RANDOM WEIGHTING APPROXIMATION OF M-ESTIMATORS WITH INCREASING DIMENSIONS OF PARAMETER

Ruixing Ming[†,1], Chengyao Yu[†,1], Min Xiao[1] and Zhangfeng Wang[*,2]

[1] *Zhejiang Gongshang University*

and [2] *University of Science and Technology of China*

Supplementary Material

Supplementary material includes proofs of relevant theorems, propositions, corollaries, and simulation details.

S1 Preliminaries

First, we introduce an important lemma based on which our theorems are obtained. Let $R_n = (m_n \log(n + m_n))^{1/2}$.

Lemma S.1. *Let $\{v_i(t), t \in R^{m_n}\}$, $1 \le i \le n$ be independent R^{p_n} valued random variables with $E(v_i(t)) = 0$ for all t. Assume that there exist $r_1 > 0$*

[†]Co-first authors and contributed equally to this work.

[*]Corresponding author.

and $r_2 > 0$ *such that for every* $s \in R^{m_n}$, $0 < d \le 1$, $1 \le i \le n$,

$$E\left(\sup_{t:||t-s||\le d}||v_i(t)-v_i(s)||\right) \le n^{r_1}d^{r_2}.$$

Define

$$B_n(t,s) = \left(\sum_{i=1}^{n} E||v_i(t)-v_i(s)||^2\right)^{1/2}, \quad V_n(t,s) = \left(\sum_{i=1}^{n} ||v_i(t)-v_i(s)||^2\right)^{1/2}.$$

Then

$$\sup_{||t||\le n^{r_3},||s||\le n^{r_3}} \frac{||\sum_{i=1}^{n}(v_i(t)-v_i(s))||}{n^{-2}+B_n(t,s)+V_n(t,s)} = O_p(R_n), \tag{S1.1}$$

for every $r_3 \ge 0$.

Proof. The proof of Lemma S.1 can be seen in that of Lemma 3.2 in He and Shao (2000). □

Lemma S.2. *Under the assumption (A2), we have*

$$\sup_{||\tau||\le n^{r_3},||\theta||\le n^{r_3}} \frac{||\sum_{i=1}^{n}\eta_i^*(\tau,\theta)||}{n^{-2}+(\sum_{i=1}^{n}E||\eta_i^*(\tau,\theta)||^2)^{1/2}+\sum_{i=1}^{n}||\eta_i^*(\tau,\theta)||^2)^{1/2}} = O_p(R_n),$$

$$\sup_{||\tau||\le n^{r_3},||\theta||\le n^{r_3}} \frac{||\sum_{i=1}^{n}(\tau-\theta)^\top\eta_i^*(\tau,\theta)||}{n^{-2}+(\sum_{i=1}^{n}E||(\tau-\theta)^\top\eta_i^*(\tau,\theta)||^2)^{1/2}+\sum_{i=1}^{n}||(\tau-\theta)^\top\eta_i^*(\tau,\theta)||^2)^{1/2}}$$
$$= O_p(R_n),$$

and

$$\sup_{||\alpha||\le n^{r_3},||\tau||\le n^{r_3},||\theta||\le n^{r_3}} \frac{||\sum_{i=1}^{n}\alpha^\top\eta_i^*(\tau,\theta)||}{n^{-2}+(\sum_{i=1}^{n}E||\alpha^\top\eta_i^*(\tau,\theta)||^2)^{1/2}+\sum_{i=1}^{n}||\alpha^\top\eta_i^*(\tau,\theta)||^2)^{1/2}}$$
$$= O_p(R_n),$$

for every $r_3 \geq 0$, *where* $\eta_i^*(\tau, \theta) = w_i\psi(z_i, \tau) - w_i\psi(z_i, \theta) - E(\psi(z_i, \tau)) + E(\psi(z_i, \theta))$.

Proof. From the assumption (A2) and the boundedness of the weight variable w_i, it follows that there exists constant c and $r \in (0, 2]$ such that for $0 < d \leq 1$,

$$\max_{i \leq n} E_\theta \left(\sup_{\tau:||\tau-\theta|| \leq d} ||\eta_i^*(\tau, \theta)||^2 \right) \leq n^c d^r.$$

Hence, $v_i(t) = w_i\psi(z_i, t) - E(\psi(z_i, t))$ satisfies the condition of Lemma S.1. Consequently, the results of Lemma S.2 are easily obtained from Lemma S.1.

□

S2 Proof of Theorems and Propositions

In this section, we present the proofs of Theorems 1–3 and Proposition 2.

