $F$ be a collection of $A$-measurable functions defined on $A$ with envelope function $F$; i.e. $|f| \leq F$ for each $f \in F$. Assume that $F \cup \{F\} \subseteq L_2(P)$ for each $P \in P$. Let $\{P_n\}$ be a sequence in $P$. Suppose that $X_1, \cdots , X_\infty$ are row independent, iid $P_n$, ($A$-valued) random variables. Assume that the triangular array is defined on a common probability space $(\Omega, \Sigma, \text{Pr})$. 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

\[ P_n(A) = n^{-1} \sum_{i=1}^n 1_A(X_i) \quad \text{for } A \in A \]

and

\[ X_n = \sqrt{n} (P_n - P) \]

Thus

\[ X_n(f) = \int f \, dX_n = \sqrt{n} \int f \, d(P_n - P), \quad f \in F. \]

If

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

then the finite dimensional distributions of the processes $X_n$ converge to those of the mean zero Gaussian process $X$ with covariance

\[ E X(f)X(g) = \int f g \, dP_0 - \int f \, dP_0 \int g \, dP_0 \quad \text{for } f, g \in F. \]

Define $X$ to be the space of all bounded continuous real functions on $F$ (with the $L_2(P_0)$ topology), and define $C(F,P_0)$ to be the set of all functions $x(\cdot)$ in $X$ that are uniformly continuous with respect to the $L_2(P_0)$ seminorm. Suppose that $X$ is equipped with the uniform norm $\|x\| = \sup_{f \in F} |x(f)|$. The $\sigma$-field $\mathcal{B}_{P_n}$ is defined to be the smallest $\sigma$-field which:

(i) Contains all closed balls with centers in $C(F,P_0)$.

(ii) Makes all the finite dimensional projections measurable.

In the following we will again assume that $F$ is a "permissible" in the terminology of Pollard (1984) -- collection of functions, and also that $F$ has a countable dense subset in the $L_2(P_0)$ seminorm.
Theorem 2. Suppose that:

(i) \( F \) is permissible and totally bounded in the \( L_2(P_0) \) seminorm.

(ii) \( \limsup_{n \to \infty} E_P F^2 \mathbb{1}_{\{ |f| \geq \lambda \}} \to 0 \) as \( \lambda \to \infty \).

(iii) \( \rho(P_n, P_0) = \sup_{f, g \in F \cup \{1\}} |(P_n - P_0)(f - g)^2| \to 0 \) as \( n \to \infty \).

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

\[
\limsup_{n \to \infty} \frac{\sup_{P \in \mathbb{P}} \{ \sup_{X_n(f - g)} > \eta \}}{n} < \varepsilon
\]

where \( \delta_n = \{ (f, g) : f, g \in F \text{ and } \|f - g\|_{L_2(P)} < \delta \} \).

Then \( X_n \Rightarrow X_0 \) as random elements of \( X \) (in the sense of weak convergence of random elements defined by Pollard (1984) page 65), and \( X_0 \) is a tight Gaussian random element of \( X \) whose sample paths all belong to \( C(F, P_0) \).

Note that (ii) holds if \( \limsup_{n \to \infty} E_P F^{2+\delta} < \infty \) for some \( \delta > 0 \).

Hypothesis (iv) of theorem 2 is implied by easily checked conditions. Pollard's (1982) entropy condition together with \( P_* \)-uniform integrability of the envelope function \( F \) suffice. In fact, the following theorem yields (iv) uniformly in \( P \in \mathbb{P} \) if \( F \) is \( P \)-uniformly square integrable.

Theorem 3. Suppose \( F \) is permissible and that:

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

(ii) \( \limsup_{n \to \infty} \int F^2 \mathbb{1}_{\{ |f| \geq \lambda \}} dP = 0 \).

Then for every \( \varepsilon > 0 \) there is a \( \delta = \delta(\varepsilon) > 0 \) for which

\[
\limsup_{n \to \infty} \frac{\sup_{P \in \mathbb{P}} \{ \sup_{X_n(f - g)} > \varepsilon \}}{n} < \varepsilon
\]

where \( \delta_P = \{ (f, g) : f, g \in F \text{ and } \|f - g\|_{L_2(P)} < \delta \} \).

Here is an immediate corollary of theorems 2 and 3:

Corollary 3. Suppose \( F \) is permissible and that:

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

(ii) \( \limsup_{n \to \infty} \int F^2 \mathbb{1}_{\{ |f| \geq \lambda \}} dP_n = 0 \).

(iii) \( \rho(P_n, P_0) \to 0 \) as \( n \to \infty \) where \( \rho(P_n, P_0) = \sup_{f, g \in F \cup \{1\}} |(P_n - P_0)(f - g)^2| \).

Then (iv) of theorem 2 holds and \( X_n \Rightarrow X_0 \) as random elements of \( (X, B_{P_0}) \).
In keeping with Pollard (1982), we say that a collection of functions $\mathbf{F}$ with envelope function $F$ is \textit{sparse} if condition (i) of theorem 3 or corollary 3 holds. A further corollary is obtained by specializing corollary 3 to Vapnik-Chervonenkis classes of sets:

\textbf{Corollary 4.} Suppose that $C$ is a permissible Vapnik-Chervonenkis class of subsets of $A$, let $\mathbf{F} = \{1_{C} : C \in C\}$, and suppose that $\{P_{n}\}$ satisfies

\begin{equation}
\rho(P_{n},P_{o}) = \sup_{B,C \in C} |P_{n}(B \cap C) - P_{o}(B \cap C)| \to 0
\end{equation}

as $n \to \infty$. Then $\mathbf{X}_{n} \Rightarrow \mathbf{X}_{0}$ as random elements of $(X,B_{P})$.

Corollary 3 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 $Q$:

\begin{equation}
H^{2}(P,Q) = \int (\frac{dP}{d\mu})^{1/2} - (\frac{dQ}{d\mu})^{1/2} d\mu
\end{equation}

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

\textbf{Definition 3.} \{\hat{P}_{n}\} = \{\hat{P}_{n}(f) : f \in F\} is a \textit{sequentially regular estimator} of $\{P\} = \{P(f) : f \in F\}$ on $P$ if $\mathbf{X}_{n} = \sqrt{n}(\hat{F}_{n} - P_{n}) \Rightarrow \mathbf{X}$ for every sequence $\{P_{n}\} \subset P$ with $H(P_{n},P_{o}) \to 0$ where the (distribution of the) limit process $\mathbf{X}$ depends on $P_{o}$ but not on the sequence $\{P_{n}\}$.

