Biometrika (XXXX), **XXX**, X, *p.* 1
Printed in Great Britain Advance Access publication on XXX

Supplementary material for "On Rosenbaum's Rank-based Matching Estimator"

By Matias D. Cattaneo

Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A.

cattaneo@princeton.edu

Fang Han

Department of Statistics, University of Washington, Seattle, Washington 98195, U.S.A.

fanghan@uw.edu

and Zhexiao Lin

Department of Statistics, University of California, Berkeley, Berkeley, California 94720, U.S.A.

zhexiaolin@berkeley.edu

A. Generalized Framework

A.1. *Setup*

This section extends Rosenbaum's rank standardization idea to a more general setting, and establishes general theory that will cover Theorem 1 as a special case. More specifically, for $\omega = 0, 1$, consider the following general mappings

$$\phi_\omega : \mathcal{X} \to \mathcal{X}_\phi \subset \mathbb{R}^m,$$

with $\mathcal{X}$ representing the support of X and m not necessarily equal to d. Note that here we allow ϕ_0 and ϕ_1 to be different. Consider the setting when ϕ_ω is possibly unknown, and we will approximate it based on the sample $\{(X_i, D_i, Y_i)\}_{i=1}^n$, leading to a generic estimator $\widehat{\phi}_\omega$ that may differ with different ω. We then define

$$U_{\phi,\omega} \equiv \phi_\omega(X) \qquad \text{and} \qquad \widehat{U}_{\phi,\omega,i} \equiv \widehat{\phi}_\omega(X_i) \qquad \text{for} \quad i \in \{1, \dots, n\}.$$

Note that, when setting $\phi_0 = \phi_1 = F$ and $\widehat{\phi}_0 = \widehat{\phi}_1 = \widehat{F}_n$, the latter of which stands for the empirical CDF, we recover the U and $\widehat{U}_i$'s introduced in Section 2.

Similar to Section 2, let $\mathcal{J}_\phi(i)$ represent the index set of M-NN matches of $\widehat{U}_{\phi,1-D_i,i}$ in $\{\widehat{U}_{\phi,1-D_i,j} : D_j = 1 - D_i\}_{j=1}^n$ with ties broken in an arbitrary way. In other words, for determining the nearest neighbors, we are going to measure the similarity based on the Euclidean distance between transformed data points with the transformation function probably also having to be learned from the same data. Additionally, let $\widehat{\mu}_{\phi,\omega}(u)$ be a mapping from $\mathcal{X}_\phi$ to $\mathbb{R}$ that estimates the conditional means of the outcomes

$$\mu_{\phi,\omega}(u) \equiv E(Y \mid U_{\phi,\omega} = u, D = \omega).$$

The general ϕ-transformation based bias-corrected matching estimator is then defined to be

$$\widehat{\tau}_\phi \equiv \frac{1}{n} \sum_{i=1}^n \big(\widehat{Y}_{\phi,i}(1) - \widehat{Y}_{\phi,i}(0)\big),$$

where

$$\widehat{Y}_{\phi,i}(\omega) \equiv \begin{cases} Y_i, & \text{if } D_i = \omega, \\ \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\big(Y_j + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})\big) & \text{if } D_i = 1-\omega \end{cases}.$$

It follows that $\widehat{\tau}_\phi$ generalizes $\widehat{\tau}$ in (1).

A.2. *General Theory*

In order to analyze $\widehat{\tau}_\phi$, we introduce some additional notation and assumptions that are in parallel to those made in Section 3. Let the residuals from fitting the outcome models be

$$\widehat{R}_{\phi,i} \equiv Y_i - \widehat{\mu}_{\phi,D_i}(\widehat{U}_{\phi,D_i,i}), \qquad i \in \{1,\ldots,n\},$$

and the estimator based on the outcome models be

$$\widehat{\tau}_\phi^{\mathrm{reg}} \equiv n^{-1}\sum_{i=1}^n \big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\big).$$

Finally, let $K_\phi(i)$ be the number of matched times for the unit i according to the distances between $\widehat{U}_{\phi,D_i,i}$'s, i.e.,

$$K_\phi(i) \equiv \sum_{j=1,D_j=1-D_i}^n \mathbb{1}\big(i \in \mathcal{J}_\phi(j)\big).$$

The first lemma corresponds to a generalization of the AIPW representation of the bias-corrected rank-based estimator given in (2) in Section 3.

LEMMA 1. *It holds true that*

$$\widehat{\tau}_\phi = \widehat{\tau}_\phi^{\mathrm{reg}} + \frac{1}{n}\sum_{i=1}^n (2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}.$$

The first two assumptions in this section parallel Assumptions 1 and 2.

Assumption 1. (i) For almost all $x \in X$, D is independent of $(Y(0), Y(1))$ conditional on $X = x$, and there exists some constant $c > 0$ such that $c < \mathrm{pr}(D = 1 \mid X = x) < 1 - c$.
(ii) $\{(X_i, D_i, Y_i)\}_{i=1}^n$ are i.i.d. following the joint distribution of (X, D, Y).
(iii) $E\{(Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega}))^2 \mid U_{\phi,\omega} = u\}$ is uniformly bounded for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.
(iv) $E(\mu_{\phi,\omega}^2(U_{\phi,\omega}))$ is bounded for $\omega = 0, 1$.

Assumption 2. (i) There exists a deterministic, possibly changing with n, function $\bar{\mu}_{\phi,\omega}(\cdot) : \mathbb{R}^m \to \mathbb{R}$ such that $E(\bar{\mu}_{\phi,\omega}^2(U_{\phi,\omega}))$ is uniformly bounded and the estimator $\widehat{\mu}_{\phi,\omega}(x)$ satisfies $\|\widehat{\mu}_{\phi,\omega} - \bar{\mu}_{\phi,\omega}\|_\infty = o_{\mathrm{P}}(1)$ for $\omega = 0, 1$.
(ii) $\max_{i\in\{1,\ldots,n\}} |\bar{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \bar{\mu}_{\phi,\omega}(U_{\phi,\omega,i})| = o_{\mathrm{P}}(1)$ for $\omega = 0, 1$.

The next assumption regulates the transformation ϕ_ω; cf. Lin et al. (2023, Section B) and Lin & Han (2022, Assumption 3.3(ii)). From a high level perspective, it roughly states that $M^{-1}K_\phi(i)$ should be a consistent density ratio estimator. A detailed discussion of this assumption is given in Section A.3 ahead.

Assumption 3. The number of matched times satisfies

$$\lim_{n\to\infty} E\Big\{\frac{K_\phi(1)}{M} - \Big(D_1\frac{1-e(X_1)}{e(X_1)} + (1-D_1)\frac{e(X_1)}{1-e(X_1)}\Big)\Big\}^2 = 0,$$

where, for any $x \in \mathcal{X}$, $e(x) \equiv \mathrm{pr}(D = 1 \mid X = x)$ is the propensity score.

The next three assumptions correspond to Assumptions 4 through 6 in Section 3.

Assumption 4. (i) The estimator $\widehat{\mu}_{\phi,\omega}(x)$ satisfies $\|\widehat{\mu}_{\phi,\omega} - \mu_{\phi,\omega}\|_\infty = o_{\mathrm{P}}(1)$ for $\omega = 0, 1$.

(ii) $\max_{i\in\{1,\ldots,n\}} |\mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,i})| = o_{\mathrm{P}}(1)$ for $\omega = 0, 1$.
(iii) $E(Y(\omega) \mid X = x) = \mu_{\phi,\omega}(\phi_\omega(x))$ for almost all $x \in \mathcal{X}$ and $\omega = 0, 1$.

Assumption 5. (i) $E\{(Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega}))^2 \mid U_{\phi,\omega} = u\}$ is uniformly bounded away from zero for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.
(ii) There exists a constant $c > 0$ such that $E(|Y(\omega) - \mu_{\phi,\omega}(U_{\phi,\omega})|^{2+c} \mid U_{\phi,\omega} = u)$ is uniformly bounded for almost all $u \in \mathcal{X}_\phi$ and $\omega = 0, 1$.
(iii) $\max_{t\in\Lambda_{\max\{\lfloor m/2\rfloor,1\}+1}} \|\partial^t \mu_{\phi,\omega}\|_\infty$ is bounded.
(iv) $E(Y(\omega) \mid X = x) = \mu_{\phi,\omega}(\phi_\omega(x))$ for almost all $x \in \mathcal{X}$ and $\omega = 0, 1$.
(v) The density of $\phi_\omega(X)$ is continuous over its support for $\omega = 0, 1$.

Assumption 6. For $\omega = 0, 1$, the estimator $\widehat{\mu}_\omega(x)$ satisfies

$$\max_{t\in\Lambda_{\max\{\lfloor m/2\rfloor,1\}+1}} \|\partial^t \widehat{\mu}_{\phi,\omega}\|_\infty = O_{\mathrm{P}}(1)$$

and

$$\max_{t\in\Lambda_\ell} \|\partial^t \widehat{\mu}_{\phi,\omega} - \partial^t \mu_{\phi,\omega}\|_\infty = O_{\mathrm{P}}(n^{-\gamma_\ell}) \ \text{ for all } \ \ell \in \{1, \ldots, \max\{\lfloor m/2\rfloor, 1\}\},$$

with some constants $\gamma_\ell > \max\{1/2 - \ell/m, 0\}$ for $\ell = 1, 2, \ldots, \max\{\lfloor m/2\rfloor, 1\}$.

The next assumption poses a Donsker-type condition on the approximation accuracy of the estimated transformation $\widehat{\phi}_\omega$ towards ϕ_ω. This assumption is usually needed when one wishes to avoid using sample splitting.

Assumption 7. For $\omega = 0, 1$, the estimator $\widehat{\phi}_\omega$ satisfies

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathrm{pr}\Big(n^{1/2} \sup_{x,y\in\mathcal{X}, \|\phi_\omega(x)-\phi_\omega(y)\|\le\delta} \|(\widehat{\phi}_\omega - \phi_\omega)(x) - (\widehat{\phi}_\omega - \phi_\omega)(y)\| \ge \epsilon\Big) = 0.$$

We are now ready to present the following theorem, which is a generalization to Theorem 1.

Theorem 1 (Generalized Main Theorem). *(i) (Double robustness of $\widehat{\tau}_\phi$) If either Assumptions 1, 2, 3 hold, or Assumptions 1 and 4 hold, then*

$$\widehat{\tau}_\phi - \tau \text{ converges in probability to } 0.$$

(ii) (Semiparametric efficiency of $\widehat{\tau}_\phi$) Assume the distribution of (X, D, Y) satisfies Assumptions 1, 3, 5, 6, 7. Define

$$\gamma = \max\left\{\left[1 - \frac{1}{2}\frac{m}{\max\{\lfloor m/2\rfloor, 1\} + 1}\right], \min_{\ell\in\{1,\ldots,\max\{\lfloor m/2\rfloor,1\}\}} \left\{1 - \left(\frac{1}{2} - \gamma_\ell\right)\frac{m}{\ell}\right\}\right\},$$

recalling that γ_ℓ's were introduced in Assumption 6. Then, if $M \to \infty$ and $M/n^\gamma \to 0$ as $n \to \infty$, we have

$$n^{1/2}(\widehat{\tau}_\phi - \tau) \text{ converges in distribution to } N(0, \sigma^2).$$

(iii) If in addition Assumption 4 holds, then $\widehat{\sigma}^2_\phi$ converges in probability to σ^2, where

$$\widehat{\sigma}^2_\phi \equiv \frac{1}{n}\sum_{i=1}^n \Big\{\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) + (2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} - \widehat{\tau}_\phi\Big\}^2.$$

A.3. *Discussion on High-Level Assumption*

It remains to decipher the high-level condition in Assumption 3. To this end, we first give additional regularizations about the population-transformed data.

Assumption 8. (i) The diameter of $\mathcal{X}_\phi$ and the surface area of $\mathcal{X}_\phi$ are bounded.
(ii) The density of $\phi_\omega(X)$ is continuous over its support for $\omega = 0, 1$.

(iii) $\mathrm{pr}(D = 1 \mid \phi_\omega(X) = \phi_\omega(x)) = \mathrm{pr}(D = 1 \mid X = x)$ for almost all $x \in \mathcal{X}$ and $\omega = 0, 1$.

Next, we give two different type of conditions for the estimator $\widehat{\phi}_\omega$ to approximate ϕ_ω so that Assumption 3 can hold.

Assumption 9. For $\omega = 0, 1$,

$$\lim_{n\to\infty} E\left\{\left(\frac{n}{M}\right)^2 \sup_{x_1,x_2\in\mathcal{X}} \|\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega\|_\infty^{2d}\right\} = 0,$$

where $\widehat{\phi}_\omega(\cdot; x_1, x_2)$ is the estimator constructed by inserting two more new points, x_1 and x_2, into the group with $D = 1 - \omega$ for some $x_1, x_2 \in \mathcal{X}$.

Assumption 10. For $\omega = 0, 1$, we assume that for any fixed $\epsilon > 0$, there exists a function $T_\epsilon(u)$ such that, for any $\delta > 0$,

$$\mathrm{pr}\left(\sup_{\delta\geq u} \delta^{-1} \sup_{\|\phi_\omega(s)-\phi_\omega(t)\|\leq\delta} \sup_{x_1,x_2\in\mathcal{X}} \|(\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega)(s) - (\widehat{\phi}_\omega(\cdot; x_1, x_2) - \phi_\omega)(t)\| > \epsilon\right) \leq T_\epsilon(u),$$

and, for any $k \in \{1, \ldots, 2\}$,

$$\lim_{n\to\infty} \left(\frac{n}{M}\right)^k \int_0^\infty u^{k-1} T_\epsilon(u^{1/m})\mathrm{d}u = 0$$

and

$$\lim_{n\to\infty} \left(\frac{n}{M}\right)^2 \mathrm{pr}\left(\|\widehat{\phi}_\omega - \phi_\omega\|_\infty > \epsilon\right) = 0.$$

THEOREM 2. *Assume that Assumption 8 holds, $M \log n/n \to 0$ and $M \to \infty$ as $n \to \infty$. If either Assumption 9 or Assumption 10 holds, then Assumption 3 holds.*

B. PROOFS OF MAIN RESULTS

B.1. *Proof of Theorem 1*

We take $\phi_0 = \phi_1 = F$ and $\widehat{\phi}_0 = \widehat{\phi}_1 = \widehat{F}_n$, where $\widehat{F}_n$ stands for the empirical CDF. Note that F is a bijective function. Then Assumption 1 implies Assumption 1. To show that Assumption 2 and Assumption 4 imply Assumption 2 and Assumption 4, respectively, it remains to show

$$\max_{i\in\{1,\ldots,n\}} |\bar{\mu}_\omega(\widehat{U}_i) - \bar{\mu}_\omega(U_i)| = o_{\mathrm{P}}(1) \;\text{ or }\; \max_{i\in\{1,\ldots,n\}} |\mu_\omega(\widehat{U}_i) - \mu_\omega(U_i)| = o_{\mathrm{P}}(1).$$

Note that $[0, 1]^d$ is compact. Then the continuity of μ_ω implies uniform continuity. Then $\max_{i\in\{1,\ldots,n\}} |\mu_\omega(\widehat{U}_i) - \mu_\omega(U_i)| = o_{\mathrm{P}}(1)$ is directly from $\max_{i\in\{1,\ldots,n\}} \|\widehat{U}_i - U_i\| = o_{\mathrm{P}}(1)$. The same holds for $\bar{\mu}_\omega$.

To verify Assumption 3, we use Theorem 2. Assumption 8 holds by Assumption 3 and the bijection of F. We verify Assumption 9 when $d \geq 2$ and Assumption 10 when $d = 1$.

Part I. $d \geq 2$.