S2.1 Proof of Theorem 1

Since $\rho(z, \theta)$ is a convex function with respect to θ for given z, then $G_n^*(\theta)$ is also convex on θ. Hence, we only need to demonstrate that for any $\epsilon > 0$, there exists a constant $B < \infty$ such that

$$P\left(\inf_{||\beta||=1} \sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0 + B(m_n/n)^{1/2}\beta) > 0 \right) > 1 - \epsilon$$

for sufficiently large n. From Lemma S.2 together with assumptions (R1) and (A4)-(A6), we have

$$\sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0 + B(m_n/n)^{1/2}\beta)$$
$$= \sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0) + B(m_n n)^{1/2} \beta^\top D_n \beta + o(n^{1/2}) + O_p((A(n, m_n) m_n \log n)^{1/2})$$

uniformly in $\beta \in S_{m_n}$. Due to $A(n, m_n) = o(n/\log n)$, for sufficiently large n, we have

$$P\left(\inf_{||\beta||=1} \sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0 + B(m_n/n)^{1/2}\beta) > 0 \right)$$
$$\geq P\left((m_n n)^{-1/2} \inf_{||\beta||=1} \sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0) > -(B/2)\lambda_{\min}(D_n) \right).$$

By the assumption (A3) and the boundedness of the weight variable, we have

$$P\left((m_n n)^{-1/2} \inf_{||\beta||=1} \sum_{i=1}^{n} w_i \beta^\top \psi(z_i, \theta_0) > -(B/2)\lambda_{\min}(D_n) \right) \to 1, \quad \text{as } n \to \infty.$$

Hence, the proof of Theorem 1 is completed.

S2.2 Proof of Theorem 2

For any α satisfying $||\alpha|| = O(1)$ as $m_n \to \infty$, we have from Lemma S.2 together with assumptions (R1) and (A4)-(A6) that

$$\sum_{i=1}^{n} w_i \alpha^\top \psi(z_i, \theta_0) + n\alpha^\top D_n(\theta_n^* - \theta_0) = o_p(n^{1/2}) + O_p((A(n, m_n) m_n \log n)^{1/2}).$$

If $A(n, m_n) = o(n/(m_n \log n))$, the right of equation has order $o_p(n^{1/2})$. Since $\liminf_{n\to\infty} \lambda_{\min}(D_n) > 0$, we have the result of Theorem 2 by replacing α with $D_n^{-1}\alpha$.

S2.3 Proof of Theorem 3

We only need to prove that Theorem 3 is valid for $\alpha \in S_{m_n}$. We have from Theorem 2 that for any $\alpha \in S_{m_n}$,

$$\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \alpha^\top D_n^{-1}\psi(z_i, \theta_0)\frac{(w_i - 1)}{\sigma} + \sqrt{n}\alpha^\top r_n, \quad \text{(S2.2)}$$

$$\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \alpha^\top D_n^{-1}\psi(z_i, \theta_0) + \sqrt{n}\alpha^\top \xi_n, \quad \text{(S2.3)}$$

where $||r_n|| = o_p(n^{-1/2})$ and $||\xi_n|| = o_p(n^{-1/2})$. According to assumption (A7), the variable $\psi(z_i, \theta_0)$ has mean zero. We have

$$\sigma_n^{*2} = \mathrm{Var}^*\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n \alpha^\top D_n^{-1}\psi(z_i, \theta_0)(1 - w_i)\right) = \frac{1}{n}\sum_{i=1}^n \left(\alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right)^2,$$

$$\sigma_n^2 = \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right) = \frac{1}{n}\sum_{i=1}^n E\left(\alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right)^2.$$

First, we prove that $\sigma_n^{*2} - \sigma_n^2 \xrightarrow{\text{a.s.}} 0$. We have

$$\sigma_n^{*2} - \sigma_n^2 = \frac{1}{n}\sum_{i=1}^n \left[\left(\alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right)^2 - E\left(\alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right)^2\right].$$

By assumption (A4), there exists a constant $0 < C < \infty$ such that

$$||\alpha^\top D_n^{-1}|| \le \sqrt{\lambda_{\max}((D_n D_n^\top)^{-1})}||\alpha|| = \frac{1}{\sqrt{\lambda_{\min}(D_n D_n^\top)}} < C \quad \text{(S2.4)}$$

for all n. By (S2.4) and Abel's summation formula, we have

$$
\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^{n}\frac{\mathrm{Var}\left(\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right)^2}{i^2}\\
\le&\lim_{n\to\infty}\sum_{i=1}^{n}\frac{E\left|\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right|^4}{i^2}\\
=&\lim_{n\to\infty}\frac{1}{(n+1)^2}\sum_{i=1}^{n}E\left|\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right|^4\\
&+\lim_{n\to\infty}\sum_{k=1}^{n-1}\left(\frac{1}{k^2}-\frac{1}{(k+1)^2}\right)\sum_{j=1}^{k}E\left|\alpha^\top D_n^{-1}\psi(z_j,\theta_0)\right|^4\\
\le&\, C\lim_{n\to\infty}\frac{\sup_{\alpha_0\in S_{m_n}}\sum_{i=1}^{n}E\left|\alpha_0^\top\psi(z_i,\theta_0)\right|^4}{(n+1)^2}\\
&+C\lim_{n\to\infty}\sum_{k=1}^{n-1}\left(\frac{1}{k^2}-\frac{1}{(k+1)^2}\right)\sup_{\alpha_0\in S_{m_n}}\sum_{j=1}^{k}E\left|\alpha_0^\top\psi(z_j,\theta_0)\right|^4\\
\le&\lim_{n\to\infty}\frac{Cc_2n^\delta}{(n+1)^2}+\lim_{n\to\infty}\sum_{k=1}^{n-1}Cc_2k^\delta\left(\frac{1}{k^2}-\frac{1}{(k+1)^2}\right)\\
\le&\lim_{n\to\infty}\frac{Cc_2n^\delta}{(n+1)^2}+\lim_{n\to\infty}\sum_{k=1}^{n-1}Cc_2\frac{2k^{\delta+1}+k^\delta}{k^4+2k^3+k^2}\\
<&\,\infty.
\end{aligned}
$$

By the strong law of large numbers, we have $\sigma_n^{*2}-\sigma_n^2\xrightarrow{\text{a.s.}}0$.