\textbf{Corollary 5.} Suppose that $F$ is permissible and that $F$ and $P$ satisfy (i) and (ii) of corollary 3. Then $H(P_{n},P_{o}) \to 0$ implies that $\rho(P_{n},P) \to 0$ and $P_{n}$ is a sequentially regular estimator of $\{P\}$ on $P$.

\textbf{Convergence of the "bootstrapped empirical process"}

The preceding theorems imply convergence of the "bootstrapped empirical process" very generally. Suppose that $X_{1}, \cdots, X_{n}$ are iid $P_{o}$ on $(A,A)$, and let

\begin{equation}
P_{n} = n^{-1} \sum_{i=1}^{n} \delta_{X_{i}}
\end{equation}

be the empirical measure of the $X_{i}$'s. Suppose further that $X_{1}^{*}, \cdots, X_{m}^{*}$ are iid $P_{n}$, let

\begin{equation}
P_{n}^{*} = m^{-1} \sum_{i=1}^{m} \delta_{X_{i}^{*}},
\end{equation}

and set

\begin{equation}
\mathbf{X}_{m}^{*} = \sqrt{m}(P_{n}^{*} - P_{n}).
\end{equation}

Thus $X_{1}^{*}, \cdots, X_{m}^{*}$ is the "bootstrap sample", $P_{n}^{*}$ is the "bootstrap empirical
measure", and \( \mathbf{X}_m^* \) is the "bootstrap empirical process". Since \( P_n \) converges to \( P_0 \) a.s., it is clear that corollary 1 may be applied to deduce weak convergence (a.s. conditional on \( P_n \)) of \( \mathbf{X}_m^* \) whenever \( m = m(n) \to \infty \). The following theorems generalize and extend the results of Bickel and Freedman (1981), Shorack (1982) (see Shorack and Wellner (1986) chapter 23), Gaenssler (1985), Beran and Millar (1986), and Beran, Le Cam, and Millar (1985).

**Theorem 4.** Suppose \( F \) is permissible and that:

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

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

If \( m = m(n) \to \infty \), then \( \mathbf{X}_m^* \Rightarrow \mathbf{X}_0^* \equiv \mathbf{X}_0 \) as random elements of \( (X,B_{P_0}) \) a.s. conditional on \( P_n \).

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

**Theorem 5.** Suppose that \( F \) is permissible and that (i) - (iii) of corollary 3 hold. If \( m = m(n) \to \infty \), then \( \mathbf{X}_m^* \Rightarrow \mathbf{X}_0^* \equiv \mathbf{X}_0 \) as random elements of \( (X,B_{P_0}) \) a.s. conditional on \( P_n \).

**Remarks:**

1. Corollary 3 can be viewed as extending some of the results of Le Cam (1983). While Le Cam allows independent but non-identically distributed \( X_{ni} \)'s, his functions \( f \in F \) must be bounded. Our results treat only the iid case (with \( P_{ni} = P_n \) changing from row to row), but allow unbounded classes \( F \).

2. Since the collection of half-spaces in \( \mathbb{R}^d \) is a Vapnik-Chervonenkis (1971) class (see also Dudley (1979)), corollary 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.

3. The notion of sequential regularity given in definition 3 is stronger (i.e. involves more regularity) than the local regularity required by Beran (1977), Begun et al. (1983), or Millar (1985) as a hypothesis for their convolution theorems for (locally) regular estimators. Local regularity would typically involve that the limit process \( \mathbf{X} \) be the same for all sequences \( \{P_n\} \) obtained from a regular parametric submodel \( \{P_{\theta}: \theta \in \Theta\} \) of \( P \) (for which \( F \) is uniformly integrable) via \( P_n = P_{\theta_n} \) and \( P_0 = P_\theta \) with \( \theta_n = \theta + h_n^{-1/2} \). This of course implies the (more restrictive) condition
4. Dudley (1986) has shown that Pollard's entropy condition, hypothesis (i) of theorem 3 is sufficient for \( F \) to be a "universal Donsker class": \( \mathbb{X}_n \Rightarrow \mathbb{X} \) for every (fixed) \( P \) on \((\mathcal{A}, \mathcal{A})\). Because Dudley (1986) insists on convergence for every \( P \), his collections of functions \( F \) are forced to be bounded. Theorem 3 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 \( P \in \mathcal{P} \).

5. It would be interesting to have a complete "uniform in \( P \)" version of theorem 5.2, page 158 of Dudley (1985). Our theorem 3 gives sufficient conditions for a "uniform in \( P \)" version of part 2 of Dudley's theorem. In particular, we conjecture that the conclusion of our theorem 3 is equivalent to the "uniform in \( P \)" version of part 3 of Dudley's theorem 5.2. A result of this type would yield Lai's (1978) "uniform in \( P \)" strong approximation of partial sum processes as a corollary in the same way that Chung's theorem 0 is still contained in our theorem 1.

2. Examples and further applications

In section 1 we have already given one major application of our results: the bootstrapped empirical process behaves correctly under the sparseness assumption on \( F \) and a square integrability assumption on the envelope function \( F \). Here we give further examples to illustrate the usefulness of the theorems of section 1 as tools for proving the regularity of estimates and for establishing power approximations for tests.

Estimation

We often want to consider estimation of some other function or parameter \( v(P) \), where \( v: P \rightarrow M \) with \((M, d)\) some metric space, instead of \( P \) itself. Just as in our definition 1.3 of a regular estimator of \( P \), we would like our estimator, often \( v(P_n) \), to be "regular". First, consider the case of a fixed \( P \), say \( P_0 \).

When \( X_1, \cdots, X_n \) are iid \( P_0 \), let \( \mathbb{X}_n^0 \) denote the corresponding empirical process

\[
\mathbb{X}_n^0 = \sqrt{n} (P_n - P_0).
\]

To show weak convergence or asymptotic normality of \( \sqrt{n} (v(P_n) - v(P_0)) \), we usually show that \( v \) is asymptotically linear:

\[
\sqrt{n} (v(P_n) - v(P_0)) = \mathbb{X}_n^0(v(\cdot, P_0)) + o_p(1)
\]

(2)

The function \( \hat{v}(\cdot, P_0) \) is called the influence function of \( v \). If (2) holds, weak convergence of the left side follows from the ordinary (multivariate) central limit

February 19, 1987
If $M$ is $R^1$, then $(M,d)$ is some collection of functions, say $(L_∞(T), \| \cdot \|_∞)$, and we typically have

(i) $\nu(y,P_0) \in M = L_∞(T)$ for fixed $y \in A$.

