Econometrica Supplementary Material

SUPPLEMENT TO "ESTIMATION BASED ON NEAREST NEIGHBOR MATCHING: FROM DENSITY RATIO TO AVERAGE TREATMENT EFFECT"

(*Econometrica*, Vol. 91, No. 6, November 2023, 2187–2217)

ZHEXIAO LIN
Department of Statistics, University of California, Berkeley

PENG DING
Department of Statistics, University of California, Berkeley

FANG HAN
Department of Statistics, University of Washington

S1. PROOFS OF THE RESULTS IN SECTIONS 3 AND 4

Additional Notation. WE USE $\boldsymbol{X}$ and $\boldsymbol{Z}$ to represent $(X_1, X_2, \dots, X_{N_0})$ and $(Z_1, Z_2, \dots, Z_{N_1})$, respectively. Let $U(0,1)$ denote the uniform distribution on $[0,1]$. Let $U \sim U(0,1)$ and $U_{(M)}$ be the Mth order statistic of N_0 independent random variables from $U(0,1)$, assumed to be mutually independent and both independent of $(\boldsymbol{X}, \boldsymbol{Z})$. It is well known that $U_{(M)} \sim \mathrm{Beta}(M, N_0 + 1 - M)$. Let $\mathrm{Bin}(\cdot,\cdot)$ denote the binomial distribution. Let $L_1(\mathbb{R}^d)$ denote the space of all functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int |f(x)|\,\mathrm{d}x < \infty$. For any $x \in \mathbb{R}^d$ and function $f : \mathbb{R}^d \to \mathbb{R}$, we say x is a Lebesgue point (Bogachev and Ruas (2007, Theorem 5.6.2)) of f if

$$\lim_{\delta \to 0^+} \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \big|f(x) - f(z)\big|\,\mathrm{d}z = 0.$$

S2. PROOFS OF THE RESULTS IN APPENDIX A

S2.1. *Proof of Theorem A.1*

PROOF OF THEOREM A.1: We consider the complexities of two algorithms separately.

Algorithm 1.

The worst-case computation complexity of building a balanced k-d tree is $O(dN_0 \log N_0)$ (cf. Brown (2015)) since the size of the k-d tree is N_0.

The average computation complexity of searching a NN is $O(\log N_0)$ from Friedman, Bentley, and Finkel (1977), and then the average computation complexity of search M-NNs in $\{X_i\}_{i=1}^{N_0}$ for all $\{Z_j\}_{j=1}^{N_1}$ is $O(MN_1 \log N_0)$.

Notice that $|S_j| = M$ for any $j \in [\![N_1]\!]$ and then $|\bigcup_{j=1}^{N_1} S_j| \le N_1 M$. Since the elements of each S_j are in $[\![N_0]\!]$, the largest integer in $\bigcup_{j=1}^{N_1} S_j$ is N_0. Then the computation complexity of counting step is $O(N_1 M + N_0)$ due to the counting sort algorithm (Cormen, Leiserson, Rivest, and Stein (2009, Section 8.2)).

Combining the above three steps completes the proof for Algorithm 1.

Zhexiao Lin: zhexiaolin@berkeley.edu
Peng Ding: pengdingpku@berkeley.edu
Fang Han: fanghan@uw.edu

 https://doi.org/10.3982/ECTA20598

Algorithm 2.
The computation complexity of building a k-d tree is $O(d(N_0+n)\log(N_0+n))$ from Algorithm 1 since the size of the k-d tree is N_0+n.

For the searching step, for each $j \in [\![N_1]\!]$, the number of NNs to be searched is $M+\sum_{i=1}^n \mathbb{1}(\|x_i - Z_j\| \le \|\mathcal{X}_{(M)}(Z_j) - Z_j\|)$. Then from (2.2), the total number of NNs searched for all $j \in [\![N_1]\!]$ is $\sum_{j=1}^{N_1}(M+\sum_{i=1}^n \mathbb{1}(\|x_i - Z_j\| \le \|\mathcal{X}_{(M)}(Z_j) - Z_j\|)) = N_1 M + \sum_{i=1}^n K_M(x_i)$. Let X, Z be two independent copies from ν_0, ν_1, respectively, and are independent of the data. Since $[Z_j]_{j=1}^{N_1}$ are i.i.d. and $[X_i]_{i=1}^{N_0} \cup [x_i]_{i=1}^n$ are i.i.d, we have $\mathrm{E}[\sum_{i=1}^n K_M(x_i)] = n\mathrm{E}[K_M(X)] = N_1 n\mathrm{E}[\nu_1(A_M(X))] = N_1 n\frac{M}{N_0+1}$ since $\mathrm{E}[\nu_1(A_M(X))] = \mathrm{P}(\|X-Z\| \le \|\mathcal{X}_{(M)}(Z) - Z\|) = \mathrm{P}(U \le U_{(M)}) = \frac{M}{N_0+1}$ by using the probability integral transform. Then the average computation complexity for the searching step is $O(N_0^{-1}N_1 M(N_0+n)\log(N_0+n))$.

For the counting step, the computation complexity for counting $\bigcup_{j=1}^{N_1} S_j$ is $O(N_0+N_1M)$ since the cardinality of $\bigcup_{j=1}^{N_1} S_j$ is at most N_1M and the largest integer is N_0. The average computation complexity for counting $\bigcup_{j=1}^{N_1} S_j'$ is $O(N_0^{-1}N_1Mn+n)$ since the average cardinality of $\bigcup_{j=1}^{N_1} S_j'$ is at most $N_0^{-1}N_1Mn$ and the largest integer is n.

Combining the above three steps completes the proof for Algorithm 2. *Q.E.D.*

S3. PROOFS OF THE RESULTS IN APPENDIX B

S3.1. *Proof of Lemma B.1*

PROOF OF LEMMA B.1: From the Lebesgue differentiation theorem, for any $f \in L_1(\mathbb{R}^d)$, x is a Lebesgue point of f for λ-almost all x. Then for ν_0-almost all x, we have $f_0(x) > 0$ and x is a Lebesgue point of f_0 and f_1 from the absolute continuity of ν_0 and ν_1. We then only need to consider those $x \in \mathbb{R}^d$ such that $f_0(x) > 0$ and x is a Lebesgue point of f_0 and f_1.

We first introduce a lemma about the Lebesgue point.

LEMMA S3.1: *Let ν be a probability measure on $\mathbb{R}^d$ admitting a density f with respect to the Lebesgue measure. Let $x \in \mathbb{R}^d$ be a Lebesgue point of f. Then for any $\epsilon \in (0,1)$, there exists $\delta = \delta_x > 0$ such that for any $z \in \mathbb{R}^d$ satisfying $\|z-x\| \le \delta$, we have*

$$\left|\frac{\nu(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})} - f(x)\right| \le \epsilon, \qquad \left|\frac{\nu(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f(x)\right| \le \epsilon.$$

Part I. This part proves the first claim. We separate the proof of Part I into two cases based on the value of $f_1(x)$.

Case I.1. $f_1(x) > 0$. Since x is a Lebesgue point of ν_0 and ν_1, by Lemma S3.1, for any $\epsilon \in (0,1)$, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathbb{R}^d$ with $\|z-x\| \le \delta$, we have for $w \in \{0,1\}$,

$$\left|\frac{\nu_w(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})} - f_w(x)\right| \le \epsilon f_w(x), \qquad \left|\frac{\nu_w(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f_w(x)\right| \le \epsilon f_w(x).$$

Accordingly, if $\|z-x\| \le \delta$, by $\lambda(B_{z,\|x-z\|}) = \lambda(B_{x,\|x-z\|})$, we have

$$\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)} \le \frac{\nu_0(B_{z,\|x-z\|})}{\lambda(B_{z,\|x-z\|})}\frac{\lambda(B_{x,\|x-z\|})}{\nu_1(B_{x,\|x-z\|})} = \frac{\nu_0(B_{z,\|x-z\|})}{\nu_1(B_{x,\|x-z\|})} \le \frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}. \tag{S3.1}$$

On the other hand, for any $z \in \mathbb{R}^d$ such that $\|z-x\| > \delta$, $\nu_0(B_{z,\|z-x\|}) \geq \nu_0(B_{z^*,\delta}) \geq (1-\epsilon)f_0(x)\lambda(B_{z^*,\delta}) = (1-\epsilon)f_0(x)\lambda(B_{0,\delta})$, where z^* is the intersection point of the surface of $B_{x,\delta}$ and the line connecting z and x.

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 $\eta_N \frac{M}{N_0} = 4\frac{M}{N_0}\log(\frac{N_0}{M}) < (1-\epsilon)f_0(x)\lambda(B_{0,\delta})$. Then for any $z \in \mathbb{R}^d$ such that $\nu_0(B_{z,\|z-x\|}) \leq \eta_N M/N_0$, we have $\|z-x\| \leq \delta$ since otherwise it would contradict the selection of N_0.

Let Z be a copy from ν_1 independent of the data. Then

$$\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] = \mathrm{P}\big(Z \in A_M(x)\big) = \mathrm{P}\big(\nu_0(B_{Z,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big). \tag{S3.2}$$

For any given $z \in \mathbb{R}^d$, $[\nu_0(B_{z,\|X_i-z\|})]_{i=1}^{N_0}$ are i.i.d. from $U(0,1)$ since $[X_i]_{i=1}^{N_0}$ are i.i.d. from ν_0 and we use the probability integral transform. Then $\nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})$ has the same distribution as $U_{(M)}$ and is independent of Z.

Upper bound. With a slight abuse of notation, we define $W = \nu_0(B_{Z,\|x-Z\|})$. We then have, from (S3.1) and (S3.2),

$$\begin{aligned}
&\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big]\\
&\quad= \mathrm{P}\big(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big)\\
&\quad\leq \mathrm{P}\bigg(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}\bigg) + \mathrm{P}\bigg(\nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) > \eta_N\frac{M}{N_0}\bigg)\\
&\quad= \mathrm{P}\bigg(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}, \|Z-x\| \leq \delta\bigg) + \mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg)\\
&\quad\leq \mathrm{P}\big(\nu_0(B_{Z,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}), \|Z-x\| \leq \delta\big) + \mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg)\\
&\quad\leq \mathrm{P}\bigg(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}), \|Z-x\| \leq \delta\bigg) + \mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg)\\
&\quad\leq \mathrm{P}\bigg(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\bigg) + \mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg)\\
&\quad= \mathrm{P}\bigg(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\bigg) + \mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg).
\end{aligned} \tag{S3.3}$$

For the second term in (S3.3), notice that $\eta_N \to \infty$ as $N_0 \to \infty$. Then from the Chernoff bound and for N_0 sufficiently large, we have

$$\begin{aligned}
\frac{N_0}{M}\mathrm{P}\bigg(U_{(M)} > \eta_N\frac{M}{N_0}\bigg) &= \frac{N_0}{M}\mathrm{P}\bigg(\mathrm{Bin}\bigg(N_0, \eta_N\frac{M}{N_0}\bigg) < M\bigg)\\
&\leq \frac{N_0}{M}\exp\big((1+\log\eta_N - \eta_N)M\big)\\
&\leq \frac{N_0}{M}\exp\bigg(-\frac{1}{2}\eta_N M\bigg) = \bigg(\frac{N_0}{M}\bigg)^{1-2M}.
\end{aligned}$$

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

$$\lim_{N_0\to\infty} \frac{N_0}{M} \mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) = 0. \tag{S3.4}$$

For the first term in (S3.3), we have

$$\begin{aligned}
&\frac{N_0}{M}\mathrm{P}\Big(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\Big)\\
&= \frac{N_0}{M}\int_0^1 \mathrm{P}\Big(U_{(M)} \geq \frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}t\Big)\,\mathrm{d}t\\
&= \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\int_0^{\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}} \mathrm{P}\Big(U_{(M)} \geq \frac{M}{N_0}t\Big)\,\mathrm{d}t \leq \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\int_0^\infty \mathrm{P}\Big(\frac{N_0}{M}U_{(M)} \geq t\Big)\,\mathrm{d}t\\
&= \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{N_0}{M}\mathrm{E}[U_{(M)}] = \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{N_0}{N_0+1}.
\end{aligned}\tag{S3.5}$$

We then obtain

$$\limsup_{N_0\to\infty} \frac{N_0}{M}\mathrm{P}\Big(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\Big) \leq \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}. \tag{S3.6}$$

Plugging (S3.4) and (S3.6) to (S3.3) then yields

$$\limsup_{N_0\to\infty} \frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] \leq \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}. \tag{S3.7}$$

Lower bound. We have, from (S3.1) and (S3.2),

$$\begin{aligned}
\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] &= \mathrm{P}\big(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big) \geq \mathrm{P}\Big(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}\Big)\\
&= \mathrm{P}\Big(W \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}, \|Z-x\| \leq \delta\Big)\\
&\geq \mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}, \|Z-x\| \leq \delta\Big)\\
&= \mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \leq \eta_N\frac{M}{N_0}\Big)\\
&\geq \mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\Big)\\
&\quad - \mathrm{P}\Big(\nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) > \eta_N\frac{M}{N_0}\Big)\\
&= \mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\Big) - \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big).
\end{aligned}\tag{S3.8}$$

The second last equality is from the fact that for $\|Z-x\|>\delta$,

$$\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \geq \frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\delta}) \geq \frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}f_1(x)(1-\epsilon)\lambda(B_{0,\delta}) > \eta_N\frac{M}{N_0},$$

and then that $\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \leq \eta_N\frac{M}{N_0}$ implies $\|Z-x\|\leq\delta$.

