Supplementary Material to "Optimal Nonparametric Inference with Two-Scale Distributional Nearest Neighbors"

Emre Demirkaya, Yingying Fan, Lan Gao, Jinchi Lv,
Patrick Vossler and Jingbo Wang

This Supplementary Material contains a bootstrap estimator for the distribution of TDNN, an application of TDNN in heterogeneous treatment effect estimation and inference, some additional simulation results, and the proofs of all main results and key lemmas, as well as some additional technical details.

A Bootstrap estimator for distribution of TDNN

We now provide an alternative bootstrap method for directly estimating the distribution of the TDNN estimator. Denote by $\mathbb{P}^*$ the distribution of a bootstrap sample $(\mathbf{Z}_1^*, \cdots, \mathbf{Z}_n^*)$ with replacement conditional on the original n observations $(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$. Let us define $\theta^* = \mathbb{E}\big[\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)\big|\mathbf{Z}_1, \cdots, \mathbf{Z}_n\big]$, where the expectation is taken with respect to the resampling distribution $\mathbb{P}^*$. Recall that $\{(\mathbf{X}_{(1)}, Y_{(1)}), \cdots, (\mathbf{X}_{(n)}, Y_{(n)})\}$ is the ascendingly ordered sample by the distance of $\mathbf{X}_i$ to the given point $\mathbf{x}$. Using the result in Biau et al. (2010), we can show that

$$\theta^* = \sum_{i=1}^{n} \left\{ w_1^* \Big[\Big(1 - \frac{i-1}{n}\Big)^{s_1} - \Big(1 - \frac{i}{n}\Big)^{s_1} \Big] + w_2^* \Big[\Big(1 - \frac{i-1}{n}\Big)^{s_2} - \Big(1 - \frac{i}{n}\Big)^{s_2} \Big] \right\} Y_{(i)}. \quad \text{(A.1)}$$

The theorem below shows that the conditional distribution of the bootstrapped TDNN estimator is asymptotically equivalent to the distribution of the TDNN estimator.

Theorem 7. *Assume that all the conditions of Theorem 6 are satisfied. Then we have that as $n \to \infty$,*

$$\sup_{u\in\mathbb{R}} \Big| \mathbb{P}^*\Big\{(s_2/n)^{-1/2}[D_n^*(s_1,s_2)(\mathbf{x}) - \theta^*] \le u\Big\} - \mathbb{P}\Big\{(s_2/n)^{-1/2}[D_n(s_1,s_2)(\mathbf{x}) - \mu(\mathbf{x}) - \Lambda] \le u\Big\}\Big| = o_p(1). \tag{A.2}$$

Theorem 7 lays the theoretical foundation for directly estimating the distribution of the TDNN estimator with the bootstrap. The Glivenko–Cantelli theorem implies that the empirical distribution of i.i.d. observations converges uniformly to the underlying true distribution almost surely as the number of observations grows to infinity. Therefore, practically, we can generate B i.i.d. bootstrap samples $\{(\mathbf{Z}_{b,1}^*, \cdots, \mathbf{Z}_{b,n}^*)\}_{1\le b\le B}$ from $(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$ with replacement for a relatively large value of B. Then we can approximate the distribution $\mathbb{P}\{D_n(s_1,s_2)(\mathbf{x}) - \mu(\mathbf{x}) - \Lambda \le u\}$ using $B^{-1}\sum_{b=1}^B \mathbb{1}\{D_n^{(b)}(s_1,s_2)(\mathbf{x}) - \theta^* \le u\}$ with $\mathbb{1}\{\cdot\}$ representing the indicator function, which is the empirical distribution of $D_n^*(s_1,s_2)(\mathbf{x})$ based on the B bootstrap samples. As a consequence, any quantile of the distribution of $D_n(s_1,s_2)(\mathbf{x})-\mu(\mathbf{x})-\Lambda$ can be approximated by that of the empirical bootstrap distribution $B^{-1}\sum_{b=1}^B \mathbb{1}\{D_n^*(s_1,s_2)(\mathbf{x}) - \theta^* \le u\}$. Accordingly, for each given $\alpha \in (0,1)$, the two-sided $(1-\alpha)$-level confidence interval for the mean regression function $\mu(\mathbf{x})$ can be constructed as $[D_n(s_1,s_2)(\mathbf{x}) - (\widehat{\xi}_{1-\alpha/2} - \theta^*), D_n(s_1,s_2)(\mathbf{x}) - (\widehat{\xi}_{\alpha/2} - \theta^*)]$, where $\widehat{\xi}_{\alpha/2}$ and $\widehat{\xi}_{1-\alpha/2}$ denote the αth and $(1-\alpha)$th sample quantiles of the bootstrap samples $\{D_n^{(b)}(s_1,s_2)(\mathbf{x})\}_{1\le b\le B}$, respectively.

B Application to heterogeneous treatment effect estimation and inference

As an application, we discuss in this section how to exploit the suggested TDNN method to estimate and infer the treatment effects in the potential outcomes model framework (Rubin, 1974; Imbens and Rubin, 2015). The problems of treatment effect estimation and inference have broad applications in a wide variety of scientific areas, ranging from economics to medical studies. In particular, the estimation and inference of the heterogeneous treatment effect (HTE) which focuses on the unit level effect by considering the treatment effect conditional on the pre-treatment covariates have received rapidly growing attention in recent years because of their ability to provide information that the average treatment effect (ATE) cannot provide. For some recent developments, see, e.g., Crump et al. (2008); Lee (2009); Wager and Athey (2018); Wager et al. (2014); Shalit et al. (2017); Hahn et al. (2020); Powers et al. (2017); Zaidi and Mukherjee (2018).

Among the existing literature, the causal k-NN (Hitsch and Misra (2018)) is most closely related to our approach. This method estimates the treatment effect function by taking the difference of two separate k-NN regression function estimates for the treatment group and control group, respectively. The tuning parameter of neighborhood size k was chosen by minimizing the squared difference between the estimated treatment effect function and the propensity score weighted response. However, there lacks theoretical justification for the causal k-NN estimator.

Let $Y_{T=1} \in \mathbb{R}$ and $Y_{T=0} \in \mathbb{R}$ represent the potential outcomes for the treatment and control groups, respectively, where T denotes the treatment indicator with $T = 1$ representing treated and $T = 0$ being untreated. Then the observed scalar response can be written as

$$Y = T\, Y_{T=1} + (1 - T)\, Y_{T=0}.$$

Denote by $\mathbf{X} \in \mathbb{R}^d$ the random feature vector for an individual. We consider the randomized experiment setting which amounts to the choice of constant treatment propensity $\mathbb{P}(T = 1|\mathbf{X}, Y_{T=1}, Y_{T=0}) = 1/2$. Here, $1/2$ can be replaced with any other constant in $(0, 1)$. Given a fixed feature vector $\mathbf{x} \in \mathbb{R}^d$, the heterogeneous treatment effect (HTE) of treatment T on response Y is defined as

$$\tau(\mathbf{x}) = \mathbb{E}\left[Y_{T=1} - Y_{T=0} | \mathbf{X} = \mathbf{x}\right]. \tag{A.3}$$

Since the setting of randomized experiments entails the unconfoundedness given by $(Y_{T=0}, Y_{T=1}) \perp\!\!\!\perp T \mid \mathbf{X}$, our goal of HTE estimation and inference for (A.3) reduces to the problem of nonparametric regression applied separately to the treatment and control groups, giving rise to

$$\begin{aligned} \tau(\mathbf{x}) &= \mathbb{E}[Y_{T=1}|\mathbf{X} = \mathbf{x}] - \mathbb{E}[Y_{T=0}|\mathbf{X} = \mathbf{x}] \\ &= \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, T = 1] - \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, T = 0]. \end{aligned} \tag{A.4}$$

Specifically, let us consider the nonparametric regression model for the treatment group

$$Y_{T=1} = \mu(\mathbf{X}) + \epsilon,$$

where $\mu(\mathbf{X}) = \mathbb{E}[Y_{T=1}|\mathbf{X}]$ denotes the true mean regression function and the model error ϵ with zero mean and finite variance is independent of d-dimensional random feature vector $\mathbf{X}$. Similarly, we can introduce the corresponding nonparametric regression model for the control group; see Section 3.2 for more detailed technical descriptions. We will separately apply TDNN to the control and treatment groups and then combine the resulting estimators together using (A.4) to estimate the heterogeneous treatment effect.

To formally present the asymptotic theory, let us first introduce some necessary notation. Denote by n_1 and n_0 the sizes of the i.i.d. samples from the treatment and control groups, respectively. The assumption of completely randomized experiments entails that $n_0/n_1 \xrightarrow{p} 1$ as $n \to \infty$ and the two samples for the treatment and control groups are independent of each other. Let $\mathbf{x} \in \operatorname{supp}(\mathbf{X}_1) \cap \operatorname{supp}(\mathbf{X}_0)$ be a fixed feature vector, where $\operatorname{supp}(\mathbf{X}_1)$ and $\operatorname{supp}(\mathbf{X}_0)$ stand for the supports of the corresponding feature distributions for the treatment and control groups, respectively. Similarly, denote by $\mu_1(\cdot)$ and $\mu_0(\cdot)$ the true mean regression functions corresponding to responses $Y_{T=1}$ and $Y_{T=0}$, respectively, and ϵ_1 and ϵ_0 the model errors, with the subscript indicating the treatment and control groups, respectively. Then we can construct two individual two-scale DNN estimators $D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x})$ and $D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})$ separately based on the treatment and control samples with pairs of subsampling scales $(s_1^{(1)}, s_2^{(1)})$ and $(s_1^{(0)}, s_2^{(0)})$, respectively.

In view of (A.4), the population version of the heterogeneous treatment effect at the fixed vector $\mathbf{x}$ is given by

$$\tau(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}). \tag{A.5}$$

We estimate $\tau(\mathbf{x})$ using the following TDNN heterogeneous treatment effect estimator

$$\widehat{\tau}(\mathbf{x}) = D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x}). \tag{A.6}$$

The theorem below characterizes the asymptotic distribution of the TDNN HTE estimator $\widehat{\tau}(\mathbf{x})$.

Theorem 8. *Assume that Conditions 1–3 with the subscripts attached hold for both treatment and control groups. Further assume that $s_2^{(i)} \to \infty$, $s_2^{(i)} = o(n)$, and there exist some constants $0 < c_1 < c_2 < 1$ such that $c_1 \leq s_1^{(i)}/s_2^{(i)} \leq c_2$ for $i = 0, 1$. Then for any fixed $\mathbf{x} \in \operatorname{supp}(\mathbf{X}_1) \cap \operatorname{supp}(\mathbf{X}_0) \subset \mathbb{R}^d$, it holds that for some positive sequence σ_n of order $\{(s_2^{(1)} + s_2^{(0)})/n\}^{1/2}$,*

$$\frac{[D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})] - \tau(\mathbf{x}) - \Lambda}{\sigma_n} \xrightarrow{\mathscr{D}} N(0,1) \tag{A.7}$$

as $n \to \infty$, where $\Lambda = O\{(s_1^{(1)})^{-4/d} + (s_2^{(1)})^{-4/d} + (s_1^{(0)})^{-4/d} + (s_2^{(0)})^{-4/d}\}$ for $d \geq 2$ and $\Lambda = O\{(s_1^{(1)})^{-3} + (s_2^{(1)})^{-3} + (s_1^{(0)})^{-3} + (s_2^{(0)})^{-3}\}$ for $d = 1$.

As explained before, the sequence σ_n in Theorem 8 above is a generic notation representing the asymptotic standard deviation of the TDNN heterogeneous treatment effect estimator. We see that the subsampling scales need to satisfy that $s_2^{(i)} \to \infty$ and $s_2^{(i)} = o(n)$

for $i = 0, 1$. The asymptotic bias of the TDNN estimator $\widehat{\tau}(\mathbf{x})$ is only of the second order $O\{(s_1^{(1)})^{-4/d} + (s_2^{(1)})^{-4/d} + (s_1^{(0)})^{-4/d} + (s_2^{(0)})^{-4/d}\}$ for $d \geq 2$ and $O\{(s_1^{(1)})^{-3} + (s_2^{(1)})^{-3} + (s_1^{(0)})^{-3} + (s_2^{(0)})^{-3}\}$ for $d = 1$. The asymptotic variance identified in Theorems 3 and 8 depends generally on the underlying distributions and the fixed vector $\mathbf{x}$, whose complicated form calls for a need to develop practical approaches to the estimation of the asymptotic variance for the TDNN estimator.

For the practical implementation of TDNN for the HTE inference, we advocate the use of the L-statistic representation. Since the single-scale DNN estimator is an L-statistic as shown in Lemma 1, the two-scale DNN estimator, which is a linear combination of a pair of single-scale DNN estimators, is still an L-statistic. We thus can construct a pair of two-scale DNN estimators separately based on the treatment and control subsamples and then take a difference. As suggested by Theorems 6 and 7, we can further bootstrap such difference by resampling within each group to provide tight heterogeneous treatment effect inference. Therefore, the two-scale procedure of TDNN coupled with the bootstrap enjoys both theoretical justifications and computational scalability.

C Additional simulation results

C.1 Comparison with k-NN

We repeat the same simulation study as in Section 5.1 of the main text using the k-NN estimator by varying the neighborhood size k from 1 to 200. The performance of the k-NN estimator is shown in Figure 2. From Figure 2, we see that the finite-sample bias of k-NN tends to increase with the neighborhood size k, which is sensible since moving further away from the fixed test point incurs naturally inflated bias. The MSE plot in Figure 2 shows a similar U-shaped pattern of the bias-variance tradeoff. In contrast, the minimum value of the MSE attained by k-NN is 0.1273, which is outperformed by both the single-scale DNN and TDNN.

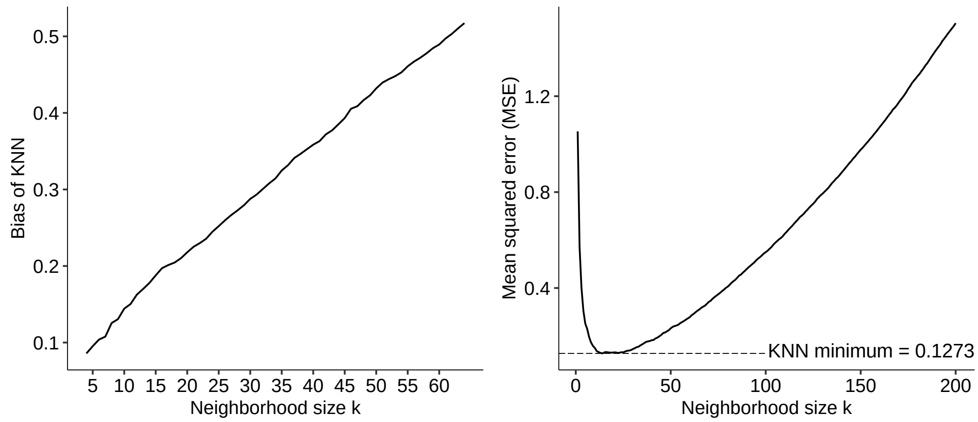

Figure 2: The bias and MSE results for k-NN in Section 5.1.

	Fixed Test Point			Random Test Points		
Method	MSE	Bias^2	Variance	MSE	Bias^2	Variance
DNN	0.0402	0.0038	0.0270	0.1337	0.0815	0.0404
k-NN	0.0488	0.0024	0.0470	0.1826	0.1305	0.0499
TDNN	0.0259	0.0005	0.0252	0.1284	0.0388	0.0649

Table 4: A modified version of the comparison of DNN, k-NN, and TDNN in simulation setting 1 as described in Section 5.2, but with the random feature vector $\mathbf{X}$ drawn from $U([0,1]^3)$ instead of $N(0, I_3)$.

C.2 Simulation setting 1 with uniform design

We repeat simulation setting 1 as described in Section 5.2, but now with random feature vector $\mathbf{X} \sim U([0,1]^p)$ as opposed to the Gaussian design used in the original model setting.

All the parameter settings stay the same as in Section 5.2. From the results in Table 4, we can see that TDNN improve substantially over both DNN and k-NN. Moreover, compared to the results in Table 1 under the Gaussian design, the average MSEs for random test points under the uniform design are now much smaller and closer to the MSE for the fixed test point.

C.3 Simulation setting 3 for HTE estimation and inference

The first two simulation examples in Section 5.2 demonstrate the estimation accuracy of TDNN for general nonparametric regression and the third one will focus on the heterogeneous treatment effect (HTE) estimation and inference with the confidence interval coverage. We use a modified version of the second simulation setting for causal inference in Wager and Athey (2018).

Setting 3. *Assume that the treatment propensity* $e(\mathbf{x}) = 0.5$*, the main effect* $m(\mathbf{x}) = \frac{1}{8}(x_1 - 1)$ *for the control group, and the treatment effect* $\tau(\mathbf{x}) = \varsigma(x_1)\varsigma(x_2)\varsigma(x_3)$ *with* $\varsigma(x) = 1 + \{1 + \exp(-20(x - \frac{1}{3}))\}^{-1}$ *for the treatment group, where* $\mathbf{x} = (x_1, \cdots, x_p)^T$*. Further assume that the feature vector* $\mathbf{X} \sim U([0,1]^p)$ *and the regression error* $\epsilon \sim N(0,1)$ *independent of* $\mathbf{X}$ *for both groups. We increase the ambient dimensionality* p *along the sequence* $\{3, 5, 10, 15, 20\}$.

As with simulation setting 2, we evaluate the performance of the three nonparametric learning and inference methods at a fixed test point chosen as $x_1 = 0.2$, $x_2 = 0.4$, $x_3 = 0.6$, and $x_j = 0.5$ for $j > 3$ as well as for a set of 100 test points randomly drawn from the hypercube $[0,1]^p$. For the TDNN estimator, the ratio $c = s_2/s_1$ is chosen from the sequence $\{2, 4, 6, 8, 10, 15, 20, 25, 30\}$ for random test points and we fix $c = 2$ for the fixed test point for simplicity. The subsampling scale s_1 is chosen from the interval $[s_{\text{sign}}, 2s_{\text{sign}}]$ for each given c, where s_{sign} is given by the sign-change tuning process introduced at the beginning of Section 5. We apply the TDNN estimator to the treatment group and control group

		Fixed Test Point					Random Test Points				
Method	p	MSE	Bias^2	Variance	Coverage	Width	MSE	Bias^2	Variance	Coverage	Width
DNN	3	0.1511	0.0414	0.0977	0.816	1.1541	0.3152	0.1580	0.1066	0.6727	1.2215
k-NN	3	0.1269	0.0517	0.0756	0.856	1.0702	0.3916	0.3130	0.0733	0.5340	1.0511
TDNN	3	0.0899	0.0145	0.0836	0.948	1.1236	0.3022	0.0672	0.1670	0.8196	1.5124
DNN	5	0.1706	0.0430	0.0967	0.801	1.1551	0.3204	0.1612	0.1061	0.6707	1.2188
k-NN	5	0.1320	0.0560	0.0752	0.852	1.0676	0.4013	0.3208	0.0731	0.5262	1.0499
TDNN	5	0.1008	0.0168	0.0833	0.915	1.1209	0.3063	0.0704	0.1668	0.8162	1.5112
DNN	10	0.1600	0.0364	0.0987	0.833	1.1647	0.3337	0.1718	0.1083	0.6635	1.2305
k-NN	10	0.1302	0.0489	0.0780	0.869	1.0866	0.4154	0.3325	0.0750	0.5251	1.0627
TDNN	10	0.1014	0.0113	0.0852	0.934	1.1318	0.3174	0.0764	0.1722	0.8143	1.5336
DNN	15	0.1687	0.0313	0.1019	0.825	1.1808	0.3428	0.1782	0.1093	0.6608	1.2361
k-NN	15	0.1287	0.0500	0.0782	0.872	1.0868	0.4291	0.3427	0.0759	0.5201	1.0682
TDNN	15	0.1021	0.0109	0.0888	0.923	1.1536	0.3237	0.0791	0.1746	0.8124	1.5445
DNN	20	0.1628	0.0382	0.0985	0.820	1.1669	0.3394	0.1757	0.1094	0.6642	1.2368
k-NN	20	0.1330	0.0497	0.0798	0.877	1.0981	0.4232	0.3366	0.0764	0.5248	1.0721
TDNN	20	0.1061	0.0125	0.0892	0.927	1.1564	0.3215	0.0772	0.1748	0.8144	1.5464

Table 5: Comparison of DNN, k-NN, and TDNN in simulation setting 3 described in Section C.3.

separately, and then take the difference between the TDNN estimators for the two groups to estimate the HTE. In addition, we also report the coverage probability of 95% confidence intervals for the HTE constructed based on the asymptotic normality results established in Section B. The DNN and k-NN estimators are similarly applied for estimation and inference of the HTE. In particular, we see from the results in Table 5 that the TDNN estimator indeed provides lower MSEs for HTE estimation and valid confidence intervals for HTE inference with higher coverage compared to the DNN and k-NN estimators.