Note that

$$\sup_{x_1,x_2\in\mathcal{X}} \|\widehat{F}_n(\cdot; x_1, x_2) - F\|_\infty^{2d} \lesssim \sup_{x_1,x_2\in\mathcal{X}} \|\widehat{F}_n(\cdot; x_1, x_2) - \widehat{F}_n\|_\infty^{2d} + \|\widehat{F}_n - F\|_\infty^{2d},$$

where $\lesssim$ means "asymptotically small than".

By the definition of $\widehat{F}_n(\cdot; x_1, x_2)$ and $\widehat{F}_n$, for any $x_1, x_2 \in \mathcal{X}$,

$$\begin{aligned}
&\|\widehat{F}_n(\cdot; x_1, x_2) - \widehat{F}_n\|_\infty \\
&\leq d^{1/2} \max_{k\in\{1,\ldots,d\}} \max_{x\in\mathbb{R}} \left|\frac{1}{n+2}\left(\sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x) + \mathbb{1}(x_1 \leq x) + \mathbb{1}(x_2 \leq x)\right) - \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_{i,k} \leq x)\right|
\end{aligned}$$

$$= d^{1/2} \max_{k\in\{1,\dots,d\}} \max_{x\in\mathbb{R}} \Big| \frac{1}{n+2}\Big(\mathbb{1}(x_1 \le x) + \mathbb{1}(x_2 \le x)\Big) - \frac{2}{n(n+2)} \sum_{i=1}^n \mathbb{1}(X_{i,k} \le x) \Big|$$
$$\le \frac{4d^{1/2}}{n+2}.$$

We then have

$$\lim_{n\to\infty} E\Big\{\Big(\frac{n}{M}\Big)^2 \sup_{x_1,x_2\in\mathcal{X}} \|\widehat{F}_n(\cdot;x_1,x_2) - \widehat{F}_n\|_\infty^{2d}\Big\} = 0.$$

Note that

$$\|\widehat{F}_n - F\|_\infty^{2d} \lesssim \max_{k\in\{1,\dots,d\}} \max_{x\in\mathbb{R}} \Big|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_{i,k} \le x) - \mathrm{pr}(X_k \le x)\Big|^{2d}$$
$$\le \sum_{k=1}^d \max_{x\in\mathbb{R}} \Big|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_{i,k} \le x) - \mathrm{pr}(X_k \le x)\Big|^{2d}.$$

By the Dvoretzky–Kiefer–Wolfowitz inequality, we have

$$E(\|\widehat{F}_n - F\|_\infty^{2d}) \lesssim n^{-d}.$$

By $d \ge 2$ and $M \to \infty$, we have

$$\lim_{n\to\infty} E\Big\{\Big(\frac{n}{M}\Big)^2 \|\widehat{F}_n - F\|_\infty^{2d}\Big\} = 0.$$

The proof is now complete.

Part II. $d = 1$.

Note that for any $s, t \in \mathbb{R}$, by Part I,

$$\sup_{x_1,x_2\in\mathcal{X}} \Big|(\widehat{F}_n(\cdot;x_1,x_2) - F)(s) - (\widehat{F}_n(\cdot;x_1,x_2) - F)(t)\Big| \le \frac{8d^{1/2}}{n+2} + \Big|(\widehat{F}_n - F)(s) - (\widehat{F}_n - F)(t)\Big|.$$

For any $\epsilon > 0$, we can take n sufficiently large such that $8d^{1/2}/(n+2) < \epsilon$, and then

$$\mathrm{pr}\Big(\sup_{\delta\ge u} \delta^{-1} \sup_{|F(s)-F(t)|\le\delta} \sup_{x_1,x_2\in\mathcal{X}} \Big|(\widehat{F}_n(\cdot;x_1,x_2) - F)(s) - (\widehat{F}_n(\cdot;x_1,x_2) - F)(t)\Big| > 2\epsilon\Big)$$
$$\le \mathrm{pr}\Big(\sup_{\delta\ge u} \delta^{-1} \sup_{|F(s)-F(t)|\le\delta} \Big|(\widehat{F}_n - F)(s) - (\widehat{F}_n - F)(t)\Big| > \epsilon\Big)$$
$$= \mathrm{pr}\Big(\sup_{\delta\ge u} \sup_{s\le t:\mathrm{pr}(s<X\le t)\le\delta} \delta^{-1}\Big|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(s < X_i \le t) - \mathrm{pr}(s < X \le t)\Big| > \epsilon\Big).$$

Fix $u \ge 0$ and ϵ. Let $T_i = \delta^{-1}u(\mathbb{1}(s < X_i \le t) - \mathrm{pr}(s < X \le t))$ for $i \in \{1,\dots,n\}$. Then for any $i \in \{1,\dots,n\}$, we have $E(T_i) = 0$, and that $|T_i| \le 1$ for $\delta \ge u$.

Note that

$$\sup_{\delta\ge u} \sup_{s\le t:\mathrm{pr}(s<X\le t)\le\delta} \sum_{i=1}^n E(T_i^2) \le \sup_{\delta\ge u} \sup_{s\le t:\mathrm{pr}(s<X\le t)\le\delta} \sum_{i=1}^n \delta^{-2}u^2\mathrm{pr}(s < X_i \le t) \le un,$$

and

$$E(\sup_{\delta\ge u} \sup_{s\le t:\mathrm{pr}(s<X\le t)\le\delta} \sum_{i=1}^n T_i^2) \le E(\sup_{\delta\ge u} \sup_{s\le t:\mathrm{pr}(s<X\le t)\le\delta} \sum_{i=1}^n \delta^{-2}u^2\mathbb{1}(s < X_i \le t)) \lesssim un,$$

by standard empirical process theory. Then by the concentration inequality for bounded processes (Boucheron et al., 2013, Theorem 12.2), we have for n sufficiently large,

$$\begin{aligned}
&\mathrm{pr}\Big(\sup_{\delta\geq u}\sup_{s\leq t:\mathrm{pr}(s<X\leq t)\leq\delta}\delta^{-1}\Big|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(s<X_i\leq t)-\mathrm{pr}(s<X\leq t)\Big|>\epsilon\Big)\\
=&\mathrm{pr}\Big(\sup_{\delta\geq u}\sup_{s\leq t:\mathrm{pr}(s<X\leq t)\leq\delta}\Big|\sum_{i=1}^{n}T_i\Big|>un\epsilon\Big)\\
\leq&\exp\Big(-\frac{u^2n^2\epsilon^2}{Cun+un\epsilon}\Big)=\exp\Big(-\frac{\epsilon^2}{C+\epsilon}un\Big).
\end{aligned}$$

The proof for this part is now complete by taking integral using $m = 1$ and $M \to \infty$.

Lastly, it is easy to see Assumption 5 and Assumption 6 imply Assumption 5 and Assumption 6, respectively. Assumption 7 is followed by the Donsker's theorem applied to empirical distribution function.

B.2. *Proof of Theorem 2*

Let $\tilde{Q} \equiv \vec{P}^\top\vec{P}/n$. Using $\|\vec{Q}^{-1/2}\|_2 = \lambda_K^{-1/2}$,

$$\|\vec{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\vec{Q}^{-1/2}\|_2=\|\vec{Q}^{-1/2}p_K(W_i)\|^2\leq\lambda_K^{-1}\zeta_{0,K}^2,$$

$$\begin{aligned}
&\|E(\vec{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\vec{Q}^{-1}p_K(W_i)p_K(W_i)^\top\vec{Q}^{-1/2})\|_2^2\\
\leq&\lambda_K^{-1}\zeta_{0,K}^2\|E(\vec{Q}^{-1/2}p_K(W_i)p_K(W_i)^\top\vec{Q}^{-1/2})\|_2^2=\lambda_K^{-1}\zeta_{0,K}^2,
\end{aligned}$$

and a standard exponential concentration inequality for random matrices (Tropp, 2012, Section 6),

$$\|\vec{Q}^{-1/2}\tilde{Q}\vec{Q}^{-1/2}-I_K\|_2=O_{\mathrm{P}}\big(\lambda_K^{-1/2}\zeta_{0,K}(\log(K)/n)^{1/2}+\lambda_K^{-1}\zeta_{0,K}^2\log(K)/n\big)=o_{\mathrm{P}}(1)$$

because $\lambda_K^{-1}\zeta_{0,K}^2\log(K)/n = o(1)$ by assumption.

Let $\tilde{Q}_n \equiv \vec{P}_n^\top\vec{P}_n/n$. Then,

$$\begin{aligned}
\|\vec{Q}^{-1/2}\vec{P}^\top\|_2^2&=\|\vec{Q}^{-1/2}\vec{P}^\top\vec{P}\vec{Q}^{-1/2}\|_2\\
&\leq n\|\vec{Q}^{-1/2}(\tilde{Q}-\vec{Q})\vec{Q}^{-1/2}\|_2+n\|\vec{Q}^{-1/2}\vec{Q}\vec{Q}^{-1/2}\|_2\\
&=O_{\mathrm{P}}(n\lambda_K^{-1/2}\zeta_{0,K}(\log(K)/n)^{1/2}+n),
\end{aligned}$$

and therefore

$$\begin{aligned}
&\|\vec{Q}^{-1/2}(\tilde{Q}_n-\tilde{Q})\vec{Q}^{-1/2}\|_2\\
=&\|\vec{Q}^{-1/2}(\vec{P}_n^\top\vec{P}_n-\vec{P}^\top\vec{P})\vec{Q}^{-1/2}\|_2/n\\
\leq&\|\vec{Q}^{-1/2}(\vec{P}_n-\vec{P})^\top(\vec{P}_n-\vec{P})\vec{Q}^{-1/2}\|_2/n+2\|\vec{Q}^{-1/2}\vec{P}^\top(\vec{P}_n-\vec{P})\vec{Q}^{-1/2}\|_2/n\\
\leq&\|(\vec{P}_n-\vec{P})\vec{Q}^{-1/2}\|_2^2/n+2\|\vec{Q}^{-1/2}\vec{P}^\top\|_2\|(\vec{P}_n-\vec{P})\vec{Q}^{-1/2}\|_2/n\\
=&O_{\mathrm{P}}(B_n+\lambda_K^{-1/4}\zeta_{0,K}(\log(K)/n)^{1/4}B_n^{1/2}+B_n^{1/2})=o_{\mathrm{P}}(1),
\end{aligned}$$

because $\lambda_K^{-1}\zeta_{0,K}^2\log(K)/n = o(1)$ and $B_n = o_{\mathrm{P}}(1)$ by assumption.

Putting the two results together,

$$\begin{aligned}
\|\vec{Q}^{-1/2}\tilde{Q}_n\vec{Q}^{-1/2}-I_K\|_2&\leq\|\vec{Q}^{-1/2}(\tilde{Q}_n-\tilde{Q})\vec{Q}^{-1/2}\|_2+\|\vec{Q}^{-1/2}\tilde{Q}\vec{Q}^{-1/2}-I_K\|_2\\
&=O_{\mathrm{P}}(B_n^{1/2}+\lambda_K^{-1/2}\zeta_{0,K}(\log(K)/n)^{1/2})=o_{\mathrm{P}}(1),
\end{aligned}$$

given the rate restrictions imposed in the theorem.

Let $\mathbb{1}_n = \mathbb{1}(\lambda_{\min}(\vec{Q}^{-1/2}\tilde{Q}_n\vec{Q}^{-1/2}) > 1/2)$. Then, $\lim_{n\to\infty}\mathrm{pr}(\mathbb{1}_n = 1) = 1$. Letting $\varepsilon \equiv \vec{Y} - \vec{\Psi}$,

$$\mathbb{1}_n\|\widehat{\psi}_K-\psi_K\|_{L^2}^2=\mathbb{1}_n\int(p_K(w)^\top\widehat{\beta}_K-p_K(w)^\top\beta_K)^2dF_W(w)$$

$$= \mathbb{1}_n(\widehat{\beta}_K - \beta_K)^\top E(p_K(W)p_K(W)^\top)(\widehat{\beta}_K - \beta_K) = \mathbb{1}_n\|\vec{Q}^{1/2}(\widehat{\beta}_K - \beta_K)\|^2$$
$$\leq 2\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top\varepsilon/n\|^2 + 2\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top(\vec{\Psi} - \vec{P}_n\beta_K)/n\|^2.$$

For the first term, we have

$$\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top\varepsilon/n\|^2 = O_\mathrm{P}(K/n)$$

because

$$\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top\varepsilon/n\|^2 \leq \mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1/2}\|_2^2\|\tilde{Q}_n^{-1/2}\vec{P}_n^\top\varepsilon/n\|^2 = O_\mathrm{P}(1)\mathbb{1}_n\|\tilde{Q}_n^{-1/2}\vec{P}_n^\top\varepsilon/n\|^2,$$

using $\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1/2}\|_2^2 = \mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{Q}^{1/2}\|_2 = O_\mathrm{P}(1)$, and

$$E(\|\tilde{Q}_n^{-1/2}\vec{P}_n^\top\varepsilon/n\|^2|\mathcal{F}_n) = \mathrm{tr}(\tilde{Q}_n^{-1/2}\vec{P}_n^\top E(\varepsilon\varepsilon^\top|\mathcal{F}_n)\vec{P}_n\tilde{Q}_n^{-1/2})/n^2 = O_\mathrm{P}(K/n).$$

We can bound the second term in different ways, depending on the approximation errors considered. The first two bounds rely on vanishing approximation errors ($\xi_K \to 0$ or $\vartheta_{0,K} \to 0$), and thus (implicitly) require $K \to \infty$ in general:

$$\begin{aligned}
&\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top(\vec{\Psi} - \vec{P}_n\beta_K)/n\|^2\\
\leq&\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top/n^{1/2}\|_2^2\|(\vec{\Psi} - \vec{P}_n\beta_K)/n^{1/2}\|^2\\
\leq&O_\mathrm{P}(1)\|\vec{\Psi} - \vec{P}_n\beta_K\|^2/n\\
=&O_\mathrm{P}(\min\{B_n + \xi_K^2\ ,\ R_n + \vartheta_{0,K}^2\}),
\end{aligned}$$

because $\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top/n^{1/2}\|_2^2 = \mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{Q}^{1/2}\|_2 = O_\mathrm{P}(1)$, and because the term $\|\vec{\Psi} - \vec{P}_n\beta_K\|^2/n$ can be bounded in two different ways:

$$\|\vec{\Psi} - \vec{P}_n\beta_K\|^2/n \leq 2\|\vec{\Psi} - \vec{\Psi}_n\|^2/n + 2\|\vec{\Psi}_n - \vec{P}_n\beta_K\|^2/n = O_\mathrm{P}(R_n + \vartheta_{0,K}^2),$$

or

$$\begin{aligned}
\|\vec{\Psi} - \vec{P}_n\beta_K\|^2/n &\leq 2\|\vec{\Psi} - \vec{P}\beta_K\|^2/n + 2\|(\vec{P} - \vec{P}_n)\beta_K\|^2/n\\
&\leq O_\mathrm{P}(\xi_K^2) + 2\|(\vec{P} - \vec{P}_n)\vec{Q}^{-1/2}\|_2^2\|E(\vec{Q}^{-1/2}p_K(W_1)\psi(W_1))\|^2/n\\
&= O_\mathrm{P}(B_n + \xi_K^2),
\end{aligned}$$

because $\|E(\vec{Q}^{-1/2}p_K(W_i)\psi(W_i)) = \beta_K^\top\vec{Q}\beta_K = E\{(p_K(W_i)^\top\beta_K)^2\} = E(\psi_K(W_i)^2) \leq E(\psi(W_i)^2) = O(1)$. Therefore, $\|(\vec{P} - \vec{P}_n)\beta_K\|^2/n = O_\mathrm{P}(B_n)$.

