Strong uniform laws of large numbers for bootstrap means and other randomly weighted sums

Neil A. Spencer*[a], Jeffrey W. Miller[b]

[a] *University of Connecticut, Department of Statistics*
Philip E. Austin Building, Storrs, CT 06269
[b] *Harvard University, Department of Biostatistics*
655 Huntington Ave, Boston MA, 02115

Abstract

This article establishes novel strong uniform laws of large numbers for randomly weighted sums such as bootstrap means. By leveraging recent advances, these results extend previous work in their general applicability to a wide range of weighting procedures and in their flexibility with respect to the effective bootstrap sample size. In addition to the standard multinomial bootstrap and the m-out-of-n bootstrap, our results apply to a large class of randomly weighted sums involving negatively orthant dependent (NOD) weights, including the Bayesian bootstrap, jackknife, resampling without replacement, simple random sampling with over-replacement, independent weights, and multivariate Gaussian weighting schemes. Weights are permitted to be non-identically distributed and possibly even negative. Our proof technique is based on extending a proof of the i.i.d. strong uniform law of large numbers to employ strong laws for randomly weighted sums; in particular, we exploit a recent Marcinkiewicz–Zygmund strong law for NOD weighted sums.

Keywords: Empirical Processes, Generalized Bootstrap, Negatively Orthant Dependence, Uniform Marcinkiewicz–Zygmund Strong Law, Resampling

1. Introduction

The bootstrap (Efron and Tibshirani, 1994) and related resampling procedures such as bagging (Breiman, 1996), the Bayesian bootstrap (Rubin, 1981), and the jackknife (Efron, 1982) are widely used general-purpose tools for statistical inference. In addition to its original purpose of approximating sampling distributions of estimators (Efron, 1979), the bootstrap and its relatives have been applied to a variety of statistical tasks, including model averaging (Breiman, 1996), approximate Bayesian inference (Newton and Raftery, 1994), outlier detection (Singh and Xie, 2003), robust Bayesian inference (Huggins and Miller, 2019, 2022), and causal inference (Little and Badawy, 2019).

Due to its versatility, extensions of the bootstrap are frequently proposed to address new statistical questions. When establishing the properties of such methods, bootstrap versions of classical asymptotic results play a key role, such as the weak law of large numbers (Athreya

*Corresponding author

URL: `neil.spencer@uconn.edu` (Neil A. Spencer*), `jwmiller@hsph.harvard.edu` (Jeffrey W. Miller)

et al., 1984, Theorem 1), the strong law of large numbers (Athreya et al., 1984, Theorem 2), and the central limit theorem (Singh, 1981) for bootstrap means.

Meanwhile, it is sometimes important to obtain convergence over an entire collection of random variables simultaneously, thus guaranteeing convergence for even the worst case in the collection. To this end, several authors have established uniform laws of large numbers for bootstrap means. Giné and Zinn (1990, Theorems 2.6 and 3.5) proved weak uniform laws of large numbers for the standard multinomial bootstrap, that is, with $\text{Multinomial}\big(n, (1/n, \ldots, 1/n)\big)$ weights. Vaart and Wellner (1996, Theorem 3.6.16) proved an analogous result for exchangeably weighted sums such as the Bayesian bootstrap. As weak laws, these show convergence in probability, but often one needs almost sure convergence, that is, a strong law.

Strong uniform laws of large numbers for bootstrap means are provided by Kosorok (2008, Section 10.2) for weighted sums with independent and identically distributed (i.i.d.) weights, with weights obtained by normalizing n i.i.d. random variables, and with $\text{Multinomial}\big(n, (1/n, \ldots, 1/n)\big)$ weights (Theorem 10.13, Corollary 10.14, and Theorem 10.15, respectively). However, these results do not apply to more general schemes such as the jackknife, resampling without replacement, and $\text{Multinomial}\big(m_n, (p_{n1}, \ldots, p_{nn})\big)$ weights.

In this article, we present new strong uniform laws of large numbers for randomly weighted sums, aiming to fill these gaps in the literature. Our first result applies to the case of $\text{Multinomial}\big(m_n, (1/n, \ldots, 1/n)\big)$ weights, known as the m-out-of-n bootstrap. Our second and third results apply more generally to a large class of randomly weighted sums that involve negatively orthant dependent (NOD) weights. This covers a wide range of weighting schemes, including the Bayesian bootstrap (Rubin, 1981), various versions of the jackknife (Efron, 1982; Chatterjee and Bose, 2005), resampling without replacement (Bickel et al., 2012), simple random sampling with over-replacement (Antal and Tillé, 2011b), $\text{Multinomial}\big(m_n, (p_{n1}, \ldots, p_{nn})\big)$ weights (Antal and Tillé, 2011a, Section 3), independent weights (Newton and Raftery, 1994), and even schemes involving negative weights such as multivariate Gaussian weights with non-positive correlations (Patak and Beaumont, 2009). All three theorems are flexible in terms of the effective bootstrap sample size m_n (that is, the sum of the weights), for instance, allowing $m_n = o(n)$ which is of particular interest for certain applications (Bickel et al., 2012; Huggins and Miller, 2022).

The article is organized as follows. We present our main results in Section 2, provide examples in Section 3, and the proofs of the main results are provided in Section 4.

