Bootstrapping general empirical measures

by

Evarist Giné[1] and Joel Zinn[2]

Texas A&M University

Running head: Bootstrapping empirical measures.

Abstract

It is proved that the bootstrapped central limit theorem for empirical processes indexed by a class of functions $\mathcal{F}$ and based on a probability measure P holds a.s. if and only if $\mathcal{F} \,\epsilon\, CLT(P)$ and $\int F^2 dP < \infty$, where $F = \sup_{f \epsilon \mathcal{F}} |f|$ and it holds in probability if and only if $\mathcal{F} \in CLT(P)$. Thus, for a large class of statistics, no local uniformity of the CLT (about P) is needed for the bootstrap to work. Consistency of the bootstrap (the bootstrapped law of large numbers) is also characterized. These results are proved under some mild measurability assumptions of $\mathcal{F}$ for P.

AMS 1980 subject classifications. Primary: 60F17, 62E20; secondary: 60B12.

Key works and phrases: bootstrapping, empirical processes, central limit theorem.

[1] Research partially supported by National Science Foundation Grant No. DMS8619411

[2] Research partially supported by National Science Foundation Grant No. DMS8601250

Introduction. B. Efron (1979) introduced the "bootstrap", a resampling method for approximating the distribution functions of statistics $H_n(X_1, ..., X_n; P)$, where the random variables X_i are independent, identically distributed with common law P (i.i.d.(P)). Since the empirical measure

$$P_n(\omega) = n^{-1} \sum_{i=1}^{n} \delta_{X_i(\omega)} \tag{1.1}$$

is (a.s.) close to P, one may hope that, if $\hat{X}_{n1}, ..., \hat{X}_{nn}$ are i.i.d.($P_n(\omega)$) (i.e. the $\hat{X}_{ni}$ are obtained by sampling from the data, with replacement), then the distribution of $\hat{H}_n(\omega) = H_n(\hat{X}_{n1}, ..., \hat{X}_{nn}; P_n(\omega))$ is ω-a.s. asymptotically close to that of $H_n(X_1, ..., X_n; P)$. In turn, the distribution of the bootstrapped statistic, $\hat{H}_n(\omega)$, can be approximated by Monte-Carlo simulation. This suggestive method has been validated with limit theorems for many particular $\hat{H}_n(\omega)$ by Efron (loc. cit.), Bickel and Freedman (1981), Singh (1981), Beran (1982, 1984), Bretagnolle (1983), Gaenssler (1986) and others. In this article we offer a justification of the bootstrap for functions H_n of a special type, namely for continuous functions of the empirical measure viewed as an element of $\ell^\infty(\mathcal{F})$, for classes of functions $\mathcal{F}$. Such H include the Kolmogorov-Smirnov and the Cramér-Von Miser statistics (in any number of dimensions) as well as the statistics considered in Beran and Millar (1986).

Let $(S, \mathcal{S}, P)$ be a probability space, let $X_i : (S^{\mathbf{N}}, \mathcal{S}^{\mathbf{N}}, P^{\mathbf{N}}) \to (S, \mathcal{S}, P)$ be the coordinate functions (i.i.d.(P)), let $P_n(\omega)$ be as in (1.1) for $\omega \in S^{\mathbf{N}}$, let $\hat{X}^\omega_{nj}$ $j = 1, ..., n$, be i.i.d.($P_n(\omega)$), let $\hat{P}_n(\omega)$ be the empirical measure based on $\{\hat{X}^\omega_{nj}\}_{j=1}^n$, i.e.

$$\hat{P}_n(\omega) = n^{-1} \sum_{j=1}^{n} \delta_{\hat{X}^\omega_{nj}}, \tag{1.2}$$

and let $\mathcal{F}$ be a class of measurable functions on $(S, \mathcal{S})$ such that

$$F = \sup_{f \in \mathcal{F}} |f| \tag{1.3}$$

is finite for all $s \in S$. We then prove that, under some measurability on $\mathcal{F}$, the conditions

$$\int F^2 dP < \infty \tag{1.4}$$

and

$$n^{1/2}(P_n - P) \to G_P \qquad \text{weakly in } \ell^\infty(\mathcal{F}) \tag{1.5}$$

are necessary and sufficient for

$$n^{1/2}(\hat{P}_n(\omega) - P_n(\omega)) \to G \qquad \text{weakly in } \ell^\infty(\mathcal{F}), \quad \omega - \text{a.s.} \tag{1.6}$$

for a centered Gaussian process G independent of ω, and then G coincides with G_P, the Gaussian limit in (1.5).

Thus, this result completely settles, modulo measurability, the question of the validity of the bootstrap for the CLT for empirical processes indexed by classes of functions (or sets).

The main feature of this theorem, aside from its generality, is that no assumptions are made on local uniformity (about P) of the CLT (1.5) for the bootstrap CLT (1.6) to hold (this was unexpected, in view of e.g. the comments in Bickel and Freedman, loc. cit. page 1209). Another new feature is necessity of the integrability condition (1.4) and the usual CLT (1.5) for the bootstrap.

The proof relies on several results and techniques from Probability in Banach spaces. Among other such results and techniques, we use symmetrization by randomization in an essential way (an idea in Pisier (1985) has been useful in connection with this), results of Le Cam (1970) on Poissonization and on the CLT in Banach spaces, integrability of Gaussian processes (e.g. Fernique (1984)), Hoffmann-Jorgensen's (1974) inequality and convergence of moments in the CLT in Banach spaces (de Acosta and Giné (1979)), results on empirical processes from Giné and Zinn (1984, 1986) and, particularly, a result of Ledoux, Talagrand and Zinn (cf. Ledoux and Talagrand (1988)) on the almost sure weak convergence of $\Sigma_{i=1}^n g_i X_i(\omega)/n^{1/2}, g_i$ i.i.d. with $\int_0^\infty (P\{|g_1| > t\})^{1/2} dt < \infty$ (i.e. $g_1 \in L_{2,1}$). Actually, it is this last result that is at the base of our proof. The Ledoux-Talagrand-Zinn result uses for its proof a recent extension of Yurinski's decomposition as applied to $E_g \|\Sigma g_i x_i\| - E\|\Sigma g_i X_i\|$. This was observed by Ledoux and Talagrand (1986) in the proof of one of the main results about the law of the iterated logarithm in Banach spaces.

The above techniques (except for the result of Ledoux, Talagrand and Zinn) can be used to obtain a similar result for the bootstrap in probability. The a.s. results are given in Section 2 and Section 3 contains the "in probability" result.

The bootstrapped law of the large numbers, much easier to prove than the CLT, is also characterized.

2. **The a.s. bootstrapped limit theorems.** Given P, a probability measure on a measurable space $(S, \mathcal{S})$, we let

$$\rho_P^2(f,g) = \int (f-g)^2 dP - \left(\int (f-g)dP\right)^2, \; f,g \in \mathcal{L}_2(P), \tag{2.1}$$

$$e_P^2(f,g) = \int (f-g)^2 dP, \; f,g \in \mathcal{L}_2(P) \tag{2.2}$$

and, given a collection $\mathcal{F}$ of P-square integrable functions on $(S, \mathcal{S})$, we let

$$\mathcal{F}'_\delta = \{f-g : f,g \in \mathcal{F}, e_P(f,g) \leq \delta\}, \quad \delta > 0, \tag{2.3}$$

$$\mathcal{F}' = \{f-g : f,g \in \mathcal{F}\},$$

$$(\mathcal{F}')^2 = \{(f-g)^2 : f,g \in \mathcal{F}\}. \tag{2.4}$$

$G_P := \{G_P(f) : f \in \mathcal{F}\}$ denotes a centered Gaussian process indexed by $\mathcal{F}$, with covariance

$$EG_P(f)G_P(g) = \int fg dP - \int f dP \int g dP, \qquad f,g \in \mathcal{F} \tag{2.5}$$

and $Z_P := \{Z_P(f) : f \in \mathcal{F}\}$ denotes the centered Gaussian process with

$$EZ_P(f)Z_P(g) = \int fg dP, \qquad f,g \in \mathcal{F}. \tag{2.6}$$

We recall Hoffmann-Jorgensen's (1984) definition of weak convergence in $\ell^\infty(\mathcal{F})$, the space of bounded functions $\mathcal{F} \to \mathbf{R}$ with the sup norm topology: a sequence $\{Y_n\}_{n=1}^\infty$ of random elements of $\ell^\infty(\mathcal{F})$ *converges weakly* in $\ell^\infty(\mathcal{F})$ if there exists a Radon probability measure γ on $\ell^\infty(\mathcal{F})$ such that for all $H : \ell^\infty(\mathcal{F}) \to \mathbf{R}$ bounded and continuous,

$$\lim_{n\to\infty} E^* H(Y_n) = \int H d\gamma.$$

Then we say that $\mathcal{F} \,\epsilon\, CLT(P)$ if the sequence $\{n^{1/2}(P_n - P)(f) : f \,\epsilon\, \mathcal{F}\}$ converges weakly in $\ell^\infty(\mathcal{F})$ to a Radon centered Gaussian probability measure γ_P on $\ell^\infty(\mathcal{F})$. γ_P is the law of G_P which, by virtue of the Radonicity of γ_P, admits a version with bounded uniformly continuous paths on $(\mathcal{F}, \rho_P)$, and $(\mathcal{F}, \rho_P)$ is totally bounded (see e.g. Giné and Zinn (1986)). We continue denoting this version by G_P.