For the first term in (S3.8), we have

$$\frac{N_0}{M}\mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\Big) = \frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)}\int_0^{\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}}\mathrm{P}\Big(U_{(M)} \geq \frac{M}{N_0}t\Big)\,\mathrm{d}t.$$

If $\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)} \geq 1$, then by $U_{(M)}\in[0,1]$, we have

$$\frac{N_0}{M}\mathrm{P}\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}U \leq U_{(M)}\Big) = \frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)}\frac{N_0}{M}\mathrm{E}[U_{(M)}] = \frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)}\frac{N_0}{N_0+1}.$$

If $\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)} < 1$, from the Chernoff bound,

$$\begin{aligned}
&\int_{\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}}^{\frac{N_0}{M}}\mathrm{P}\Big(U_{(M)} \geq \frac{M}{N_0}t\Big)\,\mathrm{d}t\\
&\quad\leq \Big[1-\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\Big]\frac{N_0}{M}\mathrm{P}\Big(U_{(M)} \geq \frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\Big)\\
&\quad\leq \Big[1-\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\Big]\frac{N_0}{M}\exp\Big[M-\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}N_0\\
&\qquad -M\log M + M\log\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}N_0\Big)\Big].
\end{aligned}$$

Since $f_0(x)>0$ and $M\log N_0/N_0 \to 0$, we obtain

$$\lim_{N_0\to\infty}\int_{\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}}^{\frac{N_0}{M}}\mathrm{P}\Big(U_{(M)} \geq \frac{M}{N_0}t\Big)\,\mathrm{d}t = 0.$$

Then we always have

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

Using the above identity along with (S3.4) to (S3.8) yields

$$\liminf_{N_0\to\infty}\frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] \geq \frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)}. \tag{S3.9}$$

Lastly, combining (S3.7) with (S3.9) and noticing that ϵ is arbitrary, we obtain

$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] = \frac{f_1(x)}{f_0(x)} = r(x). \tag{S3.10}$$

Case I.2. $f_1(x) = 0$. Again, for any $\epsilon \in (0,1)$, by Lemma S3.1, there exists some $\delta = \delta_x > 0$ such that for any $z \in \mathbb{R}^d$ with $\|z-x\| \le \delta$, we have

$$\left|\frac{\nu_0(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f_0(x)\right| \le \epsilon f_0(x), \qquad \left|\frac{\nu_1(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})}\right| \le \epsilon.$$

Recall that $W = \nu_0(B_{Z,\|x-Z\|})$. Then if $\|Z-x\| \le \delta$, we have

$$W \ge (1-\epsilon)f_0(x)\lambda(B_{Z,\|x-Z\|}) = (1-\epsilon)f_0(x)\lambda(B_{x,\|x-Z\|}) \ge \epsilon^{-1}(1-\epsilon)f_0(x)\nu_1(B_{x,\|x-Z\|}).$$

Proceeding in the same way as (S3.3), we obtain

$$\begin{aligned}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] &\le \mathrm{P}\Big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\eta_N\frac{M}{N_0}, \|Z-x\| \le \delta\Big) + \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big)\\ &\le \mathrm{P}\Big(\frac{1-\epsilon}{\epsilon}f_0(x)U \le U_{(M)}\Big) + \mathrm{P}\Big(U_{(M)} > \eta_N\frac{M}{N_0}\Big).\end{aligned}$$

For the first term above,

$$\begin{aligned}\frac{N_0}{M}\mathrm{P}\Big(\frac{1-\epsilon}{\epsilon}f_0(x)U \le U_{(M)}\Big) &= \frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\int_0^{\frac{1-\epsilon}{\epsilon}f_0(x)\frac{N_0}{M}}\mathrm{P}\Big(U_{(M)} \ge \frac{M}{N_0}t\Big)\,\mathrm{d}t\\ &\le \frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\int_0^{\infty}\mathrm{P}\Big(\frac{N_0}{M}U_{(M)} \ge t\Big)\,\mathrm{d}t\\ &= \frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\frac{N_0}{M}\mathrm{E}[U_{(M)}] = \frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\frac{N_0}{N_0+1}.\end{aligned}$$

By (S3.4) and noticing ϵ is arbitrary, we have

$$\lim_{N_0\to\infty}\frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] = 0 = r(x). \tag{S3.11}$$

Combining (S3.10) and (S3.11) completes the proof of the first claim.

Part II. This part proves the second claim. We also separate the proof of Part II into two cases based on the value of $f_1(x)$.

Case II.1. $f_1(x) > 0$. Again, for any $\epsilon \in (0,1)$, we take δ in the same way as in Case I.1. Let $\eta_N = \eta_{N,p} = 4p\log(N_0/M)$. We also take N_0 sufficiently large so that $\eta_N\frac{M}{N_0} = 4p\frac{M}{N_0}\log(\frac{N_0}{M}) < (1-\epsilon)f_0(x)\lambda(B_{0,\delta})$.

Let $\widetilde{Z}_1,\dots,\widetilde{Z}_p$ be p independent copies that are drawn from ν_1 independent of the data. Then

$$\begin{aligned}&\mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]\\ &\quad= \mathrm{P}\big(\widetilde{Z}_1,\dots,\widetilde{Z}_p \in A_M(x)\big)\\ &\quad= \mathrm{P}\big(\nu_0(B_{\widetilde{Z}_1,\|x-\widetilde{Z}_1\|}) \le \nu_0(B_{\widetilde{Z}_1,\|\mathcal{X}_{(M)}(\widetilde{Z}_1)-\widetilde{Z}_1\|}),\dots,\nu_0(B_{\widetilde{Z}_p,\|x-\widetilde{Z}_p\|}) \le \nu_0(B_{\widetilde{Z}_p,\|\mathcal{X}_{(M)}(\widetilde{Z}_p)-\widetilde{Z}_p\|})\big).\end{aligned}$$

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

Let $W_{\max} = \max_{k\in[\![p]\!]} W_k$ and $V_{\max} = \max_{k\in[\![p]\!]} V_k$. Then

$$\begin{aligned}\mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big] &\le \mathrm{P}(W_{\max} \le V_{\max}) \\ &\le \mathrm{P}\Big(W_{\max} \le V_{\max} \le \eta_N \frac{M}{N_0}\Big) + \mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big). \end{aligned} \tag{S3.12}$$

For the second term in (S3.12),

$$\mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big) \le \sum_{k=1}^{p} \mathrm{P}\Big(V_k > \eta_N \frac{M}{N_0}\Big) = p\mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big).$$

Proceeding as (S3.4),

$$\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(U_{(M)} > \eta_N \frac{M}{N_0}\Big) \le \Big(\frac{N_0}{M}\Big)^p \exp\Big(-\frac{1}{2}\eta_N M\Big) = \Big(\frac{N_0}{M}\Big)^{p(1-2M)}.$$

We then obtain

$$\lim_{N_0\to\infty} \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(V_{\max} > \eta_N \frac{M}{N_0}\Big) = 0. \tag{S3.13}$$

For the first term in (S3.12), notice that $[\nu_1(B_{x,\|\widetilde{Z}_k - x\|})]_{k=1}^p$ are i.i.d. from $U(0,1)$ since $[\widetilde{Z}_k]_{k=1}^p$ are i.i.d. We then have

$$\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max} \le V_{\max} \le \eta_N \frac{M}{N_0}\Big) \\
&\quad = \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max} \le V_{\max} \le \eta_N \frac{M}{N_0}, \max_{k\in[\![p]\!]} \|\widetilde{Z}_k - x\| \le \delta\Big) \\
&\quad \le \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)} \max_{k\in[\![p]\!]} \nu_1(B_{x,\|\widetilde{Z}_k - x\|}) \le V_{\max} \le \eta_N \frac{M}{N_0}, \max_{k\in[\![p]\!]} \|\widetilde{Z}_k - x\| \le \delta\Big) \\
&\quad \le \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)} \max_{k\in[\![p]\!]} \nu_1(B_{x,\|\widetilde{Z}_k - x\|}) \le V_{\max}\Big) \\
&\quad = \Big(\frac{N_0}{M}\Big)^p \int_0^1 p t^{p-1} \mathrm{P}\Big(V_{\max} \ge \frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)} t \,\Big|\, \max_{k\in[\![p]\!]} \nu_1(B_{x,\|\widetilde{Z}_k - x\|}) = t\Big)\,\mathrm{d}t \\
&\quad = p\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\Big)^p \\
&\qquad \times \int_0^{\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}} t^{p-1} \mathrm{P}\Big(V_{\max} \ge \frac{M}{N_0} t \,\Big|\, \max_{k\in[\![p]\!]} \nu_1(B_{x,\|\widetilde{Z}_k - x\|}) = \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{M}{N_0} t\Big)\,\mathrm{d}t
\end{aligned}$$

$$= p\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\Big)^p\Big[\int_0^1 t^{p-1}\mathrm{P}\Big(V_{\max}\geq\frac{M}{N_0}t\,\Big|\,\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde{Z}_k-x\|})=\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{M}{N_0}t\Big)\,\mathrm{d}t$$

$$+\int_1^{\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}} t^{p-1}\mathrm{P}\Big(V_{\max}\geq\frac{M}{N_0}t\,\Big|\,\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde{Z}_k-x\|})=\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{M}{N_0}t\Big)\,\mathrm{d}t\Big].$$

For the first term,

$$\int_0^1 t^{p-1}\mathrm{P}\Big(V_{\max}\geq\frac{M}{N_0}t\,\Big|\,\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde{Z}_k-x\|})=\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\frac{M}{N_0}t\Big)\,\mathrm{d}t\leq\int_0^1 t^{p-1}\,\mathrm{d}t=\frac{1}{p}.$$

For the second term, using the Chernoff bound, conditional on $\widetilde{\mathbf{Z}}=(\widetilde{Z}_1,\dots,\widetilde{Z}_p)$,

$$\begin{aligned}
&\int_1^{\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\frac{N_0}{M}} t^{p-1}\mathrm{P}\Big(V_{\max}\geq\frac{M}{N_0}t\,\Big|\,\widetilde{\mathbf{Z}}\Big)\,\mathrm{d}t\\
&\quad\leq\int_0^\infty(1+t)^{p-1}\mathrm{P}\Big(V_{\max}\geq\frac{M}{N_0}(1+t)\,\Big|\,\widetilde{\mathbf{Z}}\Big)\,\mathrm{d}t\\
&\quad\leq\int_0^\infty(1+t)^{p-1}\Big[\sum_{k=1}^p\mathrm{P}\Big(V_k\geq\frac{M}{N_0}(1+t)\,\Big|\,\widetilde{\mathbf{Z}}\Big)\Big]\,\mathrm{d}t\\
&\quad=p\int_0^\infty(1+t)^{p-1}\mathrm{P}\Big(U_{(M)}\geq\frac{M}{N_0}(1+t)\Big)\,\mathrm{d}t\\
&\quad\leq p\int_0^\infty(1+t)^{p-1}(1+t)^M\exp(-tM)\,\mathrm{d}t\leq\sqrt{2\pi}pM^{-1/2}\Big(1+\frac{1}{M}\Big)^{p-1}\big(1+o(1)\big),
\end{aligned}$$

where the last step follows from Stirling's approximation with $M\to\infty$.

Then we obtain

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{P}\Big(W_{\max}\leq V_{\max},V_{\max}\leq\eta_N\frac{M}{N_0}\Big)\leq\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\Big)^p. \tag{S3.14}$$

Plugging (S3.13) and (S3.14) into (S3.12) yields

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]\leq\Big(\frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}\Big)^p=\Big(\frac{1+\epsilon}{1-\epsilon}r(x)\Big)^p. \tag{S3.15}$$

Lastly, using Hölder's inequality,

$$\Big(\frac{N_0}{M}\Big)^p\mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]\geq\Big[\frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big]\Big]^p.$$

Employing the first claim, we have

$$\liminf_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p\mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]\geq\big[r(x)\big]^p. \tag{S3.16}$$

Combining (S3.15) with (S3.16) and noting that ϵ is arbitrary, we obtain

$$\lim_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p \mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]=\big[r(x)\big]^p. \tag{S3.17}$$

Case II.2. $f_1(x)=0$. For any $\epsilon\in(0,1)$, we take δ in the same way as in the proof of Case I.2 and take η_N as in the proof of Case II.1.

By (S3.12),

$$\Big(\frac{N_0}{M}\Big)^p \mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]\le\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max}\le V_{\max}\le\eta_N\frac{M}{N_0}\Big)+\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(V_{\max}>\eta_N\frac{M}{N_0}\Big).$$

For the first term,

$$\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max}\le V_{\max}\le\eta_N\frac{M}{N_0}\Big)\\
&\le\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(\frac{1-\epsilon}{\epsilon}f_0(x)\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde Z_k-x\|})\le V_{\max}\Big)\\
&=\Big(\frac{N_0}{M}\Big)^p\int_0^1 pt^{p-1}\mathrm{P}\Big(V_{\max}\ge\frac{1-\epsilon}{\epsilon}f_0(x)t\,\Big|\,\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde Z_k-x\|})=t\Big)\,\mathrm{d}t\\
&=p\Big(\frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\Big)^p\int_0^{\frac{1-\epsilon}{\epsilon}f_0(x)\frac{N_0}{M}} t^{p-1}\mathrm{P}\Big(V_{\max}\ge\frac{M}{N_0}t\,\Big|\,\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde Z_k-x\|})=t\Big)\,\mathrm{d}t.
\end{aligned}$$

Then proceeding in the same way as (S3.14), we have

$$\limsup_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(W_{\max}\le V_{\max}\le\eta_N\frac{M}{N_0}\Big)\le\Big(\frac{\epsilon}{1-\epsilon}\frac{1}{f_0(x)}\Big)^p.$$

Lastly, using (S3.13) and noting again that ϵ is arbitrary, we obtain

$$\lim_{N_0\to\infty}\Big(\frac{N_0}{M}\Big)^p \mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big]=0=\big[r(x)\big]^p. \tag{S3.18}$$

Combining (S3.17) and (S3.18) then completes the proof of the second claim. *Q.E.D.*

S3.2. *Proof of Theorem B.1*

PROOF OF THEOREM B.1(i): By (2.4) and that $[Z_j]_{j=1}^{N_1}$ are i.i.d,

$$\mathrm{E}\big[\widehat r_M(x)\big]=\mathrm{E}\Big[\frac{N_0}{N_1}\frac{K_M(x)}{M}\Big]=\frac{N_0}{N_1M}\mathrm{E}\Big[\sum_{j=1}^{N_1}\mathbb{1}\big(Z_j\in A_M(x)\big)\Big]=\frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big].$$

Employing Lemma B.1 then completes the proof. *Q.E.D.*

PROOF OF THEOREM B.1(ii): By Hölder's inequality, it suffices to consider the case when p is even. Because x^p is convex for $p > 1$ and $x > 0$, we have

$$\begin{aligned}&\mathrm{E}\big[\big|\widehat{r}_M(x) - r(x)\big|^p\big]\\&\quad\leq 2^{p-1}\big(\mathrm{E}\big[\big|\widehat{r}_M(x) - \mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]\big|^p\big] + \mathrm{E}\big[\big|\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big] - r(x)\big|^p\big]\big).\end{aligned} \tag{S3.19}$$

For the second term in (S3.19), Lemma B.1 implies

$$\lim_{N_0\to\infty} \mathrm{E}\big[\big|\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big] - r(x)\big|^p\big] = \lim_{N_0\to\infty} \mathrm{E}\left[\left|\frac{N_0}{M}\nu_1\big(A_M(x)\big) - r(x)\right|^p\right] = 0 \tag{S3.20}$$

by expanding the product term.