2. Main results

We present three strong uniform laws of large numbers: Theorem 2.2 for the multinomial bootstrap, Theorem 2.4 for more general randomly weighted sums, and Theorem 2.7 which establishes faster convergence rates under stronger regularity and moment conditions than Theorem 2.4. All three results are obtained via extensions of the proof of the i.i.d. strong uniform law of large numbers presented by Ghosh and Ramamoorthi (2003). Specifically, our proof of Theorem 2.2 involves replacing the traditional strong law of large numbers with a strong law of large numbers for bootstrap means as presented by Arenal-Gutiérrez et al. (1996). Similarly, Theorems 2.4 and 2.7 rely on a strong law of large numbers for randomly weighted sums of random variables presented by Chen et al. (2019).

Condition 2.1. *Suppose Θ is a compact subset of a separable metric space. Let $u \in \mathbb{N}$ and let $H(\theta, x)$ be a real-valued function on $\Theta \times \mathbb{R}^u$ such that*

(i) for each $x \in \mathbb{R}^u$, $\theta \mapsto H(\theta, x)$ is continuous on Θ, and

(ii) for each $\theta \in \Theta$, $x \mapsto H(\theta, x)$ is a measurable function on $\mathbb{R}^u$.

Theorem 2.2. *Let $X_1, X_2, \ldots \in \mathbb{R}^u$ i.i.d. and for $n \in \mathbb{N}$ independently, let*

$$(W_{n1}, \ldots, W_{nn}) \sim \text{Multinomial}\big(m_n, (1/n, \ldots, 1/n)\big)$$

independently of $(X_1, X_2, \ldots)$, where m_n is a positive integer for $n \in \mathbb{N}$. Assume Condition 2.1 and suppose there exists $\beta \geq 1$ such that

$$\lim_{n\to\infty} \frac{n^{1/\beta}\log(n)}{m_n} = 0 \quad \textit{and} \quad \mathbb{E}\left(\sup_{\theta\in\Theta} |H(\theta, X_1)|^\beta\right) < \infty. \tag{1}$$

Then

$$\sup_{\theta\in\Theta}\left|\frac{1}{m_n}\sum_{j=1}^n W_{nj} H(\theta, X_j) - \mathbb{E}(H(\theta, X_1))\right| \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{2}$$

In Theorems 2.4 and 2.7, we generalize beyond the standard multinomial bootstrap to weighting schemes involving NOD random variables (Lehmann, 1966; Chen et al., 2019).

Definition 2.3. A finite collection of random variables $X_1, \ldots, X_n \in \mathbb{R}$ is said to be *negatively orthant dependent (NOD)* if

$$\mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) \leq \prod_{i=1}^n \mathbb{P}(X_i \leq x_i)$$

and

$$\mathbb{P}\left(X_1 > x_1, \ldots, X_n > x_n\right) \leq \prod_{i=1}^n \mathbb{P}(X_i > x_i)$$

for all $x_1, \ldots, x_n \in \mathbb{R}$. An infinite collection of random variables is NOD if every finite subcollection is NOD.

Any collection of independent random variables is NOD, and many commonly used multivariate distributions are NOD including the multinomial distribution, the Dirichlet distribution, the Dirichlet-multinomial distribution, the multivariate hypergeometric distribution, convolutions of multinomial distributions, and multivariate Gaussian distributions for which the correlations are all non-positive (Joag-Dev and Proschan, 1983).

Theorem 2.4. *Let $X_1, X_2, \ldots \in \mathbb{R}^u$ i.i.d. and for $n \in \mathbb{N}$ independently, let $W_{n1}, \ldots, W_{nn} \in \mathbb{R}$ be NOD random variables, independent of $(X_1, X_2, \ldots)$. Assume Condition 2.1 and suppose $\mathbb{E}(\sup_{\theta\in\Theta} |H(\theta, X_1)|^\beta) < \infty$, $\sum_{j=1}^n \mathbb{E}(|W_{nj}|^\alpha) = O(n)$, and $\sum_{j=1}^n \mathbb{E}(|W_{nj}|) = O(n^{1/p})$ where $p \in [1, 2)$ and $\alpha > 2p$ and $\beta > 1$ satisfy $\alpha^{-1} + \beta^{-1} = p^{-1}$. Then*

$$\sup_{\theta\in\Theta}\left|\frac{1}{n^{1/p}}\sum_{j=1}^n \Big(W_{nj} H(\theta, X_j) - \mathbb{E}(W_{nj} H(\theta, X_j))\Big)\right| \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{3}$$

In particular, if $n^{-1/p}\sum_{j=1}^n \mathbb{E}(W_{nj}) \to 1$, then Equation 3 is analogous to Equation 2 with $n^{1/p}$ in place of m_n. While the moment condition on $H(\theta, X_1)$ in Theorem 2.4 is slightly more stringent than in Theorem 2.2, it applies to a more general class of resampling procedures. For instance, the distribution of W_{nj} can be different for each n and j; in particular, there is no assumption that $W_{n1}, \ldots, W_{nn}$ are exchangeable or even identically distributed. Further, the weights W_{nj} are not restricted to being non-negative; thus, random weights taking positive and negative values are permitted.

The main limitation of Theorem 2.4 is that whenever $p > 1$, the condition that $\sum_{j=1}^n \mathbb{E}(|W_{nj}|) = O(n^{1/p})$ requires the weights to be getting smaller as n increases. One can scale the weights up by a factor of $n^{1-1/p}$, but then the leading factor of $1/n^{1/p}$ in Equation 3 becomes $1/n$, effectively reverting back to the standard rate of convergence. In Theorem 2.7, we show that this condition can be dropped if we assume stronger regularity and moment conditions.