If $\mathcal{F}$ satisfies certain measurability conditions, then P_n can be randomized (i.e. we can replace $\delta_{X_i} - P$ by $\xi_i \delta_{X_i}$ with ξ_i symmetric, independent of X_i and satisfying certain integrability conditions) and Fubini's theorem can be freely used. These conditions spelled out in Giné and Zinn (1984) are that $\mathcal{F}$ be nearly linearly deviation measurable for P, $NLDM(P)$ for short, and that both $\mathcal{F}^2$ and $\mathcal{F}'^2$ are nearly linearly supremum measurable for P, $NLSM(P)$. In this paper if $\mathcal{F}$ satisfies all of the above conditions with respect to P we write $\mathcal{F} \in M(P)$. To see why $\mathcal{F} \in M(P)$ suffices we note, as in Giné and Zinn (1984) Remark 2.4 (2), p. 935, that the measurability of the

$$(a_1, \ldots, a_n, x_1, \ldots, x_n) \to \sup_{f\in G} \left\{ \sum_{j=1}^{n} a_j f(X_j) \right\}$$

implies, for example, the measurability for any $M < \infty$ of the map

$$(x_1, \ldots, x_n) \to \sup_{f\in G} \left\{ \sum_{j=1}^{n} f(x_j) I_{F(X_j)\le M} \right\}$$

by considering the composition of the map

$$(x_1, \ldots, x_n) \to (I(F(x_1) \le M), \ldots, I(F(x_n) \le M), x_1, \ldots, x_n)$$

with the measurable map given by hypothesis. Actually close consideration of the proofs shows that even weaker hypotheses suffice, but the best measurability is not our concern here. We further note that if $\mathcal{F}$ is countable, or if $\{P_n\}_{n=1}^{\infty}$ are stochastically separable in $\mathcal{F}$, or more generally, if $\mathcal{F}$ is image admissible Suslin (Dudley (1986), p. 101) then $\mathcal{F} \in M(P)$.

The following proposition is the first step in the proof of the bootstrap CLT. It is a version of Le Cam's Poissonization Lemma (Le Cam (1970); reproduced in Araujo and Giné (1980, Thm. 3.4.8)), for expectations.

2.1. Lemma. Let B be a separable Banach space and let $\|\cdot\|$ be a measurable pseudonorm on B. For some $n \in \mathbf{N}$, let $\{X_i\}_{i=1}^n$ be independent symmetric B-valued random variables and let $\{\mathcal{L}(X_i)\}_{i=1}^n$ be their laws. Then

$$E\Big\|\sum_{i=1}^n X_i\Big\| \le 8\int \|x\| d\ Pois\Big(\sum_{i=1}^n \mathcal{L}(X_i)\Big)(x). \tag{2.7}$$

(We recall that for a finite measure ν, $Pois\ \nu = e^{-\nu(B)}\Sigma_{n=0}^{\infty}\nu^n/n!$ where $\nu^n = \nu * \overset{n)}{\dots} * \nu$, that $Pois\Sigma\nu_i = (Pois\ \nu_1) * \dots * (Pois\ \nu_n)$, and that if $\nu = \frac{1}{2}(\delta_x + \delta_{-x})$ for some $x \in B$, then $Pois\ \nu = \mathcal{L}(\tilde{N}x)$ where $\tilde{N} = N - N'$ with N and N' independent Poisson real random variables with expectation 1/2; we will call $\tilde{N}$ a symmetrized Poisson random variable.) Here is a proof of inequality (2.7): If X_{ij} are independent, $X_{i0} = 0, \mathcal{L}(X_{ij}) = \mathcal{L}(X_i)$ for $j > 0$, and N_i are Poisson with parameter 1, independent and independent of $\{X_{ij}\}$, then Fubini's theorem and convexity ($E\|X + Y\| \ge E\|X\|$ if X and Y are independent and $EX = 0$) give

$$\begin{aligned}(1 - e^{-1})E\|\Sigma X_i\| &\le E\|\Sigma(N_i \wedge 1)X_{i1}\| \\ &= E_N(E_X\|\Sigma(N_i \wedge 1)X_{i1}\|) \le E_N\Big(E_X\Big\|\sum_i\sum_{j=0}^{N_i} X_{ij}\Big\|\Big) \\ &= E\Big\|\sum_i\sum_{j=0}^{N_i} X_{ij}\Big\| = \int \|x\| d\ Pois(\Sigma\mathcal{L}(X_i))(x).\end{aligned}$$

2.2. Proposition. Let B be a Banach space, let $\|\cdot\|$ be a measurable pseudonorm, let $n \in \mathbf{N}$, let $\{x_i\}_{i=1}^n \subset B$, let $\hat{X}_{nj}, j = 1, ..., n$, be i.i.d. B-valued random variables with $\mathcal{L}(\hat{X}_{nj}) = n^{-1}\Sigma_{i=1}^n \delta_{x_i}$, and let $\{\varepsilon_j\}_{j=1}^n, \{\tilde{N}_j\}_{j=1}^n$ be respectively a Rademacher sequence and a sequence of independent symmetrized Poisson real random variables with parameter 1/2, both independent of $\{\hat{X}_{nj}\}$. Then

$$\frac{1}{\sqrt{2}}(1-e^{-1})E\Big\|\sum_{i=1}^n \varepsilon_i x_i\Big\| \le E\Big\|\sum_{j=1}^n \varepsilon_j \hat{X}_{nj}\Big\| \le 8E\Big\|\sum_{i=1}^n \tilde{N}_i x_i\Big\|. \tag{2.8}$$

Proof. We can write

$$\hat{X}_{nj} = \sum_{i=1}^n x_i I_{A_{ij}}$$

where for each j, the sets $A_{1j}, A_{2j}, ..., A_{nj}$ are disjoint, the sequences $\{A_{ij}\}_{i=1}^n, j = 1, ..., n$, are independent, and $PA_{ij} = 1/n, i, j = 1, ..., n$. Let $\{\varepsilon_{ij}\}$ be a Rademacher array independent of $\{A_{ij}\}$. Then, by disjointness, the vectors

$$\varepsilon_j(x_1 I_{A_{1j}}, ..., x_n I_{A_{nj}}) \text{ and } (\varepsilon_{1j}x_1 I_{A_{1j}}, ..., \varepsilon_{nj}x_n I_{A_{nj}}), \qquad j = 1, ..., n,$$

all have the same distribution and, of course, they are independent for different $j's$. Moreover, by independence of $\{\varepsilon_{ij}\}$ and independence between $\{\varepsilon_{ij}\}$ and $\{A_{ij}\}$, the vector $(\Sigma_{j=1}^n \varepsilon_j I_{A_j}, ..., \Sigma_{j=1}^n \varepsilon_{nj} I_{A_{nj}})$ is symmetric. Let $\{\varepsilon'_j\}$ be a Rademacher sequence independent of $\{\varepsilon_{ij}\}$ and $\{A_{ij}\}$. Then these two observations give the following:

$$\begin{aligned} E\Big\|\sum_{j=1}^n \varepsilon_j \hat{X}_{nj}\Big\| &= E\Big\|\sum_{j=1}^n \varepsilon_j \sum_{i=1}^n x_i I_{A_{ij}}\Big\| = E\Big\|\sum_{j=1}^n\sum_{i=1}^n \varepsilon_{ij} x_i I_{A_{ij}}\Big\| \\ &= E\Big\|\sum_{i=1}^n \Big(\sum_{j=1}^n \varepsilon_{ij} I_{A_{ij}}\Big)x_i\Big\| = E\Big\|\sum_{i=1}^n \varepsilon'_i\Big(\sum_{j=1}^n \varepsilon_{ij} I_{A_{ij}}\Big)x_i\Big\| \\ &= E\Big\|\sum_{i=1}^n \varepsilon'_i\Big|\sum_{j=1}^n \varepsilon_{ij} I_{A_{ij}}\Big|x_i\Big\|. \end{aligned} \tag{2.9}$$

We now notice that by Khinchin's inequality (see Szarek (1976) or Haagerup (1981) for the best constant)

$$\begin{aligned} E\Big|\sum_{j=1}^n \varepsilon_{ij} I_{A_{ij}}\Big| &\ge \frac{1}{\sqrt{2}}E\Big(\sum_{j=1}^n I_{A_{ij}}\Big)^{1/2} \ge \frac{1}{\sqrt{2}}P\Big\{\sum_{j=1}^n I_{A_{ij}} \ne 0\Big\} \\ &= \frac{1}{\sqrt{2}}\Big[1-\Big(1-\frac{1}{n}\Big)^n\Big] \ge \frac{1}{\sqrt{2}}(1-e^{-1}). \end{aligned}$$

Hence, by Jensen's inequality and (2.9), and since $E\big|\sum_{j=1}^n \varepsilon_{ij} I_{A_{ij}}\big|$ does not depend on i,

$$E\Big\|\sum_{j=1}^{n}\varepsilon_j\hat{X}_{nj}\Big\| \geq \frac{1}{\sqrt{2}}(1-e^{-1})E\Big\|\sum_{i=1}^{n}\varepsilon_i' x_i\Big\|,$$

which is the first inequality in (2.8). This proof is essentially taken from Pisier (1975, proof of Proposition 5.1).