(ii) $\nu(\cdot,P_0)(t) \in L_2(P_0)$ for each fixed $t \in T$.

Thus the key first term in (2) is just the empirical process $\mathbf{X}_n^0$ indexed by the collection of functions

\[ (3) \quad F_0 = \{ \nu(\cdot,P_0)(t) : t \in T \}. \]

If $F_0$ is sparse and has envelope function $F \in L_2(P_0)$, then Pollard’s (1982) theorem 5 yields convergence of (2), and we conclude that

\[ (4) \quad \frac{1}{n} (\nu(P_n) - \nu(P_0)) \Rightarrow \mathbf{X}_0(\nu(\cdot,P_0)). \]

To examine regularity of the estimator $\nu(P_n)$, let $X_n, \cdots, X_n$ be iid $P_n$ with $H(P_n,P_0) \to 0$, and let $\mathbf{X}_n$ be the empirical process of (1.12). Then we can often show that the asymptotic linearity of (3) continues to holds in the sense that

\[ (5) \quad \frac{1}{n} (\nu(P_n) - \nu(P_0)) = \mathbf{X}_n(\nu(\cdot,P_n)) + o_{P}(1). \]

To establish convergence of the left side in (5), we would like to replace $\nu(\cdot,P_n)$ by $\nu(\cdot,P_0)$ on the right side and then apply our corollary 3 to $\mathbf{X}_n(\nu(\cdot,P_n))$ with $\nu(\cdot,P_0)(t) \in F_0$ of (3). To make this replacement it is often convenient to consider an appropriate large collection $P$ of $P$’s (including $P_0$ and $\{P_n\}$) and get rid of the dependence of $\nu$ on $P$ by simply unioning over $P \in P$. Thus we let

\[ (6) \quad F = \{ \nu(\cdot,P)(t) : t \in T, \ P \in P \}. \]

If we show that $F$ is sparse with an envelope function $F$ that is $P$-uniformly square integrable, then corollary 3 holds (note that $H(P_n,P_0) \to 0$ implies $\rho(P_n,P_0) \to 0$ by corollary 5), so that (iv) of theorem 2 yields

\[ (7) \quad \frac{1}{n} (\nu(P_n) - \nu(P_0)) = \mathbf{X}_n(\nu(\cdot,P_n)) + o_{P}(1) \]

if we show that

\[ (8) \quad \sup_{t \in T} \| \nu(\cdot,P_n)(t) - \nu(\cdot,P_0)(t) \|_{L_2(P)} \to 0. \]

Then corollary 3 also yields convergence of the first term in (7) to $\mathbf{X}_0(\nu(\cdot,P_0))$ and we conclude that $\nu(P_n)$ is regular. We record this argument in the following theorem; its application will become clear in the examples which follow.

**Theorem 6.** Suppose that $P$ is a collection of probability distributions on $(A,A)$, and that $\{P_n\}_{n \geq 0}$ is a sequence in $P$ satisfying $H(P_n,P_0) \to 0$. Suppose moreover that:

February 19, 1987
(i) $v(P_n)$ is asymptotically linear: (5) holds.

(ii) $F$ defined in (6) is sparse with $P$ uniformly square integrable envelope function $F$.

(iii) $v(\cdot, P)(t)$ is $L_2(P)$ continuous uniformly in $t \in T$; i.e. (8) holds.

Then $v(P_n)$ is a regular estimator of $v(P)$ on $P$:

$$\sqrt{n} (v(P_n) - v(P)) \Rightarrow X^0(\nu(\cdot, P_0)).$$

The following two examples illustrate the application of theorem 6. Applications to constrained estimation problems are given by Sheehy (1987).

**Example 1.** (Length biased sampling). Suppose that $X_1, \cdots, X_n$ are iid $P$ on $R^+ = [0, \infty)$ where the df corresponding to $P$ is given by

$$P(-\infty, x] = \frac{1}{\mu(G)} \int_{[0,x]} y \, dG(y) = \frac{\int_{y} y \, dG(y)}{\int_{0} y \, dG(y)}$$

where $G$ is a df on $R^+$ with

$$0 < \mu(G) = \int_{0} ^{\infty} y \, dG(y) < \infty;$$

$P(-\infty, \cdot]$ is the length biased distribution corresponding to $G$. Note that

$$G(t, P) = \frac{\int_{0} ^{\infty} y^{-1} \, dP(y)}{\int_{0} ^{\infty} y^{-1} \, dP(y)}.$$

Here we are interested in estimating the function $v(P) = G(\cdot, P)$. In view of (12), a natural estimator of $G$ based on the sample from the length biased distribution $P$ is

$$G_n(t) = \frac{\int_{[0,t]} y^{-1} \, dP_n(y)}{\int_{[0,\infty)} y^{-1} \, dP_n(y)} = \hat{\mu}_n \int_{[0,t]} y^{-1} \, dP_n(y)$$

with

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

Consider the process

$$Z_n(t) = \sqrt{n} (G_n(t) - G(t))$$

$$= \hat{\mu}_n \int_{0} ^{t} y^{-1} \{ 1_{y \leq t} - G(t) \} \, d\{\sqrt{n} (P_n - P)(y)\}$$

$$= \hat{\mu}_n X_n(f_t)$$

with

$$f_t(y) = f_t(y, P) = y^{-1} \{ 1_{y \leq t} - G(t, P) \}.$$
The estimator $E_n$ was introduced by Cox (1969), and, for a fixed $P$, the convergence of $Z_n$ was established by Vardi (1983) as a special case of a two-sample biased sampling problem. The limit process is

$$Z(t) = \mu X(f_t), \quad t \geq 0,$$

with covariance function

$$\text{Cov}[Z(s), Z(t)] = \mu \int_0^t [1_{[0,t]}(s) - G(s)][1_{[0,t]}(t) - G(t)] \frac{1}{x} dG(x).$$

The convergence of Vardi's (1985) estimators for an $s$-sample biased sampling problem and the corresponding processes $Z_n$ have been established by Gill and Wellner (1986) for a fixed $P$. Theorem 6 can be used to establish the regularity of Vardi's (1985) estimators under mild hypotheses. Here we illustrate the application of theorem 6 by considering $E_n$ and $Z_n$ for the just the special case of length biased sampling.