Condition 2.5. *Assume $\theta \mapsto H(\theta, x)$ is uniformly locally Hölder continuous, in the sense that there exist $a > 0$, $M > 0$, and $\delta > 0$ such that for all $x \in \mathbb{R}^u$, $\theta, \theta' \in \Theta$, if $d(\theta, \theta') < \delta$ then $|H(\theta, x) - H(\theta', x)| \leq M d(\theta, \theta')^a$.*

Condition 2.6. *Assume $N(r) \leq c\, r^{-D}$ for some $c > 0$ and $D > 0$, where $N(r)$ is the smallest number of open balls of radius r, centered at points in Θ, needed to cover Θ.*

Condition 2.6 holds for any compact $\Theta \subset \mathbb{R}^D$; indeed, by Shalev-Shwartz and Ben-David (2014, Example 27.1), $N_1(r) \leq c_1 r^{-D}$ where $N_1(r)$ is the number of r-balls needed to cover Θ, centered at any points in $\mathbb{R}^D$, and it is straightforward to verify that $N(r) \leq N_1(r/2)$.

Theorem 2.7. *Let $X_1, X_2, \ldots \in \mathbb{R}^u$ i.i.d. and for $n \in \mathbb{N}$ independently, let $W_{n1}, \ldots, W_{nn} \in \mathbb{R}$ be NOD random variables, independent of $(X_1, X_2, \ldots)$. Assume Conditions 2.1, 2.5, and 2.6. Let $p \in (1, 2)$ and suppose $\sup_{\theta\in\Theta} \mathbb{E}(|H(\theta, X_1)|^q) < \infty$, $\sum_{j=1}^n \mathbb{E}(|W_{nj}|^q) = O(n)$ for some $q > ((1 - 1/p)D/a + 1)/(1/p - 1/2)$. Then*

$$\sup_{\theta\in\Theta} \left| \frac{1}{n^{1/p}} \sum_{j=1}^n \Big(W_{nj} H(\theta, X_j) - \mathbb{E}(W_{nj} H(\theta, X_j)) \Big) \right| \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{4}$$

Note that as $p \to 1$ from above, the bound on the power q approaches 2.

3. Examples

A key condition of Theorems 2.4 and 2.7 is that the weights $W_{n1}, \ldots, W_{nn}$ must be NOD. It turns out that most popular resampling techniques satisfy this condition. For instance, the m-out-of-n bootstrap corresponds to Multinomial$\big(m_n, (1/n, \ldots, 1/n)\big)$ weights, unequal probability with replacement corresponds to Multinomial$\big(m_n, (p_{n1}, \ldots, p_{nn})\big)$ weights (Antal and Tillé, 2011a), the Bayesian bootstrap corresponds to Dirichlet weights, the delete-d jackknife and resampling without replacement correspond to multivariate hypergeometric weights (Chatterjee and Bose, 2005), the weighted likelihood bootstrap is equivalent to using independent weights (Newton and Raftery, 1994), and the reweighting scheme of Patak and Beaumont (2009) employs multivariate Gaussian weights with non-positive correlations. All of these distributions satisfy the NOD requirement (Joag-Dev and Proschan, 1983).

The NOD requirement is also satisfied by less standard reweighting schemes such as the downweight-d jackknife (Chatterjee and Bose, 2005) and simple random sampling with over-replacement (Antal and Tillé, 2011b). In the downweight-d jackknife, d indices $i_1, \ldots, i_d$ are selected uniformly at random to be downweighted such that $W_{ni_1} = \cdots = W_{ni_d} = d/n$, whereas the remaining $n-d$ indices are upweighted to $1+d/n$. These weights can be viewed as a monotonic transformation of the multivariate hypergeometric weights corresponding to the delete-d jackknife, and thus are NOD (Chen et al., 2019). For simple random sampling with over-replacement, the weights can be viewed as the conditional distribution of a sequence of n independent geometric random variables, given that their sum equals n. Geometric random variables satisfy the conditions of Theorem 2.6 from Joag-Dev and Proschan (1983) according to Efron (1965, 3.1), implying that the resulting weights are NOD.

3.1. Farlie–Gumbel–Morgenstern (FGM) copula

A general family of applicable NOD reweighting schemes can also be derived from the Farlie–Gumbel–Morgenstern (FGM) n-copula (Nelsen, 2006, page 108). An FGM copula is a distribution on the n-dimensional unit cube $[0,1]^n$ with cumulative distribution function

$$\mathbb{P}(U_1 \leq x_1, \ldots, U_n \leq x_n) = \Big(\prod_{i=1}^{n} x_i\Big)\Big(1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \theta_{j_1,j_2,\ldots,j_k} \prod_{\ell=1}^{k}(1 - x_{j_\ell})\Big),$$

for all $x \in [0,1]^n$, where the parameters $\theta_{j_1,j_2,\ldots,j_k} \in [-1,1]$ must satisfy the constraints that

$$1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \theta_{j_1,j_2,\ldots,j_k} \epsilon_{j_1}\epsilon_{j_2}\cdots\epsilon_{j_k} \geq 0$$

for all $(\epsilon_1, \ldots, \epsilon_n) \in \{-1,1\}^n$ to ensure non-negativity of the density. There are $2^n - n - 1$ parameters—one for each subset of $\{1, \ldots, n\}$ containing at least two elements. Marginally, each entry U_i is uniformly distributed on $[0,1]$, and thus, $\prod_{i=1}^n \mathbb{P}(U_i \leq x_i) = \prod_{i=1}^n x_i$ and $\prod_{i=1}^n \mathbb{P}(U_i > x_i) = \prod_{i=1}^n (1 - x_i)$ for all $x \in [0,1]^n$. Further, it can be shown that

$$\mathbb{P}(U_1 > x_1, \ldots, U_n > x_n) = \Big(\prod_{i=1}^{n}(1 - x_i)\Big)\Big(1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \theta_{j_1,j_2,\ldots,j_k} (-1)^k \prod_{\ell=1}^{k} x_{j_\ell}\Big)$$