For the first term in (S3.19), noticing that $[Z_j]_{j=1}^{N_1}$ are i.i.d, we have $K_M(x)\,|\,\boldsymbol{X} \sim \mathrm{Bin}(N_1, \nu_1(A_M(x)))$. Using Lemma B.1 and $MN_1/N_0 \to \infty$, for any positive integers p and q, we have

$$\lim_{N_0\to\infty}\left(\frac{N_0}{N_1 M}\right)^p \mathrm{E}\big[N_1^p \nu_1^p\big(A_M(x)\big)\big] = \big[r(x)\big]^p,$$

$$\lim_{N_0\to\infty}\left(\frac{N_0}{N_1 M}\right)^p \left(\frac{N_0}{M}\right)^q \mathrm{E}\big[N_1^p \nu_1^{p+q}\big(A_M(x)\big)\big] = \big[r(x)\big]^{p+q},$$

and then $\mathrm{E}[N_1^p \nu_1^p(A_M(x))]$ is the dominated term among $[\mathrm{E}[N_1^k \nu_1^{k+q}(A_M(x))]]_{k\leq p, q\geq 0}$.

To complete the proof, for any positive integer c and $Z \sim \mathrm{Bin}(n, p')$, let $\mu_c = \mathrm{E}[(Z - \mathrm{E}[Z])^c]$ be the cth central moment. By Romanovsky (1923), we have

$$\mu_{c+1} = p'(1-p')\left(nc\mu_{c-1} + \frac{\mathrm{d}\mu_c}{\mathrm{d}p'}\right).$$

Then for even p, we obtain

$$\mathrm{E}\big[\big(K_M(x) - N_1\nu_1\big(A_M(x)\big)\big)^p\big] \lesssim \mathrm{E}\big[N_1\nu_1\big(A_M(x)\big)\big]^{p/2} \lesssim \left(\frac{N_1 M}{N_0}\right)^{p/2}.$$

The first term in (S3.19) then satisfies

$$\mathrm{E}\big[\big|\widehat{r}_M(x) - \mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]\big|^p\big] = \left(\frac{N_0}{N_1 M}\right)^p \mathrm{E}\big[\big(K_M(x) - N_1\nu_1\big(A_M(x)\big)\big)^p\big] \lesssim \left(\frac{N_0}{N_1 M}\right)^{p/2}.$$

Since $MN_1/N_0 \to \infty$, we obtain

$$\lim_{N_0\to\infty} \mathrm{E}\big[\big|\widehat{r}_M(x) - \mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]\big|^p\big] = 0. \tag{S3.21}$$

Plugging (S3.20) and (S3.21) into (S3.19) then completes the proof. *Q.E.D.*

S3.3. *Proof of Theorem B.2*

PROOF OF THEOREM B.2: We first cite the Hardy–Littlewood maximal inequality.

LEMMA S3.2—Hardy–Littlewood Maximal Inequality (Stein (2016)): *For any locally integrable function* $f : \mathbb{R}^d \to \mathbb{R}$, *define*

$$\mathsf{M}f(x) = \sup_{\delta>0} \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|f(z)\right| \mathrm{d}z.$$

Then for $d \geq 1$, *there exists a constant* $C_d > 0$ *only depending on d such that for all* $t > 0$ *and* $f \in L_1(\mathbb{R}^d)$, *we have*

$$\lambda\left(\left\{x : \mathsf{M}f(x) > t\right\}\right) < \frac{C_d}{t} \|f\|_{L_1},$$

where $\|\cdot\|_{L_1}$ *stands for the function* L_1 *norm*.

Let $\epsilon > 0$ be given. We assume $\epsilon \leq f_L$. From Assumption B.1, S_0 and S_1 are bounded, then ν_0 and ν_1 are compactly supported. Since $f_0, f_1 \in L_1$, and the class of continuous functions are dense in the class of compactly supported L_1 functions from simple use of Lusin's theorem, we can find g_0, g_1 such that g_0, g_1 are continuous and $\|f_0 - g_0\|_{L_1} \leq \epsilon^3$ and $\|f_1 - g_1\|_{L_1} \leq \epsilon^3$.

Since g_0, g_1 are continuous with compact supports, they are uniformly continuous, that is, there exists $\delta > 0$ such that for any $x, z \in \mathbb{R}^d$ and $\|z - x\| \leq \delta$, we have $|g_0(x) - g_0(z)| \leq \frac{\epsilon^2}{3}$ and $|g_1(x) - g_1(z)| \leq \frac{\epsilon^2}{3}$.

For any $x \in \mathbb{R}^d$, we have

$$\begin{aligned}
&\frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|f_0(x) - f_0(z)\right| \mathrm{d}z \\
&\quad\leq \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left[\left|f_0(x) - g_0(x)\right| + \left|g_0(x) - g_0(z)\right| + \left|f_0(z) - g_0(z)\right|\right] \mathrm{d}z \\
&\quad= \left|f_0(x) - g_0(x)\right| + \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|g_0(x) - g_0(z)\right| \mathrm{d}z \\
&\qquad+ \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|f_0(z) - g_0(z)\right| \mathrm{d}z. \qquad \text{(S3.22)}
\end{aligned}$$

For the first term in (S3.22), using Markov's inequality, we have

$$\lambda\left(\left\{x : \left|f_0(x) - g_0(x)\right| > \epsilon^2/3\right\}\right) \leq 3\epsilon^{-2} \|f_0 - g_0\|_{L_1} \leq 3\epsilon. \qquad \text{(S3.23)}$$

For the second term in (S3.22), by the selection of δ,

$$\frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|g_0(x) - g_0(z)\right| \mathrm{d}z \leq \max_{z \in B_{x,\delta}} \left|g_0(x) - g_0(z)\right| \leq \frac{\epsilon^2}{3}. \qquad \text{(S3.24)}$$

For the third term,

$$\frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|f_0(z) - g_0(z)\right| \mathrm{d}z \leq \sup_{\delta>0} \frac{1}{\lambda(B_{x,\delta})} \int_{B_{x,\delta}} \left|f_0(z) - g_0(z)\right| \mathrm{d}z = \mathsf{M}(f_0 - g_0)(x).$$

Lemma S3.2 then yields

$$\lambda\big(\{x : \mathsf{M}(f_0 - g_0)(x) > \epsilon^2/3\}\big) < 3C_d\epsilon^{-2}\|f_0 - g_0\|_{L_1} \le 3C_d\epsilon. \tag{S3.25}$$

We can establish similar results for f_1, g_1.

Let

$$\begin{aligned} A_1 = &\{x : |f_0(x) - g_0(x)| > \epsilon^2/3\} \cup \{x : |f_1(x) - g_1(x)| > \epsilon^2/3\} \\ &\cup \{x : \mathsf{M}(f_0 - g_0)(x) > \epsilon^2/3\} \cup \{x : \mathsf{M}(f_1 - g_1)(x) > \epsilon^2/3\}. \end{aligned}$$

Plugging (S3.23), (S3.24), (S3.25) into (S3.22), for any $x \notin A_1$ and $\|z - x\| \le \delta$, we have

$$\frac{1}{\lambda(B_{x,\delta})}\int_{B_{x,\delta}} |f_0(x) - f_0(z)|\,\mathrm{d}z \le \epsilon^2, \qquad \frac{1}{\lambda(B_{x,\delta})}\int_{B_{x,\delta}} |f_1(x) - f_1(z)|\,\mathrm{d}z \le \epsilon^2,$$

and $\lambda(A_1) \le 6(C_d + 1)\epsilon$.

Let $A_2 = \{x : f_1(x) \le \epsilon\}$. We then separate the proof into three cases. In the following, it suffices to consider $f_0(x) > 0$ due to the definition of L_p risk.

Case I. $x \notin A_1 \cup A_2$. By $\epsilon \le f_L$ and the definition of A_2, for any $x \notin A_1 \cup A_2$ and $\|z - x\| \le \delta$,

$$\begin{aligned} &\frac{1}{\lambda(B_{x,\delta})}\int_{B_{x,\delta}} |f_0(x) - f_0(z)|\,\mathrm{d}z \le \epsilon^2 \le \epsilon f_L \le \epsilon f_0(x), \\ &\frac{1}{\lambda(B_{x,\delta})}\int_{B_{x,\delta}} |f_1(x) - f_1(z)|\,\mathrm{d}z \le \epsilon^2 \le \epsilon f_1(x). \end{aligned}$$

We then obtain for $w \in \{0, 1\}$,

$$\left|\frac{\nu_w(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})} - f_w(x)\right| \le \epsilon f_w(x), \qquad \left|\frac{\nu_w(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f_w(x)\right| \le \epsilon f_w(x).$$

Let $\eta_N = \eta_{N,p} = 4p\log(N_0/M)$. We also take N_0 large enough so that $\eta_N \frac{M}{N_0} = 4p\frac{M}{N_0}\log(\frac{N_0}{M}) < (1 - \epsilon)f_L\lambda(B_{0,\delta})$. Then for any $x \in \mathbb{R}^d$ such that $f_0(x) > 0$, we have $\eta_N \frac{M}{N_0} < (1 - \epsilon)f_0(x)\lambda(B_{0,\delta})$.

Proceeding as in the proof of Case II.1 of Lemma B.1 and also Theorem B.1 by using Fubini's theorem, since ϵ is arbitrary, we obtain

$$\lim_{N_0\to\infty} \mathrm{E}\left[\int_{\mathbb{R}^d} |\widehat{r}_M(x) - r(x)|^p f_0(x)\mathbb{1}(x \notin A_1 \cup A_2)\,\mathrm{d}x\right] = 0. \tag{S3.26}$$

Case II. $x \in A_2 \setminus A_1$. In this case, we have

$$\begin{aligned} &\left|\frac{\nu_0(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})} - f_0(x)\right| \le \epsilon f_0(x), \qquad \left|\frac{\nu_0(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f_0(x)\right| \le \epsilon f_0(x), \\ &\left|\frac{\nu_1(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})} - f_1(x)\right| \le \epsilon^2, \qquad \left|\frac{\nu_1(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f_1(x)\right| \le \epsilon^2. \end{aligned}$$

Take η_N and take N_0 sufficiently large as in Case I above. Proceeding as the proof of Case II.2 of Lemma B.1 and also Theorem B.1 by using Fubini's theorem, since ϵ is arbitrary, we obtain

$$\lim_{N_0\to\infty} \mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x)-r(x)\big|^p f_0(x)\mathbb{1}(x\in A_2\setminus A_1)\,\mathrm{d}x\Big]=0. \tag{S3.27}$$

Case III. $x\in A_1$. In this case, for any $x\in A_1$ and $z\in S_1$, $\nu_0(B_{z,\|z-x\|})\ge f_L\lambda(B_{z,\|z-x\|}\cap S_0)\ge af_L\lambda(B_{z,\|z-x\|})\ge \frac{af_L}{f_U}\nu_1(B_{x,\|z-x\|})$. Then for any $x\in A_1$, from (S3.12) and in the same way as (S3.14),

$$\begin{aligned}\Big(\frac{N_0}{M}\Big)^p \mathrm{E}\big[\nu_1^p\big(A_M(x)\big)\big] &\le \Big(\frac{N_0}{M}\Big)^p \mathrm{P}(W_{\max}\le V_{\max})\\ &\le \Big(\frac{N_0}{M}\Big)^p \mathrm{P}\Big(\frac{af_L}{f_U}\max_{k\in[\![p]\!]}\nu_1(B_{x,\|\widetilde{Z}_k-x\|})\le V_{\max}\Big)\\ &\le \Big(\frac{f_U}{af_L}\Big)^p\big(1+o(1)\big)=O(1).\end{aligned}$$

Proceeding as in the proof of Theorem B.1, and due to the boundedness assumptions on f_0 and f_1, for any $x\in A_1$ and p even,

$$\mathrm{E}\big[\big|\widehat{r}_M(x)-r(x)\big|^p\big]\lesssim \mathrm{E}\big[\big|\widehat{r}_M(x)-\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]\big|^p\big]+\mathrm{E}\big[\big(\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]\big)^p\big]+\big|r(x)\big|^p\lesssim 1.$$

Then

$$\mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x)-r(x)\big|^p f_0(x)\mathbb{1}(x\in A_1)\,\mathrm{d}x\Big]\lesssim f_U\lambda(A_1)\lesssim\epsilon.$$

Since ϵ is arbitrary, we have

$$\lim_{N_0\to\infty} \mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x)-r(x)\big|^p f_0(x)\mathbb{1}(x\in A_1)\,\mathrm{d}x\Big]=0. \tag{S3.28}$$

Combining (S3.26), (S3.27), and (S3.28) completes the proof. *Q.E.D.*

S3.4. *Proof of Corollary B.1*

PROOF OF COROLLARY B.1: Corollary B.1 can be established following the same way as that of Theorem B.2 but with less effort since we only have to show

$$\lim_{N_0\to\infty} \mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]-r(x)\big|^p f_0(x)\,\mathrm{d}x\Big]=0.$$

In detail, denote the Radon–Nikodym derivative of the probability measure of W with respect to ν_0 by r_W. We then have

$$\begin{aligned}&\limsup_{N_0\to\infty}\mathrm{E}\Big[\Big|\frac{N_0}{M}\nu_1\big(A_M(W)\big)-r(W)\Big|^p\Big]\\ &\quad=\limsup_{N_0\to\infty}\mathrm{E}\Big[\int_{\mathbb{R}^d}\Big|\frac{N_0}{M}\nu_1\big(A_M(x)\big)-r(x)\Big|^p r_W(x)f_0(x)\,\mathrm{d}x\Big]\end{aligned}$$

$$\lesssim \limsup_{N_0\to\infty} \mathrm{E}\Big[\int_{\mathbb{R}^d}\Big|\frac{N_0}{M}\nu_1\big(A_M(x)\big)-r(x)\Big|^p f_0(x)\,\mathrm{d}x\Big]$$
$$= \limsup_{N_0\to\infty} \mathrm{E}\Big[\int_{\mathbb{R}^d}\big|\mathrm{E}\big[\widehat{r}_M(x)\,|\,\boldsymbol{X}\big]-r(x)\big|^p f_0(x)\,\mathrm{d}x\Big]=0,$$

where the last line has been established in the proof of Theorem B.2. Noticing that $\mathrm{E}[r(W)]^p$ is bounded under Assumption B.1, the proof is thus complete. *Q.E.D.*

S3.5. *Proof of Theorem B.3*

We only have to prove the first two claims as the rest are trivial.