Let $e_1, ..., e_n$ be the canonical basis of $\mathbf{R}^n$, and for $a = \Sigma a_i e_i$, let $|||a||| := \big\|\Sigma a_i x_i\big\|$, which is a pseudonorm on $\mathbf{R}^n$. Consider now the random vectors

$$Y_j = \sum_{i=1}^{n}\varepsilon_{ij} I_{A_{ij}} e_i, \qquad j = 1, ..., n,$$

which are independent, symmetric and have probability laws

$$\mathcal{L}(Y_j) = \frac{1}{2n}\sum_{i=1}^{n}(\delta_{e_i} + \delta_{-e_i}) \tag{2.10}$$

(i.e. Y_j takes the values $\pm e_i$, $i = 1, ..., n$, each with probability $\frac{1}{2n}$). Then, $\big\|\Sigma_{i=1}^{n}\big(\Sigma_{j=1}^{n}\varepsilon_{ij} I_{A_{ij}}\big)x_i\big\| = \big|\big|\big|\Sigma_{j=1}^{n} Y_j\big|\big|\big|$. This, (2.9), (2.10), and Le Cam's Lemma (Lemma 2.1) give

$$\begin{aligned} E\Big\|\sum_{j=1}^{n}\varepsilon_j\hat{X}_{nj}\Big\| = E\Big|\Big|\Big|\sum_{j=1}^{n} Y_j\Big|\Big|\Big| &\leq 8\int |||x|||d\ Pois\Big(\frac{1}{2}\sum_{i=1}^{n}(\delta_{e_i}+\delta_{-e_i})\Big)(x) \\ &= 8E\Big|\Big|\Big|\sum_{i=1}^{n}\tilde{N}_i e_i\Big|\Big|\Big| = 8E\Big\|\sum_{i=1}^{n}\tilde{N}_i x_i\Big\|, \end{aligned}$$

which is the right hand side inequality in (2.8). □

What is needed from the result of Ledoux, Talagrand and Zinn is the main part of their proof, namely Lemma 5 in Ledoux and Talagrand (1988). In the empirical case one needs to complete the proof of tightness in a way different from the original; we incorporate this in the proof of our theorem. First, the

2.3. Lemma. Let $(S, \mathcal{S}, P)$ be a probability space $\mathcal{F}$ a $NLDM(P)$ class of functions on S with $E_P F^2 < \infty, |||\cdot|||$ any of the pseudonorms $\|\cdot\|_{\mathcal{F}}, \|\cdot\|_{\mathcal{F}'_\delta}, \delta > 0, X_i : S^{\mathbf{N}} \to S$ the coordinate functionals, and $\{\xi_i\}$ a sequence of i.i.d. symmetric real random variables with

$E\xi_1^2 < \infty$, independent of $\{X_i\}$ (actually defined on another probability space). Let E_ξ denote integration with respect to only the variables $\{\xi_i\}$. Then,

$$\text{a.s. } \limsup_n n^{-1/2} E_\xi \left|\left|\left| \sum_{i=1}^n \xi_i X_i(\omega) \right|\right|\right| \le 4 \limsup_n n^{-1/2} E \left|\left|\left| \sum_{i=1}^n \xi_i X_i \right|\right|\right|. \tag{2.11}$$

The bootstrap CLT is as follows:

2.4. Theorem. Let $\mathcal{F} \in M(P)$ and P be a probability measure on $(S, \mathcal{S})$. Further let $P_n, \hat{P}_n(\omega), \omega \,\epsilon\, S^{\mathbf{N}}$, and G_P be as defined in (1.1), (1.2) and (2.5). Then the following are equivalent:

(a) $\int F^2 dP < \infty$ and $\mathcal{F} \,\epsilon\, CLT(P)$;

(b) there exists a centered Gaussian process G on $\mathcal{F}$ whose law is Radon in $\ell^\infty(\mathcal{F})$ such that, $P^{\mathbf{N}}$-a.s., $n^{1/2}(\hat{P}_n(\omega) - P_n(\omega)) \to G$ weakly in $\ell^\infty(\mathcal{F})$.

And if either (a) or (b) hold, then $G = G_P$.

Proof. (a) $\Rightarrow$ (b). Obviously, if N is a Poisson real random variable, then $\int_0^\infty (P\{N > t\})^{1/2} dt < \infty$. So, Lemma 1.2.4 in Giné and Zinn (1986) holds for $g_k = \tilde{N}_k$, a sequence of i.i.d. symmetrized Poisson real random variables with parameter 1/2; hence Theorem 1.2.8 ((a) $\Rightarrow$ (e)) there gives:

$$(\mathcal{F}, e_P) \text{ is totally bounded} \tag{2.12}$$

and

$$\lim_{\delta \to 0} \limsup_n E \left|\left| \sum_{i=1}^n \tilde{N}_i \delta_{X_i} / n^{1/2} \right|\right|_{\mathcal{F}'_\delta} = 0. \tag{2.13}$$

(Here $\{X_i\}$ is independent of $\{\tilde{N}_i\}$, and is as defined in the introduction, i.e. for $i \,\epsilon\, \mathbf{N}, X_i$ is the i-th coordinate of $(S^{\mathbf{N}}, \mathcal{S}^{\mathbf{N}}, P^{\mathbf{N}})$.) Let E_N denote integration only with respect to $\{\tilde{N}_i\}$. Then, (2.13) and Lemma 2.3 give:

$$P^{\mathbf{N}} - \text{a.s.}, \quad \lim_{\delta \to 0} \limsup_n E_N \left|\left| \sum_{i=1}^n \tilde{N}_i \delta_{X_i(\omega)} / n^{1/2} \right|\right|_{\mathcal{F}'_\delta} = 0. \tag{2.14}$$

(2.14) and Proposition 2.2 then give (letting $E_{\varepsilon,A}$ denote integration only with respect to $\{\varepsilon_j\}$ and $\{A_{ij}\}$) :

$$P^{\mathbf{N}} - \text{a.s.} \quad \lim_{\delta\to 0} \limsup_n E_{\varepsilon,A} \Big\| \sum_{j=1}^n \varepsilon_j \delta_{\hat{X}_{nj}(\omega)}/n^{1/2} \Big\|_{\mathcal{F}'_\delta} = 0 \tag{2.15}$$

and by symmetrization (we will use without further mention that for $\{U_i\}$ independent, independent of $\{\varepsilon_i\}, E\|\Sigma(U_i - EU_i)\| \leq 2E\|\Sigma\varepsilon_i U_i\|$ and $E\|\Sigma\varepsilon_i(U_i - EU_i)\| \leq 2E\|\Sigma(U_i - EU_i)\|$),

$$P^{\mathbf{N}} - \text{a.s.} \lim_{\delta\to 0} \limsup_n E_A \|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}'_\delta} = 0. \tag{2.16}$$

If $\mathcal{F} \,\epsilon\, CLT(P)$, so does $\mathcal{F}' \,\epsilon\, CLT(P)$. Then, Theorem 1.4.6 in Giné and Zinn (1986) gives $\sup_{f\in\mathcal{F}} |(P_n(\omega) - P)(f^2)| \to 0$ and $\sup_{f,g\in\mathcal{F}} |(P_n(\omega) - P)(f-g)| \to 0$ in probability. Since $\int F^2 dP < \infty$ these limits hold a.s. (e.g. by a reverse submartingale argument as in Pollard (1981)). Therefore

$$\sup_{f,g\epsilon\mathcal{F}} |(P_n(\omega) - P)(fg)| \to 0 \;\text{ a.s.} \tag{2.17}$$

and of course

$$\|P_n(\omega) - P\|_{\mathcal{F}} \to 0 \;\text{ a.s.} \tag{2.18}$$

[We should note here that the proof of Theorem 1.4.6, loc. cit. contains a typographical error (which in the end, is of no consequence for its validity): the relation between entropies should read $N_{n,2}(\varepsilon, \overline{\mathcal{F}}(\lambda)^2) \leq N_{n,2}(\varepsilon/2\lambda, \overline{\mathcal{F}}(\lambda))$.] Call the subsets of $S^{\mathbf{N}}$ where (2.17) and (2.18) hold respectively Ω_1 and Ω_2, and let Ω_3 be the intersection for all $\alpha > 0$ rational of the subsets of $S^{\mathbf{N}}$ for which eventually $\max_{i\leq n} F(X_i(\omega)) \leq \alpha n^{1/2}$. It follows from the Lindeberg-Feller theorem (as e.g. in Singh (1981)) that for $\omega \,\epsilon\, \Omega_1 \cap \Omega_2 \cap \Omega_3, n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))(\Sigma_{\text{finite}} a_i f_i) \to \Sigma a_i G_P(f_i)$ weakly, for all $\{a_i\} \subset \mathbf{R}, \{f_i\} \subset \mathcal{F}$. This, (2.16) and (2.12) imply the bootstrap CLT (b) with $G = G_P$ by e.g. Theorem 1.1.3 in Giné and Zinn (1986) (which, although given for the i.i.d. sequence case, it holds, with the same proof, for triangular arrays as well).