for $x \in [0,1]^n$. Therefore, an FGM n-copula is NOD if and only if the inequalities

$$\sum_{k=2}^{n} \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \theta_{j_1,j_2,\ldots,j_k} \prod_{\ell=1}^{k}(1 - x_{j_\ell}) \leq 0 \tag{5}$$

$$\sum_{k=2}^{n} \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} \theta_{j_1,j_2,\ldots,j_k} (-1)^k \prod_{\ell=1}^{k} x_{j_\ell} \leq 0 \tag{6}$$

hold for all $x \in [0,1]^n$. Since NOD is preserved under monotonic transformations, any weights that can be written as $(W_{n1}, \ldots, W_{nn}) = (g_1(U_1), \ldots, g_n(U_n))$ are NOD if $g_1, \ldots, g_n$ are all monotone increasing (or all monotone decreasing) and U is distributed according to an FGM copula satisfying Equations 5 and 6.

3.2. Moment condition on the weights

The condition that $\sum_{j=1}^{n} \mathbb{E}(|W_{nj}|^q) = O(n)$ in Theorem 2.7 holds for all of the aforementioned reweighting procedures without any additional assumptions, except for a few cases. For the Multinomial$\big(m_n, (p_{n1}, \ldots, p_{nn})\big)$ case, it holds as long as there exists a constant $\kappa > 0$ such that $p_{nj} < \kappa/n$ for all j and n. For the Gaussian reweighting scheme (Patak and Beaumont, 2009), it holds as long as $\sup_{n,j} \mathrm{Var}(W_{nj}) < \infty$. For the independent weights and FGM copula cases, it holds when $\sup_{n,j} \mathbb{E}(|W_{nj}|^q) < \infty$.

4. Proofs

We begin by stating a strong law of large numbers for bootstrap means due to Arenal-Gutiérrez et al. (1996, Theorem 2.1). This lemma plays a key role in our proof of Theorem 2.2. In fact, Arenal-Gutiérrez et al. (1996) only assume pairwise independence of $Z_1, Z_2, \ldots$ rather than mutual independence, but we do not require this additional generality.

Lemma 4.1. *Let $Z_1, Z_2, \ldots \in \mathbb{R}$ i.i.d. and for $n \in \mathbb{N}$ independently, let*

$$(W_{n1}, \ldots, W_{nn}) \sim \mathrm{Multinomial}\big(m_n, (1/n, \ldots, 1/n)\big)$$

independently of $(Z_1, Z_2, \ldots)$, where m_n is a positive integer for $n \in \mathbb{N}$. Suppose there exists $\beta \geq 1$ such that

$$\lim_{n\to\infty} \frac{n^{1/\beta}\log(n)}{m_n} = 0 \quad \textit{and} \quad \mathbb{E}\big(|Z_1|^\beta\big) < \infty.$$

Then

$$\frac{1}{m_n}\sum_{j=1}^{n} W_{nj} Z_j \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}(Z_1).$$

Proof. See Arenal-Gutiérrez et al. (1996, Theorem 2.1) for the proof. □

Proof of Theorem 2.2. Our argument is based on the proof of Theorem 1.3.3 of Ghosh and Ramamoorthi (2003), except that we use Lemma 4.1 in place of the strong law of large numbers for i.i.d. random variables.

Condition 2.1 ensures that $x \mapsto \sup_{\theta\in\Theta} |H(\theta, x)|$ is measurable; this can be seen by letting $\theta_1, \theta_2, \ldots$ be a countable dense subset of Θ, verifying that $\sup_{j\in\mathbb{N}} |H(\theta_j, x)| = \sup_{\theta\in\Theta} |H(\theta, x)|$, and using Folland (2013, Proposition 2.7). Define $\mu(\theta) := \mathbb{E}\left(H(\theta, X_1)\right)$, and note that $\mu(\theta)$ is continuous by the dominated convergence theorem. Let $B_r(\theta_0) := \{\theta \in \Theta : d(\theta, \theta_0) < r\}$ denote the open ball of radius r at θ_0, where $d(\cdot, \cdot)$ is the metric on Θ. For $\theta \in \Theta$, $x \in \mathbb{R}^u$, and $r > 0$, define

$$\eta(\theta, x, r) := \sup_{\theta' \in B_r(\theta)} \Big|\big(H(\theta, x) - \mu(\theta)\big) - \big(H(\theta', x) - \mu(\theta')\big)\Big|, \tag{7}$$

and observe that by continuity and compactness,

$$\eta(\theta, x, r) \leq 2\sup_{\theta\in\Theta} |H(\theta, x)| + 2\sup_{\theta\in\Theta} |\mu(\theta)| < \infty.$$

