



Exchangeably Weighted Bootstraps of the General Empirical Process

Jens Praestgaard; Jon A. Wellner

The Annals of Probability, Vol. 21, No. 4 (Oct., 1993), 2053-2086.

Stable URL:

<http://links.jstor.org/sici?sici=0091-1798%28199310%2921%3A4%3C2053%3AEWBOTG%3E2.0.CO%3B2-W>

The Annals of Probability is currently published by Institute of Mathematical Statistics.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

EXCHANGEABLY WEIGHTED BOOTSTRAPS OF THE GENERAL EMPIRICAL PROCESS¹

BY JENS PRÆSTGAARD AND JON A. WELLNER

University of Iowa and University of Washington

We consider an exchangeably weighted bootstrap of the general function-indexed empirical process. We find sufficient conditions on the bootstrap weights for the central limit theorem to hold for the bootstrapped empirical process, almost surely and in probability. The results resemble those of Giné and Zinn for Efron's bootstrap. As a corollary we obtain a result on the almost sure convergence in distribution of the Efron-bootstrapped empirical process with arbitrary sample size. A large number of bootstrap resampling schemes emerge as special cases of our results.

1. Introduction. The topic of this article is bootstrapping the general empirical process with exchangeable weights. The bootstrap technique was introduced by Efron (1979, 1982) as a method for estimating the sampling distribution of a statistic. It may be explained briefly as follows: Let X_1, \dots, X_n be iid observations with distribution P ; let $\theta \equiv \theta(P)$ be a parameter of interest, and let $\theta_n \equiv \theta_n(X_1, \dots, X_n)$ be an estimator of θ . The bootstrap principle is to estimate the unknown distribution of θ_n by $\hat{\theta}_n$ where $\hat{\theta}_n$ is distributed as $\theta_n(\hat{X}_1, \dots, \hat{X}_n)$ and $\hat{X}_1, \dots, \hat{X}_n$ are iid from the empirical probability measure

$$\mathbb{P}_n \equiv \frac{1}{n} \sum_{j=1}^n \delta_{X_j}.$$

Although in general not expressible in a closed form, the bootstrap distribution can easily be evaluated by Monte Carlo by drawing samples with replacement $\hat{X}_1, \dots, \hat{X}_n$ from (X_1, \dots, X_n) and computing $\theta_n(\hat{X}_1, \dots, \hat{X}_n)$.

Often the quantity $\sqrt{n}(\theta_n - \theta)$ to be bootstrapped can be expressed as a function of the empirical process which is defined as

$$\mathbb{X}_n \equiv \sqrt{n}(\mathbb{P}_n - P).$$

In the modern theory of empirical processes it is customary to identify P , \mathbb{P}_n and \mathbb{X}_n with the mappings given by $f \rightarrow \int f dP = Pf$, $f \rightarrow \int f d\mathbb{P}_n = n^{-1} \sum_{j=1}^n f(X_j) = \mathbb{P}_n f$ and $f \rightarrow \int f d\mathbb{X}_n = n^{-1/2} \sum_{j=1}^n (f(X_j) - Pf) = \mathbb{X}_n(f)$, respectively. Here, $f \in \mathcal{F}$, and $F \subset L_2(P)$ is a collection of functions mapping the sample space \mathbf{X} to \mathbb{R} . In this way, \mathbb{X}_n becomes a random element of $l^\infty(\mathcal{F})$, the space of bounded real functions on \mathcal{F} . The most straightforward example is to take $\mathbf{X} = [0, 1]$ and let \mathcal{F} be the collection of indicator functions of sets of

Received February 1991; revised August 1992.

¹Supported in part by NSF Grants DMS-87-23011 and DMS-91-08409.

AMS 1991 subject classifications. Primary 60F17, 62E20; secondary 60B12.

Key words and phrases. Bootstrap, exchangeability, empirical process.

the form $[0, c]$, $0 < c \leq 1$. In this case \mathbb{P}_n becomes the ordinary empirical distribution function. However, many other classes \mathcal{F} of interest in statistics exist. We shall not go into detail with these here, but refer to Pollard (1990) and the expository paper of Wellner (1992) in which applications to statistics are listed. These include, among others, tree-structured classification and regression, and projection pursuit regression.

For the simple case stated above, Donsker's theorem [see Billingsley (1968) or Shorack and Wellner (1986)] states that the empirical process converges in distribution to a Brownian bridge on $[0, 1]$. The same holds for the function-indexed empirical process as follows: A P -Brownian bridge process G_P is a 0-mean Gaussian process indexed by \mathcal{F} with covariance function

$$\text{Cov}(G_P(f), G_P(g)) = Pfg - PfPg, \quad f, g \in \mathcal{F}.$$

Let ρ_P be the pseudometric on $L_2(P)$ given by

$$\rho_P^2(f, g) \equiv \text{Var}(f(X) - g(X)) = P((f - g)^2) - P(f - g)^2.$$

If there exists a version G_P of a P -Brownian bridge, indexed by \mathcal{F} , which has bounded and ρ_P -uniformly continuous sample paths, we say that \mathcal{F} is P -pre-Gaussian. We say that \mathcal{F} is P -Donsker or, shorter, that $\mathcal{F} \in \text{CLT}(P)$, if \mathcal{F} is P -pre-Gaussian and

$$\sqrt{n}(\mathbb{P}_n - P) \Rightarrow G_P.$$

This convergence is convergence in distribution in $l^\infty(\mathcal{F})$ in the sense of Hoffmann-Jørgensen (1984); see Andersen (1985), Dudley (1985) or Van der Vaart and Wellner (1989) for an explanation.

Letting $\hat{X}_{n1}, \dots, \hat{X}_{nn}$ denote a bootstrap sample from \mathbb{P}_n^ω , the bootstrap empirical measure and process are respectively

$$\hat{\mathbb{P}}_n \equiv \frac{1}{n} \sum_{j=1}^n \delta_{\hat{X}_{nj}}$$

and

$$\hat{\mathbb{X}}_n \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n^\omega).$$

(Superscript ω indicates that the sequence of data $X_1^\omega, X_2^\omega, \dots$ is considered fixed.) Giné and Zinn (1990) proved the following remarkable result: Let $F \equiv (\sup_{f \in \mathcal{F}} |f|)^*$ denote the envelope function of the collection \mathcal{F} : that is, the smallest measurable function that majorizes \mathcal{F} . Then, under some measurability restrictions on \mathcal{F} ,

$$\hat{\mathbb{X}}_n \Rightarrow G_P \quad \text{and} \quad PF^2 < \infty$$

is equivalent to

$$\hat{\mathbb{X}}_n \Rightarrow G_P \quad \text{for almost all data sequences } X_1^\omega, X_2^\omega, \dots$$

This result completely settles the question about the validity of Efron's bootstrap in a wide range of situations. The same authors also proved "in probability" convergence in distribution results for the bootstrap.

The idea of the exchangeably-weighted bootstrap stems from the fact that the bootstrapped empirical measure can alternatively be expressed as

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{j=1}^n M_{nj} \delta_{X_j^\omega}$$

where $\underline{M}_n \sim \text{Mult}_n(n, (n^{-1}, \dots, n^{-1}))$. As observed by Efron [(1982), Section 9.4, pages 71–72], this suggests that there are, in fact, not just one but several ways to bootstrap: let $\mathbf{W} = (W_{nj}, j = 1 \dots, n, n = 1, 2, \dots)$ denote a triangular array of nonnegative random variables with $\sum_{j=1}^n W_{nj} = n$; then

$$(1.1) \quad \hat{\mathbb{P}}_n \equiv \frac{1}{n} \sum_{j=1}^n W_{nj} \delta_{X_j^\omega}$$

defines a weighted bootstrap empirical measure. We refer to these as bootstraps with exchangeable weights to distinguish them from Efron's (multinomial) bootstrap. Bootstraps with exchangeable weights have not been considered as closely as the Efron bootstrap, and they are not yet widely used in statistical practice. The references we are aware of include Rubin (1981), Efron (1982), Lo (1987), Weng (1989), Zheng and Tu (1988), Newton (1991) and Mason and Newton (1990). The best known example (and the first to our knowledge) was the Bayesian bootstrap, Rubin (1981).

In this article we establish sufficient conditions on the weights \mathbf{W} for the exchangeably weighted bootstrap to “work” asymptotically, in the sense that $\mathcal{F} \in \text{CLT}(P)$ and $PF^2 < \infty$ (and sufficient measurability) implies that, with $\hat{\mathbb{P}}_n$ given in (1.1), $\hat{\mathbb{X}}_n \equiv \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n^\omega) \Rightarrow G_P$, a.s. The methods used are, for the finite-dimensional convergence part, Hájek's (1961) central limit theorem for linear rank statistics, which was first used for a weighted bootstrap by Mason and Newton (1990). For the asymptotic equicontinuity part we use techniques from probability in Banach spaces: in particular randomization with “ $L_{2,1}$ ” bounded multipliers, Poissonization which we use analogously to Giné and Zinn (1990) for Efron's bootstrap, and the key inequality of Ledoux, Talagrand and Zinn, Ledoux and Talagrand (1988), which we combine with a reverse martingale convergence theorem and the Hewitt–Savage 0-1 law to get a result on almost sure convergence of a permutation-type empirical process. Apart from this, we use an inequality originally due to Hoeffding (1963), the essence of which is that a finite sample “without replacement” is, in a sense, bounded by a sample “with replacement” of the same size and from the same population.

Throughout, we let X_1, X_2, \dots be an iid sequence from the probability space $(\mathbf{X}, \mathcal{X}, P)$, and we take the sequence to be defined on the canonical probability space

$$(\Omega, \mathcal{O}, \text{Pr}_x) \equiv (\mathbf{X}, \mathcal{X}, P)^N.$$

The notation X_j^ω will indicate that the data are considered fixed; a typical example would be $E\|\sum_{j=1}^n \varepsilon_j \delta_{X_j^\omega}\|$, where the expectation is over $\varepsilon_1, \dots, \varepsilon_n$,

while in $E^* \|\sum_{j=1}^n \varepsilon_j \delta_{X_j}\|$ the expectation is over $\varepsilon_1, \dots, \varepsilon_n$ and X_1, \dots, X_n jointly.

For $\delta > 0$ fixed, we define the seminorm $\|\cdot\|_{\mathcal{F}'(\rho_P, \delta)}$ on $l^\infty(\mathcal{F})$ by

$$\|H\|_{\mathcal{F}'(\rho_P, \delta)} \equiv \sup\{|H(f) - H(g)| : \rho_P(f, g) < \delta\}.$$

We furthermore assume that the collection \mathcal{F} possesses enough measurability for randomization with iid multipliers to be possible; such a set of conditions is $\mathcal{F} \in \text{NLDM}(P)$, and $\mathcal{F}^2, \mathcal{F}'^2 \in \text{NLSM}(P)$ in the terminology of Giné and Zinn (1984, 1986, 1990). Here \mathcal{F}^2 and \mathcal{F}'^2 denote the classes of squared functions and squared differences of functions from \mathcal{F} , respectively. When all of these conditions hold, we write $\mathcal{F} \in M(P)$. It is known that $\mathcal{F} \in M(P)$ if \mathcal{F} is countable, or if the empirical processes X_n are stochastically separable, or if \mathcal{F} is image admissible Suslin [see Giné and Zinn (1990), pages 853 and 854].

2. Main results. This section states the main results of our paper. The corresponding proofs can be found in Section 5.

Let $\mathbf{W} \equiv (W_{nj}, j = 1, \dots, n, n = 1, 2, \dots)$ be a triangular array of random variables defined on $([0, 1], \mathcal{B}, \text{Lebesgue})$. This array determines a weighted bootstrap empirical measure by

$$(2.2) \quad \hat{\mathbb{P}}_n \equiv \frac{1}{n} \sum_{j=1}^n W_{nj} \delta_{X_j^\omega}$$

with corresponding bootstrap empirical process

$$(2.3) \quad \hat{X}_n(\omega) = \hat{X}_n^\omega \equiv \sqrt{n} (\hat{\mathbb{P}}_n - \mathbb{P}_n^\omega) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega}.$$

For a random variable Y we define $\|Y\|_{2,1} \equiv \int_0^\infty (\Pr(|Y| > t))^{1/2} dt$. Note that for $r > 2$, $(1/2)\|Y\|_2 \leq \|Y\|_{2,1} \leq (r/(r-2))\|Y\|_r$. Under the following quite general conditions on \mathbf{W} we shall establish a central limit theorem for the weighted bootstrap:

- A1. The vectors \underline{W}_n are exchangeable, $n = 1, 2, \dots$.
- A2. $W_{nj} \geq 0$, for all n, j and $\sum_{j=1}^n W_{nj} = n$, for all n .
- A3. $\sup_n \|\underline{W}_n\|_{2,1} \equiv M(\mathbf{W}) < \infty$.
- A4. $\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{t \geq \lambda} t^2 P(W_{n1} \geq t) = 0$.
- A5. $(1/n) \sum_{j=1}^n (W_{nj} - 1)^2 \rightarrow c^2 > 0$, in probability.