Second, we prove

$$\mathcal{L}(\sqrt{n}\alpha^\top(\hat{\theta}_n-\theta_0)/\sigma_n)\to N(0,1), \tag{S2.5}$$

and

$$\mathcal{L}^*(\sqrt{n}\alpha^\top(\theta_n^*-\widehat{\theta}_n)/(\sigma\sigma_n^*))\to N(0,1), \quad \text{in pr.} \tag{S2.6}$$

To proof equation (S2.5), on one hand, we have $\sqrt{n}\alpha^\top\xi_n/\sigma_n=o_p(1)$ by (S2.3) and (A7) since

$$\liminf_{n\to\infty} \sigma_n^2 = \liminf_{n\to\infty} \alpha^\top D_n^{-1}\Sigma_n(D_n^\top)^{-1}\alpha \geq \liminf_{n\to\infty} \lambda_{\min}(D_n^{-1}\Sigma_n(D_n^\top)^{-1}) > 0. \tag{S2.7}$$

On the other hand, by Doob's inequality (Stout, 1974) and (A7), for any $\epsilon > 0$,

$$\begin{aligned}
& P\left(\max_{1\leq i\leq n}\left|\frac{1}{n}(\alpha^\top D_n^{-1}\psi(z_i,\theta_0))^2 - \frac{1}{n}\sigma_n^2\right| \geq 2\epsilon\right) \\
\leq\; & 2P\left(\max_{1\leq k\leq n}\left|\frac{1}{n}\sum_{i=1}^{k}(\alpha^\top D_n^{-1}\psi(z_i,\theta_0))^2 - \sigma_n^2\right| \geq \epsilon\right) \\
\leq\; & 2\epsilon^{-2}E\left|\frac{1}{n}\sum_{i=1}^{n}(\alpha^\top D_n^{-1}\psi(z_i,\theta_0))^2 - \sigma_n^2\right|^2 \\
=\; & 2\epsilon^{-2}n^{-2}\sum_{i=1}^{n}\mathrm{Var}\left(\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right)^2 \\
=\; & O(n^{\delta-2}).
\end{aligned} \tag{S2.8}$$

This implies that

$$\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{n}}\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right| \xrightarrow{p} 0, \text{ as } n\to\infty,$$

and further verifies the Lindeberg condition

$$\sum_{i=1}^{n} E\frac{1}{\sigma_n^2 n}(\alpha^\top D_n^{-1}\psi(z_i,\theta_0))^2 I\left(\left|\frac{1}{\sigma_n\sqrt{n}}\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right| \geq \epsilon\right) \to 0, \text{ as } n\to\infty.$$

By the central limit theorem and Slutsky's theorem, we complete the proof of (S2.5). By (S2.2), we have

$$\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/(\sigma\sigma_n^*) = -\frac{1}{\sigma_n^*\sqrt{n}}\sum_{i=1}^{n}\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\frac{(w_i-1)}{\sigma} + \frac{\sqrt{n}\alpha^\top r_n}{\sigma\sigma_n^*}.$$

Write

$$Z_n = -\frac{1}{\sigma_n^*\sqrt{n}} \sum_{i=1}^{n} \alpha^\top D_n^{-1} \psi(z_i, \theta_0) \frac{(w_i - 1)}{\sigma} \tag{S2.9}$$

and

$$\zeta_{in} = -\frac{1}{\sigma_n^*\sqrt{n}} \alpha^\top D_n^{-1} \psi(z_i, \theta_0) \frac{(w_i - 1)}{\sigma}. \tag{S2.10}$$

By Theorem 2 and the almost sure convergence, we have

$$E^* \left(\frac{\sqrt{n}\alpha^\top r_n}{\sigma\sigma_n^*} \right) \to 0 \quad \text{in pr. as } n \to \infty. \tag{S2.11}$$

As shown by Rao and Zhao (1992), in order to prove (S2.6), it suffices to prove

$$\mathcal{L}^*(Z_n) \to N(0,1) \quad \text{in pr. as } n \to \infty. \tag{S2.12}$$

In view of (S2.10), it is enough to demonstrate that for any $\epsilon > 0$, we have

$$L_n(\epsilon) \equiv \sum_{i=1}^{n} \frac{1}{n\sigma^2\sigma_n^{*2}} \left(\alpha^\top D_n^{-1} \psi(z_i, \theta_0) \right)^2 \mathbb{E}^* \left((w_i - 1)^2 I\left(|\zeta_{in}| \geq \epsilon \right) \right) \xrightarrow{p} 0, \tag{S2.13}$$

as $n \to \infty$. Write

$$T_n = \max_{1 \leq i \leq n} \left| \frac{1}{\sigma_n^*\sqrt{n}} \alpha^\top D_n^{-1} \psi(z_i, \theta_0) \right| = \frac{1}{\sigma_n^*} \max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{n}} \alpha^\top D_n^{-1} \psi(z_i, \theta_0) \right|.$$

