

## Fine Gaussian fluctuations on the Poisson space, I: contractions, cumulants and geometric random graphs

Raphaël Lachièze-Rey\*      Giovanni Peccati†

### Abstract

We study the normal approximation of functionals of Poisson measures having the form of a finite sum of multiple integrals. When the integrands are nonnegative, our results yield necessary and sufficient conditions for central limit theorems. These conditions can always be expressed in terms of contraction operators or, equivalently, fourth cumulants. Our findings are specifically tailored to deal with the normal approximation of the geometric  $U$ -statistics introduced by Reitzner and Schulte (2011). In particular, we shall provide a new analytic characterization of geometric random graphs whose edge-counting statistics exhibit asymptotic Gaussian fluctuations, and describe a new form of Poisson convergence for stationary random graphs with sparse connections. In a companion paper, the above analysis is extended to general  $U$ -statistics of marked point processes with possibly rescaled kernels.

**Keywords:** Central Limit Theorems; Contractions; Malliavin Calculus; Poisson Limit Theorems; Poisson Space; Random Graphs; Stein's Method;  $U$ -statistics; Wasserstein Distance; Wiener Chaos.

**AMS MSC 2010:** 60H07; 60F05; 60G55; 60D05.

Submitted to EJP on June 22, 2012, final version accepted on March 2, 2013.

Supersedes arXiv:1111.7312.

Supersedes HAL:hal-00646866.

## 1 Introduction

This paper concerns the normal approximation of random variables living inside a fixed sum of Wiener chaoses associated with a Poisson measure over a Borel measure space. Our main theoretical tools come from the two papers [25, 26], respectively by Peccati *et al.* and Peccati and Zheng, where the normal approximation of functional of Poisson measures is studied by combining two probabilistic techniques, namely the *Stein's method* and the *Malliavin calculus of variations*.

---

\*Université Paris Descartes, France, and University of Luxembourg.

E-mail: raphael.lachieze-rey@parisdescartes.fr

†University of Luxembourg. E-mail: giovanni.peccati@gmail.com

We shall focus on conditions implying that a given sequence of random variables satisfies a central limit theorem (CLT), where the convergence in distribution takes place in the sense of the Wasserstein distance (see Section 1.1 for definitions). Our main concern is to provide *analytic conditions* for asymptotic normality, that is, conditions only involving expressions related to the kernels in the chaotic expansion of a given random variable. In particular, our approach does not involve computations based on the method of moments and cumulants (with the exception of Theorem 4.12, where we deal with Poisson approximations).

The main contributions of our paper are the following:

- In Theorem 3.5, we shall prove that conditions for asymptotic normality can be expressed in terms of norms of *contraction operators* (see Section 2.2). These analytic objects already appear in CLTs living inside a fixed Wiener chaos (see [25, 26]), and are a crucial tool in order to effectively assess bounds based on Malliavin operators. One further important point is that the use of contraction operators allows one to neatly distinguish the contribution of each chaotic projection to the CLT, as well as to deduce joint CLTs for these projections starting from the asymptotic normality of their sum (see Proposition 3.14).
- In Theorem 3.12 we shall prove that, when specialized to random variables such that each kernel in the Wiener-Itô representation has a constant sign, our results yield *necessary and sufficient conditions* for asymptotic normality. The main tools in order to show such a result are two new analytic bounds, stated in Proposition 3.8 and Proposition 3.9. These findings extend to the Poisson framework the ‘fourth moment theorem’ proved by Nualart and Peccati (in a Gaussian setting) in [20], a result that has been the starting point of a new line of research in stochastic analysis – see the book [18], as well as the constantly updated webpage

<http://www.iecn.u-nancy.fr/~nourdin/steinmalliavin.htm>

- As discussed below, random variables having Wiener-Itô kernels with constant sign appear quite naturally in problems arising in stochastic geometry. In particular, we shall use our results in order to provide an exhaustive characterization of stationary geometric random graphs whose edge counting statistics exhibit asymptotic Gaussian fluctuations (see Theorem 4.9). This family of geometric graphs contains e.g. *interval graphs* and *disk graphs* – see e.g. [5, 6, 7, 14, 16, 28]. Our characterization of geometric random graphs involves ‘diagonal subsets’ of Cartesian products, that are reminiscent of the combinatorial conditions for CLTs used by Blei and Janson in [4], in the framework of CLTs for finite Rademacher sums (see also [19, Section 6]). As a by-product of our analysis (see Theorem 4.12), we shall illustrate a new form of Poisson convergence for random graphs with sparse connections.

We stress that one of our main motivations comes from a remarkable paper by Reitzner and Schulte [30], laying the foundations of a general theory for CLTs involving  $U$ -statistics based on Poisson point processes. In particular, one of the crucial insights of [30] concerns the use of a formula by Last and Penrose (see [11]), providing explicit expressions for Wiener-Itô chaotic decompositions in terms of difference operators (see Theorem 2.9). It is interesting to note that Last and Penrose’s formula is the Poisson analogous of the so-called ‘Stroock formula’ of Malliavin calculus, which is in turn an important tool for proving CLTs involving non-linear functionals of Gaussian measures (see e.g. [18, Corollary 2.7.8] for a discussion of this point). We shall see that our findings complement and extend the results proved in [30] in several directions. See also

Decreusefond *et al.* [6], Ferraz and Vergne [7], Last *et al.* [12], Minh [15], Schulte [32], Schulte and Thaele [33, 34], for several new findings pertaining to this line of research.

In order to keep the length of this paper within bounds, in Section 4 we will present applications that are related to a very specific setting, namely edge-counting in random geometric graphs with possibly large connections. The power and flexibility of the results proved in the present work are further illustrated in the companion paper [10], where the following applications are developed in full detail:

- (i) analytic bounds for the normal approximation of  $U$ -statistics based on marked point processes, in particular  $U$ -statistics with rescaled kernels;
- (ii) bounds for general subgraph counting in the disk graph model under any regime;
- (iii) an exhaustive characterization of the asymptotic behavior of geometric  $U$ -statistics;
- (iv) applications to the Boolean model, and to subgraph counting in disk graph models with random radius.

The rest of this section is devoted to the formal presentation of the main problems that are addressed in this paper.

### 1.1 Poisson measures

Throughout the paper  $(Z, \mathcal{Z}, \mu)$  is a measure space such that  $Z$  is a Borel space,  $\mathcal{Z}$  is the associated Borel  $\sigma$ -field, and  $\mu$  is a  $\sigma$ -finite Borel measure with no atoms. We write  $\mathcal{Z}_\mu = \{B \in \mathcal{Z} : \mu(B) < \infty\}$  to denote the subclass of  $\mathcal{Z}$  composed of sets with finite measure. Also, we shall write  $\eta = \{\eta(B) : B \in \mathcal{Z}_\mu\}$  to indicate a *Poisson measure* on  $(Z, \mathcal{Z})$  with *control*  $\mu$ . In other words,  $\eta$  is a collection of random variables defined on some probability space  $(\Omega, \mathcal{F}, P)$ , indexed by the elements of  $\mathcal{Z}_\mu$  and such that: (i) for every  $B, C \in \mathcal{Z}_\mu$  such that  $B \cap C = \emptyset$ , the random variables  $\eta(B)$  and  $\eta(C)$  are independent; (ii) for every  $B \in \mathcal{Z}_\mu$ ,  $\eta(B)$  has a Poisson distribution with mean  $\mu(B)$ . We shall also write  $\hat{\eta}(B) = \eta(B) - \mu(B)$ ,  $B \in \mathcal{Z}_\mu$ , and  $\hat{\eta} = \{\hat{\eta}(B) : B \in \mathcal{Z}_\mu\}$ . A random measure verifying property (i) is usually called “completely random” or “independently scattered” (see e.g. [23] for a general introduction to these concepts).

**Remark 1.1.** As it is customary, by a slight abuse of notation, we shall often write  $x \in \eta$  in order to indicate that the point  $x \in Z$  is charged by the random measure  $\eta(\cdot)$ .

In this paper, we shall focus on sequences of random variables  $\{F_n : n \geq 1\}$  having a finite Wiener-Itô chaotic decomposition, that is, such that

$$F_n = \sum_{i=1}^k I_{q_i}(f_i^{(n)}), \quad n \geq 1, \tag{1.1}$$

where the symbol  $I_{q_i}$  indicates a multiple Wiener-Itô integral of order  $q_i$  with respect to  $\hat{\eta}$ , the integer  $k$  does not depend on  $n$ , and each  $f_i^{(n)}$  is a non-zero symmetric kernel from  $Z^{q_i}$  to  $\mathbb{R}$  (see Section 2.1 below for details). We will be specifically concerned with the forthcoming Problem 1.2. Recall that, given random variables  $U, Y \in L^1(P)$ , the *Wasserstein distance* between the law of  $U$  and the law of  $Y$  is defined as the quantity

$$d_W(U, Y) = \sup_{f \in \text{Lip}(1)} |E[f(U)] - E[f(Y)]|,$$

where  $\text{Lip}(1)$  indicates the class of Lipschitz real-valued function with Lipschitz constant  $\leq 1$ . It is well-known that the topology induced by  $d_W$ , on the class of probability measures on the real line, is strictly stronger than the one induced by the convergence in distribution.

**Problem 1.2.** Find analytic conditions on the kernels  $\{f_i^{(n)}\}$  ensuring that the sequence

$$\tilde{F}_n := \frac{F_n}{\sqrt{\text{Var}(F_n)}}, \quad n \geq 1,$$

converges in distribution, as  $n \rightarrow \infty$ , to a standard Gaussian random variable  $N \sim \mathcal{N}(0, 1)$ , in the sense of the Wasserstein distance. Determine under which assumptions these conditions are also necessary, and find explicit upper bounds for the sequence

$$d_W(\tilde{F}_n, N), \quad n \geq 1.$$

We will deal with Problem 1.2 in Section 3, where it is shown that a convenient solution can be deduced by using *contraction operators*. Among other features, these operators provide a neat way to deal with the product of multiple stochastic integral, and virtually replace the use of diagram formulae – see e.g. [23]. As anticipated, we will see that, in the specific case of random variables as in (1.1) such that  $f_i^{(n)} \geq 0$ , our results lead to necessary and sufficient conditions that are analogous to the so-called ‘fourth moment theorems’ for sequences of multiple integrals in a Gaussian setting – see [20].