Next, for other possible bounds that do not require vanishing approximation errors ($\vartheta_{0,K} \geq \xi_K \not\to 0$), even when possibly $K \to \infty$, notice that

$$\begin{aligned}
&\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{P}_n^\top(\vec{\Psi} - \vec{P}_n\beta_K)/n\|^2\\
\leq&\mathbb{1}_n\|\vec{Q}^{1/2}\tilde{Q}_n^{-1}\vec{Q}^{1/2}\|_2^2\|\vec{Q}^{-1/2}\vec{P}_n^\top(\vec{\Psi} - \vec{P}_n\beta_K)\|^2/n^2\\
\leq&O_\mathrm{P}(1)\|\vec{Q}^{-1/2}(\vec{P}_n - \vec{P})^\top(\vec{\Psi} - \vec{P}_n\beta_K)\|^2/n^2\\
&+ O_\mathrm{P}(1)\|\vec{Q}^{-1/2}\vec{P}^\top(\vec{P}_n - \vec{P})\beta_K\|^2/n^2\\
&+ O_\mathrm{P}(1)\|\vec{Q}^{-1/2}\vec{P}^\top(\vec{\Psi} - \vec{P}\beta_K)\|^2/n^2,
\end{aligned}$$

where each of the three terms are bounded as follows. For the first term,

$$\begin{aligned}
&\|\vec{Q}^{-1/2}(\vec{P}_n - \vec{P})^\top(\vec{\Psi} - \vec{P}_n\beta_K)\|^2/n^2\\
\leq&\|\vec{Q}^{-1/2}(\vec{P}_n - \vec{P})^\top\|_2^2\|\vec{\Psi} - \vec{P}_n\beta_K\|^2/n^2\\
=&O_\mathrm{P}(B_n)O_\mathrm{P}(\min\{B_n + \xi_K^2\ ,\ R_n + \vartheta_{0,K}^2\}),
\end{aligned}$$

using the calcultions above. For the second term,

$$
\begin{aligned}
&\|\vec{Q}^{-1/2}\vec{P}^{\top}(\vec{P}_n-\vec{P})\beta_K\|^2/n^2\\
\le&\|\vec{Q}^{-1/2}\vec{P}^{\top}\|_2^2\|(\vec{P}_n-\vec{P})\beta_K\|^2/n^2\\
=&O_{\mathrm{P}}(\lambda_K^{-1/2}\zeta_{0,K}(\log(K)/n)^{1/2}+1)O_{\mathrm{P}}(B_n),
\end{aligned}
$$

also using the calculations above. Finally, for the third and the last term,

$$
\|\vec{Q}^{-1/2}\vec{P}^{\top}(\vec{\Psi}-\vec{P}\beta_K)\|^2/n^2=O_{\mathrm{P}}(\min\{\lambda_K^{-1}\zeta_{0,K}^2\xi_K^2,K\vartheta_{0,K}^2\}/n)
$$

because, by the orthogonality of the L^2 projection,

$$
\begin{aligned}
&E\{\|\vec{Q}^{-1/2}\vec{P}^{\top}(\vec{\Psi}-\vec{P}\beta_K)\|^2\}/n^2\\
=&\frac{1}{n^2}E\Big\{\Big(\sum_{i=1}^n\vec{Q}^{-1/2}p_K(W_i)(\psi(W_i)-\psi_K(W_i))\Big)^{\top}\Big(\sum_{i=1}^n\vec{Q}^{-1/2}p_K(W_i)(\psi(W_i)-\psi_K(W_i))\Big)\Big\}\\
=&\frac{1}{n}E\Big(p_K(W_i)^{\top}\vec{Q}^{-1}p_K(W_i)(\psi(W_i)-\psi_K(W_i))^2\Big).
\end{aligned}
$$

The final result in the theorem follows because $\lim_{n\to\infty}\mathrm{pr}(\mathbb{1}_n=1)=1$.

B.3. *Proof of Lemma 1*

For any $u\in\mathbb{R}^K$ with $\|u\|=1$, by the orthonormality of the basis functions, we have

$$
1=\|u\|^2=\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2\mathrm{d}w.
$$

Note that

$$
\begin{aligned}
\lambda_{\min}(E(p_K(W_1)p_K(W_1)^{\top}))&=\min_{u\in\mathbb{R}^K:\|u\|=1}u^{\top}E(p_K(W_1)p_K(W_1)^{\top})u=\min_{u\in\mathbb{R}^K:\|u\|=1}E\Big\{\Big(\sum_{k=1}^K u_kp_{kK}(W_1)\Big)^2\Big\}\\
&=\min_{u\in\mathbb{R}^K:\|u\|=1}\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathrm{d}w,
\end{aligned}
$$

and since $E(p_K(W_1)p_K(W_1)^{\top})$ is positive semidefinite,

$$
\|E(p_K(W_1)p_K(W_1)^{\top})\|_2=\max_{u\in\mathbb{R}^K:\|u\|=1}\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathrm{d}w.
$$

If f_W is bounded away from zero over the support of W, then for any $u\in\mathbb{R}^K$ with $\|u\|=1$,

$$
\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathrm{d}w\ge c\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2\mathrm{d}w=c,
$$

for some constants $c>0$. If f_W is bounded over the support of W, then

$$
\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathrm{d}w\le C\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2\mathrm{d}w=C,
$$

for some constants $C>0$.

More generally, for any $t>0$,

$$
\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathrm{d}w\ge\int\Big(\sum_{k=1}^K u_kp_{kK}(w)\Big)^2f_W(w)\mathbb{1}(f_W(w)\ge t)\mathrm{d}w
$$

$$\geq t \int \Big(\sum_{k=1}^{K} u_k p_{kK}(w)\Big)^2 \mathbb{1}(f_W(w) \geq t)\mathrm{d}w$$
$$= t\Big\{1 - \int \Big(\sum_{k=1}^{K} u_k p_{kK}(w)\Big)^2 \mathbb{1}(0 < f_W(w) < t)\mathrm{d}w\Big\}.$$

By the Cauchy-Schwarz inequality, for all sufficiently small $t > 0$,

$$\int \Big(\sum_{k=1}^{K} u_k p_{kK}(w)\Big)^2 \mathbb{1}(0 < f_W(w) < t)\mathrm{d}w \leq \int \Big(\sum_{k=1}^{K} p_{kK}^2(w)\Big)\mathbb{1}(0 < f_W(w) < t)\mathrm{d}w$$
$$\leq \zeta_{0,K}^2 \int \mathbb{1}(0 < f_W(w) < t)\mathrm{d}w \leq C\zeta_{0,K}^2 t^{\rho}.$$

Take $t = c'\zeta_{0,K}^{-2/\rho}$ for some sufficiently small $c' > 0$ such that $C\zeta_{0,K}^2 t^{\rho} < 1/2$. We then obtain $\lambda_{\min}(E(p_K(W_1)p_K(W_1)^{\top})) \geq c'\zeta_{0,K}^{-2/\rho}/2$, as desired.

B.4. *Proof of Proposition 1*

For Part (i), W follows the distribution of the Gaussian copula from the multivariate normal distribution with correlation matrix Σ, and thus the Lebesgue density of W is

$$f_W(w) = \frac{1}{(\det\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}(\Phi^{-1}(w_1),\ldots,\Phi^{-1}(w_d))(\Sigma^{-1} - I_d)(\Phi^{-1}(w_1),\ldots,\Phi^{-1}(w_d))^{\top}\Big),$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal. Then,

$$\Big\{w : 0 < f_W(w) < t\Big\}$$
$$=\Big\{w : (\Phi^{-1}(w_1),\ldots,\Phi^{-1}(w_d))(\Sigma^{-1} - I_d)(\Phi^{-1}(w_1),\ldots,\Phi^{-1}(w_d))^{\top} > 2\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}$$
$$\subset\Big\{w : \|(\Phi^{-1}(w_1),\ldots,\Phi^{-1}(w_d))\|^2 > \frac{2}{\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}$$
$$\subset\bigcup_{k=1}^{d}\Big\{w : \Phi^{-1}(w_k)^2 > \frac{2}{d\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}$$
$$=\bigcup_{k=1}^{d}\Big\{w : w_k > \Phi\Big(\Big\{\frac{2}{d\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}^{1/2}\Big)\Big\}.$$

For any $k \in \{1,\ldots,d\}$, by the Chernoff bound,

$$\mathsf{Leb}\Big(\Big\{w : w_k > \Phi\Big(\Big[\frac{2}{d\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big]^{1/2}\Big)\Big\}\Big)$$
$$=1 - \Phi\Big(\Big\{\frac{2}{d\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}^{1/2}\Big)$$
$$\leq \exp\Big\{-\frac{1}{d\lambda_{\max}(\Sigma^{-1} - I_d)}\log\Big(\frac{1}{t(\det\Sigma)^{1/2}}\Big)\Big\}$$
$$=\Big(t(\det\Sigma)^{1/2}\Big)^{\frac{1}{d\lambda_{\max}(\Sigma^{-1} - I_d)}}.$$

Then, we have

$$\mathsf{Leb}(\{w : 0 < f_W(w) < t\}) \leq d\Big(t(\det\Sigma)^{1/2}\Big)^{\frac{1}{d\lambda_{\max}(\Sigma^{-1} - I_d)}},$$

as desired.

B.5. *Proof of Lemma 1*

By simple algebra, we have

$$
\begin{aligned}
\widehat{\tau}_\phi &= \frac{1}{n}\sum_{i=1}^n \Big(\widehat{Y}_{\phi,i}(1) - \widehat{Y}_{\phi,i}(0)\Big)\\
&= \frac{1}{n}\sum_{i=1}^n D_i\Big(Y_i - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)} (Y_j + \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,j}))\Big)\\
&+ \frac{1}{n}\sum_{i=1}^n (1-D_i)\Big(\frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)} (Y_j + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,j})) - Y_i\Big)\\
&= \frac{1}{n}\sum_{i=1,D_i=1}^n \Big(\widehat{R}_{\phi,i} + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\widehat{R}_{\phi,j}\Big)\\
&+ \frac{1}{n}\sum_{i=1,D_i=0}^n \Big(\frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\widehat{R}_{\phi,j} - \widehat{R}_{\phi,i} + \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big)\\
&= \frac{1}{n}\sum_{i=1}^n \Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big) + \frac{1}{n}\Big\{\sum_{i=1,D_i=1}^n \Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} - \sum_{i=1,D_i=0}^n\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}\Big\}.
\end{aligned}
$$

This completes the proof.

B.6. *Proof of Theorem 1(i)*

Part I. Suppose the propensity score model is correct, i.e., Assumption 2 and 3 hold. For any $i \in \{1,\dots,n\}$, let $\bar{R}_{\phi,i} \equiv Y_i - \bar{\mu}_{\phi,D_i}(U_{\phi,D_i,i})$. By Lemma 1,

$$
\begin{aligned}
\widehat{\tau}_\phi &= \widehat{\tau}_\phi^{\rm reg} + \frac{1}{n}\sum_{i=1}^n (2D_i - 1)\Big(1+\frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i}\\
&= \frac{1}{n}\sum_{i=1}^n\Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big) - \frac{1}{n}\sum_{i=1}^n\Big(\widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \bar{\mu}_{\phi,0}(U_{\phi,0,i})\Big)\\
&+ \frac{1}{n}\Big\{\sum_{i=1}^n D_i\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) - \sum_{i=1}^n (1-D_i)\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,0}(U_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big)\Big\}\\
&+ \frac{1}{n}\Big\{\sum_{i=1}^n D_i\Big(1+\frac{K_\phi(i)}{M} - \frac{1}{e(X_i)}\Big)\bar{R}_{\phi,i} - \sum_{i=1}^n (1-D_i)\Big(1+\frac{K_\phi(i)}{M} - \frac{1}{1-e(X_i)}\Big)\bar{R}_{\phi,i}\Big\}\\
&+ \frac{1}{n}\Big\{\sum_{i=1}^n \Big(1-\frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i}) - \sum_{i=1}^n\Big(1-\frac{1-D_i}{1-e(X_i)}\Big)\bar{\mu}_{\phi,0}(U_{\phi,0,i})\Big\}\\
&+ \frac{1}{n}\Big(\sum_{i=1}^n \frac{D_i}{e(X_i)}Y_i - \sum_{i=1}^n \frac{1-D_i}{1-e(X_i)}Y_i\Big). \qquad (1)
\end{aligned}
$$

For each pair of terms, we only establish the first half part under treatment, and the second half under control can be established in the same way.

For the first term in (1), by Assumption 2,

$$
\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^n\Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big)\Big| \le \Big|\frac{1}{n}\sum_{i=1}^n\Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big)\Big| + \Big|\frac{1}{n}\sum_{i=1}^n\Big(\bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big)\Big|\\
&\le \|\widehat{\mu}_1 - \bar{\mu}_1\|_\infty + \max_{i\in\{1,\dots,n\}}\Big|\bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big| = o_{\rm P}(1).
\end{aligned}
$$

Then

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})-\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big)=o_{\mathrm{P}}(1). \tag{2}$$

For the second term in (1), by Assumption 2,

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i})-\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big)\Big|\\
&\leq \max_{i\in\{1,\dots,n\}}\Big|\bar{\mu}_{\phi,1}(U_{\phi,1,i})-\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big|\cdot\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}\Big)=\max_{i\in\{1,\dots,n\}}\Big|\bar{\mu}_{\phi,1}(U_{\phi,1,i})-\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big|\\
&\leq \|\widehat{\mu}_1-\bar{\mu}_1\|_\infty+\max_{i\in\{1,\dots,n\}}\Big|\bar{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})-\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big|=o_{\mathrm{P}}(1).
\end{aligned}$$

We then have

$$\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}\Big)\Big(\bar{\mu}_{\phi,1}(U_{\phi,1,i})-\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big)=o_{\mathrm{P}}(1). \tag{3}$$

For the third term in (1), by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}-\frac{1}{e(X_i)}\Big)\bar{R}_{\phi,i}\Big|\\
\leq&\Big\{\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}-\frac{1}{e(X_i)}\Big)^2\Big\}^{1/2}\Big(\frac{1}{n}\sum_{i=1}^{n}D_i\bar{R}_{\phi,i}^2\Big)^{1/2}.
\end{aligned}$$

Note that by Assumptions 1 and 2,

$$\begin{aligned}
&E\Big(\frac{1}{n}\sum_{i=1}^{n}D_i\bar{R}_{\phi,i}^2\Big)=E\Big(D_1\bar{R}_{\phi,1}^2\Big)=E\Big\{D_1\Big(Y_1(1)-\bar{\mu}_{\phi,1}(U_{\phi,1,1})\Big)^2\Big\}\\
\leq&2E\Big\{D_1\Big(\sigma_1^2(U_{\phi,1,1})+(\mu_{\phi,1}(U_{\phi,1,1})-\bar{\mu}_{\phi,1}(U_{\phi,1,1}))^2\Big)\Big\}<\infty,
\end{aligned}$$

where $\sigma_1^2(u)\equiv E\{(Y(1)-\mu_{\phi,1}(u))^2\mid U_{\phi,1}=u\}$ for $u\in\mathcal{X}_\phi$. We then obtain by Assumption 3 and the Markov inequality that

$$\frac{1}{n}\sum_{i=1}^{n}D_i\Big(1+\frac{K_\phi(i)}{M}-\frac{1}{e(X_i)}\Big)\bar{R}_{\phi,i}=o_{\mathrm{P}}(1). \tag{4}$$

For the fourth term in (1), notice that $\bar{\mu}_{\phi,1}(U_{\phi,1,i})$ is a function of X_i. Then by the definition of the propensity score and Assumption 1,

$$E\Big\{\Big(1-\frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big\}=0,\ \ E\Big\{\Big|\Big(1-\frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i})\Big|\Big\}<\infty.$$

By the i.i.d of $[(X_i,D_i)]_{i=1}^n$ and the weak law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big(1-\frac{D_i}{e(X_i)}\Big)\bar{\mu}_{\phi,1}(U_{\phi,1,i})=o_{\mathrm{P}}(1). \tag{5}$$

For the fifth term in (1), notice that $E[|Y|]$ is bounded from Assumption 1 and $[(X_i,D_i,Y_i)]_{i=1}^n$ are i.i.d.. Using the weak law of large numbers yields

$$\frac{1}{n}\Big(\sum_{i=1}^{n}\frac{D_i}{e(X_i)}Y_i-\sum_{i=1}^{n}\frac{1-D_i}{1-e(X_i)}Y_i\Big)\text{ converges in probability to }E\Big(Y_i(1)-Y_i(0)\Big)=\tau. \tag{6}$$

Plugging (2), (3), (4), (5) into (1) completes the proof.