Applying the dominated convergence theorem again, we have $\lim_{r\to 0}\mathbb{E}(\eta(\theta, X_1, r)) = 0$ for all $\theta \in \Theta$. Thus, for any $\varepsilon > 0$, by compactness of Θ there exist $K \in \mathbb{N}$, $\theta_1, \ldots, \theta_K \in \Theta$, and $r_1, \ldots, r_K > 0$ such that $\Theta = \bigcup_{i=1}^K B_{r_i}(\theta_i)$ and $\mathbb{E}\left(\eta(\theta_i, X_1, r_i)\right) < \varepsilon$ for all $i \in \{1, \ldots, K\}$. Choosing $\beta \geq 1$ according to the statement of Theorem 2.2,

$$\begin{aligned}\mathbb{E}\big(\eta(\theta_i, X_1, r_i)^\beta\big) &\leq \mathbb{E}\Big(\big(2\sup_{\theta\in\Theta}|H(\theta, X_1)| + 2\sup_{\theta\in\Theta}|\mu(\theta)|\big)^\beta\Big) &(8)\\ &\leq 4^\beta\mathbb{E}\Big(\sup_{\theta\in\Theta}|H(\theta, X_1)|^\beta\Big) + 4^\beta\sup_{\theta\in\Theta}|\mu(\theta)|^\beta < \infty &(9)\end{aligned}$$

since $(x+y)^\beta \leq (2\max\{|x|,|y|\})^\beta \leq 2^\beta(|x|^\beta + |y|^\beta)$. For all $i \in \{1, \ldots, K\}$, by applying Lemma 4.1 with $Z_j = \eta(\theta_i, X_j, r_i)$ we have that

$$\frac{1}{m_n}\sum_{j=1}^n W_{nj}\eta(\theta_i, X_j, r_i) \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big) < \varepsilon. \tag{10}$$

Similarly, by Lemma 4.1 with $Z_j = H(\theta_i, X_j)$,

$$\frac{1}{m_n}\sum_{j=1}^n W_{nj}H(\theta_i, X_j) \xrightarrow[n\to\infty]{\text{a.s.}} \mu(\theta_i). \tag{11}$$

Thus, for any $\theta \in \Theta$, by choosing i such that $\theta \in B_{r_i}(\theta_i)$, we have

$$\begin{aligned}&\left|\frac{1}{m_n}\sum_{j=1}^n W_{nj}H(\theta, X_j) - \mu(\theta)\right| &(12)\\ &\quad\leq \frac{1}{m_n}\sum_{j=1}^n W_{nj}\eta(\theta_i, X_j, r_i) + \left|\frac{1}{m_n}\sum_{j=1}^n W_{nj}H(\theta_i, X_j) - \mu(\theta_i)\right|\end{aligned}$$

by the triangle inequality and Equation 7. Letting V_{ni} denote the right-hand side of Equation 12, we have $V_{ni} \to \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big) < \varepsilon$ a.s. by Equations 10 and 11. Therefore,

$$\sup_{\theta\in\Theta}\left|\frac{1}{m_n}\sum_{j=1}^n W_{nj}H(\theta, X_j) - \mu(\theta)\right| \leq \max_{1\leq i\leq K} V_{ni} \xrightarrow[n\to\infty]{\text{a.s.}} \max_{1\leq i\leq K}\mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big) < \varepsilon. \tag{13}$$

Since $\varepsilon > 0$ is arbitrary, Equation 13 holds almost surely for $\varepsilon = \varepsilon_k = 1/k$ for all $k \in \mathbb{N}$, completing the proof. □

The following result, due to Chen et al. (2019), is a more general version of Lemma 4.1 that extends beyond the standard bootstrap to negatively orthant dependent (NOD) weights.

Lemma 4.2. *Let $Z_1, Z_2, \ldots \in \mathbb{R}$ be identically distributed NOD random variables. For each $n \in \mathbb{N}$ independently, let $W_{n1}, \ldots, W_{nn}$ be NOD random variables, independent of $(Z_1, Z_2, \ldots)$. Suppose $\mathbb{E}(|Z_1|^\beta) < \infty$ and $\sum_{j=1}^n \mathbb{E}(|W_{nj}|^\alpha) = O(n)$ where $p \in [1, 2)$ and either (a) $\alpha > 2p$ and $\beta > 1$ satisfy $\alpha^{-1} + \beta^{-1} = p^{-1}$, or (b) the weights W_{nj} are identically distributed for all n, j, and $\alpha = \beta = 2p$. Then*

$$\frac{1}{n^{1/p}}\sum_{j=1}^n \big(W_{nj}Z_j - \mathbb{E}(W_{nj}Z_j)\big) \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Proof. See Chen et al. (2019, Theorem 1 and Corollary 1) for the proof. □

Proof of Theorem 2.4. The proof is similar to the proof of Theorem 2.2, except that we use Lemma 4.2 instead of Lemma 4.1, and some modifications are needed to handle more general distributions of the weights W_{nj}.

As in the proof of Theorem 2.2, define $\eta(\theta, x, r)$ by Equation 7 and $\mu(\theta) := \mathbb{E}\left(H(\theta, X_1)\right)$. As before, for any $\varepsilon > 0$, there exist $K \in \mathbb{N}$, $\theta_1, \ldots, \theta_K \in \Theta$, and $r_1, \ldots, r_K > 0$ such that $\Theta = \bigcup_{i=1}^{K} B_{r_i}(\theta_i)$ and $\mathbb{E}\left(\eta(\theta_i, X_1, r_i)\right) < \varepsilon$ for all $i \in \{1, \ldots, K\}$. Just as in Equation 8,

$$\mathbb{E}\Big(\eta(\theta_i, X_1, r_i)^\beta\Big) \leq 4^\beta \mathbb{E}\Big(\sup_{\theta\in\Theta} |H(\theta, X_1)|^\beta\Big) + 4^\beta \sup_{\theta\in\Theta} |\mu(\theta)|^\beta < \infty.$$

For all $i \in \{1, \ldots, K\}$, by applying Lemma 4.2 with $Z_j = 1$ and $Z_j = H(\theta_i, X_j)$, respectively,

$$\frac{1}{n^{1/p}} \sum_{j=1}^{n} \big(W_{nj} - \mathbb{E}(W_{nj})\big) \xrightarrow[n\to\infty]{\text{a.s.}} 0, \tag{14}$$

$$\frac{1}{n^{1/p}} \sum_{j=1}^{n} \big(W_{nj} H(\theta_i, X_j) - \mathbb{E}\left(W_{nj}\right) \mu(\theta_i)\big) \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{15}$$

