The Annals of Applied Probability
2024, Vol. 34, No. 6, 5172–5210
https://doi.org/10.1214/24-AAP2088

AZADKIA–CHATTERJEE'S CORRELATION COEFFICIENT ADAPTS TO MANIFOLD DATA

BY FANG HAN[1,a] AND ZHIHAN HUANG[2,b]

[1]*Department of Statistics, University of Washington,* [a]*fanghan@uw.edu*
[2]*Department of Statistics and Data Science, University of Pennsylvania,* [b]*zhihanh@wharton.upenn.edu*

In their seminal work, Azadkia and Chatterjee (*Ann. Statist.* **49** (2021) 3070–3102) initiated graph-based methods for measuring variable dependence strength. By appealing to nearest neighbor graphs based on the Euclidean metric, they gave an elegant solution to a problem of Rényi (*Acta Math. Acad. Sci. Hung.* **10** (1959) 441–451). This idea was later developed in Deb, Ghosal and Sen (2020) (https://arxiv.org/abs/2010.01768) and the authors there proved that, quite interestingly, Azadkia and Chatterjee's correlation coefficient can automatically adapt to the manifold structure of the data. This paper furthers their study in terms of calculating the statistic's limiting variance under independence—showing that it only depends on the manifold dimension—and extending this distribution-free property to a class of metrics beyond the Euclidean.

1. Introduction. Consider $\boldsymbol{X} \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ to be two random variables defined over the same probability space with fixed and continuous joint distribution function $F_{\boldsymbol{X},Y}$ and marginal distributions $F_{\boldsymbol{X}}$ and F_Y, respectively. Let $(\boldsymbol{X}_1, Y_1), \dots, (\boldsymbol{X}_n, Y_n)$ be n independent copies of $(\boldsymbol{X}, Y)$, R_i be the rank of Y_i among $\{Y_1, \dots, Y_n\}$ and $N(i)$ index the nearest neighbor (NN) of $\boldsymbol{X}_i$ among $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$, measured using the Euclidean metric. Built on an earlier work of Chatterjee (2021), Azadkia and Chatterjee (2021) introduced the following graph-based correlation coefficient:

$$\xi_n = \xi_n\big(\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n\big) := \frac{6}{n^2-1}\sum_{i=1}^n \min\{R_i, R_{N(i)}\} - \frac{2n+1}{n-1}.$$

This correlation coefficient was shown in Azadkia and Chatterjee ((2021), Theorem 2.2), to converge strongly to a population dependence measure that was first introduced in Dette, Siburg and Stoimenov (2013),

$$\xi = \xi(\boldsymbol{X}, Y) := \frac{\int \operatorname{Var}\{\mathrm{E}[\mathbb{1}(Y \ge t) | \boldsymbol{X}]\}\, \mathrm{d}F_Y(t)}{\int \operatorname{Var}\{\mathbb{1}(Y \ge t)\}\, \mathrm{d}F_Y(t)}.$$

Dette, Siburg and Stoimenov (2013)'s dependence measure satisfies some of the most desirable properties discussed in Rényi (1959) including, in particular, the following three:

(1) $\xi \in [0, 1]$;
(2) $\xi = 0$ if and only if $\boldsymbol{X}$ is independent of Y;
(3) $\xi = 1$ if and only if Y is a measurable function of $\boldsymbol{X}$ almost surely.

Azadkia and Chatterjee thus outlined an elegant approach to measuring the dependence strength between $\boldsymbol{X}$ and Y, resolving many long-standing issues that surround Rényi's criteria as were recently discussed by Professor Peter Bickel (Bickel (2022)).

Received May 2024.
MSC2020 subject classifications. Primary 60F05; secondary 62G10.
Key words and phrases. Manifold, graph-based methods, dependence measure, nearest neighbor graphs, adaptivity.

The authors of this paper are interested in ξ_n's adaptivity to the manifold structure of the data, a problem that has received much interest in the NN literature (Kpotufe (2011), Kpotufe and Garg (2013), Levina and Bickel (2004)). To this end, our focus is on the limiting null distribution of ξ_n, that is, its limiting distribution under independence between $\boldsymbol{X}$ and Y. Such a result, if derived, would immediately give rise to a statistical test of the following null hypothesis:

$$H_0: \quad \boldsymbol{X} \text{ (supported on a smooth manifold) is independent of } Y.$$

Below is the main result of this paper.

THEOREM 1.1 (Central limit theorem of ξ_n for manifold data). *Let $Y \in \mathbb{R}$ be independent of $\boldsymbol{X} \in \mathbb{R}^d$ and let $F_{\boldsymbol{X},Y}$ be fixed and continuous. Further assume that the following two conditions hold*:

(i) $\boldsymbol{X} \in \mathcal{M}$, *where $\mathcal{M}$ is an m-dimensional C^∞ manifold in $\mathbb{R}^d$ with $m \le d$*;

(ii) *the law of $\boldsymbol{X}$ is absolutely continuous with respect to $\mathcal{H}^m \llcorner \mathcal{M}$, the restricted m-dimensional Hausdorff measure in $\mathbb{R}^d$ on $\mathcal{M}$.*

We then have, as $n \to \infty$,

$$\sqrt{n}\xi_n \text{ converges to } N\Big(0, \frac{2}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m\Big) \text{ in distribution},$$

where for any integer $m \ge 1$,

$$\mathfrak{q}_m := \Big\{2 - I_{3/4}\Big(\frac{m+1}{2}, \frac{1}{2}\Big)\Big\}^{-1}, \qquad I_x(a,b) := \frac{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t},$$

$$\mathfrak{o}_m := \iint_{\Gamma_{m;2}} \exp\big[-\lambda\big\{B\big(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|\big) \cup B\big(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|\big)\big\}\big]\,\mathrm{d}(\boldsymbol{w}_1, \boldsymbol{w}_2),$$

$$\Gamma_{m;2} := \big\{(\boldsymbol{w}_1, \boldsymbol{w}_2) \in \big(\mathbb{R}^m\big)^2 : \max\big(\|\boldsymbol{w}_1\|, \|\boldsymbol{w}_2\|\big) < \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|\big\},$$

$B(\boldsymbol{w}_1, r)$ represents the ball of radius r and center $\boldsymbol{w}_1$, $\|\cdot\|$ is the Euclidean norm, and $\lambda(\cdot)$ is the Lebesgue measure.

REMARK 1.1. Notably speaking, the authors of Azadkia and Chatterjee (2021) introduced two statistics: one aimed at measuring conditional correlation, and the other at measuring marginal correlation. This was indicated in Azadkia and Chatterjee ((2021), before Theorem 2.2). In this paper, we are specifically concerned with the latter statistic.

REMARK 1.2. In Theorem 1.1, we assume a constant global dimension of $\mathcal{M}$. When the data structure is more complex, the dimension may differ between different (connected or not) components of $\mathcal{M}$. In such cases, the value of ξ_n's asymptotic variance can be derived analogously as a mixture distribution of each part corresponding to one component of $\mathcal{M}$.

The essence of Theorem 1.1 is the following result, which calculates the limiting null variance of ξ_n.

THEOREM 1.2. *Suppose that the conditions in Theorem 1.1 hold. We then have, as n goes to ∞,*

$$\mathrm{Var}(\sqrt{n}\xi_n) \text{ converges to } \frac{2}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m.$$

TABLE 1
The first 10 *values of* $\mathfrak{q}_m$ *and* $\mathfrak{o}_m$

m	1	2	3	4	5	6	7	8	9	10
$\mathfrak{q}_m$	0.67	0.62	0.59	0.57	0.56	0.55	0.54	0.53	0.53	0.52
$\mathfrak{o}_m$	0.49	0.63	0.71	0.76	0.79	0.84	0.86	0.90	0.98	1.00

As we will explain later in Section 1.2, the terms $\mathfrak{q}_m$ and $\mathfrak{o}_m$ count the averaged numbers of nearest neighbor pairs and triples, respectively. The first ten $\mathfrak{q}_m$ and $\mathfrak{o}_m$ were shown in Table 1 and some basic properties are listed below.

LEMMA 1.1. *The following hold true*:

(a) $\mathfrak{q}_m \in (\frac{1}{2}, \frac{2}{3}]$ *is strictly decreasing as m increases*;
(b) $\sup_m \mathfrak{o}_m < 2$ *and* $\limsup_m \mathfrak{o}_m \le 1$.

1.1. *Literature review.* Theorem 1.1 builds a bridge between two statistical fields, the study of graph-based correlation coefficients and the study of nearest neighbor methods and their adaptivity to manifold data.

On one hand, since the pioneering work of Chatterjee (2021) and Azadkia and Chatterjee (2021), the study of graph-based correlation coefficients has quickly attracted attention; a literature is being built up rapidly and includes, among many others, Cao and Bickel (2020), Shi, Drton and Han (2022), Gamboa et al. (2022), Deb, Ghosal and Sen (2020), Huang, Deb and Sen (2022), Auddy, Deb and Nandy (2024), Shi, Drton and Han (2024), Lin and Han (2023), Fuchs (2024), Azadkia, Taeb and Bühlmann (2021), Griessenberger, Junker and Trutschnig (2022), Strothmann, Dette and Siburg (2024), Lin and Han (2022), Zhang (2023), Bickel (2022) and Chatterjee and Vidyasagar (2022).

In the following we list three existing results that are most relevant to Theorem 1.1. Readers of more interest are referred to Han (2021) and Lin and Han ((2022), Section 1.1) for a slightly more complete review.

(1) Deb, Ghosal and Sen (2020) studied a general class of graph-based correlation coefficients, to which ξ_n belongs. Their Corollary 5.1 examined the *convergence rate* for ξ_n to ξ, illustrating an interesting interplay between the intrinsic dimension of $\boldsymbol{X}$ and the smoothness of some conditional expectation functions relating Y to $\boldsymbol{X}$. They revealed, for the first time, the adaptation of graph-based correlation coefficients to the manifold structure of $\boldsymbol{X}$.

(2) Built on the work of Deb, Ghosal and Sen (2020), Shi, Drton and Han ((2024), Theorem 3.1(ii)) established a *central limit theorem (CLT)* for ξ_n under (a) independence between Y and $\boldsymbol{X}$; (b) *absolute continuity* (with respect to the Lebesgue measure) of $F_{\boldsymbol{X}}$. Under these conditions, they showed

$$\sqrt{n}\xi_n \text{ converges to } N\left(0, \frac{2}{5} + \frac{2}{5}\mathfrak{q}_d + \frac{4}{5}\mathfrak{o}_d\right) \text{ in distribution.}$$

(3) In a more recent preprint, Lin and Han ((2022), Theorem 1.1) established a CLT for ξ_n while removing both independence and absolute continuity assumptions required in Shi, Drton and Han (2024). In particular, they showed that as long as (a) $F_{\boldsymbol{X},Y}$ is fixed and continuous and (b) Y is not almost surely a measurable function of $\boldsymbol{X}$, it holds true that

$$(\xi_n - \mathrm{E}\xi_n)/\sqrt{\mathrm{Var}(\xi_n)} \text{ converges to } N(0, 1) \text{ in distribution.}$$

Theorem 1.1 can thus be viewed as a descendent of the above three results:

(1) compared to Deb, Ghosal and Sen (2020), it established a weak convergence instead of a point estimation type result;

(2) compared to Shi, Drton and Han (2024), it removed the absolute continuity assumption required therein;

(3) compared to Lin and Han ((2022), Theorem 1.1), Theorem 1.1 calculated the explicit value of the asymptotic variance.

On the other hand, in practice many data are believed to be structured, that is, they are embedded in a space that is of a much higher dimension than necessary (Amelunxen et al. (2014), Levina and Bickel (2004)). Local methods, especially the NN-based ones, are long believed to be suitable for analyzing such data, capable of automatically adapting to the data structure (Clarkson (2006), Kpotufe (2011), Kpotufe (2017), Kpotufe and Garg (2013)). We believe Theorem 1.1 also bears potential to contribute to this line of the research. In particular:

(1) As we shall show in Section 1.2, an essence of Theorem 1.1 is to calculate the averaged numbers of nearest neighbor pairs and triples; they are thus monitoring the stochastic structure of an NN graph (NNG) when the data are distributed over a manifold.

(2) Theorem 1.1 is also, to our knowledge, the *first* weak convergence type results for tracking the statistical behavior of a NN-based functional over a manifold-supported probability space.

From a technical standpoint, we utilized Lin and Han ((2022), Theorem 1.1) in our proof to demonstrate the asymptotic normality of ξ_n. However, in our perspective, this step of establishing a central limit theorem is not the most crucial aspect of Theorem 1.1. The primary focus is on the fact that the null variance limit of ξ_n can adapt automatically to the manifold dimension of $\boldsymbol{X}$. Proving this assertion required several novel calculations, which are presented in Lemmas 1.4 and 1.5 ahead. These lemmas represent genuinely new contributions.

1.2. *Proof sketch.* We first introduce some auxiliary results on the NNGs and the manifold. Recall that $[\boldsymbol{X}_i]_{i=1}^n$ comprise n independent copies of a random vector $\boldsymbol{X} \in \mathbb{R}^d$ from an unknown distribution function $F_{\boldsymbol{X}}$. Let $\mathcal{G}_n$ be the associated directed NNG with vertex set $\{1, \dots, n\}$ and edge set $\mathcal{E}(\mathcal{G}_n)$; here an edge $\{i \to j\} \in \mathcal{E}(\mathcal{G}_n)$ means $\boldsymbol{X}_j$ is the NN of $\boldsymbol{X}_i$.

We are interested in manifold data; more precisely, we are interested in such random vector $\boldsymbol{X}$ that is supported on $\mathcal{M}$, a smooth submanifold of $\mathbb{R}^d$ with manifold dimension $m \le d$. The following concepts are from Lee (2013).

DEFINITION 1.1. Let $\mathcal{M}$ be an m-dimensional smooth manifold. A *coordinate chart*, abbreviated as a *chart*, on $\mathcal{M}$ is a pair (U, ψ), where U is an open set of $\mathcal{M}$ and $\psi : U \to V$ is a homeomorphism from U to an open subset $V = \psi(U) \subset \mathbb{R}^m$.

DEFINITION 1.2. Given a smooth manifold $\mathcal{M}$ and a chart (U, ψ) of $\mathcal{M}$, U is called a *coordinate neighborhood* of each point $w \in U$.

Concerning any point $\boldsymbol{x} \in \mathcal{M}$, one can find a chart of $\mathcal{M}$ with coordinate neighborhood $U (= U_{\boldsymbol{x}})$ and corresponding homeomorphism $\psi (= \psi_{\boldsymbol{x}})$.[1] With this notion, we give an assumption on the distribution of $\boldsymbol{X}$ that will be shown to be an alternative to Theorem 1.1(ii). In the following, the law of $\boldsymbol{X}$ is denoted by μ and the restriction of μ to a set U is denoted by μ_U.

[1] There are, of course, many such neighborhoods and homeomorphisms circling $\boldsymbol{x}$; in the sequel we simply pick one of them.

ASSUMPTION 1.1 (Distribution assumption). The positive measure μ satisfies the following condition: for any point $\boldsymbol{x} \in \mathcal{M}$ and any chart (U, ψ) such that U is a coordinate neighborhood of $\boldsymbol{x}$ and $\psi : U \to V \subset \mathbb{R}^m$, the restricted pushforward measure $\psi_* \mu_U$ is absolutely continuous with respect to the Lebesgue measure $\lambda(\cdot)$ on V.

Assumption 1.1 yields an alternate description of the data generating process and will appear to be useful in the following proofs; the next lemma shows that it is equivalent to the assumption of Theorem 1.1(ii).

LEMMA 1.2 (Alternative distribution assumption). *As $\mathcal{M}$ satisfies the assumption of Theorem* 1.1(i), *Assumption* 1.1 *is equivalent to the assumption of Theorem* 1.1(ii).

We then move on to study the stochastic behavior of the NNG $\mathcal{G}_n$ as $\mathcal{M}$ satisfies Theorem 1.1(ii) and $\boldsymbol{X}$ satisfies Assumption 1.1. Since in the expression of ξ_n, one of the rank terms is indexed by its NNG, the asymptotic distribution of ξ_n has a connection with the properties of the NNG. In the following we introduce a series of lemmas on this topic. To begin with, Lemma 1.3 is a well-known result by Bickel and Breiman (1983) on the maximum number of nearest neighbors.

LEMMA 1.3 (Maximum degree in nearest neighbor graphs). *There is an upper bound for the degree of any point in NNGs. More specifically, let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be any collection of n distinct points in $\mathbb{R}^d$. Then there exists a constant $\mathfrak{C}_d$ depending only on the dimension d such that $\boldsymbol{x}_1$ is the nearest neighbor of at most $\mathfrak{C}_d$ points from $\{\boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$.*

The next two lemmas draw the average numbers of some specific structures in an NNG. They are extensions of conclusions in Devroye (1988) and Henze (1987). We first focus on the number of loops between two vertices, which we call a *nearest neighbor pair* in $\mathcal{G}_n$.

LEMMA 1.4 (Expected number of nearest-neighbor pairs). *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ with $\mathcal{M}$ and μ satisfying the assumption of Theorem* 1.1(i) *and Assumption* 1.1, *respectively. We then have, as $n \to \infty$,*

$$\mathrm{E}\Big(\frac{1}{n}\#\{(i,j) \text{ distinct} : i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\}\Big) \to \frac{V_m}{U_m} := \mathfrak{q}_m,$$

where V_m is the volume of the unit ball in $\mathbb{R}^m$, and U_m is the volume of the union of two unit balls in $\mathbb{R}^m$ whose centers are a unit distance apart.

We then turn to another structure in $\mathcal{E}(\mathcal{G}_n)$ that monitors those parent vertices that share the same child vertex. We call them a *nearest neighbor triple* in $\mathcal{G}_n$.

LEMMA 1.5. *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ with $\mathcal{M}$ and μ satisfying the assumption of Theorem* 1.1(i) *and Assumption* 1.1, *respectively. We then have, as $n \to \infty$,*

$$\mathrm{E}\Big(\frac{1}{n}\#\{(i,j,k) \text{ distinct} : i \to k, j \to k \in \mathcal{E}(\mathcal{G}_n)\}\Big) \to \mathfrak{o}_m.$$

Get back to the data $(\boldsymbol{X}_i, Y_i)$, $i = 1, \dots, n$. Let the assumptions in Theorem 1.1 hold and we construct $\mathcal{G}_n$ on the manifold data $[\boldsymbol{X}_i]_{i=1}^n$. Recall that m denotes the manifold dimension and $\mathfrak{q}_m$, $\mathfrak{o}_m$ are positive constants depending only on m. With the lemmas presented above, it is then straightforward to derive the limiting variance following the proof of Shi, Drton and Han ((2024), Theorem 3.1), outlined in Theorem 1.2.

Last, when Y is independent of $\boldsymbol{X}$, using Theorem 1.1 in Lin and Han (2022), we have

$$\frac{\xi_n}{\sqrt{\mathrm{Var}(\xi_n)}} \xrightarrow{d} N(0,1).$$

Combining the above result with Theorem 1.2 then yields Theorem 1.1.

1.3. *Extension to non-Euclidean metrics.* This section extends the above results to metrics beyond the Euclidean. Let us consider a general kernel function K over some support $\mathcal{X} \times \mathcal{X}$ that induces a kernel metric (Schölkopf (2000)),

$$D(\boldsymbol{X}_i, \boldsymbol{X}_j) := \sqrt{\frac{K(\boldsymbol{X}_i, \boldsymbol{X}_i) + K(\boldsymbol{X}_j, \boldsymbol{X}_j)}{2} - K(\boldsymbol{X}_i, \boldsymbol{X}_j)}. \tag{1.1}$$

To make $D(\boldsymbol{X}_i, \boldsymbol{X}_j)$ well defined, we require some characteristics properties of the kernel function. Two specific types of such structures are introduced below.

1.3.1. *Difference-based kernel metrics.* First, consider the difference-based kernel metric, that is,

$$K(\boldsymbol{x}, \boldsymbol{y}) = -f(\boldsymbol{x} - \boldsymbol{y}),$$

for some function $f : \mathcal{X} \to \mathbb{R}$. The function f is regulated below so that the corresponding K can indeed induce a metric.

