Normal Approximation in Geometric Probability

Mathew D. Penrose[1] and J. E. Yukich[2]

University of Bath and Lehigh University

February 1, 2008

Abstract

We use Stein's method to obtain bounds on the rate of convergence for a class of statistics in geometric probability obtained as a sum of contributions from Poisson points which are exponentially stabilizing, i.e. locally determined in a certain sense. Examples include statistics such as total edge length and total number of edges of graphs in computational geometry and the total number of particles accepted in random sequential packing models. These rates also apply to the 1-dimensional marginals of the random measures associated with these statistics.

[1] Department of Mathematical Sciences, University of Bath, Bath BA1 7AY, United Kingdom: m.d.penrose@bath.ac.uk

[2] Department of Mathematics, Lehigh University, Bethlehem PA 18015: joseph.yukich@lehigh.edu

[2] Research supported in part by NSA grant MDA904-01-1-0029 and NSF grant DMS-0203720

1 Introduction

In the study of limit theorems for functionals on Poisson or binomial spatial point processes, the notion of *stabilization* has recently proved to be a useful unifying concept [4, 11, 13]. Laws of large numbers and central limit theorems can be proved in the general setting of functionals satisfying an abstract 'stabilization' property whereby the insertion of a point into a Poisson process has only a local effect in some sense. These results can then be applied to deduce limit laws for a great variety of particular functionals, including those concerned with the minimal spanning tree, the nearest neighbor graph, Voronoi and Delaunay graphs, packing, and germ-grain models.

Several different techniques are available for proving general central limit theorems for stabilizing functionals. These include a martingale approach [11] and a method of moments [4]. In the present work, we revisit a third technique for proving central limit theorems for stabilizing functionals on Poisson point processes, which was introduced by Avram and Bertsimas [1]. This method is based on the normal approximation of sums of random variables which are 'mostly independent of one another' in a sense made precise via dependency graphs, which in turn is proved via Stein's method [14]. It has the advantage of providing explicit error bounds and rates of convergence.

We extend the work of Avram and Bertsimas in several directions. First, whereas in [1] attention was restricted to certain particular functionals, here we derive a general result holding for arbitrary functionals satisfying a stabilization condition which can then be checked rather easily for many special cases. Second, we consider non-uniform point process intensities and do not require the functionals to be translation invariant. Third, we improve on the rates of convergence in [1] by making use of the recent refinement by Chen and Shao [7] of previous normal approximation results for sums of 'mostly independent' variables. Finally, we apply the methods not only to random variables obtained by summing some quantity over Poisson points, but to the associated *random point measures*, thereby recovering many of the results of Baryshnikov and Yukich [4] on convergence of these measures, and without requiring higher order moment calculations. We add to [4] by providing information about the rate of convergence, and relaxing the continuity conditions required in [4] for test functions and point process intensities.

A brief comparison between the methods of deriving central limit theorems for functionals of spatial point processes is warranted. Only the dependency graph method used here, to date, has yielded error bounds and rates of convergence. On the other hand, our method requires bounds on the tail of the 'radius of stabilization' (i.e., on the range of the local effect of an inserted point). The martingale method, in contrast, requires only that this radius be almost surely finite, and for this reason is applicable to some examples such as those concerned with the minimal spanning tree, for which no tail bounds are known and which therefore lie beyond the scope of the present work. The moment method [4] and martingale method [10], unlike the dependency graph method, provide information about the variance of the Gaussian limits. The moment method has also been used [5] to establish moderate scale limit behavior of functionals of spatial point processes. Whereas the moment method requires exponential tail bounds for the radius of stabilization, one of our central limit theorems (Theorem 2.2) requires only that this tail $\tau(t)$ decay as a (large) negative power of t.

With regard to ease of use in applications, the dependency graph method and method of moments require checking tail bounds for the radius of stabilization, which is usually straightforward where possible at all. The method of moments requires a more complicated (though checkable) version of the bounded moments condition (2.5) below (see [4]). The dependency graph method requires some separate calculation of variances if one wishes to identify explicitly the variance of the limiting normal variable. The martingale method requires the checking of slightly more subtle versions of the stabilization conditions needed here [10, 11].

2 General results

Let $d \geq 1$ be an integer. For the sake of generality, we consider *marked* point processes in $\mathbb{R}^d$. Let $(\mathcal{M}, \mathcal{F}_{\mathcal{M}}, \mathbb{P}_{\mathcal{M}})$ be a probability space (the *mark space*). Let $\xi((x,s);\mathcal{X})$ be a measurable $\mathbb{R}$-valued function defined for all pairs $((x,s),\mathcal{X})$, where $\mathcal{X} \subset \mathbb{R}^d \times \mathcal{M}$ is finite and where $(x,s) \in \mathcal{X}$ (so $x \in \mathbb{R}^d$ and $s \in \mathcal{M}$). When $(x,s) \in (\mathbb{R}^d \times \mathcal{M}) \setminus \mathcal{X}$, we abbreviate notation and write $\xi((x,s);\mathcal{X})$ instead of $\xi((x,s);\mathcal{X} \cup \{(x,s)\})$.

Given $\mathcal{X} \subset \mathbb{R}^d \times \mathcal{M}$, $a > 0$ and $y \in \mathbb{R}^d$, we let $a\mathcal{X} := \{(ax,t) : (x,t) \in \mathcal{X}\}$ and $y + \mathcal{X} := \{(y+x,t) : (x,t) \in \mathcal{X}\}$; in other words, translation and scalar multiplication on $\mathbb{R}^d \times \mathcal{M}$ act only on the first component. For all $\lambda > 0$ let

$$\xi_\lambda((x,s);\mathcal{X}) := \xi((x,s); x + \lambda^{1/d}(-x + \mathcal{X})).$$

We say ξ is *translation invariant* if $\xi((y+x,s); y+\mathcal{X}) = \xi((x,s);\mathcal{X})$ for all $y \in \mathbb{R}^d$, all $(x,s) \in \mathbb{R}^d \times \mathcal{M}$ and all finite $\mathcal{X} \subset \mathbb{R}^d \times \mathcal{M}$. When ξ is translation invariant, the functional ξ_λ simplifies to $\xi_\lambda((x,s);\mathcal{X}) = \xi((\lambda^{1/d}x,s);\lambda^{1/d}\mathcal{X})$.

Let κ be a probability density function on $\mathbb{R}^d$ with compact support $A \subset \mathbb{R}^d$. For all $\lambda > 0$, let $\mathcal{P}_\lambda$ denote a Poisson point process in $\mathbb{R}^d \times \mathcal{M}$ with intensity measure $(\lambda\kappa(x)dx) \times \mathbb{P}_{\mathcal{M}}(ds)$. We shall assume throughout that κ is bounded with supremum denoted $\|\kappa\|_\infty$.

Let $(A_\lambda, \lambda \geq 1)$ be a family of Borel subsets of A. The simplest case, with $A_\lambda = A$ for all λ, covers all examples considered here; we envisage possibly using the general case in future work.

The following notion of exponential stabilization, adapted from [4], plays a central role in all that follows. For $x \in \mathbb{R}^d$ and $r > 0$, let $B_r(x)$ denote the Euclidean ball centered at x of radius r. Let U denote a random element of $\mathcal{M}$ with distribution $\mathbb{P}_{\mathcal{M}}$, independent of $\mathcal{P}_\lambda$.

Definition 2.1 *ξ is* exponentially stabilizing *with respect to κ and $(A_\lambda)_{\lambda\geq 1}$ if for all $\lambda \geq 1$ and all $x \in A_\lambda$, there exists an a.s. finite random variable $R := R(x,\lambda)$ (*a radius of stabilization *for ξ at x) such that for all finite $\mathcal{X} \subset (A \setminus B_{\lambda^{-1/d}R}(x)) \times \mathcal{M}$, we have*

$$\xi_\lambda\left((x,U);[\mathcal{P}_\lambda \cap (B_{\lambda^{-1/d}R}(x)\times\mathcal{M})] \cup \mathcal{X}\right) = \xi_\lambda\left((x,U);\mathcal{P}_\lambda \cap (B_{\lambda^{-1/d}R}(x)\times\mathcal{M})\right), \tag{2.1}$$

and moreover the tail probability $\tau(t)$ defined for $t > 0$ by

$$\tau(t) := \sup_{\lambda\geq 1,\ x\in A_\lambda} P[R(x,\lambda) > t] \tag{2.2}$$

satisfies

$$\limsup_{t\to\infty} t^{-1}\log\tau(t) < 0. \tag{2.3}$$

For $\gamma > 0$, we say ξ is polynomially stabilizing of order γ *if the above conditions hold with (2.3) replaced by the condition* $\limsup_{t\to\infty} t^\gamma \tau(t) < \infty$.

