A Note on Bootstrap Moment Consistency for Semiparametric M-Estimation

Guang Cheng[*]
Purdue University

Abstract: The bootstrap variance estimate is widely used in semiparametric inferences. However, its theoretical validity is a well known open problem. In this note, we provide a *first* theoretical study on the bootstrap moment estimates in semiparametric models. Specifically, we establish the bootstrap moment consistency of the Euclidean parameter which immediately implies the consistency of t-type bootstrap confidence set. It is worthy pointing out that the only additional cost to achieve the bootstrap moment consistency in contrast with the distribution consistency (obtained in Cheng and Huang (2010)) is to simply strengthen the L_1 maximal inequality condition required in the latter to the L_p maximal inequality condition for $p \geq 1$. The general L_p multiplier inequality developed in this note is the key technical tool, and is also of independent interest. These general conclusions hold for the bootstrap methods with exchangeable bootstrap weights, e.g., nonparametric bootstrap, and apply to a broad class of semiparametric models with root-n convergent nuisance parameters. Our general theory is illustrated in the celebrated Cox regression model.

Keywords and phrases: Bootstrap Variance Consistency, Semiparametric Model, M-estimation.

1. Introduction

In semiparametric models, the asymptotic variance estimate for the Euclidean parameter is required in the construction of confidence sets and test statistics based on the asymptotic normality result. For example, in the bootstrap inferences, the asymptotic variance estimate is needed to build the t-type confidence set which is known to have smaller coverage probability error than the percentile/hybrid confidence sets; see [25]. In general, the explicit variance estimation is not feasible due to the presence of an infinite dimensional nuisance parameter; see [3, 27] for numerous examples. In the literature, there are two existing estimation procedures, i.e., the profile sampler [14] and the observed profile information [19]. The former (latter) method requires a careful choice of the prior on the Euclidean parameter (of the step size in calculating discretized information estimate). Subsampling [21] is another possibility, but the optimal subsample size is difficult to choose in practice. In contrast, the bootstrap can estimate the asymptotic variance without involving any tuning parameter, and thus becomes a widely used semiparametric inference procedure, e.g., [5, 12, 17]. However, the theoretical validity of the bootstrap variance estimate is a well known open problem.

Cheng and Huang (2010) have recently proven that the bootstrap is asymptotically consistent in estimating the distribution of the M-estimate of Euclidean parameter. However, this distributional consistency does not imply the consistency of the bootstrap variance estimators. Nishiyama (2010) and Kato (2011) have shown the moment convergence of the (bootstrap) M-estimate in parametric models. Inspired by these recent developments, we provide a first theoretical study on the bootstrap moment estimates in semiparametric models. Specifically, we establish the bootstrap moment consistency of the Euclidean parameter which immediately implies the consistency of t-type bootstrap confidence set with the help of the conditional Slutsky's Lemma. It is worthy pointing out that the only additional cost to achieve the bootstrap moment consistency in contrast with the distribution

[*]Department of Statistics, Purdue University, West Lafayette, IN 47907-2066, Email: chengg@purdue.edu. Research sponsored by NSF (DMS-0906497)

consistency is to simply strengthen the L_1 maximal inequality condition required in the latter to the L_p maximal inequality condition for $p \geq 1$. The general L_p multiplier inequality developed in this note is the key technical tool, and is also of independent interest. These general conclusions hold for the bootstrap methods with exchangeable bootstrap weights, e.g., nonparametric bootstrap, and apply to a broad class of semiparametric models with root-n convergent nuisance parameters, e.g., Cox regression model, proportional odds model and case control studies with a missing covariate [23]. The classical Cox regression model is used to illustrate the practicality of the required conditions. As far as we are aware, this note presents the first theoretical studies on the bootstrap variance consistency in semiparametric models.

2. Preliminary

2.1. Semiparametric M-Estimation

The semiparametric M-estimation, including the maximum likelihood estimation as a special case, refers to a general method of estimation. Let $\theta \in \Theta$ be a Euclidean parameter of interest and $\eta \in \mathcal{H}$ be an infinite dimensional nuisance parameter with the norm $d(\cdot)$. The semiparametric M-estimator $(\widehat{\theta}, \widehat{\eta})$ is obtained by optimizing some objective function $m(\theta, \eta)$ based on the observations $(X_1, \ldots, X_n)$:

$$(\widehat{\theta}, \widehat{\eta}) = \arg\sup_{\theta\in\Theta,\eta\in\mathcal{H}} \textstyle\sum_{i=1}^n m(\theta, \eta)(X_i). \tag{1}$$

The form of the objective function depends on the context. For example, it could be the log-likelihood, quasi-likelihood [16] or some pseudo-likelihood function, e.g., [30]. Define $(\theta_0, \eta_0) = \arg\sup_{\theta\in\Theta,\eta\in\mathcal{H}} E_X m(\theta, \eta)(X)$. Under mild conditions, Cheng and Huang (2010) show that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma). \tag{2}$$

Note that $\widehat{\theta}$ is semiparametric efficient and Σ is the inverse of the efficient information matrix when $m(\theta, \eta)$ is the log-likelihood function.

2.2. Exchangeably Weighted Bootstrap

Define the bootstrap M-estimator $(\widehat{\theta}^*, \widehat{\eta}^*) = \arg\sup_{\theta\in\Theta,\eta\in\mathcal{H}} \sum_{i=1}^n m(\theta, \eta)(X_i^*)$, where $(X_1^*, \ldots, X_n^*)$ is the bootstrap sample. Note that the Efron's nonparametric bootstrap consists of independent draws with replacement from the original observations. In this case, we can re-express

$$(\widehat{\theta}^*, \widehat{\eta}^*) = \arg\sup_{\theta\in\Theta,\eta\in\mathcal{H}} \sum_{i=1}^n W_{ni} m(\theta, \eta)(X_i),$$

where $(W_{n1}, \ldots, W_{nn}) \sim \text{Mult}_n(n, (n^{-1}, \ldots, n^{-1}))$. This multinomial formulation can be naturally generalized to a class of exchangeable bootstrap weights $\{W_{ni}\}_{i=1}^n$ whose distribution corresponds to different bootstrap sampling schemes. This general bootstrap method, called exchangeably weighted bootstrap, was first proposed by Rubin (1981) and then extensively studied in [1, 18, 22]. The class of exchangeably weighted bootstrap is practically useful. For example, in Cox regression model, the nonparametric bootstrap often gives many ties when it is applied to censored survival data due to its "discreteness" while the general weighting scheme comes to the rescue. Other variations of nonparametric bootstrap are also studied in [4] using the term "generalized bootstrap".

The bootstrap weights W_{ni}'s are assumed to satisfy the following conditions given in [22]:

W1. The vector $W_n = (W_{n1}, \ldots, W_{nn})'$ is exchangeable for all $n = 1, 2, \ldots$, i.e., for any permutation $\pi = (\pi_1, \ldots, \pi_n)$ of $(1, 2, \ldots, n)$, the joint distribution of $\pi(W_n) = (W_{n\pi_1}, \ldots, W_{n\pi_n})'$ is the same as that of W_n.
W2. $W_{ni} \geq 0$ for all n, i and $\sum_{i=1}^n W_{ni} = n$ for all n.
W3. Let $\|W_{n1}\|_{2,1} = \int_0^\infty \sqrt{P_W(|W_{n1}| \geq u)}du$. Assume $\limsup_{n\to\infty} \|W_{n1}\|_{2,1} \leq C$ for some positive constant $C < \infty$.
W4. $\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 P_W(W_{n1} > t) = 0$.
W5. $(1/n)\sum_{i=1}^n (W_{ni} - 1)^2 \xrightarrow{P_W} c^2 > 0$.

Condition W3 is slightly stronger than the bounded second moment but is implied whenever $(2+\epsilon)$ absolute moment exists for any $\epsilon > 0$; see Appendix A.3. By the Markov's inequality, Condition W4 is satisfied if the $(2 + \epsilon')$ moment of W_{n1} is finite for some $\epsilon' > 0$. The value of c in W5 is independent of n and depends on the resampling method, e.g., $c = 1$ for nonparametric bootstrap. The bootstrap weights corresponding to nonparametric bootstrap satisfy W1–W5. In the below, we present several bootstrap examples satisfying W1 – W5 as shown in Praestgaard and Wellner (1993) where we can find more details on the sampling schemes.

