Nonparametric Statistics and Related Topics
A.K.Md.E. Saleh (Editor)
1992 Elsevier Science Publishers B.V.

BOOTSTRAP LIMIT THEOREMS: A PARTIAL SURVEY

Jon A. Wellner[1,2]

[1] Department of Statistics GN-22, University of Washington, Seattle, Washington, 98195

[2] Research partially supported by: National Science Foundation Grant DMS-8723011

Abstract:

Bootstrap resampling methods have earned an important place in the statistician's toolkit since their systematic introduction by Efron (1979). One useful and basic way of validating a particular bootstrap method is by proving that it is consistent: conditionally on the observed data the bootstrap distribution has the same asymptotic behavior as the (centered, standardized) sampling distribution of the original estimator either a.s. or in probability. Considerable progress has been made during the past few years in validating various bootstrap methods via such consistency limit theorems, notably the elegant results of Giné and Zinn (1990) for bootstrapping empirical processes. Here we review limit theory for a variety of bootstrap methods including Efron's bootstrap, the Bayesian bootstrap, and various parametric or model - based bootstrap methods, with emphasis on results for general bootstrap empirical processes. We also present several new results -- in particular a limit theorem for sequential bootstrap sampling which yields joint in bootstrap sample size limit distributions -- and briefly review some open problems.

1. Introduction

Bootstrap resampling methods have become an important tool in statistics since their introduction by Efron (1979), (1982). The new Cumulative Index to IMS Scientific Journals 1960 - 1989 (Trumbo and Burdick (1990)) lists 57 articles on the bootstrap in IMS publications alone, and this represents just a small fraction of the research effort on the bootstrap and variations thereon over the past 12 years.

One important variation on the bootstrap with which we will be concerned here is the "Bayesian bootstrap" introduced by Rubin (1981).

Bickel and Freedman (1981) carried out an extensive asymptotic analysis of Efron's nonparametric bootstrap for iid real - valued data, and, in particular, showed that the bootstrap empirical and quantile processes are asymptotically consistent; i.e. the asymptotic behavior of the original processes is a.s. replicated by the Efron's nonparametric bootstrap. Their consistency for empirical and quantile processes in one dimension have been considerably strengthened by Csörgő and Mason (1989). Gaenssler (1986) gave the first validation of the bootstrap for general empirical processes: he treated the empirical process indexed by a Vapnik - Chervonenkis class of sets. Lo (1987) gave an asymptotic justification for Rubin's Bayesian bootstrap in the classical one-dimensional setting. His results in one dimension have been strengthened (in the much the same way that Csörgo and Mason (1989) strengthens Bickel and Freedman (1981)) by Einmahl and Mason (1991) following preparatory work by Mason and Newton (1990).

Our object here is to give a brief survey of recent progress in validating various bootstrap methods, including Efron's nonparametric bootstrap, Rubin's (1981) Bayesian bootstrap and generalizations thereof, and parametric or model - based bootstrap methods, with emphasis on results for general bootstrap empirical processes: Instead of considering real - valued random variables, we consider, as in the theory of general empirical processes, sampling from a probability measure P on an arbitrary measurable space $(A, \mathbf{A})$.

We also include several new results: in section 2 we examine the joint behavior of several different bootstrap methods; in section 4 we present a theorem concerning the "joint in sample size" behavior of Efron's bootstrap.

For a more comprehensive review of bootstrap methods, see e.g. Swanepoel (1990).

Now we introduce some notation and terminology from the theory of empirical processes which we will use throughout the paper; see also Giné and Zinn (1984), (1986), (1990).

Let $(A, \mathbf{A}, P)$ be a probability space, and let $X_1, \cdots, X_n, \cdots$ be iid P. Suppose that $\mathbf{F}$ is a collection of real-valued measurable functions on A. Let

$$\boldsymbol{P}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \equiv \text{the empirical measure},$$

and

$$\boldsymbol{X}_n \equiv \sqrt{n}\,(\boldsymbol{P}_n - P) \equiv \text{the empirical process}$$

considered as elements of $l^\infty(\mathbf{F})$, the space of all bounded real valued functions on $\mathbf{F}$. Here

$$\boldsymbol{P}_n(f) \equiv \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad \text{and} \quad P(f) \equiv \int f\,dP, \quad f \in \mathbf{F}.$$

Note that $P \in l^\infty(\mathbf{F})$ if and only if $\|P\|_{\mathbf{F}} \equiv \sup_{f \in \mathbf{F}} |P(f)| < \infty$.

Two centered Gaussian processes in $l^\infty(\mathbf{F})$ which arise in the following are:

$$\mathbb{Z} \equiv \text{a P - Brownian motion process } (Z_P)$$

with

$$Cov(\mathbb{Z}(f), \mathbb{Z}(g)) = P(fg), \qquad f, g \in \mathbf{F};$$

and

$$\boldsymbol{X} \equiv \text{a P - Brownian bridge process } (G_P)$$

with

$$Cov(\boldsymbol{X}(f), \boldsymbol{X}(g)) = P(fg) - P(f)P(g), \qquad f, g \in \mathbf{F}.$$

Let ρ_P and e_P denote the natural Gaussian pseudometrics on $\mathbf{F}$ corresponding to $\boldsymbol{X}$ and $\mathbb{Z}$ respectively: for $f, g \in \mathbf{F}$

$$\rho_P^2(f,g) \equiv Var_P(f(X) - g(X)), \qquad e_P^2(f,g) \equiv E_P(f(X) - g(X))^2.$$

Note that since $\mathbb{Z}$ is P – Brownian motion, then

$$\mathbb{Z} - \mathbb{Z}(1)P \quad \text{is a P - Brownian bridge process } G_P,$$

and, on the other hand, if the P – Brownian bridge process $\boldsymbol{X}$ and a standard normal random variable Z are independent, then

$$\boldsymbol{X} + ZP \quad \text{is a P - Brownian motion process } Z_P.$$

We say that $\mathbf{F}$ is Z_P – *pregaussian* if $\mathbb{Z}$ can be chosen so that $\mathbb{Z} \in C(\mathbf{F}, e_P)$ a.s. where $C(\mathbf{F}, e_P)$ is the collection of e_P – uniformly continuous functions in $l^\infty(\mathbf{F})$. For short this will be written as $\mathbf{F} \in PG(Z_P)$. Similarly, $\mathbf{F}$ is

G_P – *pregaussian* if $\mathbb{X}$ can be chosen so that $\mathbb{X} \in C(\mathbf{F}, \rho_P)$ a.s. where $C(\mathbf{F}, \rho_P)$ is the collection of ρ_P – uniformly continuous functions in $l^\infty(\mathbf{F})$, and in this case we write $\mathbf{F} \in PG(G_P)$.

2. Limit Theorems for Nonparametric Bootstrapping

Let $(A, \mathbf{A}, P)$ be a probability space, and let $X_1, X_2, \cdots$ be iid P. Let

$$\mathbb{P}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \equiv \text{the empirical measure on } A,$$

and let

$$\mathbb{X}_n \equiv \sqrt{n}(\mathbb{P}_n - P) \equiv \text{the empirical process}.$$

We consider $\mathbb{P}_n$ and $\mathbb{X}_n$ as processes indexed by functions f in a class $\mathbf{F} \subset L_2(P)$.