Condition (2.1) may be cast in a more transparent form as follows. Each point of $\mathcal{X}$ is a pair (x,U), with $x \in \mathbb{R}^d$ and $U \in \mathcal{M}$, but for notational convenience we can view it as a point x in $\mathbb{R}^d$ carrying a mark $U := U_x$. Then we can view $\mathcal{X}$ as a point set in $\mathbb{R}^d$ with each point carrying a mark in $\mathcal{M}$. With this interpretation, (2.1) stipulates that for all finite (marked) $\mathcal{X} \subset A \setminus B_{\lambda^{-1/d}R}(x)$, we have

$$\xi_\lambda\left(x;(\mathcal{P}_\lambda \cap B_{\lambda^{-1/d}R}(x)) \cup \mathcal{X}\right) = \xi_\lambda\left(x;\mathcal{P}_\lambda \cap B_{\lambda^{-1/d}R}(x)\right). \tag{2.4}$$

Roughly speaking, $R := R(x,\lambda)$ is a radius of stabilization if the value of $\xi_\lambda(x;\mathcal{P}_\lambda)$ is unaffected by changes to the points outside $B_{\lambda^{-1/d}R}(x)$.

Functionals of spatial point processes often satisfy exponential stabilization (2.1) (or (2.4)); here is an example. Suppose $\mathcal{M} = [0,1]$ and $\mathbb{P}_M$ is the uniform distribution on $[0,1]$. Suppose that A is convex or polyhedral, and κ is bounded away from zero on A. Suppose a measurable function $(q(x), x \in A)$ is specified, taking values in $[0,1]$. Adopting the conventions of the preceding paragraph, for a marked point set $\mathcal{X} \subset \mathbb{R}^d$ let us denote each point $x \in \mathcal{X}$ as 'red' if $U_x \leq q(x)$ and as 'green' if $U_x > q(x)$. Let $\xi(x;\mathcal{X})$ take the value 0 if the nearest neighbor of x in $\mathcal{X}$ has the same color as x, and take the value 1 otherwise. Note that ξ is *not* translation invariant in this example, unless $q(\cdot)$ is constant. For $x \in A$ let $R := R(x,\lambda)$ denote the distance between $\lambda^{1/d}x$ and its nearest neighbor in $\lambda^{1/d}\mathcal{P}_\lambda$. Then stabilization (2.4) holds because points lying outside $B_{\lambda^{-1/d}R}(x)$ will not change the value of $\xi_\lambda(x;\mathcal{P}_\lambda)$, and it is easy to see that R has exponentially decaying tails. This example is relevant to the multivariate two-sample test described by Henze [8]. See Section 3 for further examples.

Definition 2.2 *ξ has a moment of order $p > 0$ (with respect to κ and $(A_\lambda)_{\lambda \geq 1}$) if*

$$\sup_{\lambda \geq 1,\ x \in A_\lambda} \mathbb{E}\left[|\xi_\lambda((x, U); \mathcal{P}_\lambda)|^p\right] < \infty. \tag{2.5}$$

For $\lambda > 0$, define the random weighted point measure μ_λ^ξ on $\mathbb{R}^d$ by

$$\mu_\lambda^\xi := \sum_{(x,s) \in \mathcal{P}_\lambda \cap (A_\lambda \times \mathcal{M})} \xi_\lambda((x, s); \mathcal{P}_\lambda) \delta_x$$

and the centered version $\overline{\mu}_\lambda^\xi := \mu_\lambda^\xi - \mathbb{E}[\mu_\lambda^\xi]$.

Let $B(A)$ denote the set of bounded Borel-measurable functions on A. Given $f \in B(A)$, let $\langle f, \mu_\lambda^\xi \rangle := \int_A f d\mu_\lambda^\xi$ and $\langle f, \overline{\mu}_\lambda^\xi \rangle := \int_A f d\overline{\mu}_\lambda^\xi$.

Let Φ denote the distribution function of the standard normal. Our main result is a normal approximation result for $\langle f, \overline{\mu}_\lambda^\xi \rangle$, suitably scaled.

Theorem 2.1 *Suppose $\|\kappa\|_\infty < \infty$. Suppose that ξ is exponentially stabilizing and satisfies the moments condition (2.5) for some $p > 2$. Let $q \in (2, 3]$ with $q < p$. Let $f \in B(A)$ and put $T_\lambda := \langle f, \mu_\lambda^\xi \rangle$. Then there exists a finite constant C depending on d, ξ, κ, p, q and f, such that for all $\lambda \geq 2$,*

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{T_\lambda - \mathbb{E}\, T_\lambda}{(\mathrm{Var} T_\lambda)^{1/2}} \leq t\right] - \Phi(t) \right| \leq C (\log \lambda)^{dq} \lambda (\mathrm{Var} T_\lambda)^{-q/2}. \tag{2.6}$$

Separate arguments are required to establish the asymptotic behavior of the denominator $(\mathrm{Var}(T_\lambda))^{1/2}$ in (2.6). When $A_\lambda = A$ for all λ, it is typically the case for polynomially stabilizing functionals satisfying moments conditions along the lines of (2.5) that there is a constant $\sigma^2(f, \xi, \kappa) \geq 0$ such that

$$\lim_{\lambda \to \infty} \lambda^{-1} \mathrm{Var} \langle f, \mu_\lambda^\xi \rangle = \sigma^2(f, \xi, \kappa). \tag{2.7}$$

For further information about $\sigma^2(f, \xi, \kappa)$ and precise conditions under which (2.7) holds, see Theorem 2.4(i) of [4]. When (2.7) holds, by combining it with Theorem 2.1 we obtain

$$\langle f, \lambda^{-1/2} \overline{\mu}_\lambda^\xi \rangle \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(f, \xi, \kappa)\right), \tag{2.8}$$

where $\mathcal{N}(0, \sigma^2)$ denotes a centered normal distribution with variance σ^2 if $\sigma^2 > 0$, and a unit point mass at 0 if $\sigma^2 = 0$.

In many applications (2.7) holds with $\sigma^2(f, \xi, \kappa) > 0$, showing that the case $q = 3$ of (2.6) yields a rate of convergence $O((\log \lambda)^{3d} \lambda^{-1/2})$ to the normal distribution. In other words, we will make frequent use of:

Corollary 2.1 *Suppose $\|\kappa\|_\infty < \infty$. Suppose that ξ is exponentially stabilizing and satisfies the moments condition (2.5) for some $p > 3$. Let $f \in B(A)$ and put $T_\lambda := \langle f, \mu_\lambda^\xi \rangle$. If (2.7) holds with $\sigma^2(f, \xi, \kappa) > 0$, then there exists a finite constant C depending on d, ξ, κ, p and f, such that for all $\lambda \geq 2$,*

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{T_\lambda - \mathbb{E}\, T_\lambda}{(\mathrm{Var} T_\lambda)^{1/2}} \leq t \right] - \Phi(t) \right| \leq C (\log \lambda)^{3d} \lambda^{-1/2}.$$

Our methods actually yield normal approximation and a central limit theorem when the exponential decay condition is replaced by a polynomial decay condition of sufficiently high order. We give a further result along these lines.

Theorem 2.2 *Suppose $\|\kappa\|_\infty < \infty$. Suppose for some $p > 3$ that ξ is polynomially stabilizing of order γ with $\gamma > d(150 + 6/p)$, and satisfies the moments condition (2.5). Let $f \in B(A)$ and put $T_\lambda := \langle f, \mu_\lambda^\xi \rangle$. Suppose that (2.7) holds for some $\sigma^2 \geq 0$. Then (2.8) holds and if $\sigma^2 := \sigma^2(f, \xi, \kappa) > 0$ there exists a finite constant C depending on d, ξ, κ, p and f, such that for all $\lambda \geq 2$,*

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{T_\lambda - \mathbb{E}\, T_\lambda}{(\mathrm{Var} T_\lambda)^{1/2}} \leq t \right] - \Phi(t) \right| \leq C \lambda^{(150pd + 6d - p\gamma)/2(p\gamma - 6d)}. \tag{2.9}$$

Remarks

1. Our results are stated for *marked* Poisson point processes, i.e., for Poisson processes in $\mathbb{R}^d \times \mathcal{M}$ where $\mathcal{M}$ is the mark space. These results are reduced to the corresponding results for unmarked Poisson point processes in $\mathbb{R}^d$ by taking $\mathcal{M}$ to have a single element (denoted m, say) and identifying $\mathbb{R}^d \times \mathcal{M}$ with $\mathbb{R}^d$ in the obvious way by identifying (x, m) with x for each $x \in \mathbb{R}^d$. In this case the notation (2.4) is particularly appropriate. Other treatments such as [4, 10, 11, 12] tend to concentrate on the unmarked case with commentary that the proofs carry through to the marked case; here we spell out the results and proofs in the more general marked case, which seems worthwhile since examples such as those in Section 3.3 use the results for marked point processes. Our examples in Sections 3.1, 3.2, and 3.4 refer to unmarked point processes and in these examples we identify $\mathbb{R}^d \times \{m\}$ with $\mathbb{R}^d$ as indicated above (so that $\mathcal{P}_\lambda$ is viewed as a Poisson process in $\mathbb{R}^d$).