Let $P$ be any collection with $y^{-1} P$ uniformly square integrable:

$$\lim_{\lambda \to -\infty} \sup_{x \in P} \int_{y^{-1} \geq x} y^{-2} dP(y) = 0.$$

Set

$$F = \{ f_i(\cdot, P) : t \geq 0, P \in P \}.$$

for the functions $f_i(\cdot, P)$ in (15). Note that $F$ depends on $P$ through $G$, but in a quite simple way since $0 \leq G(t, P) \leq 1$ as $P$ varies.

Furthermore, $F$ has envelope function $F(y) = y^{-1}$ and that $F$ is sparse: the first term is of the form $\{ F_{1C} : C \in C \}$ where $C$ is a Vapnik-Chervonenkis class, so it is sparse by Pollard's (1982) theorem 9; the second term is $\{ G(t)F : t \geq 0 \}$, which is sparse by Pollard's (1982) theorem 10 since $0 \leq G(t, P) \leq 1$. Hence $F$ is sparse (again by Pollard's (1982) theorem 10), and in view of (19), (ii) of theorem 6 holds.

Suppose that $X_{n1}, \ldots, X_{nn}$ are iid $P_n$ with $\{ P \} \cup \{P_n\} \subset P$ and $H(P_n, P) \to 0$. By straightforward algebra it follows that

$$Z_n(t) = \sqrt{n} (E_n(t) - G_n(t)) = \hat{\mu}_n \sqrt{n} (P_n - P_n)(f_i(\cdot, P_n)) \quad \text{for} \quad t \geq 0,$$

and hence (5) (hypothesis (i) of theorem 6) holds. To show regularity of the estimator $E_n$, we want to show that

$$Z_n \Rightarrow Z \quad \text{of (17) under} \quad P_n.$$

We do this by verifying the remaining hypothesis (iii) of theorem 6.

February 19, 1987
To see that (iii) holds, note that with $G_n = G(\cdot, P_n)$ we have

$$\sup_{t \geq 0} \left| \int f_t(y, P_n) - f_t(y, P) \right|^2 dP_n(y)$$

$$\leq \sup_{t \geq 0} \left| G_n(t) - G(t) \right| \int F^2(y) dP_n(y)$$

$$= o(1) O(1) = o(1)$$

by an easy argument using (12), (19), and $H(P_n, P) \to 0$ (which implies that $d_{TV}(P_n, P) \to 0$).

Thus theorem 6 applies, and when combined with $\mu_n \to \mu(G)$ under $P_n$ (which follows easily from theorem 0 by use of (19)), we conclude that (22) holds; $G_n$ is a regular estimator of $G$.

**Testing**

In many typical testing problems we have a test statistic $T_n$ which can be expressed as

$$T_n = g(\mathbf{X}_n^0)$$

where $g$ is a continuous function of the "null hypothesis empirical process"

$$\mathbf{X}_n^0 = \sqrt{n} (P_n - P_0).$$

Convergence of $\mathbf{X}_n^0$ to a process $\mathbf{X}^0$ under the null hypothesis $P_0$ together with continuity of $g$ yields a large-sample approximation of the distribution of $T_n$ which often allows us to implement the test in practice. For this much, convergence of $\mathbf{X}_n^0$ under a fixed $P = P_0$ suffices, so the results of Dudley (1978), or Pollard (1982), or others for a fixed $P_0$ may be applied.

Under an alternative hypothesis $P \neq P_0$, and now $T_n = g(\mathbf{X}_n^0)$ where

$$\mathbf{X}_n^0 = \sqrt{n} (P_n - P) + \sqrt{n} (P - P_0)$$

$$= \mathbf{X}_n + \delta_n.$$  

To get nondegenerate limiting power we assume that $P = P_n$ depends on $n$ in such a way that $\delta_n = \sqrt{n} (P_n - P_0)$ converges to a fixed (signed) measure $\delta$. Then our corollary 3 (after checking (iii)) handles convergence of $\mathbf{X}_n = \sqrt{n} (P_n - P_n)$ under $P_n$. The upshot is that

$$T_n \to_d g(\mathbf{X}^0 + \delta) \quad \text{under} \quad P_n,$$

and this yields an approximation for the (local asymptotic) power of the test.

Here is a simple example.

**Example 2.** (Power of a goodness of fit test). Suppose that $X_1, \ldots, X_n$ are iid $P$ on $R^k$. Consider testing $H_0: P = P_0$ versus $P \neq P_0$ where $P_0$ is known. Consider
the test statistic
\[ T_n = \iint n \{ P_n(x, y) - P(x, y) \}^2 dP_0(x) dP_0(y) \]
\[ = \iint \{ X^0_n(x, y) \}^2 dP_0(x) dP_0(y) \]
\[ = g(X^0_n) \tag{23} \]

where \( P_0(x, y) = P_0(x < X \leq y) \) and \( P_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{x, y\} = X_i} \) so that

\[ X^0_n(x, y) = \sqrt{n} \int 1_{\{x, y\}} d(P_n - P) \tag{24} \]
is the empirical process indexed by the indicators of rectangles, 
\[ F = \{ 1_C : C \in \mathcal{C} = \text{all rectangles in } R^k \}. \]

Since \( \mathcal{C} \) is a Vapnik-Chervonenkis class of sets and \( g \) defined by (23) is continuous, it follows from Dudley (1978) or Pollard (1982) (or indeed from Dudley (1966)) that under the null hypothesis we have

\[ T_n \rightarrow d T_0 = \iint \{ X^0(x, y) \}^2 dP_0(x) dP_0(y) \quad \text{under } P_0 \]

where \( X^0 \) has covariance function

\[ \text{Cov}[X^0(x_1, y_1), X^0(x_2, y_2)] = P_0(R_1 \cap R_2) - P_0(R_1)P_0(R_2) \]

with \( R_i = (x_i, y_i) \) for \( i = 1, 2 \). Hence

\[ \Pr_0(T_n \geq t) \rightarrow Pr(T \geq t) \quad \text{as } n \rightarrow \infty. \tag{25} \]

Of course the distribution of \( T \) on the right side of (25) depends on \( P_0 \) for \( k \geq 2 \) (and for \( k = 1 \) unless the corresponding df \( F_0 \) is continuous), so implementation of the test requires separate tabulation of the distribution on the right side of (25) for any given \( F_0 \), or, alternatively, bootstrapping (monte carlo sampling from \( P_n \)) or monte carlo sampling from \( P_0 \). In the two- or r-sample versions of the problem, the test could be implemented as a permutation test. For tables in the case of \( P_0 \) uniform on \([0, 1]^k \) and rectangles replaced by lower left orthants, see Cotterill and Csorgo (1982).