For any given $n \geq 1$ and a given sequence of X_i's, $X_1(\omega), X_2(\omega), \cdots$ denote the empirical measure of the first n $X_i(\omega)$'s by $\mathbb{P}_n^\omega$. Let

$$X_1^*, \cdots, X_n^*$$

be a "bootstrap sample" from $\mathbb{P}_n^\omega$. Then the bootstrap empirical process $\mathbb{X}_n^*$ is

$$\begin{aligned}
\mathbb{X}_n^* &= \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n^\omega) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^*} - \mathbb{P}_n^\omega\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} M_{ni}^* \delta_{X_i(\omega)} - \mathbb{P}_n^\omega\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} (M_{ni}^* - 1)\delta_{X_i(\omega)}
\end{aligned}$$

where

$$\underline{M}_n^* \sim Mult_n(n, (\frac{1}{n}, \cdots, \frac{1}{n})) \qquad \text{is independent of the } X_i\text{'s}.$$

Note that the components M_{ni}^* of $\underline{M}_n^*$ are marginally just Binomial(n, $1/n$) random variables, and hence they converge in distribution (marginally) to Poisson(1) random variables. Furthermore the components of $\underline{M}_n^*$ are nearly uncorrelated since $Cov[M_{ni}^*, M_{nj}^*] = -1/n^2$. Thus the bootstrap empirical process $\mathbb{X}_n^*$ can be thought of as essentially equivalent to the process

$$\mathbb{Z}_n^* \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} (Y_i - 1)\delta_{X_i(\omega)}$$

where $Y_1, Y_2, \cdots$ are iid Poisson(1) random variables. Another way of arriving at the process $\mathbb{Z}_n^*$ is via "Poissonization" of the bootstrap sample size: let $N_n^* \sim Poisson(n)$ be independent of the X_i's and the X_i^*'s. If

$$\underline{1}_j \equiv (1_{1j}, \cdots, 1_{nj}) \sim Mult_n(1, (\frac{1}{n}, \cdots, \frac{1}{n})), \quad j = 1, 2, \cdots$$

are iid, then

$$\underline{M}_k^* = \sum_{j=1}^{k} \underline{1}_j \sim Mult_n(k; (\frac{1}{n}, \cdots, \frac{1}{n})),$$

and

$$\underline{M}_{N_n^*}^* \sim (Y_1, \cdots, Y_n)$$

where $Y_1, \cdots, Y_n$ are iid Poisson(1). Thus the "Poissonized bootstrap empirical process" $\mathbb{Z}_n^*$ defined by

$$\mathbb{Z}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (M_{N_n^*,i}^* - 1)\, \delta_{X_i(\omega)}$$
$$\stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - 1)\, \delta_{X_i(\omega)} .$$

Now note that we can rewrite $\mathbb{Z}_n^*$ as

$$\mathbb{Z}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - 1)(\delta_{X_i(\omega)} - P) + \sqrt{n}\,(\overline{Y}_n - 1)P$$
$$= \mathbf{Y}_n^* + \sqrt{n}\,(\overline{Y}_n - 1)P$$

where the $(Y_i - 1)$'s are iid real-valued, mean-zero "multipliers," and the sequence $(\delta_{X_i} - P)$ satisfies the CLT in the Banach space $l^\infty(\mathbf{F})$. For "multiplier sums" such as in $\mathbf{Y}_n^*$, Ledoux and Talagrand (1988) and Praestgaard (1990) have proved an interesting "multiplier central limit theorem" which we now state.

To prepare the way, we need just a little bit more notation and terminology: Let F be $F(x) = (\sup_{f \in \mathbf{F}} |f(x)|)^*$, $x \in A$, be the envelope function of $\mathbf{F}$. Here, if $h : A \to R$ is an arbitrary function, then h^* denotes the least measurable function dominating h; see e.g. Dudley (1984) or (1985).

For our main results we also need a measurability assumption which insures that $\|\mathbf{X}_n\|_{\mathbf{F}'(\delta,\rho_P)}$ is completion measurable under P and Fubini's theorem can be applied to $\|\sum_{i=1}^n Y_i \delta_{X_i}\|_{\mathbf{F}'(\delta,\rho_P)}$ where the Y_i are iid real-valued mean zero rv's independent of the X_i's. In the terminology of Giné and Zinn (1984, 1986, 1990), we require $\mathbf{F}$ to be nearly linearly deviation measurable for P, or $\mathbf{F} \in NLDM(P)$ for short, and that both $\mathbf{F}^2$ and $\mathbf{F}'^2$ be nearly linearly supremum measurable, or $NLSM(P)$. When all of these hold, we say $\mathbf{F} \in M(P)$. It is known that $\mathbf{F} \in M(P)$ if $\mathbf{F}$ is countable, or if the empirical processes $\mathbf{X}_n$ are stochastically separable, or if $\mathbf{F}$ is image admissible Suslin (see Giné and Zinn (1990), pages 853, 854).

Theorem 2.1. (Ledoux and Talagrand (1988); Praestgaard (1990)). Suppose that:

(i) $\mathbf{F} \in M(P)$ and $\|P\|_{\mathbf{F}} < \infty$.

(ii) $\eta_1, \eta_2, \cdots$ are iid with $E\eta_1 = 0$, $Var(\eta_1) = \sigma_\eta^2 < \infty$, and $\eta_1 \in L_{2,1}$; i.e.

$$\int_0^\infty \sqrt{Pr(|\eta_1| > t)}\, dt < \infty .$$

Then the following are equivalent:

A. $\mathbf{F} \in CLT(P)$ and $PF^2 < \infty$.

B. $\mathbf{F} \in PG(P)$ and $n^{-1/2} \sum_{i=1}^n \eta_i (\delta_{X_i(\omega)} - P) \Rightarrow \sigma_\eta G_P$ in $l^\infty(\mathbf{F})$ for P^∞ – a.e. ω.

Ledoux and Talagrand (1988) prove theorem 2.1 in the case of a separable Banach space $\mathbf{B}$ and with η_i iid standard normal; they also indicated the extension to non-Gaussian, symmetric multipliers η satisfying the $L_{2,1}$ condition. Praestgaard (1990) carried out the generalization to the nonseparable Banach space $l^\infty(\mathbf{F})$ and possibly asymmetric multipliers η_i as stated in theorem 2.1. Also see Ledoux and Talagrand (1991), theorem 10.14, page 293.

In view of theorem 2.1 and the preceding discussion, the following limit theorem for the bootstrap empirical process seems very natural.

Theorem 2.2. (CLT for Efron's bootstrap, Giné and Zinn (1990)). Suppose that: $\mathbf{F} \in M(P)$, $\|P\|_{\mathbf{F}} < \infty$, and that $\mathbf{F}$ has envelope function F. Then the following statements are equivalent:

A. $\mathbf{F} \in CLT(P)$ and $P(F^2) < \infty$.

B. $\mathbb{Z}_n^* \Rightarrow \mathbb{Z}^* \sim Z_P$ a.s. P^∞ in $l^\infty(\mathbf{F})$.

C. $\mathbf{X}_n^* \Rightarrow \mathbf{X}^* \sim G_P$ a.s. P^∞ in $l^\infty(\mathbf{F})$.

The equivalence of A and C is due to Giné and Zinn (1990) (without the additional hypothesis $\|P\|_{\mathbf{F}} < \infty$), while the equivalence of A and B follows from Praestgaard's (1990) extension of the Ledoux - Talagrand theorem 2.1. Klaassen and Wellner (1992) show that B and C are equivalent. Hence there is a direct link between the multiplier CLT of Ledoux and Talagrand (1988) and the bootstrap CLT of Giné and Zinn (1990).

This "multiplier" perspective on Efron's (1979) bootstrap makes it clear that multipliers other than M_{ni}^* may be of interest and importance. In fact, another interesting choice of multipliers is as follows: let $T_1, T_2, \cdots$ be iid positive rv's with $ET_1 = 1$. For $n = 1, 2, \cdots$ define

$$W_{ni} \equiv \frac{T_i}{T_1 + \cdots + T_n}, \quad i = 1, \cdots, n. \tag{2.1}$$

Then $0 \le W_{ni} < 1$ and $\sum_{i=1}^n W_{ni} = 1$. The *general weighted bootstrap empirical measure* $\mathbf{P}_n^W$ and *general weighted bootstrap empirical process* $\mathbf{X}_n^W$ are defined by

$$\mathbf{P}_n^W \equiv \sum_{i=1}^n W_{ni}\, \delta_{X_i(\omega)}$$

and

$$\mathbf{X}_n^W \equiv \sqrt{n}\,(\mathbf{P}_n^W - \mathbf{P}_n^\omega)$$

When the T_i 's are exponential(1),

$$(W_{n1}, \cdots, W_{nn}) \overset{d}{=} (D_{n1}, \cdots, D_{nn}) \tag{2.2}$$

where $D_{ni} \equiv U_{n-1,i} - U_{n-1,i-1}$, $i = 1, \cdots, n$ are the spacings of a sample of $n-1$ uniform(0,1) random variables. Hence for this choice of the T_i 's, $\mathbf{P}_n^W$ is the "Bayesian bootstrap" of Rubin (1981). It yields the posterior distribution of P under a flat (noninformative, improper) prior for P.

To use the Ledoux and Talagrand theorem 1 to study $\mathbf{X}_n^W$, we note that it can be expressed in terms of the iid T_i 's as

$$\mathbf{X}_n^W = \frac{1}{\overline{T}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^n (T_i - 1)(\delta_{X_i(\omega)} - P) + \sqrt{n}\left(\frac{1}{\overline{T}_n} - 1\right)(\mathbf{P}_n^\omega - P) \tag{2.3}$$

where the first term is exactly of the form dealt with by theorem 2.1, and the second term is easy to handle by a Glivenko - Cantelli theorem for $\mathbf{P}_n$ and the ordinary central limit theorem. This motivates the following theorem of Praestgaard (1990):

Theorem 2.3. (CLT for the general weighted bootstrap; Praestgaard (1990)). Suppose that:

(i) $\mathbf{F} \in M(P)$ and $\|P\|_{\mathbf{F}} < \infty$.

