

Appendix

5.1 Proof of Theorem 3

Theorem 3 is the main technical result of this paper. Proofs of other utility results (Theorem 4, Theorem 5, and Corollary 6) can be obtained by simple modifications. The proof consists of the following three lemmas.

The first lemma says that under condition (A3), if $\hat{M}'(\theta)$ and $M'(\theta)$ are close to each other, then so are their zero points.

Lemma 9. *Let $\Theta_0 \subseteq \Theta$ be a bounded neighborhood of θ^* and $\nu_n \rightarrow 0$ be a positive sequence. If $\hat{M}'_n(\theta)$ is a sequence of continuously differentiable functions such that $\sup_{\Theta_0} |\hat{M}'_n - M'| = O_P(\nu_n)$, and $M(\theta)$ satisfies (A3), then there exists a sequence $\hat{\theta}_n \in \Theta_0$ such that $\hat{M}'_n(\hat{\theta}_n) = 0$ and $|\hat{\theta}_n - \theta^*| = O_P(\nu_n)$.*

Proof. The proof is elementary and we include it here for completeness. For simplicity we assume $d = 1$.

Without loss of generality, we assume $\Theta_0 = \{\theta : |\theta - \theta^*| \leq s\}$ for some positive s . By condition (A3) we have for small enough s , there exists $U > 0$ such that $M''(\theta) \geq U$ for all $\theta \in \Theta_0$. Then we have $M'(\theta^* - s) \leq -Us$ and $M'(\theta^* + s) \geq Us$. Consider event $E_n(s) := \{\sup_{\Theta_0} |\hat{M}'_n - M'| \leq Us/2\}$. By the convergence assumption on $|\hat{M}'_n - M'|$ we have $P(E_n(s)) \rightarrow 1$ as $n \rightarrow \infty$. On $E_n(s)$ we have $\hat{M}'_n(\theta^* - s) \leq -Us/2$ and $\hat{M}'_n(\theta^* + s) \geq Us/2$. By continuity of \hat{M}'_n , there exists a $\hat{\theta}_n \in \Theta_0$ such that $\hat{M}'_n(\hat{\theta}_n) = 0$. Using the lower bound of $M''(\theta)$ again, we have $|M'(\hat{\theta}_n) - \hat{M}'_n(\hat{\theta}_n)| = |M'(\hat{\theta}_n)| \geq U|\hat{\theta}_n - \theta^*|$. Putting these together, we have for any $A > 0$,

$$\begin{aligned} & P(|\hat{\theta}_n - \theta^*| \geq A\nu_n) \\ & \leq P(E_n^C(s)) + P(E_n(s), |\hat{\theta}_n - \theta^*| \geq A\nu_n) \\ & \leq P(E_n^C(s)) + P(|\hat{M}'_n(\hat{\theta}_n) - M'(\hat{\theta}_n)| \geq AU\nu_n) \\ & = P(|\hat{M}'_n(\hat{\theta}_n) - M'(\hat{\theta}_n)| \geq AU\nu_n) + o(1). \end{aligned}$$

Note that $|\hat{M}'_n(\hat{\theta}_n) - M'(\hat{\theta}_n)| = O_P(\nu_n)$. The above inequality suggests $|\hat{\theta}_n - \theta^*| = O_P(\nu_n)$. \square

Recall that

$$\begin{aligned} M(\theta) &= Em(X, \theta), \\ M_n(\theta) &= n^{-1} \sum_{i=1}^n m(X_i, \theta). \end{aligned}$$

The next lemma controls the distance between M' and M'_n , the sampling error term in eq. (8).

Lemma 10. *Under assumptions (A1) and (A2), we have*

$$\sup_{\Theta_0} |M'(\theta) - M'_n(\theta)| = O_P(1/\sqrt{n}). \quad (10)$$

Proof. Lemma 10 is a typical result in empirical process theory. By Theorem 1.3 in [11] we have

$$P\left(\sup_{\Theta_0} |M'(\theta) - M'_n(\theta)| \geq A/\sqrt{n}\right) \leq C \left(\frac{A}{\sqrt{p}}\right)^p \exp(-2A^2),$$

which immediately implies the lemma.

