Title

Normal Approximation of Stabilizing Statistics

Author

Shi, Zhaoyang

Publication Date

2024

Peer reviewed|Thesis/dissertation

Normal Approximation of Stabilizing Statistics

By

ZHAOYANG SHI
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Krishnakumar Balasubramanian, Co-Chair

Wolfgang Polonik, Co-Chair

Can Le

Committee in Charge

2024

ACKNOWLEDGMENTS

There are many people to whom I want to express my sincere gratitude. First, I want to thank my two advisors: Professor Krishna Balasubramanian and Professor Wolfgang Polonik, who have been giving me all the help and support during my PhD life. Their advice is of great value and benefit to me not only on my PhD projects, this thesis but has a profound influence on shaping my research career on statistics from the beginning when I graduated from a math major with minimum knowledge on what problems statistic researchers should focus on. Their mentorship and wisdom are essential to my PhD training on both mastering statistical theories and fostering my professional statistical thinking. And it will forever affect my career. I also extend this gratitude to our collaborator: Professor Chinmoy Bhattacharjee, who offered a lot of help on one of my PhD projects.

Secondly, I also want to thank all my friends: PhD classmates and my undergraduate friends. They have been accompanying me throughout all my PhD life during my ups and downs. We have had a lot of interesting discussions on math and statistics that broaden my knowledge as we are living in an era of interdisciplinary subjects. Particularly, I want to thank those little cute squirrels outside my apartment. They have brought me a lot of happiness and comfort.

Finally, I would like to thank my parents: Chunzhi Zhao and Shenghua Shi. Their love and support throughout my life has inspired me to become the researcher I am today. Without this, I would have experienced much more difficult choices when I was struggling on whether to pursue a postdoc for my research. Their encouragement for me during my PhD life has always been my support during my hard time.

Abstract

Quantitative normal approximation bounds are important to obtain finite-sample, non-asymptotic inferential guarantees for various statistical problems. We derived such quantitative normal approximation results for a class of stabilizing statistics. Examples of such statistics include K-nearest neighbor based Entropy estimators, Euler characteristics, minimal spanning trees and Random forests. The central challenge in these problems is to handle the delicate dependency structure arising in such statistics. We handle this by utilizing the concept of stabilization, which we combine with Stein's method to establish our results. This notion of stabilization is quite universal in characterizing the local dependencies and thus provides a powerful tool for obtaining normal approximation analysis and doing finite-sample statistical inferences. Additionally, bootstrap methodology can be considered for performing practical statistical inference for the above class of problems.

Contents

1 Introduction

Let $(\mathbb{X}, \mathcal{F})$ be a metric measure space equipped with a σ-finite measure $\mathbb{Q}$ and a metric $d : \mathbb{X} \times \mathbb{X} \to [0, \infty)$. For $s \geq 1$, let $\mathcal{P}_s$ denote the canonical Poisson process on $\mathbb{X}$ with intensity measure $\lambda := s\mathbb{Q}$, and for $\mathbb{Q}$ a probability measure, let ξ_n denote a binomial process with n i.i.d. observations from $\mathbb{Q}$. Let $d_K(Y, Z)$ denote the Kolmogorov distance between two random variables Y, Z, i.e., $d_K(Y, Z) := \sup_{t \in \mathbb{R}} | \mathbb{P}(Y \leq t) - \mathbb{P}(Z \leq t)|$. In this thesis, we study normal approximation results for real-valued functionals $F_s(\mathcal{P}_s)$ and $F_n(\xi_n)$, respectively, of the Poisson and the binomial point processes in the Kolmogorov and Wasserstein metrics under relatively flexible assumptions on the functionals. In particular, motivated by geometric and topological statistics, we focus on the case when the functionals F_s and F_n are not necessarily expressible as sums of certain score functions, and on obtaining presumably optimal bounds in this case.

Our proof techniques are based on the concept of *stabilization*, which was first used by Kesten and Lee (1996) to establish central limit theorems for Poisson-based minimal spanning tree. Since then, it has been extensively developed as a general tool to establish normal approximation rates for various functionals of Poisson and Binomial point process. We refer the interested reader to Penrose and Yukich (2001); Penrose (2005); Baryshnikov and Yukich (2005); Penrose and Yukich (2005); Schreiber (2010); Last et al. (2016); Chatterjee and Sen (2017); Lachièze-Rey and Peccati (2017) and reference therein for details. In particular, Last et al. (2016) develops normal approximation bounds for a fairly general class of functions of Poisson processes by combining Malliavin-Stein techniques (Peccati et al., 2010; Peccati and Reitzner, 2016), second-order Poincaré inequalities (Chatterjee, 2009; Nourdin et al., 2009) and stabilization concepts, and by using the iterated add-one cost operator, also called second order cost operator. Considering the case of functionals expressible as a sum of exponentially stabilizing score functions, Lachièze-Rey et al. (2019) establishes user-friendly normal approximation results based on Last et al. (2016). The work of Lachièze-Rey et al. (2022) introduces bounds for general functionals of Poisson process. Their method does not involve the hard-to-evaluate iterated add-one cost operators but uses the add-one cost operator at two different scales, an approach pioneered by Chatterjee and Sen (2017) for the case of Poisson-based minimal spanning trees. However, their generality comes at the cost of sub-optimality – in general, the bounds based on Lachièze-Rey et al. (2022) are sub-optimal compared to those of Last et al. (2016). Furthermore, for the case of functionals that are expressible as a sum of exponentially stabilizing score functions, the bounds of Lachièze-Rey et al. (2022) necessarily lead to sub-optimal rates.

Hence, the following question remains: *Can one obtain presumably optimal bounds for general functionals that automatically result in presumably optimal bounds when specialized to the case of functionals that can be expressed as sums of score functions?* Following Lachièze-Rey et al. (2019), we use the term *presumably optimal* to refer to the case when the order of the normal approximation is the same as that of a sum of i.i.d. random variables. In this thesis, we answer this question in the affirmative for a class of functionals. Similar to Lachièze-Rey et al. (2022), our approach is based on the idea of using the add-one cost operator at two scales. However, in contrast to their work, we use it to directly simplify the evaluation of the iterated add-one cost operators. When specialized to the case of sums of score functions, such an approach recovers the presumably optimal results of Lachièze-Rey et al. (2019). In summary, we make the following contributions:

- In Definition 2.7, we leverage a flexible notion of the add-one cost operator with a general set A_x that allows to (relatively) easily evaluate computations with the iterated add-one cost or second-order difference operators, for general functionals of Poisson and binomial point process that are not necessarily a sum. While this notion was proposed originally in Lachièze-Rey et al. (2022), our use of it appears to be new and different.

- In our main results, Theorems 2.1 and 2.2, we provide normal approximation results for functionals of Poisson and binomial point processes, respectively. In particular, the functionals do not necessarily need to be expressible as sums of certain score functions.
- We illustrate the applicability of our approach by deriving normal approximation results for several geometric and topological statistics. Specifically, in Theorem 2.3 and 2.4 we use our approach to derive normal approximation results for the total edge length of k-Nearest Neighbor graphs and weighted k-Nearest Neighbor graph based Shannon entropy estimators. In Theorem 2.5, we derive results for the Euler Characteristic, which is an elementary statistics widely used in the field of topological data analysis. Finally, we discuss the applicability of our approach for the minimal spanning tree problem in Theorem 2.6, by recovering existing results via our approach.

Furthermore, in this thesis, we also considered a very important statistics: random forest, which lies in this stabilization framework, where we adopted and developed a generalized version of stabilization concept: region-based stabilization. Random forest (Amit and Geman, 1997; Ho, 1998; Breiman, 2001) is an extremely successful general-purpose prediction method. It is based on aggregating randomized tree-based base-learners and has been successfully applied in a wide range of fields including remote sensing (Belgiu and Drăguţ, 2016), healthcare (Khalilia et al., 2011; Qi, 2012) and causal inference (Wager and Athey, 2018). Different versions of random forest mainly differ by their randomized tree-building processes. Broadly speaking, there are two sources of randomness while constructing the random forest.

- *Bagging or sub-sampling:* The first is due to the use of bagging or sub-sampling, where a randomly chosen subset of the entire training data is used to construct the individual (random) trees. Breiman (1996a) originally considered the random forest with bagging where the trees are constructed by bootstrap sub-samples from the original sample. It was also shown that for unstable base-learners, the bagging random forest can increase the accuracy; see Breiman (1996b) for more details. Moreover, other works including those by Peng et al. (2022), Mentch and Hooker (2016) and Wager (2014) focus on averaging all possible sub-samples of the training, with a particular fixed size.

- *Random tree construction:* The second source depends on the way in which the random-tree based base-learners are constructed. If the randomization in the tree construction only depends on the covariates, then such random forests are called non-adaptive random forests. They are also closely related to the kernel-based prediction techniques (Lin and Jeon, 2006; Scornet, 2016b). From the perspective of nearest neighbor regression methods, non-adaptive random forests could also be viewed as an implicitly adaptive nearest neighbor method. Conversely, if the randomization in the tree construction also depends on the response, such random forests are called adaptive random forests.

Despite the widespread usage, progress in the theoretical understanding of their statistical properties has been rather slow. Motivated by the works of Breiman (2000, 2004), the most theoretically well-studied version of random forest has been the non-bagging and non-adaptive random forest. Specifically Lin and Jeon (2006), Meinshausen (2006), Biau and Devroye (2010), Biau et al. (2008), Biau (2012), Scornet (2016a), Scornet (2016b), Mourtada et al. (2020), Klusowski (2021) and Klusowski and Tian (2022) studied consistency properties, with a few of them also establishing minimax rates over various smooth function classes. Biau and Scornet (2016) provide an exposition of the above results. Wager et al. (2014), Mentch and Hooker (2016), Zhou et al. (2021) and Cattaneo et al. (2023) studied asymptotic normality and developed asymptotically valid confidence intervals.

The focus in the above works are mainly about the fluctuations with respect to the bagging procedure, and are agnostic to the randomization in the base-learners. Due to the asymptotic nature, the above works do not provide any insight into when the Gaussian behavior actually kicks-in, from a finite-sample perspective.

Many practical versions of random forests are truly adaptive (Breiman, 2001). From a theoretical perspective, a condition that bridges adaptive and non-adaptive random forest is that of *honesty* considered in various forms in Biau et al. (2008); Biau (2012). Roughly speaking, an honest (random) tree is defined as a tree that avoids using the same training labels for both selecting split-points for the tree construction process, and for making the predictions. This essentially makes the random forest non-adaptive for all statistical analysis purposes. This condition was further examined in detail by Wager and Athey (2018). Specifically, they argue that honesty-type conditions are essentially *necessary* to obtain point-wise asymptotic *Gaussianity* of non-bagging random forest predictions, and provide various examples. However, we would like to remark here that Gaussian limits (and other non-Gaussian limits) might be possible under certain non-standard scalings.

Alternatively, Scornet et al. (2015) and Chi et al. (2022) studied consistency of adaptive random forest (without requiring any honesty-type conditions) under additive model assumptions on the truth and with CART splitting criterion respectively. We are not aware of any further theoretical analysis of statistical properties (e.g., minimax rates and asymptotic normality) of random forests without honest-type conditions. Apart from studying classical statistical properties, works such as Mentch and Zhou (2020) and Tan et al. (2022) have also looked at explanations for the better (or worse) performance of random forests and related methods over other approaches with similar or comparable statistical properties.

From a finite-sample inference perspective, it is essential to provide non-asymptotic Gaussian approximation bounds for random forest predictions. To the best of our knowledge, only Peng et al. (2022) establishes such bounds for sub-sampling based non-adaptive random forest predictions in the *univariate* setting. Their approach was based on directly leveraging standard results on Berry-Esseen bounds for U-statistics by Stein's method (see e.g., Chen et al., 2011, Chapter 10). They also specialized their result to the case when the base-learners are the so-called k-nearest neighbors or k-potential nearest neighbors (see Section 3.1 for details). However, as a consequence of a direct limitation of their proof techniques, they are only able to handle the case of fixed k. Handling the case of growing k is highly non-trivial and requires a very different proof technique.

In contrast to the above works, our main goal is to obtain multivariate Gaussian approximation bounds for random forest predictions (see Theorem 3.1 and Corollary 3.1), focusing on the non-bagging and non-adaptive version in the context of regression with growing k. Apart from being applicable to non-bagging and non-adaptive random forests, our results are also applicable to the so-called purely random forest such as the one studied by Mourtada et al. (2020). From a technical point-of-view, we show that random forests satisfy a certain region-based stabilization property (see Section 3.3.2), which enables us to develop and leverage tools based on Stein's method to establish a Gaussian approximation result.

Standard stabilization-based approaches, in combination with Stein's method and second-order Poincaré inequalities are well-studied to establish Gaussian approximation for functionals of Poisson and binomial point processes; see, e.g., Last et al. (2016); Lachièze-Rey et al. (2019, 2022); Schulte and Yukich (2019); Shi et al. (2023+); Schulte and Yukich (2023b) and references therein for details. Recently Bhattacharjee and Molchanov (2022) developed the notion of region-based stabilization which strictly generalizes standard stabilization, and is more widely applicable. In this thesis, we extend the univariate results in Bhattacharjee and Molchanov (2022) to the multivariate setting; Theorems 3.2 and 3.3 are widely applicable to a class of multivariate functionals of a Poisson process whose score functions satisfy the region-based stabilization property. Specializing these results to random forests, we obtain the multivariate Gaussian approximation bounds in Theorem 3.1 and Corollary 3.1.

All our results for the multivariate settings are under the Poisson sampling setting. This entails that to handle the widely studied case of independent and identically distributed (i.i.d.) observations, we need to use a de-Poissonization trick. The use of Poisson sampling setting is actually due to the technical reason that there is no natural multivariate second-order Poincaré inequality for the case of binomial point processes; see Remark 3.6 for more details. However, we would like to highlight that the proof techniques in this paper are potentially valid, with some appropriate modifications, when a univariate normal approximation of the random forest is considered under a binomial sampling regime (i.e., i.i.d. samples). Furthermore, despite the fact that we focus on a non-adaptive random forest as an example in Theorem 3.1 and Corollary 3.1, our approach of using region-based stabilization theory and Stein's method to establish Gaussian approximation results, as stated in Theorems 3.2 and 3.3, is much more widely applicable for adaptive random forests and other non-parametric regression problems such as Nadaraya-Watson and wavelets-type in which case, one would need to work with appropriate regions (depending on the procedure) and then apply our general results in Theorems 3.2 and 3.3.

2 Gaussian Approximation for General Stabilizing Statistics

In this section, we focus on general normal approximation theories on statistics that satisfy certain stabilization property with some examples including: total edge length of k-nearest neighbor graphs, Shannon entropy, Euler characteristic and edge length of the minimal spanning tree.

2.1 Preliminaries

2.1.1 Point process basics

Let $(\mathbb{X}, \mathcal{F})$ be a measure space with a σ-finite measure $\mathbb{Q}$ and a metric $d : \mathbb{X} \times \mathbb{X} \to [0, \infty)$. Let $\mathbf{N}$ be the set of all locally finite counting measures on $\mathbb{X}$, which can be interpreted as point configurations in $\mathbb{X}$. Thus, we treat the elements in $\mathbf{N}$ also as sets. The collection $\mathbf{N}$ is equipped with the smallest σ-field $\mathcal{N}$ such that the maps $m_A : \mathbf{N} \to \mathbb{N} \cup \{0, \infty\}, \mathcal{M} \mapsto \mathcal{M}(A)$ are measurable for all $A \in \mathcal{F}$; see Kallenberg and Kallenberg (1997) and Last and Penrose (2011). A point process η is a random element in $\mathbf{N}$. Denote by $\mathbf{F}(\mathbf{N})$ the class of all measurable functions $f : \mathbf{N} \to \mathbb{R}$, and by $L^0(\mathbb{X}) := L^0(\mathbb{X}, \mathcal{F})$ the class of all real-valued, measurable functions F on $\mathbb{X}$. Note that, as $\mathcal{F}$ is the completion of $\sigma(\eta)$, each $F \in L^0(\mathbb{X})$ can be written as $F = f(\eta)$ for some measurable function $f \in \mathbf{F}(\mathbf{N})$. Such a mapping f, called a *representative* of F, is $\mathbb{Q} \circ \eta^{-1}$-a.s. uniquely defined. In order to simplify the discussion, we make this convention: whenever a general function F is introduced, we will select one of its representatives and denote such a representative mapping by the same symbol F. Throughout this paper, we denote by $L^2_{\boldsymbol{\eta}} = L^2_\eta(\mathbb{X})$ the space of all square-integrable functions F of a point process η with $\mathbb{E}F^2(\eta) < \infty$. We mainly consider two different classes of point processes: Poisson point process and binomial point process.

Definition 2.1 (Poisson point process). *A Poisson point process with intensity measure λ is a point process $\mathcal{P}(\lambda)$ on $\mathbb{X}$ with the following two properties:*

1. *$\forall B \in \mathcal{F}$, $\mathcal{P}(\lambda)(B)$ is a Poisson random variable with parameter $\lambda(B)$.*
2. *$\forall m \in \mathbb{N}_+$ and for any pairwise disjoint sets $B_1, B_2, ..., B_m \in \mathcal{F}$, we have that the random variables $\mathcal{P}(\lambda)(B_1), \mathcal{P}(\lambda)(B_2), \ldots, \mathcal{P}(\lambda)(B_m)$ are independent.*

Definition 2.2 (Binomial point process)**.** *Let P be a probability distribution and n be a fixed positive integer. Let $X_1, X_2, ..., X_n$ be i.i.d. random variables sampled from P. The binomial point process ξ_n based on P and n is defined as $\xi_n := \sum_{i=1}^n \delta_{X_i}$, where δ is the Dirac measure.*

We now describe the setting for developing normal approximations of functionals of Poisson and binomial point processes. For $s \geq 1$, let $\lambda := s\mathbb{Q}$ be the intensity measure of Poisson point process $\mathcal{P}(\lambda) := \mathcal{P}_s$. For the case when $\mathbb{Q}$ is the probability measure, let ξ_n be the binomial point process based on $\mathbb{Q}$ and n. Consider square-integrable functionals of these two point processes, i.e., $F(\mathcal{P}_s) := F_s(\mathcal{P}_s) \in L^2_{\mathcal{P}_s}(\mathbb{X})$ and $F(\xi_n) := F_n(\xi_n) \in L^2_{\xi_n}(\mathbb{X})$. We then seek upper bounds for the following two quantities:

$$d_K\left(\frac{F_s(\mathcal{P}_s) - \mathbb{E}F_s(\mathcal{P}_s)}{\sqrt{\operatorname{Var} F_s(\mathcal{P}_s)}}, N\right), \qquad \text{and} \qquad d_K\left(\frac{F_n(\xi_n) - \mathbb{E}F_n(\xi_n)}{\sqrt{\operatorname{Var} F_n(\xi_n)}}, N\right),$$

where N is a standard normal random variable.

2.1.2 Stabilization

The notion of stabilization is widely used in deriving normal approximation rates for functionals of Poisson or binomial point processes (Kesten and Lee, 1996; Penrose and Yukich, 2001; Penrose, 2005; Penrose and Yukich, 2005). We start with introducing notions of stabilization for functionals that are not necessarily representable as sums of sore functions.

Definition 2.3 (Add-one cost operator)**.** *Let F be a measurable functional of a point process η on $(\mathbb{X}, \mathcal{F})$. The family of Add-One Cost Operators, $D = (D_x)_{x\in\mathbb{X}}$, are defined as*

$$D_x F(\eta) := F(\eta \cup \{x\}) - F(\eta).$$

Similarly, we can define a second-order cost operator (also called iterated add-one cost operator): for any $x_1, x_2 \in \mathbb{X}$,

$$D_{x_1,x_2} F(\eta) := F(\eta \cup \{x_1, x_2\}) - F(\eta \cup \{x_1\}) - F(\eta \cup \{x_2\}) + F(\eta).$$

In addition, we define for any $y \in \mathbb{X}$,

$$D_x F^y(\eta) := F(\eta \cup \{y\} \cup \{x\}) - F(\eta \cup \{y\}).$$

Clearly, when $\{y\} = \varnothing$, $D_x F^y$ degenerates into the add-one cost operator.

Based on the add-one cost operator introduced above, we next introduce weak and strong stabilization in the context of the functionals F of the point process η.

Definition 2.4 (Weak stabilization)**.** *The functional F is said to be weakly stabilizing at $x \in \mathbb{X}$, if and only if there exists a random variable Δ_x such that for any sequence $(W_m)_{m\geq 1}$ in $\mathcal{F}$ tending to $\mathbb{X}$, as $m \to \infty$, we have $D_x F(\eta \cap W_m) \to \Delta_x$, almost surely.*

Definition 2.5 (Strong stabilization)**.** *The functional F is said to be strongly stabilizing at $x \in \mathbb{X}$, if and only if there exists an almost surely finite random variable R_x, which is referred to as the radius of stabilization, such that for all finite $\mathbb{A} \subset \mathbb{X} \backslash B_x(R_x)$, with probability 1,*

$$D_x F((\eta \cap B_x(R_x)) \cup \mathbb{A}) = D_x F(\eta \cap B_x(R_x)),$$

where $B_x(R_x) := \{y \in \mathbb{X} : d(x, y) \leq R_x\}$.

Clearly, strong stabilization implies weak stabilization. We now discuss some special cases. In the context of Poisson or binomial point process, the functionals F_s and F_n can sometimes be represented as sums of the form

$$F_s := F_s(\mathcal{P}_s) := \sum_{x \in \mathcal{P}_s} f_s(x, \mathcal{P}_s), \qquad \text{and} \qquad F_n := F_n(\xi_n) := \sum_{x \in \xi_n} f_n(x, \xi_n),$$

where f_s, f_n are called *score functions*. In this case, there exists a useful notion of stabilization based on the score functions. For simplicity, we still state the definition for functionals of Poisson point process; the binomial case is defined similarly.

Definition 2.6 (Score-based stabilization (Lachièze-Rey et al., 2019))**.** *The score function f is said to be stabilizing at $x \in \mathbb{X}$, if and only if there exists an almost surely finite random variable r_x (the radius of stabilization) such that for all finite $\mathbb{A} \subset \mathbb{X} \backslash B_x(r_x)$, we have*

$$f(x, (\eta \cap B_x(r_x)) \cup \mathbb{A}) = f(x, \eta \cap B_x(r_x)).$$

Informally speaking, the above definition posits that the value of the score function f will not be affected by the points outside the ball centered at x with radius r_x. For discussing the relation between strong stabilization and score-based stabilization, we present the following simple result.

Proposition 2.1. *Given any $F_s(\mathcal{P}_s) = \sum_{y \in \mathcal{P}_s} f_s(y, \mathcal{P}_s)$, we have for all $x \in \mathbb{X}$,*

$$D_x F_s(\mathcal{P}_s) = f_s(x, \mathcal{P}_s \cup \{x\}) + \sum_{y \in \mathcal{P}_s} D_x f_s(y, \mathcal{P}_s).$$

A similar result also holds for F_n.

Proof of Proposition 2.1. The proof is straightforward by using the definitions. See Lachièze-Rey et al. (2019, Lemma 5.2). □

Remark 2.1.

(1) In contrast to score-based notions of stabilization, strong stabilization focuses on the cost function of the functional F itself. The above result reveals the relationship between the two notions. It plays an important role in Section 2.4.2 and its proof.

(2) Strong stabilization does not restrict the form of the functional to be expressible as a sum of scores, and, if applicable, it thus avoids finding appropriate scores that can be handled by the method.

The following example from the literature on Topological Data Analysis (TDA), further illustrates the aforementioned remarks. Readers unfamiliar with the basics of TDA are directed to the elementary definitions provided in the Supplementary Material (Shi et al., 2023). We also refer to Edelsbrunner and Harer (2010) and Boissonnat et al. (2018) for more on the basics of TDA.

Example 2.1 (Euler characteristic)**.** *Given a simplicial complex K, the Euler characteristic is defined as*

$$\chi(K) := \sum_{k=0}^{\infty} (-1)^k \#\{S_k\},$$

where $\#\{S_k\}$ is the number of simplices of dimension k.

Typically, the simplicial complex K, is taken to be the Vietoris-Rips complex (VR complex) or the Čech complex, constructed over a point cloud sampled from binomial or Poisson point processes ξ_n or $\mathcal{P}_s$, respectively. In this case, we denote the simplicial complex as $K(\xi_n)$ or $K(\mathcal{P}_s)$ to denote the dependency on the underlying point process explicitly. We now discuss the stabilization properties of the above statistic. While it is possible to express the Euler characteristic as a sum of certain score functions, it is not required to do so, as the Euler characteristic is strongly stabilizing with radius of stabilization $R_x = 2r$ for the Čech and the VR-complex; see Krebs et al. (2021).

The following example on the total edge length of a k-nearest neighbor graph is a canonical example of a geometric statistic satisfying score-based stabilization and strong stabilization.

Example 2.2 (k-nearest neighbor (k-NN) graphs)**.** *Consider a configuration of a Poisson point process $\mathcal{P}_s$, where here we represent $\mathcal{P}_s$ by a random number of (conditionally) i.i.d. points X_i, i.e. $\mathcal{P}_s = \{X_i\}_{i=1}^{|\mathcal{P}_s|}$. For some $k \in \mathbb{N}_+$, and for every integer $1 \le j \le k$, denote by $X_{j,i}$ the j-nearest neighbor [1] of X_i, i.e. $X_{j,i}$ is the j-th closest point to X_i. Furthermore, let $\rho_{j,i}$ denote the distance between $X_{j,i}$ and X_i. Then, the (undirected) k-NN graph $NG_k(\mathcal{P}_s)$ is the graph with the vertex set $V := \mathcal{P}_s$ and an edge $x \sim y$ if y is some j-nearest neighbor of X and (or) x is some j-nearest neighbor of y. For $\vartheta > 0$, we define*

$$f_s(x, \mathcal{P}_s) := \begin{cases} \sum_{x \sim y} \frac{1}{2} d(x,y)^{\vartheta}, & \text{if } x, y \text{ are mutual } k\text{-nearest neighbors,} \\ \sum_{x \sim y} d(x,y)^{\vartheta}, & \text{if } x, y \text{ are not mutual } k\text{-nearest neighbors.} \end{cases} \tag{2.1}$$

The total edge length is defined as

$$F_s := \sum_{x \in \mathcal{P}_s} f_s(x, \mathcal{P}_s).$$

According to Lachièze-Rey et al. (2019), the total edge length statistic satisfies score-based stabilization. Additionally, by the proof of Lemma 6.1 in Penrose and Yukich (2001), we also have that it satisfies strong stabilization. For instance, when $d = 2$, the radius of stabilization is $R_x = 4R$, where R is defined in the following way: for each $t > 0$, construct six disjoint equilateral triangles $T_j(t)$, $1 \le j \le 6$, such that the origin is a vertex of each triangle, such that each triangle has edge length t and such that $T_j(t) \subset T_j(u)$ whenever $t < u$. Then, define R to be the minimum t such that each triangle $T_j(t)$ contains at least $k+1$ points from $\mathcal{P}_s$.

By definition, strong stabilization only focuses on the first-order add-one cost operator. In order to deal with second-order cost operators, which are also crucial in obtaining our normal approximation results, we introduce the following flexible notion of add-one cost operators.

Definition 2.7 (Flexible add-one cost, Lachièze-Rey et al. (2022))**.** *For any point process η in $(\mathbb{X}, \mathcal{F})$, any $x \in \mathbb{X}$ and a set $A_x \in \mathcal{F}$ (that may or may not depend on x), the flexible add-one cost operator for the functional F is defined as*

$$D_x F(A_x) := D_x F(A_x)(\eta) := D_x F(\eta|_{A_x}) := F((\eta|A_x) \cup \{x\}) - F(\eta|A_x),$$

where we denote by $\eta|A_x$ the restriction of the point process η to the set A_x.

Informally speaking, the flexible add-one cost $D_x F(A_x)$ observes the point process in the 'window' A_x. Obviously, if one sets $A_x = \mathbb{X}$, the flexible add-one cost function degenerates into the

[1] We refer to Lachièze-Rey et al. (2019) for details about the the tie-breaking procedure.

classical add-one cost function in Definition 2.3. The above flexible cost functions were proposed in Lachièze-Rey et al. (2022). We here name it "flexible" as our main Theorem will allow the set A_x to be picked adaptively.

Observe that the second-order cost functions can be expressed as the telescoping sums of differences of flexible cost functions and standard add-one cost functions:

$$D_{x_1,x_2}F = (D_{x_1}F^{x_2} - D_{x_1}F^{x_2}(A_x)) + (D_{x_1}F^{x_2}(A_x) - D_{x_1}F(A_x)) + (D_{x_1}F(A_x) - D_{x_1}F)\,.$$

We use this simple observation in our proofs in a different way compared to Lachièze-Rey et al. (2022) to obtain improved rates.

2.2 Assumptions

We now discuss the assumptions to obtain the normal approximation results. Following Penrose (2007); Penrose and Yukich (2005); Yukich (2015); Lachièze-Rey et al. (2019), we make the following assumption on the measure $\mathbb{Q}$: There exist constants $\kappa > 0, \omega > 1$ such that for $r \geq 0$ and all $x \in \mathbb{X}$,

$$\limsup_{\epsilon \to 0^+} \frac{\mathbb{Q}(B_x(r+\epsilon)) - \mathbb{Q}(B_x(r))}{\epsilon} \leq \kappa\omega r^{\omega-1}. \tag{2.2}$$

This assumption implies $\mathbb{Q}$ is diffuse, i.e. $\mathbb{Q}(\{x\}) = 0$ for all $x \in \mathbb{X}$ (see Lachièze-Rey et al., 2019, Lemma 5.1 (a)). For example, one can consider a measure $\mathbb{Q}$ on $\mathbb{X}$, a full dimensional subset of $\mathbb{R}^d$, with a bounded density f with respect to the Lebesgue measure, where one can choose $\kappa = V_d \sup f$, where V_d is the volume of the unit ball, and $\omega = d$. Another example is given by Riemannian manifolds $\mathbb{X}$ with non-negative Ricci curvature, with the semi-metric d being the geodesic distance; see Lachièze-Rey et al. (2019) for more details and other examples.

We also make the following tail-bound assumption on the radius of strong stabilization. Recall that F_s refers to a functional of a Poisson process with intensity measure $s\mathbb{Q}$ while F_n stands for a functional of a binomial process associated with n i.i.d. points sampled from $\mathbb{Q}$.

Assumption 2.1 (Decay of radius of stabilization)**.** *Under the setting of strong stabilization and (2.2), we say the radius of stabilization R_x of F_s decays exponentially if and only if there exist constants $c_1, c_2, c_3 > 0$ such that for $r \geq 0$,*

$$\mathbb{P}(R_x \geq r) \leq c_1 e^{-c_2 (s^{1/\omega} r)^{c_3}}.$$

If R_x is the radius of stabilization of F_n, then the same definition holds with s replaced by n.

Yet another reason for why we refer to Definition 2.7 as "flexible" is that even when the tail probability of the radius of stabilization R_x is unknown for a specific functional, it might be possible to pick A_x "strategically" and use our approach to obtain normal approximation bounds. We illustrate this point in the proofs of the results of Section 2.4.4 by using our approach to recover existing results on normal approximation for the total edge length of the minimal spanning tree.

Assumption 2.2 ($\mathbb{K}$-exponential bound)**.** *We say the add-one cost function $D_x F_s$ satisfies a $\mathbb{K}$-exponential bound, where $\mathbb{K}$ is a measurable subset of $\mathbb{X}$, if and only if for $x, x^* \in \mathbb{X}$, there exist constants $k_1, k_2, k_3 > 0$ such that*

$$\mathbb{P}(D_x F_s \neq 0) \leq k_1 e^{-k_2 d_s(x,\mathbb{K})^{k_3}},$$

and

$$\mathbb{P}\big(|D_{x,x^*} F_s| \neq 0\big) \leq k_1 e^{-k_2 \max\{d_s(x,x^*), d_s(x,\mathbb{K}), d_s(x^*,\mathbb{K})\}^{k_3}},$$

where $d_s(\cdot,\cdot) := s^{1/\omega} d(\cdot,\cdot)$ and $d(x,\mathbb{K}) \coloneqq \inf_{y\in\mathbb{K}}\ d(x,y)$. A similar assumption holds for F_n by changing s to n.

Here, we make the assumption directly on the cost functions, which captures a more general class of functionals than that considered in Lachièze-Rey et al. (2019). Particularly, when the functional is expressed as a sum of scores, the above assumption coincides with a condition mentioned in Lachièze-Rey et al. (2019, Proof of Lemma 5.9). Moreover, a similar assumption has been made in Lachièze-Rey et al. (2019, Equations (2.8) and (2.9)) on the score functions to capture functionals whose variances exhibit surface area order scaling. If $\mathbb{K} = \mathbb{X}$, the second bound is directly obtained by the definition of strong stabilization and Assumption 2.1.

Assumption 2.3 (Moment Condition)**.** *The cost function of F_s satisfies the following moment condition: For $p > 4$,*

$$\sup_{s\geq 1} \sup_{x,x^*\in\mathbb{X}} \left(\mathbb{E}\, |D_x F_s|^p + \mathbb{E}|D_x F_s^{x^*}|^p\right) =: H < \infty. \tag{2.3}$$

A similar assumption holds for $D_x F_n$ by simply changing s to n.

Bounded moment conditions are commonly assumed when normal approximation results are derived. For related work in the context of stabilizing functionals of point process, see Lachièze-Rey et al. (2019), equations (2.6) and (2.7), and Lachièze-Rey et al. (2022), equations (1.5) and (1.8). While Lachièze-Rey et al. (2019) considers moment conditions on score functions, we directly deal with the cost functions so that it fits a more general class.

2.3 Main results

We now present our two main results on the normal approximation of a certain class of functionals of Poisson and binomial point process, Theorem 2.1 and 2.2 respectively, that are not necessarily expressible as sums of score functions. We discuss several applications in Section 2.4.

We first present our main result for the Poisson case. Note that while this result is not leveraging Assumptions 2.1, 2.2 and 2.3, the refined result presented in Corollary 2.1 does.

Theorem 2.1 (Normal approximation for functionals of Poisson point processes)**.** *Let F be a functional of the Poisson point process $\mathcal{P}(\lambda)$ with $F \in L^2_{\mathcal{P}(\lambda)}$ and $\mathbb{E}\int (D_x F)^2 \lambda(dx) < \infty$. For any $x, x_1, x_2 \in \mathbb{X}$ and sets $A_x, A_{x_1} \in \mathcal{F}$, define*

$$\mathbb{E}|D_x F - D_x F(A_x)|^4 =: b_1(x, A_x), \quad \mathbb{E}|D_x F(A_x)|^4 =: b_2(x, A_x),$$

and

$$\begin{aligned} \mathbb{E}|D_{x_1} F^{x_2} - D_{x_1} F^{x_2}(A_{x_1})|^4 &=: b_3(x_1, x_2, A_{x_1}), \\ \mathbb{E}|D_{x_1} F^{x_2}(A_{x_1}) - D_{x_1} F(A_{x_1})|^4 &=: b_4(x_1, x_2, A_{x_1}). \end{aligned}$$

Then, there is an absolute constant $C^* > 0$ such that

$$d_K\Big(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N\Big) \leq C^* \sum_{i=1}^{6} \gamma_i',$$

where

$$\gamma_1{}' := \frac{1}{\operatorname{Var} F}\Bigg(\int \Big[\sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{4}} \sum_{j=1}^{2} b_j(x_2, A_{x_2})^{\frac{1}{4}}$$

$$\times \Big(b_1(x_3, A_{x_3})^{\frac{1}{4}} + \sum_{j=3}^{4} b_j(x_3, x_1, A_{x_3})^{\frac{1}{4}}\Big)\Big(b_1(x_3, A_{x_3})^{\frac{1}{4}} + \sum_{j=3}^{4} b_j(x_3, x_2, A_{x_3})^{\frac{1}{4}}\Big)\Big]$$

$$\times\ \lambda^3(d(x_1, x_2, x_3))\Bigg)^{\frac{1}{2}},$$

$$\gamma_2' := \frac{1}{\operatorname{Var} F}\Bigg(\int \Big(b_1(x_3, A_{x_3}) + \sum_{j=3}^{4} b_j(x_3, x_1, A_{x_3})\Big)\Big(b_1(x_3, A_{x_3}) + \sum_{j=3}^{4} b_j(x_3, x_2, A_{x_3})\Big)$$

$$\times\ \lambda^3(d(x_1, x_2, x_3))\Bigg)^{\frac{1}{2}},$$

$$\gamma_3' := \frac{1}{(\operatorname{Var} F)^{\frac{3}{2}}} \int \sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}} \lambda(dx),$$

$$\gamma_4' := \frac{\int \sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}} \lambda(dx)}{(\operatorname{Var} F)^2}\Bigg(\Big(\int \sum_{j=1}^{2} b_j(x, A_x)^{\frac{1}{2}} \lambda(dx)\Big)^{\frac{1}{2}} + \Big(\int \sum_{j=1}^{2} b_j(x, A_x) \lambda(dx)\Big)^{\frac{1}{4}}$$

$$+ (\operatorname{Var} F)^{\frac{1}{2}}\Bigg),$$

$$\gamma_5' := \frac{1}{\operatorname{Var} F}\Big(\int \sum_{j=1}^{2} b_j(x, A_x) \lambda(dx)\Big)^{\frac{1}{2}},$$

$$\gamma_6' := \frac{1}{\operatorname{Var} F}\Bigg(\int \Big[\sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{2}}\Big(b_1(x_1, A_{x_1})^{\frac{1}{2}} + \sum_{j=3}^{4} b_j(x_1, x_2, A_{x_1})^{\frac{1}{2}}\Big)$$

$$+ \Big(b_1(x_1, A_{x_1}) + \sum_{j=3}^{4} b_j(x_1, x_2, A_{x_1})\Big)\Big] \lambda^2(d(x_1, x_2))\Bigg)^{\frac{1}{2}}.$$

Remark 2.2.

(i) Theorem 2.1 provides an alternative way of controlling the bounds derived in Last et al. (2016, Theorem 1.2) by leveraging the cost function $D_x F_s(A_x)$.

(ii) Theorem 2.1 is valid for deriving normal approximation rates for general functionals or stabilizing functionals not having a known tail probability bound; see Section 2.4.4. When such tail bounds are known, a more refined result is available (see Corollary 2.1 below).

With $A_x = \mathbb{X}$ and $\lambda = s\mathbb{Q}$, Assumptions 2.1, 2.2 and 2.3 can be leveraged to give upper bounds for the following crucial probabilities that appear *implicitly* in the proof of Theorem 2.1:

$$\begin{aligned} I_s(x) &:= \mathbb{P}(D_x F_s \neq 0), \\ J_s(x_1, x_2) &:= P\big(|D_{x_1,x_2} F_s| \neq 0\big). \end{aligned} \tag{2.4}$$

resulting in the following corollary.

Corollary 2.1. *Suppose $F_s \in L^2_{\mathcal{P}_s}$ and that F_s is strongly stabilizing with the radius of stabilization satisfying Assumption 2.1. Furthermore, suppose Assumption 2.2 and Assumption 2.3 hold. Then,* *there exists a constant $C_0 > 0$ depending only on the constants in (2.2) and (2.3) such that for $s \geq 1$,*

$$d_K\left(\frac{F_s - \mathbb{E}F_s}{\sqrt{\operatorname{Var} F_s}}, N\right) \leq C_0\left(\frac{\Theta_{\mathbb{K},s}^{\frac{1}{2}}}{\operatorname{Var} F_s} + \frac{\Theta_{\mathbb{K},s}}{(\operatorname{Var} F_s)^{\frac{3}{2}}} + \frac{\Theta_{\mathbb{K},s}^{\frac{5}{4}} + \Theta_{\mathbb{K},s}^{\frac{3}{2}}}{(\operatorname{Var} F_s)^2}\right),$$

where

$$\Theta_{\mathbb{K},s} := s\int_{\mathbb{X}} e^{-C_2\frac{(p-4)}{4p}\left(\frac{d_s(x,\mathbb{K})}{2}\right)^{C_3}}\mathbb{Q}(dx). \tag{2.5}$$

Corollary 2.2. *Under the conditions of Corollary 2.1, and assuming there exists a constant $C > 0$ such that*

$$\sup_{s\geq 1}\frac{\Theta_{\mathbb{K},s}}{\operatorname{Var} F_s} \leq C, \tag{2.6}$$

there exists a constant $C_0' > 0$ depending on C and the constants in (2.2) and (2.3) such that for $s \geq 1$,

$$d_K\left(\frac{F_s - \mathbb{E}F_s}{\sqrt{\operatorname{Var} F_s}}, N\right) \leq C_0'\frac{1}{\sqrt{\operatorname{Var} F_s}}. \tag{2.7}$$

Next, we consider the binomial case, which requires some different tools as compared to the Poisson case treated above. Indeed, for the Poisson case we leverage Theorem **??** for our proofs. However, due to the fact that there is no nice counterpart of the second-order Poincaré inequality (see Last et al., 2016), an analogue of Theorem **??** is not known for the binomial case. Instead, we will use Lachièze-Rey and Peccati (2017, Theorem 5.1) in this binomial setting. This then leads to the following result:

Theorem 2.2 (Normal approximation for functionals of binomial point process)**.** *Suppose $F_n \in L^2_{\xi_n}$ and F_n is strongly stabilizing with radius of stabilization satisfying Assumption 2.1. Furthermore, suppose Assumption 2.2 and Assumption 2.3 hold. Then, there exists a constant $C_0 > 0$ depending only on the constants in (2.2) and (2.3) such that for $n \geq 2$,*

$$d_K\left(\frac{F_n - \mathbb{E}F_n}{\sqrt{\operatorname{Var} F_n}}, N\right) \leq C_0\left(\frac{\Theta_{\mathbb{K},n}^{\frac{1}{2}}}{\operatorname{Var} F_n} + \frac{\Theta_{\mathbb{K},n}}{(\operatorname{Var} F_n)^{\frac{3}{2}}} + \frac{\Theta_{\mathbb{K},n} + \Theta_{\mathbb{K},n}^{\frac{3}{2}}}{(\operatorname{Var} F_n)^2}\right), \tag{2.8}$$

where

$$\Theta_{\mathbb{K},n} := n\int_{\mathbb{X}} e^{-C_2\frac{(p-4)}{4p}\left(\frac{d_n(x,\mathbb{K})}{2}\right)^{C_3}}\mathbb{Q}(dx).$$

Remark 2.3. *Compared to the Poisson case, the exponent of $\Theta_{\mathbb{K},n}$ in the third component of the sum on the right hand side of (2.8) is different. In essence, this difference can be traced back to the above mentioned fundamental fact that there is no nice counterpart of the second-order Poincaré inequality for the binomial case (see Last et al., 2016). Instead, we use the approach taken in Lachièze-Rey et al. (2019, Theorem 4.2) to prove Theorem 2.2.*

Corollary 2.3. *Under the conditions of Theorem 2.2, assume there exists a constant $C > 0$ such that*

$$\sup_{n\geq 1}\frac{\Theta_{\mathbb{K},n}}{\operatorname{Var} F_n} \leq C, \tag{2.9}$$

then there exists a constant $C_0' > 0$ *depending on* C *and the constants in* (2.2) *and* (2.3) *such that for* $n \geq 2$,

$$d_K\left(\frac{F_n - \mathbb{E}F_n}{\sqrt{\operatorname{Var} F_n}}, N\right) \leq C_0' \frac{1}{\sqrt{\operatorname{Var} F_n}}. \tag{2.10}$$

Remark 2.4. *We make the following remarks on the above main results for both Poisson and binomial cases.*

(i) *If* $\mathbb{K} = \mathbb{X}$ *and* $\mathbb{Q}(\mathbb{X}) < \infty$, *the conditions* (2.6) *and* (2.9) *can be simplified as*

$$\sup_{s \geq 1} \frac{s}{\operatorname{Var} F_s} \leq C, \tag{2.11}$$

$$\sup_{n \geq 1} \frac{n}{\operatorname{Var} F_n} \leq C, \tag{2.12}$$

(ii) Optimality: *Following Lachièze-Rey et al. (2019), we refer to cases where the bounds in* (2.7) *and* (2.10) *can be attained, as being presumably optimal. Indeed, Corollary 2.2 and 2.3 show that if the variance of the statistics* F_s, F_n *are bounded below by* $\Theta_{\mathbb{K},s}, \Theta_{\mathbb{K},n}$, *respectively, a presumably optimal normal approximation rate is achieved. To give an intuition on why the above situation is referred to as being presumably optimal, note that for the case of sums of i.i.d random variables, non-trivial i.i.d. random variables can be constructed that achieve the upper bounds of the form in* (2.7) *and* (2.10). *Formal lower bounds on the optimality are available for the case of integer-valued statistics in Englund (1981) and Peköz et al. (2013). Furthermore, in a recent work, Schulte and Yukich (2023a) established lower bounds for a large class of statistics.*

(iii) Wasserstein distance: *The quantitative bounds derived in this section are also valid for the Wasserstein distance. Recall that for two random variables* X, Y *with* $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$, *the Wasserstein distance between* X *and* Y *is defined as*

$$d_W(X, Y) := \sup_{h \in Lip(1)} |\mathbb{E}h(X) - \mathbb{E}h(Y)|,$$

where $Lip(1)$ *stands for the set of all Lipschitz functions* $h : \mathbb{R} \to \mathbb{R}$ *with Lipschitz constant at most* 1. *Indeed, as mentioned in the remark in Section 4 in Lachièze-Rey et al. (2019), the underlying bounds in Theorem 6.1 in Last et al. (2016) and Theorem 4.3 (see also Remark 4.3 in Lachièze-Rey and Peccati, 2017) are true for* d_W *as well. Consequently, our results in this section, which are derived based on the above theorems, are also immediately valid.*

Comparison to related works. We now provide some comparisons to the related work. Firstly, our proof techniques, similar to Lachièze-Rey et al. (2019), are based on several central ideas proposed in Last et al. (2016). For the case of functionals that are expressible as sums of score functions, Lachièze-Rey et al. (2019) established presumably optimal bounds under the score-based stabilization assumption, for both the binomial and Poisson cases. While they too use second-order cost operators, our Theorems 2.1 and 2.2 handle a much larger class of functionals (not necessarily as sums of scores). The work of Lachièze-Rey et al. (2022) consider general functionals (not necessarily sums) and work under strong stabilization assumption. However, they only consider the Poisson case. To get explicit bounds (e.g., their Corollary 1.5 and Proposition 1.12), they introduce a specific form of A_x in their proofs and their overall approach results in sub-optimal rates in comparison to our results, Corollary 2.1, and to Lachièze-Rey et al. (2019) in the case when the functional is expressible as sums of scores. Our Theorem 2.1 generalizes Last et al. (2016), Theorem 1.2, by introducing the flexible cost function $D_xF(A_x)$ for general functionals of Poisson point process. The work of Chatterjee (2008) also considers normal approximations of general functions (not necessarily sums). However, their approach is only valid for the binomial case. Moreover, a further investigation of the main theorems Chatterjee (2008, Theorem 2.2) and Lachièze-Rey and Peccati (2017, Theorem 4.2) reveals that their normal approximation bounds are obtained by computing certain quantities (for example, T_A, T in Chatterjee (2008, Section 2) and T, T' in Lachièze-Rey and Peccati (2017, Section 4)), which are complicated to deal with for some functionals, e.g., Euler characteristic; see Krebs et al. (2021, Proof of Theorem 3.2).

Applying our main results. We conclude this section, with the following three-step procedure illustrating how to apply our main theorems, Theorems 2.1 and 2.2.

- *Step 1:* Check whether the functional F is strongly stabilization (Definition 2.5); whether the tail probability of the radius of stabilization (Assumptions 2.1) can be computed; verify Assumption 2.2 on the cost functions.
  - If the functional is not strongly stabilizing or no upper bound of the radius of stabilization R_x is known, consider the flexible cost functions $D_xF_s(A_x)$ with appropriate choice of A_x and apply Theorem 2.1.
- *Step 2:* Check the bounded moment condition (Assumption 2.3).
- *Step 3:* In order to check for presumable optimality, one can seek to bound the variance, i.e., (2.11) and (2.12).