(ii) T_i are iid positive random variables with $E\, T_i = 1$, $Var(T_i) = \sigma^2 < \infty$, and $T_1 \in L_{2,1}$:

$$\int_0^\infty \sqrt{Pr(T_1 > t)}\, dt < \infty.$$

Then the following are equivalent:

A. $\mathbf{F} \in CLT(P)$ and $P(F^2) < \infty$.

B. $\mathbb{X}_n^W \Rightarrow \sigma G_P$ in $l^\infty(\mathbf{F})$ for $P^\infty -$ a.e. ω.

The upshot of theorems 2.2 and 2.3 is that there are many different ways of obtaining a bootstrap empirical process which mimics the behavior of the empirical process $\mathbb{X}_n$ for large n: every choice of the iid sequence $\{T_i\}$ satisfying the hypotheses (ii) of theorem 2.3 yields a different bootstrap. While theorems 2.2 and 2.3 tell us that these different bootstrap methods are in a sense the same, it remains to describe how the various methods differ, both asymptotically and for finite sample sizes.

In the classical case of $A = R$ and $\mathbf{F} = \{1_{(-\infty,t]} : t \in R\}$, Lo (1987) studied the large sample behavior of Rubin's Bayesian bootstrap. To better understand the differences between $\mathbb{P}_n^*$ and $\mathbb{P}_n^W$, we now consider Lo's construction of $\mathbb{P}_n^*$ and $\mathbb{P}_n^W$ in the classical case. Now our T_i's are iid exponential(1) random variables so that (2.2) holds, and we write $\mathbb{P}_n^{**}$ for $\mathbb{P}_n^W$. In fact, we will use a sequence of uniform(0,1) random variables and the spacings thereof to form both $\mathbb{P}_n^*$ and $\mathbb{P}_n^W \equiv \mathbb{P}_n^{**}$.

To be specific, let $X_1, \cdots, X_n$ be iid $F = P(-\infty, \cdot]$ (this F is the usual distribution function, *not* the envelope function for an indexing class of functions; here the envelope function is simply the constant function 1), let $\mathbb{F}_n \equiv n^{-1}\sum_{i=1}^n 1_{(-\infty,\cdot]}(X_i)$, and let $X_{n:1} \le \cdots \le X_{n:n}$ denote the order statistics corresponding to $X_1, \cdots, X_n$. Let $\xi_1, \cdots, \xi_n, \cdots$ be iid Uniform(0,1) random variables, let $\mathbb{G}_n(t) \equiv n^{-1}\sum_{i=1}^n 1_{[0,t]}(\xi_i)$, write $\mathbb{G}_n(s,t] \equiv \mathbb{G}_n(t) - \mathbb{G}_n(s)$ for $0 \le s \le t \le 1$, and let

$$0 \equiv \xi_{n:0} \le \xi_{n:1} \le \cdots \le \xi_{n:n} \le \xi_{n:n+1} \equiv 1$$

be the ordered ξ's. Then $D_{ni} \equiv \xi_{n-1:i} - \xi_{n-1:i-1}$, $i = 1, \cdots, n$, are the spacings of the first $n-1$ uniform random variables ξ_i. Let $\mathbb{F}_n^{**}$ be the distribution function which puts mass D_{ni} at $X_{n:i}(\omega)$, $i = 1, \cdots, n$. Thus

$$\mathbb{F}_n^{**}(t) = \sum_{i=1}^n D_{ni}\, \delta_{X_{n:i}(\omega)}(-\infty, t] = \sum_{i=1}^n \xi_{n-1:i}\, 1_{[X_{n:i}(\omega), X_{n:i+1}(\omega))}(t). \tag{2.4}$$

Furthermore, since

$$(n\mathbb{G}_n(\frac{i-1}{n}, \frac{i}{n}],\ i = 1, \cdots, n) \sim Mult_n(n, (\frac{1}{n}, \cdots, \frac{1}{n})),$$

we can construct Efron's nonparametric bootstrap empirical distribution function as

$$\mathbb{F}_n^*(t) \equiv \mathbb{P}_n^*(-\infty, t] = \sum_{i=1}^n \mathbb{G}_n(\frac{i-1}{n}, \frac{i}{n}]\, \delta_{X_{n:i}(\omega)}(-\infty, t]$$

$$= \sum_{i=1}^n \mathbb{G}_n(\frac{i}{n})\, 1_{[X_{n:i}(\omega), X_{n:i+1}(\omega))}(t). \tag{2.5}$$

Then

$$\mathbb{X}_n^*(t) \equiv \sqrt{n}\,(\mathbb{F}_n^*(t) - \mathbb{F}_n^\omega(t)) = \sum_{i=1}^n \sqrt{n}\,(\mathbb{G}_n(\frac{i}{n}) - \frac{i}{n})\, 1_{[X_{n:i}(\omega), X_{n:i+1}(\omega))}(t),$$

$$\mathbb{X}_n^{**}(t) \equiv \sqrt{n}\,(\mathbb{F}_n^{**}(t) - \mathbb{F}_n^{\omega}(t))$$
$$= \sum_{i=1}^{n} \sqrt{n}\,(\mathbb{G}_n^{-1}(\frac{i}{n-1}) - \frac{i}{n})\, 1_{[X_{n:i}(\omega),X_{n:i+1}(\omega))}(t)\,,$$

and

$$\mathbb{X}_n^{*}(t) + \mathbb{X}_n^{**}(t)$$
$$= \sum_{i=1}^{n} \{\sqrt{n}\,(\mathbb{G}_n^{-1}(\frac{i}{n-1}) - \frac{i}{n}) + \sqrt{n}\,(\mathbb{G}_n^{-1}(\frac{i}{n-1}) - \frac{i}{n})\}\, 1_{[X_{n:i}(\omega),X_{n:i+1}(\omega))}(t)\,,$$
$$= \sum_{i=1}^{n} \{\mathbb{U}_n(\frac{i}{n}) + \mathbb{V}_{n-1}(\frac{i}{n-1}) + \sqrt{n}\,(\frac{i}{n-1} - \frac{i}{n})\}\, 1_{[X_{n:i}(\omega),X_{n:i+1}(\omega))}(t) \qquad (2.6)$$

where

$$\mathbb{U}_n \equiv \sqrt{n}\,(\mathbb{G}_n - I)\,, \qquad \mathbb{V}_n \equiv \sqrt{n}\,(\mathbb{G}_n^{-1} - I)\,.$$

Now we know from Bahadur and Kiefer that

$$\limsup_{n\to\infty} \frac{n^{1/4}\|\mathbb{U}_n + \mathbb{V}_n\|}{\sqrt{b_n \log n}} = \frac{1}{\sqrt{2}} \quad \text{a.s.}\,; \qquad (2.7)$$

see theorem 15.1.2, page 586, Shorack and Wellner (1986). Hence it is not surprising that Efron's bootstrap empirical process $\mathbb{X}_n^*$ and Rubin's Bayesian bootstrap empirical process are related in the same way:

Theorem 2.4. (Connection/difference between $\mathbb{F}_n^*$ and $\mathbb{F}_n^{**}$ when $A = R$). Suppose that $\mathbb{F}_n^*$ and $\mathbb{F}_n^{**}$ are as constructed in (2.3) and (2.4). Then, with $b_n \equiv \sqrt{2\log\log n}$,

$$\limsup_{n\to\infty} \frac{n^{1/4}\|\mathbb{X}_n^* + \mathbb{X}_n^{**}\|}{\sqrt{b_n \log n}} \qquad (2.8)$$
$$= \limsup_{n\to\infty} \frac{n^{3/4}\sup_{-\infty<t<\infty} |\mathbb{F}_n^*(t) + \mathbb{F}_n^{**}(t) - 2\mathbb{F}_n^{\omega}(t)|}{\sqrt{b_n \log n}}$$
$$= \frac{1}{\sqrt{2}} \quad \text{a.s.}$$

Corollary 2.1. If $\mathbb{F}_n^*$ and $\mathbb{F}_n^{**}$ are constructed as in (2.3) and (2.4), then $(\mathbb{X}_n^*, \mathbb{X}_n^{**}) \Rightarrow (\mathbb{U}(F), -\mathbb{U}(F))$ in $l^\infty(R)$ $P^\infty -$ a.s. where $\mathbb{U}$ is a standard Brownian bridge process. Hence the "average bootstrap estimator" $\mathbb{F}_n^{ave} \equiv (\mathbb{F}_n^* + \mathbb{F}_n^{**})/2$ satisfies $\sqrt{n}\,(\mathbb{F}_n^{ave} - \mathbb{F}_n^{\omega}) \Rightarrow 0$ a.s.