Let $W_{nj}^{+} = \max\{W_{nj}, 0\}$ and $W_{nj}^{-} = \max\{-W_{nj}, 0\}$. Then $(W_{n1}^{+}, \ldots, W_{nn}^{+})$ and $(W_{n1}^{-}, \ldots, W_{nn}^{-})$ are each NOD because they are monotone transformations of the W_{nj}'s (Chen et al., 2019). Thus, Lemma 4.2 applied to $Z_j = \eta(\theta_i, X_j, r_i)$ with W_{nj}^{+} and W_{nj}^{-}, respectively, yields

$$\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(W_{nj}^{+} \eta(\theta_i, X_j, r_i) - \mathbb{E}(W_{nj}^{+}) \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big)\Big) \xrightarrow[n\to\infty]{\text{a.s.}} 0, \tag{16}$$

$$\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(W_{nj}^{-} \eta(\theta_i, X_j, r_i) - \mathbb{E}(W_{nj}^{-}) \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big)\Big) \xrightarrow[n\to\infty]{\text{a.s.}} 0. \tag{17}$$

By adding Equations 16 and 17 and using the fact that $|W_{nj}| = W_{nj}^{+} + W_{nj}^{-}$, we have

$$\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(|W_{nj}| \eta(\theta_i, X_j, r_i) - \mathbb{E}\left(|W_{nj}|\right) \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big)\Big) \xrightarrow[n\to\infty]{\text{a.s.}} 0, \tag{18}$$

Since $\mathbb{E}(\eta(\theta_i, X_1, r_i)) < \varepsilon$, Equation 18 implies that, almost surely,

$$\limsup_{n\to\infty} \frac{1}{n^{1/p}} \sum_{j=1}^{n} |W_{nj}| \eta(\theta_i, X_j, r_i) \leq \limsup_{n\to\infty} \frac{1}{n^{1/p}} \sum_{j=1}^{n} \mathbb{E}(|W_{nj}|) \mathbb{E}\big(\eta(\theta_i, X_1, r_i)\big) \leq C\varepsilon \tag{19}$$

where $C := \limsup_{n\to\infty} n^{-1/p} \sum_{j=1}^{n} \mathbb{E}(|W_{nj}|)$. Note that $C < \infty$ by the assumption that $\sum_{j=1}^{n} \mathbb{E}(|W_{nj}|) = O(n^{1/p})$. For any $\theta \in \Theta$, choosing i such that $\theta \in B_{r_i}(\theta_i)$, we can write

$$\begin{aligned} W_{nj} H(\theta, X_j) - \mathbb{E}(W_{nj})\mu(\theta) = W_{nj}\Big(&(H(\theta, X_j) - \mu(\theta)) - (H(\theta_i, X_j) - \mu(\theta_i))\Big) \\ &+ (W_{nj} - \mathbb{E}(W_{nj}))(\mu(\theta) - \mu(\theta_i)) \end{aligned} \tag{20}$$

$$+ \Big(W_{nj} H(\theta_i, X_j) - \mathbb{E}(W_{nj})\mu(\theta_i)\Big).$$

Summing Equation 20 over j, using the triangle inequality, and employing Equation 7,

$$\begin{aligned}\left|\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(W_{nj} H(\theta, X_j) - \mathbb{E}(W_{nj})\mu(\theta)\Big)\right| \leq \frac{1}{n^{1/p}} \sum_{j=1}^{n} |W_{nj}|\, \eta(\theta_i, X_j, r_i) \qquad (21)\\ + \left|\frac{1}{n^{1/p}} \sum_{j=1}^{n} (W_{nj} - \mathbb{E}(W_{nj}))\right| \sup_{\theta\in\Theta} |\mu(\theta) - \mu(\theta_i)|\\ + \left|\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(W_{nj} H(\theta_i, X_j) - \mathbb{E}(W_{nj})\mu(\theta_i)\Big)\right|.\end{aligned}$$

Letting V_{ni} denote the right-hand side of Equation 21, we have $\limsup_n V_{ni} \leq C\varepsilon$ a.s. by Equations 19, 14, and 15, along with the fact that $\mu(\theta)$ is continuous and Θ is compact. Therefore, almost surely,

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \left|\frac{1}{n^{1/p}} \sum_{j=1}^{n} \Big(W_{nj} H(\theta, X_j) - \mathbb{E}(W_{nj})\mu(\theta)\Big)\right| \leq \limsup_{n\to\infty} \max_{1\leq i\leq K} V_{ni} \leq C\varepsilon. \qquad (22)$$

As before, since $\varepsilon > 0$ is arbitrary, Equation 22 holds almost surely for $\varepsilon = \varepsilon_k = 1/k$ for all $k \in \mathbb{N}$, completing the proof. □

Proof of Theorem 2.7. First, we may assume without loss of generality that W_{nj} and $H(\theta, X_j)$ are nonnegative since we can write $W_{nj} = W_{nj}^+ - W_{nj}^-$ and $H(\theta, X_j) = H(\theta, X_j)^+ - H(\theta, X_j)^-$, and apply the result in the nonnegative case to each of $W_{nj}^+ H(\theta, X_j)^+$, $W_{nj}^+ H(\theta, X_j)^-$, $W_{nj}^- H(\theta, X_j)^+$, and $W_{nj}^- H(\theta, X_j)^-$ to obtain the result for $W_{nj} H(\theta, X_j)$ in the general case; see the proof of Theorem 1 from Chen et al. (2019) for a similar argument.