2. We are not sure if the logarithmic factors can be removed in Theorem 2.1 or Corollary 2.1. Avram and Bertsimas [1] obtain a rate of $O((\log \lambda)^{1+3/(2d)} \lambda^{-1/4})$, for the length of the k-nearest neighbors (directed) graph, the Voronoi graph, and the Delaunay graph (see Sections 3.1 and 3.2). Our method for general stabilizing functionals is based on theirs, but uses a stronger general normal approximation result (Lemma 4.1 below).

3. If (2.7) holds with $\sigma^2(f, \xi, \kappa) = 0$, then (2.8) holds trivially by Chebyshev's inequality, but Theorem 2.1 does not provide any useful information on rate of convergence. In examples of interest, it can usually be established that $\sigma^2(f, \xi, \kappa) > 0$, by further separate arguments. We do not discuss these in detail here but refer the reader to [1, 11, 4].

4. Theorems 2.1, 2.2, and Corollary 2.1 require neither the underlying density function κ nor the test function f to be continuous (both of these conditions are imposed in [4]). In particular, these three results apply when f is the indicator function of a Borel subset B of A, giving normal approximation for $\bar{\mu}_\lambda^\xi(B)$.

5. We do not have rate of convergence results in the binomial (non-Poisson) setting. For central limit theorems in the binomial setting, we refer to [11] and [4], which treat uniform and non-uniform samples respectively.

6. Some functionals, such as those defined in terms of the minimal spanning tree, stabilize without any known bounds on the rate of decay of the tail probability $\tau(t)$. In these cases univariate and multivariate central limit theorems hold [10, 11] but our Theorems 2.1 and 2.2 do not apply and explicit rates of convergence are not known.

3 Applications

Applications of Corollary 2.1 to geometric probability include functionals of proximity graphs, germ-grain models, and random sequential packing models. The following examples are for illustrative purposes only and are not meant to be encyclopedic. For simplicity we will assume that $\mathbb{R}^d$ is equipped with the usual Euclidean metric. While translation invariance is not needed in the general results in Section 2, most of the examples treated in this section involve translation invariant functionals ξ. However, the examples can be modified to treat the (non-translation-invariant) situation where $\mathbb{R}^d$ has a local metric structure.

3.1 k-nearest neighbors graph

Let k be a positive integer. Given a locally finite point set $\mathcal{X} \subset \mathbb{R}^d$, the k-nearest neighbors (undirected) graph on $\mathcal{X}$, denoted $kNG(\mathcal{X})$, is the graph with vertex set $\mathcal{X}$ obtained by including $\{x, y\}$ as an edge whenever y is one of the k nearest neighbors of x and/or x is one of the k nearest neighbors of y. The k-nearest neighbors (directed) graph on $\mathcal{X}$, denoted $kNG'(\mathcal{X})$, is the graph with vertex set $\mathcal{X}$ obtained by placing a directed edge between each point and its k nearest neighbors.

Let $N^k(\mathcal{X})$ denote the total edge length of the (undirected) k-nearest neighbors graph on $\mathcal{X}$. Note that $N^k(\mathcal{X}) = \sum_{x \in \mathcal{X}} \xi^k(x; \mathcal{X})$, where $\xi^k(x; \mathcal{X})$ denotes half the sum of the edge lengths in $kNG(\mathcal{X})$ incident to x. If A is convex or polyhedral and κ is bounded away from 0 on A, then ξ^k is exponentially stabilizing (cf. Lemma 6.1 of [11]) and has moments of all orders. Moreover, as shown in [4] (see e.g. display (2.11), Theorem 3.1), at least when f is continuous and $A_\lambda = A$ for all λ,

$$\lim_{\lambda\to\infty} \lambda^{-1}\mathrm{Var}\langle f, \mu_\lambda^\xi \rangle = V^\xi \int_A f(x)^2 \kappa(x)^{(d-2)/d} dx,$$

where V^ξ denotes the limiting variance for the total edge length of the k-nearest neighbors graph on $\lambda^{1/d}\mathcal{P}_\lambda$ when κ is the uniform distribution on the unit cube. Since V^ξ is strictly positive (Theorem 6.1 of [11]), it follows that (2.7) holds with $\sigma^2(f, \xi^k, \kappa) > 0$. We thus obtain via Corollary 2.1 the following rates in the CLT for the total edge length of $N^k(\lambda^{1/d}\mathcal{P}_\lambda)$ improving upon Avram and Bertsimas [1] and Bickel and Breiman [6]. A similar CLT holds for the total edge length of the k-nearest neighbors directed graph.

Theorem 3.1 *Suppose A is convex or polyhedral and κ is bounded away from 0 on A. Let $N_\lambda := N^k(\lambda^{1/d}\mathcal{P}_\lambda)$ denote the total edge length of the k-nearest neighbors graph on $\lambda^{1/d}\mathcal{P}_\lambda$. There exists a finite constant C depending on d, ξ^k, and κ such that*

$$\sup_{t\in\mathbb{R}} \left| P\left[\frac{N_\lambda - \mathbb{E}\, N_\lambda}{(\mathrm{Var} N_\lambda)^{1/2}} \le t \right] - \Phi(t) \right| \le C(\log\lambda)^{3d}\lambda^{-1/2}. \qquad (3.1)$$

Similarly, letting $\xi^s(x;\mathcal{X})$ be one or zero according to whether the distance between x and its nearest neighbor in $\mathcal{X}$ is less than s or not, we can verify that ξ^s is exponentially stabilizing and that the variance of $\sum_{x\in\lambda^{1/d}\mathcal{P}_\lambda} \xi^s(x;\lambda^{1/d}\mathcal{P}_\lambda)$ is bounded below by a positive multiple of λ. We thus obtain rates of convergence of $O((\log\lambda)^{3d}\lambda^{-1/2})$ in the CLT for the one-dimensional marginals of the empirical distribution function of k nearest neighbors distances on $\lambda^{1/d}\mathcal{P}_\lambda$, improving upon those implicit on p. 88 of [9].

Using the results from section 6.2 of [11], we could likewise obtain the same rates of convergence in the CLT for the number of vertices of fixed degree in the k nearest neighbors graph.

Finally in this section, we re-consider the non-translation-invariant example given in Section 2, where a point at x is colored red with probability $q(x)$ and green with probability $1 - q(x)$, and $\xi(x;\mathcal{X})$ takes the value 0 if the nearest neighbor of x in $\mathcal{X}$ has the same color as x, and takes the value 1 otherwise. We can use Corollary 2.1 to derive a central limit theorem, with $O((\log\lambda)^{3d}\lambda^{-1/2})$ rate of convergence, for $\sum_{x\in\mathcal{P}_\lambda} f(x)\xi(x;\mathcal{P}_\lambda)$, where f is a bounded measurable test function.

3.2 Voronoi and sphere of influence graphs

We will consider the Voronoi graph for $d = 2$ and the sphere of influence graph for all d. Given a locally finite set $\mathcal{X} \subset \mathbb{R}^2$ and given $x \in \mathcal{X}$, the locus of points closer to x than to any other point in $\mathcal{X}$ is called the *Voronoi* cell centered at x. The graph consisting of all boundaries of Voronoi cells is called the Voronoi graph generated by $\mathcal{X}$.

The sum of the lengths of the finite edges of the Voronoi graph on $\mathcal{X}$ admits the representation $\sum_{x\in\mathcal{X}} \xi(x;\mathcal{X})$, where $\xi(x;\mathcal{X})$ denotes one half the sum of the lengths of the finite edges in the Voronoi cell at x. If κ is bounded away from 0 and infinity and A is convex, then geometric arguments show that there is a random variable R with exponentially decaying tails such that for any $x \in \mathcal{P}_\lambda$, the value of $\xi(x;\mathcal{P}_\lambda)$ is unaffected by points outside $B_{\lambda^{-1/d}R}(x)$ [4, 11, 13]. In other words, ξ is exponentially stabilizing and satisfies the moments condition (2.5) for all $p > 1$. Also, the variance of the total edge length of these graphs on $\lambda^{1/d}\mathcal{P}_\lambda$ is bounded below by a multiple of λ. We thus obtain $O((\log\lambda)^{3d}\lambda^{-1/2})$ rates of convergence in the CLT for the total edge length functionals of these graphs on $\lambda^{1/d}\mathcal{P}_\lambda$, thereby improving and generalizing the results of Avram and Bertsimas [1].