On one hand, according to (S2.7) and the almost surely convergence of $\sigma_n^{*2} - \sigma_n^2$, the term σ_n^{*2} is also bounded away from zero almost surely. On the other hand, we have $\max_{1 \leq i \leq n} \left| \frac{1}{\sqrt{n}} \alpha^\top D_n^{-1} \psi(z_i, \theta_0) \right| \xrightarrow{p} 0$. Thus, $T_n \xrightarrow{p} 0$

as $n \to \infty$. Then we have as $n \to \infty$,

$$L_n(\epsilon) \leq \sum_{i=1}^{n} \frac{1}{n\sigma^2\sigma_n^{*2}} \left(\alpha^\top D_n^{-1}\psi(z_i, \theta_0)\right)^2 \mathbb{E}^* \left((w_i - 1)^2 I \left(T_n|w_1 - 1| \geq \epsilon\right)\right) \xrightarrow{p} 0.$$

This completes the proof of (S2.13). By the central limit theorem and (S2.11), the proof of (S2.6) is completed. By a routine argument (Rao and Zhao, 1992), we can derive from (S2.5) and (S2.6) that as $n \to \infty$,

$$\sup_u |P^*(\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/(\sigma\sigma_n^*) \leq u) - P(\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)/\sigma_n \leq u)| \xrightarrow{p} 0.$$

Finally, by $\sigma_n^{*2} - \sigma_n^2 \xrightarrow{\text{a.s.}} 0$, we have

$$\sup_u |P^*(\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma \leq u) - P(\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0) \leq u)| \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

S2.4 Proof of Proposition 2

For the common least-squares estimator, we have $\hat{\theta}_n = (\sum_{i=1}^n x_i x_i^\top)^{-1} \sum_{j=1}^n x_j y_j$. Therefore,

$$\operatorname{var}(\alpha^\top\hat{\theta}_n) = E\left(\operatorname{var}(\alpha^\top\hat{\theta}_n|\{z_i\}_{i=1}^n)\right) + 0 = \sigma_\epsilon^2 \alpha^\top E\left(\left(\sum_{i=1}^n x_i x_i^\top\right)^{-1}\right)\alpha.$$

It is well known that $\sum_{i=1}^n x_i x_i^\top$ follows a Wishart distribution and satisfies $E((\sum_{i=1}^n x_i x_i^\top)^{-1}) = \Sigma^{-1}/(n - m_n - 1)$. Therefore, we have

$$\frac{m_n \operatorname{var}(\alpha^\top\hat{\theta}_n)}{\alpha^\top\Sigma^{-1}\alpha} = \sigma_\epsilon^2 \frac{m_n}{n - m_n - 1} = \sigma_\epsilon^2 \frac{\gamma_n}{1 - \gamma_n - 1/n}.$$

As demonstrated in El Karoui and Purdom (2018), to prove the second equation in the Proposition 2, it is without loss of generality to study $\Sigma =$

I_{m_n} and $\theta_0 = \mathbf{0}$. In this case, we have $\alpha^\top \Sigma^{-1} \alpha = \|\alpha\|_2^2 = 1$. According to the proof of Theorem 2 in El Karoui and Purdom (2018) (see Page 43-48), it can be derived that

$$\sigma_\epsilon^2 \left[\gamma_n \frac{\operatorname{trace}\left(\widehat{\Sigma}_w^{-2}\right)/n}{[\operatorname{trace}\left(\widehat{\Sigma}_w^{-1}\right)/n]^2} - \frac{1}{1-\gamma_n} \right] = m_n E(\operatorname{var}(\alpha^\top \theta_n^* | \{z_i\}_{i=1}^n)) + o_p(1), \tag{S2.14}$$

$$\frac{\operatorname{trace}\left(\widehat{\Sigma}_w^{-2}\right)/n}{\left(\operatorname{trace}\left(\widehat{\Sigma}_w^{-1}\right)/n\right)^2} = \frac{1}{bE_{w_i}\left(\frac{w_i}{(1+bw_i)^2}\right)} + o_p(1), \tag{S2.15}$$

where $\hat{\Sigma}_w = \sum_{i=1}^n w_i z_i z_i^\top$, $\hat{\Sigma}_w^{-2} = (\hat{\Sigma}_w^{-1})^2$, and b is the unique solution of $\mathbf{E}_{w_i}\left(\frac{1}{1+bw_i}\right) = 1 - \gamma_n$. We have

$$bE_{w_i}\left(\frac{w_i}{(1+bw_i)^2}\right) = 1 - \gamma_n - E_{w_i}\left(\frac{1}{(1+bw_i)^2}\right). \tag{S2.16}$$