(b) $\Rightarrow$ (a). We show first that if (b) holds then $\int F^2 dP < \infty$. Note that the convergence in (b) is actually weak convergence of Radon measures (for each ω for which there is convergence) and therefore the CLT theory for separable Banach spaces applies. The system $\{Y_{nj}(\omega) = n^{-1/2}\delta_{\hat{X}^{\omega}_{nj}}\}$ is infinitesimal ω-a.s.: $P^{\mathbf{N}}$-a.s., for all $\varepsilon > 0, P_n^n\{\|f(\hat{X}^{\omega}_{n1})\|_{\mathcal{F}} > \varepsilon n^{1/2}\} = \Sigma_{i=1}^n I(F(X_i(\omega)) > \varepsilon n^{1/2})/n \to 0$ by the law of large numbers (by monotonicity, it is enough to consider rational $\varepsilon > 0$). Hence, since ω-a.s. the sequence $\{\Sigma_{j=1}^n Y_{nj}(\omega)\}$ is shift convergent in law to a Gaussian limit it follows from a result of Le Cam (1970) (see e.g. Araujo and Giné (1980, Theorem 3.5.4)) that

$$nP_n\{\|f(\hat{X}^{\omega}_{n1})\|_{\mathcal{F}} > n^{1/2}\} \to 0 \qquad \text{a.s.}$$

that is,

$$\sum_{i=1}^n I(F(X_i(\omega)) > n^{1/2}) \to 0 \qquad \text{a.s.} \tag{2.19}$$

Since if $\Sigma_{i=1}^m I(F(X_i(\omega)) > n^{1/2}) < 1$ then $\Sigma_{i=1}^m I(F(X_i(\omega)) > n^{1/2}) = 0$, (2.19) implies that ω-a.s. there is $n(\omega) < \infty$ such that for $n > n(\omega)$,

$$F(X_n(\omega))/n^{1/2} \le \max_{i \le n} F(X_i(\omega))/n^{1/2} \le 1.$$

This and the Borel Cantelli lemma give $\Sigma P\{F(X_n) > n^{1/2}\} < \infty$, that is

$$EF^2(X_1) < \infty. \tag{2.20}$$

Let $f \in \mathcal{F}' \cup \mathcal{F}$. Then by hypothesis $\mathcal{L}(n^{1/2}(\hat{P}_n f - P_n f)) \to_w \mathcal{L}(G(f))$ and by the converse CLT in $\mathbf{R}$ for triangular arrays, together with (2.19), we have

$$\lim_{n\to\infty}\left(\sum_{i=1}^n f(X_i)^2/n - \left(\sum_{i=1}^n f(X_i)/n\right)^2\right) = E(G(f))^2 \qquad a.s.$$

But, by (2.20) and the law of large numbers, this limit is $E(f(X_1))^2 - (Ef(X_1))^2$. We have thus shown

$$G = G_P. \tag{2.21}$$

Moreover, since G, hence G_P, has a Radon law, and since (2.21) holds, we also have that $(\mathcal{F}, e_P)$ is totally bounded.

Next we prove $P^{\mathbf{N}}$-a.s. uniform integrability of $\{\|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))\|_{\mathcal{F}}\}_{n=1}^{\infty}$. By Theorem 3.2 in de Acosta and Giné (1979) it is enough to show

$$\sup_n E_A \max_{j\le n} \|\delta_{\hat{X}^{\omega}_{nj}} - P_n(\omega)\|^2_{\mathcal{F}}/n < \infty \qquad \text{a.s.} \tag{2.22}$$

where E_A denotes integration with respect to $\{I_{A_{ij}}\}$. But the random variable in (2.22) is bounded by

$$\begin{aligned}\sup_n E_A\|\delta_{\hat{X}^{\omega}_{n1}} - P_n(\omega)\|^2_{\mathcal{F}} &= \sup_n \frac{1}{n}\sum_{i=1}^n \|\delta_{X_i(\omega)} - P_n(\omega)\|^2_{\mathcal{F}} \\ &\le 4\sup_n \frac{1}{n}\sum_{i=1}^n F^2(X_i(\omega)) < \infty \qquad \text{a.s.}\end{aligned}$$

(by the law of large numbers, since $\int F^2 dP < \infty$). We thus have, by uniform integrability,

$$P^{\mathbf{N}} - \text{a.s.}, \begin{cases} E_A\|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))\|_{\mathcal{F}} \to E\|G\|_{\mathcal{F}} \\ E_A\|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))\|_{\mathcal{F}'_\delta} \to E\|G_P\|_{\mathcal{F}'_\delta} \qquad \text{for all } \delta > 0. \end{cases} \tag{2.23}$$

Denote by $|||\cdot|||$ any of the pseudonorms $\|\cdot\|_{\mathcal{F}'_\delta}, \delta > 0$, or $\|\cdot\|_{\mathcal{F}}$. By Proposition 2.2 we have, with $c = (1+e^{-1})/\sqrt{2}$,

$$\begin{aligned} P^{\mathbf{N}} - \text{a.s.}, \qquad cE_\varepsilon\Big|\Big|\Big|\sum_{i=1}^n \varepsilon_i\delta_{X_i(\omega)}/n^{1/2}\Big|\Big|\Big| &\le E_{\varepsilon,A}\Big|\Big|\Big|\sum_{j=1}^n \varepsilon_j\delta_{\hat{X}^{\omega}_{nj}}/n^{1/2}\Big|\Big|\Big| \\ &\le E_{\varepsilon,A}\Big|\Big|\Big|\sum_{j=1}^n \varepsilon_j(\delta_{\hat{X}^{\omega}_{nj}} - P_n(\omega))/n^{1/2}\Big|\Big|\Big| + \Big(E\Big|\sum_{i=1}^n \varepsilon_j/n^{1/2}\Big|\Big)|||P_n(\omega)||| \\ &\le 2E_A|||n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))||| + |||P_n(\omega)|||. \end{aligned} \tag{2.24}$$

(2.23) and (2.24) give

$$\limsup_{n\to\infty} Pr\Big\{\Big|\Big|\Big|\sum_{i=1}^n \varepsilon_i\delta_{X_i}/n^{1/2}\Big|\Big|\Big| \geq M\Big\} \leq \frac{1}{M}\limsup_{n\to\infty} E\Big(E_\varepsilon\Big|\Big|\Big|\sum_{i=1}^n \varepsilon_i\delta_{X_i}/n^{1/2}\Big|\Big|\Big| \wedge M\Big)$$

$$\leq \frac{1}{M} E \limsup_{n\to\infty} E_\varepsilon \Big|\Big|\Big|\sum_{i=1}^n \varepsilon_i\delta_{X_i}/n^{1/2}\Big|\Big|\Big| \wedge M$$

$$\leq \frac{E|||\delta_{X_1}||| + 2E|||G_P|||}{cM} \to 0 \text{ as } M \to \infty.$$

The above inequality, by Hoffmann-Jørgensen's inequality and $EF^2(X_1) < \infty$, implies

$$\sup_n E\Big|\Big|\Big|\sum_{i=1}^n \varepsilon_i\delta_{X_i}/n^{1/2}\Big|\Big|\Big| < \infty. \tag{2.25}$$

In particular $E\|\Sigma_{i=1}^n \varepsilon_i\delta_{X_i}/n\|_{\mathcal{F}} \to 0$, hence $E\|P_n - P\| \to 0$, or,

$$|||P_n - P||| \to 0 \qquad \text{a.s.}$$

(cf. Pollard (1981)). Hence

$$\lim_{\delta\to 0}\lim_{n\to\infty} \|P_n(\omega)\|_{\mathcal{F}'_\delta} = \lim_{\delta\to 0} \|Ef(X_1)\|_{\mathcal{F}'_\delta} \leq \lim_{\delta\to 0} \|(Ef^2(X_1)\|_{\mathcal{F}'_\delta}^{1/2} = 0. \tag{2.26}$$

Using (2.26) in (2.24) we obtain that $P^{\mathbf{N}}$-a.s.

$$\lim_{\delta\to 0}\limsup_n E_\varepsilon\Big\|\sum_{i=1}^n \varepsilon_i\delta_{X_i(\omega)}/n^{1/2}\Big\|_{\mathcal{F}'_\delta} \leq (2 + c')c^{-1}\lim_{\delta\to 0} E\|G_P\|_{\mathcal{F}'_\delta} = 0. \tag{2.27}$$

Bounded convergence and Fatou's lemma then give $\lim_{\delta\to 0}\limsup_n E\big(\|\Sigma_{i=1}^n \varepsilon_i\delta_{X_i}/n^{1/2}\|_{\mathcal{F}'_\delta} \wedge M\big) = 0$ for all $M > 0$, which, by Theorem 1.2.8 in Giné and Zinn (1986) implies that $\mathcal{F} \in CLT(P)$. □

2.5. Remark. A corollary of Theorem 2.4 is that if X_i are i.i.d. B-valued random variables, B a separable Banach space, then

$$E\|X_1\|^2 < \infty \text{ and } X_1 \in CLT \Leftrightarrow \sum_{j=1}^n (\hat{X}_{nj} - \bar{X}_n)/n^{1/2} \to G_X \text{ weakly a.s.}$$

Actually the proof of this result is somewhat simpler than that of Theorem 2.4 since in this case $E\|X_1\| < \infty$ already implies $\||P_n - P\|| \to 0$ a.s. (see below (2.25)).