Given a locally finite set $\mathcal{X} \subset \mathbb{R}^d$, the sphere of influence graph SIG($\mathcal{X}$) is a graph with vertex set $\mathcal{X}$, constructed as follows: for each $x \in \mathcal{X}$ let B_x be a ball around x with radius equal to $\min_{y\in\mathcal{X}\setminus\{x\}}\{|y-x|\}$. Then B_x is called the *sphere of influence* of x. We put an edge between x and y iff the balls B_x and B_y overlap. The collection of such edges is the *sphere of influence graph* (SIG) on $\mathcal{X}$.

The total number of edges of the sphere of influence graph on $\mathcal{X}$ admits the representation $\sum_{x\in\mathcal{X}} \xi(x;\mathcal{X})$, where $\xi(x;\mathcal{X})$ denotes one half the degree of SIG at the vertex x. The number of vertices of fixed degree δ admits a similar representation, with $\xi(x;\mathcal{X})$ now equal to one (respectively, zero) if the degree at x is δ (respectively, if degree at x is not δ). If κ is bounded away from 0 and infinity and A is convex, then geometric arguments show that both choices of the functional ξ stabilize (see sections 7.1 and 7.3 of [11]). Also, the variance of both the total number edges and the number of vertices of fixed degree in the SIG on $\lambda^{1/d}\mathcal{P}_\lambda$ is bounded below by a multiple of λ (sections 7.1 and 7.3 of [11]). We thus obtain $O((\log\lambda)^{3d}\lambda^{-1/2})$ rates of convergence in the CLT for the total number of edges and the number of vertices of fixed degree in the sphere of influence graph on $\mathcal{P}_\lambda$.

3.3 Random sequential packing models

The following prototypical random sequential packing model is of considerable scientific interest; see [12] for references to the vast literature.

With $N(\lambda)$ standing for a Poisson random variable with parameter λ, we let $B_{\lambda,1}, B_{\lambda,2}, ..., B_{\lambda,N(\lambda)}$ be a sequence of d-dimensional balls of volume λ^{-1} whose centers are i.i.d. random d-vectors $X_1, ..., X_{N(\lambda)}$ with probability density function $\kappa : A \to [0,\infty)$. Without loss of generality, assume that the balls are sequenced in the order determined by marks (time coordinates) in $[0,1]$. Let the first ball $B_{\lambda,1}$ be *packed*, and recursively for $i = 2, 3, \ldots,$ let the i-th ball $B_{\lambda,i}$ be packed iff $B_{\lambda,i}$ does not overlap any ball in $B_{\lambda,1}, ..., B_{\lambda,i-1}$ which has already been packed. If not packed, the i-th ball is discarded.

Packing models of this type arise in diverse disciplines, including physical, chemical, and biological processes [12]. Central limit theorems for the number of accepted (i.e., packed) balls are established in [12, 4], whereas laws of large numbers are given in [13].

Let $\mathcal{M} = [0, 1]$ with $\mathbb{P}_{\mathcal{M}}$ being the uniform distribution on the unit interval. For any finite point set $\mathcal{X} \subset \mathbb{R}^d \times [0, 1]$, assume the points $(x, s) \in \mathcal{X}$ represent the locations and arrival times. Assume balls of volume λ^{-1} centered at the locations of $\mathcal{X}$ arrive sequentially in an order determined by the time coordinates, and assume as before that each ball is packed or discarded according to whether or not it overlaps a previously packed ball. Let $\xi((x, s); \mathcal{X})$ be either 1 or 0 depending on whether the ball centered at x at times s is packed or discarded. Consider the re-scaled packing functional $\xi_\lambda((x, s); \mathcal{X}) = \xi((\lambda^{1/d}x, s); \lambda^{1/d}\mathcal{X})$, where balls centered at points of $\lambda^{1/d}\mathcal{X}$ have volume one. The random measure

$$\mu_\lambda^\xi := \sum_{i=1}^{N(\lambda)} \xi_\lambda((X_i, U_i); \{(X_j, U_j)\}_{j=1}^{N(\lambda)})\ \delta_{X_i},$$

is called the random sequential packing measure induced by balls with centers arising from κ. The convergence of the finite dimensional distributions of the packing measures μ_λ^ξ is established in [3, 4]. ξ is exponentially stabilizing [12, 3] and for any continuous $f \in B([0, 1]^d)$ and κ uniform, the variance of $\langle f, \mu_\lambda^\xi\rangle$ is bounded below by a positive multiple of λ [4], showing that $\langle f, \mu_\lambda^\xi\rangle$ satisfies a CLT with an $O((\log \lambda)^{3d}\lambda^{-1/2})$ rate of convergence.

It follows easily from the stabilization analysis of [12] that many variants of the above basic packing model satisfy similar rates of convergence in the CLT. Examples include balls of bounded random radius, cooperative sequential adsorption ([12]), and monolayer ballistic deposition ([12]). In each case the number of particles accepted satisfies the CLT with an $O((\log \lambda)^{3d}\lambda^{-1/2})$ rate of convergence. The same comment applies for the number of seeds accepted in spatial birth-growth models [12].

3.4 Independence number, off-line packing

An *independent set* of vertices in a graph G is a set of vertices in G, no two of which are connected by an edge. The *independence number* of G, which we denote $\beta(G)$, is defined to be the maximum cardinality of all independent sets of vertices in G.

For $r > 0$, and for finite or countable $\mathcal{X} \subset \mathbb{R}^d$, let $G(\mathcal{X}, r)$ denote the *geometric graph* with vertex set $\mathcal{X}$ and with edges between each pair of vertices distant at most r apart. Then the independence number $\beta(G(\mathcal{X}, r))$ is the maximum number of disjoint closed balls of radius $r/2$ that can be centered at points of $\mathcal{X}$; it is an 'off-line' version of the packing functionals considered in the previous section.

Let $b > 0$ be a constant, and consider the graph $G(\mathcal{P}_\lambda, b\lambda^{-1/d})$ (or equivalently, $G(\lambda^{1/d}\mathcal{P}_\lambda, b)$). Random geometric graphs of this type are the subject of [9], although independence number is considered only briefly there (on page 135). A law of large numbers for the independence number is described in Theorem 2.7 (iv) of [13].

For $u > 0$, let $\mathcal{H}_u$ denote a homogeneous Poisson point process of intensity u on $\mathbb{R}^d$, and let $\mathcal{H}_u^0$ be the point process $\mathcal{H}_u$ with a point inserted at the origin. As on page 189 of [9], let λ_c be the infimum of all u such that the origin has a non-zero probability of being in an infinite component of $G(\mathcal{H}_u, 1)$.

If $b^d\|\kappa\|_\infty < \lambda_\mathrm{c}$, we can use Corollary 2.1 to obtain a central limit theorem for the independence number $\beta(G(\mathcal{P}_\lambda, b\lambda^{-1/d}))$, namely

$$\sup_{t\in\mathbb{R}} \left| P\left[\frac{\beta(G(\mathcal{P}_\lambda, b\lambda^{-1/d})) - \mathbb{E}\,\beta(G(\mathcal{P}_\lambda, b\lambda^{-1/d}))}{(\mathrm{Var}\beta(G(\mathcal{P}_\lambda, b\lambda^{-1/d})))^{1/2}} \le t \right] - \Phi(t) \right| \le C(\log\lambda)^{3d}\lambda^{-1/2} \quad (3.2)$$

We sketch the proof. For finite $\mathcal{X} \subset \mathbb{R}^d$ and $x \in \mathcal{X}$, let $\xi(x;\mathcal{X})$ denote the independence number of the component of $G(\mathcal{X}, b)$ containing vertex x, divided by the number of vertices in this component. Then $\sum_{x\in\mathcal{X}} \xi(x;\mathcal{X})$ is the independence number of $G(\mathcal{X}, b)$, since the independence number of any graph is the sum of the independence numbers of its components. Also, our choice of ξ is translation-invariant, and so we obtain

$$\begin{aligned}\beta(G(\mathcal{P}_\lambda, \lambda^{-1/d}b)) = \beta(G(\lambda^{1/d}\mathcal{P}_\lambda, b)) &= \sum_{x\in\mathcal{P}_\lambda} \xi(\lambda^{1/d}x; \lambda^{1/d}\mathcal{P}_\lambda) \\ &= \sum_{x\in\mathcal{P}_\lambda} \xi_\lambda(x;\mathcal{P}_\lambda) = \langle \mu_\lambda^\xi, f\rangle,\end{aligned}$$

where we here take the test function f to be identically 1 and take $A_\lambda = A$ for all λ. Thus a central limit theorem holds for $\beta(G(\mathcal{P}_\lambda, \lambda^{-1/d}b))$ by application of Corollary 2.1, if ξ and κ satisfy the conditions for that result.