The point is that depending on the way in which several different bootstrap methods are constructed jointly, we may or may not be able to combine them to obtain yet another bootstrap estimator: $\mathbb{F}_n^{ave}$ above fails miserably at the goal of mimicking the behavior of $\sqrt{n}\,(\mathbb{F}_n - F)$; yet if we had chosen the weights independently (i.e. based on two independent sequences of uniform(0,1) rv's), then the resulting "average bootstrap empirical df" works just fine. We will return to this point below.

Proof of theorem 2.4. Set $d_n \equiv n^{1/4}/\sqrt{b_n \log n}$. The right side of (2.6) may be written as $I_n + II_n + III_n$ where

$$I_n(t) \equiv \sum_{i=1}^{n} \{\mathbb{U}_{n-1}(\frac{i}{n-1}) + \mathbb{V}_{n-1}(\frac{i}{n-1})\}\, 1_{[X_{n:i}(\omega),X_{n:i+1}(\omega))}(t)\,, \qquad (a)$$

$$II_n(t) \equiv \sum_{i=1}^{n} \{\mathbb{U}_n(\frac{i}{n}) - \mathbb{U}_{n-1}(\frac{i}{n-1})\} 1_{[X_{n:i}(\omega), X_{n:i+1}(\omega))}(t) , \tag{b}$$

and

$$III_n(t) = \sum_{i=1}^{n} \{\sqrt{n}\, i\, (\frac{1}{n-1} - \frac{1}{n})\} 1_{[X_{n:i}(\omega), X_{n:i+1}(\omega))}(t) . \tag{c}$$

Now $d_n \|III_n\| \le d_n n^{3/2} n^{-1} (n-1)^{-1} = o(1)$. Furthermore,

$$d_n \|I_n\| \le d_n \|\mathbb{U}_{n-1} + \mathbb{V}_{n-1}\| , \tag{d}$$

where the limsup of the right side is exactly $1/\sqrt{2}$ by the Bahadur - Kiefer theorem (2.7), and finally,

$$\begin{aligned} d_n \|II_n\| &\le d_n \{\omega_{\mathbb{U}_{n-1}}(\frac{1}{n-1}) + \|\mathbb{U}_n - \mathbb{U}_{n-1}\|\} && \text{(e)} \\ &= d_n \{O(\frac{\log n}{\sqrt{n}\, \log\log n}) + O(\frac{1}{\sqrt{n}})\} && \text{a.s.} \\ &= o(1) && \text{a.s.} \end{aligned}$$

by known properties of the oscillation modulus of $\mathbb{U}_n$ and elementary considerations; see Shorack and Wellner (1986), remark 14.2.2, page 545, with $c_n = (\log n)^{-1}$. This shows that the limsup in (2.8) is no larger than $1/\sqrt{2}$.

To prove the reverse inequality, note that

$$\begin{aligned} d_n \|\mathbb{X}_n^* + \mathbb{X}_n^{**}\| &\ge d_n \max_{1 \le i \le n-1} |\mathbb{U}_n(\frac{i}{n}) + \mathbb{V}_{n-1}(\frac{i}{n-1})| - d_n O(\frac{1}{\sqrt{n}}) \\ &\ge d_n \max_{1 \le i \le n-1} |\mathbb{U}_{n-1}(\frac{i}{n-1}) + \mathbb{V}_{n-1}(\frac{i}{n-1})| && \text{(f)} \\ &\quad - d_n O(\frac{\log n}{\sqrt{n}\, \log\log n}) && \text{by arguing as in (e) .} \end{aligned}$$

But by known properties of the oscillation moduli of $\mathbb{U}_n$ and $\mathbb{V}_n$, $d_n \omega_{\mathbb{U}_n}(1/n) = o(1)$ and $d_n \omega_{\mathbb{V}_n}(1/n) = o(1)$ a.s., and hence

$$\begin{aligned} &\limsup_{n \to \infty} d_n \max_{1 \le i \le n-1} |\mathbb{U}_{n-1}(\frac{i}{n-1}) + \mathbb{V}_{n-1}(\frac{i}{n-1})| && \text{(g)} \\ &= \limsup_{n \to \infty} d_n \|\mathbb{U}_{n-1} + \mathbb{V}_{n-1}\| = \frac{1}{\sqrt{2}} && \text{a.s.} \end{aligned}$$

as in (2.7). Combining (f) and (g) shows that the limsup in (2.8) is $\ge \frac{1}{\sqrt{2}}$. □

These results lead to a number of open problems:

Open question 2.1. If we define the Bayesian bootstrap empirical measure by $\mathbb{P}_n^{**} \equiv \sum_{i=1}^n D_{ni} \delta_{X_i(\omega)}$ and $\mathbb{P}_n^* \equiv \sum_{i=1}^n \mathbb{G}_n((i-1)/n, i/n] \delta_{X_i(\omega)}$ based on one sequence $\xi_1, \xi_2, \cdots$ of iid $Uniform(0,1)$ rv's, then what is the joint limiting behavior of $(\mathbb{X}_n^*, \mathbb{X}_n^{**})$ in $l^\infty(\mathcal{F}) \times l^\infty(\mathcal{F})$?

Open question 2.2. For what sequence of exchangeable random weights $\{W_{ni}\}$ satisfying $\sum_{i=1}^n W_{ni} = 1$ (but not necessarily of the form (2.1)) does it hold that $\mathbb{X}_n^W \Rightarrow \sigma G_P$ in $l^\infty(\mathcal{F})$ for $P^\infty -$ a.e. ω for some $\sigma > 0$?

For the classical real - valued case $A = [0,1]$ and $\mathcal{F} = \{1_{[0,t]} : t \in [0,1]\}$, Mason and Newton (1990) have solved open question 2.2 for exchangeable weights $\{W_{ni}\}$ satisfying $W_{ni} \ge 0$, $\sum W_{ni} = 1$, $n \sum_{i=1}^n (W_{ni} - 1/n)^2 \to_p c$, and

$n \max_i W_{ni}^2 \to_p 0$. This has been generalized to weighted empirical and quantile processes in one dimension by Einmahl and Mason (1990). However, the problem remains open for general empirical processes.

Now we formulate a result related to open question 2.1 for two different general weighted bootstraps. Again, the key point is that the joint behavior of two different bootstrap methods is highly dependent on the way in which we carry out the joint construction: of course this can be chosen by the statistician! The two weighted bootstraps which we will consider here will be constructed in a very particular way from one sequence of uniform(0,1) random variables $\xi_1, \xi_2, \cdots$ as follows: Let F_1, F_2 be two df's on R^+ both with mean 1 and with both $F_1, F_2 \in L_{2,1}$. Define

$$(T_i^{(1)}, T_i^{(2)}) \equiv (F_1^{-1}(\xi_i), F_2^{-1}(\xi_i)), \quad i = 1, 2, \cdots . \tag{2.9}$$

Then the $(T_i^{(1)}, T_i^{(2)})$ pairs are iid with marginal df's F_1 and F_2 and the joint distribution of each pair is the well-known Hoeffding-Fréchet joint df H^* with maximal correlation ρ^* among all joint distributions with marginal df's F_1, F_2; see for example Whitt (1976). Now we construct weights $(W_{ni}^{(1)}, W_{ni}^{(2)})$, $i = 1, \cdots, n$ from the first n $(T_i^{(1)}, T_i^{(2)})$ pairs as in (2.1):

$$W_{ni}^{(j)} \equiv T_i^{(j)} / \sum_{i'=1}^{n} T_{i'}^{(j)}, \quad j = 1, 2, \quad i = 1, \cdots, n . \tag{2.10}$$

Finally, define the two general weighted bootstrap empirical measures by

$$\boldsymbol{P}_{nj}^{**} \equiv \sum_{i=1}^{n} W_{ni}^{(j)} \delta_{X_i(\omega)}, \quad j = 1, 2 . \tag{2.11}$$

The following theorem gives the joint distribution of the two general weighted bootstrap empirical processes

$$\boldsymbol{X}_{nj}^{**} \equiv \sqrt{n}(\boldsymbol{P}_{nj}^{**} - \boldsymbol{P}_n^{\omega}), \quad j = 1, 2 . \tag{2.12}$$

Theorem 2.5. (Joint CLT for two *particular* weighted bootstrap empirical measures). Suppose that:

(i) $\mathbf{F} \in M(P)$ and $\|P\|_{\mathbf{F}} < \infty$.