Our main result is the following:

THEOREM 2.1. *Let $\mathcal{F} \in M(P)$ be a class of $L_2(P)$ functions, and let \mathbf{W} be a triangular array of bootstrap weights satisfying assumptions A1–A5. Then*

$$(2.4) \quad \mathcal{F} \in \text{CLT}(P) \quad \text{and} \quad PF^2 < \infty$$

imply that

$$(2.5) \quad \hat{X}_n^\omega = \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega} \Rightarrow cG_P \text{ in } l^\infty(\mathcal{F}) \text{ a.s.},$$

where c in (2.5) is given by assumption A5.

When the envelope F is not square integrable, we obtain instead a central limit theorem “in probability” for the weighted bootstrap.

THEOREM 2.2. *Let $\mathcal{F} \in M(P)$ be a class of $L_2(P)$ functions, and let \mathbf{W} be a triangular array of bootstrap weights satisfying assumptions A1–A5. Then*

$$(2.6) \quad \mathcal{F} \in CLT(P)$$

implies that

$$(2.7) \quad \hat{X}_n^\omega = \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega} \Rightarrow cG_P \text{ in } l^\infty(\mathcal{F}) \text{ in probability,}$$

where c is given by assumption A5.

A precise restatement of (2.7) is that $d_{BL^*}(\hat{X}_n^\omega, G_P) \rightarrow 0$, in outer probability. Here, d_{BL^*} denotes the dual bounded Lipschitz metric [see, e.g., Dudley (1990), Theorem B or Van der Vaart and Wellner (1989), Corollary 1.5] which metrizes convergence in distribution in $l^\infty(\mathcal{F})$.

The calculations leading to Theorem 2.1 also show the following new result about the Efron bootstrap with arbitrary bootstrap sample size. Form an iid sample $\hat{X}_{n1}, \dots, \hat{X}_{nm}$ from \mathbb{P}_n^ω , and let $\hat{\mathbb{P}}_{m,n} \equiv m^{-1} \sum_{j=1}^m \delta_{\hat{X}_{nj}}$ denote the bootstrap empirical measure for the bootstrap sample of size m .

COROLLARY 2.1. *$\mathcal{F} \in CLT(P)$ and $PF^2 < \infty$ imply that*

$$\sqrt{m} (\hat{\mathbb{P}}_{m,n} - \mathbb{P}_n^\omega) = \frac{1}{\sqrt{m}} \sum_{j=1}^m (\delta_{\hat{X}_{nj}} - \mathbb{P}_n^\omega) \Rightarrow G_P \text{ in } l^\infty(\mathcal{F}) \text{ a.s. as } n \wedge m \rightarrow \infty.$$

This result for regular sequences m_n and the corresponding “in probability” result in general were known to Giné and Zinn [Arcones and Giné (1992)].

COROLLARY 2.2. *$\mathcal{F} \in CLT(P)$ implies that*

$$\sqrt{m} (\hat{\mathbb{P}}_{m,n} - \mathbb{P}_n^\omega) \Rightarrow G_P \text{ in } l^\infty(\mathcal{F}) \text{ in probability as } n \wedge m \rightarrow \infty.$$

3. Examples. This section contains some examples of weighted bootstraps which satisfy the conditions A1–A5. Our intention in including these examples is to show the scope of the class of weights that are covered by the results of the preceding section. In each case, we have postponed the verification of conditions A1–A5 to Section 5.

EXAMPLE 3.1 (iid-weighted bootstraps). Let Y_1, Y_2, \dots be iid, positive random variables where $\|Y_1\|_{2,1} < \infty$, and define bootstrap weights by $W_{nj} \equiv Y_j/\bar{Y}_n$. By taking, for instance, Y_i iid exponential(1), the weights become Dirichlet $_n(1, \dots, 1)$, and we have the Bayesian bootstrap of Rubin (1981) and Lo (1987). When the Y_i 's are iid Gamma(4, 1) [so that the W_{ni}/n are equivalent to four-spacings from a sample of $4n - 1$ Uniform(0, 1) random variables], this "iid-weighted" bootstrap is second order equivalent to Efron's multinomial bootstrap for bootstrapping the sample mean, as noted by Weng (1989). Intuitively, these bootstraps are "smoother" in some sense than the multinomial bootstrap since they put some (random) weight at *all* of the X_i^ω 's in the sample, whereas the multinomial bootstrap puts positive weight at about $1 - (1 - n^{-1})^n \rightarrow 1 - e^{-1} \approx 0.6322$ proportion of the X_i^ω 's, on the average. For further comparisons, see Table 1.

For the class of weights of this example, Præstgaard (1990) has shown that (2.5) implies (2.4), in parallel to the results for Efron's bootstrap in Giné and Zinn (1990) and the almost sure multiplier central limit theorem in Ledoux and Talagrand (1988, 1991). In Section 5 we prove that this bootstrap satisfies A1–A5 with $c^2 = \text{Var } Y_1/(EY_1)^2$.

When checking conditions A1–A5, the following lemma is useful if the bootstrap weights possess moments of a higher order than assumed in Theorem 2.1.

LEMMA 3.1. *Let \mathbf{W} be a triangular array of bootstrap weights satisfying assumptions A1 and A2. If further the following conditions are satisfied:*

- B1. $\limsup_{n \rightarrow \infty} EW_{n1}^4 < \infty$,
- B2. $EW_{n1}^2 \rightarrow 1 + c^2$,
- B3. $\text{Cov}(W_{n1}^2, W_{n2}^2) \leq 0$,

then \mathbf{W} satisfies A3, A4 and A5.

REMARK 3.1. Exchangeable random variables W_i , $i = 1, \dots, n$, with fixed sum n have negative correlations $-1/(n - 1)$; see, for example, Aldous (1983), page 8. It seems intuitive that their squares W_i^2 should also be negatively, or at least nonpositively, correlated, and hence one would suspect that assumption B3 follows from A1 and A2. The following example (due to Chris Klaassen) shows that this is not the case. Consider the triangular array of constants $\mathbf{w} = (w_n, n = 1, 2, \dots)$ where $w_{nj} = 0$, $j \leq [n/2]$; $w_{nj} = 2$, $j > [n/2]$. Let R_n be a random permutation uniformly distributed on Π_n , and let $0 < p < 1$. Define an array of n -dimensional random vectors by

$$\underline{W}_n = \begin{cases} (1, \dots, 1), & \text{with probability } 1 - p, \\ (w_{nR_n(1)}, \dots, w_{nR_n(n)}), & \text{with probability } p. \end{cases}$$

Then it is easily shown that $\text{Cov}(W_{n1}^2, W_{n2}^2) \leq 0$ if and only if

$$p(1 - p)(1 - \overline{w_n^2}) \leq \frac{p}{n - 1} (\overline{w_n^4} - (\overline{w_n^2})^2),$$

where $\overline{w_n^r} \equiv n^{-1} \sum_{j=1}^n w_{nj}^r$. For the particular choice of \mathbf{w} above, this fails when, for example, $p = 1/2$ and $n > 9$. In this example, A5 also fails because

$$\frac{1}{n} \sum_{j=1}^n (W_{nj} - 1)^2 =_d \text{Bernoulli}(p).$$

PROOF OF LEMMA 3.1. For any nonnegative r.v. Y ,

$$\begin{aligned} \|Y\|_{2,1} &= \int_0^\infty \sqrt{\Pr(Y > t)} dt \\ &\leq 1 + \int_1^\infty \sqrt{EY^4/t^4} dt \\ &= 1 + \sqrt{EY^4}, \end{aligned}$$

and hence B1 shows that A3 is satisfied. Condition A4 is straightforwardly satisfied by B1 and Chebyshev's inequality. To see finally that A5 holds, note that

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n (W_{nj} - 1)^2 - c^2 &= \frac{1}{n} \left(\sum_{j=1}^n W_{nj}^2 - 1 \right) - c^2 \quad \left(\text{since } \sum_{j=1}^n W_{nj} = n \right) \\ &= \frac{1}{n} \sum_{j=1}^n (W_{nj}^2 - EW_{nj}^2) + EW_{n1}^2 - (1 + c^2), \end{aligned}$$

and hence it suffices to show that

$$(3.8) \quad \frac{1}{n} \sum_{j=1}^n (W_{nj}^2 - EW_{nj}^2) \rightarrow 0 \quad \text{in probability.}$$

Since W_{ni}^2 and W_{nj}^2 are nonpositively correlated for $i \neq j$, it follows by Chebyshev's inequality that

$$\begin{aligned} \Pr \left(\left| \frac{1}{n} \sum_{j=1}^n (W_{nj}^2 - EW_{nj}^2) \right| > \varepsilon \right) &\leq \varepsilon^{-2} n^{-2} \text{Var} \left(\sum_{j=1}^n (W_{nj}^2 - EW_{nj}^2) \right)^2 \\ &\leq \varepsilon^{-2} n^{-2} \sum_{j=1}^n E(W_{nj}^2 - EW_{nj}^2)^2 \\ &= \varepsilon^{-2} n^{-1} E(W_{n1}^2 - EW_{n1}^2)^2 \rightarrow 0. \quad \square \end{aligned}$$

• We now present some examples of bootstrap weights satisfying B1–B3.

EXAMPLE 3.2 (Efron's bootstrap). Indeed, the weights for the Efron bootstrap satisfy B1–B3 with $c^2 = 1$. As already mentioned, the weights for this bootstrap are $\underline{W}_n \sim \text{Mult}_n(n, (n^{-1}, \dots, n^{-1}))$.

EXAMPLE 3.3 (A double bootstrap). The weights for the double bootstrap emerge by “sampling from the sample:” draw a bootstrap sample $(\hat{X}_1, \dots, \hat{X}_n)$ with replacement from the observed data (X_1, \dots, X_n) , and draw subsequently with replacement a new sample of size n from $(\hat{X}_1, \dots, \hat{X}_n)$. It follows that the resulting weights have a distribution

$$(W_{n1}, \dots, W_{nn}) \sim \text{Mult}_n\left(n, \left(\frac{M_{n1}}{n}, \dots, \frac{M_{nn}}{n}\right)\right),$$

conditional on \underline{M}_n where $\underline{M}_n \sim \text{Mult}_n(n, (n^{-1}, \dots, n^{-1}))$.

In Section 5 we prove that the double bootstrap weights satisfy B1–B3 with $c^2 = 2$.

With the double bootstrap, the bootstrap sample is more likely to concentrate on a few data points than with the ordinary (Efron) bootstrap. The next example concerns a bootstrap which is more “spread out” than the ordinary bootstrap, distributing the mass more evenly over the sample. In other words, the variance of the bootstrap weights is smaller than $\text{Var } M_{n1} = (n - 1)/n$.

EXAMPLE 3.4 (The multivariate hypergeometric bootstrap). This bootstrap emerges from the following urn scheme: Put K copies of each observed data point X_1, \dots, X_n in an urn, so that the urn contains $K \cdot n$ elements. Draw from this urn a sample of size n without replacement. The resulting weights W_{n1}, \dots, W_{nn} follow a multivariate hypergeometric distribution [see, e.g., Johnson and Kotz (1977), page 91] with density

$$\Pr(W_{n1} = w_1, \dots, W_{nn} = w_n) = \frac{\binom{K}{w_1} \cdots \binom{K}{w_n}}{\binom{nK}{n}}, \quad \sum_{j=1}^n w_j = n, 0 \leq w_j \leq K.$$

For this bootstrap, one finds that B1–B3 are satisfied with $c^2 = (K - 1)/K$.

The multivariate hypergeometric distribution is an example of a so-called “urn model.” This class of models yields a number of alternative bootstraps. Here, we shall only consider one more example.

EXAMPLE 3.5 (A Polya–Eggenberger bootstrap). Consider the following way of resampling the data X_1, \dots, X_n : Put K copies of each data point in an urn; a sample of size n is then drawn with replacement such that after each draw, a number s of points equal to the one sampled are placed in the urn. Hence, intuitively, when a point is sampled, it is likely to be sampled again, just as with the double bootstrap.

Letting $\alpha \equiv K/s$ it turns out [see, e.g., Johnson and Kotz (1977), page 196] that the resulting weights have a density which is a mixture of the form

$$(W_{n1}, \dots, W_{nn}) \sim \text{Mult}_n(n, (D_{n1}, \dots, D_{nn})),$$

conditional on (D_{n1}, \dots, D_{nn}) , where $(D_{n1}, \dots, D_{nn}) \equiv \underline{D}_n \sim \text{Dirichlet}_n(\alpha, \dots, \alpha)$.

The Polya–Eggenberger bootstrap satisfies B1–B3 and hence A1–A5 with $c^2 = (\alpha + 1)/\alpha$.

The construction in Remark 3.1 is an example of an interesting class of bootstrap weights which are generated from a deterministic array by permuting the elements randomly.

EXAMPLE 3.6 (A bootstrap generated from deterministic weights). Let $\mathbf{w} = (w_{nj}, j = 1, \dots, n, n = 1, 2, \dots)$ be a deterministic array of nonnegative numbers such that $\sum_{j=1}^n w_{nj} = n$ for all n . Let R_n be a random permutation uniformly distributed on Π_n . Then $W_{nj} \equiv w_{nR_n(j)}$ defines an array of bootstrap weights.