We take $R(x,\lambda)$ to be the distance from $\lambda^{1/d}x$ to the furthest point in the component containing $\lambda^{1/d}x$ of $G(\lambda^{1/d}\mathcal{P}_\lambda, b)$, plus $2b$. Since $\xi_\lambda(x;\mathcal{P}_\lambda)$ is determined by the component of $G(\lambda^{1/d}\mathcal{P}_\lambda, b)$ containing $\lambda^{1/d}x$, and this component is unaffected by the addition or removal of points to/from $\mathcal{P}_\lambda$ at a distance greater than $\lambda^{-1/d}R(x,\lambda)$ from x, it is indeed the case that $R(x,\lambda)$ is a radius of stabilization.

The point process $\lambda^{1/d}\mathcal{P}_\lambda$ is dominated by $\mathcal{H}_{\|\kappa\|_\infty}$ (in the sense of [9], page 189). Hence, $P[R(x,\lambda) > t]$ is bounded by the probability that the component containing x of $G(\mathcal{H}_{\|\kappa\|_\infty} \cup \{\lambda^{1/d}x\}, b)$ has at least one vertex outside $B_{t-2b}(\lambda^{1/d}x)$. This probability does not depend on x, and equals the probability that the component of $G(\mathcal{H}^0_{b^d\|\kappa\|_\infty}, 1)$ containing the origin includes a vertex outside $B_{(t/b)-2}$. By exponential decay for subcritical continuum percolation (Lemma 10.2 of [9]), this probability decays exponentially in t, and exponential stabilization of ξ follows. The moments condition (2.5) is trivial in this case, for any p, since $0 \le \xi(x;\mathcal{X}) \le 1$.

Thus, Corollary 2.1 is indeed applicable, provided that (2.7) holds in this case, with $\sigma^2 > 0$. Essentially (2.7) follows from Theorem 2.1 of [4], with strict inequality $\sigma^2 > 0$ following from (2.10) of [4]; in the case where κ is the density function of a uniform distribution on some suitable subset of $\mathbb{R}^d$, one can alternatively use Theorem 2.4 of [11]. We do not go into details here about the application of results in this example, but we do comment further on why the distribution of the 'add one cost' (see [11, 4]) of insertion of a point at the origin into a homogeneous Poisson process is nondegenerate, since this is needed to verify $\sigma^2 > 0$ and this example was not considered in [11] or [4].

The above add one cost is the variable denoted $\Delta(\infty)$ in the notation of [11], or $\Delta^\xi(u)$ in the notation of [4]. It is the independence number of the component containing the origin of $G(\mathcal{H}_u; b)$ minus the independence number of this component with the origin removed (we need only to consider the case where $b^d u$ is subcritical). This variable can take the value 1, for example if the origin is isolated in $G(\mathcal{H}_u; b)$, or zero, for example if the component containing the origin has two vertices. Both of these possibilities have strictly positive probability, and therefore $\Delta(\infty)$ has a non-degenerate distribution.

4 Proof of Theorems

4.1 A CLT for dependency graphs

We shall prove Theorem 2.1 by showing that exponential stabilization implies that a modification of $\langle f, \overline{\mu}_\lambda^\xi \rangle$ has a *dependency graph* structure, whose definition we now recall (see e.g. Chapter 2 of [9]). Let $X_\alpha,\ \alpha \in \mathcal{V}$, be a collection of random variables. The graph $G := (\mathcal{V}, \mathcal{E})$ is a *dependency graph* for $X_\alpha,\ \alpha \in \mathcal{V}$, if for any pair of disjoint sets $A_1, A_2 \subset \mathcal{V}$ such that no edge in $\mathcal{E}$ has one endpoint in A_1 and the other in A_2, the sigma-fields $\sigma\{X_\alpha, \alpha \in A_1\}$, and $\sigma\{X_\alpha, \alpha \in A_2\}$, are mutually independent. Let D denote the maximal degree of the dependency graph.

It is well known that sums of random variables indexed by the vertices of a dependency graph admit rates of convergence to a normal. The rates of Baldi and Rinott [2] and those in Penrose [9] are particularly useful; Avram and Bertsimas [1] use the former to obtain rate results for the total edge length of the nearest neighbor, Voronoi, and Delaunay graphs.

In many cases, the following theorem of Chen and Shao [7] provides superior rate results. For any random variable X and any $p > 0$, let $||X||_p = (\mathbb{E}[|X|^p])^{1/p}$.

Lemma 4.1 *(see Thm 2.7 of [7]) Let $2 < q \leq 3$. Let $W_i,\ i \in \mathcal{V}$, be random variables indexed by the vertices of a dependency graph. Let $W = \sum_{i \in \mathcal{V}} W_i$. Assume that $\mathbb{E}[W^2] = 1, \mathbb{E}[W_i] = 0$, and $||W_i||_q \leq \theta$ for all $i \in \mathcal{V}$ and for some $\theta > 0$. Then*

$$\sup_t |P[W \leq t] - \Phi(t)| \leq 75 D^{5(q-1)} |\mathcal{V}| \theta^q. \tag{4.1}$$

4.2 Auxiliary lemmas

To prepare for the proof of Theorem 2.1 we will need some auxiliary lemmas. Throughout, C denotes a generic constant depending possibly on d, ξ, and κ and whose value may vary at each occurrence. We assume $\lambda > 1$ throughout.

Let $(\rho_\lambda, \lambda > 0)$ be a function to be chosen later, in such a way that $\rho_\lambda \to \infty$ and $\lambda^{-1/d}\rho_\lambda \to 0$ as $\lambda \to \infty$. Given $\lambda > 0$, let $s_\lambda := \lambda^{-1/d}\rho_\lambda$, and let $V := V(\lambda)$ denote the number of cubes of the form $Q = \prod_{i=1}^d [j_i s_\lambda, (j_i+1)s_\lambda)$, with all $j_i \in \mathbb{Z}$, such that $\int_Q \kappa(x)dx > 0$; enumerate these cubes as $Q_1, Q_2, \ldots, Q_{V(\lambda)}$. Since κ is assumed to have bounded support, it is easy to see that $V(\lambda) = O(\lambda\rho_\lambda^{-d})$ as $\lambda \to \infty$.

For all $1 \le i \le V(\lambda)$, the number of points of $\mathcal{P}_\lambda \cap (Q_i \times \mathcal{M})$ is a Poisson random variable $N_i := N(\nu_i)$, where

$$\nu_i := \lambda \int_{Q_i} \kappa(x)dx \le \|\kappa\|_\infty \rho_\lambda^d. \tag{4.2}$$

Assuming $\nu_i > 0$, choose an ordering on the points of $\mathcal{P}_\lambda \cap (Q_i \times \mathcal{M})$ uniformly at random from all $(N_i)!$ possible such orderings. Use this ordering to list the points as $(X_{i,1}, U_{i,1}), ..., (X_{i,N_i}, U_{i,N_i})$, where conditional on the value of N_i, the random variables $X_{i,j},\ j = 1, 2, ...$ are i.i.d. on Q_i with a density $\kappa_i(\cdot) := \kappa(\cdot)/\int_{Q_i}\kappa(x)dx$, and the $U_{i,j}$ are i.i.d. in $\mathcal{M}$ with distribution $\mathbb{P}_\mathcal{M}$, independent of $\{X_{i,j}\}$. Thus we have the representation $\mathcal{P}_\lambda = \cup_{i=1}^{V(\lambda)}\{(X_{i,j}, U_{i,j})\}_{j=1}^{N_i}$. For all $1 \le i \le V(\lambda)$, let $\mathcal{P}_i := \mathcal{P}_\lambda \setminus \{(X_{i,j}, U_{i,j})\}_{j=1}^{N_i}$ and note that $\mathcal{P}_i$ is a Poisson point process on $\mathbb{R}^d \times \mathcal{M}$ with intensity $\lambda\kappa(x)\mathbf{1}_{\mathbb{R}^d\setminus Q_i}(x)dx \times \mathbb{P}_\mathcal{M}(ds)$.