PROOF OF THEOREM B.3(i): For any $z\in\mathbb{R}^d$ such that $\|z-x\|\le\delta/2$, since $B_{z,\|z-x\|}\subset B_{x,2\|z-x\|}\subset B_{x,\delta}$, we have

$$\Big|\frac{\nu_0(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})}-f_0(x)\Big|\le\frac{1}{\lambda(B_{z,\|z-x\|})}\int_{B_{z,\|z-x\|}}\big|f_0(y)-f_0(x)\big|\,\mathrm{d}y\le 2L\|z-x\|,$$
$$\Big|\frac{\nu_1(B_{x,\|z-x\|})}{\lambda(B_{x,\|z-x\|})}-f_1(x)\Big|\le\frac{1}{\lambda(B_{x,\|z-x\|})}\int_{B_{x,\|z-x\|}}\big|f_1(y)-f_1(x)\big|\,\mathrm{d}y\le L\|z-x\|.$$

Consider any $\delta_N>0$ such that $\delta_N\le\delta/2$. If $\|z-x\|\le\delta_N$ and $f_0(x)>2L\delta_N$, then

$$\frac{f_0(x)-2L\delta_N}{f_1(x)+L\delta_N}\le\frac{\nu_0(B_{z,\|x-z\|})}{\lambda(B_{z,\|x-z\|})}\frac{\lambda(B_{x,\|x-z\|})}{\nu_1(B_{x,\|x-z\|})}.$$

If further $f_1(x)>L\delta_N$, then

$$\frac{\nu_0(B_{z,\|x-z\|})}{\lambda(B_{z,\|x-z\|})}\frac{\lambda(B_{x,\|x-z\|})}{\nu_1(B_{x,\|x-z\|})}\le\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}.$$

On the other hand, if $\|z-x\|\ge\delta_N$ and $f_0(x)>2L\delta_N$, $\nu_0(B_{z,\|z-x\|})\ge(f_0(x)-2L\delta_N)\times\lambda(B_{0,\delta_N})=(f_0(x)-2L\delta_N)V_d\delta_N^d$, where V_d is the Lebesgue measure of the unit ball on $\mathbb{R}^d$.

Let $\delta_N=(\frac{4}{f_LV_d})^{1/d}(\frac{M}{N_0})^{1/d}$. Since $M/N_0\to0$, we have $\delta_N\to0$ as $N_0\to\infty$. Taking N_0 large enough so that $\delta_N<f_L/(4L)$ and $\delta_N\le\delta/2$, then $2LV_d\delta_N^{d+1}=\frac{M}{N_0}\frac{8L}{f_L}\delta_N<2\frac{M}{N_0}$. Then for any $(\nu_0,\nu_1)\in\mathcal{P}_{x,\mathrm{p}}(f_L,f_U,L,d,\delta)$,

$$\big(f_0(x)-2L\delta_N\big)V_d\delta_N^d>4\frac{f_0(x)}{f_L}\frac{M}{N_0}-2\frac{M}{N_0}\ge2\frac{M}{N_0}.$$

With a slight abuse of notation, let $W=\nu_0(B_{Z,\|x-Z\|})$. Then $W\le2\frac{M}{N_0}$ implies that $\|Z-x\|\le\delta_N$.

Depending on the value of $f_1(x)$, the proof is separated into two cases.

Case I. $f_1(x)>L\delta_N$.

Upper bound. Proceeding similar to (S3.3), we have

$$
\begin{aligned}
\mathrm{E}\big[\widehat{r}_M(x)\big] &= \frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] = \frac{N_0}{M}\mathrm{P}\big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big)\\
&\le \frac{N_0}{M}\mathrm{P}\Big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \le 2\frac{M}{N_0}\Big) + \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big)\\
&\le \frac{N_0}{M}\mathrm{P}\big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}), \|Z-x\| \le \delta_N\big) + \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big)\\
&\le \frac{N_0}{M}\mathrm{P}\Big(\frac{f_0(x)-2L\delta_N}{f_1(x)+L\delta_N}\nu_1(B_{x,\|x-Z\|}) \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}), \|Z-x\| \le \delta_N\Big)\\
&\quad + \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big)\\
&\le \frac{N_0}{M}\mathrm{P}\Big(\frac{f_0(x)-2L\delta_N}{f_1(x)+L\delta_N}U \le U_{(M)}\Big) + \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big). \qquad \text{(S3.29)}
\end{aligned}
$$

For the second term in (S3.29), since $M/\log N_0 \to \infty$, for any $\gamma > 0$,

$$
\begin{aligned}
\frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big) &= \frac{N_0}{M}\mathrm{P}\Big(\mathrm{Bin}\Big(N_0, 2\frac{M}{N_0}\Big) \le M\Big)\\
&\le \frac{N_0}{M}N_0^{-(1-\log 2)M/\log N_0} \prec N_0^{-\gamma}. \qquad \text{(S3.30)}
\end{aligned}
$$

For the first term in (S3.29), proceeding as (S3.5), we obtain

$$
\frac{N_0}{M}\mathrm{P}\Big(\frac{f_0(x)-2L\delta_N}{f_1(x)+L\delta_N}U \le U_{(M)}\Big) \le \frac{f_1(x)+L\delta_N}{f_0(x)-2L\delta_N}\frac{N_0}{N_0+1}.
$$

Then we obtain

$$
\mathrm{E}\big[\widehat{r}_M(x)\big] \le \frac{f_1(x)+L\delta_N}{f_0(x)-2L\delta_N}\frac{N_0}{N_0+1} + o\big(N_0^{-\gamma}\big). \qquad \text{(S3.31)}
$$

Lower bound. Proceeding similar to (S3.8), we have

$$
\begin{aligned}
\mathrm{E}\big[\widehat{r}_M(x)\big] &= \frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] = \frac{N_0}{M}\mathrm{P}\big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big)\\
&\ge \frac{N_0}{M}\mathrm{P}\Big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \le 2\frac{M}{N_0}\Big)\\
&= \frac{N_0}{M}\mathrm{P}\Big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \le 2\frac{M}{N_0}, \|Z-x\| \le \delta_N\Big)\\
&\ge \frac{N_0}{M}\mathrm{P}\Big(\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}\nu_1(B_{x,\|x-Z\|}) \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \le 2\frac{M}{N_0}, \|Z-x\| \le \delta_N\Big)\\
&= \frac{N_0}{M}\mathrm{P}\Big(\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}\nu_1(B_{x,\|x-Z\|}) \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|}) \le 2\frac{M}{N_0}\Big)
\end{aligned}
$$

$$\geq \frac{N_0}{M}\mathrm{P}\left(\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}U \leq U_{(M)}\right) - \frac{N_0}{M}\mathrm{P}\left(U_{(M)} > 2\frac{M}{N_0}\right)$$

$$= \frac{f_1(x)-L\delta_N}{f_0(x)+2L\delta_N}\int_0^{\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}\frac{N_0}{M}} \mathrm{P}\left(U_{(M)} \geq \frac{M}{N_0}t\right)\mathrm{d}t - \frac{N_0}{M}\mathrm{P}\left(U_{(M)} > 2\frac{M}{N_0}\right).$$

Consider the first term. If $\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N} \geq 1$, then

$$\frac{f_1(x)-L\delta_N}{f_0(x)+2L\delta_N}\int_0^{\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}\frac{N_0}{M}} \mathrm{P}\left(U_{(M)} \geq \frac{M}{N_0}t\right)\mathrm{d}t = \frac{f_1(x)-L\delta_N}{f_0(x)+2L\delta_N}\frac{N_0}{N_0+1}.$$

If $\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N} < 1$, using the Chernoff bound, for any $\gamma > 0$,

$$\int_{\frac{f_0(x)+2L\delta_N}{f_1(x)-L\delta_N}\frac{N_0}{M}}^{\frac{N_0}{M}} \mathrm{P}\left(U_{(M)} \geq \frac{M}{N_0}t\right)\mathrm{d}t$$

$$\leq \int_{\frac{f_L}{f_U}\frac{N_0}{M}}^{\frac{N_0}{M}} \mathrm{P}\left(U_{(M)} \geq \frac{M}{N_0}t\right)\mathrm{d}t \leq \left[1-\frac{f_L}{f_U}\right]\frac{N_0}{M}\mathrm{P}\left(U_{(M)} \geq \frac{f_L}{f_U}\right)$$

$$\leq \left[1-\frac{f_L}{f_U}\right]\frac{N_0}{M}\exp\left[M - \frac{f_L}{f_U}N_0 - M\log M + M\log\left(\frac{f_L}{f_U}N_0\right)\right] \prec N_0^{-\gamma}.$$

The last step is due to $M\log N_0/N_0 \to 0$. Recalling (S3.30), we then obtain

$$\mathrm{E}\left[\widehat{r}_M(x)\right] \geq \frac{f_1(x)-L\delta_N}{f_0(x)+2L\delta_N}\frac{N_0}{N_0+1} - o\left(N_0^{-\gamma}\right). \tag{S3.32}$$

Combining (S3.31) and (S3.32), and taking N_0 large enough so that $L\delta_N \leq f_U \wedge (f_L/4)$, we obtain

$$\begin{aligned}
&\left|\mathrm{E}\left[\widehat{r}_M(x)\right] - r(x)\right| \\
&\leq \left|\frac{f_1(x)+L\delta_N}{f_0(x)-2L\delta_N}\frac{N_0}{N_0+1} - \frac{f_1(x)}{f_0(x)}\right| \vee \left|\frac{f_1(x)-L\delta_N}{f_0(x)+2L\delta_N}\frac{N_0}{N_0+1} - \frac{f_1(x)}{f_0(x)}\right| \\
&\quad + o\left(N_0^{-\gamma}\right) \leq \frac{f_0(x)L\delta_N + 2f_1(x)L\delta_N}{f_0(x)\left(f_0(x)-2L\delta_N\right)} + \frac{1}{N_0+1}\frac{f_1(x)+L\delta_N}{f_0(x)-2L\delta_N} + o\left(N_0^{-\gamma}\right) \\
&\leq \left(\frac{2}{f_L} + \frac{4f_U}{f_L^2}\right)L\delta_N + \frac{4f_U}{f_L}\frac{1}{N_0+1} + o\left(N_0^{-\gamma}\right).
\end{aligned}$$

By the selection of δ_N and that the right-hand side does not depend on x, we complete the proof for this case.

Case II. $f_1(x) \leq L\delta_N$. The upper bound (S3.31) in Case I still holds for this case. Accordingly, taking N_0 large enough so that $L\delta_N \leq f_L/4$, we have

$$\begin{aligned}
\left|\mathrm{E}\left[\widehat{r}_M(x)\right] - r(x)\right| &\leq \mathrm{E}\left[\widehat{r}_M(x)\right] + r(x) \\
&\leq \frac{f_1(x)+L\delta_N}{f_0(x)-2L\delta_N}\frac{N_0}{N_0+1} + \frac{f_1(x)}{f_0(x)} + o\left(N_0^{-\gamma}\right)
\end{aligned}$$

$$\leq \frac{4}{f_L} L\delta_N + \frac{1}{f_L} L\delta_N + o\big(N_0^{-\gamma}\big).$$

We thus complete the whole proof. *Q.E.D.*

PROOF OF THEOREM B.3(ii): By the law of total variance,

$$\mathrm{Var}\big[\widehat{r}_M(x)\big] = \mathrm{E}\big[\mathrm{Var}\big[\widehat{r}_M(x) \mid \boldsymbol{X}\big]\big] + \mathrm{Var}\big[\mathrm{E}\big[\widehat{r}_M(x) \mid \boldsymbol{X}\big]\big]. \tag{S3.33}$$

For the first term in (S3.33), let Z be a copy drawn from ν_1 independently of the data. Then, since $[Z_j]_{j=1}^{N_1}$ are i.i.d,

$$\begin{aligned}
\mathrm{E}\big[\mathrm{Var}\big[\widehat{r}_M(x) \mid \boldsymbol{X}\big]\big] &= \mathrm{E}\bigg[\mathrm{Var}\bigg[\frac{N_0}{N_1 M} K_M(x) \,\bigg|\, \boldsymbol{X}\bigg]\bigg] \\
&= \bigg(\frac{N_0}{N_1 M}\bigg)^2 \mathrm{E}\Bigg[\mathrm{Var}\Bigg[\sum_{j=1}^{N_1} \mathbb{1}\big(Z_j \in A_M(x)\big) \,\Bigg|\, \boldsymbol{X}\Bigg]\Bigg] \\
&= \frac{N_0^2}{N_1 M^2} \mathrm{E}\big[\mathrm{Var}\big[\mathbb{1}\big(Z \in A_M(x)\big) \mid \boldsymbol{X}\big]\big] \\
&= \frac{N_0^2}{N_1 M^2} \mathrm{E}\big[\nu_1\big(A_M(x)\big) - \nu_1^2\big(A_M(x)\big)\big] \leq \frac{N_0^2}{N_1 M^2} \mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] \\
&= \frac{N_0}{N_1 M} \mathrm{E}\big[\widehat{r}_M(x)\big] \lesssim C \frac{N_0}{N_1 M},
\end{aligned} \tag{S3.34}$$

where $C > 0$ is a constant only depending on f_L, f_U. The last step is due to (S3.31).