Part II. Suppose the outcome model is correct, i.e., Assumption 4 holds. Using the representation (2),

$$
\begin{aligned}
\widehat{\tau}_\phi &= \widehat{\tau}_\phi^{\mathrm{reg}} + \frac{1}{n}\sum_{i=1}^n (2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} \\
&= \frac{1}{n}\sum_{i=1}^n \Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \mu_{\phi,1}(U_{\phi,1,i})\Big) - \frac{1}{n}\sum_{i=1}^n \Big(\widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big) \\
&+ \frac{1}{n}\Big\{\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(\mu_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) - \sum_{i=1}^n (1 - D_i)\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(\mu_{\phi,0}(U_{\phi,0,i}) - \widehat{\mu}_{\phi,0}(\widehat{U}_{\phi,0,i})\Big)\Big\} \\
&+ \frac{1}{n}\Big\{\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big) - \sum_{i=1}^n (1 - D_i)\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,0}(U_{\phi,0,i})\Big)\Big\} \\
&+ \frac{1}{n}\sum_{i=1}^n \Big(\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big). \qquad (7)
\end{aligned}
$$

For the first term in (7), in the same way as (2),

$$\frac{1}{n}\sum_{i=1}^n \Big(\widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i}) - \mu_{\phi,1}(U_{\phi,1,i})\Big) = o_{\mathrm{P}}(1). \qquad (8)$$

For the second term in (7), in the same way as (3),

$$\frac{1}{n}\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(\mu_{\phi,1}(U_{\phi,1,i}) - \widehat{\mu}_{\phi,1}(\widehat{U}_{\phi,1,i})\Big) = o_{\mathrm{P}}(1). \qquad (9)$$

For the third term in (7), noticing that $[K_\phi(i)]_{i=1}^n$ is a function of $\{(X_i, D_i)\}_{i=1}^n$, by Assumption 1 and Assumption 4, we can obtain for any $i \in \{1, \ldots, n\}$,

$$
\begin{aligned}
&E\Big\{D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big) \,\Big|\, \{(X_i, D_i)\}_{i=1}^n\Big\} \\
=&D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(E(Y_i \mid X_i, D_i = 1) - \mu_{\phi,1}(U_{\phi,1,i})\Big) \\
=&D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(E(Y_i(1) \mid X_i) - \mu_{\phi,1}(U_{\phi,1,i})\Big) = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
&E\Big\{\Big|\frac{1}{n}\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big)\Big|\Big\} \le E\Big\{\Big|\frac{1}{n}\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big|\Big\}\|\sigma_1\|_\infty \\
&\lesssim \|\sigma_1\|_\infty = O(1).
\end{aligned}
$$

Accordingly, by the martingale convergence theorem in the same way as Abadie & Imbens (2012), we obtain

$$\frac{1}{n}\sum_{i=1}^n D_i\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,1}(U_{\phi,1,i})\Big) = o_{\mathrm{P}}(1). \qquad (10)$$

For the fourth term in (7), notice that $E\{\mu_{\phi,\omega}^2(U_{\phi,\omega})\}$ is bounded for $\omega = 0, 1$. Using the weak law of large number, we obtain

$$\frac{1}{n}\sum_{i=1}^n \Big(\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big) \text{ converges in probability to } E\Big(\mu_1(X_1) - \mu_0(X_1)\Big) = \tau. \qquad (11)$$

Plugging (8), (9), (10), (11) into (7) completes the proof.

B.7. *Proof of Theorem 1(ii)*

We decompose $\widehat{\tau}_\phi$ as

$$\begin{aligned}
\widehat{\tau}_\phi &= \widehat{\tau}_\phi^{\text{reg}} + \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\widehat{R}_{\phi,i} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big(\mu_{\phi,1}(U_{\phi,1,i}) - \mu_{\phi,0}(U_{\phi,0,i})\Big) + \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(1 + \frac{K_\phi(i)}{M}\Big)\Big(Y_i - \mu_{\phi,D_i}(U_{\phi,D_i,i})\Big) \\
&+ \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\mu_{\phi,1-D_i}(U_{\phi,1-D_i,j})\Big) \\
&- \frac{1}{n}\sum_{i=1}^{n}(2D_i - 1)\Big(\widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i}) - \frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j})\Big) \\
&\equiv \bar{\tau}_\phi + E_n + B_n - \widehat{B}_n.
\end{aligned}$$

In the same way as Lemmas A.1 and A.2 in Lin & Han (2022), we have the following central limit theorem on $\bar{\tau}_\phi + E_n$.

LEMMA 2. *Under Assumptions 1, 3, 5,*

$$n^{1/2}\sigma^{-1}(\bar{\tau}_\phi + E_n - \tau) \text{ converges in distribution to } N(0,1).$$

For the bias term $B_M - \widehat{B}_M$, in light of the smoothness conditions on μ_ω and approximation conditions on $\widehat{\mu}_\omega$ for $\omega = 0, 1$, one can establish the following lemma.

LEMMA 3. *Under Assumptions 1, 3, 5, 6, 7,*

$$n^{1/2}(B_n - \widehat{B}_n) \text{ converges in probability to } 0.$$

Combining Lemma 2 and Lemma 3 completes the proof.

The consistency of the variance estimator can be established in a similar way as the proof of Theorem 4.1 in Lin et al. (2023).

B.8. *Proof of Theorem 2*

Similar to Lin et al. (2023), we first consider a two-sample density ratio estimation problem.

With an abuse of notation, restricted to this section let's consider two general random vectors X, Z in $\mathbb{R}^d$ that are defined on the same probability space. Let $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ be the supports of X and Z respectively with $\mathcal{Z} \subset \mathcal{X}$. Consider a general function $\phi : \mathcal{X} \to \mathbb{R}^m$. Let ν_0 and ν_1 represent the probability measures of $\phi(X)$ and $\phi(Z)$, respectively. Assume ν_0 and ν_1 are absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^m$ equipped with the Euclidean norm $\|\cdot\|$; denote the corresponding densities (Radon-Nikodym derivatives) by f_0 and f_1. Assume further that ν_1 is absolutely continuous with respect to ν_0 and write the corresponding density ratio, f_1/f_0, as r; set $0/0 = 0$ by default.

Assume $X_1, \ldots, X_{N_0}$ are N_0 independent copies of X, $Z_1, \ldots, Z_{N_1}$ are N_1 independent copies of Z, and $[X_i]_{i=1}^{N_0}$ and $[Z_j]_{j=1}^{N_1}$ are mutually independent. We aim to estimate the density ratio $r(\phi(x))$ for any $x \in \mathcal{X}$ based on $\{X_1, \ldots, X_{N_0}, Z_1, \ldots, Z_{N_1}\}$.

For any $x \in \mathcal{X}$, we consider a general estimator $\widehat{\phi}$ estimating ϕ, which may depend on

$$\{X_1, \ldots, X_{N_0}, Z_1, \ldots, Z_{N_1}\} \text{ and } x.$$

Define the catchment area of x:

$$A_\phi(x) = A_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \widehat{\phi}\big) \equiv \Big\{z \in \mathcal{Z} : \|\widehat{\phi}(x) - \widehat{\phi}(z)\| \le \widehat{\Phi}_M(z)\Big\}, \tag{12}$$

where $\widehat{\Phi}_M(z)$ is the M-th order statistics of $\{\|\widehat{\phi}(X_i) - \widehat{\phi}(z)\|\}_{i=1}^{N_0}$, and the number of matched times of x:

$$K_\phi(x) = K_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \{Z_j\}_{j=1}^{N_1}\big) \equiv \sum_{j=1}^{N_1} \mathbb{1}\Big(Z_j \in A_\phi(x)\Big). \tag{13}$$

Then the density ratio estimator is defined as:

$$\widehat{r}_\phi(x) = \widehat{r}_\phi\big(x, \{X_i\}_{i=1}^{N_0}, \{Z_j\}_{j=1}^{N_1}\big) \equiv \frac{N_0}{N_1}\frac{K_\phi(x)}{M}. \tag{14}$$

For any positive integer p, let $\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}$ be the estimator replacing $(Z_1, \dots, Z_p)$ by z for $z \in \mathcal{Z}^p$. We consider the following two assumptions, which are analogies of Assumption 9 and Assumption 10 in the two-sample problem.

Assumption 11.

$$\lim_{N_0\to\infty} E\Big\{\Big(\frac{N_0}{M}\Big)^p \sup_{z\in\mathcal{Z}^p} \|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi\|_\infty^{pd}\Big\} = 0.$$

Assumption 12. For any $\epsilon > 0$ and $\delta > 0$,

$$\mathrm{pr}\Big(\sup_{\delta\in\mathbb{R}:\delta\ge u} \delta^{-1} \sup_{s,t\in\mathcal{X}:\|\phi(s)-\phi(t)\|\le\delta}\ \sup_{z\in\mathcal{Z}^p} \|(\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi)(s) - (\widehat{\phi}_{(Z_1,\dots,Z_p)\to z} - \phi)(t)\| > \epsilon\Big) \le T_\epsilon(u),$$

for $T_\epsilon(u)$ satisfying for any $k \in \{1, \dots, p\}$,

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^k \int_0^\infty u^{k-1} T_\epsilon(u^{1/m})\mathrm{d}u = 0,$$

and

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(\|\widehat{\phi} - \phi\|_\infty > \epsilon\Big) = 0.$$

The following theorem considers the asymptotic L^p moments of $\widehat{r}_\phi$.

THEOREM 3 (ASYMPTOTIC L^p MOMENTS OF $\widehat{r}_\phi$). *Let p be any positive integer. Assume Assumption 11 or 12 holds for p. Assume $M\log N_0/N_0 \to 0$, $MN_1/N_0 \to \infty$ and $M \to \infty$ as $N_0 \to \infty$. We then have*

$$\lim_{N_0\to\infty} E\{(\widehat{r}_\phi(x))^p\} = \{r(\phi(x))\}^p$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

The proof of Theorem 3 will use the following lemma.

LEMMA 4. *Under the same conditions of Theorem 3, we have*

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(Z_1, \dots, Z_p \in A_\phi(x)\Big) = \{r(\phi(x))\}^p.$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

As a direct result of Theorem 3, we can establish the pointwise consistency of the estimator $\widehat{r}_\phi$.

COROLLARY 1 (POINTWISE CONSISTENCY). *Under the same conditions as Theorem 3, if p is even, we have*

$$\lim_{N_0\to\infty} E\{|\widehat{r}_\phi(x) - r(\phi(x))|^p\} = 0$$

holds for all $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and f_0, f_1 are continuous at $\phi(x)$.

The pointwise consistency of $\widehat{r}_\phi$ can then be generalized to global consistency under the following assumptions on $\mathcal{X}$.

Assumption 13. (i) $\mathcal{X}$ is compact and the surface areas of $\mathcal{X}$ and $\mathcal{Z}$ are bounded.
(ii) r is bounded over $\mathcal{X}$.
(iii) f_0 is continuous over $\mathcal{X}$ and f_1 is continuous over $\mathcal{Z}$.

Theorem 4 (Global consistency). *Under the same conditions of Theorem 3 and Assumption 13, if p is even, we have*

$$\lim_{N_0\to\infty} E\{|\widehat{r}_\phi(X) - r(\phi(X))|^p\} = 0.$$

Now back to the causal setting, the density ratio $r(\phi(x))$ is not necessarily equal to the density ratio we are interested in. We consider a lemma showing the equivalence of the two density ratios under additional assumptions.

Lemma 5. *Let $f_{X|D=1}$ and $f_{X|D=0}$ be the density of $X \mid D=1$ and $X \mid D=0$, respectively. Let $f_{\phi,X|D=1}$ and $f_{\phi,X|D=0}$ be the density of $\phi(X) \mid D=1$ and $\phi(X) \mid D=0$. Then for any $x \in \mathcal{X}$ such that $\mathrm{pr}(D=1 \mid \phi(X)=\phi(x)) = \mathrm{pr}(D=1 \mid X=x)$, we have*

$$\frac{f_{\phi,X|D=1}(\phi(x))}{f_{\phi,X|D=0}(\phi(x))} = \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)}.$$

Note that

$$\begin{aligned}
&E\Big\{\frac{K_\phi(1)}{M} - \Big(D_1\frac{1-e(X_1)}{e(X_1)} + (1-D_1)\frac{e(X_1)}{1-e(X_1)}\Big)\Big\}^2\\
=&E\Big[E\Big[\Big\{\frac{K_\phi(1)}{M} - \Big(D_1\frac{1-e(X_1)}{e(X_1)} + (1-D_1)\frac{e(X_1)}{1-e(X_1)}\Big)\Big\}^2 \,\Big|\, \{D_i\}_{i=1}^n\Big]\Big]\\
=&E\Big[E\Big\{\Big(\frac{K_\phi(1)}{M} - \frac{1-e(X_1)}{e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1=1\Big\}\mathbb{1}\Big(D_1=1\Big)\Big]\\
&+E\Big[E\Big\{\Big(\frac{K_\phi(1)}{M} - \frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1=0\Big\}\mathbb{1}\Big(D_1=0\Big)\Big].
\end{aligned}$$

We consider the second term for example. Conditional on $\{D_i\}_{i=1}^n$, $[X_i]_{i:D_i=0}$ and $[X_i]_{i:D_i=1}$ are two samples from $X \mid D=0$ and $X \mid D=1$, respectively. Note that

$$E\Big\{\Big(\frac{K_\phi(1)}{M} - \frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1=0\Big\} = \Big(\frac{N_1}{N_0}\Big)^2 E\Big\{\Big(\frac{N_0}{N_1}\frac{K_\phi(1)}{M} - \frac{N_0}{N_1}\frac{e(X_1)}{1-e(X_1)}\Big)^2 \,\Big|\, \{D_i\}_{i=1}^n, D_1=0\Big\}.$$

By the strong law of large number, we have $(N_0/N_1) \to \mathrm{pr}(D=0)/\mathrm{pr}(D=1)$ with probability one. Note that

$$\frac{\mathrm{pr}(D=0)}{\mathrm{pr}(D=1)}\frac{e(X_1)}{1-e(X_1)} = \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)}.$$

To apply Theorem 4 and Lemma 5, the last thing is to compare the definition of $K_\phi(1)$ with $K_\phi(x)$. Note that if we define

$$A'_\phi(x) \equiv \Big\{z \in \mathcal{Z} : \|\widehat{\phi}(x) - \widehat{\phi}(z)\| < \widehat{\Phi}_M(z)\Big\},\ \ K'_\phi(x) \equiv \sum_{j=1}^{N_1}\mathbb{1}\Big(Z_j \in A'_\phi(x)\Big),$$

as long as the ties are broken in arbitrary way, we can check that $K'_\phi(X_1) \le K_\phi(1) \le K_\phi(X_1)$. Note that all the previous results for $K_\phi(x)$ are also hold for $K'_\phi(x)$. Then the proof is complete.