For these weights, the conditions A1 and A2 are obviously satisfied. Since $\Pr(W_{n1} > t) = \{\#j: w_{nj} > t\}/n \equiv \phi_n(t)$, the remaining conditions translate to checking for the following:

- A3. $\sup \int_0^\infty \sqrt{\phi_n(t)} dt < \infty$.
 A4. $\lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} t^2 \phi_n(t) = 0$.
 A5. $\frac{1}{n} \sum_{j=1}^n (w_{nj} - 1)^2 \rightarrow c^2 > 0$.

An important special case is the grouped, or delete- h , jackknife [Efron (1982), Section 2.2; Wu (1987) and Shi (1991)]. The grouped jackknife with group block size h may be viewed as a bootstrap generated by permuting the deterministic weights

$$\underline{w}_n \equiv \left(\underbrace{\frac{n}{n-h}, \dots, \frac{n}{n-h}}_{n-h}, \underbrace{0, \dots, 0}_h \right).$$

In this case,

$$\phi_n(t) = \frac{n-h}{n} 1_{\{t < n/(n-h)\}}.$$

For condition A5 we note that $n^{-1} \sum_{j=1}^n (w_{nj} - 1)^2 = h/(n-h)$, and hence for A5 to be satisfied we must let the block size $h \equiv h_n$ depend on n in such a way that $h_n/n \rightarrow \alpha \in (0, 1)$. Under this condition it also holds that

$$\int_0^\infty \sqrt{\phi_n(t)} dt = \left(\frac{n}{n-h_n} \right)^{1/2} \rightarrow \left(\frac{1}{1-\alpha} \right)^{1/2} < \infty$$

and

$$\lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} t^2 \phi_n(t) = \lim_{t \rightarrow \infty} (1-\alpha) t^2 1_{\{t < (1-\alpha)^{-1}\}} = 0.$$

Hence the grouped jackknife with block size h_n satisfies A1–A5 with $c^2 = \lim_{n \rightarrow \infty} h_n/(n-h_n) = \alpha/(1-\alpha) \in (0, \infty)$. Note that the ordinary jackknife

TABLE 1
Properties of bootstrap weights

Example	c^2	$\Pr(W_{n1} = 0)$	$\text{Var } W_{n1}$
1	$\frac{\text{Var } Y_1}{EY_1^2}$	0	$\text{Var } \frac{Y_1}{\bar{Y}_n}$
2	1	$\left(1 - \frac{1}{n}\right)^n \rightarrow e^{-1} = 0.368$	$\frac{n-1}{n}$
3	2	$E\left(1 - \frac{M_n}{n}\right)^n \rightarrow e^{e^{-1}-1} = 0.531$	$\frac{(n-1)(2n-1)}{n^2}$
4	$\frac{K-1}{K}$	$\binom{(n-1)K}{n} / \binom{nK}{n} \rightarrow \left(1 - \frac{1}{K}\right)^K$	$\frac{(n-1)(K-1)}{nK-1}$
5	$\frac{\alpha+1}{\alpha}$	$E(1 - D_{n1})^n \rightarrow 1/2$	$(\alpha+1)\frac{n-1}{n\alpha+1}$
6	$\lim_n \frac{1}{n} \sum_{j=1}^n (w_{nj} - 1)^2$	$\frac{\#\{i \leq n: w_{ni} = 0\}}{n}$	$w_2^2 - 1$
6JK	$\frac{\alpha}{1-\alpha}$	$\frac{h}{n} \rightarrow \alpha$	$\frac{h}{n-h}$

with $h = 1$ has $\alpha = 0$ so $c^2 = 0$, and hence the ordinary jackknife fails to satisfy our conditions, in agreement with the remarks of Efron [(1982), page 39].

Table 1 summarizes some properties of Examples 3.1–3.6.

4. Tools and technical results. In this section we state briefly the key results we use in our proofs.

The most important tool for showing a.s. asymptotic equicontinuity of the weighted bootstrap empirical process is the following inequality due to Hoeffding (1963).

THEOREM 4.1 (Hoeffding’s inequality). *Let (c_1, \dots, c_N) be elements of a vector space \mathbf{V} , and let (U_1, \dots, U_n) and (V_1, \dots, V_n) denote, respectively, a sample without and with replacement of size $n \leq N$ from (c_1, \dots, c_N) . Let $\varphi: \mathbf{V} \rightarrow \mathbb{R}$ be a convex function. Then*

$$E\varphi\left(\sum_{j=1}^n U_j\right) \leq E\varphi\left(\sum_{j=1}^n V_j\right).$$

We remark that the original result of Hoeffding (1963) required $\mathbf{V} \equiv \mathbb{R}$ and φ convex and continuous. It turns out, however, that the structure of \mathbf{V} is

immaterial; we only need to ensure that addition and multiplication with scalars are well defined. In our particular application we shall take $\mathbf{V} = l^\infty(\mathcal{F})$.

PROOF OF THEOREM 4.1. The proof follows from Corollary 12.A.2.B of Marshall and Olkin (1979). These authors state their result for the particular case $\mathbf{V} = \mathbb{R}$; by inspecting their proof one sees, however, that this additional structure on \mathbf{V} is never used. \square

The following result uses Hoeffding’s inequality to bound the expectation of the randomly permuted sum by the expectation of a quantity related to the Efron-bootstrapped empirical process.

COROLLARY 4.1. *Let $\hat{X}_{n_1}, \dots, \hat{X}_{n_n}$ denote an iid sample from \mathbb{P}_n^ω , and let R be a random permutation of $\{1, \dots, n\}$. Then for any $n_0 \leq n$,*

$$E \left\| \sum_{i=1}^{n_0} (\delta_{X_{R(i)}^\omega} - \mathbb{P}_n^\omega) \right\| \leq E \left\| \sum_{i=1}^{n_0} (\delta_{\hat{X}_{n_i}} - \mathbb{P}_n^\omega) \right\|.$$

PROOF OF COROLLARY 4.1. This follows immediately from Theorem 4.1 with $n = n_0$, $N = n$, $\mathbf{V} = l^\infty(\mathcal{F})$, the population (c_1, \dots, c_N) equal to $(\delta_{X_1^\omega} - \mathbb{P}_n^\omega, \dots, \delta_{X_n^\omega} - \mathbb{P}_n^\omega)$, the sample size n_0 , and the convex function $\varphi = \|\cdot\|$, upon noting that $E\phi(\sum_{j=1}^{n_0} U_j) = E\phi(\sum_{j=1}^{n_0} c_{R(j)})$. \square

The following result shows how to bound the expectation of the norm of the bootstrap empirical process by the expectation of a “randomly permuted” sum. It is a variant of the Pisier–Fernique inequality found in Giné and Zinn [(1984), Lemma 2.9], or Giné and Zinn [(1986), Lemma 1.2.4]. We need to state it in a form which holds for possibly nonmeasurable seminorms. In Dudley (1984) it is shown that a measurable cover $\|\cdot\|^*$ of a norm $\|\cdot\|$ can be defined so that it still has the properties of the norm.

LEMMA 4.1. *Let $\underline{\xi} \equiv (\xi_1, \dots, \xi_n)$ be a nonnegative, exchangeable random vector with $\|\xi_1\|_{2,1} < \infty$, and let R denote a random permutation uniformly distributed on Π_n , the set of permutations of $1, \dots, n$. Let $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ be random elements of $l^\infty(\mathcal{F})$ so that $((\underline{\xi}, R))$ and $(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ are independent (in fact defined on a product probability space). Let $\|\cdot\|$ denote a pseudonorm on $l^\infty(\mathcal{F})$. Then for any $n_0 < n$,*

$$(4.9) \quad E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j \mathbf{Z}_j \right\| \leq \frac{n_0}{\sqrt{n}} E \left(\max_{j \leq n} \xi_j \right) \frac{1}{n} E^* \sum_{j=1}^n \|\mathbf{Z}_j\|$$

$$+ \|\xi_1\|_{2,1} \max_{n_0 < k \leq n} E^* \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k \mathbf{Z}_{R(j)} \right\|,$$

where in the second line the expectation is with respect to both $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ and R .

In most of our applications of Lemma 4.1 the \mathbf{Z}_j 's are deterministic.

PROOF OF LEMMA 4.1. Define a random permutation S of $\{1, \dots, n\}$ by demanding that

$$\xi_{S(1)} \geq \dots \geq \xi_{S(n)},$$

and if

$$\xi_{S(j)} = \xi_{S(j+1)} \text{ then } S(j) < S(j + 1).$$

This definition of the vector of antiranks of ξ_1, \dots, ξ_n is one of many which is not ambiguous in the presence of ties. Furthermore, let R be a random permutation uniformly distributed on Π_n and independent of $(\underline{\xi}, S)$. By exchangeability of $\underline{\xi}$ we have that

$$E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j \mathbf{Z}_j \right\| = E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_j \mathbf{Z}_{R(j)} \right\|,$$

which, on defining $\xi_{(j)} \equiv \xi_{S(j)}$ further equals

$$(4.10) \quad E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_{(j)} \mathbf{Z}_{R \circ S(j)} \right\|.$$

One can easily verify that $R \circ S$ is distributed as R and is independent of S . Hence, by the triangle inequality,

$$(4.11) \quad \begin{aligned} E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_{(j)} \mathbf{Z}_{R(j)} \right\| &\leq E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{n_0} \xi_{(j)} \mathbf{Z}_{R(j)} \right\| \\ &+ E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=n_0+1}^n \xi_{(j)} \mathbf{Z}_{R(j)} \right\| \\ &\equiv \text{I}(n_0, n) + \text{II}(n_0, n). \end{aligned}$$

We bound $\text{I}(n_0, n)$ by

$$(4.12) \quad \text{I}(n_0, n) \leq E^* \left(\frac{1}{\sqrt{n}} \left(\max_{j \leq n} \xi_j \right) \sum_{j=1}^{n_0} \|\mathbf{Z}_{R(j)}\| \right)$$

$$(4.13) \quad = n_0 n^{-1/2} E \left(\max_{j \leq n} \xi_j \right) n^{-1} E^* \sum_{j=1}^n \|\mathbf{Z}_j\|,$$

where the last equality follows by integrating out R , which is independent of the \mathbf{Z}_j 's.

The second term in (4.11) we write as a telescoping sum [defining $\xi_{(n+1)} \equiv 0$]:

$$\begin{aligned}
 \text{II}(n_0, n) &= E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=n_0+1}^n \left(\sum_{k=j}^n \xi_{(k)} - \xi_{(k+1)} \right) \mathbf{Z}_{R(j)} \right\| \\
 &= E^* \left\| \frac{1}{\sqrt{n}} \sum_{k=n_0+1}^n (\xi_{(k)} - \xi_{(k+1)}) \sum_{j=n_0+1}^k \mathbf{Z}_{R(j)} \right\| \\
 (4.14) \quad &\leq \frac{1}{\sqrt{n}} \sum_{k=n_0+1}^n E^* \left(\sqrt{k} (\xi_{(k)} - \xi_{(k+1)}) \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k \mathbf{Z}_{R(j)} \right\| \right) \\
 &= \frac{1}{\sqrt{n}} \sum_{k=n_0+1}^n E(\sqrt{k} (\xi_{(k)} - \xi_{(k+1)})) E^* \left(\left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k \mathbf{Z}_{R(j)} \right\| \right) \\
 &\leq \frac{1}{\sqrt{n}} \sum_{k=n_0+1}^n E(\sqrt{k} (\xi_{(k)} - \xi_{(k+1)})) \max_{n_0 < k \leq n} E^* \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k \mathbf{Z}_{R(j)} \right\|.
 \end{aligned}$$

Since finally we have the bound

$$\begin{aligned}
 \frac{1}{\sqrt{n}} \sum_{k=n_0+1}^n E\sqrt{k} (\xi_{(k)} - \xi_{(k+1)}) &\leq \frac{1}{\sqrt{n}} \sum_{k=1}^n E\sqrt{k} (\xi_{(k)} - \xi_{(k+1)}) \\
 &= \frac{1}{\sqrt{n}} \int_0^{\xi_{(1)}} E\sqrt{\#j: \xi_j > t} dt \\
 (4.15) \quad &\leq \frac{1}{\sqrt{n}} \int_0^{\xi_{(1)}} \sqrt{E(\#j: \xi_j > t)} dt \\
 &= \frac{1}{\sqrt{n}} \int_0^{\xi_{(1)}} \sqrt{n \Pr(\xi_1 > t)} dt \\
 &\leq \|\xi_1\|_{2,1},
 \end{aligned}$$

the lemma follows. \square

We use Le Cam’s Poissonization lemma in the following version, which is easily proved given the background in Giné and Zinn [(1990), Lemma 2.1]. A symmetrized Poisson variable with parameter λ is the difference of two iid Poisson (λ) variables.