(ii) $(T_i^{(1)}, T_i^{(2)})$, $i = 1, 2, \cdots$ are iid pairs of positive random variables constructed as in (2.9) with $ET_i^{(j)} = 1$ and $Var(T_i^{(j)}) = \sigma_j^2 < \infty$, $j = 1, 2$, $i = 1, 2, \cdots$, and $T_1^{(j)} \in L_{2,1}$ for $j = 1, 2$.

Then

$$(\boldsymbol{X}_{n1}^{**}, \boldsymbol{X}_{n2}^{**}) \Rightarrow (\boldsymbol{X}_1^{**}, \boldsymbol{X}_2^{**}) \quad \text{in} \quad l^\infty(\mathbf{F}) \times l^\infty(\mathbf{F}) \tag{2.13}$$

for P^∞ – a.e. ω where $(\boldsymbol{X}_1^{**}, \boldsymbol{X}_2^{**})$ is jointly Gaussian, $\boldsymbol{X}_j^{**} \sim \sigma_j G_P$, $j = 1, 2$ and for $f, g \in \mathbf{F}$,

$$Cov(\boldsymbol{X}_1^{**}(f), \boldsymbol{X}_2^{**}(g)) = \rho^* \sigma_1 \sigma_2 (P(fg) - P(f)P(g)) . \tag{2.14}$$

Corollary 2.2. Suppose that the hypotheses of theorem 2.5 hold. For any $\lambda \in [0,1]$ set $\boldsymbol{P}_{n\lambda}^{**} \equiv \lambda \boldsymbol{P}_{n1}^{**} + (1-\lambda)\boldsymbol{P}_{n2}^{**}$, and define $\boldsymbol{X}_{n\lambda}^{**} \equiv \sqrt{n}(\boldsymbol{P}_{n\lambda}^{**} - \boldsymbol{P}_n^{\omega})$. Then

$$\boldsymbol{X}_{n\lambda}^{**} \Rightarrow \sigma_\lambda G_P \quad \text{in} \quad l^\infty(\mathbf{F}) \quad \text{for} \quad P^\infty\text{– a.e. } \omega \tag{2.15}$$

where

$$\sigma_\lambda^2 \equiv \lambda^2 \sigma_1^2 + 2\lambda(1-\lambda)\rho^* \sigma_1 \sigma_2 + (1-\lambda)^2 \sigma_2^2 \geq 0 . \tag{2.16}$$

Since the number σ_λ is known to us as constructors of the two weighted bootstraps, whenever it is positive we can combine the two weighted bootstraps to obtain yet another valid bootstrap based on $\boldsymbol{P}_{n\lambda}^{**}$ with a straightforward interpretation.

Proof of theorem 2.5. By theorem 2.3, the processes $\boldsymbol{X}_{nj}^{**}$ converge weakly to $\sigma_j G_P$, $j = 1, 2$, and hence they are marginally tight. Since marginal tightness implies joint tightness (see, for example, van der Vaart and Wellner (1990) for a version of this result in the Hoffmann - Jørgensen weak convergence theory), the joint sequence $\{(\boldsymbol{X}_{n1}^{**}, \boldsymbol{X}_{n2}^{**}) : n = 1, 2, \cdots\}$ is tight. To complete the proof it suffices to verify that the finite - dimensional joint laws all converge to those of the process $(\boldsymbol{X}_1^{**}, \boldsymbol{X}_2^{**})$ as claimed. But this follows easily from the identity (2.3) and the Cramer - Wold device. □

Proof of corollary 2.2. Note that by theorem 2.5

$$\boldsymbol{X}_{n\lambda}^{**} = \lambda \boldsymbol{X}_{n1}^{**} + (1-\lambda)\boldsymbol{X}_{n2}^{**} \Rightarrow \lambda \boldsymbol{X}_1^{**} + (1-\lambda)\boldsymbol{X}_2^{**} \sim \sigma_\lambda G_P . \quad \square$$

3. Limit Theorems for Model - Based Bootstrap Empirical Processes

Let $(A, \mathbf{A})$ be a measurable space, and let $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ be a model, parametric, semiparametric, or nonparametric. We do not insist that Θ be finite - dimensional. For example, in a parametric extreme case, $\mathbf{P}$ could be the family of all normal (Gaussian) distributions on $(A, \mathbb{A}) = (R^d, \mathbf{B}^d)$. Or, to give a nonparametric example with only a smoothness restriction, $\mathbf{P}$ could be the family of all distributions on $(A, \mathbb{A}) \equiv (R^d, \mathbf{B}^d)$ with a density with respect to Lebesgue measure which is uniformly continuous.

Let $X_1, X_2, \cdots, X_n, \cdots$ be iid with distribution $P_\theta \in \mathbf{P}$. We assume that there exists an estimator $\hat\theta_n = \hat\theta_n(X_1, \cdots, X_n)$ of θ. Then Efron's parametric (or model - based) bootstrap proceeds by sampling from $\hat P_n^\omega \equiv P_{\hat\theta_n(\omega)}$: suppose that $X_{n1}^\#, \cdots, X_{nn}^\#$ are independent and identically distributed with distribution $\hat P_n^\omega$ on $(A, \mathbb{A})$, and let

$$\boldsymbol{P}_n^\# \equiv n^{-1}\sum_{i=1}^n \delta_{X_{ni}^\#} \equiv \text{the parametric bootstrap empirical measure}, \tag{3.1}$$

and

$$\boldsymbol{X}_{n,n}^\# \equiv \sqrt{n}\,(\boldsymbol{P}_n^\# - \hat P_n), \tag{3.2}$$

the *parametric bootstrap empirical process.* The key difference between this parametric bootstrap procedure and the nonparametric bootstrap discussed earlier in this section is that we are now sampling from the model - based estimator $\hat P_n \equiv P_{\hat\theta_n}$ of P rather than from the nonparametric estimator $\boldsymbol{P}_n$.

It often holds that

$$\sqrt{n}\,(\hat\theta_n - \theta) \Rightarrow Y \qquad \text{as } n \to \infty, \tag{3.3}$$

and, if $\theta \to P_\theta$ is differentiable in an appropriate sense,

$$\sqrt{n}\,(P_{\hat\theta_n} - P_\theta) \Rightarrow \dot P_\theta Y \qquad \text{as } n \to \infty \tag{3.4}$$

where $\dot P_\theta$ is a derivative map. The parametric bootstrap would proceed by forming $\hat\theta_n^\# \equiv \hat\theta_n(X_{n1}^\#, \cdots, X_{nn}^\#)$. We then want to show that: for almost all sample sequences $X_1, X_2, \cdots$

$$\sqrt{n}\,(\hat\theta_n^\# - \hat\theta_n) \Rightarrow Y^\# \sim Y \tag{3.5}$$

and

$$\sqrt{n}\,(P_{\hat\theta_n^\#} - P_{\hat\theta_n}) \Rightarrow \dot P_\theta Y^\# \sim \dot P_\theta Y \qquad \text{as } n \to \infty . \tag{3.6}$$

The following result is a useful first step toward proving (3.5) or (3.6), especially if $\hat{\theta}_n = \theta(\mathbb{P}_n)$ so that $\hat{\theta}_n^{\#} = \theta(\mathbb{P}_n^{\#})$. This type of theorem for "model - based" or "parametric" bootstrap empirical processes was also suggested by Giné and Zinn (1991).

Theorem 3.1. (Convergence of the "parametric bootstrap" empirical process). Suppose that $\mathbf{F}$ is $\mathbf{P}$ – measurable with envelope function F and that:

(i) $\mathbf{F} \in CLT_u(\mathbf{P})$.

(ii) $\|P_{\hat{\theta}_n} - P_\theta\|_{\mathbf{G}}^* \equiv \|\hat{P}_n - P_\theta\|_{\mathbf{G}}^* \rightarrow_{a.s.} 0$.

(iii) F is $\mathbf{P}$ – uniformly square integrable.