Example 1. i.i.d.-Weighted Bootstraps

In this example, the bootstrap weights are defined as $W_{ni} = \omega_i/\bar{\omega}_n$, where $\omega_1, \omega_2, \ldots, \omega_n$ are i.i.d. positive r.v.s. with $\|\omega_1\|_{2,1} < \infty$. Thus, we can choose $\omega_i \sim$ Exponential(1) or $\omega_i \sim$ Gamma(4, 1). The former corresponds to the Bayesian bootstrap. The multiplier bootstrap is often thought to be a smooth alternative to the nonparametric bootstrap; see [15]. The value of c^2 is calculated as $Var(\omega_1)/(E\omega_1)^2$.

Example 2. The delete-h Jackknife

In the delete-h jackknife [31], the bootstrap weights are generated by permuting the deterministic weights

$$w_n = \left\{\frac{n}{n-h}, \ldots, \frac{n}{n-h}, 0, \ldots, 0\right\} \quad \text{with} \ \sum_{i=1}^n w_{ni} = n.$$

Specifically, we have $W_{nj} = w_{nR_n(j)}$ where $R_n(\cdot)$ is a random permutation uniformly distributed over $\{1, \ldots, n\}$. In Condition M5, $c^2 = h/(n-h)$. Thus, we need to choose $h/n \to \alpha \in (0, 1)$ such that $c > 0$. Therefore, Condition W5 does not hold for the ordinary jackknife with $h = 1$.

Example 3. The Double Bootstrap

In the double bootstrap, the bootstrap weights have the following distribution

$$(W_{n1}, \ldots, W_{nn}) \sim \text{Mult}_n\left(n, (\widetilde{W}_{n1}/n, \ldots, \widetilde{W}_{nn}/n)\right), \tag{3}$$

conditional on $\widetilde{W}_n$ following $\text{Mult}_n(n, (n^{-1}, \ldots, n^{-1}))$. The value of c is $\sqrt{2}$ in this example.

Example 4. The Polya-Eggenberger Bootstrap

In this example, the bootstrap weights follow the multinomial distribution

$$(W_{n1}, \ldots, W_{nn}) \sim \text{Mult}_n\left(n, (D_{n1}, \ldots, D_{nn})\right), \tag{4}$$

conditional on $(D_{n1}, \ldots, D_{nn}) \sim \text{Dirichlet}_n(\alpha, \ldots, \alpha)$ with $\alpha > 0$. The value of c^2 is calculated as $(\alpha + 1)/\alpha$.

Example 5. The Multivariate Hypergeometric Bootstrap

As a particular urn-based bootstrap, the bootstrap weights follow the multivariate hypergeometric distribution with density

$$P(W_{n1} = w_1, \ldots, W_{nn} = w_n) = \frac{\binom{K}{w_1} \cdots \binom{K}{w_n}}{\binom{nK}{n}} \tag{5}$$

for some positive integer K. Condition W5 is satisfied with $c^2 = (K-1)/K$.

Under Conditions W1 – W5 and other regularity conditions, Cheng and Huang (2010) prove

$$(\sqrt{n}/c)(\widehat{\theta}^* - \widehat{\theta}) \overset{d}{\Longrightarrow} N(0,\Sigma) \qquad \text{conditional on } \mathcal{X}_n \equiv (X_1,\ldots,X_n), \tag{6}$$

$$\sup_{x\in\mathbb{R}^d} \left| P_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\theta}^* - \widehat{\theta}) \le x) - P(N(0,\Sigma)\le x)\right| = o_{P_X}(1), \tag{7}$$

where " $\overset{d}{\Longrightarrow}$ " represents the conditional weak convergence (in probability) defined in [8] and $P_{W|\mathcal{X}_n}$ is the conditional probability given $\mathcal{X}_n$. In view of (6), the bootstrap variance estimate for θ is constructed as

$$\widehat{\Sigma}^* = (n/c^2)E_{W|\mathcal{X}_n}(\widehat{\theta}^* - \widehat{\theta})^{\otimes 2}, \tag{8}$$

where $E_{W|\mathcal{X}_n}$ is the conditional expectation given the observed data $\mathcal{X}_n$. We say that the bootstrap variance estimate is consistent if $\widehat{\Sigma}^* \xrightarrow{P_X} \Sigma$. In practice, $\widehat{\Sigma}^*$ can be well approximated as follows:

$$\widehat{\Sigma}^* \approx \widetilde{\Sigma}^* \equiv (n/Bc^2)\textstyle\sum_{b=1}^B \left(\widehat{\theta}^*(b) - B^{-1}\sum_{b=1}^B \widehat{\theta}^*(b)\right)^{\otimes 2},$$

where $\widehat{\theta}^*(b)$ is computed based on the b-th bootstrap sample, for sufficiently large number B of bootstrap repetitions. The bootstrap variance estimate $\widehat{\Sigma}^*$ is widely used in semiparametric inferences. However, its theoretical validity remains an open problem.

3. Main Result: Bootstrap Moment Consistency

In this section, we will establish the bootstrap moment consistency of θ which directly implies the consistency of $\widehat{\Sigma}^*$ and t-type bootstrap confidence set. To obtain the p-th moment consistency comparing to the distribution consistency, the only additional cost is to strengthen the L_1 maximal inequality condition required in the latter to the $L_{p'}$ maximal inequality condition for $p' > p$, i.e., Condition M2. A simple sufficient condition for M2 i.e., (18), is also given in terms of the bootstrap weights, and is verified in the above bootstrap examples.

It is well known that the convergence in distribution implies the convergence in moment under the uniform integrability condition. Lemma 2.1 of Kato (2011) further shows that the above argument is also valid for the conditional weak convergence in the case of nonparametric bootstrap. In fact, his arguments (after minor modifications) can also be applied to the above class of exchangeably weighted bootstrap; see below Lemma 1.

Lemma 1. *Let T_n^* be a scalar statistic of $(X_1,\ldots,X_n)$ and $(W_{n1},\ldots,W_{nn})$. Suppose that bootstrap weight W_n satisfies W1 – W5 and the conditional distribution of T_n^* given $\mathcal{X}_n$ converges weakly to some fixed distribution μ in P_X-probability. If $E_{W|\mathcal{X}_n}|T_n^*|^{q'} = O_{P_X}(1)$ for some $q' > 1$, then $E_{W|\mathcal{X}_n}(T_n^*)^q \xrightarrow{P_X} \int t^q d\mu(t)$ for any integer $1 \le q < q'$.*

Let " $\lesssim$ " (" $\gtrsim$ ") denote smaller (greater) than, up to an universal constant. Denote E_{XW} and P_{XW} as the joint expectation and joint probability, respectively. Let $P_X f = \int f dP_X$, $\mathbb{P}_n f = \sum_{i=1}^n f(X_i)/n$ and $\mathbb{P}_n^* f = \sum_{i=1}^n f(X_i^*)/n = \sum_{i=1}^n W_{ni} f(X_i)/n$. Define the (bootstrap) empirical process and its norm as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P_X)f$ ($\mathbb{G}_n^* f = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)f$) and $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|\mathbb{G}_n f|$ ($\|\mathbb{G}_n^*\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|\mathbb{G}_n^* f|$), respectively. For any class of functions $\mathcal{A}$ under the metric l, we define $\log N_{[]}(\epsilon,\mathcal{A},l)$ and $\log N(\epsilon,\mathcal{A},l)$ as the ϵ-bracketing entropy number and ϵ-entropy number, respectively. The related bracketing entropy integral and uniform entropy integral are thus

$$\begin{aligned} J_{[]}(\delta,\mathcal{A},l) &= \int_0^\delta \sqrt{1+\log N_{[]}(\epsilon,\mathcal{A},l)}d\epsilon, \\ J(\delta,\mathcal{A}) &= \sup_Q \int_0^\delta \sqrt{1+\log N(\epsilon\|A\|_{L_2(Q)},\mathcal{A},L_2(Q))}d\epsilon, \end{aligned}$$

where A is the envelop function of $\mathcal{A}$, and the supreme is taken over all discrete probability measures Q with $\|A\|_{L_2(Q)} > 0$.