LEMMA 4.2. *Let $\hat{X}_{n1}, \dots, \hat{X}_{nm}$ be iid $\mathbb{P}_n^\omega \equiv n^{-1} \sum_{j=1}^n \delta_{X_j^\omega}$, where we do not assume $n = m$. Let $\tilde{N}_1(m/2n), \dots, \tilde{N}_n(m/2n)$ be iid symmetrized Poisson($m/2n$). Then, for any seminorm $\|\cdot\|$,*

$$E \left\| \sum_{j=1}^m (\delta_{\hat{X}_{nj}} - \mathbb{P}_n^\omega) \right\| \leq 4E \left\| \sum_{i=1}^n \tilde{N}_i \left(\frac{m}{2n} \right) \delta_{X_i^\omega} \right\|.$$

The following lemma is necessary for using the Ledoux, Talagrand and Zinn inequality [see Giné and Zinn (1990), Lemma 2.3] when the empirical point masses have been permuted by a random permutation. It can be viewed as a variant of Hunt’s lemma; see for example, Dellacherie and Meyer [(1980), Chapters V–VIII, page 43].

LEMMA 4.3. *Let Z_n be a sequence of nonnegative random variables defined on $(\mathbf{X}, \mathcal{X}, P)^N$ such that Z_n is $\sigma(X_1, \dots, X_n)$ measurable for each n . Assume that*

$$\limsup_{n \rightarrow \infty} Z_n \equiv Z \leq C < \infty \quad \text{a.s.}$$

and

$$E \sup_{n \geq 1} Z_n < \infty.$$

Let \mathcal{S}_n be the σ -field generated by all \mathcal{X}^N -measurable functions $f: \mathbf{X}^N \rightarrow \mathbb{R}$ that are symmetric in the first n coordinates, and set $\tilde{Z}_{ln} \equiv E(Z_l | \mathcal{S}_n)$. Then

$$\limsup_{n \wedge n_1 \rightarrow \infty} \max_{n_1 < l \leq n} \tilde{Z}_{ln} \leq C \quad \text{a.s.}$$

PROOF OF LEMMA 4.3. We write

$$\tilde{Z}_{ln} = E(Z_l - Z | \mathcal{S}_n) + E(Z | \mathcal{S}_n).$$

The sequence $E(Z | \mathcal{S}_n)$ is a reverse martingale with respect to the decreasing σ -fields \mathcal{S}_n , and by the convergence theorem for reverse martingales, (e.g., Dudley [(1989), Theorem 10.6.4, page 241])

$$E(Z | \mathcal{S}_n) \rightarrow E(Z | \mathcal{S}) \quad \text{a.s.}$$

where \mathcal{S} is the symmetric σ -field on \mathbf{X}^N . By the Hewitt–Savage 0-1 law (c.f. Hewitt and Savage (1955) or Dudley [(1989), page 213]), \mathcal{S} contains only sets of probability 0 or 1. Hence $E(Z | \mathcal{S}) = EZ \leq C$ a.s., and the claim of the lemma will follow if we can show that

$$\limsup_{n \wedge n_1 \rightarrow \infty} \max_{n_1 \leq l \leq n} E(Z_l - Z | \mathcal{S}_n) \leq 0 \quad \text{a.s.}$$

Define $M_l \equiv \max_{m \geq l} Z_m$; then

$$\begin{aligned} \limsup_{n \wedge n_1 \rightarrow \infty} \max_{n_1 \leq l \leq n} E(Z_l - Z | \mathcal{S}_n) &\leq \limsup_{n \wedge n_1 \rightarrow \infty} \max_{n_1 < l \leq n} E(M_l - Z | \mathcal{S}_n) \\ (4.16) \qquad \qquad \qquad &\leq \limsup_{n \rightarrow \infty} E(M_m - Z | \mathcal{S}_n) \end{aligned}$$

for any $m = 1, 2, \dots$,

where the last inequality follows because $\{M_l\}$ is a decreasing sequence. By the reverse martingale theorem and the Hewitt–Savage 0-1 law again, it follows that (4.16) equals $E(M_m - Z | \mathcal{S}) = E(M_m - Z)$ a.s. Finally, $M_m \rightarrow Z$, a.s. and $E(M_m - Z) \rightarrow 0$ as $m \rightarrow \infty$ by the dominated convergence theorem. \square

The following two results give moment bounds for symmetrized Poisson random variables. Statements of these results appear in Arcones and Giné (1992); their proofs were communicated to us by these authors.

LEMMA 4.4. *Let $\tilde{N}(\lambda)$ be distributed as symmetrized Poisson(λ). Then*

$$\|\tilde{N}(\lambda)/\sqrt{\lambda}\|_{2,1} \equiv \int_0^\infty \sqrt{\Pr(|\tilde{N}(\lambda)|/\sqrt{\lambda} > t)} dt \leq 4$$

for all $\lambda > 0$.

PROOF OF LEMMA 4.4. Recall that $\tilde{N}(\lambda)$ is distributed as the difference of two iid Poisson (λ) variables N and N' . By direct calculation we have the identity

$$(4.17) \quad E\tilde{N}(\lambda)^4 = 12\lambda^2 + 2\lambda \quad \text{for all } \lambda > 0.$$

Furthermore, for any $t > 0$,

$$\begin{aligned} \Pr(|\tilde{N}(\lambda)| > t) &\leq 1 - \Pr(\tilde{N}(\lambda) = 0) \\ &\leq 1 - \Pr(N(\lambda) = 0, N'(\lambda) = 0) \\ &= 1 - e^{-2\lambda} \leq 2\lambda \end{aligned}$$

and hence by (4.17) and Markov's inequality we obtain the bounds

$$(4.18) \quad \Pr\left(\frac{|\tilde{N}(\lambda)|}{\sqrt{\lambda}} > t\right) \leq \begin{cases} 1 \wedge 2\lambda, \\ (12\lambda + 2)/\lambda t^4. \end{cases}$$

Use the bounds (4.18) and Markov's inequality to show that, for any $c > 0$,

$$(4.19) \quad \begin{aligned} \int_0^\infty \sqrt{\Pr(|\tilde{N}(\lambda)|/\sqrt{\lambda} > t)} dt &\leq c\sqrt{1 \wedge 2\lambda} + \left(\frac{12\lambda + 2}{\lambda}\right)^{1/2} \int_c^\infty \frac{1}{t^2} dt \\ &= c\sqrt{1 \wedge 2\lambda} + \frac{1}{c} \left(\frac{12\lambda + 2}{\lambda}\right)^{1/2}. \end{aligned}$$

Choosing

$$c \equiv \left(\frac{12\lambda + 2}{\lambda(1 \wedge 2\lambda)}\right)^{1/4}$$

shows that (4.19) is bounded by $2((12\lambda + 2)(1 \wedge 2\lambda)/\lambda)^{1/4} \leq 4$. \square

LEMMA 4.5. *Let $\tilde{N}_1(\lambda), \dots, \tilde{N}_n(\lambda)$ be iid symmetrized Poisson(λ). Then*

$$E\left(\max_{i \leq n} |\tilde{N}_i(\lambda)|/\sqrt{\lambda}\right) \leq 2\sqrt{n} \left(\frac{12}{n} + \frac{2}{\lambda n}\right)^{1/4}.$$

PROOF OF LEMMA 4.5. Let $N_j \equiv \tilde{N}_j(\lambda)$. Determine the integer $c > 0$ such that $\Pr(|N_j| > c) \leq n^{-1} \leq \Pr(|N_j| \geq c)$. Then, first of all, by (4.17) it holds if

$c > 0$ that $n^{-1} \leq \Pr(|N_j| \geq c) \leq EN_j^4/c^4 = (12\lambda^2 + 2\lambda)/c^4$, and hence

$$(4.20) \quad c \leq n^{1/4}(12\lambda^2 + 2\lambda)^{1/4},$$

which holds even when $c = 0$.

We now have the bound

$$\begin{aligned} E\left(\lambda^{-1/2} \max_{j \leq n} |N_j|\right) &\leq c\lambda^{-1/2} + E\left(\lambda^{-1/2} \max_{j \leq n} |N_j| 1_{\{|N_j| > c\}}\right) \\ &\leq c\lambda^{-1/2} + n\lambda^{-1/2} E(|N_1| 1_{\{|N_1| > c\}}) \\ &\leq c\lambda^{-1/2} + n\lambda^{-1/2} (EN_1^4 1_{\{|N_1| > c\}})^{1/4} \Pr(|N_1| > c)^{3/4} \\ &\leq c\lambda^{-1/2} + n\lambda^{-1/2} (12\lambda^2 + 2\lambda)^{1/4} n^{-3/4}, \end{aligned}$$

where we have used Hölder’s inequality, (4.18), and the definition of c . Finally insert (4.20) for c to get the bound

$$n^{1/4}\lambda^{-1/2}(12\lambda^2 + 2\lambda)^{1/4} + n^{-1/4}\lambda^{-1/2}(12\lambda^2 + 2\lambda)^{1/4} = 2\sqrt{n} \left(\frac{12}{n} + \frac{2}{\lambda n}\right)^{1/4}.$$

□

For proving finite-dimensional convergence in distribution we use the following lemma. It is essentially a variation on a central limit theorem for linear rank statistics due to Hájek (1961). A similar argument is used by Mason and Newton (1990); we state it in detail.

LEMMA 4.6. *Let $\{m\}$ be a sequence of natural numbers, let $\{a_{mj}\}$ be a triangular array of constants, and let B_{mj} , $j = 1, \dots, m$, $m \in \{m\}$ be a triangular array of row-exchangeable random variables such that*

$$(4.21) \quad \frac{1}{m} \sum_{j=1}^m (a_{mj} - \bar{a}_m)^2 \rightarrow \sigma^2 > 0; \quad \frac{1}{m} \max_{j \leq m} (a_{mj} - \bar{a}_m)^2 \rightarrow 0,$$

$$(4.22) \quad \frac{1}{m} \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2 \rightarrow \alpha^2 > 0 \quad \text{in probability,}$$

$$(4.23) \quad \lim_{K \rightarrow \infty} \limsup_{m \rightarrow \infty} \| |B_{m1} - \bar{B}_m| 1_{\{|B_{m1} - \bar{B}_m| > K\}} \|_2 = 0.$$

Then

$$\frac{1}{\sqrt{m}} \sum_{j=1}^m (a_{mj} B_{mj} - \bar{a}_m \bar{B}_m) \Rightarrow N(0, \alpha^2 \sigma^2).$$

[Condition (4.23) means, of course, that the sequence $\{B_{m1} - \bar{B}_m\}$ is uniformly square-integrable.]

PROOF OF LEMMA 4.6. In the following proof it shall be tacitly understood that m runs only in the sequence $\{m\}$.

Define

$$u_{mi}^2 \equiv \frac{(a_{mi} - \bar{a}_m)^2}{\sum_{i=1}^m (a_{mi} - \bar{a}_m)^2}; \quad V_{mj}^2 \equiv \frac{(B_{mj} - \bar{B}_m)^2}{\sum_{j=1}^m (B_{mj} - \bar{B}_m)^2}.$$

From (4.21) we notice that

$$(4.24) \quad \max_{i \leq m} u_{mi}^2 \rightarrow 0,$$

and by (4.22), (4.23) and Lemma 4.7 (the second implication) applied to the exchangeable array $\{|B_{mj} - \bar{B}_m|\}$, it follows that

$$(4.25) \quad \max_{j \leq m} V_{mj}^2 \rightarrow 0 \quad \text{in probability.}$$

Defining

$$\delta_{mij}^2 \equiv m U_{mi}^2 V_{mj}^2 \equiv m \frac{(a_{mi} - \bar{a}_m)^2 (B_{mj} - \bar{B}_m)^2}{\sum_{i=1}^m (a_{mi} - \bar{a}_m)^2 \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2},$$

we shall next show that for all $\tau > 0$,

$$(4.26) \quad \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \delta_{mij}^2 \mathbf{1}_{\{\delta_{mij}^2 > \tau^2\}} = 0 \quad \text{in probability.}$$

For any $\varepsilon > 0$ we can, by (4.24), find N_ε such that $\max_{i \leq m} u_{mi}^2 < \varepsilon$ when $m > N_\varepsilon$, and setting

$$A_m \equiv \left\{ \frac{1}{m} \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2 > \frac{\alpha^2}{2} \right\},$$

it follows that for $m > N_\varepsilon$,

$$\begin{aligned} \{\delta_{mij}^2 > \tau^2\} \cap A_m &\subset \{m V_{mj}^2 > \tau^2 \varepsilon^{-1}\} \cap A_m \\ &\subset \left\{ (B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2 \right\}. \end{aligned}$$

Hence, for any $a > 0$, when $m > N_\varepsilon$, on the set A_m ,

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \delta_{mij}^2 \mathbf{1}_{\{\delta_{mij}^2 > \tau^2\}} &\leq \sum_{i=1}^m \sum_{j=1}^m u_{mi}^2 V_{mj}^2 \mathbf{1}_{\{(B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}} \\ &= \sum_{j=1}^m V_{mj}^2 \mathbf{1}_{\{(B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}} \quad \left(\text{since } \sum_{i=1}^m u_{mi}^2 = 1 \right) \\ &= \frac{(1/m) \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2 \mathbf{1}_{\{(B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}}}{(1/m) \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2} \\ &\leq \frac{2}{\alpha^2} \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2 \mathbf{1}_{\{(B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}}. \end{aligned}$$

By (4.22) we can pick M_ε such that $\Pr(A_m) > 1 - \varepsilon$ when $m > M_\varepsilon$, and it follows that when $m > M_\varepsilon \vee N_\varepsilon$,

$$\begin{aligned}
 & \Pr\left(\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \delta_{mij}^2 \mathbf{1}_{\{\delta_{mij}^2 > \tau^2\}} > \alpha\right) \\
 (4.27) \quad & \leq \varepsilon + \Pr\left(\frac{1}{m} \sum_{j=1}^m (B_{mj} - \bar{B}_m)^2 \mathbf{1}_{\{(B_{mj} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}} > \alpha\right) \\
 & \leq \varepsilon + \frac{2}{\alpha^2 \alpha} E(B_{m1} - \bar{B}_m)^2 \mathbf{1}_{\{(B_{m1} - \bar{B}_m)^2 > \tau^2 \varepsilon^{-1} \alpha^2 / 2\}},
 \end{aligned}$$

where the last line follows by Markov’s inequality and exchangeability. By (4.23) it follows that (4.27) can be made arbitrarily small by picking ε small which proves (4.26).