D Proofs of main results

D.1 Proof of Theorem 1

Let us investigate the higher-order asymptotic expansion for the bias term of the single-scale distributional nearest neighbors (DNN) estimator $D_n(s)(\mathbf{x})$ introduced in (5) under the asymptotic setting when the subsampling scale $s \to \infty$ as the sample size n increases. Recall that the target point $\mathbf{x}$ is a given vector inside the domain $\text{supp}(\mathbf{X}) \subset \mathbb{R}^d$ of the covariate distribution, where the feature dimensionality d is assumed to be fixed for simplifying the technical presentation of our work. The main idea of the proof is to first consider the specific case of $s = n$ in Lemma 5 in Section E.4, and then analyze the general case of $s \to \infty$ by exploiting the projection of the mean function $\mu(\mathbf{X}) = \mathbb{E}(Y|\mathbf{X})$ onto the positive half line $\mathbb{R}_+ = [0, \infty)$ given by $\|\mathbf{X} - \mathbf{x}\|$ in Lemma 6 in Section E.5.

Since $\{i_1, \cdots, i_s\}$ is a random subsample of $\{1, \cdots, n\}$ with subsampling scale s, in view of (4) and (5) we have

$$\begin{aligned}\mathbb{E}\, D_n(s)(\mathbf{x}) &= \mathbb{E}\, \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_s}) \\ &= \mathbb{E}\, [Y_{(1)}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_s})] \\ &= \mathbb{E}\, [m(r_{(1)})(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_s})], \end{aligned} \tag{A.8}$$

where the kernel $\Phi(\mathbf{x};\cdot)$ in the U-statistic representation of the DNN estimator is simply the 1-nearest neighbor (1NN) estimator $Y_{(1)}(\cdot)$ given by the response for the closest neighbor $\mathbf{X}_{i_{(1)}}$ of $\mathbf{x}$ in the random subsample $\{\mathbf{X}_{i_1},\cdots,\mathbf{X}_{i_s}\}$ with $\mathbf{Z}_{i_j}$ denoting $(\mathbf{X}_{i_j},Y_{i_j})$, $m(r)=\mathbb{E}(Y\mid\|\mathbf{X}-\mathbf{x}\|=r)$ is the projection of the mean function $\mu(\mathbf{X})$ onto the positive half line introduced in (A.111) in Lemma 6, and $r_{(1)}=\|\mathbf{X}_{i_{(1)}}-\mathbf{x}\|$. The representation in (A.8) provides a useful starting point for our technical analysis.

From (A.8) above, we see that it is necessary to first study the asymptotic behavior of term $r_{(1)}$. Without loss of generality, for this step we can simply replace parameter s with parameter n since both subsample size s and full sample size n are assumed to diverge simultaneously. With such a notational simplification, the 1NN $\mathbf{X}_{i_{(1)}}$ of $\mathbf{x}$ in the subsample becomes the 1NN $\mathbf{X}_{(1)}$ of $\mathbf{x}$ in the full sample and thus $r_{(1)}=\|\mathbf{X}_{(1)}-\mathbf{x}\|$. We see from Lemma 5 that $\mathbb{E}r_{(1)}^2=\mathbb{E}\|\mathbf{X}_{(1)}-\mathbf{x}\|^2$ admits a higher-order asymptotic expansion with explicit constants provided for the first two leading orders, which are $n^{-2/d}$ and $n^{-4/d}$, respectively, for $d\geq 2$ as shown in (A.97) and (A.98), and n^{-2} and n^{-3}, respectively, for $d=1$ as shown in (A.98). To apply such an asymptotic expansion in Lemma 5 to the term $r_{(1)}=\|\mathbf{X}_{i_{(1)}}-\mathbf{x}\|$ in (A.8), we now need to replace parameter n back with parameter s, which also diverges by assumption.

A natural next step is to consider the expectation on the right-hand side of (A.8) by conditioning on $r_{(1)}=\|\mathbf{X}_{i_{(1)}}-\mathbf{x}\|$. Indeed, this motivates us to investigate the higher-order asymptotic expansion of the projected mean function $m(r)=\mathbb{E}(Y\mid\|\mathbf{X}-\mathbf{x}\|=r)$ in Lemma 6, where $r\to 0$ and some constants are given for the first two leading orders r^2 and r^4 in (A.113). Observe that the asymptotic regime of $r\to 0$ is reasonable since it has been shown by Lemma 2.2 in Biau and Devroye (2015) that $r_{(1)}=\|\mathbf{X}_{i_{(1)}}-\mathbf{x}\|\to 0$ almost surely as $s\to\infty$.

Based on the expansion of $\mathbb{E}\|\mathbf{X}_{(1)}-\mathbf{x}\|^2$ under different regimes of d provided in (A.96)–(A.98) in Lemma 5, we can see that there are two cases for the expansion of $\mathbb{E}\|\mathbf{X}_{(1)}-\mathbf{x}\|^2$.

Specifically, the first two leading orders are n^{-2} and n^{-3} for $d = 1$, while the first two leading orders are $n^{-2/d}$ and $n^{-4/d}$ for $d \geq 2$. Thus, we calculate $\mathbb{E}\, D_n(s)(\mathbf{x})$ for $d \geq 2$ and $d = 1$, separately.

First, for the case of $d = 1$, combining the arguments above using (A.96) and Lemma 6, from (A.8) we can deduce that

$$\begin{aligned}
\mathbb{E}\, D_n(s)(\mathbf{x}) &= \mu(\mathbf{x}) + \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})}\,\mathbb{E}r_{(1)}^2 + O_4\mathbb{E}r_{(1)}^4 \\
&= \mu(\mathbf{x}) + \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})} \\
&\quad \times \left(\frac{\Gamma(2/d+1)}{(f(\mathbf{x})V_d)^{2/d}} s^{-2/d} - \left(\frac{\Gamma(2/d+2)}{d(f(\mathbf{x})V_d)^{2/d}} \right) s^{-(1+2/d)} \right) \\
&\quad + O_4 \frac{\Gamma(4/d+1)}{(f(\mathbf{x})V_d)^{4/d}} s^{-4/d} + o(s^{-(1+2/d)}) \\
&= \mu(\mathbf{x}) + \Gamma(2/d+1) \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,V_d^{2/d}\,f(\mathbf{x})^{2/d+1}} s^{-2/d} + R(s),
\end{aligned} \tag{A.9}$$

where $R(s) = O(s^{-3})$. In addition, $\Gamma(\cdot)$ denotes the gamma function, $V_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$, $f'(\mathbf{x})$ and $\mu'(\mathbf{x})$ represent the first-order gradients of $f(\mathbf{x})$ and $\mu(\mathbf{x})$ at $\mathbf{x}$, respectively, $\mu''(\mathbf{x})$ denotes the Hessian matrix of $\mu(\cdot)$ at $\mathbf{x}$, O_4 is some constant given in Lemma 6, and $\mathrm{tr}(\cdot)$ stands for the trace operator.

We proceed to prove for the case of $d \geq 2$. In the same fashion of deriving (A.9), applying (A.97)–(A.98) and Lemma 6, from (A.8) we can obtain that

$$\begin{aligned}
\mathbb{E}\, D_n(s)(\mathbf{x}) &= \mu(\mathbf{x}) + \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})}\,\mathbb{E}r_{(1)}^2 + O_4\mathbb{E}r_{(1)}^4 \\
&= \mu(\mathbf{x}) + \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})} \times \left(\frac{\Gamma(2/d+1)}{(f(\mathbf{x})V_d)^{2/d}} s^{-2/d} - C(d,f,\mu,\mathbf{x}) s^{-4/d} \right) \\
&\quad + O_4 \frac{\Gamma(4/d+1)}{(f(\mathbf{x})V_d)^{4/d}} s^{-4/d} + o(s^{-4/d}) \\
&= \mu(\mathbf{x}) + \Gamma(2/d+1) \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,V_d^{2/d}\,f(\mathbf{x})^{2/d+1}} s^{-2/d} + R(s),
\end{aligned} \tag{A.10}$$

where $R(s) = O(s^{-4/d})$.

Therefore, combining the above results, we obtain the desired higher-order asymptotic expansion for the bias term of the single-scale DNN estimator $B(s) = \mathbb{E}\, D_n(s)(\mathbf{x}) - \mu(\mathbf{x})$. This completes the proof of Theorem 1.

D.2 Proof of Theorem 2

We now proceed to prove the asymptotic normality of the single-scale DNN estimator $D_n(s)(\mathbf{x})$. Recall that in Theorem 1, the higher-order asymptotic expansion for the bias term $B(s)$ of $D_n(s)(\mathbf{x})$ requires the assumption that the subsampling scale $s \to \infty$ as sample size n increases. As shown in the proof of Theorem 1 in Section D.1, the single-scale DNN estimator $D_n(s)(\mathbf{x})$ reduces to the 1NN estimator when we choose $s = n$, since in such a case, there is a single subsample with size $s = n$, i.e., the full sample. We immediately realize that although the choice of $s = n$ satisfies the need on the bias side, it does not make the variance shrink asymptotically. Intuitively, we would need to form the empirical average over a diverging number of such individual estimates in order to establish the desired asymptotic normality. This naturally calls for the assumption of $s = o(n)$, which entails that the total number of these individual estimates $\binom{n}{s}$ diverges as sample size n increases. Thus we will work with the asymptotic regime of subsampling scale with $s \to \infty$ and $s = o(n)$.

In view of the U-statistic representation of $D_n(s)(\mathbf{x})$ given in (5), a natural idea of the proof for the asymptotic normality of the single-scale DNN estimator is to exploit the asymptotic theory of the U-statistic framework. However, the classical U-statistic asymptotic theory is not readily applicable due to the common assumption of *fixed* subsampling scale s. In contrast, as discussed above, our asymptotic analysis needs the opposite assumption of *diverging* subsampling scale s, i.e., $s \to \infty$. Such a discrepancy causes additional technical challenges when we derive the asymptotic normality.

Let us first exploit Hoeffding's canonical decomposition introduced in Hoeffding (1948), which is an extension of the projection idea. For each $1 \leq i \leq s$, we define the centered conditional expectation

$$\begin{aligned}\widetilde{\Phi}_i(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) = \mathbb{E}[\Phi(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i, \mathbf{Z}_{i+1}, \cdots, \mathbf{Z}_s) \,|\, \mathbf{z}_1, \cdots, \mathbf{z}_i] \\ - \mathbb{E}\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s),\end{aligned} \tag{A.11}$$

where $\Phi(\mathbf{x}; \cdot)$ is the kernel defined in (4) for the U-statistic representation of the single-scale DNN estimator. Then in light of (A.11), for each $1 \leq i \leq s$ we can successively define the canonical term

$$g_i(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) = \widetilde{\Phi}_i(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) - \sum_{j=1}^{i-1} \sum_{1 \leq \alpha_1 < \cdots < \alpha_j \leq i} g_j(\mathbf{x}; \mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_j}), \tag{A.12}$$

where $g_1(\mathbf{x}; \mathbf{z}_1) = \widetilde{\Phi}_1(\mathbf{x}; \mathbf{z}_1)$ by definition. Combining (4), (A.11), and (A.12), we see that the kernel $\Phi(\mathbf{x}; \cdot)$ can be rewritten as a sum of the canonical terms

$$\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s) - \mathbb{E}\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s) = \sum_{j=1}^{s} \sum_{1 \leq \alpha_1 < \cdots < \alpha_j \leq s} g_j(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_j}). \tag{A.13}$$

Moreover, it holds that

$$\mathrm{Var}(\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s)) = \sum_{j=1}^{s} \binom{s}{j} \mathrm{Var}(g_j(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_j)). \tag{A.14}$$

The above Hoeffding's canonical decomposition in (A.13) plays an important role in establishing the asymptotic normality.

In view of (5), (A.11), and (A.13), we can deduce that

$$\begin{aligned} D_n(s) - \mathbb{E}\, D_n(s) &= \binom{n}{s}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_s \leq n} \widetilde{\Phi}_s(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_s}) \\ &= \binom{n}{s}^{-1} \Big\{ \binom{n-1}{s-1} \sum_{i_1=1}^{n} g_1(\mathbf{x}; \mathbf{Z}_{i_1}) + \binom{n-2}{s-2} \sum_{1 \leq i_1 < i_2 \leq n} g_2(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) + \cdots \\ &\quad + \binom{n-s}{s-s} \sum_{1 \leq i_1 < i_2 < \cdots < i_s \leq n} g_s(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_s}) \Big\}. \end{aligned} \tag{A.15}$$

From the above Hoeffding's canonical decomposition in (A.15) for the single-scale DNN estimator, we see that the Hájek projection introduced in Hájek (1968) of the centered DNN estimator $D_n(s) - \mathbb{E}D_n(s)$ is given by

$$\widehat{D}_n(s) = \binom{n}{s}^{-1} \binom{n-1}{s-1} \sum_{i=1}^{n} g_1(\mathbf{x}; \mathbf{Z}_i), \tag{A.16}$$

which is the first-order part of the decomposition in (A.15).

A useful observation is that the Hájek projection given in (A.16) involves the sum of some independent and identically distributed (i.i.d.) terms. Denote by σ_n^2 the variance of the Hájek projection. Then it follows from $g_1(\mathbf{x}; \mathbf{z}_1) = \widetilde{\Phi}_1(\mathbf{x}; \mathbf{z}_1)$ and (A.11) that

$$\begin{aligned} \sigma_n^2 = \mathrm{Var}(\widehat{D}_n(s)) &= \frac{s^2}{n} \mathrm{Var}(\widetilde{\Phi}_1(\mathbf{x}; \mathbf{Z}_1)) \\ &= \frac{s^2}{n} \mathrm{Var}(\Phi_1(\mathbf{x}; \mathbf{Z}_1)) = \frac{s^2}{n} \eta_1, \end{aligned} \tag{A.17}$$

where the non-centered conditional expectation $\Phi_1(\mathbf{x}; \mathbf{Z}_1)$ is defined later in (A.128) and η_1 is defined as the variance of $\Phi_1(\mathbf{x}; \mathbf{Z}_1)$. From (A.11), we see that each term $g_1(\mathbf{x}; \mathbf{Z}_i) = \widetilde{\Phi}_1(\mathbf{x}; \mathbf{Z}_i)$ of the i.i.d. sum in (A.16) has zero mean. Thus by (A.17), an application of the Lindeberg–Lévy central limit theorem in Borovkov (2013) leads to

$$\frac{\widehat{D}_n(s)}{\sigma_n} \xrightarrow{\mathscr{D}} N(0, 1), \tag{A.18}$$

which establishes the asymptotic normality of the Hájek projection $\widehat{D}_n(s)$.

Finally, we aim to show that similar asymptotic normality as above holds when the Hájek projection $\widehat{D}_n(s)$ in the numerator on the left-hand side of (A.18) is replaced with the centered single-scale DNN estimator $D_n(s)(\mathbf{x}) - \mathbb{E}\, D_n(s)(\mathbf{x}) = D_n(s)(\mathbf{x}) - \mu(\mathbf{x}) - B(s)$, where $B(s)$ is the bias term identified in Theorem 1. With the aid of Slutsky's lemma, we see that it suffices to show that

$$\frac{D_n(s) - \mathbb{E}D_n(s) - \widehat{D}_n(s)}{\sigma_n} = o_P(1). \tag{A.19}$$

Following Lemma 3.3 in Wager and Athey (2018) and replacing "tree in forest" with "kernel in U-statistics" in the proof, we can easily see that

$$\mathbb{E}[D_n(s) - \mathbb{E}D_n(s) - \widehat{D}_n(s)]^2 \leq \frac{s^2}{n^2}\operatorname{Var}(\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s)). \tag{A.20}$$

It remains to bound the variance term $\operatorname{Var}(\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s))$ above.

By Lemma 7 in Section E.6, we have an important result that

$$\operatorname{Var}(\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_s)) = o(n\eta_1). \tag{A.21}$$

Combining (A.17), (A.20), and (A.21), it holds that

$$\begin{aligned}\mathbb{E}\left[\frac{D_n(s) - \mathbb{E}D_n(s) - \widehat{D}_n(s))}{\sigma_n}\right]^2 &= o\left\{\frac{1}{\sigma_n^2}\frac{s^2}{n^2}(n\eta_1)\right\} \\ = o\left\{\frac{n}{s^2\eta_1}\frac{s^2}{n^2}(n\eta_1)\right\} &= o(1). \end{aligned}\tag{A.22}$$

Therefore, we are ready to see that (A.22) entails the desired claim (A.19). Finally, by (A.17) and (A.133) obtained in the proof of Lemma 7 in Section E.6, we see that σ_n is of order $(s/n)^{1/2}$, which concludes the proof of Theorem 2.

D.3 Proof of Theorem 3

We further prove the asymptotic normality of the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ introduced in (11). It is worth mentioning that Theorem 3 is not a simple consequence of Theorem 2 since the marginal asymptotic normalities do not necessarily lead to the joint asymptotic normality. This means that we need to analyze the two single-scale DNN estimators involved in the definition of the two-scale DNN estimator in a joint fashion. To this end, we will exploit the ideas in the proof of Theorem 2 in Section D.2. To facilitate the technical analysis, some key technical tools are provided in Lemmas 8–10 in Sections E.7–E.9, respectively.

Without loss of generality, let us assume that $s_1 < s_2$ for the two subsampling scales. In particular, we make the assumptions that $s_1, s_2 \to \infty$, $s_1, s_2 = o(n)$, and $c_1 \leq s_1/s_2 \leq c_2$ for some constants $0 < c_1 < c_2 < 1$. From Lemma 8 in Section E.7, we see that the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ is also a U-statistic of order s_2 with a new kernel $\Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2})$ introduced later in (A.135). Thus Hoeffding's canonical decomposition for U-statistics can be applied to derive the asymptotic normality of the two-scale DNN estimator. For each $1 \leq i \leq s_2$, let us define

$$\Phi_i^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) = \mathbb{E}[\Phi^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i, \mathbf{Z}_{i+1}, \cdots, \mathbf{Z}_{s_2}) \,|\, \mathbf{z}_1, \cdots, \mathbf{z}_i], \tag{A.23}$$

$$\begin{aligned} g_i^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) = \Phi_i^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) - \mathbb{E}\Phi_i^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_i) \\ - \sum_{j=1}^{i-1} \sum_{1 \leq \alpha_1 < \cdots < \alpha_j \leq i} g_j^*(\mathbf{x}; \mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_j}), \end{aligned} \tag{A.24}$$

where $g_1^*(\mathbf{x}; \mathbf{z}_1) = \Phi_1^*(\mathbf{x}; \mathbf{z}_1) - \mathbb{E}\Phi_1^*(\mathbf{x}; \mathbf{Z}_1)$ by definition. We further define

$$\operatorname{Var} \Phi^* = \operatorname{Var}(\Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2})) \quad \text{and} \quad \eta_1^* = \operatorname{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1)). \tag{A.25}$$

In view of (11), (A.23), and (A.24), an application of similar U-statistic and Hoeffding's canonical decomposition arguments to those in the proof of Theorem 2 in Section D.2 entails that

$$(n^{-1} s_2^2 \eta_1^*)^{-1/2} \big(D_n(s_1, s_2)(\mathbf{x}) - \mathbb{E}[D_n(s_1, s_2)(\mathbf{x})]\big) \tag{A.26}$$

can be approximated by the first-order part of Hoeffding's canonical decomposition that converges to a normal distribution with the remainders asymptotically negligible, where η_1^* is given in (A.25). More specifically, denote by

$$\widehat{D}_n(s_1, s_2) = \frac{s_2}{n} \sum_{i=1}^{n} g_1^*(\mathbf{x}; \mathbf{Z}_i), \tag{A.27}$$

where $g_1^*(\mathbf{x};\mathbf{Z}_i)$ is defined in (A.24). It follows from (A.25), (A.27), and the classical central limit theorem for i.i.d. random variables that

$$\frac{\widehat{D}_n(s_1,s_2)}{\sqrt{n^{-1}s_2^2\eta_1^*}} \xrightarrow{\mathscr{D}} N(0,1), \tag{A.28}$$

since it holds that $\mathrm{Var}(g_1^*(\mathbf{x};\mathbf{Z}_1)) = \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)) = \eta_1^*$.

