CONFIDENCE INTERVALS FOR A CLASS OF RESIDUAL VARIANCE ESTIMATORS (DRAFT)

ELIA LIITIÄINEN, FRANCESCO CORONA[1], AND BELLE SELENE XIA

Abstract. We analyze the asymptotic distribution of a general class of nearest neighbor based residual variance estimators providing a central limit theorem with asymptotic variance characterized as a linear combination of residual moments. Based on the theoretical results, a numerical algorithm for estimating confidence intervals is provided in conjunction with a proof of asymptotic consistency.

1. Introduction

The problem of residual variance estimation has been widely investigated as an elegant approach for addressing the limitations of linear dependency measures such as Pearson's correlation coefficient [7, 13, 19]. Given the i.i.d. observations $\{X_i, Y_i\}_{i=1}^N$, it concerns the estimation of

$$\mathrm{E}[(Y_1 - E[Y_1|X_1])^2] \tag{1}$$

corresponding to the optimal mean squared error obtainable via non-linear regression. While not obvious at first sight, estimating the residual variance is often a significantly easier task than performing an actual non-linear regression. As a matter of fact, there exists a diversity of non-parametric techniques providing estimates without explicitly reconstructing a full mapping from X to Y.

Among a plurality of approaches, we will examine methods based on using nearest neighbors to build upon the assumption of local regularity. Such techniques often admit excellent consistency properties under realistic assumptions while remaining simple and understandable [4, 7, 13]. However, in order to position them alongside classical statistical measures, it is necessary to establish sufficiently strong

[1]Department of Chemical and Metallurgical Engineering, Aalto University, 02150 Espoo, Finland

[2]A draft containing the algorithm and theoretical results was first submitted on August 2021.

asymptotic normality results together with practical methods for estimating confidence intervals. In this regard, existing literature covers specific estimators (see e.g. [4]), but does not provide a generic theoretical framework in the context of multidimensional randomized design. Relatedly, confidence intervals have not been widely discussed in a broader setting, while some attempts towards this direction exist [10].

Motivated by the aforementioned considerations, a principal aim of this paper is to establish an asymptotic variance and normality result for a class residual variance estimators with the proposed framework of sufficient generality to capture a number of well-established nearest neighbor based methods including [19, 8, 9]. Moreover, as a technique for constructing confidence intervals, we provide a simple algorithm for evaluating estimation variance with consistency guaranteed in the infinite sample limit under realistic assumptions.

Our approach necessitates the use of auxiliary residual statistics opening a new direction for future research due to their convergence properties having been relatively scarcely investigated even if some existing literature relates to our work closely in this aspect [8]. Relatedly, the finite sample accuracy of the approximative confidence intervals in different settings leaves room for further contributions.

2. Estimators

2.1. **Basic Definitions and Assumptions.** We denote by $B(x,r)$ the ball with center x and radius r in $\Re^n$, where the fixed integer $n > 0$ presents the dimensionality of the input space in our analysis. For any subset $A \subset \Re^n$, scalar $a > 0$ and vector $x \in \Re^n$, we adopt the standard set arithmetic convention

$$aA + x = \{ay + x : \ y \in A\}.$$

The notation $|A|$ is used to refer to the cardinality of A. For integers $k > 0$, we denote by A^k the set $A \times \ldots \times A$ (Cartesian product of k sets).

We will adopt the notation $\tilde{\mathcal{X}}_n$ to indicate the space of non-empty finite subsets of $\Re^n$. To impose some regularity for technical reasons, we remark that the elements of such a space correspond to finite measures on $\Re^n$ with the total variation norm immediately providing a standard topology.

Given a random event $\mathcal{A}$, we denote by $I(\mathcal{A})$ the random variable with the value 1 if the event occurs and 0 otherwise. Generally, the event is expressed as a logical expression involving random variables.

We will consider independent identically distributed (i.i.d.) random variables $\{X_i, Y_i\}_{i=1}^{\infty}$, where each random input vector X_i takes values in $\Re^n$ and output Y_i in $\Re$; specifically, we derive asymptotics w.r.t. the subsamples $\{X_i, Y_i\}_{i=1}^{N}$ for integers $N > 0$. For the reason of conciseness, we adopt the shorthand notation

$$\Xi_N = \{X_i\}_{i=1}^{N}.$$

Given a sequence of random variables $(Z_i)_{i=1}^{\infty}$, the sequence is said to converge in distribution to the normal distribution $N(0,1)$ of zero mean and unit variance conditionally on Ξ_N, if for any $t > 0$,

$$P\left(Z_N \leq t | \Xi_N\right) \to \Phi(t) \tag{2}$$

in probability in the limit $N \to \infty$, where $\Phi(t)$ is the cumulative distribution function of $N(0,1)$. As another type of convergence, $(Z_i)_{i=1}^{\infty}$ is said to converge in mean to a random variable Z if

$$\mathrm{E}[|Z_i - Z|] \to 0$$

in the limit $i \to \infty$.

We define the index of the nearest neighbor of X_i by

$$N[i,1] = \operatorname{argmin}_{1 \leq j \leq N, j \neq i} \|X_j - X_i\|$$

and that of the k-th nearest neighbor by

$$N[i,k] = \operatorname{argmin}_{j \in \{1,\ldots,N\} \setminus \{i, N[i,1], \ldots, N[i,k-1]\}} \|X_j - X_i\|.$$

In this definition, ties may be broken in an arbitrary way; however, as we will only consider points $\{X_i\}_{i=1}^{N}$ distributed according to a density function w.r.t. the Lebesgue measure, the nearest neighbors will be almost surely unique. For notational convenience, we set

$$N[i,0] = i.$$

The conditional expected output is defined by

$$m(x) = \mathrm{E}[Y_1 | X_1 = x]$$

and the residual variables by $r_i = Y_i - m(X_i)$; moreover, we define the conditional residual standard deviation function by

$$\sigma(x) = \sqrt{\mathrm{E}[r_1^2|X_1 = x]};$$

for integers $l \geq 2$, we set

$$V_l(x) = \mathrm{E}[r_1^l|X_1 = x]$$

and for $l \geq 3$,

$$V_l'(x) = \mathrm{E}[r_1^l|X_1 = x] - \mathrm{E}[r_1^2|X_1 = x]\mathrm{E}[r_1^{l-2}|X_1 = x]. \tag{3}$$

Note that in this notation, $\sigma(x)^2 = V_2(x)$. In order to state our principal assumptions, we need the following variation of Hölder continuity: for a function $f(x)$ from $\Re^n$ to $\Re$, our continuity condition imposes the requirement

$$|f(x) - f(y)| \leq c(1 + \|x\| + \|y\|)^\alpha \|x - y\|^\xi \tag{4}$$

for all $x, y \in \Re^n$ and some constants $c > 0$, $\xi > 0$ and $\alpha \geq 0$. Note that Equation (4) with $\xi = 1$ is implied by the simpler condition

$$\|\nabla f(x)\| \leq c(1 + \|x\|^\alpha)$$

for some constants $\alpha, c > 0$ and all $x \in \Re^n$. Our main assumptions are formulated as follows:

(A1) *The random variables $\{X_i, Y_i\}_{i=1}^\infty$ are i.i.d. with the points $\{X_i\}_{i=1}^\infty$ distributed according to a density $p(x)$ w.r.t. the Lebesgue measure on $\Re^n$. Moreover, for $l = 2, \ldots, 4$, the conditional moment functions $m(x)$ and $V_l(x)$ are well-defined and continuous in the sense of Equation (4).*

(A2)

$$\mathrm{E}[Y_1^8] < \infty$$

and

$$\mathrm{E}[\|X_1\|^\alpha] < \infty$$

for all $\alpha > 0$.

Assumption (A1) adopts a slightly weaker continuity condition than Lipschitz continuity covering most cases of practical interest including polynomial functions. Note that it implies the growth condition

$$|m(x)| + \mathrm{E}[r_1^4|X_1 = x] \leq c(1 + \|x\|^\gamma)$$

for some constants $c, \gamma > 0$ independent of $x \in \Re^n$.

In order to establish an additional theoretical tool by extending the concept of nearest neighbors, for a fixed $K > 0$, we define

$$\mathcal{N}_{i,0} = \{N[i,0], \dots, N[i,K]\} \tag{5}$$

and for $l \geq 1$, by recursion

$$\mathcal{N}_{i,l} = \bigcup_{\{1 \leq j \leq N:\ \mathcal{N}_{i,l-1} \cap \mathcal{N}_{j,l-1} \neq \emptyset\}} \mathcal{N}_{j,l-1}. \tag{6}$$

As a notational convention, we will generally employ K to denote a fixed positive constant determined by the specific estimator in question.

2.2. **The Estimation Framework.** As a general context aimed to capture a large part of the existing literature, we will analyze residual variance estimators of the form

$$S_N = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{K} \sum_{j'=0}^{K} W_{i,j,j'} Y_{N[i,j]} Y_{N[i,j']}, \tag{7}$$

where each weight $W_{i,j,j'}$ can be represented as

$$W_{i,j,j'} = f_{j,j'}(X_{N[i,0]}, \dots, X_{N[i,K]}) \tag{8}$$

for some bounded measurable function $f_{j,j'}$ (independent of i) implying that the value of S_N is determined by the observations in the K nearest neighborhood of each point X_i. Note that while the constant K is fully determined by the choice of the specific estimator, our theoretical results are based on varying the number of samples N setting $N \to \infty$ for an asymptotic analysis.