ASSUMPTION 1.2. Assume $f(\boldsymbol{0}) = 0$ and $f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$. Further, assume that $f(\boldsymbol{x})$ is C^2 in the neighborhood of $\boldsymbol{0}$, $\nabla^2 f(\boldsymbol{0})$ is positive definite, and for some $\varepsilon_0 > 0$ and any $0 < \varepsilon < \varepsilon_0$, there exists a $\delta > 0$, $f(\boldsymbol{x}) > \varepsilon$ for any $\|\boldsymbol{x}\| > \delta$. In addition, we require f to induce a nondegenerate metric, that is, the number of points in $\mathcal{X} \subset \mathbb{R}^d$ with the same NN can be upper bounded by some constant $C_{d,K}$ that only depend on the dimension d and the chosen kernel K.

The corresponding kernel metric can then be written as

$$D(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{-K(\boldsymbol{x}, \boldsymbol{y})} = \sqrt{f(\boldsymbol{x} - \boldsymbol{y})}. \tag{1.2}$$

In particular, the first part of Assumption 1.2 requires the kernel function to be smooth and characteristic, which is reasonable for kernels used for expressing correlation. The second part (nondegeneracy condition) holds automatically for the Euclidean metric and in Bickel and Breiman ((1983), Page 211), it is further shown that this condition holds for any norm such that the unit sphere under the norm is compact. Assumption 1.2 holds true for any kernel with sufficient smoothness to induce a norm with compact unit sphere in $\mathcal{X}$. It also holds for the (centered) Gaussian-type kernel (e.g., $f(\boldsymbol{x}) \propto 1 - \exp(\alpha\|\boldsymbol{x}\|^2), \alpha < 0$), inverse multiquadratic (IMQ) kernel (e.g., $f(\boldsymbol{x}) \propto \gamma^\beta - (\gamma + \|\boldsymbol{x}\|^2)^\beta$, $\gamma > 0$, $\beta < 0$) and their variants.

The next theorem gives the first generalization of Theorem 1.1 to such $\xi_n = \xi_{n,D}$ whose NNs are determined using the aforementioned metric D.

THEOREM 1.3. *Suppose that the conditions in Theorem 1.1 and Assumption 1.2 hold, and $\xi_n = \xi_{n,D}$ is calculated with the NNs of $\boldsymbol{X}_i$'s decided using the metric D in (1.2). We then have, as $n \to \infty$,*

$$\mathrm{Var}(\sqrt{n}\xi_n) \text{ converges to } \frac{2}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m,$$

and $\xi_n/\sqrt{\mathrm{Var}(\xi_n)} \xrightarrow{d} N(0,1)$.

It is worth noticing that the above limiting variance of ξ_n is identical to that under the Euclidean case, and the value does not depend on the choice of the kernel K. This observation enables us to consider flexible kernels and an independence test built on such difference-based kernels is direct.

REMARK 1.3. The distribution-free property appears because of the local isotropy property, that is, the studied kernel metric can be locally well approximated by the Euclidean metric (up to linear transformations), which is rotationally invariant. It offsets the anisotropy brought by the data generating distribution as well as the manifold structure. On the contrary, the Assumption 1.2 generally does not hold for L^p norms; when a point moves over a manifold, the shape of the intersection of its L^p neighborhood and the manifold can change dramatically.

1.3.2. *Geodesic distance and general kernel metrics.* For a fixed kernel K, there could exist a feature map $\varphi : \mathcal{H} \to \mathcal{H}$ such that $K(\boldsymbol{x}, \boldsymbol{y}) = \langle \varphi(\boldsymbol{x}), \varphi(\boldsymbol{y}) \rangle_{\mathcal{H}}$, where $\mathcal{H}$ is the reproducing kernel Hilbert space (RKHS) and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on it (Berlinet and Thomas-Agnan (2004)).

It is direct to define a kernel metric in the feature space, that is, $D(\boldsymbol{x}, \boldsymbol{y}) = \|\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{y})\|_{\mathcal{H}}$, and the following proposition shows that it coincides with the definition in equation (1.1). The proof can be found in Schölkopf (2000).

PROPOSITION 1.1. *Consider the real-valued kernel function K to be conditionally positive definite* Schölkopf ((2000), *Definition* 2) *such that $K(\boldsymbol{x}, \boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathcal{X}$. Then there exists a feature map φ satisfying*

$$D(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{-K(\boldsymbol{x}, \boldsymbol{y})} = \|\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{y})\|_{\mathcal{H}},$$

where $\|\cdot\|_{\mathcal{H}}$ is a semi-metric in the RKHS $\mathcal{H}$. If we additionally assume $K(\boldsymbol{x}, \boldsymbol{y}) < 0$ for all $\boldsymbol{x} \neq \boldsymbol{y}$, then $\|\cdot\|_{\mathcal{H}}$ is a metric.

In the Riemannian manifold context, the feature map has a close connection to the geodesic distance. In particular, denote the geodesic distance between arbitrary $\boldsymbol{x}$ and $\boldsymbol{y}$ over the manifold $\mathcal{M}$ as $d_g(\boldsymbol{x}, \boldsymbol{y})$. The geodesic distance can be locally captured by a kernel, in the following sense: For any fixed $\boldsymbol{x} \in \mathcal{M}$, there exists a coordinate chart (U, ψ) equipped with the *natural parameterization* such that, if we choose $\varphi := \psi^{-1}$, then for any $\boldsymbol{y} \in U$,

$$d_g(\boldsymbol{x}, \boldsymbol{y}) = \|\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{y})\|, \tag{1.3}$$

where $\|\cdot\|$ is the Euclidean distance. A more refined discussion on the relationship between geodesic distance and the exponential map of a manifold can be found in Lee ((2013), Chapter 13, Page 337).

With the relationship shown in (1.3), we are ready to adjust the kernel metric framework to the analysis of the correlation coefficient under the geodesic distance. Before presenting the result, we again make some regularity assumptions on the data generating structure.

ASSUMPTION 1.3. Suppose $m \geq 2$. For any $\boldsymbol{x} \in \mathcal{M}$, the number of points whose NN (measured using the geodesic metric) is $\boldsymbol{x}$ is bounded by some constant $C_{\mathcal{M}}$ that only depends on the manifold $\mathcal{M}$.

Assumption 1.3 is nontrivial and, for unbounded manifolds, may no longer hold. However, it is indeed true for most manifolds with a global smoothness property, for example, the linear space $\mathbb{R}^m$ and the sphere $\mathbb{S}^m$ in $\mathbb{R}^d$. Specifically, we have the following proposition in case the manifold is compact.

PROPOSITION 1.2. *Assumption 1.3 holds if $\mathcal{M}$ is compact in the metric space $(\mathcal{M}, d_g)$.*

With the above assumptions, the next theorem then gives the generalization of Theorem 1.1 to the geodesic metric case.

THEOREM 1.4. *Suppose that the conditions in Theorem 1.1 and Assumption 1.3 hold. Furthermore, let $\xi_n = \xi_{n,D}$ be calculated with the NNs of $\boldsymbol{X}_i$'s decided using the geodesic distance d_g induced by the Euclidean metric. We then have, as $n \to \infty$,*

$$\operatorname{Var}(\sqrt{n}\xi_n) \text{ converges to } \frac{2}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m,$$

and $\xi_n / \sqrt{\operatorname{Var}(\xi_n)} \xrightarrow{d} N(0, 1)$.

Again, the limiting variance in this metric space is distribution-free and only depends on the manifold dimension.

1.4. *Some finite-sample studies.* This section contains some finite-sample simulation results to examine the independence test powers, comparing the performance of ξ_n to that of distance correlation (Székely, Rizzo and Bakirov (2007)). We examine the sizes and powers of the proposed tests when the data are supported on a manifold satisfying theorem assumptions and with manifold dimension *known to us*.[2] Power comparisons are carried out with sample size $n = 100$. In each case, 5000 simulations are used to calculate the empirical size/power. For simplicity, we only study those ξ_n's calculated based on the Euclidean metric.

We first generate the raw data $(Y_i, \boldsymbol{Z}_i)$, $i = 1, \dots, n$. Here $(Y_1, \boldsymbol{Z}_1), \dots, (Y_n, \boldsymbol{Z}_n)$ constitute a sample of points independently drawn from a certain distribution on $\mathbb{R} \times \mathbb{R}^m$. The value of m will change in simulations.

- Case 1 (Gaussian): $(Y, \boldsymbol{Z})$ is Gaussian distributed with mean $\mathbf{0}$ and equi-correlation ρ between Y and each component of $\boldsymbol{Z}$, that is, $(Y, \boldsymbol{Z}) \sim N(\mathbf{0}, \Sigma)$ with

$$\Sigma := \begin{pmatrix} 1 & \rho\mathbf{1}_m^\top \\ \rho\mathbf{1}_m & I_m \end{pmatrix},$$

  where $\mathbf{1}_m := (\underbrace{1, \dots, 1}_{m})^\top$ and I_m represents the m-dimensional identity matrix.

In the following five cases, we set $\boldsymbol{Z} = (Z_1, \dots, Z_m)^\top$, where $Z_j \sim \text{Unif}[-1, 1]$ is independent of each other, and consider an additive model:

$$Y = \rho \sum_{i=1}^{m} f(Z_i) + C\epsilon,$$

where $\epsilon \sim N(0, 1)$ is independent of $\boldsymbol{Z}$. We fix C as a constant in each case to modify the noise intensity.

- Case 2 (Linear): $f(x) = x$, $C = 0.2$;
- Case 3 (Quadratic): $f(x) = x^2$, $C = 0.1$;
- Case 4 (Cosine): $f(x) = \cos(8\pi x)$, $C = 0.1$;
- Case 5 (W-shape): $f(x) = |x + 0.5| I_{\{x<0\}} + |x - 0.5| I_{\{x \geq 0\}}$, $C = 0.025$.

[2] In practice, if the manifold dimension is unknown, one could either estimate it or use permutation to obtain the test threshold.

Regarding each of the five cases, we then conduct the following two types of transformation to obtain the manifold data:

(1) Linear transformation: $\boldsymbol{Z} \mapsto R\boldsymbol{Z} =: \boldsymbol{X}$, where R is a pre-selected $5m$ by m matrix. For each dimension m, we randomly generate R from a standard Gaussian random matrix $(R_{ij})_{5m \times m}$, where all its elements are independent, and for $i = 1, \dots, 5m$, $j = 1, \dots, m$, $R_{ij} \sim N(0, 1)$. The sample of transformed points lies on a m-dimensional linear subspace in $\mathbb{R}^{5m}$.

(2) Manifold transformation: $\boldsymbol{Z} \mapsto M(\boldsymbol{Z}) =: \boldsymbol{X}$, which is a map from $\mathbb{R}^m$ to a pre-specific m-dimensional smooth manifold in $\mathbb{R}^{5m}$. In the following simulations, the mapping takes the following specific forms:

$$M(\boldsymbol{Z}) = \big(M_1(\boldsymbol{Z}), M_2(\boldsymbol{Z}), M_3(\boldsymbol{Z}), M_4(\boldsymbol{Z}), M_5(\boldsymbol{Z})\big)^\top,$$

where

$$\begin{aligned}
M_1(\boldsymbol{Z}) &:= (Z_1, \dots, Z_m) = \boldsymbol{Z}, \\
M_2(\boldsymbol{Z}) &:= \big(Z_1^2, \dots, Z_m^2\big), \\
M_3(\boldsymbol{Z}) &:= \big(\sin(8\pi Z_1), \dots, \sin(8\pi Z_m)\big), \\
M_4(\boldsymbol{Z}) &:= \big(\cos(4\pi Z_1), \dots, \cos(4\pi Z_m)\big), \\
M_5(\boldsymbol{Z}) &:= \big(\exp(Z_1), \dots, \exp(Z_m)\big).
\end{aligned}$$

We perform tests of independence with parameters $m = 1, 2, 3, 5, 10$ and $\rho = 0, 0.05, 0.10, 0.15, 0.20$ for both ξ_n and distance correlation. Nominal level is set to be $\alpha = 0.05$ for all tests. In all the cases, $\rho = 0$ corresponds to the null hypothesis

$$H_0: \quad Y \text{ and } \boldsymbol{X} \text{ are independent},$$

while the rest of the values of ρ yield powers in accordance with different degrees of dependence. The thresholds of ξ_n and distance correlation are determined by Theorem 1.1 and permutation, respectively. Table 2 and Table 3 illustrate test powers for Gaussian (Case 1) with linear and manifold transformation, respectively. Tables 4–11 analogously illustrate test powers for the additive model cases (Case 2–6) with two transformations in sequence.

Three observations are in line.

(i) All the tests considered have empirical sizes close to 0.05, indicating that they are all size valid.

(ii) When the joint distribution of $(Y, \boldsymbol{Z})$ is a multi-dimension normal distribution, the test power increases as m or ρ increases. The distance correlation based tests exhibit higher power compared to ξ_n-based, indicating that the distance correlation could potentially also adapt to the manifold structure of $\boldsymbol{X}$, an interesting phenomenon largely untouched in literature before.

TABLE 2
Case 1, *linear transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.050	0.051	0.048	0.053	0.054	0.053	0.066	0.144	0.271	0.455
$m=2$	0.044	0.048	0.042	0.046	0.052	0.044	0.080	0.208	0.405	0.609
$m=3$	0.050	0.049	0.055	0.056	0.079	0.052	0.074	0.229	0.454	0.766
$m=5$	0.057	0.058	0.049	0.065	0.112	0.041	0.141	0.331	0.604	0.844
$m=10$	0.066	0.060	0.057	0.099	0.215	0.043	0.106	0.420	0.842	0.995

TABLE 3
Case 1, *manifold transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.058	0.051	0.051	0.050	0.059	0.041	0.051	0.114	0.269	0.336
$m=2$	0.058	0.055	0.059	0.058	0.065	0.048	0.088	0.122	0.287	0.548
$m=3$	0.075	0.068	0.065	0.071	0.079	0.060	0.056	0.106	0.448	0.652
$m=5$	0.063	0.061	0.059	0.086	0.115	0.051	0.085	0.159	0.575	0.792
$m=10$	0.065	0.061	0.068	0.082	0.155	0.064	0.115	0.248	0.735	0.956

TABLE 4
Case 2, *linear transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.058	0.065	0.070	0.119	0.268	0.060	0.263	0.767	0.981	1.000
$m=2$	0.052	0.054	0.091	0.275	0.603	0.048	0.368	0.952	0.999	1.000
$m=3$	0.046	0.066	0.153	0.440	0.791	0.047	0.515	0.986	1.000	1.000
$m=5$	0.061	0.065	0.270	0.664	0.913	0.049	0.661	0.997	1.000	1.000
$m=10$	0.053	0.069	0.259	0.528	0.690	0.052	0.613	0.991	1.000	1.000

TABLE 5
Case 2, *manifold transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.053	0.051	0.061	0.118	0.267	0.055	0.199	0.671	0.949	0.998
$m=2$	0.059	0.054	0.086	0.206	0.447	0.046	0.310	0.841	0.995	1.000
$m=3$	0.068	0.069	0.115	0.284	0.539	0.049	0.345	0.937	0.999	1.000
$m=5$	0.077	0.073	0.138	0.317	0.529	0.047	0.477	0.978	1.000	1.000
$m=10$	0.080	0.074	0.127	0.236	0.334	0.047	0.651	0.996	1.000	1.000

TABLE 6
Case 3, *linear transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.045	0.053	0.071	0.134	0.278	0.050	0.072	0.240	0.586	0.914
$m=2$	0.052	0.044	0.080	0.222	0.493	0.049	0.080	0.211	0.499	0.815
$m=3$	0.056	0.048	0.073	0.189	0.402	0.051	0.078	0.196	0.461	0.648
$m=5$	0.054	0.064	0.056	0.046	0.050	0.059	0.099	0.141	0.375	0.532
$m=10$	0.059	0.254	0.476	0.573	0.620	0.042	0.056	0.161	0.327	0.441

TABLE 7
Case 3, *manifold transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.047	0.052	0.068	0.123	0.264	0.055	0.068	0.239	0.514	0.834
$m=2$	0.063	0.058	0.056	0.082	0.141	0.047	0.099	0.247	0.554	0.886
$m=3$	0.068	0.088	0.081	0.069	0.062	0.049	0.096	0.285	0.569	0.799
$m=5$	0.070	0.130	0.184	0.204	0.190	0.040	0.162	0.221	0.647	0.819
$m=10$	0.078	0.291	0.519	0.610	0.673	0.065	0.059	0.271	0.604	0.832

TABLE 8
Case 4, *linear transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.047	0.080	0.436	0.892	0.994	0.050	0.049	0.062	0.052	0.070
$m=2$	0.051	0.044	0.052	0.057	0.059	0.051	0.050	0.048	0.058	0.046
$m=3$	0.055	0.056	0.054	0.051	0.055	0.051	0.044	0.055	0.044	0.043
$m=5$	0.056	0.053	0.053	0.052	0.053	0.058	0.041	0.037	0.041	0.046
$m=10$	0.058	0.054	0.059	0.057	0.055	0.044	0.037	0.081	0.046	0.056

TABLE 9
Case 4, *manifold transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.051	0.073	0.381	0.864	0.994	0.044	0.077	0.170	0.329	0.580
$m=2$	0.055	0.063	0.062	0.089	0.123	0.051	0.073	0.093	0.114	0.157
$m=3$	0.081	0.141	0.166	0.171	0.168	0.055	0.056	0.083	0.073	0.093
$m=5$	0.079	0.295	0.404	0.456	0.474	0.049	0.093	0.070	0.095	0.089
$m=10$	0.085	0.462	0.581	0.621	0.629	0.042	0.082	0.055	0.073	0.067

TABLE 10
Case 5, *linear transformation*

	ξ_n based					Distance correlation based				
	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$	$\rho=0$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.20$
$m=1$	0.048	0.066	0.272	0.721	0.971	0.048	0.068	0.149	0.323	0.585
$m=2$	0.047	0.061	0.334	0.778	0.955	0.051	0.046	0.082	0.106	0.096
$m=3$	0.061	0.055	0.097	0.211	0.311	0.041	0.050	0.055	0.058	0.077
$m=5$	0.052	0.091	0.107	0.111	0.106	0.066	0.069	0.080	0.067	0.064
$m=10$	0.055	0.091	0.121	0.124	0.138	0.040	0.037	0.048	0.041	0.064

TABLE 11
Case 5, *manifold transformation*

	ξ_n based					Distance correlation based				
	$\rho = 0$	$\rho = 0.05$	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0$	$\rho = 0.05$	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$
$m = 1$	0.047	0.062	0.267	0.694	0.956	0.055	0.061	0.167	0.348	0.805
$m = 2$	0.063	0.056	0.093	0.158	0.252	0.047	0.060	0.114	0.183	0.303
$m = 3$	0.068	0.072	0.070	0.070	0.072	0.049	0.054	0.102	0.136	0.244
$m = 5$	0.070	0.090	0.110	0.112	0.110	0.040	0.101	0.150	0.130	0.209
$m = 10$	0.078	0.148	0.186	0.197	0.196	0.065	0.038	0.052	0.126	0.074

(iii) When the function f exhibits oscillatory properties, the test power increases as ρ increases. However, when m increases, in most cases, the power of our proposed tests shows a U-shape, that is, decreasing first and then increasing; yet the distance correlation based tests shows monotonically decreasing power. Thus, for high-dimensional data, our proposed test might be more powerful. When $m = 1$, our proposed test is also more powerful in some cases.

2. Proofs.

Table 12 lists all the symbols used in the following proofs.

2.1. *Proof of Theorem* 1.2.