Similar to (A.20), by (A.24), (A.25), and (A.27) we can deduce that

$$\begin{aligned}&\frac{\mathbb{E}\big[D_n(s_1,s_2)(\mathbf{x}) - \mathbb{E}D_n(s_1,s_2)(\mathbf{x}) - \widehat{D}_n(s_1,s_2)\big]^2}{n^{-1}s_2^2\eta_1^*}\\ &\quad\le \frac{n^{-2}s_2^2\,\mathrm{Var}\,\Phi^*}{n^{-1}s_2^2\eta_1^*} = \frac{\mathrm{Var}\,\Phi^*}{n\eta_1^*}.\end{aligned} \tag{A.29}$$

Moreover, it follows from the upper bound on $\mathrm{Var}\,\Phi^*$ obtained in Lemma 9 in Section E.8 and the asymptotic order of η_1^* established in Lemma 10 in Section E.9 that

$$\mathrm{Var}\,\Phi^*/(n\eta_1^*) \to 0 \tag{A.30}$$

since $s_2/n \to 0$ by assumption. Therefore, combining (A.28)–(A.30), an application of Slutsky's lemma yields the desired claim in (A.26), that is,

$$\frac{D_n(s_1,s_2)(\mathbf{x}) - \mathbb{E}D_n(s_1,s_2)(\mathbf{x})}{\sigma_n} \xrightarrow{\mathscr{D}} N(0,1), \tag{A.31}$$

where we define $\sigma_n^2 = n^{-1}s_2^2\eta_1^*$. Finally, we see from Lemma 10 that $\sigma_n = (n^{-1}s_2^2\eta_1^*)^{1/2}$ is of order $(s_2/n)^{1/2}$, and from the higher-order asymptotic expansion of the bias term in Theorem 1 that

$$\Lambda = \mathbb{E}D_n(s_1,s_2)(\mathbf{x}) - \mu(\mathbf{x}) = \begin{cases} O(s_1^{-4/d} + s_2^{-4/d}), & d \ge 2,\\ O(s_1^{-3} + s_2^{-3}), & d = 1.\end{cases}$$

This together with (A.31) completes the proof of Theorem 3.

D.4 Proof of Theorem 4

The main idea of the proof is to apply the bias-variance decomposition for the mean-squared error. Recall that $D_n(s_1, s_2)(\mathbf{x}) = w_1^* D_n(s_1)(\mathbf{x}) + w_2^* D_n(s_2)(\mathbf{x})$ and

$$\mathbb{E}[D_n(s_1, s_2)(\mathbf{x})] = w_1^* \mathbb{E}[\mu(\mathbf{X}_{(1)}(s_1))] + w_2^* \mathbb{E}[\mu(\mathbf{X}_{(1)}(s_2))],$$

where $\mathbf{X}_{(1)}(s_1) = \mathbf{X}_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_1})$ denotes the 1-nearest neighbor of $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_{s_1}\}$ and similarly, $\mathbf{X}_{(1)}(s_2) = \mathbf{X}_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_2})$. Then we have the bias-variance decomposition

$$\begin{aligned}
&\mathbb{E}\big([D_n(s_1, s_2)(\mathbf{x}) - \mu(\mathbf{x})]^2\big) \\
&= \mathbb{E}\Big\{\big(D_n(s_1, s_2)(\mathbf{x}) - w_1^* \mathbb{E}[\mu(\mathbf{X}_{(1)}(s_1))] - w_2^* \mathbb{E}[\mu(\mathbf{X}_{(1)}(s_2))]\big)^2\Big\} \\
&\quad + \big[\mathbb{E}(D_n(s_1, s_2)(\mathbf{x})) - \mu(\mathbf{x})\big]^2 \\
&:= I_1(\mathbf{x}) + I_2(\mathbf{x}).
\end{aligned} \tag{A.32}$$

Let us first deal with the bias term $I_2(\mathbf{x})$. Using the similar arguments to those in the proofs of Lemmas 5 and 6, we can deduce that

$$I_2(\mathbf{x}) \leq \begin{cases} \frac{R_1^2(\mathbf{X}, d, f, \mu)}{(c-1)^2} c^{-1} s_2^{-6}, & d = 1, \\ \\ \frac{R_2^2(\mathbf{X}, d, f, \mu)}{(c-1)^2} c^{-2} s_2^{-8/d}, & d \geq 2, \end{cases}$$

where $R_1(\mathbf{x}, d, f, \mu)$ and $R_2(\mathbf{x}, d, f, \mu)$ are some constants depending on the bounds for the first four derivatives of $f(\cdot)$ and $\mu(\cdot)$ in a neighborhood of $\mathbf{x}$.

We now analyze the variance term $I_1(\mathbf{x})$. It holds that

$$\begin{aligned}
I_1(\mathbf{x}) \leq (w_1^*)^2 \mathbb{E}\Bigg\{\binom{n}{s_1}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{s_1} \leq n} \Big(Y_{(1)}(\mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}) - \mathbb{E}\mu(\mathbf{X}_{(1)}(s_1))\Big)\Bigg\}^2 \\
+ (w_2^*)^2 \mathbb{E}\Bigg\{\binom{n}{s_2}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{s_2} \leq n} \Big(Y_{(1)}(\mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}}) - \mathbb{E}\mu(\mathbf{X}_{(1)}(s_2))\Big)\Bigg\}^2.
\end{aligned} \tag{A.33}$$

By the variance decomposition for the U-statistics shown in the proof of Theorem 2 in Section D.2, we can obtain that

$$\begin{aligned} I_1(\mathbf{x}) \leq (w_1^*)^2 \Big(\frac{s_1^2}{n^2} \mathrm{var}(Y_{(1)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})) + \frac{s_1^2}{n} \mathrm{var}(\mathbb{E}[Y_{(1)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1}) | \mathbf{X}_1]) \Big) \\ + (w_2^*)^2 \Big(\frac{s_2^2}{n^2} \mathrm{var}(Y_{(1)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})) + \frac{s_2^2}{n} \mathrm{var}(\mathbb{E}[Y_{(1)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2}) | \mathbf{X}_1]) \Big). \end{aligned} \tag{A.34}$$

Observe that we have shown in the proof of Lemma 7 that

$$\begin{aligned} \mathrm{var}(Y_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_1})) &= \mathrm{var}(\mu(X_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_1})) + \epsilon) \\ &\leq \mu^2(\mathbf{x}) + \sigma^2 + o(1). \end{aligned} \tag{A.35}$$

Moreover, it follows from (A.132) that

$$\begin{aligned} \mathrm{var}(\mathbb{E}[Y_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_1}) | \mathbf{X}_1]) &\leq s_1^{-1} \mathrm{var}(Y_{(1)}(\mathbf{X}_1, \cdots, \mathbf{X}_{s_1})) \\ &\leq s_1^{-1} (\mu^2(\mathbf{x}) + \sigma^2 + o(1)). \end{aligned}$$

Similar results also hold for terms related to s_2. Thus, we have

$$I_1(\mathbf{x}) \leq (\mu^2(\mathbf{x}) + \sigma^2 + o(1)) \Big[(w_1^*)^2 \cdot \frac{s_1}{n} + (w_2^*)^2 \cdot \frac{s_2}{n} \Big]. \tag{A.36}$$

Finally, the desired results can be derived by combining the above bounds for the bias and variance. This concludes the proof of Theorem 4.

D.5 Proof of Theorem 5

We now aim to establish the consistency of the jackknife estimator $\widehat{\sigma}_J^2$ introduced in (24) for the variance σ_n^2 of the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ as defined in (22). We will build on the technique in Arvesen (1969) that expands and reorganizes the jackknife estimator $\widehat{\sigma}_J^2$. However, a major theoretical challenge is that instead of an application of the classical asymptotic theory for the case of fixed order, a more delicate technical analysis of the remainders is essential to proving the consistency under our current assumption of diverging order $s_2 \to \infty$.

More specifically, we will show that the jackknife estimator $\widehat{\sigma}_J^2$ can be written as a weighted sum of a sequence of U-statistics $\{U_c\}_{0 \leq c \leq s_2}$ to be introduced in (A.42) later, where U_0 and U_1 are the dominating terms and the remaining ones are asymptotically negligible under the assumption of $s_2 = o(n^{1/3})$. Since U-statistics are symmetric with respect to the input arguments, it follows from (21) and (23) that

$$\sum_{i=1}^{n} \binom{n-1}{s_2} U_{n-1}^{(i)} = (n - s_2) \binom{n}{s_2} D_n(s_1, s_2)(\mathbf{x}),$$

which entails that

$$n^{-1} \sum_{i=1}^{n} U_{n-1}^{(i)} = D_n(s_1, s_2)(\mathbf{x}). \tag{A.37}$$

Thus, in light of the definition of the jackknife estimator $\widehat{\sigma}_J^2$ in (24) and (A.37), we can deduce that

$$\begin{aligned} n\widehat{\sigma}_J^2 &= (n-1)\left\{ \sum_{i=1}^{n} \left(U_{n-1}^{(i)}\right)^2 - n(D_n(s_1, s_2)(\mathbf{x}))^2 \right\} \\ &= (n-1)\left\{ \binom{n-1}{s_2}^{-2} \sum_{i=1}^{n} \sum_{i} \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1^i}, \cdots, \mathbf{Z}_{\alpha_{s_2}^i}) \Phi^*(\mathbf{x}; \mathbf{Z}_{\beta_1^i}, \cdots, \mathbf{Z}_{\beta_{s_2}^i}) \right. \\ &\quad \left. - n\binom{n}{s_2}^{-2} \sum \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_{s_2}}) \Phi^*(\mathbf{x}; \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2}}) \right\}, \end{aligned} \tag{A.38}$$

where we use the shorthand notation $\sum_{i}$ for

$$\sum_{\substack{1 \leq \alpha_1^i < \alpha_2^i < \cdots < \alpha_{s_2}^i \leq n \\ 1 \leq \beta_1^i < \beta_2^i < \cdots < \beta_{s_2}^i \leq n \\ \alpha_1^i, \alpha_2^i, \cdots, \alpha_{s_2}^i \neq i;\, \beta_1^i, \beta_2^i, \cdots, \beta_{s_2}^i \neq i}} \tag{A.39}$$

and $\sum$ for

$$\sum_{\substack{1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{s_2} \leq n \\ 1 \leq \beta_1 < \beta_2 < \cdots < \beta_{s_2} \leq n}} \tag{A.40}$$

to simplify the technical presentation.

For each $0 \leq c \leq s_2$, by calculating the number of terms with c overlapping components in $\Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_{s_2}})\Phi^*(\mathbf{x}; \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2}})$, we can obtain from (A.38)–(A.40) that

$$\begin{aligned} n\widehat{\sigma}_J^2 &= (n-1)\Bigg\{ \binom{n-1}{s_2}^{-2} \sum_{c=0}^{s_2} (n - 2s_2 + c) \sum \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2-c}}) \\ &\quad \cdot \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\gamma_1}, \cdots, \mathbf{Z}_{\gamma_{s_2-c}}) \\ &\quad - n\binom{n}{s_2}^{-2} \sum_{c=0}^{s_2} \sum \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2-c}}) \\ &\quad \cdot \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\gamma_1}, \cdots, \mathbf{Z}_{\gamma_{s_2-c}}) \Bigg\} \\ &= \frac{n-1}{n} \binom{n-1}{s_2}^{-2} \sum_{c=0}^{s_2} (cn - s_2^2) \binom{n}{2s_2 - c} \binom{2s_2 - c}{s_2} \binom{s_2}{c} U_c, \end{aligned} \tag{A.41}$$

where we introduce a sequence of U-statistics $\{U_c\}_{0 \leq c \leq s_2}$ defined as

$$\begin{aligned} U_c &= \left\{ \binom{n}{2s_2 - c} \binom{2s_2 - c}{s_2} \binom{s_2}{c} \right\}^{-1} \\ &\quad \cdot \sum \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2-c}}) \Phi^*(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\gamma_1}, \cdots, \mathbf{Z}_{\gamma_{s_2-c}}). \end{aligned} \tag{A.42}$$

Here, with slight abuse of notation, $\sum$ is short for denoting the summation over all possible combinations of distinct $\alpha_1, \cdots, \alpha_c, \beta_1, \cdots, \beta_{s_2-c}, \gamma_1, \cdots, \gamma_{s_2-c}$ satisfying that $1 \leq \alpha_1 < \cdots < \alpha_c \leq n$, $1 \leq \beta_1 < \cdots < \beta_{s_2-c} \leq n$, and $1 \leq \gamma_1 < \cdots < \gamma_{s_2-c} \leq n$.

Observe that by symmetrization, U_c defined in (A.42) is indeed a U-statistic that can be represented as

$$U_c = \binom{n}{2s_2 - c}^{-1} \sum_{C_{2s_2-c}} K^{(c)}(\mathbf{x}; \mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_c}, \mathbf{Z}_{\beta_1}, \cdots, \mathbf{Z}_{\beta_{s_2-c}}, \mathbf{Z}_{\gamma_1}, \cdots, \mathbf{Z}_{\gamma_{s_2-c}}), \tag{A.43}$$

where $\sum_{C_{2s_2-c}}$ represents the summation taken over all combinations of $1 \leq \alpha_1 < \cdots < \alpha_c < \beta_1 < \cdots < \beta_{s_2-c} < \gamma_1 < \cdots < \gamma_{s_2-c} \leq n$, and the symmetrized kernel function $K^{(c)}$ is given by

$$K^{(c)}(\mathbf{x};\mathbf{Z}_{\alpha_1},\cdots,\mathbf{Z}_{\alpha_c},\mathbf{Z}_{\beta_1},\cdots,\mathbf{Z}_{\beta_{s_2-c}},\mathbf{Z}_{\gamma_1},\cdots,\mathbf{Z}_{\gamma_{s_2-c}})$$
$$=\left\{\binom{2s_2-c}{c}\binom{2s_2-2c}{s_2-c}\right\}^{-1}\sum_{\Pi_{2s_2-c}}\Phi^*(\mathbf{x};\mathbf{Z}_{i_1},\cdots,\mathbf{Z}_{i_c},\mathbf{Z}_{i_{c+1}},\cdots,\mathbf{Z}_{i_{s_2}})$$
$$\cdot\,\Phi^*(\mathbf{x};\mathbf{Z}_{i_1},\cdots,\mathbf{Z}_{i_c},\mathbf{Z}_{i_{s_2+1}},\cdots,\mathbf{Z}_{i_{2s_2-c}}) \tag{A.44}$$

with $\sum_{\Pi_{2s_2-c}}$ standing for the summation over all the $\binom{2s_2-c}{c}\binom{2s_2-2c}{s_2-c}$ possible permutations of $(\alpha_1,\cdots,\alpha_c,\beta_1,\cdots,\beta_{s_2-c},\gamma_1,\cdots,\gamma_{s_2-c})$ that are not permuted within sets $(\alpha_1,\cdots,\alpha_c)$, $(\beta_1,\cdots,\beta_{s_2-c})$, and $(\gamma_1,\cdots,\gamma_{s_2-c})$.

From (A.41)–(A.44) above, we can further deduce that as long as $s_2=o(\sqrt{n})$, it holds that

$$n\widehat{\sigma}_J^2=\sum_{c=0}^{s_2}(cn-s_2^2)\frac{(n-s_2-1)(n-s_2-2)\cdots(n-2s_2+c+1)}{(n-2)(n-3)\cdots(n-s_2)c!}$$
$$\cdot\,[s_2(s_2-1)\cdots(s_2-c+1)]^2U_c$$
$$=-s_2^2\Big[1+O\big(\frac{s_2^2}{n}\big)\Big]U_0+s_2^2[1+O\big(\frac{s_2^2}{n}\big)]U_1+\sum_{c=2}^{s_2}O\big(\frac{s_2^2}{n}\big)^{c-1}\frac{s_2^2}{c!}U_c$$
$$=s_2^2(U_1-U_0)+O\big(\frac{s_2^4}{n}\big)(U_0+U_1)+\sum_{c=2}^{s_2}O\big(\frac{s_2^2}{n}\big)^{c-1}\frac{s_2^2}{c!}U_c,$$

which leads to

$$\frac{n}{s_2^2}\widehat{\sigma}_J^2=U_1-U_0+O\big(\frac{s_2^2}{n}\big)(U_0+U_1)+\sum_{c=2}^{s_2}O\big(\frac{s_2^2}{n}\big)^{c-1}\frac{U_c}{c!}. \tag{A.45}$$

By (A.42), for the mean we have

$$\mathbb{E}U_c=\mathbb{E}\big[\Phi^*(\mathbf{x};\mathbf{Z}_{\alpha_1},\cdots,\mathbf{Z}_{\alpha_c},\mathbf{Z}_{\beta_1},\cdots,\mathbf{Z}_{\beta_{s_2-c}})\Phi^*(\mathbf{x};\mathbf{Z}_{\alpha_1},\cdots,\mathbf{Z}_{\alpha_c},\mathbf{Z}_{\gamma_1},\cdots,\mathbf{Z}_{\gamma_{s_2-c}})\big]$$
$$=\mathbb{E}\big([\Phi_c^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_c)]^2\big), \tag{A.46}$$

where $\Phi_c^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_c)=\mathbb{E}[\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2})|\mathbf{Z}_1,\cdots,\mathbf{Z}_c]$.

As for the variance, it follows from Lemmas 2 and 3 in Sections E.1 and E.2, respectively, that for each $0 \leq c \leq s_2$ and fixed $\mathbf{x}$, we have

$$\mathrm{Var}(U_c) = O(s_2/n). \tag{A.47}$$

Moreover, in view of (A.46) and Jensen's inequality, it holds that for each $2 \leq c \leq s_2$,

$$\mathbb{E}U_c \leq \mathbb{E}[(\Phi^*)^2]. \tag{A.48}$$

Consequently, it follows from (A.45)–(A.48) that

$$\begin{aligned}
&\mathbb{E}\Big(\Big[\frac{n}{s_2^2}\widehat{\sigma}_J^2 - \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))\Big]^2\Big) \\
&\leq C\Big\{\mathrm{Var}(U_1) + \mathrm{Var}(U_0) + \frac{s_2^4}{n^2}\big[(\mathbb{E}\Phi^*)^4 + \big(\mathbb{E}[\Phi_1^*(\mathbf{x};\mathbf{Z}_1)]^2\big)^2\big] \\
&\quad + \sum_{j=2}^{s_2}\sum_{i=2}^{s_2}\big(\frac{s_2^2}{n}\big)^{i+j-2}\big[\mathrm{Var}(U_i) + (\mathbb{E}(\Phi^*)^2)^2\big]^{1/2}\big[\mathrm{Var}(U_j) + (\mathbb{E}(\Phi^*)^2)^2\big]^{1/2}\Big\},
\end{aligned} \tag{A.49}$$

where C is some positive constant. Recall the facts that $\mathbb{E}[\Phi^*] = O(1)$ and $\mathbb{E}[(\Phi^*)^2] = O(1)$, which have been shown previously in the proof of Theorem 3 in Section D.3. Combining (A.49) with these facts yields

$$\mathbb{E}\Big(\Big[\frac{n}{s_2^2}\widehat{\sigma}_J^2 - \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))\Big]\Big)^2 \leq C\Big(\frac{s_2}{n} + \frac{s_2^4}{n^2}\Big). \tag{A.50}$$

Furthermore, it has been shown in the proof of Theorem 3 in Section D.3 that

$$\mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)) \geq C s_2^{-1} \tag{A.51}$$

with C some positive constant. Thus, when $s_2 = o(n^{1/3})$, we can obtain from (A.50) and (A.51) that

$$\frac{\widehat{\sigma}_J^2}{\frac{s_2^2}{n}\mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))} \xrightarrow{p} 1. \tag{A.52}$$

In addition, it follows from (A.29) and the decomposition for the variance of the U-statistic that as long as $s_2 = o(n)$, we have

$$\frac{\sigma^2}{\frac{s_2^2}{n}\operatorname{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1))} \to 1. \tag{A.53}$$

Therefore, combining (A.52) and (A.53) results in $\widehat{\sigma}_J^2/\sigma_n^2 \xrightarrow{p} 1$, which establishes the desired consistency of the jackknife estimator $\widehat{\sigma}_J^2$. This completes the proof of Theorem 5.

D.6 Proof of Theorem 6

We now proceed with establishing the consistency of the bootstrap estimator $\widehat{\sigma}_{B,n}^2$ introduced in (26) for the variance σ_n^2 of the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ as defined in (22). Let us define the bootstrap version of the quantity σ_n^2 conditional on the given sample $\{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$ as

$$\widehat{\sigma}_n^2 = \operatorname{Var}(D_n^*(s_1, s_2)(\mathbf{x})|\mathbf{Z}_1, \cdots, \mathbf{Z}_n), \tag{A.54}$$

where $D_n^*(s_1, s_2)$ defined in (25) denotes the two-scale DNN estimator constructed as in (21) using the bootstrap sample $\{\mathbf{Z}_1^*, \cdots, \mathbf{Z}_n^*\}$. In fact, the quantity introduced in (A.54) above provides a crucial bridge. The main ingredients of the proof consist of two parts. First, we will show that the bootstrap estimator $\widehat{\sigma}_{B,n}^2$ is asymptotically close to $\widehat{\sigma}_n^2$ given in (A.54) as the number of bootstrap samples $B \to \infty$. Second, we will prove that the bootstrap version $\widehat{\sigma}_n^2$ is further asymptotically close to the population quantity σ^2 under the assumption of $s_2 = o(n^{1/3})$. It is worth mentioning that the technical analysis for the second part relies on the consistency of the jackknife estimator $\widehat{\sigma}_J^2$ established in Theorem 5.