C. Proofs of Auxiliary Results

C.1. *Proof of Lemma 3*

Note that for any $i \in \{1,\ldots,n\}$ and $\omega = 0,1$,

$$|\mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})|$$

$$\begin{aligned}
&\leq |\mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j})| \\
&+ |\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})|. \qquad (1)
\end{aligned}$$

We can also decompose it in another way:

$$\begin{aligned}
&|\mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})| \\
\leq &|\mu_{\phi,\omega}(U_{\phi,\omega,i}) - \mu_{\phi,\omega}(U_{\phi,\omega,j}) - \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})| \\
&+ |\mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) - \mu_{\phi,\omega}(\widehat{U}_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})|.
\end{aligned}$$

We consider the proof under (1), and the proof under the second decomposition is similar. For the first term in (1), by Taylor expansion to k-th order with $k = \max\{\lfloor m/2 \rfloor, 1\} + 1$,

$$\begin{aligned}
&\Big|\mu_{\phi,\omega}(U_{\phi,\omega,j}) - \mu_{\phi,\omega}(U_{\phi,\omega,i}) - \sum_{\ell=1}^{k-1}\sum_{t\in\Lambda_\ell}\frac{1}{t!}\partial^t\mu_{\phi,\omega}(U_{\phi,\omega,i})(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t\Big| \\
&\leq \max_{t\in\Lambda_k}\|\partial^t\mu_{\phi,\omega}\|_\infty \sum_{t\in\Lambda_k}\frac{1}{t!}\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^k.
\end{aligned}$$

In the same way,

$$\begin{aligned}
&\Big|\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \sum_{\ell=1}^{k-1}\sum_{t\in\Lambda_\ell}\frac{1}{t!}\partial^t\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i})(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t\Big| \\
&\leq \max_{t\in\Lambda_k}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty \sum_{t\in\Lambda_k}\frac{1}{t!}\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^k.
\end{aligned}$$

We also have

$$\begin{aligned}
&\Big|\sum_{\ell=1}^{k-1}\sum_{t\in\Lambda_\ell}\frac{1}{t!}(\partial^t\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \partial^t\mu_{\phi,\omega}(U_{\phi,\omega,i}))(U_{\phi,\omega,j} - U_{\phi,\omega,i})^t\Big| \\
&\leq \sum_{\ell=1}^{k-1}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \partial^t\mu_{\phi,\omega}(U_{\phi,\omega,i})\| \sum_{t\in\Lambda_\ell}\frac{1}{t!}\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|^\ell.
\end{aligned}$$

For the second term in (1), by Taylor expansion,

$$\begin{aligned}
&|\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})| \\
&= |\partial\widehat{\mu}_{\phi,\omega}(\bar{u}_j)^\top(\widehat{U}_{\phi,\omega,j} - U_{\phi,\omega,j}) - \partial\widehat{\mu}_{\phi,\omega}(\bar{u}_i)^\top(\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i})| \\
&\leq |(\partial\widehat{\mu}_{\phi,\omega}(\bar{u}_j) - \partial\widehat{\mu}_{\phi,\omega}(\bar{u}_i))^\top(\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i})| + |\partial\widehat{\mu}_{\phi,\omega}(\bar{u}_j)^\top(\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i})| \\
&\lesssim \max_{t\in\Lambda_2}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty\|\bar{u}_j - \bar{u}_i\|\|\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}\| + \|\partial\widehat{\mu}_{\phi,\omega}\|_\infty\|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\|,
\end{aligned}$$

where $\bar{u}_i$ is between $U_{\phi,\omega,i}$ and $\widehat{U}_{\phi,\omega,i}$, and $\bar{u}_j$ is between $U_{\phi,\omega,j}$ and $\widehat{U}_{\phi,\omega,j}$. Since $\|\widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,i}\| \leq \|\widehat{\phi}_\omega - \phi_\omega\|_\infty$ and $\|\bar{u}_j - \bar{u}_i\| \leq \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\| + \|\widehat{\phi}_\omega - \phi_\omega\|_\infty$, we have

$$\begin{aligned}
&|\widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,i}) - \widehat{\mu}_{\phi,\omega}(U_{\phi,\omega,j}) - \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,i}) + \widehat{\mu}_{\phi,\omega}(\widehat{U}_{\phi,\omega,j})| \\
&\lesssim \max_{t\in\Lambda_2}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty(\|\widehat{\phi}_\omega - \phi_\omega\|_\infty^2 + \|\widehat{\phi}_\omega - \phi_\omega\|_\infty\|U_{\phi,\omega,j} - U_{\phi,\omega,i}\|) + \|\partial\widehat{\mu}_{\phi,\omega}\|_\infty\|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\|.
\end{aligned}$$

Since we have $|\mathcal{J}_\phi(i)| = M$ for any $i \in \{1,\ldots,n\}$, then

$$\begin{aligned}
&|B_n - \widehat{B}_n| \\
&\leq \frac{1}{n}\sum_{i=1}^n\frac{1}{M}\sum_{j\in\mathcal{J}_\phi(i)}\Big|\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i}) - \mu_{\phi,1-D_i}(U_{\phi,1-D_i,j}) - \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i}) + \widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j})\Big|
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\Big|\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i})-\mu_{\phi,1-D_i}(U_{\phi,1-D_i,j})-\widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,i})+\widehat{\mu}_{\phi,1-D_i}(\widehat{U}_{\phi,1-D_i,j})\Big|\\
&\lesssim \max_{\omega\in\{0,1\}}\Big(\max_{t\in\Lambda_k}\|\partial^t\mu_{\phi,\omega}\|_\infty+\max_{t\in\Lambda_k}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^k\Big)\\
&+\sum_{\ell=1}^{k-1}\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,1-D_i}(U_{\phi,1-D_i,i})-\partial^t\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i})\|\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^\ell\Big)\\
&+\max_{\omega\in\{0,1\}}\max_{t\in\Lambda_2}\|\partial^t\widehat{\mu}_{\phi,\omega}\|_\infty\Big\{\|\widehat{\phi}_\omega-\phi_\omega\|_\infty^2+\|\widehat{\phi}_\omega-\phi_\omega\|_\infty\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|\Big)\Big\}\\
&+\max_{\omega\in\{0,1\}}\|\partial\widehat{\mu}_{\phi,\omega}\|_\infty\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j}-\widehat{U}_{\phi,1-D_i,i}-U_{\phi,1-D_i,j}+U_{\phi,1-D_i,i}\|\Big). \qquad (2)
\end{aligned}
$$

For any $i\in\{1,\dots,n\}$, let $\tilde{\mathcal{J}}_\phi(i)$ be the index set of M-NNs of $U_{\phi,1-D_i,i}$ in $\{U_{\phi,1-D_i,j}:D_j=1-D_i\}_{j=1}^n$ with ties broken in arbitrary way. Then

$$
\begin{aligned}
\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\| &\leq \max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j}-\widehat{U}_{\phi,1-D_i,i}\|+2\|\widehat{\phi}_{1-D_i}-\phi_{1-D_i}\|_\infty\\
&\leq \max_{j\in\tilde{\mathcal{J}}_\phi(i)}\|\widehat{U}_{\phi,1-D_i,j}-\widehat{U}_{\phi,1-D_i,i}\|+2\|\widehat{\phi}_{1-D_i}-\phi_{1-D_i}\|_\infty\\
&\leq \max_{j\in\tilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|+4\|\widehat{\phi}_{1-D_i}-\phi_{1-D_i}\|_\infty.
\end{aligned}
$$

By Li & Racine (2023, Lemma 14.1), as long as the density of $U_{\phi,\omega}$ is continuous for $\omega=0,1$, we have for any positive integer p,

$$
E\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\tilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^p\Big)\lesssim\Big(\frac{M}{n}\Big)^{p/m}.
$$

Then for any positive integer p, by Assumption 7, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^p\\
\lesssim&\frac{1}{n}\sum_{i=1}^{n}\Big(\max_{j\in\tilde{\mathcal{J}}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^p+\|\widehat{\phi}_{1-D_i}-\phi_{1-D_i}\|_\infty^p\Big)\\
=&O_{\mathrm{P}}((M/n)^{p/m}+n^{-p/2}). \qquad (3)
\end{aligned}
$$

For any positive integer $\ell\in\{1,\dots,k-1\}$, by Assumption 6, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,1-D_i}(U_{\phi,1-D_i,i})-\partial^t\mu_{\phi,1-D_i}(U_{\phi,1-D_i,i})\|\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^\ell\\
&\leq\max_{\omega\in\{0,1\}}\max_{t\in\Lambda_\ell}\|\partial^t\widehat{\mu}_{\phi,\omega}-\partial^t\mu_{\phi,\omega}\|_\infty\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|U_{\phi,1-D_i,j}-U_{\phi,1-D_i,i}\|^\ell\Big)\\
&=O_{\mathrm{P}}(n^{-\gamma_\ell}((M/n)^{\ell/m}+n^{-\ell/2})). \qquad (4)
\end{aligned}
$$

For any $\epsilon>0$ and $\omega=0,1$, we have

$$
\begin{aligned}
&\mathrm{pr}\Big(\frac{1}{n}\sum_{i=1}^{n}\max_{j\in\mathcal{J}_\phi(i)}\|\widehat{U}_{\phi,\omega,j}-\widehat{U}_{\phi,\omega,i}-U_{\phi,\omega,j}+U_{\phi,\omega,i}\|\geq n^{-\frac{1}{2}}\epsilon\Big)\\
\leq&\mathrm{pr}\Big(n^{1/2}\sup_{x,y\in\mathcal{X},\|\phi_\omega(x)-\phi_\omega(y)\|\leq\delta}\|(\widehat{\phi}_\omega-\phi_\omega)(x)-(\widehat{\phi}_\omega-\phi_\omega)(y)\|\geq\epsilon\Big)
\end{aligned}
$$

$$+ \operatorname{pr}\Big(\max_{i\in\{1,\dots,n\}} \max_{j\in\mathcal{J}_\phi(i)} \|U_{\phi,\omega,j} - U_{\phi,\omega,i}\| \ge \delta \Big),$$

which holds for any $\delta > 0$.

Taking $n \to \infty$ and then $\delta \to 0$, by Assumption 7 and $M/n \to 0$, we have

$$\frac{1}{n}\sum_{i=1}^n \max_{j\in\mathcal{J}_\phi(i)} \|\widehat{U}_{\phi,\omega,j} - \widehat{U}_{\phi,\omega,i} - U_{\phi,\omega,j} + U_{\phi,\omega,i}\| = o_{\mathrm{P}}(n^{-1/2}). \tag{5}$$

Plugging (3), (4), (5) into (2) and using Assumption 6 yields

$$\begin{aligned}
&|B_n - \widehat{B}_n| \\
&\lesssim O_{\mathrm{P}}((M/n)^{k/m} + n^{-k/2}) + \sum_{\ell=1}^{k-1} O_{\mathrm{P}}(n^{-\gamma_\ell}((M/n)^{\ell/m} + n^{-\ell/2})) + O_{\mathrm{P}}((M/n)^{1/m} n^{-1/2}) + o_{\mathrm{P}}(n^{-1/2}).
\end{aligned}$$

This completes the proof by the selection of M.

C.2. *Proof of Theorem 3*

Note that by the multinomial theorem,

$$\begin{aligned}
E\{(\widehat{r}_\phi(x))^p\} &= E\Big\{\Big(\frac{N_0}{N_1}\frac{K_\phi(x)}{M}\Big)^p\Big\} = \Big(\frac{N_0}{N_1 M}\Big)^p E\Big\{\Big(\sum_{j=1}^{N_1} \mathbb{1}\Big(Z_j \in A_\phi(x)\Big)\Big)^p\Big\} \\
&= \Big(\frac{N_0}{N_1 M}\Big)^p \sum_{p_1+\cdots+p_{N_1}=p;\ p_1,\dots,p_{N_1}\ge 0} \binom{p}{p_1,\dots,p_{N_1}} E\Big(\prod_{j=1}^{N_1} \mathbb{1}\Big(Z_j \in A_\phi(x)\Big)^{p_j}\Big) \\
&= \Big(\frac{N_0}{N_1 M}\Big)^p \sum_{p_1+\cdots+p_{N_1}=p;\ p_1,\dots,p_{N_1}\ge 0} \binom{p}{p_1,\dots,p_{N_1}} \operatorname{pr}\Big(Z_j \in A_\phi(x) : p_j > 0\Big).
\end{aligned}$$

Then by Lemma 4, we have

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^{\sum_{j=1}^{N_1}\mathbb{1}(p_j>0)} \operatorname{pr}\Big(Z_j \in A_\phi(x) : p_j > 0\Big) = \{r(\phi(x))\}^{\sum_{j=1}^{N_1}\mathbb{1}(p_j>0)}.$$

Note that for $p_1,\dots,p_{N_1} \ge 0$ with $p_1+\cdots+p_{N_1} = p$, the number of terms such that $\sum_{j=1}^{N_1}\mathbb{1}(p_j>0) = k$ is of order N_1^k for any $k \in \{1,\dots,p\}$. Also note that $\binom{p}{p_1,\dots,p_{N_1}}$ is bounded. Therefore if $MN_1/N_0 \to \infty$, we have

$$\lim_{N_0\to\infty} E\{(\widehat{r}_\phi(x))^p\} = \lim_{N_0\to\infty} \frac{1}{N_1^p}\binom{N_1}{p}\binom{p}{1,\dots,1}\{r(\phi(x))\}^p = \{r(\phi(x))\}^p.$$

This completes the proof.

C.3. *Proof of Lemma 4*

We only consider those $x \in \mathcal{X}$ such that $f_0(\phi(x)) > 0$ and $\phi(x)$ is a continuous point of f_0 and f_1. We seperate the proof into two cases depending on whether $f_1(\phi(x))$ is zero.

Part I. We first consider the simple case where $p = 1$ and Assumption 11 holds for p.

Case I. $f_1(\phi(x)) > 0$. Since $\phi(x)$ is a continuous point of f_0 and f_1, for any $\epsilon \in (0,1)$, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $|f_0(\phi(z)) - f_0(\phi(x))| \le \epsilon f_0(\phi(x))$ and $|f_1(\phi(z)) - f_1(\phi(x))| \le \epsilon f_1(\phi(x))$. Denote the closed ball in $\mathbb{R}^m$ centered at x with radius δ by $B_{x,\delta}$, and the Lebesgue measure by λ. Then for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le \delta$, we have

$$\Big|\frac{\nu_0(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \le \epsilon f_0(\phi(x)),\ \ \Big|\frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \le \epsilon f_0(\phi(x)),$$

$$\left|\frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})} - f_1(\phi(x))\right| \le \epsilon f_1(\phi(x)), \quad \left|\frac{\nu_1(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_1(\phi(x))\right| \le \epsilon f_1(\phi(x)).$$

Accordingly, if $\|\phi(z)-\phi(x)\| \le \delta$, we have

$$\frac{1-\epsilon}{1+\epsilon}\frac{f_0(\phi(x))}{f_1(\phi(x))} \le \frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})}\frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})} \le \frac{1+\epsilon}{1-\epsilon}\frac{f_0(\phi(x))}{f_1(\phi(x))}.$$

Since $\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|}) = \lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})$, we then have

$$\frac{1-\epsilon}{1+\epsilon}\frac{f_0(\phi(x))}{f_1(\phi(x))} \le \frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})} \le \frac{1+\epsilon}{1-\epsilon}\frac{f_0(\phi(x))}{f_1(\phi(x))}.$$

On the other hand, consider any $\epsilon' \in (0,1)$. For any $z \in \mathcal{X}$ such that $\|\phi(z)-\phi(x)\| > \delta$, as long as ϵ' small enough such that $\epsilon'\mathrm{diam}(\mathcal{X}) < \delta/2$, where $\mathrm{diam}(\mathcal{X})$ is the diameter of $\mathcal{X}$, we have $B_{y,\delta/2} \subset B_{\phi(z),\|\phi(z)-\phi(x)\|-\delta/2} \subset B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}$, where $y \in \mathbb{R}^m$ is taken such that y is the intersection point of the surface of $B_{\phi(x),\delta}$ and the line connecting $\phi(z)$ and $\phi(x)$. Then

$$\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}) \ge \nu_0(B_{y,\delta/2}) \ge (1-\epsilon)f_0(\phi(x))\lambda(B_{y,\delta/2}) = (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Let $\eta_N = 4\log(N_0/M)$. Since $M\log N_0/N_0 \to 0$, we can take N_0 large enough so that

$$\eta_N\frac{M}{N_0} = 4\frac{M}{N_0}\log\left(\frac{N_0}{M}\right) < (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Then for any $z \in \mathcal{X}$ such that $\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|}) \le \eta_N M/N_0$, we have $\|\phi(z)-\phi(x)\| < \delta$ since otherwise it would contradict the selection of η_N.