If the above three steps are satisfied, apply Corollary 2.1 and Theorem 2.2 for the Poisson and binomial settings respectively.

2.4 Applications

In this section, we illustrate the applicability of our bounds in Theorem 2.1, Corollary 2.1 and Theorem 2.2 on several geometric and topological statistics thereby following the agenda outlined above.

2.4.1 Total edge length of k-nearest neighbor graphs

Recall the definition of k-NN Graphs in Example 2.2 and define the total edge length of a k-NN graph based on a Poisson process $\mathcal{P}_s$ as:

$$F_s^{k\text{-NN}}(\mathcal{P}_s) := \sum_{x \in \mathcal{P}_s} f_s(x, \mathcal{P}_s), \tag{2.13}$$

with f_s as defined in (2.1) for some $\vartheta > 0$. Similarly, for an underlying binomial point process ξ_n, we define $F_n^{k-\mathrm{NN}}(\xi_n)$.

Theorem 2.3. *Assume there exists a constant $c > 0$ such that, for $r \leq diam(\mathbb{X}) < \infty$,*

$$\inf_{x \in \mathbb{X}} \mathbb{Q}(B_x(r)) \geq c r^{\omega}. \tag{2.14}$$

If there exists a constant $C > 0$ such that

$$\sup_{s \geq 1} \frac{s}{\operatorname{Var} F_s^{k\text{-}NN}(\mathcal{P}_s)} \leq C, \tag{2.15}$$

then there exists a constant $C_0 > 0$ *such that for* $s \geq 1$,

$$d_K\left(\frac{F_s^{k\text{-}NN}(\mathcal{P}_s) - \mathbb{E}F_s^{k\text{-}NN}(\mathcal{P}_s)}{\sqrt{\mathrm{Var}F_s^{k\text{-}NN}(\mathcal{P}_s)}}, N\right) \leq C_0\frac{1}{\sqrt{s}}.$$

Moreover, if there exists a constant $C > 0$ *such that*

$$\sup_{n\geq 1} \frac{n}{\mathrm{Var}\, F_n^{k\text{-}NN}(\xi_n)} \leq C, \tag{2.16}$$

then for $n \geq 2$,

$$d_K\left(\frac{F_n^{k\text{-}NN}(\xi_n) - \mathbb{E}F_n^{k\text{-}NN}(\xi_n)}{\sqrt{\mathrm{Var}F_n^{k\text{-}NN}(\xi_n)}}, N\right) \leq C_0\frac{1}{\sqrt{n}}.$$

Remark 2.5.

(i) Condition (2.14), is required in addition to (2.2) for the k-NN statistic; see Lachièze-Rey et al. (2019) for details. Note that the total edge length of a k-NN graph (2.13) is expressible as a sum of score functions. Hence, the results in Lachièze-Rey et al. (2019) already provide presumably optimal bounds. Our results above also recover the same bounds.

(ii) Now we compare our results to Lachièze-Rey et al. (2022) in the Poisson setting. Recall that similar to our results, they considered general functionals (not necessarily expressible as sums of scores). However, their generality comes at the cost of not having presumably optimal bounds in the setting of the total edge length of a k-NN graph. Specifically, Lachièze-Rey et al. (2022, Proposition 1.12), term $\sqrt{b_n/n}$ *with* $b_n \to \infty$ *implies that it has a slower rate than* $1/\sqrt{n}$*. This highlights the benefit of our approach: despite its generality, we can still obtain presumably optimal bounds.*

(iii) For the binomial setting, Lachièze-Rey and Peccati (2017) obtained rates in the Kolmogorov metric for the same statistic. However, as discussed in Lachièze-Rey et al. (2019, Remark (i) below Theorem 3.1), their results are sub-optimal and involve additional logarithmic factors, that we avoid.

(iv) For $\mathbb{X}$ *a full-dimensional compact convex subset of* $\mathbb{R}^d$ *and* $\omega = d$*, as shown in Penrose and Yukich (2001, Proof of Theorem 6.1)*[2]*, the conditions (2.15) and (2.16) are satisfied.*

2.4.2 Shannon entropy

Given an i.i.d. sample $X_1, X_2, ..., X_n$ from a density q on $\mathbb{R}^d$, the differential (Shannon) entropy is defined as $H(q) := -\mathbb{E}_{X\sim q} \log q(X) = -\int_{\mathbb{R}^d} q(x)\log q(x)dx$. The nearest neighbor entropy estimate, also known as the Kozachenko-Leonenko estimator, was first proposed in Kozachenko and Leonenko (1987) based on the 1-NN density estimator. A generalization of this estimator based on k-NN density estimator is given by

$$\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{(n-1)V_d\rho_{k,i}^d}{e^{\Psi(k)}}\right),$$

where $\rho_{k,i}$ is the distance between X_i and its k-nearest neighbor among $X_1, X_2, ..., X_n$, $V_d := \pi^{\frac{d}{2}}/\Gamma(1+\frac{d}{2})$ is the volume of a unit d-dimensional Euclidean ball, $\Psi(k) = -\gamma + \sum_{i=1}^{k-1} 1/i$ is the

[2] Penrose and Yukich (2001) consider the case of $\vartheta = 1$. However, a closer examination of the proof shows that it can be easily extended for any $\vartheta > 0$.

digamma function with γ the Euler-Mascheroni constant (Penrose and Yukich, 2013; Berrett et al., 2019).

Consistency and a CLT for the above estimator in a manifold setting were shown in Penrose and Yukich (2013) by stabilization theory. However, a non-trivial bias term arises for $d \geq 4$, rendering the above estimator asymptotically inefficient in the sense of Van der Vaart (2000, page 367). To have an (asymptotically) unbiased and efficient estimator, the following weighted k-NN estimator was proposed in Berrett et al. (2019). Defining ξ_n as the binomial point process associated with $X_1, X_2, ..., X_n$ the proposed estimator could be viewed as a functional of ξ_n, and is given by

$$F_n^{\mathrm{SE}}(\xi_n) := \sum_{i=1}^{n} f_n^w(X_i, \xi_n) \quad \text{where} \quad f_n^w(X_i, \xi_n) := \frac{1}{n}\sum_{j=1}^{k} w_j \log\left(\frac{(n-1)V_d \rho_{j,i}^d}{e^{\Psi(j)}}\right),$$

where w_j are weights (such that $\sum_{j=1}^n w_j = 1$) chosen to cancel the dominant bias term and to make F_n^{SE} asymptotically efficient. We now provide our normal approximation results for the above estimator, based on a slightly modified set of assumptions considered in Berrett et al. (2019).

Theorem 2.4. *Let q be a density of $\mathbb{Q}$ with respect to Lebesgue measure on $\mathbb{R}^d$, and let $\mathbb{X} \subset \mathbb{R}^d$ be the support of q. Assume that q is bounded, and that $m := \lceil \beta \rceil - 1$ times differentiable for $\beta > 0$. Let $a : (0, \infty) \to [1, \infty)$ be a decreasing function such that $a(\delta) = o(\delta^{-\epsilon})$ as $\delta \searrow 0$, for every $\epsilon > 0$. For $x \in \mathbb{X}$, let $r_a(x) := (8d^{\frac{1}{2}} a(q(x)))^{-\frac{1}{\beta \wedge 1}}$ and define:*

$$M_{q,a,\beta}(x) := \max\left\{\max_{t=1,\dots,m} \frac{\|q^{(t)}(x)\|}{q(x)}, \sup_{y \in B_x^o(r_a(x))} \frac{\|q^{(m)}(y) - q^{(m)}(x)\|}{q(x)\|y - x\|^{\beta - m}}\right\},$$

where $B_x^o(r) := B_x(r) \backslash \{x\}$, and assume that

$$\sup_{x : q(x) \geq \delta} M_{q,a,\beta}(x) \leq a(\delta), \ \forall \delta > 0.$$

Define the class of weights as follows: for $k \in \mathbb{N}$, let

$$\mathcal{W}^k := \left\{ w \in \mathbb{R}^k : \textstyle\sum_{j=1}^k w_j \frac{\Gamma(j + \frac{2l}{d})}{\Gamma(j)} = 0, \ \text{for } l = 1, ..., \lfloor \frac{d}{4} \rfloor, \sum_{j=1}^k w_j = 1, w_j = 0, \ \text{if } j \notin \{\lfloor \frac{k}{d} \rfloor, \lfloor \frac{2k}{d} \rfloor, ..., k\} \right\}. \tag{2.17}$$

Then under the conditions of Berrett et al. (2019, Theorem 1), that is, for any $\alpha > d$, $\beta > \frac{d}{2}$ and for any two deterministic sequences of positive integers $k_{0,n}^* = k_0^*$, $k_{1,n}^* = k_1^*$ with $k_0^* \leq k_1^*$, $k_0^* / \log^5 n \to \infty$, $k_1^* = O(n^{\tau_1})$ and $k_1^* = o(n^{\tau_2})$, where, with $\beta^* := \beta \wedge 1$,

$$\tau_1 < \min\left\{\frac{2\alpha}{5\alpha + 3d}, \frac{\alpha - d}{2\alpha}, \frac{4\beta^*}{4\beta^* + 3d}\right\}, \quad \text{and} \quad \tau_2 := \min\left\{1 - \frac{\frac{d}{4}}{1 + \lfloor \frac{d}{4} \rfloor}, 1 - \frac{d}{2\beta}\right\},$$

as well as the assumption (2.14), there exist constants $C_0 > 0, \tau > 0$ (independent of k, n) such that

$$d_K\left(\frac{F_n^{SE}(\xi_n) - H(q)}{\sqrt{\mathrm{Var} F_n^{SE}(\xi_n)}}, N\right) \leq C_0 \left(\frac{k}{n}\right)^{\tau},$$

for $k_0^* \leq k \leq k_1^*$, where τ only depends on α, d and β.

Remark 2.6.

(i) *Asymptotic limit theorems for estimators of the Shannon entropy have been obtained, for example, in Penrose and Yukich (2013), Berrett et al. (2019). The result in Penrose and Yukich (2013) is a central limit theorem not addressing the bias. The result in Berrett et al. (2019), is a central limit theorem characterizing the fluctuations around the population entropy $H(q)$, and was established under the case that the density is supported on $\mathbb{R}^d$. However, no normal convergence rate results were provided in the above works. To our best knowledge, the above result, is the first normal convergence rate result with the true center $H(q)$.*

(ii) *It is also possible to obtain a similar result using the method in Lachièze-Rey et al. (2019) since the estimator $F_n^{SE}(\xi_n)$ is expressible as a sum of score functions.*

(iii) *Furthermore, the result in Theorem 2.4 is provided for the binomial case. To the best of our knowledge, the asymptotic unbiasedness and efficiency of the weighted k-NN estimator of the Shannon entropy based on Poisson point process are still open problems.*

2.4.3 Euler characteristic

Recall Example 2.1. Following the setting of Krebs et al. (2021), consider a bounded density q on $[0,1]^d$. Let ξ_n be a binomial point process associated with n i.i.d. samples according to the density q and let $\mathcal{P}_n$ be a Poisson point process with intensity measure $n\mathbb{Q}$, where $\mathbb{Q}$ has a density q with respect to the Lebesgue measure, i.e., we set $s = n$.

Construct the Čech complex or the Vietoris-Rips complex $K_r(n^{\frac{1}{d}}\mathcal{P}_n), K_r(n^{\frac{1}{d}}\xi_n)$, see Definition 6.2 and 6.3 based on the Poisson point process $\mathcal{P}_n$ and the binomial point process ξ_n respectively with $r > 0$ as the filtration time. Here, K_r represents both complexes for simplicity. The factor $n^{\frac{1}{d}}$ corresponds to the thermodynamic/critical regime (Goel et al., 2019; Owada and Thomas, 2020; Trinh, 2017) such that this is equivalent to the case $nr_n^d \to r \in (0,\infty)$ with r_n as the filtration time. With the above construction, the Euler characteristic is given by

$$F_n^{\mathrm{EC}}(\mathcal{P}_n) := \chi(K_r(n^{\frac{1}{d}}\mathcal{P}_n)) \quad \text{and} \quad F_n^{\mathrm{EC}}(\xi_n) := \chi(K_r(n^{\frac{1}{d}}\xi_n))$$

where the Euler characteristic $\chi(K(\eta))$ for a filtration K constructed from a point cloud sampled from a point process η is defined in Example 2.1.

Theorem 2.5. *Under the above setting, for some $T > 0$ such that $0 < r \leq T$, there exists a constant $C_0 > 0$ such that for $n \geq 1$,*

$$d_K\left(\frac{F_n^{EC}(\mathcal{P}_n) - \mathbb{E}F_n^{EC}(\mathcal{P}_n)}{\sqrt{\mathrm{Var}F_n^{EC}(\mathcal{P}_n)}}, N\right) \leq C_0\frac{1}{\sqrt{n}},$$

and for $n \geq 2$,

$$d_K\left(\frac{F_n^{EC}(\xi_n) - \mathbb{E}F_n^{EC}(\xi_n)}{\sqrt{\mathrm{Var}F_n^{EC}(\xi_n)}}, N\right) \leq C_0\frac{1}{\sqrt{n}}.$$

Remark 2.7. *We make the following remarks about the above result.*

(i) *CLTs and functional limit theorems for the Euler characteristic have been studied in Thomas and Owada (2021) by viewing the Euler characteristic as a process indexed by r. Normal approximation rates of the Euler characteristic under binomial and Poisson sampling were obtained in Krebs et al. (2021) by computing certain geometric quantities (see proof of Theorem 3.2) appearing in the general result in Lachièze-Rey and Peccati (2017). Our flexible stabilization method has the advantage of avoiding such computations.*

(ii) *For the Poisson case, Lachièze-Rey et al. (2022) require a specific form of A_x, rendering their result sub-optimal, i.e., the term $\sqrt{b_n/n}$ in Lachièze-Rey et al. (2022, Proposition 1.12), leads to a slower rate than $1/\sqrt{n}$ that we obtain above.*

(iii) *While Euler characteristic could also be expressible as a sum of score functions, one could possibly leverage the results of Lachièze-Rey et al. (2019) to derive normal convergence rate. Our goal in this example is to demonstrate the flexibility of our general result.*

(iv) *As mentioned in Remark 2.4, part (ii), since the Euler characteristic is integer-valued, the bounds in Theorem 2.5 are optimal.*

All above applications consider stabilizing statistics when there are known tail bounds for the radius of stabilization R_x, i.e., quantities in (2.4). Our Theorem 2.1, however, can deal with the case when we do not have immediate bounds for those probabilities based on the flexible cost function $D_x F_s(A_x)$. We illustrate the above mentioned idea by the following application concerning the minimal spanning tree.

2.4.4 Edge length of the minimal spanning tree

Consider a finite set $V \subset \mathbb{R}^d$ (usually it is embedded in an underlying graph $G := (V, E)$). A spanning tree T of V is a connected graph with the vertex set V. Define

$$M(V) := \min_{T} \sum_{e \text{ is an edge of } T} |e|,$$

where the minimum is taken over all possible spanning trees T of V. According to Penrose (2005) and Chatterjee and Sen (2017), the total edge length statistic $M(V)$ does satisfy certain required stabilization properties. However, to the best of our knowledge, there is no result on the rates of stabilization including quantitative bound on the tail probability of the radius of stabilization. This fact poses a major difficulty when attempting to derive normal approximation rate for $M(V)$ by using other methods such as the approach in Lachièze-Rey et al. (2019). Our flexible stabilization method can yield the following theorem by picking A_x "strategically" to make use of some other existing bounds.

Theorem 2.6. *Following the Euclidean setting in Lachièze-Rey et al. (2022), consider B_0 as the unit hypercube in $\mathbb{R}^d$ centered at the origin and let $B_n := nB_0$, $n \in \mathbb{N}_+$. Given a homogeneous Poisson process $\mathcal{P}(\lambda)$ on $\mathbb{R}^d$ with intensity $\lambda > 0$, let*

$$F_{B_n}^{MST}(\mathcal{P}(\lambda)) := M(\mathcal{P}(\lambda)|_{B_n}).$$

Then, there exist constants $C_0 > 0, 1 > D_1 > 0$ and $D_2 > 0$ not depending on n such that

$$d_K\left(\frac{F_{B_n}^{MST}(\mathcal{P}(\lambda)) - \mathbb{E}F_{B_n}^{MST}(\mathcal{P}(\lambda))}{\sqrt{\operatorname{Var} F_{B_n}^{MST}(\mathcal{P}(\lambda))}}, N\right) \leq \begin{cases} C_0 n^{-D_1}, & \text{if } d = 2, \\ C_0 (\log n)^{-D_2}, & \text{if } d \geq 3. \end{cases}$$

Remark 2.8. *The normal approximation rate of the edge length statistic of the minimal spanning tree has been derived previously in Chatterjee and Sen (2017) and in Lachièze-Rey et al. (2022) with similar results as Theorem 2.6. However, Chatterjee and Sen (2017) only focused on the minimal spanning tree therefore it is hard to generalize for other stabilizing functionals. Lachièze-Rey et al. (2022) used the similar idea of introducing the set A_x but their bounds usually give sub-optimal normal convergence rates (for e.g., see Remark 2.5 regarding the total edge length of k-nearest neighbor graphs in Section 2.4.1) than ours due to the specific form of their set A_x lacking flexibility.*

3 Random Forest and Region-based Stabilization

In this section, we particular focus on one important statistic: random forest, to which we apply our generalized region-based stabilization to obtain normal approximation.

3.1 Random forests and k-potential nearest neighbors

We consider the following regression model:

$$\boldsymbol{y} = r(\boldsymbol{x}, \boldsymbol{\varepsilon}), \quad (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R},\ \boldsymbol{\varepsilon} \in \mathbb{R}, \tag{3.1}$$

where $\boldsymbol{x} \sim \mathbb{Q}$ with a.e. continuous density g on $\mathbb{R}^d$, $d \in \mathbb{N}$ and the noise $\boldsymbol{\varepsilon} \sim P_{\boldsymbol{\varepsilon}}$ independent of $\boldsymbol{x}$. We define the true regression function as $r_0(\cdot) := \mathbb{E}_{P_{\boldsymbol{\varepsilon}}}[r(\cdot, \boldsymbol{\varepsilon})] = \mathbb{E}_{P_{\boldsymbol{\varepsilon}}}[r(\boldsymbol{x}, \boldsymbol{\varepsilon})|\boldsymbol{x} = \cdot]$. Note that we do not necessarily assume that the function r is additive with respect to $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$. For the special case of additive-noise models, $r(\boldsymbol{x}, \boldsymbol{\varepsilon})$ assumes the form $r(\boldsymbol{x}, \boldsymbol{\varepsilon}) = r_0(\boldsymbol{x}) + \boldsymbol{\varepsilon}$. We also define $\sigma^2(\cdot) := \mathrm{Var}_{P_{\boldsymbol{\varepsilon}}}[r(\cdot, \boldsymbol{\varepsilon})]$, and highlight that we allow for heteroscedastic variance in the regression model.

We model the distribution of the training samples $\{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^N$, where $\boldsymbol{y}_i := \boldsymbol{y}_{\boldsymbol{x}_i} = r(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_{\boldsymbol{x}_i})$ for $1 \le i \le N$, by assuming that the pairs $\{(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_{\boldsymbol{x}_i})\}_{i=1}^N$ are being drawn from an underlying marked Poisson process $\mathcal{P}_{n\check{g}}$ with intensity measure $n(\mathbb{Q} \otimes P_{\boldsymbol{\varepsilon}})$ (see Section 3.3 for definition and additional details). Here, N is a Poisson random variable with mean n, and $\boldsymbol{\varepsilon}_{\boldsymbol{x}}$'s are independent marks associated to each point $\boldsymbol{x}$ in the sample. We refer to this sampling as the Poisson sampling setting. This sample, as a collection of points in the product space $\mathbb{R}^d \times \mathbb{R}$, can be thought of as a mixed binomial point process $\sum_{i=1}^N \delta_{(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_i)}$, where N is a Poisson random variable with mean n and $\{(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_i)\}_{i=1}^N$ are i.i.d. samples from $\mathbb{Q} \otimes P_{\boldsymbol{\varepsilon}}$, independent of the N; here $\delta_{(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_i)}$ is the Dirac measure at $(\boldsymbol{x}_i, \boldsymbol{\varepsilon}_i)$. In other words, we have an infinite i.i.d. sequence $(\boldsymbol{x}_1, \boldsymbol{\varepsilon}_1), (\boldsymbol{x}_2, \boldsymbol{\varepsilon}_2), \ldots$ with $\boldsymbol{x}_i \sim \mathbb{Q}$ and independent noise $\boldsymbol{\varepsilon}_i \sim P_{\boldsymbol{\varepsilon}}$ for $i \in \mathbb{N}$, and the sample size is given by an independent Poisson random variable N with mean n. Furthermore, we denote by $\mathcal{P}_{ng}$ the process obtained by projecting the marked Poisson process $\mathcal{P}_{n\check{g}}$ on $\mathbb{R}^d$ consisting of the Poisson sample $\{\boldsymbol{x}_i\}_{i=1}^N$.

Before we introduce the specific form of the random forest we study in this thesis, we introduce a geometric concept, the so-called k-Potential Nearest Neighbors (k-PNNs), which can be interpreted as a generalization of the classical k-nearest neighbors (k-NNs). The k-PNNs share a close connection with random forest as we explain subsequently.

For any $x_1 = (x_1^{(1)}, \ldots, x_1^{(d)}), x_2 = (x_2^{(1)}, \ldots, x_2^{(d)}) \in \mathbb{R}^d$, we define the hyperrectangle $\mathrm{Rect}(x_1, x_2)$ defined by x_1, x_2, and its volume respectively as

$$\mathrm{Rect}(x_1, x_2) := \prod_{i=1}^d [x_1^{(i)} \wedge x_2^{(i)}, x_1^{(i)} \vee x_2^{(i)}], \qquad \text{and} \qquad |x_1 - x_2| := \prod_{i=1}^d |x_1^{(i)} - x_2^{(i)}|.$$

Definition 3.1 (k-PNN)**.** *Given a target point x_0, and a locally finite point configuration μ in $\mathbb{R}^d$, a point $x \in \mu$ is said to be a k-PNN to x_0 (with respect to μ) if there are fewer than k points from $\mu \setminus \{x\}$ in Rect(x, x_0).*

The number of k-PNNs to a target point x_0 is always larger than or equal to k, provided that the underlying configuration μ has at least k points. Figure 1 illustrates an example of 2-PNNs to a point x_0 in a given configuration. One can also interpret k-PNNs in terms of monotone metrics. A metric d on $\mathbb{R}^d$ is said to be monotone if for any two points x_1, x_2, and any point x in $\mathrm{Rect}(x_1, x_2)$, one has $\mathsf{d}(x, x_1) \le \mathsf{d}(x_1, x_2)$. For instance, the Euclidean distance in $\mathbb{R}^d$ is one such metric. Then, given a collection of points μ, a point $x \in \mu$ is a k-PNN of a target point x_0, if and only if there

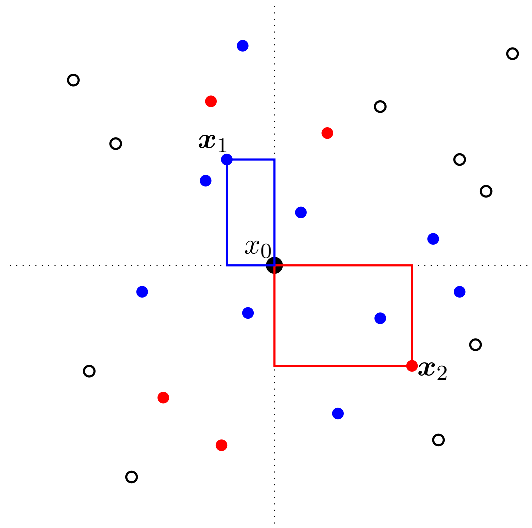

Figure 1: The set of 2-PNNs around a point $x_0 \in \mathbb{R}^2$. The point configuration includes all points in the figure except x_0. The blue and red points *together* are the 2-PNNs to x_0. The red ones such as $\boldsymbol{x}_2$ has exactly 1 point in its corresponding rectangle. The blue ones, such as $\boldsymbol{x}_1$, are also a 1-PNN, or LNN, with no other point in the rectangle formed by x_0 and those points.

exists a monotone metric under which x is among the k closest points in μ from x_0. Obviously, the classical k-NN is a special case of k-PNN with some chosen monotone metric. The case $k = 1$ is a special case, and 1-PNNs are also called layered nearest neighbors (LNNs). It has been observed that nearest neighbor methods with adaptively chosen metrics that are monotone demonstrate good empirical performance (Hastie and Tibshirani, 1995; Domeniconi et al., 2002).

The notion of k-PNN is also intrinsically linked with the notions of "dominance" and "number of maximal/minimal points"; see, e.g., Bai et al. (2005, 2006) and references therein. A point $x_1 = (x_1^{(1)}, \ldots, x_1^{(d)}) \in \mathbb{R}^d$ is said to dominate a point $x_2 = (x_2^{(1)}, \ldots, x_2^{(d)}) \in \mathbb{R}^d$ if $x_1 - x_2 \in \mathbb{R}_+^d \backslash \{0\}$, i.e., $x_1^{(i)} > x_2^{(i)}$ for all $i \in [d]$, represented by the binary relations $x_1 \succ x_2$ or $x_2 \prec x_1$. Furthermore, points in the sample not dominating any other points are called minimal (or Pareto optimal) points of the sample, and points that are not dominated by any other sample points are called maximal. Thus, LNNs to a point $x_0 \in \mathbb{R}^d$ can be thought of as a collection consisting of 2^d independent copies (one copy for each quadrant) of the classical minimal points w.r.t. x_0.

With this background, we now describe the non-bagging and non-adaptive random forest predictors that we analyze here. For a given target point, all its k-PNNs in the training set are also called its *voting points*. The prediction for that target point is then expressed as a (randomly) weighted linear combination of the labels corresponding to the voting points; see, e.g., Lin and Jeon (2006); Biau and Devroye (2010). The non-adaptiveness comes from the fact that both the weights and the randomized splitting scheme used to construct the base decision trees of the random forest do not depend on the labels. Furthermore, as discussed by Lin and Jeon (2006), regardless of the tree generating schemes, as long as the terminal nodes of each randomized tree define rectangular areas, voting points are all k-PNNs to x_0 and all k-PNNs to x_0 can be voting points. Particularly, if $k = 1$, the above procedure is also called as Layered Nearest Neighbor based prediction (Biau and Devroye, 2010; Wager, 2014).

For a given test point $x_0 \in \mathbb{R}^d$, the random forest type estimator studied in this paper is of the form

$$r_{n,k,w}(x_0) := \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}(\check{\lambda})} W_{n\boldsymbol{x}}(x_0)\boldsymbol{y_x}. \tag{3.2}$$

Here the weights $W_{n\boldsymbol{x}}(x_0) \equiv W_{n\boldsymbol{x}}(x_0, \mathcal{P}(\check{\lambda}))$ are nonnegative Borel measurable functions of x_0, of the samples $\{\boldsymbol{x}_i\}_{i=1}^N$ and of the random variables used to generate randomized trees independent of the sample. Note here that we assume the weights to not depend on the marks $\varepsilon_{\boldsymbol{x}}$; see Remark 3.12 for discussions on removing this assumption. The subscript n indicates the dependence of the weights on the given configuration of the Poisson process, which is such that $W_{n\boldsymbol{x}}(x_0) = 0$ if $\boldsymbol{x} \notin \mathcal{L}_{n,k}(x_0)$, where $\mathcal{L}_{n,k}(x_0) \equiv \mathcal{L}_{n,k}(x_0, \mathcal{P}(\check{\lambda}))$ is the set of all k-PNNs to x_0 in the Poisson sample, and

$$\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}(\check{\lambda})} W_{n\boldsymbol{x}}(x_0) = \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}(\check{\lambda})} W_{n\boldsymbol{x}}(x_0)\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)} = 1. \tag{3.3}$$

In other words, one can view the weights $\{W_{n\boldsymbol{x}}(x_0)\}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}$ as a probability mass function of a distribution over all k-PNNs of x_0. If this distribution is uniform, we have the following so-called k-PNN estimator:

$$r_{n,k}(x_0) := \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}(\check{\lambda})} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)}\boldsymbol{y_x}, \tag{3.4}$$

where $L_{n,k}(x_0) \equiv L_{n,k}(x_0, \mathcal{P}(\check{\lambda})) = |\mathcal{L}_{n,k}(x_0, \mathcal{P}(\check{\lambda}))|$. Here by convention, the sum is zero when $\mathcal{P}(\check{\lambda})$ is empty, which by properties of Poisson processes happens with the exponentially small probability $e^{-\check{\lambda}(\mathbb{R}^d\times\mathbb{R})} = e^{-n}$. Unlike k-NNs, the number of k-PNNs is usually larger than k and actually, it is increasing both in k and in n. For instance, Lin and Jeon (2006) shows that if the density g of the distribution $\mathbb{Q}$ is bounded from above and below on $[0,1]^d$, $\mathbb{E}L_{n,k}(x_0)$ is of the order $k\log^{d-1} n$.

3.2 Main results: Rates of multivariate Gaussian approximation for Random Forest

3.2.1 Probability metrics

We now introduce the integral probability metrics that we use in this thesis to quantify the error in Gaussian approximations.

Let $\boldsymbol{z}_1 = (\boldsymbol{z}_1^{(1)}, \ldots, \boldsymbol{z}_1^{(d)})$ and $\boldsymbol{z}_2 = (\boldsymbol{z}_2^{(1)}, \ldots, \boldsymbol{z}_2^{(d)})$ be two d-dimensional random vectors. Denote by $\mathcal{H}_d^{(2)}$ the class of all C^2-functions $h : \mathbb{R}^d \to \mathbb{R}$ such that

$$|h(x_1) - h(x_2)| \le \|x_1 - x_2\|, \quad x_1, x_2 \in \mathbb{R}^d, \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} \|\text{Hess } h(x)\|_{\text{op}} \le 1,$$

where Hess h is the Hessian of h, and let $\mathcal{H}_d^{(3)}$ be the class of all C^3-functions such that the absolute values of the second and third derivatives are bounded by 1. The d_2- and d_3-distances between the laws of $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are given respectively by

$$\mathsf{d}_2(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sup_{h\in\mathcal{H}_d^{(2)}} |\mathbb{E}(h(\boldsymbol{z}_1)) - \mathbb{E}(h(\boldsymbol{z}_2))|,$$
$$\mathsf{d}_3(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sup_{h\in\mathcal{H}_d^{(3)}} |\mathbb{E}(h(\boldsymbol{z}_1)) - \mathbb{E}(h(\boldsymbol{z}_2))|.$$

Here as well as for the metric $\mathsf{d}_{\mathsf{cvx}}$ below, to simplify notation, with a slight abuse of notation we write the distances between the random vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, while they are indeed distances between their laws. The distances d_2 and d_3 are well-defined for random vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ that satisfy $\mathbb{E}(\|\boldsymbol{z}_1\|), \mathbb{E}(\|\boldsymbol{z}_2\|) < \infty$, and $\mathbb{E}(\|\boldsymbol{z}_1\|^2), \mathbb{E}(\|\boldsymbol{z}_2\|^2) < \infty$ respectively. We also use the following non-smooth integral probability metric given by

$$\mathsf{d}_{\mathsf{cvx}}(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sup_{h \in \mathcal{I}} |\mathbb{E}(h(\boldsymbol{z}_1)) - \mathbb{E}(h(\boldsymbol{z}_2))|,$$

where $\mathcal{I}$ is the set of all indicators of measurable convex sets in $\mathbb{R}^d$. If we restrict the set $\mathcal{I}$ to sets of the form $\mathbb{1}_{\Pi_{i=1}^d(-\infty, t_i]}$ for all $t_i \in \mathbb{R}$, $i \geq 1$, where Π here means the Cartesian product, the distance becomes the so-called Kolmogorov distance given by

$$\mathsf{d}_K(\boldsymbol{z}_1, \boldsymbol{z}_2) := \sup_{(t_1,\ldots,t_d)\in\mathbb{R}^d} |\mathbb{P}(\boldsymbol{z}_1^{(1)} \leq t_1, \ldots, \boldsymbol{z}_1^{(d)} \leq t_d) - \mathbb{P}(\boldsymbol{z}_2^{(1)} \leq t_1, \ldots, \boldsymbol{z}_2^{(d)} \leq t_d)|.$$

Note here that we trivially have $\mathsf{d}_K \leq \mathsf{d}_{\mathsf{cvx}}$. Thus, any bound on the $\mathsf{d}_{\mathsf{cvx}}$ distance also holds for the Kolmogorov distance. The above probability metrics are widely used in the literature on quantitative bounds for Gaussian and non-Gaussian approximations.

3.2.2 Gaussian approximation bounds for random forests

We now present our main result providing rates for the multivariate Gaussian approximation of the random forest type estimator given by (3.2) for multiple test points $x_{0,1}, x_{0,2}, \ldots, x_{0,m} \in \mathbb{R}^d$ for some $m \in \mathbb{N}$. For notational convenience, when $m = 1$, we simply refer to $x_{0,1}$ as x_0.

Recall the regression model in (3.1), and the Poisson processes $\mathcal{P}(\lambda)$ and $\mathcal{P}(\check{\lambda})$ in Section 3.1. For $m \in \mathbb{N}$ and $x_{0,i} \in \mathbb{R}^d, i = 1, \ldots, m$, let

$$\mathbf{r}_{n,k,w} := (r_{n,k,w}(x_{0,1}), \ldots, r_{n,k,w}(x_{0,m}))^T \tag{3.5}$$

denote the vector of corresponding random forest predictions as defined in (3.2). Below, we write $\mathcal{P}_{x,\eta} := \mathcal{P}_{n\breve{g}} + \delta_{(x,\boldsymbol{\varepsilon}_x)} + \eta$ for the marked Poisson process $\mathcal{P}_{n\breve{g}}$ with additional point $(x, \boldsymbol{\varepsilon}_x)$ and an additional finite collection of points $\eta \subset \mathbb{R}^d \times \mathbb{R}$.

Theorem 3.1. *Assume there exist $p > 0$ and $\sigma^2 > 0$ such that*

$$\mathbb{E}(|r(\boldsymbol{x}, \boldsymbol{\varepsilon})|^{6+p}) < \infty \qquad \text{and} \qquad \sigma^2 := \inf_x \ \sigma^2(x) > 0.$$

For $m \in \mathbb{N}$ and $x_{0,i} \in \mathbb{R}^d, i = 1, \ldots, m$, let $\mathbf{r}_{n,k,w}$ be as in (3.5), with covariance matrix Σ_m. Then, for $d, n \geq 2$ and $k = \mathcal{O}(n^\alpha)$ for $0 < \alpha < 1$, there exists $c_g > 0$ depending on d, σ^2, g, α and $p > 0$, such that for $\mathbb{Q}^m$-almost all $(x_{0,1}, \ldots, x_{0,m})$,

$$\mathsf{d}\left(\Sigma_m^{-1/2}(\mathbf{r}_{n,k,w} - \mathbb{E}(\mathbf{r}_{n,k,w})), \mathcal{N}\right) \leq c_g m^{43/6} k^{\tau+1} \max_{j \in \{1,4\}} \left\{W(n,k)^{1/2+1/j} \log^{(d-1)/j} n\right\}$$

for $\mathsf{d} \in \{\mathsf{d}_2, \mathsf{d}_3, \mathsf{d}_{\mathsf{cvx}}, \mathsf{d}_K\}$, where,

$$W(n,k) := \frac{\max\limits_{i=1,\ldots,m} \left(\sup_{x,|\eta|\leq 9} \ \|W_{nx}(x_{0,i}, \mathcal{P}_{x,\eta})\|_{L_{6+p}}\right)^2}{\min\limits_{i=1,\ldots,m} \mathbb{E}\left(\sum_{\boldsymbol{x}\in\mathcal{P}(\lambda)} W_{n\boldsymbol{x}}(x_{0,i})^2\right)}, \tag{3.6}$$

and

$$\tau := 6\zeta\beta + 6\beta + 1/2 + \lceil 21(1+\zeta)/(6+p/2) \rceil, \tag{3.7}$$

with

- $\beta := \frac{p}{32+4p}$, *and* $\zeta := \frac{p}{40+10p}$, *for* $\mathsf{d} \in \{\mathsf{d}_2, \mathsf{d}_3\}$, *resulting in* $\tau \in \left(\frac{63}{20}, \frac{9}{2}\right)$.
- $\beta := \frac{p}{72+6p}$, *and* $\zeta := \frac{p}{84+14p}$, *for* $\mathsf{d} \in \{\mathsf{d}_{cvx}, \mathsf{d}_K\}$, *resulting in* $\tau \in \left(\frac{18}{7}, \frac{9}{2}\right)$.

In both cases, τ *is a decreasing function of* p.

The main ingredient of the proof of Theorem 3.1 is a general multivariate Gaussian approximation bound (see Theorems 3.2 and 3.3) for certain classes of functionals of Poisson process. Specifically the classes of functionals are expressed as sums of score functions (as in (3.9)), with the scores themselves satisfying the so-called region-based stabilization property (see Definition 3.3 and Section 3.3.2). Intuitively speaking, bounds on the region of stabilization control the level of dependency in the statistic, thereby enabling the statistic to converge to a Gaussian limit. We show that the random forest statistic satisfies region-based stabilization and leverage Theorems 3.2 and 3.3 to prove Theorem 3.1. We defer the proof of the result to Appendix 5.2.4.

Remark 3.1. *The ceiling function in the exponent* τ *is due to the proof technique required in Lemmas 5.15, 5.17 and 5.18. While we believe the exponent could be improved by removing the ceiling functions, it may require additional tedious arguments in the proof, which we do not pursue for the sake avoiding a more complicated exposition.*

Remark 3.2 (De-localization of weights)**.** *The factor* $W(n,k)$ *is a function of* n *and* k*, and it depends on the distribution of the random weights used to weight the set of* k*-PNNs for a given test point. To have a meaningful normal approximation bound, it is required to decay to zero, as suggested by Theorem 3.1. In effect, this entails that the weight distribution needs to be de-localized. Indeed, even in the case of CLTs for weighted sums of general random variable, the weights need to be sufficiently de-localized to get Gaussian limits.*

Note also that using the Cauchy-Schwarz inequality followed by (3.3) and Jensen's inequality, for the denominator term (considering $m = 1$ *for simplicity), we have that*

$$\mathbb{E}\left(\sum_{\boldsymbol{x}\in\mathcal{P}(\lambda)} W_{n\boldsymbol{x}}(x_0)^2\right) \geq \mathbb{E}\left(\frac{\left(\sum_{\boldsymbol{x}\in\mathcal{P}(\lambda)} W_{n\boldsymbol{x}}(x_0)\right)^2}{L_{n,k}(x_0)}\right) \geq \left(\mathbb{E}[L_{n,k}(x_0)]\right)^{-1} \gtrsim k^{-1}\log^{-(d-1)} n,$$

where the last inequality is due to Lemma 5.19. From the definition of $W(n,k)$*, this implies that a smaller supremum of the weights in the numerator (meaning the weights are more equally distributed) will result in a tighter upper bound, see Corollary 3.1.*

Remark 3.3. *The fractional powers* $1/j$*, for* $j = 1, 4$*, in the bound corresponds to the fractional powers in* $\Gamma_1 - \Gamma_6$ *from Theorem 3.3, which is used to prove Theorem 3.1. These in turn come from the use of the multivariate second order Poincaré inequality, see Theorem 5.2.*

As a corollary to Theorem 3.1, we obtain the rates of convergence for multivariate Gaussian approximation in the case of uniform weights, i.e., for the k-PNN estimator given by (3.4). See Appendix 5.2.4 for its proof.

Corollary 3.1. *Under the setting of Theorem 3.1, let* $\mathbf{r}_{n,k}$ *be as in (3.4) with* $k \geq 11$*. Then, there exists* $c_u > 0$ *depending on* d*,* σ^2*,* α *and* $p > 0$ *such that for* $\mathbb{Q}^m$*-almost all* $(x_{0,1}, \dots, x_{0,m})$*,*

$$\mathsf{d}\left(\Sigma_m^{-1/2}(\mathbf{r}_{n,k} - \mathbb{E}(\mathbf{r}_{n,k})), \mathcal{N}\right) \leq c_u \frac{m^{43/6}\, k^{\tau}}{\log^{(d-1)/2} n}, \qquad \mathsf{d} \in \{\mathsf{d}_2, \mathsf{d}_3, \mathsf{d}_{cvx}, \mathsf{d}_K\}, \tag{3.8}$$

where Σ_m *is the covariance matrix of* $\mathbf{r}_{n,k}$ *and* τ *is as in (3.7). In particular, for* m *fixed, if* $k = o\big(\log^{(d-1)/(2\tau)} n\big)$*, then* $\mathbf{r}_{n,k}$ *is asymptotically normal.*

Remark 3.4. *As a corollary of Theorem 3.1, one could just replace the general weights in* (3.2) *by the uniform weights given in* (3.4) *to compute the quantity* $W(n,k)$ *in* (3.6). *However, simply doing so will result in a slightly worse bound in the power of* k *as Theorem 3.1 is proved without knowing any additional specific information on the weights thus* $W(n,k)$. *We therefore adopted a more refined proof argument for Corollary 3.1 with the uniform weights.*

We now make a few additional remarks pertaining to both Theorem 3.1 and Corollary 3.1.

Remark 3.5 (Moment condition)**.** *The assumption of* $(6+p)$ *moments in Theorem 3.1 is only needed for a multivariate normal approximation result. It can be relaxed to a* $(4+p)$*-moment condition when considering a univariate normal approximation, utilizing results by Bhattacharjee and Molchanov (2022).*

Remark 3.6 (Binomial Point Processes)**.** *Apart from the Poisson setting considered here, the case of i.i.d. sampling (i.e., binomial point processes) is also of interest. According to Bhattacharjee and Molchanov (2022), for the i.i.d. case, a univariate normal approximation result for region-based stabilization can be achieved by adapting the scheme elaborated by Lachièze-Rey et al. (2019, Theorem 4.3) and bounding the required terms similarly as done in the proof of Theorem 3.3. Furthermore, replacing the Poisson cumulative distribution function (c.d.f.) in* (5.40) *by a binomial c.d.f., one can follow the subsequent line of argument to derive a univariate version of rates of normal approximation paralleling our Theorems 3.1 and Corollary 3.1 for the i.i.d. sampling case. Specifically, if we let* $m=1$ *and consider the uniform weights based predictor* $r_{n,k}(x_0)$ *as in* (3.4), *we obtain*

$$\mathsf{d}_K\left(\frac{r_{n,k}(x_0)-\mathbb{E}(r_{n,k}(x_0))}{\sqrt{\mathrm{Var}(r_{n,k}(x_0))}},\mathcal{N}\right)\leq c_u\frac{k^{\tau}}{\log^{(d-1)/2} n},\qquad \textit{for } \mathbb{Q}-\textit{almost all } x_0\in\mathbb{R}^d.$$

According to Schulte and Yukich (2023b), second-order Poincaré inequalities for the multivariate normal approximation of Poisson functionals have no available counterparts for binomial point processes. Thus, there are no immediate versions of multivariate (i.e., for $m>1$*) normal approximations (i.e., analogs of Theorem 3.2 and Theorem 3.3) of region-based stabilizing functionals under i.i.d. samples. This remains an open problem, with applications beyond the scope of this thesis.*

Remark 3.7 (Comparison to MSE rates)**.** *It has been emphasized, for instance, by Lin and Jeon (2006), that as* k *increases the mean squared error (MSE) of the random forest estimator* $r_{n,k,w}$ *is at least of the order* $k^{-1}\log^{-(d-1)} n$. *Hence, by picking* k *appropriately, one can obtain (near) optimal rates of convergence for the MSE in regression problems under various assumptions. This is in contrast to the dependence on* k *in Corollary 3.1; see also the following remarks.*

Remark 3.8 (Dependence on k)**.** *Recall the random forest predictor with uniform weights as in* (3.4). *As mentioned previously, we show later in Section 3.3 that the statistic in* (3.4) *is a sum of certain score functions that satisfies a region-based stabilization property. In particular, as* k *increases, the region of stabilization for each summand becomes large resulting in increased dependency between the scores at different points, deviating much further away from a i.i.d. setup, which has a negative effect for obtaining Gaussian limiting distributions.*

In particular, a main part of our proof of Theorem 3.1 and Corollary 3.1 is related to bounding the cumulative distribution function (in terms of k*) of the Poisson probability that* x *is in the set of* k*-PNNs to* x_0 *denoted by* $\mathcal{L}_{n,k}(x_0)$. *This plays a key role in controlling the size of the region of stabilization and results in the* k^{τ} *term in the numerator of* (3.8). *In Lemma 5.16, we establish a lower bound in* k *matching our upper bound. This is due to the fact that, as* k *increases, for points* x *in a set of substantial measure, the probability that* x *is in the region of stabilization* *given by $\mathbb{P}(\text{Poi}(\varpi) < k)$, where $\varpi = n\int_{Rect(x_0,x)} g(z)dz$, is close to 1, meaning that the region of stabilization is substantially large; see Lemma 5.16 for more details. Hence, the k^τ term cannot be further improved using the current proof technique (i.e., using region-based stabilization and Stein's method). Resolving this question of optimal k dependency, either by demonstrating that the order of k is necessary or by improving the k dependency is thus an important open question.*

Remark 3.9 (Optimality in n)**.** *In terms of n, our bounds are presumably optimal. This can be noted, by considering the response $\boldsymbol{y_x}$ to be integer-valued. Peköz et al. (2013, Lemma 4.1) prove that for integer-valued functionals (3.9), it leads to a lower bound matching the classical Berry-Esseen upper bound mentioned above in terms of n, we then have the claimed optimality result.*

Remark 3.10 (Extension to bagging random forests)**.** *As mentioned in Section **??**, Wager (2014), Mentch and Hooker (2016) and Peng et al. (2022) studied the bagging random forest based on subsampling from the entire training data to construct the base-learners. Particularly, this random forest is in essence expressible in the form of a U-statistic, given by*

$$U_{n,s} = \binom{n}{s}^{-1} \sum_{(n,s)} h(x_{i1}, \ldots, x_{is}; w_i),$$

where $\{x_{i1}, \ldots, x_{is}\}$ are subsamples of size s from n i.i.d. samples $\{x_1, \ldots, x_n\}$ according to some randomness w_i and h is an estimator that is permutation invariant in its arguments. Typically, the estimators h are tree-style base-learners such as k-NN estimators and k-PNN estimators discussed in Section 3.1.

Among the aforementioned works, only Peng et al. (2022) derived Gaussian approximation bounds in the univariate setting, albeit for fixed k. An approach to improve the rate in k in our results is to combine the U-statistics based sub-sampling approach with the region-based stabilization proof technique that we introduce in this thesis. We believe this is an intricate problem and requires further non-trivial efforts.

Remark 3.11 (Generalization to metric-valued data)**.** *Although Theorem 3.1 is stated in the context of the input data taking values in the Euclidean space, the proof techniques and the concept of k-PNNs actually do not rely on the geometry or topology of Euclidean spaces as it only requires a monotone metric. Therefore, the above result could potentially be generalized to other metric spaces of the inputs $\boldsymbol{x}$ by considering k-PNNs under other metrics (Haghiri et al., 2018). Indeed, our main probabilistic results (see Theorem 3.3 and Theorem 3.2) used to prove Theorem 3.1 are derived for general metric spaces.*

Remark 3.12 (Extension to adaptive random forest)**.** *In Theorem 3.1 we consider non-adaptive random forest, i.e., the weights $W_{n\boldsymbol{x}}(x_0)$ in (3.2) are not depending on the response $\boldsymbol{y_x}$. Indeed, the independence between $W_{n\boldsymbol{x}}(x_0)$ and $\boldsymbol{y_x}$ due to non-adaptivity is used in (5.38). However, we would like to highlight that our general result, stated later in Theorem 3.3, can be used to derive multivariate Gaussian approximation of the adaptive random forest. Indeed, it might be possible to use honesty-type assumptions to directly bound the left-hand side of (5.38) resulting in a more complicated analysis. A detailed examination of this is left as future work.*

3.2.3 Towards statistical inference

Note that the bounds in Theorem 3.1 and Corollary 3.1, could be leveraged to obtained rates for constructing non-asymptotically valid confidence intervals for the expected random forest predictions.