Our results in section 1 yield approximations to the power of this type of test. Suppose that \( P_n \) is a sequence of local alternatives to \( P_0 \) in the sense that

\[ \| \sqrt{n} (P_n - P_0) - \delta \|_F \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{26} \]

Then we can write the process \( X^0_n \) in (24) as

\[ X^0_n = X_n + \sqrt{n} (P_n - P_0) \tag{27} \]

where convergence of the process \( X_n = \sqrt{n} (P_n - P_0) \) (as in (1.12)) is handled easily under \( P_n \) by our corollary 4. The conclusion from corollary 4, (26), and (27) is that under \( P_n \)

February 19, 1987
and, since \( g \) is continuous

\begin{equation}
T_n \rightarrow_d T(\delta) = \int \int \{X^0(x, y) + \delta(x, y)\}^2 dP_0(x) dP_0(y) \quad \text{under } P_n
\end{equation}

Hence, with \( t = c(F_0, \alpha) \) on the right side in (25) chosen so that the probability of type I error is \( \alpha \),

\begin{equation}
\text{Power}(T_n, P_n) = Pr_{P_n}(T_n \geq c(F_0, \alpha)) \rightarrow Pr(T(\delta) \geq c(F_0, \alpha)).
\end{equation}

Power approximations for other weighted versions of \( T_n \) analogous to the classical Anderson Darling statistic and the statistic \( T_n^2 \) of Shorack and Wellner (1986) page 627 are also easily treated using our results.

Similar considerations apply to a wide variety of other tests in both one and several sample problems, e.g. the chi-square tests with random cells as in Pollard (1979) or Moore (1971), the minimum distance tests of Pollard (1980), or the Kolmogorov type tests and confidence regions based half - spaces of Beran and Millar (1986).

An alternative approach to proving (28) or (29) is via contiguity theory; see e.g. Neuhaus (1974).

3. Proofs
We begin with some preliminary results.

Proposition 1. Suppose that
(i) \( P \) is totally bounded in the \( \rho \) pseudo-metric.
(ii) \( H^{(2)}(\delta, F) < \infty \) for every \( \delta > 0 \).
(iii) \( F \in L^2(P) \) for each \( P \in \mathcal{P} \).

Then \( F \) is totally bounded in the \( L^2(P) \) pseudo-metric uniformly in \( P \in \mathcal{P} \); i.e. for every \( \epsilon > 0 \) there exist \( f_1, \cdots, f_k \) such that \( \min_{1 \leq i \leq k} \|f-f_i\|_{L^2(P)} < \epsilon \) for all \( P \in \mathcal{P} \).

Proof. Pollard (1982) shows that \( F \) is totally bounded in \( L^2(P) \) for each \( P \in \mathcal{P} \); i.e. given any \( \epsilon > 0 \) there exists a finite collection of functions \( F_{\epsilon, P} \) so that for any \( f \in F \)

\[
\min_{f' \in F_{\epsilon, P}} \|f - f'\|_{L^2(P)} < \epsilon.
\]

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

\[
\min_{f' \in F(\epsilon)} \|f - f'\|_{L^2(P_i)} < \epsilon, \quad \text{for } i = 1, \cdots, m.
\]

Now cover \( \mathcal{P} \) by \( \rho \) - balls of radius \( \epsilon^2 \) with centers \( \{P_1, \cdots, P_m\} \) and let \( F(\epsilon) \) be the finite collection of functions defined above. If \( P \in \mathcal{P} \), then \( \rho(P, P_i) < \epsilon^2 \) for some

February 19, 1987
i = 1, \ldots, m \text{ and if } f \in F, \|f - f'\|_{E_2(P_i)} < \varepsilon \text{ for some } f' \in F(e), \text{ so that }

\|f - f'\|_{E_2(P_i)} \leq \|f - f'\|_{E_2(P_i)}^2 + \varepsilon^2 \leq 2\varepsilon^2.

Hence

$$\min_{f' \in F(e)} \|f - f'\|_{E_2(P_i)} \leq \sqrt{2} \varepsilon \quad \text{for all } P \in P;$$

i.e. F is totally bounded uniformly in $P \in P$. \(\square\)

Our proofs of theorem 1 and its corollaries depend on the following lemmas.

**Lemma 1.** If $N_f^{(d)}(\delta, F) < \infty$ and $F_K = \{f 1_{[F \leq K]} : f \in F\}$, then $N_f^{(d)}(\delta, F_K) \leq N_f^{(d)}(\delta, F) < \infty$.

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

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

where

$$m \leq N_f^{(d)}(\delta, S \cap [F(x) \leq K], F_K) \leq N_f^{(d)}(\delta, F) < \infty. \quad \square$$

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

**Proof.** We show, in fact, that

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

Let $\{f_1, \ldots, f_m\} = F_\delta$ 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 $m \leq N_f^{(d)}(\delta, F)$. 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 $n = \#(S)$,

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

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

February 19, 1987
Lemma 3. Suppose that \( F \) is permissible and has envelope function \( K \), a constant, and that \( N_F^{(1)}(\delta, F) < \infty \) for all \( \delta > 0 \). Then \( D_n = \| P_n - P \|_F \to_{a.s.} 0 \) as \( n \to \infty \) uniformly in \( P \in \mathcal{P} \) where \( \| P_n - P \|_F = \sup_{f \in F} |(P_n - P)(f)| \):

\[
\sup_{P \in \mathcal{F}} \text{Pr}_P \left\{ \max_{m \geq n} \| P_m - P \|_F > \varepsilon \right\} \to 0 \quad \text{as} \quad n \to \infty
\]

for every \( \varepsilon > 0 \).