The law of large numbers has a proof similar to that of Theorem 2.4 but simpler since in this case the lemma of Ledoux, Talagrand and Zinn is not needed and some further simplifications are also possible.

2.6. Theorem. Let $\mathcal{F}$ be $NLDM(P)$. Then the following are equivalent:

(a) $\int F dP < \infty$ and $\|P_n - P\|_{\mathcal{F}} \to 0$ in probability.

(b) $P^{\mathbf{N}}$-a.s., $\quad \|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}} \to 0$ in probability.

Proof (Sketch). $\int F dP < \infty, \|P_n - P\|_{\mathcal{F}} \to 0$ pr. $\Rightarrow \|P_n - P\|_{\mathcal{F}} \to 0$ a.s. (e.g. Pollard (1981)) $\|\Sigma_{i=1}^n \varepsilon_i \delta_{X_i}/n\|_{\mathcal{F}} \to 0$ a.s. (p. 980, Giné and Zinn (1984) $\Rightarrow E\|\Sigma_{i=1}^n \widetilde{N}_i \delta_{X_i}/n\|_{\mathcal{F}} \to 0$ (as noted in Giné, Marcus and Zinn (1986), by a proof similar to that in Lemma 2.9 of Giné and Zinn (1984), since $E|\widetilde{N}| < \infty$) $\Rightarrow \|\Sigma_{i=1}^n \widetilde{N}_i \delta_{X_i}/n\|_{\mathcal{F}} \to 0$ a.s. (by, e.g., a reverse martingale argument as in Pollard, loc. cit.) $\Rightarrow P^{\mathbf{N}}$-a.s. $\|\Sigma_{i=1}^n \widetilde{N}_i \delta_{X_i(\omega)}/n\|_{\mathcal{F}} \to 0$ a.s. (Fubini) $\Rightarrow P^{\mathbf{N}}$-a.s. $E_N\|\Sigma_{i=1}^n \widetilde{N}_i \delta_{X_i(\omega)}/n\|_{\mathcal{F}} \to 0$. (To see this we use Hoffmann-Jørgensen's inequality (Hoffmann-Jørgensen (1974)) to reduce to showing $E_N \max_{i \le n} |\widetilde{N}_i| \|\delta_{X_i(\omega)}\|_{\mathcal{F}}/n \to 0, P^{\mathbf{N}}$-a.s. But, for any $c > 0$

$$E_N \max_{i \le n} |\widetilde{N}_i| \ \|\delta_{X_i(\omega)}/n\|_{\mathcal{F}}/n \le c \max_{i \le n} \|\delta_{X_i(\omega)}\|_{\mathcal{F}}/n + E_N \sum_{i=1}^n |\widetilde{N}_i| I_{|\widetilde{N}_i| > c} \|\delta_{X_i(\omega)}\|_{\mathcal{F}}/n.$$

The first term goes to zero since $F \in L_1$ and the second equals

$$\left(\sum_{i=1}^n \|\delta_{X_i(\omega)}\|_{\mathcal{F}}/n \right) E|\widetilde{N}| I_{|\widetilde{N}| > c}.$$

But the first term in this last quantity is $P^{\mathbf{N}}$-a.s. bounded by the strong law of large numbers and the fact that $F \in L_1$. The second can be made arbitrarily small by taking c large.) $\Rightarrow P^{\mathbf{N}}$-a.s., $E_{\varepsilon,A} \left\| \Sigma_{j=1}^n \varepsilon_j \delta_{\hat{X}_{nj}^{\omega}}/n \right\|_{\mathcal{F}} \to 0$ (Proposition 2.2) $\Rightarrow P^{\mathbf{N}}$-a.s., $E_A \|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}} \to 0$ (desymmetrization).

For the converse, observe first that, as in Theorem 2.4,

$$\|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}} \to 0 \quad \omega-\text{a.s.} \quad \Rightarrow \int F dP < \infty \text{ and } E_A \|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}} \to 0 \ \omega-\text{a.s.}$$

But, by symmetrization, as in (2.24),

$$E_{\varepsilon,A}\Big\|\sum_{j=1}^{n}\varepsilon_j\delta_{\hat{X}_{nj}}/n\Big\|_{\mathcal{F}} \le 2E_A\|\hat{P}_n(\omega)-P_n(\omega)\|_{\mathcal{F}} + \Big(E\Big|\sum_{j=1}^{n}\varepsilon_j/n\Big|\Big)\|P_n(\omega)\|_{\mathcal{F}}$$

and these two variables tend to zero a.s. (note that, since $\int FdP < \infty, \|P_n(\omega)\|_{\mathcal{F}}$ is a.s. bounded). Hence Proposition 2.2 implies $E_\varepsilon\big\|\Sigma_{i=1}^n\varepsilon_i\delta_{X_i}/n\big\|_{\mathcal{F}} \to 0$ a.s. So for all $M > 0, E\big(\big\|\Sigma_{i=1}^n\varepsilon_i\delta_{X_i}/n\big\| \wedge M\big) \to 0$, i.e. $\big\|\Sigma_{i=1}^n\varepsilon_i\delta_{X_i}/n\big\|_{\mathcal{F}} \to 0$ in pr., which, since $\int FdP < \infty$, implies $\|P_n - P\|_{\mathcal{F}} \to 0$ a.s. (Giné and Zinn (1984) page 980). □

3. The bootstrapped (in probability) limit theorems. We first give the appropriate notion of bootstrap in probability in the context of empirical processes and show how it can be used.

In Giné and Zinn (1986), Theorem 1.1.3, we give a natural and short proof of: $\mathcal{F} \in CLT(P)$ iff $(\mathcal{F}, \rho_P)$ is totally bounded and the usual eventual equicontinuity condition holds. This proof actually shows that $\mathcal{F} \in CLT(P)$ iff $\mathcal{F}$ is P-pregaussian and

$$sup_{H\in BL_1(\ell^\infty(\mathcal{F}))}|E^*H(n^{1/2}(P_n-P)) - EH(G_P)| \to 0, \tag{3.1}$$

where $BL_1(\ell^\infty(\mathcal{F})) = \{H\colon\ \ell^\infty(\mathcal{F}) \to \mathbf{R}, |H(x)-H(y)| \le \|x-y\|_{\mathcal{F}}, \|H\|_\infty \le 1\}$. With some abuse of notation, we may call the quantity in (3.1),

$$d_{BL^*}(\mathcal{L}(n^{1/2}(P_n-P)), \mathcal{L}(G_P))$$

as in the case when these are true probability laws ($n^{1/2}(P_n - P)$ may not be measurable as a $\ell^\infty(\mathcal{F})$-valued random element). The above observation extends also to more general limit theorems (e.g. non-i.i.d., different normings). In particular $n^{1/2}(\hat{P}_n(\omega) - P_n(\omega)) \to G_P$ weakly in $\ell^\infty(\mathcal{F}), \omega$-a.s. iff

$$d_{BL^*}(\mathcal{L}((n^{1/2}(\hat{P}_n(\omega)-P_n(\omega)), \mathcal{L}(G_P)) \to 0 \quad \text{a.s.} \tag{3.2}$$

So, it is justifiable to say that the bootstrapped $CLT(P)$ holds in *probability* iff the limit (3.2) takes place in outer probability.