To establish the consistency of the estimator, we require that

$$\sum_{j=0}^{K} W_{i,j,j} = 1 \qquad \text{and} \tag{9}$$

$$\sum_{j'=0}^{K} W_{i,j',j} = \sum_{j'=0}^{K} W_{i,j,j'} = 0 \tag{10}$$

almost surely for $i = 1, \dots, N$ and $j = 0, \dots, K$. Moreover, the weights are assumed to be translation invariant with respect to the nearest neighbors in the sense of being invariant w.r.t. the transformations

$$sX_{N[i,0]} + t, \dots, sX_{N[i,K]} + t$$

for any $s, t \in \Re$, where the sum of a vector and a scalar is interpreted component-wise.

2.3. **Examples.** In this section, we present some examples of estimators of the form (7) with some practical guidelines on the choice of the estimator.

2.3.1. *Locally Constant Estimators.* The well-known K-NN estimator of residual variance can be formulated as

$$S_N = \frac{1}{K(K+1)N} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} Y_{N[i,j]} - Y_i \right)^2, \tag{11}$$

where K is an arbitrary fixed constant. In this case,

$$W_{i,0,0} = \frac{K}{K+1},$$

while

$$W_{i,j,j'} = \frac{1}{K(K+1)} \qquad \text{and}$$
$$W_{i,0,j} = W_{i,j,0} = -\frac{1}{K+1}$$

for $1 \leq j, j' \leq K$. As a matter of fact, the choice $K = 1$ is often made in practical applications resulting in the simple estimator

$$S_N = \frac{1}{2N} \sum_{i=1}^{N} \left(Y_{N[i,1]} - Y_i \right)^2. \tag{12}$$

However, despite its intuitive form, the estimator (11) does not possess optimal convergence properties as it does not achieve a uniform rate of convergence for non-smooth variance functions $\sigma(x)^2$. The asymmetric product estimator introduced in [8] elegantly addresses this issue taking the simple form

$$S_N = \frac{1}{N} \sum_{i=1}^{N} (Y_{N[i,1]} - Y_i)(Y_{N[i,2]} - Y_i). \tag{13}$$

This corresponds to the choice $K = 2$ with $W_{i,0,0} = 1$, $W_{i,1,2} = 1$, $W_{i,0,2} = -1$, $W_{i,1,0} = -1$ for $i = 1, \ldots, N$ setting the remaining weights equal to 0. Some of the theoretical properties of this estimator have been analyzed in [14, 12], where it is shown that the somewhat surprising form of the estimator ensures consistency for non-smooth standard deviation functions $\sigma(x)^2$, while also providing a surprisingly fast rate of convergence.

Towards another direction, it is also of interest to consider linear combinations of the form

$$S_N = \frac{1}{2N} \sum_{j=1}^{K} w_j \sum_{i=1}^{N} \left(Y_{N[i,j]} - Y_i \right)^2 \tag{14}$$

for some fixed constants $w_1, \ldots, w_K$, where the constraint

$$\sum_{j=1}^{K} w_j = 1$$

would ensure that conditions (9) and (10) hold. However, the optimal choice and benefit of such weights remains a relatively unexplored topic closely related to the Gamma test [7].

2.3.2. *Local Linear Estimators.* The local linear estimators introduced in [19] invoke the idea of applying linear regression for points in the proximity of each observed regressor vector. While the added complexity might be expected to increase variance and risk instability, such concerns are somewhat mitigated by the boundedness of the weights in the representation (7).

In order to demonstrate the variable bandwidth estimator according to [19], we define the augmented vectors

$$\tilde{X}_{i,j} = \begin{pmatrix} 1 \\ X_{N[i,j]} - X_i \end{pmatrix}$$

for $1 \leq i \leq N$ and $j = 1, \ldots, n+1$. Moreover, we define the $n+1 \times n+1$ matrix

$$\hat{X}_i = \left[\tilde{X}_{i,1}, \ldots, \tilde{X}_{i,n+1} \right].$$

Let e be the $n+1$-dimensional vector with $e^{(1)} = 1$ and $e^{(j)} = 0$ for $1 < j \leq n+1$. Set

$$z_i = \begin{pmatrix} -1 \\ \hat{X}_i^{-1} e \end{pmatrix},$$

where it can be shown that Assumption (A1) suffices to ensure the almost sure invertibility of $\hat{X}_i$ (in contrast, [19] adopts a more general formulation covering the non-invertible case). Finally, for $0 \leq j \leq n+1$, we define $\hat{w}_{i,j}$ as the component $j+1$ of the vector $z_i / \|z_i\|$. The estimator takes the form

$$S_N = \frac{1}{N} \sum_{i=0}^{N} \left(\sum_{j=0}^{n+1} \hat{w}_{i,j} Y_{N[i,j]} \right)^2, \tag{15}$$

where we have

$$\sum_{j=0}^{n+1} \hat{w}_{i,j}^2 = 1 \quad \text{and} \tag{16}$$

$$\sum_{j=0}^{n+1} \hat{w}_{i,j} \tilde{X}_{i,j} = 0 \tag{17}$$

for any $1 \leq i \leq N$. The local linear estimator can be viewed in the framework (7) by setting $W_{i,j,j'}$ equal to

$$\hat{w}_{i,j} \hat{w}_{i,j'}$$

and observing that by Equations (16) and (17), the weights are bounded and the constraints (9) and (10) hold.

2.4. **Note on the Choice of Estimator.** At present, profound asymptotic analysis of residual variance estimators in a general context remains a rather unexplored topic. Consequently, it is hard to give reliable recommendations about the choice of algorithm in practical applications. It seems likely that further theoretical work on the bias/variance trade-off (with as exact characterization of the asymptotic behavior as possible) would provide useful insight in this respect with major influencing factors expected to include the dimensionality of the input space, boundary effect and smoothness of the mappings $m(x)$ and $\sigma(x)$. In this regard, different applications might pose varying requirements on the method used; as an example, concerns about bias are somewhat mitigated if the goal is merely to compute a p-value for the hypothesis $m(x) = 0$. Nevertheless, based on the present knowledge, some general guidelines may be outlined:

- As shown in [12], the bias of the estimator (13) is expected to be of order $N^{-3/n} log^{\alpha} N$ for some $\alpha \geq 0$ (we believe $\alpha = 0$ is provable) with variance being of order $N^{-1/2}$. This suggests that the method often performs well for low to moderate dimensionalities (up to $n = 5$ or even $n = 6$).
- For high dimensional problems ($n \geq 7$), we recommend considering the local linear estimator discussed in Section 2.3.2 to compensate for the curse of dimensionality by making use of a higher order approximator.

As a particular method meriting a remark, [4] presents an exceptionally concise estimator with satisfactory theoretical properties, which, however, falls outside the framework formulated in this paper. The simple form of the method seems beneficial in terms of variance while, at present, the bias is not yet formally fully understood. As a conjecture, we believe that it is an attractive choice in particular when the dimensionality of the input space is in the range $n \leq 3$.

2.5. **Asymptotic Distribution and Variance.** The following theorem states the asymptotic unbiasedness of residual variance estimators of the form (7) under our assumptions. The result is expected by the corresponding results in earlier literature (see e.g. [5]).

Theorem 1. *Assume that Assumptions (A1)-(A2) hold. Then*

$$\mathrm{E}[S_N] \to \mathrm{E}[\sigma(X_1)^2]$$

in the limit $N \to \infty$.

We compute the asymptotic variance of S_N by employing the decomposition

$$\mathrm{Var}[S_N] = \mathrm{E}[\mathrm{Var}[S_N|\Xi_N]] + \mathrm{Var}[\mathrm{E}\,[S_N|\Xi_N]].$$

In this context, the asymptotic distribution of $\sqrt{N}(S_N - \mathrm{E}[S_N])$ may be characterized as follows:

Theorem 2. *Assume that Assumptions (A1)-(A2) hold. Then the variable*

$$\sqrt{N}(S_N - \mathrm{E}[S_N])$$

is asymptotically normal with

$$N\mathrm{Var}[\mathrm{E}\,[S_N|\Xi_N]] \to \mathrm{Var}[\sigma(X_1)^2] \tag{18}$$

and

$$N\mathrm{Var}[S_N|\Xi_N] \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)] \tag{19}$$

in mean in the limit $N \to \infty$ for some constants A_1 and A_2 determined by n and the estimator S_N.

Remarkably, the constants A_1 and A_2 in Theorem 2 are universal in the sense that they are independent of the probability distribution of the observations (X_i, Y_i). As a strategy for constructing approximative confidence intervals, the asymptotic normality signifies that it remains to formulate a statistic guaranteed to approach the sum of the limits (18) and (19) when the number of samples increases to infinity. To this end, for $1 \leq i, j \leq N$ and $0 \leq l_1, l_2, l_3, l_4 \leq K$, we define

$$\delta_{i,j,l_1,l_2,l_3,l_4} = W_{i,l_1,l_2} W_{j,l_3,l_4} I(N[i,l_1] = N[j,l_3]) I(N[i,l_2] = N[j,l_4]) \tag{20}$$

and the random variables

$$b^{(1)}_{l_1,l_2}(X_i, \Xi_N) = \sum_{j=1}^{N} \sum_{\substack{l_3,l_4=0 \\ l_3 \neq l_4}}^{K} \delta_{i,j,l_1,l_2,l_3,l_4} + \delta_{i,j,l_1,l_2,l_4,l_3} \qquad \text{and} \tag{21}$$

$$b^{(2)}_{l}(X_i, \Xi_N) = \sum_{j=1}^{N} \sum_{l'=0}^{K} \delta_{i,j,l,l,l',l'}. \tag{22}$$

Employing these notations, we adopt a two-step approach by formulating the intermediate statistic

$$Q_N = \frac{1}{N} \sum_{i=1}^{N} \left(\sigma(X_i)^4 \sum_{\substack{l_1,l_2=0 \\ l_1 \neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i, \Xi_N) + V_4'(X_i) \sum_{l=0}^{K} b^{(2)}_{l}(X_i, \Xi_N) \right) \tag{23}$$

capable of approximating the limit (19):

Theorem 3. *Assume that Assumptions (A1)-(A2) hold. Then the random variables (23) relate to the constants A_1 and A_2 in Theorem 2 by the convergence*

$$Q_N \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)]$$

in mean in the limit $N \to \infty$.