Upper bound. Let $\Phi_M(z)$ be the M-th order statistics of $\{\|\phi(X_i)-\phi(z)\|\}_{i=1}^{N_0}$. By the definition of $A_\phi(x)$, we have for any $\epsilon' \in (0,1)$,

$$\begin{aligned}
&\mathrm{pr}\Big(Z_1 \in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\Big)\\
\le&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 2\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1) + 2\|\widehat{\phi}-\phi\|_\infty\Big)\\
=&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
&+\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big). \qquad (6)
\end{aligned}$$

For the first term in (6), note that $[\phi(X_i)]_{i=1}^{N_0}$ are i.i.d from ν_0, and then $\nu_0(B_{\phi(Z_1),\|\phi(X_i)-\phi(Z_1)\|})$ are i.i.d from $U(0,1)$ and are independent of Z_1 by the probability integral transform. Then

$$\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\mathrm{pr}\Big((1-\epsilon')\|\phi(x)-\phi(Z_1)\| \le \Phi_M(Z_1)\Big)\\
=&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)\\
\le&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \le \eta_N\frac{M}{N_0}\Big) + \mathrm{pr}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big),
\end{aligned}$$

where $U_{(M)}$ is the M-th order statistic of N_0 independent random variables from $U(0,1)$.

By the selection of η_N, and taking ϵ' small and N_0 large enough, we have

$$\begin{aligned}
&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \le \eta_N\frac{M}{N_0}\Big)\\
\le&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x)-\phi(Z_1)\| \le \delta\Big).
\end{aligned}$$

Under the event $\{\|\phi(x)-\phi(Z_1)\|\le\delta\}$, we have

$$
\begin{aligned}
&\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|})-\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|})\\
=&\int_{B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}\setminus B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}} f_0(y)\mathrm{d}y\\
\le&(1+\epsilon)f_0(\phi(x))\lambda(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}\setminus B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|})\\
=&(1+\epsilon)f_0(\phi(x))V_m[1-(1-\epsilon')^d]\|\phi(x)-\phi(Z_1)\|^d\\
\le&(1+\epsilon)f_0(\phi(x))V_m d\epsilon'\|\phi(x)-\phi(Z_1)\|^d\\
=&(1+\epsilon)f_0(\phi(x))d\epsilon'\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})\\
\le&\frac{(1+\epsilon)f_0(\phi(x))d\epsilon'}{(1-\epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}),
\end{aligned}
$$

where V_m is the Lebesgue measure of the $m-$dimensional unit ball, and

$$
\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|})\ge\frac{(1-\epsilon)f_0(\phi(x))}{(1+\epsilon)f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}).
$$

From the probability integral transform, we have $\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})$ is from $U(0,1)$ and then for $U\sim U(0,1)$,

$$
\begin{aligned}
&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|})\le\nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}),\|\phi(x)-\phi(Z_1)\|\le\delta\Big)\\
\le&\mathrm{pr}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|})\le\nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)\\
=&\mathrm{pr}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}U\le U_{(M)}\Big).
\end{aligned}
$$

We can check that

$$
\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{pr}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}U\le U_{(M)}\Big)=\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.
$$

Note that $\eta_N\to\infty$ as $N_0\to\infty$ since $M/N_0\to 0$. Then from the Chernoff bound and for N_0 sufficiently large, we have

$$
\begin{aligned}
&\frac{N_0}{M}\mathrm{pr}\Big(U_{(M)}>\eta_N\frac{M}{N_0}\Big)=\frac{N_0}{M}\mathrm{pr}\Big(\mathrm{Bin}\Big(N_0,\eta_N\frac{M}{N_0}\Big)\le M\Big)\\
\le&\frac{N_0}{M}\exp\Big((1+\log\eta_N-\eta_N)M\Big)\le\frac{N_0}{M}\exp\Big(-\frac{1}{2}\eta_N M\Big)=\Big(\frac{N_0}{M}\Big)^{1-2M}.
\end{aligned}
$$

Since $M/N_0\to 0$ and $M\ge 1$, we then obtain

$$
\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{pr}\Big(U_{(M)}>\eta_N\frac{M}{N_0}\Big)=0.
$$

Then we obtain

$$
\begin{aligned}
&\limsup_{N_0\to\infty}\frac{N_0}{M}\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\|-4\|\widehat{\phi}-\phi\|_\infty\le\Phi_M(Z_1),4\|\widehat{\phi}-\phi\|_\infty\le\epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.
\end{aligned}
$$

For the second term in (6), we have for any $\delta>0$,

$$
\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\|-4\|\widehat{\phi}-\phi\|_\infty\le\Phi_M(Z_1),4\|\widehat{\phi}-\phi\|_\infty>\epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\|<4\|\widehat{\phi}-\phi\|_\infty/\epsilon'\Big)
\end{aligned}
$$

$$\leq \mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \leq \delta\Big) + \mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big).$$

Note that $\widehat{\phi}$ may depend on Z_1. Recall that $\widehat{\phi}_{Z_1\to z}$ is the estimator of ϕ replacing Z_1 by some $z \in \mathcal{Z}$. Then $\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty$ is independent of Z_1 and then we have

$$\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \leq \delta\Big)\\
\leq&\mathrm{pr}\Big(\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq V_m(4/\epsilon')^d\|\widehat{\phi}-\phi\|_\infty^d, \|\phi(x)-\phi(Z_1)\|\leq\delta\Big)\\
\leq&\mathrm{pr}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq (1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\|\widehat{\phi}-\phi\|_\infty^d\Big)\\
\leq&\mathrm{pr}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq (1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty^d\Big)\\
=&E[\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty^d\}\wedge 1].
\end{aligned}$$

By the Markov inequality,

$$\mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big) \leq \delta^{-d}(4/\epsilon')^d E(\|\widehat{\phi}-\phi\|_\infty^d) \leq \delta^{-d}(4/\epsilon')^d E(\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty^d).$$

By Assumption 11, we have

$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\|-4\|\widehat{\phi}-\phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big) = 0.$$

By (6) and ϵ, ϵ' are arbitrary, we obtain

$$\limsup_{N_0\to\infty}\frac{N_0}{M}\mathrm{pr}\Big(Z_1 \in A_\phi(x)\Big) \leq \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Lower bound. For any $\epsilon' \in (0,1)$, we have

$$\begin{aligned}
&\mathrm{pr}\Big(Z_1\in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \leq \widehat{\Phi}_M(Z_1)\Big)\\
&\geq \mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| + 2\|\widehat{\phi}-\phi\|_\infty \leq \Phi_M(Z_1) - 2\|\widehat{\phi}-\phi\|_\infty\Big)\\
&\geq \mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| + 4\|\widehat{\phi}-\phi\|_\infty \leq \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \leq \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
&\geq \mathrm{pr}\Big((1+\epsilon')\|\phi(x)-\phi(Z_1)\| \leq \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \leq \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
&\geq \mathrm{pr}\Big((1+\epsilon')\|\phi(x)-\phi(Z_1)\| \leq \Phi_M(Z_1)\Big) - \mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
&= \mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big) - \mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
&\geq \mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N\frac{M}{N_0}\Big) - \mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big).
\end{aligned}$$

Note that by the selection of η_N, and taking ϵ' small and N_0 large enough, we have

$$\begin{aligned}
&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N\frac{M}{N_0}\Big)\\
=&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N\frac{M}{N_0}, \|\phi(x)-\phi(Z_1)\|\leq\delta\Big).
\end{aligned}$$

Under the event $\{\|\phi(x)-\phi(Z_1)\|\leq\delta\}$, we have $B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|} \subset B_{\phi(x),3\delta}$, and then

$$\begin{aligned}
&\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) - \nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|})\\
=&\int_{B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}\setminus B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}} f_0(y)\mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
&\leq (1+\epsilon) f_0(\phi(x)) \lambda(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|} \setminus B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}) \\
&= (1+\epsilon) f_0(\phi(x)) V_m [(1+\epsilon')^d - 1] \|\phi(x)-\phi(Z_1)\|^d \\
&\leq (1+\epsilon) f_0(\phi(x)) V_m d\epsilon'(1+\epsilon')^{d-1} \|\phi(x)-\phi(Z_1)\|^d \\
&= (1+\epsilon) f_0(\phi(x)) d\epsilon'(1+\epsilon')^{d-1} \lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \\
&\leq \frac{(1+\epsilon) f_0(\phi(x)) d\epsilon'(1+\epsilon')^{d-1}}{(1-\epsilon) f_1(\phi(x))} \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}),
\end{aligned}
$$

and

$$
\nu_0(B_{\phi(Z_1),\|\phi(x)-\phi(Z_1)\|}) \leq \frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))} \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}).
$$

Then

$$
\begin{aligned}
&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_1),(1+\epsilon')\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}, \|\phi(x)-\phi(Z_1)\| \leq \delta\Big) \\
&\geq \mathrm{pr}\Big(\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}, \|\phi(x)-\phi(Z_1)\| \leq \delta\Big) \\
&= \mathrm{pr}\Big(\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}) \leq \eta_N \frac{M}{N_0}\Big) \\
&\geq \mathrm{pr}\Big(\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \leq \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big) - \mathrm{pr}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) \\
&= \mathrm{pr}\Big(\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)U \leq U_{(M)}\Big) - \mathrm{pr}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big).
\end{aligned}
$$

The second last equality is from the fact that for $z \in \mathcal{X}$ such that $\|\phi(z)-\phi(x)\| > \delta$,

$$
\begin{aligned}
&\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)\nu_1(B_{\phi(x),\|\phi(x)-\phi(z)\|}) \geq \frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\nu_1(B_{\phi(x),\delta}) \\
\geq\ &\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))} f_1(\phi(x))(1-\epsilon)\lambda(B_{0,\delta}) > \eta_N \frac{M}{N_0}
\end{aligned}
$$

by the selection of η_N.

We can check that

$$
\lim_{N_0\to\infty} \frac{N_0}{M}\mathrm{pr}\Big(\frac{(1+\epsilon) f_0(\phi(x))}{(1-\epsilon) f_1(\phi(x))}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)U \leq U_{(M)}\Big) = \frac{1-\epsilon}{1+\epsilon}\Big(1+d\epsilon'(1+\epsilon')^{d-1}\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}.
$$

By ϵ, ϵ' are arbitrary, we obtain

$$
\liminf_{N_0\to\infty} \frac{N_0}{M}\mathrm{pr}\Big(Z_1 \in A_\phi(x)\Big) \geq \frac{f_1(\phi(x))}{f_0(\phi(x))}.
$$

Combining the upper bound and the lower bound yields

$$
\lim_{N_0\to\infty} \frac{N_0}{M}\mathrm{pr}\Big(Z_1 \in A_\phi(x)\Big) = \frac{f_1(\phi(x))}{f_0(\phi(x))}.
$$

Case II. $f_1(\phi(x)) = 0$.

For any $\epsilon \in (0,1)$, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathcal{X}$ with $\|\phi(z)-\phi(x)\| \leq 3\delta$, we have $|f_0(\phi(z)) - f_0(\phi(x))| \leq \epsilon f_0(\phi(x))$ and $f_1(\phi(z)) \leq \epsilon$. Then for any $z \in \mathcal{X}$ with $\|\phi(z)-\phi(x)\| \leq \delta$, we have

$$
\Big|\frac{\nu_0(B_{\phi(z),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \leq \epsilon f_0(\phi(x)), \quad \Big|\frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}\Big| \leq \epsilon.
$$

We consider the same decomposition as (6). For the first term in (6), we still have

$$\begin{aligned}
&\operatorname{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\operatorname{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x)-\phi(Z_1)\|\le\delta\Big) + \operatorname{pr}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big).
\end{aligned}$$

Note that for any $z \in \mathcal{X}$ with $\|\phi(x)-\phi(z)\|\le\delta$,

$$\Big|\frac{\nu_0(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})} - f_0(\phi(x))\Big| \le \epsilon f_0(\phi(x)),$$

and

$$\begin{aligned}
&\frac{\nu_1(B_{\phi(x),\|\phi(x)-\phi(z)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})} = \frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}\frac{\nu_1(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}\\
\le&\ \frac{\lambda(B_{\phi(x),\|\phi(z)-\phi(x)\|})}{\lambda(B_{\phi(z),(1-\epsilon')\|\phi(z)-\phi(x)\|})}\epsilon = (1-\epsilon')^{-d}\epsilon.
\end{aligned}$$

Then

$$\begin{aligned}
&\operatorname{pr}\Big(\nu_0(B_{\phi(Z_1),(1-\epsilon')\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)}), \|\phi(x)-\phi(Z_1)\|\le\delta\Big)\\
\le&\operatorname{pr}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \nu_0(B_{\phi(Z_1),\Phi_M(Z_1)})\Big)\\
=&\operatorname{pr}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))U \le U_{(M)}\Big).
\end{aligned}$$

We can check that

$$\lim_{N_0\to\infty}\frac{N_0}{M}\operatorname{pr}\Big((1-\epsilon')^d\epsilon^{-1}(1-\epsilon)f_0(\phi(x))U \le U_{(M)}\Big) = \epsilon(1-\epsilon')^{-d}(1-\epsilon)^{-1}\frac{1}{f_0(\phi(x))}.$$

Then we obtain

$$\begin{aligned}
&\limsup_{N_0\to\infty}\frac{N_0}{M}\operatorname{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\epsilon(1-\epsilon')^{-d}(1-\epsilon)^{-1}\frac{1}{f_0(\phi(x))}.
\end{aligned}$$

For the second term in (6), we still have

$$\begin{aligned}
&\operatorname{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_1), 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_1)\|\Big)\\
\le&\ \operatorname{pr}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta\Big) + \operatorname{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big).
\end{aligned}$$

Note that

$$\begin{aligned}
&\operatorname{pr}\Big(\|\phi(x)-\phi(Z_1)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta\Big)\\
\le&\operatorname{pr}\Big(\lambda(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le V_m(4/\epsilon')^d\|\widehat{\phi}-\phi\|_\infty^d, \|\phi(x)-\phi(Z_1)\|\le\delta\Big)\\
\le&\operatorname{pr}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \epsilon V_m(4/\epsilon')^d\|\widehat{\phi}-\phi\|_\infty^d\Big)\\
\le&\operatorname{pr}\Big(\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) \le \epsilon V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty^d\Big)\\
=&E[\{\epsilon V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}}\|\widehat{\phi}_{Z_1\to z}-\phi\|_\infty^d\}\wedge 1].
\end{aligned}$$

By ϵ, ϵ' are arbitrary, we obtain

$$\lim_{N_0\to\infty}\frac{N_0}{M}\operatorname{pr}\Big(Z_1 \in A_\phi(x)\Big) = 0 = \frac{f_1(\phi(x))}{f_0(\phi(x))}.$$

Part II. We then consider the general case where p is a fixed positive integer and Assumption 11 holds for p. We only consider the case where $f_1(\phi(x)) > 0$. The case with $f_1(\phi(x)) = 0$ can be established in a similar way.