For each $1 \le b \le B$, denote by $D_n^{(b)}(s_1, s_2)(\mathbf{x})$ the two-scale DNN estimator $D_n^*(s_1, s_2)$ constructed using the bth bootstrap sample. It is easy to see from (A.54) that for each

$1 \leq b \leq B$,

$$\mathrm{Var}(D_n^{(b)}(s_1, s_2)(\mathbf{x})|\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = \widehat{\sigma}_n^2. \tag{A.55}$$

Since the sample variance defined in (26) is an unbiased estimator for the population variance, by (A.55) it holds that

$$\mathbb{E}[\widehat{\sigma}_{B,n}^2|\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_n] = \widehat{\sigma}_n^2. \tag{A.56}$$

Thus, in view of (26) and (A.56), we can obtain

$$\mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \sigma^2)^2] = \mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \widehat{\sigma}_n^2)^2] + \mathbb{E}[(\widehat{\sigma}_n^2 - \sigma^2)^2]. \tag{A.57}$$

Without loss of generality, let us assume that $\mathbb{E}[D_n(s_1, s_2)(\mathbf{x})] = 0$ to ease our technical presentation; otherwise we can subtract the mean first.

We begin with considering the first term $\mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \widehat{\sigma}_n^2)^2]$ on the right-hand side of (A.57). Since $\{D_n^{(b)}(s_1, s_2)(\mathbf{x})\}_{1 \leq b \leq B}$ are i.i.d. random variables conditional on the given sample $\{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$, we can deduce that

$$
\begin{aligned}
&\mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \widehat{\sigma}_n^2)^2 | \mathbf{Z}_1, \cdots, \mathbf{Z}_n] \\
&= \mathbb{E}\Big[\frac{1}{(B-1)^2}\Big(\sum_{b=1}^{B}\big([D_n^{(b)}(s_1,s_2)(\mathbf{x})]^2 - \widehat{\sigma}_n^2\big) - (B\bar{D}_{B,n}^2 - \widehat{\sigma}_n^2)\Big)^2 \Big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\Big] \\
&\leq \frac{2}{(B-1)^2}\Big\{\mathbb{E}\Big[\Big(\sum_{b=1}^{B}\big([D_n^{(b)}(s_1,s_2)(\mathbf{x})]^2 - \widehat{\sigma}_n^2\big)\Big)^2 \Big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\Big] \\
&\quad + \mathbb{E}\big[(B\bar{D}_{B,n}^2 - \widehat{\sigma}_n^2)^2 | \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big]\Big\} \\
&\leq \frac{2}{(B-1)^2}\Big\{B\mathbb{E}\big[\big([D_n^{(1)}(s_1,s_2)(\mathbf{x})]^2 - \widehat{\sigma}_n^2\big)^2 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big] \\
&\quad + \frac{2B}{B^2}\mathbb{E}\big[\big([D_n^{(1)}(s_1,s_2)(\mathbf{x})]^2 - \widehat{\sigma}_n^2\big)^2 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big] \\
&\quad + \frac{4}{B^2}\sum_{1\leq i\neq j\leq B}\mathbb{E}\big[\big(D_n^{(i)}(s_1,s_2)(\mathbf{x})\big)^2\big(D_n^{(j)}(s_1,s_2)(\mathbf{x})\big)^2 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big]\Big\} \\
&\leq \frac{C}{B}\mathbb{E}\big[\big([D_n^{(1)}(s_1,s_2)(\mathbf{x})]^2 - \widehat{\sigma}_n^2\big)^2 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big] \\
&\quad + \frac{C}{B^2}\Big(\mathbb{E}\big[[D_n^{(1)}(s_1,s_2)(\mathbf{x})]^2 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big]\Big)^2 \\
&\leq \frac{C}{B}\mathbb{E}\big[[D_n^{(1)}(s_1,s_2)(\mathbf{x})]^4 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big],
\end{aligned}
\tag{A.58}
$$

where the last inequality follows from the conditional Jensen's inequality.

Let $k = [n/s_2]$ be the integer part of the number n/s_2. We define

$$
\begin{aligned}
h(\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = k^{-1}\big(&\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2}) + \Phi^*(\mathbf{x}; \mathbf{Z}_{s_2+1}, \cdots, \mathbf{Z}_{2s_2}) \\
&+ \cdots + \Phi^*(\mathbf{x}; \mathbf{Z}_{ks_2-s_2+1}, \cdots, \mathbf{Z}_{ks_2})\big).
\end{aligned}
\tag{A.59}
$$

Note that it has been shown in (2.1.15) in Korolyuk and Borovskich (1994) that

$$
\mathbb{E}\big[[D_n^{(1)}(s_1,s_2)(\mathbf{x})]^4 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big] \leq \mathbb{E}[h^4(\mathbf{Z}_1^*, \cdots, \mathbf{Z}_n^*) | \mathbf{Z}_1, \cdots, \mathbf{Z}_n] \tag{A.60}
$$

with the functional $h(\cdot)$ given in (A.59). Moreover, with an application of Rosenthal's inequality for independent random variables, we can obtain that

$$\mathbb{E}[h^4(\mathbf{Z}_1^*, \cdots, \mathbf{Z}_n^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n]$$
$$\leq Ck^{-4}k^2\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4\big|\mathbf{Z}_1, \cdots, \mathbf{Z}_n\big), \tag{A.61}$$

where C is some positive constant. Then in light of (A.61), it remains to bound the quantity $\mathbb{E}([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4)$, which has been dealt with in Lemma 4 in Section E.3. Thus, it follows from (A.58), (A.60)–(A.61), and Lemma 4 that

$$\mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \widehat{\sigma}_n^2)^2] \leq \frac{C}{B}\frac{s_2^2}{n^2}n^{-s_2}\sum_{i_1=1}^{n}\cdots\sum_{i_{s_2}=1}^{n}\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}})]^4\big)$$
$$\leq \frac{CMs_2^2}{Bn^2}, \tag{A.62}$$

where M is some positive constant given in Lemma 4.

We next proceed with analyzing the second term $\mathbb{E}[(\widehat{\sigma}_n^2 - \sigma^2)^2]$ on the right-hand side of (A.57). Recall the definition of the bootstrap version $\widehat{\sigma}_n^2$ for the population quantity σ^2 introduced in (A.54). Let us define

$$m_n = \mathbb{E}[\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n] \tag{A.63}$$

and

$$h_1(\mathbf{z}) = \mathbb{E}[\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*) - m_n|\mathbf{Z}_1^* = \mathbf{z}]. \tag{A.64}$$

Then applying similar arguments as for (A.20) in the proof of Theorem 2 in Section D.2, we can deduce that

$$\widehat{\sigma}_n^2 = \frac{s_2^2}{n}\mathbb{E}[h_1^2(\mathbf{Z}_1^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n] + \Delta_1, \tag{A.65}$$

where $0 \leq \Delta_1 \leq \frac{s_2^2}{n^2}\,\mathrm{Var}(\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$ and function $h_1(\cdot)$ is given in (A.64) and (A.63). Similarly, it holds that

$$\sigma_n^2 = \frac{s_2^2}{n}\mathbb{E}[g_1^2(\mathbf{Z}_1)] + \Delta_2, \tag{A.66}$$

where $g_1(\mathbf{Z}_1) = \mathbb{E}[\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_n)|\mathbf{Z}_1]$ and $0 \leq \Delta_2 \leq \frac{s_2^2}{n^2} \operatorname{Var}(\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2}))$. Hence, by (A.65) and (A.66) we can obtain that

$$\begin{aligned}
&\mathbb{E}[(\widehat{\sigma}_n^2 - \sigma_n^2)^2] \\
&\leq C\mathbb{E}\Big(\frac{s_2^4}{n^2}\big[\mathbb{E}[h_1^2(\mathbf{Z}_1^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n] - \mathbb{E}[g_1^2(\mathbf{Z}_1)]\big]^2 + \Delta_1^2 + \Delta_2^2\Big),
\end{aligned} \tag{A.67}$$

where C is some positive constant.

Observe that

$$\Delta_2^2 = O(\frac{s_2^4}{n^4}) \tag{A.68}$$

and

$$\begin{aligned}
\mathbb{E}(\Delta_1^2) &\leq \frac{s_2^4}{n^4}\mathbb{E}\big[\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^2|\mathbf{Z}_1, \cdots, \mathbf{Z}_n\big)\big]^2 \\
&\leq \frac{s_2^4}{n^4}\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4\big) \\
&\leq \frac{Ms_2^4}{n^4},
\end{aligned} \tag{A.69}$$

where the last inequality follows from Lemma 4 with M some positive constant. In addition, it holds that

$$\mathbb{E}[h_1^2(\mathbf{Z}_1^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n] = \frac{1}{n}\sum_{i=1}^n h_1^2(\mathbf{Z}_i) \tag{A.70}$$

and

$$\begin{aligned}
h_1(\mathbf{Z}_i) = n^{-s_2+1}&\sum_{i_2=1}^n \cdots \sum_{i_{s_2}=1}^n \Phi^*(\mathbf{x}; \mathbf{Z}_i, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_{s_2}}) \\
&- n^{-s_2}\sum_{i_1=1}^n \cdots \sum_{i_{s_2}=1}^n \Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}}).
\end{aligned} \tag{A.71}$$

Let us further define

$$S_i = \binom{n-1}{s_2-1}^{-1} \sum_{\substack{1\leq j_1<j_2<\cdots<j_{s_2-1}\leq n \\ j_1, j_2, \cdots, j_{s_2-1}\neq i}} \Phi^*(\mathbf{x}; \mathbf{Z}_i, \mathbf{Z}_{j_1}, \cdots, \mathbf{Z}_{j_{s_2-1}}). \tag{A.72}$$

From the equality $\binom{n-1}{s_2}U_{n-1}^{(i)} + \binom{n-1}{s_2-1}S_i = \binom{n}{s_2}D_n(s_1, s_2)(\mathbf{x})$ in view of (A.72), it is easy to see that the jackknife estimator $\widehat{\sigma}_J^2$ introduced in (24) satisfies that

$$\frac{n\widehat{\sigma}_J^2}{s_2^2} = \frac{n-1}{(n-s_2)^2}\sum_{i=1}^{n}\big(S_i - D_n(s_1, s_2)(\mathbf{x})\big)^2. \tag{A.73}$$

Then the main idea of the remaining proof is to show that under the assumption of $s_2 = o(n^{1/3})$, $h_1(\mathbf{Z}_i)$ is asymptotically close to $S_i - D_n(s_1, s_2)(\mathbf{x})$ and thus $\mathbb{E}[h_1^2(\mathbf{Z}_1^*)|\mathbf{Z}_1, \cdots, \mathbf{Z}_n]$ is asymptotically close to $\frac{n\widehat{\sigma}_J^2}{s_2^2}$. Observe that

$$n^{-s_2+1}\binom{n-1}{s_2-1}(s_2-1)! = 1 + O(s_2^2/n)$$

and

$$n^{-s_2}\binom{n}{s_2}s_2! = 1 + O(s_2^2/n),$$

which entail that

$$\Big(n^{s_2-1} - \binom{n-1}{s_2-1}(s_2-1)!\Big)n^{-s_2+1} = O(s_2^2/n)$$

and

$$\Big(n^{s_2} - \binom{n}{s_2}s_2!\Big)n^{-s_2} = O(s_2^2/n).$$

Thus, it follows from (A.71) and these facts that

$$\begin{aligned} h_1(\mathbf{Z}_i) &= (1 + O(s_2^2/n))\big[S_i - D_n(s_1, s_2)(\mathbf{x})\big] + n^{-s_2+1}\sum_{\mathscr{D}_1}\Phi^*(\mathbf{x}; \mathbf{Z}_i, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_{s_2}}) \\ &\quad - n^{-s_2}\sum_{\mathscr{D}_2}\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_{s_2}}), \end{aligned} \tag{A.74}$$

where $\mathscr{D}_1 = \{(i_2, \cdots, i_{s_2})$: there is at least one pair that are equal or there is a component that is equal to $i\}$ and $\mathscr{D}_2 = \{(i_1, \cdots, i_{s_2})$: there is at least one pair of components that are equal$\}$.

With an application of similar arguments as in the proof of Lemma 4 in Section E.3, we can obtain that

$$\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_{s_2}})]^4\big) \le M \tag{A.75}$$

with M some positive constant, regardless of how many components of $(i_1, i_2, \cdots, i_{s_2})$ are equal. As a consequence, by (A.75) it holds that

$$\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n}\Big(n^{-s_2+1}\sum_{\mathscr{D}_1}\Phi^*(\mathbf{x};\mathbf{Z}_i,\mathbf{Z}_{i_2},\cdots,\mathbf{Z}_{i_{s_2}})\Big)^2\Big)^2\Big]$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\Big(n^{-s_2+1}\sum_{\mathscr{D}_1}\Phi^*(\mathbf{x};\mathbf{Z}_i,\mathbf{Z}_{i_2},\cdots,\mathbf{Z}_{i_{s_2}})\Big)^4\Big] \leq \frac{CMs_2^8}{n^4} \tag{A.76}$$

and similarly,

$$\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n}\Big(n^{-s_2}\sum_{\mathscr{D}_2}\Phi^*(\mathbf{x};\mathbf{Z}_{i_1},\mathbf{Z}_{i_2},\cdots,\mathbf{Z}_{i_{s_2}})\Big)\Big)^2\Big] \leq \frac{CMs_2^8}{n^4}, \tag{A.77}$$

where C represents some positive constant whose value may change from line to line. Hence, combining (A.70), (A.74), and (A.76)–(A.77), we can deduce that as long as $s_2 = o(n^{1/3})$, it holds that

$$\mathbb{E}\Big(\big[\mathbb{E}[h_1^2(\mathbf{Z}_1^*)|\mathbf{Z}_1,\cdots,\mathbf{Z}_n] - \mathbb{E}[g_1^2(\mathbf{Z}_1)]\big]^2\Big)$$
$$\leq C\mathbb{E}\Big[\Big(\frac{1}{n}\sum_{i=1}^{n}(1+O(s_2^2/n))^2[S_i - D_n(s_1,s_2)(\mathbf{x})]^2 - \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))\Big)^2\Big]$$
$$+\frac{CMs_2^8}{n^4}$$
$$\leq C\mathbb{E}\Big[\Big(\frac{(n-s_2)^2}{n(n-1)}(1+O(s_2^2/n))\frac{n}{s_2^2}\widehat{\sigma}_J^2 - \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))\Big)^2\Big] + \frac{CMs_2^8}{n^4}$$
$$\leq C(1+O(s_2^2/n))\mathbb{E}\big([\frac{n}{s_2^2}\widehat{\sigma}_J^2 - \mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))]^2\big) + \frac{Cs_2^4}{n^2}(\mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)))^2$$
$$+\frac{CMs_2^8}{n^4}$$
$$\leq \frac{Cs_2}{n} + \frac{s_2^2}{n^2} + \frac{CMs_2^8}{n^4} \leq \frac{C(M+1)s_2}{n}, \tag{A.78}$$

where the second to the last inequality comes from (A.50) and (A.146) in the proof of Lemma 10 in Section E.9.

Substituting the above bounds in (A.68)–(A.69) and (A.78) into (A.67) leads to

$$\mathbb{E}[(\widehat{\sigma}_n^2 - \sigma_n^2)^2] = O\Big(\frac{s_2^5}{n^3} + \frac{s_2^4}{n^4}\Big). \tag{A.79}$$

Thus, combining (A.62) and (A.79), we can obtain that

$$\mathbb{E}[(\widehat{\sigma}_{B,n}^2 - \sigma_n^2)^2] = O\Big(\frac{s_2^5}{n^3} + \frac{s_2^2}{Bn^2}\Big). \tag{A.80}$$

Recall the fact that $\sigma^2 = O(\frac{s_2}{n})$ under the assumption of $s_2 = o(n)$. Consequently, such fact along with (A.80) entails that

$$\mathbb{E}\Big[\Big(\frac{\widehat{\sigma}_{B,n}^2}{\sigma_n^2} - 1\Big)^2\Big] = O(\frac{s_2^3}{n} + \frac{1}{B}). \tag{A.81}$$

Therefore, combining (A.81) and the assumptions of $s_2 = o(n^{1/3})$ and $B \to \infty$ yields $\widehat{\sigma}_{B,n}^2/\sigma_n^2 \xrightarrow{p} 1$, which establishes the desired consistency of the bootstrap estimator $\widehat{\sigma}_{B,n}^2$. This concludes the proof of Theorem 6.

D.7 Proof of Theorem 7

The main idea of the proof is to show that both the TDNN estimator $D_n(s_1, s_2)(\mathbf{x})$ and its bootstrap version $D_n^*(s_1, s_2)(\mathbf{x})$ are asymptotically normal and in addition, their asymptotic variances are close to each other. Then the conditional distribution of $D_n^*(s_1, s_2)(\mathbf{x})$ given $(\mathbf{Z}_1, \cdots, \mathbf{Z}_n)$ approaches the distribution of $D_n(s_1, s_2)(\mathbf{x})$ as the sample size n increases. To this end, let us first recall that it has been shown in Theorem 3 that $D_n(s_1, s_2)(\mathbf{x})$ is asymptotically normal. Since the normal distribution $\Phi(\cdot)$ is continuous, it follows that

$$\sup_{u \in \mathbb{R}} \Big|\mathbb{P}(\sigma_n^{-1}(D_n(s_1, s_2)(\mathbf{x}) - \mu(\mathbf{x}) - \Lambda) \le u) - \Phi(u)\Big| = o(1), \tag{A.82}$$

where $\sigma_n^2 = \mathrm{Var}(D_n(s_1, s_2)(\mathbf{x}))$.

We next deal with the bootstrapped statistic $D_n^*(s_1, s_2)(\mathbf{x})$. In light of Hoeffding's decomposition for the U-statistic, we have

$$D_n^*(s_1, s_2)(\mathbf{x}) - \theta^* = \frac{s_2}{n} \sum_{i=1}^n \widehat{g}_1(\mathbf{x}; \mathbf{Z}_i^*) + R_n^*,$$

where $\widehat{g}_1^*(\mathbf{x};\mathbf{z}) = \mathbb{E}^*\big[\Phi^*(\mathbf{x};\mathbf{Z}_1^*,\cdots,\mathbf{Z}_{s_2}^*)|\mathbf{Z}_1^* = \mathbf{z}\big] - \theta^*$ with the expectation $\mathbb{E}^*$ taken with respect to the bootstrap resampling distribution of $(\mathbf{Z}_2^*,\cdots,\mathbf{Z}_n^*)$ given $(\mathbf{Z}_1,\cdots,\mathbf{Z}_n)$, and R_n^* is the higher-order remainder. Given $(\mathbf{Z}_1,\cdots,\mathbf{Z}_n)$, the Berry–Esseen theorem for the sum of i.i.d. random variables (Berry, 1941) leads to

$$\begin{aligned}&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}^*\Big([n\operatorname{Var}^*(\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*))]^{-1/2}\sum_{i=1}^n\widehat{g}_1(\mathbf{x};\mathbf{Z}_i^*)\le u\Big)-\Phi(u)\Big|\\&\quad\le\frac{\mathbb{E}^*(|\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*)|^3)}{\sqrt{n}\operatorname{Var}^*(\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*))},\end{aligned}$$

where the variance Var^* is again taken with respect to the bootstrap resampling distribution given $(\mathbf{Z}_1,\cdots,\mathbf{Z}_n)$.