We now strengthen (4.25) and (4.26) to almost sure results. From the sequence $\{m\}$ of natural numbers we can by a diagonal argument extract a further subsequence $\{m_k\}$ such that, as $k \rightarrow \infty$,

$$(4.28) \quad \frac{1}{m_k} \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} \delta_{m_kij}^2 \mathbf{1}_{\{\delta_{m_kij}^2 > 2^{-k}\}} \rightarrow 0 \quad \text{a.s.}$$

Since the sum in the left-hand side of (4.26) is decreasing in τ , it follows from (4.28) that along the sequence $\{m_k\}$, for all $\tau > 0$,

$$\frac{1}{m_k} \sum_{i=1}^{m_k} \sum_{j=1}^{m_k} \delta_{m_kij}^2 \mathbf{1}_{\{\delta_{m_kij}^2 > \tau\}} \rightarrow 0 \quad \text{a.s.}$$

We can also assume that (4.25) holds almost surely along the subsequence $\{m_k\}$. Conditioning on the values $B_{m_kj} = b_{m_kj}$, we may now apply Theorem 4.1 of Hájek (1961) to show that the linear rank statistic

$$\rho_{m_k} \equiv \sum_{j=1}^{m_k} a_{m_kj} b_{m_k R_{m_k}(j)}$$

is asymptotically normally distributed with mean and variance

$$\begin{aligned}
 E\rho_{m_k} &= m_k \bar{a}_{m_k} \bar{b}_{m_k}, \\
 \text{Var } \rho_{m_k} &= \frac{1}{m_k} \sum_{j=1}^{m_k} (b_{m_kj} - \bar{b}_{m_k})^2 \sum_{j=1}^{m_k} (a_{m_kj} - \bar{a}_{m_k})^2.
 \end{aligned}$$

Since along the chosen subsequence $m_k^{-1} \sum_{j=1}^{m_k} (b_{m_kj} - \bar{b}_{m_k})^2 \rightarrow \alpha^2$, we notice by (4.21) that $\text{Var } m_k^{-1/2} \rho_{m_k} \rightarrow \alpha^2 \sigma^2$. Hence, conditionally on the array $\{B_{mj}\}$,

$$(4.29) \quad \frac{1}{\sqrt{m_k}} (\rho_{m_k} - m_k \bar{a}_{m_k} \bar{b}_{m_k}) \Rightarrow N(0, \alpha^2 \sigma^2) \quad \text{a.s.}$$

Finally, since any subsequence of $\{m\}$ has a further subsequence with the property (4.29), (4.29) holds in probability, and also unconditionally. Since by

exchangeability $(B_{mR_m(1)}, \dots, B_{mR_m(m)}) = (B_{m1}, \dots, B_{mm})$ in distribution, the conclusion of the lemma follows. \square

The last lemma contains two auxiliary results for arrays of exchangeable random variables.

LEMMA 4.7. *Let \mathbf{W} be a triangular array of nonnegative, row-exchangeable random variables. Then*

(4.30) \mathbf{W} satisfies conditions A3 and A4

implies that

(4.31) *the sequence $\{W_{n1}\}$ is uniformly square-integrable; that is,*

$$\lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} EW_{n1}^2 1_{\{W_{n1} > t\}} = 0.$$

Furthermore, A3 and A4 also imply that

(4.32)
$$\frac{1}{\sqrt{n}} E\left(\max_{1 \leq i \leq n} W_{ni}\right) \rightarrow 0.$$

PROOF OF LEMMA 4.7. We prove that (4.31) implies (4.32) first. Let $\varepsilon > 0$ and choose n sufficiently large that $\limsup_{n \rightarrow \infty} t^2 \Pr(W_{n1} > t) \leq \limsup_{n \rightarrow \infty} EW_{n1}^2 1_{\{W_{n1} > t\}} \leq \varepsilon^2$ for $t \geq \varepsilon\sqrt{n}$. Then

$$\begin{aligned} E\left(\max_{i \leq n} W_{ni}\right) &= \int_0^{\varepsilon\sqrt{n}} \Pr\left(\max_{i \leq n} W_{ni} > t\right) dt + \int_{\varepsilon\sqrt{n}}^{\infty} \Pr\left(\max_{i \leq n} W_{ni} > t\right) dt \\ &\leq \varepsilon\sqrt{n} + n \int_{\varepsilon\sqrt{n}}^{\infty} t^2 \Pr(W_{n1} > t) t^{-2} dt \\ &\leq \varepsilon\sqrt{n} + n\varepsilon^2 \frac{1}{\varepsilon\sqrt{n}} = 2\varepsilon\sqrt{n}. \end{aligned}$$

To prove that (4.30) implies (4.31), choose λ large enough that $t^2 \Pr(W_{n1} \geq t) \leq \varepsilon^2$ for all n when $t \geq \lambda$. Then

$$\begin{aligned} EW_{n1}^2 1_{\{W_{n1} \geq \lambda\}} &= \int_0^{\infty} 2t \Pr(W_{n1} \geq t \vee \lambda) dt \\ &= \lambda^2 \Pr(W_{n1} \geq \lambda) + 2 \int_{\lambda}^{\infty} t \Pr(W_{n1} \geq \lambda) dt \\ &\leq \varepsilon^2 + 2 \sup_{t \geq \lambda} t \sqrt{\Pr(W_{n1} \geq t)} \int_{\lambda}^{\infty} \sqrt{\Pr(W_{n1} \geq t)} dt \\ &\leq \varepsilon^2 + 2\varepsilon \|W_{n1}\|_{2,1}, \end{aligned}$$

and by virtue of A3 this completes the proof. \square

5. Proofs for Sections 2 and 3.

PROOF OF THEOREM 2.1. To prove Theorem 2.1, it suffices by Pollard [(1990), Theorem 10.3] to show that the following two conditions are satisfied.

C1 $(\hat{X}_n^\omega(f_1), \dots, \hat{X}_n^\omega(f_l)) \Rightarrow c \cdot (G_P(f_1), \dots, G_P(f_l))$, for all $f_1, \dots, f_l \in \mathcal{F}$,
 a.s.

C2 $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} E \|\hat{X}_n^\omega\|_{\mathcal{F}'(\rho_P, \delta)} = 0$ a.s.

PROOF OF C1 (Finite-dimensional convergence). By the Cramér–Wold device and linearity of \hat{X}_n^ω it suffices to show that

(5.33)
$$\hat{X}_n^\omega(\phi) \Rightarrow c \cdot G_P(\phi),$$

for all ϕ of the form

(5.34)
$$\phi \equiv \sum_{p=1}^l c_p f_p, \quad c_1, \dots, c_l \in \mathbb{R}, f_1, \dots, f_l \in \mathcal{F}, \text{ a.s.}$$

To be precise, we must find a set Ω_0 of full measure such that for any $\omega \in \Omega_0$ the convergence in (5.33) holds for every function ϕ of the form (5.34). Under the hypothesis (2.4) the class of products of functions in \mathcal{F} is a Glivenko–Cantelli class for P , that is,

$$\left(\sup_{f, g \in \mathcal{F}} |\mathbb{P}_n(fg) - P(fg)| \right)^* \rightarrow 0, \quad \text{a.s.}$$

(see, e.g., Giné and Zinn [(1990), equation (2.17)]). Hence

(5.35)
$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n (\phi(X_j^\omega) - \mathbb{P}_n^\omega(\phi))^2 &= \mathbb{P}_n^\omega(\phi)^2 - (\mathbb{P}_n^\omega \phi)^2 \\ &\rightarrow P\phi^2 - (P\phi)^2 = \text{Var } G_P(\phi), \end{aligned}$$

for all ϕ of the form (5.34), a.s. Since further

$$n^{-1} \max_{j \leq n} (\phi(X_j^\omega) - \mathbb{P}_n^\omega(\phi))^2 \leq 4n^{-1} \max_{j \leq n} \phi(X_j^\omega)^2,$$

it also holds that

(5.36)
$$\frac{1}{n} \max_{j \leq n} (\phi(X_j^\omega) - \mathbb{P}_n^\omega(\phi))^2 \rightarrow 0$$

for all ϕ of the form (5.34), for almost all ω .

It follows from (5.35), (5.36), assumptions A4 and A5 and Lemma 4.7 that the conditions of Lemma 4.6 are satisfied with the sequence $\{m\}$ equal to $\{n\}$; $\alpha_{m,j} \equiv \phi(X_j^\omega)$; $\bar{\alpha}_m \equiv \mathbb{P}_n^\omega(\phi)$; $B_{m,j} \equiv W_{n,j}$; $\bar{B}_m \equiv 1$; $\alpha^2 \equiv c^2$, and, by (5.35),

$$\sigma^2 \equiv P\phi^2 - (P\phi)^2.$$

Hence by Lemma 4.6,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \phi(X_j^\omega)(W_{n,j} - 1) \Rightarrow N(0, c^2(P\phi^2 - (P\phi)^2))$$

for all ϕ of the form (5.34) a.s., which we identify as the distribution of $c \cdot G_P(\phi)$. This completes the proof of C1. \square

PROOF OF C2 (Asymptotic equicontinuity). Let $\|\cdot\| \equiv \|\cdot\|_{\mathcal{F}'(\rho_P, \delta)}$ for some $\delta > 0$, and let \mathbf{W}' be an independent copy of \mathbf{W} . Then $\sum_{j=1}^n (W_{n,j} - W'_{n,j}) = 0$, and by Jensen's inequality,

$$\begin{aligned} E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - 1) \delta_{X_j^\omega} \right\| &\leq E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - W'_{n,j}) \delta_{X_j^\omega} \right\| \\ (5.37) \qquad \qquad \qquad &= E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - W'_{n,j}) (\delta_{X_j^\omega} - \mathbb{P}_n^\omega) \right\| \\ &\leq 2E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n W_{n,j} (\delta_{X_j^\omega} - \mathbb{P}_n^\omega) \right\|. \end{aligned}$$

Apply Lemma 4.1 to (5.37) for fixed n with $\xi_j \equiv W_{n,j}$, and $\mathbf{Z}_j \equiv \delta_{X_j^\omega} - \mathbb{P}_n^\omega$ (which may be regarded here as deterministic since ω is fixed) to get the following bound, which holds for any n and $n_0 < n$:

$$\begin{aligned} E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - 1) \delta_{X_j^\omega} \right\| \\ (5.38) \qquad \qquad \leq 2n_0 n^{-1/2} E \left(\max_{j \leq n} W_{n,j} \right) n^{-1} \sum_{j=1}^n \|\delta_{X_j^\omega} - \mathbb{P}_n^\omega\| \\ \qquad \qquad \qquad + 2\|W_{n,1}\|_{2,1} \max_{n_0 < k \leq n} E \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k (\delta_{X_{R_n(j)}^\omega} - \mathbb{P}_n^\omega) \right\|. \end{aligned}$$

Since $\|W_{n,1}\|_{2,1} \leq M(\mathbf{W}) < \infty$, and $n^{-1} \sum_{j=1}^n \|\delta_{X_j^\omega} - \mathbb{P}_n^\omega\| \leq 4\mathbb{P}_n^\omega F$, we can bound (5.38) by

$$\begin{aligned} 8n_0 n^{-1/2} E \left(\max_{j \leq n} W_{n,j} \right) \mathbb{P}_n^\omega F \\ (5.39) \qquad \qquad \qquad + 2M(\mathbf{W}) \max_{n_0 < k \leq n} E \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k (\delta_{X_{R_n(j)}^\omega} - \mathbb{P}_n^\omega) \right\|. \end{aligned}$$

Let $\hat{X}_{n,1}, \dots, \hat{X}_{n,n}$ be iid from \mathbb{P}_n^ω . Then Corollary 4.1 (Hoeffding's finite-sampling inequality) and Lemma 4.2 (Poissonization) applied to the second term

on the right-hand side of (5.39) yield the bound

$$\begin{aligned}
 (5.40) \quad \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k (\delta_{X_{R_n(j)}^\omega} - \mathbb{P}_n^\omega) \right\| &\leq \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k (\delta_{\tilde{X}_{n_j}} - \mathbb{P}_n^\omega) \right\| \\
 &\leq 4 \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^n \tilde{N}_j \left(\frac{k-n_0}{2n} \right) \delta_{X_j^\omega} \right\|,
 \end{aligned}$$

where $\tilde{N}_j((k-n_0)/2n)$, $j=1, \dots, n$ are iid, symmetrized Poisson($(k-n_0)/2n$). By Jensen's inequality and closure of the Poisson family under convolution, we can increase the Poisson parameter in (5.40) to $k/2n$ (a matter of keeping notation simple), and the right-hand side of (5.40) is further bounded by

$$(5.41) \quad 4 \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^n \tilde{N}_j \left(\frac{k}{2n} \right) \delta_{X_j^\omega} \right\|.$$