**Remark 1.3.** Problem 1.2 is also explicitly studied in [30, Section 4]. In particular, Theorem 4.1 in [30] provides bounds in the Wasserstein distance for random variables having a finite chaotic decomposition, where the bounds are expressed in terms of expectations of inner products of multiple integral stochastic processes. On the other hand, Theorem 4.7 in [30] provides an analytic bound, involving sums over partitions, for the normal approximation of absolutely convergent  $U$ -statistics. Here, we call ‘analytic bound’ any upper bound only involving deterministic transformations of the kernel determining the  $U$ -statistic, without any additional probabilistic component.

## 1.2 Random graphs

As anticipated, we shall now apply our main theoretical results to the study of geometric random graphs whose edge-counting statistics satisfy a CLT. The class of geometric random graphs considered below allow for long connections, in the sense that the geometric rule used to define edges is based on the use of arbitrarily large sets and therefore is not local. It is worth noting by now that our setting represents a natural generalization of the so called *Gilbert graphs* – see Example 1.5 below. Also, as explained in Remark 1.8 below, part of the models we consider cannot be dealt with by directly using the powerful theory of *stabilization* (see e.g. [13]).

Now let the notation introduced in the previous section prevail. In what follows, we shall denote by  $W$  (as in ‘window’) a measurable subset of  $Z$  such that  $\mu(W) < \infty$ . We first introduce the notion of a *geometric random graph* based on the restriction of the Poisson measure  $\eta$  to  $W$ , and on some symmetric set  $H \subset Z \times Z$ .

**Definition 1.4** (Random geometric graphs). Let  $H \subset Z \times Z$  be such that  $\mu^2(H) < \infty$ ,  $H$  is symmetric (that is, for every  $(x, y) \in H$ , one also has  $(y, x) \in H$ ) and  $H$  is non-diagonal (that is,  $H$  does not contain any pair of the type  $(x, x)$ ).

- (a) The *random geometric graph* based on  $\eta$ ,  $W$  and  $H$  is the undirected random graph

$$G = G(\eta, W, H) = (V, \mathcal{E}),$$

such that: (i) the vertices of  $G$  are given by the class  $V = \eta \cap W = \{x \in \eta : x \in W\}$ , and (ii) a pair  $\{x, y\}$  belongs to the set  $\mathcal{E}$  of the edges of  $G$  if and only if  $(x, y) \in H$ . We observe that, since  $H$  is non-diagonal, then  $G$  has no loops, that is:  $G$  does not contain any edge of the type  $\{x, x\}$ .

- (b) Assume in addition that  $Z$  is a vector space. The random geometric graph at Point (a) is said to be *stationary* if there exists a set  $\overline{H} \subset Z$  such that

$$H = \{(x, y) \in Z^2 : x - y \in \overline{H}\}.$$

Note that, since  $H$  is symmetric, one has necessarily that  $\overline{H} = -\overline{H}$ ; moreover, since  $H$  has no diagonal components,  $0 \notin \overline{H}$ .

- Example 1.5.** (i) The class of random geometric graphs introduced above generalizes the notion of a *Gilbert graph*, obtained by taking  $Z$  equal to some metric space (endowed with a distance  $d$ ) and  $H = \{(x_1, x_2) \in Z^2 : 0 < d(x_1, x_2) < \delta\}$ ,  $\delta > 0$ . In this case, the random geometric graph  $G$  is obtained by connecting two vertices  $v_1, v_2 \in \eta \cap W$  if and only if  $d(v_1, v_2) < \delta$ . See e.g. [28].
- (ii) If  $Z = \mathbb{R}^d$  and  $H = \{(x_1, x_2) \in Z^2 : 0 < \|x_1 - x_2\|_{\mathbb{R}^d} < \delta\}$ , then the corresponding geometric random graph is stationary with  $\overline{H} = B(0, \delta) \setminus \{0\}$ , where  $B(0, \delta) \subset Z$  stands for the open ball of radius  $\delta$  centered at the origin. Graphs of this type are customarily called *interval graphs* when  $Z = \mathbb{R}$ , and *disk graphs* when  $Z = \mathbb{R}^2$  – see e.g. [5, 6, 7, 14, 16] for recent developments on the subject.

In Section 4, we shall use our general results in order to deal with the following problem.

**Problem 1.6.** Fix a set  $W \subset Z$ , as well as a symmetric function  $g : W \times W \rightarrow \mathbb{R}$ . Consider the following objects: (i) a collection of Poisson measures  $\eta_\lambda$ ,  $\lambda > 0$ , with  $\sigma$ -finite and non atomic control measures  $\mu_\lambda$  such that  $\mu_\lambda(W) < \infty$  and  $\mu_\lambda(W) \uparrow \infty$ , as  $\lambda \rightarrow \infty$ , and (ii) a collection of non-diagonal non-negligible symmetric sets  $H_\lambda \subset Z \times Z$ ,  $\lambda > 0$ . Define  $G_\lambda = (V_\lambda, \mathcal{E}_\lambda)$  to be the geometric random graph based on  $\eta_\lambda$ ,  $W$  and  $H_\lambda$ . Characterize those classes  $\{\mu_\lambda, H_\lambda : \lambda > 0\}$  such that the random variables

$$F(g, W; \eta_\lambda, H_\lambda) = \sum_{\{x, y\} \in \mathcal{E}_\lambda} g(x, y), \quad \lambda > 0, \tag{1.2}$$

verify the limit relation

$$\tilde{F}(g, W; \eta_\lambda, H_\lambda) := \frac{F(g, W; \eta_\lambda, H_\lambda) - E[F(g, W; \eta_\lambda, H_\lambda)]}{\sqrt{\text{Var}(F(g, W; \eta_\lambda, H_\lambda))}} \xrightarrow{\text{Law}} N \sim \mathcal{N}(0, 1), \tag{1.3}$$

as  $\lambda \rightarrow \infty$ .

**Remark 1.7.** (i) At this stage, the role of the window  $W$  might seem immaterial, and indeed the substance of Problem 1.6 does not change if one takes  $W = Z$ . However, the above formulation allows for the more general case of a window  $W = W_\lambda$  possibly depending on  $\lambda$ . Moving windows of this type appear in Section 4, as well as in the paper [10].

- (ii) In many examples and applications, one considers sets  $H_\lambda$  such that  $\alpha(H_\lambda \cap (W \times W)) \downarrow 0$ , as  $\lambda \rightarrow \infty$ , for some fixed measure  $\alpha$  on  $W \times W$ . Heuristically, the fact that  $\mu_\lambda(W) \uparrow \infty$  and  $\alpha(H_\lambda \cap (W \times W)) \downarrow 0$  ensures that the following phenomenon takes place: as  $\lambda$  grows, more and more vertices and edges are added to the geometric graph, whereas old edges are deleted as a consequence of the asymptotic negligibility of  $H_\lambda \cap (W \times W)$ . Solving Problem 1.6 in this framework is equivalent to characterizing all sequences of random geometric graphs such that the addition of vertices and the cancellation of edges compensate, thus generating asymptotic Gaussian fluctuations.

When specialized to the case of Gilbert graphs on  $Z = \mathbb{R}^d$ , Problem 1.6 is tackled in the classic reference [28, Chapter 3] as a special case of general subgraph counting. A comparison with the results of [28, Chapter 3] is provided in Section 4.3.1 below. A complete solution of Problem 1.6 for general subgraph counting in Gilbert graphs, based on the techniques developed in this paper, is presented in [10, Section 3]. See also [30, Section 6.2].

**Remark 1.8.** Assume that, for every  $x \in \eta$ , there exists a random radius  $R_x$  such that all the  $y$  connected to  $x$  in the random graph lie in the ball with center  $x$  and radius  $R_x$ . Then, the variable  $F = F(1, W; \eta_\lambda, H_\lambda)$  in (1.2) is *stabilizing*, meaning that  $F$  can be written in the form

$$F = \sum_{x \in \eta} \xi(x, \eta),$$

where  $\xi$  is such that  $\xi(x, \eta)$  is not modified by adding or removing a finite number of points to  $\eta$  outside the ball with center  $x$  and radius  $R_x$  (see [13] for more details on this topic). In our case, to fit the framework of formula (1.2) in the case  $g = 1$ ,  $\xi(x, \eta)$  should be defined as

$$\xi(x, \eta) = \frac{1}{2} \#(\{y \neq x : \{x, y\} \in \mathcal{E}_\lambda\}),$$

where  $\#A$  indicates the cardinality of  $A$ . The CLTs presented for instance in [1, 29] cover well this case. Remark that in this particular framework of a deterministic connection rule, stabilization theory only allows for a bounded length, while we consider here models where points can have arbitrarily long connections.

The rest of the paper is organized as follows. In Section 2, we discuss several background results concerning Poisson measures, Wiener chaos and  $U$ -statistics. Section 3 contains our main abstract results concerning the normal approximation of random variables having a finite chaotic decomposition. Section 4 focuses on random graphs and on several analytical characterizations of associated CLTs. An Appendix (see Section 5) provides some basic definitions and results of Malliavin calculus.

## 2 Preparation

### 2.1 Multiple integrals and chaos

As before,  $(Z, \mathcal{Z}, \mu)$  is a non-atomic Borel measure space, and  $\eta$  is a Poisson measure on  $Z$  with control  $\mu$ .

**Remark 2.1.** For every  $z \in Z$ , we denote by  $\delta_z$  the Dirac mass at  $z$ . By virtue of the assumptions on the space  $(Z, \mathcal{Z}, \mu)$ , and to simplify the discussion, we will assume throughout the paper that  $(\Omega, \mathcal{F}, P)$  and  $\eta$  are such that

$$\Omega = \left\{ \omega = \sum_{j=1}^n \delta_{z_j}, n \in \mathbb{N} \cup \{\infty\}, z_j \in Z \right\},$$

and  $\eta$  is defined as the *canonical mapping*

$$(\omega, B) \mapsto \eta(B)(\omega) = \omega(B), \quad B \in \mathcal{Z}_\mu, \quad \omega \in \Omega.$$

Also, the  $\sigma$ -field  $\mathcal{F}$  will be always supposed to be the  $P$ -completion of the  $\sigma$ -field generated by  $\eta$ .