Let $\eta_N = \eta_{N,p} = 4p\log(N_0/M)$. We also take N_0 sufficiently large so that

$$\eta_N \frac{M}{N_0} = 4p\frac{M}{N_0}\log\left(\frac{N_0}{M}\right) < (1-\epsilon)f_0(\phi(x))\lambda(B_{0,\delta/2}).$$

Then

$$\begin{aligned}
&\mathrm{pr}\Big(Z_1,\dots,Z_p \in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_k)\| \le \widehat{\Phi}_M(Z_k), \forall k \in \{1,\dots,p\}\Big)\\
&\le \mathrm{pr}\Big(\|\phi(x)-\phi(Z_k)\| - 2\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_k) + 2\|\widehat{\phi}-\phi\|_\infty, \forall k \in \{1,\dots,p\}\Big)\\
&= \sum_{S\subset\{1,\dots,p\}} \mathrm{pr}\Big(\|\phi(x)-\phi(Z_k)\| - 4\|\widehat{\phi}-\phi\|_\infty \le \Phi_M(Z_k), 4\|\widehat{\phi}-\phi\|_\infty \le \epsilon'\|\phi(x)-\phi(Z_k)\| \text{ for } k\in S,\\
&4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\|\phi(x)-\phi(Z_k)\| \text{ for } k \notin S\Big)\\
&\le \sum_{S\subset\{1,\dots,p\}} \mathrm{pr}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \le \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k\in S, 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\Big).
\end{aligned} \tag{7}$$

Let $W_k = \nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})$ and $V_k = \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)})$ for any $k \in \{1,\dots,p\}$. Then $[W_k]_{k=1}^p$ are i.i.d. since $[Z_k]_{k=1}^p$ are i.i.d.. For any $k\in\{1,\dots,p\}$ and $Z_k \in \mathcal{X}$ given, $V_k \mid Z_k$ has the same distribution as $U_{(M)}$. Then for any $k \in \{1,\dots,p\}$, V_k has the same distribution as $U_{(M)}$, and V_k is independent of Z_k.

Fix $S \subset \{1,\dots,p\}$. Let $W_{\max} = \max_{k\in S} W_k$ and $V_{\max} = \max_{k\in S} V_k$. Then

$$\begin{aligned}
&\mathrm{pr}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \le \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k\in S, 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\Big)\\
&\le \mathrm{pr}\Big(W_{\max} < V_{\max}, 4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\Big)\\
&\le \mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N\frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta\Big)\\
&+ \mathrm{pr}\Big(V_{\max} > \eta_N\frac{M}{N_0}\Big) + \mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big).
\end{aligned} \tag{8}$$

For the first term in (8), by the selection of η_N, and taking $\epsilon' < 1/2$ and N_0 large enough, we have

$$\begin{aligned}
&\mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N\frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta\Big)\\
&\le \mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N\frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta, \max_{k\in S}\|\phi(x)-\phi(Z_k)\| \le \delta\Big).
\end{aligned}$$

Let $W'_k = \nu_0(B_{\phi(Z_k),\|\phi(x)-\phi(Z_k)\|})$ and $W'_{\max} = \max_{k\in S} W'_k$. Under the event $\{\max_{k\in S}\|\phi(x)-\phi(Z_k)\| \le \delta\}$, we have

$$\begin{aligned}
W'_{\max} - W_{\max} &\le \max_{k\in S}\{\nu_0(B_{\phi(Z_k),\|\phi(x)-\phi(Z_k)\|}) - \nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})\}\\
&\le \frac{(1+\epsilon)f_0(\phi(x))d\epsilon'}{(1-\epsilon)f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}).
\end{aligned}$$

and

$$W'_{\max} \ge \frac{(1-\epsilon)f_0(\phi(x))}{(1+\epsilon)f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}).$$

On the other hand, recall that $\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}$ is the estimator replacing $(Z_1,\dots,Z_p)$ by z for $z\in\mathcal{Z}^p$. Then $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty$ is independent of $(Z_1,\dots,Z_p)$. Note that

$$\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|<4\|\widehat{\phi}-\phi\|_\infty/\epsilon'\le\delta$$

implies that

$$\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})\le(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d.$$

Then

$$\begin{aligned}
&\mathrm{pr}\Big(W_{\max}<V_{\max}\le\eta_N\frac{M}{N_0},\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|<4\|\widehat{\phi}-\phi\|_\infty/\epsilon'\le\delta,\max_{k\in S}\|\phi(x)-\phi(Z_k)\|\le\delta\Big)\\
&\le\mathrm{pr}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})<V_{\max},\\
&\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})\le(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\Big)\\
&=E\Big\{\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})<V_{\max}\Big)\\
&\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\Big\},
\end{aligned}$$

since $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty$, $V_{\max}$ and $[Z_k]_{k\in S}$ are all independent with $[Z_k]_{k\notin S}$.

Note that

$$\begin{aligned}
&E\Big\{\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})<V_{\max}\Big)\\
&\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\Big\}\\
&=\int_0^1|S|u^{|S|-1}E\Big\{\mathbb{1}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}u<V_{\max}\Big)\\
&\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\,\Big|\,\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})=u\Big\}\mathrm{d}u\\
&=|S|\Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}\Big\}^{|S|}\int_0^{(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon')\frac{f_0(\phi(x))}{f_1(\phi(x))}\frac{N_0}{M}}u^{|S|-1}E\Big\{\mathbb{1}\Big(V_{\max}>\frac{M}{N_0}u\Big)\\
&\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\,\Big|\\
&\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})=\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big\}\mathrm{d}u.
\end{aligned}\tag{9}$$

We split the above integral into two parts using 1. For the first part, note that $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty$ is independent with $\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})$. Then

$$\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1u^{|S|-1}E\Big\{\mathbb{1}\Big(V_{\max}>\frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\,\Big|\\
&\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})=\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big\}\mathrm{d}u\\
&\le\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1u^{|S|-1}E\Big\{\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge1\Big)^{p-|S|}\Big\}\mathrm{d}u.
\end{aligned}$$

If $|S| = p$, we have

$$
\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 u^{|S|-1}E\Big\{\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\Big\}\mathrm{d}u\\
&=\int_0^1 u^{p-1}\mathrm{d}u=\frac{1}{p}. \qquad (10)
\end{aligned}
$$

If $|S| < p$, by Assumption 11, we have

$$
\begin{aligned}
&\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 u^{|S|-1}E\Big\{\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\Big\}\mathrm{d}u\\
&\lesssim \limsup_{N_0\to\infty}E\Big\{\Big(\frac{N_0}{M}\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\Big)^{p-|S|}\Big\}=0. \qquad (11)
\end{aligned}
$$

For the second part, we have

$$
\begin{aligned}
&\int_1^{(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon')\frac{f_0(\phi(x))}{f_1(\phi(x))}\frac{N_0}{M}} u^{|S|-1}E\Big\{\mathbb{1}\Big(V_{\max}>\frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\ \Big|\\
&\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})=\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big\}\mathrm{d}u\\
&\le\int_1^\infty u^{|S|-1}E\Big\{\mathbb{1}\Big(V_{\max}>\frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\ \Big|\\
&\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|})=\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\frac{M}{N_0}u\Big\}\mathrm{d}u\\
&\le\sum_{k\in S}\int_1^\infty u^{|S|-1}E\Big\{\mathbb{1}\Big(V_k>\frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\Big\}\mathrm{d}u, \qquad (12)
\end{aligned}
$$

where the last step is from the fact that V_k and $\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty$ are independent of $[Z_k]_{k\in S}$ for any $k\in S$.

For any $k\in S$, by the Hölder inequality,

$$
\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_1^\infty u^{|S|-1}E\Big\{\mathbb{1}\Big(V_k>\frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d \sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\Big\}\mathrm{d}u\\
&\lesssim\int_1^\infty u^{|S|-1}E\Big\{\mathbb{1}\Big(V_k>\frac{M}{N_0}u\Big)\Big(\frac{N_0}{M}\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\Big)^{p-|S|}\Big\}\mathrm{d}u\\
&\le\int_1^\infty u^{|S|-1}\Big\{\mathrm{pr}\Big(V_k>\frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\Big[E\Big\{\Big(\frac{N_0}{M}\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\Big)^p\Big\}\Big]^{\frac{p-|S|}{p}}\mathrm{d}u\\
&=\Big[E\Big\{\Big(\frac{N_0}{M}\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\Big)^p\Big\}\Big]^{\frac{p-|S|}{p}}\int_1^\infty u^{|S|-1}\Big\{\mathrm{pr}\Big(V_k>\frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u.
\end{aligned}
$$

Using the Chernoff bound,

$$
\begin{aligned}
&\int_1^\infty u^{|S|-1}\Big\{\mathrm{pr}\Big(V_k>\frac{M}{N_0}u\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u=\int_0^\infty(1+u)^{|S|-1}\Big\{\mathrm{pr}\Big(U_{(M)}>\frac{M}{N_0}(1+u)\Big)\Big\}^{\frac{|S|}{p}}\mathrm{d}u\\
&\le\int_0^\infty(1+u)^{|S|-1}(1+u)^{M|S|/p}\exp(-uM|S|/p)\mathrm{d}u\\
&=\exp(M|S|/p)\int_1^\infty u^{M|S|/p+|S|-1}\exp(-uM|S|/p)\mathrm{d}u
\end{aligned}
$$

$$
\begin{aligned}
&\le \exp(M|S|/p)\int_0^\infty u^{M|S|/p+|S|-1}\exp(-uM|S|/p)\mathrm{d}u\\
&= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}}\Gamma(M|S|/p+|S|)\\
&= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}}(M|S|/p+1)^{|S|-1}\Gamma(M|S|/p+1)(1+o(1))\\
&= \frac{\exp(M|S|/p)}{(M|S|/p)^{M|S|/p+|S|}}(M|S|/p+1)^{|S|-1}(2\pi M|S|/p)^{1/2}\Big(\frac{M|S|/p}{e}\Big)^{M|S|/p}(1+o(1))\\
&= (2\pi)^{1/2}(M|S|/p)^{-1/2}\Big(1+\frac{p}{M|S|}\Big)^{|S|-1}(1+o(1)),
\end{aligned}
$$

where the last three steps are from Stirling's approximation using $M \to \infty$.

By $M \to \infty$ and Assumption 11, we have

$$
\lim_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^{p-|S|}\int_1^\infty u^{|S|-1}E\Big\{\mathbb{1}\Big(V_k > \frac{M}{N_0}u\Big)\Big(\{(1+\epsilon)f_1(\phi(x))V_m(4/\epsilon')^d\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^d\}\wedge 1\Big)^{p-|S|}\Big\}\mathrm{d}u
$$
$$
= 0. \tag{13}
$$

Combining (10), (11), (12), (13) by (9) yields

$$
\begin{aligned}
&\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N\frac{M}{N_0}, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\| < 4\|\widehat{\phi}-\phi\|_\infty/\epsilon' \le \delta\Big)\\
&\le \frac{1}{p}|S|\Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^{|S|}\mathbb{1}(|S|=p) = \Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^{p}\mathbb{1}(|S|=p).
\end{aligned}
$$

For the second term in (8), by the Chernoff bound,

$$
\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(V_{\max} > \eta_N\frac{M}{N_0}\Big) \le |S|\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big) = 0.
$$

For the third term in (8),

$$
\begin{aligned}
\mathrm{pr}(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta) &\le \delta^{-pd}(4/\epsilon')^{pd}E(\|\widehat{\phi}-\phi\|_\infty^{pd})\\
&\le \delta^{-pd}(4/\epsilon')^{pd}E(\sup_{z\in\mathcal{Z}^p}\|\widehat{\phi}_{(Z_1,\dots,Z_p)\to z}-\phi\|_\infty^{pd}).
\end{aligned}
$$

By Assumption 11,

$$
\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(4\|\widehat{\phi}-\phi\|_\infty/\epsilon' > \delta\Big) = 0.
$$

Then we obtain for any $S \subset \{1,\dots,p\}$,

$$
\begin{aligned}
&\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(\nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|}) \le \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)}) \text{ for } k\in S,\\
&4\|\widehat{\phi}-\phi\|_\infty > \epsilon'\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\Big)\\
&\le \Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^p\mathbb{1}(|S|=p),
\end{aligned}
$$

and then by (7),

$$
\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{pr}\Big(Z_1,\dots,Z_p\in A_\phi(x)\Big) \le \Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^p.
$$

By ϵ, ϵ' are arbitrary, we obtain

$$\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(Z_1, \ldots, Z_p \in A_\phi(x)\Big) \le \Big(\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big)^p.$$

A matched lower bound is directly from the Hölder inequality.

Then we obtain

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(Z_1, \ldots, Z_p \in A_\phi(x)\Big) = \Big(\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big)^p.$$

Part III. We consider the simple case where $p = 1$ and Assumption 12 holds. We only consider the case where $f_1(\phi(x)) > 0$, while the case where $f_1(\phi(x)) = 0$ is similar.

Upper bound. Note that

$$\begin{aligned}
&\mathrm{pr}\Big(Z_1 \in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\Big)\\
&\le \mathrm{pr}\Big(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1)\Big).
\end{aligned}$$

The inequality is from the fact that under the event $\{\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\}$, if $\|\phi(x) - \phi(Z_1)\| - 2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \Phi_M(Z_1)$, then there exists a set $S \subset \{1, \ldots, N_0\}$ such that $|S| \ge M$ and for any $i \in S$, $\|\phi(x) - \phi(Z_1)\| - 2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \|\phi(X_i) - \phi(Z_1)\|$. For these $i \in S$, we then have $\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| > \|\widehat{\phi}(X_i) - \widehat{\phi}(Z_1)\|$ since $\|\phi(x) - \phi(Z_1)\| > \|\phi(X_i) - \phi(Z_1)\|$. Then $\|\widehat{\phi}(x) - \widehat{\phi}(Z_1)\| > \widehat{\Phi}_M(Z_1)$ using $|S| \ge M$, which contradicts the event we assume.

For any $\epsilon \in (0, 1)$, we decompose as

$$\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1)\Big)\\
&= \mathrm{pr}\Big(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\\
&\quad 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \epsilon\|\phi(x) - \phi(Z_1)\|\Big)\\
&\quad + \mathrm{pr}\Big(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\\
&\quad 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon\|\phi(x) - \phi(Z_1)\|\Big).
\end{aligned}\tag{14}$$

For the first term in (14),

$$\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x) - \phi(Z_1)\| - 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_1),\\
&\quad 2 \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \epsilon\|\phi(x) - \phi(Z_1)\|\Big)\\
&\le \mathrm{pr}\Big((1-\epsilon)\|\phi(x) - \phi(Z_1)\| \le \Phi_M(Z_1)\Big),
\end{aligned}$$

and then can be handled in the same way as the upper bound part in Part I.

For the second term in (14), note that for any $u \in (0, 1)$, conditional on $\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u$ and under $\|\phi(x) - \phi(Z_1)\| \le \delta$, we have $u \le (1+\epsilon) f_1(\phi(x)) V_m \|\phi(x) - \phi(Z_1)\|^m$, and then $\|\phi(x) - \phi(Z_1)\| \ge [u/\{(1+\epsilon) f_1(\phi(x)) V_m\}]^{1/m}$. Then for any $u \in (0, 1)$,

$$\mathrm{pr}\Big(\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon\|\phi(x) - \phi(Z_1)\|, \|\phi(x) - \phi(Z_1)\| \le \delta$$

$$
\begin{aligned}
&\Big| \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u\Big) \\
&\le \mathrm{pr}\Big(\sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_1)\|} \sup_{z\in\mathcal{Z}} \|(\widehat{\phi}_{Z_1\to z} - \phi)(s) - (\widehat{\phi}_{Z_1\to z} - \phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\| \le \delta \\
&\Big| \nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_1)\|}) = u\Big) \\
&\le T_\epsilon([u/\{(1+\epsilon)f_1(\phi(x))V_m\}]^{1/m}),
\end{aligned}
$$

where the last step is from the fact that $\sup_{z\in\mathcal{Z}}\|(\widehat{\phi}_{Z_1\to z}-\phi)(s)-(\widehat{\phi}_{Z_1\to z}-\phi)(t)\|$ does not depend on Z_1 together with Assumption 12.