Let $\varepsilon_1, \dots, \varepsilon_n$ be iid Rademacher random variables, independent of the symmetrized Poisson variables. Then it holds by symmetry that

$$\mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^n \tilde{N}_j \left(\frac{k}{2n} \right) \delta_{X_j^\omega} \right\| = \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^n \left| \tilde{N}_j \left(\frac{k}{2n} \right) \right| \varepsilon_j \delta_{X_j^\omega} \right\|.$$

Fix n and k and apply again Lemma 4.1, this time with $\xi_j \equiv |\tilde{N}_j(k/2n)|$ and $\mathbf{Z}_j \equiv \varepsilon_j \delta_{X_j^\omega}$, noting that $\|\varepsilon_j \delta_{X_j^\omega}\| \leq F(X_j^\omega)$ and using exchangeability of the ε_i , to show that for any $n_1 < n$,

$$\begin{aligned}
 (5.42) \quad \mathbf{E} \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^n |\tilde{N}_j \left(\frac{k}{2n} \right)| \varepsilon_j \delta_{X_j^\omega} \right\| &= \sqrt{\frac{n}{k}} \mathbf{E} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n |\tilde{N}_j \left(\frac{k}{2n} \right)| \varepsilon_j \delta_{X_j^\omega} \right\| \\
 &\leq \sqrt{\frac{n}{k}} \left(n_1 n^{-1/2} \mathbf{E} \left(\max_{j \leq n} \left| \tilde{N}_j \left(\frac{k}{2n} \right) \right| \right) \right) \mathbb{P}_n^\omega F \\
 &\quad + \left\| \tilde{N}_1 \left(\frac{k}{2n} \right) \right\|_{2,1} \max_{n_1 < j \leq n} \mathbf{E} \left\| \frac{1}{\sqrt{j}} \sum_{l=n_1+1}^j \varepsilon_l \delta_{X_{R_n(l)}^\omega} \right\|.
 \end{aligned}$$

By Lemma 4.5 we have the bound

$$\mathbf{E} \left(k^{-1/2} \max_{j \leq n} \left| \tilde{N}_j \left(\frac{k}{2n} \right) \right| \right) \leq \sqrt{2} \left(\frac{12}{n} + \frac{4}{k} \right)^{1/4} \leq 2\sqrt{2} n_0^{-1/4} \quad \text{when } n, k > n_0,$$

and by Lemma 4.4, $\sqrt{n/k} \|\tilde{N}_1(k/2n)\|_{2,1} \leq 4/\sqrt{2}$ for all n, k . By the triangle inequality,

$$\max_{n_1 < j \leq n} \mathbf{E} \left\| \frac{1}{\sqrt{j}} \sum_{l=n_1+1}^j \varepsilon_l \delta_{X_{R_n(l)}^\omega} \right\| \leq 2 \max_{n_1 \leq j \leq n} \mathbf{E} \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_n(l)}^\omega} \right\|,$$

and hence a simpler bound in (5.42) is, for any $n_0 \leq k \leq n, n_1 \leq n,$

$$(5.43) \quad 4\sqrt{2} n_0^{-1/4} n_1 \mathbb{P}_n^\omega F + \frac{8}{\sqrt{2}} \max_{n_1 \leq j \leq n} E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_n^\omega(l)}} \right\|.$$

When we combine (5.37)–(5.43), we get the following bound for the expectation of the norm of the weighted bootstrap empirical process, which holds for any n and $n_0, n_1 < n$:

$$(5.44) \quad E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega} \right\| \leq 8n_0 n^{-1/2} E \left(\max_{j \leq n} W_{nj} \right) \mathbb{P}_n^\omega F + 8M(\mathbf{W}) \left(4\sqrt{2} n_0^{-1/4} n_1 \mathbb{P}_n^\omega F + \frac{8}{\sqrt{2}} \max_{n_1 \leq j \leq n} E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_n^\omega(l)}} \right\| \right).$$

Now we apply Lemma 4.3 to the last term in the right-hand side of (5.44). Define

$$Z_j = \left(E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_l^\omega} \right\| \right)^*$$

(since the norm may be nonmeasurable), and notice that

$$(5.45) \quad \begin{aligned} \left(E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_n^\omega(l)}} \right\| \right)^* &= \left(\frac{1}{n!} \sum_{\pi \in \Pi_n} E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{\pi(l)}^\omega} \right\| \right)^* \\ &\leq \frac{1}{n!} \sum_{\pi \in \Pi_n} \left(E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{\pi(l)}^\omega} \right\| \right)^* \\ &= E \left(\left(E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_l^\omega} \right\| \right)^* \middle| \mathcal{S}_n \right) \\ &= E(Z_j | \mathcal{S}_n). \end{aligned}$$

By the inequality of Ledoux, Talagrand and Zinn [Giné and Zinn (1990), Lemma 2.3],

$$\limsup_{n \rightarrow \infty} Z_n \leq 4 \limsup_{n \rightarrow \infty} E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j \delta_{X_j} \right\| = 4E \|G_P\| < \infty \quad \text{a.s.,}$$

where the equality follows because $\mathcal{F} \in \text{CLT}(P)$ implies that

$$\left\| n^{-1/2} \sum_{j=1}^n \varepsilon_j \delta_{X_j} \right\|^p$$

is uniformly integrable when $p < 2$; this follows from Giné and Zinn [(1986),

Theorem 1.2.8] and Andersen [(1985), Proposition 3.7, page 449]. Furthermore, the proof of the Ledoux, Talagrand, and Zinn almost sure bound (see Ledoux and Talagrand [(1988), pages 40 and 41]) with their M replaced by

$$K \equiv \sup_{n \geq 1} E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j \delta_{X_j} \right\|,$$

and computing a bound in the truncation step, yields

$$\Pr^* \left(\sup_{n \geq 1} Z_n > 2(2K + 5\varepsilon) \right) \leq 2 \frac{\sqrt{2}}{\sqrt{2} - 1} PF^2 \left(\frac{4K}{\varepsilon^3} + \frac{1}{\varepsilon^2} \right)$$

for every $\varepsilon > 0$. Here the second term comes from the truncation step in their proof. This implies that $E(\sup_{n \geq 1} Z_n) < \infty$.

Hence by Lemma 4.3 and (5.45),

$$(5.46) \quad \limsup_{n_1 \wedge n \rightarrow \infty} \max_{n_1 \leq j \leq n} E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_{n(l)}}} \right\| \leq \limsup_{n_1 \wedge n \rightarrow \infty} \max_{n_1 \leq j \leq n} E(Z_j | \mathcal{I}_n) \\ \leq 4E\|G_P\| \quad \text{a.s.}$$

Returning to (5.44), we notice that by Lemma 4.7,

$$n^{-1/2} E \left(\max_{j \leq n} W_{n,j} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and $\mathbb{P}_n^\omega F \rightarrow PF$ a.s. follows by the SLLN. Define momentarily

$$a_n \equiv n^{-1/2} E \left(\max_{j \leq n} W_{n,j} \right),$$

then $a_n \rightarrow 0$, and if we then choose the sequences $n_0 = n_0(n) \equiv [a_n]^{-1/2} \wedge [n/2]$ and $n_1 = n_1(n) \equiv [n_0(n)^{1/5}]$, we have that $n_0(n) \rightarrow \infty$, $n_1(n) \rightarrow \infty$, $n_0(n)^{-1/4} n_1(n) \rightarrow 0$, and

$$n_0(n) n^{-1/2} E \left(\max_{j \leq n} W_{n,j} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

By considering (5.44), we now see that

$$\limsup_{n \rightarrow \infty} E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - 1) \delta_{X_j^\omega} \right\| \\ \leq \frac{64}{\sqrt{2}} M(\mathbf{W}) \limsup_{n \rightarrow \infty} \max_{n_1(n) \leq j \leq n} E \left\| \frac{1}{\sqrt{j}} \sum_{l=1}^j \varepsilon_l \delta_{X_{R_{n(l)}}} \right\| \\ \leq \frac{256}{\sqrt{2}} M(\mathbf{W}) E\|G_P\| \quad \text{a.s.}$$

Recall that $\|\cdot\| \equiv \|\cdot\|_{\mathcal{G}'(\rho_P, \delta)}$ for $\delta > 0$ to obtain by ρ_P -continuity of G_P

sample paths that $\lim_{\delta \rightarrow 0} E \|G_P\|_{\mathcal{F}'(\rho_P, \delta)} = 0$, and hence, finally,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{nj} - 1) \delta_{X_j^\omega} \right\|_{\mathcal{F}'(\rho_P, \delta)} = 0 \quad \text{a.s.}$$

This proves C2 and hence also Theorem 2.1. \square

PROOF OF COROLLARY 2.1. We shall only prove almost sure asymptotic equicontinuity since the finite-dimensional convergence part follows by standard methods. But the proof of Theorem 2.1 actually shows that

$$\lim_{\delta \rightarrow 0} \limsup_{n_0 \wedge n \rightarrow \infty} \max_{n_0 < m \leq n} E \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^n \tilde{N}_j \left(\frac{m}{2n} \right) \delta_{X_j^\omega} \right\|_{\mathcal{F}'(\rho_P, \delta)} = 0 \quad \text{a.s.}$$

[this follows by considering (5.42), (5.43), (5.44) and (5.46)]. By Lemma 4.2,

$$E \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^m (\delta_{\hat{X}_{nj}} - \mathbb{P}_n^\omega) \right\| \leq 4 E \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^n \tilde{N}_j \left(\frac{m}{2n} \right) \delta_{X_j^\omega} \right\|$$

and it follows that

$$\lim_{\delta \rightarrow 0} \limsup_{m \wedge n \rightarrow \infty} E \left\| \frac{1}{\sqrt{m}} \sum_{j=1}^m (\delta_{\hat{X}_{nj}} - \mathbb{P}_n^\omega) \right\|_{\mathcal{F}'(\rho_P, \delta)} = 0 \quad \text{a.s.}$$

by using Lemmas 4.1, 4.4, 4.5 and 4.3, just as in the proof of Theorem 2.1. \square

PROOF OF THEOREM 2.2. When $\mathcal{F} \in \text{CLT}(P)$, \mathcal{F} is in particular P -pre-Gaussian and hence totally bounded in the ρ_P -seminorm (see, e.g., Pollard [(1984), Problem VII.3]). Given $\delta > 0$, let $\{f_1, \dots, f_{k(\delta)}\}$ be a δ -net in \mathcal{F} , and, for $Z \in l^\infty(\mathcal{F})$, let Z^δ be defined by $Z^\delta(f) = Z(f_i)$ if and only if $\rho_P(f, f_i) < \delta$ and, $\rho_P(f, f_j) \geq \delta, j \neq i$. (This terminology is in accordance with Giné and Zinn [(1986), page 59].) In this way we can bound the dual bounded Lipschitz distance by

$$\begin{aligned} d_{BL^*}(\hat{X}_n(\omega), cG_P) &\leq \sup_{H \in BL_1} |EH(\hat{X}_n(\omega)) - EH(\hat{X}_n(\omega)^\delta)| \\ &\quad + \sup_{H \in BL_1} |EH(\hat{X}_n(\omega)^\delta) - EH(cG_P^\delta)| \\ &\quad + \sup_{H \in BL_1} |EH(cG_P^\delta) - EH(cG_P)| \\ &\equiv \text{I}(n, \delta, \omega) + \text{II}(n, \delta, \omega) + \text{III}(\delta). \end{aligned}$$

Hence, in order to show (5.33), it suffices to show:

D1. Any sequence $\{n\}$ has a subsequence $\{n_k\}$ along which

$$(\hat{X}_{n_k}^\omega(f_1), \dots, \hat{X}_{n_k}^\omega(f_l)) \Rightarrow (G_P(f_1), \dots, G_P(f_l)), \quad \text{for all } f_1, \dots, f_l \in \mathcal{F} \text{ a.s.}$$

D2. $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr^*(E \|\hat{X}_n\|_{\mathcal{F}'(\rho_P, \delta)} > \eta) = 0$, for all $\eta > 0$.

When condition D1 is satisfied, it follows that

$$\text{II}(n, \delta, \omega) \rightarrow 0 \text{ in outer probability, for all } \delta > 0,$$

and D2 implies that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr^*(\text{I}(n, \delta) > \eta) = 0 \text{ for all } \delta > 0.$$

Finally, $\text{III}(\delta) \rightarrow 0, \delta \rightarrow 0$ follows by ρ_P -continuity of G_P paths.