Throughout the paper, for  $p \in [1, \infty)$ , the symbol  $L^p(\mu)$  is shorthand for  $L^p(Z, \mathcal{Z}, \mu)$ . For an integer  $q \geq 2$ , we shall write  $L^p(\mu^q) := L^p(Z^q, \mathcal{Z}^{\otimes q}, \mu^q)$ , whereas  $L_s^p(\mu^q)$  stands

for the subspace of  $L^p(\mu^q)$  composed of functions that are  $\mu^q$ -almost everywhere symmetric. Also, we adopt the convention  $L^p(\mu) = L^p_s(\mu) = L^p(\mu^1) = L^p_s(\mu^1)$  and use the following standard notation: for every  $q \geq 1$  and every  $f, g \in L^2(\mu^q)$ ,

$$\langle f, g \rangle_{L^2(\mu^q)} = \int_{Z^q} f(z_1, \dots, z_q)g(z_1, \dots, z_q)\mu^q(dz_1, \dots, dz_q), \quad \|f\|_{L^2(\mu^q)} = \langle f, f \rangle_{L^2(\mu^q)}^{1/2}.$$

For every  $f \in L^2(\mu^q)$ , we denote by  $\tilde{f}$  the canonical symmetrization of  $f$ , that is,

$$\tilde{f}(x_1, \dots, x_q) = \frac{1}{q!} \sum_{\sigma} f(x_{\sigma(1)}, \dots, x_{\sigma(q)}),$$

where  $\sigma$  runs over the  $q!$  permutations of the set  $\{1, \dots, q\}$ . Note that  $\|\tilde{f}\|_{L^2(\mu^q)} \leq \|f\|_{L^2(\mu^q)}$  (to see this, use for instance the triangle inequality).

**Definition 2.2.** For every deterministic function  $h \in L^2(\mu)$ , we write

$$I_1(h) = \hat{\eta}(h) = \int_Z h(z)\hat{\eta}(dz)$$

to indicate the *Wiener-Itô integral* of  $h$  with respect to  $\hat{\eta}$ . For every  $q \geq 2$  and every  $f \in L^2_s(\mu^q)$ , we denote by  $I_q(f)$  the *multiple Wiener-Itô integral*, of order  $q$ , of  $f$  with respect to  $\hat{\eta}$ . We also set  $I_q(f) = I_q(\tilde{f})$ , for every  $f \in L^2(\mu^q)$  (not necessarily symmetric), and  $I_0(b) = b$  for every real constant  $b$ .

The reader is referred for instance to the monograph [23], by Peccati and Taqqu, for a complete discussion of multiple Wiener-Itô integrals and their properties (including the forthcoming Proposition 2.3 and Proposition 2.4).

**Proposition 2.3.** *The following equalities hold for every  $q, m \geq 1$ , every  $f \in L^2_s(\mu^q)$  and every  $g \in L^2_s(\mu^m)$ :*

1.  $E[I_q(f)] = 0$ ,
2.  $E[I_q(f)I_m(g)] = q! \langle f, g \rangle_{L^2(\mu^q)} \mathbf{1}_{(q=m)}$  (isometric property).

The Hilbert space composed of the random variables with the form  $I_q(f)$ , where  $q \geq 1$  and  $f \in L^2_s(\mu^q)$ , is called the  $q$ th *Wiener chaos* associated with the Poisson measure  $\eta$ . The following well-known *chaotic representation property* is an essential feature of Poisson random measures. Recall that  $\mathcal{F}$  is assumed to be generated by  $\eta$ .

**Proposition 2.4** (Wiener-Itô chaotic decomposition). *Every random variable*

$$F \in L^2(\Omega, \mathcal{F}, P) := L^2(P)$$

*admits a (unique) chaotic decomposition of the type*

$$F = E[F] + \sum_{i=1}^{\infty} I_i(f_i), \tag{2.1}$$

*where the series converges in  $L^2(P)$  and, for each  $i \geq 1$ , the kernel  $f_i$  is an element of  $L^2_s(\mu^i)$ .*

**2.2 Star contractions and multiplication formulae**

We shall now introduce *contraction operators*, and succinctly discuss some of their properties. As anticipated in the Introduction, these objects are at the core of our main results.

The kernel  $f \star_r^l g$  on  $Z^{p+q-r-l}$ , associated with functions  $f \in L_s^2(\mu^p)$  and  $g \in L_s^2(\mu^q)$ , where  $p, q \geq 1, r = 1, \dots, p \wedge q$  and  $l = 1, \dots, r$ , is defined as follows:

$$\begin{aligned}
 & f \star_r^l g(\gamma_1, \dots, \gamma_{r-l}, t_1, \dots, t_{p-r}, s_1, \dots, s_{q-r}) \\
 = & \int_{Z^l} \mu^l(dz_1, \dots, dz_l) f(z_1, \dots, z_l, \gamma_1, \dots, \gamma_{r-l}, t_1, \dots, t_{p-r}) \\
 & \times g(z_1, \dots, z_l, \gamma_1, \dots, \gamma_{r-l}, s_1, \dots, s_{q-r}).
 \end{aligned}
 \tag{2.2}$$

Roughly speaking, the star operator ' $\star_r^l$ ' reduces the number of variables in the tensor product of  $f$  and  $g$  from  $p+q$  to  $p+q-r-l$ : this operation is realized by first identifying  $r$  variables in  $f$  and  $g$ , and then by integrating out  $l$  among them. To deal with the case  $l = 0$  for  $r = 0, \dots, p \wedge q$ , we set

$$\begin{aligned}
 & f \star_r^0 g(\gamma_1, \dots, \gamma_r, t_1, \dots, t_{p-r}, s_1, \dots, s_{q-r}) \\
 = & f(\gamma_1, \dots, \gamma_r, t_1, \dots, t_{p-r}) g(\gamma_1, \dots, \gamma_r, s_1, \dots, s_{q-r}),
 \end{aligned}$$

and

$$f \star_0^0 g(t_1, \dots, t_p, s_1, \dots, s_q) = f \otimes g(t_1, \dots, t_p, s_1, \dots, s_q) = f(t_1, \dots, t_p) g(s_1, \dots, s_q).$$

By using the Cauchy-Schwarz inequality, one sees immediately that  $f \star_r^r g$  is square-integrable for any choice of  $r = 0, \dots, p \wedge q$ , and every  $f \in L_s^2(\mu^p), g \in L_s^2(\mu^q)$ .

**Remark 2.5.** For every  $1 \leq p \leq q$  and every  $r = 1, \dots, p$ ,

$$\int_{Z^{p+q-r}} (f \star_r^0 g)^2 d\mu^{p+q-r} = \int_{Z^r} (f \star_p^{p-r} f)(g \star_q^{q-r} g) d\mu^r,
 \tag{2.3}$$

for every  $f \in L_s^2(\mu^p)$  and every  $g \in L_s^2(\mu^q)$

The next statement contains an important *product formula* for Poisson multiple integrals (see e.g. [23] for a proof).

**Proposition 2.6** (Product formula). *Let  $f \in L_s^2(\mu^p)$  and  $g \in L_s^2(\mu^q)$ ,  $p, q \geq 1$ , and suppose moreover that  $f \star_r^l g \in L^2(\mu^{p+q-r-l})$  for every  $r = 1, \dots, p \wedge q$  and  $l = 1, \dots, r$  such that  $l \neq r$ .*

Then,

$$I_p(f) I_q(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} \sum_{l=0}^r \binom{r}{l} I_{p+q-r-l}(\widetilde{f \star_r^l g}),
 \tag{2.4}$$

with the tilde  $\sim$  indicating a symmetrization, that is,

$$\widetilde{f \star_r^l g}(x_1, \dots, x_{p+q-r-l}) = \frac{1}{(p+q-r-l)!} \sum_{\sigma} f \star_r^l g(x_{\sigma(1)}, \dots, x_{\sigma(p+q-r-l)}),$$

where  $\sigma$  runs over all  $(p+q-r-l)!$  permutations of the set  $\{1, \dots, p+q-r-l\}$ .

### 2.3 About the Malliavin formalism

For the rest of the paper, we shall use definitions and results related to Malliavin-type operators defined on the space of functionals of the Poisson measure  $\eta$ . Our formalism coincides with the one introduced by Nualart and Vives in [21]. In particular, we shall denote by  $D$ ,  $\delta$ ,  $L$  and  $L^{-1}$ , respectively, the *Malliavin derivative*, the *divergence operator*, the *Ornstein-Uhlenbeck generator* and its *pseudo-inverse*. The domains of  $D$ ,  $\delta$  and  $L$  are written  $\text{dom}D$ ,  $\text{dom}\delta$  and  $\text{dom}L$ . The domain of  $L^{-1}$  is given by the subclass of  $L^2(P)$  composed of centered random variables. For the convenience of the reader we have collected some crucial definitions and results in the Appendix (see Section 5). Here, we just recall that, since the underlying probability space  $\Omega$  is assumed to be the collection of discrete measures described in Remark 2.1, then one can meaningfully define the random variable  $\omega \mapsto F_z(\omega) = F(\omega + \delta_z)$ ,  $\omega \in \Omega$ , for every given random variable  $F$  and every  $z \in Z$ , where  $\delta_z$  is the Dirac mass at  $z$ . One can therefore prove that the following neat representation of  $D$  as a *difference operator* is in order.

**Lemma 2.7.** *For each  $F \in \text{dom}D$ ,*

$$D_z F = F_z - F, \text{ a.e. } \mu(dz).$$

A complete proof of Lemma 2.7 can be found in [21].

### 2.4 $U$ -statistics

Following [30, Section 3.1], we now introduce the concept of a  $U$ -statistic associated with the Poisson measure  $\eta$ .

**Definition 2.8** ( *$U$ -statistics*). Fix  $k \geq 1$ . A random variable  $F$  is called a  *$U$ -statistic of order  $k$* , based on the Poisson measure  $\eta$ , if there exists a kernel  $f \in L^1_s(\mu^k)$  such that

$$F = \sum_{(x_1, \dots, x_k) \in \eta_{\neq}^k} f(x_1, \dots, x_k), \tag{2.5}$$

where the symbol  $\eta_{\neq}^k$  indicates the class of all  $k$ -dimensional vectors  $(x_1, \dots, x_k)$  such that  $x_i \in \eta$  and  $x_i \neq x_j$  for every  $1 \leq i \neq j \leq k$ . As made clear in [30, Definition 3.1], the possibly infinite sum appearing in (2.5) must be regarded as the  $L^1(P)$  limit of objects of the type  $\sum_{(x_1, \dots, x_k) \in \eta_{\neq}^k \cap A_n} f(x_1, \dots, x_k)$ ,  $n \geq 1$ , where the sets  $A_n \in Z^k$  are such that  $\mu^k(A_n) < \infty$  and  $A_n \uparrow Z^k$ , as  $n \rightarrow \infty$ .

Plainly, a  $U$ -statistic of order one is just a linear functional of  $\eta$ , with the form

$$\sum_{x \in \eta} f(x) = \int_Z f(x) \eta(dx),$$

for some  $f \in L^1(\mu)$ . The following statement, based on the results proved by Reitzner and Schulte in [30], collects two crucial properties of  $U$ -statistics.