The main condition required is the covering number condition. For a pair of functions $g_l(x)$ and $g_u(x)$, define the bracket

$$[g_l, g_u] = \{g(x) : g_l \leq g \leq g_u\}.$$

The condition needed by our proof is that for any $\epsilon > 0$, the set

$$\mathcal{G}_0 = \left\{ \frac{\partial m(x, \theta)}{\partial \theta} : \theta \in \Theta_0 \right\}$$

can be covered by at most $(V/\epsilon)^p$ brackets $[g_l, g_u]$ such that $E(g_l(X) - g_u(X))^2 \leq \epsilon^2$. It is easy to check this condition using (A1) and (A2). More details can be found in [9, Example 19.7]. \square

The third lemma controls the sum of additive Laplacian noises in the first term of the right hand side of eq. (8). Without loss of generality, we assume $\Theta_0 = \{\theta \in \mathbf{R}^p : \|\theta - \theta^*\|_\infty \leq s\}$ for some $0 < s \leq 1/2$.

Lemma 11. Define $\mathbb{F}(\theta) = z_{\mathbf{r}} g(a_{\mathbf{r}}, \theta)$, where $g(x, \theta) = \frac{\partial m(x, \theta)}{\partial \theta}$. Under assumptions (A1) and (A2), if $h_n \asymp (\sqrt{\log n}/n)^{2/(2+d)}$, we have,

$$\sup_{\Theta_0} n |\mathbb{F}(\theta)| = O_P \left((\sqrt{\log n}/n)^{2/(2+d)} \right).$$

Proof. The basic idea is to establish convergence rate for a finite set of θ 's in Θ_0 and then extend the convergence to the whole Θ_0 . For each $\delta_n > 0$, Θ_0 can be partitioned into $L_n \leq \delta_n^{-p}$ cubes. Let $\theta_1, \dots, \theta_L$ be the centers of these cubes, then for each $\theta \in \Theta_0$ there exists an ℓ such that $\|\theta - \theta_\ell\|_2 \leq \sqrt{p} \delta_n/2$.

For a pair of positive numbers A and B , define events:

$$E_{n,A} := \left\{ \max_{\mathbf{r}} |z_{\mathbf{r}}| \leq A \log(h_n^{-d}) \right\}, \quad (11)$$

$$E_{n,B} := \left\{ \max_{1 \leq \ell \leq L} |\mathbb{F}(\theta_\ell)| \leq B h_n^{-d/2} \sqrt{\log(n)} \right\}. \quad (12)$$

First we have, by union bounds

$$\begin{aligned} P(E_{n,A}^C) &\leq \sum_{\mathbf{r}} P(|z_{\mathbf{r}}| \geq A \log(h_n^{-d})) \\ &= h_n^{-d} \exp(-\alpha A \log(h_n^{-d})/2) \\ &= h_n^{-d} (h_n^{-d})^{-\alpha A/2}. \end{aligned} \quad (13)$$

From Lemma 12 introduced below we have for each θ ,

$$P \left(|\mathbb{F}(\theta)| \geq B h_n^{-d/2} \sqrt{\log(n)} \right) \leq 2 \exp \left(-C \min \left(\frac{B \alpha h_n^{-d/2} \sqrt{\log(n)}}{2C_1}, \frac{B^2 \alpha^2 \log(n)}{4C_1^2} \right) \right), \quad (14)$$

where C is a universal constant. Since h_n decays polynomially in n , the second term dominates the rate for large n :

$$P \left(|\mathbb{F}(\theta)| \geq B h_n^{-d/2} \sqrt{\log(n)} \right) \leq 2n^{-CB^2\alpha^2/4C_1^2}.$$

Therefore we have a bound on the probability of $E_{n,B}$:

$$P(E_{n,B}^C) \leq 2L_n n^{-CB^2\alpha^2/4C_1^2} \leq 2\delta_n^{-p} n^{-CB^2\alpha^2/4C_1^2}. \quad (15)$$