In the following, we provide a set of sufficient conditions for bootstrap moment consistency.

M1. For any $(\theta,\eta)\in\Theta\times\mathcal{H}$, we have

$$E_X(m(\theta,\eta)-m(\theta_0,\eta_0)) \lesssim -\|\theta-\theta_0\|^2 - d^2(\eta,\eta_0). \tag{9}$$

M2. Define $\mathcal{N}_\delta = \{m(\theta,\eta)-m(\theta_0,\eta_0) : \|\theta-\theta_0\|\le\delta, d(\eta,\eta_0)\le\delta, (\theta,\eta)\in\Theta\times\mathcal{H}\}$. We assume that, for some $p'\ge 1$ and every $\delta>0$,

$$\Big(E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}^{p'}\Big)^{1/p'} \lesssim \delta, \tag{10}$$

$$\Big(E_{XW}\|\mathbb{G}_n^*\|_{\mathcal{N}_\delta}^{p'}\Big)^{1/p'} \lesssim \delta. \tag{11}$$

M3. Assume that $d(\widehat{\eta},\eta_0)=O_{P_X}(n^{-1/2})$ and $d(\widehat{\eta}^*,\eta_0)=O_{P_{XW}}(n^{-1/2})$.

Condition M1 assumes the quadratic behavior of the criterion function $(\theta,\eta)\mapsto E_X m(\theta,\eta)$. Condition M2 assumes two maximal inequalities in terms of $L_{p'}$-norm for $p'\ge 1$. Both conditions are assumed in the global sense which is absolutely needed to achieve the moment consistency. The convergence rate of the bootstrap estimate in Condition M3, i.e., $d(\widehat{\eta}^*,\eta_0)=O_{P_{XW}}(n^{-1/2})$, can also be understood in the following way: for any $\delta>0$, there exists a $0<L<\infty$ such that

$$P_X\Big(P_{W|\mathcal{X}_n}(\sqrt{n}d(\widehat{\eta}^*,\eta_0)\ge L)>\delta\Big)\longrightarrow 0 \quad \text{as } n\to\infty.$$

We can verify Condition M3 using Theorem 2 of [6] under very weak model assumptions. In the current framework, we find that relaxing the root-n rate required in M3 to the slower one appears to be quite challenging. However, we conjecture that our bootstrap moment consistency results still hold when the infinite dimensional nuisance parameter has slower than root-n rate of convergence, and will leave this as future topics.

In the below, we discuss three different approaches for verifying (10). Lemma 2.14.1 in [27] implies that

$$\Big(E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}^{p'}\Big)^{1/p'} \lesssim J(1,\mathcal{N}_\delta)\|N_\delta\|_{L_{2\vee p'}(P_X)}, \tag{12}$$

where N_δ is the envelop function of $\mathcal{N}_\delta$. Thus, Condition (10) holds if

$$J(1,\mathcal{N}_\delta) < \infty, \tag{13}$$

$$\|N_\delta\|_{L_{2\vee p'}(P_X)} \lesssim \delta. \tag{14}$$

The typical function classes with finite uniform entropy integral include the VC class and the related larger VC-hull class; see their definitions in Section 2.6 of [27]. Under the (global) Lipschitz continuous condition:

$$|m(\theta,\eta)(x)-m(\theta_0,\eta_0)(x)|\le M(x)(\|\theta-\theta_0\|+d(\eta,\eta_0)) \quad \text{for any } (\theta,\eta)\in\Theta\times\mathcal{H}, \tag{15}$$

we can show (14) if $E_X M^{2\vee p'}(X)<\infty$. Alternatively, by decomposing $(m(\theta,\eta)-m(\theta_0,\eta_0))$ as the sum of $(m(\theta,\eta)-m(\theta_0,\eta))$ and $(m(\theta_0,\eta)-m(\theta_0,\eta_0))$, we can also verify (10) if the following holds:

$$\left(E_X\|\mathbb{G}_n\|_{\dot{\mathcal{N}}_\delta}^{p'}\right)^{1/p'}<\infty\quad\text{and}\quad\left(E_X\|\mathbb{G}_n\|_{\mathcal{N}_{\delta 2}}^{p'}\right)^{1/p'}\lesssim\delta,$$

where $\dot{\mathcal{N}}_\delta=\{(\partial/\partial\theta)m(\theta,\eta):\|\theta-\theta_0\|\le\delta,d(\eta,\eta_0)\le\delta\}$ and $\mathcal{N}_{\delta 2}=\{m(\theta_0,\eta)-m(\theta_0,\eta_0):d(\eta,\eta_0)\le\delta\}$. Again, Lemma 2.14.1 in [27] can be applied here. Our third approach is to bound the higher moments $(E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}^{p'})^{1/p'}$ for $p'>1$ by $E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}$ plus some norm of N_δ, based on the following two inequalities:

$$\left(E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}^{p'}\right)^{1/p'}\lesssim E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}+n^{1/2-1/p'}\|N_\delta\|_{\psi_{p'}}\quad\text{for } 1<p'<2,\tag{16}$$

$$\left(E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}^{p'}\right)^{1/p'}\lesssim E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}+n^{-1/2+1/p'}\left(E_X|N_\delta|^{p'}\right)^{1/p'}\quad\text{for } p'\ge 2,\tag{17}$$

where $\|\cdot\|_{\psi_p}$ is the Orlicz norm with $\psi_p(t)=\exp(t^p)-1$. The above two inequalities are derived based on Theorem 2.14.5 in [27] and the fact that the ψ_p-norm dominates the L_p-norm for each p. Now, we assume (15). When $p'>1$ but $\neq 2$, the second term in the right hand side of (16) ((17)) converges to zero as $n\to\infty$ if $\|M\|_{\psi_{p'}}<\infty$ ($\|M\|_{L_{p'}(P_X)}<\infty$). When $p'=2$, the second term in the right hand side of (17) is of the order $O(\delta)$ if $\|M\|_{L_2(P_X)}<\infty$. Thus, if $E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}\lesssim\delta$, we can show (10). Fortunately, several technical tools are available to compute the upper bound of $E_X\|\mathbb{G}_n\|_{\mathcal{N}_\delta}$ in terms of the bracketing entropy integral (using Theorem 2.14.2 or Lemma 3.4.2 in [27]) or the uniform entropy integral (see van der Vaart and Wellner (2011)). For example, in view of the above analysis and Theorem 2.14.2 in [27], a simple sufficient condition for (10) is

$$J_{[]}(1,\mathcal{N}_\delta,L_2(P_X))+\|M\|_{\psi_{p'\vee 2}}<\infty\text{ and Condition (15)}$$

due to the fact that the ψ_p-norm dominates the L_p-norm for each p, and ψ_q-norm for any $q\le p$.

To verify (11), we will employ the general L_p multiplier inequality developed in Appendix A.4 to bound $(E_{XW}\|\mathbb{G}_n^*\|_{\mathcal{N}_\delta}^{p'})^{1/p'}$. According to Appendix A.5, it suffices to show the following bootstrap weight condition

$$W_{n1}^{p'}\ \text{satisfies Conditions}\ W3\ \&\ W4;\tag{18}$$

if (10) holds. Condition (18) is essentially very weak; see discussions in *Examples 1 – 5* below. In the end, we want to point out that Conditions W1 – W5 and M1 – M3 (when $p'=1$) are also needed in showing the bootstrap distribution consistency (6) & (7); see Theorems 1 & 3 of [6]. In view of the above discussions, it appears that we only need to strengthen the L_1 maximal inequalities to the $L_{p'}$ maximal inequalities for $p'\ge 1$ to achieve the bootstrap moment consistency beyond the distribution consistency.

Let ET^p represent the p-th moment of any random vector T.

Theorem 1. *Suppose that Conditions W1 – W5 and M1 – M3 hold. If $\widehat{\theta}^*$ is distribution consistent, i.e., (6), then we have*

$$E_{W|\mathcal{X}_n}(\sqrt{n}(\widehat{\theta}^*-\widehat{\theta}))^p\xrightarrow{P_X}ET^p,\tag{19}$$

where $T\sim N(0,\Sigma)$, for any integer p satisfying $1\le p<p'$.