**Theorem 2.9** (See [30]). *Let  $F \in L^1(P)$  be a  $U$ -statistic as in (2.5). Then, the following two properties hold.*

(a) *The expectation of  $F$  is given by*

$$E[F] = \int_{Z^k} f(z_1, \dots, z_k) \mu^k(dz_1, \dots, dz_k). \tag{2.6}$$

- (b) If  $F$  is also square-integrable, then necessarily  $f \in L_s^2(\mu^k)$ , and the Wiener-Itô representation (2.1) of  $F$  is such that  $f_i = 0$ , for  $i \geq k + 1$ , and

$$f_i(x_1, \dots, x_i) = \binom{k}{i} \int_{Z^{k-i}} f(x_1, \dots, x_i, z_1, \dots, z_{k-i}) \mu^{k-i}(dz_1, \dots, dz_{k-i}) \quad (2.7)$$

for every  $i = 1, \dots, k$ . In particular,  $f_k = f$ . For each  $i = 1, \dots, k$ , one has that  $f_i \in L_s^1(\mu^i) \cap L_s^2(\mu^i)$ .

One should note that formula (2.7) follows from an application of the results proved by Last and Penrose in [11].

### 2.5 $U$ -statistics and random graphs

In this paper, we will be interested in characterizing the Gaussian fluctuations of  $U$ -statistics having a specific *support*. In particular this allows one to deal with the set of ‘local  $U$ -statistics’ introduced by Reitzner and Schulte in [30, Section 6]. Recall that a set  $H \subset Z^k$  is called *symmetric* if the following implication holds: if  $(x_1, \dots, x_k) \in H$ , then  $(x_{\sigma(1)}, \dots, x_{\sigma(k)}) \in H$  for every permutation  $\sigma$  of  $\{1, \dots, k\}$ .

**Definition 2.10** (Support of a  $U$ -statistic). Let  $k \geq 2$ , and let  $H \subset Z^k$  be a measurable symmetric set. A  $U$ -statistic  $F$  as in (2.5) is said to have *support in  $H$*  if the function  $f$  is such that

$$f(x_1, \dots, x_k) = 0, \quad \forall (x_1, \dots, x_k) \notin H.$$

**Example 2.11** (Local  $U$ -statistics). Let  $Z$  be a metric space. Then, the class of *local  $U$ -statistics*, as defined in [30, Section 6], coincides with the family of  $U$ -statistics having support in a set of the type  $H = \{(x_1, \dots, x_k) : \text{diam}(\{x_1, \dots, x_k\}) < \delta\}$  for some  $\delta > 0$ . Here, the symbol  $\text{diam}(B)$  is shorthand for the diameter of  $B$ .

We shall now point out a well-known connection between  $U$ -statistics and hypergraphs. Recall that a *hypergraph* of order  $k \geq 2$  is a pair  $(V, \mathcal{E})$ , where  $V = (v_1, \dots, v_m)$  is a set of vertices, and  $\mathcal{E} = (E_1, \dots, E_s)$  is a collection of (possibly non-disjoint) subsets of  $V$  (called *edges*), such that each  $E_i$  contains exactly  $k$  elements; in particular a hypergraph of order 2 is an undirected graph.

**Remark 2.12** ( $U$ -statistics as graph statistics). (i) Let  $k \geq 2$ , let  $F$  be a  $U$ -statistic as in (2.5), and assume that  $f = \mathbf{1}_W \times \mathbf{1}_H$ , where  $W \subset Z$  is some set (usually called a ‘window’) such that  $\mu(W) < \infty$ . Then, the random variable  $\frac{1}{k!}F$  counts the number of edges in the random hypergraph  $(V, \mathcal{E})$ , obtained as follows:  $V = \eta \cap W$ , and the class of edges  $\mathcal{E}$  is composed of all subsets  $\{x_1, \dots, x_k\} \subset V$  such that  $(x_1, \dots, x_k) \in H$ .

- (ii) If  $k = 2$ ,  $Z$  is some metric space (endowed with a distance  $d$ ) and  $H = \{(x_1, x_2) \in Z^2 : d(x_1, x_2) < \delta\}$  and  $f = \mathbf{1}_W \times \mathbf{1}_H$ , then the random variable  $\frac{1}{2}F$  counts the number of edges in the undirected graph whose vertices  $V$  are given by the points of  $W$  charged by  $\eta$  and such that two vertices  $v_1, v_2$  are connected by an edge if and only if  $0 < d(v_1, v_2) < \delta$ . These are the ‘Gilbert random graphs’ discussed in Example 1.5(i).

To conclude, we present the notion of a *stationary  $U$ -statistic*. It will play an important role in Section 4.

**Definition 2.13** (Stationary  $U$ -statistics). Fix  $k \geq 2$ , assume that  $Z$  is a vector space, and let  $F$  be a  $U$ -statistic of the type (2.5), having support in a symmetric set  $H$ . We shall say that  $F$  is *stationary* if there exists  $\overline{H} \subset Z^{k-1}$  such that

$$H = \{(x_1, \dots, x_k) : (x_2 - x_1, x_3 - x_1, \dots, x_k - x_1) \in \overline{H}\}. \quad (2.8)$$

**Example 2.14.** Consider the case  $k = 2$ ,  $Z = \mathbb{R}^d$  and  $H = \{(x_1, x_2) \in Z^2 : \|x_1 - x_2\|_{\mathbb{R}^d} < \delta\}$ . Then, the corresponding  $U$ -statistic  $F$  is stationary, with  $\overline{H} = B(0, \delta)$ , where  $B(0, \delta) \subset Z$  stands for the open ball of radius  $\delta$  centered at the origin. See Example 1.5(ii).

### 3 Normal approximations for finite chaotic expansions

#### 3.1 Framework

We shall tackle Problem 1.2, by focussing on the normal approximation of random variables  $F$  having the form

$$F = E[F] + \sum_{i=1}^k I_{q_i}(f_i), \tag{3.1}$$

where:

- $k \geq 1$  is an integer;
- the integers  $q_i$ ,  $i = 1, \dots, k$ , are such that  $1 \leq q_1 < q_2 < \dots < q_k$ ;
- the symbol  $I_q$  indicates a multiple Wiener-Itô integral of order  $q$ , with respect to a centered Poisson measure  $\hat{\eta} = \eta - \mu$ , where  $\eta$  is a Poisson measure on the Borel measurable space  $(Z, \mathcal{Z})$ , with deterministic and  $\sigma$ -finite non-atomic control measure  $\mu$ ;
- each kernel  $f_i$  is a nonzero element of  $L_s^2(\mu^{q_i})$ , and the class  $\{f_i : i = 1, \dots, k\}$  verifies in addition the forthcoming Assumption 3.1.

**Assumptions 3.1** (Technical assumptions on integrands). Let the notation of Section 2.2 prevail. Every random variable of the type (3.1) considered in the sequel of this paper is such that the following properties (i)-(iii) are verified.

- (i) For every  $i = 1, \dots, k$  and every  $r = 1, \dots, q_i$ , the kernel  $f_i \star_r^{q_i-r} f_i$  is an element of  $L^2(\mu^r)$ .
- (ii) For every  $i$  such that  $q_i \geq 2$ , every contraction of the type  $(z_1, \dots, z_{2q_i-r-l}) \mapsto |f_i| \star_r^l |f_i|(z_1, \dots, z_{2q_i-r-l})$  is well-defined and finite for every  $r = 1, \dots, q_i$ , every  $l = 1, \dots, r$  and every  $(z_1, \dots, z_{2q_i-r-l}) \in Z^{2q_i-r-l}$ .
- (iii) For every  $i, j = 1, \dots, d$  such that  $\max(q_i, q_j) > 1$ , for every  $k = |q_i - q_j| \vee 1, \dots, q_i + q_j - 2$  and every  $(r, l)$  verifying  $k = q_i + q_j - 2 - r - l$ ,

$$\int_Z \left[ \sqrt{\int_{Z^k} (f_i(z, \cdot) \star_r^l f_j(z, \cdot))^2 d\mu^k} \right] \mu(dz) < \infty,$$

where, for every fixed  $z \in Z$ , the symbol  $f_i(z, \cdot)$  denotes the mapping  $(z_1, \dots, z_{q-1}) \mapsto f_i(z, z_1, \dots, z_{q-1})$ .

**Remark 3.2.** According to [26, Lemma 2.9 and Remark 2.10], Point (i) in Assumption 3.1 implies that the following properties (a)-(c) are verified:

- (a) for every  $1 \leq i < j \leq k$ , for every  $r = 1, \dots, q_i \wedge q_j$  and every  $l = 1, \dots, r$ , the contraction  $f_i \star_r^l f_j$  is a well-defined element of  $L^2(\mu^{q_i+q_j-r-l})$ ;
- (b) for every  $1 \leq i \leq j \leq k$  and every  $r = 1, \dots, q_i$ ,  $f_i \star_r^0 f_j$  is an element of  $L^2(\mu^{q_i+q_j-r})$ ;
- (c) for every  $i = 1, \dots, k$ , for every  $r = 1, \dots, q_i$ , and every  $l = 1, \dots, r \wedge (q_i - 1)$ , the kernel  $f_i \star_r^l f_i$  is a well-defined element of  $L^2(\mu^{2q_i-r-l})$ .

In particular, the multiplication formula (2.4) implies that every random variable  $F$  verifying Assumption 3.1 is such that  $I_{q_i}(f_i)^2 \in L^2(P)$  for every  $i = 1, \dots, k$ , yielding in turn that  $E[F^4] < \infty$ . Following a similar route, one can also show that, under Assumption 3.1, the random variable  $\langle DF, -DL^{-1}F \rangle_{L^2(\mu)}$  is square-integrable (and not merely an element of  $L^1(P)$ ).

**Remark 3.3.** Assumption 3.1 imply that Assumptions A-B-C in [26] are satisfied, so that the computations therein can be directly applied in our framework.

**Remark 3.4.** For instance, Assumption 3.1 is verified whenever each  $f_i$  is a bounded function with support in a rectangle of the type  $B \times \dots \times B$ , where  $\mu(B) < \infty$ .