For example, as a consequence of Corollary 3.1, for a significance level $0 < \alpha < 1$, we immediately have

$$\mathbb{P}\left(\mathbb{E}r_{n,k}(x_0) \in \left(r_{n,k}(x_0) - \sqrt{\operatorname{Var} r_{n,k}(x_0)}z_{1-\alpha/2}, r_{n,k}(x_0) + \sqrt{\operatorname{Var} r_{n,k}(x_0)}z_{1-\alpha/2}\right)\right)$$
$$\geq 1 - \alpha - c_u k^{\tau} \log^{-(d-1)/2} n,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. However, there are two key issue towards using the above result in practice: (i) we need to estimate the unknown mean $\mathbb{E}(r_{n,k}(x_0))$ and variance $\operatorname{Var}(r_{n,k}(x_0))$ and, (ii) we only get the confidence interval for the expected random forest prediction (i.e., $\mathbb{E}r_{n,k}(x_0)$) and not the true regression function value at x_0 (i.e, $r_0(x_0)$). Replacing the mean parameter by the true regression at x_0 will result in an error of bias. Hence in order to make practical statistical inferences on the true regression function, it is required to do data-driven bias correction, along with variance (and expectation) estimates, which is beyond the scope of this thesis.

In the following, we give a quantitative analysis on the order of bias and show that the bias is not small enough to be negligible compared to the standard deviation. We state a proposition (see Appendix 5.2.6 for its proof) providing a rate of convergence to zero for the bias of the random forest estimator (3.4) with uniform weights under some regularity conditions. Recall that a function $\psi : \mathbb{R}^d \to \mathbb{R}$ is said to be Hölder continuous at $x_0 \in \mathbb{R}^d$ if there exist constants $L_\psi > 0$ and $\gamma_\psi > 0$ such that for $x \in \mathbb{R}^d$,

$$|\psi(x) - \psi(x_0)| \leq L_\psi |x - x_0|^{\gamma_\psi}.$$

Proposition 3.1. *Let the assumptions required for Corollary 3.1 prevail. In addition, assume the density $g(x)$ and the function $r_0(x)$ are Hölder continuous at x_0 with parameters $L_g, \gamma_g > 0$ and $L_1, \gamma_1 > 0$ respectively and assume $y_x = r(x, \varepsilon)$ is uniformly bounded. Then for any $0 < \zeta < 1$ and $x_0 \in \mathbb{R}^d$, the bias satisfies*

$$|\mathbb{E}r_{n,k}(x_0) - r_0(x_0)| \lesssim \left((\log^{-(\gamma_g \wedge \gamma_1)\zeta} n) \vee (k^{-1/4} \log^{-(d-1)/4} n)\right)$$

for n large enough.

The bias of the subsampling random forest has been considered by Wager and Athey (2018, Theorem 3.2). Our result above considers the non-subsampling version associated with k-PNN estimators and allow k to increase with n. We want to emphasize that the bias bound here is consistent with that derived by Biau and Devroye (2010, Lemma 3 and section 5.3), and we assume Hölder continuity to quantify the convergence of the bias to zero as mentioned above.

As we mentioned before, the bias given above turns out to be relatively large compared to the standard deviation. Indeed, according to (5.68) and Lemma 5.19, note that the standard deviation is lower bounded by $(k \log^{d-1} n)^{-1/2}$ such that $(\operatorname{Var} r_{n,k}(x_0))^{-1/2} |\mathbb{E}r_{n,k}(x_0) - r_0(x_0)| \to \infty$. A bias reduction technique is hence of great importance for inferences on the true regression function r_0, for instance, deriving consistent confidence intervals. While some preliminary work is undertaken by Mentch and Hooker (2016) with bootstrap in the context of bagging random forests, it remains an open problem how to reduce the bias to get non-asymptotically valid confidence intervals. We leave a detailed methodological study of this, including performing large scale simulations, for future work.

3.3 Region-based stabilizing functionals and Gaussian approximation

In this section, we introduce some preliminaries about point processes, region-based stabilizing functionals, and some related notation. We refer to works by Schulte and Yukich (2023b) and Bhattacharjee and Molchanov (2022) for additional details and background. As mentioned earlier, the general multivariate Gaussian approximation results (i.e., Theorems 3.2 and 3.3) established in this section form the backbone for establishing our main result in Theorem 3.1 for random forest.

3.3.1 Functionals of point processes

Let $(\mathbb{X}, \mathcal{F})$ be a measure space with a σ-finite measure $\mathbb{Q}$. We will generally consider marked Poisson processes (see recent work of Schulte and Yukich (2023b) for more details) where the points live in $\mathbb{X}$ while their marks live in a probability space $(\mathbb{M}, \mathcal{F}_{\mathbb{M}}, \mathbb{Q}_{\mathbb{M}})$. Let $\check{\mathbb{X}} := \mathbb{X} \times \mathbb{M}$ with $\check{\mathcal{F}}$ as the product σ-field of $\mathcal{F}$ and $\mathcal{F}_{\mathbb{M}}$ and $\check{\mathbb{Q}} := \mathbb{Q} \otimes \mathbb{Q}_{\mathbb{M}}$ is the product measure. When $(\mathbb{M}, \mathcal{F}_{\mathbb{M}}, \mathbb{Q}_{\mathbb{M}})$ is a singleton endowed with a Dirac point mass, then the measure $\check{\mathbb{Q}}$ reduces to $\mathbb{Q}$. For $\check{x} \in \check{\mathbb{X}}$, we shall use the representation $\check{x} := (x, m_x)$ with $x \in \mathbb{X}$ and $m_x \in \mathbb{M}$. Let $\mathbf{N}$ be the set of σ-finite counting measures on $(\check{\mathbb{X}}, \check{\mathcal{F}})$, which can be interpreted as point configurations in $\check{\mathbb{X}}$. The set $\mathbf{N}$ is equipped with the smallest σ-field $\mathscr{N}$ such that the maps $m_A : \mathbf{N} \to \mathbb{N} \cup \{0, \infty\}, \mathcal{M} \mapsto \mathcal{M}(A)$ are measurable for all $A \in \check{\mathcal{F}}$. A point process is a random element in $\mathbf{N}$. For $\eta \in \mathbf{N}$, we write $\check{x} \in \eta$ if $\eta(\{\check{x}\}) \geq 1$. Furthermore, denote by η_A the restriction of η onto the set $A \in \check{\mathcal{F}}$. For $\eta_1, \eta_2 \in \mathbf{N}$, we write $\eta_1 \leq \eta_2$ if the difference $\eta_1 - \eta_2$ is non-negative. Denote by $\mathcal{P}(\lambda)$ and $\mathcal{P}(\check{\lambda})$ the Poisson processes with intensity measures $\lambda := n\mathbb{Q}$ (resp. $\check{\lambda} := n\check{\mathbb{Q}}$) on $(\mathbb{X}, \mathcal{F})$ (resp. $(\check{\mathbb{X}}, \check{\mathcal{F}})$).

To proceed, we need additional definitions and notation. Denote by $\mathbf{F}(\mathbf{N})$ the class of all measurable functions $f : \mathbf{N} \to \mathbb{R}$, and by $L^0(\check{\mathbb{X}}) = L^0(\check{\mathbb{X}}, \check{\mathcal{F}})$ the class of all real-valued, measurable functions F on $\check{\mathbb{X}}$. Note that, as $\check{\mathcal{F}}$ is the completion of $\sigma(\eta)$, each $F \in L^0(\check{\mathbb{X}})$ can be written as $F = f(\eta)$ for some measurable function $f \in \mathbf{F}(\mathbf{N})$. Such a mapping f, called a *representative* of F, is $\check{\mathbb{Q}} \circ \eta^{-1}$-a.s. uniquely defined. In order to simplify the presentation, we make this convention: whenever a general function F is introduced, we will select one of its representatives and denote such a representative mapping by the same symbol F. We denote by $L^2_{\mathcal{P}(\check{\lambda})}(\check{\mathbb{X}}) = L^2_{\mathcal{P}(\check{\lambda})}(\check{\mathbb{X}}, \check{\mathcal{F}})$ the space of all square-integrable functions F of the Poisson process $\mathcal{P}(\check{\lambda})$ with $\mathbb{E}(F^2) < \infty$.

For $n, m \in \mathbb{N}$, and $i \in [m]$, consider a collection of real-valued $\check{\mathcal{F}} \otimes \mathscr{N}$-measurable score functions $\xi_n^{(i)}(\cdot, \cdot)$ defined on each pair $(\check{x}, \eta)$, where $\check{x} \in \eta$ and $\eta \in \mathbf{N}$. We are interested in the following functionals of the Poisson process $\mathcal{P}(\check{\lambda})$:

$$F_n^{(i)} = F_n^{(i)}(\mathcal{P}(\check{\lambda})) := \sum_{\check{x} \in \mathcal{P}(\check{\lambda})} \xi_n^{(i)}((x, m_x), \mathcal{P}(\check{\lambda})). \tag{3.9}$$

We define $\bar{F}_n^{(i)} := F_n^{(i)} - \mathbb{E}[F_n^{(i)}]$ and seek to have a result on rates of multivariate normal approximation for the m-vector $\mathbf{F}_n = (F_n^{(1)}, \dots, F_n^{(m)})$ or $\bar{\mathbf{F}}_n = (\bar{F}_n^{(1)}, \dots, \bar{F}_n^{(m)})$ with a appropriate normalizer and $m \geq 1$.

Definition 3.2 (Cost/Difference Operators)**.** *Let F be a measurable function on $\mathbf{N}$. The family of add-one cost operators, $D = (D_{\check{x}})_{\check{x} \in \check{\mathbb{X}}}$, are defined as*

$$D_{\check{x}} F(\eta) := F(\eta + \delta_{\check{x}}) - F(\eta), \quad \check{x} \in \check{\mathbb{X}}, \eta \in \mathbf{N}.$$

Similarly, we can define a second-order cost operator (also called iterated add-one cost operator): for any $\check{x}_1, \check{x}_2 \in \check{\mathbb{X}}$ and $\eta \in \mathbf{N}$,

$$D^2_{\check{x}_1, \check{x}_2} F(\eta) := F(\eta + \delta_{\check{x}_1} + \delta_{\check{x}_2}) - F(\eta + \delta_{\check{x}_1}) - F(\eta + \delta_{\check{x}_2}) + F(\eta).$$

We say that F belongs to the domain of the difference operator $F \in \mathbf{dom}\ D$ if $F \in L^2_{\mathcal{P}(\check{\lambda})}(\check{\mathbb{X}})$ and

$$\int_{\check{\mathbb{X}}} \mathbb{E}((D_{\check{x}} F)^2) \check{\lambda}(d\check{x}) < \infty.$$

Definition 3.3 (Region of Stabilization)**.** *For $n, m \in \mathbb{N}$ we consider the class of $\check{\mathcal{F}} \otimes \mathscr{N}$-measurable score functions $\xi_n^{(i)}(\check{x}, \eta)$ for $i \in [m]$. Throughout the paper, we will always assume that if $\xi_n^{(i)}(\check{x}, \eta_1) = \xi_n^{(i)}(\check{x}, \eta_2)$ for some $\eta_1, \eta_2 \in \boldsymbol{N}$ with $0 \neq \eta_1 \leq \eta_2$ then*

$$\xi_n^{(i)}(\check{x}, \mu_1) = \xi_n^{(i)}(\check{x}, \eta') \quad \textit{for all } \eta' \in \boldsymbol{N} \textit{ with } \eta_1 \leq \eta' \leq \eta_2. \tag{3.10}$$

This is a form of monotonicity property that is natural to any reasonable choice of score functions.

We now introduce some additional assumptions on the score functions that are sufficient to derive our Gaussian approximation results. Specifically, we assume that for each $i \in [m]$, the score functions $\xi_n^{(i)}(\check{x}, \eta)$ are region-stabilizing (Bhattacharjee and Molchanov, 2022), i.e., for all $n \geq 1$,

(**R1**) there exists a map $R_n^{(i)}$ from $\{(\check{x}, \eta) \in \check{\mathbb{X}} \times \mathbf{N} : \check{x} \in \eta\}$ to $\check{\mathcal{F}}$ such that for all $\eta \in \mathbf{N}$ and $\check{x} \in \eta$, we have that

$$\xi_n(\check{x}, \eta) = \xi_n(\check{x}, \eta_{R_n^{(i)}(\check{x}, \eta)}); \tag{3.11}$$

(**R2**) the set

$$\{(\check{x}, \check{y}_1, \check{y}_2, \eta) : \{\check{y}_1, \check{y}_2\} \subseteq R_n^{(i)}(\check{x}, \eta + \delta_{\check{x}})\}$$

is measurable with respect to the σ-field on $\check{\mathbb{X}}^3 \times \mathbf{N}$;

(**R3**) the map $R_n^{(i)}$ is monotonically decreasing in the second argument:

$$R_n^{(i)}(\check{x}, \eta_1) \supseteq R_n^{(i)}(\check{x}, \eta_2), \quad \eta_1 \leq \eta_2,\ \check{x} \in \eta_1;$$

(**R4**) for all $\eta \in \mathbf{N}$ and $\check{x} \in \eta$, we have that

$$\eta_{R_n^{(i)}(\check{x}, \eta)} \neq 0 \implies (\eta + \delta_{\check{y}})_{R_n^{(i)}(\check{x}, \eta + \delta_{\check{y}})} \neq 0, \text{ for all } \check{y} \notin R_n^{(i)}(\check{x}, \eta).$$

Before moving on with our further assumptions, we note here that the notion of region-stabilization is a generalization of the idea of stabilization radius. In particular, while classically it is assumed that a stabilizing score function at a point is determined by the configuration inside a ball around the point, our Assumption (**R1**) only requires a local region R_n, which is not necessarily a ball, on which the score function ξ_n can be determined. Thus, the dependency between the score functions at different points could be measured only by the size of regions around those points alone, which leads to a Gaussian limit when the regions are small enough. An example where classical stabilization works well is the k-NN distance based for entropy estimation (Berrett et al., 2019; Shi et al., 2023+), where the ball formed by the point and its k-th nearest neighbor determines the k-NNs.

On the other hand, if we consider the k-PNNs (see Definition 3.1), it turns out that considering balls is vastly suboptimal, and one needs to consider general regions to prove Gaussian convergence with presumably optimal rates. Our Assumption (**R3**) is a geometric condition that roughly says that if we add more points to our configuration, the stabilization region $R_n^{(i)}$ can only get smaller, i.e., one would need to explore the configuration in a smaller region to determine the value of the score function. Such a property is very natural and is satisfied for most stabilizing functionals. The Assumptions **(R2)** and **(R4)** are rather technical ones, in particular, as noted by Bhattacharjee and Molchanov (2022, Section 2), Assumption **(R2)** ensures that

$$\{\eta \in \mathbf{N} : \check{y} \in R_n(\check{x}, \eta + \delta_{\check{x}})\} \in \mathscr{N}$$

for all $(\check{x}, \check{y}) \in \check{\mathbb{X}}^2$, and that

$$\mathbb{P}(\check{y} \in R_n(\check{x}, \eta + \delta_{\check{x}})) \quad \text{and} \quad \mathbb{P}(\{\check{y}_1, \check{y}_2\} \in R_n(\check{x}, \eta + \delta_{\check{x}}))$$

are measurable functions of $(\check{x}, \check{y}) \in \check{\mathbb{X}}^2$ and $(\check{x}, \check{y}_1, \check{y}_2) \in \check{\mathbb{X}}^3$ respectively.

3.3.2 Connection to random forest

We now connect the terminology above with the random forest notation introduced in Section 3.1. We have $\mathbb{X} = \mathbb{R}^d$ and the measure $\mathbb{Q}$ is taken to be a probability with an a.e. continuous density g with respect to the Lebesgue measure λ_d on $\mathbb{R}^d$. Hence, the intensity measure $\lambda = ng$. Moreover, the mark space $\mathbb{M}$ which represents in this case the domain of the noise $\boldsymbol{\varepsilon}$, is taken to be $\mathbb{R}$ with $\mathbb{Q}_{\mathbb{M}}$ being its distribution $P_{\boldsymbol{\varepsilon}}$. Correspondingly the marked version of the intensity measure is $\check{\lambda} = n\check{g}$. We now let $\mathcal{P}_{n\check{g}}$ and $\mathcal{P}_{ng}$ denote the canonical Poisson process on $\check{\mathbb{X}}$ (resp. $\mathbb{X}$) with intensity measure $n\check{\mathbb{Q}}$ (resp. $n\mathbb{Q}$) for $n \geq 1$. The random forest predictor in (3.2), as well as the one in (3.4) with uniform weights, are given respectively by

$$r_{n,k,w}(x_0) := \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon}_{\boldsymbol{x}}) \in \mathcal{P}_{n\check{g}}} W_{n\boldsymbol{x}}(x_0) \mathbb{1}_{\boldsymbol{x} \in \mathcal{L}_{n,k}(x_0)} \boldsymbol{y}_{\boldsymbol{x}}, \qquad \text{and}$$

$$r_{n,k}(x_0) := \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon}_{\boldsymbol{x}}) \in \mathcal{P}_{n\check{g}}} \frac{\mathbb{1}_{\boldsymbol{x} \in \mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)} \boldsymbol{y}_{\boldsymbol{x}}.$$

Now, observe that for $r_{n,k,w}(x_0)$, is a region-based stabilizing functional with the region of stabilization given by

$$R_n(\check{x}, \mathcal{P}_{n\check{g}}) := \begin{cases} \text{Rect}(x_0, x) \times \mathbb{R}, & \text{if } \mathcal{P}_{n\check{g}}((\text{Rect}(x_0, x) \backslash \{x\}) \times \mathbb{R}) < k, \\ \varnothing, & \text{otherwise.} \end{cases}$$

This region is similar to the one considered by Bhattacharjee and Molchanov (2022, Theorem 2.2) in the context of minimal points, and is indeed a generalized version exploiting the connection between k-PNNs and minimal points. In particular, for most k-PNNs, this region is thin in some directions and long in the other directions, which makes it suboptimal for it to be enclose by a ball. Consequently, standard results on multivariate Gaussian approximation (e.g., Schulte and Yukich, 2023b) are not immediately applicable in this example due to the fact that they require a ball with a small radius as the region of stabilization.

3.3.3 Tail condition

Going back to our general model, it is clear that with a general stabilization region as ours, we need some control on its size, so that the score functions are only locally dependent facilitating a Gaussian limit. This motivates the following assumption. Below, for $x \in \mathbb{X}$, we write $\boldsymbol{m}_x$ to denote the random mark associated to x independent of all else.

(**T**) For each $i \in [m]$, assume that there exists a measurable function $r_n^{(i)} : \mathbb{X} \times \mathbb{X} \to [0, \infty]$ such that

$$\mathbb{P}((y, \boldsymbol{m}_y) \in R_n^{(i)}((x, \boldsymbol{m}_x), \mathcal{P}(\check{\lambda}) + \delta_{(x,\boldsymbol{m}_x)}) \le e^{-r_n^{(i)}(x,y)}, \quad x, y \in \mathbb{X} \text{ a.e.} \tag{3.12}$$

When $r_n^{(i)}$ does not vanish, Assumption (**T**) is an analog of the usual exponential stabilization condition by Schulte and Yukich (2023b). Note that $r_n^{(i)}$ is allowed to be infinity and the probability (3.12) is well-defined due to Assumption (**R2**).

3.3.4 Moment condition

(**M**) For some $p_0 > 0$, there exists $p > 0$ such that for all $i \in [m]$ and $\eta \in \mathbf{N}$ with $\eta(\check{\mathbb{X}}) \le 3 + p_0$,

$$\|\xi_n^{(i)}((x, \boldsymbol{m}_x), \mathcal{P}(\check{\lambda}) + \delta_{(x,\boldsymbol{m}_x)} + \eta)\|_{L_{p_0+p}} \le M_{n,p_0,p}^{(i)}(x), \quad n \ge 1, \ x \in \mathbb{X} \text{ a.e.}, \tag{3.13}$$

where $M_{n,p_0,p}^{(i)} : \mathbb{X} \to \mathbb{R}, \ n, m \ge 1, \ i \in [m]$ are measurable functions. When $M_{n,p_0,p}^{(i)}(x)$ is a constant not depending on x, it recovers the standard case by Schulte and Yukich (2023b) with uniformly bounded moments. For brevity of notation, in the sequel we will always write $M_n^{(i)}$ instead of $M_{n,p_0,p}^{(i)}$, and generally drop p_0, p from all subscripts.

3.3.5 Gaussian approximation For Region-based Stabilization

We will require a few additional quantities to present our main results. For $i \in [m]$, let

$$q_n^{(i)}(x_1, x_2) := n \int_{\check{\mathbb{X}}} \mathbb{P}(\{(x_1, \boldsymbol{m}_{x_1}), (x_2, \boldsymbol{m}_{x_2})\} \subseteq R_n^{(i)}(\check{z}, \mathcal{P}(\check{\lambda}) + \delta_{\check{z}}))\check{\mathbb{Q}}(d\check{z}). \tag{3.14}$$

For $\zeta > 0$, $y \in \mathbb{X}$ and $i \in [m]$, let

$$g_n^{(i)}(y) := n \int_{\mathbb{X}} e^{-\zeta r_n^{(i)}(x,y)} \mathbb{Q}(dx), \quad h_n^{(i)}(y) := n \int_{\mathbb{X}} M_n^{(i)}(x)^{p_0+p/2} e^{-\zeta r_n^{(i)}(x,y)} \mathbb{Q}(dx), \tag{3.15}$$

and

$$G_n^{(i)}(y) := M_n^{(i)}(y) + h_n^{(i)}(y)^{1/(p_0+p/2)}(1 + g_n(y)^{p_0})^{1/(p_0+p/2)}. \tag{3.16}$$

For $\alpha > 0$, $i, j, l, t \in [m]$, and $\alpha_i, \alpha_j, \alpha_l \ge 0$, define for $y \in \mathbb{X}$,

$$\begin{aligned} f_{\alpha_i,\alpha_j,\alpha_l,\alpha}^{(i,j,l,t)}(y) := {} & n \int_{\mathbb{X}} G_n^{(i)}(x)^{\alpha_i} G_n^{(j)}(x)^{\alpha_j} G_n^{(l)}(x)^{\alpha_l} e^{-\alpha r_n^{(t)}(x,y)} \mathbb{Q}(dx) \\ & + n \int_{\mathbb{X}} G_n^{(i)}(x)^{\alpha_i} G_n^{(j)}(x)^{\alpha_j} G_n^{(l)}(x)^{\alpha_l} e^{-\alpha r_n^{(t)}(y,x)} \mathbb{Q}(dx) \\ & + n \int_{\mathbb{X}} G_n^{(i)}(x)^{\alpha_i} G_n^{(j)}(x)^{\alpha_j} G_n^{(l)}(x)^{\alpha_l} q_n^{(t)}(x,y)^{\alpha} \mathbb{Q}(dx). \end{aligned} \tag{3.17}$$

Moreover, we define for $x \in \mathbb{X}$ and $i \in [m]$,

$$\kappa_n^{(i)}(x) := \mathbb{P}(\xi_n^{(i)}((x, \boldsymbol{m}_x), \mathcal{P}(\check{\lambda}) + \delta_{(x,\boldsymbol{m}_x)}) \ne 0). \tag{3.18}$$

The above quantities are essential to our multivariate Gaussian approximation of region-based stabilizing functionals, where $q_n^{(i)}(x_1,x_2), g_n^{(i)}(y)$ and $\kappa_n^{(i)}(x)$ correspond to the tail probability condition **(T)**, i.e., the "size" of the region of stabilization (see Lemma 5.10), and $h_n^{(i)}(x), G_n^{(i)}(y)$ are associated with the moment condition **(M)**.

For $i \in [m]$, assume $F_n^{(i)} \in \mathbf{dom}\ D$ defined in Section 3.2. We also define

$$\mathrm{P}_n^{-1} := \mathrm{diag}(1/\varrho_n^{(1)}, \dots, 1/\varrho_n^{(m)}),$$

as the normalizer for $\bar{\mathbf{F}}_n$. Let $\Sigma := (\sigma_{ij})_{i,j=1}^m \in \mathbb{R}^{m\times m}$ be any given positive definite matrix and recall that $\mathcal{N}_\Sigma$ be the m-dimensional normal random vector with mean $\mathbf{0}$ and covariance matrix Σ. Define

$$\Gamma_0 := \sum_{i,j=1}^m |\sigma_{ij} - \mathrm{Cov}((\bar{F}_n^{(i)}/\varrho_n^{(i)}), (\bar{F}_n^{(j)}/\varrho_n^{(j)})|, \tag{3.19}$$

$$\Gamma_1 := \left(\sum_{i,j=1}^m \frac{n\mathbb{Q}\big(f_{1,1,0,\beta}^{(i,j,i,i)}\big)^2}{(\varrho_n^{(i)}\varrho_n^{(j)})^2} \right)^{\frac{1}{2}} \tag{3.20}$$

$$\Gamma_2 := \sum_{i=1}^m \frac{n\mathbb{Q}\big((\kappa_n^{(i)} + g_n^{(i)})^{3\beta}(G_n^{(i)})^3\big)}{(\varrho_n^{(i)})^3}. \tag{3.21}$$

The following two theorems provide rates for the multivariate normal approximation of region-based stabilizing functionals measured by d_2-, d_3- and $\mathsf{d}_{\mathsf{cvx}}$ distances defined in Section 3.2.1. These results are generalizations of their univariate versions proved by Bhattacharjee and Molchanov (2022). See Appendix 5.2.1 for the proofs.

Theorem 3.2 (Multivariate Normal Approximation in d_2- and d_3-distances)**.** *For $i \in [m]$, suppose the functional $F_n^{(i)} \in$ **dom** D assumes the form (3.9) with the score function $\xi_n^{(i)}$ satisfying Assumptions **(R1)**-**(R4)**, **(T)** and **(M)** for $p_0 = 4$ and $p > 0$. Let $\zeta := p/(40+10p)$ in (3.15) and $\beta := p/(32+4p)$ in (3.19)-(3.21). Then for a positive definite matrix Σ as above, we have*

(a) for all $n \geq 1$, there exists a constant $c_3 > 0$ depending only on p, such that

$$\mathsf{d}_3\left(\mathrm{P}_n^{-1}\bar{\mathbf{F}}_n, \mathcal{N}_\Sigma\right) \leq c_3 m\bigg(\Gamma_0 + \Gamma_1 + m\Gamma_2\bigg),$$

and

(b) for all $n \geq 1$, there exists a constant $c_2 > 0$ depending only on p, such that

$$\mathsf{d}_2\left(\mathrm{P}_n^{-1}\bar{\mathbf{F}}_n, \mathcal{N}_\Sigma\right) \leq c_2\bigg(\|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}^{\frac{1}{2}}\Gamma_0 + \|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}^{\frac{1}{2}}\Gamma_1 + m^2\|\Sigma^{-1}\|_{op}^{\frac{3}{2}}\|\Sigma\|_{op}\Gamma_2\bigg).$$

In order to state our next theorem for the $\mathsf{d}_{\mathsf{cvx}}$ distance, we introduce the following additional terms. Define

$$\Gamma_3 := \sum_{i=1}^m \frac{n\mathbb{Q}\big((\kappa_n^{(i)} + g_n^{(i)})^{6\beta}(G_n^{(i)})^3\big)}{(\varrho_n^{(i)})^3}, \tag{3.22}$$

$$\Gamma_4 := \left(m \sum_{i=1}^{m} \frac{n\mathbb{Q}\big((\kappa_n^{(i)} + g_n^{(i)})^{6\beta}(G_n^{(i)})^4\big)}{(\varrho_n^{(i)})^4} \right)^{\frac{1}{2}} + \left(\sum_{i,j=1}^{m} \frac{n\mathbb{Q}\big(f_{2,2,0,3\beta}^{(i,j,i,i)}\big)}{(\varrho_n^{(i)}\varrho_n^{(j)})^2} \right)^{\frac{1}{2}}, \tag{3.23}$$

$$\Gamma_5 := \sqrt{m}\Big(\sum_{i,j,l,t=1}^{m} \sum_{s=1}^{m} \frac{n\mathbb{Q}\big(f_{1,1,1/2,\beta}^{(i,j,l,t)}\big)^2}{\varrho_n^{(s)}(\varrho_n^{(i)}\varrho_n^{(j)})^2} \Big)^{\frac{1}{3}}, \tag{3.24}$$

$$\Gamma_6 := m^{3/4}\Big(\sum_{i,j,l,t=1}^{m} \sum_{s=1}^{m} \frac{n\mathbb{Q}\left(f_{1,1,1,\beta}^{(i,j,l,t)}\right)^2}{(\varrho_n^{(s)}\varrho_n^{(i)}\varrho_n^{(j)})^2} \Big)^{\frac{1}{4}}. \tag{3.25}$$

Theorem 3.3 (Multivariate Normal Approximation in $\mathsf{d}_{\mathsf{cvx}}$-distance)**.** *For $i \in [m]$, suppose the functional $F_n^{(i)} \in$ **dom** D assumes the form (3.9) with the score function $\xi_n^{(i)}$ satisfying Assumptions **(R1)**-**(R4)**, **(T)** and **(M)** with $p_0 = 6$. Let $\zeta := p/(84 + 14p)$ in (3.15) and $\beta := p/(72 + 6p)$ in (3.19), (3.20) and (3.22)-(3.25). Then, for any positive definite matrix $\Sigma := (\sigma_{ij})_{i,j=1}^m \in \mathbb{R}^{m\times m}$, we have*

$$\mathsf{d}_{cvx}\left(\mathrm{P}_n^{-1}\bar{\mathbf{F}}_n, \mathcal{N}_\Sigma\right) \le c_{cvx} m^5 \left(\|\Sigma^{-1/2}\|_{op} \vee \|\Sigma^{-1/2}\|_{op}^3 \right) \cdot (\Gamma_0 \vee \Gamma_1 \vee \Gamma_3 \vee \ldots \vee \Gamma_6),$$

for all $n \ge 1$, where the constant c_{cvx} depends only on p and m.

Remark 3.13. *Theorems 3.2 and 3.3 admittedly involve several complicated quantities which may seem hard to interpret. The backbone of these results is a multivariate second-order Poincaré inequality (see Theorem 5.2) proved by Schulte and Yukich (2019, Theorems 1.1 and 1.2). Just as the classical Poincaré inequality provides concentration bounds for Poisson functionals in terms of their first order difference (Wu, 2000), a second order Poincaré inequality provides a central limit theorem with non-asymptotic bounds on various distances between a Poisson functional and a Gaussian random variable/vector in terms of certain moments of the first and second order differences. In many examples, one needs to make some additional assumptions on the functionals to be able to optimally bound these moments. One such simplification is the assumption of stabilization, or as in our case, region-stabilization. All the quantities in our bounds, except Γ_0, are essentially upper bounds on these quantities involving various moments of the differences, under the additional assumption of region-stabilization. On the other hand, the term Γ_0 simply measures the error in approximation incurred due to replacing the sample covariance matrix by Σ.*

We would like to highlight when the region of stabilization $R_n^{(i)}$ is taken to be a ball with its radius having an exponentially decaying tail, and the bound $M_n^{(i)}(x)$ in our moment condition **(M)** is constant independent of x, Theorem 3.3 simplifies to the following bound proved by Schulte and Yukich (2023b):

$$\mathsf{d}_{\mathsf{cvx}}\left(n^{-1/2}\bar{\mathbf{F}}_n, \mathcal{N}_\Sigma\right) \le c_{\mathsf{cvx}} n^{-1/d}, \tag{3.26}$$

where Σ is taken to be the limiting covariance matrix. The relative complexity of our bounds compared to a result such as in (3.26) is mainly because of the fact that our setting and assumptions are more general. Specifically, compared to the work by Schulte and Yukich (2023b) including: (a) we assume general regions of stabilization instead of a ball, which is essential in our main applications Theorem 3.1 and Corollary 3.1; (b) in our moment condition, a non-uniform bound is assumed, which is often necessary to obtain optimal rates. It should be noted that in many statistical problems, a ball as region of stabilization and the uniformly bounded moments suffice to obtain presumably optimal rates of convergence, see Lachièze-Rey et al. (2019); Shi et al. (2023+); Schulte and Yukich

(2023b) for several such statistical applications. Despite their seemingly complicated forms, our Theorems 3.2 and 3.3 are often required to obtain optimal rates, or even to prove Gaussian limits in some case, with the random forest results being a concrete example.

4 Discussion

Quantitative normal approximation bounds are important to obtain finite-sample, non-asymptotic inferential guarantees for various statistical problems. A central problem in normal approximation theory is to go beyond independence to non-independent assumptions. Motivated by the classical Berry-Esseen theorem, which is applicable mainly to a sum of independent score functions $F_n(\mathcal{X}_n) = \sum_{i=1}^n f_n(X_i)$ based on i.i.d. data $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, one question of interest in this thesis is to allow non-independent structures among scores functions, i.e., $F_n(\mathcal{X}_n) = \sum_{i=1}^n f_n(X_i; \mathcal{X}_n)$. By this extension, many geometric and topological statistics including those related to k-nearest neighbor graphs, the minimal spanning trees and the Euler characteristics could be handled. Furthermore, there are also statistics that are not (or hard to be) expressed into sums of score functions. Hence, extending Berry-Esseen bounds to a general class of statistics is of great importance.

Towards this goal, a flexible concept of *stabilization* was introduced here. Our proposed concept, generalizes related notions of stabilization proposed in Lachièze-Rey et al. (2019) and references therein. It consists of many variants: weak stabilization, score-based stabilization and strong stabilization. While they all focus on many statistics of interest having mostly local dependence, among them, only strong stabilization can quantify the local dependence of the statistics F_n while not necessarily assuming F_n to be expressed as sum of score functions. Based on this and Stein's method, we then derive presumably optimal rates of normal approximation for strong stabilizing statistics in terms of recovering the Berry-Esseen bounds. This improves the recent sub-optimal rates in Lachièze-Rey et al. (2022).

Concretely, we provided Gaussian approximation bounds for a class of statistics that satisfy certain stabilization properties. Examples of such statistics include k-nearest neighbor based Entropy estimators, Euler characteristics, minimal spanning trees and Random forests. The proof technique is based on our observation that such statistics could be expressed as functions which subsequently satisfy a certain stabilization property. Based on this observation, the Gaussian approximation bounds are derived by using Malliavin-Stein's method. Additionally, our bounds are also applied to many other statistics of certain geometric local properties, for instance, average treatment effect (ATE) estimators, wavelet estimators, etc.

Furthermore, as a very important example which falls in our stabilization framework, random forest is an extremely successful general-purpose prediction method. Despite their widespread usage, theoretical understanding of their statistical properties has been rather slow. Among many variants of random forest, the most theoretically well-studied version of random forest is the non-bagging and non-adaptive random forest. We refer to Biau and Devroye (2010) for an exposition of this kind. One of the key questions that has remained open regarding non-bagging and non-adaptive random forest is its (multivariate) normal approximation, with applications to statistical inference.

Motivated by the link of random forest with k-potential nearest neighbors in Lin and Jeon (2006), which is an adaptive version of classical k-nearest neighbors, in this thesis, we discover that random forest falls in the category of the region-based stabilization. Region-based stabilization allows general regions to capture the aforementioned local dependence instead of only ball-shaped regions. Based on this and Stein's method, we derived rates of multivariate normal approximation of random forest while allowing k to grow with n, which is the first in the literature. Furthermore, a smoothed Bootstrap procedure has been discussed for possible statistical inferences such as confidence intervals.

Specially, for random forest application, we conclude our thesis with the following potential future direction. First note that combining our multivariate result with standard tightness arguments will entail that trained non-adaptive and non-bagging random forests are Gaussian processes (with a particular covariance function) in the limit. As discussed by Athey et al. (2019), honesty-type conditions (which essentially make random forest to be non-adaptive for all statistical analysis purpose) appear to be necessary to have Gaussian limits. It is interesting to explore limit theorems and distributional approximation bounds (both at multivariate and process level) for adaptive random forests, so as to reveal the advantages of adaptivity, when considering true regression functions to be coming from a more general function class, and also make further advances towards developing rigorous non-asymptotically valid statistical inference for random forest predictions. Finally, bootstrap procedures such as smoothed bootstrap can be considered for constructing practically confidence intervals.

Finally, we want to highlight that the approach of using region-based stabilization theory and Stein's method to establish Gaussian approximation results, as stated in Theorems 3.2 and 3.3, is much more widely applicable for other non-parametric regression problems such as Nadaraya-Watson and wavelets-type in which case, one would need to work with appropriate regions (depending on the procedure) and then apply our general results in Theorems 3.2 and 3.3. This reveals a very important tool to handle the dedicated dependence structure in modern statistics: Controlling dependence within a small region and moment conditions lead to rates of normal approximation.

5 Proof

5.1 Proof of Section 2

Proof of Theorem 2.1. The idea is to use the above theorem, and to bound the first and second order cost functions appearing in the quantities $\gamma_i, i = 1, \dots, 6$ by using our flexible approach. To this end, we rewrite the first and second order cost functions as:

$$D_x F = (D_x F - D_x F(A_x)) + D_x F(A_x),$$

and

$$D_{x_1,x_3} F = (D_{x_3} F^{x_1} - D_{x_3} F^{x_1}(A_{x_3})) + (D_{x_3} F^{x_1}(A_{x_3}) - D_{x_3} F(A_{x_3})) + (D_{x_3} F(A_{x_3}) - D_{x_3} F).$$

We start with γ_3. By the fact that $(a+b)^3 \le 4(a^3 + b^3)$ for any $a \ge 0, b \ge 0$, we have

$$\begin{aligned}(\operatorname{Var} F)^{\frac{3}{2}} \gamma_3 &= \int \mathbb{E}|(D_x F - D_x F(A_x)) + D_x F(A_x)|^3 \lambda(dx) \\ &\le \int 4\mathbb{E}|D_x F - D_x F(A_x)|^3 \lambda(dx) + \int 4\mathbb{E}|D_x F(A_x)|^3 \lambda(dx).\end{aligned}$$

By Hölder's inequality and the assumptions in Theorem 2.1, we then have

$$\mathbb{E}|D_x F(A_x)|^3 \le (\mathbb{E}|D_x F(A_x)|^4)^{\frac{3}{4}} 1^{1-\frac{3}{4}} \le b_2(x, A_x)^{\frac{3}{4}}. \tag{5.1}$$

Similarly,

$$\mathbb{E}|D_x F - D_x F(A_x)|^3 \le b_1(x, A_x)^{\frac{3}{4}}.$$

Therefore,

$$(\operatorname{Var} F)^{\frac{3}{2}}\gamma_3 \leq C\int\sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}}\lambda(dx) =: C(\operatorname{Var} F)^{\frac{3}{2}}\gamma_3'.$$

Next, we turn to γ_4. According to Last et al. (2016, Lemma 4.3), we have

$$\mathbb{E}(F-\mathbb{E}F)^4 \leq \max\left\{256\left(\int(\mathbb{E}(D_xF)^4)^{\frac{1}{2}}\lambda(dx)\right)^2, 4\int\mathbb{E}(D_xF)^4\lambda(dx) + 2(\operatorname{Var} F)^2\right\}.$$

Consequently,

$$\gamma_4 \leq \frac{\int(\mathbb{E}(D_xF)^4)^{\frac{3}{4}}\lambda(dx)}{2(\operatorname{Var} F)^2}\left(4\left(\int(\mathbb{E}(D_xF)^4)^{\frac{1}{2}}\lambda(dx)\right)^{\frac{1}{2}} + \sqrt{2}\left(\int\mathbb{E}(D_xF)^4\lambda(dx)\right)^{\frac{1}{4}} + 2^{\frac{1}{4}}(\operatorname{Var} F)^{\frac{1}{2}}\right).$$

By calculations similar to γ_3, we have

$$\int(\mathbb{E}(D_xF)^4)^{\frac{1}{2}}\lambda(dx) \leq C\int\sum_{j=1}^{2} b_j(x, A_x)^{\frac{1}{2}}\lambda(dx).$$

Similarly,

$$\int(\mathbb{E}(D_xF)^4)^{\frac{3}{4}}\lambda(dx) \leq C\int\sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}}\lambda(dx),$$

$$\int\mathbb{E}(D_xF_s)^4\lambda(dx) \leq C\int\sum_{j=1}^{2} b_j(x, A_x)\lambda(dx).$$

Combining all above, we have

$$\gamma_4 \leq \frac{C\int\sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}}\lambda(dx)}{(\operatorname{Var} F)^2}\left(\left(\int\sum_{j=1}^{2} b_j(x, A_x)^{\frac{1}{2}}\lambda(dx)\right)^{\frac{1}{2}} + \left(\int\sum_{j=1}^{2} b_j(x, A_x)\lambda(dx)\right)^{\frac{1}{4}} + (\operatorname{Var} F)^{\frac{1}{2}}\right) =: C\gamma_4'.$$

Furthermore, by similar arguments on bounding γ_3, it leads to

$$\gamma_5 \leq \frac{C}{\operatorname{Var} F}\left(\int\sum_{j=1}^{2} b_j(x, A_x)\lambda(dx)\right)^{\frac{1}{2}} =: C\gamma_5'.$$

We next move on to bounding the remaining part: γ_1, γ_2 and γ_6, that involve the second order cost functions. Before we proceed, for notational convenience, we would like to define for all $x_1, x_2 \in \mathbb{X}$ and $A_{x_1} \in \mathcal{F}$,

$$b_5(x_1, x_2, A_{x_1}) := b_1(x_1, A_{x_1}).$$

Although it is true that b_5 actually doesn't depend on x_2, this slight abuse of notation allows us to write

$$\mathbb{E}(D_{x_1,x_2}F)^4 \leq C\sum_{j=3}^{4} b_j(x_1, x_2, A_{x_1}) + b_1(x_1, A_{x_1})$$

$$= C\sum_{j=3}^{5} b_j(x_1, x_2, A_{x_1}), \tag{5.2}$$

as a compact form, similar to the first order cost function.