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

\[
\sup_{P \in \mathcal{F}} \text{Pr}_P \left\{ \max_{m \geq n} \| P_m - P \|_F > \varepsilon \right\} < \varepsilon \quad \text{for} \quad n \geq n(\varepsilon).
\]

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

\[
(2) \quad n(\varepsilon) \geq \max \left\{ \frac{2K^2}{\varepsilon^2}, \frac{256K^2}{\varepsilon^2}, H_F^{(1)}(\varepsilon/8K, F), n(\varepsilon, K) \right\}
\]

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

\[
(3) \quad 8 \sum_{m=n(\varepsilon, K)}^{\infty} \exp\left(-\frac{m \varepsilon^2}{256K^2}\right) < \varepsilon.
\]

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 \( \sigma_1, \sigma_2, \ldots \) 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

\[
(4) \quad \sup_{P \in \mathcal{F}} \text{Pr}_P \left\{ \| P_n - P \|_F > \varepsilon \right\} \leq 4 \sup_{P \in \mathcal{F}} \text{Pr}_P \left\{ \| P_n^{0} \|_F > \frac{\varepsilon}{4} \right\}
\]

\[
= 4 \sup_{P \in \mathcal{F}} E_P \left[ \text{Pr}_P \left\{ \| P_n^{0} \|_F > \frac{\varepsilon}{4} \mid X_n \right\} \right]
\]

for \( n \geq 2K^2/\varepsilon^2 \); this depends on

\[
\sup_{P \in \mathcal{F}} \sup_{f \in F} \text{Pr}_P \left\{ |(P_n - P)(f)| > \varepsilon \right\}
\]

\[
\leq \sup_{P \in \mathcal{F}} \sup_{f \in F} \frac{E_P f^2}{n \varepsilon^2} \leq \frac{K^2}{n \varepsilon^2} \leq \frac{1}{2} \quad \text{for} \quad n \geq 2K^2/\varepsilon^2
\]

to obtain the factor of 4 on the right side. Given \( X_n \), choose functions \( g_1, \ldots, g_M \), \( M = N_F^{(1)}(\varepsilon/8K, F) \) so that

\[
\min_{j} P_n |f - g_j| \leq \frac{\varepsilon}{8} \quad \text{for each} \quad f \in F.
\]

February 19, 1987
Write $f^*$ for a $g_j$ at which the minimum is achieved. For any function $g$

$$ |P_n g| = |n^{-1} \sum_{i=1}^{n} \sigma_i g(X_i)| \leq n^{-1} \sum_{i=1}^{n} |g(X_i)| = P_n|g| . $$

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),

$$ \Pr \{ \sup \{\sup |P_n^0 f_j > \epsilon \} \leq \Pr \{ \sup \{ |P_n^0 f^*| + P_n|f - f^*| \} > \epsilon 4 |X_n | \} $$

$$ \leq \Pr \{ \max |P_n^0 g_j| > \epsilon 8 |X_n | \} \quad \text{because } P_n|f - f^*| \leq \epsilon 8 $$

$$ \leq N_K(\epsilon 8 K, F) \max \Pr \{ |P_n^0 g_j| > \epsilon 8 |X_n | \} . $$

$$ \leq N_K(\epsilon 8 K, F) \max \Pr \{ |\sum_{i=1}^{n} \sigma_i g_j(X_i)| > n\epsilon 8 |X_n | \} $$

$$ \leq 2 \exp \{ H_K(\epsilon 8 K, F) - 2(n\epsilon 8)^2/\sum_{i=1}^{n} (g_j(X_i))^2 \} \quad \text{by Hoeffding's inequality} $$

(e) $$ \leq 2 \exp \{-n \epsilon^2/256 K^2 \} \quad \text{using } |g_j| \leq K \text{ and (b)} . $$

Combining (d) and (e) yields, for $n \geq n(e)$,

(f) $$ \sup_{P \in \mathcal{P}} \Pr \{ \max_{m \geq n} \|P_m - P\|_F > \epsilon \} $$

$$ \leq 4 \sum_{m=n}^{\infty} \sup_{P \in \mathcal{P}} \Pr \{ \|P_m^0\|_F > \epsilon 4 \} $$

$$ \leq 8 \sum_{m=n}^{\infty} \exp \left( - \frac{n \epsilon^2}{256 K^2} \right) < \epsilon , $$

and hence (a) holds.

With these three lemmas as preparation we can now prove theorem 1.

**Proof of theorem 1.** In view of (ii) we can choose $K$ so large that

(a) $$ \sup_{P \in \mathcal{P}} \Pr \{ F_1 |F > K| \} < \frac{\epsilon}{4} . $$

Then, since

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

$$ \leq |P_m - P|_{F_K} + P_m F 1_{|F > K|} + P F 1_{|F > K|} $$

$$ \leq |P_m - P|_{F_K} + |(P_m - P) F 1_{|F > K|}| + 2 P F 1_{|F > K|} $$

$$ \leq |P_m - P|_{F_K} + |(P_m - P) F 1_{|F > K|}| + \frac{\epsilon}{2} \quad \text{by (a)} . $$

February 19, 1987
it follows that
\[ \sup_{P \in \mathcal{P}} \Pr_P \left\{ \max_{m \geq n} \| P_m - P \|_F > \varepsilon \right\} \]
\[ \leq \sup_{P \in \mathcal{P}} \Pr_P \left\{ \max_{m \geq n} \| P_m - P \|_F > \frac{\varepsilon}{4} \right\} \]
\[ + \sup_{P \in \mathcal{P}} \Pr_P \left\{ \max_{m \geq n} |(P_m - P) 1_{[F > K]}| > \frac{\varepsilon}{4} \right\} \]
\[ \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \quad \text{for} \quad n \geq \text{some} \ N(\varepsilon) \]

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

**Proof of corollary 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 \( F = K = 1 \); see e.g. Pollard (1982, 1984) or Dudley (1984, 1986). Condition (ii) holds trivially. Thus (1.7) follows from theorem 1. □

**Proof of corollary 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 1. By lemma 2 this implies that \( N_{K^{(1)}(\delta, F_K)} < \infty \) for all \( \delta > 0 \) where \( F_K^* = \{ f g 1_{[F^* \leq K]} : f, g \in F \} \). Hence the conditions of theorem 1 are satisfied with

\[ F^* = \{ f g : f, g \in F \} \]

replacing \( F \). Thus (1.8) holds. Since \( D_{n^*}^* \leq 4D_{n^*}^*, \) (1.9) follows immediately from (1.8). □

**Proof of theorem 2.** The proof is a slight variation on Pollard's (1984) proof of his theorem 21, page 157, so we omit it. The key difference is the use of condition (iii) to argue that \( [\delta/2] \leq [\delta] \) for \( n \) large so that by (iv) we have: given any \( \eta, \varepsilon > 0 \) there exists \( \delta > 0 \) for which

\[ \limsup_{n \to \infty} \Pr_P \left\{ \sup_{[\delta/2]_n} |X_n(f - g)| > \eta \right\} < \varepsilon. \]

The remainder of the argument is just the same Pollard's proof. □

**Proof of theorem 3.** Our proof of theorem 3 is a modification of Pollard's (1982) proof of his theorem 7. None of the basic techniques are new, but in order to clearly show how (i) and (ii) yield (14), we give the proof in detail.