### 3.2 A general bound

Let  $F$  be a random variable as in (3.1) such that  $E[F^2] = \sigma^2 > 0$  ( $\sigma > 0$ ) and  $E[F] = m \in \mathbb{R}$ , and consider a Gaussian random variable  $N \sim \mathcal{N}(m, \sigma^2)$  with the same mean and variance. Define

$$B_1(F; \sigma) = \frac{1}{\sigma} E [|\sigma^2 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)}|] \tag{3.2}$$

$$+ \frac{1}{\sigma^2} \int_Z E[(D_z F)^2 | D_z L^{-1}F] \mu(dz),$$

$$B_2(F; \sigma) = \frac{1}{\sigma} \sqrt{E [(\sigma^2 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)})^2]} \tag{3.3}$$

$$+ \frac{1}{\sigma^2} \int_Z E[(D_z F)^2 | D_z L^{-1}F] \mu(dz).$$

Then, a slight modification of [25, Theorem 3.1] (the modification resides in the fact that we consider an arbitrary variance  $\sigma^2$ ) yields the following estimates:

$$d_W(F, N) \leq B_1(F; \sigma) \leq B_2(F; \sigma), \tag{3.4}$$

The next statement shows that  $B_2(F; \sigma^2)$  can be further bounded in terms of the contractions introduced in Section 2.2.

**Theorem 3.5.** *Let  $F$  and  $N$  be the random variables appearing in (3.4)-(3.3). Then, there exists a universal constant  $C_0 = C_0(q_1, \dots, q_k) \in (0, \infty)$ , depending only on  $q_1, \dots, q_k$ , such that*

$$B_2(F; \sigma) \leq C_0 \times B_3(F; \sigma), \tag{3.5}$$

where

$$B_3(F; \sigma) = \frac{1}{\sigma} \left\{ \max_1 \|f_i \star_r^l f_i\|_{L^2(\mu^{2q_i-r-l})} + \max_2 \|f_i \star_r^l f_j\|_{L^2(\mu^{q_i+q_j-r-l})} + \max_{i=1, \dots, k} \|f_i\|_{L^4(\mu^{q_i})}^2 \right\}. \tag{3.6}$$

In the previous expression,  $\max_1$  ranges over all  $1 \leq i \leq k$  such that  $q_i > 1$ , and all pairs  $(r, l)$  such that  $r \in \{1, \dots, q_i\}$  and  $1 \leq l \leq r \wedge (q_i - 1)$ , whereas  $\max_2$  ranges over all  $1 \leq i < j \leq d$  and all pairs  $(r, l)$  such that  $r \in \{1, \dots, q_i\}$  and  $l \in \{1, \dots, r\}$ . When  $q_1 = 1$ , one can replace  $\|f_1\|_{L^4(\mu)}^2$  in the previous bound by the smaller quantity

$$\frac{\|f_1\|_{L^3(\mu)}^3}{\|f_1\|_{L^2(\mu)}}. \tag{3.7}$$

*Proof of Theorem 3.5.* Without loss of generality, we can assume that  $m = 0$ . Also, throughout this proof, we write  $F_i = I_{q_i}(f_i)$  and  $\sigma_i^2 = E[I_{q_i}(f_i)^2] = q_i! \|f_i\|_{L^2(\mu^{q_i})}^2$ ,  $i = 1, \dots, k$ , in such a way that  $\sigma^2 = E[F^2] = \sum_{i=1}^k E[F_i^2] = \sum_{i=1}^k \sigma_i^2$ . Now write

$$\begin{aligned} \sqrt{E[(\sigma^2 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)})^2]} &\leq \sum_{i=1}^k \mathbf{1}_{q_i > 1} \sqrt{E[(\sigma_i^2 - \langle DF_i, -DL^{-1}F_i \rangle_{L^2(\mu)})^2]} \\ &+ \sum_{1 \leq i \neq j \leq k} \sqrt{E[(\langle DF_i, -DL^{-1}F_j \rangle_{L^2(\mu)})^2]}, \end{aligned}$$

so that one can directly apply [26, Proposition 5.5] and deduce that there exists a constant  $a = a(q_1, \dots, q_k)$  such that

$$\begin{aligned} &\sqrt{E[(\sigma^2 - \langle DF, -DL^{-1}F \rangle_{L^2(\mu)})^2]} \\ &\leq a \left\{ \max_1 \|f_i \star_r^l f_i\|_{L^2(\mu^{2q_i - r - l})} + \max_2 \|f_i \star_r^l f_j\|_{L^2(\mu^{q_i + q_j - r - l})} \right\}. \end{aligned}$$

To conclude, observe that

$$\int_Z E[(D_z F)^2 |D_z L^{-1}F|] \mu(dz) \leq \int_Z E \left[ \left( \sum_{i=1}^k |D_z F_i| \right)^2 \left( \sum_{i=1}^k |D_z L^{-1}F_i| \right) \right] \mu(dz),$$

so that [26, Proposition 5.6] implies that there exists a constant  $b = b(q_1, \dots, q_k)$  such that

$$\int_Z E[(D_z F)^2 |D_z L^{-1}F|] \mu(dz) \leq b\sigma \left\{ \max_1 \|f_i \star_r^l f_i\|_{L^2(\mu^{2q_i - r - l})} + \max_{i=1, \dots, k} \|f_i\|_{L^4(\mu^{q_i})}^2 \right\}.$$

Taking  $C_0 = a + b$  yields the desired conclusion. The last assertion in the statement comes from the fact that, when  $q_1 = 1$ ,  $\int_Z E[(D_z F)^2 |D_z L^{-1}F|] \mu(dz) = \|f_1\|_{L^3(\mu)}^3$ .  $\square$

**Remark 3.6.** According to [26, Lemma 2.9], for every quadruple  $(i, j, r, l)$  entering the expression of  $\max$  in (3.6), the following estimate holds:

$$\|f_i \star_r^l f_j\|_{L^2(\mu^{q_i + q_j - r - l})}^2 \leq \|f_i \star_r^l f_i\|_{L^2(\mu^{2q_i - r - l})} \times \|f_j \star_r^l f_j\|_{L^2(\mu^{2q_j - r - l})}. \quad (3.8)$$

### 3.3 Estimates for positive kernels

We shall now specialize Theorem 3.5 to the case on random variables having the form (3.1) and such that  $f_i \geq 0$ . In particular, we shall unveil some useful connections between the quantity  $B_3(F; \sigma)$  and the fourth cumulant of  $F$ .

**Remark 3.7.** Random variables admitting a Wiener-Itô chaotic expansion with positive kernels appear rather naturally in stochastic geometry. For instance, an application of (2.7) shows that any  $U$ -statistic with a positive kernel admits a Wiener-Itô chaotic expansion of this type. Note that many *counting statistics* have the form of  $U$ -statistics with a non-negative integer-valued kernel – such as for instance the subgraph-counting statistics in random geometric graphs (see e.g. [28, Chapter 3] and the references therein), or the statistics associated with hyperplane tessellations considered in [8].

The following statement concerns random variables of the form (3.1), in the special case where  $E[F] = 0$ ,  $k = 1$  and the multiple stochastic integral has a nonnegative kernel. One should note that, whenever  $F$  is a centered random variable with finite

fourth moment and variance  $\sigma^2 > 0$ , then the quantity  $E[F^4] - 3\sigma^4$  coincides with the fourth cumulant of  $F$ , customarily denoted by  $\chi_4(F)$ , defined as

$$\chi_4(F) = (-i)^n \frac{\partial^n}{\partial z^n} \ln g_F(z)|_{z=0}, \quad \text{with } g_F(z) := E[e^{izF}].$$

**Proposition 3.8** (Fourth moment bound, I). *Consider a random variable  $F$  as in (3.1), with the special form  $F = I_q(f)$ ,  $q \geq 1$ , where  $f \geq 0$  and  $E[F^2] = q! \|f\|_{L^2(\mu^q)}^2 = \sigma^2 > 0$ . Then,*

$$B_3(F; \sigma) = \frac{1}{\sigma} \left\{ \max_1 \|f \star_r^l f\|_{L^2(\mu^{2q-r-l})} + \|f\|_{L^4(\mu^q)}^2 \right\},$$

and there exist universal constants  $c_1 = c_1(q) < C_1 = C_1(q)$ , depending only on  $q$ , such that

$$c_1 \times B_3(F; \sigma) \leq \sqrt{E[F^4] - 3\sigma^4} \leq C_1 \times B_3(F; \sigma). \tag{3.9}$$

*Proof.* Using the multiplication formula (2.4), together with (2.3) and the positivity assumptions on  $f$ , one sees that  $F^2$  can be written in the following form

$$F^2 = \sigma^2 + I_{2q}(f \star_0^0 f) + R,$$

where  $R$  is a random variable orthogonal to the constants and to  $I_{2q}(f \star_0^0 f)$  and such that

$$J \left\{ \max_1 \|f \star_r^l f\|_{L^2(\mu^{2q-r-l})}^2 + \|f\|_{L^4(\mu^q)}^4 \right\} \leq E[R^2] \leq K \left\{ \max_1 \|f \star_r^l f\|_{L^2(\mu^{2q-r-l})}^2 + \|f\|_{L^4(\mu^q)}^4 \right\},$$

for some universal positive constants  $J, K$  depending uniquely on  $q$ . The conclusion is obtained by using the relation

$$E[I_{2q}(f \star_0^0 f)^2] = (2q)! \widetilde{\|f \star_0^0 f\|_{L^2(\mu^{2q})}^2} = 2\sigma^4 + \sum_{p=1}^{q-1} \frac{q!^4}{p!^2(q-p)!^2} \widetilde{\|f \star_p^p f\|_{L^2(\mu^{2q-2p})}^2},$$

where we have used [23, formula (11.6.30)]. □

The following general bound deals with random variables of the form (3.1) and with positive kernels. It will be used in the proof of Theorem 3.12.

**Proposition 3.9** (Fourth moment bound, II). *Let  $F$  be as in (3.1), with  $k \geq 1$ , and assume moreover that  $E[F] = 0$ , and  $E[F^2] = \sigma^2 > 0$ , and  $f_i \geq 0$  for every  $i$ . Then, there exists a universal constant  $C_2 = C_2(q_1, \dots, q_k)$ , depending uniquely on  $q_1, \dots, q_k$ , such that*

$$\left\{ \max_1 \|f_i \star_r^l f_i\|_{L^2(\mu^{2q_i-r-l})} + \max_{i=1, \dots, k} \|f_i\|_{L^4(\mu^{q_i})}^2 \right\} \leq C_2 \sqrt{E[F^4] - 3\sigma^4}. \tag{3.10}$$

*Proof.* Write as before  $F_i = I_{q_i}(f_i)$ , and  $\sigma_i^2 = E[F_i^2]$ ,  $i = 1, \dots, k$ . We can now write

$$\begin{aligned} E[F^4] - 3\sigma^4 &= \sum_{i=1}^k \{E[F_i^4] - 3\sigma_i^4\} \\ &+ 6 \sum_{1 \leq i < j \leq k} \{E[F_i^2 F_j^2] - \sigma_i^2 \sigma_j^2\} + \sum_{(i_1, i_2, i_3, i_4) \in V_d} E[F_{i_1} F_{i_2} F_{i_3} F_{i_4}] \\ &:= W + Y + Z, \end{aligned}$$

where  $V_d$  stands for the collection of those  $(i_1, i_2, i_3, i_4) \in \{1, \dots, k\}^4$ , such that one of the following conditions is satisfied: (a)  $i_1 \neq i_2 = i_3 = i_4$ , (b)  $i_1 \neq i_2 = i_3 \neq i_4$  and  $i_4 \neq i_1$ , (c) the elements of  $(i_1, i_2, i_3, i_4)$  are all distinct. Applying the multiplication formula (2.4) and exploiting the fact that each  $f_i$  is nonnegative, we immediately deduce that  $Y \geq 0$  and  $Z \geq 0$ , so that the desired conclusion follows from Proposition 3.8. □

### 3.4 Conditions for asymptotic Gaussianity

This section contains a general statement (Theorem 3.12) about the normal approximation of random variables admitting a finite chaotic decomposition. The first part of such a result provides sufficient conditions for Central Limit Theorems, that are directly based on Theorem 3.5. As indicated in the subsequent parts, these conditions become necessary whenever the involved kernels are nonnegative. Theorem 3.12 is one of the main results of the paper, and is the main tool used to deduce the CLTs appearing in Section 4 and in [10].