An obvious implication of Theorem 1 is that the bootstrap moment estimate of arbitrary order is consistent if Condition M2 is valid for all $p' \geq 1$. It is worthwhile to remark that the uniform integrability of $\widehat{\theta}$, i.e., $E_X\|\sqrt{n}(\widehat{\theta}-\theta_0)\|^p < \infty$, is also proven in the proof of Theorem 1. Thus, under the same set of conditions, the moment convergence of $\widehat{\theta}$ also follows. In addition, Theorem 1 is also valid even for the approximate maximizer, i.e.,

$$\begin{aligned}
\mathbb{P}_n m(\widehat{\theta},\widehat{\eta}) &\geq \mathbb{P}_n m(\theta_0,\eta_0) - O_{P_X}(n^{-1}),\\
\mathbb{P}_n^* m(\widehat{\theta}^*,\widehat{\eta}^*) &\geq \mathbb{P}_n^* m(\theta_0,\eta_0) - O_{P_{XW}}(n^{-1})
\end{aligned}$$

after slightly modifying its proof.

The distribution consistency result (7) directly implies the consistency of bootstrap hybrid and percentile confidence sets. Given the consistent variance estimate $\widehat{\Sigma}$ based on $(X_1,\ldots,X_n)$, the more accurate t-type bootstrap confidence set is constructed as

$$BC_t(\alpha) = \left[\widehat{\theta} - \frac{\widehat{\Sigma}^{1/2}\omega^*_{n(1-\alpha/2)}}{\sqrt{n}}, \widehat{\theta} - \frac{\widehat{\Sigma}^{1/2}\omega^*_{n(\alpha/2)}}{\sqrt{n}}\right],$$

where $\omega^*_{n\alpha}$ satisfies $P_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\Sigma}^*)^{-1/2}(\widehat{\theta}^*-\widehat{\theta}) \leq \omega^*_{n\alpha}) = \alpha$. Note that $\omega^*_{n\alpha}$ is not unique when θ is a vector. The following Corollary theoretically justifies the widely used bootstrap variance estimate $\widehat{\Sigma}^*$, and further establishes the consistency of t-type confidence set $BC_t(\alpha)$.

Corollary 1. *Suppose that Conditions in Theorem 1 hold. If we further require that Condition M2 holds for some $p' > 2$, then we have*

$$\widehat{\Sigma}^* \xrightarrow{P_X} \Sigma, \tag{20}$$

$$P_{XW}(\theta_0 \in BC_t(\alpha)) \longrightarrow 1-\alpha \tag{21}$$

as $n \to \infty$.

The variance consistency (20) directly follows from Theorem 1. To show the consistency of t-type confidence set, i.e., (21), we apply the Slutsky's Lemma and its conditional version given in Appendix A.2 (together with Lemma 4.6 of [22]) to (2) and (6). Thus, for any fixed $x \in \mathbb{R}^d$, we obtain that

$$P_X(\sqrt{n}\widehat{\Sigma}^{-1/2}(\widehat{\theta}-\theta_0) \leq x) \longrightarrow \Psi(x), \tag{22}$$

$$P_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\Sigma}^*)^{-1/2}(\widehat{\theta}^*-\widehat{\theta}) \leq x) \xrightarrow{P_X} \Psi(x), \tag{23}$$

where $\Psi(x) = P(N(0,I) \leq x)$. A straightforward application of Lemma 23.3 in [28] concludes the proof of (21) based on (22) & (23).

In the end of this section, we will verify the bootstrap weight condition (18) in six different types of bootstraps introduced in Section 2.2.

Example 1. i.i.d.-Weighted Bootstraps (Cont')

We will show that (18) holds under the assumption that ω_i has bounded $(2+\epsilon)p'$-th moment for some $\epsilon > 0$. This assumption implies that

$$\|\omega_i^{p'}\|_{2,1} < \infty \tag{24}$$

based on Appendix A.3. The derivations in Page 2080 of [22] give that

$$P_W(W_{n1}^{p'} > t) \leq P(\omega_1 > t^{1/p'}(1-\epsilon)) + t^{-p/(2p')}n^{p/2}\rho(\epsilon)^{n/2} \tag{25}$$

for any $0 < \epsilon < 1$, $p > 0$ and some $0 < \rho(\epsilon) < 1$, which further implies that

$$\begin{aligned}\|W_{n1}^{p'}\|_{2,1} &\leq 1 + \int_1^\infty \sqrt{P_W(W_{n1}^{p'} > t)}dt \\ &\leq 1 + \int_1^\infty \sqrt{P(\omega_1^{p'} > t(1-\epsilon)^{p'})}dt + n^{p/4}\rho(\epsilon)^{n/4}\int_1^\infty t^{-p/4p'}dt \\ &\leq 1 + \frac{1}{(1-\epsilon)^{p'}}\|\omega_1^{p'}\|_{2,1} + n^{p/4}\rho(\epsilon)^{n/4}\int_1^\infty t^{-p/4p'}dt.\end{aligned}$$

By choosing $p > 4p'$, we know that $\limsup_{n\to\infty}\|W_{n1}^{p'}\|_{2,1} < \infty$ due to (24). To see that $W_{n1}^{p'}$ satisfies Condition W4, it suffices to show that $\lim_{t\to\infty} t^2 P(\omega_1^{p'} > t) = 0$ according to (25). This is implied by the Markov's inequality and the bounded moment assumption on ϵ.

Example 2. The delete-h Jackknife (Cont')

Recall that the bootstrap weight $W_{nj} = w_{nR_n(j)}$. Then, we have

$$P_W(W_{n1} > t) = \sharp\{j : w_{nj} > t\} = \frac{n-h}{n}1\{t < n/(n-h)\}. \tag{26}$$

In view of (26), Condition (18) can be verified as follows

$$\limsup_{n\to\infty}\int_0^\infty \sqrt{P_W(W_{n1} > u)}du^{p'} = \limsup_{n\to\infty}\left(\frac{n}{n-h}\right)^{p'-1/2} = \left(\frac{1}{1-\alpha}\right)^{p'-1/2} < \infty, \tag{27}$$

$$\lim_{t\to\infty}\limsup_{n\to\infty}\frac{n-h}{n}t^2 1\left\{t < \frac{n^{p'}}{(n-h)^{p'}}\right\} = \lim_{t\to\infty}(1-\alpha)t^2 1\{t < (1-\alpha)^{-p'}\} = 0. \tag{28}$$

A sufficient condition for (18) is

$$\limsup_{n\to\infty} E_W W_{n1}^{(2+\epsilon)p'} < \infty \quad \text{for some } \epsilon > 0. \tag{29}$$

This can be proven based on the Appendix A.3 and Chebyshev's inequality as remarked above. Thus, to guarantee the bootstrap variance consistency, i.e. Corollary 1, we only need to require

$$\limsup_{n\to\infty} E_W W_{n1}^5 < \infty \tag{30}$$

since we can always choose $p' = 5/(2+\epsilon) > 2$ for small enough $\epsilon > 0$. Assuming $W_n = (W_{n1}, \ldots, W_{nn})' = \text{Mult}_n(n, (p_1, \ldots, p_n))$, we have

$$E_W W_{n1}^5 = np_1 + 15n^{(2)}p_1^2 + 25n^{(3)}p_1^3 + 10n^{(4)}p_1^4 + n^{(5)}p_1^5, \tag{31}$$

where $n^{(k)} = n(n-1)\cdots(n-k+1)$, according to Page 33 in [9]. If $p_i = 1/n$ for $i = 1, \ldots, n$, we know $E_W W_{n1}^5 < 52$. Thus, Condition (30) (also (18)) is trivially satisfied in the *Efron's nonparametric bootstrap*. Condition (30) can be easily verified in the examples 3 – 5 discussed before.

Example 3. The Double Bootstrap (Cont')

Based on (3) & (31), we can compute $E_W W_{n1}^5$ as

$$\begin{aligned}& E(E_W(W_{n1}^5|\widetilde{W}_n)) \\ = \;& E\left(\widetilde{W}_{n1} + 15(n^{(2)}/n^2)\widetilde{W}_{n1}^2 + 25(n^{(3)}/n^3)\widetilde{W}_{n1}^3 + 10(n^{(4)}/n^4)\widetilde{W}_{n1}^4 + (n^{(5)}/n^5)\widetilde{W}_{n1}^5\right),\end{aligned}$$

which implies Condition (30) since $E\widetilde{W}_{n1}^5 < 52$.