Then for any $\theta \in \Theta_0$, let $\|\theta_\ell - \theta\|_2 \leq \sqrt{p} \delta_n/2$. We have

$$\begin{aligned} |\mathbb{F}(\theta)| &\leq |\mathbb{F}(\theta_\ell)| + |\mathbb{F}(\theta) - \mathbb{F}(\theta_\ell)| \\ &= |\mathbb{F}(\theta_\ell)| + \left| \sum_{\mathbf{r}} z_{\mathbf{r}} (g(a_{\mathbf{r}}, \theta) - g(a_{\mathbf{r}}, \theta_\ell)) \right| \\ &\leq |\mathbb{F}(\theta_\ell)| + C_2 \sqrt{p} \delta_n h_n^{-d} \max_{\mathbf{r}} |z_{\mathbf{r}}|/2, \end{aligned} \quad (16)$$

where the last inequality follows from the uniform bound on $|z_{\mathbf{r}}|$ and the Lipschitz condition of $g(x, \theta)$.

Then on $E_{n,A} \cap E_{n,B}$ we have,

$$\sup_{\Theta_0} |\mathbb{F}(\theta)| \leq B h_n^{-d/2} \sqrt{\log(n)} + C_2 \sqrt{p} \delta_n h_n^{-d} A \log(h_n^{-d})/2. \quad (17)$$

Ignoring constants, let $h_n = (\sqrt{\log n}/n)^{2/(d+2)}$ and $\delta_n = n^{-(d+1)/(d+2)}$, then combine (17) with (13) and (15) we have,

$$\begin{aligned} & P\left(\sup_{\theta \in \Theta_0} |\mathbb{F}(\theta)| \geq B\sqrt{\log(n)}n^{d/(d+2)} + \frac{C_2\sqrt{pd}A}{d+2}\tilde{L}n^{(d-1)/(d+2)}\right) \\ & \leq P(E_{n,A}^C) + P(E_{n,B}^C) \\ & \leq n^{-(\alpha A-2)d/(d+2)} + 2n^{-(CB^2\alpha^2/4C_1^2-p(d+1)/(d+2))}, \end{aligned} \quad (18)$$

where \tilde{L} is a polynomial of $\log n$ and $\log \log n$. The above probability goes to zero polynomially in n if

$$A > \frac{1}{\alpha} \left(3 + \frac{2}{d}\right) \quad \text{and} \quad B > \frac{2C_1\sqrt{p(d+1)}}{\alpha\sqrt{C(d+2)}}.$$

The proof concludes by observing that $\sqrt{\log nn}^{d/(d+2)}$ dominates $\tilde{L}n^{(d-1)/(d+2)}$ for large n . \square

The following lemma provides a concentration inequality for sums of double-exponential random variables, which is used to establish equation (14).

Lemma 12 ([11]). *Let z_1, \dots, z_K be i.i.d double-exponential random variables with density $\frac{1}{2}e^{-|z|}$. For every real-valued function $f : \mathbf{R}^K \mapsto \mathbf{R}^1$, such that*

$$\sum_{k=1}^K |\partial_k f|^2 \leq \lambda^2, \quad \text{and} \quad \max_{1 \leq k \leq K} |\partial_k f| \leq \eta, \quad (19)$$

we have

$$P(f(z_1, \dots, z_K) \geq \text{Med}(f) + t) \leq \exp\left(-C \min\left(\frac{t}{\eta}, \frac{t^2}{\lambda^2}\right)\right),$$

where $\text{Med}(f)$ is the median of $f(z_1, \dots, z_k)$ and $C \geq 0$ is some numerical constant.

For any given θ , consider $\mathbb{F}(\theta)$ the LHS of (14) as a linear function of $(z_{\mathbf{r}} : \mathbf{r} \in \{1, \dots, k_n\}^d)$. Note that $z = \alpha z_{\mathbf{r}}/2$ has density $\frac{1}{2}e^{|z|}$. Then it is easy to check that $\mathbb{F}(\theta)$ satisfies (19) with $\lambda^2 = 4C_1^2 h_n^{-d} \alpha^{-2}$ and $\eta = 2C_1 \alpha^{-1}$.