First note that (ii) implies that there is a constant \( M < \infty \) so that

\[ \sup_{P \in \mathcal{P}} \| F \|_{L_2(P)}^2 = \sup_{P \in \mathcal{P}} E_P F^2 \leq M < \infty. \]
This constant enters repeatedly in the remainder of the proof.

Now let \( \delta_j = 2^{-j} \) for \( j \geq 1 \), and set \( H_j = H_{(j)}(2^{-j}, F) \) 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

(b) \( \sum_{j=1}^{\infty} \eta_j < \infty \),

(c) \( \eta_j \geq (144M \delta_j^2 H_j)^{1/2} \), \( \text{so that } H_j \leq \frac{\eta_j^2}{144M \delta_j^2} \),

(d) \( \sum_{j=1}^{\infty} \exp\left(-\frac{\eta_j^2}{72 \delta_j^2 M}\right) < \infty \).

This is possible because of the growth condition (i) on \( H_{(j)}(\cdot, F) \). For example, \( \eta_j = \max\{f \delta_j, (144\|F\|^2 \delta_j^2 H_j)^{1/2}\} \) works.

We now give our choices of \( \delta = \delta(\epsilon) > 0 \) and \( n = n(\epsilon, \delta) \) (not dependent on \( P \in P \)) which yield (14). Choose an integer \( r = r(\epsilon) \) so large that, with \( \eta = \epsilon/8 \),

(e) \( \sum_{j=r+1}^{\infty} \exp\left(-\frac{\eta_j^2}{72 M \delta_j^2}\right) < \frac{\epsilon}{16} \), \( \text{by (d)} \)

(f) \( \sum_{j=r+1}^{\infty} \eta_j < \frac{\epsilon}{16} \), \( \text{by (b)} \)

(g) \( 2 \exp\left(-\frac{\eta_r^2}{72 \delta_r^2 M}\right) < \frac{\epsilon}{16} \), \( \text{since } \delta_r \to 0 \)

(h) \( \eta_r^2 \geq 144M H_r \delta_r^2 \) \( \text{by condition (i)} \) \( \text{so that } H_r \leq \frac{\eta_r^2}{144M \delta_r^2} \)

all hold. Now choose \( \delta = \delta(\epsilon) > 0 \) so that

(i) \( \delta \leq \min\{\sqrt{\frac{2}{3}} \delta_r(\epsilon) M^{1/2}, \frac{\epsilon}{\sqrt{2}}\} \).

With this \( \delta = \delta(\epsilon) \) we choose \( n = n(\epsilon, \delta) \) so large that

(j) \( \sup_{P \in P} \text{Pr}_P \left\{ \sup_{f \neq g \in F} \|P_n - P\| f - g \| > 8^2/2 \right\} \leq \frac{\epsilon}{16} \)

and

(k) \( \sup_{P \in P} \text{Pr}_P \left\{ \|F\|_n^2 > \|F\|_{L_2(p)}^2 + M \right\} < \frac{\epsilon}{16} \);

such a choice is possible in view of corollary 2 and theorem 0 respectively.

To prove (14) we will show that this choice of \( \delta \) and \( n(\epsilon, \delta) \) imply that

(l) \( \sup_{P \in P} \text{Pr}_P \left\{ \sup_{[S]} |X_n(f - g)| > \epsilon \right\} \leq \epsilon \) \( \text{for all } n \geq n(\epsilon, \delta) \).

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

\[ (m) \quad \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} X_n(f - g) > \varepsilon \} \leq 4 \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} P_n^0(f - g) > \frac{\varepsilon}{4\sqrt{n}} \} \]

for all \( n \geq 1 \) where \( P_n^0 \) is the symmetrized empirical measure as in lemma 3. The validity of \( (m) \) depends on

\[ \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} (P_n - P)(f - g) > \varepsilon \} \leq \sup_{P \in \mathcal{P}} \sup_{[\delta]} E_P (f - g)^2 \leq \frac{\delta^2}{\varepsilon^2} \leq \frac{1}{2} \]

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

We use (j) to replace \([\delta]\) on the right side in \( (m) \) by \([3^{1/2}\delta/2^{1/2}]\) where \([\delta]\) = \{(f,g): f, g \in F \text{ and } \|f - g\|_n < \delta \} \text{ and } \|f\|_n^2 = \int f^2 dP_n \). Let \( D_n^{**} \) be defined as in (1.9), and let \( B_n^* \) denote the event \([\|f\|_n^2 > \|f\|_{\delta(P)}^2 + M]\) in (k). It follows from our choice of \( n \) and (j) and (k) that the right side of \( (m) \) is bounded by

\[ \frac{4}{\varepsilon} \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} P_n^0(f - g) > \frac{\varepsilon}{4\sqrt{n}}, D_n^{**} > \frac{\delta^2}{2} \} \]

\[ + \frac{4}{\varepsilon} \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} P_n^0(f - g) > \frac{\varepsilon}{4\sqrt{n}}, D_n^{**} \leq \frac{\delta^2}{2} \} \]

\[ \leq \frac{4}{\varepsilon} \sup_{P \in \mathcal{P}} \sup_{[\delta]} P_P \{ \sup_{[\delta]} P_n^0(f - g) > \frac{\varepsilon}{4\sqrt{n}} \} \]

\[ (n) \quad \leq \frac{4}{\varepsilon} \sup_{P \in \mathcal{P}} E_P \left\{ P_P \{ \sup_{[\delta]} P_n^0(f - g) > \frac{\varepsilon}{4\sqrt{n}} \} 1_{B_n^*} \right\} \]

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

\[ \text{Prob} \{ \sup_{[\delta]} |P_n^0(f - g)| > \varepsilon \sqrt{n} \} < \frac{\varepsilon}{8} \quad \text{on } B_n^* \]

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

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

By definition 1, \( F(i) \) need contain at most \( \exp(H_i) \) functions (recall that \( \delta_i = 2^{-i} \) and \( H_i = H_{f^{(2)}(2^{-i}, F)} \)). 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 \( r \)

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

pointwise on \( X_n \).