Using Cauchy-Schwarz inequality, we have

$$(\mathbb{E}(D_{x_1}F)^2(D_{x_2}F)^2)^{\frac{1}{2}} \le \mathbb{E}(D_{x_1}F)^4)^{\frac{1}{4}}\mathbb{E}(D_{x_2}F)^4)^{\frac{1}{4}}, \tag{5.3}$$

$$(\mathbb{E}(D_{x_1,x_3}F)^2(D_{x_2,x_3}F)^2)^{\frac{1}{2}} \le \mathbb{E}(D_{x_1,x_3}F)^4)^{\frac{1}{4}}\mathbb{E}(D_{x_2,x_3}F)^4)^{\frac{1}{4}}. \tag{5.4}$$

With all results above, according to (5.1) to (5.4), we give an upper bound for γ_1 in a similar way by Hölder's inequality:

$$\begin{aligned}\gamma_1 \le \frac{C}{\operatorname{Var} F}\Bigg(\int \Bigg(&\sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{4}} \sum_{j=1}^{2} b_j(x_2, A_{x_2})^{\frac{1}{4}} \\ &\sum_{j=3}^{5} b_j(x_3, x_1, A_{x_3})^{\frac{1}{4}} \sum_{j=3}^{5} b_j(x_3, x_2, A_{x_3})^{\frac{1}{4}}\Bigg)\lambda^3(d(x_1,x_2,x_3))\Bigg)^{\frac{1}{2}} \\ =:&\, C\gamma_1'.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\gamma_2 &\le \frac{C}{\operatorname{Var} F}\left(\int \sum_{j=3}^{5} b_j(x_3, x_1, A_{x_3}) \sum_{j=3}^{5} b_j(x_3, x_2, A_{x_3})\lambda^3(d(x_1,x_2,x_3))\right)^{\frac{1}{2}} \\ &:= C\gamma_2', \\ \gamma_6 &\le \frac{C}{\operatorname{Var} F}\Bigg(\int \sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{2}} \sum_{j=3}^{5} b_j(x_1, x_2, A_{x_1})^{\frac{1}{2}} + \sum_{j=3}^{5} b_j(x_1, x_2, A_{x_1})\lambda^2(d(x_1,x_2))\Bigg)^{\frac{1}{2}} \\ &=: C\gamma_6'.\end{aligned}$$

Combining the obtained bounds above for γ_i, $1 \le i \le 6$, we complete the proof. □

5.1.1 Proof of Corollary 2.1

Without loss of generality, we assume $c_1 = k_1 := C_1, c_2 = k_2 := C_2$ and $c_3 = k_3 := C_3$. In Theorem 2.1, we set $A_x = \mathbb{X}$ and $\lambda = s\mathbb{Q}$. Then, we can set

$$\begin{aligned} b_1(x, A_x) &= 0, \\ b_3(x_1, x_2, A_{x_1}) &= 0, \\ b_5(x_1, x_2, A_{x_1}) &= 0. \end{aligned}$$

Recall that as mentioned in Section **??**, for notational convenience, we would like to define for all $x_1, x_2 \in \mathbb{X}$ and $A_{x_1} \in \mathcal{F}$,

$$b_5(x_1, x_2, A_{x_1}) := b_1(x_1, A_{x_1}).$$

According to Assumption 2.3, by Hölder's inequality, we have:

$$b_2(x, \mathbb{X}) := \mathbb{E}|D_xF_s|^4 \le \mathbb{E}|D_xF_s|^p\mathbb{P}(D_xF_s \ne 0)^{1-\frac{4}{p}} \le C\mathbb{P}(D_xF_s \ne 0)^{1-\frac{4}{p}},$$

and

$$b_4(x_1,x_2,\mathbb{X}) := \mathbb{E}|D_{x_1}F_s^{x_2} - D_{x_1}F_s|^4 \le \mathbb{E}|D_{x_1}F_s^{x_2} - D_{x_1}F_s|^p \mathbb{P}(|D_{x_1}F_s^{x_2} - D_{x_1}F_s| \neq 0)^{1-\frac{4}{p}}$$
$$\le C\mathbb{P}(|D_{x_1}F_s^{x_2} - D_{x_1}F_s| \neq 0)^{1-\frac{4}{p}}.$$

Define

$$\phi_s := s\int_{\mathbb{X}} \mathbb{P}(D_xF_s \neq 0)^{\frac{p-4}{2p}}\mathbb{Q}(dx),$$
$$\psi_s(x_1,x_2) := \mathbb{P}(|D_{x_1}F_s^{x_2} - D_{x_1}F_s| \neq 0)^{\frac{p-4}{4p}}. \tag{5.5}$$

With all b_i, $1 \le i \le 5$, given above, we will bound all γ_i', $1 \le i \le 6$ in Theorem 2.1. We again start with γ_3':

$$\begin{aligned}\gamma_3' &= \frac{1}{(\operatorname{Var} F_s^{\frac{3}{2}}} \int \sum_{j=1}^{2} b_j(x,A_x)^{\frac{3}{4}}\lambda(dx)\\ &= Cs\int_{\mathbb{X}} \mathbb{P}(D_xF_s \neq 0)^{(1-\frac{4}{p})\frac{3}{4}}\mathbb{Q}(dx)\\ &\le C\phi_s.\end{aligned}$$

Similarly, as for γ_3', it holds that

$$\begin{aligned}\gamma_5' &= \frac{1}{\operatorname{Var} F_s}\left(\int \sum_{j=1}^{2} b_j(x,A_x)\lambda(dx)\right)^{\frac{1}{2}}\\ &= \frac{1}{\operatorname{Var} F_s}\left(s\int_{\mathbb{X}} \mathbb{P}(D_xF_s \neq 0)^{1-\frac{4}{p}}\mathbb{Q}(dx)\right)^{\frac{1}{2}}\\ &\le C\phi_s^{\frac{1}{2}}.\end{aligned}$$

For γ_4', we have

$$\begin{aligned}\gamma_4' = \frac{\int \sum_{j=1}^{2} b_j(x,A_x)^{\frac{3}{4}}\lambda(dx)}{(\operatorname{Var} F_s)^2}\Bigg(&\left(\int \sum_{j=1}^{2} b_j(x,A_x)^{\frac{1}{2}}\lambda(dx)\right)^{\frac{1}{2}}\\ &+\left(\int \sum_{j=1}^{2} b_j(x,A_x)\lambda(dx)\right)^{\frac{1}{4}} + (\operatorname{Var} F_s)^{\frac{1}{2}}\Bigg).\end{aligned} \tag{5.6}$$

Respectively,

$$\begin{aligned}\left(\int \sum_{j=1}^{2} b_j(x,A_x)^{\frac{1}{2}}\lambda(dx)\right)^{\frac{1}{2}} &= \left(s\int_{\mathbb{X}} \mathbb{P}(D_xF_s \neq 0)^{(1-\frac{4}{p})\frac{1}{2}}\mathbb{Q}(dx)\right)^{\frac{1}{2}}\\ &\le C\phi_s^{\frac{1}{2}},\end{aligned}$$

$$\left(\int \sum_{j=1}^{2} b_j(x,A_x)\lambda(dx)\right)^{\frac{1}{4}} = \left(s\int_{\mathbb{X}} \mathbb{P}(D_xF_s \neq 0)^{1-\frac{4}{p}}\mathbb{Q}(dx)\right)^{\frac{1}{4}}$$

$$\leq C\phi_s^{\frac{1}{4}},$$

and

$$\int \sum_{j=1}^{2} b_j(x, A_x)^{\frac{3}{4}} \lambda(dx) = s \int_{\mathbb{X}} \mathbb{P}(D_x F_s \neq 0)^{(1-\frac{4}{p})\frac{3}{4}} \mathbb{Q}(dx)$$
$$\leq C\phi_s.$$

From the above calculations, we see that the exponent $(p-4)/2p$ is set so that indeed ϕ_s provides an upper bound for all the terms appearing in the right hand of (5.6). Hence, we have

$$\gamma_4' \leq C\left(\frac{\phi_s^{\frac{3}{2}} + \phi_s^{\frac{5}{4}}}{(\operatorname{Var} F_s)^2} + \frac{\phi_s}{(\operatorname{Var} F_s)^{\frac{3}{2}}}\right).$$

For terms that include b_j, $3 \leq j \leq 5$, we have

$$\gamma_1' = \frac{1}{\operatorname{Var} F_s}\Bigg(\int \Bigg(\sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{4}} \sum_{j=1}^{2} b_j(x_2, A_{x_2})^{\frac{1}{4}}$$
$$\sum_{j=3}^{5} b_j(x_3, x_1, A_{x_3})^{\frac{1}{4}} \sum_{j=3}^{5} b_j(x_3, x_2, A_{x_3})^{\frac{1}{4}}\Bigg)\lambda^3(d(x_1, x_2, x_3))\Bigg)^{\frac{1}{2}}$$
$$\leq C\frac{s^{\frac{3}{2}}}{\operatorname{Var} F_s}\Bigg(\int_{\mathbb{X}^3}\Big(\mathbb{P}(|D_{x_1}F_s^{x_2} - D_{x_1}F_s| \neq 0)^{\frac{p-4}{4p}}$$
$$\mathbb{P}(|D_{x_2}F_s^{x_3} - D_{x_2}F_s| \neq 0)^{\frac{p-4}{4p}}\Big)\mathbb{Q}^3(d(x_1, x_2, x_3))\Bigg)^{\frac{1}{2}}$$
$$= C\frac{s^{\frac{3}{2}}}{\operatorname{Var} F_s}\sqrt{\int_{\mathbb{X}}\left(\int_{\mathbb{X}} \psi_s(x_1, x_2)\mathbb{Q}(dx_2)\right)^2 \mathbb{Q}(dx_1)}.$$

Similarly,

$$\gamma_2' = \frac{C}{\operatorname{Var} F_s}\left(\int \sum_{j=3}^{5} b_j(x_3, x_1, A_{x_3}) \sum_{j=3}^{5} b_j(x_3, x_2, A_{x_3})\lambda^3(d(x_1, x_2, x_3))\right)^{\frac{1}{2}}$$
$$\leq C\frac{s^{\frac{3}{2}}}{\operatorname{Var} F_s}\Bigg(\int_{\mathbb{X}^3}\Big(\mathbb{P}(|D_{x_1}F_s^{x_2} - D_{x_1}F_s| \neq 0)^{\frac{p-4}{p}}$$
$$\mathbb{P}(|D_{x_2}F_s^{x_3} - D_{x_2}F_s| \neq 0)^{\frac{p-4}{p}}\Big)\mathbb{Q}^3(d(x_1, x_2, x_3))\Bigg)^{\frac{1}{2}}$$
$$= C\frac{s^{\frac{3}{2}}}{\operatorname{Var} F_s}\sqrt{\int_{\mathbb{X}}\left(\int_{\mathbb{X}} \psi_s(x_1, x_2)\mathbb{Q}(dx_2)\right)^2 \mathbb{Q}(dx_1)},$$

and

$$\gamma_6' = \frac{1}{\operatorname{Var} F_s}\Bigg(\int \sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{2}} \sum_{j=3}^{5} b_j(x_1, x_2, A_{x_1})^{\frac{1}{2}} + \sum_{j=3}^{5} b_j(x_1, x_2, A_{x_1})\lambda^2(d(x_1, x_2))\Bigg)^{\frac{1}{2}}$$

$$\le C\frac{s}{\operatorname{Var} F_s}\Big(\int_{\mathbb{X}^2}\mathbb{P}(|D_{x_2}F_s^{x_3}-D_{x_2}F_s|\neq 0)^{\frac{p-4}{2p}}$$

$$+\,\mathbb{P}(|D_{x_2}F_s^{x_3}-D_{x_2}F_s|\neq 0)^{\frac{p-4}{p}}\mathbb{Q}^2(d(x_1,x_2))\Big)^{\frac{1}{2}}$$

$$\le C\frac{s}{\operatorname{Var} F_s}\left(\int_{\mathbb{X}^2}\mathbb{P}(|D_{x_2}F_s^{x_3}-D_{x_2}F_s|\neq 0)^{\frac{p-4}{2p}}\mathbb{Q}^2(d(x_1,x_2))\right)^{\frac{1}{2}}$$

$$\le C\frac{s}{\operatorname{Var} F_s}\sqrt{\int_{\mathbb{X}^2}\psi_s(x_1,x_2)^2\mathbb{Q}^2(d(x_1,x_2))}.$$

Therefore, combining all bounds for γ_i', $1\le i\le 6$, we have by Theorem 2.1,

$$d_K\left(\frac{F_s-\mathbb{E}F_s}{\sqrt{\operatorname{Var} F_s}},N\right)\le C(\theta_1+\theta_2+\theta_3), \tag{5.7}$$

where

$$\theta_1 := \frac{s}{\operatorname{Var} F_s}\sqrt{\int_{\mathbb{X}^2}\psi_s(x_1,x_2)^2\mathbb{Q}^2(d(x_1,x_2))}, \tag{5.8}$$

$$\theta_2 := \frac{s^{\frac{3}{2}}}{\operatorname{Var} F_s}\sqrt{\int_{\mathbb{X}}\left(\int_{\mathbb{X}}\psi_s(x_1,x_2)\mathbb{Q}(dx_2)\right)^2\mathbb{Q}(dx_1)}, \tag{5.9}$$

$$\theta_3 := \frac{(\phi_s)^{\frac{1}{2}}}{\operatorname{Var} F_s}+\frac{\phi_s}{(\operatorname{Var} F_s)^{\frac{3}{2}}}+\frac{\phi_s^{\frac{5}{4}}+\phi_s^{\frac{3}{2}}}{(\operatorname{Var} F_s)^2}. \tag{5.10}$$

We next proceed to obtain refined bounds for ϕ_s and $\psi_s(x_1,x_2)$. Before proceeding, we recall the following definitions from (2.4):

$$I_s(x) := \mathbb{P}(D_xF_s\neq 0),$$
$$J_s(x_1,x_2) := P\big(|D_{x_1,x_2}F_s|\neq 0\big).$$

Lemma 5.1. *Assume that all conditions in Corollary 2.1 hold, and recall that R_x denotes the radius of stabilization. Then,*

$$\phi_s\le Cs\int_{\mathbb{X}}e^{-C_2\frac{p-4}{2p}d_s(x,\mathbb{K})^{C_3}}\mathbb{Q}(dx),$$

and

$$\psi_s(x_1,x_2)\le Ce^{-C_2\frac{p-4}{4p}d_s(x_1,x_2)^{C_3}}.$$

Proof of Lemma 5.1. Based on the Assumptions 2.1 and 2.2, we immediately obtain

$$I_s(x)\le C_1e^{-C_2d_s(x,\mathbb{K})^{C_3}},$$
$$J_s(x_1,x_2)\le C_1e^{-C_2\max\{d_s(x_1,x_2),d_s(x_1,\mathbb{K}),d_s(x_2,\mathbb{K})\}^{C_3}}.$$

Recall the definition of ϕ_s in (5.5). Applying the above bound for $I_s(x)$, we have

$$\phi_s\le Cs\int_{\mathbb{X}}e^{-C_2\frac{p-4}{2p}d_s(x,\mathbb{K})^{C_3}}\mathbb{Q}(dx).$$

Similarly, applying the bound for $J_s(x_1,x_2)$, we obtain the stated upper bound for $\psi_s(x_1,x_2)$. □

Lemma 5.2. *Suppose the condition* (2.2) *holds. Then, for any $x \in \mathbb{X}$ and $r \geq 0$, we have*

$$\mathbb{Q}(B_x(r)) \leq \kappa r^{\omega}.$$

Proof of Lemma 5.2. For any $x \in \mathbb{X}$ fixed, consider $Q(r) := \mathbb{Q}(B_x(r)), r \geq 0$ as an increasing function of r. According to Lebesgue's theorem for the differentiability of monotone functions, the derivative $Q'(r)$ exists almost everywhere and then with the condition (2.2),

$$\mathbb{Q}(B_x(r)) - \mathbb{Q}(B_x(0)) \leq \int_0^r Q'(u)du \leq \int_0^r \kappa\omega u^{\omega-1}du = \kappa r^{\omega}.$$

Note that $\mathbb{Q}(B_x(0)) = 0$. Therefore, we obtain the desired result. $\square$

Lemma 5.3. *Suppose the condition* (2.2) *holds. For any $x \in \mathbb{X}$, $r \geq 0$ and $\alpha > 0$, there exists a constant $C > 0$ such that*

$$\int_{\mathbb{X}\backslash B_x(r)} e^{-\alpha d_s(x,y)^{C_3}}\mathbb{Q}(dy) \leq \frac{C}{s}e^{-\frac{1}{2}\alpha(s^{1/\omega}r)^{C_3}}.$$

Proof of Lemma 5.3. Let $\{r_n\}_{n=1}^{\infty}$ be an increasing sequence satisfying

$$r_1 := r, \qquad \lim_{n\to\infty} r_n = \infty, \qquad \text{and} \qquad \lim_{n\to\infty}\sup_{n\geq 2}|r_n - r_{n-1}| = 0.$$

Then,

$$\begin{aligned}
\int_{\mathbb{X}\backslash B_x(r)} e^{-\alpha d_s(x,y)^{C_3}}\mathbb{Q}(dy) &\leq \sum_{n=2}^{\infty} e^{-\alpha(s^{1/\omega}r_{n-1})^{C_3}}\mathbb{Q}(B_x(r_n)\backslash B_x(r_{n-1})) \\
&\leq \sum_{n=2}^{\infty} e^{-\alpha(s^{1/\omega}r_{n-1})^{C_3}}\kappa\omega r_{n-1}^{\omega-1}(r_n - r_{n-1}) \\
&= \int_r^{\infty} e^{-\alpha(s^{1/\omega}u)^{C_3}}\kappa\omega u^{\omega-1}du.
\end{aligned}$$

Therefore, it suffices to show

$$\zeta(s,r) := se^{\frac{\alpha}{2}(s^{1/\omega}r)^{C_3}}\int_r^{\infty} e^{-\alpha(s^{1/\omega}u)^{C_3}}u^{\omega-1}du$$

is bounded on $\Theta := \{(s,r) : s \geq 1, r \geq 0\}$. Since $\zeta(s,r)$ is a continuous function, then we only need to show

$$\lim_{(s,r)\to\partial\Theta}\zeta(s,r) < \infty.$$

Note that

$$\zeta(s,r) \leq se^{(\frac{\alpha}{2}-\frac{2\alpha}{3})(s^{1/\omega}r)^{C_3}}\int_r^{\infty} e^{-\frac{\alpha}{3}(s^{1/\omega}u)^{C_3}}u^{\omega-1}du.$$

Moreover, let

$$\int_r^{\infty} e^{-\frac{\alpha}{3}(s^{1/\omega}u)^{C_3}}u^{\omega-1}du \leq \int_0^{\infty} e^{-\frac{\alpha}{3}(s^{1/\omega}u)^{C_3}}u^{\omega-1}du =: \eta(s),$$

with

$$\lim_{s\to\infty}\eta(s) < \infty.$$

Then,

$$\zeta(s,r) \leq se^{(\frac{\alpha}{2}-\frac{2\alpha}{3})(s^{1/\omega}r)^{C_3}}\ \eta(s).$$

Consequently, noting

$$\frac{\alpha}{2}-\frac{2\alpha}{3}=-\frac{1}{6}\alpha<0,$$

we have

$$\lim_{(s,r)\to\partial\Theta}\zeta(s,r)<\infty,$$

giving us the desired result. □

Lemma 5.4. *For any $s \geq 1$, $r \geq 0$ and $\beta > 0$, there exists a constant $C > 0$ such that*

$$sr^{\omega}e^{-\beta(s^{1/\omega}r)^{C_3}} \leq Ce^{-\frac{1}{2}\beta(s^{1/\omega}r)^{C_3}}.$$

Proof of Lemma 5.4. Let

$$\mu(s,r) := sr^{\omega}e^{-\frac{1}{2}\beta(s^{1/\omega}r)^{C_3}}.$$

It suffices to prove

$$\sup_{s\geq 1, r\geq 0}\mu(s,r)<\infty.$$

Similar to $\zeta(s,r)$ in Lemma 5.3, we only need to show

$$\lim_{(s,r)\to\partial\Theta}\mu(s,r)<\infty,$$

where $\Theta := \{(s,r) : s \geq 1, r \geq 0\}$. We note that the above claim follows by calculations similar to that in the proof of Lemma 5.3 and the fact that $-\beta/2 < 0$, thus providing the desired result. □

Before proceeding, we recall the definition of $\Theta_{\mathbb{K},s}$ from (2.5) for convenience:

$$\Theta_{\mathbb{K},s} := s\int_{\mathbb{X}} e^{-C_2\frac{(p-4)}{4p}\left(\frac{d_s(x,\mathbb{K})}{2}\right)^{C_3}}\mathbb{Q}(dx).$$

Note that by Lemma 5.1, we also have that $\phi_s \leq C\Theta_{\mathbb{K},s}$, for a constant $C > 0$.

Lemma 5.5. *Suppose the conditions in Corollary 2.1 hold. Then, there exists a constant $C > 0$ such that*

$$s^2\int_{\mathbb{X}^2}\psi_s(x_1,x_2)^2\mathbb{Q}^2(d(x_1,x_2)) \leq C\Theta_{\mathbb{K},s}.$$

Proof of Lemma 5.5. Without loss of generality, we assume $d_s(x_1,\mathbb{K}) \geq d_s(x_2,\mathbb{K})$. Similar reasoning can be used for the other case. According to Lemma 5.1,

$$\psi_s(x_1,x_2) \leq Ce^{-C_2\frac{p-4}{4p}\max\{d_s(x_1,x_2),d_s(x_1,\mathbb{K})\}^{C_3}}.$$

Let

$$L_{x_2,s} := s\int_{\mathbb{X}}\psi_s(x_1,x_2)^2\mathbb{Q}(dx_1).$$

It suffices to show there exists a constant $C > 0$ such that

$$L_{x_2,s} \leq Ce^{-C_2 \frac{p-4}{2p} \left(\frac{d_s(x_2,\mathbb{K})}{2} \right)^{C_3}}.$$

Let $r := \frac{1}{2} d(x_2, \mathbb{K})$ and note the fact that $\max\{x, y\} \geq x, \max\{x, y\} \geq y$ for any x, y, then

$$\begin{aligned} L_{x_2,s} &\leq Cs \int_{\mathbb{X}} e^{-C_2 \frac{p-4}{2p} \max\{d_s(x_1,x_2), d_s(x_1,\mathbb{K})\}^{C_3}} \mathbb{Q}(dx_1) \\ &\leq Cs \int_{B_{x_2}(r)} e^{-C_2 \frac{p-4}{2p} d_s(x_1,\mathbb{K})^{C_3}} \mathbb{Q}(dx_1) + Cs \int_{\mathbb{X} \setminus B_{x_2}(r)} e^{-C_2 \frac{p-4}{2p} d_s(x_1,x_2)^{C_3}} \mathbb{Q}(dx_1) \\ &:= CL_1 + CL_2. \end{aligned}$$

By the triangle inequality, $2r \leq d(x_2, \mathbb{K}) \leq d(x_1, x_2) + d(x_1, \mathbb{K})$, then when $d(x_2, x_1) \leq r$, $d(x_1, \mathbb{K}) \geq r$. Therefore, according to Lemma 5.2 and lemma 5.4, there exists a constant $C > 0$ such that

$$L_1 \leq s \int_{B_{x_2}(r)} e^{-C_2 \frac{p-4}{2p} (s^{1/\omega} r)^{C_3}} \mathbb{Q}(dx_1) \leq s\kappa r^{\omega} e^{-C_2 \frac{p-4}{4p} (s^{1/\omega} r)^{C_3}} \leq Ce^{-C_2 \frac{p-4}{4p} (s^{1/\omega} r)^{C_3}}.$$

According to Lemma 5.3, there exists a constant $C > 0$ such that

$$L_2 \leq s \cdot \frac{C}{s} e^{-C_2 \frac{p-4}{4p} \left(s^{1/\omega} r \right)^{C_3}} = Ce^{-C_2 \frac{p-4}{4p} \left(s^{1/\omega} r \right)^{C_3}}.$$

Then,

$$L_{x_2,s} \leq Ce^{-C_2 \frac{p-4}{4p} \left(s^{1/\omega} r \right)^{C_3}} = Ce^{-C_2 \frac{p-4}{4p} \left(\frac{d_s(x_2,\mathbb{K})}{2} \right)^{C_3}},$$

giving us the desired result. □

Lemma 5.6. *Suppose the conditions in Corollary 2.1 hold. Then, there exists a constant $C > 0$ such that*

$$s^3 \int_{\mathbb{X}} \left(\int_{\mathbb{X}} \psi_s(x_1, x_2) \mathbb{Q}(dx_2) \right)^2 \mathbb{Q}(dx_1) \leq C\Theta_{\mathbb{K},s}.$$

Proof of Lemma 5.6. We can prove this lemma in a similar way as Lemma 5.5. Let

$$L'_{x_1,s} := s \int_{\mathbb{X}} \psi_s(x_1, x_2) \mathbb{Q}(dx_2).$$

Similar to $L_{x_2,s}$, one can show there exists a constant $C > 0$ such that

$$L'_{x_1,s} \leq Ce^{-C_2 \frac{p-4}{8p} \left(\frac{d_s(x_1,\mathbb{K})}{2} \right)^{C_3}}.$$

Therefore,

$$\begin{aligned} s^3 \int_{\mathbb{X}} \left(\int_{\mathbb{X}} \psi_s(x_1, x_2) \mathbb{Q}(dx_2) \right)^2 \mathbb{Q}(dx_1) &= s \int_{\mathbb{X}} \left(s \int_{\mathbb{X}} \psi_s(x_1, x_2) \mathbb{Q}(dx_2) \right)^2 \mathbb{Q}(dx_1) \\ &\leq Cs \int_{\mathbb{X}} e^{-C_2 \frac{p-4}{4p} \left(\frac{d_s(x_1,\mathbb{K})}{2} \right)^{C_3}} \mathbb{Q}(dx_1) \\ &\leq C\Theta_{\mathbb{K},s}. \end{aligned}$$

□

With the above ingredients in place, we are finally in a position to prove Corollary 2.1.

Proof of Corollary 2.1. Recall the definition of θ_1, θ_2 and θ_3 respectively in (5.8), (5.9) and (5.10). According to Lemma 5.1, Lemma 5.5 and Lemma 5.6, there exists a constant $C > 0$ such that

$$\theta_1 \leq C\frac{(\Theta_{\mathbb{K},s})^{\frac{1}{2}}}{\operatorname{Var} F_s}, \quad \theta_2 \leq C\frac{(\Theta_{\mathbb{K},s})^{\frac{1}{2}}}{\operatorname{Var} F_s}, \quad \phi_s \leq C\Theta_{\mathbb{K},s}.$$

Therefore, according to (5.7), we complete the proof. □

5.1.2 Proof of Theorem 2.2

The proof of the Poisson case applies mutatis mutandis to the binomial case. We now highlight the main changes. First, recall that we do not have a similar result like Theorem 2.1 for Poisson case due to the fact that there is no nice counterpart of the second order Poincaré inequality (see Last et al. (2016)). Hence, Theorem 2.2 cannot follow from Theorem 2.1. Instead, we use (Lachièze-Rey et al., 2019, Theorem 4.2) that provides an auxiliary result for the binomial case. While Lachièze-Rey et al. (2019) provided a general result for marked binomial point process, we state the following result for the unmarked binomial point process.

Theorem 5.1 (Lachièze-Rey et al., 2019)**.** *Let $n \geq 3$ and let F be a functional of a binomial point process η_n with $\mathbb{E}F(\xi_n)^2 < \infty$. Assume that there are constants $c, \rho \in (0, \infty)$ such that*

$$\mathbb{E}|D_x F(\xi_{n-1-|\mathbb{A}|} \cup \mathbb{A})|^{4+\rho} \leq c, \quad \mathbb{Q}-a.e., x \in \mathbb{X}, \mathbb{A} \subset \mathbb{X}, |\mathbb{A}| \leq 2.$$

Then, there is a constant $C := C(c, \rho) \in (0, \infty)$ such that

$$d_K\left(\frac{F - \mathbb{E}F}{\sqrt{\operatorname{Var} F}}, N\right) \leq C(S_1' + S_2' + S_3'),$$

with

$$\begin{aligned}
{\Gamma_n}' &:= n\int_{\mathbb{X}} \mathbb{P}(D_x F(\xi_{n-1}) \neq 0)^{\frac{\rho}{8+2\rho}} \mathbb{Q}(dx), \\
\psi_n'(x, x') &:= \sup_{\mathbb{A} \subset \mathbb{X}: |\mathbb{A}| \leq 1} \mathbb{P}(D_{x,x'} F(\xi_{n-2-|\mathbb{A}|}) \neq 0)^{\frac{\rho}{8+2\rho}}, \\
S_1' &:= \frac{n}{\operatorname{Var} F}\sqrt{\int_{\mathbb{X}^2} \psi_n'(x, x') \mathbb{Q}^2(d(x, x'))}, \\
S_2' &:= \frac{n^{\frac{3}{2}}}{\operatorname{Var} F}\sqrt{\int_{\mathbb{X}} \left(\int_{\mathbb{X}} \psi_n'(x, x') \mathbb{Q}(dx')\right)^2 \mathbb{Q}(dx)}, \\
S_3' &:= \frac{(\Gamma_n')^{\frac{1}{2}}}{\operatorname{Var} F} + \frac{\Gamma_n'}{(\operatorname{Var} F)^{\frac{3}{2}}} + \frac{\Gamma_n' + (\Gamma_n')^{\frac{3}{2}}}{(\operatorname{Var} F)^2}.
\end{aligned} \tag{5.11}$$

Remark 5.1. *The exponent of Γ_n' in the third component of the sum on the right hand side of (5.11) is different than that of Poisson case, due to the fact that it is not derived by the counterpart of the second order Poincaré inequality. In fact, it is obtained by Lachièze-Rey et al. (2019, Theorem 4.3).*

Now, based on Theorem 5.1, we start to prove Theorem 2.2.

Proof of Theorem 2.2. The proof of the binomial point process is similar to that of the Poisson case, based on Theorem 5.1 for binomial case. We treat Theorem 5.1 for binomial as a counterpart of (5.7) for Poisson noting $D_{x,x'}F_n = D_x F_n^{x'} - D_x F_n$. Starting with this, one can follow the same procedure to get the required counterparts of Lemma 5.1 to Lemma 5.6 by changing s as n. This provides the desired result. □

5.1.3 Total edge length of k-NN

Proof of Theorem 2.3. We begin with the Poisson point process case and consider the statistic $F_s^{k\text{-NN}}(\mathcal{P}_s) \coloneqq \sum_{x\in\mathcal{P}_s} f_s(x,\mathcal{P}_s)$, with f_s as defined in (2.1).

Step 1: From Example 2.2, $F_s^{k\text{-NN}}(\mathcal{P}_s)$ is a strongly stabilizing functional with the radius of stabilization $R_x = 4R$, with R being defined as follows: for each $t > 0$, construct six disjoint equilateral triangles $T_j(t)$, $1 \le j \le 6$, such that the origin is a vertex of each triangle, such that each triangle has edge length t and such that $T_j(t) \subset T_j(u)$ whenever $t < u$. Then, define R to be the minimum t such that each triangle $T_j(t)$ contains at least $k+1$ points from $\mathcal{P}_s$. Consequently, we have

$$\mathbb{P}(R > r) \le \mathbb{P}\left(\mathcal{P}_s\left(\cup_{j=1}^6 T_j(r)\right) \le 6k\right) \le \mathbb{P}\left(\mathcal{P}_s\left(B_x\left(r\sqrt{3}/2\right)\right) \le 6k\right),$$

where $\mathcal{P}_s\left(B_x\left(r\sqrt{3}/2\right)\right)$ follows Poisson distribution with parameter $s\mathbb{Q}\left(B_x\left(r\sqrt{3}/2\right)\right)$. According to the assumption (2.14) and a Chernoff bound for Poisson tail, (Penrose, 2003, Lemma 1.2), we have that there exits a constant $c' > 0$ such that

$$\begin{aligned}\mathbb{P}(R > r) &\le \mathbb{P}\left(\mathcal{P}_s\left(B_x\left(\frac{r\sqrt{3}}{2}\right)\right) \le 6k\right)\\ &\le \mathbb{P}\left(\mathrm{Poi}\left(cs\left(\frac{r\sqrt{3}}{2}\right)^{\omega}\right) \le 6k\right)\\ &\le 6ke^{-c'sr^{\omega}}.\end{aligned}$$

This implies the radius of stabilization $R_x = 4R$ decays exponentially. Furthermore, we set $\mathbb{K} = \mathbb{X}$ for the $\mathbb{K}-$exponential bound.

Step 2: As for the bounded moment condition, according to Lachièze-Rey et al. (2019, Lemma 5.5), for some $p > 4$, the bounded moment condition holds.

Step 3: As for the variance condition, by the assumption (2.15), it is satisfied.

Therefore, according to Corollary 2.2, we end the proof for the Poisson case. The proof for the binomial case is similar to the Poisson case by considering a Chernoff bound for the binomial distribution as in Penrose (2003, Lemma 1.1) and Lachièze-Rey et al. (2019, Lemma 5.6). □

5.1.4 Shannon entropy estimation

Proof of Theorem 2.4. We start by replacing the biased center $\mathbb{E}F_n^{\mathrm{SE}}(\xi_n)$ by the true parameter $H(q)$. By triangle inequality, we have that

$$d_K\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - H(q)}{\sqrt{\mathrm{Var}\,F_n^{\mathrm{SE}}(\xi_n)}}, N\right) \le \underbrace{d_K\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\mathrm{Var}\,F_n^{\mathrm{SE}}(\xi_n)}}, N\right)}_{=:d_1}$$

$$+ \underbrace{d_K\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - H(q)}{\sqrt{\mathrm{Var}\, F_n^{\mathrm{SE}}(\xi_n)}}, \frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\mathrm{Var}\, F_n^{\mathrm{SE}}(\xi_n)}}\right)}_{=:d_2}.$$

We first apply Corollary 2.3 to bound d_1, using the three step approach.

Step 1: Similar to the total edge length of k-NN, following the proof of Penrose and Yukich (2001, Lemma 6.1), $F_n^{\mathrm{SE}}(\xi_n)$ is strongly stabilizing with the radius of stabilization R_x decaying exponentially and we take $\mathbb{K} = \mathbb{X}$.

Step 2: As for the moment condition, again, by Lachièze-Rey et al. (2019, Lemma 5.6), the bounded moment condition holds for $p > 4$.

Step 3: For the variance, note that for d_1, we have

$$\frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\mathrm{Var}\, F_n^{\mathrm{SE}}(\xi_n)}} \quad \stackrel{d}{=} \quad \frac{nF_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}\left(nF_n^{\mathrm{SE}}(\xi_n)\right)}{\sqrt{\mathrm{Var}\left(nF_n^{\mathrm{SE}}(\xi_n)\right)}},$$

where $\stackrel{d}{=}$ means equal in distribution. Therefore, we can consider the variance condition for the $nF_n^{\mathrm{SE}}(\xi_n)$ instead. According to Berrett et al. (2019, Lemma 7), there exits a constant $C > 0$ such that for $k_0^* \le k \le k_1^*$,

$$\sup_{n>0} \frac{n}{\mathrm{Var}\left(nF_n^{\mathrm{SE}}(\xi_n)\right)} \le C. \tag{5.12}$$

By Corollary 2.3, we immediately have that $d_1 \le C_0'(k,p)/\sqrt{n}$. Since k also diverges as n goes to ∞, it is necessary to calculate the constant $C_0'(k,p)$ more explicitly. Therefore, we derive the following lemma, which is a refined version of Theorem 5.1 as it reveals how the constant $C_0'(k,p)$ is related to the constant p and the functional F, and thus k.

Lemma 5.7. *Assume there are constants $c > 0, p_0 > 0$ such that*

$$\mathbb{E}|D_x F(\xi_{n-1-|\mathbb{A}|} \cup \mathbb{A})|^{4+p_0} \le c, \ \ |\mathbb{A}| \le 2.$$

Then, there exists some constant C not depending on n nor F such that

$$d_K\left(\frac{F - \mathbb{E}F}{\sqrt{\mathrm{Var}\, F}}, N\right) \le C(S_1 + S_2 + S_3 + S_4 + S_5),$$

where

$$S_1 := c^{\frac{2}{4+p_0}} \frac{n}{\mathrm{Var}\, F} \sqrt{\int_{\mathbb{X}^2} \psi_n(x,x')\mathbb{Q}^2(d(x,x'))},$$

$$S_2 := c^{\frac{2}{4+p_0}} \frac{n^{\frac{3}{2}}}{\mathrm{Var}\, F} \sqrt{\int_{\mathbb{X}} \left(\int_{\mathbb{X}} \psi_n(x,x')\mathbb{Q}(dx')\right)^2 \mathbb{Q}(dx)},$$

$$S_3 := c^{\frac{2}{4+p_0}} \frac{\sqrt{\Gamma_n}}{\mathrm{Var}\, F},$$

$$S_4 := \left(\sqrt{3}\max\left\{4\sqrt{2}c^{\frac{1}{4+p_0}} \frac{\Gamma_n^{\frac{1}{2}}}{(\mathrm{Var}\, F)^{\frac{1}{2}}}, \sqrt{2}c^{\frac{1}{4+p_0}} \frac{\Gamma_n^{\frac{1}{4}}}{(\mathrm{Var}\, F)^{\frac{1}{2}}} + 1\right\}\right.$$
$$\left. + c^{\frac{1}{4+p_0}} \frac{\Gamma_n^{\frac{1}{4}}}{n^{\frac{1}{4}}(\mathrm{Var}\, F)^{\frac{1}{2}}}\right) c^{\frac{3}{4+p_0}} \frac{\Gamma_n}{(\mathrm{Var}\, F)^{\frac{3}{2}}} + c^{\frac{4}{4+p_0}} \frac{\Gamma_n}{(\mathrm{Var}\, F)^2},$$

$$S_5 := c^{\frac{3}{4+p_0}} \frac{\Gamma_n}{(\operatorname{Var} F)^{\frac{3}{2}}}.$$

Here,

$$\Gamma_n := n \int_{\mathcal{X}} \mathbb{P}(D_x F(\xi_{n-1}) \neq 0)^{\frac{p}{8+2p_0}} \mathbb{Q}(dx),$$
$$\psi_n(x, x') := \sup_{\mathbb{A} \subset \mathbb{X}: |\mathbb{A}| \leq 1} \mathbb{P}(D_{x,x'} F(\xi_{n-2-|\mathbb{A}|} \cup \mathbb{A}) \neq 0)^{\frac{4}{8+2p_0}}.$$

The proof of Lemma 5.7 follows in a straightforward manner by Lachièze-Rey et al. (2019, Proof of Theorem 4.2). Now, we proceed to calculate moment bounds to see how k is related to the constant C_0'. Let

$$\zeta_i := \sum_{j=1}^{k} w_j \log \left(\frac{(n-1) V_d \rho_{j,i}^d}{e^{\Psi(j)}} \right).$$

Following Lachièze-Rey et al. (2019, Lemma 5.6), by Jensen's inequality, for $p = 4 + p_0$, $p_0 > 0$,

$$\begin{aligned}
&\mathbb{E}|D_y F_n^{\mathrm{SE}}(\xi_{n-1-|\mathbb{A}|} \cup \mathbb{A})|^{4+p_0} \\
&= \mathbb{E}\left|\zeta_k(y, \xi_{n-1-|\mathbb{A}|} \cup \{y\} \cup \mathbb{A}) + \sum_{x \in \xi_{n-1-|\mathbb{A}|} \cup \mathbb{A}} D_y \zeta_k(x, \xi_{n-1-|\mathbb{A}|})\right|^{4+p_0} \\
&\leq 4^{3+p_0} \mathbb{E}|\zeta_k(y, \xi_{n-1-|\mathbb{A}|} \cup \{y\} \cup \mathbb{A})|^{4+p_0} + 4^{3+p_0} \sum_{x \in \mathbb{A}} \mathbb{E}|D_y \zeta_k(x, \xi_{n-1-|\mathbb{A}|})|^{4+p_0} \\
&+ 4^{3+p_0} \sum_{x \in \xi_{n-1-|\mathbb{A}|}} \mathbb{E}|D_y \zeta_k(x, \xi_{n-1-|\mathbb{A}|})|^{4+p_0}.
\end{aligned}$$

Then, following Lachièze-Rey et al. (2019, Proof of Lemma 5.6), the constant c in Lemma 5.7 satisfies

$$c \lesssim 4^{3+p_0} c_a + 4^{3+p_0} c_b + 4^{5+\frac{3}{2}p_0} c_b^{\frac{4+p_0}{4+2p_0}} k^{\frac{4}{4+2p_0}},$$

where $A \lesssim B$ means there is a constant $C > 0$ such that $A \leq CB$. Note that according to the definition of weights (2.17), there only exist finitely many terms in the sum of ζ_i and according to Singh and Póczos (2016, Section 3.2),

$$c_a := \mathbb{E}|\zeta_k(y, \xi_{n-1-|\mathbb{A}|} \cup \{y\} \cup \mathbb{A})|^{4+p_0} \lesssim \|w\|_\infty^{4+p_0} \mathbb{E} \left| \log \left(\frac{(n-1) V_d \rho_{k,i}^d}{e^{\Psi(k)}} \right) \right|^{4+p_0} < \infty,$$

and similarly,

$$c_b : = \mathbb{E}|D_y \zeta_k(x, \xi_{n-1-|\mathbb{A}|})|^{4+p_0} < \infty.$$

Then,

$$c \lesssim k^{\frac{p_0}{4+2p_0}},$$

and by Lachièze-Rey et al. (2019, Theorem 4.3) and Lachièze-Rey and Peccati (2017, Theorem 5.1), the constant C in Lemma 5.7 does not depend on either n or k.

Therefore, according to Corollary 2.3 and Lemma 5.7,

$$d_1 \lesssim \frac{1}{\sqrt{n}} \times c^{\frac{4}{4+p_0}} \lesssim \frac{1}{\sqrt{n}} k^{\frac{4p_0}{(4+p_0)(4+2p_0)}}. \tag{5.13}$$

Next, we focus on d_2, which is related to the bias, $\mathbb{E}F_n^{\mathrm{SE}}(\xi_n) - H(q)$. Let

$$h^w := \frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}} \quad \text{and} \quad \Delta h := \frac{H(q) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}}.$$

Then we have

$$\begin{aligned}
d_2 :&= d_K\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - H(q)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}}, \frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}}\right) \\
&= \sup_{t\in\mathbb{R}} \left|\mathbb{P}\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - H(q)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}} \le t\right) - \mathbb{P}\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - \mathbb{E}F_n^{\mathrm{SE}}(\xi_n)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}} \le t\right)\right| \\
&= \sup_{t\in\mathbb{R}} |\mathbb{P}(h^w \le t + \Delta h) - \mathbb{P}(h^w \le t)| \\
&\le 2d_1 + \sup_{t\in\mathbb{R}} |\Phi(t+\Delta h) - \Phi(t)| \\
&= 2d_1 + 2\left(\Phi\left(\left|\frac{\Delta h}{2}\right|\right) - \frac{1}{2}\right) \\
&\le 2d_1 + 1 - e^{-\frac{1}{2}(\Delta h)^2 - \sqrt{\frac{2}{\pi}}|\Delta h|} \\
&\lesssim d_1 + |\Delta h|,
\end{aligned}$$

where Φ is the c.d.f. of the standard normal distribution and actually, we applied a tight bound for standard normal distribution function according to Mastin and Jaillet (2013). Therefore,

$$d_K\left(\frac{F_n^{\mathrm{SE}}(\xi_n) - H(q)}{\sqrt{\operatorname{Var} F_n^{\mathrm{SE}}(\xi_n)}}, N\right) \lesssim \max\{d_1, |\Delta h|\}.$$

The bias of $F_n^{\mathrm{SE}}(\xi_n)$ satisfies the following bound according to Berrett et al. (2019, Corollary 4): for every $\epsilon > 0$,

$$\sup_{f\in\mathcal{F}_{d,\theta}} |\mathbb{E}_f \hat{H}_n^w - H(f)| = O\left(\max\left\{\left(\frac{k}{n}\right)^{\frac{\alpha}{\alpha+d}-\epsilon}, \left(\frac{k}{n}\right)^{\frac{2(\lfloor\frac{d}{4}\rfloor+1)}{d}}, \left(\frac{k}{n}\right)^{\frac{\beta}{d}}\right\}\right). \tag{5.14}$$

Note that $\alpha > d$ and $\beta > \frac{d}{2}$. With (5.12) and (5.14) , by elementary algebraic manipulation we have that

$$|\Delta h| \lesssim \max\left\{\left(\frac{k}{n}\right)^{\frac{\alpha}{\alpha+d}-\frac{1}{2}-\epsilon}, \left(\frac{k}{n}\right)^{\frac{2(\lfloor\frac{d}{4}\rfloor+1)}{d}-\frac{1}{2}}, \left(\frac{k}{n}\right)^{\frac{\beta}{d}-\frac{1}{2}}\right\}. \tag{5.15}$$

Comparing (5.13) and (5.15), we obtain the desired result.

□

5.1.5 Euler characteristic

Proof of Theorem 2.5. We follow the 3-step procedure.

Step 1: According to Example 2.1, the Euler characteristic is strongly stabilizing with the radius of stabilization $R_x = 2r$. Clearly, as a constant, R_x can be bounded exponentially as one can choose c_1 large enough and $c_2, c_3 > 0$ such that for $0 \le t \le 2r$,

$$1 \le c_1 e^{-c_2 (n^{1/d} t)^{c_3}}.$$

The similar argument also holds for Poisson case. Also, we take $\mathbb{K} = [0,1]^d$.

Step 2: Moreover, for the bounded moment condition, note that for $p > 4$, there exists a constant $C(d,T) > 0$ such that

$$\begin{aligned}
\sup_{n>0, x\in[0,1]^d} \mathbb{E}|D_x F_n^{\mathrm{EC}}(\xi_n)|^p \le& \mathbb{E}\Bigg| \sum_{\ell=0}^{n} \#\{\sigma \in K_r\left(n^{\frac{1}{d}}(\xi_n \cup \{x\})\right) : \\
& \qquad \sigma \text{ is an } \ell\text{-simplex intersecting with } \{x\}\} \Bigg|^p \\
\le& \left| \sum_{\ell=0}^{n} \sum_{\{j_1,\dots,j_\ell\}\subset\{1,2,\dots,n\}} \left(\mathbb{P}\left(x_{j_1} \in B_x\left(n^{-\frac{1}{d}} r\right)\right)\right)^\ell \right|^p \\
\le& \left| \sum_{\ell=0}^{n} \binom{n}{\ell} \frac{(C(d,T)\|q\|_\infty r^d)^\ell}{n^\ell} \right|^p \\
\le& \left| \sum_{\ell=0}^{n} \frac{(C(d,T)\|q\|_\infty r^d)^\ell}{\ell!} \right|^p \\
\le& (e^{C(d,T)\|q\|_\infty r^d})^p < \infty.
\end{aligned} \tag{5.16}$$

Also, by Yogeshwaran et al. (2017, Lemma 4.1) and Krebs et al. (2021, Proof of Lemma 4.2),

$$\sup_{n>0, x\in[0,1]^d} \mathbb{E}|F_n^{\mathrm{EC}}(\mathcal{P}_n)|^p < \infty.$$

Moreover, consider

$$\begin{aligned}
D_x F_n^{\mathrm{EC}}(\xi_n)^{x^*} =& \chi\left(K_r\left(n^{\frac{1}{d}}(\xi_n \cup \{x\} \cup \{x^*\})\right)\right) - \chi\left(K_r\left(n^{\frac{1}{d}}\xi_n\right)\right) \\
&+ \chi\left(K_r\left(n^{\frac{1}{d}}\xi_n\right)\right) - \chi\left(K_r\left(n^{\frac{1}{d}}(\xi_n \cup \{x^*\})\right)\right),
\end{aligned}$$

and following a similar argument like (5.16), we have

$$\sup_{n>0, x\in[0,1]^d, x^*\in[0,1]^d} \mathbb{E}|D_x F_n^{\mathrm{EC}}(\xi_n)^{x^*}|^p < \infty.$$

The bounded moment condition is satisfied and similar arguments can be utilized to show for Poisson case.

Step 3: According to Penrose and Yukich (2001, Theorem 2.1) and Krebs et al. (2021, Proposition 4.6), there exists a constant $C > 0$ such that

$$\sup_{n>0} \frac{n}{\operatorname{Var} F_n^{\mathrm{EC}}(\xi_n)} \le C,$$

and it also holds for Poisson case.

Therefore, according to Corollary 2.2 and Corollary 2.3, we complete the proof. □

5.1.6 Edge length statistic of the minimum spanning tree

Proof of Theorem 2.6. According to Penrose (2005, Theorem 3.3), the total edge length $M(V)$ of the minimal spanning tree satisfies the strong stabilization with a radius of stabilization R_x almost surely finite. Without knowing the tail probability of R_x, we need to use Theorem 2.1. Set

$$A_{x_1} = B_n \cap \{x_1 + n^\alpha B_0\},$$

where $0 < \alpha < 1$ and $\{x + A\} := \{x + y : y \in A\}$ for any set A. According to Lachièze-Rey et al. (2022, Proposition 3.7), for any $q_0 > 0$ and $x \in B_n$, there exists a constant $C_{q_0} > 0$ independent of n, x such that uniformly

$$\mathbb{E}|D_x F_{B_n}^{\mathrm{MST}}(\mathcal{P}(\lambda))|^{q_0} \le C_{q_0}. \tag{5.17}$$

The same bound also holds for $D_x F_{B_n}^{y,\mathrm{MST}}$

As for the second order cost, according to Lachièze-Rey et al. (2022, Proposition 3.11)[3], for any $q_0 \ge 1$, there exist constants $E_0 > 0, E_1 > 0, E_2 > 0$ such that for any $x_1 \in B_n$ with $d(x_1, \partial B_n) > n^\alpha$,

$$\mathbb{E}|D_{x_1} F_{B_n}^{\mathrm{MST}}(A_{x_1}) - D_{x_1} F_{B_n}^{\mathrm{MST}}(\mathcal{P}(\lambda))|^{q_0} \le \begin{cases} E_0 n^{-E_1}, & \text{if } d = 2, \\ E_0 (\log n)^{-E_2}, & \text{if } d \ge 3. \end{cases} \tag{5.18}$$

The same bound also holds for $D_{x_1} F_{B_n}^{y,\mathrm{MST}}$. On the other hand, the variance bound is directly from Lachièze-Rey et al. (2022, Propostion 3.9): there exists a constant $c > 0$,

$$\mathrm{Var}\, F_{B_n}^{\mathrm{MST}}(\mathcal{P}(\lambda)) \ge c|B_n| \asymp n^d.$$

Then, plugging these two bounds (5.17) and (5.18) in Theorem 2.1, we get that there exist constants $C_{\text{first}} > 0$ and $C_{\text{second}} > 0$ such that,

$$\sum_{j=1}^{2} b_j(x_1, A_{x_1})^{\frac{1}{4}} \le C_{\text{first}},$$

and for $x_3 \in B_n$, $d(x_3, \partial B_n) > n^\alpha$ and $d(x_1, x_3) \ge n^\alpha$,

$$\sum_{j=3}^{4} b_j(x_3, x_1, A_{x_3})^{\frac{1}{4}} + b_1(x_3, A_{x_3})^{\frac{1}{4}} \le \begin{cases} 2E_0 n^{-E_1}, & \text{if } d = 2, \\ 2E_0 (\log n)^{-E_2}, & \text{if } d \ge 3, \end{cases}$$

and for other $(x_1, x_3) \in B_n \times B_n$, the points near the boundary of B_n, similar to Lachièze-Rey et al. (2022, proof of Proposition 3.11), we just apply uniform moment bounds for the flexible cost functions and get

$$\sum_{j=3}^{4} b_j(x_3, x_1, A_{x_3})^{\frac{1}{4}} + b_1(x_3, A_{x_3})^{\frac{1}{4}} \le C_{\text{second}}.$$

Therefore, by the fact that

$$|\{(x, y) \in B_n^2 : A_x \cap B_n \cap A_y \ne \varnothing\}| \asymp n^d n^{d\alpha},$$

[3] This proposition is provided for $q_0 = 1$. However, a closer examination of the proof reveals that it can be extended to any $q_0 \ge 1$ in a straightforward manner.

then we have,

$$\gamma_1' \lesssim \begin{cases} \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 n^{-E_1}, & \text{if } d = 2, \\ \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 (\log n)^{-E_2}, & \text{if } d \geq 3. \end{cases}$$

Similarly, one can derive

$$\gamma_2' \lesssim \begin{cases} \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 n^{-E_1}, & \text{if } d = 2, \\ \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 (\log n)^{-E_2}, & \text{if } d \geq 3, \end{cases}$$

and

$$\gamma_3', \gamma_4', \gamma_5' \lesssim \frac{1}{n^{\frac{d}{2}}},$$

and

$$\gamma_6' \lesssim \begin{cases} \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 n^{-E_1}, & \text{if } d = 2, \\ \frac{n^{\frac{d\alpha}{2}}}{n^{\frac{d}{2}}} C_{\text{second}} + 2E_0 (\log n)^{-E_2}, & \text{if } d \geq 3, \end{cases}$$

Therefore, we complete the proof by invoking Theorem 2.1. □

5.2 Proof of Section 3

5.2.1 Proofs of results in Section 3.3

For multivariate normal approximation of Poisson functionals, the second order Poincaré inequalities (see, e.g., Schulte and Yukich, 2019, Theorem 1.1 and 1.2) serve as a key tool. Even though the result therein is stated for an unmarked Poisson process, we can obviously apply it to a marked point process by simply considering the marked space as the underlying space. Since we consider the marked Poisson process $\mathcal{P}(\check{\lambda})$ with independent marks distributed as $\mathbb{Q}_{\mathbb{M}}$ and intensity measure $\check{\lambda} = \lambda \otimes \mathbb{Q}_{\mathbb{M}} = n\mathbb{Q} \otimes \mathbb{Q}_{\mathbb{M}}$, where $\mathbb{Q}$ is a σ-finite measure, we state in Theorem 5.2 below the marked version of the results. According to Schulte and Yukich (2019, Theorem 1.1 and 1.2), and upon using the Cauchy-Schwarz inequality, we have the following result; see the proof of Theorem 4.5 therein for more details.