More precisely, in what follows we shall fix integers  $k \geq 1$  and  $1 \leq q_1 < q_2 < \dots < q_k$  (not depending on  $n$ ), and consider a sequence  $\{F^{(n)} : n \geq 1\}$  of random variables with the form

$$F^{(n)} = \sum_{i=1}^k I_{q_i}(f_i^{(n)}), \quad n \geq 1 \tag{3.11}$$

each verifying the same assumptions as the random variable  $F$  appearing in (3.1) (in particular, Assumption 3.1 is satisfied for each  $n$ ). We also use the following additional notation: (i)  $\sigma^2(n) = E[(F^{(n)})^2]$ , (ii)  $F_i^{(n)} = I_{q_i}(f_i^{(n)})$ ,  $i = 1, \dots, k$ , and (iii)  $\sigma_i^2(n) = E[(F_i^{(n)})^2]$ .

**Remark 3.10.** One of the main achievements of the forthcoming Theorem 3.12 is the ‘fourth moment theorem’ appearing at Point 3 in the statement, which only holds for random variables such that the kernels in the chaotic decomposition are nonnegative. As first proved in [20] (see also [18, Chapter 5]) an analogous result holds for general sequences of multiple Wiener-Itô integrals with respect to a Gaussian process. In particular, in the Gaussian framework one does not need to assume that the integrands have a constant sign. Proving analogous statements in a Poisson setting is quite a demanding task, one of main reasons being the rather intricate multiplication formula (2.4). Some previous partial findings in the Poisson case can be found in Peccati and Taqqu [22] (for sequences of double integrals) and in Peccati and Zheng [27] (for sequences of multiple integrals having the form of homogeneous sums).

**Remark 3.11.** In the statement of Theorem 3.12, we implicitly allow that the underlying Poisson measure  $\eta$  also changes with  $n$ . In particular, one can assume that the associated control measure  $\mu = \mu_n$  explicitly depends on  $n$ . This general framework is needed for the geometric applications developed in Section 4.

**Theorem 3.12.** *Let  $\{F^{(n)}\}$  be a sequence of random variables as in (3.11), and assume that there exists  $\sigma^2 > 0$  such that  $\lim_{n \rightarrow \infty} \sigma^2(n) = \sigma^2$ . Let  $N \sim \mathcal{N}(0, \sigma^2)$ .*

1. *For every  $n$ , one has the estimate*

$$d_W(F^{(n)}, N) \leq C_0 \times B_3(F^{(n)}; \sigma(n)) + \frac{\sqrt{2/\pi}}{\sigma(n) \vee \sigma} |\sigma^2(n) - \sigma^2|. \tag{3.12}$$

*In particular, if  $B_3(F^{(n)}; \sigma(n)) \rightarrow 0$ , as  $n \rightarrow \infty$ , then  $d_W(F^{(n)}, N) \rightarrow 0$  and therefore  $F^{(n)} \xrightarrow{\text{Law}} N$ .*

2. *Assume that  $f_i^{(n)} \geq 0$  for every  $i, n$ . Then, a sufficient condition in order to have that  $B_3(F^{(n)}; \sigma(n)) \rightarrow 0$  is that  $E[(F^{(n)})^4] - 3\sigma^4(n) \rightarrow 0$ .*
3. *Assume that  $f_i^{(n)} \geq 0$  for every  $i, n$ , and also that the sequence  $(F^{(n)})^4$ ,  $n \geq 1$ , is uniformly integrable. Then, the following conditions (a)–(c) are equivalent, as  $n \rightarrow \infty$ : (a)  $d_W(F^{(n)}, N) \rightarrow 0$ , (b)  $B_3(F^{(n)}; \sigma(n)) \rightarrow 0$ , and (c)  $E[(F^{(n)})^4] - 3\sigma^4(n) \rightarrow 0$ .*

*Proof.* 1. Let  $N_n \sim \mathcal{N}(0, \sigma^2(n))$ ,  $n \geq 1$ . Then, one has that (see e.g. [18, Proposition 3.6.1])

$$d_W(N_n, N) \leq \frac{\sqrt{2/\pi}}{\sigma(n) \vee \sigma} |\sigma^2(n) - \sigma^2|,$$

so that the desired conclusion follows from Theorem 3.5 as well as the inequality

$$d_W(F^{(n)}, N) \leq d_W(F^{(n)}, N_n) + d_W(N_n, N).$$

2. If  $E[(F^{(n)})^4] - 3\sigma^4(n) \rightarrow 0$ , then relation (3.10) implies that

$$\left\{ \max_1 \|f_i^{(n)} \star_r^l f_i^{(n)}\|_{L^2(\mu^{2q_i-r-l})} + \max_{i=1, \dots, k} \|f_i^{(n)}\|_{L^4(\mu^{q_i})}^2 \right\} \rightarrow 0.$$

Using (3.8), we see that the last relation implies that  $B_3(F^{(n)}; \sigma(n)) \rightarrow 0$ , so that the desired conclusion follows from Point 1 in the statement.

3. In view of Point 1 and Point 2 in the statement, we shall only prove that (a)  $\Rightarrow$  (c). To prove this implication, just observe that  $F^{(n)} \xrightarrow{L^{\text{aw}}} N$  and  $\{(F^{(n)})^4\}$  is uniformly integrable, then necessarily  $E[(F^{(n)})^4] \rightarrow E[N^4] = 3\sigma^4$ , so that the conclusion follows from the fact that  $\sigma^2(n) \rightarrow \sigma^2$ .  $\square$

**Remark 3.13.** A sufficient condition in order to have that the sequence  $\{(F^{(n)})^4\}$  is uniformly integrable is the following: there exists some  $\epsilon > 0$  such that

$$\sup_{n \geq 1} E|F^{(n)}|^{4+\epsilon} < \infty.$$

We shall use some estimates taken from [26] (see, in particular, Table 2, p. 1505, therein). Given a thrice differentiable function  $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$ , we set

$$\|\varphi''\|_\infty = \max_{1 \leq i_1 \leq i_2 \leq k} \sup_{x \in \mathbb{R}^k} \left| \frac{\partial^2}{\partial x_{i_1} \partial x_{i_2}} \varphi(x) \right|.$$

$$\|\varphi'''\|_\infty = \max_{1 \leq i_1 \leq i_2 \leq i_3 \leq k} \sup_{x \in \mathbb{R}^k} \left| \frac{\partial^3}{\partial x_{i_1} \partial x_{i_2} \partial x_{i_3}} \varphi(x) \right|.$$

**Proposition 3.14.** *Let the assumptions and notation of Theorem 3.12 prevail, and suppose that  $B_3(F^{(n)}; \sigma) \rightarrow 0$ , as  $n \rightarrow \infty$ . Let  $N_{n,i} \sim \mathcal{N}(0, \sigma_i^2(n))$ ,  $i = 1, \dots, k$ . Then, for every thrice differentiable function  $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$ , such that  $\|\varphi''\|_\infty, \|\varphi'''\|_\infty < \infty$ , one has that*

$$E[\varphi(F_1^{(n)}, \dots, F_k^{(n)})] - E[\varphi(N_{n,1}, \dots, N_{n,k})] \rightarrow 0, \quad n \rightarrow \infty.$$

*Proof.* According to [26], the following estimate takes place:

$$\begin{aligned} & |E[\varphi(F_1^{(n)}, \dots, F_k^{(n)})] - E[\varphi(N_{n,1}, \dots, N_{n,k})]| \\ & \leq \frac{k}{2} \|\varphi''\|_\infty \sqrt{\sum_{i,j=1}^k E[(\sigma_i^2(n) \mathbf{1}_{i=j} - \langle DF_i, -DL^{-1}F_j \rangle_{L^2(\mu)})^2]} \\ & + \frac{1}{4} \|\varphi'''\|_\infty \int_Z E \left[ \left( \sum_{i=1}^k |D_z F_i| \right)^2 \left( \sum_{i=1}^k |D_z L^{-1} F_i| \right) \right] \mu(dz), \end{aligned}$$

so that the conclusion follows from (3.5).  $\square$

## 4 Edge-counting in random geometric graphs: from Gaussian fluctuations to clustering

### 4.1 Framework

Our aim is now to tackle Problem 1.6. Throughout this section we shall work under the following slightly more restrictive setting (we use the notation of Problem 1.6).

- For every  $\lambda > 0$ , the control measure of  $\eta_\lambda$  is  $\mu_\lambda(dx) = \lambda \times \theta(dx)$ , where  $\theta$  is a  $\sigma$ -finite non-atomic measure on  $(Z, \mathcal{Z})$ .
- The symmetric function  $g : W \times W \rightarrow \mathbb{R}$  is bounded (this assumption can be relaxed – see the discussion below).

In the forthcoming Section 4.2, we will show that the normal approximation of the random variables  $F(g, W; \eta_\lambda, H_\lambda)$  (as defined in (1.2)) can be completely characterized in terms of some diagonal restrictions of Cartesian products of the sets  $H_\lambda$ . Among several consequences, this remarkable phenomenon implicitly provides a new geometric interpretation of contraction operators.