We show that the condition (2.5), which bounds the moments of the value of ξ at points inserted into $\mathcal{P}_\lambda$,

also yields bounds on $\mathbb{E}\,[|\xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda)|^p \mathbf{1}_{A_\lambda}(X_{i,j})\mathbf{1}_{j\le N_i}]$. More precisely, we have:

Lemma 4.2 *If (2.5) holds for some $p > 0$, then there is a constant $C := C(p)$ such that for all $\lambda > 1$, all $j \ge 1$ and $1 \le i \le V(\lambda)$,*

$$\mathbb{E}\,[|\xi_\lambda(X_{i,j}; \mathcal{P}_\lambda) \cdot \mathbf{1}_{A_\lambda}(X_{i,j})\mathbf{1}_{j\le N_i}|^p] \le C\rho_\lambda^d. \tag{4.3}$$

Proof. If $N_i = n$, then denote $\{(X_{i,1}, U_{i,1}), ..., (X_{i,N_i}, U_{i,N_i})\}$ by $\mathcal{X}_n$. By definition,

$$\mathbb{E}\,[|\xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda)\cdot\mathbf{1}_{A_\lambda}(X_{i,j})\mathbf{1}_{j\le N_i}|^p] = \sum_{n=j}^\infty \int_{Q_i\cap A_\lambda} \mathbb{E}\,[|\xi_\lambda((x, U); \mathcal{X}_{n-1}\cup\mathcal{P}_i)|^p]\kappa_i(x)dx\cdot P[N_i = n],$$

where the expectation on the right hand side is with respect to U, $\mathcal{X}_{n-1}$ and $\mathcal{P}_i$. The above is bounded by

$$\le \nu_i \sum_{n=1}^\infty \int_{Q_i\cap A_\lambda} \mathbb{E}\,[|\xi_\lambda((x, U); \mathcal{X}_{n-1}\cup\mathcal{P}_i)|^p]\kappa_i(x)dx \cdot \frac{e^{-\nu_i}\nu_i^{n-1}}{(n-1)!}$$

$$= \nu_i \sum_{m=0}^{\infty} \int_{Q_i \cap A_\lambda} \mathbb{E}\left[|\xi_\lambda((x,U);\mathcal{P}_\lambda)|^p \;\middle|\; |\mathcal{P}_\lambda \cap (Q_i \times \mathcal{M})| = m \right] \kappa_i(x) dx \cdot P[|\mathcal{P}_\lambda \cap (Q_i \times \mathcal{M})| = m]$$

$$= \nu_i \int_{Q_i \cap A_\lambda} \mathbb{E}\left[|\xi_\lambda((x,U);\mathcal{P}_\lambda)|^p \right] \kappa_i(x) dx \leq \text{const.} \times \nu_i,$$

where the last inequality follows by (2.5). By (4.2), this shows (4.3). □

For $1 \leq i \leq V$, and for $j \in \{1, 2, \ldots\}$, we define

$$\xi_{i,j} := \begin{cases} \xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda) & \text{if } N_i \geq j, X_{i,j} \in A_\lambda \\ 0 & \text{otherwise} \end{cases}$$

Lemma 4.3 *If (2.5) holds for some $p > q > 1$, then there exists $C := C(p,q)$ such that for $1 \leq i \leq V(\lambda)$,*

$$\left\| \sum_{j=1}^{\infty} |\xi_{i,j}| \right\|_q \leq C \rho_\lambda^{d(p+1)/p}. \tag{4.4}$$

Proof. Fix $i \leq V(\lambda)$ and write ξ_j for $\xi_{i,j}$. Clearly, with $N := N_i$ and $\nu := \nu_i$,

$$\left\| \sum_{j=1}^{\infty} |\xi_j| \right\|_q = \left\| \sum_{j=1}^{\infty} |\xi_j| \left(\mathbf{1}_{N \leq \nu} + \sum_{k=0}^{\infty} \mathbf{1}_{2^k \nu < N \leq 2^{k+1} \nu} \right) \right\|_q$$

$$\leq \left\| \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} |\xi_j| \cdot \mathbf{1}_{2^k \nu < N \leq 2^{k+1} \nu} \right\|_q + \left\| \sum_{j=1}^{\infty} |\xi_j| \cdot \mathbf{1}_{N \leq \nu} \right\|_q .$$

Since a.s. only finitely many summands in the double sum are non-zero, by subadditivity of the norm, the above is bounded by

$$\leq \sum_{k=0}^{\infty} \left\| \sum_{j=1}^{\infty} |\xi_j| \cdot \mathbf{1}_{2^k \nu < N \leq 2^{k+1} \nu} \right\|_q + \left\| \sum_{j=1}^{\lfloor \nu \rfloor} |\xi_j| \cdot \mathbf{1}_{N \leq \nu} \right\|_q$$

$$\leq \sum_{k=0}^{\infty} \left\| \sum_{j=1}^{\lfloor 2^{k+1} \nu \rfloor} |\xi_j| \cdot \mathbf{1}_{N \geq 2^k \nu} \right\|_q + \left\| \sum_{j=1}^{\lfloor \nu \rfloor} |\xi_j| \cdot \mathbf{1}_{N \leq \nu} \right\|_q$$

$$\leq \sum_{k=0}^{\infty} \sum_{j=1}^{\lfloor 2^{k+1} \nu \rfloor} \left\| \xi_j \cdot \mathbf{1}_{N \geq 2^k \nu} \right\|_q + \sum_{j=1}^{\lfloor \nu \rfloor} \left\| \xi_j \cdot \mathbf{1}_{N \leq \nu} \right\|_q , \tag{4.5}$$

where here and elsewhere $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

With $\eta := (1/q) - (1/p)$, Hölder's inequality followed by (4.3) yields

$$||\xi_j \cdot \mathbf{1}_{N \geq 2^k \nu}||_q \leq ||\xi_j||_p \cdot (P[N \geq 2^k \nu])^\eta \leq C \rho_\lambda^{d/p} (P[N \geq 2^k \nu])^\eta. \tag{4.6}$$

Substituting into (4.5) we obtain

$$\left|\left| \sum_{j=1}^{\infty} |\xi_j| \right|\right|_q \leq C \rho_\lambda^{d/p} \sum_{k=0}^{\infty} \nu 2^{k+1} \cdot (P[N \geq 2^k \nu])^\eta + \sum_{j=1}^{\lfloor \nu \rfloor} ||\xi_j \cdot \mathbf{1}_{N \leq \nu}||_p. \tag{4.7}$$

By tail estimates for the Poisson distribution (see e.g. (1.12) in [9]), since $e^2 < 8$ we have

$$(P[N \geq 2^k \nu])^\eta \leq \left(\exp\left(-2^{k-1} \nu \log(2^k)\right)\right)^\eta = \exp(-k 2^{k-1} \eta \nu \log 2), \quad k \geq 3.$$

Hence,

$$(P[N \geq 2^k \nu])^\eta \leq 2^{-2k}, \quad k \geq \max(3, 2 - \log_2(\eta \nu)).$$

Hence, since η is a constant, for $\nu > 0$ we have

$$\begin{aligned} \sum_{k=0}^{\infty} 2^{k+1} \cdot (P[N \geq 2^k \nu])^\eta &\leq \sum_{k < \max(3, 2 - \log_2(\eta\nu))} 2^{k+1} + \sum_{k \geq \max(3, 2 - \log_2(\eta\nu))} 2^{1-k} \\ &\leq C \max(1/\nu, 1) + 2 \end{aligned}$$

where C does not depend on ν. Thus by (4.2), the first term in the right hand side of (4.7) is $O(\rho_\lambda^{d+d/p})$. Also, the second term in the right hand side of (4.7) is $O(\rho_\lambda^{d+d/p})$ by Lemma 4.2. Hence, (4.7) implies (4.4). □

4.3 Proof of Theorems 2.1 and 2.2

We prove Theorems 2.1 and 2.2 together. When proving Theorem 2.1 we assume that ξ is exponentially stabilizing and (2.5) holds for some $p > 2$, and we choose $q \in (2, 3]$ with $q < p$. When proving Theorem 2.2 we assume that ξ is polynomially stabilizing of order γ with $\gamma > d(150 + 6/p)$, and that (2.5) holds for some $p > 3$, and we set $q = 3$.