Let $\mathbf{H} := (H^{(1)}, \ldots, H^{(m)})$ be a vector of functionals of the Poisson process $\mathcal{P}(\check{\lambda})$ with $\mathbb{E}[H^{(i)}] = 0$ and $H^{(i)} \in \textbf{dom}\ D$ for all $i \in [m]$. Denote $D_{\check{x}}\mathbf{H} := (D_{\check{x}}H^{(1)}, \ldots, D_{\check{x}}H^{(m)})$ and $D^2_{\check{x}_1,\check{x}_2}\mathbf{H} := (D^2_{\check{x}_1,\check{x}_2}H^{(1)}, \ldots, D^2_{\check{x}_1,\check{x}_2}H^{(m)})$ for $\check{x}, \check{x}_1, \check{x}_2 \in \check{\mathbb{X}}$ and define

$$\gamma_1 := \Bigg(\sum_{i,j=1}^{m} \int_{\mathbb{X}^3} (\mathbb{E}(D_{\check{x}_1}H^{(j)})^2 (D_{\check{x}_2}H^{(j)})^2)^{\frac{1}{2}} (\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}H^{(i)})^2 \\ \times (D^2_{\check{x}_2,\check{x}_3}H^{(i)})^2)^{\frac{1}{2}} \lambda^3(d(x_1,x_2,x_3)) \Bigg)^{\frac{1}{2}},$$

$$\gamma_2 := \Bigg(\sum_{i,j=1}^{m} \int_{\mathbb{X}^3} (\mathbb{E}(D^2_{\check{x}_1,\check{x}_3} H^{(i)})^2 (D^2_{\check{x}_2,\check{x}_3} H^{(i)})^2)^{\frac{1}{2}}$$
$$\times (\mathbb{E}(D^2_{\check{x}_1,\check{x}_3} H^{(j)})^2 (D^2_{\check{x}_2,\check{x}_3} H^{(j)})^2)^{\frac{1}{2}} \lambda^3(d(x_1,x_2,x_3)) \Bigg)^{\frac{1}{2}},$$
$$\gamma_3 := \sum_{i=1}^{m} \int_{\mathbb{X}} \mathbb{E}|D_{\check{x}} H^{(i)}|^3 \lambda(dx),$$
$$\gamma_4 := \Bigg(\sum_{i,j=1}^{m} \int_{\mathbb{X}} \mathbb{E}(D_{\check{x}} H^{(i)})^4 \lambda(dx) + 6 \int_{\mathbb{X}^2} (\mathbb{E}(D^2_{\check{x}_1,\check{x}_2} H^{(i)})^4)^{\frac{1}{2}} (\mathbb{E}(D_{\check{x}_1} H^{(j)})^4)^{\frac{1}{2}}$$
$$\lambda^2(d(x_1,x_2)) + 3 \int_{\mathbb{X}^2} (\mathbb{E}(D^2_{\check{x}_1,\check{x}_2} H^{(i)})^4)^{\frac{1}{2}} (\mathbb{E}(D^2_{\check{x}_1,\check{x}_2} H^{(j)})^4)^{\frac{1}{2}} \lambda^2(d(x_1,x_2)) \Bigg)^{\frac{1}{2}},$$
$$\gamma_5 := \Bigg(3 \sum_{i,j=1}^{m} \int_{\mathbb{X}^3} \Big(\mathbb{E} \mathbb{1}_{D^2_{\check{x}_1,\check{x}_3}\mathbf{H} \neq \mathbf{0}, D^2_{\check{x}_2,\check{x}_3}\mathbf{H} \neq \mathbf{0}} (\|D_{\check{x}_1}\mathbf{H}\| + \|D^2_{\check{x}_1,\check{x}_3}\mathbf{H}\|)^{\frac{3}{4}}$$
$$\times (\|D_{\check{x}_2}\mathbf{H}\| + \|D^2_{\check{x}_2,\check{x}_3}\mathbf{H}\|)^{\frac{3}{4}} |D_{\check{x}_1} H^{(i)}|^{\frac{3}{2}} |D_{\check{x}_2} H^{(i)}|^{\frac{3}{2}} \Big)^{\frac{2}{3}}$$
$$\times (\mathbb{E}|D_{\check{x}_1} H^{(j)}|^3 |D_{\check{x}_2} H^{(j)}|^3)^{\frac{1}{3}} \lambda^3(d(x_1,x_2,x_3))$$
$$+ \sum_{i,j=1}^{m} \int_{\mathbb{X}^3} \Big(\mathbb{E}(\|D_{\check{x}_1}\mathbf{H}\| + \|D^2_{\check{x}_1,\check{x}_3}\mathbf{H}\|)^{\frac{3}{2}} (\|D_{\check{x}_2}\mathbf{H}\| + \|D^2_{\check{x}_2,\check{x}_3}\mathbf{H}\|)^{\frac{3}{2}} \Big)^{\frac{1}{3}}$$
$$\times \Big(\frac{45}{2} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(i)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(i)}|^3)^{\frac{1}{3}} (\mathbb{E}|D_{\check{x}_1} H^{(j)}|^3 |D_{\check{x}_2} H^{(j)}|^3)^{\frac{1}{3}}$$
$$+ \frac{9}{2} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(i)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(i)}|^3)^{\frac{1}{3}} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(j)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(j)}|^3)^{\frac{1}{3}} \Big)$$
$$\lambda^3(d(x_1,x_2,x_3)) \Bigg)^{\frac{1}{3}},$$
$$\gamma_6 := \Bigg(3 \sum_{i,j=1}^{m} \int_{\mathbb{X}^3} \Big(\mathbb{E} \mathbb{1}_{D^2_{\check{x}_1,\check{x}_3}\mathbf{H} \neq \mathbf{0}, D^2_{\check{x}_2,\check{x}_3}\mathbf{H} \neq \mathbf{0}} (\|D_{\check{x}_1}\mathbf{H}\|^2 + \|D^2_{\check{x}_1,\check{x}_3}\mathbf{H}\|^2)^{\frac{3}{4}}$$
$$\times (\|D_{\check{x}_2}\mathbf{H}\|^2 + \|D^2_{\check{x}_2,\check{x}_3}\mathbf{H}\|^2)^{\frac{3}{4}} |D_{\check{x}_1} H^{(i)}|^{\frac{3}{2}} |D_{\check{x}_2} H^{(i)}|^{\frac{3}{2}} \Big)^{\frac{2}{3}}$$
$$(\mathbb{E}|D_{\check{x}_1} H^{(j)}|^3 |D_{\check{x}_2} H^{(j)}|^3)^{\frac{1}{3}} \lambda^3(d(x_1,x_2,x_3))$$
$$+ \sum_{i,j=1}^{m} \int_{\mathbb{X}^3} \Big(\mathbb{E}(\|D_{\check{x}_1}\mathbf{H}\|^2 + \|D^2_{\check{x}_1,\check{x}_3}\mathbf{H}\|^2)^{\frac{3}{2}} (\|D_{\check{x}_2}\mathbf{H}\|^2 + \|D^2_{\check{x}_2,\check{x}_3}\mathbf{H}\|^2)^{\frac{3}{2}} \Big)^{\frac{1}{3}}$$
$$\times \Big(\frac{135}{8} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(i)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(i)}|^3)^{\frac{1}{3}} (\mathbb{E}|D_{\check{x}_1} H^{(j)}|^3 |D_{\check{x}_2} H^{(j)}|^3)^{\frac{1}{3}}$$
$$+ \frac{27}{8} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(i)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(i)}|^3)^{\frac{1}{3}} (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3} H^{(j)}|^3 |D^2_{\check{x}_2,\check{x}_3} H^{(j)}|^3)^{\frac{1}{3}} \Big)$$
$$\lambda^3(d(x_1,x_2,x_3)) \Bigg)^{\frac{1}{4}}.$$

We note here that by a slight abuse of notation for the sake of brevity, in the expressions of γ_i, $1 \leq i \leq 6$ above as well as in integrals in the rest of this section, we write $\check{x}_1, \check{x}_2$ and $\check{x}_2$ to mean

$(x_1, \boldsymbol{m}_{x_1})$, $(x_2, \boldsymbol{m}_{x_2})$ and $(x_3, \boldsymbol{m}_{x_3})$ respectively, i.e., we integrate only over x_1, x_2 and x_3, while their marks are random, and are integrated as part of the various expectations in the integrands.

Theorem 5.2 (Schulte and Yukich (2019) Theorems 1.1 and 1.2, and 4.5)**.** *For* $\mathbf{H} := (H^{(1)}, \ldots, H^{(m)})$ *a vector of functionals* $\mathcal{P}(\check{\lambda})$ *with* $\mathbb{E}[H^{(i)}] = 0$ *and* $H^{(i)} \in \mathbf{dom}\ D$ *for all* $i \in [m]$*, let* γ_i*,* $1 \leq i \leq 6$ *be as above. Then, for a positive semi-definite matrix* $\Sigma := (\sigma_{ij})_{i,j=1}^m \in \mathbb{R}^m \times \mathbb{R}^m$*, we have*

$$\mathsf{d}_3\left(\mathbf{H}, N_\Sigma\right) \leq \frac{m}{2} \sum_{i,j=1}^m |\sigma_{ij} - \mathrm{Cov}(H^{(i)}, H^{(j)})| + m\gamma_1 + \frac{m}{2}\gamma_2 + \frac{m^2}{4}\gamma_3.$$

Additionally, if Σ *is positive definite, we have*

$$\mathsf{d}_2(\mathbf{H}, N_\Sigma) \leq \|\Sigma^{-1}\|_{op} \|\Sigma\|_{op}^{\frac{1}{2}} \sum_{i,j=1}^m \left|\sigma_{ij} - \mathrm{Cov}(H^{(i)}, H^{(j)})\right| + 2\|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}^{\frac{1}{2}}\gamma_1$$
$$+ \|\Sigma^{-1}\|_{op}\|\Sigma\|_{op}^{\frac{1}{2}}\gamma_2 + \frac{\sqrt{2\pi}m^2}{8}\|\Sigma^{-1}\|_{op}^{\frac{3}{2}}\|\Sigma\|_{op}\gamma_3,$$

and

$$\mathsf{d}_{cvx}(\mathbf{H}, N_\Sigma) \leq 941 m^5 \left(\|\Sigma^{-1/2}\|_{op} \vee \|\Sigma^{-1/2}\|_{op}^3\right)$$
$$\times \Big(\sum_{i,j=1}^m \left|\sigma_{ij} - \mathrm{Cov}(H^{(i)}, H^{(j)})\right| \vee \gamma_1 \vee \ldots \vee \gamma_6\Big),$$

where $\Sigma^{1/2}$ *is the positive definite matrix such that* $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$ *and* $\Sigma^{-1/2} := (\Sigma^{1/2})^{-1}$*.*

5.2.2 Proof of Theorem 3.2

The key idea of proving Theorem 3.2 is to bound the terms γ_i, $1 \leq i \leq 6$, appearing in Theorem 5.2 by combining properties of region of stabilization, Assumptions **(R1)**-**(R4)**, **(T)** and **(M)** with $p_0 = 4$. We start by noicing that by Hölder's inequality, for $q \in (0, 4 + p/2)$ and $\check{y}, \check{y}_1, \check{y}_2 \in \check{\mathbb{X}}$, we have

$$\mathbb{E}|D_{\check{y}} F_n^{(i)}|^q \leq (\mathbb{E}|D_{\check{y}} F_n^{(i)}|^{4+p/2})^{\frac{q}{4+p/2}} \mathbb{P}(D_{\check{y}} F_n^{(i)} \neq 0)^{\frac{4+p/2-q}{4+p/2}}, \tag{5.19}$$

and

$$\mathbb{E}|D^2_{\check{y}_1,\check{y}_2} F_n^{(i)}|^q \leq (\mathbb{E}|D^2_{\check{y}_1,\check{y}_2} F_n^{(i)}|^{4+p/2})^{\frac{q}{4+p/2}} \mathbb{P}(D^2_{\check{y}_1,\check{y}_2} F_n^{(i)} \neq 0)^{\frac{4+p/2-q}{4+p/2}}. \tag{5.20}$$

The following result, whose proof is immediate from definitions, reveals the connection between the cost functions $D_{\check{y}} F_n^{(i)}$, $D^2_{\check{y}_1,\check{y}_2} F_n^{(i)}$ and the score function $\xi_n^{(i)}$ for $i \in [m]$ and $n \geq 1$.

Lemma 5.8. *For* $\check{y}, \check{y}_1, \check{y}_2 \in \check{\mathbb{X}}$*,* $i \in [m]$ *and* $n \geq 1$*,*

$$D_{\check{y}} F_n^{(i)}(\mathcal{P}(\check{\lambda})) = \xi_n^{(i)}(\check{y}, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}}) + \sum_{\check{z} \in \mathcal{P}(\check{\lambda})} D_{\check{y}} \xi_n^{(i)}(\check{z}, \mathcal{P}(\check{\lambda})),$$

and

$$D^2_{\check{y}_1,\check{y}_2} F_n^{(i)}(\mathcal{P}(\check{\lambda})) = D_{\check{y}_1} \xi_n^{(i)}(\check{y}_2, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}_2}) + D_{\check{y}_2} \xi_n^{(i)}(\check{y}_1, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}_1})$$
$$+ \sum_{\check{z} \in \mathcal{P}(\check{\lambda})} D^2_{\check{y}_1,\check{y}_2} \xi_n^{(i)}(\check{z}, \mathcal{P}(\check{\lambda})).$$

The next lemma implies the cost operator $D_{\check{y}}$ vanishes if $\check{y}$ is outside the region of stabilization.

Lemma 5.9. *Let Assumptions (R1)-(R4) hold. For $\check{y}, \check{y}_1, \check{y}_2 \in \check{\mathbb{X}}$, $i \in [m]$ and $n \geq 1$, we have that if $\check{y} \notin R_n^{(i)}(\check{x}, \mathcal{P}(\check{\lambda}) + \delta_{\check{x}})$,*

$$D_{\check{y}}\xi_n^{(i)}(\check{x}, \mathcal{P}(\check{\lambda}) + \delta_{\check{x}}) = 0,$$

and if $\{\check{y}_1, \check{y}_2\} \subsetneq R_n^{(i)}(\check{x}, \mathcal{P}(\check{\lambda}) + \delta_{\check{x}})$,

$$D^2_{\check{y}_1,\check{y}_2}\xi_n^{(i)}(\check{x}, \mathcal{P}(\check{\lambda}) + \delta_{\check{x}}) = 0.$$

Proof. According to Bhattacharjee and Molchanov (2022, Lemma 5.3), for the marked Poisson process $\mathcal{P}_{n\check{g}}$ and all $i \in [m]$, the result follows. □

Lemma 5.10. *Let Assumptions (R1)-(R4) and (T) hold. For $y, y_1, y_2 \in \mathbb{X}$, we have*

$$\mathbb{P}(D_{\check{y}}F_n^{(i)} \neq 0) \leq \kappa_n^{(i)}(y) + g_n^{(i)}(y),$$

and

$$\mathbb{P}(D^2_{\check{y}_1,\check{y}_2}F_n^{(i)} \neq 0) \leq e^{-r_n^{(i)}(y_2,y_1)} + e^{-r_n^{(i)}(y_1,y_2)} + q_n^{(i)}(y_1, y_2).$$

Proof. According to the Mecke formula, (3.15), (3.18), Lemma 5.8 and Lemma 5.9, we have

$$\begin{aligned}\mathbb{P}(D_{\check{y}}F_n^{(i)}(\mathcal{P}(\check{\lambda})) \neq 0) &\leq \mathbb{P}(\xi_n^{(i)}(\check{y}, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}}) \neq 0) + \mathbb{E}\sum_{\check{z}\in\mathcal{P}(\check{\lambda})} \mathbb{1}_{D_{\check{y}}\xi_n^{(i)}(\check{z},\mathcal{P}(\check{\lambda}))\neq 0}\\ &\leq \kappa_n^{(i)}(y) + n\int_{\check{\mathbb{X}}} \mathbb{P}(D_{\check{y}}\xi_n^{(i)}(\check{z}, \mathcal{P}(\check{\lambda}) + \delta_{\check{z}}) \neq 0)\check{\mathbb{Q}}(d\check{z})\\ &\leq \kappa_n^{(i)}(y) + g_n^{(i)}(y),\end{aligned}$$

where the last inequality follows from the fact that $0 < \zeta = p/(40 + 10p) < 1$. Similarly, the Mecke formula, (3.12), (3.14), Lemma 5.8 and Lemma 5.9 yield

$$\begin{aligned}&\mathbb{P}(D_{\check{y}_1,\check{y}_2}F_n^{(i)}(\mathcal{P}(\check{\lambda})) \neq 0)\\ &\leq \mathbb{P}(D_{\check{y}_1}\xi_n^{(i)}(\check{y}_2, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}_2}) \neq 0) + \mathbb{P}(D_{\check{y}_2}\xi_n^{(i)}(\check{y}_1, \mathcal{P}(\check{\lambda}) + \delta_{\check{y}_1}) \neq 0)\\ &+ \mathbb{E}\sum_{\check{z}\in\mathcal{P}(\check{\lambda})} \mathbb{1}_{D^2_{\check{y}_1,\check{y}_2}\xi_n^{(i)}(\check{z},\mathcal{P}(\check{\lambda}))\neq 0}\\ &\leq e^{-r_n^{(i)}(y_2,y_1)} + e^{-r_n^{(i)}(y_1,y_2)} + n\int_{\check{\mathbb{X}}} \mathbb{P}(D_{\check{y}_1,\check{y}_2}\xi_n^{(i)}(\check{z}, \mathcal{P}(\check{\lambda}) + \delta_{\check{z}}) \neq 0)\check{\mathbb{Q}}(d\check{z})\\ &\leq e^{-r_n^{(i)}(y_2,y_1)} + e^{-r_n^{(i)}(y_1,y_2)} + q_n^{(i)}(y_1, y_2).\end{aligned}$$

□

Lemma 5.11. *Under Assumptions (R1)-(R4), (T) and (M) with $p_0 = 4$, there exits a constant $C_p \in (0, \infty)$ depending only on p such that for all $i \in [m]$, $n \geq 1$, $y \in \mathbb{X}$ and $\eta \in \mathbf{N}$ with $\eta(\check{\mathbb{X}}) \leq 1$, we have*

$$\begin{aligned}\mathbb{E}|D_{\check{y}}F_n^{(i)}(\mathcal{P}(\check{\lambda}) + \eta)|^{4+p/2} &\leq C_p\left(M_n^{(i)}(y)^{4+p/2} + h_n^{(i)}(y)(1 + g_n^{(i)}(y)^4)\right)\\ &\leq C_pG_n^{(i)}(y)^{4+p/2}.\end{aligned}$$

Proof of Lemma 5.11. The proof of the first inequality is an extension of a result by Bhattacharjee and Molchanov (2022, Lemma 5.5) to the marked Poisson process $\mathcal{P}(\check{\lambda})$, noting additionally that the intensity measure $n\check{\mathbb{Q}} = n\mathbb{Q} \otimes \mathbb{Q}_{\mathbb{M}}$ assumes a product form due to independent marks, and hence the marks can be integrated using the Cauchy-Schwarz inequality. The second inequality is stratghforward from Lemma 6.1. □

Now, we are in a position to prove Theorem 3.2.

Proof of Theorem 3.2. According to Theorem 5.2, d_2- and d_3-distances only involve γ_1, γ_2 and γ_3. We begin by bounding γ_1.

For $i \in [m]$, let $H^{(i)} := (\varrho_n^{(i)})^{-1}\bar{F}_n^{(i)}$. By Hölder inequality, for $i, j \in [m]$,

$$(\mathbb{E}(D_{\check{x}_1}F_n^{(j)})^2(D_{\check{x}_2}F_n^{(j)})^2)^{\frac{1}{2}} \leq (\mathbb{E}(D_{\check{x}_1}F_n^{(j)})^4)^{\frac{1}{4}}(\mathbb{E}(D_{\check{x}_2}F_n^{(j)})^4)^{\frac{1}{4}}, \tag{5.21}$$

and

$$(\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)})^2(D^2_{\check{x}_2,\check{x}_3}F_n^{(i)})^2)^{\frac{1}{2}} \leq (\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)})^4)^{\frac{1}{4}}(\mathbb{E}(D^2_{\check{x}_2,\check{x}_3}F_n^{(i)})^4)^{\frac{1}{4}}. \tag{5.22}$$

According to (5.19) and Lemma 5.11, we have

$$\begin{aligned}(\mathbb{E}(D_{\check{x}_1}F_n^{(j)})^4)^{\frac{1}{4}} &\leq (\mathbb{E}|D_{\check{x}_1}F_n^{(j)}|^{4+p/2})^{\frac{1}{4+p/2}}\mathbb{P}(D_{\check{x}_1}F_n^{(j)} \neq 0)^{\frac{p}{32+4p}} \\ &\leq C_p^{\frac{1}{4+p/2}}G_n^{(j)}(x_1).\end{aligned} \tag{5.23}$$

Similarly,

$$(\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)})^4)^{\frac{1}{4}} \leq (\mathbb{E}|D^2_{\check{x}_1,\check{x}_3}F_n^{(i)}|^{4+p/2})^{\frac{1}{4+p/2}}\mathbb{P}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)} \neq 0)^{\frac{p}{32+4p}}. \tag{5.24}$$

By definition,

$$\begin{aligned}D^2_{\check{x}_1,\check{x}_3}F_n^{(i)}(\mathcal{P}(\check{\lambda})) &= D_{\check{x}_1}F_n^{(i)}(\mathcal{P}(\check{\lambda}) + \delta_{\check{x}_3}) - D_{\check{x}_1}F_n^{(i)}(\mathcal{P}(\check{\lambda})) \\ &= D_{\check{x}_3}F_n^{(i)}(\mathcal{P}(\check{\lambda}) + \delta_{\check{x}_1}) - D_{\check{x}_3}F_n^{(i)}(\mathcal{P}(\check{\lambda})).\end{aligned}$$

Consequently, by Lemmas 6.1 and 5.11, we have

$$\begin{aligned}&\mathbb{E}|D^2_{\check{x}_1,\check{x}_3}F_n^{(i)}(\mathcal{P}(\check{\lambda}))|^{4+p/2} \\ &\leq 2^{3+p/2}(\mathbb{E}|D_{\check{x}_1}F_n^{(i)}(\mathcal{P}(\check{\lambda}) + \delta_{\check{x}_3})|^{4+p/2} + \mathbb{E}|D_{\check{x}_1}F_n^{(i)}(\mathcal{P}(\check{\lambda}))|^{4+p/2}) \\ &\leq 2^{4+p/2}C_pG_n^{(i)}(x_1)^{4+p/2},\end{aligned}$$

as well as

$$\mathbb{E}|D^2_{\check{x}_1,\check{x}_3}F_n^{(i)}|^{4+p/2} \leq 2^{4+p/2}C_pG_n^{(i)}(x_3)^{4+p/2}.$$

Therefore, according to (5.24) and Lemma 5.10, we obtain

$$\begin{aligned}(\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)})^4)^{\frac{1}{4}} &\leq 2C_p^{\frac{1}{4+p/2}}(G_n^{(i)}(x_1) \wedge G_n^{(i)}(x_3))\mathbb{P}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)} \neq 0)^{\frac{p}{32+4p}} \\ &\leq 2C_p^{\frac{1}{4+p/2}}(G_n^{(i)}(x_1) \wedge G_n^{(i)}(x_3))\Big(e^{-\frac{p}{32+4p}r_n^{(i)}(x_3,x_1)} + e^{-\frac{p}{32+4p}r_n^{(i)}(x_1,x_3)} \\ &\qquad + q_n^{(i)}(x_1,x_3)^{\frac{p}{32+4p}}\Big).\end{aligned} \tag{5.25}$$

Combining (5.23) and (5.25), and recalling (3.17), we obtain

$$\gamma_1^2 \leq 4C_p^{\frac{4}{4+p/2}} \sum_{i,j=1}^m (\varrho_n^{(i)})^{-2}(\varrho_n^{(j)})^{-2} n \int_{\mathbb{X}} \left(n \int_{\mathbb{X}} G_n^{(j)}(x_1) G_n^{(i)}(x_1) \left(e^{-\frac{p}{32+4p} r_n^{(i)}(x_3,x_1)} \right.\right.$$
$$\left.\left. + e^{-\frac{p}{32+4p} r_n^{(i)}(x_1,x_3)} + q_n^{(i)}(x_1,x_3)^{\frac{p}{32+4p}} \right) \mathbb{Q}(dx_1) \right)^2 \mathbb{Q}(dx_3)$$
$$\leq 4C_p^{\frac{4}{4+p/2}} \sum_{i,j=1}^m \frac{n\mathbb{Q}\left(f_{1,1,0,\beta}^{(i,j,i,i)} \right)^2}{(\varrho_n^{(i)} \varrho_n^{(j)})^2}. \tag{5.26}$$

Using (5.25) again, we obtain

$$\gamma_2^2 \leq 16C_p^{\frac{4}{4+p/2}} \sum_{i,j=1}^m \frac{n\mathbb{Q}\left(f_{1,1,0,\beta}^{(i,j,i,i)} \right)^2}{(\varrho_n^{(i)} \varrho_n^{(j)})^2}. \tag{5.27}$$

As for γ_3, note that by letting $q = 3$ in (5.19), we have for $i \in [m]$,

$$\mathbb{E}|D_{\check{x}} F_n^{(i)}|^3 \leq (\mathbb{E}|D_{\check{x}} F_n^{(i)}|^{4+p/2})^{\frac{3}{4+p/2}} \mathbb{P}(D_{\check{x}} F_n^{(i)} \neq 0)^{\frac{1+p/2}{4+p/2}}.$$

Arguing similarly as for (5.23), by Lemma 5.11 we have

$$\begin{aligned}\mathbb{E}|D_{\check{x}} F_n^{(i)}|^3 &\leq (\mathbb{E}|D_{\check{x}} F_n^{(j)}|^{4+p/2})^{\frac{3}{4+p/2}} \mathbb{P}(D_{\check{x}} F_n^{(i)} \neq 0)^{\frac{2+p}{8+p}} \\ &\leq C_p^{\frac{3}{4+p/2}} G_n^{(i)}(x)^3 \mathbb{P}(D_{\check{x}} F_n^{(i)} \neq 0)^{\frac{p}{8+p}} \\ &\leq C_p^{\frac{3}{4+p/2}} G_n^{(i)}(x)^3 (\kappa_n^{(i)}(x) + g_n^{(i)}(x))^{4\beta}.\end{aligned}$$

Therefore, we obtain

$$\gamma_3 \leq C_p^{\frac{3}{4+p/2}} \sum_{i=1}^m (\varrho_n^{(i)})^{-3} n \int_{\mathbb{X}} G_n^{(i)}(x)^3 (\kappa_n^{(i)}(x) + g_n^{(i)}(x))^{4\beta} \mathbb{Q}(dx)$$
$$\leq C_p^{\frac{3}{4+p/2}} \sum_{i=1}^m \frac{n\mathbb{Q}\left((\kappa_n^{(i)} + g_n^{(i)})^{4\beta} \left(G_n^{(i)} \right)^3 \right)}{(\varrho_n^{(i)})^3}. \tag{5.28}$$

Combining (5.26), (5.27) and (5.28) and invoking Theorem 5.2, the result follows. □

5.2.3 Proof of Theorem 3.3

Since $\mathsf{d}_{\mathsf{cvx}}$ is a distance with non-smooth test functions, it requires relatively stronger assumptions. We now take $p_0 = 6$ in (3.13). Again, we make use of Theorem 5.2 with $H^{(i)} := (\varrho_n^{(i)})^{-1} \bar{F}_n^{(i)}$ for $i \in [m]$. Note that the bound on $\mathsf{d}_{\mathsf{cvx}}$ involves three additional γ_i, $i = 4, 5, 6$ compared to the bound on d_2 and d_3. We will need a slightly modified version of Lemma 5.11 as stated below in Lemma 5.12, whose proof follows from that of Lemma 5.11 by observing that $6 + p/2 = 4 + (2 + p/2)$.

Lemma 5.12. *Under Assumptions **(R1)**-**(R4)**, **(T)** and **(M)** with $p_0 = 6$, there exits a constant C_p depending only on p such that for all $i \in [m]$, $n \geq 1$, $y \in \mathbb{X}$ and $\eta \in \boldsymbol{N}$ with $\eta(\check{\mathbb{X}}) \leq 1$, we have*

$$\mathbb{E}|D_{\check{y}} F_n^{(i)}(\mathcal{P}_{n\check{y}} + \eta)|^{6+p/2} \leq C_p \left(M_n^{(i)}(y)^{6+p/2} + h_n^{(i)}(y)(1 + g_n^{(i)}(y)^6) \right) \leq C_p G_n^{(i)}(y)^{6+p/2}.$$

Proof of Theorem 3.3. We start by bounding γ_i, $i = 1, 2, 3$ in a similar way as in the proof of Theorem 3.2. By considering similar (changing $p_0 = 4$ to $p_0 = 6$) Hölder inequalities as for (5.19), (5.20), (5.21) and (5.22) in the proof of Theorem 3.2 and using Lemma 5.12, we have for $i, j \in [m]$,

$$\begin{aligned}(\mathbb{E}(D_{\check{x}_1}F_n^{(j)})^4)^{\frac{1}{4}} &\leq (\mathbb{E}|D_{\check{x}_1}F_n^{(j)}|^{6+p/2})^{\frac{1}{6+p/2}}\mathbb{P}(D_{\check{x}_1}F_n^{(j)} \neq 0)^{\frac{4+p}{48+4p}} \\ &\leq C_p^{\frac{1}{6+p/2}} G_n^{(j)}(x_1),\end{aligned}$$

$$\begin{aligned}(\mathbb{E}(D^2_{\check{x}_1,\check{x}_3}F_n^{(i)})^4)^{\frac{1}{4}} &\leq 2C_p^{\frac{1}{6+p/2}}(G_n^{(i)}(x_1) \wedge G_n^{(i)}(x_3))\mathbb{P}(D_{\check{x}_1,\check{x}_3}F_n^{(i)} \neq 0)^{\frac{4+p}{48+4p}} \\ &\leq 2C_p^{\frac{1}{6+p/2}}(G_n^{(i)}(x_1) \wedge G_n^{(i)}(x_3))\Bigg(e^{-\frac{4+p}{48+4p}r_n^{(i)}(x_3,x_1)} + e^{-\frac{4+p}{48+4p}r_n^{(i)}(x_1,x_3)} \\ &\qquad + q_n^{(i)}(x_1,x_3)^{\frac{4+p}{48+4p}}\Bigg), \end{aligned} \tag{5.29}$$

and

$$\begin{aligned}\mathbb{E}|D_{\check{x}}F_n^{(i)}|^3 &\leq \mathbb{E}|D_{\check{x}}F_n^{(j)}|^{6+p/2})^{\frac{3}{6+p/2}}\mathbb{P}(D_{\check{x}}F_n^{(i)} \neq 0)^{\frac{6+p}{12+p}} \\ &\leq C_p^{\frac{3}{6+p/2}} G_n^{(i)}(x)^3\mathbb{P}(D_{\check{x}}F_n^{(i)} \neq 0)^{6\beta} \\ &\leq C_p^{\frac{3}{6+p/2}} G_n^{(i)}(x)^3(\kappa_n^{(i)}(x) + g_n^{(i)}(x))^{6\beta}.\end{aligned}$$

Combining all the bounds above and recalling (3.17), we obtain

$$\gamma_1^2, \gamma_2^2 \leq 16C_p^{\frac{4}{6+p/2}}\sum_{i,j=1}^m \frac{n\mathbb{Q}\left(f_{1,1,0,\beta}^{(i,j,i,i)}\right)^2}{(\varrho_n^{(i)}\varrho_n^{(j)})^2},$$

and

$$\gamma_3 \leq C_p^{\frac{3}{6+p/2}}\sum_{i=1}^m \frac{n\mathbb{Q}\left((\kappa_n^{(i)} + g_n^{(i)})^{6\beta}\left(G_n^{(i)}\right)^3\right)}{(\varrho_n^{(i)})^3}.$$

Next, we proceed to bounding the remaining terms γ_i, $i = 4, 5, 6$. First, we focus on γ_4. Appplying Hölder's inequality as before and using Lemma 5.12, we have

$$\begin{aligned}\mathbb{E}(D_{\check{x}}F_n^{(i)})^4 &\leq (\mathbb{E}|D_{\check{x}}F_n^{(i)}|^{6+p/2})^{\frac{4}{6+p/2}}\mathbb{P}(D_{\check{x}}F_n^{(i)} \neq 0)^{\frac{4+p}{12+p}} \\ &\leq C_p^{\frac{4}{6+p/2}} G_n^{(i)}(x)^4\mathbb{P}(D_{\check{x}}F_n^{(i)} \neq 0)^{6\beta} \\ &\leq C_p^{\frac{4}{6+p/2}} G_n^{(i)}(x)^4(\kappa_n^{(i)}(y) + g_n^{(i)}(x))^{6\beta}. \end{aligned} \tag{5.30}$$

Thus, by (5.29) and (5.30), we obtain

$$\gamma_4^2 \lesssim C_p^{\frac{4}{6+p/2}} m\sum_{i=1}^m \frac{n\mathbb{Q}\left((\kappa_n^{(i)} + g_n^{(i)})^{6\beta}\left(G_n^{(i)}\right)^4\right)}{(\varrho_n^{(i)})^4} + C_p^{\frac{4}{6+p/2}}\sum_{i,j=1}^m \frac{n\mathbb{Q}f_{2,2,0,3\beta}^{(i,j,l,i)}}{(\varrho_n^{(i)}\varrho_n^{(j)})^2}.$$

Next, we turn to γ_5 and γ_6. By Lemmas 6.1 and 5.12, we have

$$\begin{aligned}\mathbb{E}\|D_{\check{x}}\mathbf{F}_n\|^{6+p/2} &\le m^{\frac{4+p/2}{2}}\sum_{i=1}^m \mathbb{E}|D_{\check{x}}F_n^{(i)}|^{6+p/2}\\ &\le m^{\frac{4+p/2}{2}}C_p\sum_{i=1}^m G_n^{(i)}(x)^{6+p/2}, \end{aligned}\tag{5.31}$$

and

$$\begin{aligned}\mathbb{E}\|D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\|^{6+p/2} &\le m^{\frac{4+p/2}{2}}\sum_{i=1}^m \mathbb{E}|D^2_{\check{x}_1,\check{x}_3}F_n^{(i)}|^{6+p/2}\\ &\le m^{\frac{4+p/2}{2}}2^{6+p/2}C_p\sum_{i=1}^m \Big(G_n^{(i)}(x_1)\wedge G_n^{(i)}(x_3)\Big)^{6+p/2}. \end{aligned}\tag{5.32}$$

Then, by the Hölder inequality (noting that $\frac{p}{12+p}+2\times\frac{3}{24+2p}+2\times\frac{3}{12+p}+\frac{3}{12+p}=1$, the last power being for the factor one), we have

$$\begin{aligned}&\mathbb{E}\mathbb{1}_{D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0}}(\|D_{\check{x}_1}\mathbf{F}_n\|+\|D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\|)^{\frac34}\\ &\quad\times(\|D_{\check{x}_2}\mathbf{F}_n\|+\|D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\|)^{\frac34}|D_{\check{x}_1}F_n^{(i)}|^{\frac32}|D_{\check{x}_2}F_n^{(i)}|^{\frac32}\\ &\le \mathbb{E}\mathbb{1}_{D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0}}(\|D_{\check{x}_1}\mathbf{F}_n\|^{\frac34}+\|D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\|^{\frac34})\\ &\quad\times(\|D_{\check{x}_2}\mathbf{F}_n\|^{\frac34}+\|D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\|^{\frac34})|D_{\check{x}_1}F_n^{(i)}|^{\frac32}|D_{\check{x}_2}F_n^{(i)}|^{\frac32}\\ &\lesssim \mathbb{P}(D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0})^{\frac{p}{12+p}}\\ &\qquad\times\Big((\mathbb{E}\|D_{\check{x}_1}\mathbf{F}_n\|^{6+p/2})^{\frac{3}{4(6+p/2)}}+(\mathbb{E}\|D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\|^{6+p/2})^{\frac{3}{4(6+p/2)}}\Big)\\ &\qquad\times\Big((\mathbb{E}\|D_{\check{x}_2}\mathbf{F}_n\|^{6+p/2})^{\frac{3}{4(6+p/2)}}+(\mathbb{E}\|D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\|^{6+p/2})^{\frac{3}{4(6+p/2)}}\Big)\\ &\qquad\times(\mathbb{E}|D_{\check{x}_1}F_n^{(i)}|^{6+p/2})^{\frac{3}{2(6+p/2)}}(\mathbb{E}|D_{\check{x}_2}F_n^{(i)}|^{6+p/2})^{\frac{3}{2(6+p/2)}}.\end{aligned}$$

Plugging in (5.31) and (5.32) and by Lemma 5.12, we have

$$\begin{aligned}&\mathbb{E}\mathbb{1}_{D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0}}(\|D_{\check{x}_1}\mathbf{F}_n\|+\|D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\|)^{\frac34}\\ &\quad\times(\|D_{\check{x}_2}\mathbf{F}_n\|+\|D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\|)^{\frac34}|D_{\check{x}_1}F_n^{(i)}|^{\frac32}|D_{\check{x}_2}F_n^{(i)}|^{\frac32}\\ &\lesssim \mathbb{P}(D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0})^{\frac{p}{12+p}}\\ &\qquad\times m^{\frac34}C_p^{\frac{3}{2(6+p/2)}}\sum_{l=1}^m G_n^{(l)}(x_1)^{\frac34}\sum_{l=1}^m G_n^{(l)}(x_2)^{\frac34}\\ &\qquad\times C_p^{\frac{3}{(6+p/2)}}G_n^{(i)}(x_1)^{\frac32}G_n^{(i)}(x_2)^{\frac32}\\ &\le m^{\frac34}C_p^{\frac{9}{2(6+p/2)}}\mathbb{P}(D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n\neq\mathbf{0},D^2_{\check{x}_2,\check{x}_3}\mathbf{F}_n\neq\mathbf{0})^{\frac{p}{12+p}}G_n^{(i)}(x_1)^{\frac32}G_n^{(i)}(x_2)^{\frac32}\\ &\qquad\qquad\times\sum_{l=1}^m G_n^{(l)}(x_1)^{\frac34}\sum_{l=1}^m G_n^{(l)}(x_2)^{\frac34}.\end{aligned}\tag{5.33}$$

Also, we have by Lemma 5.12,

$$(\mathbb{E}|D_{\check{x}_1}F_n^{(j)}|^3|D_{\check{x}_2}F_n^{(j)}|^3)^{\frac13}\le(\mathbb{E}|D_{\check{x}_1}F_n^{(j)}|^{6+p/2})^{\frac{1}{6+p/2}}(\mathbb{E}|D_{\check{x}_2}F_n^{(j)}|^{6+p/2})^{\frac{1}{6+p/2}}$$

$$\leq C_p^{\frac{2}{6+p/2}} G_n^{(j)}(x_1) G_n^{(j)}(x_2). \tag{5.34}$$

Therefore, by writing $\gamma_5^3 := \gamma_{5.1} + \gamma_{5.2}$ for the first term and the second term in γ_5, we have by (5.33) and (5.34),

$$\gamma_{5.1} \lesssim \sqrt{m} C_p^{\frac{5}{6+p/2}} \sum_{i,j,l=1}^{m} \Big(\sum_{s=1}^{m} (\varrho_n^{(s)})^{-1} \Big) (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2}$$
$$\times n \int_{\mathbb{X}} \Big(n \int_{\mathbb{X}} G_n^{(i)}(x_1) G_n^{(j)}(x_1) G_n^{(l)}(x_1)^{\frac{1}{2}} \mathbb{P}(D^2_{\check{x}_1,\check{x}_3} \mathbf{F}_n \neq \mathbf{0})^{\frac{p}{36+3p}} \mathbb{Q}(dx_1) \Big)^2 \mathbb{Q}(dx_3), \tag{5.35}$$

where we have pulled the sum out of the integral, and used the fact that

$$\|D_{\check{x}}((\mathrm{P}_n^{-1}\mathbf{F}_n)\| = \Big(\sum_{i=1}^{m} (\varrho_n^{(i)})^{-2} (D_{\check{x}} F_n^{(i)})^2 \Big)^{\frac{1}{2}} \leq \bigg(\Big(\sum_{i=1}^{m} (\varrho_n^{(i)})^{-4} \Big)^{\frac{1}{2}} \Big(\sum_{i=1}^{m} (D_{\check{x}} F_n^{(i)})^4 \Big)^{\frac{1}{2}} \bigg)^{\frac{1}{2}}$$
$$\leq \Big(\sum_{i=1}^{m} (\varrho_n^{(i)})^{-2} \sum_{i=1}^{m} (D_{\check{x}} F_n^{(i)})^2 \Big)^{\frac{1}{2}} \leq \Big(\sum_{i=1}^{m} (\varrho_n^{(i)})^{-1} \Big) \|D_{\check{x}} \mathbf{F}_n\|,$$

which also applies to $D^2_{\check{x}_1,\check{x}_3}\mathbf{F}_n$. Moreover, note that

$$\mathbb{P}(D^2_{\check{x}_1,\check{x}_3} \mathbf{F}_n \neq \mathbf{0})^{\frac{p}{36+3p}} \leq \sum_{i=1}^{m} \mathbb{P}(D_{\check{x}_1,\check{x}_3} F_n^{(i)} \neq 0)^{\frac{p}{36+3p}}. \tag{5.36}$$

Consequently, according to (5.35) and (5.36), we have

$$\gamma_{5.1} \lesssim m^{\frac{3}{2}} C_p^{\frac{5}{6+p/2}} \sum_{i,j,l,t=1}^{m} \Big(\sum_{s=1}^{m} (\varrho_n^{(s)})^{-1} \Big) (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2}$$
$$\times n \int_{\mathbb{X}} \Big(n \int_{\mathbb{X}} G_n^{(i)}(x_1) G_n^{(j)}(x_1) G_n^{(l)}(x_1)^{\frac{1}{2}} \mathbb{P}(D^2_{\check{x}_1,\check{x}_3} F_n^{(t)} \neq 0)^{\frac{p}{36+3p}} \mathbb{Q}(dx_1) \Big)^2 \mathbb{Q}(dx_3),$$

where we have pulled the sum (over the probabilities) out of the integral resulting in the extra factor m by Lemma 6.1. Similar arguments also yield that

$$\gamma_{5.2} \lesssim m^{\frac{3}{2}} C_p^{\frac{5}{6+p/2}} \sum_{i,j,l,t=1}^{m} \Big(\sum_{s=1}^{m} (\varrho_n^{(s)})^{-1} \Big) (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2}$$
$$\times n \int_{\mathbb{X}} \Big(n \int_{\mathbb{X}} G_n^{(i)}(x_1) G_n^{(j)}(x_1) G_n^{(l)}(x_1)^{\frac{1}{2}} \mathbb{P}(D^2_{\check{x}_1,\check{x}_3} F_n^{(t)} \neq 0)^{\frac{p}{72+6p}} \mathbb{Q}(dx_1) \Big)^2 \mathbb{Q}(dx_3).$$

Combining the bounds on γ_{51} and γ_{52}, we obtain

$$\gamma_5^3 \lesssim m^{\frac{3}{2}} C_p^{\frac{5}{6+p/2}} \sum_{i,j,l,t=1}^{m} \Big(\sum_{s=1}^{m} (\varrho_n^{(s)})^{-1} \Big) (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2}$$
$$\times n \int_{\mathbb{X}} \Big(n \int_{\mathbb{X}} G_n^{(i)}(x_1) G_n^{(j)}(x_1) G_n^{(l)}(x_1)^{\frac{1}{2}} \mathbb{P}(D^2_{\check{x}_1,\check{x}_3} F_n^{(t)} \neq 0)^{\frac{p}{72+6p}} \mathbb{Q}(dx_1) \Big)^2 \mathbb{Q}(dx_3)$$
$$\lesssim m^{\frac{3}{2}} C_p^{\frac{5}{6+p/2}} \sum_{i,j,l,t=1}^{m} \sum_{s=1}^{m} \frac{n\mathbb{Q}\left(f_{1,1,1/2,\beta}^{(i,j,l,t)} \right)^2}{\varrho_n^{(s)} (\varrho_n^{(i)} \varrho_n^{(j)})^2}.$$

Similar to γ_5, one can derive

$$\begin{aligned}
\gamma_6^4 &\lesssim m^2 C_p^{\frac{6}{6+p/2}} \sum_{i,j,l,t=1}^m \Big(\sum_{s=1}^m (\varrho_n^{(s)})^{-1}\Big)^2 (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2} \\
&\quad \times n \int_{\mathbb{X}} \left(n \int_{\mathbb{X}} G_n^{(i)}(x_1) G_n^{(j)}(x_1) G_n^{(l)}(x_1) \mathbb{P}(D^2_{\check{x}_1,\check{x}_3} F_n^{(t)} \neq 0)^{\frac{p}{72+6p}} \mathbb{Q}(dx_1) \right)^2 \mathbb{Q}(dx_3) \\
&\lesssim m^2 C_p^{\frac{6}{6+p/2}} \sum_{i,j,l,t=1}^m \Big(\sum_{s=1}^m (\varrho_n^{(s)})^{-1}\Big)^2 (\varrho_n^{(i)})^{-2} (\varrho_n^{(j)})^{-2} n\mathbb{Q}\left(f_{1,1,1,\beta}^{(i,j,l,t)}\right)^2 \\
&\lesssim m^3 C_p^{\frac{6}{6+p/2}} \sum_{i,j,l,t=1}^m \sum_{s=1}^m \frac{n\mathbb{Q}\left(f_{1,1,1,\beta}^{(i,j,l,t)}\right)^2}{(\varrho_n^{(s)}\varrho_n^{(i)}\varrho_n^{(j)})^2},
\end{aligned}$$

where in the last step, we use the Lemma 6.1. Putting together all the bounds on γ_1 to γ_6 above yields the desired conclusion. □

5.2.4 Proofs of results in Section 3.2

Additional Notation: For $a \in \mathbb{R}$, $\mathbf{b} := (b^{(1)}, \dots, b^{(d)}) \in \mathbb{R}^d$ and $\psi : \mathbb{R} \to \mathbb{R}$, we write $a\mathbf{b} := (ab^{(1)}, \dots, ab^{(d)})$, $a + \mathbf{b} := (a + b^{(1)}, \dots, a + b^{(d)})$ and $\psi(\mathbf{b}) := (\psi(b^{(1)}), \dots, \psi(b^{(d)}))$. Let I be a subset of $[d]$ and denote by $x^I := (x^{(i)})_{i\in I}$ the subvector indexed by the set I. We also write $I^c = [d]\backslash I$.