**Definition 4.1** (Sets with diagonal restrictions). Given the sets  $H_\lambda$ ,  $\lambda > 0$ , defining the random variables  $F(g, W; \eta_\lambda, H_\lambda)$  and  $\tilde{F}(g, W; \eta_\lambda, H_\lambda)$ , respectively in (1.2) and (1.3), we define the following four sets

- $H_\lambda^{(1)} := \{(x_1, x_2, x_3) \in Z^3 : (x_1, x_2) \in H_\lambda, (x_1, x_3) \in H_\lambda\} \subset Z^3$ ;
- $H_\lambda^{(2)} := \{(x_1, x_2, x_3, x_4, x_5) \in Z^5 : (x_1, x_i) \in H_\lambda, \forall i = 2, \dots, 5\} \subset Z^5$ ;
- $H_\lambda^{(3)} = \{(x_1, x_2, x_3, x_4) \in Z^4 : (x_1, x_2) \in H_\lambda, (x_2, x_3) \in H_\lambda, (x_3, x_4) \in H_\lambda, (x_4, x_1) \in H_\lambda\} \subset Z^4$ ;
- $H_\lambda^{(4)} := \{(x_1, x_2, x_3, x_4, x_5) \in Z^5 : (x_1, x_2) \in H_\lambda, (x_2, x_3) \in H_\lambda, (x_3, x_4) \in H_\lambda, (x_4, x_5) \in H_\lambda\} \subset Z^5$ .

We shall also use the following quantities, defined for every  $\lambda > 0$ :

(i)

$$V_\lambda^2(g) = V_{1,\lambda}^2(g) + V_{2,\lambda}^2(g) \\ := 4\lambda^3 \int_{W^3 \cap H_\lambda^{(1)}} g \star_1^0 g(x_1, x_2, x_3) \theta^3(dx_1, dx_2, dx_3) + 2\lambda^2 \int_{W^2 \cap H_\lambda} g^2(x_1, x_2) \theta^2(dx_1, dx_2);$$

(ii)  $A_\lambda(g) = \lambda^{5/2} \sqrt{\int_{W^5 \cap H_\lambda^{(2)}} \prod_{i=2}^5 g(x_1, x_i) \theta^5(dx_1, \dots, dx_5)}$ ;

(iii)  $B_\lambda(g) = \lambda^2 \sqrt{\int_{W^4 \cap H_\lambda^{(3)}} g(x_1, x_2) g(x_3, x_4) g(x_1, x_4) g(x_2, x_3) \theta^4(dx_1, \dots, dx_4)}$ ;

(iv)  $C_\lambda(g) = \lambda^{3/2} \sqrt{\int_{W^3 \cap H_\lambda^{(1)}} g^2(x_1, x_2) g^2(x_1, x_3) \theta^3(dx_1, dx_2, dx_3)}$ ;

(v)  $D_\lambda(g) = \lambda \sqrt{\int_{W^2 \cap H_\lambda} g^4(x_1, x_2) \theta^2(dx_1, dx_2)}$ ;

(vi)  $E_\lambda(g) = \lambda^{5/2} \sqrt{\int_{W^5 \cap H_\lambda^{(4)}} g(x_1, x_2) g(x_1, x_3) g(x_2, x_4) g(x_3, x_5) \theta^5(dx_1, \dots, dx_5)}$ .

The following expressions are obtained by setting  $g = 1$ :

(i')  $V_\lambda^2(1) = V_{1,\lambda}^2(1) + V_{2,\lambda}^2(1) = 4\lambda^3 \theta(W^3 \cap H_\lambda^{(1)}) + 2\lambda^2 \theta^2(W^2 \cap H_\lambda)$ ;

(ii')  $A_\lambda(1) = \lambda^{5/2} \sqrt{\theta^5(W^5 \cap H_\lambda^{(2)})}$ ;

(iii')  $B_\lambda(1) = \lambda^2 \sqrt{\theta^4(W^4 \cap H_\lambda^{(3)})}$ ;

(iv')  $C_\lambda(1) = \lambda^{3/2} \sqrt{\theta^3(W^3 \cap H_\lambda^{(1)})}$ ;

(v')  $D_\lambda(1) = \lambda \sqrt{\theta^2(W^2 \cap H_\lambda)}$ ;

(vi')  $E_\lambda(1) = \lambda^{5/2} \sqrt{\theta^5(W^5 \cap H_\lambda^{(4)})}$ .

**4.2 General conditions and bounds**

We start with a general estimate.

**Theorem 4.2** (General bound for geometric graphs). *Let the previous assumptions and notation prevail, and let  $N \sim \mathcal{N}(0, 1)$ . Then, there exists a universal constant  $C$ , not depending on  $\lambda$ , such that, for every  $\lambda > 0$ ,*

$$d_W(\tilde{F}(g, W; \eta_\lambda, H_\lambda), N) \leq \frac{C_0}{V_\lambda^2(g)} \times \max\{A_\lambda(g), B_\lambda(g), C_\lambda(g), D_\lambda(g), E_\lambda(g)\}. \tag{4.1}$$

If the class  $\{\tilde{F}(g, W; \eta_\lambda, H_\lambda)^4 : \lambda > 0\}$  is uniformly integrable, then the RHS of (4.1) converges to zero, as  $\lambda \rightarrow \infty$ , if and only if the CLT (1.3) takes place.

*Proof.* In what follows, we write  $F_\lambda = F(g, W; \eta_\lambda, H_\lambda)$  and  $\tilde{F}_\lambda = \tilde{F}(g, W; \eta_\lambda, H_\lambda)$  to simplify the notation. Last and Penrose’s formula (2.7) implies that the random variable  $F_\lambda$  admits the following chaotic decomposition

$$F_\lambda = E[F_\lambda] + I_1(f_{1,\lambda}) + I_2(f_{2,\lambda}) := E[F_\lambda] + F_{1,\lambda} + F_{2,\lambda}, \tag{4.2}$$

where  $f_{1,\lambda}(x) = 2\lambda \int_Z \mathbf{1}_{\{H_\lambda \cap W \times W\}}(x, y)g(x, y)\theta(dy)$  and

$$f_{2,\lambda}(x_1, x_2) = \mathbf{1}_{H_\lambda \cap (W \times W)}(x_1, x_2)g(x_1, x_2).$$

Routine computations imply then that

$$V_{i,\lambda}^2(g) = \text{Var}(F_{i,\lambda}), \quad i = 1, 2. \tag{4.3}$$

It follows that

$$\tilde{F}_\lambda = I_1(f_1/V_\lambda(g)) + I_2(f_2/V_\lambda(g)).$$

The upper bound (4.1) is now obtained by using (3.6), as well as the following relations, that can be proved by a standard use of the Fubini Theorem:

$$\begin{aligned} 4A_\lambda(g) &= \|f_{1,\lambda}\|_{L^4(\mu_\lambda)}^2, \\ B_\lambda(g) &= \|f_{2,\lambda} \star_1^1 f_{2,\lambda}\|_{L^2(\mu_\lambda^2)}, \\ C_\lambda(g) &= \|f_{2,\lambda} \star_2^1 f_{2,\lambda}\|_{L^2(\mu_\lambda)}, \\ D_\lambda(g) &= \|f_{2,\lambda}\|_{L^4(\mu_\lambda^2)}^2, \\ 2E_\lambda(g) &= \|f_{1,\lambda} \star_1^1 f_{2,\lambda}\|_{L^2(\mu_\lambda)}. \end{aligned}$$

The last assertion in the statement follows from a direct application of Theorem 3.12.□

The next two statements provide simplified bounds in case one of the two elements of the chaotic decomposition of  $\tilde{F}(g, W; \eta_\lambda, H_\lambda)$  converges to zero, as  $\lambda \rightarrow \infty$ . The proof (which is standard and left to the reader) uses (4.3) as well as the following basic estimate: if  $Q, R, S$  are three random variables in  $L^1(P)$ , then

$$d_W(Q + R, S) \leq d_W(R, S) + E[|Q|].$$

**Proposition 4.3** (Dominating first chaos). *If*

$$\frac{V_{2,\lambda}(g)}{V_{1,\lambda}(g)} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \tag{4.4}$$

then there exists a constant  $C_1$ , independent of  $\lambda$ , such that, for  $\lambda$  large enough,

$$d_W(\tilde{F}(g, W; \eta_\lambda, H_\lambda), N) \leq C_1 \left\{ \frac{V_{2,\lambda}(g)}{V_{1,\lambda}(g)} + \frac{1}{V_{1,\lambda}^2(g)} \times \max\{A_\lambda(g)\} \right\}. \tag{4.5}$$

If the class  $\{\tilde{F}(g, W; \eta_\lambda, H_\lambda)^4 : \lambda > 0\}$  is uniformly integrable, then the RHS of (4.5) converges to zero, as  $\lambda \rightarrow \infty$ , if and only if the CLT (1.3) takes place.

**Proposition 4.4** (Dominating second chaos). *If*

$$\frac{V_{1,\lambda}(g)}{V_{2,\lambda}(g)} \rightarrow 0, \quad \text{as } \lambda \rightarrow \infty, \tag{4.6}$$

*then there exists a constant  $C_2$ , independent of  $\lambda$ , such that, for  $\lambda$  large enough,*

$$d_W(\tilde{F}(g, W; \eta_\lambda, H_\lambda), N) \leq C_2 \left\{ \frac{V_{1,\lambda}(g)}{V_{2,\lambda}(g)} + \frac{1}{V_{2,\lambda}^2(g)} \times \max\{B_\lambda(g), C_\lambda(g), D_\lambda(g)\} \right\}. \tag{4.7}$$

*If the class  $\{\tilde{F}(g, W; \eta_\lambda, H_\lambda)^4 : \lambda > 0\}$  is uniformly integrable, then the RHS of (4.7) converges to zero, as  $\lambda \rightarrow \infty$ , if and only if the CLT (1.3) takes place.*

**4.3 Edge counting in stationary graphs**

For the rest of the section, we fix an integer  $d \geq 1$ , and focus on the case of a random Poisson measure on  $\mathbb{R}^d$ ; the symbol  $\ell$  will denote the Lebesgue measure on  $\mathbb{R}^d$ . For every  $\lambda > 0$ , we define the set  $Q_\lambda = [-\frac{1}{2}\lambda^{1/d}, \frac{1}{2}\lambda^{1/d}]^d$ . We specialize the framework of the previous two sections to the following setting

$$Z = \mathbb{R}^d, \quad W = Q_1, \quad g = 1, \quad \mu_\lambda = \lambda\ell, \tag{4.8}$$

We shall assume that, for every  $\lambda > 0$ , the symmetric non-diagonal set  $H_\lambda$  has the form

$$H_\lambda = \{(x, y) : x - y \in \overline{H}_\lambda\},$$

for some set  $\overline{H}_\lambda$  verifying

$$\ell(\overline{H}_\lambda \cap Q_1) > 0. \tag{4.9}$$

We insist that the novelty here (with respect to the usual setting of disk graphs – see e.g. [28, Chapter 3] and the references therein) is that  $\overline{H}_\lambda$  need not be bounded, allowing for arbitrarily distant points to be connected. This is especially relevant whenever  $\overline{H}_\lambda = \alpha_\lambda \overline{H}_1$ , where  $\alpha_\lambda$  is a scaling factor and  $\overline{H}_1$  is a fixed unbounded geometric connection rule. Unlike in the classical literature of stochastic geometry, e.g. in stabilization theory, this allows for models with unbounded interactions, such as between distant particles. As already recalled, our approach is further applied in [10], where  $U$ -statistics with general stationary kernels (not only taking values 0 or 1), and general order  $k \geq 2$ , are considered.