To see the usefulness of this notion, suppose that $\|P_n - P\|_{\mathcal{F}}$ is measurable, that $\|G_P\|_{\mathcal{F}}$ has a continuous distribution and that $\mathcal{F}$ satisfies both the $CLT(P)$ and the bootstrapped $CLT(P)$ in probability. Since $\overline{H} = H \circ \|\cdot\|_{\mathcal{F}} \in BL_1(\ell^\infty(\mathcal{F}))$ if $H \in BL_1(R)$, we have

$$d_{BL^*}(\mathcal{L}(n^{1/2}\|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}}), \mathcal{L}(\|G_P\|_{\mathcal{F}})) \to 0 \text{ in } pr. \tag{3.3}$$

By passing back and forth to a.s. convergent subsequences, since d_{BL^*} metrizes weak convergence in $\mathbf{R}$, we get from (3.3) that

$$\sup_{x \in \mathbf{R}} |F_{n^{1/2}\|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}}}(x) - F_{\|G_P\|_{\mathcal{F}}}(x)| \to 0 \text{ in } pr \tag{3.4}$$

(where F_ξ denotes the distribution function of the real random variable ξ). By the assumptions, we also have

$$\sup_{x \in \mathbf{R}} |F_{n^{1/2}\|P_n - P\|_{\mathcal{F}}}(x) - F_{\|G_P\|_{\mathcal{F}}}(x)| \to 0. \tag{3.5}$$

So, if $c_n(\alpha) = c_n(\alpha, \omega)$ is defined by

$$c_n(\alpha) = \inf\{t\colon\ F_{n^{1/2}\|\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}}}(t) \geq 1 - \alpha\}$$

then (3.4) and (3.5) give

$$F_{n^{1/2}\|P_n - P\|_{\mathcal{F}}}(c_n(\alpha)) \to 1 - \alpha \text{ in } pr. \tag{3.6}$$

Or one can consider P_n and $c_n(\alpha)$ defined on different probability spaces, say P_n on (Ω_1, Pr_1) and $c_n(\alpha)$ on (Ω_2, Pr_2). By (3.6) and boundedness of distribution functions, we have $E_{Pr_2} F_{n^{1/2}\|P_n - P\|_{\mathcal{F}}}(c_n(\alpha)) \to 1 - \alpha$. Therefore

$$(Pr_1 \times Pr_2)\{n^{1/2}\|P_n - P\|_{\mathcal{F}} \leq c_n(\alpha)\} \to 1 - \alpha \tag{3.7}$$

In conclusion the bootstrap in probability as described above allows the construction of asymptotic confidence regions for P.

3.1. Theorem. Assuming $\mathcal{F} \in M(P)$, the following are equivalent:

(a) $\mathcal{F} \in CLT(P)$,

(b) there exists a centered Gaussian process G on $\mathcal{F}$ whose law is Radon in $\ell^\infty(\mathcal{F})$ such that

$$d_{BL_1^*}(\mathcal{L}(n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))), \mathcal{L}(G)) \to 0 \text{ in } pr^*. \tag{3.8}$$

and if either (a) or (b) hold, then $G = G_P$, i.e. $\mathcal{F}$ satisfies the bootstrapped $CLT(P)$ in probability.

Proof. (a) $\Rightarrow$ (b). Using the decomposition (1.13) in Theorem 1.1.3, Giné and Zinn (1986), of

$$E^* H(n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))) - EH(G_P), \qquad H \in BL_1(\ell^\infty(\mathcal{F}))$$

and the bootstrapped CLT in probability of Athreya (1986), it follows that, in order to establish (3.8) it suffices to prove that

$$\lim_{\delta\to 0} \limsup_n Pr^*\{E_A \|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega)\|_{\mathcal{F}'_\delta} > \varepsilon\} = 0 \text{ for all } \varepsilon > 0. \tag{3.9}$$

Symmetrization and Proposition 2.2 give

$$\begin{aligned} E_A \|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))\|_{\mathcal{F}'_\delta} &\le 2E_{\varepsilon,A} \left\| \sum_{j=1}^n \varepsilon_j \delta_{\hat{X}^\omega_{nj}} / n^{1/2} \right\|_{\mathcal{F}'_\delta} \\ &\le 16 E_N \left\| \sum_{i=1}^n \tilde{N}_i \delta_{X_i(\omega)} / n^{1/2} \right\|_{\mathcal{F}'_\delta} \end{aligned}$$

Now, by the multiplier Lemma 1.2.4 and Theorem 1.1.8 in Giné and Zinn (loc. cit.), the above inequality yields

$$\begin{aligned} &\lim_{\delta\to 0} \limsup_n EE_A \|n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))\|_{\mathcal{F}'_\delta} \le \\ &\le 16 \|\tilde{N}\|_{2,1} \lim_{\delta\to 0} \limsup_n E \left\| \sum_{i=1}^n \varepsilon_i \delta_{X_i} / n^{1/2} \right\|_{\mathcal{F}'_\delta} = 0. \end{aligned}$$

This gives (3.9), hence (3.8) with $G = G_P$.

b) $\Rightarrow$ a). If (b) holds, for every subsequence of $\mathbf{N}$ there is a further subsequence, say $\{n_k\}$ such that

$$d_{BL^*}(\mathcal{L}(n_k^{1/2}(\hat{P}_{n_k}(\omega) - P_n(\omega), \mathcal{L}(G)) \to 0 \ \omega\text{−a.s.} \tag{3.10}$$

Then, by infinitesimality and Gaussian limits, we have, as in the proof of Theorem 2.4, for all $\delta > 0$,

$$\sum_{i=1}^{n_k} I(F(X_i(\omega)) > \delta n_k^{1/k}) \to 0 \quad \text{a.s.} \tag{3.11}$$

(= 0 eventually a.s.). This implies

$$\sum_{i=1}^{n} I(F(X_i(\omega)) > \delta n^{1/2}) \to 0 \ \text{ in } pr.$$

Now, previous arguments show that this limit holds in expectation, i.e.

$$nPr\{F(X) > \delta n^{1/2}\} \to 0. \tag{3.12}$$

For every subsequence $\{n_k\}$ for which (3.10) holds, we can use (3.11) and the converse CLT in $\mathbf{R}$ to obtain, as in the proof of Theorem 2.4,

$$\lim_{n_k \to 0} \left(\frac{\sum_{i=1}^{n_k} f(X_i)^2}{n_k} - \left(\frac{\sum_{i=1}^{n_k} f(X_i)}{n_k} \right)^2 \right) = E(G(f))^2 \quad \text{a.s.}$$

for all $f \in \mathcal{F}' \cup \mathcal{F}$. Hence this limit holds for the whole sequence $\mathbf{N}$ in probability. If $Ef^2(X) < \infty$ the limit is actually $E(G_P(f))^2$ by the law of large numbers. If $Ef^2(X) = \infty$ then by Lemma 2 in Giné and Zinn (1988) the empirical second moment dominates the square of the empirical first (absolute) moment, and we get

$$\lim_{n\to\infty} \sum_{i=1}^{n} f^2(X_i)/n = E(G(f))^2 \text{ in } pr.$$

Then, by the converse CLT (centering part), the truncated centers must converge, i.e. $Ef^2(X_1)I(|f(X_1)| \le \sqrt{n})$ converges, implying $Ef^2(X) < \infty$, contradiction. We have just proved $Ef^2(X) < \infty, f \in \mathcal{F}$, and

$$G = G_P. \tag{3.13}$$

Consider now a subsequence $\{n_k\}$ for which (3.10) holds. Then, for any $p > 0$ and $a > 0$,

$$\begin{aligned} E_A \max_{j \le n_k} (\|\delta_{\hat{X}^\omega_{nj}} - P_{n_k}(\omega)\|_{\mathcal{F}}/n_k^{1/2})^p \le \\ \le 2^p \max_{j \le n_k} (F(X_i(\omega))/n_k^{1/2})^p \\ \le 2^p \left[a + \sum_{i=1}^{n_k} \frac{F(X_i(\omega))}{n_k^{1/2}} I(F(X_i(\omega)) > a n_k^{1/2}) \right]^p \end{aligned}$$

and by (3.11) this last quantity is eventually $(2a)^p$ a.s. Hence

$$\sup_k E_A \max_{j \le n_k} \|\delta_{\hat{X}^\omega_{n_k j}} - P_{n_k}(\omega)\|^p_{\mathcal{F}}/n^p < \infty \text{ a.s.}$$

This allows us to follow for $\{n_k\}$ exactly the same steps as in the proof of (b) $\Rightarrow$ (a) in Theorem 2.4, from inequality (2.22) on, to conclude that

$$d_{BL_1^*}(\mathcal{L}\Big(\sum_{i=1}^{n_k} \varepsilon_i \delta_{X_i}/n_k^{1/2}\Big), \mathcal{L}(Z_P)) \to 0.$$

Hence, since every subsequence has a further subsequence $\{n_k\}$ for which this limit holds, we obtain

$$d_{BL_1^*}(\mathcal{L}\Big(\sum_{i=1}^{n} \varepsilon_i \delta_{X_i}/n^{1/2}\Big), \mathcal{L}(Z_P)) \to 0,$$

i.e. $\mathcal{F} \in CLT(P)$). □

3.2. Remark. A similar result holds in the case of normings $a_n \ne n^{1/2}$ and *Gaussian limits:* $\mathcal{F} \in CLT(P; a_n)$ with limit G iff $\mathcal{L}\{\sum_{j=1}^{n} (f(\hat{X}^\omega_{nj}) - P_n(\omega)(f))/a_n \colon f \in \mathcal{F}\} \to_w \mathcal{L}(G)$ in probability. The proof is analogous to that of Theorem 3.1 and is omitted. However, such a result cannot hold in the case of a stable non-Gaussian limit (Giné and Zinn (1988)).

3.3. Remark. Note that the proof of Theorem 3.1 is more elementary than the proof of Theorem 2.4: the deeper Lemma 2.3 is not needed for the bootstrap in probability.