According to equation (S2.14), (S2.15), (S2.16), we have

$$m_n \frac{E\left(\operatorname{var}\left(\alpha^\top \theta_n^* / \sigma | \{z_i\}_{i=1}^n\right)\right)}{\alpha^\top \Sigma^{-1} \alpha} = \frac{\sigma_\epsilon^2}{\sigma^2} \left[\frac{\gamma_n}{1 - \gamma_n - E\left(\frac{1}{(1+bw_i)^2}\right)} - \frac{1}{1-\gamma_n} \right] + o(1).$$

S2.5 Derivation of Equation (2.3)

Let

$$m_n \frac{\operatorname{var}(\alpha^\top \hat{\theta}_n)}{\alpha^\top \Sigma^{-1} \alpha} = \frac{\sigma_\epsilon^2}{\sigma^2} \left[\frac{\gamma_n}{1 - \gamma_n - f(\gamma_n)} - \frac{1}{1-\gamma_n} \right].$$

By straightforward caculations, we have

$$f(\gamma_n) = \frac{(1-\gamma_n)^2(\sigma^2\gamma_n - 1/n) + (1-\gamma_n)^3}{(1-\gamma_n)(\sigma^2\gamma_n + 1) - 1/n}.$$

By ignoring the term of $1/n$, we have

$$E(\frac{1}{(1+bw_i)^2}) = \frac{(1-\sigma^2)\gamma_n^2 + (\sigma^2-2)\gamma_n + 1}{1+\gamma_n\sigma^2}.$$

This completes the derivation.

S3 Proof of Corollaries

S3.1 Proof of Corollary 1

For smooth scores, by Taylor's expansion and Dominated Convergence Theorem, we have

$$\begin{aligned}
&\left|\alpha^\top \sum_{i=1}^n E_{\theta_0}(\psi(z_i,\theta) - \psi(z_i,\theta_0)) - \alpha^\top E(\phi^{'}(\epsilon_i)) \sum_{i=1}^n x_i x_i^\top (\theta-\theta_0)\right| \\
&= \left|\alpha^\top \sum_{i=1}^n \frac{1}{2}(\theta-\theta_0)^\top E(\phi^{''}(\epsilon_i)) x_i x_i^\top (\theta-\theta_0) x_i\right| \\
&\leq \frac{1}{2}\sup_r |\phi^{''}(r)| \sum_{i=1}^n |x_i^\top(\theta-\theta_0)|^2 |\alpha^\top x_i| \\
&= O(\sum_{i=1}^n |x_i^\top(\theta-\theta_0)|^2 |\alpha^\top x_i|)
\end{aligned}$$

by (B2) for any $\alpha \in S_m$. For jump scores, we also have

$$\begin{aligned}
&\left|\alpha^\top \sum_{i=1}^n E_{\theta_0}(\psi(z_i,\theta) - \psi(z_i,\theta_0)) + \alpha^\top \int_{-\infty}^{\infty} \phi(r) f'(r)\ dr \sum_{i=1}^n x_i x_i^\top (\theta-\theta_0)\right| \\
&= O(\sum_{i=1}^n |x_i^\top(\theta-\theta_0)|^2 |\alpha^\top x_i|).
\end{aligned}$$

Further, by conditions (B3) and Young inequality,

$$O\left(\sum_{i=1}^n |x_i^\top(\theta-\theta_0)|^2 |\alpha^\top x_i|\right) \leq \frac{B^2 m_n}{2n} O\left(\sum_{i=1}^n |x_i^\top \alpha|^4 + |x_i^\top \gamma|^2\right) = O(m_n).$$

Therefore, if $m_n^2 = o(n)$, then assumption (A4) holds. To verify $\|\theta_n^* - \theta_0\|^2 = O_p(m_n/n)$, the error bound of (A4) can be relaxed to $o((nm)^{1/2})$, therefore, $m = o(n)$ is sufficient. By He and Shao (2000), if $m_n(\log m_n)^3/n \to 0$, then conditions (B2)-(B3) imply assumptions (A1)-(A3) and (A5)-(A6) with $A(n, m_n) = o(n/\log n)$. Furthermore, if $m_n^2 \log m_n/n \to 0$ for smooth scores or $m_n^3(\log m_n)^2/n \to 0$ for jump scores, then assumptions (A1)-(A6) are satisfied with $A(n, m_n) = o(n/(m_n \log n))$. Thus, we have $||\theta_n^* - \theta_0||^2 = O_p(m_n/n)$ if $m_n(\log m_n)^3/n \to 0$. In addition, by condition (B3), we have

$$\sup_{\alpha \in S_{m_n}} \sum_{i=1}^{n} |x_i^\top \alpha|^4 \leq \sup_{\alpha, \beta \in S_{m_n}} \sum_{i=1}^{n} |x_i^\top \alpha|^2 |x_i^\top \beta|^2 = O(n).$$

By (B2), $E|\phi(\epsilon_i)|^4$ is bounded for both smooth and jump scores. Therefore, for both types of scores, we have

$$\frac{1}{n} \sum_{i=1}^{n} E|\alpha^\top \psi(z_i, \theta_0)|^4 = \frac{1}{n} \sum_{i=1}^{n} \left|\alpha^\top x_i\right|^4 E|\phi(\epsilon_i)|^4 = O(1).$$