Throughout this section, we fix $f \in B(A)$ and set $T_\lambda := \langle f, \mu_\lambda^\xi \rangle$. We follow the setup of the preceding section, with the support of κ covered by cubes of side $\lambda^{-1/d} \rho_\lambda$, and we now choose ρ_λ. With the tail probability $\tau(t)$ defined at (2.2), we choose ρ_λ in such a way that there is a constant C such that for all $\lambda \geq 1$,

$$\rho_\lambda^{d/p} (\lambda \tau(\rho_\lambda))^{(q-2)/(2q)} < C \lambda^{-4} \quad \text{and} \quad \tau(\rho_\lambda) < C \lambda^{-3} \tag{4.8}$$

and also

$$\rho_\lambda^d < C \lambda^{p/(p+2)}. \tag{4.9}$$

In the exponentially stabilizing case (Theorem 2.1) we achieve this by taking $\rho_\lambda = \alpha \log \lambda$ for some suitably large constant α. In the polynomially stabilizing case of order γ (Theorem 2.2), we take $\rho_\lambda = C\lambda^a$ with

$$a = \frac{25p}{p\gamma - 6d} \quad \text{so} \quad a\left(\frac{\gamma}{6} - \frac{d}{p}\right) = \frac{25}{6}. \tag{4.10}$$

which implies that (4.8) holds with $q = 3$, and that (4.9) holds (to obtain the last conclusion we use our assumption on γ, which implies $\gamma > d(25 + 56/p)$.)

For all $1 \le i \le V$ and all $j = 1, 2, ...$, let $R_{i,j}$ denote the radius of stabilization of ξ at $(X_{i,j}, U_{i,j})$ if $1 \le j \le N_i$ and $X_{i,j} \in A_\lambda$; let $R_{i,j}$ be zero otherwise.

Let $E_{i,j} := \{R_{i,j} \le \rho_\lambda\}$. Let $E_\lambda := \cap_{i=1}^V \cap_{j=1}^\infty E_{i,j}$ Then by Markov's inequality and standard Palm theory (e.g. Theorem 1.6 in [9])

$$P[E_\lambda^c] \le \mathbb{E}\left[\sum_{i=1}^V \sum_{j=1}^{N_i} \mathbf{1}_{E_{i,j}^c}\right] = \lambda \int_{A_\lambda} P[R(x,\lambda) > \rho_\lambda]\kappa(x)dx \le \lambda\tau(\rho_\lambda). \tag{4.11}$$

For each λ, and $x \in \mathbb{R}^d$, set $f_\lambda(x) := f(x)\mathbf{1}_{A_\lambda}(x)$. Recalling the representation $\mathcal{P}_\lambda = \cup_{i=1}^{V(\lambda)}\{X_{i,j}\}_{j=1}^{N_i}$, we have

$$T_\lambda = \sum_{i=1}^{V(\lambda)} \sum_{j=1}^{N_i} \xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda) \cdot f_\lambda(X_{i,j}).$$

To obtain rates of normal approximation for T_λ, it will be be convenient to consider a closely related sum enjoying more independence between terms, namely

$$T'_\lambda := \sum_{i=1}^{V(\lambda)} \sum_{j=1}^{N_i} \xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda) \cdot \mathbf{1}_{E_{i,j}} \cdot f_\lambda(X_{i,j}).$$

For all $1 \le i \le V(\lambda)$ define

$$S_i := S_{Q_i} := (\mathrm{Var}T'_\lambda)^{-1/2} \sum_{j=1}^{N_i} \xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda) \cdot \mathbf{1}_{E_{i,j}} \cdot f_\lambda(X_{i,j})$$

and put $S := (\mathrm{Var}T'_\lambda)^{-1/2}(T'_\lambda - \mathbb{E}\, T'_\lambda) = \sum_{i=1}^{V(\lambda)}(S_i - \mathbb{E}\, S_i)$. Clearly $\mathrm{Var}S = \mathbb{E}\, S^2 = 1$.

We define a graph $G_\lambda := (\mathcal{V}_\lambda, \mathcal{E}_\lambda)$ as follows. The set $\mathcal{V}_\lambda$ consists of the sub-cubes $Q_1, ..., Q_{V(\lambda)}$ and edges (Q_i, Q_j) belong to $\mathcal{E}_\lambda$ if $d(Q_i, Q_j) \le 2\alpha\lambda^{-1/d}\rho_\lambda$, where $d(Q_i, Q_j) := \inf\{|x - y| : x \in Q_i, y \in Q_j\}$. By definition of the radius of stabilization $R(x, \lambda)$, the value of S_i is determined by the restriction of $\mathcal{P}_\lambda$ to the $\lambda^{-1/d}\rho_\lambda$-neighborhood of the cube Q_i. By the independence property of the Poisson point process, if A_1 and A_2 are disjoint collections of cubes in $\mathcal{V}_\lambda$ such that no edge in $\mathcal{E}_\lambda$ has one endpoint in A_1 and one endpoint in A_2, then the random variables $\{S_{Q_i}, Q_i \in A_1\}$ and $\{S_{Q_j}, Q_j \in A_2\}$ are independent. Thus G_λ is a dependency graph for $\{S_i\}_{i=1}^{V(\lambda)}$.

To prepare for an application of Lemma 4.1, we make five observations:

(i) $V(\lambda) := |\mathcal{V}_\lambda| = O(\lambda \rho_\lambda^{-d})$ as $\lambda \to \infty$.

(ii) Since the number of cubes in $Q_1, ..., Q_V$ distant at most $2\lambda^{-1/d}\rho_\lambda$ from a given cube is bounded by 5^d, it follows that the maximal degree D satisfies $D := D_\lambda \le 5^d$.

(iii) The definitions of S_i and $\xi_{i,j}$ and Lemma 4.3 tell us that for all $1 \le i \le V(\lambda)$

$$\|S_i\|_q \le C(\mathrm{Var}T'_\lambda)^{-1/2} \left\| \sum_{j=1}^{\infty} |\xi_{i,j}| \right\|_q \le C(\mathrm{Var}T'_\lambda)^{-1/2} \rho_\lambda^{d(p+1)/p}. \tag{4.12}$$

(iv) We can bound $\mathrm{Var}[T'_\lambda]$ as follows. Observe that T'_λ is the sum of $V(\lambda)$ random variables, which by the case $q = 2$ of Lemma 4.3 each have a second moment bounded by a constant multiple of $\rho_\lambda^{2d(p+1)/p}$. Thus the variance of each of the $V(\lambda)$ random variables is also bounded by a constant multiple of $\rho_\lambda^{2d(p+1)/p}$. Moreover, the covariance of any pair of the $V(\lambda)$ random variables is zero when the indices of the random variables correspond to non-adjacent cubes. For adjacent cubes, by the Cauchy-Schwarz inequality the covariance is also bounded by a constant multiple of $\rho_\lambda^{2d(p+1)/p}$. This shows that

$$\mathrm{Var}[T'_\lambda] = O(\rho_\lambda^{d(p+2)/p}\lambda). \tag{4.13}$$

(v) $\mathrm{Var}[T'_\lambda]$ is close to $\mathrm{Var}[T_\lambda]$ for λ large. We require more estimates to show this. Note that $|T'_\lambda - T_\lambda| = 0$ except possibly on the set E_λ^c. Lemma 4.3, along with Minkowski's inequality, yields the upper bound

$$\left\| \sum_{i=1}^{V(\lambda)} \sum_{j=1}^{N_i} |\xi_\lambda((X_{i,j}, U_{i,j}); \mathcal{P}_\lambda)| \mathbf{1}_{A_\lambda}(X_{i,j}) \right\|_q \le CV(\lambda)\rho_\lambda^{d(p+1)/p} \le C\lambda\rho_\lambda^{d/p}. \tag{4.14}$$

Since $T_\lambda = T'_\lambda$ on event E_λ, the Hölder and Minkowski inequalities yield

$$\begin{aligned} \|T_\lambda - T'_\lambda\|_2 &\le \|T_\lambda - T'_\lambda\|_q P[E_\lambda^c]^{(1/2)-(1/q)} \\ &\le (\|T_\lambda\|_q + \|T'_\lambda\|_q) P[E_\lambda^c]^{(q-2)/(2q)}. \end{aligned}$$

Hence, by (4.8), (4.11), and (4.14),

$$\|T_\lambda - T'_\lambda\|_2 \le C\lambda\rho_\lambda^{d/p}(\lambda\tau(\rho_\lambda))^{(q-2)/(2q)} \le C\lambda^{-3} \tag{4.15}$$

which implies that

$$\mathbb{E}\left[|T'_\lambda - T_\lambda|\right] \le C\lambda^{-3}, \tag{4.16}$$

which we use later. Since

$$\mathrm{Var}[T_\lambda] = \mathrm{Var}[T'_\lambda] + \mathrm{Var}(T_\lambda - T'_\lambda) + 2\mathrm{Cov}(T'_\lambda, T_\lambda - T'_\lambda),$$

by (4.15), (4.13), (4.9) and the Cauchy-Schwarz inequality, we obtain

$$|\mathrm{Var}[T_\lambda] - \mathrm{Var}[T'_\lambda]| \leq C\lambda^{-2}. \tag{4.17}$$