3.4. Remark. Beran, Le Cam and Millar (1987) show that whenever a bootstrapped limit theorem holds a probability, then the empirical distributions of the bootstrapped laws also converge weakly in probability. This justifies using Montecarlo to approximate the bootstrapped distributions. Concretely Theorem 3.1 above and the Corollary in Section 4 of their paper give:

Let $\hat{\nu}_n^\omega = n^{1/2}(\hat{P}_n(\omega) - P_n(\omega))$, which is a $\ell^\infty(\mathcal{F})$-valued random variable, and for $j_n \to \infty$ consider i.i.d. copies of $\hat{\nu}_n^\omega$, say $\{\hat{\nu}_{n,j}^\omega\}_{j=1}^{j_n}$. Then, if $\mathcal{F} \in CLT(P)$, we have

$$d_{BL^*}\left(\mathcal{L}\left(\frac{1}{j_n}\sum_{j=1}^{j_n}\delta_{\hat{\nu}_{n,j}^\omega}\right), \mathcal{L}(G_P)\right) \to 0 \quad \text{in probability}$$

(in probability refers to $(\mathcal{L}(\hat{\nu}_n^\omega))^{j_n} \otimes Pr$, for each n).

Finally we show that the weak law of large numbers for empirical processes, can also be bootstrapped in probability. It may be worth mentioning that an example of $\mathcal{F}$ and P for which the $WLLN$ holds but the strong law does not hold is: $P =$ uniform distribution on [0,1), $\mathcal{F} = \{w(t)I_{(0,t]}\colon t \in (0, 1/2]\}$ with w decreasing, $tw(t) \to 0$ but $\int_0^{1/2} w(t)dt = \infty$, i.e. the weighted empirical process (Theorem 7.3 in Andersen, Giné and Zinn (1988)). Some additonal notation for Theorem 3.5: Given random variables ξ, η, d_{pr} denotes their Ky Fan distance, which metrizes convergence in probability, $d_{pr}(\xi, \eta) = \inf[\varepsilon\colon Pr\{|\xi - \eta| > \varepsilon\} < \varepsilon]$. If the random variables involve $\widehat{X}_{nj}^\omega, \varepsilon_j, N_j$, then $d_{pr_A}, d_{pr_{\varepsilon,A}}$ and d_{pr_N} indicate that the distance d_{pr} is taken with respect to the conditional probability given $X_1(\omega), \ldots, X_n(\omega)$.

3.5. Theorem. Let $\mathcal{F}$ be $NLM(P)$. The following are equivalent:

(i) $\|\Sigma_1^n(f(X_i) - PfI(F \le n))/n\|_{\mathcal{F}} \to 0$ in pr.

(ii) $d_{pr_A}(\|\Sigma_{j=1}^n(\delta_{\widehat{X}_{nj}^\omega} - P_n(\omega))/n\|_{\mathcal{F}}, 0) \to 0$ in pr.

and if (i) or (ii) holds then also

$$E_A\left\|\sum_{j=1}^n(\delta_{\widehat{X}_{nj}^\omega} - P_n(\omega))/n\right\|_{\mathcal{F}} \to 0 \quad \text{in } pr.$$

Proof. (a). (i) $\Rightarrow$ (ii): We first show (i) $\Rightarrow \|\Sigma_{i=1}^n\varepsilon_i\delta_{X_i}/n\|_{\mathcal{F}} \to 0$ in probability. To this end we note that

$$\left\| \sum_{i=1}^{n} (f(X_i) - PfI_{F\le n})/n \right\|_{\mathcal{F}} \to 0 \quad \text{in } pr.$$

implies

$$\left\| \sum_{i=1}^{n} (f(X_i) - f(X_i'))/n \right\|_{\mathcal{F}} \to 0 \quad \text{in } pr.$$

by the triangle inequality for $\| \cdot \|_{\mathcal{F}}$, where $\{X_i, X_j'\}_{i,j=1}^{\infty}$ are i.i.d. And this implies (see the proof of Corollary 2.13 in Giné and Zinn (1984) that

$$nPr(\|\delta_{X_1} - \delta_{X_1'}\|_{\mathcal{F}} > n) \to 0 \quad \text{as} \quad n \to \infty.$$

But then

$$nPr(\|\delta_{X_1}\|_{\mathcal{F}} > 2n)Pr(\|\delta_{X_1'}\|_{\mathcal{F}} \le n) \to 0 \quad \text{as} \quad n \to \infty$$

and hence

$$tPr(\|\delta_{X_1}\|_{\mathcal{F}} > t) \to 0 \quad \text{as} \quad t \to \infty.$$

Also, from symmetrization procedures (Lemma 2.7, Giné and Zinn (1984)) we know

$$\begin{aligned} Pr\Big\{ \Big\| \sum_{1}^{n} \varepsilon_i (f(X_i) - PfI(F \le n)) \Big\|_{\mathcal{F}} > \varepsilon n \Big\} \\ \le 2 \max_{k \le r} Pr\Big\{ \Big\| \sum_{1}^{k} (f(X_i) - PfI(F \le n)) \Big\|_{\mathcal{F}} > \frac{\varepsilon n}{2} \Big\} \\ + 2 \max_{r < k \le n} Pr\Big\{ \Big\| \sum_{1}^{k} (f(X_i) - PfI(F \le n)) \Big\|_{\mathcal{F}} > \frac{\varepsilon n}{2} \Big\}. \end{aligned}$$

The first term on the right goes to zero since $n \to \infty$. The second term can be made less than any $\varepsilon > 0$ if r (and therefore k) is large enough, since the $WLLN$ (i.e. (i)) is assumed to hold. Further, since $tPr(\|\delta_{X_1}\|_{\mathcal{F}} > t) \to 0$ as $t \to \infty$,

$$|PfI(F \le n)| \le \int_0^{\infty} Pr(|f(X)|I(F(X) \le n) > t)dt \le 1 + \int_1^n \frac{K}{t} dt \le K' \ln n,$$

where K and K' are fixed constants. But then

$$\left| \sum_{i=1}^{n} \frac{\varepsilon_i PfI(F \le n)}{n} \right| \le K' \frac{|\sum\limits_{i=1}^{n} \varepsilon_i|}{(n/\ln n)},$$

which converges to zero a.s. by, e.g., the Marcinkiewicz-Zymund $SLLN$. Hence, for all $\varepsilon > 0$, $Pr(\|\Sigma_{i=1}^n \varepsilon_i f(X_i)\|_{\mathcal{F}} > \varepsilon n) \to 0$ as $n \to \infty$.

(b) Le Cam's Poissonization Lemma (Le Cam, (1970); see also Araujo and Giné (1980) Lemma 3.4.8)) in probability gives

$$d_{pr_{\varepsilon,A}}(\| \sum_{j=1}^{n} \varepsilon_j \delta_{\widehat{X}_{nj}^{\omega}}/n\|_{\mathcal{F}}, 0) \le 2 d_{pr_N}(\| \sum_{i=1}^{n} \widetilde{N}_i \delta_{X_i(\omega)}/n\|_{\mathcal{F}}, 0).$$

(c) If $\| \sum\limits_{i=1}^{n} \varepsilon_i \delta_{X_i}/n\|_{\mathcal{F}} \to 0$ pr. then, as pointed out in Giné, Marcus and Zinn (1988), Remark 4.2, $\| \sum\limits_{i=1}^{n} \widetilde{N}_i \delta_{X_i}/n\|_{\mathcal{F}} \to 0$ in pr. because $E\widetilde{N}^{1+\delta} < \infty$. Hence, by (a), for all $\varepsilon > 0$

$$E_X Pr_N \left\{ \left\| \sum_{i=1}^{n} \widetilde{N}_i \delta_{X_i}(w)/n \right\|_{\mathcal{F}} > \varepsilon \right\} \to 0. \tag{3.14}$$

But, $d_{pr_N}(\| \sum\limits_{i=1}^{n} \widetilde{N}_i \delta_{X_i(\omega)}/n\|_{\mathcal{F}}, 0) \le \varepsilon \vee Pr_N\{\| \sum\limits_{i=1}^{n} \widetilde{N}_i \delta_{X_i(\omega)}/n\|_{\mathcal{F}} > \varepsilon\}$, for all $\varepsilon > 0$ by definition of the Ky Fan distance. Therefore (3.14) implies

$$E_X dpr_N \left(\left\| \sum_{i=1}^{n} \widetilde{N}_i \delta_{X_i(\omega)}/n \right\|_{\mathcal{F}}, 0 \right) \to 0.$$

Now, (b) and (c) give

$$E_X dpr_{\varepsilon,A} \left(\left\| \sum_{i=1}^{n} \varepsilon_j \delta_{\widehat{X}_{nj}^{\omega}}/n\|_{\mathcal{F}}, 0 \right) \to 0. \tag{3.15}$$

(d) Now we must desymmetrize in (3.15).