Then

$$
\begin{aligned}
&\mathrm{pr}\Big(\|\phi(x)-\phi(Z_1)\| - 2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| \le \Phi_M(Z_1), \\
&2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|\Big) \\
&\le \mathrm{pr}\Big(2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|\Big) \\
&\le \mathrm{pr}\Big(2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\|\le\delta\Big) \\
&+ \mathrm{pr}\Big(2\sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_1)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon\|\phi(x)-\phi(Z_1)\|, \|\phi(x)-\phi(Z_1)\|>\delta\Big) \\
&\le \int_0^1 T_{\epsilon/2}([u/\{(1+\epsilon)f_1(\phi(x))V_m\}]^{1/m})\mathrm{d}u + \mathrm{pr}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon\delta/4\Big) \\
&= (1+\epsilon)f_1(\phi(x))V_m \int_0^{1/\{(1+\epsilon)f_1(\phi(x))V_m\}} T_{\epsilon/2}(u^{1/m})\mathrm{d}u + \mathrm{pr}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon\delta/4\Big).
\end{aligned}
$$

By Assumption 12 and ϵ is arbitrary, using (14), we obtain

$$
\limsup_{N_0\to\infty} \frac{N_0}{M}\mathrm{pr}\Big(Z_1\in A_\phi(x)\Big) \le \frac{f_1(\phi(x))}{f_0(\phi(x))}.
$$

Lower bound. For any $\epsilon\in(0,1)$,

$$
\begin{aligned}
&\mathrm{pr}\Big(Z_1\in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)\Big) \\
&\ge \mathrm{pr}\Big(\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1), \sup_{\delta\ge\|\phi(x)-\phi(Z_1)\|} \delta^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\delta} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| \le \epsilon\Big) \\
&\ge \mathrm{pr}\Big((1+\epsilon)\|\phi(x)-\phi(Z_1)\| \le (1-\epsilon)\Phi_M(Z_1), \sup_{\delta\ge\|\phi(x)-\phi(Z_1)\|} \delta^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\delta} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| \le \epsilon\Big).
\end{aligned}
$$

The last inequality is from the fact that under the event $\{(1+\epsilon)\|\phi(x)-\phi(Z_1)\| \le (1-\epsilon)\Phi_M(Z_1)\}$, there exists a set $S\subset\{1,\ldots,N_0\}$ such that $|S|\ge N_0-M$ and for any $i\in S$, $(1+\epsilon)\|\phi(x)-\phi(Z_1)\| \le (1-\epsilon)\|\phi(X_i)-\phi(Z_1)\|$. Under the event that $\{\sup_{\delta\ge\|\phi(x)-\phi(Z_1)\|}\delta^{-1}\sup_{\|\phi(s)-\phi(t)\|\le\delta}\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| \le \epsilon\}$, for these $i\in S$, we then have $\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \le \|\widehat{\phi}(X_i)-\widehat{\phi}(Z_1)\|$ since $\|\phi(x)-\phi(Z_1)\| \le \|\phi(X_i)-\phi(Z_1)\|$ for $i\in S$. Then $\|\widehat{\phi}(x)-\widehat{\phi}(Z_1)\| \le \widehat{\Phi}_M(Z_1)$.

Then

$$
\begin{aligned}
\mathrm{pr}\Big(Z_1\in A_\phi(x)\Big) \ge{}& \mathrm{pr}\Big((1+\epsilon)\|\phi(x)-\phi(Z_1)\| \le (1-\epsilon)\Phi_M(Z_1)\Big) \\
&- \mathrm{pr}\Big(\sup_{\delta\ge\|\phi(x)-\phi(Z_1)\|} \delta^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\delta} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon\Big).
\end{aligned}
$$

The first term can be handled in the same way as the lower bound part in Part I. The second term can be handled in the same way as the second term in (14) in this proof. Then we can obtain a matched lower bound.

Part IV. We then consider the general case where p is a fixed positive integer and Assumption 12 holds. We only consider the case where $f_1(\phi(x)) > 0$, while the case where $f_1(\phi(x)) = 0$ is similar.

For any $\epsilon' \in (0, 1)$, we have

$$
\begin{aligned}
&\mathrm{pr}\Big(Z_1, \ldots, Z_p \in A_\phi(x)\Big) = \mathrm{pr}\Big(\|\widehat{\phi}(x) - \widehat{\phi}(Z_k)\| \le \widehat{\Phi}_M(Z_k), \forall k \in \{1, \ldots, p\}\Big)\\
&\le \mathrm{pr}\Big(\|\phi(x) - \phi(Z_k)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_k), \forall k \in \{1, \ldots, p\}\Big)\\
&= \sum_{S \subset \{1,\ldots,p\}} \mathrm{pr}\Big(\|\phi(x) - \phi(Z_k)\| - 2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \Phi_M(Z_k),\\
&2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| \le \epsilon' \|\phi(x) - \phi(Z_k)\| \text{ for } k \in S,\\
&2 \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon' \|\phi(x) - \phi(Z_k)\| \text{ for } k \notin S\Big)\\
&\le \sum_{S \subset \{1,\ldots,p\}} \mathrm{pr}\Big((1 - \epsilon')\|\phi(x) - \phi(Z_k)\| \le \Phi_M(Z_k) \text{ for } k \in S,\\
&2\|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon' \text{ for } k \notin S\Big).
\end{aligned}
$$

If $|S| = p$, we have in the same way as the upper bound part in Part III that for any $\epsilon \in (0, 1)$,

$$
\begin{aligned}
&\limsup_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big((1 - \epsilon')\|\phi(x) - \phi(Z_k)\| \le \Phi_M(Z_k) \text{ for } k \in S,\\
&2\|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon' \text{ for } k \notin S\Big)\\
&\le \Big\{\Big(\frac{1-\epsilon}{1+\epsilon} - \frac{1+\epsilon}{1-\epsilon} d\epsilon'\Big)^{-1} \frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^p.
\end{aligned}
$$

Now we consider $|S| < p$. Recall that $W_k = \nu_0(B_{\phi(Z_k),(1-\epsilon')\|\phi(x)-\phi(Z_k)\|})$ and $V_k = \nu_0(B_{\phi(Z_k),\Phi_M(Z_k)})$ for any $k \in \{1, \ldots, p\}$. Fix $S \subset \{1, \ldots, p\}$. Recall that $W_{\max} = \max_{k\in S} W_k$ and $V_{\max} = \max_{k\in S} V_k$. We have

$$
\begin{aligned}
&\mathrm{pr}\Big((1 - \epsilon')\|\phi(x) - \phi(Z_k)\| \le \Phi_M(Z_k) \text{ for } k \in S,\\
&2\|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon' \text{ for } k \notin S\Big)\\
&\le \mathrm{pr}\Big(W_{\max} < V_{\max}, 2 \min_{k\notin S} \|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon'\Big)\\
&\le \mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N \frac{M}{N_0}, 2 \min_{k\notin S} \|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon'\Big)\\
&+ \mathrm{pr}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big).
\end{aligned}
$$

Note that

$$
2 \min_{k\notin S} \|\phi(x) - \phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon'
$$

implies that

$$
(\max_{k\notin S} \|\phi(x) - \phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\| \le \max_{k\notin S} \|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi} - \phi)(s) - (\widehat{\phi} - \phi)(t)\| > \epsilon'/2.
$$

Then

$$\mathrm{pr}\Big(W_{\max} < V_{\max} \le \eta_N \frac{M}{N_0}, 2\min_{k\notin S}\|\phi(x)-\phi(Z_k)\|^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon'\Big)$$
$$\le \mathrm{pr}\Big(\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)\frac{f_0(\phi(x))}{f_1(\phi(x))}\max_{k\in S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) < V_{\max},$$
$$(\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon'/2\Big).$$

In the same way as the upper bound part in Part II, it suffices to show that

$$\lim_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{pr}\Big((\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon'/2\Big) = 0.$$

For any $u\in(0,1)$, conditional on $\max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = u$ and under $\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\le\delta$, we have

$$\max_{k\notin S}\|\phi(x)-\phi(Z_k)\| \ge [u/\{(1+\epsilon)f_1(\phi(x))V_m\}]^{1/m}.$$

Then by Assumption 12,

$$\Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{pr}\Big((\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|} \|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon'/2\Big)$$
$$= \Big(\frac{N_0}{M}\Big)^{p-|S|}\int_0^1 (p-|S|)u^{p-|S|-1}\mathrm{pr}\Big((\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|)^{-1} \sup_{\|\phi(s)-\phi(t)\|\le\max_{k\notin S}\|\phi(x)-\phi(Z_k)\|}$$
$$\|(\widehat{\phi}-\phi)(s)-(\widehat{\phi}-\phi)(t)\| > \epsilon'/2, \max_{k\notin S}\|\phi(x)-\phi(Z_k)\|\le\delta \,\Big|\, \max_{k\notin S}\nu_1(B_{\phi(x),\|\phi(x)-\phi(Z_k)\|}) = u\Big)\mathrm{d}u$$
$$+ \Big(\frac{N_0}{M}\Big)^{p-|S|}\mathrm{pr}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon'\delta/4\Big)$$
$$\le \Big(\frac{N_0}{M}\Big)^{p-|S|}\Big\{\int_0^1 (p-|S|)u^{p-|S|-1}T_{\epsilon'}(\{u/(\|f_1\|_\infty V_m)\}^{1/m})\mathrm{d}u + \mathrm{pr}\Big(\|\widehat{\phi}-\phi\|_\infty > \epsilon'\delta/4\Big)\Big\} = o(1).$$

Then

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(Z_1,\ldots,Z_p\in A_\phi(x)\Big) \le \Big\{\Big(\frac{1-\epsilon}{1+\epsilon}-\frac{1+\epsilon}{1-\epsilon}d\epsilon'\Big)^{-1}\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big\}^p.$$

By ϵ,ϵ' are arbitrary, we obtain

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p \mathrm{pr}\Big(Z_1,\ldots,Z_p\in A_\phi(x)\Big) \le \Big(\frac{f_1(\phi(x))}{f_0(\phi(x))}\Big)^p.$$

A matched lower bound is directly from the Hölder inequality.

C.4. *Proof of Theorem 4*

Consider any $\epsilon\in(0,1)$ be given. From Assumption 13, $\mathcal{X}$ is compact, and then $\mathcal{Z}$ is also compact. Since f_0, f_1 are continuous over their compact supports, they are uniformly continuous, that is, there exists $\delta>0$ such that for any $x,z\in\mathcal{Z}$ with $\|\phi(z)-\phi(x)\|\le 3\delta$, we have $|f_1(\phi(z))-f_1(\phi(x))|\le\epsilon^2$, and for any $x,z\in\mathcal{X}$ with $\|\phi(z)-\phi(x)\|\le 3\delta$, we have $|f_0(\phi(z))-f_0(\phi(x))|\le\epsilon^2$.

Let $\mathcal{E}_1 = \{x : f_1(\phi(x))\le\epsilon\}$, $\mathcal{E}_2 = \{x : f_0(\phi(x))\le\epsilon \text{ or } \mathrm{dist}(x,\partial\mathcal{Z})\vee\mathrm{dist}(x,\partial\mathcal{X})\le 3\delta\}$. We then seperate the proof into three cases. In the following, it suffices to consider x such that $f_0(\phi(x))>0$ since we are considering L_p risk.

Case I. $x\notin\mathcal{E}_1\cup\mathcal{E}_2$. In this case we have $f_0(\phi(x)), f_1(\phi(x))>\epsilon$. Then for any $z\in\mathcal{X}$ with $\|\phi(z)-\phi(x)\|\le 3\delta$, we have $z\in\mathcal{Z}$ by the definition of $\mathcal{E}_2$ and then $|f_0(\phi(z))-f_0(\phi(x))|\le\epsilon f_0(\phi(x))$ and $|f_1(\phi(z))-f_1(\phi(x))|\le\epsilon f_1(\phi(x))$.

Proceeding as in the proof of Case I in Lemma 4, we obtain

$$\lim_{N_0\to\infty} \sup_{x\notin\mathcal{E}_1\cup\mathcal{E}_2} E\left(\left|\widehat{r}_\phi(x) - r(\phi(x))\right|^p\right) = 0,$$

and then

$$\lim_{N_0\to\infty} E\left\{\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right\} = \lim_{N_0\to\infty} E\left\{E\left(\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \,\middle|\, X = x\right)\mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right\}$$
$$\le \lim_{N_0\to\infty} E\left\{\sup_{x\notin\mathcal{E}_1\cup\mathcal{E}_2} E\left(\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \,\middle|\, X = x\right)\mathbb{1}\left(X \notin \mathcal{E}_1 \cup \mathcal{E}_2\right)\right\} = 0.$$

Case II. $x \in \mathcal{E}_1 \setminus \mathcal{E}_2$. In this case we have $f_0(\phi(x)) > \epsilon$. Then for any $z \in \mathcal{X}$ with $\|\phi(z) - \phi(x)\| \le 3\delta$, we have $z \in \mathcal{Z}$ by the definition of $\mathcal{E}_2$ and then $|f_0(\phi(z)) - f_0(\phi(x))| \le \epsilon f_0(\phi(x))$ and $f_1(\phi(z)) \le \epsilon + \epsilon^2$.

Proceeding as in the proof of Case II in Lemma 4, we obtain

$$\lim_{N_0\to\infty} \sup_{x\in\mathcal{E}_1\setminus\mathcal{E}_2} E\left(\left|\widehat{r}_\phi(x) - r(\phi(x))\right|^p\right) = 0,$$

and then

$$\lim_{N_0\to\infty} E\left\{\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \notin \mathcal{E}_1 \setminus \mathcal{E}_2\right)\right\} = 0.$$

Case III. $x \in \mathcal{E}_2$. In this case we have $f_0(\phi(x)) \le \epsilon$ or $\mathrm{dist}(x, \partial\mathcal{Z}) \vee \mathrm{dist}(x, \partial\mathcal{X}) \le 3\delta$. Since $\mathcal{X}$ is compact, the surface areas of $\mathcal{X}$ and $\mathcal{Z}$ are bounded, and r is bounded uniformly, we have

$$\limsup_{N_0\to\infty} E\left\{\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \in \mathcal{E}_2\right)\right\} \lesssim \mathrm{pr}\left(X \in \mathcal{E}_2\right) \lesssim \epsilon + \delta.$$

Since ϵ is arbitrary and δ can be taken arbitrary small, we obtain

$$\lim_{N_0\to\infty} E\left\{\left|\widehat{r}_\phi(X) - r(\phi(X))\right|^p \mathbb{1}\left(X \in \mathcal{E}_2\right)\right\} = 0.$$

Combining the above three cases completes the proof.

C.5. *Proof of Lemma 5*

Note that for any $x \in \mathcal{X}$,

$$\frac{f_{\phi,X\mid D=1}(\phi(x))}{f_{\phi,X\mid D=0}(\phi(x))} = \frac{\mathrm{pr}(D = 1 \mid \phi(X) = \phi(x))\mathrm{pr}(D = 0)}{\mathrm{pr}(D = 0 \mid \phi(X) = \phi(x))\mathrm{pr}(D = 1)},$$

and

$$\frac{f_{X\mid D=1}(x)}{f_{X\mid D=0}(x)} = \frac{\mathrm{pr}(D = 1 \mid X = x)\mathrm{pr}(D = 0)}{\mathrm{pr}(D = 0 \mid X = x)\mathrm{pr}(D = 1)}.$$

This completes the proof.

REFERENCES

ABADIE, A. & IMBENS, G. W. (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association* **107**, 833–843.

BOUCHERON, S., LUGOSI, G. & MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

LI, Q. & RACINE, J. S. (2023). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

LIN, Z., DING, P. & HAN, F. (2023). Estimation based on nearest neighbor matching: from density ratio to average treatment effect. *Econometrica* **91**, 2187–2217.

LIN, Z. & HAN, F. (2022). On regression-adjusted imputation estimators of the average treatment effect. *arXiv preprint arXiv:2212.05424* .

TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* **12**, 389–434.

[*Received on* X *XX* XXX. *Editorial decision on* X *XX* XXX]