For the second term in (S3.33), notice that

$$\mathrm{Var}\big[\mathrm{E}\big[\widehat{r}_M(x) \mid \boldsymbol{X}\big]\big] = \mathrm{Var}\bigg[\mathrm{E}\bigg[\frac{N_0}{N_1 M} K_M(x) \,\bigg|\, \boldsymbol{X}\bigg]\bigg] = \bigg(\frac{N_0}{M}\bigg)^2 \mathrm{Var}\big[\nu_1\big(A_M(x)\big)\big].$$

Recalling that $W = \nu_0(B_{Z,\|x-Z\|})$, we have the following lemma about the density of W near 0.

LEMMA S3.3: *Denote the density of W by f_W. Then for any $(\nu_0, \nu_1) \in \mathcal{P}_{x,\mathrm{p}}(f_L, f_U, L, d, \delta)$,*

$$f_W(0) = r(x).$$

Furthermore, for any $\epsilon > 0$ and N_0 sufficiently large, we have for all $0 \leq w \leq 2M/N_0$,

$$\sup_{(\nu_0,\nu_1)\in\mathcal{P}_{x,\mathrm{p}}(f_L,f_U,\delta,L,d)} f_W(w) \leq (1+\epsilon)\frac{f_U}{f_L}.$$

Due to Lemma S3.3, we can take N_0 sufficiently large so that for any $0 \leq w \leq 2M/N_0$,

$$\sup_{(\nu_0,\nu_1)\in\mathcal{P}_{x,\mathrm{p}}(f_L,f_U,\delta,L,d)} f_W(w) \leq 2\frac{f_U}{f_L}.$$

Let Z, $\widetilde{Z}$ be two independent copies from ν_1 that are further independent of the data. Let $W=\nu_0(B_{Z,\|x-Z\|})$ and $\widetilde{W}=\nu_0(B_{\widetilde{Z},\|x-\widetilde{Z}\|})$. Let $V=\nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})$ and $\widetilde{V}=\nu_0(B_{\widetilde{Z},\|\mathcal{X}_{(M)}(\widetilde{Z})-\widetilde{Z}\|})$. We then have

$$\begin{aligned}\operatorname{Var}\bigl[\nu_1\bigl(A_M(x)\bigr)\bigr]&=\mathrm{E}\bigl[\nu_1^2\bigl(A_M(x)\bigr)\bigr]-\bigl(\mathrm{E}\bigl[\nu_1\bigl(A_M(x)\bigr)\bigr]\bigr)^2\\&=\mathrm{P}\bigl(Z\in A_M(x),\widetilde{Z}\in A_M(x)\bigr)-\mathrm{P}\bigl(Z\in A_M(x)\bigr)\mathrm{P}\bigl(\widetilde{Z}\in A_M(x)\bigr)\\&=\mathrm{P}(W\le V,\widetilde{W}\le\widetilde{V})-\mathrm{P}(W\le V)\mathrm{P}(\widetilde{W}\le\widetilde{V}).\end{aligned}$$

Due to the independence between Z and $\widetilde{Z}$, W and $\widetilde{W}$ are independent. Notice that $V\mid Z$ have the same distribution as $U_{(M)}$ for any $Z\in\mathbb{R}^d$, then V and Z are independent, so are $\widetilde{V}$ and $\widetilde{Z}$.

Let us expand the variance further as

$$\begin{aligned}&\operatorname{Var}\bigl[\nu_1\bigl(A_M(x)\bigr)\bigr]\\&=\biggl[\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},W\le 2\frac{M}{N_0},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\\&\quad-\mathrm{P}\biggl(W\le V,W\le 2\frac{M}{N_0}\biggr)\mathrm{P}\biggl(\widetilde{W}\le\widetilde{V},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\biggr]\\&\quad+\biggl[\mathrm{P}(W\le V,\widetilde{W}\le\widetilde{V})-\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},W\le 2\frac{M}{N_0},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\biggr]\\&\quad-\biggl[\mathrm{P}(W\le V)\mathrm{P}(\widetilde{W}\le\widetilde{V})\\&\quad-\mathrm{P}\biggl(W\le V,W\le 2\frac{M}{N_0}\biggr)\mathrm{P}\biggl(\widetilde{W}\le\widetilde{V},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\biggr].\end{aligned}\tag{S3.35}$$

For the first term in (S3.35), we have the following lemma.

LEMMA S3.4: *We have*

$$\begin{aligned}&\biggl(\frac{N_0}{M}\biggr)^2\biggl[\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},W\le 2\frac{M}{N_0},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\\&\quad-\mathrm{P}\biggl(W\le V,W\le 2\frac{M}{N_0}\biggr)\mathrm{P}\biggl(\widetilde{W}\le\widetilde{V},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\biggr]\le C\frac{1}{M},\end{aligned}$$

where $C>0$ is a constant only depending on f_L, f_U.

For the second term in (S3.35),

$$\begin{aligned}&\mathrm{P}(W\le V,\widetilde{W}\le\widetilde{V})-\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},W\le 2\frac{M}{N_0},\widetilde{W}\le 2\frac{M}{N_0}\biggr)\\&\quad\le\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},W>2\frac{M}{N_0}\biggr)+\mathrm{P}\biggl(W\le V,\widetilde{W}\le\widetilde{V},\widetilde{W}>2\frac{M}{N_0}\biggr)\end{aligned}$$

$$\le \mathrm{P}\Big(V > 2\frac{M}{N_0}\Big) + \mathrm{P}\Big(\widetilde{V} > 2\frac{M}{N_0}\Big) = 2\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big).$$

Using the Chernoff bound and $M/\log N_0 \to \infty$, for any $\gamma > 0$,

$$\Big(\frac{N_0}{M}\Big)^2 \mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big) \le \Big(\frac{N_0}{M}\Big)^2 \exp\big[-(1-\log 2)M\big] \prec N_0^{-\gamma}.$$

We then have

$$\Big(\frac{N_0}{M}\Big)^2 \Big[\mathrm{P}(W \le V, \widetilde{W} \le \widetilde{V}) - \mathrm{P}\Big(W \le V, \widetilde{W} \le \widetilde{V}, W \le 2\frac{M}{N_0}, \widetilde{W} \le 2\frac{M}{N_0}\Big)\Big]$$
$$\le 2\Big(\frac{N_0}{M}\Big)^2 \mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big) \prec N_0^{-\gamma}. \tag{S3.36}$$

For the third term in (S3.35), we can check

$$\Big[\mathrm{P}(W \le V)\mathrm{P}(\widetilde{W} \le \widetilde{V}) - \mathrm{P}\Big(W \le V, W \le 2\frac{M}{N_0}\Big)\mathrm{P}\Big(\widetilde{W} \le \widetilde{V}, \widetilde{W} \le 2\frac{M}{N_0}\Big)\Big] \ge 0.$$

Plugging Lemma S3.4 and (S3.36) into (S3.35) by taking $\gamma > 1$, we obtain

$$\Big(\frac{N_0}{M}\Big)^2 \mathrm{Var}\big[\nu_1\big(A_M(x)\big)\big] \lesssim C\frac{1}{M}, \tag{S3.37}$$

where $C > 0$ is a constant only depending on f_L, f_U.

Plugging (S3.34) and (S3.37) into (S3.33) completes the proof. *Q.E.D.*

S3.6. *Proof of Proposition B.1*

PROOF OF PROPOSITION B.1: We take ν_0 and ν_1 to share the same support, and assume x to be the origin of $\mathbb{R}^d$ without loss of generality.

When $N_1 \lesssim N_0$, we take ν_0 to be the uniform distribution with density f_L on $[-f_L^{-1/d}/2, f_L^{-1/d}/2]^d$. Then the MSE is lower bounded by the density estimation over Lipchitz class with N_1 samples.

When $N_0 \lesssim N_1$, we take ν_1 to be the uniform distribution with density f_U on $[-f_U^{-1/d}/2, f_U^{-1/d}/2]^d$. Notice that $1/f_0$ is also local Lipchitz from the lower boundness condition and local Lipchitz condition on f_0. Then the MSE is lower bounded by the density estimation over Lipchitz class with N_0 samples.

We then complete the proof by combining the above two lower bounds and then using the famous minimax lower bound in Lipschitz density estimation (Tsybakov (2009, Exercise 2.8)), *Q.E.D.*

S3.7. *Proof of Theorem B.4*

PROOF OF THEOREM B.4: We only have to prove the first claim as the second is trivial.

Take $\delta_N = (\frac{4}{f_L V_d})^{1/d}(\frac{M}{N_0})^{1/d}$ as in the proof of Theorem B.3(i). Take $\delta_N' = (\frac{2}{a f_L V_d})^{1/d} \times (\frac{M}{N_0})^{1/d}$. For any $x \in \mathbb{R}^d$, denote the distance of x to the boundary of S_1 by $\Delta(x)$, that is, $\Delta(x) = \inf_{z \in \partial S_1} \|z - x\|$.

Depending on the position of x and the value of $\Delta(x)$, we separate the proof into three cases.

Case I. $x \in S_1$ and $\Delta(x) > 2\delta_N$. In this case, since $\Delta(x) > 2\delta_N$, for any $\|z-x\| \le \delta_N$, we have $B_{z,\|z-x\|} \subset S_1$. From the smoothness conditions on f_0 and f_1, similar to the proof of Theorem B.3, we have

$$
\begin{aligned}
&\mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x) - r(x)\big| f_0(x) \mathbb{1}\big(x \in S_1, \Delta(x) > 2\delta_N\big)\,\mathrm{d}x\Big] \\
&\le \int_{\mathbb{R}^d} \big(\mathrm{E}\big[\widehat{r}_M(x) - r(x)\big]^2\big)^{1/2} f_0(x) \mathbb{1}\big(x \in S_1, \Delta(x) > 2\delta_N\big)\,\mathrm{d}x \\
&\le C\Big[\Big(\frac{M}{N_0}\Big)^{1/d} + \Big(\frac{1}{M}\Big)^{1/2} + \Big(\frac{N_0}{MN_1}\Big)^{1/2}\Big] \int_{\mathbb{R}^d} f_0(x) \mathbb{1}\big(x \in S_1, \Delta(x) > 2\delta_N\big)\,\mathrm{d}x \\
&\le C\Big[\Big(\frac{M}{N_0}\Big)^{1/d} + \Big(\frac{1}{M}\Big)^{1/2} + \Big(\frac{N_0}{MN_1}\Big)^{1/2}\Big], \qquad \text{(S3.38)}
\end{aligned}
$$

where the constant $C > 0$ only depends on f_L, f_U, L, d.

Case II. $x \in S_0 \setminus S_1$ and $\Delta(x) > \delta'_N$. In this case, $r(x) = 0$ and for any $z \in S_1$,

$$
\nu_0(B_{z,\|z-x\|}) \ge f_L \lambda(B_{z,\|z-x\|} \cap S_0) \ge a f_L \lambda(B_{z,\|z-x\|}) > a f_L V_d \delta_N'^{d} \ge 2\frac{M}{N_0}.
$$

Then for any $\gamma > 0$,

$$
\begin{aligned}
\mathrm{E}\big[\big|\widehat{r}_M(x) - r(x)\big|\big] &= \mathrm{E}\big[\widehat{r}_M(x)\big] = \frac{N_0}{M}\mathrm{E}\big[\nu_1\big(A_M(x)\big)\big] \\
&= \frac{N_0}{M}\mathrm{P}\big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big) \le \frac{N_0}{M}\mathrm{P}\Big(U_{(M)} > 2\frac{M}{N_0}\Big) \prec N_0^{-\gamma}.
\end{aligned}
$$

We then obtain

$$
\begin{aligned}
&\mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x) - r(x)\big| f_0(x) \mathbb{1}\big(x \notin S_1, \Delta(x) > \delta'_N\big)\,\mathrm{d}x\Big] \\
&\prec N_0^{-\gamma} \int_{\mathbb{R}^d} f_0(x) \mathbb{1}\big(x \in S_0 \setminus S_1, \Delta(x) > \delta'_N\big)\,\mathrm{d}x \le N_0^{-\gamma}. \qquad \text{(S3.39)}
\end{aligned}
$$

Case III. $x \in S_0$ and $\Delta(x) \le (2\delta_N) \vee \delta'_N$. In this case, for any $z \in S_1$,

$$
\nu_0(B_{z,\|z-x\|}) \ge f_L \lambda(B_{z,\|z-x\|} \cap S_0) \ge a f_L \lambda(B_{z,\|z-x\|}) \ge \frac{a f_L}{f_U}\nu_1(B_{x,\|z-x\|}).
$$

Accordingly,

$$
\begin{aligned}
\mathrm{E}\big[\big|\widehat{r}_M(x) - r(x)\big|\big] &\le \mathrm{E}\big[\widehat{r}_M(x)\big] + r(x) = \frac{N_0}{M}\mathrm{P}\big(W \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\big) + r(x) \\
&\le \frac{N_0}{M}\mathrm{P}\Big(\frac{a f_L}{f_U}\nu_1(B_{x,\|x-Z\|}) \le \nu_0(B_{Z,\|\mathcal{X}_{(M)}(Z)-Z\|})\Big) + r(x) \\
&\le \frac{N_0}{M}\mathrm{P}\Big(\frac{a f_L}{f_U} U \le U_{(M)}\Big) + r(x) = \frac{f_U}{a f_L}\big(1 + o(1)\big) + \frac{f_U}{f_L}.
\end{aligned}
$$

From the definition of δ_N, δ'_N, and $M/N_0 \to 0$, we have $\delta_N, \delta'_N \to 0$ as $N_0 \to \infty$. Since the surface area of S_1 is bounded by H, we have $\lambda(\{x : \Delta(x) \le (2\delta_N) \vee \delta'_N\}) \lesssim H\{(2\delta_N) \vee \delta'_N\}$. Then we obtain

$$
\begin{aligned}
&\mathrm{E}\Big[\int_{\mathbb{R}^d} \big|\widehat{r}_M(x) - r(x)\big| f_0(x) \mathbb{1}\big(\Delta(x) \le (2\delta_N) \vee \delta'_N\big)\,\mathrm{d}x\Big] \\
&\quad \le \left(\frac{f_U}{af_L}\big(1 + o(1)\big) + \frac{f_U}{f_L}\right) \int_{\mathbb{R}^d} f_0(x) \mathbb{1}\big(\Delta(x) \le (2\delta_N) \vee \delta'_N\big)\,\mathrm{d}x \\
&\quad \le \left(\frac{f_U}{af_L}\big(1 + o(1)\big) + \frac{f_U}{f_L}\right) f_U \lambda\big(\{x : \Delta(x) \le (2\delta_N) \vee \delta'_N\}\big) \\
&\quad \lesssim \left(\frac{f_U}{af_L} + \frac{f_U}{f_L}\right) f_U H\big(\delta_N + \delta'_N\big) \le C\left(\frac{M}{N_0}\right)^{1/d},
\end{aligned} \tag{S3.40}
$$

where the constant $C > 0$ only depends on f_L, f_U, a, H, d.