We have proved $D_n = n^{-1} E(\phi^{'}(\epsilon_i)) \sum_{i=1}^n x_i x_i^\top$ for smooth scores and $D_n = n^{-1} \int_{-\infty}^{\infty} \phi(r) f'(r)\, dr \sum_{i=1}^n x_i x_i^\top$ for jump scores. By conditions (B2)-(B3) and Cauchy–Schwarz inequality, we have $\lambda_{\max}(D_n) = O(1)$ for both scores. By (B1), we have

$$n^{-1} \sum_{i=1}^{n} E\left|\alpha^\top \psi(z_i, \theta_0)\right|^2 \geq \lambda_{\min}\left(n^{-1} \sum_{i=1}^{n} x_i x_i^\top\right) E|\phi(\epsilon_n)|^2 > 0.$$

Therefore, (A7) is verified. This completes the proof of Corollary 1.

S3.2 Proof of Corollary 2

The derivative of $\rho(z,\theta)$ is

$$\psi(z,\theta) = \left(\frac{e^{x^\top\theta}}{1+e^{x^\top\theta}} - y\right)x.$$

Thus, we have $E(\psi(z_i,\theta_0)) = 0$. Assumption (A1) is true since ρ is differentiable in θ. By (C1), we have $\sup_{\alpha\in S_{m_n}}\sum_{i=1}^n|\alpha^\top x_i|^4 = O(n)$. This implies

$$\sum_{i=1}^n E|\alpha^\top\psi(z_i,\theta_0)|^4 = \sum_{i=1}^n E\left|\alpha^\top x_i\left(\frac{e^{x_i^\top\theta_0}}{1+e^{x_i^\top\theta_0}} - y_i\right)\right|^4 \le \sum_{i=1}^n\left|\alpha^\top x_i\right|^4 = O(n).$$

By Cauchy-Schwarz Inequality, we have

$$\alpha^\top D_n\alpha \le n^{-1}\sum_{i=1}^n\left|\alpha^\top x_i\right|^2 \le n^{-1/2}\sqrt{\sup_{\alpha,\beta\in S_m}\sum_{i=1}^n|\alpha^\top x_i|^2|\beta^\top x_i|^2} = O(1).$$

This implies that $\lambda_{\max}(D_n)$ has a finite upper bound. By (C2),

$$\alpha^\top D_n\alpha = n^{-1}\sum_{i=1}^n\frac{e^{x_i^\top\theta_0}}{1+e^{x_i^\top\theta_0}}\left|\alpha^\top x_i\right|^2 \ge \lambda_{\min}(D_n) > 0.$$

Together with (C1), the term

$$n^{-1}\sum_{i=1}^n E\left|\alpha^\top\psi(z_i,\theta_0)\right|^2 = n^{-1}\sum_{i=1}^n\frac{e^{x_i^\top\theta_0}}{1+e^{x_i^\top\theta_0}}\left(1-\frac{e^{x_i^\top\theta_0}}{1+e^{x_i^\top\theta_0}}\right)\left|\alpha^\top x_i\right|^2$$

is also bounded away from zero. Thus, assumption (A7) is verified. According to the result of He and Shao (2000), assumptions (A2), (A5) and (A6) are satisfied with $r=2$ and $A(n,m_n)=m_n$ by (C1), and assumption (A4) is implied by (C1) and (C2). Assumption (A2) is straightforward by

(C1). Thus, we have $||\theta_n^* - \theta_0||^2 = O_p(m_n/n)$ by $m_n \log m_n/n \to 0$, and we have equation (2.2) by $m_n^2 \log m_n/n \to 0$. This completes the proof of Corollary 2.

S3.3 Proof of Corollary 3

The loss function $\rho(z,\theta) = ||z-\theta||$ is convex and its derivative is $\psi(z,\theta) = -(z-\theta)/||z-\theta||$ for $\theta \neq z$. For $z = \theta$, the subdifferential set is

$$\partial_\theta \rho(z,\theta) = \partial_\theta \|z-\theta\| = \{u \in \mathbb{R}^{m_n} : \|u\| \leq 1\}.$$

We consider $\psi(z,\theta) = 0$ for $z = \theta$. Therefore, we have

$$\left\| \sum_{i=1}^n \psi(z_i, \hat{\theta}_n) \right\| = o(n^{1/2}),$$

which verifies assumption (A1). To verify (A3), we have

$$E\left(\left\| \sum_{i=1}^n \psi(z_i,\theta_0) \right\|^2 \right) = E\left(\sum_{i=1}^n \sum_{j=1}^n \psi(z_i,\theta_0)^\top \psi(z_j,\theta_0) \right) = n,$$

which further leads to

$$P\left(\left\| \sum_{i=1}^n \psi(z_i,\theta_0) \right\| \geq C(nm_n)^{1/2} \right) \leq \frac{n}{C^2 n m_n} = \frac{1}{m_n},$$

by Chebyshev inequality. Since

$$\sup_{\tau:\|\tau-\theta\|\leqslant d} \|\psi(z,\tau) - \psi(z,\theta)\| \leqslant 2d/\|z-\theta\|,$$

we have assumption (A2) holds with $r = 2$, and assumptions (A5)-(A6) hold with $A(n, m_n) = m_n$. Assumption (A4) holds with $D = D_n = E[||z -$