Then, for P^∞ almost all sample sequences $X_1, X_2, \cdots$,

$$\mathbf{X}_{n,n}^{\#} \equiv \sqrt{n}\,(\mathbf{P}_n^{\#} - P_{\hat{\theta}_n}) \Rightarrow \mathbf{X}_0^{\#} \sim G_{P_0} \quad \text{in} \quad l^\infty(\mathbf{F}) \tag{3.7}$$

as $n \to \infty$.

Proof. First note that (i) and (iii) imply that $\mathbf{F} \in AEC_u(\mathbf{P}, \rho)$ and $(\mathbf{F}, \rho_P)$ is totally bounded uniformly in $P \in \mathbf{P}$ by Sheehy and Wellner (1991), theorem 2.2. Hence, in particular, $\mathbf{F} \in AEC_u(\{\hat{P}_n\}, \rho)$ and $(\mathbf{F}, \rho_{P_\theta})$ is totally bounded. Furthermore, (iii) implies that for P^∞ a.e. ω the envelope F is $\{\hat{P}_n^\omega\}$ – uniformly square integrable. Thus, for P^∞ – a.e. ω, the hypotheses of Sheehy and Wellner (1991) theorem 3.1 are satisfied by $\mathbf{F}$ for the sequence $\{\hat{P}_n^\omega\} \equiv \{P_{\hat{\theta}_n(\omega)}\}$. Then the conclusion follows from theorem 3.1 with $P_0 \equiv P_\theta$. □

To give an example where this result is immediately useful, consider the non-parametric example mentioned briefly above:

Example 3.1. (Bootstrapping from a "smoothed empirical measure"; or, the "smoothed bootstrap"). Suppose that

$$\mathbf{P} = \{P \text{ on } (R^d, \mathbf{B}^d) : \; p \equiv \frac{dP}{d\lambda} \text{ exists and is uniformly continuous}\}.$$

Suppose $\mathbf{C}$ is a measurable Vapnik - Chervonenkis class of subsets of R^d. Then $\mathbf{F} = \{1_C : C \in \mathbf{C}\} \in CLT_u \equiv CLT_u(\mathbf{M}) \subset CLT_u(\mathbf{P})$, so (i) holds. Suppose that $\hat{\mathbb{P}}_n$ is defined for $C \in \mathbb{A}$ by

$$\hat{\mathbb{P}}_n(C) = \int 1_C(x)\hat{p}_n(x)dx$$

where

$$\hat{p}_n(x) = \frac{1}{b_n^d}\int k\left(\frac{y-x}{b_n}\right)d\mathbb{P}_n(y)$$

where $k : R^d \to \mathbb{R}$ is a uniformly continuous density function. It follows that $\hat{\mathbb{P}}_n \in \mathbf{P}$ and, if $b_n \to 0$ and $nb_n^d \to \infty$, then

$$\int |\hat{p}_n(x) - p(x)|\,dx \rightarrow_{a.s.} 0;$$

see Devroye (1983), theorem 1. When $\mathbf{F}$ is all indicators of a subclass of Borel sets $\mathbf{C}$ the supremum in (ii) is bounded by the total variation distance between $\hat{\mathbb{P}}_n$ and P, which, in turn, is well-known to equal half the L_1 – distance between the respective densities (see e.g. the statement of Scheffé's theorem in Billingsley (1968), page 224)). Hence

$$\|\hat{\mathbb{P}}_n - P\|_{\mathbf{G}}^* \le \|\hat{\mathbb{P}}_n - P\|_{\mathbf{B}^d} = \frac{1}{2}\int |\hat{p}_n(x) - p(x)|\,dx \rightarrow_{a.s.} 0,$$

so (ii) holds. Since (iii) holds trivially (with $F \equiv 1$), theorem 4.5 shows that "the bootstrap from $\hat{\mathbb{P}}_n$ works:" i.e. for P^∞ almost all sample sequences

$X_1, X_2, \cdots,\quad \mathbf{X}^{\#}_{n,n} \Rightarrow \mathbf{X}^{\#}_0 \sim G_P$ in $l^{\infty}(\mathbf{F})$.

For more general classes $\mathbf{F}$, the results of Yukich (1989) could be used to verify hypothesis (ii) of theorem 4.5.

Silverman and Young (1987) have studied several smoothed bootstrap methods, and give criteria for determining when $\alpha(P_{\hat\theta_n})$ will give a better estimator of $\alpha(P)$ than $\alpha(\boldsymbol{P}_n)$ for functionals $\alpha : \mathbf{P} \to R$; see also Hall, DiCiccio, and Romano (1989) for further work in this direction.

Example 3.2. (Bootstrapping from a monotone density). Suppose that $\mathbf{P} = \{P$ on $R^+ : P$ has nonincreasing Lebesgue density $f\}$. Corresponding to $P \in \mathbf{P}$, the distribution function is $F(x) = P(X \le x)$ for $x \in R^+$. Since f is monotone decreasing, F is concave. Grenander (1956) showed that if $X_1, \cdots, X_n$ are iid $P \in \mathbf{P}$, then the least concave majorant $\hat{\boldsymbol{F}}_n$ of the empirical df $\boldsymbol{F}_n$ is the maximum likelihood estimator of F. It is known that $\hat{\boldsymbol{F}}_n$ is consistent:

$$\|\hat{\boldsymbol{F}}_n - F\| \equiv \sup_x |\hat{\boldsymbol{F}}_n(x) - F(x)| \to_{a.s.} 0 ;$$

see e.g. Barlow et al. (1972) and Marshall (1970), who shows that $\|\hat{\boldsymbol{F}}_n - F\| \le \|\boldsymbol{F}_n - F\|$. Kiefer and Wolfowitz (1976) show that if F is twice differentiable and $\gamma(F) \equiv \sup_{0<t<T} |f'(t)|/\inf_{0<t<T} f^2(t) < \infty$ and $\beta(F) \equiv \inf_{0<t<T} |-f'(t)|/f^2(t) > 0$ where $T \equiv \inf\{t : F(t) = 1\}$, then

$$\sqrt{n}\,(\hat{\boldsymbol{F}}_n - F) \Rightarrow \boldsymbol{U}(F) \quad \text{in } l^{\infty}(\mathbf{F}) \supset D[0,T] ; \tag{3.8}$$

it is clear from Wang (1986) that (3.8) continues to hold if the conditions of Kiefer and Wolfowitz (1976) are relaxed to "uniform convexity" in the sense of assumption A of Wang (1976), page 1116. Furthermore, Prakasa Rao (1969) and Groeneboom (1985) show that if $\hat f_n(t)$ is the derivative of $\hat{\boldsymbol{F}}_n$ at t (which exists at all t except for a subset of the observations), then, for t such that $-f'(t) > 0$,

$$n^{1/3}(\hat f_n(t) - f(t)) \to_d (-\frac{1}{2} f(t) f'(t))^{1/3} 2Z \tag{3.9}$$

where Z is distributed as the location of the maximum of the process $\{W(t) - t^2 : t \in R\}$; here W is a two - sided standard Brownian motion starting from 0.

The question is which (if any) of these results can be bootstrapped successfully?

We first apply theorem 3.1 to bootstrap sampling from $\hat{\boldsymbol{F}}_n$. Let $\mathbf{F} = \{1_{[0,t]} : t \in [0,T]\}$. The collection of sets used to define $\mathbf{F}$ is trivially a Vapnik - Červonenkis class of sets, and hence is easily shown to be a $\mathbf{M}$ – uniform Donsker collection; see e.g. Giné and Zinn (1991) or Sheehy and Wellner (1990), theorem 2.4. Thus $\mathbf{F} \in CLT_u(\mathbf{M}) \subset CLT_u(\mathbf{P})$. It follows from theorem 3.1 that

$$\mathbf{X}^{\#}_{n,n} \equiv \sqrt{n}\,(\boldsymbol{F}^{\#}_n - \hat{\boldsymbol{F}}^{\omega}_n) \Rightarrow \boldsymbol{U}(F) \quad \text{a.s. } P^{\infty}. \tag{3.10}$$

Of course this can also be proved directly by arguing as in Shorack (1982); see Shorack and Wellner (1986), section 23.1: If $\xi^{\#}_1, \xi^{\#}_2, \cdots$ are iid uniform(0,1) with empirical df $\boldsymbol{G}^{\#}_n$, then we can represent the bootstrap sample by $X^{\#}_i = \hat{\boldsymbol{F}}_n^{-1}(\xi^{\#}_i)$, $i = 1, \cdots, n$. Hence

$$\sqrt{n}\,(\boldsymbol{F}^{\#}_n - \hat{\boldsymbol{F}}^{\omega}_n) \overset{d}{=} \boldsymbol{U}^{\#}_n(\hat{\boldsymbol{F}}^{\omega}_n) \Rightarrow \boldsymbol{U}^{\#}(F) \quad \text{a.s. } P^{\infty}.$$

But another interesting question is: if $\hat{\boldsymbol{F}}^{\#}_n$ is the least concave majorant of $\boldsymbol{F}^{\#}_n$, does

$$\sqrt{n}\,(\hat{\boldsymbol{F}}^{\#}_n - \hat{\boldsymbol{F}}^{\omega}_n) \Rightarrow \boldsymbol{U}(F) \quad \text{a.s. } P^{\infty}\,? \tag{3.11}$$

(Note that the conditions of Kiefer and Wolfowitz (1976) fail.) Or, to go further still, if $\hat{f}_n^{\#}(t)$ is the derivative of $\hat{F}_n^{\#}$ at t where $f'(t) < 0$, does it hold that

$$n^{1/3}(\hat{f}_n^{\#}(t) - \hat{f}_n^{\omega}(t)) \rightarrow_d 2|2^{-1}f(t)f'(t)|^{1/3}Z \quad P^{\infty} - \text{a.s. ?} \tag{3.12}$$

We list these as open questions:

Open question 3.1. Does (3.11) hold?