Combining (S3.38), (S3.39), (S3.40) completes the proof. *Q.E.D.*

S3.8. *Proof of Proposition B.2*

PROOF OF PROPOSITION B.2: We take ν_0 and ν_1 to be of the same support.

When $N_1 \lesssim N_0$, we take ν_0 to be the uniform distribution with density f_L on $[-f_L^{-1/d}/2, f_L^{-1/d}/2]^d$. Then the L_1 risk is lower bounded by the L_1 risk over support of density estimation over Lipchitz class with N_1 samples.

When $N_0 \lesssim N_1$, we take ν_1 to be the uniform distribution with density f_U on $[-f_U^{-1/d}/2, f_U^{-1/d}/2]^d$. Notice $1/f_0$ is also Lipchitz from the lower boundness condition and Lipchitz condition on f_0. From the lower boundness condition on f_0, the L_1 risk is lower bounded by the L_1 risk over support of density estimation over Lipchitz class with N_0 samples.

We then complete the proof by combining the above two lower bounds and then using then the minimax lower bound of L_1 risk for density estimation over Lipchitz class (Zhao and Lai (2022, Theorem 1)). *Q.E.D.*

S4. PROOFS OF THE RESULTS IN APPENDIX C

S4.1. *Proof of Lemma C.1*

PROOF OF LEMMA C.1: For any $x \in \mathbb{X}$, define $\sigma_\omega^2(x) = \mathrm{E}[U_\omega^2 \,|\, X = x] = \mathrm{E}[[Y(\omega) - \mu_\omega(X)]^2 \,|\, X = x]$ for $\omega \in \{0, 1\}$. Let

$$
V^\tau = \mathrm{E}\big[\mu_1(X) - \mu_0(X) - \tau\big]^2 \quad \text{and} \quad V^E = \frac{1}{n}\sum_{i=1}^n \left(1 + \frac{K_M(i)}{M}\right)^2 \sigma_{D_i}^2(X_i).
$$

From the central limit theorem (Billingsley (2008, Theorem 27.1)), we have

$$
\sqrt{n}\big(\bar{\tau}(\boldsymbol{X}) - \tau\big) \xrightarrow{\mathrm{d}} N\big(0, V^\tau\big). \tag{S4.1}
$$

Let $E_{M,i} = (2D_i - 1)(1 + K_M(i)/M)\epsilon_i$ for any $i \in [\![n]\!]$. Conditional on $\boldsymbol{X}, \boldsymbol{D}$, $[E_{M,i}]_{i=1}^n$ are independent. Notice that $\mathrm{E}[E_{M,i} \,|\, \boldsymbol{X}, \boldsymbol{D}] = 0$ and $\sum_{i=1}^n \mathrm{Var}[E_{M,i} \,|\, \boldsymbol{X}, \boldsymbol{D}] = nV^E$. To apply the

Lindeberg–Feller central limit theorem (Billingsley (2008, Theorem 27.2)), it suffices to verify that: for a given $(\boldsymbol{X}, \boldsymbol{D})$,

$$\frac{1}{nV^E}\sum_{i=1}^n \mathrm{E}\big[(E_{M,i})^2\mathbb{1}\big(|E_{M,i}| > \delta\sqrt{nV^E}\big) \,|\, \boldsymbol{X}, \boldsymbol{D}\big] \to 0,$$

for all $\delta > 0$.

Let $C_\sigma = \sup_{x\in\mathbb{X},\omega\in\{0,1\}}\{\mathrm{E}[|U_\omega|^{2+\kappa} \,|\, X = x] \vee \mathrm{E}[U_\omega^2 \,|\, X = x]\} < \infty$. Let $p_1 = 1 + \kappa/2$ and p_2 be the constant such that $p_1^{-1} + p_2^{-1} = 1$. By Hölder's inequality and Markov's inequality,

$$\begin{aligned}
&\frac{1}{nV^E}\sum_{i=1}^n \mathrm{E}\big[(E_{M,i})^2\mathbb{1}\big(|E_{M,i}| > \delta\sqrt{nV^E}\big) \,|\, \boldsymbol{X}, \boldsymbol{D}\big] \\
&\quad\le \frac{1}{nV^E}\sum_{i=1}^n \big(\mathrm{E}\big[|E_{M,i}|^{2+\kappa} \,|\, \boldsymbol{X}, \boldsymbol{D}\big]\big)^{1/p_1}\big(\mathrm{P}\big(|E_{M,i}| > \delta\sqrt{nV^E} \,|\, \boldsymbol{X}, \boldsymbol{D}\big)\big)^{1/p_2} \\
&\quad\le \frac{1}{nV^E}\sum_{i=1}^n \big(\mathrm{E}\big[|E_{M,i}|^{2+\kappa} \,|\, \boldsymbol{X}, \boldsymbol{D}\big]\big)^{1/p_1}\bigg(\frac{1}{\delta^2 nV^E}\mathrm{E}\big[(E_{M,i})^2 \,|\, \boldsymbol{X}, \boldsymbol{D}\big]\bigg)^{1/p_2} \\
&\quad\le \frac{C_\sigma}{nV^E}\bigg(\frac{1}{\delta^2 nV^E}\bigg)^{1/p_2}\sum_{i=1}^n\bigg(1 + \frac{K_M(i)}{M}\bigg)^{2(1+1/p_2)}.
\end{aligned}$$

Notice that $\mathrm{E}[1 + K_M(i)/M]^{2(1+1/p_2)} < \infty$ from Theorem B.2. Let $c_\sigma = \inf_{x\in\mathbb{X},\omega\in\{0,1\}}\mathrm{E}[U_\omega^2 \,|\, X = x] > 0$. From the definition of V^E, we have $V^E \ge c_\sigma$ for almost all $\boldsymbol{X}$, $\boldsymbol{D}$. Then

$$\mathrm{E}\Bigg[\frac{1}{nV^E}\sum_{i=1}^n \mathrm{E}\big[(E_{M,i})^2\mathbb{1}\big(|E_{M,i}| > \delta\sqrt{nV^E}\big) \,|\, \boldsymbol{X}, \boldsymbol{D}\big]\Bigg] = O\big(n^{-1/p_2}\big) = o(1).$$

We thus obtain

$$\frac{1}{nV^E}\sum_{i=1}^n \mathrm{E}\big[(E_{M,i})^2\mathbb{1}\big(|E_{M,i}| > \delta\sqrt{nV^E}\big) \,|\, \boldsymbol{X}, \boldsymbol{D}\big] = o_{\mathrm{P}}(1).$$

Applying the Lindeberg–Feller central limit theorem then yields

$$\sqrt{n}\big(V^E\big)^{-1/2}E_M = \big(nV^E\big)^{-1/2}\sum_{i=1}^n E_{M,i} \xrightarrow{\mathsf{d}} N(0, 1). \tag{S4.2}$$

Noticing that $\sqrt{n}(\bar{\tau}(\boldsymbol{X}) - \tau)$ and $\sqrt{n}(V^E)^{-1/2}E_M$ are asymptotically independent, leveraging the same argument as made in Abadie and Imbens (2006, Proof of Theorem 4, p. 267) and then combining (S4.1) and (S4.2) reaches

$$\sqrt{n}\big(V^\tau + V^E\big)^{-1/2}\big(\bar{\tau}(\boldsymbol{X}) + E_M - \tau\big) \xrightarrow{\mathsf{d}} N(0, 1). \tag{S4.3}$$

We decompose V^E as

$$
\begin{aligned}
V^E &= \frac{1}{n}\sum_{i=1,D_i=1}^{n}\left(1+\frac{K_M(i)}{M}\right)^2\sigma_1^2(X_i) + \frac{1}{n}\sum_{i=1,D_i=0}^{n}\left(1+\frac{K_M(i)}{M}\right)^2\sigma_0^2(X_i) \\
&= \left[\frac{1}{n}\sum_{i=1,D_i=1}^{n}\left(\frac{1}{e(X_i)}\right)^2\sigma_1^2(X_i) + \frac{1}{n}\sum_{i=1,D_i=0}^{n}\left(\frac{1}{1-e(X_i)}\right)^2\sigma_0^2(X_i)\right] \\
&\quad + \frac{1}{n}\sum_{i=1,D_i=1}^{n}\left[\left(1+\frac{K_M(i)}{M}\right)^2 - \left(\frac{1}{e(X_i)}\right)^2\right]\sigma_1^2(X_i) \\
&\quad + \frac{1}{n}\sum_{i=1,D_i=0}^{n}\left[\left(\frac{1}{1-e(X_i)}\right)^2 - \left(1+\frac{K_M(i)}{M}\right)^2\right]\sigma_0^2(X_i). \qquad \text{(S4.4)}
\end{aligned}
$$

For the first term in (S4.4), notice that $[(X_i, D_i, Y_i)]_{i=1}^n$ are i.i.d. and $\mathrm{E}[D_i(e(X_i))^{-2} \times \sigma_1^2(X_i)], \mathrm{E}[(1-D_i)(1-e(X_i))^{-2}\sigma_0^2(X_i)] < \infty$. Using the weak law of large numbers, we have

$$
\frac{1}{n}\sum_{i=1,D_i=1}^{n}\left(\frac{1}{e(X_i)}\right)^2\sigma_1^2(X_i) + \frac{1}{n}\sum_{i=1,D_i=0}^{n}\left(\frac{1}{1-e(X_i)}\right)^2\sigma_0^2(X_i) \xrightarrow{\mathsf{p}} \mathrm{E}\left[\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right].
$$

For the second term in (S4.4), using the Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\mathrm{E}\left|\frac{1}{n}\sum_{i=1,D_i=1}^{n}\left[\left(1+\frac{K_M(i)}{M}\right)^2 - \left(\frac{1}{e(X_i)}\right)^2\right]\sigma_1^2(X_i)\right| \\
&\le C_\sigma \mathrm{E}\left[D_i\left|\left(1+\frac{K_M(i)}{M}\right)^2 - \left(\frac{1}{e(X_i)}\right)^2\right|\right] \\
&= C_\sigma \mathrm{E}\left[D_i \mathrm{E}\left[\left|\left(1+\frac{K_M(i)}{M}\right)^2 - \left(\frac{1}{e(X_i)}\right)^2\right| \,\Big|\, \boldsymbol{D}\right]\right] \\
&\le C_\sigma \mathrm{E}\left[D_i\left(\mathrm{E}\left[\left(\frac{K_M(i)}{M} - \frac{1-e(X_i)}{e(X_i)}\right)^2 \,\Big|\, \boldsymbol{D}\right]\mathrm{E}\left[\left(2+\frac{K_M(i)}{M} + \frac{1-e(X_i)}{e(X_i)}\right)^2 \,\Big|\, \boldsymbol{D}\right]\right)^{1/2}\right] \\
&= o(1),
\end{aligned}
$$

where the last step is due to Theorem B.2. Then we obtain

$$
\frac{1}{n}\sum_{i=1,D_i=1}^{n}\left[\left(1+\frac{K_M(i)}{M}\right)^2 - \left(\frac{1}{e(X_i)}\right)^2\right]\sigma_1^2(X_i) \xrightarrow{\mathsf{p}} 0.
$$

For the third term in (S4.4), we can establish in the same way that

$$
\frac{1}{n}\sum_{i=1,D_i=0}^{n}\left[\left(\frac{1}{1-e(X_i)}\right)^2 - \left(1+\frac{K_M(i)}{M}\right)^2\right]\sigma_0^2(X_i) \xrightarrow{\mathsf{p}} 0.
$$

Then from (S4.4),

$$V^E \xrightarrow{\mathrm{p}} \mathrm{E}\left[\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1-e(X)}\right].$$

By (S4.3), Slutsky's lemma (van der Vaart (1998, Theorem 2.8)), and the definition of σ^2, we complete the proof. *Q.E.D.*

S4.2. *Proof of Lemma C.2*

PROOF OF LEMMA C.2: From Assumption B.1 and Assumption 4.1, let $R = \mathrm{diam}(\mathbb{X}) < \infty$ and $f_L = \inf_{x\in\mathbb{X},\omega\in\{0,1\}} f_\omega(x) > 0$. For any $x \in \mathbb{X}$, $\omega \in \{0,1\}$, and $u \le R$, from Assumption B.1, $\nu_\omega(B_{x,u} \cap \mathbb{X}) \ge f_L \lambda(B_{x,u} \cap \mathbb{X}) \ge f_L a\lambda(B_{x,u}) = f_L a V_d u^d$, where V_d is the Lebesgue measure of the unit ball on $\mathbb{R}^d$.