As a matter of fact, the definition (23) is motivated by non-asymptotic considerations in the proof of Theorem 2 aiming to ensure that the value is close to the real value also in the non-asymptotic range. However, the functions $\sigma(x)$ and $V_4'(x)$ are generally unknown posing an auxiliary approximation problem closely related to estimating the right side of (18).

3. Confidence Intervals

In this section, we complete Theorems 2 and 3 by providing estimators for

$$\mathrm{Var}[\sigma(X_1)^2] = \mathrm{E}[\sigma(X_1)^4] - \mathrm{E}[\sigma(X_1)^2]^2 \tag{24}$$

and Q_N in Equation (23) sufficient for evaluating approximative confidence intervals. To this end, we propose an empirical method by adapting the statistical techniques presented in [8] concerning higher order residual moments.

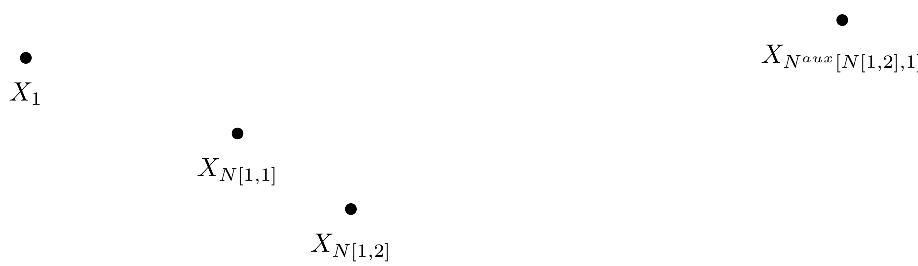
To outline the estimator for the first term in Equation (23),

$$\frac{1}{N}\sum_{i=1}^{N}\sigma(X_i)^4 \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N), \tag{25}$$

we define

$$N^{aux}[j,i] = \operatorname{argmin}_{1\leq l\leq N, l\notin\{j,i,N[i,1]\}} \|X_l - X_j\|;$$

this definition is demonstrated by the following example involving a set of four points:



Defining

$$\hat{V}^2_{2,i} = \frac{1}{4}(Y_{N[i,1]} - Y_i)^2 (Y_{N^{aux}[N[i,2],i]} - Y_{N[i,2]})^2, \tag{26}$$

the proposed estimator of (25) is then formulated as an average of the terms

$$\hat{V}^2_{2,i} \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N).$$

To demonstrate the intuition behind the definition, we resort to setting $m=0$ and observe that for a sufficiently large sample size, the relevant quantity to examine is the expectation

$$\mathrm{E}\left[\hat{V}^2_{2,i} \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N)\right] = \mathrm{E}\left[\mathrm{E}\left[\hat{V}^2_{2,1}\middle|\Xi_N\right] \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N)\right].$$

In this context, using the conditional independence of the residuals, we have

$$\begin{aligned}\mathrm{E}\left[\hat{V}_{2,1}^2 \middle| \Xi_N\right] &= \frac{1}{4}\mathrm{Var}\left[\left(r_{N[1,1]} - r_1\right)^2 \left(r_{N^{aux}[N[1,2],1]} - r_{N[1,2]}\right)^2 \middle| \Xi_N\right] \\ &= \frac{1}{4}\left(\sigma(X_1)^2 + \sigma(X_{N[1,1]})^2\right)\left(\sigma(X_{N[1,2]})^2 + \sigma(X_{N^{aux}[N[1,2],1]})^2\right),\end{aligned}$$

which is expected to be asymptotically close to the desired value $\sigma(X_1)^4$.

Recalling that the second term in (23) takes the form

$$\mathrm{E}\left[V_4'(X_1)\sum_{l=0}^{K} b_l^{(2)}(X_1, \Xi_N)\right] = \mathrm{E}\left[(V_4(X_1) - \sigma(X_1)^4)\sum_{l=0}^{K} b_l^{(2)}(X_1, \Xi_N)\right],$$

we suggest making use of Equation (26) and the product estimator for the fourth moment of the residual variance presented in [8] to define

$$\hat{V}_{4,i}' = \prod_{j=1}^{4}(Y_{N[i,j]} - Y_i) - \hat{V}_{2,i}^2.$$

The complete approximation of Q_N is then given by

$$\hat{Q}_N = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{V}_{2,i}^2 \sum_{\substack{l_1,l_2=0 \\ l_1 \neq l_2}}^{K} b_{l_1,l_2}^{(1)}(X_i, \Xi_N) + \hat{V}_{4,i}' \sum_{l=0}^{K} b_l^{(2)}(X_i, \Xi_N)\right). \tag{27}$$

At the same time, (24) can be estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\hat{V}_{2,i}^2 - S_N^2 \tag{28}$$

for any consistent residual variance estimator S_N.

The following result demonstrates the asymptotic validity of the estimates (27) and (28).

Theorem 4. *Assume that Assumptions (A1)-(A2) hold. Then*

$$Q_N - \hat{Q}_N \to 0$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\hat{V}_{2,i}^2 - S_N^2 \to \mathrm{E}[\sigma(X_1)^4] - \mathrm{E}[\sigma(X_1)^2]^2 \tag{29}$$

in mean in the limit $N \to \infty$.

Computing (27) and (28) based on empirical data is straightfoward and at once provides approximative confidence intervals. It should be noted that while the proof of Theorem 4 demonstrates the asymptotic correctness of the approximations, it does not address finite sample behaviour and rate of convergence. As a related topic of future work, it is also pertinent to consider the direction of the finite sample bias, that is, whether the true value is over- or underestimated.

4. Empirical Demonstration

In order to assess the practical validity of the algorithm introduced in Section 3, we computed relative estimation errors for the method (13) by computing empirical averages using simulated data. More specifically, the input vector X was taken as uniform on $[0,1]^n$, whereas the output was generated according to

$$Y = \sqrt{12}X^{(1)}U + X^{(1)}, \tag{30}$$

where U denotes an independent uniform random variable on $[-1/2, 1/2]$ and $X^{(1)}$ the first component of the vector valued random variable X. In the empirical simulations, the size N of the samples drawn from (30) is varied from 1000 to 11000; for each value of N, an error measure for the discrepancy between the estimated variance of S_N and the actual variance is computed.

Fixing a value for N and considering the residual variance estimates $\hat{S}_{j,N}$ ($j = 1, \ldots, L$) obtained by drawing L samples from the model (30), where L is a large integer in the constraints of the computational resources available, the true empirical variance of the estimator was estimated by

$$\mathrm{Var}_{exp,L}\left[S_N\right] = \frac{1}{L}\sum_{j=1}^{L} \hat{S}_{j,N}^2 - \left(\frac{1}{L}\sum_{j=1}^{L}\hat{S}_{j,N}\right)^2. \tag{31}$$

Specifically, L was fixed as 10000.

The generated datasets were also used to compute the variance estimates $\mathrm{Var}_j\left[S_N\right]$ ($j = 1, \ldots, L$) using the method described in Section 3. As an implementation detail, we decided to substitute $\hat{S}_{j,N}$ for S_N in Equation (28) even if in principle, another estimator could have been used. Assuming that $\mathrm{Var}_{exp,L}\left[S_N\right]$ is close to

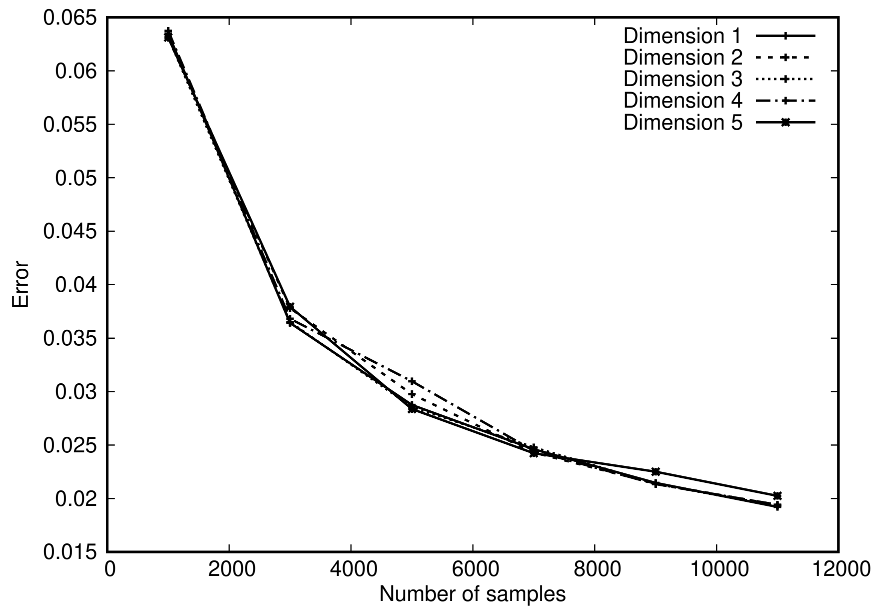


FIGURE 1. Expected relative errors of the standard deviation estimates as a function of the number of samples (N).

the true estimator variance, the discrepancy measure

$$\frac{1}{L}\sum_{j=1}^{L}\frac{\left|\sqrt{\mathrm{Var}_{exp,L}\left[S_N\right]}-\sqrt{\mathrm{Var}_j\left[S_N\right]}\right|}{\sqrt{\mathrm{Var}_{exp,L}\left[S_N\right]}} \tag{32}$$

was evaluated to assess the validity of the predicted estimation variances. Note the square roots in (32) employed in order to focus on assessing standard deviations instead of variances.