PROOF OF THEOREM 1.2. Recall that we write $\mathcal{G}_n$ for the random directed nearest neighbor graph (NNG) corresponding to n sample points $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ and use $\mathcal{E}(\mathcal{G}_n)$ to denote the

TABLE 12
Common symbols and notation

$\boldsymbol{X}$	Random vector in $\mathbb{R}^d$ indicating the distribution of samples
$\boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{X}_k \dots$	i.i.d. copies of $\boldsymbol{X}$
Y	Random vector in $\mathbb{R}$
$Y_i, Y_j, Y_k \dots$	i.i.d. copies of Y
$\mathcal{M}$	m-dimensional smooth manifold in the Euclidean space $\mathbb{R}^d$
m	Manifold dimension of $\mathcal{M}$
d	Dimension of ambient space
n	Number of data points
λ	Lebesgue measure
$\\|\cdot\\|$	Euclidean norm
$\mathcal{H}^m$	m-dimensional Hausdorff measure
μ	Probability measure of $\boldsymbol{X}$
$g(\boldsymbol{x})$	Density function of pushforward $\psi_*\mu$ at point $\boldsymbol{x}$
$\mathcal{G}_n$	Directed nearest neighbor graph (NNG) constructed on n sample points $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$
$\mathcal{E}(\mathcal{G}_n)$	Edge set of $\mathcal{G}_n$
N_{total}	Total number of nearest pairs in $\mathcal{G}_n$
$N(\boldsymbol{X}_i)$	Total number of nearest pairs in $\mathcal{G}_n$
$D(\boldsymbol{X}_i)$	Out-degree of $\boldsymbol{X}_i$ in $\mathcal{G}_n$
M_{total}	$\#\{(i, j, k)\ \textit{distinct} : j \to i, k \to i \in \mathcal{E}(\mathcal{G}_n)\}$
$M(\boldsymbol{X}_i)$	$\#\{(j, k)\ \textit{distinct} : j \to i, k \to i \in \mathcal{E}(\mathcal{G}_n)\}$
π	Tangent plane of a given point on $\mathcal{M}$
$U^d(\boldsymbol{x}_1, \boldsymbol{x}_2)$	Union of two balls whose centers are $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and radius are both $\\|\boldsymbol{x}_1 - \boldsymbol{x}_2\\|$ in $\mathbb{R}^d$
$B^k(\boldsymbol{x}, r)$	k-dimensional ball with center $\boldsymbol{x}$ and radius r
V_m	Volume of unit ball in $\mathbb{R}^m$
U_m	Volume of $U^m(0, \rho)$, $\\|\rho\\| = 1$

edge set of $\mathcal{G}_n$. According to Lemmas 1.4 and 1.5,

$$\lim_{n\to\infty} \mathrm{E}\bigg(\frac{1}{n}\#\{(i,j)\ distinct: i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\}\bigg) = \mathfrak{q}_m,$$

$$\lim_{n\to\infty} \mathrm{E}\bigg(\frac{1}{n}\#\{(i,j,k)\ distinct: i \to k, j \to k \in \mathcal{E}(\mathcal{G}_n)\}\bigg) = \mathfrak{o}_m.$$

Pursuing the idea in Shi, Drton and Han (2024), we resort to Hájek representation for calculating ξ_n's asymptotic variance. In Lin and Han (2022), the intermediate statistic $\breve{\xi}_n$ is defined as

$$\breve{\xi}_n := \frac{6n}{n^2-1}\Bigg(\sum_{i=1}^n \min\{F_Y(Y_i), F_Y(Y_{N(i)})\} - \frac{1}{n-1}\sum_{\substack{i,j=1\\ i\neq j}}^n \min\{F_Y(Y_i), F_Y(Y_j)\} + \sum_{i=1}^n g(Y_i)$$
$$+ \frac{1}{n-1}\sum_{\substack{i,j=1\\ i\neq j}}^n E\big[\min\{F_Y(Y_i), F_Y(Y_j)\}|\boldsymbol{X}_i, \boldsymbol{X}_j\big] - E\Bigg[\sum_{i=1}^n g(Y_i)|\boldsymbol{X}_i\Bigg] + \sum_{i=1}^n g_0(\boldsymbol{X}_i)\Bigg),$$

where specific to the case that Y is independent of $\boldsymbol{X}$, we have

$$F_Y(t) := \mathrm{P}(Y \le t),$$
$$G_{\boldsymbol{X}}(t) := \mathrm{P}(Y \ge t|\boldsymbol{X}) = \mathrm{P}(Y \ge t) = 1 - F_Y(t),$$
$$g(t) := \mathrm{Var}_{\boldsymbol{X}}\big[G_{\boldsymbol{X}}(t)\big] = 0,$$
$$g_0(\boldsymbol{x}) := \int \mathrm{E}\big[G_{\boldsymbol{X}}(t)\big]^2 \,\mathrm{d}F_{Y|\boldsymbol{X}=\boldsymbol{x}}(t) = \int \big(1 - F_Y(t)\big)^2 \,\mathrm{d}F_Y(t) = \frac{1}{3}.$$

Leveraging these expressions, $\breve{\xi}_n$ under independence then takes the form

$$\breve{\xi}_n = \frac{6n}{n^2-1}\Bigg(\sum_{i=1}^n \min\{F_Y(Y_i), F_Y(Y_{N(i)})\} - \frac{1}{n-1}\sum_{\substack{i,j=1\\ i\neq j}}^n \min\{F_Y(Y_i), F_Y(Y_j)\}\Bigg) + C_0,$$

where C_0 is a fixed constant.

In Lin and Han (2022), it is proved that

$$\lim_{n\to\infty} n\,\mathrm{Var}[\xi_n - \breve{\xi}_n] = 0.$$

Thus we have

$$\limsup_{n\to\infty}\bigg|\frac{\mathrm{Cov}(\xi_n, \xi_n - \breve{\xi}_n)}{\mathrm{Var}[\xi_n]}\bigg| \le \limsup_{n\to\infty}\bigg(\frac{\mathrm{Var}[\xi_n - \breve{\xi}_n]}{\mathrm{Var}[\xi_n]}\bigg)^{\frac{1}{2}} \le \bigg(\frac{\limsup_{n\to\infty}\mathrm{Var}[\xi_n - \breve{\xi}_n]}{\limsup_{n\to\infty}\mathrm{Var}[\xi_n]}\bigg)^{\frac{1}{2}} = 0.$$

Then we have

$$\lim_{n\to\infty}\frac{n\,\mathrm{Var}[\breve{\xi}_n]}{n\,\mathrm{Var}[\xi_n]} = \lim_{n\to\infty}\frac{\mathrm{Var}[\xi_n - (\xi_n - \breve{\xi}_n)]}{\mathrm{Var}[\xi_n]} = 1. \tag{2.1}$$

Since (2.1) holds, it suffices to calculating the variance of the intermediate statistic $\breve{\xi}_n$, that is, we only have to calculate

$$n\,\mathrm{Var}[\breve{\xi}_n] = \frac{36n^3}{(n^2-1)^2}\mathrm{Var}\Bigg[\sum_{i=1}^n \min\{F_Y(Y_i), F_Y(Y_{N(i)})\} - \frac{1}{n-1}\sum_{\substack{i,j=1\\ i\neq j}}^n \min\{F_Y(Y_i), F_Y(Y_j)\}\Bigg]. \tag{2.2}$$

For the sake of presentation clearness, introduce

$$A_{ij} := 6\min\{F_Y(Y_i), F_Y(Y_j)\} - 2,$$
$$V_i := n^{-\frac{1}{2}}\Big(\sum_{j:i\to j} A_{ij} - \frac{1}{n-1}\sum_{j:j\neq i} A_{ij}\Big), \quad \text{and} \quad S_n := \sum_{i=1}^n V_i.$$

Then we can reformulate (2.2) as

$$n\operatorname{Var}[\breve{\xi}_n] = \Big(\frac{n^2}{n^2-1}\Big)^2 \operatorname{Var}\Big[\sum_{i=1}^n V_i\Big] = \Big(\frac{n^2}{n^2-1}\Big)^2 \operatorname{Var}[S_n]. \tag{2.3}$$

Since $Y_1, \dots, Y_n$ are independent and identically distributed (i.i.d.), for every $i \in \{1, \dots, n\}$, $F_Y(Y_i) \overset{i.i.d.}{\sim} \text{Unif}[0, 1]$. Thus, the expectation about A_{ij} can be derived as

$$\mathrm{E}A_{ij} = 0, \qquad \mathrm{E}(A_{ij})^2 = 2 := \gamma_1, \qquad \mathrm{E}(A_{ij}A_{ik}) = \frac{4}{5} := \gamma_2, \quad \text{and} \quad \mathrm{E}V_i = 0.$$

Get back to the calculation of $\operatorname{Var}[\breve{\xi}_n]$. According to (2.3), it remains to calculate

$$\operatorname{Var}[S_n] = \mathrm{E}\Big(\sum_{i=1}^n V_i\Big)^2 = \sum_{i=1}^n \mathrm{E}V_i^2 + \sum_{i\neq j}\mathrm{E}V_iV_j.$$

For the first term, we have

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}V_i^2 &= n^{-1}\sum_{i=1}^n \mathrm{E}\Big(\sum_{j:i\to j} A_{ij} - \frac{1}{n-1}\sum_{j:i\neq j} A_{ij}\Big)^2 \\
&= n^{-1}\Big(\sum_{i=1}^n \mathrm{E}\Big(\sum_{i\to j} A_{ij}\Big)^2 + \sum_{i=1}^n \mathrm{E}\Big(\frac{1}{n-1}\sum_{j:i\neq j} A_{ij}\Big)^2 \\
&\quad - \sum_{i=1}^n 2\mathrm{E}\Big(\frac{1}{n-1}\Big(\sum_{i\to j} A_{ij}\Big)\Big(\sum_{j:i\neq j} A_{ij}\Big)\Big)\Big) \\
&= \gamma_1 + \Big(\frac{1}{n-1}\gamma_1 + \frac{n-2}{n-1}\gamma_2\Big) - \Big(\frac{2}{n-1}\gamma_1 + \frac{2(n-2)}{n-1}\gamma_2\Big) \\
&= \Big(1 - \frac{1}{n-1}\Big)\gamma_1 - \Big(1 - \frac{1}{n-1}\Big)\gamma_2 \\
&\to \gamma_1 - \gamma_2.
\end{aligned}$$

For the second term, we have

$$\begin{aligned}
\sum_{i\neq j}\mathrm{E}V_iV_j &= n^{-1}\sum_{i\neq j}\mathrm{E}\Big(\sum_{k:i\to k} A_{ik} - \frac{1}{n-1}\sum_{l:i\neq l} A_{il}\Big)\Big(\sum_{m:j\to m} A_{jm} - \frac{1}{n-1}\sum_{p:j\neq p} A_{jp}\Big) \\
&= n^{-1}\Big(\sum_{i\neq j}\mathrm{E}\Big(\sum_{k:i\to k} A_{ik}\sum_{m:j\to m} A_{jm}\Big) + \sum_{i\neq j}\mathrm{E}\Big(\frac{1}{(n-1)^2}\sum_{l:i\neq l} A_{il}\sum_{p:j\neq p} A_{jp}\Big) \\
&\quad - \sum_{i\neq j}\mathrm{E}\Big(\frac{2}{n-1}\sum_{k:i\to k} A_{ik}\sum_{p:j\neq p} A_{jp}\Big)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{i\to j, j\to i} \gamma_1 + \sum_{\substack{i\to j, j\to i \\ or\ i\to j, j\to k \\ or\ j\to i, i\to k}} \gamma_2 \right) \\
&\quad + \left(\frac{n}{n-1}\gamma_1 + \frac{3(n-2)}{n-1}\gamma_2 \right) - \left(\frac{n+1}{n-1}\gamma_1 + \frac{6(n-2)}{n-1}\gamma_2 \right) \\
&= -\frac{1}{n-1}\gamma_1 - 3\left(1 - \frac{1}{n-1}\right)\gamma_2 + \gamma_1 \mathfrak{q}_m + \gamma_2(\mathfrak{o}_m + 2 - 2\mathfrak{q}_m) + o(1) \\
&\to (\gamma_1 - 2\gamma_2)\mathfrak{q}_m + \gamma_2 \mathfrak{o}_m - \gamma_2.
\end{aligned}
$$

In conclusion, we have

$$
\begin{aligned}
\lim_{n\to\infty} n \operatorname{Var}[\check{\xi}_n] &= \lim_{n\to\infty} \operatorname{Var}[S_n] \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E} V_i^2 + \lim_{n\to\infty} \sum_{i\neq j} \mathrm{E} V_i V_j \\
&= \frac{2}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m.
\end{aligned}
$$

This is the form of Shi, Drton and Han ((2024), Theorem 3.1(ii)) when $\boldsymbol{X}_i$'s are sampled from an absolutely continuous distribution in $\mathbb{R}^m$. □

2.2. *Proof of Lemma* 1.2.

PROOF OF LEMMA 1.2. Recall that $\mathcal{H}^m \llcorner \mathcal{M}$ represents the restricted m-dimensional Hausdorff measure on $\mathcal{M}$ and λ represents the m-dimensional Lebesgue measure.

To obtain the global property, that is, μ is absolutely continuous with respect to $\mathcal{H}^m \llcorner \mathcal{M}$, it suffices to show the local property, that is, for any point $\boldsymbol{x} \in \mathcal{M}$ and the corresponding coordinate chart (U, ψ), μ_U is absolutely continuous with respect to $\mathcal{H}^m \llcorner U$. In the following proof, we focus on a sufficiently small coordinate neighborhood U. Recall the definition of absolutely continuity of μ with respect to some base measure, which states that any null set with respect to the base measure is a μ-null set. With this definition, to prove Lemma 1.2, it suffices to prove that any $\mathcal{H}^m$-null set $\psi^{-1}E$ is equivalent to a λ-null set E.

We refer to Chapter 3 of Evans and Gariepy (2015) for the following lemma and its proof. Assume a Lipschitz map $f : \mathbb{R}^m \to \mathbb{R}^n$, where $m \le n$. Thus, f is differentiable λ-a.e. by Rademacher's theorem (Theorem 3.2 in Evans and Gariepy (2015)). At any point $\boldsymbol{y} \in \mathbb{R}^m$ of differentiability we denote by $Jf(\boldsymbol{y})$ the Jacobian of f. Also, we denote by $\mathcal{H}^0$ the 0-dimensinal Hausdorff measure, which is equivalent to counting measure.

LEMMA 2.1 (Area formula). *We have $\boldsymbol{y} \mapsto Jf(\boldsymbol{y})$ is Lebesgue measurable. In addition, for any Lebesgue measurable set $A \subset \mathbb{R}^m$, the map $z \mapsto f^{-1}(\{z\})$ is $\mathcal{H}^m$-measurable and the following equality holds*:

$$
\int_{\mathbb{R}^m} \mathcal{H}^0\big(A \cap f^{-1}(z)\big)\, \mathrm{d}\mathcal{H}^m(z) = \int_A Jf(\boldsymbol{y})\, \mathrm{d}\lambda(\boldsymbol{y}).
$$

Back to the proof, since $\mathcal{M}$ is a C^∞ manifold, the homeomorphisms ψ and ψ^{-1} are C^∞ maps, thus locally Lipschitz continuous. Applying Lemma 2.1, for any $E \subset V$ and $\psi^{-1}(E) \subset U$ which is $\mathcal{H}^m$-measurable, we have

$$
\mathcal{H}^m\big(\psi^{-1}(E)\big) = \int_E Jf(\boldsymbol{y})\, \mathrm{d}\lambda(\boldsymbol{y}). \tag{2.4}
$$

Here for a smooth manifold, $Jf(\mathbf{y})$ can be written specifically as $Jf(\mathbf{y}) = \sqrt{\det g}$, where g is the metric tensor of the Riemannian submanifold, that is, $g_{ij} = \partial_i \psi^{-1} \partial_j \psi^{-1}$. Due to the existence of ψ^{-1}, the differentials of ψ and ψ^{-1} have maximum ranks everywhere locally, which implies that for every $\mathbf{y} \in E$, $Jf(\mathbf{y}) > 0$. Using equation (2.4), one directly obtains

$$\mathcal{H}^m\big(\psi^{-1}(E)\big) = 0 \quad \text{if and only if} \quad \lambda(E) = 0,$$

which completes the proof. □

2.3. *Proof of Lemma* 1.4.

PROOF OF LEMMA 1.4. Recall that μ is the (induced) probability measure of $\mathbf{X}$, that is, $\mu(A) = \mathbb{P}(\mathbf{X} \in A)$. Let $[\mathbf{X}_i]_{i=1}^n$ be a sample comprised of n independent copies of $\mathbf{X}$. Let $N(\mathbf{X}_i)$ denote the number of NN pairs containing $\mathbf{X}_i$ and N_{total} denote the total number of NN pairs (double edges) in NNG; in other words,

$$N_{\text{total}} := \#\big\{(i, j)\ \textit{distinct} : i \to j,\ j \to i \in \mathcal{E}(\mathcal{G}_n)\big\},$$
$$N(\mathbf{X}_i) := \#\big\{j : i \to j,\ j \to i \in \mathcal{E}(\mathcal{G}_n)\big\}.$$

Observe that

$$N_{\text{total}} = \sum_{i=1}^n N(\mathbf{X}_i);$$

it thus suffices to calculate the value of $\mathrm{E}(N(\mathbf{X}_i))$. Using Lemma 1.3, there exists an upper bound of the number of points, whose nearest neighbor is $\mathbf{X}_i$. For any point $\mathbf{x}_i$, we use constant $\mathfrak{C}_d$ to denote this upper bound:

$$N(\mathbf{x}_i) \le \mathfrak{C}_d.$$

To obtain the asymptotic expecatation of $\mathrm{E}(N(X_i))$, it suffices to prove that for μ-a.e. $\mathbf{x}_i \in \mathcal{M}$,

$$\lim_{n\to\infty} \mathrm{E}\big(N(\mathbf{x}_i)\big) = \frac{V_m}{U_m},$$

since the Lebesgue dominated convergence theorem can be used as

$$\begin{aligned} \lim_{n\to\infty} \mathrm{E}N(\mathbf{X}_i) &= \lim_{n\to\infty} \int \mathrm{E}\big(N(\mathbf{x}_i)\big)\, \mathrm{d}\mu(\mathbf{x}_i) \\ &= \int \lim_{n\to\infty} \mathrm{E}\big(N(\mathbf{x}_i)\big)\, \mathrm{d}\mu(\mathbf{x}_i) = \frac{V_m}{U_m}. \end{aligned}$$

We use $U^d(\mathbf{x}_1, \mathbf{x}_2)$ to denote the union of two balls in $\mathbb{R}^d$ whose centers are $\mathbf{x}_1$ and $\mathbf{x}_2$ and radius are both $\|\mathbf{x}_1 - \mathbf{x}_2\|$. Thus, $\mathbf{x}_1$ and $\mathbf{x}_2$ form a nearest neighbor pair if and only if there are no other sample points in $U^d(\mathbf{x}_1, \mathbf{x}_2)$.

We first consider the nearest neighbor of one fixed sample point $\mathbf{X}_i$ and write $N(\mathbf{x}_i) := N(\mathbf{X}_i | \mathbf{X}_i = \mathbf{x}_i)$ for convenience. Applying the absolute continuity assumption to the specific point $\mathbf{x}_i \in \mathcal{M}$, we can write (recalling the notation of Assumption 1.1)

$$g = \mathrm{d}(\psi_* \mu)/\mathrm{d}\lambda$$

for the density of the pushforward, that is, $\psi_*\mu(A) = \int_A g\, \mathrm{d}\lambda$, for any μ-measurable set $A \in V$.

We assume that $\boldsymbol{x}_i \in \operatorname{supp}(\mu)$ and $g(\psi(\boldsymbol{x}_i)) > 0$ throughout the manuscript. Observe that for a fixed $\boldsymbol{x}_i$,

$$\begin{aligned}\mathrm{E}\big(N(\boldsymbol{x}_i)\big) &= \sum_{i\neq j} \mathrm{P}\Big(\bigcup_{k\neq i,j}\{\boldsymbol{X}_k \notin U^d(\boldsymbol{x}_i, \boldsymbol{X}_j)\}\Big)\\ &= (n-1)\mathrm{P}\Big(\bigcup_{k\neq i,j}\{\boldsymbol{X}_k \notin U^d(\boldsymbol{x}_i, \boldsymbol{X}_j)\}\Big)\\ &= (n-1)\int_{\mathcal{M}} \big(1-\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\big)\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j),\end{aligned} \tag{2.5}$$

which is the expression used for the following calculation.