Given a target point $x_0 \in \mathbb{R}^d$, recall from (3.2) the random forest estimator at x_0 associated to k-PNNs for some $k \geq 1$, which is given by

$$r_{n,k,w}(x_0) = \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} W_{n\boldsymbol{x}}(x_0) \mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)} \boldsymbol{y_x},$$

Particularly, for the case (3.4) with uniform weights, we have,

$$r_{n,k}(x_0) = \sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)} \boldsymbol{y_x},$$

where $\boldsymbol{y_x} := r(\boldsymbol{x}, \boldsymbol{\varepsilon_x})$. As mentioned briefly in Section 3.3, $r_{n,k,w}(x_0), r_{n,k}(x_0)$ can be viewed as sums of score functions of the marked Poisson process $\mathcal{P}_{n\breve{g}}$ with intensity measure $n\mathbb{Q} \otimes P_{\boldsymbol{\varepsilon}}$, where $\mathbb{Q}$ has an a.e. continuous density g on $\mathbb{R}^d$. Given a target point $x_0 \in \mathbb{R}^d$, we can consider the score function ξ_n associated to $r_{n,k,w}(x_0)$ given by

$$\xi_n(\check{x}, \eta) = W_{nx}(x_0, \eta) \mathbb{1}_{x\in\mathcal{L}_{n,k}(x_0,\eta)}\, y_x, \quad 0 \neq \eta \in \mathbf{N}, \check{x} \in \eta,$$

while in the special case of uniform weights in $r_{n,k}(x_0)$, the score function becomes

$$\xi_n(\check{x}, \eta) = \frac{\mathbb{1}_{x\in\mathcal{L}_{n,k}(x_0,\eta)}}{L_{n,k}(x_0, \eta)}\, y_x, \quad 0 \neq \eta \in \mathbf{N}, \check{x} \in \eta.$$

By the definition of k-PNNs, it is straightforward to see that the score functions above are region-stabilizing in the sense of (3.11) with the region of stabilization given by

$$R_n(\check{x}, \eta) := \begin{cases} \mathrm{Rect}(x_0, x) \times \mathbb{R}, & \text{if } \eta((\mathrm{Rect}(x_0, x)\backslash\{x\}) \times \mathbb{R}) < k, \\ \varnothing, & \text{otherwise} \end{cases} \tag{5.37}$$

for $\eta \in \mathbf{N}$ and $\check{x} \in \eta$. Therefore, we aim to apply the theorems in Section 3.3. Throughout this section, we shall omit λ_d in integrals and simply write dx instead of $\lambda_d(dx)$. It is straightforward to check that Assumptions (3.10) and **(R1)**-**(R4)** are satisfied for this score function and the region (5.37). For $\eta \in \mathbf{N}$ with $\eta(\mathbb{R}^d \times \mathbb{R}) \leq 9$ (since $p_0 = 6$ here), and $x \in \mathbb{R}^d$, recall that $\mathcal{P}_{x,\eta} := \mathcal{P}_{n\breve{g}} + \delta_{(x,\boldsymbol{\varepsilon}_x)} + \eta$. By independence, for any $p > 0$ we have

$$\left\| W_{nx}(x_0, \mathcal{P}_{x,\eta}) \mathbb{1}_{x \in \mathcal{L}_{n,k}(x_0, \mathcal{P}_{x,\eta})} y_x \right\|_{L_{6+p}} \leq \left(\mathbb{E}_{P_{\boldsymbol{\varepsilon}}} |r(\boldsymbol{\varepsilon}_x, x)|^{6+p} \right)^{1/(6+p)} \sup_{(x,\eta):|\eta| \leq 9} \| W_{nx}(x_0, \mathcal{P}_{x,\eta}) \|_{L_{6+p}},$$

and

$$\left\| \frac{\mathbb{1}_{x \in \mathcal{L}_{n,k}(x_0, \mathcal{P}_{x,\eta})}}{L_{n,k}(x_0, \mathcal{P}_{x,\eta})} y_x \right\|_{L_{6+p}} \leq \left(\mathbb{E}_{P_{\boldsymbol{\varepsilon}}} |r(\boldsymbol{\varepsilon}_x, x)|^{6+p} \right)^{\frac{1}{(6+p)}} \sup_{(x,\eta):|\eta| \leq 9} \left\| \frac{1}{L_{n,k}(x_0, \mathcal{P}_{x,\eta})} \right\|_{L_{6+p}}. \tag{5.38}$$

Since $L_{n,k}(x_0, \mathcal{P}_{x,\eta}) \geq 1$ for all $x \neq x_0$, Assumption **(M)** is satisfied for $x \neq x_0$ with $p_0 = 6$ and

$$M_n(x) := \Omega_n r^*_{6+p}(x), \tag{5.39}$$

where

$$\Omega_n := \begin{cases} \sup_{(x,\eta):|\eta| \leq 9} \| W_{nx}(x_0, \mathcal{P}_{x,\eta}) \|_{L_{6+p}} & \text{for } r_{n,k,w}(x_0), \\ \sup_{(x,\eta):|\eta| \leq 9} \| L_{n,k}(x_0, \mathcal{P}_{x,\eta})^{-1} \|_{L_{6+p}} & \text{for } r_{n,k}(x_0), \end{cases}$$

which does not depend on x, and $r^*_{6+p}(x) := \left(\mathbb{E}_{P_{\boldsymbol{\varepsilon}}} |r(\epsilon_x, x)|^{6+p} \right)^{1/(6+p)}$. Also, from (5.37), we have for $\check{x}, \check{y} \in \mathbb{R}^d \times \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(\check{y} \in R_n(\check{x}, \mathcal{P}_{n\breve{g}} + \delta_{\check{x}})) &= \sum_{j=0}^{k-1} \mathbb{1}_{y \in \mathrm{Rect}(x_0, x)} e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z) dz} \frac{\left(n \int_{\mathrm{Rect}(x,x_0)} g(z) dz \right)^j}{j!} \\ &= \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} \psi(n, k, x_0, x), \end{aligned} \tag{5.40}$$

where for notational convenience, for $n, k \geq 1$ and $x_0, x \in \mathbb{R}^d$, we define

$$\begin{aligned} \psi(n, k, x_0, x) &:= \mathbb{P}\left(\mathrm{Poi}\left(n \int_{\mathrm{Rect}(x_0,x)} g(z) dz \right) < k \right) \\ &= \sum_{j=0}^{k-1} e^{-n \int_{\mathrm{Rect}(x_0,x)} g(z) dz} \frac{\left(n \int_{\mathrm{Rect}(x,x_0)} g(z) dz \right)^j}{j!}. \end{aligned} \tag{5.41}$$

Noting that for any $t > 0$ and $0 \leq j \leq k-1$, $e^{-t} t^j \leq j!$, so that $e^{-(1-(j+2)^{-1})t}((1-(j+2)^{-1})t)^j \leq j!$, we obtain

$$e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z) dz} \frac{\left(n \int_{\mathrm{Rect}(x,x_0)} g(z) dz \right)^j}{j!} \leq (1 - (j+2)^{-1})^{-j} e^{-\frac{1}{j+2} n \int_{\mathrm{Rect}(x,x_0)} g(z) dz}.$$

Note that $(1-(j+2)^{-1})^{-j} \le e^{j/(j+2)} \le e$. Therefore, we can upper bound the probability in Assumption (T) as

$$\mathbb{P}(\check{y} \in R_n(\check{x}, \mathcal{P}_{n\check{g}} + \delta_{\check{x}})) \le e \sum_{j=0}^{k-1} e^{-\frac{1}{j+2} n \int_{\mathrm{Rect}(x,x_0)} g(z)dz} =: e^{-r_n(x,y)}, \tag{5.42}$$

when $y \in \mathrm{Rect}(x_0, x)$, while we take $r_n(x,y) = \infty$ otherwise. Therefore, Assumption (T) is also satisfied. It remains to estimate the quantities appearing in Theorems 3.2 and 3.3.

To this end, we first introduce a function that will play a key role in the estimation. For an a.e. continuous function $\phi(x) : \mathbb{R}^d \to \mathbb{R}_+$ with $\int_{\mathbb{R}^d} \phi(x)g(x)dx < \infty$, $\alpha, s > 0$, $d \ge 1$ and $x_0 \in \mathbb{R}^d$, define the function $c_{\alpha,s,x_0} : \mathbb{R}^d \to \mathbb{R}$ as

$$c_{\alpha,s,x_0}(y) := s \int_{\mathbb{R}^d} \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} e^{-\alpha s \int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx, \tag{5.43}$$

where we suppress the dependence on ϕ for ease of notation. Observe that c_{α,s,x_0} has the following scaling property: for all $\alpha, s > 0$,

$$c_{\alpha,s,x_0}(y) = \alpha^{-1} c_{1,\alpha s,x_0}(y). \tag{5.44}$$

Therefore, we will often take $\alpha = 1$ without loss of generality. While a similar function was also studied by Bhattacharjee and Molchanov (2022, Section 3) in the context of minimal points, it is important to emphasize here that we relax several assumptions made therein. For instance, we do not require a uniform density g on a compact set $[0,1]^d$, and consider instead an a.e. continuous density g on $\mathbb{R}^d$. The way we deal with such a general density is to divide the integral over $\mathbb{R}^d$ into one over a suitable compact hyperrectangle A which we choose to be a neighborhood of the point x_0, and another integral over its complement A^c. For the integral over A, we can apply similar arguments as done by Bhattacharjee and Molchanov (2022, Section 3), since up to finitely many rotations, translations and scalings, any hyperrectangle in $\mathbb{R}^d$ is "equivalent" to $[0,1]^d$. On A^c, we bound the integral over the coordinates that are within the neighborhood and those outside the neighborhood separately.

For $\epsilon > 0$, we write the m-dimensional vector $\boldsymbol{\epsilon} := (\epsilon, \dots, \epsilon)$. Note that we can choose a set $A \subseteq \mathbb{R}^d$ with $\mathbb{Q}(A) = 1$ such that for $x_0 \in A$ we have $g(x_0) > 0$ and $\phi(x), g(x)$ are continuous at x_0. For such an x_0 and $\delta \in (0, 1/2\, g(x_0))$, by continuity, there exists $\epsilon := \epsilon(x_0, \delta) > 0$ such that for $x \in \mathrm{Rect}(x_0 - \boldsymbol{\epsilon}, x_0 - \boldsymbol{\epsilon})$,

$$|\phi(x) - \phi(x_0)| < \delta \quad \text{and} \quad |g(x) - g(x_0)| < \delta. \tag{5.45}$$

Also recall that for $x \in \mathbb{R}^d$ and $j \in [d]$, we denote $x^{[j]} := (x^{(1)}, \dots, x^{(j)})$. For $j \in \{0\} \cup [d]$ let $\mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\bar{j}}) := x_0^{\bar{j}} + [-\epsilon, \epsilon]^j$, where $\bar{j}$ denotes a j-tuple with elements in $\{0\} \cup [d]$.

Lemma 5.13. *Let $x_0 \in A$ be as above and $\delta \in (0, 1/2\, g(x_0))$. Then there exists $\epsilon > 0$ such that for all $\alpha > 0$ and $n, d, k \ge 1$, we have*

$$c_{\alpha(k+1)^{-1},n,x_0}(y) \le n e^{-\alpha(k+1)^{-1} n (g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x)g(x)dx$$
$$+ \sum_{j=1}^{d} \frac{\Lambda_j D_j}{\alpha(g(x_0)-\delta)} \left(\frac{k+1}{\epsilon^{d-j}}\right) \sum_{\underline{j}} e^{-\alpha\left(\frac{n(g(x_0)-\delta)\epsilon^{d-j}}{k+1}\right)|y^{\underline{j}} - x_0^{\underline{j}}|/2}$$

$$\times \left(\left| \log \left(\alpha \left(\frac{n(g(x_0)-\delta)\epsilon^{d-j}}{k+1} \right) |y^{\underline{j}} - x_0^{\underline{j}}| \right) \right|^{j-1} + 1 \right) \mathbb{1}_{y^{\underline{j}} \in \mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})},$$

where $\sum_{\underline{j}}$ *denotes the sum over all of* j*-tuples in* $[d]$*, and for* $j \in [d]$*,* $D_j > 0$ *is a constant depending only on* j*, while* $\Lambda_j \equiv \Lambda(j,\epsilon) > 0$ *is a constant depending on* x_0*,* j*,* ϵ, ϕ *and* g*. In particular, we can take* $\Lambda_d \coloneqq (\phi(x_0)+\delta)(g(x_0)+\delta)$*.*

Proof. Fix $x_0 \in A$, $\delta \in (0, 1/2\ g(x_0))$. By arguments similar to Biau and Devroye (2010, Proof of Theorem 2.2), for $n > 0$,

$$\begin{aligned} c_{1,n,x_0}(y) &= n \int_{\mathbb{R}^d} \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x) g(x) dx \\ &= n \int_{\mathcal{C}_{\boldsymbol{\epsilon}}(x_0)} \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x) g(x) dx \\ &\quad + n \int_{\mathbb{R}^d \setminus \mathcal{C}_{\boldsymbol{\epsilon}}(x_0)} \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x) g(x) dx \\ &=: c_1(y) + c_2(y), \end{aligned}$$

where $\mathcal{C}_{\boldsymbol{\epsilon}}(x_0) = \mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{d}}) = \mathrm{Rect}(x_0 - \boldsymbol{\epsilon}, x_0 + \boldsymbol{\epsilon})$. We first bound $c_1(y)$. Fix $\epsilon \in (0, \epsilon(x_0, \delta))$, and denote $\Delta_d = (g(x_0) - \delta)^{\frac{1}{d}}$. Let $\Lambda_d \in (0, \infty)$, depending only on x_0, j, ϵ, ϕ and g, be such that ϕg is uniformly bounded by Λ_d on $\mathcal{C}_{\boldsymbol{\epsilon}}(x_0)$. In particular, By (5.45) we see that we can take $\Lambda_d \leq (\phi(x_0)+\delta)(g(x_0)+\delta)$. Thus we obtain

$$\begin{aligned} c_1(y) &\leq \Lambda_d n \int_{\mathcal{C}_{\boldsymbol{\epsilon}}(x_0)} \mathbb{1}_{y \in \mathrm{Rect}(x_0,x)} e^{-n(g(x_0)-\delta)|x-x_0|} dx \\ &= \Lambda_d n \int_{\mathrm{Rect}(-\boldsymbol{\epsilon},\boldsymbol{\epsilon})} \mathbb{1}_{y-x_0 \in \mathrm{Rect}(0,x)} e^{-n(g(x_0)-\delta)|x|} dx \\ &= \frac{\Lambda_d}{g(x_0)-\delta} n \int_{\mathrm{Rect}(-\Delta_d \boldsymbol{\epsilon}, \Delta_d \boldsymbol{\epsilon})} \mathbb{1}_{\Delta_d(y-x_0) \in \mathrm{Rect}(0,x)} e^{-n|x|} dx. \end{aligned}$$

First note that $c_1(y) = 0$ for $y \notin \mathcal{C}_{\boldsymbol{\epsilon}}(x_0)$, since the indicator $\mathbb{1}_{y \in \mathrm{Rect}(x_0,x)}$ in the first step is then always zero. Also note that the indicator $\mathbb{1}_{\Delta_d(y-x_0) \in \mathrm{Rect}(0,x)}$ enforces that x in the integral can only be in one of the 2^d orthants. Let $\mathrm{abs}(y) := (|y^i|)_{i \in [d]}$ denote the vector of absolute values of the coordinates of $y \in \mathbb{R}^d$. Then, by symmetry we obtain

$$\begin{aligned} c_1(y) &\leq \frac{\Lambda_d}{g(x_0)-\delta} n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} \mathbb{1}_{\mathrm{abs}(\Delta_d(y-x_0)) \in \mathrm{Rect}(0,x)} e^{-n|x|} dx \\ &\leq \frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta} n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} \mathbb{1}_{\mathrm{abs}(\Delta_d(y-x_0)) \in \mathrm{Rect}(0,x)} e^{-n|x|} dx. \end{aligned} \tag{5.46}$$

Next, we note the following inequality which is due to Bhattacharjee and Molchanov (2022, Lemma 3.1): for $\alpha > 0$ and $n \geq 1$, there exists a constant $D > 0$ depending only on d such that

$$n \int_{[0,1]^d} \mathbb{1}_{x \succ y} e^{-\alpha n|x|} dx \leq \frac{D}{\alpha} e^{-\alpha n|y|/2} (1 + |\log(\alpha n|y|)|^{d-1}). \tag{5.47}$$

Using the transformation $\tilde{x} = (\Delta_d \epsilon)^{-1} x$ in the first step and (5.47) in the second (replacing n by $n|\Delta_d \boldsymbol{\epsilon}|$ and taking $y = \epsilon^{-1}(y - x_0)$), from (5.44) we obtain

$$n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} \mathbb{1}_{\mathrm{abs}(\Delta_d(y-x_0)) \in \mathrm{Rect}(0,x)} e^{-n|x|} dx$$

$$
\begin{aligned}
&= n|\Delta_d \boldsymbol{\epsilon}| \int_{[0,1]^d} \mathbb{1}_{0 \prec \epsilon^{-1} \operatorname{abs}(y-x_0) \prec \tilde{x}} e^{-n|\Delta_d \boldsymbol{\epsilon}||\tilde{x}|} d\tilde{x} \\
&\leq D e^{-n|\Delta_d \boldsymbol{\epsilon}||\epsilon^{-1}(y-x_0)|/2} (1 + |\log(n|\Delta_d \boldsymbol{\epsilon}||\epsilon^{-1}(y-x_0)|)|^{d-1}) \\
&= D e^{-n|\Delta_d (y-x_0)|/2} \left(1 + |\log(n|\Delta_d (y-x_0)|)|^{d-1}\right). \qquad (5.48)
\end{aligned}
$$

To bound c_2, we argue similar to Biau and Devroye (2010, Proof of Theorem 2.2). Note that

$$
\mathbb{R}^d \backslash \mathcal{C}_{\boldsymbol{\epsilon}}(x_0) = \bigcup_{j=0}^{d-1} \mathcal{C}_j, \qquad (5.49)
$$

where, $\mathcal{C}_j$, $j \in \{0\} \cup [d-1]$ denotes the collection of all $y \in \mathbb{R}^d \backslash \mathcal{C}_{\boldsymbol{\epsilon}}(x_0)$ which have exactly j of the d coordinates within an ϵ-neighborhood of the corresponding coordinates of x_0. By symmetry, for each $j \in \{0\} \cup [d-1]$,

$$
\mathcal{C}_j = \bigcup_{\underline{j}} \mathcal{C}_{\underline{j}}, \qquad (5.50)
$$

where the index $\underline{j}$ runs over all $\binom{d}{j}$ possible j-tuples in $[d]$, and $\mathcal{C}_{\underline{j}} \equiv \mathcal{C}_{\underline{j}}^{x_0}$ denotes the collection of points for which the coordinates in $\underline{j}$ are within an ϵ-neighborhood of those coordinates of x_0. Denote the function

$$
(\phi g)_{\underline{j}}(x^{\underline{j}}) := \int_{\mathbb{R}^{d-j}} \phi(x) g(x) dx^{[d] \backslash \underline{j}}.
$$

For $\epsilon = \epsilon(x_0, \delta)$, note that for each $j \in \{0\} \cup [d-1]$, there exists $\Lambda_j \in (0, \infty)$ depending only on x_0, j, ϵ, ϕ and g such that the functions $(\phi g)_{\underline{j}}$ are uniformly bounded over $C_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})$ by Λ_j.
Considering the integral in c_2 over $\mathcal{C}_0$, by (5.45) we have

$$
\begin{aligned}
& n \int_{\mathcal{C}_0} \mathbb{1}_{y \in \operatorname{Rect}(x_0, x)} e^{-n \int_{\operatorname{Rect}(x, x_0)} g(z) dz} \phi(x) g(x) dx \\
& \leq n \int_{\mathcal{C}_0} \mathbb{1}_{y \in \operatorname{Rect}(x_0, x)} e^{-n \int_{\mathcal{C}_{\boldsymbol{\epsilon}}(x_0)} g(z) dz} \phi(x) g(x) dx \\
& \leq n e^{-n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x) g(x) dx. \qquad (5.51)
\end{aligned}
$$

Similarly, for $j \in [d-1]$, one may write

$$
\begin{aligned}
& n \int_{\mathcal{C}_j} \mathbb{1}_{y \in \operatorname{Rect}(x_0, x)} e^{-n \int_{\operatorname{Rect}(x, x_0)} g(z) dz} \phi(x) g(x) dx \\
& = s \sum_{\underline{j}} \int_{\mathcal{C}_{\underline{j}}} \mathbb{1}_{y \in \operatorname{Rect}(x_0, x)} e^{-n \int_{\operatorname{Rect}(x, x_0)} g(z) dz} \phi(x) g(x) dx \\
& \leq n \sum_{\underline{j}} \int_{\mathcal{C}_{\underline{j}}} \mathbb{1}_{y \in \operatorname{Rect}(x_0, x)} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l \in \underline{j}} |x^{(l)} - x_0^{(l)}|} \phi(x) g(x) dx \\
& \leq n \sum_{\underline{j}} \int_{\mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})} \mathbb{1}_{y^{\underline{j}} \in \operatorname{Rect}(x_0^{\underline{j}}, x^{\underline{j}})} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l \in \underline{j}} |x^{(l)} - x_0^{(l)}|} (\phi g)_{\underline{j}}(x^{\underline{j}}) dx^{\underline{j}},
\end{aligned}
$$

where in the final step, we have integrated $\phi(x)g(x)$ over the coordinates $[d] \backslash \underline{j}$. We again note that the integral inside the sum is zero when $y^{\underline{j}} \in \mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})^c$. Thus, the sum is zero when the number of coordinates $i \in [d]$ where $y^{(i)} \in [x_0^{(i)} - \epsilon, x_0^{(i)} + \epsilon]$ is less than j. Since each $(\phi g)_{\underline{j}}$ is uniformly bounded by Λ_j over $\mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})$, letting $\Delta_j := (g(x_0) - \delta)^{\frac{1}{j}}$, arguing sumilarly as for c_1, we obtain

$$\begin{aligned}
&n \int_{\mathcal{C}_{\underline{j}}} \mathbb{1}_{y \in \mathrm{Rect}(x_0, x)} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l \in \underline{j}} |x^{(l)} - x_0^{(l)}|} \phi(x) g(x) dx \\
&\le \Lambda_j n \int_{\mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})} \mathbb{1}_{y^{\underline{j}} \in \mathrm{Rect}(x_0^{\underline{j}}, x^{\underline{j}})} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l \in \underline{j}} |x^{(l)} - x_0^{(l)}|} dx^{\underline{j}} \\
&= \Lambda_j n \int_{[0,\epsilon]^j} \mathbb{1}_{\mathrm{abs}(y^{\underline{j}} - x_0^{\underline{j}}) \in \mathrm{Rect}(0, x^{\underline{j}})} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l \in \underline{j}} |x^{(l)}|} dx^{\underline{j}} \\
&= \frac{\Lambda_j}{(g(x_0) - \delta)\epsilon^{d-j}} n \int_{[0, \Delta_j \epsilon^{d/j}]^j} \mathbb{1}_{\mathrm{abs}(y^{\underline{j}} - x_0^{\underline{j}}) \in \mathrm{Rect}(0, \Delta_j^{-1} \epsilon^{-(d-j)/j} x^{\underline{j}})} e^{-n|x^{\underline{j}}|} dx^{\underline{j}}. \qquad (5.52)
\end{aligned}$$

Note that the integral in the last step in (5.52), when we take $j = d$, is exactly the same as the integral in (5.46). Therefore, a similar argument as used in bounding c_1 can be applied here, and we obtain for $j \in [d-1]$,

$$\begin{aligned}
&n \int_{\mathcal{C}_j} \mathbb{1}_{y \in \mathrm{Rect}(x_0, x)} e^{-n \int_{\mathrm{Rect}(x, x_0)} g(z) dz} \phi(x) g(x) dx \\
&\le \frac{\Lambda_j D_j}{(g(x_0) - \delta)\epsilon^{d-j}} \sum_{\underline{j}} e^{-n|\Delta_j \epsilon^{(d-j)/j}(y^{\underline{j}} - x_0^{\underline{j}})|/2} \left(|\log(n|\Delta_j \epsilon^{(d-j)/j}(y^{\underline{j}} - x_0^{\underline{j}})|)|^{j-1} + 1 \right), \qquad (5.53)
\end{aligned}$$

where the constant $D_j \in (0, \infty)$ depends only on j by (5.47). Now, combining (5.51) and (5.53), we have

$$\begin{aligned}
c_2(y) \le\ & n e^{-n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x) g(x) dx + \sum_{j=1}^{d-1} \frac{\Lambda_j D_j}{(g(x_0) - \delta)\epsilon^{d-j}} \sum_{\underline{j}} e^{-n|\Delta_j \epsilon^{(d-j)/j}(y^{\underline{j}} - x_0^{\underline{j}})|/2} \\
&\times \left(|\log(n|\Delta_j \epsilon^{(d-j)/j}(y^{\underline{j}} - x_0^{\underline{j}})|)|^{j-1} + 1 \right) \mathbb{1}_{y^{\underline{j}} \in \mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})}. \qquad (5.54)
\end{aligned}$$

Combining (5.46), (5.48) and (5.54), the result now follows by (5.44) upon replacing n as $\alpha(k+1)^{-1}n$. ☐

Remark 5.2. *We will often make use of this translation $y^{\underline{j}} - x_0^{\underline{j}}$ and the scaling $1/\epsilon$, as done in the proof of Lemma 5.13 above, in various similar integrals in later proofs: up to the term $ne^{-\alpha(k+1)^{-1}n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x) g(x) dx$, for $y \notin \mathcal{C}_{\boldsymbol{\epsilon}}(x_0)$, the terms in the sum in the bound in Lemma 5.13 are obtained from lower dimensional versions of c_{α,n,x_0}, where we upper bound the integral of ϕg over coordinates that are not within an ϵ neighbourhood of the corresponding coordinates of x_0. This approach enables us to make use of some results by Bhattacharjee and Molchanov (2022) in the setting where $\mathbb{X} = [0,1]^d$ and g is a uniform density.*

For $\alpha, s > 0$, $d \in \mathbb{N}$, define another function $\tilde{c}_{\alpha,s} : \mathbb{R}^d \to \mathbb{R}_+$ as

$$\tilde{c}_{\alpha,s}(x_0) := s \int_{\mathbb{R}^d} e^{-\alpha s \int_{\mathrm{Rect}(x, x_0)} g(z) dz} \phi(x) g(x) dx,$$

where the function $\phi(x) : \mathbb{R}^d \to \mathbb{R}$ as before is a.e. continuous with $\int_{\mathbb{R}^d} \phi(x) g(x) dx < \infty$. As for c_{α,s,x_0}, the function $\tilde{c}_{\alpha,s}$ also satisfies a scaling property. Consequently, we will often take $\alpha = 1$. Below we use the notation $\mathcal{O}_\delta$ to mean that the constant in the $\mathcal{O}$ term may depend on δ (it may also depend on other parameters and functions, such as α, g, ϕ, x_0, which remain fixed for us).

Lemma 5.14. *Under the same setting of Lemma 5.13, for all $n, d \geq 1$ and $\alpha > 0$, we have*

$$\tilde{c}_{\alpha(k+1)^{-1},n}(x_0) \leq ne^{-\alpha(k+1)^{-1}n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x)g(x)dx + \sum_{j=1}^{d} \frac{\binom{d}{j}\Lambda_j}{\alpha(g(x_0)-\delta)} \left(\frac{k+1}{\epsilon^{d-j}}\right) \mathcal{O}\left(\log^{j-1}\left(\frac{n(g(x_0)-\delta)\epsilon^d}{k+1}\right)\right),$$

where Λ_j for $j \in [d]$ is as in Lemma 5.13. Moreover, for fixed $0 < \delta < 1/2\ g(x_0)$, we have

$$\tilde{c}_{\alpha(k+1)^{-1},n}(x_0) = (k+1)\mathcal{O}_\delta\left(\log^{d-1}\left(\frac{n}{k+1}\right)\right). \tag{5.55}$$

Proof. We follow a very similar argument as in the proof of Lemma 5.13 and also use notation introduced there. For $s > 0$, write

$$\begin{aligned}
\tilde{c}_{1,n}(x_0) &= n \int_{\mathbb{R}^d} e^{-n\int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx \\
&= n \int_{\mathcal{C}_\epsilon(x_0)} e^{-n\int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx \\
&\quad + n \int_{\mathbb{R}^d \setminus \mathcal{C}_\epsilon(x_0)} e^{-n\int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx \\
&:= \tilde{c}_1(x_0) + \tilde{c}_2(x_0).
\end{aligned}$$

Arguing as for bounding c_1 in the proof of Lemma 5.13, with Λ_j as therein, we have

$$\begin{aligned}
\tilde{c}_1(x_0) &\leq 2^d \frac{\Lambda_d}{g(x_0)-\delta} n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} e^{-n|x|}dx \\
&\leq 2^d \frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta} n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} e^{-n|x|}dx,
\end{aligned} \tag{5.56}$$

where the additional 2^d is due to identical integrals over the 2^d different orthants. It is well known that (see, e.g., Bai et al., 2005) for $n \geq 1$ and $d \in \mathbb{N}$.

$$n \int_{[0,1]^d} e^{-b|x|}dx = \mathcal{O}(\log^{d-1} n),$$

Similar to the proof of Lemma 5.13, by the transformation $\tilde{x} = (\Delta_d \epsilon)^{-1}x$, we obtain

$$n \int_{\mathbf{0} \prec x \prec \Delta_d \boldsymbol{\epsilon}} e^{-n|x|}dx = \mathcal{O}(\log^{d-1}(n\Delta_d^d \epsilon^d)) = \mathcal{O}_\delta(\log^{d-1} n). \tag{5.57}$$

Next, we bound $\tilde{c}_2(x_0)$. Note using the same notation as in (5.49) and (5.50), arguing similarly as in (5.51) we have

$$n \int_{\mathcal{C}_0} e^{-n\int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx \leq ne^{-n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x)g(x)dx. \tag{5.58}$$

For $j \in [d-1]$, the argument for $\mathcal{C}_j$ also mimics the same in Lemma 5.13. As in there, one may write upon integrating $\phi(x)g(x)$ over the coordinates $[d]\backslash \underline{j}$ and bounding $(\phi g)_{\underline{j}}$ by Λ_j and letting $\Delta_j := (g(x_0)-\delta)^{\frac{1}{j}}$,

$$n \int_{\mathcal{C}_j} e^{-n\int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x)g(x)dx$$

$$\leq n \sum_{\underline{j}} \int_{\mathcal{C}_\epsilon(x_0^{\underline{j}})} e^{-n(g(x_0)-\delta)\epsilon^{d-j} \prod_{l\in \underline{j}} |x^{(l)}-x_0^{(l)}|} (\phi g)_{\underline{j}}(x^{\underline{j}}) dx^{\underline{j}}$$

$$\leq 2^j \Lambda_j n \sum_{\underline{j}} \int_{[0,\epsilon]^j} e^{-n(g(x_0)-\delta)\epsilon^{d-j}|x^{\underline{j}}|} dx^{\underline{j}}$$

$$\leq \frac{2^j \binom{d}{j} \Lambda_j}{(g(x_0)-\delta)\epsilon^{d-j}} n \int_{[0,\Delta_j \epsilon^{d/j}]^j} e^{-n|x^{[j]}|} dx^{[j]}. \tag{5.59}$$

Note that the integral on the last inequality (5.59) is exactly the same as the integral in (5.56) when $j = d$. Therefore, arguing as for bounding $\tilde{c}_1$ above, we have for $j = 1, \ldots, d-1$,

$$n \int_{\mathcal{C}_j} e^{-n \int_{\mathrm{Rect}(x,x_0)} g(z)dz} \phi(x) g(x) dx \leq \frac{2^j \binom{d}{j} \Lambda_j}{(g(x_0)-\delta)\epsilon^{d-j}} \mathcal{O}(\log^{j-1}(n \Delta_j^j \epsilon^d)). \tag{5.60}$$

Combining (5.58) and (5.60), we have

$$\tilde{c}_2(x_0) \leq n e^{-n(g(x_0)-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x) g(x) dx + \sum_{j=1}^{d-1} \frac{2^j \binom{d}{j} \Lambda_j}{(g(x_0)-\delta)\epsilon^{d-j}} \mathcal{O}(\log^{j-1}(n \Delta_j^j \epsilon^d)). \tag{5.61}$$

The result now follows from noting that $\tilde{c}_{\alpha(k+1)^{-1},n}(x_0) = (k+1)\alpha^{-1} \tilde{c}_{1,\alpha(k+1)^{-1}n}(x_0)$, by replacing n as $\alpha(k+1)^{-1}n$ and combining the bounds in (5.56), (5.57) and (5.61). □

Denote by $I_+ \subseteq [d]$ the coordinates $i \in [d]$ with $x^{(i)} \geq 0$, so that for $j \in [d] \backslash I_+$, $x^{(j)} < 0$. Define $\mathrm{Rect}(x, \partial\mathbb{R}^d) := \prod_{i \in I_+, j \in I_+^c} (-\infty, x^{(j)}] \times [x^{(i)}, \infty)$ as the "hyperrectangle" defined by x and the boundary of $\mathbb{R}^d$. Let $A_{x_0}(x)$ denote the set of points in $\mathbb{R}^d$ which are in the same orthant as $x \in \mathbb{R}^d$ w.r.t. x_0. For $x_1, x_2 \in \mathbb{R}^d$ with $x_2 \in A_{x_0}(x_1)$ (i.e., x_1 and x_2 are in the same orthant w.r.t. x_0), denote by $(x_1 \vee x_2)_{x_0}$ the unique point in $\mathrm{Rect}(x_1, \partial\mathbb{R}^d) \cap \mathrm{Rect}(x_2, \partial\mathbb{R}^d) \neq \varnothing$ having the minimal distance to x_0. In particular, when $x_0 = \mathbf{0}$ and $x_0 \prec x_1, x_2$, we have $(x_1 \vee x_2)_{x_0} = x_1 \vee x_2$.

In the setting when g is uniform on $\mathbb{X} = [0,1]^d$ and $\phi \equiv 1$ on $\mathbb{X}$, all three bounds in the following result follow according to Bhattacharjee and Molchanov (2022, Lemma 3.2). Here, we extend these results to the general setting we consider.

Lemma 5.15. *For all $i \in \mathbb{N}$, $\alpha, t > 0$, $n, d \geq 2$ and x_0, δ as in Lemma 5.13, when $k < n-1$, we have*

$$n \int_{\mathbb{R}^d} c_{\alpha(k+1)^{-1},n,x_0}(y)^t \phi(y) g(y) dy = (k+1)^{t+1} \mathcal{O}_\delta \left(\log^{d-1} \left((k+1)^{-1} n \right) \right), \tag{5.62}$$

$$n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_0}(y)} c_{\alpha(k+1)^{-1},n,x_0}((x \vee y)_{x_0}) \phi(x) g(x) dx \right)^i g(y) dy$$
$$= (k+1)^{2i+1} \mathcal{O}_\delta \left(\log^{d-1} \left((k+1)^{-1} n \right) \right),$$

$$n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_0}(y)} e^{-\alpha(k+1)^{-1} n \int_{Rect(x_0,(x \vee y)_{x_0})} g(z)dz} \phi(x) g(x) dx \right)^i g(y) dy$$
$$= (k+1)^{i+1} \mathcal{O}_\delta \left(\log^{d-1} \left((k+1)^{-1} n \right) \right). \tag{5.63}$$

Proof. The arguments employed to prove the bounds are similar to that used by Bhattacharjee and Molchanov (2022, Lemma 3.2), upon using the approach outlined in Remark 5.2, and demonstrated in the proof of Lemma 5.13. Therefore, we will only give a very brief outline of the proofs here by using notation introduced in the proofs above. We start by proving the first two bounds. Note that for all $t > 0$, $n \geq 1$, $\mathbf{j} \in [\mathbf{n}]$ and $\alpha > 0$, we trivially have

$$n^{t+1} e^{-j\alpha(k+1)^{-1} n(g(x_0)-\delta)\epsilon^d} = (k+1)^{t+1} \mathcal{O}_\delta \left(\log^{d-1}((k+1)^{-1} n) \right).$$

On the other hand, as argued in Remark 5.2, for the terms in the sum in the upper bound in Lemma 5.13, for each $\underline{j}$ with $y^{\underline{j}} \in \mathcal{C}_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})$, we can first integrate over the other $d - j$ coordinates and then upper bound $(\phi g)_{\underline{j}}$ uniformly over $C_{\boldsymbol{\epsilon}}(x_0^{\underline{j}})$ by Λ_j, and finally arguing as in the proof of (Bhattacharjee and Molchanov, 2022, Equation (3.10) in Lemma 3.2) with s there replaced by $\frac{n(g(x_0)-\delta)\epsilon^{d-j}}{k+1}$, we can obtain an upper bound for the integrals of each of these summands. Since $\int_{\mathbb{R}^d} \phi(x) g(x) dx < \infty$, the first conclusion follows by a simple application of Lemma 6.1. The second conclusion now also follows arguing exactly as in (Bhattacharjee and Molchanov, 2022, Equation (3.12) in Lemma 3.2). Arguing similarly using Lemma 5.14 instead of Lemma 5.13, the last bound also follows mimicking arguments in the proof of (Bhattacharjee and Molchanov, 2022, Equation (3.11) in Lemma 3.2). □

Before we proceed to more results related to the function c_{α,n,x_0}, which serves as an upper bound on the probability (5.40) that a point $\check{y}$ is in the region of stabilization of another point $\check{x}$, we present the following lemma providing a lower bound to this probability. The result indeed shows that the upper bounds in Lemma 5.15 which are polynomial in k are tight by our method.

Lemma 5.16. *Let ϕ be bounded from below by a constant $C_\phi > 0$. Then for $t > 0$, there exists a constant $C_{low} > 0$ depending only on d, α and t such that for $y \in Rect(x_0, x)$, when $k \leq 2n$, we have:*

$$n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} \left(\mathbb{P}(\check{y} \in R_n(\check{x}, \mathcal{P}_{n\check{g}} + \delta_{\check{x}})) \right)^\alpha \phi(x) g(x) dx \right)^t \phi(y) g(y) dy \geq C_{low} k^{t+1}.$$

Proof. Since by (5.45), the density g can be lower bounded by a positive constant in some rectangle around x_0, without loss of generality, we consider the density $g = \mathbb{1}_{[0,1]^d}$ and $x_0 = \mathbf{0}$. Thus from (5.40) we have

$$\begin{aligned} & n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} \left(\mathbb{P}(\check{y} \in R_n(\check{x}, \mathcal{P}_{n\check{g}} + \delta_{\check{x}})) \right)^\alpha \phi(x) g(x) dx \right)^t \phi(y) g(y) dy \\ & = n \int_{[0,1]^d} \left(n \int_{[0,1]^d} \mathbb{1}_{y \in \mathrm{Rect}(\mathbf{0},x)} \psi(n,k,\mathbf{0},x)^\alpha \phi(x) dx \right)^t \phi(y) dy \\ & \geq C_\phi^{t+1} n \int_{[0,1]^d} \left(n \int_{[0,1]^d} \mathbb{1}_{y \in \mathrm{Rect}(\mathbf{0},x)} \psi(n,k,\mathbf{0},x)^\alpha dx \right)^t dy, \end{aligned}$$

where $\psi(n, k, \mathbf{0}, x)$ is defined at (5.41). Note that for g being uniform and $x_0 = \mathbf{0}$, the parameter of the Possion distribution in (5.41) simply becomes $n|x|$. By Poisson concentration (see Lemma 6.2), we have for x with $n|x| \leq \frac{1}{2}k$,

$$\psi(n, k, \mathbf{0}, x) \geq 1 - e^{-\frac{k}{8}}.$$

Therefore, by restricting the integral over y to region $A := \{(k/(4n))^{1/d}\mathbf{1} \prec x \prec (k/(2n))^{1/d}\mathbf{1}\}$, and noting that the volume of this region is larger than $(1-d(2^{1/d}-1))k/(4n)$, we obtain the lower bound

$$\begin{aligned} & n\int_{[0,1]^d}\left(n\int_{[0,1]^d}\mathbb{1}_{y\in \mathrm{Rect}(\mathbf{0},x)}\psi(n,k,\mathbf{0},x)^{\alpha}dx\right)^t dy \\ & \geq n\int_{[0,(k/4n)^{1/d}]^d}\left(n\int_A \psi(n,k,\mathbf{0},x)^{\alpha}dx\right)^t dy \\ & \geq (1-e^{-\frac{k}{8}})^{\alpha t}(1-d(2^{1/d}-1))^t\frac{1}{4^{t+1}}k^{t+1}. \end{aligned}$$

□

Remark 5.3. *By Lemmas 5.15 and 5.16, it follows that the bound on the double integral therein of the tail probability $\mathbb{P}(\check{y}\in R_n(\check{x},\mathcal{P}_{n\check{g}}+\delta_{\check{x}}))$ of the region of stabilization is tight in k, i.e, the rate k^{j+1} cannot be improved by our method. This is due to the fact that the Poisson distribution concentrates around its mean. This along with (5.42) results in a tail bound that is exponential decaying in n, at the cost of having a polynomial growth in k.*

Lemma 5.17. *For $\alpha_1,\alpha_2>0$, $0<\zeta<\beta$, $i\in\mathbb{N}$, $j\in\{1,2\}$ and $n\geq 2$,*

$$\begin{aligned} n\int_{\mathbb{R}^d}&\left(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_0}(y)}c_{\alpha_1(k+1)^{-1},n,x_0}(x)^i e^{-\alpha_2(k+1)^{-1}n\int_{Rect(x_0,(x\vee y)_{x_0})}g(z)dz}g(x)dx\right)^j g(y)dy \\ &= (k+1)^{(i+1)j+1}\mathcal{O}_\delta\left(\log^{d-1}\left((k+1)^{-1}n\right)\right), \end{aligned}$$

$$\begin{aligned} n\int_{\mathbb{R}^d}&\left(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_0}(y)}c_{\zeta(k+1)^{-1},n,x_0}(x)^i c_{\beta(k+1)^{-1},n,x_0}((x\vee y)_{x_0})g(x)dx\right)^j g(y)dy \\ &= (k+1)^{(i+2)j+1}\mathcal{O}_\delta\left(\log^{d-1}\left((k+1)^{-1}n\right)\right). \end{aligned}$$

Proof. The proof of the first bound for $j=2$ is similar to the proof of Lemma 5.15 by applying Lemma 3.4 (rather than Lemma 3.2) in Bhattacharjee and Molchanov (2022) and replacing s by $\frac{n(g(x_0)-\delta)\epsilon^{d-j'}}{k+1}$ for $j'\in[n]$. The bounds of A_2 from the proof of Lemma 3.5 in Bhattacharjee and Molchanov (2022), yield the second bound in Lemma 5.17 for $j=2$. For $j=1$, the desired two bounds follow by mimicking the derivation of the bounds of A_1 and A_2, respectively, in the proof of Lemma 3.3 in Bhattacharjee and Molchanov (2022) with s replaced by $\frac{n(g(x_0)-\delta)\epsilon^{d-j'}}{k+1}$, $j'\in[n]$. □

Lemma 5.18. *For $\alpha>0$, $p>0$, $i\in\{1,2\}$ and $n\geq 2$,*

$$\begin{aligned} n\int_{\mathbb{R}^d}&\left(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in Rect(x_0,y)}e^{-\alpha(k+1)^{-1}n\int_{Rect(y,x_0)}g(z)dz}\phi(x)g(x)dx\right)^i g(y)dy \\ &= (k+1)^{i+1}\mathcal{O}_\delta\left(\log^{d-1}\left((k+1)^{-1}n\right)\right). \end{aligned}$$

Proof. Note that when $x\in \mathrm{Rect}(x_0,y)$, we have $(x\vee y)_{x_0}=y$. Thus, replacing $\mathrm{Rect}(y,x_0)$ by $\mathrm{Rect}((x\vee y)_{x_0},x_0)$ in the exponent in the integral in Lemma 5.18 and dropping the indicator $\mathbb{1}_{x\in \mathrm{Rect}(x_0,y)}$, the double integral is upper bounded by the integral in (5.63), and hence the result follows by invoking Lemma 5.15. □

Recall the definiton (5.41) of the c.d.f. of the Poisson distribution. In the following lemma, we show that the function $\psi(n,k,x_0,x)$ has a localizing effect, "forcing" the integral of the product $\psi(n,k,x_0,x)\phi(x)g(x)$, where ϕ is an a.e. continuous and integrable function, to converge to $\phi(x_0)$ with the rate $k\log^{d-1}n$.

Lemma 5.19. *Under the setting of Lemma 5.13, for $n,d \geq 2$, $k \geq 1$ and $k = O(n^\alpha)$ with $0 < \alpha < 1$, we have*

$$\begin{aligned}
&\frac{2^d}{(d-1)!}\frac{(\phi(x_0)-\delta)(g(x_0)-\delta)}{g(x_0)+\delta}k\mathcal{O}_\delta(\log^{d-1}n) \\
&\leq n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx \\
&\leq \frac{2^d}{(d-1)!}\frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta}k\mathcal{O}_\delta(\log^{d-1}n).
\end{aligned}$$

Particularly, taking $\phi(x)\equiv 1$, there exist constants $C_1 > C_2 > 0$ (depending on the parameters δ,ϵ and g,x_0) such that

$$C_2 k\log^{d-1}n \leq \mathbb{E}L_{n,k}(x_0) \leq C_1 k\log^{d-1}n.$$

Proof. Arguing as in the proof of Lemma 5.14, we have

$$\begin{aligned}
&n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx \\
&=\sum_{j=0}^{k-1}n\int_{\mathcal{C}_\epsilon(x_0)}e^{-n\int_{\mathrm{Rect}(x_0,x)}g(z)dz}\frac{\left(n\int_{\mathrm{Rect}(x_0,x)}g(z)dz\right)^j}{j!}\phi(x)g(x)dx \\
&\quad+\sum_{j=0}^{k-1}n\int_{\mathbb{R}^d\setminus\mathcal{C}_\epsilon(x_0)}e^{-n\int_{\mathrm{Rect}(x_0,x)}g(z)dz}\frac{\left(n\int_{\mathrm{Rect}(x_0,x)}g(z)dz\right)^j}{j!}\phi(x)g(x)dx \\
&\leq\sum_{j=0}^{k-1}n\int_{\mathcal{C}_\epsilon(x_0)}e^{-n(g(x_0)-\delta)|x-x_0|}\frac{(n(g(x_0)-\delta)|x-x_0|)^j}{j!}\phi(x)g(x)dx \\
&\quad+\sum_{j=0}^{k-1}n\int_{\mathbb{R}^d\setminus\mathcal{C}_\epsilon(x_0)}e^{-n(g(x_0)-\delta)|x-x_0|}\frac{(n(g(x_0)-\delta)|x-x_0|)^j}{j!}\phi(x)g(x)dx \\
&=:\sum_{j=0}^{k-1}(e_{j,1}+e_{j,2}),
\end{aligned}$$

where the inequality follows due to the decresingness of the c.d.f. of Poisson distribution with respect to the Poisson parameter.