Let $c_0 = f_L a V_d$. For any $i \in [\![n]\!]$, $x \in \mathbb{X}$, $M \le n_{1-D_i}$, if $0 \le u \le R n_{1-D_i}^{1/d}$, we have

$$\begin{aligned}
&\mathrm{P}\big(\|X_j - X_i\| \ge u n_{1-D_i}^{-1/d} \,|\, \boldsymbol{D}, X_i = x, j = j_M(i)\big) \\
&\quad\le \mathrm{P}\big(\mathrm{Bin}\big(n_{1-D_i}, \nu_{1-D_i}(B_{x,un_{1-D_i}^{-1/d}} \cap \mathbb{X})\big) \le M \,|\, \boldsymbol{D}\big) \\
&\quad\le \mathrm{P}\big(\mathrm{Bin}\big(n_{1-D_i}, c_0 u^d n_{1-D_i}^{-1}\big) \le M \,|\, \boldsymbol{D}\big).
\end{aligned}$$

Using the Chernoff bound, if $M < c_0 u^d$, then

$$\mathrm{P}\big(\mathrm{Bin}\big(n_{1-D_i}, c_0 u^d n_{1-D_i}^{-1}\big) \le M \,|\, \boldsymbol{D}\big) \le \exp\left(M - c_0 u^d + M \log\left(\frac{c_0 u^d}{M}\right)\right).$$

Notice that the above upper bound does not depend on x. We then obtain

$$\begin{aligned}
&\mathrm{P}\big(\|X_j - X_i\| \ge u n_{1-D_i}^{-1/d} \,|\, \boldsymbol{D}, j = j_M(i)\big) \\
&\quad\le \mathbb{1}\big(M < c_0 u^d\big) \exp\left(M - c_0 u^d + M \log\left(\frac{c_0 u^d}{M}\right)\right) + \mathbb{1}\big(M \ge c_0 u^d\big).
\end{aligned}$$

On the other hand, if $u > R n_{1-D_i}^{1/d}$, then the probability is zero from the definition of R. Accordingly, the above bound holds for any $u \ge 0$.

For any $i \in [\![n]\!]$, we thus have

$$\begin{aligned}
&n_{1-D_i}^{p/d} \mathrm{E}\big[\|U_{M,i}\|^p \,|\, \boldsymbol{D}\big] \\
&\quad= p \int_0^\infty \mathrm{P}\big(\|X_j - X_i\| \ge u n_{1-D_i}^{-1/d} \,|\, \boldsymbol{D}, j = j_M(i)\big) u^{p-1} \,\mathrm{d}u \\
&\quad\le p \int_0^\infty \left[\mathbb{1}\big(M < c_0 u^d\big) \exp\left(M - c_0 u^d + M \log\left(\frac{c_0 u^d}{M}\right)\right) + \mathbb{1}\big(M \ge c_0 u^d\big)\right] u^{p-1} \,\mathrm{d}u \\
&\quad= p c_0^{-p/d} d^{-1} \left[\int_M^\infty \left(\frac{e}{M}\right)^M t^{M+\frac{p}{d}-1} e^{-t} \,\mathrm{d}t + \int_0^M t^{\frac{p}{d}-1} \,\mathrm{d}t\right],
\end{aligned} \tag{S4.5}$$

where the last step is through taking $t = c_0 u^d$.

For the first term in (S4.5), from Stirling's formula and $M \to \infty$,

$$\int_M^\infty \left(\frac{e}{M}\right)^M t^{M+\frac{p}{d}-1} e^{-t} \, \mathrm{d}t \le \int_0^\infty \left(\frac{e}{M}\right)^M t^{M+\frac{p}{d}-1} e^{-t} \, \mathrm{d}t \sim \sqrt{2\pi} M^{\frac{p}{d}-\frac{1}{2}},$$

where $\sim$ means asymptotic convergence.

For the second term in (S4.5), $\int_0^M t^{\frac{p}{d}-1} \, \mathrm{d}t = \frac{d}{p} M^{\frac{p}{d}}$. Combining the above two terms then completes the proof. *Q.E.D.*

S4.3. *Proof of Lemma C.3*

PROOF OF LEMMA C.3: We bound $B_M - \widehat{B}_M$ by

$$\begin{aligned}
&|B_M - \widehat{B}_M| \\
&\quad = \left| \frac{1}{n} \sum_{i=1}^n (2D_i - 1) \left[\frac{1}{M} \sum_{m=1}^M \big(\mu_{1-D_i}(X_i) - \mu_{1-D_i}(X_{j_m(i)}) - \widehat{\mu}_{1-D_i}(X_i) + \widehat{\mu}_{1-D_i}(X_{j_m(i)}) \big) \right] \right| \\
&\quad \le \frac{1}{n} \sum_{i=1}^n \max_{m \in [\![M]\!]} \big| \mu_{1-D_i}(X_i) - \mu_{1-D_i}(X_{j_m(i)}) - \widehat{\mu}_{1-D_i}(X_i) + \widehat{\mu}_{1-D_i}(X_{j_m(i)}) \big| \\
&\quad \le \frac{1}{n} \sum_{i=1}^n \max_{m \in [\![M]\!], \omega \in \{0,1\}} \big| \mu_\omega(X_i) - \mu_\omega(X_{j_m(i)}) - \widehat{\mu}_\omega(X_i) + \widehat{\mu}_\omega(X_{j_m(i)}) \big|. \qquad \text{(S4.6)}
\end{aligned}$$

Let $k = \lfloor d/2 \rfloor + 1$. For any $\omega \in \{0, 1\}$, by Taylor expansion to kth order,

$$\left| \mu_\omega(X_{j_m(i)}) - \mu_\omega(X_i) - \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \sum_{t \in \Lambda_\ell} \partial^t \mu_\omega(X_i) U_{m,i}^t \right| \le \max_{t \in \Lambda_k} \big\| \partial^t \mu_\omega \big\|_\infty \frac{1}{k!} \sum_{t \in \Lambda_k} \|U_{m,i}\|^k. \qquad \text{(S4.7)}$$

In the same way,

$$\left| \widehat{\mu}_\omega(X_{j_m(i)}) - \widehat{\mu}_\omega(X_i) - \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \sum_{t \in \Lambda_\ell} \partial^t \widehat{\mu}_\omega(X_i) U_{m,i}^t \right| \le \max_{t \in \Lambda_k} \big\| \partial^t \widehat{\mu}_\omega \big\|_\infty \frac{1}{k!} \sum_{t \in \Lambda_k} \|U_{m,i}\|^k. \qquad \text{(S4.8)}$$

We also have

$$\left| \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \sum_{t \in \Lambda_\ell} \big(\partial^t \widehat{\mu}_\omega(X_i) - \partial^t \mu_\omega(X_i) \big) U_{m,i}^t \right| \le \sum_{\ell=1}^{k-1} \max_{t \in \Lambda_\ell} \big\| \partial^t \widehat{\mu}_\omega - \partial^t \mu_\omega \big\|_\infty \frac{1}{\ell!} \sum_{t \in \Lambda_\ell} \|U_{m,i}\|^\ell. \qquad \text{(S4.9)}$$

Notice that $\|U_{M,i}\| = \max_{m \in [\![M]\!]} \|U_{m,i}\|$ for any $i \in [\![n]\!]$, $\omega \in \{0, 1\}$. Then for any $\omega \in \{0, 1\}$, plugging (S4.7), (S4.8), (S4.9) into (S4.6), we obtain

$$\begin{aligned}
|B_M - \widehat{B}_M| &\lesssim \left(\max_{\omega \in \{0,1\}} \max_{t \in \Lambda_k} \big\| \partial^t \mu_\omega \big\|_\infty + \max_{\omega \in \{0,1\}} \max_{t \in \Lambda_k} \big\| \partial^t \widehat{\mu}_\omega \big\|_\infty \right) \left(\frac{1}{n} \sum_{i=1}^n \|U_{M,i}\|^k \right) \\
&\quad + \sum_{\ell=1}^{k-1} \left(\max_{\omega \in \{0,1\}} \max_{t \in \Lambda_\ell} \big\| \partial^t \widehat{\mu}_\omega - \partial^t \mu_\omega \big\|_\infty \right) \left(\frac{1}{n} \sum_{i=1}^n \|U_{M,i}\|^\ell \right).
\end{aligned}$$

From Lemma C.2, all moments of $(n/M)^{p/d}\|U_{M,i}\|^p$ are bounded. Then for any positive integer p, using Markov's inequality, we have

$$\frac{1}{n}\sum_{i=1}^{n}\|U_{M,i}\|^p = O_{\mathrm{P}}\Big(\Big(\frac{M}{n}\Big)^{p/d}\Big).$$

By Assumption 4.4 and Assumption 4.5, we then obtain

$$\begin{aligned} B_M - \widehat{B}_M &= O_{\mathrm{P}}(1)O_{\mathrm{P}}\Big(\Big(\frac{M}{n}\Big)^{k/d}\Big) + \max_{\ell\in[\![k-1]\!]} O_{\mathrm{P}}\big(n^{-\gamma_\ell}\big)O_{\mathrm{P}}\Big(\Big(\frac{M}{n}\Big)^{\ell/d}\Big) \\ &= O_{\mathrm{P}}\Big(\Big(\frac{M}{n}\Big)^{k/d}\Big) + \max_{\ell\in[\![k-1]\!]} O_{\mathrm{P}}\Big(n^{-\gamma_\ell}\Big(\frac{M}{n}\Big)^{\ell/d}\Big). \end{aligned}$$

The proof is thus complete by noticing the definition of γ and $M \prec n^{\gamma}$. *Q.E.D.*

S5. PROOFS OF RESULTS IN SUPPLEMENT

S5.1. *Proof of Lemma S3.1*

PROOF OF LEMMA S3.1: The first inequality is directly from the definition of Lebesgue points. The second inequality follows by

$$\begin{aligned} \Big|\frac{\nu(B_{z,\|z-x\|})}{\lambda(B_{z,\|z-x\|})} - f(x)\Big| &\le \frac{1}{\lambda(B_{z,\|z-x\|})}\int_{B_{z,\|z-x\|}} \big|f(y)-f(x)\big|\,\mathrm{d}y \\ &\le \frac{1}{\lambda(B_{z,\|z-x\|})}\int_{B_{x,2\|z-x\|}} \big|f(y)-f(x)\big|\,\mathrm{d}y \\ &= \frac{\lambda(B_{x,2\|z-x\|})}{\lambda(B_{z,\|z-x\|})}\frac{1}{\lambda(B_{x,2\|z-x\|})}\int_{B_{x,2\|z-x\|}} \big|f(y)-f(x)\big|\,\mathrm{d}y \\ &= 2^d\frac{1}{\lambda(B_{x,2\|z-x\|})}\int_{B_{x,2\|z-x\|}} \big|f(y)-f(x)\big|\,\mathrm{d}y, \end{aligned}$$

and then the definition of Lebesgue points. *Q.E.D.*

S5.2. *Proof of Lemma S3.3*

PROOF OF LEMMA S3.3: Fix any $(\nu_0,\nu_1)\in\mathcal{P}_{x,\mathrm{p}}(f_L,f_U,L,d,\delta)$.

We first prove the first claim. First, consider $f_1(x)>0$. For any $\epsilon>0$, there exists $\delta'>0$ such that for any $z\in\mathbb{R}^d$ satisfying $\|z-x\|\le 2\delta'$, we have $|f_0(z)-f_0(x)|\le\epsilon f_0(x)$ and $|f_1(z)-f_1(x)|\le\epsilon f_1(x)$ from the local Lipschitz assumption. We take $w>0$ sufficiently small such that $w<(1-\epsilon)f_0(x)\lambda(B_{0,\delta'})$. Then $W\le w$ implies $\|x-Z\|\le\delta'$. Then for $w>0$ sufficiently small,

$$\mathrm{P}(W\le w) = \mathrm{P}\big(W\le w, \|x-Z\|\le\delta'\big) \le \mathrm{P}\Big(\frac{1-\epsilon}{1+\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|})\le w\Big) = \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}w,$$

and

$$\mathrm{P}(W \le w) = \mathrm{P}\big(W \le w, \|x - Z\| \le \delta'\big) \ge \mathrm{P}\left(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \le w, \|x - Z\| \le \delta'\right)$$

$$= \mathrm{P}\left(\frac{1+\epsilon}{1-\epsilon}\frac{f_0(x)}{f_1(x)}\nu_1(B_{x,\|x-Z\|}) \le w\right) = \frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)} w.$$

Then we have

$$\frac{1-\epsilon}{1+\epsilon}\frac{f_1(x)}{f_0(x)} \le \liminf_{w\to 0} w^{-1}\mathrm{P}(W \le w) \le \limsup_{w\to 0} w^{-1}\mathrm{P}(W \le w) \le \frac{1+\epsilon}{1-\epsilon}\frac{f_1(x)}{f_0(x)}.$$

Since ϵ is arbitrary, we obtain

$$f_W(0) = \lim_{w\to 0} w^{-1}\mathrm{P}(W \le w) = \frac{f_1(x)}{f_0(x)} = r(x).$$

The case for $f_1(x) = 0$ can be established in the same way. This completes the proof of the first claim.