LEMMA 2.2. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart $\psi(=\psi_{\boldsymbol{x}})$, it holds true that*

$$\lim_{\|\boldsymbol{x}_j-\boldsymbol{x}\|\to 0} \frac{\mu(U^d(\boldsymbol{x},\boldsymbol{x}_j))}{g(\psi(\boldsymbol{x}))\|\boldsymbol{x}_j-\boldsymbol{x}\|^m U_m} = 1.$$

LEMMA 2.3. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart $\psi(=\psi_{\boldsymbol{x}})$, it holds true that*

$$\lim_{\|\boldsymbol{x}_j-\boldsymbol{x}\|\to 0} \frac{\mu(B^d(\boldsymbol{x},r))}{g(\psi(\boldsymbol{x}))r^m V_m} = 1.$$

Lemma 2.2 informs that when fixing $\boldsymbol{x}_i \in \mathcal{M}$, for every $\epsilon > 0$, it is almost sure that there exists $\delta_1 > 0$ such that for every $x_j : \|\boldsymbol{x}_j - \boldsymbol{x}_i\| < \delta_1$, we have

$$\frac{\mu(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j))}{g(\psi(\boldsymbol{x}_i))\|\boldsymbol{x}_j-\boldsymbol{x}_i\|^m U_m} \in [1-\epsilon, 1+\epsilon]. \tag{2.6}$$

Similarly, Lemma 2.3 informs that there exists $\delta_2 > 0$ such that for every $r < \delta_2$,

$$\frac{\mu(B^d(\boldsymbol{x}_i,r))}{g(\psi(\boldsymbol{x}_i))r^m V_m} \in [1-\epsilon, 1+\epsilon]. \tag{2.7}$$

In the following part of this section, we take $0 < \delta < \min\{\delta_1, \delta_2\}$, which implies that both (2.6) and (2.7) hold for every $\boldsymbol{x}_j$ such that $\|\boldsymbol{x}_j - \boldsymbol{x}_i\| < \delta$.

Back to the expression (2.5) for $\mathrm{E}(N(\boldsymbol{x}_i))$, we first consider the lower bound,

$$\begin{aligned}\mathrm{E}\big(N(\boldsymbol{x}_i)\big) &\geq (n-1)\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\leq\delta} \big(1-\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\\ &\geq (n-1)\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\leq\delta} \big(1-(1+\epsilon)g\big(\psi(\boldsymbol{x}_i)\big)\|\boldsymbol{x}_j-\boldsymbol{x}_i\|^m U_m\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j).\end{aligned} \tag{2.8}$$

Next, we introduce a nonincreasing function

$$\phi_1(r) := \big(1-(1+\epsilon)g\big(\psi(\boldsymbol{x}_i)\big)r^m U_m\big)^{n-2} I_{\{r\leq\delta\}}, \tag{2.9}$$

where $I_{\{r\leq\delta\}}$ denotes the indicator function. Notice that $\phi_1(0) = 1$.

With this function, we can reformulate equation (2.8) as follows:

$$\begin{aligned}(2.8) &= (n-1)\int_{\mathcal{M}} \phi_1\big(\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\big)\,\mathrm{d}\mu(\boldsymbol{x}_j)\\ &= (n-1)\mathrm{E}\big(\phi_1\big(\|\boldsymbol{X}_j-\boldsymbol{x}_i\|\big)\big)\\ &= (n-1)\int_0^1 \mathrm{P}\big(\phi_1\big(\|\boldsymbol{X}_j-\boldsymbol{x}_i\|\big) > t\big)\,\mathrm{d}t\\ &= (n-1)\int_0^1 \Big(\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\leq\phi_1^{-1}(t)} \mathrm{d}\mu(\boldsymbol{x}_j)\Big)\,\mathrm{d}t.\end{aligned} \tag{2.10}$$

For handling (2.10), we select the range of t to make $\boldsymbol{x}_j$ close enough to $\boldsymbol{x}_i$. Applying then the approximation derived above in (2.7) yields

$$\begin{aligned}(2.10) &\geq (n-1)\int_{\phi_1(\delta)}^{1}\left(\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\leq\phi_1^{-1}(t)} \mathrm{d}\mu(\boldsymbol{x}_j)\right)\mathrm{d}t \\ &= (n-1)\int_{\phi_1(\delta)}^{1}\mu\big(B^d(\boldsymbol{x}_i,\phi_1^{-1}(t))\big)\,\mathrm{d}t \\ &\geq (n-1)\int_{\phi_1(\delta)}^{1}(1-\epsilon)V_m\big(\phi_1^{-1}(t)\big)^m g\big(\psi_1(\boldsymbol{x}_i)\big)\,\mathrm{d}t.\end{aligned} \tag{2.11}$$

The expression of $\phi_1^{-1}(t)$ can be solved from equation (2.9). Plugging it into (2.11) gives

$$\begin{aligned}(2.11) &= (n-1)\int_{\phi_1(\delta)}^{1}\frac{1-\epsilon}{1+\epsilon}\frac{V_m}{U_m}\big(1-t^{\frac{1}{n-2}}\big)\,\mathrm{d}t \\ &\geq \frac{1-\epsilon}{1+\epsilon}\frac{V_m}{U_m}\big(1-(n-1)\phi_1(\delta)\big).\end{aligned} \tag{2.12}$$

Consider the limit of (2.12). First let n go to infinity. Since

$$\lim_{n\to\infty}(n-1)\phi_1(\delta) \leq \lim_{n\to\infty}(n-1)\big(1-(1+\epsilon)g\big(\psi(\boldsymbol{x}_i)\big)\delta^m U_m\big)^{n-2} = 0,$$

we have

$$\lim_{n\to\infty} E\big(N(\boldsymbol{x}_i)\big) \geq \frac{1-\epsilon}{1+\epsilon}\frac{V_m}{U_m}$$

holds for arbitrary $\epsilon > 0$. Thus,

$$\lim_{n\to\infty} E\big(N(\boldsymbol{x}_i)\big) \geq \frac{V_m}{U_m}.$$

Now we turn to the inequality in the other direction to find the upper bound. Using the fact that

$$e^{-x} \geq 1-x \quad \text{for any } x \in [0,1],$$

we have

$$\begin{aligned}\mathrm{E}\big(N(\boldsymbol{x}_i)\big) &\leq (n-1)\int_{\mathcal{M}}\exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j) \\ &= (n-1)\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|>\delta}\exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j) \\ &\quad + (n-1)\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\leq\delta}\exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j) \\ &=: I_{n,1} + I_{n,2}.\end{aligned} \tag{2.13}$$

Here the domain of integration is partitioned into two sets. We use $I_{n,1}$ to denote the integral on $\mathcal{M}\cap\{\boldsymbol{x}_j : \|\boldsymbol{x}_j-\boldsymbol{x}_i\| > \delta\}$ and $I_{n,2}$ to denote the integral on $\mathcal{M}\cap\{\boldsymbol{x}_j : \|\boldsymbol{x}_j-\boldsymbol{x}_i\| \leq \delta\}$. We then process these two terms in (2.13) respectively. For term $I_{n,1}$, we have

$$\begin{aligned}I_{n,1} &= (n-1)\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|>\delta}\exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j) \\ &\leq (n-1)\sup_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\geq\delta}\big\{\exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big)\big\}\int_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|>\delta}\mathrm{d}\mu(\boldsymbol{x}_j) \\ &\leq (n-1)\exp\Big(-(n-2)\inf_{\|\boldsymbol{x}_j-\boldsymbol{x}_i\|\geq\delta}\big\{\mu\big(U^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\cap\mathcal{M}\big)\big\}\Big)\int_{\mathcal{M}}\mathrm{d}\mu(\boldsymbol{x}_j).\end{aligned} \tag{2.14}$$

Since for $\nu > 0$, $U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \subset U^d(\boldsymbol{x}_i, \boldsymbol{x}_j + \nu(\boldsymbol{x}_j - \boldsymbol{x}_i))$ and $\int_{\mathcal{M}} \mathrm{d}\mu = 1$,

$$(2.14) \le (n-1)\exp\Big(-(n-2) \inf_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| = \delta} \mu\big(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}\big)\Big)$$
$$:= (n-1)\exp\big(-(n-2)h(\delta)\big).$$

Here $h(\delta)$ is a positive function of δ. Thus for any $\epsilon, \delta > 0$, $\lim_{n\to\infty} I_{n,1} = 0$.

For the second term, using the approximation (2.6), we have

$$\begin{aligned} I_{n,2} &= (n-1)\int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \le \delta} \exp\big(-(n-2)\mu\big(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}\big)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j) \\ &\le (n-1)\int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \le \delta} \exp\big(-(n-2)(1-\epsilon)g\big(\psi(\boldsymbol{x}_i)\big)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^m U_m\big)\,\mathrm{d}\mu(\boldsymbol{x}_j). \end{aligned} \tag{2.15}$$

Again, introduce a nonincreasing function

$$\phi_2(r) := \exp\big(-(n-2)(1-\epsilon)^2 g\big(\psi(\boldsymbol{x}_i)\big)r^m U_m\big)I_{\{r \le \delta\}}$$

and reformulate the expression as before:

$$\begin{aligned} (2.15) &= (n-1)\mathrm{E}\big(\phi_2\big(\|\boldsymbol{X}_j - \boldsymbol{x}_i\|\big)\big) \\ &= (n-1)\int_0^1 \mathrm{P}\big(\phi_2\big(\|\boldsymbol{X}_j - \boldsymbol{x}_i\| > t\big)\big)\,\mathrm{d}t \\ &= (n-1)\int_0^1 \bigg(\int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \le \phi_2^{-1}(t)} \mathrm{d}\mu(\boldsymbol{x}_j)\bigg)\,\mathrm{d}t \\ &\le (n-1)\int_0^{\phi_2(\delta)} \mathrm{d}t + (n-1)\int_{\phi_2(\delta)}^1 \mu\big(B^d\big(\boldsymbol{x}_i, \phi_2^{-1}(t)\big)\big)\,\mathrm{d}t \\ &\le (n-1)\phi_2(\delta) + (n-1)\int_{\phi_2(\delta)}^1 (1+\epsilon)V_m\big(\phi_2^{-1}(t)\big)^m\,\mathrm{d}t. \end{aligned} \tag{2.16}$$

The expression of $\phi_2^{-1}(t)$ can be solved from (2.16). Then the integral above can be specifically calculated as

$$(2.16) = (n-1)\phi_2(\delta) + \frac{n-1}{n-2}\frac{1+\epsilon}{1-\epsilon}\frac{V_m}{U_m}\big(1 - \phi_2(\delta) + \phi_2(\delta)\log\phi_2(\delta)\big).$$

For fixed $\epsilon > 0$ and $\delta > 0$, one can use the definition of $\phi_2(r)$ and get

$$\lim_{n\to\infty} n\phi_2(\delta) = 0.$$

Then we have

$$\lim_{n\to\infty} \mathrm{E}\big(N(\boldsymbol{x}_i)\big) \le \frac{1+\epsilon}{1-\epsilon}\frac{V_m}{U_m}$$

holds for arbitrary $\epsilon > 0$. Thus,

$$\lim_{n\to\infty} \mathrm{E}\big(N(\boldsymbol{x}_i)\big) \le \frac{V_m}{U_m}.$$

Combining the upper and lower bounds yields

$$\lim_{n\to\infty} \mathrm{E}\big(N(\boldsymbol{x}_i)\big) = \frac{V_m}{U_m}.$$

Finally, we obtain the conclusion

$$\lim_{n\to\infty} \frac{\mathrm{E}N_{\text{total}}}{n} = \lim_{n\to\infty} \sum_{i=1}^n \frac{\mathrm{E}(N(\boldsymbol{X}_i))}{n} = \frac{V_m}{U_m} = \mathfrak{q}_m.$$

This completes the proof. □

2.4. *Proof of Lemma* 1.5.

PROOF OF LEMMA 1.5. Analogously to the proof of Lemma 1.4, we only have to consider one specific sample point $\boldsymbol{X}_i$. Let Q_j denote the event that $\boldsymbol{X}_i$ is the nearest neighbor of $\boldsymbol{X}_j$, that is, $Q_j := \{j \to i \in \mathcal{E}(\mathcal{G}_n)\}$ for all $j \neq i$ and put

$$M_{\text{total}} = \#\{(i, j, k)\ \textit{distinct} : j \to i, k \to i \in \mathcal{E}(\mathcal{G}_n)\},$$
$$D(\boldsymbol{X}_i) = \sum_{j \neq i} I_{Q_j} = \#\{j : j \to i \in \mathcal{E}(\mathcal{G}_n)\},$$
$$M(\boldsymbol{X}_i) = \#\{(j, k)\ \textit{distinct} : j \to i, k \to i \in \mathcal{E}(\mathcal{G}_n)\}.$$

Here $M(\boldsymbol{X}_i)$ is the coefficient we are interested in. Observe that

$$M_{\text{total}} = \sum_{i=1}^{n} M(\boldsymbol{X}_i),$$
$$M(\boldsymbol{X}_i) = D(\boldsymbol{X}_i)\big(D(\boldsymbol{X}_i) - 1\big),$$

so that we have

$$\begin{aligned} \mathrm{E}M(\boldsymbol{X}_i) &= \mathrm{E}D(\boldsymbol{X}_i)\big(D(\boldsymbol{X}_i) - 1\big) \\ &= \mathrm{E} \sum_{\substack{(j,k)\ \textit{distinct} \\ j,k \neq i}} \mathrm{P}(Q_j \cap Q_k) \\ &= (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k). \end{aligned}$$

We first consider a fixed sample point $\boldsymbol{X}_i = \boldsymbol{x}_i$. Using Lemma 1.3, we have

$$\begin{aligned} \mathfrak{C}_d^2 &\geq \mathrm{E}D(\boldsymbol{X}_i | \boldsymbol{X}_i = \boldsymbol{x}_i)\big(D(\boldsymbol{X}_i | \boldsymbol{X}_i = \boldsymbol{x}_i) - 1\big) \\ &= (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}_i). \end{aligned}$$

Then applying the Lebesgue dominated convergence theorem, we have

$$\begin{aligned} \lim_{n\to\infty} (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k) &= \lim_{n\to\infty} \int_{\mathcal{M}} (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}_i)\, \mathrm{d}\mu(\boldsymbol{x}_i) \\ &= \int_{\mathcal{M}} \lim_{n\to\infty} (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}_i)\, \mathrm{d}\mu(\boldsymbol{x}_i). \end{aligned}$$

It thus remains to prove that

$$\lim_{n\to\infty} n^2 \mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}_i) = \lim_{n\to\infty} (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}_i) = \mathfrak{o}_m,$$

where $\boldsymbol{x}_i$ is any point with positive density, which is held fixed in what follows.

We first introduce the notation to simplify the integral. For a fixed point $\boldsymbol{x} \in \mathcal{M}$, let

$$\Gamma_x := \big\{(\boldsymbol{x}_j, \boldsymbol{x}_k) \in (\mathbb{R}^d)^2 : \max\{\|\boldsymbol{x}_j - \boldsymbol{x}\|, \|\boldsymbol{x}_k - \boldsymbol{x}\|\} \leq \|\boldsymbol{x}_j - \boldsymbol{x}_k\|\big\},$$
$$S_j := B\big(\boldsymbol{x}_j, \|\boldsymbol{x}_j - \boldsymbol{x}\|\big) \cap \mathcal{M}, \qquad S_k := B\big(\boldsymbol{x}_k, \|\boldsymbol{x}_k - \boldsymbol{x}\|\big) \cap \mathcal{M}.$$

Applying the above notation, we have that $\boldsymbol{X} = \boldsymbol{x}$ is the nearest neighbor of $\boldsymbol{X}_j = \boldsymbol{x}_j$ if and only if there are no other sample points in S_j. In the following part, we record $\boldsymbol{x}_i$ as $\boldsymbol{x}$ for notation simplicity. It then holds true that

$$n^2 \mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}) = \iint_{\Gamma_x} n^2 \big(1 - \mu(S_j \cup S_k)\big)^{n-2}\, \mathrm{d}\mu(\boldsymbol{x}_j)\, \mathrm{d}\mu(\boldsymbol{x}_k),$$

where we use $\mathrm{d}\mu(\boldsymbol{x}_j)$ and $\mathrm{d}\mu(\boldsymbol{x}_k)$ to denote the measure and the corresponding random variables which the integral corresponds to. The next step is to split the region of integral that $\Gamma_{\boldsymbol{x}} = \Gamma^1_{\boldsymbol{x},\delta} \cup \Gamma^2_{\boldsymbol{x},\delta}$. Here

$$\Gamma^1_{\boldsymbol{x},\delta} := \Gamma_{\boldsymbol{x}} \cap \{(\boldsymbol{x}_1, \boldsymbol{x}_2) \in (\mathbb{R}^d)^2 : \|\boldsymbol{x}_1 - \boldsymbol{x}\|, \|\boldsymbol{x}_2 - \boldsymbol{x}\| \le \delta\},$$

$$\Gamma^2_{\boldsymbol{x},\delta} := \Gamma_{\boldsymbol{x}} \setminus \Gamma^1_{\boldsymbol{x},\delta},$$

where δ is a sufficiently small but a positive real number. Specifically, $\delta = \min\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}$, in which δ_i's are to be defined in the proof below. Taking this idea, we partition the integral into two parts:

$$\begin{aligned} n^2\mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}) &= \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2(1 - \mu(S_j \cup S_k))^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k) \\ &\quad + \iint_{\Gamma^2_{\boldsymbol{x},\delta}} n^2(1 - \mu(S_j \cup S_k))^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k) \\ &=: J_{n,1} + J_{n,2}. \end{aligned}$$

We first prove that

$$\lim_{n\to\infty} J_{n,2} = 0.$$

LEMMA 2.4. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart ψ, it holds true that*

$$\lim_{\|\boldsymbol{x}_j - \boldsymbol{x}\|\to 0} \frac{\mu(B^d(\boldsymbol{x}_j, \|\boldsymbol{x} - \boldsymbol{x}_j\|)))}{g(\psi(\boldsymbol{x}))\lambda(B^m(\boldsymbol{x}_j, \|\boldsymbol{x} - \boldsymbol{x}_j\|))} = 1.$$

Lemma 2.4 shows that for any $\epsilon > 0$, there exists $\delta_1 > 0$ such that for μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and for any $\boldsymbol{x}_j$ such that $\|\boldsymbol{x} - \boldsymbol{x}_j\| \le \delta_1$,

$$\frac{\mu(B^d(\boldsymbol{x}_j, \|\boldsymbol{x} - \boldsymbol{x}_j\|)))}{g(\psi(\boldsymbol{x}))\lambda(B^m(\boldsymbol{x}_j, \|\boldsymbol{x} - \boldsymbol{x}_j\|))} \in [1-\epsilon, 1+\epsilon].$$

By definition of $\Gamma^2_{\boldsymbol{x},\delta}$, for any $(\boldsymbol{x}_j, \boldsymbol{x}_k) \in \Gamma^2_{\boldsymbol{x},\delta}$, there exists $\boldsymbol{x}_l = \boldsymbol{x}_j$ or $\boldsymbol{x}_k$ satisfying $\|\boldsymbol{x}_l - \boldsymbol{x}\| > \delta$. Without loss of generality, we assume that $\boldsymbol{x}_l = \boldsymbol{x}_j$. For the point

$$\boldsymbol{y} := \boldsymbol{x} - \frac{\delta}{2}\frac{\boldsymbol{x} - \boldsymbol{x}_j}{\|\boldsymbol{x} - \boldsymbol{x}_j\|},$$

we have $\|\boldsymbol{x} - \boldsymbol{y}\| = \delta/2$. For every $\boldsymbol{x}^* \in B(\boldsymbol{y}, \|\boldsymbol{x} - \boldsymbol{y}\|) = B(\boldsymbol{y}, \frac{\delta}{2})$,

$$\|\boldsymbol{x}^* - \boldsymbol{x}_j\| \le \|\boldsymbol{x}^* - \boldsymbol{y}\| + \|\boldsymbol{y} - \boldsymbol{x}_j\| < \frac{\delta}{2} + \left(\|\boldsymbol{x} - \boldsymbol{x}_j\| - \frac{\delta}{2}\right) = \|\boldsymbol{x} - \boldsymbol{x}_j\|,$$

yielding

$$B\left(\boldsymbol{y}, \frac{\delta}{2}\right) \subset B(\boldsymbol{x}, \delta) \cap B(\boldsymbol{x}_j, \|\boldsymbol{x}_j - \boldsymbol{x}\|).$$

This in turn implies

$$\begin{aligned} \mu(S_j \cup S_k) &\ge \mu(B(\boldsymbol{x}_j, \|\boldsymbol{x}_j - \boldsymbol{x}\|) \cap \mathcal{M}) \\ &\ge \mu(B(\boldsymbol{y}, \delta/2) \cap \mathcal{M}) \\ &\ge (1-\epsilon) g(\psi(\boldsymbol{x}))\lambda(B^m(0, \delta/2)) \\ &= (1-\epsilon) g(\psi(\boldsymbol{x})) V_m \left(\frac{\delta}{2}\right)^m. \end{aligned}$$

Back to the integral $J_{n,2}$, we then have the upper bound

$$n^2(1-\mu(S_j \cup S_k))^{n-2} \le n^2\left(1-(1-\epsilon)g(\psi(\boldsymbol{x}))V_m\left(\frac{\delta}{2}\right)^m\right)^{n-2},$$

which is a constant with respect to $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ and thus

$$0 \le \lim_{n\to\infty} J_{n,2} \le \lim_{n\to\infty} n^2\left(1-(1-\epsilon)g(\psi(\boldsymbol{x}))V_m\left(\frac{\delta}{2}\right)^m\right)^{n-2} = 0.$$

This finishes the proof of the first part.