We first consider $e_{j,1}$ for $0 \leq j \leq k-1$. Note by (5.45) that

$$\begin{aligned}
e_{j,1} &\leq (\phi(x_0)+\delta)(g(x_0)+\delta) \\
&\qquad\times n\int_{\mathcal{C}_\epsilon(x_0)}e^{-n(g(x_0)-\delta)|x-x_0|}\frac{(n(g(x_0)-\delta)|x-x_0|)^j}{j!}dx \\
&= 2^d\frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta}n\int_{\mathbf{0}\prec x\prec\Delta_d\boldsymbol{\epsilon}}e^{-n|x|}\frac{(n|x|)^j}{j!}dx.
\end{aligned}$$

Now, with similar calculations as in Bai et al. (2005), we have

$$
\begin{aligned}
& n \int_{[0,1]^d} e^{-n|x|} \frac{(n|x|)^j}{j!} dx \\
&= \frac{1}{j!} \int_{[0,n^{\frac{1}{d}}]^d} e^{-|u|} |u|^j du \qquad (x = n^{-1/d} u) \\
&= \frac{1}{j!} \int_{[-d^{-1}\log n, \infty)^d} \exp\left\{ e^{-\sum_{j=1}^d z_j} - (j+1) \sum_{j=1}^d z_j \right\} dz \qquad (z_j = -\log u_j) \\
&= \frac{1}{j!(d-1)!} \int_{-\log n}^{\infty} (\log n + x)^{d-1} \exp(-(j+1)x - e^{-x}) dx \qquad \left(x = \sum_{j=1}^d z_j \right) \\
&= \frac{1}{j!(d-1)!} \int_0^n (\log n - \log y)^{d-1} e^{-y} y^j dy \qquad (y = e^{-x}) \\
&= \frac{\log^{d-1} n}{j!(d-1)!} \sum_{0 \le l \le d-1} \binom{d-1}{l} \frac{(-1)^l}{\log^l n} \int_0^{\infty} (\log^l y) e^{-y} y^j dy + \mathcal{O}(e^{-n} n^j \log^{d-1} n) \\
&= \frac{\log^{d-1} n}{(d-1)!} + \mathcal{O}(\log j \log^{d-2} n),
\end{aligned}
$$

where, noting that $\int_0^\infty (\log^l y) e^{-y} \frac{y^j}{j!} dy$ is the l-th moment of $\log X$ with X following a gamma distribution Gamma$(j+1, 1)$ for $0 < l < d$, according to the moment generating function of log-Gamma distribution, we have

$$
\int_0^\infty (\log^l y) e^{-y} \frac{y^j}{j!} dy = \frac{\Gamma^{(l)}(j+1)}{\Gamma(j+1)} = \mathcal{O}(\log^l j),
$$

with $\Gamma(\cdot)$ as the gamma function and $\Gamma^{(l)}(\cdot)$ denotes its l-th derivative.

Therefore, with the transformation $\tilde{x} = (\Delta_d \epsilon)^{-1} x$, we obtain

$$
\begin{aligned}
\sum_{j=0}^{k-1} e_{j,1} &\le \frac{2^d}{(d-1)!} \frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta} (k \log^{d-1}(n\Delta_d^d \epsilon^d) + k\mathcal{O}\left(\log k \log^{d-2}(n\Delta_d^d \epsilon^d) \right) \\
&= \frac{2^d}{(d-1)!} \frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta} k \; \mathcal{O}_\delta \left(\log^{d-1} n \right). \qquad (5.64)
\end{aligned}
$$

Next, for $e_{j,2}$, $0 \le j \le k-1$, similar to bounding $\tilde{c}_2$ in Lemma 5.14, we have

$$
\begin{aligned}
\sum_{j=0}^{k-1} e_{j,2} \le & \sum_{j=0}^{k-1} n e^{-n(g(x_0)-\delta)\epsilon^d} \frac{(n(g(x_0)-\delta)\epsilon^d)^j}{j!} \int_{\mathbb{R}^d} \phi(x) g(x) dx \\
& + \sum_{j=0}^{k-1} \sum_{l=1}^{d-1} \frac{2^l \binom{d}{l} \Lambda_l}{(g(x_0)-\delta)\epsilon^{d-l}} \mathcal{O}\left(\log j \log^{l-1}(n\Delta_l^l \epsilon^d) \right) \\
= & \; \mathcal{O}_\delta \left(k \log k \log^{d-2} n \right). \qquad (5.65)
\end{aligned}
$$

Combining (5.64) and (5.65), we obtain the upper bound.

As for the lower bound, we trivially have

$$n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx\geq\sum_{j=0}^{k-1}e_{j,1}.$$

Using a similar argument as for the upper bound, letting $\Delta_d'=(g(x_0)+\delta)^{1/d}$, we have for $0\leq j\leq k-1$,

$$\begin{aligned}
e_{j,1}\geq&(\phi(x_0)-\delta)(g(x_0)-\delta)n\int_{\mathcal{C}_{\epsilon}(x_0)}e^{-n(g(x_0)+\delta)|x-x_0|}\frac{(n(g(x_0)+\delta)|x-x_0|)^j}{j!}dx\\
=&2^d\frac{(\phi(x_0)-\delta)(g(x_0)-\delta)}{g(x_0)+\delta}n\int_{\mathbf{0}\prec x\prec\Delta_d'\boldsymbol{\epsilon}}e^{-n|x|}\frac{(n|x|)^j}{j!}dx\\
=&\frac{2^d}{(d-1)!}\frac{(\phi(x_0)-\delta)(g(x_0)-\delta)}{g(x_0)+\delta}(\log^{d-1}(n\Delta_d'^d\epsilon^d)+\mathcal{O}(\log j\log^{d-2}(n\Delta_d'^d\epsilon^d))).
\end{aligned}\tag{5.66}$$

Consequently, it yields

$$\sum_{j=0}^{k-1}e_{j,1}\geq\frac{2^d}{(d-1)!}\frac{(\phi(x_0)-\delta)(g(x_0)-\delta)}{g(x_0)+\delta}k\mathcal{O}_{\delta}\left(\log^{d-1}n\right).$$

This proves the first assertion. By taking $\phi(x)\equiv 1$, we have the second assertion. □

Remark 5.4. *A slightly more careful computation in the proof of Lemma 5.19 above (first fixing $0<\delta<g(x_0)/2$ and letting $n\to\infty$, and then letting $\delta\to 0$ with $k=o(\log^{(d-1)/(2\tau)}n)$ in Corollary 3.1) gives the following limit:*

$$\lim_{n\to\infty}\frac{n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx}{k\log^{d-1}n}=\frac{2^d}{(d-1)!}\phi(x_0).$$

Remark 5.5. *It was shown by Lin and Jeon (2006) that if the density g is bounded above and away from zero from below, the expected number of k-PNNs $\mathbb{E}L_{n,k}(x_0)$ to a target point $x_0\in\mathbb{R}^d$ is of the order $k\log^{d-1}n$. For general g, one might expect that the size of $\mathbb{E}L_{n,k}(x_0)$ depends on the smoothness of g. Lemma 5.19 shows that the same order holds for any a.e. continuous density g.*

5.2.5 Proofs of Theorem 3.1 and Corollary 3.1

We will employ Theorems 3.2 and 3.3 to prove the results. In view of this, we take $\bar{\mathbf{F}}_n$ therein as $\mathbf{r}_{n,k,w}$, we pick the normalizer $\rho_n^{(i)}:=\sqrt{\operatorname{Var}r_{n,k,w}(x_{0,i})}$ for all $i\in[m]$ and take

$$\sigma_{ij}:=\frac{\operatorname{Cov}(r_{n,k,w}(x_{0,i}),r_{n,k,w}(x_{0,j}))}{\sqrt{\operatorname{Var}r_{n,k,w}(x_{0,i})}\sqrt{\operatorname{Var}r_{n,k,w}(x_{0,j})}},$$

for all $i,j\in[m]$ so that $\Sigma=\operatorname{Cov}(\mathrm{P}_n^{-1}\bar{\mathbf{F}}_n)$ in Theorems 3.2 and 3.3. We define $\rho_n^{(i)}$ and σ_{ij} similarly for the uniform weights case $r_{n,k}(x_{0,i})$. We have already checked the Assumptions **(R1)**-**(R4)**, **(T)** and **(M)** with $p_0\in(0,6]$ for the score functions associated to $r_{n,k,w}$ and $r_{n,k}$. We fix $p_0=6$ in the sequel. Thus we can apply Theorems 3.2 and 3.3. By our choice of Σ, we have that $\Gamma_0=0$ in (3.19). Also, letting $\sigma^2:=\inf_x\operatorname{Var}_{P_{\boldsymbol{\varepsilon}}}r(\boldsymbol{\varepsilon}_x,x)$, note that for all $i\in[m]$, by the law of total variance, we have

$$\left(\rho_n^{(i)}\right)^2=\operatorname{Var}r_{n,k,w}(x_{0,i})$$

$$\geq \mathbb{E}\left(\operatorname{Var}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} W_{n\boldsymbol{x}}(x_{0,i})\boldsymbol{y_x}\middle|\mathcal{P}_{ng}\right)\right)$$

$$\geq \sigma^2\mathbb{E}\left(\sum_{\boldsymbol{x}\in\mathcal{P}_{ng}} W_{n\boldsymbol{x}}(x_{0,i})^2\right). \tag{5.67}$$

Specializing to the uniform weights case $r_{n,k}(x_{0,i})$, $i\in[m]$, with σ^2 as before, by Jensen's inequality we obtain the lower bound

$$\begin{aligned}\left(\rho_n^{(i)}\right)^2 &\geq \sigma^2\mathbb{E}\left(\frac{1}{L_{n,k}(x_{0,i})^2}\sum_{\boldsymbol{x}\in\mathcal{P}_{ng}} \mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_{0,i})}\right)\\ &= \sigma^2\mathbb{E}\left(\frac{1}{L_{n,k}(x_{0,i})}\right)\\ &\geq \sigma^2\frac{1}{\mathbb{E}L_{n,k}(x_{0,i})}.\end{aligned} \tag{5.68}$$

Recall from (5.39) the moment bound $M_n^{(i)}(x) = \Omega_{i,n} r^*_{6+p}(x)$ for $i\in[m]$, with $r^*_{6+p}(x) := \left(\mathbb{E}_{P_{\boldsymbol{\varepsilon}}}|r(\epsilon_x,x)|^{6+p}\right)^{1/(6+p)}$ and

$$\Omega_{i,n} := \begin{cases}\sup_{(x,\eta):|\eta|\leq 9}\|W_{nx}(x_{0,i},\mathcal{P}_{x,\eta})\|_{L_{6+p}} & \text{for } r_{n,k,w}(x_{0,i}),\\ \sup_{(x,\eta):|\eta|\leq 9}\|L_{n,k}(x_{0,i},\mathcal{P}_{x,\eta})^{-1}\|_{L_{6+p}} & \text{for } r_{n,k}(x_{0,i}),\end{cases} \tag{5.69}$$

where $\mathcal{P}_{x,\eta} := \mathcal{P}_{n\breve{g}} + \delta_{(x,\boldsymbol{m}_x)} + \eta$.

Recall from (5.42) that

$$e^{-r_n^{(1)}(x,y)} := e\sum_{j=1}^{k-1} e^{-\frac{1}{j+2}n\int_{\mathrm{Rect}(x,x_{0,1})} g(z)dz} \lesssim ke^{-\frac{1}{k+1}n\int_{\mathrm{Rect}(x,x_{0,1})} g(z)dz},\ y\in\mathrm{Rect}(x_{0,1},x), \tag{5.70}$$

and $r_n^{(1)}(x,y)=\infty$ otherwise. Then, from (3.15) we have

$$\begin{aligned}h_n^{(1)}(y) &= n\int_{\mathbb{R}^d} M_n^{(1)}(x)^{6+p/2}e^{-\zeta r_n^{(1)}(x,y)}g(x)dx\\ &\lesssim \Omega_{1,n}^{6+p/2}k^\zeta n\int_{\mathbb{R}^d}\mathbb{1}_{y\in\mathrm{Rect}(x_{0,1},x)}e^{-\zeta(k+1)^{-1}n\int_{\mathrm{Rect}(x,x_{0,1})}g(z)dz} r^*_{6+p}(x)^{6+p/2}g(x)dx\\ &\leq \Omega_{1,n}^{6+p/2}k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y),\end{aligned}$$

and

$$g_n^{(1)}(y) = n\int_{\mathbb{R}^d} e^{-\zeta r_n^{(1)}(x,y)}g(x)dx \lesssim k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y), \tag{5.71}$$

with $c_{\zeta(k+1)^{-1},n,x_{0,1}}$ is defined in (5.43) with ϕ as above. Plugging these in (3.16), we obtain

$$\begin{aligned}&G_n^{(1)}(y)\\ =&\Omega_{1,n}r^*_{6+p}(y) + \Omega_{1,n}(k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y))^{1/(6+p/2)}(1+(k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y))^6)^{1/(6+p/2)}\end{aligned}$$

$$\leq \Omega_{1,n} r^*_{6+p}(y) + \Omega_{1,n}(1 + (k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y))^{7/(6+p/2)})$$
$$\lesssim \Omega_{1,n}(r^*_{6+p}(y) \vee 1) + \Omega_{1,n}(k^\zeta c_{\zeta(k+1)^{-1},n,x_{0,1}}(y))^{7/(6+p/2)}. \tag{5.72}$$

Also, from (3.14) we obtain

$$q_n^{(1)}(x_1, x_2) = n \int_{\mathbb{R}^d \times \mathbb{R}} \mathbb{P}(\{\check{x}_1, \check{x}_2\} \subseteq R_n^{(1)}(\check{z}, \mathcal{P}_{n\check{g}} + \delta_{\check{z}}))\mathbb{Q}(d\check{z})$$
$$\lesssim k c_{(k+1)^{-1},n,x_{0,1}} \left((x_1 \vee x_2)_{x_{0,1}}\right) \mathbb{1}_{x_1 \in A_{x_{0,1}}(x_2)}. \tag{5.73}$$

Finally, according to (3.18), we have

$$\kappa_n^{(1)}(x) = \mathbb{P}(\xi_n^{(1)}(\check{x}, \mathcal{P}_{n\check{g}} + \delta_{\check{x}}) \neq 0) \lesssim k e^{-(k+1)^{-1} n \int_{\mathrm{Rect}(x_{0,1},x)} g(z)dz}. \tag{5.74}$$

While it is complicated to deal with general weights $\Omega_{i,n}$ from (5.69), we can start with the following estimate of $\Omega_{i,n}$ in the uniform case.

Lemma 5.20. *In the case of uniform weights, for $k \geq 11$, $k = O(n^\alpha)$ with $0 < \alpha < 1$ and $i \in [m]$, we have*

$$\Omega_{i,n} = \sup_{(x,\eta):|\eta| \leq 9} \|L_{n,k}(x_{0,i}, \mathcal{P}_{x,\eta})^{-1}\|_{L_{6+p}} \lesssim \frac{1}{k \log^{d-1} n} \asymp \frac{1}{\mathbb{E} L_{n,k}(x_{0,i})}. \tag{5.75}$$

Proof. Fix $i \in [m]$, $x \in \mathbb{R}^d$, η with $|\eta| \leq 9$. According to Adamczak et al. (2022, Proposition 4.20) and Lemma 6.1, it follows that for any integer $r \geq 1$,

$$\mathbb{E}(L - \mathbb{E}L)^{2r} \lesssim_r \left[\mathbb{E}\left(\int (D_{\check{y}} L)^2 \check{\lambda}(d\check{y})\right)^r + \mathbb{E}\left(\sum_{\check{y} \in \mathcal{P}_{n\check{g}}} \left(D^-_{\check{y}} L\right)^2\right)^r\right],$$

where $L \equiv L_{n,k}(x_{0,i}, \mathcal{P}_{x,\eta})$ and $D^-_{\check{x}} L(\eta) := L(\eta) - L(\eta - \delta_{\check{x}})$ for $\check{x} \in \eta$ is the remove-one cost operator, similar to the add-one cost function in Definition 3.2. Then for the first summand above, using Hölder's inequality we obtain

$$\mathbb{E}\left(\int (D_{\check{y}} L)^2 \check{\lambda}(d\check{y})\right)^r = \int \mathbb{E}\left[\prod_{i=1}^r (D_{\check{y}_i} L)^2\right] \check{\lambda}^r(d\check{y}_1, \ldots, d\check{y}_r)$$
$$\leq \left(\int \left[\mathbb{E}(D_{\check{y}} L)^{2r}\right]^{1/r} \check{\lambda}(d\check{y})\right)^r. \tag{5.76}$$

Now, we also have by Hölder's inequality that

$$\left[\mathbb{E}(D_{\check{y}} L)^{2r}\right]^{1/r} \leq (\mathbb{E}|D_{\check{y}} L|^{2r+1})^{\frac{2}{2r+1}} \mathbb{P}(D_{\check{y}} L \neq 0)^{\frac{1}{r(2r+1)}}.$$

Notice, that $D_{\check{y}} L \neq 0$ implies that $\mathrm{Rect}(y, x_{0,i})$ has at most $k-1$ points from the configuration $\mathcal{P}_{x,\eta}$ (since if not, then adding $\check{y}$ won't change the value of L), and hence also from the configuration $\mathcal{P}_{n\check{g}}$. Thus, by (5.70),

$$\mathbb{P}(D_{\check{y}} L \neq 0) \lesssim k e^{-\frac{1}{k+1} n \int_{\mathrm{Rect}(y,x_{0,i})} g(z)dz}.$$

On the other hand, $\mathbb{E}|D_{\check{y}} L|^{2r+1}$ can be bounded similarly as in Lemma 5.11. Indeed, letting

$$L' = \sum_{x \in \mathcal{P}_{n\check{g}}} \mathbb{1}_{x \text{ is a } k\text{-PNN to } x_{0,i} \text{ in } \mathcal{P}_{x,\eta}} =: \sum_{x \in \mathcal{P}_{n\check{g}}} \xi'(x, \mathcal{P}_{n\check{g}}),$$

we have $|L - L'| \le 10$, so that $\mathbb{E}|D_{\check{y}} L|^{2r+1} \lesssim_r 1 + \mathbb{E}|D_{\check{y}} L'|^{2r+1}$. Also, the scores ξ' has region of stabilization as defined at (5.37). Now arguing as in Lemma 5.11 with $4+p/2$ replaced by $2r+1$ (i.e., p replaced by $4r-6$), since L' is a sum of indicators, taking the bound on the L^{4r-2} (in place of L^{4+p}) norm in Lemma 5.11 trivially as 1, we obtain

$$\mathbb{E}|D_{\check{y}} L'|^{2r+1} \lesssim_r 1 + g_n(\check{y})^5$$

with g_n defined as in (3.15) with $\zeta = \zeta_0 := (2r-3)/(2r+17)$ (in place of $p/(40+p)$) and r_n as in (5.42). Thus, by (5.71), we have

$$\begin{aligned}\left[\mathbb{E}(D_{\check{y}} L)^{2r}\right]^{1/r} &\lesssim_r (1 + g_n(\check{y})^{\frac{10}{2r+1}}) \mathbb{P}(D_{\check{y}} L \neq 0)^{\frac{1}{r(2r+1)}} \\ &\lesssim_r \left(1 + k^{\frac{10\zeta_0}{2r+1}} c_{\zeta_0 (k+1)^{-1}, n, x_{0,i}}(y)^{\frac{10}{2r+1}}\right) \left(k e^{-\frac{1}{k+1} n \int_{\mathrm{Rect}(y, x_{0,i})} g(z) dz}\right)^{\frac{1}{r(2r+1)}}.\end{aligned}$$

Hence,

$$\begin{aligned}\int \left[\mathbb{E}(D_{\check{y}} L)^{2r}\right]^{1/r} \check{\lambda}(d\check{y}) &\lesssim_r k^{\frac{1}{r(2r+1)}} n \int_{\mathbb{R}^d} e^{-\frac{1}{r(2r+1)(k+1)} n \int_{\mathrm{Rect}(y, x_{0,i})} g(z) dz} g(y) dy \\ &+ k^{\frac{10\zeta_0}{2r+1} + \frac{1}{r(2r+1)}} n \int_{\mathbb{R}^d} c_{\zeta_0 (k+1)^{-1}, n, x_{0,i}}(y)^{\frac{10}{2r+1}} e^{-\frac{1}{r(2r+1)(k+1)} n \int_{\mathrm{Rect}(y, x_{0,i})} g(z) dz} g(y) dy \\ &\le k^{\frac{1}{r(2r+1)}} n \int_{\mathbb{R}^d} e^{-\frac{1}{r(2r+1)(k+1)} n \int_{\mathrm{Rect}(y, x_{0,i})} g(z) dz} g(y) dy \\ &+ k^{\frac{10\zeta_0}{2r+1} + \frac{1}{r(2r+1)}} \left(n \int_{\mathbb{R}^d} c_{\zeta_0 (k+1)^{-1}, n, x_{0,i}}(y)^{\frac{20}{2r+1}} g(y) dy\right)^{\frac{1}{2}} \\ &\times \left(n \int_{\mathbb{R}^d} e^{-\frac{2}{r(2r+1)(k+1)} n \int_{\mathrm{Rect}(y, x_{0,i})} g(z) dz} g(y) dy\right)^{\frac{1}{2}} \\ &\lesssim_r k^{\frac{10\zeta_0}{2r+1} + \frac{1}{r(2r+1)} + 1 + \frac{10}{2r+1}} \log^{d-1} n,\end{aligned}$$

where the final step is due to (5.55) and (5.62). Thus, (5.76) yields

$$\mathbb{E}\left(\int (D_{\check{y}} L)^2 \check{\lambda}(d\check{y})\right)^r \lesssim_r k^{r(\tau_{\mathrm{conc}}+1)} \log^{r(d-1)} n,$$

where we have $\tau_{\mathrm{conc}} := \frac{10}{2r+1}(1+\zeta_0) + \frac{1}{r(2r+1)} \le \frac{20r+1}{r(2r+1)} \le \frac{10}{r}$.

For the second summand, using Lemma 6.1, followed by an application of the Mecke formula, a similar argument as above yields

$$\begin{aligned}\mathbb{E}\left(\sum_{\check{y} \in \mathcal{P}_{n\check{g}}} \left(D^-_{\check{y}} L\right)^2\right)^r &\lesssim_r \mathbb{E}\left[\sum_{\check{y} \in \mathcal{P}_{n\check{g}}} \left(D^-_{\check{y}} L\right)^{2r} + \cdots + \sum_{\check{y}_1 \neq \cdots \neq \check{y}_r \in \mathcal{P}_{n\check{g}}} \left(D^-_{\check{y}_1} L\right)^2 \cdots \left(D^-_{\check{y}_r} L\right)^2\right] \\ &\lesssim k^{r(\tau_{\mathrm{conc}}+1)} \log^{r(d-1)} n.\end{aligned}$$

Now by Chebyshev's inequality,

$$\mathbb{P}(L \le \mathbb{E}L/2) \lesssim \frac{\mathbb{E}(L - \mathbb{E}L)^{2r}}{(\mathbb{E}L)^{2r}}.$$

Next we need a lower bound on $\mathbb{E}L$. Since $k > 10$, even in the presence of additional points x and η (which together are at most 10), we have $L = L_{n,k}(x_{0,i}, \mathcal{P}_{x,\eta}) \ge L_{n,k-10}(x_{0,i}, \mathcal{P}_{n\check{g}})$, so that by Lemma 5.19,

$$\mathbb{E}L \ge \mathbb{E}L_{n,k-10}(x_{0,i}, \mathcal{P}_{n\check{g}}) \asymp k \log^{d-1} n.$$

Combining this with the above tail bound, we obtain

$$\mathbb{P}(L \le \mathbb{E}L/2) \lesssim_r (k \log^{d-1} n)^{-2r} k^{r(\tau_{\text{conc}}+1)} \log^{r(d-1)} n = k^{-r(1-\tau_{\text{conc}})} \log^{-r(d-1)} n.$$

Thus, since $L \ge 1$, we have:

$$\mathbb{E}L^{-(6+p)} \lesssim \frac{1}{(\mathbb{E}L)^{6+p}} + \mathbb{P}(L \le \mathbb{E}L/2) \lesssim_r (k \log^{d-1} n)^{-(6+p)} + k^{-r(1-\tau_{\text{conc}})} \log^{-r(d-1)} n,$$

so that

$$\|L^{-1}\|_{L_{6+p}} = (\mathbb{E}L^{-(6+p)})^{\frac{1}{6+p}} \lesssim_r (k \log^{d-1} n)^{-1} + k^{-\frac{r(1-\tau_{\text{conc}})}{6+p}} \log^{-\frac{r}{6+p}(d-1)} n.$$

Now since $\tau_{\text{conc}} \le 10/r$, we have $\frac{r(1-\tau_{\text{conc}})}{6+p} \ge \frac{1}{7}(r-10)$. Thus, choosing $r = 17$ yields

$$\|L^{-1}\|_{L_{6+p}} \lesssim \frac{1}{k \log^{d-1} n}.$$

Since the choice of s and η were arbitrary and the upper bound above doesn't depend on this choice, taking a supremum yields the desired bound. The final part of the result is due to Lemma 5.19. □

Remark 5.6. *Note that although Lemma 5.20 focuses on the first moment of $L_{n,k}(x_{0,i})$ for $i \in [m]$, the proof arguments provided above are actually valid for generalizing it to any moment, i.e, $(k \log^{d-1} n)^q \asymp \mathbb{E}L^q_{n,k}(x_{0,i})$ for $i \in [m]$ and $q \in \mathbb{Z}$.*

In the following, we will bound the Γ_i's in Theorems 3.2 and 3.3. Throughout, we take $\phi(x) = r^*_{6+p}(x)^{6+p/2} \vee 1$ in (5.43). Write

$$f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha}(y) =: f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,1}(y) + f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,2}(y) + f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,3}(y),$$

with

$$f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,1}(y) := n \int_{\mathbb{X}} G^{(i)}_n(x)^{\alpha_i} G^{(j)}_n(x)^{\alpha_j} G^{(l)}_n(x)^{\alpha_l} e^{-\alpha r^{(t)}_n(x,y)} \mathbb{Q}(dx), \tag{5.77}$$

$$f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,2}(y) := n \int_{\mathbb{X}} G^{(i)}_n(x)^{\alpha_i} G^{(j)}_n(x)^{\alpha_j} G^{(l)}_n(x)^{\alpha_l} e^{-\alpha r^{(t)}_n(y,x)} \mathbb{Q}(dx), \tag{5.78}$$

$$f^{(i,j,l,t)}_{\alpha_i,\alpha_j,\alpha_l,\alpha,3}(y) := n \int_{\mathbb{X}} G^{(i)}_n(x)^{\alpha_i} G^{(j)}_n(x)^{\alpha_j} G^{(l)}_n(x)^{\alpha_l} q^{(t)}_n(x,y)^{\alpha} \mathbb{Q}(dx). \tag{5.79}$$

We first consider then case when $i = j = l = t$, and without loss of generality, we fix $i, j, l, t = 1$. We start with Γ_1 defined at (3.20). By (5.72), (5.77), (5.78), (5.79) and Lemma 6.1, we have

$$\begin{aligned} f^{(1,1,1,1)}_{1,1,0,\beta,1}(y) &= n \int_{\mathbb{R}^d} (G^{(1)}_n(x))^2 e^{-\beta r^{(1)}_n(x,y)} g(x) dx \\ &\lesssim \Omega^2_{1,n} \Bigg(n \int_{\mathbb{R}^d} (r^*_{6+p}(x)^2 \vee 1) e^{-\beta r^{(1)}_n(x,y)} g(x) dx \\ &\quad + k^{14\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} e^{-\beta r^{(1)}_n(x,y)} g(x) dx \Bigg), \end{aligned} \tag{5.80}$$

$$f^{(1,1,1,1)}_{1,1,0,\beta,2}(y) = n \int_{\mathbb{R}^d} (G^{(1)}_n(x))^2 e^{-\beta r^{(1)}_n(y,x)} g(x) dx$$

$$\lesssim \Omega_{1,n}^2 \left(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^2 \vee 1) e^{-\beta r_n^{(1)}(y,x)} g(x) dx \right.$$
$$\left. + k^{14\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} e^{-\beta r_n^{(1)}(y,x)} g(x) dx \right), \tag{5.81}$$

and

$$f_{1,1,0,\beta,3}^{(1,1,1,1)}(y) = n \int_{\mathbb{R}^d} (G_n^{(1)}(x))^2 q_n^{(1)}(x,y)^\beta g(x) dx$$
$$\lesssim \Omega_{1,n}^2 \left(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^2 \vee 1) q_n^{(1)}(x,y)^\beta g(x) dx \right.$$
$$\left. + k^{14\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} q_n^{(1)}(x,y)^\beta g(x) dx \right). \tag{5.82}$$

Also, from (5.80), we have that

$$\begin{aligned} \big(f_{1,1,0,\beta,1}^{(1,1,1,1)}(y)\big)^2 &\lesssim \Omega_{1,n}^4 \Big(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^2 \vee 1) e^{-\beta r_n^{(1)}(x,y)} g(x) dx \Big)^2 \\ &\quad + \Omega_{1,n}^4 k^{28\zeta/(6+p/2)} \Big(n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} e^{-\beta r_n^{(1)}(x,y)} g(x) dx \Big)^2. \end{aligned} \tag{5.83}$$

For the rest of this proof, we fix $\delta \in (0, 1/2 \min_i g(x_{0,i}))$ and ϵ as in Lemma 5.13 such that the conclusion therein holds for all $x_{0,i}$, $i \in [m]$. To simplify notation, we also drop the dependence on δ and simply write $\mathcal{O} \equiv \mathcal{O}_\delta$. Recalling that $\phi(x) = r_{6+p}^*(x)^{6+p/2} \vee 1$ and using (5.70), by (5.62) we have

$$n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^2 \vee 1) e^{-\beta r_n^{(1)}(x,y)} g(x) dx \right)^2 g(y) dy$$
$$\lesssim k^{2\beta} n \int_{\mathbb{R}^d} c_{\beta(k+1)^{-1},n,x_{0,1}}(y)^2 g(y) dy = (k+1)^{2\beta+3} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.84}$$

Also, using that $c_{\alpha,n,x_{0,1}}(x) \leq c_{\alpha,n,x_{0,1}}(y)$ for $y \in \mathrm{Rect}(x_{0,1}, x)$, and that $\zeta < \beta$, again by (5.70) and (5.62),

$$\begin{aligned} & n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} e^{-\beta r_n^{(1)}(x,y)} g(x) dx \right)^2 g(y) dy \\ & \leq n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(y)^{28/(6+p/2)} \left(n \int_{\mathbb{R}^d} e^{-\beta r_n^{(1)}(x,y)} g(x) dx \right)^2 g(y) dy \\ & \lesssim k^{2\beta} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(y)^{28/(6+p/2)+2} g(y) dy \\ & = (k+1)^{2\beta+3+28/(6+p/2)} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \end{aligned} \tag{5.85}$$

Combining (5.84) and (5.85), from (5.80) we obtain

$$n\mathbb{Q}\left(f_{1,1,0,\beta,1}^{(1,1,1,1)}\right)^2 = \Omega_{1,n}^4 (k+1)^{2\beta+3+28(1+\zeta)/(6+p/2)} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.86}$$

Next, we note by (5.82) that

$$\left(f_{1,1,0,\beta,3}^{(1,1,1,1)}(y)\right)^2 \lesssim \Omega_{1,n}^4 \left(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^2 \vee 1) q_n^{(1)}(x,y)^\beta g(x) dx \right)^2$$

$$+\,\Omega_{1,n}^4 k^{28\zeta/(6+p/2)}\left(n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} q_n^{(1)}(x,y)^\beta g(x)dx\right)^2.$$

For $0<\alpha<1$ and $y\in\mathbb{R}^d$, we observe that

$$\begin{aligned}
c_{1,n,x_{0,1}}(y)^\alpha &= \Bigg(n\int_{\mathbb{R}^d} \mathbb{1}_{y\in \mathrm{Rect}(x_{0,1},x)} e^{-n\left(\int_{\mathrm{Rect}(x_{0,1},x)} g(z)dz - \int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz\right)}\\
&\qquad \times e^{-n\int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz}(r^*_{6+p}(x)^{6+p/2}\vee 1)g(x)dx\Bigg)^\alpha\\
&= e^{-\alpha n\int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz}\Bigg(n\int_{\mathbb{R}^d} \mathbb{1}_{\mathrm{Rect}(x_{0,1},x)} e^{-n\left(\int_{\mathrm{Rect}(x_{0,1},x)} g(z)dz - \int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz\right)}\\
&\qquad \times (r^*_{6+p}(x)^{6+p/2}\vee 1)g(x)dx\Bigg)^\alpha\\
&\le e^{-\alpha n\int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz}\Bigg(1+n\int_{\mathbb{R}^d} \mathbb{1}_{\mathrm{Rect}(x_{0,1},x)}\\
&\qquad \times e^{-n\left(\int_{\mathrm{Rect}(x_{0,1},x)} g(z)dz - \int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz\right)}(r^*_{6+p}(x)^{6+p/2}\vee 1)g(x)dx\Bigg)\\
&= e^{-\alpha n\int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz} + c_{\alpha,n,x_{0,1}}(y).
\end{aligned}\tag{5.87}$$

Thus, using (5.73) in the first step and (5.87) in the second, for $0<\beta<1$), Lemma 5.15 yields

$$\begin{aligned}
&n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d}(r^*_{6+p}(x)^2\vee 1)q_n^{(1)}(x,y)^\beta g(x)dx\right)^2 g(y)dy\\
&\lesssim k^{2\beta} n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_{0,1}}(y)} c_{(k+1)^{-1},n,x_{0,1}}((x\vee y)_{x_{0,1}})^\beta (r^*_{6+p}(x)^2\vee 1)g(x)dx\right)^2 g(y)dy\\
&\lesssim k^{2\beta} n\int_{\mathbb{R}^d}\Bigg(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_{0,1}}(y)}(k+1)^\beta e^{-\beta(k+1)^{-1}n\int_{\mathrm{Rect}\left(x_{0,1},(x\vee y)_{x_{0,1}}\right)} g(z)dz}\\
&\qquad\qquad \times (r^*_{6+p}(x)^2\vee 1)g(x)dx\Bigg)^2 g(y)dy\\
&\quad + k^{2\beta} n\int_{\mathbb{R}^d}\Bigg(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_{0,1}}(y)}(k+1)^{\beta-1} c_{\beta(k+1)^{-1},n,x_{0,1}}((x\vee y)_{x_{0,1}})\\
&\qquad\qquad \times (r^*_{6+p}(x)^2\vee 1)g(x)dx\Bigg)^2 g(y)dy\\
&= (k+1)^{4\beta+3}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)).
\end{aligned}\tag{5.88}$$

On the other hand, using (5.87) again, arguing same as above yields

$$\begin{aligned}
&n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} q_n^{(1)}(x,y)^\beta g(x)dx\right)^2 g(y)dy\\
&\lesssim k^{2\beta} n\int_{\mathbb{R}^d}\Bigg(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_{0,1}}(y)}(k+1)^\beta c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}\\
&\qquad\qquad \times e^{-\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},(x\vee y)_{x_{0,1}})} g(z)dz} g(x)dx\Bigg)^2 g(y)dy\\
&\quad + k^{2\beta} n\int_{\mathbb{R}^d}\Bigg(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in A_{x_{0,1}}(y)}(k+1)^{\beta-1} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}\\
&\qquad\qquad \times c_{\beta(k+1)^{-1},n,x_{0,1}}((x\vee y)_{x_{0,1}})g(x)dx\Bigg)^2 g(y)dy
\end{aligned}$$

$$=: A_1 + A_2. \tag{5.89}$$

Since for any $\alpha, t \geq 1$, we have $c_{\alpha,n,x_{0,1}}(x)^t \leq \max\{c_{\alpha,n,x_{0,1}}(x)^{\lfloor t \rfloor}, c_{\alpha,n,x_{0,1}}(x)^{\lceil t \rceil}\}$, by Lemma 5.17, we have that $A_1, A_2 = (k+1)^{4\beta+3+\lceil 28/(6+p/2)\rceil}\mathcal{O}(\log^{d-1}((k+1)^{-1}n))$. Therefore, by (5.88) and (5.89), we obtain

$$n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,0,\beta,3}\right)^2 = \Omega^4_{1,n}(k+1)^{4\beta+3+\lceil 28(1+\zeta)/(6+p/2)\rceil}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.90}$$

From (5.81), we have

$$\left(f^{(1,1,1,1)}_{1,1,0,\beta,2}(y)\right)^2 \lesssim \Omega^4_{1,n}\left(n\int_{\mathbb{R}^d}(r^*_{6+p}(x)^2 \vee 1)e^{-\beta r^{(1)}_n(y,x)}g(x)dx\right)^2$$
$$+ \Omega^4_{1,n}k^{28\zeta/(6+p/2)}\left(n\int_{\mathbb{R}^d}c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}e^{-\beta r^{(1)}_n(y,x)}g(x)dx\right)^2.$$

For the first term, using (5.70) and Lemma 5.18 yields

$$\begin{aligned}
&n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d}e^{-\beta r^{(1)}_n(y,x)}(r^*_{6+p}(x)^2 \vee 1)g(x)dx\right)^2 g(y)dy\\
&\lesssim k^{2\beta}n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in\mathrm{Rect}(x_{0,1},y)}e^{-\beta(k+1)^{-1}n\int_{\mathrm{Rect}(y,x_{0,1})}g(z)dz}(r^*_{6+p}(x)^2 \vee 1)g(x)dx\right)^2 g(y)dy\\
&= (k+1)^{2\beta+3}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)).
\end{aligned} \tag{5.91}$$

As for the second term, changing the order of integration in the second step and using the Cauchy-Schwarz inequality in the third, (5.62) and Lemma 5.17 yield

$$\begin{aligned}
&n\int_{\mathbb{R}^d}\left(n\int_{\mathbb{R}^d}c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}e^{-\beta r^{(1)}_n(y,x)}g(x)dx\right)^2 g(y)dy\\
&\lesssim k^{2\beta}n\int_{\mathbb{R}^d}\Bigg(n\int_{\mathbb{R}^d}\mathbb{1}_{x\in\mathrm{Rect}(x_{0,1},y)}c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}e^{-\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},y)}g(z)dz}\\
&\qquad\qquad\times g(x)dx\Bigg)^2 g(y)dy\\
&\leq k^{2\beta}n^2\int_{\mathbb{R}^d\times\mathbb{R}^d}\Big(\mathbb{1}_{x\in A_{x_{0,1}}(y)}c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)}c_{\zeta(k+1)^{-1},n,x_{0,1}}(y)^{14/(6+p/2)}\\
&\qquad\qquad\times c_{\beta(k+1)^{-1},n,x_{0,1}}((x\vee y)_{x_{0,1}})g(x)g(y)\Big)dxdy\\
&\leq k\left(n\int_{\mathbb{R}^d}c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)}g(x)dx\right)^{\frac{1}{2}}A_2^{\frac{1}{2}}\\
&= (k+1)^{2\beta+3+28/(6+p/2)}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)),
\end{aligned} \tag{5.92}$$

where A_2 is defined at (5.89).

Combining (5.91) and (5.92), we have

$$n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,0,\beta,2}\right)^2 = \Omega^4_{1,n}(k+1)^{2\beta+3+\lceil 28(1+\zeta)\rceil/(6+p/2)}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.93}$$

Now, putting together (5.86), (5.90) and (5.93), we obtain

$$n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,0,\beta}\right)^2 = \Omega^4_{1,n}(k+1)^{4\beta+3+\lceil 28(1+\zeta)/(6+p/2)\rceil}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)).$$

Therefore, by (5.67), (5.69) ((5.68) and (5.75), respectively, in the uniform case) and Lemma 5.19, with

$$\varsigma := \frac{1+\zeta}{6+p/2}, \quad \text{and} \quad W_1(n,k) := \frac{\left(\sup_{(x,\eta):|\eta|\le 9} \|W_{nx}(x_{0,1},\mathcal{P}_{x,\eta})\|_{L_{6+p}}\right)^2}{\mathbb{E}\left(\sum_{\boldsymbol{x}\in\mathcal{P}_{ng}} W_{n\boldsymbol{x}}(x_{0,1})^2\right)}, \tag{5.94}$$

we have

$$\left(\frac{n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,0,\beta}\right)^2}{(\varrho_n^{(1)}\varrho_n^{(1)})^2}\right)^{\frac{1}{2}} = \begin{cases} k^{2\beta+3/2+\lceil 14\varsigma\rceil} W_1(n,k)\mathcal{O}(\log^{(d-1)/2} n), & \text{for general weights,}\\ k^{2\beta+1/2+\lceil 14\varsigma\rceil}\mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights.}\end{cases} \tag{5.95}$$

Next, we focus on Γ_3 and Γ_4 defined at (3.22) and (3.23), respectively. For $i \in \{3,4\}$, by Lemma 6.1 we have

$$G_n^{(1)}(x)^i(\kappa_n^{(1)}(x)+g_n^{(1)}(x))^{6\beta} \lesssim G_n^{(1)}(x)^i\kappa_n^{(1)}(x)^{6\beta} + G_n^{(1)}(x)^i g_n^{(1)}(x)^{6\beta}.$$

Using (5.72) in the first step, (5.74) in the second, and (5.55) and (5.62) in the final one, for $i \in \{3,4\}$ we obtain

$$\begin{aligned}
& n\int_{\mathbb{R}^d} G_n^{(1)}(x)^i\kappa_n^{(1)}(x)^{6\beta}g(x)dx\\
&\lesssim \Omega_{1,n}^i n\int_{\mathbb{R}^d}\kappa_n^{(1)}(x)^{6\beta}(r^*(x)^i\vee 1)g(x)dx\\
&\qquad + \Omega_{1,n}^i k^{7i\zeta/(6+p/2)} n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{7i/(6+p/2)}\kappa_n^{(1)}(x)^{6\beta}g(x)dx\\
&\lesssim \Omega_{1,n}^i k^{6\beta} n\int_{\mathbb{R}^d} e^{-6\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},x)}g(z)dz}(r^*(x)^i\vee 1)g(x)dx\\
&+ \Omega_{1,n}^i k^{6\beta+7i\zeta/(6+p/2)} n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{7i/(6+p/2)} e^{-6\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},x)}g(z)dz}g(x)dx\\
&\le \Omega_{1,n}^i k^{6\beta} n\int_{\mathbb{R}^d} e^{-6\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},x)}g(z)dz}(r^*(x)^i\vee 1)g(x)dx\\
&\qquad + \Omega_{1,n}^i k^{6\beta+7i\zeta/(6+p/2)}\left(n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14i/(6+p/2)}g(x)dx\right)^{\frac{1}{2}}\\
&\qquad\qquad \times\left(n\int_{\mathbb{R}^d} e^{-12\beta(k+1)^{-1}n\int_{\mathrm{Rect}(x_{0,1},x)}g(z)dz}g(x)dx\right)^{\frac{1}{2}}\\
&= \Omega_{1,n}^i (k+1)^{6\beta+1+\lceil 7i\varsigma\rceil}\mathcal{O}(\log^{d-1}((k+1)^{-1}n)).
\end{aligned} \tag{5.96}$$

Applying (5.62) in Lemma 5.15, from (5.71) and (5.72) we also have

$$\begin{aligned}
& n\int_{\mathbb{R}^d} G_n^{(1)}(x)^i g_n^{(1)}(x)^{6\beta}g(x)dx\\
&\lesssim \Omega_{1,n}^i k^{6\zeta\beta} n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{6\beta}(r^*(x)^i\vee 1)g(x)dx\\
&\qquad + \Omega_{1,n}^i k^{6\zeta\beta+7i\zeta/(6+p/2)} n\int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{7i/(6+p/2)} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{6\beta}g(x)dx
\end{aligned}$$

$$= \Omega_{1,n}^i (k+1)^{6\zeta\beta+6\beta+1+\lceil 7i\varsigma\rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)), \tag{5.97}$$

where the last step is by (5.55) and (5.62). Combining (5.96) and (5.97), we obtain

$$n\mathbb{Q}\left((\kappa_n^{(1)} + g_n^{(1)})^{6\beta} \left(G_n^{(1)} \right)^i \right) = \Omega_{1,n}^i (k+1)^{6\zeta\beta+6\beta+1+\lceil 7i\varsigma\rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)).$$

Therefore, (5.67), (5.69) ((5.68) and (5.75), respectively, in the uniform case) and Lemma 5.19 yield that for $i \in \{3,4\}$,

$$\begin{aligned} &\frac{n\mathbb{Q}\left((\kappa_n^{(1)} + g_n^{(1)})^{6\beta} \left(G_n^{(1)} \right)^i \right)}{(\varrho_n^{(1)})^i} \\ &= \begin{cases} k^{6\zeta\beta+6\beta+1+\lceil 7i\varsigma\rceil} W_1(n,k)^{i/2} \mathcal{O}(\log^{d-1} n), & \text{for general weights,} \\ k^{6\zeta\beta+6\beta+1-i/2+\lceil 7i\varsigma\rceil} \mathcal{O}(\log^{(d-1)(1-i/2)} n), & \text{for uniform weights,} \end{cases} \end{aligned} \tag{5.98}$$

yielding a bound for Γ_3 and the first summand of Γ_4.