Open question 3.2. Does (3.12) hold?

Theorem 3.1 clearly has applications to a wide variety of model - based bootstrap methods: for example, semiparametric models such as the Cox model, but we forego these applications here.

4. A "joint in sample size" bootstrap limit theorem

One interesting feature of Efron's bootstrap is that it allows the possibility of estimation of the sampling distribution of some statistic T_k based on a sample size k different than the sample size n actually observed: simply draw a bootstrap sample of size k ! This was observed by Bickel and Freedman (1981). In fact, we can easily estimate the sampling distribution of T_k for several different sample sizes $0 < k_1 < \cdots < k_r$ simultaneously by simply drawing bootstrap samples of these same sizes. If we do this in the natural way by successively sampling

$$X_1^*, X_2^*, \cdots, X_{k_r}^*, \cdots \quad \text{iid} \quad \mathbb{P}_n^{\omega}, \tag{4.1}$$

then it is possible to establish the *joint* limiting behavior of our estimators

$$(\hat{P}(T_{k_1} \in A), \cdots, \hat{P}(T_{k_r} \in A))$$

for different sample sizes.

To set the stage for our "joint in bootstrap sample size" limit theorem, for each $n = 1, 2, \cdots$ let

$$\underline{1}_j \equiv \underline{1}_j^{(n)} \equiv (1_{1j}, \cdots, 1_{nj}), \quad j = 1, 2, \cdots \tag{4.2}$$

be iid Multinomial$_n(1; (\frac{1}{n}, \cdots, \frac{1}{n}))$. Thus for $k = 1, 2, \cdots$

$$\underline{M}_k^* \equiv \underline{M}_k^{*(n)} \equiv \sum_{j=1}^{k} \underline{1}_j \sim Multinomial_n(k; (\frac{1}{n}, \cdots, \frac{1}{n})), \tag{4.3}$$

and

$$\mathbb{P}_k^* \equiv \frac{1}{k}\sum_{i=1}^{n} M_{ki}^* \delta_{X_i(\omega)} = \frac{1}{k}\sum_{j=1}^{k}(\sum_{i=1}^{n} 1_{ji}\, \delta_{X_i(\omega)}) \tag{4.4}$$

represents the empirical measure of a sample of size k from $\mathbb{P}_n^{\omega}$.

Let the bootstrap empirical process $\mathbb{X}_{n,k}^*$ for sample size k be defined by

$$\begin{aligned} \mathbb{X}_{n,k}^* &= \sqrt{k}\,(\mathbb{P}_k^* - \mathbb{P}_n^{\omega}) \\ &= \frac{1}{\sqrt{k}}\sum_{i=1}^{n}(M_{ki}^* - \frac{k}{n})\delta_{X_i(\omega)} \\ &= \frac{1}{\sqrt{k}}\sum_{i=1}^{n}(M_{ki}^* - \frac{k}{n})(\delta_{X_i(\omega)} - P), \end{aligned} \tag{4.5}$$

and define the *sequential bootstrap empirical process* $\mathbb{K}_n^*$ on $[0,1] \times \mathcal{F}$ by

$$\mathbb{K}_n^*(t\,f) \equiv \sqrt{\frac{[nt]}{n}}\,\mathbb{X}_{n,[nt]}^*(f)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (M^*_{[nt],i} - \frac{[nt]}{n})(f(X_i(\omega)) - Pf),$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \sum_{i=1}^{n} (1_{ji} - \frac{1}{n})(f(X_i(\omega)) - Pf). \tag{4.6}$$

Note that for $0 \le s \le t \le 1$ and $f, g \in \mathbf{F}$,

$$Cov(K_n^*(s,f), K_n^*(t,g) \mid \underline{X}) = \frac{[ns]}{n} \{ \mathbf{P}_n^{\omega}(fg) - \mathbf{P}_n^{\omega}(f)\mathbf{P}_n^{\omega}(g) \}$$

$$\rightarrow_{a.s.} (s \wedge t)\{P(fg) - Pf\,Pg\} \tag{4.7}$$

$$= Cov(K^*(s,f), K^*(t,g))$$

where K^* is a P – Kiefer process K_P on $[0,1] \times \mathbf{F} \equiv \tilde{\mathbf{F}}$. This leads naturally to the following theorem, phrased in the spirit of the Giné and Zinn theorem 3.2.

Theorem 4.1. Suppose that $\mathbf{F} \in M(P)$. Then the following are equivalent:

A. $\mathbf{F} \in CLT(P)$ and $P(F^2) < \infty$.

B. $\mathbf{X}_n^* \Rightarrow \mathbf{X}^* \sim G_P$ a.s. P^∞ in $l^\infty(\mathbf{F})$.

C. $K_n^* \Rightarrow K^* \sim K_P$ a.s. P^∞ in $l^\infty(\tilde{\mathbf{F}})$.

As an immediate corollary, we obtain the joint behavior of the bootstrap empirical processes $\mathbf{X}^*_{n,[ns]}$ and $\mathbf{X}^*_{n,[nt]}$ for sample sizes $k_1 = [ns]$ and $k_2 = [nt]$.

Corollary 4.1. Suppose that $\mathbf{F} \in M(P)$, $\mathbf{F} \in CLT(P)$, $P(F^2) < \infty$, and let $0 < s < t < \infty$. Then

$$(\mathbf{X}^*_{n,[ns]}, \mathbf{X}^*_{n,[nt]}) = (\sqrt{\frac{n}{[ns]}} K_n^*(s,\cdot), \sqrt{\frac{n}{[nt]}} K_n^*(t,\cdot))$$

$$\Rightarrow (\frac{1}{\sqrt{s}} K^*(s,\cdot), \frac{1}{\sqrt{t}} K^*(t,\cdot)) \equiv (G_P^s, G_P^t)$$

where G_P^s, G_P^t are two P – Brownian bridge processes on $\mathbf{F}$ with

$$Cov(G_P^s, G_P^t) = \sqrt{\frac{s}{t}} \{P(fg) - Pf\,Pg\}.$$

As a second corollary we obtain the limiting behavior of the sequential bootstrap process corresponding to any one fixed function $f \in L_2(P)$ by taking $\mathbf{F} = \{f\}$:

Corollary 4.2. Suppose that $f \in L_2(P)$. Then

$$K_n^*(\cdot,f) \Rightarrow K(\cdot,f) \overset{d}{=} \sigma_f B \quad a.s. \quad P^\infty \quad \text{in } l^\infty([0,1]) \supset D[0,1].$$

where $\sigma_f^2 \equiv Var_P(f(X))$ and B is standard Brownian motion.

Sketch of the proof of theorem 4.1. A is equivalent to B by the Giné and Zinn theorem 2.2, and C implies B trivially (by noting that $K_n^*(1,\cdot) = \mathbf{X}_n^*$). Hence it remains only to show that B implies C.