We then prove that

$$\lim_{n\to\infty} J_{n,1} = \mathfrak{o}_m,$$

that is,

$$\lim_{n\to\infty} \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2(1-\mu(S_j \cup S_k))^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k) = \mathfrak{o}_m.$$

Here δ is a constant we will select later. Similar to the process in the proof of Lemma 1.4, we will then derive the upper and lower bounds of $J_{n,1}$.

For the upper bound, first leveraging the fact that $e^{-x} \ge 1-x$ for all $x \in (0,1)$, we have

$$J_{n,1} \le \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp(-(n-2)\mu(S_j \cup S_k))\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k). \tag{2.17}$$

We set $z_j = \psi(\boldsymbol{x}_j)$, $z_k = \psi(\boldsymbol{x}_k)$ for simplicity and thus, $\boldsymbol{x}_j = \psi^{-1}(z_j)$, $\boldsymbol{x}_k = \psi^{-1}(z_k)$.

LEMMA 2.5. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart ψ, we have*

$$\lim_{\substack{\|\boldsymbol{x}_j-\boldsymbol{x}\|\to 0\\ \|\boldsymbol{x}_k-\boldsymbol{x}\|\to 0}} \frac{\mu((B^d(\boldsymbol{x}_j, \|\boldsymbol{x}_j-\boldsymbol{x}\|) \cup B^d(\boldsymbol{x}_k, \|\boldsymbol{x}_k-\boldsymbol{x}\|)))}{g(\psi(\boldsymbol{x}))\lambda(B^m(z_j, \|z_j-\psi(\boldsymbol{x})\|) \cup B^m(z_k, \|z_k-\psi(\boldsymbol{x})\|))} = 1.$$

Lemma 2.5 implies that for μ-a.e $\boldsymbol{x} \in \Gamma_{\boldsymbol{x}}$ and any $\epsilon > 0$, there exists $\delta_3 > 0$ such that for any $(\boldsymbol{x}_j, \boldsymbol{x}_k)$ satisfying $\|\boldsymbol{x}_j-\boldsymbol{x}\| \le \delta_3$, $\|\boldsymbol{x}_k-\boldsymbol{x}\| \le \delta_3$, we have

$$\frac{\mu((B^d(\boldsymbol{x}_j, \|\boldsymbol{x}_j-\boldsymbol{x}\|) \cup B^d(\boldsymbol{x}_k, \|\boldsymbol{x}_k-\boldsymbol{x}\|)))}{g(\psi(\boldsymbol{x}))\lambda(B^m(z_j, \|z_j-\psi(\boldsymbol{x})\|) \cup B^m(z_k, \|z_k-\psi(\boldsymbol{x})\|))} \in [1-\epsilon, 1+\epsilon].$$

Plugging them into the righthand side of (2.17), we obtain

$$\begin{aligned}
&\iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp(-(n-2)\mu(S_j \cup S_k))\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\
&\quad\le \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp\big(-(n-2)(1-\epsilon)g(\psi(\boldsymbol{x}))\\
&\qquad\times \lambda(B^m(z_j, \|z_j-\psi(\boldsymbol{x})\|) \cup B^m(z_k, \|z_k-\psi(\boldsymbol{x})\|))\big)\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k).\\
&\quad= \iint_{\psi(\Gamma^1_{\boldsymbol{x},\delta})} n^2 \exp\big(-(n-2)(1-\epsilon)g(\psi(\boldsymbol{x}))\\
&\qquad\times \lambda(B^m(z_j, \|z_j-\psi(\boldsymbol{x})\|) \cup B^m(z_k, \|z_k-\psi(\boldsymbol{x})\|))\big)\\
&\qquad\times g(z_j)g(z_k)\,\mathrm{d}\lambda(z_j)\,\mathrm{d}\lambda(z_k).
\end{aligned} \tag{2.18}$$

LEMMA 2.6. *Assume all the conditions of the Lebesgue differential theorem hold, and f, g are integrable functions with respect to Lebesgue measure λ. We then have*

$$\lim_{U\to x, U\in\mathcal{V}} \frac{1}{|U|}\int_U fg\,\mathrm{d}\lambda = f(\boldsymbol{x}) \lim_{U\to x, U\in\mathcal{V}} \frac{1}{|U|}\int_U g\,\mathrm{d}\lambda.$$

Using Lemma 2.6, for fixed $\epsilon > 0$, there exists $\delta_4 > 0$ such that for every $\delta < \delta_4$, we have

$$\begin{aligned}(2.18) \le (1+\epsilon)n^2 g\big(\psi(\boldsymbol{x})\big)^2 \iint_{\psi(\Gamma^1_{\boldsymbol{x},\delta})} \exp\big(-(n-2)(1-\epsilon)g\big(\psi(\boldsymbol{x})\big)\\ \times \lambda\big(B^m\big(z_j, \|z_j - \psi(\boldsymbol{x})\|\big) \cup B^m\big(z_k, \|z_k - \psi(\boldsymbol{x})\|\big)\big)\big)\,\mathrm{d}\lambda(z_j)\,\mathrm{d}\lambda(z_k).\end{aligned} \tag{2.19}$$

LEMMA 2.7. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart ψ, it holds true that*

$$\lim_{\delta\to 0} \frac{\lambda(\psi(\Gamma^1_{\boldsymbol{x},\delta}))}{\lambda(\Gamma_{m,\delta})} = 1.$$

Considering Lemma 2.7 and the uniform property of Lebesgue measure, we can just do a translation and make the origin point contained in the domain of integration so that

$$\begin{aligned}(2.19) \le (1+\epsilon)^2 n^2 g\big(\psi(\boldsymbol{x})\big)^2 \iint_{\Gamma_{m,\delta}} \exp\big(-(n-2)(1-\epsilon)g\big(\psi(\boldsymbol{x})\big)\\ \times \lambda\big(B^m\big(z_j, \|z_j\|\big) \cup B^m\big(z_k, \|z_k\|\big)\big)\,\mathrm{d}\lambda(z_j)\,\mathrm{d}\lambda(z_k),\end{aligned}$$

where

$$\Gamma_{m,\delta} := \Gamma_m \cap \big\{(z_j, z_k) \in \big(\mathbb{R}^m\big)^2 : \|z_j\|, \|z_k\| \le \delta\big\}.$$

Denote $B^m(z_j, \|z_j\|)$ and $B^m(z_k, \|z_k\|)$ to be m-dimensional balls in $\mathbb{R}^m$. We then introduce a translation

$$z^*_j := l_m z_j, \qquad z^*_k := l_m z_k.$$

We thus have

$$\lambda(B^m\big(z^*_j, \|z^*_j\|\big) \cup \big(B^m\big(z^*_k, \|z^*_k\|\big)\big) = (l_m)^m \lambda(B^m\big(z_j, \|z_j\|\big) \cup \big(B^m\big(z_k, \|z_k\|\big)\big).$$

Set $l_m = ((n-2)(1-\epsilon)g(\psi(\boldsymbol{x})))^{1/m}$ and the corresponding Jacobi matrix satisfies $|J| = (n-2)(1-\epsilon)g(\psi(\boldsymbol{x}))$. As $n \to \infty$, $l_m \to \infty$ and $\frac{n}{n-2} \to 1$, and thus

$$\begin{aligned}\lim_{n\to\infty} J_{n,1} &\le \lim_{n\to\infty} \frac{(1+\epsilon)n^2}{(1-\epsilon)^2(n-2)^2}\\ &\quad\times \iint_{\Gamma_{m,l_m\delta}} \exp\big(-\lambda\big(B^m\big(z^*_j, \|z^*_j\|\big) \cup \big(B^m\big(z^*_k, \|z^*_k\|\big)\big)\big)\,\mathrm{d}\lambda\big(z^*_j\big)\,\mathrm{d}\lambda\big(z^*_k\big)\\ &= \frac{1+\epsilon}{(1-\epsilon)^2} \iint_{\Gamma_m} \exp\big(-\lambda\big(B^m\big(z^*_j, \|z^*_j\|\big) \cup \big(B^m\big(z^*_k, \|z^*_k\|\big)\big)\big)\,\mathrm{d}\lambda\big(z^*_j\big)\,\mathrm{d}\lambda\big(z^*_k\big)\end{aligned} \tag{2.20}$$

holds for arbitrary $\epsilon > 0$. We thus obtain the upper bound.

For the lower bound, the proof is similar. Use the fact that, for any $\epsilon > 0$, there exists $\sigma > 0$ such that for every $x \in [0, \sigma]$, $e^{-x} \le (1+\epsilon)(1-x)$. We have there exists δ_5 such that for any $\delta < \delta_5$,

$$J_{n,1} \ge \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp\big(-(n-2)(1+\epsilon)\mu(S_j \cup S_k)\big)\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k).$$

We then follow the same process as above: for a sufficiently small $\delta > 0$,

$$\begin{aligned}
&\iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp\bigl(-(n-2)(1+\epsilon)\mu(S_j \cup S_k)\bigr)\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\
&\quad\ge \iint_{\Gamma^1_{\boldsymbol{x},\delta}} n^2 \exp\bigl(-(n-2)(1+\epsilon)^2 g\bigl(\psi(\boldsymbol{x})\bigr)\\
&\qquad \times \lambda\bigl(B^m\bigl(z_j, \|z_j - \psi(\boldsymbol{x})\|\bigr) \cup B^m\bigl(z_k, \|z_k - \psi(\boldsymbol{x})\|\bigr)\bigr)\bigr)\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\
&\quad\ge (1-\epsilon)n^2 g\bigl(\psi(\boldsymbol{x})\bigr)^2 \iint_{\psi(\Gamma^1_{\boldsymbol{x},\delta})} \exp\bigl(-(n-2)(1+\epsilon)^2 g\bigl(\psi(\boldsymbol{x})\bigr)\\
&\qquad \times \lambda\bigl(B^m\bigl(z_j, \|z_j - \psi(\boldsymbol{x})\|\bigr) \cup B^m\bigl(z_k, \|z_k - \psi(\boldsymbol{x})\|\bigr)\bigr)\bigr) g(z_j) g(z_k)\,\mathrm{d}\lambda(z_j)\,\mathrm{d}\lambda(z_k)\\
&\quad\ge (1-\epsilon)^2 n^2 g\bigl(\psi(\boldsymbol{x})\bigr)^2 \iint_{\Gamma_{m,\delta}} \exp(-(n-2)(1+\epsilon)^2 g\bigl(\psi(\boldsymbol{x})\bigr)\\
&\qquad \times \lambda\bigl(B^m\bigl(z_j, \|z_j\|\bigr) \cup \bigl(B^m\bigl(z_k, \|z_k\|\bigr)\bigr)\bigr)\,\mathrm{d}\lambda(z_j)\,\mathrm{d}\lambda(z_k).
\end{aligned}$$

Again use a translation

$$z_j^{**} := r_m z_j, \qquad z_k^{**} := r_m z_k,$$

and thus the corresponding scaling result

$$\lambda(B^m\bigl(z_j^{**}, \|z_j^{**}\|\bigr) \cup \bigl(B^m\bigl(z_k^{**}, \|z_k^{**}\|\bigr)\bigr) = (r_m)^m \lambda(B^m\bigl(z_j, \|z_j\|\bigr) \cup \bigl(B^m\bigl(z_k, \|z_k\|\bigr)\bigr).$$

Setting $r_m := ((n-2)(1+\epsilon)^2 g(\psi(\boldsymbol{x})))^{1/m}$ and noting that the corresponding Jacobi matrix satisfies $|J| = (n-2)(1+\epsilon)^2 g(\psi(\boldsymbol{x}))$, we have

$$\begin{aligned}
J_{n,1} &\ge \frac{(1-\epsilon)n^2}{(1+\epsilon)^4(n-2)^2}\\
&\quad \times \iint_{\Gamma_{m,l_m\delta}} \exp(-\lambda\bigl(B^m\bigl(z_j^{**}, \|z_j^{**}\|\bigr) \cup \bigl(B^m\bigl(z_k^{**}, \|z_k^{**}\|\bigr)\bigr)\bigr)\,\mathrm{d}\lambda\bigl(z_j^{**}\bigr)\,\mathrm{d}\lambda\bigl(z_k^{**}\bigr).
\end{aligned} \tag{2.21}$$

Following the same procedure as we discussed before, as n goes to infinity, we then obtain the same lower bound.

Matching the upper and lower bounds then yields

$$\lim_{n\to\infty} \mathrm{E}M(\boldsymbol{X}_i) = \lim_{n\to\infty} n^2 \mathrm{P}(Q_j \cap Q_k | \boldsymbol{X}_i = \boldsymbol{x}) = \lim_{n\to\infty} I_{n,2} = \mathfrak{o}_m.$$

Finally,

$$\lim_{n\to\infty} \frac{\mathrm{E}M_{\text{total}}}{n} = \lim_{n\to\infty} \mathrm{E}\sum_{i=1}^n \frac{M(\boldsymbol{X}_i)}{n} = \lim_{n\to\infty} \sum_{i=1}^n \frac{\mathrm{E}M(\boldsymbol{X}_i)}{n} = \mathfrak{o}_m.$$

This completes the proof. □

2.5. *Proof of Theorem* 1.3.

PROOF OF THEOREM 1.3. Our proof of Theorem 1.3 is similar to that of Theorem 1.1. In detail, the proof of Theorem 1.1 reveals that the asymptotic variance is determined by the limiting proportion of two graph structures shown in Lemmas 1.4 and 1.5. The proof of Theorem 1.3 is based on the same idea. The following lemmas are in parallel to Lemmas 1.4 and 1.5.

LEMMA 2.8 (Expected number of nearest-neighbor pairs under difference-based kernel metric). *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ to be the NNG induced by $f(\cdot)$, and $\mathcal{M}$ and μ satisfying the assumptions of Theorem* 1.1(i) *and Assumption* 1.1, *respectively. We then have, as $n \to \infty$,*

$$\mathrm{E}\bigg(\frac{1}{n}\#\{(i,j)\ \textit{distinct}: i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\}\bigg) \to \mathfrak{q}_m.$$

LEMMA 2.9. *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ to be the NNG induced by $f(\cdot,\cdot)$, and $\mathcal{M}$ and μ satisfying the assumptions of Theorem* 1.1(i) *and Assumption* 1.1, *respectively. We then have, as $n \to \infty$,*

$$\mathrm{E}\bigg(\frac{1}{n}\#\{(i,j,k)\ \textit{distinct}: i \to k, j \to k \in \mathcal{E}(\mathcal{G}_n)\}\bigg) \to \mathfrak{o}_m.$$

PROOF OF LEMMA 2.8. We first recall the notation used in the previous proofs. Let $N(\boldsymbol{X}_i)$ denote the number of NN pairs containing $\boldsymbol{X}_i$ and N_{total} denote the total number of NN pairs (double edges) in NNG; in other words,

$$N_{\text{total}} := \#\{(i,j)\ \textit{distinct}: i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\},$$
$$N(\boldsymbol{X}_i) := \#\{j : i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\}.$$

Assumption 1.2 implies an upper bound on $N(\boldsymbol{X}_i)$, that is,

$$N(\boldsymbol{X}_i) \le C_{d,K} < \infty.$$

Since by the dominant convergence theorem, we have

$$\lim_{n\to\infty} \mathrm{E}N(\boldsymbol{X}_i) = \lim_{n\to\infty} \int \mathrm{E}\big(N(\boldsymbol{x}_i)\big)\,\mathrm{d}\mu(\boldsymbol{x}_i) = \int \lim_{n\to\infty} \mathrm{E}\big(N(\boldsymbol{x}_i)\big)\,\mathrm{d}\mu(\boldsymbol{x}_i),$$

it suffices to find the limit of $\mathrm{E}(N(\boldsymbol{x}_i))$ given the sequence of graph $\mathcal{G}_n$. Denote $U_K^d(\boldsymbol{x}_1,\boldsymbol{x}_2)$ to be the union of two balls under metric $D(\cdot,\cdot)$ induced by $K(\cdot,\cdot)$ in $\mathbb{R}^d$ whose centers are $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and radius are both $D(\boldsymbol{x}_1,\boldsymbol{x}_2)$. The same analysis as in equation (2.5) gives us that

$$\mathrm{E}\big(N(\boldsymbol{x}_i)\big) = (n-1)\int_{\mathcal{M}} \big(1 - \mu\big(U_K^d(\boldsymbol{x}_i,\boldsymbol{x}_j)\big)\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j). \tag{2.22}$$

To get a good approximation of $\mu(U_K^d(\boldsymbol{x}_i,\boldsymbol{x}_j))$, we consider the linearization version of the kernel-based matric. Define $H(\boldsymbol{x}) := \nabla^2 f(\boldsymbol{x})$ for $\boldsymbol{x}$ in a neighborhood of $\boldsymbol{0}$. By Assumption 1.2, $H(\boldsymbol{0})$ is reversible, and we define $\Sigma = H(\boldsymbol{0})^{-1}$. The corresponding Mahalanobis distance is defined as $\|\boldsymbol{x}\|_\Sigma := \sqrt{\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}}$. We then approximate the kernel-based metric based on this Mahalanobis distance. Recall that λ is denoted to be the m-dimensional Lebesgue measure.

LEMMA 2.10. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart $\psi (= \psi_{\boldsymbol{x}})$, it holds true that*

$$\lim_{\|\boldsymbol{x}_j - \boldsymbol{x}\|_\Sigma \to 0} \frac{\mu(U_K^d(\boldsymbol{x},\boldsymbol{x}_j))}{g(\psi(\boldsymbol{x}))\|\boldsymbol{x}_j - \boldsymbol{x}\|_\Sigma^m U_{m,\Sigma,\boldsymbol{x}}} = 1,$$

where $U_{m,\Sigma,\boldsymbol{x}}$ is the volumn (m-dimensional Lebesgue measure) of $U_K^d(\boldsymbol{x},\boldsymbol{y}) \cap \pi_{\boldsymbol{x}}$ with $\|\boldsymbol{x} - \boldsymbol{y}\|_\Sigma = 1$, and $\pi_{\boldsymbol{x}}$ denoting the tangent space of $\mathcal{M}$ at $\boldsymbol{x}$.

It can be explicitly shown that $U_{m,\Sigma,\boldsymbol{x}} = U_m \cdot (\prod_{i=1}^m \lambda_i(\mathrm{Proj}_{\pi_{\boldsymbol{x}}}\Sigma))^{1/2}$; here we use $\mathrm{Proj}_\pi(A)$ to denote the projection of the matrix A onto π, that is, for a projection matrix P_π (projecting onto π), $\mathrm{Proj}_\pi(A) := P_\pi A P_\pi^T$, and $\lambda_i(A)$ to denote the ith largest eigenvalue of A.