$\theta_0||^2 I - (z-\theta_0)(z-\theta_0)^\top/||z-\theta_0||^3]$ if $m_n^2/n \to 0$, see He and Shao (2000). By (D1), $E_{\theta_0}(1/||z-\theta_0||^2) < \infty$ implies that $E_{\theta_0}(1/||z-\theta_0||) < \infty$. We have

$$\lambda_{\max}(D_n) = \max_{\alpha \in S_{m_n}} \alpha^\top E\left[\frac{||z-\theta_0||^2 I - (z-\theta_0)(z-\theta_0)^\top}{||z-\theta_0||^3}\right]\alpha \le E||z-\theta_0||^{-1}.$$

We also have

$$\sum_{i=1}^n E|\alpha^\top \psi(z_i,\theta_0)|^4 = \sum_{i=1}^n |\alpha^\top (z_i-\theta_0)|^4/||z_i-\theta_0||^4 \le n.$$

Meanwhile, by (D3), the term $\sum_{i=1}^n E|\alpha^\top\psi(z_i,\theta_0)|^2/n$ is bounded away from zero for any $\alpha \in S_{m_n}$ and any n. Thus, assumption (A7) holds. This completes the proof of Corollary 3.

S4 Simulation Details

S4.1 The reversed percentile method

In this subsection, we show how the RW method constructs confidence intervals for $\alpha^\top\theta_0$ via the reversed percentile method for any α with a bounded norm.

Given samples $z_1,\ldots,z_n$ and a random weight distribution with $\sigma^2>0$, we compute the RW estimator B times. Specifically, for $b\in\{1,\ldots,B\}$, we generate weights $\{w_i^b\}_{i=1}^n$ and compute $\theta_n^{*b} = \arg\min_{\theta\in R^{m_n}} \sum_{i=1}^n w_i^b \rho(z_i,\theta)$.

The β-th quantile of $\{\alpha^\top \theta_n^{*b}/\sigma\}_{b=1}^B$ is defined by

$$\hat{Q}_\beta = \inf\left\{ t \in \{\alpha^\top \theta_n^{*1}, \dots, \alpha^\top \theta_n^{*B}\} : \frac{1}{B}\sum_{b=1}^{B} 1(\alpha^\top \theta_n^{*b}/\sigma \le t) \ge \beta \right\}.$$

Let the target coverage level be $1-\beta_0$. The reversed percentile method constructs confidence interval for $\alpha^\top \theta_0$ by

$$C_{\beta_0} = [(1+1/\sigma)\alpha^\top \hat{\theta}_n - \hat{Q}_{1-\beta_0/2}, (1+1/\sigma)\alpha^\top \hat{\theta}_n - \hat{Q}_{\beta_0/2}]. \tag{S4.17}$$

Proposition S.1. *Under the same conditions of Theorem 3, for any* $\beta_0 \in (0,1)$*, we have*

$$\lim_{n\to\infty, B\to\infty} P(\alpha^\top \theta_0 \in C_{\beta_0}) = 1-\beta_0,$$

where C_{β_0} *is defined in equation (S4.17).*

Proof. Rewrite C_{β_0} by

$$\left[\alpha^\top \hat{\theta}_n - \frac{\sqrt{n}(\hat{Q}_{1-\beta_0/2} - \alpha^\top \hat{\theta}_n/\sigma)}{\sqrt{n}}, \alpha^\top \hat{\theta}_n - \frac{\sqrt{n}(\hat{Q}_{\beta_0/2} - \alpha^\top \hat{\theta}_n/\sigma)}{\sqrt{n}} \right].$$

Let $\tilde{Q}_\beta$ denote the β-th quantile of the conditional distribution of $\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ given $\{z_i\}_{i=1}^n$. By Glivenko-Cantelli theorem, for any fixed $n > 0$ and $\beta \in (0,1)$, we have $\sqrt{n}(\hat{Q}_\beta - \alpha^\top \hat{\theta}_n/\sigma) - \tilde{Q}_\beta \to 0$ as $B \to \infty$ almost surely. Let Q_β denote the β-th quantile of $\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)$. We have

$$P\left(\alpha^\top \theta_0 \in \left[\alpha^\top \hat{\theta} - \frac{Q_{1-\beta_0/2}}{\sqrt{n}}, \alpha^\top \hat{\theta} - \frac{Q_{\beta_0/2}}{\sqrt{n}} \right] \right) = 1-\beta_0.$$

By Theorem 3 and the Continuous Mapping Theorem, we have $\tilde{Q}_\beta - Q_\beta \xrightarrow{p} 0$. This completes the proof. $\square$

Table S.1: MASDE for the spatial median.

m_n	Average SD	Exp	Gamma	Gamma2	Gamma3	Pair
36	2.820	0.043	0.035	0.098	0.116	0.049
47	2.686	0.027	0.025	0.047	0.060	0.031
55	2.682	0.051	0.047	0.019	0.074	0.050

Table S.2: MACE and MCIW for the spatial median.