Example 4. The Polya-Eggenberger Bootstrap (Cont')
Following similar analysis in double bootstrap and (4), we have

$$E_W W_{n1}^5 = E\left(nD_{n1} + 15n^{(2)}D_{n1}^2 + 25n^{(3)}D_{n1}^3 + 10n^{(4)}D_{n1}^4 + n^{(5)}D_{n1}^5\right).$$

We can verify (30) if we can show

$$\limsup_{n\to\infty} n^{(p)}ED_{n1}^p < \infty$$

for $p = 1, \ldots, 5$. This is essentially true for all p based on the below derivations

$$n^{(p)}ED_{n1}^p = n^{(p)}\frac{\alpha\cdots(\alpha+p-1)}{n\alpha\cdots(n\alpha+p-1)} \longrightarrow \prod_{k=1}^{p-1}\frac{\alpha+k}{\alpha} \quad \text{as } n\to\infty,$$

where the formula for calculating ED_{n1}^p is given in Page 96 of [10].
Example 5. The Multivariate Hypergeometric Bootstrap (Cont')
According to (5) and Page 96 of [10], we have

$$E_W W_{n1}^5 = a_{n,K}(1) + 15a_{n,K}(2) + 25a_{n,K}(3) + 10a_{n,K}(4) + a_{n,K}(5),$$

where $a_{n,K}(r) = n^{(r)}K^{(r)}/(nK)^{(r)}$. Since $a_{n,K}(r) < K^{(r)}$, we can show $\limsup_{n\to\infty} E_W W_{n1}^5 < \infty$.

4. Cox Regression Model with Right Censored Data

We use the following Cox regression model to illustrate the practicality of the stated conditions M1 – M3. Indeed, the advantages of using bootstrap inferences in this model were considered in the literature, e.g., [7]. In the Cox regression model, the hazard function of the survival time T of a subject with covariate Z is modelled as:

$$\lambda(t|z) \equiv \lim_{\Delta\to 0}\frac{1}{\Delta}P(t \le T < t+\Delta | T \ge t, Z = z) = \lambda(t)\exp(\theta' z), \tag{32}$$

where λ is an unspecified baseline hazard function and θ is a regression vector. In this model, we are usually interested in θ while treating the cumulative hazard function $\eta(y) = \int_0^y \lambda(t)dt$ as the nuisance parameter. With right censoring of survival time, the data observed is $X = (Y, \delta, Z)$, where $Y = T\wedge C$, C is a censoring time, $\delta = I\{T \le C\}$, and Z is a regression covariate belonging to a compact set $\mathbb{Z}\subset\mathbb{R}^d$. We assume that C is independent of T given Z. The log-likelihood is thus

$$m(\theta,\eta)(x) = \delta\theta' z - \exp(\theta' z)\eta(y) + \delta\log\eta\{y\}, \tag{33}$$

where $\eta\{y\} = \eta(y) - \eta(y-)$ is a point mass that denotes the jump of η at point y. The parameter space $\mathcal{H}$ is restricted to a set of nondecreasing cadlag functions on the interval $[0,\tau]$ with $\eta(\tau) \le M$ for some constant M. It is well known that the MLE $\widehat{\theta}$ is semiparametric efficient with the asymptotic variance obtained in [2]:

$$\Sigma = \widetilde{I}_0^{-1} \equiv \left\{E\widetilde{\ell}_{\theta_0,\eta_0}(X)\widetilde{\ell}'_{\theta_0,\eta_0}(X)\right\}^{-1}, \tag{34}$$

where the efficient information matrix $\widetilde{I}_0$ is computed via the efficient score function

$$\widetilde{\ell}_{\theta,\eta}(x) = \delta\left(z - \frac{Ee^{\theta' Z}Z1\{Y\ge y\}}{Ee^{\theta' Z}1\{Y\ge y\}}\right) - e^{\theta' z}\int_0^y\left(z - \frac{Ee^{\theta' Z}Z1\{Y\ge t\}}{Ee^{\theta' Z}1\{Y\ge t\}}\right)d\eta(t).$$

The negative second derivative of the partial likelihood can be used to estimate Σ^{-1}. This is a special case of the observed profile information defined as the negative second numerical derivative of the profile likelihood; see [19]. In general, this approach requires a careful choice of the step size and crucially depends on the curvature structure of the profile likelihood which may not behave well under small sample.

Cheng and Huang (2010) have shown that the exchangeably weighted bootstrap is consistent in estimating the limiting distribution of $\widehat{\theta}$. Below, we will verify that Conditions M1 – M3 hold for this model such that the bootstrap is also consistent for estimating Σ. Since the true value (θ_0, η_0) is the maximizer of $(\theta, \eta) \mapsto P_X m(\theta, \eta)$ (under certain identifiability condition), it is not difficult to verify Condition M1 by defining $d(\eta, \eta_0) = \|\eta - \eta_0\|_\infty$, where $\|\cdot\|_\infty$ denotes the supreme norm. The convergence rates of $\widehat{\eta}$ ($\widehat{\eta}^*$) is established in Theorem 3.1 of [19] (Theorem 2 of [6]), i.e.,

$$\|\widehat{\eta} - \eta_0\|_\infty = O_{P_X}(n^{-\frac{1}{2}}) \quad \text{and} \quad \|\widehat{\eta}^* - \eta_0\|_\infty = O_{P_{XW}}(n^{-\frac{1}{2}}). \tag{35}$$

Thus, we have verified Condition M3. To verify (10) in M2, we apply the first approach by showing (13) & (14). Note that the class of bounded monotone functions, e.g., $\eta(y)$ and $\eta(y-)$, is VC-hull class. Considering the form of $m(\theta, \eta)$ (writing $\eta\{y\} = \eta(y) - \eta(y-)$), we know that (13) is satisfied by the stability property of the BUEI function class, i.e., Lemma 9.14 of [13]. Note that (14) trivially holds since we can show (15) with $M(x)$ as some finite constant due to the compactness of $\mathbb{Z}$ and $\mathcal{H}$. This also justifies $\|N_\delta\|_{L_{p'}(P_X)} < \infty$. Thus, (11) holds according to Appendix A.5.

Acknowledgment. The author thanks Professor Yoichi Nishiyama for sending me his technical note attached to Nishiyama (2010) and thanks Professor Jon Wellner for helpful discussions.

Appendix

For simplicity, we denote $\|f\|_{Q,r}$ as the $L_r(Q)$-norm of the function f. Let T_n^* be a random vector composed of $(X_1, \ldots, X_n)$ and $(W_{n1}, \ldots, W_{nn})$. According to [8], we say that the conditional distribution of T_n^* given $\mathcal{X}_n$ converges weakly to some fixed distribution T in P_X-probability, denoted as "$T_n^* \Longrightarrow T$", if

$$\sup_{f \in BL_1} \left| E_{W|\mathcal{X}_n} f(T_n^*) - \int f dT \right| \xrightarrow{P_X} 0, \tag{A.1}$$

where BL_1 is the class of Lispchitz functions bounded by 1 and with Lipschitz norm 1.