LEMMA 2.11. *For μ-a.e. $\boldsymbol{x} \in \mathcal{M}$ and the corresponding chart $\psi (= \psi_{\boldsymbol{x}})$, it holds true that*

$$\lim_{\|\boldsymbol{x}_j - \boldsymbol{x}\|_\Sigma \to 0} \frac{\mu(B^d(\boldsymbol{x}, r))}{g(\psi(\boldsymbol{x}))r^m V_{m,\Sigma,\boldsymbol{x}}} = 1,$$

where $V_{m,\Sigma,\boldsymbol{x}}$ is defined to be the volumn (m-dimensional Lebesgue measure) of $B^d(\boldsymbol{x}, 1) \cap \pi_{\boldsymbol{x}}$, and $B^d(\boldsymbol{x}, r)$ is the ball under the Mahalanobis distance with center $\boldsymbol{x}$ and radius r.

Similarly, we have $V_{m,\Sigma,\boldsymbol{x}} = V_m \cdot (\prod_{i=1}^m \lambda_i(\mathrm{Proj}_{\pi_{\boldsymbol{x}}} \Sigma))^{1/2}$.

With Lemmas 2.10 and 2.11, we can then first get the lower bound of the integral in equation (2.22) following the same procedure as in inequalities (2.8), (2.10), and (2.12). For arbitrary $\epsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned}
\mathrm{E}(N(\boldsymbol{x}_i)) &= (n-1) \int_{\mathcal{M}} \bigl(1 - \mu(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j))\bigr)^{n-2} \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&\geq (n-1) \int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_\Sigma \leq \delta} \bigl(1 - (1+\epsilon) g(\psi(\boldsymbol{x}_i)) \|\boldsymbol{x}_j - \boldsymbol{x}_i\|_\Sigma^m U_{m,\Sigma,\boldsymbol{x}_i}\bigr)^{n-2} \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&\geq (n-1) \int_{\phi_{1,\Sigma}(\delta)}^1 \frac{1-\epsilon}{1+\epsilon} \frac{V_{m,\Sigma}}{U_{m,\Sigma,\boldsymbol{x}_i}} \bigl(1 - t^{\frac{1}{n-2}}\bigr) \,\mathrm{d}t \\
&\geq \frac{1-\epsilon}{1+\epsilon} \frac{V_{m,\Sigma,\boldsymbol{x}_i}}{U_{m,\Sigma,\boldsymbol{x}_i}} \bigl(1 - (n-1)\phi_{1,\Sigma}(\delta)\bigr),
\end{aligned}$$

where

$$\phi_{1,\Sigma}(\delta) := \bigl(1 - (1+\epsilon) g(\psi(\boldsymbol{x}_i)) r^m U_{m,\Sigma,\boldsymbol{x}_i}\bigr)^{n-2} I_{\{r \leq \delta\}}.$$

Let n go to infinity, by direct calculation we then have

$$\lim_{n\to\infty} E(N(\boldsymbol{x}_i)) \geq \frac{1-\epsilon}{1+\epsilon} \frac{V_{m,\Sigma,\boldsymbol{x}_i}}{U_{m,\Sigma,\boldsymbol{x}_i}}.$$

Since ϵ is arbitrary, we then have

$$\lim_{n\to\infty} E(N(\boldsymbol{x}_i)) \geq V_{m,\Sigma,\boldsymbol{x}_i} / U_{m,\Sigma,\boldsymbol{x}_i} = V_m / U_m.$$

To get the upper bound of the integral in equation (2.22), we can follow the same procedure as in inequalities (2.13), (2.14), and (2.16). For arbitrary $\epsilon > 0$, we choose $\delta > 0$ such that

$$\begin{aligned}
\mathrm{E}(N(\boldsymbol{x}_i)) &= (n-1) \int_{\mathcal{M}} \bigl(1 - \mu(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j))\bigr)^{n-2} \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&\leq (n-1) \int_{\mathcal{M}} \exp\bigl(-(n-2)\mu(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})\bigr) \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&= (n-1) \int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_\Sigma > \delta} \exp\bigl(-(n-2)\mu(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})\bigr) \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&\quad + (n-1) \int_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_\Sigma \leq \delta} \exp\bigl(-(n-2)\mu(U_k^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})\bigr) \,\mathrm{d}\mu(\boldsymbol{x}_j) \\
&\leq (n-1) \exp\Bigl(-(n-2) \inf_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\|_\Sigma \geq \delta} \bigl\{\mu(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})\bigr\}\Bigr) \int_{\mathcal{M}} \mathrm{d}\mu(\boldsymbol{x}_j) \\
&\quad + (n-1)\phi_{2,\Sigma}(\delta) + (n-1) \int_{\phi_{2,\Sigma}(\delta)}^1 (1+\epsilon) V_{m,\Sigma,\boldsymbol{x}_i} \bigl(\phi_{2,\Sigma}^{-1}(t)\bigr)^m \,\mathrm{d}t,
\end{aligned}$$

where

$$\phi_{2,\Sigma}(r) := \exp\big(-(n-2)(1-\epsilon)^2 g(\psi(\boldsymbol{x}_i)) r^m U_{m,\Sigma,\boldsymbol{x}_i}\big) I_{\{r\le\delta\}}.$$

Direct calculation then yields

$$\lim_{n\to\infty} E\big(N(\boldsymbol{x}_i)\big) \le \frac{1+\epsilon}{1-\epsilon}\frac{V_{m,\Sigma,\boldsymbol{x}_i}}{U_{m,\Sigma,\boldsymbol{x}_i}}.$$

Since ϵ is arbitrary, we then have

$$\lim_{n\to\infty} E\big(N(\boldsymbol{x}_i)\big) \le V_{m,\Sigma,\boldsymbol{x}_i}/U_{m,\Sigma,\boldsymbol{x}_i} = V_m/U_m.$$

Combining the upper and lower bounds of $\mathrm{E}(N(\boldsymbol{x}_i))$, we then have

$$\lim_{n\to\infty} E\big(N(\boldsymbol{x}_i)\big) = \frac{V_{m,\Sigma,\boldsymbol{x}_i}}{U_{m,\Sigma,\boldsymbol{x}_i}} = \frac{V_m}{U_m} = \mathfrak{q}_m,$$

and consequently,

$$\lim_{n\to\infty} \mathrm{E}N(\boldsymbol{X}_i) = \int_{\mathcal{M}} \lim_{n\to\infty} \mathrm{E}N(\boldsymbol{x}_i)\,\mathrm{d}\mu(\boldsymbol{x}_i) = \mathfrak{q}_m.$$

Finally, we conclude that

$$\lim_{n\to\infty} \frac{\mathrm{E}N_{\text{total}}}{n} = \lim_{n\to\infty}\sum_{i=1}^n \frac{\mathrm{E}(N(\boldsymbol{X}_i))}{n} = \mathfrak{q}_m$$

and thus complete the proof. □

PROOF OF LEMMA 2.9. The proof is similar to that of Lemma 1.5. Denote

$$Q_j := \{j \to i \in \mathcal{E}(\mathcal{G}_n)\}$$

for any $j \neq i$ and put

$$M_{\text{total}} = \#\{(i,j,k)\ \textit{distinct} : j\to i, k\to i \in \mathcal{E}(\mathcal{G}_n)\},$$
$$D(\boldsymbol{X}_i) = \sum_{j\neq i} I_{Q_j} = \#\{j : j\to i \in \mathcal{E}(\mathcal{G}_n)\},$$
$$M(\boldsymbol{X}_i) = \#\{(j,k)\ \textit{distinct} : j\to i, k\to i \in \mathcal{E}(\mathcal{G}_n)\}.$$

Here $M(\boldsymbol{X}_i)$ is the coefficient we are interested in. Observe that

$$M_{\text{total}} = \sum_{i=1}^n M(\boldsymbol{X}_i),$$
$$M(\boldsymbol{X}_i) = D(\boldsymbol{X}_i)\big(D(\boldsymbol{X}_i)-1\big),$$

so that we have

$$\begin{aligned}\mathrm{E}M(\boldsymbol{X}_i) &= \mathrm{E}D(\boldsymbol{X}_i)\big(D(\boldsymbol{X}_i)-1\big)\\ &= \mathrm{E}\sum_{\substack{(j,k)\ \textit{distinct}\\ j,k\neq i}} \mathrm{P}(Q_j\cap Q_k)\\ &= (n-1)(n-2)\mathrm{P}(Q_j\cap Q_k).\end{aligned}$$

We first consider a fixed sample point $X_i = x_i$. Using Assumption 1.2, we have

$$C_{d,K}^2 \geq \mathrm{E}D(X_i|X_i = x_i)\big(D(X_i|X_i = x_i) - 1\big)$$
$$= (n-1)(n-2)\mathrm{P}(Q_j \cap Q_k | X_i = x_i).$$

Then applying the Lebesgue dominated convergence theorem, we have

$$\lim_{n\to\infty}(n-1)(n-2)\mathrm{P}(Q_j \cap Q_k) = \lim_{n\to\infty}\int_{\mathcal{M}}(n-1)(n-2)\mathrm{P}(Q_j \cap Q_k|X_i = x_i)\,\mathrm{d}\mu(x_i)$$
$$= \int_{\mathcal{M}}\lim_{n\to\infty}(n-1)(n-2)\mathrm{P}(Q_j \cap Q_k|X_i = x_i)\,\mathrm{d}\mu(x_i).$$

It remains to analyze the limit term

$$\lim_{n\to\infty}(n-1)(n-2)\mathrm{P}(Q_j \cap Q_k|X_i = x_i).$$

We first introduce the notation to simplify the integral. For a fixed point $x \in \mathcal{M}$, let

$$\Gamma_{x,\Sigma} := \big\{(x_j, x_k) \in (\mathbb{R}^d)^2 : \max\{\|x_j - x\|_\Sigma, \|x_k - x\|_\Sigma\} \leq \|x_j - x_k\|_\Sigma\big\},$$
$$S_{j,\Sigma} := B(x_j, \|x_j - x\|_\Sigma) \cap \mathcal{M}, \quad \text{and} \quad S_{k,\Sigma} := B(x_k, \|x_k - x\|_\Sigma) \cap \mathcal{M}.$$

We then have

$$n^2\mathrm{P}(Q_j \cap Q_k|X_i = x) = \iint_{\Gamma_x} n^2\big(1 - \mu(S_{j,\Sigma} \cup S_{k,,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(x_j)\,\mathrm{d}\mu(x_k).$$

We then perform a similar operation to split the region of integral so that $\Gamma_{x,\Sigma} = \Gamma^1_{x,\delta,\Sigma} \cup \Gamma^2_{x,\delta,\Sigma}$. Here

$$\Gamma^1_{x,\delta,\Sigma} := \Gamma_{x,\Sigma} \cap \big\{(x_1, x_2) \in (\mathbb{R}^d)^2 : \|x_1 - x\|_\Sigma, \|x_2 - x\|_\Sigma \leq \delta\big\},$$
$$\Gamma^2_{x,\delta,\Sigma} := \Gamma_{x,\Sigma} \setminus \Gamma^1_{x,\delta},$$

where δ is a sufficiently small but positive real number to be defined. Taking this idea, we partition the integral into two parts:

$$n^2\mathrm{P}(Q_j \cap Q_k|X_i = x) = \iint_{\Gamma^1_{x,\delta,\Sigma}} n^2\big(1 - \mu(S_{j,\Sigma} \cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(x_j)\,\mathrm{d}\mu(x_k)$$
$$+ \iint_{\Gamma^2_{x,\delta,\Sigma}} n^2\big(1 - \mu(S_{j,\Sigma} \cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(x_j)\,\mathrm{d}\mu(x_k).$$

Analogously to Lemma 2.4, we can directly prove a parallel version of it.

LEMMA 2.12. *For μ-a.e. $x \in \mathcal{M}$ and the corresponding chart ψ, it holds true that*

$$\lim_{\|x_j - x\|_\Sigma \to 0} \frac{\mu(B^d(x_j, \|x - x_j\|_\Sigma)))}{g(\psi(x))\lambda(B^d(x_j, \|x - x_j\|_\Sigma) \cap \pi_x)} = 1,$$

where π_x is the tangent space of $\mathcal{M}$ at x.

With Lemma 2.12, the same argument as in Lemma 1.5 gives us that

$$0 \leq \iint_{\Gamma^2_{x,\delta,\Sigma}} n^2\big(1 - \mu(S_{j,\Sigma} \cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(x_j)\,\mathrm{d}\mu(x_k)$$
$$\leq n^2\bigg(1 - (1-\epsilon)g\big(\psi(x)\big)V_{m,\Sigma}\bigg(\frac{\delta}{2}\bigg)^m\bigg)^{n-2} \to 0.$$

To bound the other term

$$\iint_{\Gamma^1_{\boldsymbol{x},\delta,\Sigma}} n^2\big(1-\mu(S_{j,\Sigma}\cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k),$$

we apply the following lemma (cf. Lemma 2.5).

LEMMA 2.13. *For μ-a.e. $\boldsymbol{x}\in\mathcal{M}$ and the corresponding chart ψ, we have*

$$\lim_{\substack{\|\boldsymbol{x}_j-\boldsymbol{x}\|_\Sigma\to 0\\ \|\boldsymbol{x}_k-\boldsymbol{x}\|_\Sigma\to 0}} \frac{\mu((B^d(\boldsymbol{x}_j,\|\boldsymbol{x}_j-\boldsymbol{x}\|_\Sigma)\cup B^d(\boldsymbol{x}_k,\|\boldsymbol{x}_k-\boldsymbol{x}\|_\Sigma)))}{g(\psi(\boldsymbol{x}))\lambda(B^d(z_j,\|z_j-\psi(\boldsymbol{x})\|_\Sigma)\cup B^d(z_k,\|z_k-\psi(\boldsymbol{x})\|_\Sigma)\cap\pi_{\boldsymbol{x}})}=1,$$

where $z_j=\psi(\boldsymbol{x}_j)$, $z_k=\psi(\boldsymbol{x}_k)$, and B^d represents the ball under the Mahalanobis distance.

The same argument as in the derivation of inequality (2.20) shows

$$\begin{aligned}&\lim_{n\to\infty}\iint_{\Gamma^1_{\boldsymbol{x},\delta,\Sigma}} n^2\big(1-\mu(S_{j,\Sigma}\cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\&\quad\le\frac{1+\epsilon}{(1-\epsilon)^2}\iint_{\Gamma_{m,\Sigma,\boldsymbol{x}}}\exp(-\lambda\big(B^d\big(z_j^*,\|z_j^*\|_\Sigma\big)\cup\big(B^d\big(z_k^*,\|z_k^*\|_\Sigma\big)\cap\pi_{\boldsymbol{x}}\big)\big)\,\mathrm{d}\lambda\big(z_j^*\big)\,\mathrm{d}\lambda\big(z_k^*\big),\end{aligned}$$

where

$$\Gamma_{m,\Sigma,\boldsymbol{x}}:=\big\{(\boldsymbol{w}_1,\boldsymbol{w}_2)\in(\pi_{\boldsymbol{x}})^2:\max\big(\|\boldsymbol{w}_1-\boldsymbol{x}\|_\Sigma,\|\boldsymbol{w}_2-\boldsymbol{x}\|_\Sigma\big)<\|\boldsymbol{w}_1-\boldsymbol{w}_2\|_\Sigma\big\}.$$

Here with a little abuse of notion, we use $\pi_{\boldsymbol{x}}$ to denote the tangent plane of $\mathcal{M}$ at $\boldsymbol{x}$. Since z_j^* and z_k^* are supported on a m-diemnsional subspace of $\mathbb{R}^d$, we still use m in the index of $\Gamma_{m,\Sigma,\boldsymbol{x}}$. In the following, we write $\Gamma_{m,\Sigma}$ as a shorthand of $\Gamma_{m,\Sigma,\mathbf{0}}$. Again we use the derivation of inequality (2.21) to get the lower bound:

$$\begin{aligned}&\lim_{n\to\infty}\iint_{\Gamma^1_{\boldsymbol{x},\delta,\Sigma}} n^2\big(1-\mu(S_{j,\Sigma}\cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\&\quad\ge\frac{(1-\epsilon)}{(1+\epsilon)^4}\\&\qquad\times\iint_{\Gamma_{m,\Sigma,\boldsymbol{x}}}\exp(-\lambda\big(B^d\big(z_j^{**},\|z_j^{**}\|_\Sigma\big)\cup\big(B^d\big(z_k^{**},\|z_k^{**}\|_\Sigma\big)\big)\cap\pi_{\boldsymbol{x}}\big)\,\mathrm{d}\lambda\big(z_j^{**}\big)\,\mathrm{d}\lambda\big(z_k^{**}\big).\end{aligned}$$

Finally, since $\epsilon>0$ is arbitrarily chosen, we conclude that

$$\begin{aligned}&\lim_{n\to\infty}\iint_{\Gamma^1_{\boldsymbol{x},\delta,\Sigma}} n^2\big(1-\mu(S_{j,\Sigma}\cup S_{k,\Sigma})\big)^{n-2}\,\mathrm{d}\mu(\boldsymbol{x}_j)\,\mathrm{d}\mu(\boldsymbol{x}_k)\\&\quad=\iint_{\Gamma_{m,\Sigma}}\exp(-\lambda\big(B^d\big(z_j,\|z_j\|_\Sigma\big)\cup\big(B^d\big(z_k,\|z_k\|_\Sigma\big)\cap\pi_{\boldsymbol{x}}\big)\big)\,\mathrm{d}\lambda\big(z_j\big)\,\mathrm{d}\lambda\big(z_k\big),\end{aligned}$$

where $\pi_{\boldsymbol{x}}$ is the tangent space of $\mathcal{M}$ at $\boldsymbol{x}$. It can be regarded as the m-dimensional subspace in $\mathbb{R}^d$ which is tangent to $\mathcal{M}$ at $\boldsymbol{x}$. Thus, we have

$$\begin{aligned}\lim_{n\to\infty}\mathrm{E}M(\boldsymbol{X}_i)&=\int_{\mathcal{M}}\lim_{n\to\infty}n^2\mathrm{P}(Q_j\cap Q_k|\boldsymbol{X}_i=\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})\\&=\int_{\mathcal{M}}\iint_{\Gamma_{m,\Sigma}}\exp(-\lambda\big(B^d\big(z_j,\|z_j\|_\Sigma\big)\cup\big(B^d\big(z_k,\|z_k\|_\Sigma\big)\cap\pi_{\boldsymbol{x}}\big)\big)\,\mathrm{d}\lambda\big(z_j\big)\,\mathrm{d}\lambda\big(z_k\big)\,\mathrm{d}\mu(\boldsymbol{x})\\&=:\Sigma_1'(m,K,P_X).\end{aligned}$$

Now we go one step further to calcuate $\Sigma_1'(m, K, P_X)$. Denote the eigendecomposition of $P_{\pi_x}^T \Sigma P_{\pi_x}$ as $P_{\pi_x}^T \Sigma P_{\pi_x} = U \Lambda U^T$, where U is the orthogonal matrix and Λ is the diagonal matrix with eigenvalues on the diagonal. Since Σ is a positive definite matrix, we have $\Lambda_{ii} > 0$ for all $i = 1, \dots, m$, and $\Lambda_{ii} = 0$ for $i > m$. Consider the change of variables in the integral as

$$\boldsymbol{w}_i = \lambda \Lambda^{-\frac{1}{2}} U^{-1} z_i, \qquad \boldsymbol{w}_j = \lambda \Lambda^{-\frac{1}{2}} U^{-1} z_j,$$

where $\lambda = \prod_{i=1}^m \Lambda_{ii}^{1/2m}$ is a normalized constant to force the Jacobian of the transformation to be 1. It can be checked that $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$ are supported on $\pi_{\boldsymbol{x}}$, and also,

$$\lambda(B^m(\boldsymbol{w}_j, \|\boldsymbol{w}_j\|) \cup (B^m(\boldsymbol{w}_k, \|\boldsymbol{w}_k\|))) = \lambda(B^d(z_j, \|z_j\|_\Sigma) \cup (B^d(z_k, \|z_k\|_\Sigma) \cap \pi_{\boldsymbol{x}}), \tag{2.23}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^m$. Plugging equation (2.23) into the expression of $\Sigma_1'(m, K, P_X)$, we have

$$\begin{aligned}
&\Sigma_1'(m, K, P_X) \\
&= \int_{\mathcal{M}} \iint_{\Gamma_{m,\Sigma}} \exp(-\lambda(B^d(z_j, \|z_j\|_\Sigma) \cup (B^d(z_k, \|z_k\|_\Sigma) \cap \pi_{\boldsymbol{x}})) \, \mathrm{d}\lambda(z_j) \, \mathrm{d}\lambda(z_k) \, \mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathcal{M}} \iint_{\Gamma_{m;2}} \exp(-\lambda(B^m(\boldsymbol{w}_j, \|\boldsymbol{w}_j\|_2) \cup (B^m(\boldsymbol{w}_k, \|\boldsymbol{w}_k\|_2)))) \, \mathrm{d}\lambda(\boldsymbol{w}_j) \, \mathrm{d}\lambda(\boldsymbol{w}_k) \, \mathrm{d}\mu(\boldsymbol{x}) \\
&= \mathfrak{o}_m,
\end{aligned}$$

so that

$$\lim_{n\to\infty} \frac{\mathrm{E}M_{\text{total}}}{n} = \lim_{n\to\infty} \sum_{i=1}^n \frac{\mathrm{E}M(\boldsymbol{X}_i)}{n} = \mathfrak{o}_m.$$

This completes the proof. □

With Lemmas 2.8 and 2.9, we are now ready to prove Theorem 1.3. Plugging the limits of $\mathrm{E}M_{\text{total}}/n$ and $\mathrm{E}N_{\text{total}}/n$ to the expression of asymptotic variance, we obtain

$$\lim_{n\to\infty} \operatorname{Var}(\sqrt{n}\xi_n) = \frac{4}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m.$$

We then show the asymptotic normality of $\sqrt{n}\xi_n$ by leveraging Theorem 4.1 in Deb, Ghosal and Sen (2020). Since here we consider the directed nearest neighbor graph, the out-degree of each node is 1. To check the asymptotic normality, we need to verify all the assumptions in Theorem 4.1 in Deb, Ghosal and Sen (2020). To verify (A2) and (A3), adopting the notation in Deb, Ghosal and Sen (2020), we choose $t_n = r_n := 1$ and $q_n := C_{d,K} + 1$ for all n. Since the out-degree of the node indexed by i is $d_i = 1$. Thus, (A2) and (A3) are satisfied with these chosen parameters. Noticing that

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{d_i}{(\log n)^\gamma} = 0,$$

for any $\gamma > 0$, we can just choose $D := C_{d,K} + 2 > 0$ to ensure $r_n^{-1}(q_n + t_n) \le D$, which makes sure that the additional conditions in Theorem 4.1 in Deb, Ghosal and Sen (2020) are satisfied in our setting.