The empirical results displayed in Figure 1 are well-aligned with the theoretical results of the paper exhibiting a rate of convergence independent of the dimensionality.

5. Auxiliary Results

5.1. **Nearest Neighbors.** It is well-known that in general finite dimensional settings, nearest neighbor distances approach zero when the number of points tends to infinity [11]. In our context, we need to the following restatement of this fact:

Lemma 1. *Assume that Assumptions (A1)-(A2) hold. Then for any $\epsilon > 0$ and $0 \leq \alpha < n$,*

$$N^{\alpha/n-\epsilon}\mathrm{E}[d_{1,K}^{\alpha}] \to 0$$

in the limit $N \to \infty$.

Proof. For $\gamma > 0$ and $N > 0$, define the random variable

$$M = \sum_{i=1}^{N} I(X_i \in [-N^{\gamma}, N^{\gamma}]^n).$$

The results in [11] imply that when $0 < \alpha < n$, we have

$$\sum_{i=1}^{N} d_{i,K}^{\alpha} \leq cN^{1+\gamma\alpha-\alpha/n}$$

on the event $M = N$ for a constant $c > 0$ independent N,

By Assumption (A2), the probability of the event $M \neq N$ approaches zero faster than $N^{-\beta}$ for any $\beta > 0$ finalizing the proof by an application of Hölder's inequality and the inequality

$$d_{i,K} \leq \|X_i\| + \sum_{j=1}^{K+1} \|X_j\|.$$

$\square$

The following result builds upon [5] by extending the corresponding result for nearest neighbor graphs to cover the extended neighborhoods defined in Equations (5)-(6).

Lemma 2. *Assume that (A1) holds. Then for any fixed integer $l \geq 0$, there exists constants $N_0 > 0$ and $c > 0$ such that all $N > N_0$, the inequalities*

$$|\{1 \leq i \leq N : \ 1 \in \mathcal{N}_{i,l}\}| \leq c \qquad \textit{and} \tag{33}$$

$$|\mathcal{N}_{1,l}| \leq c \tag{34}$$

hold almost surely.

Proof. Proceeding by induction, let us assume the induction hypothesis that the claim holds for some constant $c > 0$ and integers $l \geq 0$ and $N_0 > 0$:

$$\sum_{j=1}^{N} I(i \in \mathcal{N}_{j,l}) \leq c \tag{35}$$

and

$$|\mathcal{N}_{i,l}| \leq c \tag{36}$$

almost surely for all $N > N_0$ and $1 \leq i \leq N$. As a matter of fact, Lemma 3.2 in [7] establishes the upper bound (35) for $l = 0$. Then, we have

$$|\mathcal{N}_{i,l+1}| \leq \sum_{j \in \mathcal{N}_{i,l}} \sum_{i'=1}^{N} I(j \in \mathcal{N}_{i',l}) |\mathcal{N}_{i',l}| \leq c^3 \tag{37}$$

and

$$\begin{aligned} \sum_{i=1}^{N} I(j \in \mathcal{N}_{i,l+1}) &\leq \sum_{i=1}^{N} \sum_{j'=1}^{N} \sum_{j''=1}^{N} I(j' \in \mathcal{N}_{i,l}) I(j' \in \mathcal{N}_{j'',l}) I(j \in \mathcal{N}_{j'',l}) \\ &\leq c \sum_{j''=1}^{N} \sum_{j'=1}^{N} I(j' \in \mathcal{N}_{j'',l}) I(j \in \mathcal{N}_{j'',l}) \\ &\leq c^2 \sum_{j''=1}^{N} I(j \in \mathcal{N}_{j'',l}) \leq c^3 \end{aligned} \tag{38}$$

for any $1 \leq j \leq N$. □

Lemma 2 can be used to prove the following result, which will turn out to be useful for bounding the moments of random variables expressible as a sum over an extended neighborhood.

Lemma 3. *Assume that Assumption (A1) holds. For any $N > 0$, let $(a_i)_{i=1}^N$ be a sequence of positive numbers and fix arbitrary integers $l \geq 1$ and $s \geq 1$. Then for some $N_0 > 0$ and all $N > N_0$,*

$$\sum_{i=1}^{N} \left(\sum_{j \in \mathcal{N}_{i,l}} a_j \right)^s \leq c \sum_{i=1}^{N} a_i^s$$

for some constant $c > 0$ independent of N and $(a_i)_{i=1}^N$.

Proof. We observe that

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,l}} a_j = \sum_{i=1}^{N} a_i \left(\sum_{j=1}^{N} I(i \in \mathcal{N}_{j,l}) \right).$$

By Lemma 2 we know that

$$\sum_{j=1}^{N} I(i \in \mathcal{N}_{j,l}) \leq c_1$$

and consequently

$$\sum_{i=1}^{N} \sum_{j \in \mathcal{N}_{i,l}} a_j \leq c_1 \sum_{i=1}^{N} a_i$$

for some constant $c_1 > 0$. Moreover, by Jensen's inequality [18] and Lemma 2,

$$\left(\sum_{j \in \mathcal{N}_{i,l}} a_j \right)^s \leq |\mathcal{N}_{i,l}|^{s-1} \sum_{j \in \mathcal{N}_{i,l}} a_j^s \leq c_2 \sum_{j \in \mathcal{N}_{i,l}} a_j^s$$

for some constant $c_2 > 0$ independent of N and $1 \leq i \leq N$. Consequently, the case $s > 1$ follows from the proof for $s = 1$. □

The following lemma is a corollary of Lemma 3 and Assumption (A2).

Lemma 4. *Assume that Assumptions (A1)-(A2) hold. Then for any $\alpha > 0$ and integer $l \geq 0$,*

$$\limsup_{N \to \infty} \mathrm{E} \left[\sum_{j \in \mathcal{N}_{1,l}} \|X_j\|^\alpha \right] < \infty.$$

Lemmas 1 and 4 in conjunction with Hölder's inequality yield the following result.

Lemma 5. *Assume that Assumptions (A1)-(A2) hold. Then for any $\xi \geq 0$, $l \geq 0$ and $\alpha > 0$,*

$$\mathrm{E} \left[\left(1 + \sum_{j \in \mathcal{N}_{1,l}} \|X_{N[1,j]}\| \right)^\xi d_{1,K}^\alpha \right] \to 0 \tag{39}$$

in the limit $N \to \infty$.

In order to prove our main results, we will invoke the theory in [17] based on stabilizing random functionals. To this end, we must establish a radius of stabilization result in the case of the spatial Poisson process (also known as the Poisson point process) of unit intensity on $\Re^n$ (denoted by $\mathcal{P}$). Consider the undirected K-nearest neighbor graph (a widely studied structure in computational geometry, see e.g. [16]) with vertices given by

$$\mathcal{A} \cup \{0\}$$

for a set of points $\mathcal{A} \subset \Re^n$, where for any $r > 0$, $\mathcal{A} \cap B(0,r)$ is a finite set with unique pointwise distances (to ensure that the graph is well-defined); for each point, there is an edge between the point and its K nearest neighbors. Closely related to Equations (5)-(6) with a slightly different perspective, we define

$$\mathcal{Z}_K^l[\mathcal{A}] = \{x \in \mathcal{A} \cup \{0\} : \text{the shortest path between } x \text{ and } 0 \text{ has at most } l \text{ edges}\}.$$

As an example, the set $\mathcal{Z}_K^1[\mathcal{A}]$ contains the zero vector 0, the K nearest neighbors of 0 and conversely, the points for which 0 is among the K nearest neighbors. Our stability result is a variation of Lemma 6.1 in [15] with an adaptation for our slightly different conceptual framework.

Lemma 6. *For any integer $l \geq 1$, there exists a random variable R acting as a radius of stabilization, i.e., associating a positive number with each realization of $\mathcal{P}$ in such a way that almost surely*

$$\mathcal{Z}_K^l[(\mathcal{P} \cap B(0,R)) \cup \mathcal{Y}] = \mathcal{Z}_K^l[\mathcal{P}] \tag{40}$$

for all finite point sets $\mathcal{Y} \subset B(0,R)^C$.

Proof. We define $\mathcal{Z}_K^0[\mathcal{P}] = \{0\}$ and state the inductive hypothesis that the claim holds for an integer $l \geq 0$ with a radius of stabilization R. For formal correctness, it should be remarked that in the following, we will take the notational liberty to denote by $\mathcal{P}$ and R fixed realizations of the actual random variables.

Fix a small number $0 < \epsilon < 1$ and choose the cones $S_1, \dots, S_L \subset \Re^n$ such that for all $1 \leq j \leq L$ and $x, y \in S_j$, we have

$$x^T y \geq (1-\epsilon)\|x\|\|y\|$$

and $\cup_{j=1}^L S_j = \Re^n$. For $0 < r_1 < r_2$, we define $S_{j,r_1,r_2} = S_j \cap B(0,r_2) \setminus B(0,r_1)$.