		MACE					MCIW				
β_0	m_n	Exp	Gamma	Gamma2	Gamma3	Pair	Exp	Gamma	Gamma2	Gamma3	Pair
0.05	36	0.004	0.005	0.009	0.004	0.005	0.485	0.487	0.510	0.472	0.484
	47	0.006	0.005	0.007	0.005	0.005	0.329	0.330	0.338	0.324	0.328
	55	0.010	0.011	0.013	0.009	0.010	0.265	0.266	0.270	0.263	0.265
0.1	36	0.013	0.013	0.010	0.008	0.012	0.404	0.405	0.418	0.395	0.403
	47	0.006	0.006	0.007	0.008	0.006	0.275	0.276	0.280	0.272	0.275
	55	0.012	0.010	0.011	0.012	0.011	0.222	0.222	0.224	0.221	0.222

S4.2 Simulation for spatial median estimation

Let $z_1, \ldots, z_n$ be independent sampled from the mixture normal distribution $0.6N(\mathbf{0}, \Sigma_1) + 0.4(\mathbf{4}, 2I_{m_n})$, where $(\Sigma_1)_{ij} = 0.8^{|i-j|}$ for $1 \leq i, j \leq m_n$. Consider the order $m_n = \lfloor 3n^{0.4} \rfloor$ and choose $(n, m_n) = (500, 36)$, $(1000, 47)$, $(1500, 55)$. For each dimension of z_i, we take $\hat{\theta}_n$ derived from simulated data with a large sample size ($n = 100000$) as a proxy for θ_0, where $\hat{\theta}_n = \arg\min_{\theta \in R^{m_n}} \sum_{i=1}^n \|z_i - \theta\|$. Other settings are identical to those in the logistic regression.

Table S.3: Results of $\|\theta_n^* - \theta_0\|^2/m_n$ in the linear model.

m_n	Error	Exp	Gamma	Gamma2	Gamma3	Pair	Residual
19	Normal	0.161	0.169	0.288	0.105	0.178	0.123
	Mixture	0.046	0.049	0.102	0.025	0.051	0.029
	Double exp	0.057	0.061	0.119	0.033	0.064	0.038
24	Normal	0.073	0.077	0.129	0.048	0.078	0.059
	Mixture	0.017	0.019	0.038	0.010	0.018	0.012
	Double exp	0.073	0.077	0.129	0.048	0.078	0.059
32	Normal	0.027	0.029	0.048	0.018	0.028	0.023
	Mixture	0.005	0.006	0.011	0.003	0.006	0.004
	Double exp	0.007	0.008	0.014	0.004	0.007	0.006
39	Normal	0.013	0.014	0.023	0.009	0.013	0.011
	Mixture	0.002	0.003	0.005	0.002	0.003	0.002
	Double exp	0.003	0.003	0.006	0.002	0.003	0.003

Generally speaking, according to Table S.1 and Table S.2, the RW methods perform well in variance estimation and interval estimation. The RW method with Gamma weights has a slight advantage in estimating the standard deviation of $\sqrt{n}(\hat{\theta}_{nj} - \theta_{0j})$.

Table S.4: Results of $\|\theta_n^* - \theta_0\|^2/m_n$ in the logistic model.

m_n	Exp	Gamma	Gamma2	Gamma3	Pair
18	0.338	0.360	0.661	0.215	0.358
23	0.149	0.158	0.280	0.096	0.154
27	0.095	0.101	0.176	0.062	0.098

Table S.5: Results of $\|\theta_n^* - \theta_0\|^2/m_n$ for the spatial median.

m_n	Exp	Gamma	Gamma2	Gamma3	Pair
36	0.033	0.035	0.064	0.022	0.033
47	0.015	0.016	0.027	0.010	0.015
55	0.010	0.010	0.017	0.006	0.010

S4.3 Results of the squared l_2 norm error

For completeness of the paper, we also compute $\|\theta_n^* - \theta_0\|^2/m_n$ for different methods to show the validity of Theorem 1. Note that the task of RW method is not parameter estimation, but approximates the sampling distribution of the M-estimator $\hat{\theta}_n$ for statistical inference (e.g., interval estimation).

Table S.3 - S.5 show that $\|\theta_n^* - \theta_0\|^2/m_n$ tends to zero as n increases, which demonstrates Theorem 1. For RW methods, the lighter the tail of the random weight w_i, the smaller $\|\theta_n^* - \theta_n\|^2/m_n$ is achieved. This is because the weight w_i is more concentrated around 1 when its tail is light, in which cases the RW estimator θ_n^* is closer to θ_0 because the weighted samples is the closest to the original samples.

Bibliography

El Karoui, N. and E. Purdom (2018). Can we trust the bootstrap in high-dimensions? the case of linear models. *Journal of Machine Learning Research 19*(5), 1–66.

He, X. and Q. Shao (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis 73*(1), 120–135.

Rao, C. R. and L. Zhao (1992). Approximation to the distribution of m-estimates in linear models by randomly weighted bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A 54*(3), 323–331.

Stout, W. F. (1974). *Almost Sure Convergence.* New York: Academic Press.