A.1. Proof of Theorem 1

Chose some p'' satisfying $p < p'' < p'$. According to Lemma 1, it suffices to show that

$$\sup_n E_{XW} \|\sqrt{n}(\widehat{\theta}^* - \theta_0)\|^{p''} < \infty \quad \text{and} \quad \sup_n E_X \|\sqrt{n}(\widehat{\theta} - \theta_0)\|^{p''} < \infty.$$

The latter result is a special case of the former since we may take $W_{ni} = 1$ a.s. for $i = 1, \ldots, n$. To show the former, it suffices to show

$$\sup_n E_{XW} \{\sqrt{n}[\|\widehat{\theta}^* - \theta_0\| + d(\widehat{\eta}^*, \eta_0)]\}^{p''} < \infty. \tag{A.2}$$

To show (A.2), we need to partition the parameter space $\Theta \times \mathcal{H}$ into "shells" $S_{j,n}$, i.e.,

$$S_{j,n} = \left\{(\theta, \eta) \in \Theta \times \mathcal{H} : 2^{j-1} < \sqrt{n}(\|\theta - \theta_0\| + d(\eta, \eta_0)) \le 2^j\right\}$$

with j ranging over integers, and then bound the probability of each shell under Conditions M1-M2. For any fixed $j_0 > 0$, we have

$$\begin{aligned}
& E_{XW}\{\sqrt{n}[\|\widehat{\theta}^* - \theta_0\| + d(\widehat{\eta}^*, \eta_0)]\}^{p''} \\
\leq\ & 2^{(j_0-1)p''} P_{XW}(\sqrt{n}(\|\widehat{\theta}^* - \theta_0\| + d(\widehat{\eta}^*, \eta_0)) \leq 2^{j_0-1}) \\
& + \sum_{j=j_0}^{\infty} 2^{jp''} P_{XW}(2^{j-1} < \sqrt{n}(\|\widehat{\theta}^* - \theta_0\| + d(\widehat{\eta}^*, \eta_0)) \leq 2^{j}) \\
\leq\ & 2^{(j_0-1)p''} + \sum_{j=j_0}^{\infty} 2^{jp''} P_{XW}\left(\sup_{(\theta,\eta)\in S_{j,n}} \mathbb{P}_n^* \left(m(\theta,\eta) - m(\theta_0,\eta_0)\right) \geq 0\right) \\
\leq\ & 2^{(j_0-1)p''} + \sum_{j=j_0}^{\infty} 2^{jp''} P_{XW}\left(\sup_{(\theta,\eta)\in S_{j,n}} (\mathbb{P}_n^* - P_X)\left(m(\theta,\eta) - m(\theta_0,\eta_0)\right) \gtrsim \frac{2^{2j-2}}{n}\right),
\end{aligned}$$

where the last inequality follows from Condition M1. By the decomposition that $(\mathbb{P}_n^* - P_X)f = n^{-1/2}(\mathbb{G}_n^* + \mathbb{G}_n)f$, we can further bound the second term in the above by

$$\begin{aligned}
& \sum_{j=j_0}^{\infty} 2^{jp''} P_{XW}\left(\sup_{(\theta,\eta)\in S_{j,n}} \mathbb{G}_n^*(m(\theta,\eta) - m(\theta_0,\eta_0)) \gtrsim \frac{2^{2j-2}}{\sqrt{n}}\right) \\
+\ & \sum_{j=j_0}^{\infty} 2^{jp''} P_{X}\left(\sup_{(\theta,\eta)\in S_{j,n}} \mathbb{G}_n(m(\theta,\eta) - m(\theta_0,\eta_0)) \gtrsim \frac{2^{2j-2}}{\sqrt{n}}\right) \\
\leq\ & \sum_{j=j_0}^{\infty} 2^{jp''} \left[\left(\frac{2^j/\sqrt{n}}{2^{2j-2}/\sqrt{n}}\right)^{p'} + \left(\frac{2^j/\sqrt{n}}{2^{2j-2}/\sqrt{n}}\right)^{p'}\right] \\
\lesssim\ & \sum_{j=j_0}^{\infty} 2^{j(p''-p')}
\end{aligned}$$

The first inequality follows from Markov's inequality and Condition M2. Now, we can conclude that

$$\begin{aligned}
& E_{XW}\{\sqrt{n}[\|\widehat{\theta}^* - \theta_0\| + d(\widehat{\eta}^*, \eta_0)]\}^{p'} \\
\leq\ & 2^{(j_0-1)p'} + \sum_{j=j_0}^{\infty} 2^{j(p''-p')} < \infty
\end{aligned}$$

since we assume that $p'' < p'$. This concludes the proof. $\square$

A.2. Conditional Slutsky's Lemma

Let T_n^* and C_n be random vectors composed of $(X_1, \ldots, X_n)$ and $(W_{n1}, \ldots, W_{nn})$, and $(X_1, \ldots, X_n)$, respectively. If $T_n^* \Longrightarrow T$ and $C_n \xrightarrow{P_X} C$ for some vector C, then we have

(i) $T_n + C_n \Longrightarrow T + C$;
(ii) $C_n T_n \Longrightarrow CT$;
(iii) $C_n^{-1} T_n \Longrightarrow C^{-1} T$ provided $C \neq 0$,

where the vector C in (i) must be of the same dimension as T and C in (ii) & (iii) can be a matrix.

Proof: Without loss of generality, we assume C to be a vector. If C is a matrix, the conclusions in (ii) and (iii) are still valid since the matrix multiplication and matrix inversion are both continuous operations. We first show the conditional weak convergence $(T_n, C_n) \Longrightarrow (T, C)$, and then apply the conditional version of the continuous mapping Theorem, i.e., Theorem 10.8 in [13], to conclude the proof. We first show the following result:

$$\text{if } U_n^* \Longrightarrow U \text{ and } \|U_n^* - V_n^*\| \xrightarrow{P_X} 0, \text{ then } V_n^* \Longrightarrow U, \tag{A.3}$$

where U_n^* and V_n^* are random vectors composed of $(X_1, \ldots, X_n)$ and $(W_{n1}, \ldots, W_{nn})$. For any $f \in BL_1$, we have

$$|E_{W|\mathcal{X}_n} f(U_n^*) - E_{W|\mathcal{X}_n} f(V_n^*)| \leq \epsilon E_{W|\mathcal{X}_n} 1\{\|U_n^* - V_n^*\| \leq \epsilon\} + 2E_{W|\mathcal{X}_n} 1\{\|U_n^* - V_n^*\| > \epsilon\} \tag{A.4}$$

for every $\epsilon > 0$. The first term in the right hand side of (A.4) can be made arbitrarily small by choice of ϵ while the second term converges to zero in P_X-probability as $n \to \infty$. Thus, we claim

$$\sup_{f \in BL_1} \left| E_{W|\mathcal{X}_n} f(U_n^*) - E_{W|\mathcal{X}_n} f(V_n^*) \right| \xrightarrow{P_X} 0.$$

Considering the definition (A.1) and $U_n^* \Longrightarrow U$, we complete the proof of (A.3). According to (A.3), it suffices to show $(T_n^*, C) \Longrightarrow (T, C)$ since $\|(T_n^*, C_n) - (T_n^*, C)\| = \|C_n - C\| \xrightarrow{P_X} 0$. It is easy to show that for every bounded Lipschitz function $(x, y) \mapsto f(x, y)$, the function $x \mapsto f(x, c)$ is also bounded and Lipschitz continuous. Thus, if $T_n^* \Longrightarrow T$, then we have

$$\sup_{f \in BL_1} \left| E_{W|\mathcal{X}_n} f(T_n^*, C) - Ef(T, C) \right| \leq \sup_{f \in BL_1} \left| E_{W|\mathcal{X}_n} f(T_n^*) - Ef(T) \right| \xrightarrow{P_X} 0.$$

Again, an application of (A.1) completes the whole proof. □

A.3. An Inequality for $\|\cdot\|_{2,1}$-norm

For any $Y > 0$ and $r > 2$, we have

$$\frac{1}{2}\|Y\|_2 \leq \|Y\|_{2,1} \leq \frac{r}{r-2}\|Y\|_r, \tag{A.5}$$

where $\|Y\|_r = (EY^r)^{1/r}$.