Without losing generality, we assume that

$$|S_{j,2R,4R} \cap \mathcal{P}| \geq K \tag{41}$$

for $j = 1, \dots, L$. Consider two vectors $x \in S_{j,4R,\infty}$ and $y \in S_{j,2R,4R}$ for some $1 \le j \le L$. Then

$$\begin{aligned}
&\|x-y\|^2 - \min_{z \in B(0,R)} \|x-z\|^2 \\
&\qquad \le \|x-y\|^2 - (\|x\| - R)^2 \\
&\qquad \le -2x^T y + 2R\|x\| + \|y\|^2 - R^2 \\
&\qquad \le -2(1-\epsilon)\|y\|\|x\| + 2R\|x\| + \|y\|^2 - R^2 \\
&\qquad \le (4\epsilon - 2)R\|x\| + 3R^2 \le (16\epsilon - 5)R^2 \le 0 \qquad (42)
\end{aligned}$$

for all $0 < \epsilon < 5/16$ (setting $\|y\| = 2R$ to maximize the right side of the third inequality). Aiming to demonstrate that $12R$ serves as a radius of stabilization for $\mathcal{Z}_K^{l+1}[\mathcal{P}]$, let us consider

$$\mathcal{A} = (\mathcal{P} \cap B(0,12R)) \cup \mathcal{Y} \cup \{0\}$$

for an arbitrary finite set $\mathcal{Y} \subset B(0,12R)^C$. Equations (41) and (42) imply that no point in $\mathcal{A}$ located outside $B(0,4R)$ contains a point in $B(0,R)$ among its K nearest neighbors. To the other direction, recalling that by construction, we certainly have

$$|B(0,4R) \cap \mathcal{P}| \ge K + 1$$

ensuring that the K nearest neighbors of the points in $B(0,4R)$ are contained in $B(0,12R)$.

To complete the proof, we observe that $\mathcal{Z}_K^{l+1}[\mathcal{A}]$ constitutes of the points in $\mathcal{Z}_K^{l}[\mathcal{A}]$, their K nearest neighbors and the points in $\mathcal{A}$ with a K nearest neighbors in $\mathcal{Z}_K^{l}[\mathcal{A}]$. We have demonstrated that these are determined by the points in $B(0,4R)$ and their K nearest neighbors, which in turn are determined by the points in $B(0,12R)$ regardless of $\mathcal{Y}$. We conclude that

$$\mathcal{Z}_K^{l+1}[\mathcal{A}] = \mathcal{Z}_K^{l+1}[\mathcal{P} \cap B(0,12R)]$$

establishing $12R$ as a radius of stabilization for $\mathcal{Z}_K^{l+1}[\mathcal{P}]$. □

5.2. **Results for Local Random Variables.** In this section, we consider functions of the points in the K nearest neighborhood of a given point. More specifically, we will examine random variables $\{h_{N,i}\}_{i=1}^N$ of the form

$$h_{N,i} = h(X_{N[i,0]}, Y_{N[i,0]}, \dots, X_{N[i,K]}, Y_{N[i,K]}) \qquad (43)$$

for a fixed measurable function h mapping $(\Re^n)^{K+1}$ to $\Re$.

As a fundamental result of statistics, under general conditions, the standard deviation of the mean of a sequence of independent identically distributed random variables approaches zero proportionally to $N^{-1/2}$. However, if the terms in the average exhibit a local dependency structure instead of full independence, a more elaborate analysis is necessary. To prove such results, the Efron-Stein inequality has turned out to be a convenient tool [6].

Lemma 7. *Assume that Assumption (A1) hold with random variables of the form (43). Then we have*

$$\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i}\right] \leq c\left(\mathrm{E}[h_{N,1}^2] + \mathrm{E}[h_{N-1,1}^2]\right)$$

for some constant $c > 0$ independent of $N > K + 1$.

Prior to proceeding to the actual proof, we formulate the following restatement of the Efron-Stein inequality in a form suitable for the proof of Lemma 7. See [1] for a derivation of the result among other concentration inequalities.

Lemma 8. *Let $\{Z_i\}_{i=1}^N$ be a set of i.i.d. random vectors and for $1 \leq i \leq N$, define the random variables*

$$G = g_N(Z_1, \dots, Z_N) \qquad \text{and}$$
$$G_i = g_{N-1}(Z_1 \dots, Z_{i-1}, Z_{i+1}, \dots, Z_N),$$

where g_N and g_{N-1} are measurable functions taking values in $\Re$. Assuming that the random variables G and G_i are square integrable, we have

$$\mathrm{Var}[G] \leq \sum_{i=1}^{N} \mathrm{E}[(G - G_i)^2].$$

Proof of Lemma 7. Observe that the random variables $h_{N-1,i}$ $(1 \leq i \leq N-1)$ are computed using $\{X_i, Y_i\}_{i=1}^{N-1}$ corresponding to the removal of the N-th observation. Setting $h_{N-1,N} = 0$, we observe that for any $1 \leq i \leq N$,

$$h_{N,i} - h_{N-1,i} = 0 \tag{44}$$

unless

$$N \in \{N[i,0], \dots, N[i,K]\} \tag{45}$$

implying the more general condition

$$i \in \mathcal{N}_{N,1}. \tag{46}$$

Lemma 8, Hölder's inequality and Equations (44)-(46) imply the following inequality:

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i}\right] &\le \mathrm{E}\left[\left(\sum_{i=1}^{N} h_{N,i} - h_{N-1,i}\right)^2\right] \le \mathrm{E}\left[\left(\sum_{i\in\mathcal{N}_{N,1}} |h_{N,i} - h_{N-1,i}|\right)^2\right] \\
&\le \mathrm{E}\left[|\mathcal{N}_{N,1}| \sum_{i\in\mathcal{N}_{N,1}} (h_{N,i} - h_{N-1,i})^2\right] \\
&\le 3\mathrm{E}\left[|\mathcal{N}_{N,1}| \sum_{i\in\mathcal{N}_{N,1}} h_{N,i}^2 + h_{N-1,i}^2\right],
\end{aligned}$$

where in the last inequality the fact that $(a+b)^2 \le 3a^2 + 3b^2$ for any $a, b \in \Re$ was used. Moreover, by invoking Lemma 2, Lemma 3 and a symmetry argument, we have

$$\begin{aligned}
\mathrm{E}\left[|\mathcal{N}_{N,1}| \sum_{i\in\mathcal{N}_{N,1}} h_{N,i}^2 + h_{N-1,i}^2\right] &\le c_1 \mathrm{E}\left[\sum_{i\in\mathcal{N}_{N,1}} h_{N,i}^2 + h_{N-1,i}^2\right] \\
&= \frac{c_1}{N}\mathrm{E}\left[\sum_{j=1}^{N}\sum_{i\in\mathcal{N}_{j,1}} h_{N,i}^2 + h_{N-1,i}^2\right] \\
&\le \frac{c_2}{N}\mathrm{E}\left[\sum_{i=1}^{N} h_{N,i}^2 + h_{N-1,i}^2\right] \\
&\le c_2 \mathrm{E}[h_{N,1}^2 + h_{N-1,1}^2]
\end{aligned}$$

for some constants $c_1, c_2 > 0$ independent of N. □

Recall that in addition to investigating variance, we also aim to prove asymptotic normality in the class of estimators covered by our framework. We will attain this goal by establishing a sufficiently strong central limit theorem for functions of the form (43). The following lemma demonstrates a local dependency structure alike to those in [2] and [3], where various local dependency structures were shown to imply a central limit theorem together with non-asymptotic bounds on the approximation error by a normal distribution.

Lemma 9. *Assume that the random variables $h_{N,i}$ are of the form (43) and Assumption (A1) holds. Conditioning on Ξ_N and fixing arbitrary $l \geq 0$, $N > K$ and $1 \leq i \leq N$, $\{h_{N,j}\}_{j \in \mathcal{N}_{i,l}}$ is independent of $\{h_{N,j}\}_{j \in \{1,\ldots,N\} \setminus \mathcal{N}_{i,l+2}}$.*

Proof. Let $\{i_j\}_{j=1}^{l_1}$ and $\{\tilde{i}_j\}_{j=1}^{l_2}$ be two sets of indices in $\{1, \ldots, N\}$. We consider conditional independence on the events

$$\bigcup_{j \in \mathcal{N}_{i,l}} \mathcal{N}_{j,0} = \{i_j\}_{j=1}^{l_1}$$

and

$$\bigcup_{j \in \{1,\ldots,N\} \setminus \mathcal{N}_{i,l+2}} \mathcal{N}_{j,0} = \{\tilde{i}_j\}_{j=1}^{l_2}$$

observing that by Equation (6), the two sets of indices must be disjoint. However, on this event, each random variable in $\{h_{N,j}\}_{j \in \mathcal{N}_{i,l}}$ may be represented as

$$f(X_{i_1}, Y_{i_1}, \ldots, X_{i_{l_1}}, Y_{i_{l_1}}) \tag{47}$$

for a measurable function f and analogously, the random variables in

$$\{h_{N,j}\}_{j \in \{1,\ldots,N\} \setminus \mathcal{N}_{i,l+2}}$$

take the form

$$g(X_{\tilde{i}_1}, Y_{\tilde{i}_1}, \ldots, X_{\tilde{i}_{l_2}}, Y_{\tilde{i}_{l_2}}) \tag{48}$$

for a measurable function g. The claim of the lemma follows due to the fact that

$$\{X_{i_j}, Y_{i_j}\}_{j=1}^{l_1}$$

is independent from

$$\{X_{\tilde{i}_j}, Y_{\tilde{i}_j}\}_{j=1}^{l_2}.$$

□

The analogy between Lemma 9 and conditions (LD1)-(LD2) in Section 4.7 of [2] (alternatively, the results in [3] could be used) provides us the appropriate tool for a concise proof of conditional asymptotic normality. It should be stressed that similar results are well-known in the existing literature on local geometry and nearest neighbors [5].

Lemma 10. *Assume that the random variables $h_{N,i}$ are of the form (43) and Assumption (A1) holds with*

$$\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i}\middle|\Xi_N\right] \to \sigma^2$$

in probability for some constant $\sigma > 0$ when $N \to \infty$, while

$$\limsup_{N\to\infty} \mathrm{E}[|h_{N,1}|^3] < \infty.$$

Then (in the terminology introduced in Section 2.1) we have the convergence in distribution

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i} - \mathrm{E}\left[h_{N,i}|\Xi_N\right] \to N(0,\sigma^2) \tag{49}$$

conditionally on Ξ_N in the limit $N \to \infty$.