The proof of D1 only differs slightly from that of C1: By the Cramér–Wold device and linearity of \hat{X}_n^ω , it suffices to show that every sequence has a subsequence $\{n_k\}$ such that

$$(5.47) \quad \hat{X}_{n_k}^\omega(\phi) \Rightarrow cG_P(\phi),$$

for all ϕ of the form (5.34), almost surely. [Cf. this with (5.33) and (5.34).] To be precise, we must find a set Ω_0 of full measure such that for any $\omega \in \Omega_0$ the convergence in (5.33) holds for every function ϕ of the form (5.34). When $\mathcal{F} \in \text{CLT}(P)$, the class of products of functions from \mathcal{F} is a weak Glivenko–Cantelli class for P , that is,

$$(5.48) \quad \sup_{f, g \in \mathcal{F}} |\mathbb{P}_n(fg) - P(fg)| \rightarrow 0 \text{ in outer probability.}$$

Hence any sequence $\{n\}$ has a subsequence $\{n_k\}$ along which

$$(5.49) \quad \begin{aligned} \frac{1}{n_k} \sum_{j=1}^{n_k} (\phi(X_j^\omega) - P_{n_k}^\omega(\phi))^2 &= P_{n_k}^\omega(\phi)^2 - (P_{n_k}^\omega \phi)^2 \\ &\rightarrow P\phi^2 - (P\phi)^2 = \text{Var } G_P(\phi), \end{aligned}$$

for all ϕ of the form (5.34), a.s. Since further

$$n_k^{-1} \max_{j \leq n_k} (\phi(X_j^\omega) - P_{n_k}^\omega(\phi))^2 \leq 4n_k^{-1} \max_{j \leq n_k} \phi(X_j^\omega)^2,$$

it also holds that

$$(5.50) \quad \frac{1}{n_k} \max_{j \leq n_k} (\phi(X_j^\omega) - P_{n_k}^\omega(\phi))^2 \rightarrow 0$$

for all ϕ of the form (5.34), a.s. As for C1, it follows from (5.49), (5.50), assumptions A4 and A5 that the conditions of Lemma 4.6 are satisfied with the sequence $\{m\}$ equal to $\{n_k\}$; $a_{mj} \equiv \phi(X_j^\omega)$; $\bar{a}_m \equiv P_{n_k}^\omega(\phi)$; $B_{mj} \equiv W_{n_k j}$; $\bar{B}_m \equiv 1$; $\alpha^2 \equiv c^2$, and, by (5.49),

$$\sigma^2 \equiv P\phi^2 - (P\phi)^2.$$

Hence

$$\frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \phi(X_j^\omega)(W_{n_k j} - 1) \Rightarrow N(0, c^2(P\phi^2 - (P\phi)^2)),$$

which we identify as the distribution of $c \cdot G_P(\phi)$. Since any sequence has a subsequence with this property, (5.47) and in turn D1 follows.

For D2, it suffices by Markov's inequality to show that

$$(5.51) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} E^* \|\hat{X}_n\|_{\mathcal{F}'(\rho_P, \delta)} = 0.$$

Let $\|\cdot\| \equiv \|\cdot\|_{\mathcal{F}'(\rho_P, \delta)}$ for a $\delta > 0$, and let (X'_1, \dots, X'_n) be an iid copy of (X_1, \dots, X_n) . Then

$$(5.52) \quad \begin{aligned} E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n (W_{n,j} - 1) \delta_{X_j} \right\| &\leq E^* \left\| \frac{1}{\sqrt{n}} \tilde{W}_{n,j} (\delta_{X_j} - \delta_{X'_j}) \right\| \\ &= E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n |\tilde{W}_{n,j}| (\delta_{X_j} - \delta_{X'_j}) \right\|, \end{aligned}$$

where $\tilde{W}_{n,j}$ are, as usual, symmetrized weights. Apply Lemma 4.1 with $\xi = (|\tilde{W}_{n,1}|, \dots, |\tilde{W}_{n,n}|)$ and $(\mathbf{Z}_1, \dots, \mathbf{Z}_n) = (\delta_{X_1} - \delta_{X'_1}, \dots, \delta_{X_n} - \delta_{X'_n})$ to bound (5.52) by

$$(5.53) \quad \begin{aligned} &n_0 n^{-1/2} E \left(\max_{j \leq n} |\tilde{W}_{n,j}| \right) \frac{1}{n} E^* \sum_{j=1}^n \|\delta_{X_j} - \delta_{X'_j}\| \\ &+ \|\tilde{W}_{n,j}\|_{2,1} \max_{n_0 < k \leq n} E^* \left\| \frac{1}{\sqrt{k}} \sum_{j=n_0+1}^k (\delta_{X_{R_n(j)}} - \delta_{X'_{R_n(j)}}) \right\|. \end{aligned}$$

Now $E^* \|\delta_{X_j} - \delta_{X'_j}\| \leq 2E^* \|\delta_{X_j} - P\| \leq 4E^* \|\delta_{X_j} - P\|_{\mathcal{F}} < \infty$, when $\mathcal{F} \in \text{CLT}(P)$ (the last inequality follows from, e.g., the argument in Pisier [(1975), page III.10]), and we can bound (5.53) by

$$\begin{aligned} &4n_0 n^{-1/2} E \left(\max_{j \leq n} |\tilde{W}_{n,j}| \right) E^* \|\delta_{X_1} - P\|_{\mathcal{F}} \\ &+ 8M(\mathbf{W}) \max_{n_0 \leq k \leq n} E^* \left\| \frac{1}{\sqrt{k}} \sum_{j=1}^k \varepsilon_j \delta_{X_j} \right\|_{\mathcal{F}'(\rho_P, \delta)}. \end{aligned}$$

By Theorem 1.2.8 of Giné and Zinn (1986), it follows that when $\mathcal{F} \in \text{CLT}(P)$,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} E^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j \delta_{X_j} \right\|_{\mathcal{F}'(\rho_P, \delta)} = 0,$$

and (5.51) and D2 follow. \square

The rest of this section contains calculations related to the examples of bootstrap weights given in Section 3.

EXAMPLE 3.1 (The iid-weighted bootstraps). It follows by the law of large numbers that A5 is satisfied with $c^2 = \text{Var } Y_1 / (EY_1)^2$. For A3 and A4 we need the following bounds for the tail probabilities of the weights. Without loss of

generality we can assume that $EY_1 = 1$ in the rest of the argument. Let $0 < \varepsilon < 1$. Then

$$(5.54) \quad \Pr(Y_1/\bar{Y}_n > t) \leq \Pr(Y_1 > t(1 - \varepsilon)) + \Pr(Y_1/\bar{Y}_n > t, \bar{Y}_n < 1 - \varepsilon).$$

Furthermore, for any $p > 0$,

$$(5.55) \quad \begin{aligned} \Pr(Y_1/\bar{Y}_n > t, \bar{Y}_n < 1 - \varepsilon) &\leq \sqrt{\Pr(Y_1/\bar{Y}_n > t) \Pr(\bar{Y}_n < (1 - \varepsilon))} \\ &\leq t^{-p/2} E(Y_1/\bar{Y}_n)^{p/2} \sqrt{\Pr(\bar{Y}_n < (1 - \varepsilon))} \\ &\leq t^{-p/2} n^{p/2} \sqrt{\Pr(\bar{Y}_n < (1 - \varepsilon))}, \end{aligned}$$

where we have used the bound $Y_1/\bar{Y}_n < n$. To bound the left-hand side of (5.55) we finally apply the following large deviations result; see, for example, Shorack and Wellner [(1986), equation (A.4.17)]. Define the Laplace transform of $1 - Y_1$ by

$$\phi(s) \equiv Ee^{s(1-Y_1)};$$

since $1 - Y_1 \leq 1$ a.s., this is finite for all $s \in \mathbb{R}$. Moreover, $\phi'(0) = 0$ since $E(1 - Y_1) = 0$, and hence

$$\frac{d}{ds} [e^{-se}\phi(s)]_{s=0} = -\varepsilon < 0,$$

for any $\varepsilon > 0$. This shows that $\rho(\varepsilon) \equiv \inf_{s>0} e^{-se}\phi(s) < 1$. Then

$$(5.56) \quad \begin{aligned} \Pr(1 - \bar{Y}_n > \varepsilon) &\leq \inf_{r>0} \frac{Ee^{r(1-\bar{Y}_n)}}{e^{r\varepsilon}} \\ &= \inf_{r>0} e^{-r\varepsilon} \phi\left(\frac{r}{n}\right)^n \\ &= \inf_{s>0} (e^{-se}\phi(s))^n \equiv \rho(\varepsilon)^n. \end{aligned}$$

Combining (5.54), (5.55) and (5.56), we get the following bound for the tail probabilities of the weights:

$$(5.57) \quad \Pr(Y_1/\bar{Y}_n > t) \leq \Pr(Y_1 > t(1 - \varepsilon)) + t^{-p/2} n^{p/2} \rho(\varepsilon)^{n/2}.$$

By choosing $p > 4$ this shows that

$$\left\| \frac{Y_1}{\bar{Y}_n} \right\|_{2,1} \leq \frac{1}{1 - \varepsilon} \|Y_1\|_{2,1} + \int_0^\infty t^{-p/4} dt n^{p/4} \rho(\varepsilon)^{n/4},$$

and A3 is satisfied. To see that A4 holds, by virtue of (5.57) we need only check that $\|Y_1\|_{2,1} < \infty$ implies $\lim_{t \rightarrow \infty} t^2 \Pr(Y_1 > t) = 0$. This holds since $L_{2,1} \subset L_2$ and L_2 implies weak L_2 by Markov's inequality. (Recall that $(1/2)\|Y\|_2 \leq \|Y\|_{2,1}$.)

We now turn to the remaining examples 3.2–3.5.

Let $M \sim \text{Mult}_n(n, (p_1, \dots, p_n))$. The following moments can, except for the mixed terms, be found in Johnson and Kotz [(1969), page 51]. We use the notation $n^{(k)} \equiv n(n-1) \cdots (n-k+1)$.

$$\begin{aligned} EM_1 &= np_1, \\ EM_1^2 &= np_1 + n^{(2)}p_1^2, \\ EM_1M_2 &= n(n-1)p_1p_2, \\ \text{Cov}(M_1, M_2) &= -np_1p_2, \\ EM_1^4 &= np_1 + 7n^{(2)}p_1^2 + 6n^{(3)}p_1^3 + n^{(4)}p_1^4, \\ EM_1^2M_2^2 &= n^{(2)}p_1p_2 + n^{(3)}(p_1^2p_2 + p_1p_2^2) + n^{(4)}p_1^2p_2^2, \\ \text{Cov}(M_1^2, M_2^2) &= (n^{(4)} - (n^{(2)})^2)p_1^2p_2^2 + (n^{(3)} - nn^{(2)}) \\ &\quad \times (p_1^2p_2 + p_1p_2^2) + (n^{(2)} - n^2)p_1p_2, \\ EM_1M_2^2 &= n^{(2)}p_1p_2 + n^{(3)}p_1p_2^2, \\ \text{Cov}(M_1, M_2^2) &= (n^{(3)} - nn^{(2)})p_1p_2^2 + (n^{(2)} - n^2)p_1p_2. \end{aligned}$$

Notice that all the covariances above are negative. Also, if in particular $p_1 = \cdots = p_n = n^{-1}$, it follows that $EM_{n1}^2 = 1 + (n-1)/n \rightarrow 2$, and that $EM_1^4 < 15$ (a crude bound).

EXAMPLE 3.2 (Efron's bootstrap). The weights

$$\underline{M}_n \sim \text{Mult}_n(n, (n^{-1}, \dots, n^{-1}))$$

satisfy B1–B3: $EM_{n1}^2 = 1 + (n-1)/n \rightarrow 2$; $EM_{n1}^4 < 15$, and

$$\text{Cov}(M_{n1}^2, M_{n2}^2) < 0$$

can be checked in the list above.

EXAMPLE 3.3 (The double bootstrap). By the definition of the double bootstrap weights,

$$\begin{aligned} EW_{n1}^2 &= EE(W_{n1}^2 | \underline{M}_n) \\ &= E\left(M_n + \frac{n^{(2)}}{n^2} M_n^2\right) \\ &= 1 + \frac{n^{(2)}}{n^2} \left(1 + \frac{n^{(2)}}{n^2}\right) \\ &\rightarrow 3 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

In the same way,

$$\begin{aligned} EW_{n1}^4 &= EE(W_{n1}^4 | \underline{M}_n) \\ &= E\left(M_{n1} + 7 \frac{n^{(2)}}{n^2} M_{n1}^2 + 6 \frac{n^{(3)}}{n^3} M_{n1}^3 + \frac{n^{(4)}}{n^4} M_{n1}^4\right), \end{aligned}$$

and in order to show that $\sup_n EW_{n1}^4 < \infty$, it suffices to know that $EM_{n1}^4 < \text{Const} < \infty$. As remarked above, 15 is such a (crude) constant.