Proof: The first inequality is established as follows:

$$\|Y\|_2^2 = 2\int_0^\infty tP(|Y| > t)du = 2\int_0^\infty t\sqrt{P(|Y| > t)}\sqrt{P(|Y| > t)}dt \leq 2\|Y\|_{2,1}\|Y\|_2$$

by Markov's inequality. For the second inequality, we have

$$\begin{aligned} \|Y\|_{2,1} &= \left(\int_0^a + \int_a^\infty\right)\sqrt{P(|Y| > t)}dt \\ &\leq a + \|Y\|_r^{r/2}\int_a^\infty t^{-r/2}dt \\ &\leq a + \|Y\|_r^{r/2}\frac{2a^{1-r/2}}{r-2} \equiv U(a) \end{aligned}$$

for any $a > 0$. It is easy to show that the minimal of $U(a)$ is just $[r/(r-2)]\|Y\|_r$ when $a = \|Y\|_r$. This completes the proof of the second inequality in (A.5). □

A.4. The L_p Multiplier Inequality

Let $W_n = (W_{n1}, \ldots, W_{nn})'$ be non-negative exchangeable random variables on $(\mathcal{W}, \Omega, P_W)$ such that, for every n, $R_n = \int_0^\infty \sqrt{P_W(W_{n1} \geq u)}du < \infty$. Let Z_{ni}, $i = 1, 2, \ldots, n$, be i.i.d. random elements in $(\mathcal{X}^\infty, \mathcal{A}^\infty, P_X^\infty)$ with values in $\ell^\infty(\mathcal{F}_n)$, and write $\|\cdot\|_n = \sup_{f\in\mathcal{F}_n}|Z_{ni}(f)|$. It is assumed that Z_{ni}'s are independent of W_n. Then for any n_0 such that $1 \leq n_0 < \infty$ and any $n > n_0$, the following inequality holds for any $p \geq 1$:

$$\begin{aligned}\left\|\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni}Z_{ni}\right\|_n\right\|_{P_{XW},p} &\leq n_0 \,\|\|Z_{n1}\|_n\|_{P_X,p}\cdot\frac{\|\max_{1\leq i\leq n}W_{ni}\|_{P_W,p}}{\sqrt{n}}\\ &\quad + R_n^{1/p}\cdot\left\|\max_{n_0<i\leq n}\left\|\frac{1}{\sqrt{i}}\sum_{j=n_0+1}^i Z_{nj}\right\|_n\right\|_{P_X,p}. \qquad \text{(A.6)}\end{aligned}$$

Proof: This Lemma generalizes the results in Lemma 4.1 of [29] where $p = 1$. By the triangle inequality, we have

$$\left\|\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni}Z_{ni}\right\|_n\right\|_{P_{XW},p} \leq \left\|\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n_0} W_{ni}Z_{ni}\right\|_n\right\|_{P_{XW},p} + \left\|\left\|\frac{1}{\sqrt{n}}\sum_{i=n_0+1}^{n} W_{ni}Z_{ni}\right\|_n\right\|_{P_{XW},p}.$$

The first term in the above is trivially bounded by

$$n_0\,\|\|Z_{n1}\|_n\|_{P_X,p}\cdot\frac{\|\max_{1\leq i\leq n}W_{ni}\|_{P_W,p}}{\sqrt{n}}.$$

Denote $W_{n(i)}$ as the ith ordered values of W_{ni}, i.e., $W_{n(1)} \geq W_{n(2)} \geq \cdots \geq W_{n(n)}$. Note that $\|\|\sum_{i=1}^n W_{ni}Z_{ni}\|_n\|_{P_{XW},p} = \|\|\sum_{i=1}^n W_{n(i)}Z_{ni}\|_n\|_{P_{XW},p}$ since W_n is assumed to be exchangeable and P_X^∞ is permutation invariant. We write the second term as the following telescoping sum,

$$\frac{1}{\sqrt{n}}\sum_{i=n_0+1}^n W_{n(i)}Z_{ni} = \frac{1}{\sqrt{n}}\sum_{i=n_0+1}^n \sqrt{i}(W_{n(i)} - W_{n(i+1)})T_i,$$

where $T_i \equiv i^{-1/2}\sum_{j=n_0+1}^i Z_{nj}$ and $W_{n(n+1)} \equiv 0$. Thus, we obtain that

$$\begin{aligned}\left\|\left\|\sum_{i=n_0+1}^n W_{n(i)}Z_{ni}\right\|_n\right\|_{P_{XW},p} &\leq \left\|\sum_{i=n_0+1}^n \sqrt{i}(W_{n(i)} - W_{n(i+1)})\|T_i\|_n\right\|_{P_{XW},p}\\ &\leq \left\|\max_{n_0<i\leq n}\|T_i\|_n\right\|_{P_X,p}\cdot\left\|\sum_{i=n_0+1}^n\sqrt{i}(W_{n(i)} - W_{n(i+1)})\right\|_{P_W,p}.\end{aligned}$$

Recalling the definition of T_i, it remains to show

$$E_W\left(\sum_{i=n_0+1}^n \sqrt{i}(W_{n(i)} - W_{n(i+1)})\right)^p \leq n^{p/2}R_n. \qquad \text{(A.7)}$$

By some algebra, the left hand side of (A.7) can be re-written as

$$E_W\left(\sum_{i=n_0+1}^n\int_{W_{n(i+1)}}^{W_{n(i)}}\sqrt{i}du\right)^p = E_W\left(\int_0^{W_{n(n_0+1)}}\sqrt{\sharp\{r\geq 1: W_{n(r)}\geq u\}}du\right)^p,$$

which is bounded by

$$\begin{aligned} E_W \int_0^{W_{n(n_0+1)}} \left\{ \sharp\{r \geq 1 : W_{n(r)} \geq u\} \right\}^{p/2} du &\leq n^{p/2} E_W \int_0^{\infty} \left\{ \sharp\{r \geq 1 : W_{n(r)} \geq u\}/n \right\}^{p/2} du \\ &\leq n^{p/2} E_W \int_0^{\infty} \left\{ \sharp\{r \geq 1 : W_{n(r)} \geq u\}/n \right\}^{1/2} du \\ &\leq n^{p/2} \int_0^{\infty} \sqrt{P_W(W_{n1} \geq u)} du = n^{p/2} R_n \end{aligned}$$

based on the Jensen's inequality. This completes the whole proof. □

A.5. Verification of Condition (11)

Suppose that the $L_{p'}$ maximal inequality (10) and bootstrap weight condition (18) hold. If $\|N_\delta\|_{P_X,p'} < \infty$, then we have Condition (11) for each $p' \geq 1$.

Proof: We first apply the symmetrization argument to show

$$\big\| \|\mathbb{G}_n^*\|_{\mathcal{N}_\delta} \big\|_{P_{XW},p'} \lesssim 2 \left\| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{ni}(\delta_{X_i} - P_X) \right\|_{\mathcal{N}_\delta} \right\|_{P_{XW},p'}. \tag{A.8}$$

Note that

$$\mathbb{G}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - 1)\delta_{X_i} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - 1)(\delta_{X_i} - P_X)$$

by Condition W2. Let $W_n' = (W_{n1}', \ldots, W_{nn}')$ be exchangeable bootstrap weights generated from $P_{W'}$, an independent copy of P_W. The bootstrap weight conditions $W1$ and $W2$ imply that $E_{W'} W_{ni}' = 1$ for $i = 1, \ldots, n$. Then, we have

$$\begin{aligned} E_{XW} \|\mathbb{G}_n^*\|_{\mathcal{N}_\delta}^{p'} &= E_{XW} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - 1)(\delta_{X_i} - P_X) \right\|_{\mathcal{N}_\delta}^{p'} \\ &= E_{XW} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - E_{W'} W_{ni}')(\delta_{X_i} - P_X) \right\|_{\mathcal{N}_\delta}^{p'} \\ &\leq E_{XW} E_{W'} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_{ni} - W_{ni}')(\delta_{X_i} - P_X) \right\|_{\mathcal{N}_\delta}^{p'} \end{aligned}$$

based on the Jensen's inequality and the reverse Fatou's Lemma. In the end, a typical application of the symmetrization argument and Minkowski's inequality concludes (A.8).