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

February 20, 1987
(r) \[ \text{Prob}\{ \sup_{f'} |P_n^0(f_{r'-f'})| > \frac{\epsilon}{16 \sqrt{n}} |X_n| \} < \frac{\epsilon}{16} \quad \text{on } B_n = \{ \|F\|^2_n \leq \|F\|^2_{L_2(P)} + M \}. \]

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

(s) \[ \text{Prob}\{ \sup_{f',f''} |P_n^0(f_{r'-f''})| > \frac{\epsilon}{8 \sqrt{n}} |X_n| \} < \frac{\epsilon}{16} \quad \text{on } B_n. \]

Since

(t) \[ \sup_{f',f''} |P_n^0(f_{r'-f''})| \leq 2 \sup_{f'} |P_n^0(f_{r'-f'})| + \sup_{f',f''} |P_n^0(f_{r'-f''})| \]

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

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

\[ \text{Prob}\{ \sup_{f'} |P_n^0(f_{r'-f'})| > \frac{1}{\sqrt{n}} \sum_{j=r+1}^{\infty} \eta_j |X_n| \} \]
\[ \leq \sum_{j=r+1}^{\infty} \text{Prob}\{ \sup_{f'} |P_n^0(f_{r'-f'})| > \frac{\eta_j}{\sqrt{n}} |X_n| \} \]
\[ \leq \sum_{j=r+1}^{\infty} |F_j| \text{Prob}\{ |P_n^0(f_{r-j})| > \frac{\eta_j}{\sqrt{n}} |X_n| \}. \]

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

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

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

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

where

\[ \sum h_i^2 = n \|f_{r-j-1}\|^2_n \leq n (\|f-f_{r-j}\|_n + \|f-f_{r-j-1}\|_n)^2 \]
\[ \leq n \|F\|^2_n (\delta_j + \delta_{j-1})^2 \]
\[ \leq 9 n \delta_j^2 (\|F\|^2_{L_2(P)} + M) \quad \text{on } B_n \]
\[ \leq 18 n \delta_j^2 M \quad \text{by (a)}. \]

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

\[ \sum_{j=r+1}^{\infty} \exp(2 H_j) \exp(- \frac{\eta_j^2}{36 \delta_j^2 M}) \leq 2 \sum_{j=r+1}^{\infty} \exp(- \frac{\eta_j^2}{72 \delta_j^2 M}) \quad \text{by (c)} \]

February 20, 1987
hence (r) holds.

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

\[
\|f' - f''\|_n \leq \|f' - f\|_n + \|f'' - f\|_n + \|f'' - f''\|_n \\
\leq \sqrt{3/2} \delta + 2 \delta_r \|F\|_n \\
\leq \sqrt{3/2} \delta + 2 \delta_r \|F\|^2_{L(P)} + M^{1/2} \quad \text{on} \ B_n
\]

by (a) and our choice of \( \delta \) in (i). Let \( \eta = \epsilon / 8 \). Use of Hoeffding's inequality again allows us to bound the left side of (s) on \( B_n \) by

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

Hence (s) holds. \( \square \)

**Proof of corollary 3.** First note that (i) and (iii) imply the hypotheses of proposition 1 with \( P = \{P_n, n \geq 1, P_0\} \). Thus proposition 1 yields in particular that \( F \) is totally bounded in the \( L_2(P_0) \) seminorm; and hence (i) of theorem 2 holds. Conditions (ii) and (iii) of theorem 2 hold by hypotheses (ii) and (iii), while (iv) of theorem 2 follows from (i), (ii), and theorem 3. Hence theorem 2 yields the conclusion. \( \square \)

**Proof of corollary 4.** It suffices to verify the hypotheses of corollary 3. Hypothesis (i) follows from theorem 9 of Pollard (1982); or see Dudley (1984), (1986). Hypothesis (ii) holds trivially since \( F = 1 \) is bounded in this case, and (iii) holds by hypothesis. Hence corollary 3 applies. \( \square \)

**Proof of corollary 5.** By corollary 3 it suffices to show that \( H(P_n, P_0) \to 0 \) implies \( \rho(P_n, P_0) \to 0 \) for \( \{P_n\} \subset P \). But, by Cauchy - Schwarz, for \( f, g \in F \), with \( \mu_n = P_n + P_0, \ P_n = \frac{dP_n}{d(P_n + P_0)}, \ s_n = \sqrt{P_n}, \ p_0 = \frac{dP_0}{d(P_n + P_0)}, \ s_0 = \sqrt{P_0} \), it follows that

\[
\left\{ \int g d(P_n - P_0) \right\}^2 = \left\{ \int g [s_n - s_0][s_n + s_0] d\mu_n \right\}^2
\]

February 20, 1987
where
\[ \int g^2[s_n - s_0]^2d\mu_n \leq 2\int g^2d(P_n + P_0) \]
by the uniform integrability hypothesis (ii), and, for \( \lambda > 0 \),
\[ \int f^2[s_n - s_0]^2d\mu_n \]
\[ = \int f^21_{|f|>\lambda}[s_n - s_0]^2d\mu_n + \int f^21_{|f|\leq\lambda}[s_n - s_0]^2d\mu_n \]
\[ \leq 2\int f^21_{|f|>\lambda}[s_n^2 + s_0^2]d\mu_n + \lambda^2H^2(P_n, P_0) \]
\[ \leq 2\{E_{P_n}F^21_{|F|>\lambda} + E_{P_0}F^21_{|F|>\lambda} + \lambda^2H^2(P_n, P_0) \}
Hence, combining (a), (b), and (c) yields
\[ \rho^2(P_n, P_0) \leq M \left\{ 2\{E_{P_n}F^21_{|F|>\lambda} + E_{P_0}F^21_{|F|>\lambda} + \lambda^2H^2(P_n, P_0) \} \right\} \]
\[ \rightarrow 0 \]
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 \( n \rightarrow \infty \).

Proofs of theorems 4 and 5. In both cases it suffices to check that \( \rho(P_n, P_0) \rightarrow_{a.s.} 0 \) as \( n \rightarrow \infty \). In the case of theorem 4, this follows easily from corollary 2 applied to \( P = \{P_0\} \). In the case of theorem 5 we have
\[ \rho(P_n, P_0) \leq \rho(P_n, P_n) + \rho(P_n, P_0) \]
\[ \rightarrow_{a.s.} 0 + 0 = 0 \]
by (9) of corollary 2 and by hypothesis. The conclusion then follows from corollary 3.

Acknowledgment: We owe thanks to a referee for suggestions which lead to a clearer presentation of the proof of theorem 3.