An application of similar arguments as in the proof of Lemma 10 yields

$$\operatorname{Var}^*(\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*))\sim O_p(s_2^{-1}).$$

It follows from Jensen's inequality that

$$\mathbb{E}^*(|\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*)|^3)\le\mathbb{E}^*(|\Phi^*(\mathbf{x};\mathbf{Z}_1^*,\cdots,\mathbf{Z}_{s_2}^*)|^3).$$

Similar to Lemma 9, we can deduce that

$$\mathbb{E}(|\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2})|^3)\le M$$

for some positive constant M. Hence, it holds that $\mathbb{E}^*(|\Phi^*(\mathbf{x};\mathbf{Z}_1^*,\cdots,\mathbf{Z}_{s_2}^*)|^3)=O_p(1)$ and $\mathbb{E}^*(|\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*)|^3)=O_p(1)$. Consequently, the approximation error satisfies that

$$\begin{aligned}&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}^*\Big([n\operatorname{Var}^*(\widehat{g}_1(\mathbf{x};\mathbf{Z}_1^*))]^{-1/2}\sum_{i=1}^n\widehat{g}_1(\mathbf{x};\mathbf{Z}_i^*)\le u\Big)-\Phi(u)\Big|\\&\quad=O_p(s_2/\sqrt{n})=o_p(1)\end{aligned}\tag{A.83}$$

since $s_2=o(n^{1/3})$.

Let us define $\widehat{\sigma}_n^2 = \mathrm{Var}[D_n^*(s_1, s_2)(\mathbf{x})|\mathbf{Z}_1, \cdots, \mathbf{Z}_n]$. Then the variance decomposition of the U-statistic implies that

$$\widehat{\sigma}_n^2 = \frac{s_2^2}{n} \mathrm{Var}^*(\widehat{g}_1(\mathbf{x}; \mathbf{Z}_1^*)) + \mathrm{Var}^*(R_n^*).$$

Note that from the similar argument as in (A.20), we see that the remainder R_n^* above satisfies that

$$\mathrm{Var}^*(R_n^*) \leq \frac{s_2^2}{n^2} \mathrm{Var}^*(\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)).$$

Since it has been shown in Section E.8 that the second moment $\mathbb{E}\{[\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]^2\} \leq M$ for some positive constant M, we have

$$\mathrm{Var}^*(\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)) = O_p(1),$$

and thus $\mathrm{Var}^*(R_n^*) = O_p(s_2^2/n^2)$. Furthermore, it follows from $\mathrm{Var}^*(\widehat{g}_1(\mathbf{x}; \mathbf{Z}_1^*)) \sim O_p(s_2^{-1})$ that

$$\frac{s_2^2}{n} \mathrm{Var}^*(\widehat{g}_1(\mathbf{x}; \mathbf{Z}_1^*))/\widehat{\sigma}_n^2 \xrightarrow{p} 1.$$

Hence, (A.83) entails that

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}^*\Big(\widehat{\sigma}_n^{-1}(D_n^*(s_1, s_2)(\mathbf{x}) - \theta^*) \leq u\Big) - \Phi(u) \right| = o_p(1). \tag{A.84}$$

Moreover, we have shown in Section D.3 that σ_n is of order $(s_2/n)^{1/2}$ and in Section D.6 that $\widehat{\sigma}_n/\sigma_n \xrightarrow{p} 1$. Therefore, combining (A.82) and (A.83) results in

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}^*\Big(\sigma_n^{-1}(D_n^*(s_1, s_2)(\mathbf{x}) - \theta^*) \leq u\Big) - \mathbb{P}\Big(\sigma_n^{-1}(D_n(s_1, s_2)(\mathbf{x}) - \mu(\mathbf{x}) - \Lambda)\Big) \right|$$
$$= o_p(1).$$

Since σ_n is of order $(s_2/n)^{1/2}$ and unknown in practice, we can rewrite the above approximation error as

$$\sup_{u \in \mathbb{R}} \left| \mathbb{P}^*\Big((s_2/n)^{-1/2}(D_n^*(s_1, s_2)(\mathbf{x}) - \theta^*) \leq u\Big) - \mathbb{P}((s_2/n)^{-1/2}(D_n(s_1, s_2)(\mathbf{x}) - \mu(\mathbf{x}) - \Lambda)) \right|$$
$$= o_p(1).$$

This completes the proof of Theorem 7.

D.8 Proof of Theorem 8

We now aim to prove the asymptotic normality of the HTE estimator

$$\widehat{\tau}(\mathbf{x}) = D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})$$

introduced in (A.6), where $D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x})$ and $D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})$ denote the two-scale DNN estimators constructed using the treatment sample of size n_1 and the control sample of size n_0, respectively. Denote by $n = n_0 + n_1$ the total sample size. By the assumption $P(T = 1|\mathbf{X}, Y_{T=0}, Y_{T=1}) = 1/2$, it is easy to see that $n_0/n_1 \xrightarrow{p} 1$ as $n \to \infty$. For each of the treatment and control groups in the randomized experiment, by the assumptions a separate application of Theorem 3 shows that there exist some positive numbers σ_{n_1} of order $(s_2^{(1)}/n_1)^{1/2}$ and σ_{n_0} of order $(s_2^{(0)}/n_0)^{1/2}$ such that

$$\frac{D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - \mathbb{E}[D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x})]}{\sigma_{n_1}} \xrightarrow{\mathscr{D}} N(0,1) \tag{A.85}$$

and

$$\frac{D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x}) - \mathbb{E}[D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})]}{\sigma_{n_0}} \xrightarrow{\mathscr{D}} N(0,1). \tag{A.86}$$

In view of the randomized experiment assumption, the treatment sample and control sample are independent of each other, which entails that the two separate two-scale DNN estimators $D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x})$ and $D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})$ are independent. Thus it follows from (A.85) and (A.86) that

$$\begin{aligned}&\frac{D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x}) - \mathbb{E}[D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x}) - D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})]}{\sigma_n} \\ &\quad \xrightarrow{\mathscr{D}} N(0,1), \end{aligned} \tag{A.87}$$

where we define $\sigma_n = (\sigma_{n_1}^2 + \sigma_{n_0}^2)^{1/2}$. Moreover, from the higher-order asymptotic expansion of the bias term in Theorem 1 applied to the potential treatment and control responses, respectively, and the definition of the heterogeneous treatment effect (HTE) $\tau(\mathbf{x})$ introduced in (A.5), we see that

$$\mathbb{E}[D_{n_1}^{(1)}(s_1^{(1)}, s_2^{(1)})(\mathbf{x})] - \mathbb{E}[D_{n_0}^{(0)}(s_1^{(0)}, s_2^{(0)})(\mathbf{x})] = \tau(\mathbf{x}) + \Lambda, \tag{A.88}$$

where $\Lambda = O\{(s_1^{(1)})^{-4/d} + (s_2^{(1)})^{-4/d} + (s_1^{(0)})^{-4/d} + (s_2^{(0)})^{-4/d}\}$ for $d \geq 2$ and $\Lambda = O\{(s_1^{(1)})^{-3} + (s_2^{(1)})^{-3} + (s_1^{(0)})^{-3} + (s_2^{(0)})^{-3}\}$ for $d = 1$. Therefore, combining (A.87) and (A.88) yields the desired asymptotic normality of the HTE estimator $\widehat{\tau}(\mathbf{x})$ based on the two-scale DNN estimators. This concludes the proof of Theorem 8.

E Some key lemmas and their proofs

E.1 Lemma 2 and its proof

Lemma 2. *Under the conditions of Theorem 5, we have that for each $0 \leq c \leq s_2$ and fixed $\mathbf{x}$,*

$$\mathrm{Var}(U_c) \leq \frac{2s_2 - c}{n} \mathrm{Var}(K^{(c)}), \tag{A.89}$$

where U_c is the U-statistic defined in (A.42) and $K^{(c)}$ is the symmetrized kernel function given in (A.44).

Proof. For notational simplicity, we will drop the dependence of all the functionals on the fixed vector $\mathbf{x}$ whenever there is no confusion. For each $1 \leq j \leq 2s_2 - c$, let us define

$$K_j^{(c)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_j) = \mathbb{E}[K^{(c)}|\mathbf{Z}_1, \cdots, \mathbf{Z}_j],$$
$$g_j^{(c)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_j) = K_j^{(c)} - \mathbb{E}[K^{(c)}] - \sum_{i=1}^{j-1} \sum_{1 \leq \alpha_1 < \cdots < \alpha_i \leq j} g_i^{(c)}(\mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_i}),$$

and $V_j = \mathrm{Var}(g_j^{(c)}(\mathbf{Z}_1, \cdots, \mathbf{Z}_j))$. Then it follows from Hoeffding's decomposition that

$$U_c = \mathbb{E}[K^{(c)}] + \binom{n}{2s_2 - c}^{-1} \sum_{i=1}^{2s_2 - c} \binom{n - i}{2s_2 - c - i} \sum_{1 \leq \alpha_1 < \cdots < \alpha_i \leq n} g_i^{(c)}(\mathbf{Z}_{\alpha_1}, \cdots, \mathbf{Z}_{\alpha_i}). \tag{A.90}$$

Observe that $\mathrm{Var}(K^{(c)}) = \sum_{i=1}^{2s_2-c} \binom{2s_2-c}{i} V_i$. Thus, in view of (A.90), we can deduce that

$$\begin{aligned}
\mathrm{Var}(U_c) &= \sum_{i=1}^{2s_2-c} \binom{n}{2s_2-c}^{-2} \binom{n-i}{2s_2-c-i}^2 \binom{n}{i} V_j \\
&= \sum_{i=1}^{2s_2-c} \frac{(2s_2-c)!(n-i)!}{n!(2s_2-c-i)!} \binom{2s_2-c}{i} V_i \\
&\le \frac{2s_2-c}{n} \sum_{i=1}^{2s_2-c} \binom{2s_2-c}{i} V_i \\
&= \frac{2s_2-c}{n} \mathrm{Var}(K^{(c)}),
\end{aligned}$$

which establishes the desired upper bound in (A.89). This completes the proof of Lemma 2.

E.2 Lemma 3 and its proof

Lemma 3. *Under the conditions of Theorem 5, it holds that for each* $0 \le c \le s_2$ *and fixed* $\mathbf{x}$*,*

$$\mathrm{Var}(K^{(c)}) \le C[(w_1^*)^4 + (w_2^*)^4]\big(\mu^4(\mathbf{x}) + 6\mu^2(\mathbf{x})\sigma_\epsilon + 4\mu(\mathbf{x}) + \mathbb{E}[\epsilon_1^4]\big), \tag{A.91}$$

where $K^{(c)}$ *is the symmetrized kernel function given in (A.44) and* C *is some positive constant.*

Proof. By the Cauchy–Schwarz inequality, we can deduce that

$$\begin{aligned}
&\mathrm{Var}(K^{(c)}) \leq \mathbb{E}[(K^{(c)})^2] \\
&= \left[\binom{2s_2-c}{c}\binom{2s_2-2c}{s_2-c}\right]^{-2} \sum_{\Pi_{2s_2-c}} \sum_{\Pi_{2s_2-c}} \\
&\quad \mathbb{E}\big\{\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_c}, \mathbf{Z}_{i_{c+1}}, \cdots, \mathbf{Z}_{i_{s_2}})\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_c}, \mathbf{Z}_{i_{s_2+1}}, \cdots, \mathbf{Z}_{i_{2s_2-c}}) \\
&\quad \times \Phi^*(\mathbf{x}; \mathbf{Z}_{j_1}, \cdots, \mathbf{Z}_{j_c}, \mathbf{Z}_{j_{c+1}}, \cdots, \mathbf{Z}_{j_{s_2}})\Phi^*(\mathbf{x}; \mathbf{Z}_{j_1}, \cdots, \mathbf{Z}_{j_c}, \mathbf{Z}_{j_{s_2+1}}, \cdots, \mathbf{Z}_{j_{2s_2-c}})\big\} \\
&\leq \mathbb{E}\big\{\big[\Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2})\big]^4\big\}, \qquad \text{(A.92)}
\end{aligned}$$

where $\sum_{\Pi_{2s_2-c}}$ denotes the summation introduced in (A.44). In light of the definition of Φ^* in (A.135), we have

$$\begin{aligned}
\mathbb{E}\big\{\big[\Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2})\big]^4\big\} &\leq 8(w_1^*)^4\mathbb{E}[\Phi^4(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] \\
&\quad + 8(w_2^*)^4\mathbb{E}[\Phi^4(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]. \qquad \text{(A.93)}
\end{aligned}$$

Let us make some useful observations. Note that

$$\begin{aligned}
\mathbb{E}[\Phi^4(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] &= \mathbb{E}\Big[\Big(\sum_{i=1}^n y_i\zeta_{i,s_1}\Big)^4\Big] \\
&= \sum_{i=1} \mathbb{E}[y_i^4\zeta_{i,s_1}] = s_1\mathbb{E}[y_1^4\zeta_{1,s_1}]
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[y_1^4\zeta_{1,s_1}] &= \mathbb{E}\big([\mu(\mathbf{X}_1) + \epsilon_1]^4\zeta_{1,s_1}\big) \\
&= \mathbb{E}[\mu^4(\mathbf{X}_1)\zeta_{1,s_1}] + 6\mathbb{E}[\mu^2(\mathbf{X}_1)\zeta_{1,s_1}]\sigma_\epsilon^2 + 4\mathbb{E}[\mu(\mathbf{X}_1)\zeta_{1,s_1}] + \mathbb{E}[\epsilon_1^4],
\end{aligned}$$

where $\zeta_{i,s}$ represents the indicator function for the event that $\mathbf{X}_i$ is the 1NN of $\mathbf{x}$ among $\mathbf{X}_1, \cdots, \mathbf{X}_s$. Moreover, it follows from Lemma 13 in Section F.3 that as $s_1 \to \infty$,

$$s_1\mathbb{E}[\mu^k(\mathbf{X}_1)\zeta_{1,s_1}] \to \mu^k(\mathbf{x})$$

for $k = 1, 2, 4$. Hence, it holds that

$$\mathbb{E}[\Phi^4(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] = s_1 \mathbb{E}[y_1^4 \zeta_{1,s_1}]$$
$$\to \mu^4(\mathbf{x}) + 6\mu^2(\mathbf{x})\sigma_\epsilon + 4\mu(\mathbf{x}) + \mathbb{E}[\epsilon_1^4]$$

as $s_1 \to \infty$.

Using similar arguments, we can show that as $s_2 \to \infty$,

$$\mathbb{E}[\Phi^4(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})] \to \mu^4(\mathbf{x}) + 6\mu^2(\mathbf{x})\sigma_\epsilon + 4\mu(\mathbf{x}) + \mathbb{E}[\epsilon_1^4].$$

Therefore, combining the asymptotic limits obtained above, (A.92), and (A.93) results in

$$\mathrm{Var}(K^{(c)}) \leq C[(w_1^*)^4 + (w_2^*)^4]\big(\mu^4(\mathbf{x}) + 6\mu^2(\mathbf{x})\sigma_\epsilon + 4\mu(\mathbf{x}) + \mathbb{E}[\epsilon_1^4]\big),$$

where C is some positive constant. This concludes the proof of Lemma 3.

E.3 Lemma 4 and its proof

Lemma 4. *Under the conditions of Theorem 6, there exists some constant $M > 0$ depending upon w_1^*, w_2^*, $\mathbf{x}$, and the distribution of ϵ such that*

$$\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4\big) \leq M. \tag{A.94}$$

Proof. Since the observations in the bootstrap sample $\{\mathbf{Z}_1^*, \cdots, \mathbf{Z}_n^*\}$ are selected independently and uniformly from the original sample $\{\mathbf{Z}_1, \cdots, \mathbf{Z}_n\}$, we have

$$\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4\big) = \mathbb{E}\Big(\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1^*, \cdots, \mathbf{Z}_{s_2}^*)]^4 \big| \mathbf{Z}_1, \cdots, \mathbf{Z}_n\big)\Big)$$
$$= n^{-s_2} \sum_{i_1=1}^{n} \cdots \sum_{i_{s_2}=1}^{n} \mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}})]^4\big).$$

Observe that for distinct $i_1, \cdots, i_{s_2}$, we have shown in the proof of Lemma 3 in Section E.2 that as $s_2 \to \infty$,

$$\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]^4\big) \to A$$

for some positive constant A that depends upon w_1^*, w_2^*, $\mathbf{x}$, and the distribution of ϵ.

Furthermore, note that if $i_1 = i_2 = \cdots = i_c$ and the remaining arguments are distinct, then it holds that

$$\Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}}) = \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_{c+1}}, \cdots, \mathbf{Z}_{i_{s_2}}).$$

Therefore, there exists some positive constant M depending upon w_1^*, w_2^*, $\mathbf{x}$, and the distribution of ϵ such that

$$\mathbb{E}\big([\Phi^*(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}})]^4\big) \leq M$$

for any $1 \leq i_1 \leq n, \cdots, 1 \leq i_{s_2} \leq n$. This completes the proof of Lemma 4.

E.4 Lemma 5 and its proof

In Lemma 5 below, we will provide the asymptotic expansion of $\mathbb{E}\, \|\mathbf{X}_{(1)} - \mathbf{x}\|^k$ with $k \geq 1$ and its higher-order asymptotic expansion for the case of $k = 2$ as the sample size $n \to \infty$.

Lemma 5. *Assume that Conditions 1–3 hold and $\mathbf{x} \in \operatorname{supp}(\mathbf{X}) \subset \mathbb{R}^d$ is fixed. Then the 1-nearest neighbor (1NN) $\mathbf{X}_{(1)}$ of $\mathbf{x}$ in the i.i.d. sample $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ satisfies that for any $k \geq 1$,*

$$\mathbb{E}\, \|\mathbf{X}_{(1)} - \mathbf{x}\|^k = \frac{\Gamma(k/d+1)}{(f(\mathbf{x})V_d)^{k/d}} n^{-k/d} + o(n^{-k/d}) \tag{A.95}$$

as $n \to \infty$, where $\Gamma(\cdot)$ is the gamma function and $V_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$. In particular, when $k = 2$, there are three cases. If $d = 1$, we have

$$\mathbb{E}\, \|\mathbf{X}_{(1)} - \mathbf{x}\|^2 = \frac{\Gamma(2/d+1)}{(f(\mathbf{x})V_d)^{2/d}} n^{-2/d} - \left(\frac{\Gamma(2/d+2)}{d(f(\mathbf{x})V_d)^{2/d}}\right) n^{-(1+2/d)} + o(n^{-(1+2/d)}). \tag{A.96}$$

If $d = 2$, we have

$$\mathbb{E}\, \|\mathbf{X}_{(1)} - \mathbf{x}\|^2 = \frac{\Gamma(2/d+1)}{(f(\mathbf{x})V_d)^{2/d}} n^{-2/d} - \left(\frac{\operatorname{tr}(f''(\mathbf{x}))\Gamma(4/d+1)}{f(\mathbf{x})(f(\mathbf{x})V_d)^{4/d} d(d+2)} + \frac{\Gamma(2/d+2)}{d(f(\mathbf{x})V_d)^{2/d}}\right) n^{-4/d} + o(n^{-4/d}), \tag{A.97}$$

where $f''(\cdot)$ stands for the Hessian matrix of the density function $f(\cdot)$. If $d \geq 3$, we have

$$\mathbb{E}\,\|\mathbf{X}_{(1)} - \mathbf{x}\|^2 = \frac{\Gamma(2/d+1)}{(f(\mathbf{x})V_d)^{2/d}} n^{-2/d} - \left(\frac{\mathrm{tr}(f''(\mathbf{x}))\Gamma(4/d+1)}{f(\mathbf{x})(f(\mathbf{x})V_d)^{4/d} d(d+2)}\right) n^{-4/d} + o(n^{-4/d}). \tag{A.98}$$

Proof. Denote by φ the probability measure on $\mathbb{R}^d$ given by random vector $\mathbf{X}$. We begin with obtaining an approximation of $\varphi(B(\mathbf{x}, r))$, where $B(\mathbf{x}, r)$ represents a ball in the Euclidean space $\mathbb{R}^d$ with center $\mathbf{x}$ and radius $r > 0$. Recall that by Condition 2, the density function $f(\cdot)$ of measure φ with respect to the Lebesgue measure λ is four times continuously differentiable with bounded corresponding derivatives in a neighborhood of $\mathbf{x}$. Then using the Taylor expansion, we see that for any $\boldsymbol{\xi} \in S^{d-1}$ and $0 < \rho < r$,

$$f(\mathbf{x} + \rho\boldsymbol{\xi}) = f(\mathbf{x}) + f'(\mathbf{x})^T \boldsymbol{\xi}\rho + \frac{1}{2}\boldsymbol{\xi}^T f''(\mathbf{x})\boldsymbol{\xi}\rho^2 + o(\rho^2), \tag{A.99}$$

where S^{d-1} denotes the unit sphere in $\mathbb{R}^d$, and $f'(\cdot)$ and $f''(\cdot)$ stand for the gradient vector and the Hessian matrix, respectively, of the density function $f(\cdot)$. With the aid of the representation in (A.99), an application of the spherical integration leads to

$$\begin{aligned}
\varphi(B(\mathbf{x}, r)) &= \int_0^r \int_{S^{d-1}} f(\mathbf{x} + \rho\boldsymbol{\xi})\rho^{d-1}\nu(\mathrm{d}\boldsymbol{\xi})\,\mathrm{d}\rho \\
&= \int_0^r \int_{S^{d-1}} \left(f(\mathbf{x}) + f'(\mathbf{x})^T\boldsymbol{\xi}\rho + \frac{1}{2}\boldsymbol{\xi}^T f''(\mathbf{x})\boldsymbol{\xi}\rho^2 + o(\rho^2)\right)\rho^{d-1}\nu(\mathrm{d}\boldsymbol{\xi})\,\mathrm{d}\rho \\
&= \int_0^r \left[f(\mathbf{x}) d V_d \rho^{d-1} + \frac{\mathrm{tr}(f''(\mathbf{x}))V_d}{2}\rho^{d+1} + o(\rho^{d+1})\right]\mathrm{d}\rho \\
&= f(\mathbf{x})V_d r^d + \frac{\mathrm{tr}(f''(\mathbf{x}))V_d}{2(d+2)} r^{d+2} + o(r^{d+2}),
\end{aligned} \tag{A.100}$$

where ν denotes a measure constructed on the unit sphere $\mathbb{S}^{d-1}$ as characterized in Lemma 11 in Section F.1 and $\mathrm{d}\cdot$ stands for the differential of a given variable hereafter.