For the second claim, for any $0 < \epsilon < f_L$, there exists $\delta' > 0$ such that for any $z \in \mathbb{R}^d$ satisfying $\|z - x\| \le 2\delta'$, we have $|f_0(z) - f_0(x)| \le \epsilon$ and $|f_1(z) - f_1(x)| \le \epsilon$ from the local Lipschitz assumption. We take N_0 sufficiently large such that $2\frac{M}{N_0} < (f_L - \epsilon)\lambda(B_{0,\delta'})$. Then for any $0 < w \le 2\frac{M}{N_0}$, we have $w < (f_L - \epsilon)\lambda(B_{0,\delta'})$. We take $t > 0$ such that $w + t < (f_L - \epsilon)\lambda(B_{0,\delta'})$. Then for any $(\nu_0, \nu_1) \in \mathcal{P}_{x,\mathrm{p}}(f_L, f_U, L, d, \delta)$,

$$\mathrm{P}(w \le W \le w + t) = \nu_1\big(\big\{z \in \mathbb{R}^d : \nu_0(B_{z,\|x-z\|}) \in [w, w+t]\big\}\big)$$
$$\le \frac{f_1(x)+\epsilon}{f_0(x)-\epsilon}\nu_0\big(\big\{z \in \mathbb{R}^d : \nu_0(B_{z,\|x-z\|}) \in [w, w+t]\big\}\big).$$

Notice that f_0 is lower bounded by f_L. Then for N_0 sufficiently large,

$$\limsup_{t\to 0} t^{-1}\mathrm{P}(w \le W \le w + t) \le \frac{f_1(x)+\epsilon}{f_0(x)-\epsilon}(1+\epsilon).$$

This then completes the proof. *Q.E.D.*

S5.3. *Proof of Lemma S3.4*

PROOF OF LEMMA S3.4: Due to the i.i.d.-ness of Z and $\widetilde{Z}$,

$$\left(\frac{N_0}{M}\right)^2\bigg[\mathrm{P}\left(W \le V, \widetilde{W} \le \widetilde{V}, W \le 2\frac{M}{N_0}, \widetilde{W} \le 2\frac{M}{N_0}\right)$$
$$- \mathrm{P}\left(W \le V, W \le 2\frac{M}{N_0}\right)\mathrm{P}\left(\widetilde{W} \le \widetilde{V}, \widetilde{W} \le 2\frac{M}{N_0}\right)\bigg]$$
$$= \left(\frac{N_0}{M}\right)^2 \int_0^{2\frac{M}{N_0}}\int_0^{2\frac{M}{N_0}} \big[\mathrm{P}(V \ge w_1, \widetilde{V} \ge w_2) - \mathrm{P}(V \ge w_1)\mathrm{P}(\widetilde{V} \ge w_2)\big]$$
$$\times f_W(w_1) f_W(w_2)\, \mathrm{d}w_1\, \mathrm{d}w_2$$

$$
\begin{aligned}
&\le 4\left(\frac{f_U}{f_L}\right)^2\left(\frac{N_0}{M}\right)^2\int_0^{2\frac{M}{N_0}}\int_0^{2\frac{M}{N_0}}\left|\mathrm{P}(V\ge w_1,\widetilde{V}\ge w_2)-\mathrm{P}(V\ge w_1)\mathrm{P}(\widetilde{V}\ge w_2)\right|\mathrm{d}w_1\,\mathrm{d}w_2\\
&=4\left(\frac{f_U}{f_L}\right)^2\int_{-1}^{1}\int_{-1}^{1}\bigg|\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\\
&\quad-\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1)\Big)\mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\bigg|\,\mathrm{d}t_1\,\mathrm{d}t_2,
\end{aligned}
$$

where the last step is from taking $w_1=\frac{M}{N_0}(1+t_1)$ and $w_2=\frac{M}{N_0}(1+t_2)$.

Let

$$
\begin{aligned}
S(t_1,t_2)=\bigg|&\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\\
&-\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1)\Big)\mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\bigg|.
\end{aligned}
$$

If $t_1\ge t_2\ge 0$, $S(t_1,t_2)\le \mathrm{P}(V\ge\frac{M}{N_0}(1+t_1))=\mathrm{P}(U_{(M)}\ge\frac{M}{N_0}(1+t_1))$. If $t_2\ge t_1\ge 0$, $S(t_1,t_2)\le \mathrm{P}(\widetilde{V}\ge\frac{M}{N_0}(1+t_2))=\mathrm{P}(U_{(M)}\ge\frac{M}{N_0}(1+t_2))$. Then for $t_1,t_2\ge 0$,

$$
S(t_1,t_2)\le \mathrm{P}\Big(U_{(M)}\ge\frac{M}{N_0}(1+t_1\vee t_2)\Big).
$$

If $t_1\le t_2\le 0$ and $\mathrm{P}(V\ge\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2))\ge \mathrm{P}(V\ge\frac{M}{N_0}(1+t_1))\mathrm{P}(\widetilde{V}\ge\frac{M}{N_0}(1+t_2))$,

$$
\begin{aligned}
S(t_1,t_2)&\le \mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)-\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1)\Big)\mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\\
&=\mathrm{P}\Big(V\le\frac{M}{N_0}(1+t_1)\Big)\mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\le \mathrm{P}\Big(V\le\frac{M}{N_0}(1+t_1)\Big)\\
&=\mathrm{P}\Big(U_{(M)}\le\frac{M}{N_0}(1+t_1)\Big).
\end{aligned}
$$

If $t_1\le t_2\le 0$ and $\mathrm{P}(V\ge\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2))\le \mathrm{P}(V\ge\frac{M}{N_0}(1+t_1))\mathrm{P}(\widetilde{V}\ge\frac{M}{N_0}(1+t_2))$,

$$
\begin{aligned}
S(t_1,t_2)&\le \mathrm{P}\Big(\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)-\mathrm{P}\Big(V\ge\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\\
&=\mathrm{P}\Big(V\le\frac{M}{N_0}(1+t_1),\widetilde{V}\ge\frac{M}{N_0}(1+t_2)\Big)\le \mathrm{P}\Big(V\le\frac{M}{N_0}(1+t_1)\Big)\\
&=\mathrm{P}\Big(U_{(M)}\le\frac{M}{N_0}(1+t_1)\Big).
\end{aligned}
$$

If $t_2\le t_1\le 0$, we can establish in the same way that

$$
S(t_1,t_2)\le \mathrm{P}\Big(U_{(M)}\le\frac{M}{N_0}(1+t_2)\Big).
$$

Then for $t_1, t_2 \le 0$,

$$S(t_1, t_2) \le \mathrm{P}\Big(U_{(M)} \le \frac{M}{N_0}(1 + t_1 \wedge t_2)\Big).$$

For $t_1 \ge 0 \ge t_2$, if $t_1 + t_2 \ge 0$, $S(t_1, t_2) \le \mathrm{P}(U_{(M)} \ge \frac{M}{N_0}(1 + t_1))$, and if $t_1 + t_2 \le 0$, $S(t_1, t_2) \le \mathrm{P}(U_{(M)} \le \frac{M}{N_0}(1 + t_2))$. Then

$$\begin{aligned}
&\Big(\frac{N_0}{M}\Big)^2 \Big[\mathrm{P}\Big(W \le V, \widetilde{W} \le \widetilde{V}, W \le 2\frac{M}{N_0}, \widetilde{W} \le 2\frac{M}{N_0}\Big) \\
&\quad - \mathrm{P}\Big(W \le V, W \le 2\frac{M}{N_0}\Big)\mathrm{P}\Big(\widetilde{W} \le \widetilde{V}, \widetilde{W} \le 2\frac{M}{N_0}\Big)\Big] \\
&\le 4\Big(\frac{f_U}{f_L}\Big)^2 \int_{-1}^{1}\int_{-1}^{1} S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&= 4\Big(\frac{f_U}{f_L}\Big)^2 \Big[\int_0^1\int_0^1 S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 + \int_{-1}^{0}\int_{-1}^{0} S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&\quad + 2\int_0^1\int_{-1}^{0} S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\Big],
\end{aligned} \tag{S5.1}$$

where the last step is from the symmetry of $S(t_1, t_2)$.

For the first term in (S5.1), by the symmetry of $S(t_1, t_2)$ and the Chernoff bound,

$$\begin{aligned}
&\int_0^1\int_0^1 S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&\le \int_0^\infty\int_0^\infty S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 = 2\int_0^\infty\int_0^\infty S(t_1, t_2)\mathbb{1}(t_1 \ge t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&\le 2\int_0^\infty\int_0^\infty \mathrm{P}\Big(U_{(M)} \ge \frac{M}{N_0}(1 + t_1 \vee t_2)\Big)\mathbb{1}(t_1 \ge t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\
&= 2\int_0^\infty t\mathrm{P}\Big(U_{(M)} \ge \frac{M}{N_0}(1 + t)\Big)\,\mathrm{d}t \le 2\int_0^\infty t(1+t)^M e^{-Mt}\,\mathrm{d}t.
\end{aligned}$$

Notice that since $M \to \infty$, by Stirling's approximation,

$$\int_0^\infty t(1+t)^M e^{-Mt}\,\mathrm{d}t = \frac{1}{M} + \frac{e^M}{M}\int_1^\infty t^M e^{-Mt}\,\mathrm{d}t \le \frac{1}{M}\big(1 + o(1)\big). \tag{S5.2}$$

We then obtain

$$\int_0^1\int_0^1 S(t_1, t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \le \frac{2}{M}\big(1 + o(1)\big). \tag{S5.3}$$

For the second term in (S5.1),

$$\begin{aligned}\int_{-1}^{0}\int_{-1}^{0} S(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 &= 2\int_{-1}^{0}\int_{-1}^{0} S(t_1,t_2)\mathbb{1}(t_1\le t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\\ &\le 2\int_{-1}^{0}\int_{-1}^{0} \mathrm{P}\Big(U_{(M)}\le \frac{M}{N_0}(1+t_1\wedge t_2)\Big)\mathbb{1}(t_1\le t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\\ &= 2\int_0^1 t\mathrm{P}\Big(U_{(M)}\le \frac{M}{N_0}(1-t)\Big)\,\mathrm{d}t \le 2\int_0^1 t(1-t)^M e^{Mt}\,\mathrm{d}t.\end{aligned}$$

Notice that

$$\int_0^1 t(1-t)^M e^{Mt}\,\mathrm{d}t \le \frac{1}{M}. \tag{S5.4}$$

We then obtain

$$\int_{-1}^{0}\int_{-1}^{0} S(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \le \frac{2}{M}. \tag{S5.5}$$

For the third term in (S5.1),

$$\begin{aligned}&\int_0^1\int_{-1}^{0} S(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2\\ &= \int_0^1\int_{-t_1}^{0} \mathrm{P}\Big(U_{(M)}\ge \frac{M}{N_0}(1+t_1)\Big)\,\mathrm{d}t_1\,\mathrm{d}t_2 + \int_0^1\int_{-1}^{-t_1} \mathrm{P}\Big(U_{(M)}\le \frac{M}{N_0}(1+t_2)\Big)\,\mathrm{d}t_1\,\mathrm{d}t_2\\ &= \int_0^1 t\mathrm{P}\Big(U_{(M)}\ge \frac{M}{N_0}(1+t)\Big)\,\mathrm{d}t + \int_{-1}^{0}(-t)\mathrm{P}\Big(U_{(M)}\le \frac{M}{N_0}(1+t)\Big)\,\mathrm{d}t\\ &\le \int_0^\infty t\mathrm{P}\Big(U_{(M)}\ge \frac{M}{N_0}(1+t)\Big)\,\mathrm{d}t + \int_{-1}^{0}(-t)\mathrm{P}\Big(U_{(M)}\le \frac{M}{N_0}(1+t)\Big)\,\mathrm{d}t\\ &\le \frac{1}{M}\big(1+o(1)\big) + \frac{1}{M} = \frac{2}{M}\big(1+o(1)\big),\end{aligned}$$

where the last step is from (S5.2) and (S5.4).

We then obtain

$$\int_0^1\int_{-1}^{0} S(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \le \frac{2}{M}\big(1+o(1)\big). \tag{S5.6}$$

Plugging (S5.3), (S5.5), (S5.6) into (S5.1) yields

$$\begin{aligned}&\Big(\frac{N_0}{M}\Big)^2\Big[\mathrm{P}\Big(W\le V, \widetilde{W}\le \widetilde{V}, W\le 2\frac{M}{N_0}, \widetilde{W}\le 2\frac{M}{N_0}\Big)\\ &\quad - \mathrm{P}\Big(W\le V, W\le 2\frac{M}{N_0}\Big)\mathrm{P}\Big(\widetilde{W}\le \widetilde{V}, \widetilde{W}\le 2\frac{M}{N_0}\Big)\Big] \le 32\Big(\frac{f_U}{f_L}\Big)^2\frac{1}{M}\big(1+o(1)\big),\end{aligned}$$

and thus completes the proof. *Q.E.D.*

REFERENCES

ABADIE, ALBERTO, AND GUIDO W. IMBENS (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. [22]

BILLINGSLEY, PATRICK (2008): *Probability and Measure*. John Wiley and Sons. [21,22]

BOGACHEV, VLADIMIR I., AND MARIA A. S. RUAS (2007): *Measure Theory*, Vol. 1. Springer. [1]

BROWN, RUSSELL A. (2015): "Building k-d Tree in $O(kn\log n)$ Time," *Journal of Computer Graphics Techniques*, 4 (1), 50–68. [1]

CORMEN, THOMAS H., CHARLES E. LEISERSON, RONALD L. RIVEST, AND CLIFFORD STEIN (2009): *Introduction to Algorithms*. MIT press. [1]

FRIEDMAN, JEROME H., JON L. BENTLEY, AND RAPHAEL A. FINKEL (1977): "An Algorithm for Finding Best Matches in Logarithmic Expected Time," *ACM Transactions on Mathematical Software*, 3 (3), 209–226. [1]

ROMANOVSKY, VLADIMIR (1923): "Note on the Moments of a Binomial $(p+q)^n$ About Its Mean," *Biometrika*, 15 (3/4), 410–412. [10]

STEIN, ELIAS M. (2016): *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press. [11]

TSYBAKOV, ALEXANDRE B. (2009): *Introduction to Nonparametric Estimation*. Springer. [19]

VAN DER VAART, AAD W. (1998): *Asymptotic Statistics*. Cambridge University Press. [24]

ZHAO, PUNING, AND LIFENG LAI (2022): "Analysis of KNN Density Estimation," *IEEE Transactions on Information Theory*, 68 (12), 7971–7995. [21]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 22 February, 2022; final version accepted 7 September, 2023; available online 7 September, 2023.