To further bound the right hand side of (A.8), we next apply the L_p multiplier inequality (A.6) with $Z_{ni} = (\delta_{X_i} - P_X)$ and $\mathcal{F}_n = \mathcal{N}_\delta$. This gives, due to Condition W3,

$$\big\| \|\mathbb{G}_n^*\|_{\mathcal{N}_\delta} \big\|_{P_{XW},p'} \lesssim \Big\| \|Z_{n1}\|_{\mathcal{N}_\delta} \Big\|_{P_X,p'} \cdot \frac{1}{\sqrt{n}} \left\| \max_{1 \leq i \leq n} W_{ni} \right\|_{P_W,p'} + \left\| \max_{n_0 < i \leq n} \left\| \frac{1}{\sqrt{i}} \sum_{j=n_0+1}^{i} Z_{nj} \right\|_{\mathcal{N}_\delta} \right\|_{P_X,p'}$$

for any $1 \le n_0 < \infty$ and $n > n_0$. For the last term in the above, we can bound it by

$$\begin{aligned}
& \left\| \max_{n_0 < i \le n} \left\| \frac{1}{\sqrt{i}} \sum_{j=1}^{i} Z_{nj} \right\|_{\mathcal{N}_\delta} \right\|_{P_X, p'} + \left\| \left\| \frac{1}{\sqrt{n_0}} \sum_{j=1}^{n_0} Z_{nj} \right\|_{\mathcal{N}_\delta} \right\|_{P_X, p'} \\
\le & \left\| \max_{n_0 < k \le n} \|\mathbb{G}_k\|_{\mathcal{N}_\delta} \right\|_{P_X, p'} + \left\| \|\mathbb{G}_{n_0}\|_{\mathcal{N}_\delta} \right\|_{P_X, p'} \\
\le & \; 2 \left\| \max_{n_0 \le k \le n} \|\mathbb{G}_k\|_{\mathcal{N}_\delta} \right\|_{P_X, p'}
\end{aligned}$$

by the triangular inequality. In addition, we can bound $\|\|Z_{n1}\|_{\mathcal{N}_\delta}\|_{P_X,p'}$ as

$$\|\|Z_{n1}\|_{\mathcal{N}_\delta}\|_{P_X,p'} = \|\|\delta_{X_1} - P_X\|_{\mathcal{N}_\delta}\|_{P_X,p'} \le \|\|\delta_{X_1}\|_{\mathcal{N}_\delta}\|_{P_X,p'} + \|\|P_X\|_{\mathcal{N}_\delta}\|_{P_X,p'} \le 2\|N_\delta\|_{P_X,p'}$$

due to the reverse Fatou's Lemma. Thus, we obtain that

$$\begin{aligned}
\|\|\mathbb{G}_n^*\|_{\mathcal{N}_\delta}\|_{P_{XW},p'} & \lesssim \|N_\delta\|_{P_X,p'} \cdot \frac{1}{\sqrt{n}} \left\| \max_{1 \le i \le n} W_{ni} \right\|_{P_W, p'} + \left\| \max_{n_0 \le k \le n} \|\mathbb{G}_k\|_{\mathcal{N}_\delta} \right\|_{P_X, p'} \\
& \lesssim I + II.
\end{aligned}$$

Considering Condition (18) and Lemma 4.7 of [22], we have $n^{-1/2} E_W(\max_{1\le i\le n} W_{ni}^{p'}) \longrightarrow 0$. The inequality that $\|\max_{1\le i\le n} W_{ni}\|_{P_W,p'} \le E_W(\max_{1\le i\le n} W_{ni}^{p'})$ (due to $\max_{1\le i\le n} W_{ni}^{p'} \ge 1$) implies

$$\frac{1}{\sqrt{n}} \left\| \max_{1 \le i \le n} W_{ni} \right\|_{P_W, p'} = o(1). \tag{A.9}$$

Since $\|N_\delta\|_{P_X,p'}$ is assumed to be finite, the above term I converges to zero, and thus is smaller than arbitrary $\delta > 0$ for sufficiently large n. For any positive r.v. Y, it is easy to prove that

$$EY^q = \int_0^\infty q t^{q-1} P(Y > t) dt \quad \text{for any } q > 0.$$

The Lévy's inequality, i.e., Proposition A.1.2 in [27], implies that

$$P\left(\max_{k \le n} \|\mathbb{G}_k\|_{\mathcal{N}_\delta} > \lambda \right) \le 2P\left(\|\mathbb{G}_n\|_{\mathcal{N}_\delta} > \lambda \right) \quad \text{for every } \lambda > 0.$$

Thus, we have that $II \le 2^{1/p'} \|\|\mathbb{G}_n\|_{\mathcal{N}_\delta}\|_{P_X,p'}$. This concludes the whole proof. □

References

[1] Barbe, P. and Bertail, P. (1995), *The Weighted Bootstrap.* Lecture Notes in Statistics, Vol 98, Springer-Verlag, New York.

[2] Begun, J.M., Hall, W.J., Huang, W.M. and Wellner, J.A. (1983), Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics* 11 432-452.

[3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998) *Efficient and Adaptive Estimation for Semiparametric Models.* Springer-Verlag, New York.

[4] Chatterjee, S. and Bose, A. (2005) Generalized Bootstrap for Estimating Equations. *Annals of Statistics* 33 414-436.

[5] Chen, X. and Pauzo, D. (2009) Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals. *Journal of Econometrics* 152 46-60.

[6] Cheng,G. and Huang, J.Z. (2010). Bootstrap Consistency for General Semiparametric M-estimate. *Annals of Statistics*, **38**, 2884–2915.
[7] Efron, B. and Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, *Statistical Science* 1 54-75.
[8] Hoffmann-Jorgensen, I. (1984) Stochastic Processes on Polish Spaces. Unpublished Manuscript.
[9] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1997) Discrete Multivariate Distributions, Wiley-Interscience.
[10] Johnson, N.L. and Kotz, S. (1977) *Urn Models and Their Applications*. Wiley, New York.
[11] Kato, K. (2011) A Note On Moment Convergence Of Bootstrap M-Estimators. *Statistics & Decisions* 28 51-61.
[12] Kosorok, M., Lee, B.L. and Fine, J.P. (2004) Robust Inference for Univariate Proportional Hazards Frailty Regression Models. *Annals of Statistics* 32 1448-1491.
[13] Kosorok, M. (2008a) *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.
[14] Lee, B.L., Kosorok, M.R. and Fine, J.P. (2005) The profile sampler, *Journal of the American Statistical Association* 100 960-969.
[15] Lo, A.Y. (1993) A Bayesian bootstrap for Censored data, *Annals of Statistics* 21 100-123.
[16] Mammen, E. and van de Geer, S. (1997) Penalized Quasi-Likelihood Estimation in Partial Linear Models *Annals of Statistics* 25 1014-1035
[17] Ma, S. and Kosorok, M. (2005a) Robust Semiparametric M-estimation and the Weighted Bootstrap, *Journal of Multivariate Analysis* 96 190-217.
[18] Mason, D. and Newton, M. (1992) A rank statistic approach to the consistency of a general bootstrap, *Annals of Statistics* 20 1611-1624.
[19] Murphy, S.A. and van der Vaart, A.W. (1999) Observed Information in Semiparametric Models, *Bernoulli* 5 381-412.
[20] Nishiyama, Y. (2010) Moment Convergence of M-Estimators, *Statistica Neerlandica* 64 505-507.
[21] Politis, D.N. and Romano, J.P. (1994). Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions. *Annals of Statistics*, **22**, 2031–2050.
[22] Praestgaard, J. and Wellner, J. (1993) Exchangeably Weighted Bootstraps of the General Empirical Process, *Annals of Probability* 21 2053-2086.
[23] Roeder, K., Carroll, R. J., and Lindsay, B. G. (1996) A Semiparametric Mixture Approach to Case-Control Studies with Errors in Covariables, *Journal of the American Statistical Association* 91 722-732.
[24] Rubin, D. (1981) The Bayesian bootstrap, *Annals of Statistics* 9 130-134.
[25] Shao, J. and Tu, D. (1996). The Jackknife and Bootstrap. Springer, New York.
[26] van der Vaart, A. W., and Wellner, J. A. (2011) A Local Maximal Inequality under Uniform Entropy. *Electronic Journal of Statistics*. 5 192-203.
[27] van der Vaart, A. W., and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.
[28] van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge.
[29] Wellner, J.A. and Zhan, Y. (1996) Bootstrapping Z-estimators. Technical report 308, University of Washington.
[30] Wellner, J.A. and Zhang, Y. (2007) Two Likelihood-Based Semiparametric Estimation Methods for Panel Count Data with Covariates. *Annals of Statistics* 35 2106-2142.
[31] Wu, C.F.J. (1987) On the asymptotic property of the jackknife histogram. Technical report, Dept. of Statistics, Univ. of Wisconsin, Madison.