We now turn our attention to the target quantity $\mathbb{E}\,\|\mathbf{X}_{(1)} - \mathbf{x}\|^k$ for any $k \geq 1$. It holds that

$$\begin{aligned}
\mathbb{E}\,\|\mathbf{X}_{(1)}-\mathbf{x}\|^k &= \int_0^\infty \mathbb{P}\left(\|\mathbf{X}_{(1)}-\mathbf{x}\|^k > t\right)\,\mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\left(\|\mathbf{X}_{(1)}-\mathbf{x}\| > t^{1/k}\right)\,\mathrm{d}t \\
&= \int_0^\infty [1-\varphi(B(\mathbf{x},t^{1/k}))]^n\,\mathrm{d}t \\
&= n^{-k/d}\int_0^\infty \left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n\,\mathrm{d}t. \qquad \text{(A.101)}
\end{aligned}$$

To evaluate the integration in (A.101), we need to analyze the term $\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n$. It follows from the asymptotic expansion of $\varphi(B(\mathbf{x},r))$ in (A.100) that

$$\begin{aligned}
&\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n \\
&= \left[1-\frac{f(\mathbf{x})V_d t^{d/k}}{n}-\frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)}t^{(d+2)/k}}{n^{1+2/d}}\quad + o(n^{-(1+2/d)})\right]^n. \qquad \text{(A.102)}
\end{aligned}$$

From (A.102), we see that for each fixed $t>0$,

$$\lim_{n\to\infty}\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n = \exp(-f(\mathbf{x})V_d t^{d/k}).$$

Moreover, by Condition 1, we have

$$\begin{aligned}
\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n &\le \left[\exp\left(-\alpha\frac{t^{1/k}}{n^{1/d}}\right)\right]^n \\
&\le \exp\left(-\alpha t^{1/k}\right).
\end{aligned}$$

Thus, an application of the dominated convergence theorem yields

$$\begin{aligned}
\lim_{n\to\infty}\int_0^\infty \left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n\,\mathrm{d}t &= \int_0^\infty \lim_{n\to\infty}\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/k}}{n^{1/d}}\right)\right)\right]^n\,\mathrm{d}t \\
&= \int_0^\infty \exp(-f(\mathbf{x})V_d t^{d/k})\,\mathrm{d}t \\
&= \frac{\Gamma(k/d+1)}{(f(\mathbf{x})V_d)^{k/d}}, \qquad \text{(A.103)}
\end{aligned}$$

which establishes the desired asymptotic expansion in (A.95) for any $k \geq 1$.

We further investigate higher-order asymptotic expansion for the case of $k = 2$. The leading term of the asymptotic expansion for $\mathbb{E}\,\|\mathbf{X}_{(1)} - \mathbf{x}\|^2$ has been identified in (A.103) with the choice of $k = 2$. But we now aim to conduct a higher-order asymptotic expansion. To do so, we will resort to the higher-order asymptotic expansion given in (A.102). In view of (A.102), we can deduce from the Taylor expansion for function $\log(1-x)$ around 0 that

$$\begin{aligned}
&\left[1 - \varphi\left(B\left(\mathbf{x}, \frac{t^{1/2}}{n^{1/d}}\right)\right)\right]^n - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{n \log\left[1 - \frac{f(\mathbf{x})V_d t^{d/2}}{n} - \frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)} t^{(d+2)/2}}{n^{1+2/d}} + o(n^{-(1+2/d)})\right]\right\} \\
&\quad - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2} - \frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)} t^{(d+2)/2}}{n^{2/d}} - \frac{f^2(\mathbf{x})V_d^2 t^d}{2n} + o(n^{-(2/d)})\right\} \\
&\quad - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \qquad \text{(A.104)}
\end{aligned}$$

as $n \to \infty$. To determine the order of the above remainders, there are three separate cases, that is, $d = 1$, $d = 2$, and $d \geq 3$.

First, for the case of $d = 1$, it follows from (A.104) that

$$\begin{aligned}
&\left[1 - \varphi\left(B\left(\mathbf{x}, \frac{t^{1/2}}{n^{1/d}}\right)\right)\right]^n - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2} - \frac{f^2(\mathbf{x})V_d^2 t^d}{2n} + o(n^{-1})\right\} - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(\exp\left\{-\frac{f^2(\mathbf{x})V_d^2 t^d}{2n} + o(n^{-1})\right\} - 1\right) \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(-\frac{f^2(\mathbf{x})V_d^2 t^d}{2n} + o(n^{-1})\right) \qquad \text{(A.105)}
\end{aligned}$$

as $n \to \infty$. Furthermore, it holds that

$$\int_0^\infty \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(-\frac{f^2(\mathbf{x})V_d^2 t^d}{2}\right)\, \mathrm{d}t = -\frac{\Gamma(2/d+2)}{d(f(\mathbf{x})V_d)^{2/d}}, \qquad \text{(A.106)}$$

where we have used the fact that for any $a > 0$ and $b > 0$,

$$\int_0^\infty x^{a-1} \exp(-bx^p)\, \mathrm{d}x = \frac{1}{p} b^{-a/p} \Gamma(\frac{a}{p}). \tag{A.107}$$

Therefore, combining (A.101), (A.103), (A.105), and (A.106) results in the desired higher-order asymptotic expansion in (A.96) for the case of $k = 2$ and $d = 1$.

When $d = 2$, noting that $2/d = 1$, it follows from (A.104) that

$$\begin{aligned}
&\left[1 - \varphi\left(B\left(\mathbf{x}, \frac{t^{1/2}}{n^{1/d}}\right)\right)\right]^n - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2} - \frac{\frac{\operatorname{tr}(f''(\mathbf{x}))V_d}{2(d+2)} t^{(d+2)/2}}{n^{2/d}} - \frac{f^2(\mathbf{x})V_d^2 t^d}{2n^{2/d}} + o(n^{-(2/d)})\right\} \\
&\quad - \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\} \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(\exp\left\{-\frac{\frac{\operatorname{tr}(f''(\mathbf{x}))V_d}{2(d+2)} t^{(d+2)/2}}{n^{2/d}} - \frac{f^2(\mathbf{x})V_d^2 t^d}{2n^{2/d}} + o(n^{-(2/d)})\right\} - 1\right) \\
&= \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(-\frac{\frac{\operatorname{tr}(f''(\mathbf{x}))V_d}{2(d+2)} t^{(d+2)/2}}{n^{2/d}} - \frac{f^2(\mathbf{x})V_d^2 t^d}{2n^{2/d}} + o(n^{-(2/d)})\right)
\end{aligned} \tag{A.108}$$

as $n \to \infty$. Applying equality (A.107) again yields

$$\begin{aligned}
&\int_0^\infty \exp\left\{-f(\mathbf{x})V_d t^{d/2}\right\}\left(-\frac{\operatorname{tr}(f''(\mathbf{x}))V_d}{2(d+2)n^{2/d}} t^{(d+2)/2}\right)\, \mathrm{d}t \\
&= -\left(\frac{\operatorname{tr}(f''(\mathbf{x}))\Gamma(4/d+1)}{d(d+2)f(\mathbf{x})(f(\mathbf{x})V_d)^{4/d}}\right) n^{-2/d}.
\end{aligned} \tag{A.109}$$

Hence, combining (A.101), (A.103), (A.106), (A.108), and (A.109) leads to the desired higher-order asymptotic expansion in (A.97) for the case of $k = 2$ and $d = 2$.

Finally, it remains to investigate the case of $d \geq 3$. In view of $n^{-1} = o(n^{-2/d})$ for $d \geq 3$, we can obtain from (A.104) that

$$\left[1-\varphi\left(B\left(\mathbf{x},\frac{t^{1/2}}{n^{1/d}}\right)\right)\right]^n-\exp\left\{-f(\mathbf{x})V_dt^{d/2}\right\}$$
$$=\exp\left\{-f(\mathbf{x})V_dt^{d/2}-\frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)}t^{(d+2)/2}}{n^{2/d}}+o(n^{-(2/d)})\right\}-\exp\left\{-f(\mathbf{x})V_dt^{d/2}\right\}$$
$$=\exp\left\{-f(\mathbf{x})V_dt^{d/2}\right\}\left(\exp\left\{-\frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)}t^{(d+2)/2}}{n^{2/d}}+o(n^{-(2/d)})\right\}-1\right)$$
$$=\exp\left\{-f(\mathbf{x})V_dt^{d/2}\right\}\left(-\frac{\frac{\mathrm{tr}(f''(\mathbf{X}))V_d}{2(d+2)}t^{(d+2)/2}}{n^{2/d}}+o(n^{-(2/d)})\right). \tag{A.110}$$

Consequently, combining (A.101), (A.103), (A.109), and (A.110) yields the desired higher-order asymptotic expansion in (A.98) for the case of $k = 2$ and $d \geq 3$. This concludes the proof of Lemma 5.

E.5 Lemma 6 and its proof

As in Biau and Devroye (2015), we define the projection of the mean function $\mu(\mathbf{X}) = \mathbb{E}(Y|\mathbf{X})$ onto the positive half line $\mathbb{R}_+ = [0, \infty)$ given by $\|\mathbf{X}-\mathbf{x}\|$ as

$$m(r) = \lim_{\delta\to 0+}\mathbb{E}\left[\mu(\mathbf{X}) \mid r \leq \|\mathbf{X}-\mathbf{x}\| \leq r+\delta\right] = \mathbb{E}\left[Y \mid \|\mathbf{X}-\mathbf{x}\| = r\right] \tag{A.111}$$

for any $r \geq 0$. Clearly, the definition in (A.111) entails that

$$m(0) = \mathbb{E}\left[Y \mid \mathbf{X} = \mathbf{x}\right] = \mu(\mathbf{x}). \tag{A.112}$$

We will show in Lemma 6 below that the projection $m(\cdot)$ admits an explicit higher-order asymptotic expansion as the distance $r \to 0$.

Lemma 6. *For each fixed $\mathbf{x} \in \mathrm{supp}(\mathbf{X}) \subset \mathbb{R}^d$, we have*

$$m(r) = m(0) + \frac{f(\mathbf{x})\,\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})}\,r^2 + O_4 r^4 \tag{A.113}$$

as $r \to 0$, where O_4 is some bounded quantity depending only on d and the fourth-order partial derivatives of the underlying density function $f(\cdot)$ and regression function $\mu(\cdot)$. Here $g'(\cdot)$ and $g''(\cdot)$ stand for the gradient vector and the Hessian matrix, respectively, of a given function $g(\cdot)$.

Proof. We will exploit the spherical coordinate integration in our proof. Let us first introduce some necessary notation. Denote by $B(\mathbf{0}, r)$ the ball centered at $\mathbf{0}$ and with radius r in the Euclidean space $\mathbb{R}^d$, $\mathbb{S}^{d-1}$ the unit sphere in $\mathbb{R}^d$, ν a measure constructed on the unit sphere $\mathbb{S}^{d-1}$ as in (A.100), and $\boldsymbol{\xi} = (\xi_i) \in \mathbb{S}^{d-1}$ an arbitrary point on the unit sphere. Let V_d be the volume of the unit ball in $\mathbb{R}^d$ as given in (A.95). The integration with the spherical coordinates is equivalent to the standard integration through the identity

$$\int_{B(\mathbf{0},r)} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_0^r u^{d-1} \int_{\mathbb{S}^{d-1}} f(u\,\boldsymbol{\xi})\,\nu(\mathrm{d}\boldsymbol{\xi})\,\mathrm{d}u. \tag{A.114}$$

From Lemma 11 in Section F.1, we have the following integration formulas with the spherical coordinates

$$\int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi}) = d\,V_d, \tag{A.115}$$

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}\,\nu(\mathrm{d}\boldsymbol{\xi}) = \mathbf{0}, \tag{A.116}$$

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\,\boldsymbol{\xi}\,\nu(\mathrm{d}\boldsymbol{\xi}) = \mathrm{tr}(A)\,V_d, \tag{A.117}$$

$$\int_{\mathbb{S}^{d-1}} \xi_i\xi_j\xi_k\nu(\mathrm{d}\boldsymbol{\xi}) = 0 \quad \text{for any } 1 \le i, j, k \le d, \tag{A.118}$$

where A is any $d \times d$ symmetric matrix. We will make use of the identities in (A.115)–(A.118) in our technical analysis.

Let us decompose $m(r)$ into two terms that we will analyze separately

$$\begin{aligned} m(r) &= \lim_{\delta \to 0+} \mathbb{E}\left[\mu(\mathbf{X}) \mid r \le \|\mathbf{X} - \mathbf{x}\| \le r + \delta\right] \\ &= \lim_{\delta \to 0+} \frac{\mathbb{E}\left[\mu(\mathbf{X})\mathbb{1}(r \le \|\mathbf{X} - \mathbf{x}\| \le r + \delta)\right]}{\mathbb{P}\left(r \le \|\mathbf{X} - \mathbf{x}\| \le r + \delta\right)}, \end{aligned} \tag{A.119}$$

where $\mathbb{1}(\cdot)$ stands for the indicator function. In view of (A.114), we can obtain the spherical coordinate representations for the denominator and numerator in (A.119)

$$\mathbb{P}\left(r \le \|\mathbf{X}-\mathbf{x}\| \le r+\delta\right) = \int_r^{r+\delta} u^{d-1} \int_{\mathbb{S}^{d-1}} f(\mathbf{x}+u\,\boldsymbol{\xi})\,\nu(\mathrm{d}\boldsymbol{\xi})\,\mathrm{d}u \tag{A.120}$$

and

$$\begin{aligned}&\mathbb{E}\left[\mu(\mathbf{X})\mathbb{1}(r \le \|\mathbf{X}-\mathbf{x}\| \le r+\delta)\right]\\ &= \int_r^{r+\delta} u^{d-1} \int_{\mathbb{S}^{d-1}} \mu(\mathbf{x}+u\,\boldsymbol{\xi}) f(\mathbf{x}+u\,\boldsymbol{\xi})\,\nu(\mathrm{d}\boldsymbol{\xi})\,\mathrm{d}u.\end{aligned} \tag{A.121}$$

Note that in light of (A.119)–(A.121), an application of L'Hôpital's rule leads to

$$\begin{aligned} m(r) &= \lim_{\delta\to 0+} \frac{\mathbb{E}\left[\mu(\mathbf{X})\mathbb{1}(r \le \|\mathbf{X}-\mathbf{x}\| \le r+\delta)\right]}{\mathbb{P}\left(r \le \|\mathbf{X}-\mathbf{x}\| \le r+\delta\right)}\\ &= \frac{\int_{\mathbb{S}^{d-1}} \mu(\mathbf{x}+r\,\boldsymbol{\xi}) f(\mathbf{x}+r\,\boldsymbol{\xi})\;\nu(\mathrm{d}\boldsymbol{\xi})}{\int_{\mathbb{S}^{d-1}} f(\mathbf{x}+r\,\boldsymbol{\xi})\;\nu(\mathrm{d}\boldsymbol{\xi})}. \end{aligned} \tag{A.122}$$

First let us expand the denominator. Using the spherical coordinate integration, we can deduce that

$$\begin{aligned} &\int_{\mathbb{S}^{d-1}} f(\mathbf{x}+r\boldsymbol{\xi})\;\nu(\mathrm{d}\boldsymbol{\xi})\\ &= \int_{\mathbb{S}^{d-1}} \Big(f(\mathbf{x}) + f'(\mathbf{x})^T\,\boldsymbol{\xi}\,r + \frac{1}{2}\,\boldsymbol{\xi}^T\,f''(\mathbf{x})\,\boldsymbol{\xi}\,r^2 + \frac{1}{6}\sum_{1\le i,j,k\le d} \frac{\partial^3 f(\mathbf{x})}{\partial \mathbf{x}_i \partial \mathbf{x}_j \partial \mathbf{x}_k} \xi_i\xi_j\xi_k r^3 \\ &\quad + \frac{1}{24}\sum_{1\le i,j,k,l\le d} \frac{\partial^4 f(\mathbf{x}+\theta r\boldsymbol{\xi})}{\partial \mathbf{x}_i \partial \mathbf{x}_j \partial \mathbf{x}_k \partial \mathbf{x}_l} \xi_i\xi_j\xi_k\xi_l r^4 \Big)\;\nu(\mathrm{d}\boldsymbol{\xi}), \end{aligned} \tag{A.123}$$

where $0 < \theta < 1$. Note that the fourth-order partial derivatives of f are bounded in some neiborghhood of $\mathbf{x}$ by Condition 2, and

$$\begin{aligned} \int_{\mathbb{S}^{d-1}} \sum_{1\le i,j,k,l\le d} |\xi_i\xi_j\xi_k\xi_l|\;\nu(\mathrm{d}\boldsymbol{\xi}) &= \int_{\mathbb{S}^{d-1}} \Big(\sum_{i=1}^d |\xi_i|\Big)^4\;\nu(\mathrm{d}\boldsymbol{\xi})\\ &\le \int_{\mathbb{S}^{d-1}} d^2\Big(\sum_{i=1}^d \xi_i^2\Big)^2\;\nu(\mathrm{d}\boldsymbol{\xi})\\ &= d^2\int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi}) = d^3 V_d. \end{aligned} \tag{A.124}$$

Thus, from (A.115)–(A.118) and (A.124) we can obtain

$$\int_{\mathbb{S}^{d-1}} f(\mathbf{x}+r\boldsymbol{\xi})\ \nu(\mathrm{d}\boldsymbol{\xi}) = f(\mathbf{x})\, d\, V_d + \frac{1}{2}\,\mathrm{tr}(f''(\mathbf{x}))V_d\, r^2 + R_1(d,f,\mathbf{x})\, r^4, \tag{A.125}$$

where the coefficient $R_1(d,f,\mathbf{x})$ in the remainder term is bounded and depends only on the fourth-order partial derivatives of f and dimensionality d.