For every subsequence of $\mathbf{N}$, there exists a further subsequence, $\{n_k\}$, such that

$$\left\| \sum_{i=1}^{n_k} \varepsilon_j \delta_{\hat{X}_{n_k j}^{\omega}}/n_k \right\|_{\mathcal{F}} \to 0 \text{ in } pr_{\varepsilon,A} \quad \omega - a.s.$$

Hence $\sum\limits_{i=1}^{n_k} I(\|X_i\| > a n_k) = 0$ eventually a.s., for all $a > 0$. Therefore,

$$E_A \max_{j\le n_k} \frac{\|\delta_{\hat X^\omega_{nj}}\|_{\mathcal F}}{n_k} \le a + \int_a^\infty \sum_1^{n_k} I(\|X_i\| > n_k t)dt \le a \text{ eventually a.s.}$$

$$\Rightarrow E_A \max_{j\le n_k} \|\delta_{\hat X^\omega_{n_k j}}\|_{\mathcal F}/n_k \to 0\omega\text{-a.s.}$$

$$\Rightarrow E_{\varepsilon,A}\Big\|\sum_{j=1}^{n_k} \varepsilon_j \delta_{\hat X^\omega_{n_k j}}/n_k\Big\|_{\mathcal F} \to 0, \omega\text{-a.s. (by (3.15) and Hoffmann-Jørgensen's inequality)}$$

$$\Rightarrow E_A\Big\|\sum_1^{n_k}(\delta_{\hat X^\omega_{n_k j}} - P_{n_k}(\omega))/n_k\Big\|_{\mathcal F} \to 0 \quad \omega\text{–a.s. (see the inequalities following (2.15))}$$

$$\Rightarrow E_A\Big\|\sum_{j=1}^{n}(\delta_{\hat X^\omega_{nj}} - P_n(\omega))/n\Big\|_{\mathcal F} \to 0 \text{ in probability}$$

which is even more than the actual statement (ii).

(ii) $\Rightarrow$ (i). If (ii) holds, we obtain as in the CLT that

$$nPr(F > n) \to 0. \tag{3.16}$$

Recall that for any $\{n_k\}$ for which $d_{pr_A}\big(\|\Sigma_1^{n_k}(\delta_{\hat X^\omega_{n_k j}} - P_{n_k}(\omega)/n_k\|_{\mathcal F}, 0\big) \to 0$ a.s.

$$\sum_{i=1}^{n_k} I(F(X_i) > an_k) = 0 \text{ eventually, a.s.} \tag{3.17}$$

So, as above, $E_A \max_{j\le n_k} \Big\|\frac{\delta_{\widehat{X^\omega_{nj}}}}{n}\Big\|_{\mathcal F} \to 0$ a.s. And also, $\|P_{n_k}(w)/n_k\|_{\mathcal F} = \|\sum_{i=1}^{n_k} f(X_i)/n_k^2\|_{\mathcal F} \to 0$ (since, eventually, this norm is $\le a$). So, $E_A \max_{j\le n_k} \Big\|\frac{\delta_{\hat X^\omega_{nj}} - P_{n_k}(w)}{n_k}\Big\|_{\mathcal F} \to 0$ a.s. Hence, by Hoffmann-Jørgensen's inequality,

$$E_A\Big\|\frac{\sum_1^{n_k}(\delta_{\hat X^\omega_{n_k j}} - P_{n_k}(w))}{n_k}\Big\| \to 0 \quad \text{a.s.} \tag{3.18}$$

Now, as in (2.24),

$$cE_\varepsilon\Big\|\sum_{i=1}^{n_k}\varepsilon_i\delta_{X_i}/n_k\Big\|_{\mathcal F} \le 2E_A\|(\hat P_{n_k}(w) - P_{n_k}(w)\|_{\mathcal F} + \|P_{n_k}(w)\|_{\mathcal F}/n_k^{1/2}.$$

So,

$$\overline{\lim} Pr\Big\{\Big\|\sum_{i=1}^{n_k}\varepsilon_i\delta_{X_i}/n_k\Big\|_{\mathcal{F}} \geq \varepsilon\Big\} \leq \frac{1}{\varepsilon}\overline{\lim} E[(E_A\|(\hat{P}_{n_k}(w) - P_{n_k}(w)\|_{\mathcal{F}} \wedge \varepsilon]$$

$$+ \frac{1}{\varepsilon}E\Big(\frac{\|P_{n_k}(w)\|_{\mathcal{F}}}{n_k^{1/2}} \wedge \varepsilon\Big) = \text{(I)} + \text{(II)}.$$

(I) $\to 0$ by (3.18) and the dominated convergence theorem, and (II) $\to 0$ because, by (3.16), $\sum_{i=1}^{n} F(X_i)/n^{3/2} \to 0$ in probability. Hence,

$$\Big\|\sum_{i=1}^{n}\varepsilon_i\delta_{X_i}/n\Big\|_{\mathcal{F}} \to 0 \text{ in probability.} \tag{3.19}$$

Finally, (i) follows by a stanadrd desymmetrization: $\|\sum_{i=1}^{n} I(F(X_i) > n)\delta_{X_i}/n\|_{\mathcal{F}} \to 0$ in probability by (3.16), hence we can truncate in (3.19) and then take expectations and use the symmetrization inequalities given immediately after (2.15) to obtain $E\|\sum_{i=1}^{n}(f(X_i)I(F(X_i) \leq n) - PfI(F \leq n))/n\|_{\mathcal{F}} \to 0$. Again, using (3.16) we obtain (i). □

3.6. Remark. The weak law of large numbers with normings other then n (i.e. $n^{1/p}$ or even more general a_k's) can also be bootstrapped in probability, in complete analogy with Theorem 3.6. (See e.g. Andersen et al. (1988) for examples of Marcinkiewicz type laws of large numbers for empirical processes.)

Acknowledgement. The authors would like to thank Professor Richard M. Dudley for a careful reading of this paper and for motivating us to clarify the exposition.

References

de Acosta, A. and Giné, E. (1979). Convergence of moments and related functionals in the general central limit theorem in Banach spaces. Zeits. Wahrs. verw. Geb. 48, 213-231.

Alexander, K. (1987). The central limit theorem for empirical processes on Vapnik-Cervonenkis classes. Ann. Probability 15, 178-203.

Andersen, N.T., Giné, E. and Zinn, J. (1986). The central limit theorem for empirical processes under local conditions: the case of Radon limits without Gaussian component. Trans. Amer. Math. Soc., 308, 603-635.

Araujo, A. and Giné, E. (1980). The central limit theorem for real and Banach valued random variables. Wiley, New York.

Beran, R. (1982). Estimated sampling distributions: the bootstrap and competitors. Ann. Statistics 10, 212-225.

Beran, R., Le Cam, L. and Millar, P.W. (1987), Convergence of Stochastic Empirical Measures. J. Multivariate Analysis 23, 159-168.

Beran, R. and Millar, P.W. (1986). Confidence sets for a multivariate distribution. Ann. Statistics 14, 431-443.

Bickel, P.J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. Ann. Statistics 9, 1196-1216.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionelles. Ann. Inst. H. Poincaré 19, 281-296.

Dudley, R.M. (1984). A course on empirical processes. Lect. Notes in Math. 1097, 2-142. Springer, Berlin.

Efron, B. (1979). Bootstrap methods: another look at the Jackknife. Ann. Statistics 7, 1-26.

Fernique, X. (1974). Regularité des trajectoires des fonctions aléatoires gaussiennes. Lect. Notes in Math. 480, 1-96. Springer, Berlin.

Gaenssler, P. (1986). Bootstrapping empirical measures indexed by Vapnik-Cervonenkis classes of sets. Prob. Theory and Math. Stat. 467-481. Prohorov et. al. (eds.) VNU Press, The Netherlands.

Giné, E., Marcus, M.B. and Zinn, J. (1986). On random multipliers in the central limit theorem with p-stable limit $0 < p < 2$. Lect. Notes in Math., to appear.

Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. Ann. Probability 12, 929-989.

Giné, E. and Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. Lect. Notes in Math. 1221, 50-113. Springer, Berlin.

Giné, E. and Zinn, J. (1988). Necessary conditions for the bootstrap of the mean. Ann. Statistics, to appear.

Haagerup, U. (1981). The best constants in the Khinchine inequality. Studia. Math. 70, 231-283.

Hoffmann-Jørgensen, J. (1974). Sums of independent Banach space valued random variables. Studia Math. 52, 159-186.

Hoffmann-Jørgensen, J. (1984). Stochastic processes on Polish spaces. To appear.

Le Cam, L. (1970). Remarques sur le théorème limite central dans les espaces localement convexes. In Les Probabilités sur les Structures Algébriques, pp. 233-249. CNRS, Paris.

Ledoux, M. and Talagrand, M. (1988). Characterization of the law of the iterated logarithm in Banach spaces. Ann. Probability, 16, 1242-1264.

Ledoux, M. and Talagrand, M. (1988). Un critère sur les petites boules dans le théorème limite central. Probability Theory and Rel. Fields, 77, 29-47.

Pisier, G. (1975). Le théorème limite central et la loi du logarithme iteré dans les espaces de Banach. Séminaire Maurey-Schwartz 1975-1976, exp. III et IV. Ecole Polytechnique, Paris.

Pollard, D. (1981). Limit theorems for empirical processes. Zeits. Wahrs. verw. Geb. 57, 181-185.

Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. Ann. Statist. 9, 1187-1195.

Szarek, S.J. (1976). On the best constant in the Khinchine inequality. Studia Math. 58, 197-208.

The College of Staten Island of CUNY
Department of Mathematics
l30 Stuyvesant Place
Staten Island, NY 10301

Texas A&M University
Department of Mathematics
College Station, TX 77843