Let $\varepsilon > 0$. As before, define $\eta(\theta, x, r)$ by Equation 7 and $\mu(\theta) := \mathbb{E}\left(H(\theta, X_1)\right)$. For $n \in \mathbb{N}$, define $r_n = (\varepsilon/n^{(1-1/p)})^{1/a}$ and $K_n = N(r_n)$, where $N(r)$ is defined in Condition 2.6. Let $\theta_{n1}, \ldots, \theta_{nK_n} \in \Theta$ be the centers of K_n balls of radius r_n that cover Θ, that is, $\bigcup_{i=1}^{K_n} B_{r_n}(\theta_{ni})$.

Let $C_W = 1 + \sup_n \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}(W_{nj}^q) < \infty$. Note that $q > 2p$ since $p > 1$. Thus, by Lemma 4.2 with $Z_j = 1$,

$$\frac{1}{n} \sum_{j=1}^{n} W_{nj} \leq C_W + \frac{1}{n} \sum_{j=1}^{n} (W_{nj} - \mathbb{E}(W_{nj})) \xrightarrow[n\to\infty]{\text{a.s.}} C_W$$

For all n sufficiently large that $r_n < \delta$, we have $\eta(\theta, x, r_n) \leq 2M r_n^a = 2M\varepsilon/n^{(1-1/p)}$ by Condition 2.5, and thus,

$$\limsup_{n\to\infty} \max_{i\in[K_n]} \frac{1}{n^{1/p}} \sum_{j=1}^{n} W_{nj} \eta(\theta_{ni}, X_j, r_n) \leq 2M\varepsilon \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} W_{nj} \overset{\text{a.s.}}{\leq} 2MC_W\varepsilon \qquad (23)$$

where $[K_n] = \{1, \ldots, K_n\}$. By another application of Lemma 4.2 with $Z_j = 1$,

$$\limsup_{n\to\infty} \max_{i\in[K_n]} \left|\frac{1}{n^{1/p}} \sum_{j=1}^{n} (W_{nj} - \mathbb{E}(W_{nj}))\right| \sup_{\theta\in\Theta} |\mu(\theta) - \mu(\theta_{ni})| \overset{\text{a.s.}}{=} 0 \qquad (24)$$

since $\max_{i\in[K_n]} \sup_\theta |\mu(\theta) - \mu(\theta_{ni})| \leq 2\sup_\theta |\mu(\theta)| < \infty$. Defining $Y_{nij} := W_{nj}H(\theta_{ni}, X_j) - \mathbb{E}(W_{nj})\mu(\theta_{ni})$, we claim that

$$\limsup_{n\to\infty} \max_{i\in[K_n]} \Big|\frac{1}{n^{1/p}} \sum_{j=1}^n Y_{nij}\Big| \overset{\text{a.s.}}{\leq} \varepsilon. \tag{25}$$

Assuming Equation 25 for the moment, we use the same decomposition as in Equation 21 and plug in Equations 23–25 to obtain

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \Big|\frac{1}{n^{1/p}} \sum_{j=1}^n \Big(W_{nj}H(\theta, X_j) - \mathbb{E}(W_{nj})\mu(\theta)\Big)\Big| \leq 2MC_W\varepsilon + 0 + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrarily small, this will yield the result of the theorem.

To complete the proof, we need to show Equation 25. For each n and i, by Chen et al. (2019, Lemma 1), $Y_{ni1}, \ldots, Y_{nin}$ is an NOD sequence since we have assumed without loss of generality that W_{nj} and $H(\theta, X_j)$ are nonnegative, and adding constants preserves the NOD property. Asadian et al. (2006) and Rivaz et al. (2007) provide moment inequalities that are useful in this context. By Asadian et al. (2006, Corollary 2.2) (also see Rivaz et al., 2007, Corollary 3), since $\mathbb{E}(Y_{nij}) = 0$ and $q > 2$,

$$\mathbb{E}\Big(\Big|\sum_{j=1}^n Y_{nij}\Big|^q\Big) \leq C_A(q) \sum_{j=1}^n \mathbb{E}(|Y_{nij}|^q) + C_A(q)\Big(\sum_{j=1}^n \mathbb{E}(|Y_{nij}|^2)\Big)^{q/2} \tag{26}$$

where $C_A(q)$ is a universal constant that depends only on q. Letting $C_H = 1 + \sup_{\theta\in\Theta} \mathbb{E}(|H(\theta, X_1)|^q) < \infty$, we have

$$\sum_{j=1}^n \mathbb{E}(|Y_{nij}|^q) \leq \sum_{j=1}^n 2^{q+1}\mathbb{E}\big(|W_{nj}|^q |H(\theta_{ni}, X_j)|^q\big) \leq 2^{q+1} C_H C_W n. \tag{27}$$

Likewise, since $\mathrm{Var}(X) \leq \mathbb{E}(X^2)$ for any random variable X,

$$\sum_{j=1}^n \mathbb{E}(|Y_{nij}|^2) \leq \sum_{j=1}^n \mathbb{E}\big(|W_{nj}|^2 |H(\theta_{ni}, X_j)|^2\big) \leq C_H C_W n. \tag{28}$$