Proof. Set

$$M_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i} - \mathrm{E}\left[h_{N,i}|\Xi_N\right]$$

and

$$\sigma_N^2 = \mathrm{Var}[M_N|\Xi_N].$$

By Lemma 3, we have

$$\limsup_{N\to\infty} \frac{1}{N}\mathrm{E}\left[\sum_{i=1}^{N}\left(\sum_{j\in\mathcal{N}_{i,4}} |h_{N,j}|\right)^3\right] < \infty$$

from which it follows that

$$\frac{1}{\sigma_N^3 N^{3/2}}\sum_{i=1}^{N} \mathrm{E}\left[\left(\sum_{j\in\mathcal{N}_{i,4}} |h_{N,j}|\right)^3\middle|\Xi_N\right] \to 0 \tag{50}$$

in probability in the limit $N \to \infty$. In this context, recalling the dependency structure in Lemma 9 and associating $\mathcal{N}_{i,2}$ with the sets A_i and $\mathcal{N}_{i,4}$ with B_i in Theorem 4.13 of [2] implies the convergence in distribution

$$P\left(\frac{M_N}{\sigma_N} \le t\middle|\Xi_N\right) \to \Phi(t)$$

in probability for any fixed $t > 0$ in the limit $N \to \infty$, where Φ refers to the cumulative distribution function of the normal distribution of mean zero and unit variance.

However, as (by the statement of the lemma) $\sigma_N \to \sigma$ in probability, we also have

$$P\left(\frac{M_N}{\sigma} \leq t \middle| \Xi_N\right) \to \Phi(t).$$

□

5.3. **Asymptotic Variance of a Weighted Average of Residuals.** In the following, we consider random variables of the form

$$h_{N,i} = \sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} r_{N[i,j]} r_{N[i,j']} \tag{51}$$

for the weights introduced in (7). It will turn out that (51) is the most important random variable to examine in the proof of Theorems 2 and 3 after excluding asymptotically negligible terms.

As a preliminary remark, observe that recalling the notation $\tilde{\mathcal{X}}_n$ in Section 2.1, $b^{(1)}_{l_1,l_2}$ and $b^{(2)}_{l}$ in Equations (21) and (22) may be considered as bounded (by Lemma 2) scalar functions $\Re^n \times \tilde{\mathcal{X}}_n \mapsto \Re$. As a matter of fact, by depending only on the positions of the points and not their indexes, the definitions (21) and (22) uniquely define functions for all arguments $(x, \mathcal{Y}) \in \Re \times \tilde{\mathcal{X}}_n$ belonging to the range of (X_i, Ξ_N) for some $N > K$ with Ξ_N viewed as a member of $\tilde{\mathcal{X}}_n$. This construction implicitly assumes that $x \in \mathcal{Y}$, whereas in the opposite case we simply consider $(x, \{x\} \cup \mathcal{Y})$. Moreover, in the absence of ties, the definition trivially extends to any $(x, \mathcal{Y}) \in \Re \times \tilde{\mathcal{X}}_n$, while in the non-pertinent case of ties we may always define the value as 0.

Lemma 11. *Suppose that Assumptions (A1)-(A2) hold. Then, considering the random variables in Equation (23),*

$$Q_N \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)]$$

in mean in the limit $N \to \infty$ with the constants A_1 and A_2 determined by the specific form of the functions $f_{j,j'}$ $(0 \leq j, j' \leq K)$ in Equation (8).

Proof. By the invariance of the weights (9) and (10) together with the fact that nearest neighbor indices are invariant with respect to scaling and translation, we have

$$b^{(1)}_{l_1,l_2}(x, \mathcal{Y}) = b^{(1)}_{l_1,l_2}(0, a(\mathcal{Y} - x)) \tag{52}$$

for $a > 0$, $x \in \Re^n$ and $\mathcal{Y} \in \tilde{\mathcal{X}}_n$ with a similar equation for $b_l^{(2)}$. Furthermore, the right side of (52) is invariant to setting $a = 1$ and letting $\mathcal{Y}$ vary in such a way that $\mathcal{Z}_K^3[\mathcal{Y} - x]$ stays intact. As a matter of fact, the value is determined by the points in $\mathcal{Z}_K^2[\mathcal{Y} - x]$ and their K nearest neighbors, both of which stay unchanged. Consequently, Lemma 6 with $l = 3$ provides the radius of stabilization R as defined in [17]: for almost all realizations of the spatial Poisson process $\mathcal{P}$ and the associated radius R,

$$\lim_{r\to\infty} b^{(1)}_{l_1,l_2}(0, B(0,r) \cap \mathcal{P}) = b^{(1)}_{l_1,l_2}(0, (B(0,R) \cap \mathcal{P}) \cup \mathcal{Y}) \tag{53}$$

and

$$\lim_{r\to\infty} b^{(2)}_{l}(0, B(0,r) \cap \mathcal{P}) = b^{(2)}_{l}(0, (B(0,R) \cap \mathcal{P}) \cup \mathcal{Y}) \tag{54}$$

for all finite sets of points $\mathcal{Y} \subset B(0,R)^C$.

Employing the radius of stabilization formulated in Equations (53) and (54) in conjunction with the invariance (52), Assumption (A2) (which ensures the finiteness of the moments involved) and Theorem 2.1 in [17], we have

$$\begin{aligned} Q_N = \frac{1}{N}\sum_{i=1}^{N}\left(\sigma(X_i)^4 \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i, \Xi_N) + V_4'(X_i)\sum_{l=0}^{K} b^{(2)}_{l}(X_i,\Xi_N)\right) \\ \to A_1 \mathrm{E}[\sigma(X_1)^4] + A_2\mathrm{E}[V_4'(X_1)] \end{aligned}$$

in mean in the limit $N \to \infty$, where the constants A_1 and A_2 are defined by

$$A_1 = E\left[\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(0,\mathcal{P})\right] \quad \text{and}$$

$$A_2 = E\left[\sum_{l=0}^{K} b^{(2)}_{l}(0,\mathcal{P})\right] \tag{55}$$

for the spatial Poisson process $\mathcal{P}$ of unit intensity on $\Re^n$ adopting the notation

$$b^{(1)}_{l_1,l_2}(0,\mathcal{P}) = \lim_{r\to\infty} b^{(1)}_{l_1,l_2}(0, \mathcal{P}\cap B(0,r))$$

for $b^{(1)}_{l_1,l_2}(0,\mathcal{P})$ with a similar definition for $b^{(2)}_{l}(0,\mathcal{P})$. □

The asymptotic variance for random variables of the form (51) may now be characterized as follows.

Lemma 12. *Suppose that Assumptions (A1)-(A2) hold and the terms $h_{N,i}$ are of the form (51). Then, setting $N \to \infty$, we have*

$$\operatorname{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} h_{N,i}\middle|\Xi_N\right] \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)]$$

in mean for the constants A_1 and A_2 in Lemma 11.

Proof. Observe that by the conditional independence of the residuals,

$$\operatorname{Cov}\left[r_{N[i,l_1]} r_{N[i,l_2]}, r_{N[j,l_3]} r_{N[j,l_4]}\middle|\Xi_N\right] = 0$$

unless $N[i,l_1] = N[j,l_3]$ and $N[i,l_2] = N[j,l_4]$, or alternatively, $N[i,l_1] = N[j,l_4]$ and $N[i,l_2] = N[j,l_3]$. Recalling the definition (20), while distinguishing the cases $l_1 = l_2$ and $l_1 \neq l_2$, a short computation reveals that

$$\begin{aligned}
&\frac{1}{N}\operatorname{Var}\left[\sum_{i=1}^{N} h_{N,i}\middle|\Xi_N\right] \\
&\quad = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg(\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K}\sum_{\substack{l_3,l_4=0\\ l_3\neq l_4}}^{K}(\delta_{i,j,l_1,l_2,l_3,l_4} + \delta_{i,j,l_1,l_2,l_4,l_3})\sigma(X_{N[i,l_1]})^2\sigma(X_{N[i,l_2]})^2 \\
&\qquad + \sum_{l_1=0}^{K}\sum_{l_2=0}^{K}\delta_{i,j,l_1,l_1,l_2,l_2}V_4'(X_{N[i,l_1]})\Bigg) \\
&\quad = \frac{1}{N}\sum_{i=1}^{N}\Bigg(\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N)\sigma(X_{N[i,l_1]})^2\sigma(X_{N[i,l_2]})^2 \\
&\qquad + \sum_{l=0}^{K} b^{(2)}_{l}(X_i,\Xi_N)V_4'(X_{N[i,l]})\Bigg).
\end{aligned}$$

Due to Lemma 11, the claim follows by observing that the continuity assumptions stated in Assumption (A1) together with the boundedness of $b^{(1)}_{l_1,l_2}$ and $b^{(2)}_{l}$ allow us to invoke Lemma 5 to show that

$$\begin{aligned}
&E\Bigg[\Bigg|\frac{1}{N}\sum_{i=1}^{N}\Bigg(\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N)\sigma(X_{N[i,l_1]})^2\sigma(X_{N[i,l_2]})^2 \\
&\qquad + \sum_{l=0}^{K} b^{(2)}_{l}(X_i,\Xi_N)V_4'(X_{N[i,l]})\Bigg) - Q_N\Bigg|\Bigg] \\
&\quad \le \frac{c}{N}\sum_{i=1}^{N}\mathrm{E}\left[\left(1 + \sum_{l=0}^{K}\|X_{N[i,l]}\|^{\beta}\right) d_{i,K}^{\beta'}\right] \to 0
\end{aligned}$$

in the limit $N \to \infty$, where c, β and β' are positive constants independent of N.