For the numerator, it holds that

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}} \mu(\mathbf{x}+r\,\boldsymbol{\xi})f(\mathbf{x}+r\,\boldsymbol{\xi})\ \nu(\mathrm{d}\boldsymbol{\xi})\\
&= \int_{\mathbb{S}^{d-1}} \Big[\mu(\mathbf{x}) + \mu'(\mathbf{x})^T\boldsymbol{\xi}\, r + \frac{1}{2}\,\boldsymbol{\xi}^T\mu''(\mathbf{x})\,\boldsymbol{\xi}\, r^2\\
&\quad + \frac{1}{6}\sum_{1\le i,j,k\le d}\frac{\partial^3\mu(\mathbf{x})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k}\xi_i\xi_j\xi_k r^3 + \frac{1}{24}\sum_{1\le i,j,k,l\le d}\frac{\partial^4\mu(\mathbf{x}+\theta_1 r\boldsymbol{\xi})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k\partial\mathbf{x}_l}\xi_i\xi_j\xi_k\xi_l r^4\Big]\\
&\quad\times\Big[f(\mathbf{x}) + f'(\mathbf{x})^T\boldsymbol{\xi}\, r + \frac{1}{2}\boldsymbol{\xi}^T f''(\mathbf{x})\,\boldsymbol{\xi}\, r^2\\
&\quad + \frac{1}{6}\sum_{i,j,k}\frac{\partial^3 f(\mathbf{x})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k}\xi_i\xi_j\xi_k r^3 + \frac{1}{24}\sum_{1\le i,j,k,l\le d}\frac{\partial^4 f(\mathbf{x}+\theta_2 r\boldsymbol{\xi})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k\partial\mathbf{x}_l}\xi_i\xi_j\xi_k\xi_l r^4\Big]\ \nu(\mathrm{d}\boldsymbol{\xi}),
\end{aligned}\tag{A.126}$$

where $0<\theta_1<1$ and $0<\theta_2<1$. In the same manner as deriving (A.124), we can bound the integrals associated with r^4 and the higher-orders r^5, r^6, r^7, and r^8 under Condition 2 that the fourth-order partial derivatives of $f(\cdot)$ and $\mu(\cdot)$ are bounded in a neighborhood of $\mathbf{x}$. Hence, we can deduce that

$$\begin{aligned}
&\int_{\mathbb{S}^{d-1}} \mu(\mathbf{x}+r\,\boldsymbol{\xi})f(\mathbf{x}+r\,\boldsymbol{\xi})\ \nu(\mathrm{d}\boldsymbol{\xi})\\
&= \mu(\mathbf{x})f(\mathbf{x})\int_{\mathbb{S}^{d-1}}\ \nu(\mathrm{d}\boldsymbol{\xi}) + \frac{\mu(\mathbf{x})r^2}{2}\int_{\mathbb{S}^{d-1}}\boldsymbol{\xi}^T f''(\mathbf{x})\boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi})\\
&\quad + r^2\int_{\mathbb{S}^{d-1}}\boldsymbol{\xi}^T\mu'(\mathbf{x})f'(\mathbf{x})^T\boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi}) + \frac{f(\mathbf{x})r^2}{2}\int_{\mathbb{S}^{d-1}}\boldsymbol{\xi}^T\mu''(\mathbf{x})\boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi})\\
&\quad + R_2(d,f,\mathbf{x})r^4 + o(r^4)\\
&= \mu(\mathbf{x})f(\mathbf{x})dV_d + \frac{1}{2}\left[f(\mathbf{x})\,\mathrm{tr}(\mu''(\mathbf{x})) + \mu(\mathbf{x})\ \mathrm{tr}(f''(\mathbf{x}))\right]V_d\, r^2\\
&\quad + \mu'(\mathbf{x})^T f'(\mathbf{x})V_d\, r^2 + R_2(d,f,\mathbf{x})r^4 + o(r^4),
\end{aligned}\tag{A.127}$$

where the coefficient $R_2(d, f, \mathbf{x})$ in the remainder term is bounded and depends only on the fourth-order partial derivatives of f and dimensionality d. The last equality in (A.127) follows from (A.115)–(A.118). Therefore, substituting (A.125) and (A.127) into (A.122) leads to

$$m(r) = \mu(\mathbf{x}) + \frac{f(\mathbf{x})\mathrm{tr}(\mu''(\mathbf{x})) + 2\,\mu'(\mathbf{x})^T f'(\mathbf{x})}{2\,d\,f(\mathbf{x})}\,r^2 + O_4 r^4$$

as $r \to 0$, where O_4 is a bounded quantity depending only on d and the fourth-order partial derivatives of $f(\cdot)$ and $\mu(\cdot)$. This completes the proof of Lemma 6.

E.6 Lemma 7 and its proof

Lemma 7 below provides us with the order of the variance for the first-order Hájek projection. To simplify the technical presentation, we use $\mathbf{Z}_i$ as a shorthand notation for $(\mathbf{X}_i, Y_i)$. Given any fixed vector $\mathbf{x}$, the projection of $\Phi(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_s)$ onto $\mathbf{Z}_1$ is denoted as $\Phi_1(\mathbf{x}; \mathbf{z}_1)$ given by

$$\begin{aligned}\Phi_1(\mathbf{x}; \mathbf{z}_1) &= \mathbb{E}\,[\Phi(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_s) | \mathbf{Z}_1 = \mathbf{z}_1] \\ &= \mathbb{E}\,[\Phi(\mathbf{x}; \mathbf{z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_s)]. \end{aligned} \tag{A.128}$$

Denote by $\mathbb{E}_i$ and $\mathbb{E}_{i:s}$ the expectations with respect to $\mathbf{Z}_i$ and $\{\mathbf{Z}_i, \mathbf{Z}_{i+1}, \cdots, \mathbf{Z}_s\}$, respectively.

Lemma 7. *For any fixed* $\mathbf{x}$*, the variance* η_1 *of* $\Phi_1(\mathbf{x}; \mathbf{Z}_1)$ *defined in (A.128) satisfies that when* $s \to \infty$ *and* $s = o(n)$*,*

$$\lim_{n\to\infty} \frac{\mathrm{Var}(\Phi)}{n\eta_1} = 0. \tag{A.129}$$

Proof. A main ingredient of the proof is to decompose $\mathrm{Var}(\Phi)$ and η_1 using the conditioning arguments. Denote by $\zeta_{i,s}$ the indicator function for the event that $\mathbf{X}_i$ is the 1NN of $\mathbf{x}$

among $\{\mathbf{X}_1, \cdots, \mathbf{X}_s\}$. By symmetry, we can see that $\zeta_{i,s}$ are identically distributed with mean

$$\mathbb{E}\zeta_{i,s} = s^{-1}.$$

In addition, observe that $\Phi(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_s) = \sum_{i=1}^s y_i \zeta_{i,s}$. Then we can obtain an upper bound of $\operatorname{Var}\Phi$ as

$$\begin{aligned}\operatorname{Var}(\Phi) \le \mathbb{E}[\Phi^2] &= \mathbb{E}\Big[\Big(\sum_{i=1}^s y_i \zeta_{i,s}\Big)^2\Big] = \sum_{i=1}^s \mathbb{E}[y_i^2 \zeta_{i,s}] \\ &= s\mathbb{E}[y_1^2 \zeta_{1,s}],\end{aligned}$$

where we have used the fact that $\zeta_{i,s}\zeta_{j,s} = 0$ with probability one when $i \neq j$.

Since $\mathbb{E}[\epsilon|\mathbf{X}] = 0$ by assumption, it holds that

$$\begin{aligned}s\mathbb{E}[y_1^2 \zeta_{1,s}] &= s\mathbb{E}[\mu^2(\mathbf{X}_1)\zeta_{1,s}] + \sigma_\epsilon^2 s\mathbb{E}[\zeta_{1,s}] \\ &= \mathbb{E}_1[\mu^2(\mathbf{X}_1) s\mathbb{E}_{2:s}[\zeta_{1,s}]] + \sigma_\epsilon^2.\end{aligned}$$

A key observation is that $\mathbb{E}_{2:s}[\zeta_{1,s}] = \{1 - \varphi(B(\mathbf{x}, \|\mathbf{X}_1 - \mathbf{x}\|))\}^{s-1}$ and $\mathbb{E}_1[s\mathbb{E}_{2:s}[\zeta_{1,s}]] = 1$. See Lemma 12 in Section F.2 for a list of properties for the indicator functions $\zeta_{i,s}$. Thus, $s\mathbb{E}_{2:s}[\zeta_{1,s}]$ behaves like a Dirac measure at $\mathbf{x}$ as $s \to \infty$. Such observation leads to Lemma 13 in Section F.3, which entails that

$$\operatorname{Var}(\Phi) \le \mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1) \tag{A.130}$$

as $s \to \infty$.

To derive a lower bound for η_1, we exploit the idea in Theorem 3 of Peng et al. (2019). Let B be the event that $\mathbf{X}_1$ is the nearest neighbor of $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_s\}$. Denote by

$\mathbf{X}_1^*$ the nearest point to $\mathbf{x}$ and y_1^* the corresponding response. Then we can deduce that

$$\begin{aligned}\Phi_1(\mathbf{x};\mathbf{Z}_1) &= \mathbb{E}[y_1\mathbb{1}_B|\mathbf{Z}_1] + \mathbb{E}[y_1^*\mathbb{1}_{B^c}|\mathbf{Z}_1] \\ &= y_1\mathbb{E}[\mathbb{1}_B|\mathbf{Z}_1] + \mathbb{E}[y_1^*\mathbb{1}_{B^c}|\mathbf{Z}_1] \\ &= \epsilon_1\mathbb{E}[\mathbb{1}_B|\mathbf{X}_1] + \mu(\mathbf{X}_1)\mathbb{E}[\mathbb{1}_B|\mathbf{X}_1] + \mathbb{E}[\mu(\mathbf{X}_1^*)\mathbb{1}_{B^c}|\mathbf{X}_1] \\ &= \epsilon_1\mathbb{E}[\mathbb{1}_B|\mathbf{X}_1] + \mathbb{E}[\mu(\mathbf{X}_1^*)|\mathbf{X}_1].\end{aligned}$$

Since ϵ is an independent model error term with $\mathbb{E}[\epsilon|\mathbf{X}] = 0$ by assumption, it holds that

$$\begin{aligned}\eta_1 &= \mathrm{Var}(\Phi_1(\mathbf{x};\mathbf{Z}_1)) = \mathrm{Var}(\epsilon_1\mathbb{E}[\mathbb{1}_B|\mathbf{X}_1]) + \mathrm{Var}(\mathbb{E}[\mu(\mathbf{X}_1^*)|\mathbf{X}_1]) \\ &\geq \mathrm{Var}(\epsilon_1\mathbb{E}[\mathbb{1}_B|\mathbf{X}_1]) = \sigma_\epsilon^2\mathbb{E}\left[\mathbb{E}^2[\mathbb{1}_B|\mathbf{X}_1]\right] \\ &= \frac{\sigma_\epsilon^2}{2s-1}, \end{aligned} \tag{A.131}$$

where we have used the fact that

$$\mathbb{E}[\mathbb{E}^2[\mathbb{1}_B|\mathbf{X}_1]] = \mathbb{E}[\mathbb{1}_{B'}|\mathbf{X}_1] = \frac{1}{2s-1}$$

with B' representing the event that $\mathbf{X}_1$ is the nearest neighbor of $\mathbf{x}$ among the i.i.d. observations $\{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_s, \mathbf{X}'_2, \cdots, \mathbf{X}'_s\}$.

We now turn to the upper bound for η_1. From the variance decomposition for $\mathrm{Var}(\Phi)$ given in (A.14), we can obtain

$$\begin{aligned}\mathrm{Var}(\Phi) &= \sum_{j=1}^{s}\binom{s}{j}\mathrm{Var}(g_j(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_j)) \\ &= s\eta_1 + \sum_{j=2}^{s}\binom{s}{j}\mathrm{Var}(g_j(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_j)),\end{aligned}$$

which along with (A.130) entails that

$$s\eta_1 \leq \mathrm{Var}(\Phi) \leq \mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1). \tag{A.132}$$

Consequently, combining (A.131) and (A.132) leads to

$$\eta_1 \sim s^{-1}, \tag{A.133}$$

where $\sim$ denotes the asymptotic order. Finally, recall that it has been shown that $\mathrm{Var}(\Phi) \leq C$ for some positive constant depending upon $\mu(\mathbf{x})$ and σ_ϵ. Therefore, we see that as long as $s \to \infty$ and $s = o(n)$,

$$\frac{\mathrm{Var}(\Phi)}{n\eta_1} = O(\frac{s}{n}) \to 0,$$

which yields the desired conclusion in (A.129). This concludes the proof of Lemma 7.

E.7 Lemma 8 and its proof

Assume that $s_1 < s_2$ for the two subsampling scales. Let us define

$$\Phi^{(1)}(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}) = \binom{s_2}{s_1}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{s_1} \leq s_2} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \cdots, \mathbf{Z}_{i_{s_1}}) \tag{A.134}$$

and

$$\Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}) = w_1^* \Phi^{(1)}(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}) + w_2^* \Phi(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}), \tag{A.135}$$

where w_1^* and w_2^* are determined by the system of linear equations (9)–(10).

Lemma 8. *The two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ admits a U-statistic representation given by*

$$D_n(s_1, s_2)(\mathbf{x}) = \binom{n}{s_2}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{s_2} \leq n} \Phi^*(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}), \tag{A.136}$$

where the kernel function $\Phi^*(\mathbf{x}; \cdot)$ is defined in (A.135).

Proof. From the definition of the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$ introduced in (11), we have

$$D_n(s_1, s_2)(\mathbf{x}) = w_1^* \binom{n}{s_1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_1} \le n} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}) + w_2^* \binom{n}{s_2}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_2} \le n} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}}).$$

Thus, to establish the U-statistic representation for the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$, it suffices to show that

$$\begin{aligned} &\binom{n}{s_1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_1} \le n} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}) \\ &= \binom{n}{s_2}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_2} \le n} \Phi^{(1)}(\mathbf{x}; \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_{s_2}) \\ &= \binom{n}{s_2}^{-1} \binom{s_2}{s_1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_2} \le n} \sum_{1 \le j_1 < j_2 < \cdots < j_{s_1} \le s_2} \Phi(\mathbf{x}; \mathbf{Z}_{i_{j_1}}, \mathbf{Z}_{i_{j_2}}, \cdots, \mathbf{Z}_{i_{j_{s_1}}}). \end{aligned} \tag{A.137}$$

Observe that for each given tuple $1 \le u_1 < u_2 < \cdots < u_{s_1} \le n$, it will appear a total of $\binom{n-s_1}{s_2-s_1}$ times in the summation

$$\sum_{1 \le i_1 < i_2 < \cdots < i_{s_2} \le n} \sum_{1 \le j_1 < j_2 < \cdots < j_{s_1} \le s_2} \Phi(\mathbf{x}; \mathbf{Z}_{i_{j_1}}, \mathbf{Z}_{i_{j_2}}, \cdots, \mathbf{Z}_{i_{j_{s_1}}}).$$

Indeed, if $(i_{j_1}, i_{j_2}, \cdots, i_{j_{s_1}}) = (u_1, u_2, \cdots, u_{s_1})$ are fixed, then there exist $\binom{n-s_1}{s_2-s_1}$ options for the remaining $s_2 - s_1$ places in $(i_1, i_2, \cdots, i_{s_2})$. Consequently, it holds that

$$\begin{aligned} &\binom{n}{s_2}^{-1} \binom{s_2}{s_1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_2} \le n} \sum_{1 \le j_1 < j_2 < \cdots < j_{s_1} \le s_2} \Phi(\mathbf{x}; \mathbf{Z}_{i_{j_1}}, \mathbf{Z}_{i_{j_2}}, \cdots, \mathbf{Z}_{i_{j_{s_1}}}) \\ &= \binom{n}{s_2}^{-1} \binom{s_2}{s_1}^{-1} \binom{n-s_1}{s_2-s_1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_1} \le n} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}) \\ &= \binom{n}{s_1}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_{s_1} \le n} \Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}), \end{aligned}$$

which establishes the desired claim in (A.137). This completes the proof of Lemma 8.

E.8 Lemma 9 and its proof

We provide in Lemma 9 below the order of the variance of the kernel function Φ^* defined in (A.135) for the two-scale DNN estimator $D_n(s_1, s_2)(\mathbf{x})$, which states that the variance of the kernel function is bounded from above by some positive constant depending upon the underlying distributions. Denote by $\mathrm{Var}(\Phi^*) = \mathrm{Var}[\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]$ for simplicity.

Lemma 9. *Under the conditions of Theorem 3, there exists some positive constant C depending upon c_1 and c_2 such that*

$$\mathrm{Var}(\Phi^*) \leq C\big(\mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1)\big) \tag{A.138}$$

as $s_1 \to \infty$ and $s_2 \to \infty$.

Proof. Since $\mathrm{Var}(\Phi^*) \leq \mathbb{E}[(\Phi^*)^2]$, it suffices to bound $\mathbb{E}[(\Phi^*)^2]$. It follows that

$$\begin{aligned}
\mathbb{E}[(\Phi^*)^2] &\leq 2(w_1^*)^2 \mathbb{E}\big\{[\Phi^{(1)}(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]^2\big\} \\
&\quad + 2(w_2^*)^2 \mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})] \\
&\leq 2(w_1^*)^2 \mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] \\
&\quad + 2(w_2^*)^2 \mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})],
\end{aligned} \tag{A.139}$$

where the last inequality holds since

$$\begin{aligned}
&\mathbb{E}\big\{[\Phi^{(1)}(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})]^2\big\} \\
&= \binom{s_2}{s_1}^{-2} \sum_{\substack{1 \leq i_1 < \cdots < i_{s_1} \leq s_2 \\ 1 \leq j_1 < \cdots < j_{s_1} \leq s_2}} \mathbb{E}\big\{\Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}})\Phi(\mathbf{x}; \mathbf{Z}_{j_1}, \cdots, \mathbf{Z}_{j_{s_1}})\big\} \\
&\leq \binom{s_2}{s_1}^{-2} \sum_{\substack{1 \leq i_1 < \cdots < i_{s_1} \leq s_2 \\ 1 \leq j_1 < \cdots < j_{s_1} \leq s_2}} \mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] \\
&= \mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})].
\end{aligned}$$

Since $\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1}) = \sum_{i=1}^{s_1} y_i \zeta_{i,s_1}$ and $\zeta_{i,s_1}\zeta_{j,s_1} = 0$ with probability one when $i \neq j$, we can deduce that

$$\begin{aligned}\mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] &= \mathbb{E}\Big[\Big(\sum_{i=1}^{s_1} y_i \zeta_{i,s_1}\Big)^2\Big] = \sum_{i=1}^{s_1}\sum_{j=1}^{s_1} y_i y_j \zeta_{i,s_1}\zeta_{j,s_1} \\ &= \sum_{i=1}^{s_1} y_i^2 \zeta_{i,s_1} = s_1 \mathbb{E}[y_1^2 \zeta_{1,s_1}] \\ &= s_1 \mathbb{E}[\mu^2(\mathbf{X}_1)\zeta_{1,s_1}] + \sigma_\epsilon^2 s_1 \mathbb{E}[\zeta_{1,s_1}].\end{aligned}$$

Note that $s_1 \mathbb{E}[\zeta_{1,s_1}] = \sum_{i=1}^{n} \zeta_{i,s_1}$. Furthermore, it follows from Lemma 13 in Section F.3 that

$$s_1 \mathbb{E}[\mu^2(\mathbf{X}_1)\zeta_{1,s_1}] \to \mu^2(\mathbf{x})$$

as $s_1 \to \infty$. Thus, we have that as $s_1 \to \infty$,

$$\mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] = \mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1). \tag{A.140}$$

Similarly, we can show that as $s_2 \to \infty$,

$$\mathbb{E}[\Phi^2(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})] = \mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1). \tag{A.141}$$

Consequently, combining (A.139), (A.140), and (A.141) results in

$$\mathbb{E}[(\Phi^*)^2] \leq 2\big[(w_1^*)^2 + (w_2^*)^2\big]\big[\mu^2(\mathbf{x}) + \sigma_\epsilon^2 + o(1)\big]. \tag{A.142}$$

Since $c_1 \leq s_1/s_2 \leq c_2$ by assumption, it holds that

$$(w_1^*)^2 \leq C \quad \text{and} \quad (w_2^*)^2 \leq C$$

for some absolute positive constant C depending upon c_1 and c_2, which together with (A.142) entails the desired upper bound in (A.138). This concludes the proof of Lemma 9.

E.9 Lemma 10 and its proof

Lemma 10 below establishes the order of the variance for the first-order Hájek projection of the kernel function Φ^* defined in (A.135). Recall that in the proof of Theorem 3 in Section D.3, we have defined that for each $1 \leq i \leq s_2$,

$$\Phi_i^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) = \mathbb{E}[\Phi^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i, \mathbf{Z}_{i+1}, \cdots, \mathbf{Z}_{s_2}) \,|\, \mathbf{z}_1, \cdots, \mathbf{z}_i],$$

$$\begin{aligned} g_i^*(\mathbf{z}_1, \cdots, \mathbf{z}_i) = \Phi_i^*(\mathbf{x}; \mathbf{z}_1, \cdots, \mathbf{z}_i) - \mathbb{E}\Phi^*(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_i) \\ - \sum_{j=1}^{i-1} \sum_{1 \leq \alpha_1 < \cdots < \alpha_j \leq i} g_j^*(\mathbf{z}_{\alpha_1}, \cdots, \mathbf{z}_{\alpha_j}), \end{aligned}$$

and $\eta_1^* = \mathrm{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1))$.

Lemma 10. *Under the conditions of Theorem 3, it holds that*

$$\eta_1^* \sim s_2^{-1}, \tag{A.143}$$

where $\sim$ denotes the asymptotic order.