For the second summand in Γ_4, note that by (5.77) and (5.72), we have

$$\begin{aligned} f_{2,2,0,3\beta,1}^{(1,1,1,1)}(y) \lesssim \Omega_{1,n}^4 \Bigg(& n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^4 \vee 1) e^{-3\beta r_n^{(1)}(x,y)} g(x) dx \\ & + k^{28\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} e^{-3\beta r_n^{(1)}(x,y)} g(x) dx \Bigg). \end{aligned}$$

By (5.70) and (5.62),

$$\begin{aligned} & n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} e^{-3\beta r_n^{(1)}(x,y)} (r_{6+p}^*(x)^4 \vee 1) g(x) dx g(y) dy \\ & \lesssim k^{3\beta} n \int_{\mathbb{R}^d} c_{3\beta(k+1)^{-1},n,x_{0,1}}(y) g(y) dy = (k+1)^{3\beta+2} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \end{aligned} \tag{5.99}$$

Using that $\zeta < 3\beta$, we have by (5.85), (5.70) and (5.62) that

$$\begin{aligned} & n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} e^{-3\beta r_n^{(1)}(x,y)} g(x) dx g(y) dy \\ & \leq n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(y)^{28/(6+p/2)} \left(n \int_{\mathbb{R}^d} e^{-\zeta r_n^{(1)}(x,y)} g(x) dx \right) g(y) dy \\ & \lesssim k^{3\beta} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(y)^{28/(6+p/2)+1} g(y) dy \\ & = (k+1)^{3\beta+2+28/(6+p/2)} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \end{aligned} \tag{5.100}$$

Together, (5.99) and (5.100) yield

$$n\mathbb{Q} f_{2,2,0,3\beta,1}^{(1,1,1,1)} = \Omega_{1,n}^4 (k+1)^{3\beta+2+28\varsigma} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.101}$$

Next, note that by (5.79) and (5.72),

$$f_{2,2,0,3\beta,3}^{(1,1,1,1)}(y) \lesssim \Omega_{1,n}^4 \Bigg(n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^4 \vee 1) q_n^{(1)}(x,y)^{3\beta} g(x) dx$$

$$+k^{28\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} q_n^{(1)}(x,y)^{3\beta} g(x)dx \Big).$$

Since $0 < 3\beta < 1$, using (5.73) and arguing as for (5.87) in the second step, Lemma 5.15 yields

$$\begin{aligned}
& n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} q_n^{(1)}(x,y)^{3\beta} (r_{6+p}^*(x)^4 \vee 1) g(x) dx g(y) dy \\
& \lesssim k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} c_{(k+1)^{-1},n,x_{0,1}}((x \vee y)_{x_{0,1}})^{3\beta} (r_{6+p}^*(x)^4 \vee 1) g(x) dx g(y) dy \\
& \lesssim k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} (k+1)^{3\beta-1} c_{3\beta(k+1)^{-1},n,x_{0,1}}((x \vee y)_{x_{0,1}}) (r_{6+p}^*(x)^4 \vee 1) \\
& \qquad \times g(x) dx g(y) dy \\
& \quad + k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} (k+1)^{3\beta} e^{-3\beta(k+1)^{-1} n \int_{\mathrm{Rect}\left(x_{0,1},(x \vee y)_{x_{0,1}}\right)} g(z) dz} (r_{6+p}^*(x)^4 \vee 1) \\
& \qquad \times g(x) dx g(y) dy \\
& = (k+1)^{6\beta+2} \mathcal{O}(\log^{d-1}((k+1)^{-1} n)). \qquad (5.102)
\end{aligned}$$

Using (5.73) and arguing similarly, we also have

$$\begin{aligned}
& n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} q_n^{(1)}(x,y)^{3\beta} g(x) dx g(y) dy \\
& \le k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} c_{(k+1)^{-1},n,x_{0,1}}((x \vee y)_{x_{0,1}})^{3\beta} \\
& \qquad \times g(x) dx g(y) dy \\
& \lesssim k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} (k+1)^{3\beta} \\
& \qquad \times e^{-3\beta(k+1)^{-1} n \int_{z \in \mathrm{Rect}(x_{0,1},(x \vee y)_{x_{0,1}})} g(z) dz} g(x) dx g(y) dy \\
& \quad + k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in A_{x_{0,1}}(y)} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} (k+1)^{3\beta-1} \\
& \qquad \times c_{3\beta(k+1)^{-1},n,x_{0,1}}((x \vee y)_{x_{0,1}}) g(x) dx g(y) dy \\
& := B_1 + B_2. \qquad (5.103)
\end{aligned}$$

By Lemma 5.17, we have $B_1, B_2 = (k+1)^{6\beta+2+\lceil 28/(6+p/2) \rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1} n))$. Therefore, combining (5.102) and (5.103), we obtain

$$n \mathbb{Q} f_{2,2,0,3\beta,3}^{(1,1,1,1)} = \Omega_{1,n}^4 (k+1)^{6\beta+2+\lceil 28\varsigma \rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1} n)). \qquad (5.104)$$

Note also that by (5.78) and (5.72),

$$\begin{aligned}
f_{2,2,0,3\beta,2}^{(1,1,1,1)}(y) \lesssim \Omega_{1,n}^4 \Big(& n \int_{\mathbb{R}^d} (r_{6+p}^*(x)^4 \vee 1) e^{-3\beta r_n^{(1)}(y,x)} g(x) dx \\
& + k^{28\zeta/(6+p/2)} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} e^{-3\beta r_n^{(1)}(y,x)} g(x) dx \Big).
\end{aligned}$$

By (5.70) and Lemma 5.18, we have

$$n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} e^{-3\beta r_n^{(1)}(y,x)} (r_{6+p}^*(x)^4 \vee 1) g(x) dx g(y) dy$$

$$\lesssim k^{3\beta} n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} \mathbb{1}_{x \in \mathrm{Rect}(x_{0,1},y)} e^{-3\beta(k+1)^{-1} n \int_{\mathrm{Rect}(x_{0,1},y)} g(z)dz} (r^*_{6+p}(x)^4 \vee 1) g(x) dx g(y) dy$$
$$= (k+1)^{3\beta+2} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.105}$$

Again using that $\zeta < 3\beta$, (5.70) and (5.62) yield

$$n \int_{\mathbb{R}^d} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} e^{-3\beta r_n^{(1)}(y,x)} g(x) dx g(y) dy$$
$$\leq n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)} \left(n \int_{\mathbb{R}^d} e^{-\zeta r_n^{(1)}(y,x)} g(y) dy \right) g(x) dx$$
$$\lesssim k^{3\beta} n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{28/(6+p/2)+1} g(x) dx$$
$$= (k+1)^{3\beta+2+28/(6+p/2)} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.106}$$

From (5.105) and (5.106), we have

$$n\mathbb{Q} f^{(1,1,1,1)}_{2,2,0,3\beta,2} = \Omega^4_{1,n} (k+1)^{3\beta+2+\lceil 28\varsigma \rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.107}$$

Combining (5.101), (5.104) and (5.107), we obtain

$$n\mathbb{Q} f^{(1,1,1,1)}_{2,2,0,3\beta} = \Omega^4_{1,n} (k+1)^{6\beta+2+\lceil 28\varsigma \rceil} \mathcal{O}(\log^{d-1}((k+1)^{-1}n)). \tag{5.108}$$

Now (5.98) and (5.108) together with (5.67), (5.69) ((5.68) and (5.75), respectively, in the uniform case) and Lemma 5.19 yield

$$\left(\frac{n\mathbb{Q}\left((\kappa_n^{(1)} + g_n^{(1)})^{6\beta} \left(G_n^{(1)} \right)^4 \right)}{(\varrho_n^{(1)})^4} \right)^{\frac{1}{2}} + \left(\frac{n\mathbb{Q} f^{(1,1,1,1)}_{2,2,0,3\beta}}{(\varrho_n^{(1)} \varrho_n^{(1)})^2} \right)^{\frac{1}{2}}$$
$$= \begin{cases} k^{3\zeta\beta+3\beta+1+\lceil 14\varsigma \rceil} W_1(n,k) \mathcal{O}(\log^{(d-1)/2} n), & \text{for general weights,} \\ k^{3\zeta\beta+3\beta+\lceil 14\varsigma \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights.} \end{cases} \tag{5.109}$$

Finally, we are left to bound Γ_5 and Γ_6 defined at (3.24) and (3.25) respectively. By similar arguments as those used in bounding Γ_1 (with $i = j = l = t = 1$ and $s = 1$), one can show

$$\left(\frac{n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,1/2,\beta} \right)^2}{\varrho_n^{(1)} (\varrho_n^{(1)} \varrho_n^{(1)})^2} \right)^{\frac{1}{3}} = \begin{cases} k^{4\beta/3+1+\lceil 35\varsigma/3 \rceil} W_1(n,k)^{5/6} \mathcal{O}(\log^{(d-1)/3} n), \\ \qquad \text{for general weights,} \\ k^{4\beta/3+1/6+\lceil 35\varsigma/3 \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights.} \end{cases} \tag{5.110}$$

and

$$\left(\frac{n\mathbb{Q}\left(f^{(1,1,1,1)}_{1,1,1,\beta} \right)^2}{(\varrho_n^{(1)} \varrho_n^{(1)} \varrho_n^{(1)})^2} \right)^{\frac{1}{4}} = \begin{cases} k^{\beta+3/4+\lceil 21\varsigma/2 \rceil} W_1(n,k)^{3/4} \mathcal{O}(\log^{(d-1)/4} n), \\ \qquad \text{for general weights,} \\ k^{\beta+\lceil 21\varsigma/2 \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights.} \end{cases} \tag{5.111}$$

Therefore, combining (5.95), (5.98), (5.109), (5.110) and (5.111) and recalling that $\Gamma_0 = 0$, we conclude that the sums of all the contributions from Γ_s, $s \in \{0,1,3,4,5,6\}$ with $i = j = l = t$ (after writing the power of the sums as sums of powers, up to constants, using Lemma 6.1) is of the order

$$\begin{cases} k^{6\zeta\beta+6\beta+3/2+\lceil 21\varsigma\rceil} \max_{j\in\{1,4\}} \left(W(n,k)^{1/2+1/j} \mathcal{O}(\log^{(d-1)/j} n) \right), \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for general weights,} \\ k^{6\zeta\beta+6\beta+1/2+\lceil 21\varsigma\rceil} \mathcal{O}(\log^{-(d-1)/2} n), \qquad \text{for uniform weights,} \end{cases} \tag{5.112}$$

where we recall from (3.6) that

$$W(n,k) := \frac{\max_{i=1,\dots,m} \left(\sup_{(x,\eta):|\eta|\le 9} \|W_{nx}(x_{0,i}, \mathcal{P}_{x,\eta})\|_{L_{6+p}} \right)^2}{\min_{i=1,\dots,m} \mathbb{E}\left[\sum_{\boldsymbol{x}} W_{n\boldsymbol{x}}(x_{0,i})^2\right]}.$$

Next, we will consider the case when i, j, l, t are not all equal. According to Theorems 3.2 and 3.3, only $\Gamma_1, \Gamma_4, \Gamma_5$ and Γ_6 will change. We first focus on the case when $i \neq j$, and without loss of generality, take $i = 1$ and $j = 2$.

Again, we start with Γ_1. Note that by (5.77), (5.72) and Cauchy's inequality,

$$\begin{aligned} &\left(f^{(1,2,1,1)}_{1,1,0,\beta,1}(y) \right)^2 \\ &\lesssim (\Omega^4_{1,n} + \Omega^4_{2,n}) \left(n \int_{\mathbb{R}^d} (r^*_{6+p}(x)^2 \vee 1) e^{-\beta r^{(1)}_n(x,y)} g(x) dx \right)^2 \\ &\quad + \Omega^4_{1,n} k^{28\zeta/(6+p/2)} \left(n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,1}}(x)^{14/(6+p/2)} e^{-\beta r^{(1)}_n(x,y)} g(x) dx \right)^2 \\ &\quad + \Omega^4_{2,n} k^{28\zeta/(6+p/2)} \left(n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,2}}(x)^{14/(6+p/2)} e^{-\beta r^{(1)}_n(x,y)} g(x) dx \right)^2 . \end{aligned}$$

Therefore, it suffices to derive bounds for the final additional term compared to (5.83). Denote

$$\mathcal{I}_{\text{add}} := n \int_{\mathbb{R}^d} \left(n \int_{\mathbb{R}^d} c_{\zeta(k+1)^{-1},n,x_{0,2}}(x)^{14/(6+p/2)} e^{-\beta r^{(1)}_n(x,y)} g(x) dx \right)^2 g(y) dy.$$

Fix $\delta \in (0, 1/2(g(x_{0,1}) \wedge g(x_{0,2})))$ and ϵ as in Lemma 5.13 such that the conclusion therein holds for both $x_{0,1}, x_{0,2}$. We argue as in Lemma 5.13 using the partition (5.49) for the inside integral in $\mathcal{I}_{\text{add}}$. For any $\underline{j}_1 \subseteq [d]$ with $|\underline{j}_1| = j_1 \geq 1$, we have

$$\begin{aligned} &n \int_{\mathbb{R}^d} \left(n \int_{\mathcal{C}^{x_{0,1}}_{\underline{j}_1}} c_{\zeta(k+1)^{-1},n,x_{0,2}}(x)^{14/(6+p/2)} e^{-\beta r^{(1)}_n(x,y)} g(x) dx \right)^2 g(y) dy \\ &\lesssim n \int_{\mathbb{R}^d} \left(n \int_{\mathcal{C}^{x_{0,1}}_{\underline{j}_1}} c_{\zeta(k+1)^{-1},n,x_{0,2}}(x)^{14/(6+p/2)} e^{-\frac{\beta n}{k+1}(g(x_{0,1})-\delta)\epsilon^{d-j} \prod_{l\in\underline{j}_1} |x^{(l)}-x^{(l)}_{0,1}|} g(x) \right. \\ &\qquad\qquad \left. \times \mathbb{1}_{y^{\underline{j}_1} \in \text{Rect}(x^{\underline{j}_1}, x^{\underline{j}_1}_{0,1})} dx \right)^2 g(y) dy. \end{aligned}$$

Now, we further bound $c_{\zeta(k+1)^{-1},n,x_{0,2}}(x)$ using Lemma 5.13. In the upper bound therein, the first term

$$n e^{-\alpha(k+1)^{-1} n (g(x_{0,1})-\delta)\epsilon^d} \int_{\mathbb{R}^d} \phi(x) g(x) dx,$$

is exponentially small, so consider the summand corresponding to $\underline{j}_2$ with $|\underline{j}_2| = j_2 \leq d-1$, i.e., $\mathbb{1}_{x^{\underline{j}_2} \in \mathcal{C}_\epsilon(x_{0,2}^{\underline{j}_2})} = 1$. First assume that $\underline{j}_1 \cap \underline{j}_2 = \varnothing$. Writing $a = 14/(6+p/2)$ for ease, we have

$$
\begin{aligned}
& n \int_{\mathbb{R}^d} \left(n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-\zeta\left(\frac{an(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1}\right)|x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}|/2} \right. \\
& \qquad \times \left(\left| \log \left(\zeta \left(\frac{n(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1} \right) |x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}| \right) \right|^{a(j_2-1)} + 1 \right) \\
& \qquad\qquad \left. \times e^{-\frac{\beta n}{k+1}(g(x_{0,1})-\delta)\epsilon^{d-j}|x^{\underline{j}_1} - x_{0,1}^{\underline{j}_1}|} g(x) \mathbb{1}_{y^{\underline{j}_1} \in \mathrm{Rect}(x^{\underline{j}_1}, x_{0,1}^{\underline{j}_1})} dx \right)^2 g(y) dy \\
& \lesssim n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-2\zeta\left(\frac{an(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1}\right)|x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}|/2} \\
& \qquad \times \left(\left| \log \left(\zeta \left(\frac{n(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1} \right) |x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}| \right) \right|^{2a(j_2-1)} + 1 \right) g(x) dx \\
& \quad \times n \int_{\mathbb{R}^{j_1}} \left(n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-2\frac{\beta n}{k+1}(g(x_{0,1})-\delta)\epsilon^{d-j}|x^{\underline{j}_1} - x_{0,1}^{\underline{j}_1}|} g(x) \mathbb{1}_{y^{\underline{j}_1} \in \mathrm{Rect}(x^{\underline{j}_1}, x_{0,1}^{\underline{j}_1})} dx \right) g^{\underline{j}_1}(y^{\underline{j}_1}) dy^{\underline{j}_1} \\
& \lesssim k^3 \mathcal{O}(\log^{j_1+j_2-2} n),
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality in the penultimate step. Since $j_1 + j_2 \leq d$, this is at most of the order $k^3 \log^{d-2} n$.

Now, we are left with the case when $\underline{j}_1 \cap \underline{j}_2 \neq \varnothing$. Let us illustrate with the computation when $\underline{j}_2 \setminus \underline{j}_1 \neq \varnothing$. Let $\underline{j}_2' = \underline{j}_2 \setminus \underline{j}_1$ with $|\underline{j}_2'| = j_2'$. Writing $a = 14/(6+p/2)$ for ease, we have

$$
\begin{aligned}
& n \int_{\mathbb{R}^d} \left(n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-\zeta\left(\frac{an(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1}\right)|x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}|/2} \right. \\
& \qquad \left(\left| \log \left(\zeta \left(\frac{n(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1} \right) |x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}| \right) \right|^{a(j_2-1)} + 1 \right) \\
& \qquad\qquad \left. \times e^{-\frac{\beta n}{k+1}(g(x_{0,1})-\delta)\epsilon^{d-j}|x^{\underline{j}_1} - x_{0,1}^{\underline{j}_1}|} g(x) \mathbb{1}_{y^{\underline{j}_1} \in \mathrm{Rect}(x^{\underline{j}_1}, x_{0,1}^{\underline{j}_1})} dx \right)^2 g(y) dy \\
& \lesssim n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-2\zeta\left(\frac{an(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1}\right)|x^{\underline{j}_2'} - x_{0,2}^{\underline{j}_2'}|/2} \\
& \qquad \left(\left| \log \left(\zeta \left(\frac{n(g(x_{0,2})-\delta)\epsilon^{d-j_2}}{k+1} \right) |x^{\underline{j}_2} - x_{0,2}^{\underline{j}_2}| \right) \right|^{2a(j_2-1)} + 1 \right) g(x) dx \\
& \qquad \times n \int_{\mathbb{R}^{j_1}} \left(n \int_{\mathcal{C}_{\underline{j}_1}^{x_{0,1}}} e^{-\frac{2a' n}{k+1}|x^{\underline{j}_1} - x_{0,1}^{\underline{j}_1}|} g(x) \mathbb{1}_{y^{\underline{j}_1} \in \mathrm{Rect}(x^{\underline{j}_1}, x_{0,1}^{\underline{j}_1})} dx \right) g^{\underline{j}_1}(y^{\underline{j}_1}) dy^{\underline{j}_1} \\
& \lesssim k^3 \mathcal{O}(\log^{j_1+j_2'-2} n),
\end{aligned}
$$

where we have used the Cauchy-Schwarz inequality, and

$$
a' = \min\left\{ \zeta\left(a(g(x_{0,2})-\delta)\epsilon^{d-j_2}/2 \right), \beta(g(x_{0,1})-\delta)\epsilon^{d-j} \right\}.
$$

Since $j_1 + j_2' \leq d$, this is again at most of the order $k^3 \log^{d-2} n$. The case when $\underline{j}_1 \cap \underline{j}_2 \neq \varnothing$, but $\underline{j}_2 \backslash \underline{j}_1 = \varnothing$ can be dealt with similarly as above, by putting the exponential factor corresponding to the common coordinates in one of the exponential factors, instead of in both. Finally, the case when $j_1 = 0$ is trivial, since in this case,

$$e^{-\beta r_n^{(1)}(x,y)} \lesssim k^\beta e^{-\frac{\beta}{k+1} n\epsilon^d (g(x_{0,1})-\delta)}.$$

Combining all the above analysis, we conclude that the additional term appeared $\mathcal{I}_{\text{add}}$ is of lower order compared to the corresponding bound for $f_{1,1,0,\beta,1}^{(1,1,1,1)}$.

Using similar arguments as for $\left(f_{1,1,0,\beta,2}^{(1,2,1,1)}(y)\right)^2$ and $\left(f_{1,1,0,\beta,3}^{(1,2,1,1)}(y)\right)^2$, combining the cases $i = j$ in (5.95) and the case when $i \neq j$, from (3.20) we finally obtain

$$\Gamma_1 = \begin{cases} mk^{2\beta+3/2+\lceil 14\varsigma \rceil} W(n,k) \mathcal{O}(\log^{(d-1)/2} n), & \text{for general weights,} \\ mk^{2\beta+1/2+\lceil 14\varsigma \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights,} \end{cases}$$

where ς is given in (5.94) and $W(n,k)$ is defined in (3.6).

Similar arguments using (5.109), (5.110) and (5.111) as well as (3.23), (3.24) and (3.25) yield

$$\Gamma_4 = \begin{cases} mk^{3\zeta\beta+3\beta+1+\lceil 14\varsigma \rceil} W(n,k) \mathcal{O}(\log^{(d-1)/2} n), & \text{for general weights,} \\ mk^{3\zeta\beta+3\beta+\lceil 14\varsigma \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights,} \end{cases}$$

$$\Gamma_5 = \begin{cases} m^{13/6} k^{4\beta/3+1+\lceil 35\varsigma/3 \rceil} W(n,k)^{5/6} \mathcal{O}(\log^{(d-1)/3} n), & \text{for general weights,} \\ m^{13/6} k^{4\beta/3+1/6+\lceil 35\varsigma/3 \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights,} \end{cases}$$

and

$$\Gamma_6 = \begin{cases} m^2 k^{\beta+3/4+\lceil 21\varsigma/2 \rceil} W(n,k)^{3/4} \mathcal{O}(\log^{(d-1)/4} n), & \text{for general weights,} \\ m^2 k^{\beta+\lceil 21\varsigma/2 \rceil} \mathcal{O}(\log^{-(d-1)/2} n), & \text{for uniform weights.} \end{cases}$$

Proofs of Theorem 3.1 and Corollary 3.1. Putting together all bounds above on Γ_s for $s \in \{0, 1, 3, 4, 5, 6\}$ as well as a similar bound for Γ_2, and Γ_1 with $p_0 = 4$, we obtain the proof of Theorem 3.1 and Corollary 3.1 by invoking Theorems 3.2 and 3.3. □

5.2.6 Proof of Proposition 3.1

Proof of Proposition 3.1. We start with studying the following general integral more carefully. By the Mecke formula, for any function $\phi(x)$ as considered in Lemma 5.19 and Remark 5.4, we have

$$\mathbb{E}\Bigg(\sum_{\boldsymbol{x} \in \mathcal{P}_{ng}} \phi(\boldsymbol{x}) \mathbb{1}_{\boldsymbol{x} \in \mathcal{L}_{n,k}(x_0)} \Bigg) = n \int_{\mathbb{R}^d} \psi(n,k,x_0,x) \phi(x) g(x) dx.$$

When g and ϕ are assumed to be continuous, by Remark 5.4, we in particular have that

$$\lim_{n\to\infty} \frac{n \int_{\mathbb{R}^d} \psi(n,k,x_0,x) \phi(x) g(x) dx}{k \log^{d-1} n} = \frac{2^d}{(d-1)!} \phi(x_0).$$

The rate of convergence for this limit, which we estimate below in (5.116), plays a key role in analyzing the bias.

Let ϕ and g be a Hölder continuous functions at x_0 with parameters $L_\phi, \gamma_\phi > 0$ and $L_g, \gamma_g > 0$, respectively. Note for any $\delta \in (0, 1/2\ g(x_0))$ and $\epsilon > 0$ as in (5.45), according to the proof of Lemma 5.19, by (5.64), (5.65) and (5.66), we have

$$\begin{aligned}
&\frac{2^d}{(d-1)!}\frac{(\phi(x_0)-\delta)(g(x_0)-\delta)}{g(x_0)+\delta}\left(k\log^{d-1}(n\Delta_d'^d\epsilon^d)+k\mathcal{O}\left(\log k\log^{d-2}(n\Delta_d'^d\epsilon^d)\right)\right)\\
&\leq n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx\\
&\leq \frac{2^d}{(d-1)!}\frac{(\phi(x_0)+\delta)(g(x_0)+\delta)}{g(x_0)-\delta}\left(k\log^{d-1}(n\Delta_d^d\epsilon^d)+k\mathcal{O}\left(\log k\log^{d-2}(n\Delta_d^d\epsilon^d)\right)\right)\\
&\quad+\sum_{j=0}^{k-1}ne^{-n(g(x_0)-\delta)\epsilon^d}\frac{(n(g(x_0)-\delta)\epsilon^d)^j}{j!}\int_{\mathbb{R}^d}\phi(x)g(x)dx\\
&\qquad+\sum_{l=1}^{d-1}\frac{2^l\binom{d}{l}\Lambda(l,\epsilon)}{(g(x_0)-\delta)\epsilon^{d-l}}k\mathcal{O}\left(\log k\log^{l-1}(n\Delta_l^l\epsilon^d)\right). \qquad (5.113)
\end{aligned}$$

Now to obtain a rate of convergence, instead of fixing the pair (δ, ϵ) as in the proofs of Theorem 3.1 and Corollary 3.1, we use the assumed Hölder continuity to determine (δ, ϵ) more explicitly as a function of n. We have for any $x \in \mathcal{C}_\epsilon(x_0)$,

$$|g(x_0)-g(x)| \leq L_g(d\epsilon^2)^{\frac{\gamma_g}{2}} = L_g d^{\frac{\gamma_g}{2}}\epsilon^{\gamma_g} =: \delta_g, \qquad (5.114)$$

and

$$|\phi(x_0)-\phi(x)| \leq L_\phi(d\epsilon^2)^{\frac{\gamma_\phi}{2}} = L_\phi d^{\frac{\gamma_\phi}{2}}\epsilon^{\gamma_\phi} =: \delta_\phi. \qquad (5.115)$$

To make sure that $\epsilon^{-(d-l)}\log^{l-1} n = o(\log^{d-1} n)$ in (5.113), we pick $\epsilon \equiv \epsilon(n) = \log^{-\zeta} n$ for some $0 < \zeta < 1$. Thus for $1 \leq l \leq d-1$,

$$\frac{1}{\epsilon^{d-l}}\log^{l-1} n = \log^{(d-l)\zeta+(l-1)} \leq \log^{d-2+\zeta} n = o(\log^{d-1} n),$$

while for $l = d-1$, we have

$$\frac{1}{\epsilon^{d-l}}\log^{l-1} n = \log^{d-2+\zeta} n > \log^{d-2} n.$$

Now, from (5.114), we have $\delta_g = L_g d^{\frac{\gamma_g}{2}}\log^{-\zeta\gamma_g} n$ so that choosing $n \geq \exp\left[\left(\frac{2L_g d^{\frac{\gamma_g}{2}}}{g(x_0)}\right)^{\frac{1}{\zeta\gamma_g}}\right]$, ensures that $\delta_g < 1/2\ g(x_0)$. Also, noting that $\delta_\phi = L_\phi d^{\frac{\gamma_\phi}{2}}\log^{-\zeta\gamma_\phi} n$, starting with $\delta \equiv \delta(n) = \delta_g \vee \delta_\phi$, we have that the above choice of $\epsilon \equiv \epsilon(n)$ satisfies (5.45). Thus from (5.113) we obtain

$$\begin{aligned}
&\left|\frac{n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx}{k\log^{d-1} n}-\frac{2^d}{(d-1)!}\phi(x_0)\right|\\
&=\mathcal{O}\left((\log^{-\zeta(\gamma_g\wedge\gamma_\phi)} n)\vee(\log k\log^{-(1-\zeta)} n)\right). \qquad (5.116)
\end{aligned}$$

Moreover, when $\phi(x_0) = 0$, we have

$$\left|\frac{n\int_{\mathbb{R}^d}\psi(n,k,x_0,x)\phi(x)g(x)dx}{k\log^{d-1} n}\right| = \mathcal{O}(\delta_g \vee \delta_\phi) = \mathcal{O}(\log^{-\zeta(\gamma_g\wedge\gamma_\phi)} n). \qquad (5.117)$$

The above is now being applied for the actual proof of Proposition 3.1. From (3.4), by Fubini's theorem, we have

$$
\begin{aligned}
&\mathbb{E} r_{n,k}(x_0) - r_0(x_0) \\
&= \mathbb{E}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)} (\boldsymbol{y_x} - r_0(x_0)) \right) \\
&= \mathbb{E}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)} [(\boldsymbol{y_x} - r_0(\boldsymbol{x})) + (r_0(\boldsymbol{x}) - r_0(x_0))] \right) \\
&= \mathbb{E}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{L_{n,k}(x_0)} (r_0(\boldsymbol{x}) - r_0(x_0)) \right),
\end{aligned}
$$

where the final step is due to the fact that $\mathbb{E}[y_{\boldsymbol{x}}|\boldsymbol{x}] = r_0(x)$. In order to apply (5.117) with $\phi(x) = r_0(x) - r_0(x_0)$, we aim to substitute $L_{n,k}(x_0)$ with $\mathbb{E}L_{n,k}(x_0)$ and bound the error term. To this end, by the triangle inequality we can write

$$
\begin{aligned}
|\mathbb{E} r_{n,k}(x_0) - r_0(x_0)| &\le \mathbb{E}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \frac{\mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)}}{\mathbb{E}L_{n,k}(x_0)} (r_0(\boldsymbol{x}) - r_0(x_0)) \right) \\
&+ \mathbb{E}\left(\sum_{(\boldsymbol{x},\boldsymbol{\varepsilon_x})\in\mathcal{P}_{n\breve{g}}} \left| \frac{1}{L_{n,k}(x_0)} - \frac{1}{\mathbb{E}L_{n,k}(x_0)} \right| \mathbb{1}_{\boldsymbol{x}\in\mathcal{L}_{n,k}(x_0)} (r_0(\boldsymbol{x}) - r_0(x_0)) \right) \\
&=: \mathcal{J}_1 + \mathcal{J}_2.
\end{aligned} \tag{5.118}
$$

Plugging $\phi(x) = r_0(x) - r_0(x_0)$, which is Hölder continuous at x_0 with parameters $L_1, \gamma_1 > 0$ by our assumption, in (5.117) and using Lemma 5.19, we conclude that for any $\zeta \in (0,1)$,

$$
\mathcal{J}_1 = \mathcal{O}(\log^{-\zeta(\gamma_g \wedge \gamma_1)} n).
$$

To estimate $\mathcal{J}_2$, we proceed with the following two subcases. First, on $\mathcal{E} := \{|L_{n,k}(x_0) - \mathbb{E}L_{n,k}(x_0)| < (\mathbb{E}L_{n,k}(x_0))^{3/4}\}$, we have

$$
\left| \frac{1}{\mathbb{E}L_{n,k}(x_0)} - \frac{1}{L_{n,k}(x_0)} \right| = \frac{|L_{n,k}(x_0) - \mathbb{E}L_{n,k}(x_0)|}{L_{n,k}(x_0)\mathbb{E}L_{n,k}(x_0)} < \frac{1}{L_{n,k}(x_0)(\mathbb{E}L_{n,k}(x_0))^{\frac{1}{4}}}.
$$

Next, on $\mathcal{E}^c = \{|L_{n,k}(x_0) - \mathbb{E}L_{n,k}(x_0)| \ge (\mathbb{E}L_{n,k}(x_0))^{3/4}\}$, we simply bound

$$
\left| \frac{1}{\mathbb{E}L_{n,k}(x_0)} - \frac{1}{L_{n,k}(x_0)} \mathbb{1}_{L_{n,k}(x_0)\ge 1} \right| \lesssim 1, \tag{5.119}
$$

which holds since $\mathbb{E}L_{n,k}(x_0) = \mathcal{O}(k \log^{d-1} n)$ by Remark 5.4. On the other hand, arguing as in the proof of Lemma 5.20, we have for any $r \ge 1$ that

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}^c) &= \mathbb{P}(|L_{n,k}(x_0) - \mathbb{E}L_{n,k}(x_0)| \ge (\mathbb{E}L_{n,k}(x_0))^{3/4}) \\
&\lesssim \frac{\mathbb{E}(L_{n,k}(x_0) - \mathbb{E}L_{n,k}(x_0))^{2r}}{(\mathbb{E}L_{n,k}(x_0))^{3r/2}} \\
&\lesssim_r (k \log^{d-1} n)^{-3r/2} k^{r(\tau_{\text{conc}}+1)} \log^{r(d-1)} n
\end{aligned}
$$

$$= k^{r(\tau_{\text{conc}}+1)-3r/2} \log^{-(d-1)r/2} n, \tag{5.120}$$

where τ_{conc} are defined in the proof of Lemma 5.20 and it satisfies $\tau_{\text{conc}} \leq 10/r$. Taking $r = 29$ so that $\tau_{\text{conc}} \leq 10/29$, we thus have

$$\mathbb{P}(\mathcal{E}^c) \lesssim k^{-9/2} \log^{-29(d-1)/2} n.$$

Using the above bounds with the fact that both $r_0(\boldsymbol{x})$ and $r_0(x_0)$ are uniformly bounded (almost surely) by our assumption, we now obtain

$$\begin{aligned}
\mathcal{J}_2 &\lesssim \left| \mathbb{E}\left(\mathbb{1}_{\mathcal{E}} \left| \frac{1}{\mathbb{E}L_{n,k}(x_0))} - \frac{1}{L_{n,k}(x_0)} \right| \sum_{\boldsymbol{x} \in \mathcal{P}_{ng}} \mathbb{1}_{\boldsymbol{x} \in \mathcal{L}_{n,k}(x_0)} \right) \right| \\
&\quad + \left| \mathbb{E}\left(\mathbb{1}_{\mathcal{E}^c} \left| \frac{1}{\mathbb{E}L_{n,k}(x_0)} - \frac{1}{L_{n,k}(x_0)} \right| \sum_{\boldsymbol{x} \in \mathcal{P}_{ng}} \mathbb{1}_{\boldsymbol{x} \in \mathcal{L}_{n,k}(x_0)} \right) \right| \\
&\lesssim (\mathbb{E}L_{n,k}(x_0))^{-1/4} + \mathbb{E}\Big(L_{n,k}(x_0) \mathbb{1}_{\mathcal{E}^c} \Big) \\
&\lesssim (k \log^{d-1} n)^{-1/4} + \sqrt{\mathbb{E}L_{n,k}(x_0)^2 \mathbb{P}(\mathcal{E}^c)} \\
&\lesssim (k \log^{d-1} n)^{-1/4} + k^{-5/4} \log^{-25(d-1)/4} n = \mathcal{O}((k \log^{d-1} n)^{-1/4}),
\end{aligned}$$

where in the penultimate step, we have also used Remark 5.6. Putting the bounds for $\mathcal{J}_1$ and $\mathcal{J}_2$ together in (5.118) now yields

$$|\mathbb{E}r_{n,k}(x_0) - r_0(x_0)| = \mathcal{O}\left((\log^{-\zeta(\gamma_g \wedge \gamma_1)} n) \vee (k^{-1/4} \log^{-(d-1)/4} n) \right) \tag{5.121}$$

completing the proof. □

6 Auxillary Results

Lemma 6.1. *The two inequalities below, follow by elementary analysis:*

- *For* $\{a_i\}_{i=1}^{\ell} \subset \mathbb{R}$ *and* $\iota \geq 1$, $\left(\sum_{i=1}^{\ell} a_i\right)^{\iota} \leq \ell^{\iota-1} \sum_{i=1}^{\ell} a_i^{\iota}$.
- *For* $\{a_i\}_{i=1}^{\iota} \subset \mathbb{R}_+$ *and* $0 < \iota < 1$, *we have that* $\left(\sum_{i=1}^{\ell} a_i\right)^{\iota} \leq \sum_{i=1}^{\ell} a_i^{\iota}$.

Lemma 6.2 (Poisson concentration). *Let* $h : [-1, \infty) \to \mathbb{R}$ *be given by* $h(x) := \frac{2(1+x)\log(1+x)-x}{x^2}$, *and let* $\text{Poi}(\lambda_0)$ *be a Poisson random variable with parameter* $\lambda_0 > 0$. *Then, for any* $x > 0$, *we have*

$$\mathbb{P}(\text{Poi}(\lambda_0) \geq \lambda_0 + x) \leq e^{-\frac{x^2}{2\lambda_0} h\left(\frac{x}{\lambda_0}\right)},$$

and, for any $0 < x < \lambda_0$,

$$\mathbb{P}(\text{Poi}(\lambda_0) \leq \lambda_0 - x) \leq e^{-\frac{x^2}{2\lambda_0} h\left(-\frac{x}{\lambda_0}\right)}.$$

In particular, this implies that for $x > 0$,

$$\max\left[\mathbb{P}(\text{Poi}(\lambda_0) \geq \lambda_0 + x), \mathbb{P}(\text{Poi}(\lambda_0) \leq \lambda_0 - x)\right] \leq e^{-\frac{x^2}{2(\lambda_0 + x)}}.$$

The proof of Lemma 6.2 follows by a standard use of Chernoff's bound, and is hence omitted.

6.1 Background on TDA

Definition 6.1 (Simplicial Complexes)**.** *An* abstract simplicial complex *over a (finite) vertex set* $\mathbb{A}_n := \{a_1, \ldots, a_n\}$ *is a collection* S *of subsets of* $\mathbb{A}_n$ *with the properties that*

(i) $\{a_i\} \in S, i = 1, \ldots, n,$

(ii) $\sigma \in S$ *and* $\tau \subset \sigma$ *implies that* $\tau \in S$.

Every $\tau \subset \sigma$ *is called a* face *of* σ. *Every* $\sigma \in S$ *with* $|\sigma| = \ell + 1, \ell \geq 0,$ *is called* ℓ-simplex.

Note that the vertices do not necessarily have to be elements of a Euclidean space. If they are (affinely independent) elements of $\mathbb{R}^d$, one can think of every simplex of order $\ell \leq d$ as a convex hull of $\ell + 1$ (affinely independent) vertices, so that 0-simplices are points, 1-simplices are lines, 2-simplices are triangles, etc. The following two types of simplicial complexes, the Vietoris-Rips complex (VR complex) and the Čech complex, are widely used in TDA.

Definition 6.2 (VR Complex)**.** *Following the definition 6.1, let the vertex set* $\mathbb{A}_n$ *be in a metric space with the metric* d. *Then, the VR complex* $VR_r(\mathbb{A}_n)$ *for a given postive real number* $r > 0$ *is a collection of simplices, where a simplex* $\sigma \in VR_r(\mathbb{A}_n)$ *if and only if for any pair of vertices* $a_i, a_j \in \sigma$, $d(a_i, a_j) < r$.

Definition 6.3 (Čech Complex)**.** *Following the definition 6.1, let the vertex set* $\mathbb{A}_n$ *be in a metric space with the metric* d. *Then, the Čech complex* $C_r(\mathbb{A}_n)$ *for a given postive real number* $r > 0$ *is a collection of simplices, where a simplex* $\sigma := \{a_i\}_{i \in I} \in C_r(\mathbb{A}_n)$ *for some* $I \subset \{1, 2, ..., n\}$ *if and only if for* $\underset{i \in I}{\cap} B_{a_i}\left(\frac{r}{2}\right) \neq \varnothing$.

Definition 6.4 (Filtrations)**.** *A filtration* $\mathcal{S}$ *of a simplicial complex* S *is a nested sequence of simplicial complexes* $\varnothing = S_0 \subset S_1 \subset \cdots \subset S_I = S$, *where* $S_i = S_{i-1} \cup \sigma_i, i = 1, \ldots, I$ *for some* $\sigma_i \in S$. *A filtration is thus equivalent to an ordering of the simplices in the complex. Usually, a filtration is given in form of a filtration function* $\psi : S \to \mathbb{R}$ *that assigns a real value* $\psi(\sigma)$ *to each simplex* $\sigma \in S$. *The filtration itself is then defined via* $S(r) = \{\sigma \in S : \psi(\sigma) \leq r\}$. *Note that while* r *is a continuous parameter, there are only finitely many values of* r *at which the complex is changing for a simplex over a finite set of vertices as considered here. Additionally, the parameter* r *here is called as the filtration parameter or the filtration time.*

Remark 6.1. *The real number* r *in Definition 6.2 and 6.3 is called the filtration parameter/time for the VR complex and the Čech complex respectively.*

References

R. Adamczak, B. Polaczyk, and M. Strzelecki. Modified log-Sobolev inequalities, Beckner inequalities and moment estimates. *J. Funct. Anal.*, 282(7):Paper No. 109349, 2022. ISSN 0022-1236,1096-0783.

Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Comput.*, 9(7):1545–1588, 1997.

S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. *Ann. Stat.*, 47(2):1148–1178, 2019.

Z.-D. Bai, L. Devroye, H.-K. Hwang, and T.-H. Tsai. Maxima in hypercubes. *Random Structures & Algorithms*, 27(3):290–309, 2005.

Z.-D. Bai, S. Lee, and M. D. Penrose. Rooted edges of a minimal directed spanning tree on random points. *Adv. Appl. Probab.*, 38(1):1–30, 2006.

Y. Baryshnikov and J. Yukich. Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.*, 15(1A):213–253, 2005.

M. Belgiu and L. Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm.*, 114:24–31, 2016.

T. Berrett, R. Samworth, and M. Yuan. Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist*, 47(1):288–318, 2019.

C. Bhattacharjee and I. Molchanov. Gaussian approximation for sums of region-stabilizing scores. *Electron. J. Probab.*, 27:1–27, 2022.

G. Biau. Analysis of a random forests model. *J. Mach. Learn. Res.*, 13(1):1063–1095, 2012.

G. Biau and L. Devroye. On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. *J. Multivar. Anal.*, 101 (10):2499–2518, 2010.

G. Biau and E. Scornet. A random forest guided tour. *Test*, 25:197–227, 2016.

G. Biau, L. Devroye, and G. Lugosi. Consistency of random forests and other averaging classifiers. *J. Mach. Learn. Res.*, 9(9), 2008.

J.-D. Boissonnat, F. Chazal, and M. Yvinec. *Geometric and Topological Inference*, volume 57. Cambridge University Press, 2018.

L. Breiman. Bagging predictors. *Mach. Learn.*, 24:123–140, 1996a.

L. Breiman. Heuristics of instability and stabilization in model selection. *Ann. Stat.*, 24(6):2350–2383, 1996b.

L. Breiman. Some infinity theory for predictor ensembles. 2000. URL https://api.semanticscholar.org/CorpusID:10844956.

L. Breiman. Random forests. *Mach. Learn.*, 45:5–32, 2001.

L. Breiman. Consistency for a simple model of random forests. 2004. URL https://api.semanticscholar.org/CorpusID:123042984.

M. D. Cattaneo, J. M. Klusowski, and W. G. Underwood. Inference with mondrian random forests. *arXiv preprint arXiv:2310.09702*, 2023.

S. Chatterjee. A new method of normal approximation. *Ann. Probab.*, 36(4):1584–1610, 2008.

S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1):1–40, 2009.

S. Chatterjee and S. Sen. Minimal spanning trees and Stein's method. *Ann. Appl. Probab.*, 27(3): 1588–1645, 2017.

L. H. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*, volume 2. Springer, 2011.

C.-M. Chi, P. Vossler, Y. Fan, and J. Lv. Asymptotic properties of high-dimensional random forests. *Ann. Stat.*, 50(6):3415–3438, 2022.

C. Domeniconi, J. Peng, and D. Gunopulos. Locally adaptive metric nearest-neighbor classification. *IEEE T. Pattern Anal. Mach. Intell.*, 24(9):1281–1285, 2002.

H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction.* American Mathematical Society, 2010.

G. Englund. A remainder term estimate for the normal approximation in classical occupancy. *Ann. Probab.*, pages 684–692, 1981.

A. Goel, K. D. Trinh, and K. Tsunoda. Strong law of large numbers for Betti numbers in the thermodynamic regime. *J. Stat. Phys*, 174(4):865–892, 2019.

S. Haghiri, D. Garreau, and U. Luxburg. Comparison-based random forests. In *International Conference on Machine Learning*, pages 1871–1880. PMLR, 2018.

T. Hastie and R. Tibshirani. Discriminant adaptive nearest neighbor classification and regression. *Adv. Neural. Inf. Process. Syst.*, 8, 1995.

T. K. Ho. The random subspace method for constructing decision forests. *IEEE T. Pattern Anal. Mach. Intell.*, 20(8):832–844, 1998.

O. Kallenberg and O. Kallenberg. *Foundations of Modern Probability*, volume 2. Springer, 1997.

H. Kesten and S. Lee. The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.*, 6(2):495–527, 1996.

M. Khalilia, S. Chakraborty, and M. Popescu. Predicting disease risks from highly imbalanced data using random forest. *BMC Med. Inform. Decis.*, 11:1–13, 2011.

J. Klusowski. Sharp analysis of a simple model for random forests. In *Int. Conf. Artif. Intell. Stat.*, pages 757–765. PMLR, 2021.

J. M. Klusowski and P. Tian. Large scale prediction with decision trees. *J. Am. Stat. Assoc.*, 2022.

L. Kozachenko and N. Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23(2):9–16, 1987.

J. Krebs, B. Roycraft, and W. Polonik. On approximation theorems for the euler characteristic with applications to the bootstrap. *Electron. J. Stat.*, 15(2):4462–4509, 2021.

R. Lachièze-Rey and G. Peccati. New Berry–Esseen bounds for functionals of binomial point processes. *Ann. Appl. Probab.*, 27(4):1992–2031, 2017.

R. Lachièze-Rey, M. Schulte, and J. Yukich. Normal approximation for stabilizing functionals. *Ann. Appl. Probab.*, 29(2):931–993, 2019.

R. Lachièze-Rey, G. Peccati, and X. Yang. Quantitative Two-scale Stabilization on the Poisson Space. *Ann. Appl. Probab.*, 32(4):3085–3145, 2022.

G. Last and M. Penrose. Poisson process Fock space representation, chaos expansion and covariance inequalities. *Probab. Theory Related Fields*, 150(3):663–690, 2011.

G. Last, G. Peccati, and M. Schulte. Normal Approximation on Poisson Spaces: Mehler's Formula, Second Order Poincaré Inequalities and Stabilization. *Probab. Theory Related Fields*, 165(3): 667–723, 2016.

Y. Lin and Y. Jeon. Random forests and adaptive nearest neighbors. *J. Am. Stat. Assoc.*, 101(474): 578–590, 2006.

A. Mastin and P. Jaillet. Log-quadratic bounds for the Gaussian Q-function. *arXiv preprint arXiv:1304.2488*, 2013.

N. Meinshausen. Quantile regression forests. *J. Mach. Learn. Res.*, 7(6), 2006.

L. Mentch and G. Hooker. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.*, 17(1):841–881, 2016.

L. Mentch and S. Zhou. Randomization as regularization: A degrees of freedom explanation for random forest success. *J. Mach. Learn. Res.*, 21(1):6918–6953, 2020.

J. Mourtada, S. Gaïffas, and E. Scornet. Minimax optimal rates for mondrian trees and forests. *Ann. Stat.*, 48(4):2253–2276, 2020.

I. Nourdin, G. Peccati, and G. Reinert. Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.*, 257(2):593–609, 2009.

T. Owada and A. Thomas. Limit theorems for process-level Betti numbers for sparse and critical regimes. *Adv. in Appl. Probab.*, 52(1):1–31, 2020.

G. Peccati and M. Reitzner. *Stochastic Analysis for Poisson Point Processes: Malliavin Calculus, Wiener-Itô Chaos Expansions and Stochastic Geometry*, volume 7. Springer, 2016.

G. Peccati, J. L. Solé, M. S. Taqqu, and F. Utzet. Stein's method and normal approximation of Poisson functionals. *Ann. Probab.*, 38(2):443–478, 2010.

E. A. Peköz, A. Röllin, and N. Ross. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.*, 23(3):1188–1218, 2013.

W. Peng, T. Coleman, and L. Mentch. Rates of convergence for random forests via generalized u-statistics. *Electron. J. Stat.*, 16(1):232–292, 2022.

M. Penrose. *Random Geometric Graphs*, volume 5. OUP Oxford, 2003.

M. Penrose. Multivariate spatial central limit theorems with applications to percolation and spatial graphs. *Ann. Probab.*, 33(5):1945–1991, 2005.

M. Penrose. Gaussian limts for random geometric measures. *Electron. J. Probab.*, 12:989–1035, 2007.

M. Penrose and J. Yukich. Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.*, pages 1005–1041, 2001.

M. Penrose and J. Yukich. Normal approximation in geometric probability. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 37–58, 2005.

M. Penrose and J. Yukich. Limit theory for point processes in manifolds. *Ann. Appl. Probab.*, 23 (6):2161–2211, 2013.

Y. Qi. Random forest for bioinformatics. In *Ensemble Machine Learning: Methods and applications*, pages 307–323. Springer, 2012.

T. Schreiber. Limit theorems in stochastic geometry. *New Perspect. Stoch. Geom.*, pages 111–144, 2010.

M. Schulte and J. Yukich. Rates of multivariate normal approximation for statistics in geometric probability. *Ann. Appl. Probab.*, 33(1):507 – 548, 2023a.

M. Schulte and J. Yukich. Rates of multivariate normal approximation for statistics in geometric probability. *Ann. Appl. Probab.*, 33(1):507–548, 2023b.

M. Schulte and J. E. Yukich. Multivariate second order poincaré inequalities for poisson functionals. *Electron. J. Probab.*, 24:1–42, 2019.

E. Scornet. On the asymptotics of random forests. *J. Multivar. Anal.*, 146:72–83, 2016a.

E. Scornet. Random forests and kernel methods. *IEEE. T. Inform. Theory*, 62(3):1485–1500, 2016b.

E. Scornet, G. Biau, and J.-P. Vert. Consistency of random forests. *Ann. Stat.*, 43(4):1716–1741, 2015.

Z. Shi, K. Balasubramanian, and W. Polonik. Supplement to "a flexible approach for normal approximation of geometric and topological statistics". 2023.

Z. Shi, K. Balasubramanian, and W. Polonik. A flexible approach for normal approximation of geometric and topological statistics. *Bernoulli (to appear)*, 2023+.

S. Singh and B. Póczos. Finite-sample analysis of fixed-k nearest neighbor density functional estimators. *Adv. Neural Inf. Process. Syst.*, 29, 2016.

Y. S. Tan, A. Agarwal, and B. Yu. A cautionary tale on fitting decision trees to data from additive models: Generalization lower bounds. In *Int. Conf. Artif. Intell. Stat.*, pages 9663–9685. PMLR, 2022.

A. Thomas and T. Owada. Functional limit theorems for the Euler characteristic process in the critical regime. *Adv. in Appl. Probab.*, 53(1):57–80, 2021.

K. D. Trinh. A remark on the convergence of Betti numbers in the thermodynamic regime. *Pacific J. Math. Ind.*, 9(1):1–7, 2017.

A. Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge university press, 2000.

S. Wager. Asymptotic theory for random forests. *arXiv preprint arXiv:1405.0352*, 2014.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.*, 113(523):1228–1242, 2018.

S. Wager, T. Hastie, and B. Efron. Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.*, 15(1):1625–1651, 2014.

L. Wu. A new modified logarithmic sobolev inequality for poisson point processes and several applications. *Probab. Theory Related Fields*, 118(3):427–438, 2000.

D. Yogeshwaran, E. Subag, and R. Adler. Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields*, 167(1):107–142, 2017.

J. Yukich. Surface order scaling in stochastic geometry. *Ann. Appl. Probab.*, 25(1):177–210, 2015.

Z. Zhou, L. Mentch, and G. Hooker. V-statistics and variance estimation. *J. Mach. Learn. Res.*, 22 (1):13112–13159, 2021.