Suppose that B holds. It follows from (4.6) that, for each fixed n and fixed ω, K_n^* is just the partial sum process corresponding to the iid $l^\infty(\mathbf{F})$ – valued random elements

$$T_{nj} \equiv \sum_{j=1}^{n} (1_{ji} - \frac{1}{n})(\delta_{X_i(\omega)} - P), \quad j = 1, 2, \cdots. \tag{a}$$

Thus we are in the domain of theorem 1.1, page 511, Dudley and Philipp (1983). The key conditions (1.4) - (1.7) of their theorem 1.1 hold with $X_j \equiv T_{nj}$, $j \ge 1$, and with $Y_j \equiv Y_{nj}$, $j \ge 1$, iid Gaussian with covariance

$$Cov[Y_{n1}(f), Y_{n1}(g)] = Cov[T_{n1}(f), T_{n1}(g)] \qquad \text{(b)}$$
$$= \mathbf{P}_n^{\omega}(fg) - \mathbf{P}_n^{\omega}(f)\mathbf{P}_n^{\omega}(g)$$

for $f, g \in \mathbf{F}$. The condition (1.5) follows from a.s. asymptotic equicontinuity of the bootstrap empirical process $\mathbf{X}_n^*$ which is guaranteed by B; see Giné and Zinn (1990), equation (2.16), page 857. Thus we conclude that

$$n^{-1/2} \max_{k \le n} \| \sum_{j \le k} (T_{nj} - Y_{nj}) \|_{\mathbf{F}} \to 0 \qquad \text{(c)}$$

in probability and in L_p for $p < 2$ a.s. P^{∞}. It follows easily that, by defining the Gaussian processes $\mathbf{L}_n^*$ by

$$\mathbf{L}_n^*(t, f) \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} Y_{nj}(f), \quad t \in [0,1], \ f \in \mathbf{F},$$

we have

$$d_{BL^*}(K_n^*, \mathbf{L}_n^*) \equiv \sup_{h \in BL_1(\tilde{\mathbf{F}})} |E\, h(K_n^*) - E\, h(\mathbf{L}_n^*)| \to 0 \qquad \text{(d)}$$

a.s. P^{∞}. Furthermore the Gaussian processes $\mathbf{L}_n^*$ have covariances

$$Cov[\mathbf{L}_n^*(s, f), \mathbf{L}_n^*(t, g)] = \frac{[ns] \wedge [nt]}{n} \{\mathbf{P}_n^{\omega}(fg) - \mathbf{P}_n^{\omega}(f)\mathbf{P}_n^{\omega}(g)\} \qquad \text{(e)}$$

for $s, t \in [0,1]$ and $f, g \in \mathbf{F}$, where the right side converges P^{∞} a.s. to $(s \wedge t)\{P(fg) - PfPg\} = Cov[K^*(s,f), K^*(t,g)]$. Hence it follows from Fernique (1985), corollary 2.2 (much as in the proof of E implies F of theorem 1.1 of Sheehy and Wellner (1990)) that

$$d_{BL^*}(\mathbf{L}_n^*, K^*) \equiv \sup_{h \in BL_1(\tilde{\mathbf{F}})} |E\, h(\mathbf{L}_n^*) - E\, h(K^*)| \to 0 \qquad \text{(f)}$$

a.s. P^{∞}. Combining (d) and (f) completes the proof. □

Comparing theorem 4.1 with the results of section 2 leads to the following problem:

Open question 4.1. What is the appropriate analogue of theorem 4.1 for the general weighted bootstrap (as in theorem 2.3)?

Acknowledgments: For the results on parametric bootstrap methods in section 3, I owe thanks to both Anne Sheehy and E. Giné. The development in section 3 was suggested E. Giné, and was originally a part of Sheehy and Wellner (1990).

References

Arcones, M. A. and Giné, E. (1990). On the bootstrap of M - estimators and other statistical functionals *Preprint*, Department of Statistics, University of Connecticut, Storrs.

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

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Csörgő, S. and Mason, D. (1990). Bootstrapping empirical functions. *Ann. Statist.* **17**, 1447 - 1471.

Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.* **11**, 896 - 904.

Dudley, R. M. (1984). A course on empirical processes; Ecole d'Eté de Probabilités de St. Flour. *Lecture Notes in Math.* **1097**, 2 - 142, Springer Verlag, New York.

Dudley, R. M. (1985). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions. *Lecture Notes in Math.* **1153**, 141 - 178, Springer - Verlag, New York.

Dudley, R. M. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Geb.* **62**, 509-552.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1 - 16.

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. *CBMS-NSF Regional Conference Series in Applied Mathematics* **38**, Society for Industrial and Applied Mathematics, Philadelphia.

Einmahl, U. and Mason, D. M. (1990). Approximations to permutation and exchangeable processes. *Technical Report*, University of Delaware.

Fernique, X. (1985). Sur la convergence etroite des mesures gaussiennes. *Z. Wahrsch. v. Geb.* **68**, 331 - 336.

Gaenssler, P. (1986). Bootstrapping empirical measures indexed by Vapnik - Chervonenkis classes of sets. *Proceedings IV Vilnius Conference (1985) Prob. Theory and Math. Statist.*, **1**, 467 - 481.

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. (1990). Bootstrapping general empirical measures. *Ann. Probability* **18**, 851 - 869.

Giné, E. and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probability*, **19**, 758 - 782.

Grenander, U. (1956). On the theory of mortality measurement, part II *Skand. Akt.* **39**, 125 - 153.

Groeneboom, P. (1985). Estimating a monotone density *Proceedings of the Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II.* Wadsworth, Belmont, L. M. Le Cam and R. A. Olshen, editors.

Hall, P., DiCiccio, T. J., and Romano, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17**, 692 - 704.

Hoffmann-Jørgensen, J. (1984). *Stochastic Processes on Polish Spaces.* Unpublished manuscript.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **34**, 73 - 85.

Klaassen, C. A. J. and Wellner, J. A. (1992). Kac empirical processes and the bootstrap. To appear, *Proceedings of the Eight International Conference on Probability in Banach Spaces, Bowdoin College, July 1991*, Birkhauser.

Ledoux, M. and Talagrand, M. (1986). Conditions d'integrabilite pour les multiplicateurs dans le TLC Banachique. *Ann. Probability* **14**, 916 - 921.

Ledoux, M. and Talagrand, M. (1988). Un critere sur les petites boules dan le theoreme limite central. *Probab. Th. Rel. Fields* **77**, 29 - 47.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer-Verlag, New York.

Lo, A. (1987). A large sample study of the Bayesian bootstrap *Ann. Statist.* **15,** 360 - 375.

Marshall, A. W. (1970). Discussion of Barlow and van Zwet's papers in M. L. Puri (Ed.) *Nonparametric Techniques in Statistical Inference*, 175 - 176, Cambridge University Press, Cambridge.

Mason, D. M. and Newton, M. A. (1990). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.*, to appear.

Praestgaard, J. (1990). Bootstrap with general weights and multiplier central limit theorems *Technical Report* **195**, Department of Statistics, University of Washington, Submitted to *Ann. Prob.*

Prakasa Rao, B.L.S. (1969). Estimation of a unimodal density. *Sankhya* **Ser. A 31,** 23 - 36.

Romano, J. P. (1988). A bootstrap revival of some nonparametric distance tests. *J. Amer. Statist. Assoc.* **83**, 698 - 708.

Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17,** 141 - 159.

Rubin, D. (1981). The Bayesian bootstrap. *Ann. Statist.* **9**, 130 - 134.

Sheehy, A. and Wellner, J. A. (1991). Uniform Donsker classes of functions *Technical Report* **189**, Department of Statistics, University of Washington. Submitted to *Ann. Prob.*.

Shorack, G. R. (1982). Bootstrapping robust regression *Commun. Statist.* **A11**, 961 - 972.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Silverman, B. and Young, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74**, 469 - 479.

Swanepoel, J.W.H. (1990). A review of bootstrap methods. *S. Afri. Statist. J.* **24**, 1 - 34.

Trumbo, B. E., and Burdick, R. K. (1990). *Cumulative Index to IMS Scientific Journals 1960 - 1989.* Institute of Mathematical Statistics, Hayward.

Vaart, A. W. van der and Wellner, J. A. (1990). Prohorov and continuous mappings theorems in the Hoffmann-Jørgensen weak convergence theory with applications to convolution and asymptotic minimax theorems *Technical Report* **157**, Department of Statistics, University of Washington, Seattle. Submitted to *Theory of Probability and Related Fields.*

Wang, J.-L. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. *Ann. Statist.* **14**, 1113 - 1131.

Weng, C. (1989). On a second - order property of the Bayesian bootstrap mean. *Ann. Statist.* **17**, 705 - 710.

Whitt, W. (1976). Bivariate distributions with given marginals. *Ann. Statist.* **4**, 1280 - 1289.

Yukich, J. E. (1989). A note on limit theorems for perturbed empirical processes. *Stoch. Processes and Their Applications* **33**, 163 - 174.