□

6. Proofs

6.1. **Proof of Theorems 1, 2 and 3.** Reflecting different aspects of common asymptotic considerations, the proofs of our three main results will be jointly demonstrated. For notational brevity, we introduce the notation

$$\Delta_{i,j} = m(X_i) - m(X_j) \tag{56}$$

and decompose the residual variance estimator (7) as

$$S_N = R_1 + R_2 + R_3, \tag{57}$$

where, applying Equations (9) and (10), we have

$$\begin{aligned}
R_1 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} r_{N[i,j]} r_{N[i,j']}, \\
R_2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} m(X_{N[i,j]}) r_{N[i,j']} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} r_{N[i,j]} m(X_{N[i,j']}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} \Delta_{N[i,j],i} r_{N[i,j']} \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} r_{N[i,j]} \Delta_{N[i,j'],i} \qquad \text{and} \\
R_3 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} m(X_{N[i,j]}) m(X_{N[i,j']}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'} \Delta_{N[i,j],i} \Delta_{N[i,j'],i}.
\end{aligned}$$

In order to verify Theorem 1, we will examine $\mathrm{E}[R_1]$, $\mathrm{E}[R_2]$ and $\mathrm{E}[R_3]$. The more challenging task of providing the asymptotic formulas (18) and (19) is achieved by first establishing the asymptotic negligibility result

$$N\mathrm{Var}[R_2] + N\mathrm{Var}[R_3] \to 0 \tag{58}$$

and consequently,

$$N\mathrm{Var}[\mathrm{E}\left[R_2|\Xi_N\right]] + N\mathrm{Var}[\mathrm{E}\left[R_3|\Xi_N\right]] + N\mathrm{E}[\mathrm{Var}[R_2|\Xi_N]] + N\mathrm{E}[\mathrm{Var}[R_3|\Xi_N]] \to 0 \tag{59}$$

in the limit $N \to \infty$. Observe that by Equation (58)

$$\limsup_{N\to\infty} N\mathrm{Var}[\mathrm{E}\left[R_1|\Xi_N\right]] < \infty$$

implies that

$$\begin{aligned}
\mathrm{E}[|\mathrm{Var}[S_N|\Xi_N] - \mathrm{Var}[R_1|\Xi_N]|] &\le \mathrm{E}[|(S_N - \mathrm{E}\left[S_N|\Xi_N\right])^2 - (R_1 - \mathrm{E}\left[R_1|\Xi_N\right])^2|] \\
&\le \mathrm{E}[\mathrm{Var}[R_2 + R_3|\Xi_N]] \\
&\quad + 2\sqrt{\mathrm{E}[\mathrm{Var}[R_1|\Xi_N]]}\sqrt{\mathrm{E}[\mathrm{Var}[R_2 + R_3|\Xi_N]]} \to 0
\end{aligned}$$

and

$$N\mathrm{Var}[\mathrm{E}\left[S_N|\Xi_N\right]] - N\mathrm{Var}[\mathrm{E}\left[R_1|\Xi_N\right]] \to 0$$

in the limit $N \to \infty$. Consequently, once Equation (58) has been proven to hold, it will remain to consider the terms $\mathrm{Var}[\mathrm{E}\left[R_1|\Xi_N\right]]$ and $\mathrm{Var}[R_1|\Xi_N]$.

Concerning $\mathrm{E}[R_1]$, by the properties of conditional expectations together with the independence of the residuals, we have

$$\begin{aligned}
\mathrm{E}\left[R_1|\Xi_N\right] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'}\mathrm{E}\left[r_{N[i,j]}r_{N[i,j']}\middle|\Xi_N\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K} W_{i,j,j}\sigma(X_{N[i,j]})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\sigma(X_i)^2 + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K} W_{i,j,j}(\sigma(X_{N[i,j]})^2 - \sigma(X_i)^2), \qquad (60)
\end{aligned}$$

where, the weights $W_{i,j,j'}$ being bounded, Lemma 5 and Equation (9) yield

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{K} W_{i,j,j}(\sigma(X_{N[i,j]})^2 - \sigma(X_i)^2) \to 0 \tag{61}$$

in mean in the limit $N \to \infty$ keeping in mind that $\sigma(x)^2$ is a continuous function in the sense of Equation (4).

To advance further towards a proof of Theorem 1, we observe that by the properties of conditional expectations,

$$\mathrm{E}[R_2] = \mathrm{E}[\mathrm{E}\left[R_2|\Xi_N\right]] = 0.$$

Moreover, we impose Jensen's inequality as follows:

$$
\begin{aligned}
&\mathrm{E}\left[\left(\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'}\Delta_{N[i,j],i}r_{N[i,j']}\right)^2 \middle| \Xi_N\right] \\
&\leq K^3\mathrm{E}\left[\sum_{j=0}^{K}\sum_{j'=0}^{K} W_{i,j,j'}^2\Delta_{N[i,j],i}^2 r_{N[i,j']}^2 \middle| \Xi_N\right] \\
&\leq K^3\left(\max_{1\leq j,j'\leq K} W_{i,j,j'}^2\right)\mathrm{E}\left[\sum_{j=0}^{K} r_{N[i,j]}^2 \middle| \Xi_N\right]\left(\sum_{j=0}^{K}\Delta_{N[i,j],i}^2\right) \\
&\leq c_1\left(1+\sum_{j=0}^{K}\|X_{N[i,j]}\|^\beta\right) d_{i,K}^{\beta'}
\end{aligned}
$$

for some positive constants c_1, β and β' independent of N and $1 \leq i \leq N$ with an analogous derivation for the other terms present in R_2 and R_3. Lemmas 5 and 7 then imply that Equation (58) holds and $\mathrm{E}[R_3] \to 0$ in the limit $N \to \infty$ finalizing the proof of Theorem 1.

Having demonstrated the negligibility of R_2 and R_3 both in terms of expectations and variances, we proceed towards determining the asymptotic behaviour of $\mathrm{Var}[\mathrm{E}\,[R_1|\Xi_N]]$ and $\mathrm{Var}[R_1|\Xi_N]$. Observing that since by Lemmas 5 and 7, there exists a positive number $c_2 \geq 0$ with

$$
\begin{aligned}
&\limsup_{N\to\infty}\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j=0}^{K} W_{i,j,j}(\sigma(X_{N[i,j]})^2-\sigma(X_i)^2)\right] \\
&\leq c_2\limsup_{N\to\infty}\mathrm{E}\left[\left(\sum_{j=0}^{K} W_{1,j,j}(\sigma(X_{N[1,j]})^2-\sigma(X_1)^2)\right)^2\right] = 0,
\end{aligned}
\tag{62}
$$

we may make use of Equation (60) to establish the limit

$$N\mathrm{Var}[\mathrm{E}\,[R_1|\Xi_N]] \to \mathrm{Var}[\sigma(X_1)^2],$$

while by Lemma 12,

$$N\mathrm{Var}[R_1|\Xi_N] \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)] \tag{63}$$

in mean for some constants A_1 and A_2 determined by n and the specific form of the residual variance estimator completing the proofs of the limits (18) and (19). To address the proof of Theorem 3, we observe that Lemma 12 also implies the limit

$$Q_N \to A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)]$$

in mean when $N \to \infty$.

In order to establish the asymptotic normality required in Theorem 2, we observe that

$$\begin{aligned}\sqrt{N}(R_1 - \mathrm{E}[R_1]) &= \sqrt{N}(R_1 - \mathrm{E}\left[R_1|\Xi_N\right]) + \sqrt{N}(\mathrm{E}\left[R_1|\Xi_N\right] - \mathrm{E}[R_1]) \\ &= \sqrt{N}(R_1 - \mathrm{E}\left[R_1|\Xi_N\right]) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\sigma(X_i)^2 - \mathrm{E}[\sigma(X_1)^2]) + \Xi, \qquad (64)\end{aligned}$$

where by Equations (60) and (62), the term

$$\Xi = \sqrt{N}(\mathrm{E}\left[R_1|\Xi_N\right] - \mathrm{E}[R_1]) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\sigma(X_i)^2 - \mathrm{E}[\sigma(X_1)^2])$$

is asymptotically negligible due to the fact that

$$\mathrm{E}[\Xi^2] = \mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{j=0}^{K} W_{i,j,j}(\sigma(X_{N[i,j]})^2 - \sigma(X_i)^2)\right].$$

Let us introduce the notations

$$Z_N = \sqrt{N}(R_1 - \mathrm{E}\left[R_1|\Xi_N\right])$$

and

$$Z_N' = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sigma(X_i)^2 - \mathrm{E}[\sigma(X_1)^2].$$

Taking into account that the addends in Z_N certainly possess bounded third moments, Equation (63) and a representation of Lemma 10 in terms of characteristic functions yield

$$\mathrm{E}\left[e^{itZ_N}\middle|\Xi_N\right] \to e^{-\frac{1}{2}(A_1 E[\sigma(X_1)^4] + A_2 E[V_4'(X_1)])t^2} \qquad (65)$$

for any $t \in \Re$ in probability when $N \to \infty$. Moreover, since Z_N' is asymptotically (unconditionally) normally distributed by the standard central limit theorem, the asymptotic normality of (64) follows from

$$\mathrm{E}\left[e^{it(Z_N + Z_N')}\right] = \mathrm{E}\left[\mathrm{E}\left[e^{itZ_N}\middle|\Xi_N\right] e^{itZ_N'}\right]$$

by invoking the limit (65) as the convergence in probability suffices in this case due to the boundedness of characteristics functions.