Given the five observations (i)-(v), we are now ready to apply Lemma 4.1 to prove Theorem 2.1. By (4.13) and (4.17), $\mathrm{Var}[T_\lambda]$, as a function of λ, is bounded on bounded intervals. Hence, it suffices to prove that there exists $\lambda_0 \geq 2$ such that (2.6) holds for all $\lambda \geq \lambda_0$, since we can then extend (2.6) to all $\lambda \geq 2$ by changing C if necessary. Trivially, (2.6) holds for large enough λ when $\mathrm{Var}[T_\lambda] < 1$, and so without loss of generality we now assume $\mathrm{Var}[T_\lambda] \geq 1$. To establish the error bound (2.6) in this case, we apply the bound (4.1) to $W_i := S_i - \mathbb{E}\, S_i,\ 1 \leq i \leq V_\lambda$, with

$$\theta := C(\mathrm{Var}T'_\lambda)^{-1/2}\rho_\lambda^{d(p+1)/p}.$$

Our choice of θ is applicable by (4.12). We clearly have $\mathbb{E}\,[W_i] = 0$ and $\mathbb{E}\,[(\sum_{i=1}^{V(\lambda)} W_i)^2] = 1$. With $S = \sum_{i=1}^{V(\lambda)} W_i$, Lemma 4.1 along with observation (i) above yields

$$\begin{aligned}\sup_t |P[S \leq t] - \Phi(t)| &\leq C\lambda\rho_\lambda^{-d}(\mathrm{Var}T'_\lambda)^{-q/2}\rho_\lambda^{dq(p+1)/p} \\ &\leq C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq},\end{aligned} \tag{4.18}$$

where the last line makes use of the fact that $\mathrm{Var}[T'_\lambda] \geq \mathrm{Var}[T_\lambda]/2$, which follows (for λ large) from (4.17).

Now if $\beta > 0$ is a constant and Z any random variable then by (4.18) we have for all $t \in \mathbb{R}$

$$\begin{aligned}P[Z \leq t] &\leq P[S \leq t + \beta] + P[|Z - S| \geq \beta] \\ &\leq \Phi(t+\beta) + C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} + P[|Z - S| \geq \beta] \\ &\leq \Phi(t) + C\beta + C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} + P[|Z - S| \geq \beta]\end{aligned}$$

by the Lipschitz property of Φ. Similarly for all $t \in \mathbb{R}$,

$$P[Z \leq t] \geq \Phi(t) - C\beta - C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} - P[|Z - S| \geq \beta].$$

In other words

$$\sup_t |P[Z \leq t] - \Phi(t)| \leq C\beta + C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} + P[|Z - S| \geq \beta]. \tag{4.19}$$

Now by definition of S,

$$|(\mathrm{Var}T'_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\, T_\lambda) - S| = |(\mathrm{Var}T'_\lambda)^{-1/2}\{(T_\lambda - \mathbb{E}\, T_\lambda) - (T'_\lambda - \mathbb{E}\, T'_\lambda)\}|$$

$$\leq (\mathrm{Var} T'_\lambda)^{-1/2}\{|T_\lambda - T'_\lambda| + \mathbb{E}\,[|T_\lambda - T'_\lambda|]\}$$

which by (4.16) is bounded by $C\lambda^{-3}$ except possibly on the set E_λ^c which has probability less than $C\lambda^{-2}$ by (4.11) and (4.8). Thus by (4.19) with $Z = (\mathrm{Var}T'_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\,T_\lambda)$ and $\beta = C\lambda^{-3}$

$$\sup_t |P[(\mathrm{Var}T'_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\,T_\lambda) \leq t] - \Phi(t)| \leq C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} + C\lambda^{-2}. \quad (4.20)$$

Moreover, by the triangle inequality

$$\sup_t \left|P[(\mathrm{Var}T_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\,T_\lambda) \leq t] - \Phi(t)\right| \leq$$

$$\leq \sup_t \left| P\left[(\mathrm{Var}T'_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\,T_\lambda) \leq t\cdot(\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda})^{1/2}\right] - \Phi\left(t(\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda})^{1/2}\right)\right| +$$

$$+ \sup_t \left|\Phi\left(t(\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda})^{1/2}\right) - \Phi(t)\right|. \quad (4.21)$$

Since for all $s \leq t$, we have $|\Phi(s) - \Phi(t)| \leq (t-s)\max_{s\leq u\leq t}\phi(u)$ where ϕ denotes the standard normal density, and since by (4.17) there is a constant $0 < C < \infty$ such that for all $\lambda > 0$ and all $t \in \mathbb{R}$

$$\left|t(\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda})^{1/2} - t\right| \leq |t|\left|\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda} - 1\right| \leq \frac{C|t|}{\lambda^2}$$

we get

$$\sup_t \left|\Phi\left(t(\frac{\mathrm{Var}T_\lambda}{\mathrm{Var}T'_\lambda})^{1/2}\right) - \Phi(t)\right| \leq C\sup_t \left(\left(\frac{|t|}{\lambda^2}\right)\left(\max_{u\in[t-tC/\lambda^2,\ t+tC/\lambda^2]}\phi(u)\right)\right) \leq \frac{C}{\lambda^2}.$$

Thus by (4.20) and (4.21),

$$\sup_t |P[(\mathrm{Var}T_\lambda)^{-1/2}(T_\lambda - \mathbb{E}\,T_\lambda) \leq t] - \Phi(t)| \leq C\lambda(\mathrm{Var}T_\lambda)^{-q/2}\rho_\lambda^{dq} + C\lambda^{-2}. \quad (4.22)$$

Finally we can deduce from (4.13) and (4.17) that $\mathrm{Var}T_\lambda = O(\lambda\rho_\lambda^{d(p+2)/p})$. Hence, under the assumptions of Theorem 2.1, in (4.22) the first term in the right hand side dominates, thus yielding the desired bound (2.6), and the proof of Theorem 2.1 is complete.

Under the assumptions of Theorem 2.2, provided $\sigma^2(f,\xi,\kappa) > 0$ the right hand side of (4.22) is bounded by $C\lambda^{-1/2}\rho_\lambda^{dq}$, and since in this case we set $\rho_\lambda = \lambda^a$ with a given by (4.10), some elementary algebra yields (2.9). Our assumption that $\gamma > d(150 + 6/p)$ then yields the central limit theorem behavior (2.8), which is also trivially true in the case where $\sigma^2(f,\xi,\kappa) = 0$. This completes the proof of Theorem 2.2. □

Acknowledgments. We began this work while visiting the Institute for Mathematical Sciences at the National University of Singapore, and continued it while visiting the Isaac Newton Institute for Mathematical Sciences at Cambridge. We thank both institutions for their hospitality.

References

[1] F. Avram and D. Bertsimas (1993), On central limit theorems in geometrical probability. *Ann. Appl. Probab.* **3**, 1033-1046.

[2] P. Baldi and Y. Rinott (1989), Asymptotic normality of some graph related statistics. *J. Appl. Probab.* **26**, 171-175.

[3] Yu. Baryshnikov and J.E. Yukich (2003), Gaussian fields and random packing. *J. Statist. Phys.* **111**, 443-463.

[4] Yu. Baryshnikov and J.E. Yukich (2004), Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.*, to appear.

[5] Yu. Baryshnikov, P. Eichelsbacher, T. Scheiber, and J. E. Yukich (2004), Moderate deviations and cluster measures in geometric probability, preprint.

[6] P. J. Bickel and L. Breiman (1983), Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11,** 185-214.

[7] L. Chen and Q.-M. Shao (2003), Normal approximation under local dependence, preprint, *Ann. Probab.*, to appear July 2004.

[8] N. Henze (1988), A multivariate two-sample test based on the number of nearest-neighbor type coincidences. *Ann. Statist.* **16**, 772-783.

[9] M.D. Penrose (2003), Random Geometric Graphs, Oxford University Press.

[10] M.D. Penrose (2004). Multivariate spatial central limit theorems with applications to percolation and spatial graphs. Preprint, University of Bath. Electronically available via `http://www.bath.ac.uk/math-sci/`

[11] M.D. Penrose and J.E. Yukich (2001), Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11**, 1005-1041.

[12] M.D. Penrose and J.E. Yukich (2002), Limit theory for random sequential packing and deposition. *Ann. Appl. Probab.* **12**, 272-301.

[13] M.D. Penrose and J.E. Yukich (2003), Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.*, **13**, 277-303.

[14] C. Stein (1972), Approximate Computation of Expectations. IMS, Hayward, CA.