To wrap up, $\sqrt{n}\xi_n$ is asymptotically normal with mean 0 and variance $\frac{4}{5} + \frac{2}{5}\mathfrak{q}_m + \frac{4}{5}\mathfrak{o}_m$. □

2.6. *Proof of Theorem* 1.4.

PROOF OF THEOREM 1.4. The proof of Theorem 1.4 is similar to that of Theorem 1.3. Therefore, we do not repeat the whole calculation process, but only present the major modifications.

Since we study a smooth manifold $\mathcal{M}$, the natural parameterization mapping ψ is smooth and is locally bijective. This argument makes sure the existence of its inverse mapping $\varphi = \psi^{-1}$. Also, φ is a smooth function guaranteed by the definition of smooth manifold. Denote

$$\Sigma_{\boldsymbol{x}} = \big((\nabla\varphi_{\boldsymbol{x}}(\boldsymbol{x}))^T \nabla\varphi_{\boldsymbol{x}}(\boldsymbol{x})\big)^{-1},$$

where $\varphi_{\boldsymbol{x}}$ is the inverse mapping of natural parametrization mapping at the point $\boldsymbol{x} \in \mathcal{M}$. Here, by Tietze extension theorem (Dugundji (1951)), we can consider $\varphi_{\boldsymbol{x}}$ as a function on $\mathbb{R}^d$ without loss of generality, and the derivative operator is defined in this sense. Since the mapping is bijective, $\nabla\varphi_{\boldsymbol{x}}(\boldsymbol{x})$ is full-rank, so the inverse of the matrix is well defined.

With this locally defined matrix, we then introduce the Mahalanobis distance on the neighborhood of $\boldsymbol{x}$ as

$$\|\boldsymbol{y}\|_{\Sigma_x} = \sqrt{\boldsymbol{y}^T \Sigma_{\boldsymbol{x}}^{-1} \boldsymbol{y}},$$

for arbitrary fixed point $\boldsymbol{x} \in \mathcal{M}$. The proofs of Lemmas 2.10 and 2.11 can then be directly applied to the geodesic distance context. We state the geodesic distance version of Lemmas 2.10 and 2.11 below; their proofs are omitted.

LEMMA 2.14 (Expected number of nearest-neighbor pairs under goedesic distance). *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ to be the NNG induced by the geodesic distance $d_g(\cdot,\cdot)$, and $\mathcal{M}$ and μ satisfying the assumptions of Theorem* 1.1(i), *respectively. We then have*, *as $n \to \infty$*,

$$\mathrm{E}\bigg(\frac{1}{n}\#\big\{(i,j)\ distinct : i \to j, j \to i \in \mathcal{E}(\mathcal{G}_n)\big\}\bigg) \to \mathfrak{q}_m.$$

LEMMA 2.15. *Consider $\mathcal{G}_n$ in $\mathbb{R}^d$ to be the NNG induced by metric $f(\cdot,\cdot)$, and $\mathcal{M}$ and μ satisfying the assumptions of Theorem* 1.1(i), *respectively. We then have*, *as $n \to \infty$*,

$$\mathrm{E}\bigg(\frac{1}{n}\#\big\{(i,j,k)\ distinct : i \to k, j \to k \in \mathcal{E}(\mathcal{G}_n)\big\}\bigg) \to \mathfrak{o}_m.$$

With Lemmas 2.14 and 2.15, we get the asymptotic variance of the geodesic distance NNG estimator, and the CLT is a direct consequence of Deb, Ghosal and Sen ((2020), Theorem 4.1). □

3. Proofs of the rest results.

3.1. *Proof of Lemma* 1.1(a).

PROOF OF LEMMA 1.1(a). Recall the expression for $\mathfrak{q}_m$:

$$\mathfrak{q}_m := \bigg\{2 - I_{3/4}\bigg(\frac{m+1}{2}, \frac{1}{2}\bigg)\bigg\}^{-1}, \qquad I_x(a,b) := \frac{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}.$$

To prove the monotone of $\mathfrak{q}_m$ as the dimension m increases, it suffices to show that $I_x(a,b)$ decreases as argument a increases when $x \in [0,1]$ and $a, b > 0$. For some $\epsilon > 0$, we directly compare the values of $I_x(a,b)$ and $I_x(a+\epsilon,b)$. We have

$$\big(I_x(a,b)\big)^{-1} = 1 + \frac{\int_x^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t},$$

$$\big(I_x(a+\epsilon,b)\big)^{-1} = 1 + \frac{\int_x^1 t^{a-1}t^{\epsilon}(1-t)^{b-1}\,\mathrm{d}t}{\int_0^x t^{a-1}t^{\epsilon}(1-t)^{b-1}\,\mathrm{d}t}.$$

Observe that

$$\int_x^1 t^{a-1}t^{\epsilon}(1-t)^{b-1}\,\mathrm{d}t > x^{\epsilon}\int_x^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t,$$

$$\int_0^x t^{a-1}t^{\epsilon}(1-t)^{b-1}\,\mathrm{d}t < x^{\epsilon}\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t,$$

which implies that $(I_x(a,b))^{-1} < (I_x(a+\epsilon,b))^{-1}$. Equivalently, $I_x(a+\epsilon,b) < I_x(a,b)$ for every $\epsilon > 0$.

With monotony, we have $\mathfrak{q}_m \in (\mathfrak{q}_\infty, \mathfrak{q}_1]$, where $\mathfrak{q}_\infty := \lim_{m\to\infty}\mathfrak{q}_m$. Here, $\mathfrak{q}_1 = 2/3$, and for $\mathfrak{q}_\infty$, notice that for some sufficient small $\delta > 0$,

$$\frac{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_x^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t} \le \frac{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_{x+\delta}^{1-\delta} t^{a-1}(1-t)^{b-1}\,\mathrm{d}t} \le \frac{x^a}{(1-x-2\delta)(x+\delta)^{a-1}\delta^{b-1}}.$$

Thus, $\lim_{a\to\to} I_x(a,b) = 0$, and hence $\mathfrak{q}_\infty = 1/2$.

In conclusion, $\mathfrak{q}_m \in (1/2, 2/3]$ is strictly decreasing as m increases. □

3.2. *Proof of Lemma* 1.1(b).

PROOF OF LEMMA 1.1(b). Recall the expression for $\mathfrak{o}_m$:

$$\mathfrak{o}_m := \iint_{\Gamma_{m;2}} \exp\big[-\lambda\big\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cup B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\big\}\big]\,\mathrm{d}(\boldsymbol{w}_1, \boldsymbol{w}_2),$$

$$\Gamma_{m;2} := \big\{(\boldsymbol{w}_1, \boldsymbol{w}_2) \in (\mathbb{R}^m)^2 : \max(\|\boldsymbol{w}_1\|, \|\boldsymbol{w}_2\|) < \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|\big\}.$$

Considering the symmetry of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ is the integral, we have

$$\mathfrak{o}_m = 2\iint_{\Gamma^*_{m;2}} \exp\big[-\lambda\big\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cup B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\big\}\big]\,\mathrm{d}(\boldsymbol{w}_1, \boldsymbol{w}_2),$$

$$\text{where } \Gamma^*_{m;2} := \Gamma_{m;2} \cap \big\{(\boldsymbol{w}_1, \boldsymbol{w}_2) \in (\mathbb{R}^m)^2 : \|\boldsymbol{w}_1\| > \|\boldsymbol{w}_2\|\big\}.$$

We first prove that for any positive integer m, $\mathfrak{o}_m < 2$. The expression for $\mathfrak{o}_m$ means that

$$\begin{aligned}
\mathfrak{o}_m &< 2\iint_{\Gamma^*_{m;2}} \exp\big[-\lambda\big\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|)\big\}\big]\,\mathrm{d}\boldsymbol{w}_2\,\mathrm{d}\boldsymbol{w}_1 \\
&= 2\int_{\mathbb{R}^m} \exp\big[-\lambda\big\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|)\big\}\big] V_m \|\boldsymbol{w}_1\|^m\,\mathrm{d}\boldsymbol{w}_1 \\
&= 2\int_0^\infty \exp(-V_m t^m) V_m t^m \cdot (m V_m t^{m-1}\,\mathrm{d}t) \\
&= \int_0^\infty \exp(-V_m t^m)\,\mathrm{d}(V_m^2 t^{2m}) = 2,
\end{aligned}$$

in which we apply polar coordinates transformation and denote $t = \|\boldsymbol{w}_1\|$ in the transformation.

Then we consider the limit behavior and prove that $\limsup_m \mathfrak{o}_m \le 1$. The definition of $\Gamma^*_{m;2}$ shows that, for $(\boldsymbol{w}_1, \boldsymbol{w}_2) \in \Gamma^*_{m;2}$, we have $\boldsymbol{w}_1 \notin B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)$ and $\|\boldsymbol{w}_2\| < \|\boldsymbol{w}_1\|$. Thus, for fixed $\boldsymbol{w}_1$, the Lebesgue measure of $B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cap B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)$ can be bounded by the restrictions above, that is,

$$\begin{aligned}\lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cap B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\} &< \lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cap B(\boldsymbol{0}, \|\boldsymbol{w}_1\|)\} \\ &= \left(2 - \frac{2}{\mathfrak{q}_m}\right)\lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|)\}.\end{aligned}$$

We denote $\epsilon_m = 2 - 2/\mathfrak{q}_m$. According to Lemma 1.1(a), it is known that $\lim_{m\to\infty} \epsilon_m = 0$. Applying the estimation, we have the bound

$$\begin{aligned}&\lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cup B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\} \\ &\quad = \lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|)\} + \lambda\{B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\} - \lambda\{B(\boldsymbol{w}_1, \|\boldsymbol{w}_1\|) \cap B(\boldsymbol{w}_2, \|\boldsymbol{w}_2\|)\} \\ &\quad > (1 - \epsilon_m) V_m \|\boldsymbol{w}_1\|^m + V_m \|\boldsymbol{w}_2\|^m.\end{aligned}$$

We then get back to the expression for $\mathfrak{o}_m$:

$$\begin{aligned}\mathfrak{o}_m &< 2 \iint_{\Gamma^*_{m;2}} \exp[-(1-\epsilon_m) V_m \|\boldsymbol{w}_1\|^m - V_m \|\boldsymbol{w}_2\|^m] \,\mathrm{d}(\boldsymbol{w}_1, \boldsymbol{w}_2) \\ &= 2 \iint_{\Gamma^*_{m;2}} (\exp[-V_m \|\boldsymbol{w}_2\|^m] \,\mathrm{d}\boldsymbol{w}_2) \exp[-(1-\epsilon_m) V_m \|\boldsymbol{w}_1\|^m] \,\mathrm{d}\boldsymbol{w}_1 \\ &< 2 \int_{\mathbb{R}^m} \left(\int_{\boldsymbol{w}_2 : \|\boldsymbol{w}_2\| < \|\boldsymbol{w}_1\|} \exp[-V_m \|\boldsymbol{w}_2\|^m] \,\mathrm{d}\boldsymbol{w}_2 \right) \exp[-(1-\epsilon_m) V_m \|\boldsymbol{w}_1\|^m] \,\mathrm{d}\boldsymbol{w}_1 \\ &= 2 \int_{\mathbb{R}^m} (1 - e^{-V_m \|\boldsymbol{w}_1\|^m}) \exp[-(1-\epsilon_m) V_m \|\boldsymbol{w}_1\|^m] \,\mathrm{d}\boldsymbol{w}_1 \\ &= 2\left(\frac{1}{1-\epsilon_m} - \frac{1}{2\sqrt{1-\epsilon_m}}\right) \to 1.\end{aligned}$$

The calculation follows from the polar coordinates transformation as used before. In conclusion, we have $\limsup_m \mathfrak{o}_m \le 1$. Combining this with $\mathfrak{o}_m < 2$, we obtain $\sup_m \mathfrak{o}_m < 2$. □

3.3. *Proof of Lemma* 2.2.

PROOF OF LEMMA 2.2. We consider approximating $\mu(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j))$ as

$$\mu(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j)) = \psi_*\mu(\psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})) = \int_{\psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M})} g \,\mathrm{d}\lambda(\boldsymbol{x}_j).$$

Using the Lebesgue differentiation theorem (LDT), there exists some $\delta_1 > 0$ such that for μ-a.e. $\boldsymbol{x}_i$ and every $\boldsymbol{x}_j$ satisfying $\|\boldsymbol{x}_j - \boldsymbol{x}_i\| < \delta_1$, we have

$$\frac{\mu(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j))}{g(\psi(\boldsymbol{x}_i))\lambda(\psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}))} \in \left[1 - \frac{\epsilon}{3}, 1 + \frac{\epsilon}{3}\right]. \tag{3.1}$$

For every $\alpha > 0$, we first define

$$U^m_\alpha(\boldsymbol{x}, \boldsymbol{y}) := B^m(\boldsymbol{x}, \alpha\|\boldsymbol{x} - \boldsymbol{y}\|) \cup B^m(\boldsymbol{y}, \alpha\|\boldsymbol{x} - \boldsymbol{y}\|).$$

Taking a specific form of ψ, for example, orthogonal projection onto the tangent plane π, we have for every $\gamma > 0$ and $\boldsymbol{x}_j$ satisfying $\|\boldsymbol{x}_j - \boldsymbol{x}_i\| < \delta_2$, we have

$$U^m_{1-\gamma}(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)) \subset \psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}) \subset U^m_{1-\gamma}(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)). \tag{3.2}$$

In fact, considering the properties of orthogonal projection, we have for every $\boldsymbol{x} \in U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}$,

$$\|\psi(\boldsymbol{x}) - \psi(\boldsymbol{x}_i)\| \le \|\boldsymbol{x} - \boldsymbol{x}_i\| \le \|\boldsymbol{x}_i - \boldsymbol{x}_j\|.$$

LEMMA 3.1. (*i*) *For any $\boldsymbol{x} \in \mathcal{M}$ and any $\alpha > 0$, there exists $\delta > 0$ such that for every $\boldsymbol{x}_i \in \mathcal{M}$ satisfying $\|\boldsymbol{x}_i - \boldsymbol{x}\| < \delta$, we have that the angle between the vector $\boldsymbol{x}_i - \boldsymbol{x}$ and its projection onto π, which is the tangent plane of $\mathcal{M}$ at point $\boldsymbol{x}$, is less than α.*

(*ii*) *Furthermore, there exists $\delta^* > 0$ such that for every $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{M}$ satisfying* $\max\{\|\boldsymbol{x}_i - \boldsymbol{x}\|, \|\boldsymbol{x}_j - \boldsymbol{x}\|\} < \delta^*$, *we have that the angle between the vector $\boldsymbol{x}_i - \boldsymbol{x}_j$ and its projection onto the tangent plane π is less than α.*

Applying Lemma 3.1, we can select a sufficiently small δ_2 such that

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \le (1+\gamma)\|\psi(\boldsymbol{x}_i) - \psi(\boldsymbol{x}_j)\|.$$

This reveals that for every

$$\boldsymbol{x} \in B(\boldsymbol{x}_i, \|\boldsymbol{x}_i - \boldsymbol{x}_j\|) \cap \mathcal{M},$$

it holds true that

$$\psi(\boldsymbol{x}) \in B\big(\psi(\boldsymbol{x}_i), (1+\gamma)\|\psi(\boldsymbol{x}_i) - \psi(\boldsymbol{x}_j)\|\big).$$

The above process can be repeated for the pair $(\boldsymbol{x}, \boldsymbol{x}_j)$ similarly. Putting them together, we get the second "$\subset$" of equation (3.2):

$$\psi\big(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}\big) \subset U^m_{1+\gamma}\big(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\big).$$

For the first "$\subset$", the proof is similar. In detail, for every

$$z \in B\big(\psi(\boldsymbol{x}_i), (1-\gamma)\|\psi(\boldsymbol{x}_i) - \psi(\boldsymbol{x}_j)\|\big) \cap V,$$

we can select a sufficiently small δ_2 such that

$$\|\psi^{-1}(z) - \boldsymbol{x}_i\| \le \left(1 + \frac{\gamma}{2}\right)\|z - \psi(\boldsymbol{x}_i)\| \le \left(1 - \frac{\gamma}{2}\right)\|\psi(\boldsymbol{x}_i) - \psi(\boldsymbol{x}_j)\| \le \|\boldsymbol{x}_i - \boldsymbol{x}_j\|,$$

which implies that

$$\psi^{-1}\big(U^m_{1-\gamma}\big(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\big)\big) \subset U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}.$$

Applying the mapping ψ to both sides and repeating the process for $\boldsymbol{x}$ and $\boldsymbol{x}_j$, we obtain the first "$\subset$" of equation (3.2):

$$U^m_{1-\gamma}\big(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)\big) \subset \psi\big(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}\big).$$

With equation (3.2) and noticing that

$$\lim_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \to 0} \frac{\|\psi(\boldsymbol{x}_i) - \psi(\boldsymbol{x}_j)\|}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|} = 1, \tag{3.3}$$

one can directly obtain that

$$\lim_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \to 0} \frac{\lambda(\psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}))}{\lambda(U^m(\boldsymbol{x}_i, \boldsymbol{x}_j))} = \lim_{\|\boldsymbol{x}_j - \boldsymbol{x}_i\| \to 0} \frac{\lambda(\psi(U^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}))}{\lambda(U^m(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)))} = 1. \tag{3.4}$$

Combining (3.1) and (3.4) completes the proof of Lemma 2.2. □

3.4. *Proofs of Lemma* 2.3, 2.4, 2.5, *and* 2.7.