Consider finally

$$\text{Cov}(W_{n1}^2, W_{n2}^2) = E \text{Cov}(W_{n1}^2, W_{n2}^2 | \underline{M}_n) + \text{Cov}(E(W_{n1}^2 | \underline{M}_n), E(W_{n2}^2 | \underline{M}_n)).$$

We calculate

$$\begin{aligned} E \text{Cov}(W_{n1}^2, W_{n2}^2 | \underline{M}_n) &= E\left(\frac{n^{(4)} - (n^{(2)})^2}{n^4} M_{n1}^2 M_{n2}^2 - \frac{n^{(3)} - nn^{(2)}}{n^3} (M_{n1}^2 M_{n2} + M_{n1} M_{n2}^2) \right. \\ &\quad \left. - \frac{n^{(2)} - n^2}{n^2} M_{n1} M_{n2}\right) \\ &\leq 0, \end{aligned}$$

since all the coefficients are negative. Furthermore,

$$\begin{aligned} \text{Cov}(E(W_{n1}^2 | \underline{M}_n), E(W_{n2}^2 | \underline{M}_n)) &= \text{Cov}\left(M_{n1} + \frac{n^{(2)}}{n^2} M_{n1}^2, M_{n2} + \frac{n^{(2)}}{n^2} M_{n2}^2\right) \\ &= \text{Cov}(M_{n1}, M_{n2}) + \left(\frac{n^{(2)}}{n^2}\right)^2 \text{Cov}(M_{n1}^2, M_{n2}^2) \\ &\quad + \frac{n^{(2)}}{n^2} \text{Cov}(M_{n1}^2, M_{n2}) + \frac{n^{(2)}}{n^2} \text{Cov}(M_{n1}, M_{n2}^2) \\ &< 0, \end{aligned}$$

since the covariances are all negative. From this we can conclude that W_{n1}^2 and W_{n2}^2 are nonpositively correlated. Hence the double bootstrap weights satisfy D1-D3 with $c^2 = 2$.

EXAMPLE 3.4 (The multivariate hypergeometric bootstrap). The factorial moments of these bootstrap weights are given by [see, e.g., Johnson and Kotz (1977), equation 2.55, page 92].

$$(5.58) \quad \mu_{(r_1, \dots, r_n)} \equiv E(W_{n1}^{(r_1)} \cdots W_{nn}^{(r_n)}) = \frac{n^{(r)} K^{(r_1)} \cdots K^{(r_n)}}{(nK)^{(r)}}$$

where $r = \sum_{j=1}^n r_j$. We can express the absolute moments of interest in terms

of factorial moments as follows [see, e.g., Stuart and Ord (1983), page 82]:

$$(5.59) \quad EW_{n1}^2 = \mu_{(2,0,\dots,0)} + \mu_{(1,0,\dots,0)},$$

$$(5.60) \quad EW_{n1}^4 = \mu_{(4,0,\dots,0)} + 6\mu_{(3,0,\dots,0)} + 7\mu_{(2,0,\dots,0)} + \mu_{(1,0,\dots,0)}$$

and

$$(5.61) \quad EW_{n1}^2 W_{n2}^2 = \mu_{(2,2,0,\dots,0)} + 2\mu_{(2,1,0,\dots,0)} + \mu_{(1,1,0,\dots,0)},$$

from which it follows that

$$EW_{n1}^2 = 1 + \frac{(n-1)(K-1)}{nK-1} \rightarrow 1 + \frac{K-1}{K} \text{ as } n \rightarrow \infty.$$

We also have the crude bound $\mu_{(r_1,0,\dots,0)} < K^{(r_1)}$, and hence by (5.60), $\sup_n EW_{n1}^4 \leq K^{(4)} + 6K^{(3)} + 7K^{(2)} + K < \infty$, for all K . Finally, (5.59) and (5.61) show that

$$\begin{aligned} \text{Cov}(W_{n1}^2, W_{n2}^2) &= \mu_{(2,2,0,\dots,0)} + 2\mu_{(2,1,0,\dots,0)} + \mu_{(1,1,0,\dots,0)}(\mu_{(2,0,\dots,0)} + \mu_{(1,0,\dots,0)})^2 \\ &= (\mu_{(2,2,0,\dots,0)} - \mu_{(2,0,\dots,0)}^2) + 2(\mu_{(2,1,0,\dots,0)} - \mu_{(2,0,\dots,0)}\mu_{(1,0,\dots,0)}) \\ &\quad + (\mu_{(1,1,0,\dots,0)} - \mu_{(1,0,\dots,0)}^2). \end{aligned}$$

By (5.58), it will follow that $\text{Cov}(W_{n1}^2, W_{n2}^2) < 0$ if we can show that $\mu_{(r+s)} \leq \mu_{(r)}\mu_{(s)}$, that is,

$$(5.62) \quad \frac{n^{(r+s)}K^{(r_1+s_1)} \dots K^{(r_n+s_n)}}{(nK)^{(r+s)}} \leq \frac{n^{(r)}K^{(r_1)} \dots K^{(r_n)}}{(nK)^{(r)}} \frac{n^{(s)}K^{(s_1)} \dots K^{(s_n)}}{(nK)^{(s)}}$$

for any sets of integers (r_1, \dots, r_n) and (s_1, \dots, s_n) with sum r and s , respectively. But (5.62) is satisfied if in particular

$$\frac{n^{(r+s)}}{(nK)^{(r+s)}} \leq \frac{n^{(r)}}{(nK)^{(r)}} \frac{n^{(s)}}{(nK)^{(s)}},$$

which follows because for positive numbers $x < a < b$, $(a-x)/(b-x) < a/b$. Hence, B1–B3 are satisfied with $c^2 = (K-1)/K$.

EXAMPLE 3.5 (The Polya–Eggenberger bootstrap). By the calculations above of multinomial moments, it follows that for these weights

$$EW_{n1}^2 = EE(W_{n1}^2 | D_n) = E(nD_{n1} + n^{(2)}D_{n1}^2).$$

The Dirichlet distribution with parameter (α, \dots, α) has moments [see, e.g., Johnson and Kotz (1977), page 96]

$$(5.63) \quad E(D_{n1}^{p_1}, \dots, D_{nn}^{p_n}) = \frac{\alpha^{[p]}}{(n\alpha)^{[p]}} \equiv \frac{\alpha \cdots (\alpha + p - 1)}{n\alpha \cdots (n\alpha + p - 1)}$$

where $p = p_1 + \dots + p_n$. It follows that

$$EW_{n1}^2 = 1 + (\alpha + 1) \frac{n - 1}{n\alpha + 1} \rightarrow 1 + \frac{\alpha + 1}{\alpha}.$$

Furthermore,

$$\begin{aligned} EW_{n1}^4 &= EE(W_{n1}^4 | \underline{D}_n) \\ &= E(nD_{n1} + 7n^{(2)}D_{n1}^2 + 6n^{(3)}D_{n1}^3 + n^{(4)}D_{n1}^4), \end{aligned}$$

and $\sup_n EW_{n1}^4 < \infty$ will follow if we can show that

$$\sup_n n^{(p)}ED_{n1}^p < \infty \quad \text{for all } p = 1, 2, \dots$$

Now by (5.63),

$$\begin{aligned} n^{(p)}ED_{n1}^p &= n^{(p)} \frac{\alpha \cdots (\alpha + p - 1)}{n\alpha \cdots (n\alpha + p - 1)} \\ &= \prod_{k=0}^{p-1} \frac{(\alpha + k)(n - k)}{n\alpha + k} \\ &\rightarrow \prod_{k=0}^{p-1} \frac{\alpha + k}{\alpha}, \quad n \rightarrow \infty. \end{aligned}$$

Finally, we must show that $\text{Cov}(W_{n1}^2, W_{n2}^2) < 0$. We write

$$\text{Cov}(W_{n1}^2, W_{n2}^2) = E \text{Cov}(W_{n1}^2, W_{n2}^2 | \underline{D}_n) + \text{Cov}(E(W_{n1}^2 | \underline{D}_n), E(W_{n2}^2 | \underline{D}_n))$$

and notice that since (W_{n1}, \dots, W_{nn}) is multinomially distributed conditional on \underline{D}_n , it can be seen in the list of multinomial moments above that

$$\text{Cov}(W_{n1}^2, W_{n2}^2 | \underline{D}_n) < 0 \quad \text{a.s.}$$

Hence, we need only show that

$$\text{Cov}(E(W_{n1}^2 | \underline{D}_n), E(W_{n2}^2 | \underline{D}_n)) < \infty.$$

We note that

$$\text{Cov}(E(W_{n1}^2 | \underline{D}_n), E(W_{n2}^2 | \underline{D}_n)) = \text{Cov}(nD_{n1} + n^{(2)}D_{n1}^2, nD_{n2} + n^{(2)}D_{n2}^2),$$

and by (5.63) one finds that

$$ED_{n1}^p D_{n2}^q = \frac{\alpha^{[p]}\alpha^{[q]}}{(n\alpha)^{[p+q]}} \leq \frac{\alpha^{[p]}}{(n\alpha)^{[p]}} \frac{\alpha^{[q]}}{(n\alpha)^{[q]}} = ED_1^p ED_2^q.$$

Hence, the Polya–Eggenberger bootstrap satisfies B1–B3 with $c^2 = (\alpha + 1)/\alpha$.

✧ **Acknowledgments.** We owe thanks to Evarist Giné, Joel Zinn and Miguel Arcones for helpful conversations concerning bootstrap limit theorems and especially for help with Lemmas 4.4 and 4.5. Chris Klaassen provided the nice example in Remark 3.1. Our thanks also go to Aad van der Vaart for pointing out a serious error in our first formulation of Lemma 4.3.

REFERENCES

- ALDOUS, D. J. (1983). Exchangeability and related topics. *École d'Été de Probabilités de Saint-Flour XIII. Lecture Notes in Math.* Springer, New York.
- ARCONES, M. and GINÉ, E. (1992). On the bootstrap of M estimators and other statistical functionals. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, ed.) 14–47. Wiley, New York.
- ANDERSEN, N. T. (1985). The central limit theorem for non-separable valued functions. *Z. Wahrsch. Verw. Gebiete* **70** 445–455.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
- DELLACHERIE, C. and MEYER, P. A. (1980). *Probabilités et Potentiel. Theorie des Martingales* Chapitres V–VIII. Hermann, Paris.
- DUDLEY, R. M. (1984). A course on empirical processes. *Lecture Notes in Math.* **1097** 2–142. Springer, New York.
- DUDLEY, R. M. (1985). An extended Wichura theorem, definitions of Donsker classes, and weighted empirical distributions. *Lecture Notes in Math.* **1143** 141–178. Springer, New York.
- DUDLEY, R. M. (1989). *Real Analysis and Probability.* Wadsworth and Brooks/Cole, Monterey.
- DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. *Probability in Banach Spaces, Progress in Probability* **21** 7 63–82. Birkhäuser, Boston.
- EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.
- EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.
- GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.
- GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes in Math.* **1221** 50–113. Springer, New York.
- GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical functions. *Ann. Probab.* **18** 851–869.
- HÁJEK, J. (1961). Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann. Math. Statist.* **32** 506–523.
- HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
- HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501.
- HOFFMANN-JØRGENSEN, J. (1984). *Stochastic Process on Polish Spaces.* Unpublished manuscript.
- JOHNSON, N. and KOTZ, S. (1969). *Distributions in Statistics. Discrete Distributions.* Houghton Mifflin, Boston.
- JOHNSON, N. and KOTZ, S. (1977). *Urn Models and Their Application.* Wiley, New York.
- LEDOUX, M. and TALAGRAND, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probab. Theory Related Fields* **77** 29–47.
- LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, New York.
- LO, A. Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* **15** 360–375.
- MASON, D. and NEWTON, M. (1990). A rank statistics approach to the consistency of a general bootstrap. Technical Report 190, Dept. Statistics, Univ. Washington.
- MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.
- NEWTON, M. (1991). The weighted likelihood bootstrap and an algorithm for pre pivoting. Ph.D. dissertation, Dept. Statistics, Univ. Washington.
- PISIER, G. (1975). Le théorème de la limite centrale et la loi du logarithme itere dans les espaces de Banach. Séminaire Maurey–Schwartz 1975–1976. Exposé IV, Ecole Polytechnique Paris.
- POLLARD, D. (1984). *Convergence in Distribution of Stochastic Processes.* Springer, New York.
- POLLARD, D. (1990). *Empirical processes: Theory and Applications.* IMS, Hayward, CA.
- PRÆSTGAARD, J. T. (1990). Bootstrap with general weights and multiplier central limit theorems. Technical Report 195, Dept. Statistics, Univ. Washington.
- RUBIN, D. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134.

- SHI, X. (1991). Some asymptotic results for jackknifing the sample quantile. *Ann. Statist.* **19** 496–503.
- SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- STUART, A. and ORD, J. K. (1983). *Kendall's Advanced Theory of Statistics*. Griffin and Company, London.
- VAN DER VAART, A. W. and WELLNER, J. A. (1989). Prohorov and continuous mapping theorems in the Hoffmann-Jørgensen weak convergence theory with application to convolution and asymptotic minimax theorems. Technical Report 157, Dept. Statistics, Univ. Washington.
- WELLNER, J. A. (1992). Empirical processes in action: a review. *Internat. Statist. Rev.* **60** 247–269.
- WENG, C. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.* **17** 705–710.
- WU, C. F. J. (1987). On the asymptotic properties of the jackknife histogram. Technical report, Dept. Statistics, Univ. Wisconsin, Madison.
- ZHENG, Z. and TU, D. (1988). Random weighting method in regression models. *Sci. Sinica Ser. A* **31** 1442–1459.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
MACLEAN HALL
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

DEPARTMENT OF STATISTICS, GN-22
B313 PADELDFORD HALL
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195