6.2. **Proof of Theorem 4.** Theorem 3 implicitly implies that

$$Q_N - \mathrm{E}[Q_N] \to 0$$

in mean in the limit $N \to 0$. For our proof, it thus suffices to show that

$$\hat{Q}_N - Q_N \to 0 \tag{66}$$

in mean in the limit $N \to \infty$.

We recall the definition (56) and introduce the further notations

$$\delta_{i,j} = r_i - r_j \tag{67}$$

and

$$\mathcal{S}_i = \hat{V}_{2,i}^2 \sum_{\substack{l_1,l_2=0 \\ l_1 \neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i, \Xi_N). \tag{68}$$

Assuming without losing generality that $K \geq 3$ to ensure that

$$\|X_{N^{aux}[N[i,2],i]} - X_{N[i,2]}\| \leq \|X_{N[i,3]} - X_{N[i,2]}\| \leq 2d_{i,K} \tag{69}$$

and

$$\{i, N[i,1], N[i,2], N^{aux}[N[i,2],i]\} \subset \mathcal{N}_{i,1} \tag{70}$$

for $1 \leq i \leq N$, consider the first term in the right side of Equation (27),

$$\frac{1}{N}\sum_{i=1}^{N} \mathcal{S}_i = \frac{1}{4N}\sum_{i=1}^{N} \delta^2_{N[i,1],i}\delta^2_{N^{aux}[N[i,2],i],N[i,2]} \sum_{\substack{l_1,l_2=0 \\ l_1 \neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i, \Xi_N) + R, \tag{71}$$

where, by Equations (69) and (70) together with the boundedness of the random variables in Equations (21) and (22) (guaranteed by Lemma 2), the residual term is bounded by

$$|R| \leq \frac{c_1}{N}\sum_{i=1}^{N}\left(\sup_{x,y\in B(X_i,2d_{i,K})} |m(x) - m(y)|\right)\left(\sum_{j\in\mathcal{N}_{i,1}} |m(X_j)| + |r_j|\right)^3$$

for a constant $c_1 > 0$ independent of N. In this context, Lemma 5 together with the continuity requirement in Assumption (A1) ensures the asymptotic negligibility

$$E[|R|] \to 0$$

in the limit $N \to \infty$.

Concerning the first term in the right side of (71), introducing the notation

$$\Upsilon_i = (\sigma(X_i)^2 + \sigma(X_{N[i,1]})^2)(\sigma(X_{N[i,2]})^2 + \sigma(X_{N^{aux}[N[i,2],i]})^2),$$

we have

$$\mathrm{E}\left[\delta^2_{N[i,1],i}\delta^2_{N^{aux}[N[i,2],i],N[i,2]}\middle|\Xi_N\right] = \Upsilon_i,$$

whereas by Assumption (A1) and Lemma 5,

$$\frac{1}{4N}\sum_{i=1}^{N}\Upsilon_i\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N) - \frac{1}{N}\sum_{i=1}^{N}\sigma(X_i)^4\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N) \to 0$$

in mean in the limit $\to\infty$. Putting the preceding logical steps together, we have shown that

$$\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i\middle|\Xi_N\right] - \frac{1}{N}\sum_{i=1}^{N}\sigma(X_i)^4\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N) \to 0 \tag{72}$$

in mean in the limit $N\to\infty$.

By the independence of the observations and Equation (70),

$$\mathrm{Cov}\left[\mathcal{S}_i,\mathcal{S}_j|\Xi_N\right] = 0$$

unless $j\in\mathcal{N}_{i,2}$. Based on this observation, the inclusion (70) and Lemma 2, we have

$$\begin{aligned}\mathrm{E}\left[\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{S}_i\middle|\Xi_N\right]\right] &= \mathrm{E}\left[\mathrm{Cov}\left[\mathcal{S}_1,\sum_{i\in\mathcal{N}_{1,2}}\mathcal{S}_i\middle|\Xi_N\right]\right] \le \mathrm{E}\left[\sum_{i\in\mathcal{N}_{1,2}}\mathcal{S}_1^2+\mathcal{S}_i^2\right]\\ &\le c_2\mathrm{E}\left[\sum_{i\in\mathcal{N}_{1,3}}Y_i^8\right] \le c_3\mathrm{E}[Y_1^8] < \infty \end{aligned}\tag{73}$$

for some constants $c_2,c_3>0$ independent of N, where we have used the fact that for any indices $i_1,\ldots,i_8\in\mathcal{N}_{1,3}$, the inequality

$$\prod_{j=1}^{8}Y_{i_j} \le \sum_{i\in\mathcal{N}_{1,3}}Y_i^8$$

holds, while when expanding the terms $\mathcal{S}_1^2+\mathcal{S}_i^2$ in the right side of the first inequality in (73), the output variables appear as products of such form. Taking into account

Equation (72), this establishes the limit

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i - \frac{1}{N}\sum_{i=1}^{N}\sigma(X_i)^4 \sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N) \to 0$$

in mean in the limit $N\to\infty$.

The proof for the expectation of the second term in $\hat{Q}_N$ is analogous by considering that we may decompose in a similar fashion to Equation (71):

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\hat{V}'_{4,i}\sum_{l=0}^{K} b^{(2)}_l(X_i,\Xi_N)\\
&\quad= \frac{1}{N}\sum_{i=1}^{N}\left(\prod_{l=1}^{4}\delta_{N[i,l],i} - \frac{1}{4}\delta^2_{N[i,1],i}\delta^2_{N^{aux}[N[i,2],i],N[i,2]}\right)\sum_{l=0}^{K} b^{(2)}_l(X_i,\Xi_N) + o(1),
\end{aligned}$$

where $o(1)$ denotes an asymptotically negligible residual term and

$$\begin{aligned}
&\mathrm{E}\left[\left(\prod_{l=1}^{4}\delta_{N[i,l],i} - \frac{1}{4}\delta^2_{N[i,1],i}\delta^2_{N^{aux}[N[i,2],i],N[i,2]}\right)\sum_{l=0}^{K} b^{(2)}_l(X_i,\Xi_N)\middle|\Xi_N\right]\\
&\quad= \left(V_4(X_i) - \frac{1}{4}\Upsilon_i\right)\sum_{l=0}^{K} b^{(2)}_l(X_i,\Xi_N).
\end{aligned}$$

The second part of the theorem, establishing the limit (29), is essentially included in the first part of the proof setting

$$\sum_{\substack{l_1,l_2=0\\ l_1\neq l_2}}^{K} b^{(1)}_{l_1,l_2}(X_i,\Xi_N)$$

equal to 1.

References

[1] S. Boucheron, G. Lugosi, and O. Bousquet. Concentration inequalities. In O. Bousquet, U. von Luxburg, and G. Rätsch, editors, *Advanced Lectures on Machine Learning*, volume 3176 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2004.

[2] L.H.Y. Chen, L. Goldstein, and Q.M. Shao. *Normal Approximation by Stein's Method*. Springer Verlag, 2010.

[3] Louis H. Y. Chen and Qi-Man Shao. Normal approximation under local dependence. *The annals of probability*, 32(3):1985–2028, 2004.

[4] L. Devroye, L. Györfi, G. Lugosi, and H. Walk. A nearest neighbor estimate of the residual variance. *Electronic Journal of Statistics*, 12:1752–1778, 2018.

[5] Dafydd Evans. *The Gamma Test: Data derived estimates of noise for unknown smooth models using near neighbour asymptotics*. PhD thesis, Cardiff University, 2002.

[6] Dafydd Evans. A law of large numbers for nearest neighbour statistics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 464(2100):3175–3192, 2008.

[7] Dafydd Evans and Antonia J. Jones. A proof of the gamma test. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 458(2027):2759–2799, 2002.

[8] Dafydd Evans and Antonia J Jones. Non-parametric estimation of residual moments and covariance. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 464(2099):2831–2846, 2008.

[9] Theo Gasser, Lothar Sroka, and Christine Jennen-Steinmetz. Residual variance and residual pattern in nonlinear regression. *Biometrika*, 73(3):625–633, 1986.

[10] Antonia Jones and Samuel Kemp. Heuristic confidence intervals for the gamma test. In *Proceedings of the 2006 International Conference on Artificial Intelligence, ICAI 2006*, volume 2, pages 522–526, 2006.

[11] S. R. Kulkarni and S. E. Posner. Rates of convergence of nearest neighbor estimation under arbitrary sampling. *Rates of convergence of nearest neighbor estimation under arbitrary sampling*, 41(4):1028–1039, 1995.

[12] Elia Liitiäinen. *Advances in the Theory of Nearest Neighbor Distributions*. PhD thesis, Aalto University School of Science and Technology, 2010.

[13] Elia Liitiäinen, Francesco Corona, and Amaury Lendasse. On nonparametric residual variance estimation. *Neural Processing Letters*, 28(3):155–167, Dec 2008.

[14] Elia Liitiäinen, Francesco Corona, and Amaury Lendasse. Residual variance estimation using a nearest neighbor statistic. *Journal of Multivariate Analysis*, 101(4):811 – 823, 2010.

[15] M. Penrose and J. Yukich. Central limit theorems for some graphs in computational geometry. *Annals of Applied Probability*, 11:1005–1041, 2001.

[16] M. Penrose and J. Yukich. Weak laws of large numbers in geometric probability. *Annals of Applied Probability*, 13(1), 2003.

[17] M. D. Penrose. Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli*, 13(4):1124–1150, 2007. ID number: ISI:000251280100012.

[18] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., New York, NY, USA, 1987.

[19] Vladimir Spokoiny. Variance estimation for high-dimensional regression models. *Journal of Multivariate Analysis*, 82(1):111 – 133, 2002.