Plugging Equations 27 and 28 into Equation 26, we have

$$\mathbb{E}\Big(\Big|\sum_{j=1}^n Y_{nij}\Big|^q\Big) \leq C n^{q/2} \tag{29}$$

where $C = C_A(q)2^{q+1}C_H C_W + C_A(q)(C_H C_W)^{q/2}$. By Condition 2.6, $K_n = N(r_n) \leq c\, r_n^{-D} = c\, n^{(1-1/p)D/a}/\varepsilon^{D/a}$. Along with Markov's inequality and Equation 29, this implies

$$\mathbb{P}\Big(\max_{i\in[K_n]} \Big|\frac{1}{n^{1/p}} \sum_{j=1}^n Y_{nij}\Big| > \varepsilon\Big) \leq \sum_{i=1}^{K_n} \mathbb{P}\Big(\Big|\frac{1}{n^{1/p}} \sum_{j=1}^n Y_{nij}\Big| > \varepsilon\Big) \leq \frac{1}{\varepsilon^q n^{q/p}} \sum_{i=1}^{K_n} \mathbb{E}\Big(\Big|\sum_{j=1}^n Y_{nij}\Big|^q\Big)$$

$$\leq \frac{K_n C n^{q/2}}{\varepsilon^q n^{q/p}} \leq \frac{c\,C}{\varepsilon^q \varepsilon^{D/a}} n^{(1-1/p)D/a}\, n^{q(1/2-1/p)} = \frac{c\,C}{\varepsilon^q \varepsilon^{D/a}} n^{-\gamma}$$

where $\gamma := q(1/p - 1/2) - (1 - 1/p)D/a > 1$ because $q > ((1 - 1/p)D/a + 1)/(1/p - 1/2)$ by assumption. Hence,

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\max_{i \in [K_n]} \Big| \frac{1}{n^{1/p}} \sum_{j=1}^{n} Y_{nij} \Big| > \varepsilon \Big) \leq \frac{c\,C}{\varepsilon^q \varepsilon^{D/a}} \sum_{n=1}^{\infty} n^{-\gamma} < \infty.$$

Therefore, Equation 25 holds by the Borel–Cantelli lemma. This completes the proof. □

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Antal, E., Tillé, Y., 2011a. A direct bootstrap method for complex sampling designs from a finite population. Journal of the American Statistical Association 106 (494), 534–543.

Antal, E., Tillé, Y., 2011b. Simple random sampling with over-replacement. Journal of Statistical Planning and Inference 141 (1), 597–601.

Arenal-Gutiérrez, E., Matrán, C., Cuesta-Albertos, J. A., 1996. On the unconditional strong law of large numbers for the bootstrap mean. Statistics & Probability Letters 27 (1), 49–60.

Asadian, N., Fakoor, V., Bozorgnia, A., 2006. Rosenthal's type inequalities for negatively orthant dependent random variables. Journal of the Iranian Statistical Society 5 (1 and 2), 69–75.

Athreya, K. B., Ghosh, M., Low, L. Y., Sen, P. K., 1984. Laws of large numbers for bootstrapped u-statistics. Journal of Statistical Planning and Inference 9 (2), 185–194.

Bickel, P. J., Götze, F., van Zwet, W. R., 2012. Resampling fewer than n observations: gains, losses, and remedies for losses. In: Selected works of Willem van Zwet. Springer, pp. 267–297.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

Chatterjee, S., Bose, A., 2005. Generalized bootstrap for estimating equations. The Annals of Statistics 33 (1), 414–436.

Chen, P., Zhang, T., Sung, S. H., 2019. Strong laws for randomly weighted sums of random variables and applications in the bootstrap and random design regression. Statistica Sinica 29 (4), 1739–1749.

Efron, B., 1965. Increasing properties of Pólya frequency function. The Annals of Mathematical Statistics, 272–279.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 1–26.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. SIAM.

Efron, B., Tibshirani, R. J., 1994. An introduction to the bootstrap. CRC press.

Folland, G. B., 2013. Real Analysis: Modern Techniques and Their Applications. John Wiley & Sons.

Ghosh, J. K., Ramamoorthi, R. V., 2003. Bayesian Nonparametrics. Springer Series in Statistics. Springer, New York, NY.

Giné, E., Zinn, J., 1990. Bootstrapping general empirical measures. The Annals of Probability, 851–869.

Huggins, J. H., Miller, J. W., 2019. Robust inference and model criticism using bagged posteriors. arXiv preprint arXiv:1912.07104.

Huggins, J. H., Miller, J. W., 2022. Reproducible model selection using bagged posteriors. Bayesian Analysis, 1–26.

Joag-Dev, K., Proschan, F., 1983. Negative association of random variables with applications. The Annals of Statistics, 286–295.

Kosorok, M. R., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer.

Lehmann, E. L., 1966. Some concepts of dependence. The Annals of Mathematical Statistics 37 (5), 1137–1153.

Little, M. A., Badawy, R., 2019. Causal bootstrapping. arXiv preprint arXiv:1910.09648.

Nelsen, R. B., 2006. An Introduction to Copulas. Springer.

Newton, M. A., Raftery, A. E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society: Series B (Methodological) 56 (1), 3–26.

Patak, Z., Beaumont, J.-F., 2009. Generalized bootstrap for prices surveys. 57th Session of the International Statistical Institute, Durban, South-Africa.

Rivaz, F., Amini, M., Bozorgnia, A. G., 2007. Moment inequalities and applications for negative dependence random variables 26 (4), 7–11.

Rubin, D. B., 1981. The Bayesian bootstrap. The Annals of Statistics, 130–134.

Shalev-Shwartz, S., Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge university press.

Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. The Annals of Statistics, 1187–1195.

Singh, K., Xie, M., 2003. Bootlier-plot: bootstrap based outlier detection plot. Sankhyā: The Indian Journal of Statistics, 532–559.

Vaart, A. W., Wellner, J. A., 1996. Weak Convergence and Empirical Processes. Springer.