PROOFS OF LEMMA 2.3, 2.4, AND 2.5. All these lemmas share a similar proof to Lemma 2.2 that we have proved above. Details are hence omitted. □

PROOF OF LEMMA 2.7. We consider the pushforward $\psi_*^{-1}\lambda$ and Lebesgue measure λ instead of μ and the pushforward $\psi_*\mu$ in this lemma. Hence, we just set $g \equiv 1$ and its proof can be regarded as a special case of the proof sketch of Lemma 2.4. □

3.5. *Proof of Lemma* 2.10. We consider approximating $\mu(U^d(\boldsymbol{x}, \boldsymbol{x}_j))$ as

$$\mu(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j)) = \psi_*\mu(\psi(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j) \cap \mathcal{M})) = \int_{\psi(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j) \cap \mathcal{M})} g \,\mathrm{d}\lambda(\boldsymbol{x}_j).$$

Using the Lebesgue differentiation theorem (LDT), there exists some $\delta > 0$ such that for μ-a.e. $\boldsymbol{x}$ and every $\boldsymbol{x}_j$ satisfying $\|\boldsymbol{x}_j - \boldsymbol{x}\| < \delta$, we have

$$\frac{\mu(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j))}{g(\psi(\boldsymbol{x}))\lambda(\psi(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j) \cap \mathcal{M}))} \in \left[1 - \frac{\epsilon}{3}, 1 + \frac{\epsilon}{3}\right]. \tag{3.5}$$

Since $H(\mathbf{0})$ is assumed to be positive definite, we can choose constants $0 < c < C < \infty$ such that

$$c < \lambda_{\min}(H(\mathbf{0})) < \lambda_{\max}(H(\mathbf{0})) < C$$

for the eigenvalues of $H(\mathbf{0})$. Thus, the Mahalanobis distance

$$\|\boldsymbol{x}\|_\Sigma = \|\boldsymbol{x}\|_{H(\mathbf{0})^{-1}} = \sqrt{\boldsymbol{x}^T(H(\mathbf{0}))\boldsymbol{x}}$$

is equivalent to the Euclidean distance $\|\boldsymbol{x}\|$. With the same argument, Lemma 3.1 holds for the Mahalanobis distance $\|\cdot\|_\Sigma$ as well. As in the proof of Lemma 2.2, Lemma 3.1 directly yields that, for any fixed $\gamma > 0$, we can choose a sufficiently small $\delta > 0$ such that

$$U^m_{1-\gamma,K}(\psi(\boldsymbol{x}), \psi(\boldsymbol{x}_j)) \subset \psi(U_K^d(\boldsymbol{x}, \boldsymbol{x}_j) \cap \mathcal{M}) \subset U^m_{1+\gamma,K}(\psi(\boldsymbol{x}), \psi(\boldsymbol{x}_j)),$$

where

$$U^m_{\alpha,K}(\boldsymbol{x}, \boldsymbol{y}) := B^d(\boldsymbol{x}, \alpha\|\boldsymbol{x} - \boldsymbol{y}\|_\Sigma) \cup B^d(\boldsymbol{y}, \alpha\|\boldsymbol{x} - \boldsymbol{y}\|_\Sigma) \cap \pi_{\boldsymbol{x}}$$

and $\pi_{\boldsymbol{x}}$ denotes the tangent plane of $\mathcal{M}$ at point $\boldsymbol{x}$. Letting $\gamma \to 0$, we have

$$\lim_{\|x - x_j\|_\Sigma \to 0} \frac{\lambda(\psi(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}))}{\lambda(U^m_{1,K}(\psi(\boldsymbol{x}_i), \psi(\boldsymbol{x}_j)))} = 1. \tag{3.6}$$

The same argument as in the derivation of (3.3) gives us

$$\lim_{\|\boldsymbol{x} - \boldsymbol{x}_j\|_\Sigma \to 0} \frac{f(\psi(\boldsymbol{x}) - \psi(\boldsymbol{x}_j))}{f(\boldsymbol{x} - \boldsymbol{x}_j)} = 1. \tag{3.7}$$

Also, the smoothness of f gives us that $\nabla f(\mathbf{0}) = 0$ and $\nabla^2 f(\mathbf{0}) = \Sigma^{-1}$, and the Taylor expansion of f around $\mathbf{0}$ gives us

$$\lim_{\|\boldsymbol{x} - \boldsymbol{x}_j\|_\Sigma \to 0} \frac{f(\boldsymbol{x} - \boldsymbol{x}_j)}{\|\boldsymbol{x} - \boldsymbol{x}_j\|_\Sigma} = 1. \tag{3.8}$$

Combining (3.6), (3.7), and (3.8), we obtain

$$\lim_{\|x - x_j\|_\Sigma \to 0} \frac{\lambda(\psi(U_K^d(\boldsymbol{x}_i, \boldsymbol{x}_j) \cap \mathcal{M}))}{\|\boldsymbol{x} - \boldsymbol{x}_j\|_\Sigma^m U_{m,K,\boldsymbol{x}}} = 1,$$

which completes the proof of Lemma 2.10.

3.6. *Proofs of Lemmas* 2.11, 2.12 *and* 2.13.

PROOFS OF LEMMAS 2.11, 2.12 AND 2.13. The proofs are nearly identical to that of Lemma 2.10. We thus omit the details. □

3.7. *Proof of Lemma* 3.1.

PROOF OF LEMMA 3.1. We denote by π the tangent plane of $\mathcal{M}$ at point $\boldsymbol{x}$ and $\pi^{\perp}$ the orthogonal complement of π in $\mathbb{R}^d$. Moreover, we denote by P_V, Q_V the orthogonal projection operators on π and $\pi^{\perp}$, respectively. The orthogonal decomposition of a vector z with respect to π can then be shown as $z = P_V(z) + Q_V(z)$.

Define $\theta(z)$ as the angle between z and its projection onto the tangent plane π. When $z = \mathbf{0}$, we define $\theta(z) = 0$. Using the notation above, one has

$$\tan\bigl(\theta(\boldsymbol{x}_i - \boldsymbol{x})\bigr) = \frac{\|Q_V(\boldsymbol{x}_i - \boldsymbol{x})\|}{\|P_V(\boldsymbol{x}_i - \boldsymbol{x})\|}, \quad \boldsymbol{x}_i \neq \boldsymbol{x}.$$

We now consider $\theta(\boldsymbol{x}_i - \boldsymbol{x})$ as a function of $\boldsymbol{x}_i$. Since $\mathcal{M}$ is a smooth manifold, $Q_V(\boldsymbol{x}_i - \boldsymbol{x})$ and $P_V(\boldsymbol{x}_i - \boldsymbol{x})$ are both continuous functions. With the definition of the tangent plane π, we know that $\theta(\boldsymbol{x}_i - \boldsymbol{x})$ is continuous at point $\boldsymbol{x}_i = \boldsymbol{x}$. Thus $\theta(\boldsymbol{x}_i - \boldsymbol{x})$ is a continuous function on $\mathcal{M}$. We can select a fixed $\delta_1 > 0$ so that $\theta(\boldsymbol{x}_i - \boldsymbol{x})$ is uniformly continuous on $\mathcal{M} \cap B^d(\boldsymbol{x}, \delta_1)$. Since $\theta(\boldsymbol{x} - \boldsymbol{x}) = 0$, there exists $0 < \delta < \delta_1$ such that for any $\boldsymbol{x}_i \in \mathcal{M}$ and $\|\boldsymbol{x}_i - \boldsymbol{x}\| < \delta$, we have $\theta(\boldsymbol{x}_i - \boldsymbol{x}) < \alpha$, which is the first claim.

For the second claim, we just have to modify $\theta(z)$ as follows: we define $\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to be the angle between the vector $\boldsymbol{x}_i - \boldsymbol{x}_j$ and its projection onto the tangent plane π so that

$$\tan\bigl(\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)\bigr) = \frac{\|Q_V(\boldsymbol{x}_i - \boldsymbol{x}_j)\|}{\|P_V(\boldsymbol{x}_i - \boldsymbol{x}_j)\|}, \quad \boldsymbol{x}_i \neq \boldsymbol{x}_j.$$

Additionally, define $\theta(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ when $\boldsymbol{x}_i = \boldsymbol{x}_j$. We select δ_1^* to make sure that $\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$ exists. The second claim can then be obtained similarly by following the proof above. □

3.8. *Proof of Lemma* 2.6.

PROOF OF LEMMA 2.6. It is a simple corollary to the Lebesgue differential theorem, which can be derived by applying the LDT to f and g respectively:

$$\lim_{U \to \boldsymbol{x}, U \in \mathcal{V}} \frac{1}{|U|} \int_U fg \,\mathrm{d}\lambda = f(\boldsymbol{x})g(\boldsymbol{x}) \lim_{U \to \boldsymbol{x}, U \in \mathcal{V}} \frac{1}{|U|} \int_U \mathrm{d}\lambda = f(\boldsymbol{x}) \lim_{U \to \boldsymbol{x}, U \in \mathcal{V}} \frac{1}{|U|} \int_U g \,\mathrm{d}\lambda.$$

This completes the proof. □

3.9. *Proof of Proposition* 1.2.

PROOF OF PROPOSITION 1.2. We first show that for any fixed point $\boldsymbol{x} \in \mathcal{M}$, the number of points whose NN is $\boldsymbol{x}$ is finite. Consider all the geodesic cones

$$\{C_\alpha\}_{\alpha \in I} \subset \mathcal{M}$$

with $\boldsymbol{x}$ as their common peak, which satisfies that, for any $\alpha \in I$, $\boldsymbol{x}', \boldsymbol{x}'' \in C_\alpha$, we have

$$d_g(\boldsymbol{x}' - \boldsymbol{x}'') < \max\bigl\{d_g(\boldsymbol{x}' - \boldsymbol{x}), d_g(\boldsymbol{x}'' - \boldsymbol{x})\bigr\}.$$

We can directly check that $C_\alpha \setminus \{\boldsymbol{x}\}$ is an open set for any $\alpha \in I$, and $\bigcup_{\alpha \in I} C_\alpha = \mathcal{M}$. In fact, for any $\boldsymbol{x}' \neq \boldsymbol{x}$, $\boldsymbol{x}' \in \mathcal{M}$, we can find a small enough $\epsilon > 0$ such that

$$\epsilon < d_g(\boldsymbol{x}, \boldsymbol{x}')/3.$$

Thus, $B(\boldsymbol{x}', \epsilon) \subset C_\alpha$ for some $\alpha \in I$. Here $B(\boldsymbol{x}', \epsilon)$ denotes the geodesic ball with center $\boldsymbol{x}'$ and radius ϵ. Since $\mathcal{M}$ is compact and $\{C_\alpha \cup B(\boldsymbol{x}, \delta)\}_{\alpha \in I}$ is an open cover of $\mathcal{M}$ for any $\delta > 0$, we can find a finite subcover of $\mathcal{M}$, denoting as $\{C_\alpha \cup B(\boldsymbol{x}, \delta)\}_{\alpha \in I_1}$, where I_1 is a finite index set, that is,

$$\mathcal{M} \subset \bigcup_{\alpha \in I_1} \{C_\alpha \cup B(\boldsymbol{x}, \delta)\}. \tag{3.9}$$

Then, we choose δ small enough so that we can find a local smooth exponential map ψ satisfying

$$\psi^{-1} : B(\boldsymbol{x}, \delta) \to \mathbb{R}^m, \quad \psi(\mathbf{0}) = \boldsymbol{x},$$

and

$$\begin{aligned} \|\psi^{-1}(\boldsymbol{x}')\| &= d_g(\boldsymbol{x}, \boldsymbol{x}'), \\ (1-\eta)\|\psi^{-1}(\boldsymbol{x}') - \psi^{-1}(\boldsymbol{x}'')\| &\le d_g(\boldsymbol{x}', \boldsymbol{x}'') \\ &\le (1+\eta)\|\psi^{-1}(\boldsymbol{x}') - \psi^{-1}(\boldsymbol{x}'')\| \quad \text{for any } \boldsymbol{x}', \boldsymbol{x}'' \neq \boldsymbol{x}, \end{aligned}$$

which holds true for the natural exponential map in Riemannian geometry with a sufficiently small $\delta > 0$. By the construction in Bickel and Breiman (1983), we can find a finite set of cones in $\mathbb{R}^m$ with $\mathbf{0}$ as their common peak satisfying

$$(1+\epsilon)\|\boldsymbol{y}' - \boldsymbol{y}''\| < (1-\epsilon)\max\{\|\boldsymbol{y}'\|, \|\boldsymbol{y}''\|\} \quad \text{for any } \boldsymbol{y}', \boldsymbol{y}'' \neq \mathbf{0}$$

in the same cone. Consider the map of these cones under exponential map ψ. Each image of the cone is a subset of a geodesic cone in $\mathcal{M}$ with $\boldsymbol{x}$ as its common peak. Therefore, we can find a finite set of geodesic cones indexed by I_2 such that

$$B(\boldsymbol{x}, \delta) \subset \bigcup_{\alpha \in I_2} C_\alpha. \tag{3.10}$$

Combining (3.9) and (3.10), we have $\mathcal{M} \subset \bigcup_{\alpha \in I_1 \cup I_2} C_\alpha$. For any two points whose NN are both $\boldsymbol{x}$, it is straightforward to get that they cannot be in the same geodesic cone, which implies that the number of points whose NN is $\boldsymbol{x}$ is finite.

Then we consider the set

$$T_K := \{\boldsymbol{x} \in \mathcal{M} : \text{the maximum number of points whose NN is } \boldsymbol{x} \text{ equals } K\}.$$

For any $\boldsymbol{x} \in T_K$, we can find a realization of K points whose NN is $\boldsymbol{x}$, denoting this set as $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_K\}$. Denote $\delta_i = \min_{j \neq i} d_g(\boldsymbol{x}_i, \boldsymbol{x}_j)$. By definition of NN, we have

$$\delta_i - d_g(\boldsymbol{x}_i, \boldsymbol{x}) > 0 \quad \text{for any } i = 1, \dots, K.$$

Denote

$$\delta^* = \min_{i=1,\dots,K} \delta_i - d_g(\boldsymbol{x}_i, \boldsymbol{x}) > 0.$$

It can be checked by triangle inequality that for any $\boldsymbol{x}' \in B(\boldsymbol{x}, \delta^*)$, we have $\boldsymbol{x}' \in T_K$. Thus, T_K is an open set in $\mathcal{M}$. For any $\boldsymbol{x} \in \mathcal{M}$, since the number of points whose NN is $\boldsymbol{x}$ is finite, there must be $\boldsymbol{x} \in T_K$ for some $K \in \mathbb{N}^*$. Therefore, $\mathcal{M} = \bigcup_{K=1}^{\infty} T_K$. Again, since $\mathcal{M}$ is compact, there exists a finite index set I_3 such that $\mathcal{M} = \bigcup_{K \in I_3} T_K$. Now we set

$$C_{\mathcal{M}} = \max_{K \in I_3} K < \infty,$$

which implies that $\mathcal{M} = \bigcup_{K=1}^{C_{\mathcal{M}}} T_K$. Thus, for any $\boldsymbol{x} \in \mathcal{M}$, the number of points whose NN is $\boldsymbol{x}$ is at most $C_{\mathcal{M}}$. $\square$

REFERENCES

AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005

AUDDY, A., DEB, N. and NANDY, S. (2024). Exact detection thresholds for Chatterjee's correlation. *Bernoulli* **30** 1640–1668.

AZADKIA, M. and CHATTERJEE, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. MR4352523 https://doi.org/10.1214/21-aos2073

AZADKIA, M., TAEB, A. and BÜHLMANN, P. (2021). A fast non-parametric approach for causal structure learning in polytrees. Available at arXiv:2111.14969.

BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

BICKEL, P. J. (2022). Measures of independence and functional dependence. Available at arXiv:2206.13663.

BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214. MR0682809

CAO, S. and BICKEL, P. J. (2020). Correlations with tailored extremal properties. Available at arXiv:2008.10177v2.

CHATTERJEE, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 2009–2022. MR4353729 https://doi.org/10.1080/01621459.2020.1758115

CHATTERJEE, S. and VIDYASAGAR, M. (2022). Estimating large causal polytree skeletons from small samples. Available at arXiv:2209.07028.

CLARKSON, K. L. (2006). Nearest-neighbor searching and metric space dimensions. In *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*. 15–59.

DEB, N., GHOSAL, P. and SEN, B. (2020). Measuring association on topological spaces using kernels and geometric graphs. Available at arXiv:2010.01768v2.

DETTE, H., SIBURG, K. F. and STOIMENOV, P. A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030 https://doi.org/10.1111/j.1467-9469.2011.00767.x

DEVROYE, L. (1988). The expected size of some graphs in computational geometry. *Comput. Math. Appl.* **15** 53–64. MR0937563 https://doi.org/10.1016/0898-1221(88)90071-5

DUGUNDJI, J. (1951). An extension of Tietze's theorem. *Pacific J. Math.* **1** 353–367. MR0044116

EVANS, L. C. and GARIEPY, R. F. (2015). *Measure Theory and Fine Properties of Functions*, Revised ed. *Textbooks in Mathematics*. CRC Press, Boca Raton, FL. MR3409135

FUCHS, S. (2024). Quantifying directed dependence via dimension reduction. *J. Multivar. Anal.* **201** 105266.

GAMBOA, F., GREMAUD, P., KLEIN, T. and LAGNOUX, A. (2022). Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli* **28** 2345–2374. MR4474546 https://doi.org/10.3150/21-bej1421

GRIESSENBERGER, F., JUNKER, R. R. and TRUTSCHNIG, W. (2022). On a multivariate copula-based dependence measure and its estimation. *Electron. J. Stat.* **16** 2206–2251. MR4401220 https://doi.org/10.1214/22-ejs2005

HAN, F. (2021). On extensions of rank correlation coefficients to multivariate spaces. *Bernoulli News* **28** 7–11.

HENZE, N. (1987). On the fraction of random points with specified nearest-neighbour interrelations and degree of attraction. *Adv. in Appl. Probab.* **19** 873–895. MR0914597 https://doi.org/10.2307/1427106

HUANG, Z., DEB, N. and SEN, B. (2022). Kernel partial correlation coefficient—a measure of conditional dependence. *J. Mach. Learn. Res.* **23** 1–58.

KPOTUFE, S. (2011). k-nn regression adapts to local intrinsic dimension. In *Advances in Neural Information Processing Systems* **24**.

KPOTUFE, S. (2017). Lipschitz density-ratios, structured data, and data-driven tuning. In *International Conference on Artificial Intelligence and Statistics*.

KPOTUFE, S. and GARG, V. (2013). Adaptivity to local smoothness and dimension in kernel regression. In *Advances in Neural Information Processing Systems* **26**.

LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. MR2954043

LEVINA, E. and BICKEL, P. (2004). Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems* **17**.

LIN, Z. and HAN, F. (2022). Limit theorems of Chatterjee's rank correlation. Available at arXiv:2204.08031.

LIN, Z. and HAN, F. (2023). On boosting the power of Chatterjee's rank correlation. *Biometrika* **110** 283–299. MR4589063 https://doi.org/10.1093/biomet/asac048

RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hung.* **10** 441–451. MR0115203 https://doi.org/10.1007/BF02024507

SCHÖLKOPF, B. (2000). The kernel trick for distances. In *Advances in Neural Information Processing Systems* **13**.

SHI, H., DRTON, M. and HAN, F. (2022). On the power of Chatterjee's rank correlation. *Biometrika* **109** 317–333. MR4430960 https://doi.org/10.1093/biomet/asab028

SHI, H., DRTON, M. and HAN, F. (2024). On Azadkia–Chatterjee's conditional dependence coefficient. *Bernoulli* **30** 851–877. MR4699537 https://doi.org/10.3150/22-bej1529

STROTHMANN, C., DETTE, H. and SIBURG, K. F. (2024). Rearranged dependence measures. *Bernoulli* **30** 1055–1078. MR4699545 https://doi.org/10.3150/23-bej1624

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

ZHANG, Q. (2023). On the asymptotic null distribution of the symmetrized Chatterjee's correlation coefficient. *Statist. Probab. Lett.* **194** Paper No. 109759, 7. MR4525660 https://doi.org/10.1016/j.spl.2022.109759