Proof. We begin with the lower bound for η_1^*. The proof follows the ideas used in the proof of Lemma 7 in Section E.6. By definition, it holds that

$$\begin{aligned} \Phi_1^*(\mathbf{x}; \mathbf{Z}_1) = w_1^* \binom{s_2}{s_1}^{-1} \sum_{1 \leq i_1 < \cdots < i_{s_1} \leq s_2} \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_1}}) | \mathbf{Z}_1] \\ + w_2^* \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_{i_1}, \cdots, \mathbf{Z}_{i_{s_2}}) | \mathbf{Z}_1] \\ = w_1^* \frac{s_2 - s_1}{s_2} \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})] + w_1^* \frac{s_1}{s_2} \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1}) | \mathbf{Z}_1] \\ + w_2^* \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2}) | \mathbf{Z}_1]. \end{aligned}$$

Since the first term on the right-hand side of the above equality is a constant, we have

$$\begin{aligned} \mathrm{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1)) = \mathrm{Var}\Big(w_1^* \frac{s_1}{s_2} \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1}) | \mathbf{Z}_1] \\ + w_2^* \mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2}) | \mathbf{Z}_1] \Big). \end{aligned}$$

Denote by A_1 the event that $\mathbf{X}_1$ is the nearest neighbor of $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_{s_1}\}$ and A_2 the event that $\mathbf{X}_1$ is the nearest neighbor of $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_{s_2}\}$. Let $\mathbf{X}_1^*$ be the nearest point to $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_{s_1}\}$ and y_1^* the corresponding value of the response. Similarly, we define $\breve{\mathbf{X}}_1$ as the nearest point to $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_{s_2}\}$ and $\breve{y}_1$ as the corresponding value of the response. Since $\epsilon_i \perp\!\!\!\perp \mathbf{X}_i$ and $\mathbb{E}[\epsilon_i] = 0$ by assumption, we can write

$$\begin{aligned}
\mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_1})|\mathbf{Z}_1] &= \mathbb{E}[y_1 \mathbb{1}_{A_1}|\mathbf{Z}_1] + \mathbb{E}[y_1^* \mathbb{1}_{A_1^c}|\mathbf{Z}_1] \\
&= \epsilon_1 \mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + \mathbb{E}[\mu(\mathbf{X}_1)\mathbb{1}_{A_1}|\mathbf{X}_1] + \mathbb{E}[\mu(\mathbf{X}_1^*)\mathbb{1}_{A_1^c}|\mathbf{X}_1] \\
&= \epsilon_1 \mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + \mathbb{E}[\mu(\mathbf{X}_1^*)|\mathbf{X}_1].
\end{aligned}$$

Similarly, we can show that

$$\mathbb{E}[\Phi(\mathbf{x}; \mathbf{Z}_1, \cdots, \mathbf{Z}_{s_2})|\mathbf{Z}_1] = \epsilon_1 \mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1] + \mathbb{E}[\mu(\breve{\mathbf{X}}_1)|\mathbf{X}_1].$$

Thus, we can obtain

$$\begin{aligned}
\operatorname{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1)) = \operatorname{Var}\Big\{&\epsilon_1\Big(w_1^* \frac{s_1}{s_2} \mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + w_2^* \mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\Big) \\
&+ w_1^* \frac{s_1}{s_2} \mathbb{E}[\mu(\mathbf{X}_1^*)|\mathbf{X}_1] + w_2^* \mathbb{E}[\mu(\breve{\mathbf{X}}_1)|\mathbf{X}_1]\Big\},
\end{aligned}$$

which along with the assumption of $\epsilon_1 \perp\!\!\!\perp \mathbf{X}_1$ and $\mathbb{E}[\epsilon_1] = 0$ yields

$$\begin{aligned}
\operatorname{Var}(\Phi_1^*(\mathbf{x}; \mathbf{Z}_1)) &= \operatorname{Var}\Big\{\epsilon_1\Big(w_1^* \frac{s_1}{s_2} \mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + w_2^* \mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\Big)\Big\} \\
&\quad + \operatorname{Var}\Big\{w_1^* \frac{s_1}{s_2} \mathbb{E}[\mu(\mathbf{X}_1^*)|\mathbf{X}_1] + w_2^* \mathbb{E}[\mu(\breve{\mathbf{X}}_1)|\mathbf{X}_1]\Big\} \\
&\geq \operatorname{Var}\Big\{\epsilon_1\Big(w_1^* \frac{s_1}{s_2} \mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + w_2^* \mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\Big)\Big\}.
\end{aligned}$$

Furthermore, we can deduce that

$$\begin{aligned}
&\operatorname{Var}\left\{\epsilon_1\Big(w_1^*\frac{s_1}{s_2}\mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + w_2^*\mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\Big)\right\} \\
&= \sigma_\epsilon^2\mathbb{E}\Big\{\Big(w_1^*\frac{s_1}{s_2}\mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1] + w_2^*\mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\Big)^2\Big\} \\
&= \sigma_\epsilon^2\Big\{\Big(w_1^*\frac{s_1}{s_2}\Big)^2\mathbb{E}\big[\mathbb{E}^2[\mathbb{1}_{A_1}|\mathbf{X}_1]\big] + 2w_1^*w_2^*\frac{s_1}{s_2}\mathbb{E}\big[\mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1]\mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\big] \\
&\quad + (w_2^*)^2\mathbb{E}\big[\mathbb{E}^2[\mathbb{1}_{A_2}|\mathbf{X}_1]\big]\Big\}.
\end{aligned}$$

Let us make use of the following basic facts

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}^2[\mathbb{1}_{A_1}|\mathbf{X}_1]\big] &= \frac{1}{2s_1-1}, \\
\mathbb{E}\big[\mathbb{E}[\mathbb{1}_{A_1}|\mathbf{X}_1]\mathbb{E}[\mathbb{1}_{A_2}|\mathbf{X}_1]\big] &= \frac{1}{s_1+s_2-1}, \\
\mathbb{E}\big[\mathbb{E}^2[\mathbb{1}_{A_2}|\mathbf{X}_1]\big] &= \frac{1}{2s_2-1}.
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\operatorname{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)) \geq \sigma_\epsilon^2\Big\{\Big(w_1^*\frac{s_1}{s_2}\Big)^2\frac{1}{2s_1-1} + 2w_1^*w_2^*\frac{s_1}{s_2}\frac{1}{s_1+s_2-1} \\
+ (w_2^*)^2\frac{1}{2s_2-1}\Big\}.
\end{aligned} \tag{A.144}$$

By (A.144) and the assumption of $c_1 \leq s_1/s_2 \leq c_2$, we can obtain

$$\operatorname{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)) \geq C\sigma_\epsilon^2 s_2^{-1} \tag{A.145}$$

for some positive constant C depending upon c_1 and c_2.

We next proceed to show the upper bound for $\operatorname{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1))$. Since

$$\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2}) - \mathbb{E}\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2}) = \sum_{j=1}^{s_2}\sum_{1\leq\alpha_1<\cdots<\alpha_j\leq s_2} g_j^*(\mathbf{Z}_{\alpha_1},\cdots,\mathbf{Z}_{\alpha_j}),$$

we see that

$$\operatorname{Var}(\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2})) = \sum_{j=1}^{s_2}\binom{s_2}{j}\operatorname{Var}(g_j^*(\mathbf{Z}_1,\cdots,\mathbf{Z}_j)).$$

Then it follows that

$$\mathrm{Var}(\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2})) \geq s_2\,\mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)).$$

Recall that it has been shown in Lemma 9 in Section E.8 that

$$\mathrm{Var}(\Phi^*(\mathbf{x};\mathbf{Z}_1,\cdots,\mathbf{Z}_{s_2})) \leq C,$$

where C is some positive constant depending upon c_1, c_2, and the underlying distributions. Therefore, we can deduce that

$$\mathrm{Var}(\Phi_1^*(\mathbf{x};\mathbf{Z}_1)) \leq C s_2^{-1}, \tag{A.146}$$

which together with (A.145) entails the desired asymptotic order in (A.143). This completes the proof of Lemma 10.

F Additional technical details

F.1 Lemma 11 and its proof

We present in Lemma 11 below some useful spherical integration formulas.

Lemma 11. *Let $\mathbb{S}^{d-1}$ be the unit sphere in $\mathbb{R}^d$, ν some measure constructed specifically on the unit sphere $\mathbb{S}^{d-1}$, and $\boldsymbol{\xi} = (\xi_i) \in \mathbb{S}^{d-1}$ an arbitrary point on the unit sphere. Then for any $d \times d$ symmetric matrix A, it holds that*

$$\int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi}) = d\,V_d, \tag{A.147}$$

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}\,\nu(\mathrm{d}\boldsymbol{\xi}) = \mathbf{0}, \tag{A.148}$$

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\,\boldsymbol{\xi}\,\nu(\mathrm{d}\boldsymbol{\xi}) = \mathrm{tr}(A)\,V_d, \tag{A.149}$$

$$\int_{\mathbb{S}^{d-1}} \xi_i\xi_j\xi_k\nu(\mathrm{d}\boldsymbol{\xi}) = 0 \quad \textit{for any } 1 \leq i,j,k \leq d, \tag{A.150}$$

where $V_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$ *denotes the volume of the unit ball in* $\mathbb{R}^d$.

Proof. It is easy to see that identities (A.148) and (A.150) hold. This is because for each of them, the integrand is an odd function of variable $\boldsymbol{\xi}$, which entails that the integral is zero. Identity (A.147) can be derived using the iterated integral

$$\begin{aligned} V_d &= \int_0^1 \int_{\mathbb{S}^{d-1}} \rho^{d-1} \nu(\mathrm{d}\boldsymbol{\xi})\, \mathrm{d}\rho = \left(\int_0^1 \rho^{d-1}\, \mathrm{d}\rho\right) \left(\int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi})\right) \\ &= \frac{1}{d} \int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi}). \end{aligned}$$

To prove (A.149), we first represent the integral in (A.149) as a sum of integrals by expanding the quadratic expression in the integrand

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\, \boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi}) = \sum_{1 \le i,j \le d} A_{ij} \int_{\mathbb{S}^{d-1}} \xi_i \xi_j\ \nu(\mathrm{d}\boldsymbol{\xi}). \tag{A.151}$$

For $i \neq j$, we have by symmetry that

$$\int_{\mathbb{S}^{d-1}} \xi_i \xi_j\ \nu(\mathrm{d}\boldsymbol{\xi}) = \int_{\mathbb{S}^{d-1}} -\xi_i \xi_j\ \nu(\mathrm{d}\boldsymbol{\xi}) = 0. \tag{A.152}$$

Thus, it holds that

$$\begin{aligned} \int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\, \boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi}) &= \sum_{i=1}^{d} A_{ii} \int_{\mathbb{S}^{d-1}} \xi_i^2\ \nu(\mathrm{d}\boldsymbol{\xi}) \\ &= \mathrm{tr}(A) \int_{\mathbb{S}^{d-1}} \xi_1^2\ \nu(\mathrm{d}\boldsymbol{\xi}). \end{aligned} \tag{A.153}$$

When $d = 1$, $\mathbb{S}^{d-1}$ reduces to the trivial case of two points, 1 and -1. Then we can obtain that for $d = 1$,

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\, \boldsymbol{\xi}\ \nu(\mathrm{d}\boldsymbol{\xi}) = 2\mathrm{tr}(A) = \mathrm{tr}(A) V_d, \tag{A.154}$$

where the last equality comes from the fact that $V_d = 2$ for $d = 1$. When $d \ge 2$, we now use the spherical coordinates: $\xi_1 = \cos(\phi_1)$, $\xi_k = \cos(\phi_k) \prod_{i=1}^{k-1} \sin(\phi_i)$ for $1 \le k \le d-1$, and $\xi_d = \prod_{i=1}^{d-1} \sin(\phi_i)$, where $0 \le \phi_{d-1} < 2\pi$ and $0 \le \phi_i < \pi$ for $1 \le i \le d-2$. Then the volume element becomes

$$\nu(\mathrm{d}\boldsymbol{\xi}) = \left(\prod_{i=1}^{d-2} \sin^{d-1-i}(\phi_i)\right) \prod_{i=1}^{d} \mathrm{d}\phi_i.$$

It follows that

$$\begin{aligned}
\int_{\mathbb{S}^{d-1}} \xi_1^2 \, \nu(\mathrm{d}\boldsymbol{\xi}) &= \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \cos^2(\phi_1) \left(\prod_{i=1}^{d-2} \sin^{d-1-i}(\phi_i)\right) \prod_{i=1}^{d} \mathrm{d}\phi_i \\
&= \frac{\int_0^{\pi} \cos^2(\phi_1) \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i}{\int_0^{\pi} \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \left(\prod_{i=1}^{d-2} \sin^{d-1-i}(\phi_i)\right) \prod_{i=1}^{d} \mathrm{d}\phi_i \\
&= \frac{\int_0^{\pi} \cos^2(\phi_1) \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i}{\int_0^{\pi} \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i} \int_{\mathbb{S}^{d-1}} \nu(\mathrm{d}\boldsymbol{\xi}) \\
&= \frac{\int_0^{\pi} \cos^2(\phi_1) \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i}{\int_0^{\pi} \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_i} dV_d. \qquad \text{(A.155)}
\end{aligned}$$

By applying the integration by parts twice to the numerator from the above expression, we can obtain

$$\int_0^{\pi} \cos^2(\phi_1) \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_1 = \frac{1}{d-1} \int_0^{\pi} \sin^{d}(\phi_1)\, \mathrm{d}\phi_1.$$

In addition, using the trigonometric integration formulas, we can show that

$$\frac{\int_0^{\pi} \sin^{d}(\phi_1)\, \mathrm{d}\phi_1}{\int_0^{\pi} \sin^{d-2}(\phi_1)\, \mathrm{d}\phi_1} = \frac{d-1}{d},$$

which along with (A.153) and (A.155) leads to

$$\int_{\mathbb{S}^{d-1}} \boldsymbol{\xi}^T A\, \boldsymbol{\xi}\, \nu(\mathrm{d}\boldsymbol{\xi}) = \mathrm{tr}(A) V_d$$

for the case of $d = 2$. Also, it is easy to see that the same formula holds for the case of $d = 1$ by (A.154). This concludes the proof of Lemma 11.

F.2 Lemma 12 and its proof

Let us define $\zeta_{i,s}$ as the indicator function for the event that $\mathbf{X}_i$ is the 1NN of $\mathbf{x}$ among $\{\mathbf{X}_1, \cdots, \mathbf{X}_s\}$. We provide in Lemma 12 below a list of properties for these indicator functions $\zeta_{i,s}$.

Lemma 12. *The indicator functions $\zeta_{i,s}$ satisfy that*

1) *For any $i \neq j$, we have $\zeta_{i,s}\zeta_{j,s} = 0$ with probability one;*
2) *$\sum_{i=1}^{s} \zeta_{i,s} = 1$;*
3) *$\mathbb{E}[\zeta_{i,s}] = s^{-1}$;*
4) *$\mathbb{E}_{2:s}[\zeta_{1,s}] = \{1-\varphi(B(\mathbf{x}, \|\mathbf{X}_1-\mathbf{x}\|))\}^{s-1}$, where $\mathbb{E}_{i:s}$ denotes the expectation with respect to $\{\mathbf{Z}_i, \mathbf{Z}_{i+1}, \cdots, \mathbf{Z}_s\}$.*

The proof of Lemma 12 involves some standard calculations and thus we omit it here for simplicity. Let us make some remarks on $\mathbb{E}_{2:s}[\zeta_{1,s}]$ that can be regarded as a function of $\mathbf{X}_1$. The last property in Lemma 12 above shows that $\mathbb{E}_{2:s}[\zeta_{1,s}]$ vanishes asymptotically as s tends to infinity, unless $\mathbf{X}_1$ is equal to $\mathbf{x}$. Moreover, we see that

$$\mathbb{E}_1[\mathbb{E}_{2:s}[\zeta_{1,s}]] = s^{-1}.$$

These two facts suggest that $\mathbb{E}_{2:s}[\zeta_{1,s}]$ tends to approximate the Dirac delta function at $\mathbf{x}$, which will be established formally in Lemma 13 in Section F.3.

F.3 Lemma 13 and its proof

Lemma 13. *For any L^1 function f that is continuous at $\mathbf{x}$, it holds that*

$$\lim_{s\to\infty} \mathbb{E}_1[f(\mathbf{X}_1)s\mathbb{E}_{2:s}[\zeta_{1,s}]] = f(\mathbf{x}). \tag{A.156}$$

Proof. We will show that the absolute difference $|\mathbb{E}_1[f(\mathbf{X}_1)s\mathbb{E}_{2:s}[\zeta_{1,s}]] - f(\mathbf{x})|$ converges to zero as $s \to \infty$. By property 3) in Lemma 12 in Section F.2, we have

$$\mathbb{E}_1[s\mathbb{E}_{2:s}[\zeta_{1,s}]] = 1.$$

Thus, we can deduce that

$$\begin{aligned} |\mathbb{E}_1[f(\mathbf{X}_1)s\mathbb{E}_{2:s}[\zeta_{1,s}]] - f(\mathbf{x})| &= |\mathbb{E}_1[(f(\mathbf{X}_1) - f(\mathbf{x}))\, s\mathbb{E}_{2:s}[\zeta_{1,s}]]| \\ &\leq \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|s\mathbb{E}_{2:s}[\zeta_{1,s}]]. \end{aligned} \tag{A.157}$$

Let $\epsilon > 0$ be arbitrarily given. By the continuity of function f at point $\mathbf{x}$, there exists a neighborhood $B(\mathbf{x}, \delta)$ of $\mathbf{x}$ with some $\delta > 0$ such that

$$|f(\mathbf{X}_1) - f(\mathbf{x})| < \epsilon$$

for all $\mathbf{X}_1 \in B(\mathbf{x}, \delta)$. We will decompose the above expectation in (A.157) into two parts: one inside and the other outside of $B(\mathbf{x}, \delta)$ as

$$\begin{aligned} \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|s\mathbb{E}_{2:s}[\zeta_{1,s}]] &= \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|s\mathbb{E}_{2:s}[\zeta_{1,s}]\mathbb{1}_{B(\mathbf{x},\delta)}(\mathbf{X}_1)] \\ &\quad + \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|s\mathbb{E}_{2:s}[\zeta_{1,s}]\mathbb{1}_{B^c(\mathbf{x},\delta)}(\mathbf{X}_1)], \end{aligned} \tag{A.158}$$

where the superscript c stands for set complement in $\mathbb{R}^d$.

The first term on the right-hand side of (A.158) is bounded by ϵ since

$$\begin{aligned} \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|s\mathbb{E}_{2:s}[\zeta_{1,s}]\mathbb{1}_{B(\mathbf{x},\delta)}(\mathbf{X}_1)] &\leq \mathbb{E}_1[\epsilon s\mathbb{E}_{2:s}[\zeta_{1,s}]\mathbb{1}_{B(\mathbf{x},\delta)}(\mathbf{X}_1)] \\ &\leq \mathbb{E}_1[\epsilon s\mathbb{E}_{2:s}[\zeta_{1,s}]] = \epsilon. \end{aligned} \tag{A.159}$$

To bound the second term on the right-hand side of (A.158), observe that

$$B(\mathbf{x}, \delta) \subset B(\mathbf{x}, \|\mathbf{X}_1 - \mathbf{x}\|)$$

when $\mathbf{X}_1 \in B^c(\mathbf{x}, \delta)$. Then an application of Lemma 12 gives

$$\mathbb{E}_{2:s}[\zeta_{1,s}] \leq (1 - \varphi(B(\mathbf{x}, \delta)))^{s-1}$$

when $\mathbf{X}_1 \in B^c(\mathbf{x}, \delta)$. Thus, we can deduce that

$$\begin{aligned}
&\mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})| s \mathbb{E}_{2:s}[\zeta_{1,s}] \mathbb{1}_{B^c(\mathbf{x},\delta)}(\mathbf{X}_1)] \\
&\leq \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})| s(1 - \varphi(B(\mathbf{x}, \delta)))^{s-1} \mathbb{1}_{B^c(\mathbf{x},\delta)}(\mathbf{X}_1)] \\
&\leq s(1 - \varphi(B(\mathbf{x}, \delta)))^{s-1} \mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})|] \\
&\leq s(1 - \varphi(B(\mathbf{x}, \delta)))^{s-1} \left(\|f\|_{L^1} + f(\mathbf{x})\right),
\end{aligned} \tag{A.160}$$

where $\|\cdot\|_{L^1}$ denotes the L^1-norm of a given function.

Finally, we see that the right-hand side of the last equation in (A.160) tends to 0 as $s \to \infty$. Therefore, for large enough s, the quantity

$$\mathbb{E}_1[|f(\mathbf{X}_1) - f(\mathbf{x})| s \mathbb{E}_{2:s}[\zeta_{1,s}] \mathbb{1}_{B^c(\mathbf{x},\delta)}(\mathbf{X}_1)]$$

can be bounded from above by 2ϵ. Since the choice of $\epsilon > 0$ is arbitrary, combining such upper bound, (A.157), (A.158), and (A.159) yields the desired limit in (A.156) as $s \to \infty$. This completes the proof of Lemma 13.