For every  $\lambda > 0$ , we shall write

$$F_\lambda = F(1, Q_1; \eta_\lambda, H_\lambda) \quad \text{and} \quad \tilde{F}_\lambda = \tilde{F}(1, Q_1; \eta_\lambda, H_\lambda), \tag{4.10}$$

where we used the notation introduced in (1.2)–(1.3). With this notation, each  $\frac{1}{2}F_\lambda$  is a stationary  $U$ -statistic (see Definition 2.13), counting the number of edges in the stationary random graph based on  $H_\lambda$  (see Definition 1.4). The chaotic decomposition of  $F_\lambda$  is written

$$F_\lambda = E[F_\lambda] + F_{1,\lambda} + F_{2,\lambda},$$

where we have adopted the same notation as in (4.2). Since  $g = 1$ , Problem 1.6 becomes the following: *characterize all collections of sets  $\{\overline{H}_\lambda\}$  such that the CLT (1.3) takes place, and assess the rate of convergence in the Wasserstein distance.*

**Remark 4.5.** For every  $\lambda > 0$ , one has the equality in law

$$F_\lambda \stackrel{\text{Law}}{=} \sum_{x,y \in \eta \cap Q_\lambda, x \neq y} \mathbf{1}_{x-y \in \overline{G}_\lambda}, \lambda > 0, \tag{4.11}$$

where  $\eta$  is a random Poisson measure with Lebesgue intensity, and  $\overline{G}_\lambda$  is a measurable subset of  $\mathbb{R}^d$  defined by the relation

$$\overline{H}_\lambda = \lambda^{-1/d} \overline{G}_\lambda, \tag{4.12}$$

so that

$$\ell(\overline{G}_\lambda \cap Q_\lambda) > 0. \tag{4.13}$$

Given two mappings  $\lambda \mapsto \gamma_\lambda, \lambda \mapsto \delta_\lambda$ , we write  $\gamma_\lambda \asymp \delta_\lambda$  if there are two positive constants  $C, C' > 0$  such that  $C\gamma_\lambda \leq \delta_\lambda \leq C'\gamma_\lambda$  for  $\lambda$  sufficiently large. We write  $\gamma_\lambda \sim \delta_\lambda$  if  $\delta_\lambda > 0$  for  $\lambda$  sufficiently large and  $\gamma_\lambda/\delta_\lambda \rightarrow 1$ .

One of the main points developed in the present section is that the asymptotic Gaussianity of the class  $\{\tilde{F}_\lambda\}$  results can be effectively studied by using the *occupation coefficient* of  $\overline{H}_\lambda$ , defined as

$$\psi(\lambda) := \ell(\overline{H}_\lambda \cap Q_1) = \frac{\ell(\overline{G}_\lambda \cap Q_\lambda)}{\ell(Q_\lambda)}, \quad \lambda > 0. \tag{4.14}$$

We also write  $\hat{W} = W - W = Q_{2^d}$  and  $\check{W} = Q_{1/2^d}$  (this is the largest set such that  $\hat{W} - \check{W} \subseteq W$ ) and define, for  $\lambda > 0$ ,

$$\check{\psi}(\lambda) := \ell(\overline{H}_\lambda \cap \check{W}) = \frac{\ell(\overline{G}_\lambda \cap Q_{\lambda/2^d})}{\ell(Q_\lambda)}, \quad \lambda > 0, \tag{4.15}$$

$$\hat{\psi}(\lambda) := \ell(\overline{H}_\lambda \cap \hat{W}) = \frac{\ell(\overline{G}_\lambda \cap Q_{2^d\lambda})}{\ell(Q_\lambda)}, \quad \lambda > 0. \tag{4.16}$$

In order to obtain necessary and sufficient conditions for asymptotic normality, we will often work under the additional assumption that

$$\check{\psi}(\lambda) \asymp \hat{\psi}(\lambda). \tag{4.17}$$

In this case, one trivially has that  $\check{\psi}(\lambda) \asymp \hat{\psi}(\lambda) \asymp \psi(\lambda)$ , and the value of  $\psi$  is only relevant up to a fixed multiplicative constant.

**Remark 4.6** (O-regularity). Assume that the geometric rule defined by  $\overline{G}_\lambda$  does not depend on  $\lambda$ , i.e.:  $\overline{G}_\lambda = \overline{G}$  for some fixed measurable set  $\overline{G}$ , in such a way that each set  $\overline{H}_\lambda$  is obtained by rescaling  $\overline{G}$  by a factor  $\lambda^{-1/d}$ . Then, condition (4.17) is implied by the following stronger assumption:  $\psi(a\lambda) \asymp \psi(\lambda)$  for every  $a > 0$ . In the terminology of [3, Section 2.2], this is equivalent to saying that  $\psi$  is *O-regular*.

In view of using the bounds appearing in Theorem 4.2, we have the following crucial estimates:

**Theorem 4.7.** *Let the previous notation and assumption prevail, set  $V_{i,\lambda}(1) = V_{i,\lambda}$ ,  $i = 1, 2$ , as well as  $A_\lambda = A_\lambda(1), B_\lambda = B_\lambda(1), \dots, E_\lambda = E_\lambda(1)$  – see Section 4.1. The following estimates are in order for every fixed  $\lambda > 0$ :*

$$\begin{aligned} 2^{-d}\check{\psi}(\lambda) &\leq \lambda^{-2}V_{2,\lambda}^2 = 2\lambda^{-2}D_\lambda^2 \leq 2^d\hat{\psi}(\lambda), \\ 2^{-d}\check{\psi}(\lambda)^2 &\leq \lambda^{-3}V_{1,\lambda}^2 = \lambda^{-3}C_\lambda^2 \leq 2^d\hat{\psi}(\lambda)^2, \\ 2^{-d}\check{\psi}(\lambda)^4 &\leq \lambda^{-5}A_\lambda^2 \leq 2^d\hat{\psi}(\lambda)^4, \\ \lambda^{-4}B_\lambda^2 &\leq 2^d\hat{\psi}(\lambda)^3, \\ 2^{-d}\check{\psi}(\lambda)^4 &\leq \lambda^{-5}E_\lambda^2 \leq 2^d\hat{\psi}(\lambda)^4. \end{aligned}$$

*Proof.* We introduce the changes of variables denoted by  $\varphi^{(i)}$ ,  $i = 0, 1, 2, 3, 4$ , where

$$\begin{aligned}\varphi^{(0)}(x_1, x_2) &= (x_1, u = x_1 - x_2), \\ \varphi^{(1)}(x_1, x_2, x_3) &= (x_1, u = x_1 - x_2, v = x_1 - x_3), \\ \varphi^{(2)}(x_1, x_2, x_3, x_4, x_5) &= (x_1, u = x_1 - x_2, v = x_1 - x_3, w = x_1 - x_4, z = x_1 - x_5), \\ \varphi^{(3)}(x_1, x_2, x_3, x_4) &= (x_1, u = x_1 - x_2, v = x_2 - x_3, w = x_3 - x_4), \\ \varphi^{(4)}(x_1, x_2, x_3, x_4, x_5) &= (x_1, u = x_1 - x_2, v = x_2 - x_3, w = x_3 - x_4, z = x_4 - x_5).\end{aligned}$$

Using the notation introduced in Definition 4.1 we have

$$\begin{aligned}\ell(W^2 \cap H_\lambda) &= \int_{W^2} \mathbf{1}_{x_1-x_2 \in \overline{H}_\lambda} dx_1 dx_2 = \int_{\varphi^{(0)}(W^2)} \mathbf{1}_{u \in \overline{H}_\lambda} dx_1 du, \\ \ell(W^3 \cap H_\lambda^{(1)}) &= \int_{W^3} \mathbf{1}_{x_1-x_2 \in \overline{H}_\lambda} \mathbf{1}_{x_1-x_3 \in \overline{H}_\lambda} dx_1 dx_2 dx_3 = \int_{\varphi^{(1)}(W^3)} \mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} dx_1 dudv, \\ \ell(W^5 \cap H_\lambda^{(2)}) &= \int_{W^5} \mathbf{1}_{x_1-x_2 \in \overline{H}_\lambda} \mathbf{1}_{x_1-x_3 \in \overline{H}_\lambda} \mathbf{1}_{x_1-x_4 \in \overline{H}_\lambda} \mathbf{1}_{x_1-x_5 \in \overline{H}_\lambda} dx_1 dx_2 dx_3 dx_4 dx_5 \\ &= \int_{\varphi^{(2)}(W^5)} \mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} \mathbf{1}_{w \in \overline{H}_\lambda} \mathbf{1}_{z \in \overline{H}_\lambda} dx_1 dudvdwdz, \\ \ell(W^4 \cap H_\lambda^{(3)}) &= \int_{W^4} \mathbf{1}_{x_1-x_2 \in \overline{H}_\lambda} \mathbf{1}_{x_2-x_3 \in \overline{H}_\lambda} \mathbf{1}_{x_3-x_4 \in \overline{H}_\lambda} \mathbf{1}_{x_4-x_1 \in \overline{H}_\lambda} dx_1 dx_2 dx_3 dx_4 \\ &= \int_{\varphi^{(3)}(W^4)} \mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} \mathbf{1}_{w \in \overline{H}_\lambda} \mathbf{1}_{u+v+w \in \overline{H}_\lambda} dx_1 dudvdw, \\ \ell(W^5 \cap H_\lambda^{(4)}) &= \int_{W^5} \mathbf{1}_{x_1-x_2 \in \overline{H}_\lambda} \mathbf{1}_{x_2-x_3 \in \overline{H}_\lambda} \mathbf{1}_{x_3-x_4 \in \overline{H}_\lambda} \mathbf{1}_{x_4-x_5 \in \overline{H}_\lambda} dx_1 dx_2 dx_3 dx_4 dx_5 \\ &= \int_{\varphi^{(4)}(W^5)} \mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} \mathbf{1}_{w \in \overline{H}_\lambda} \mathbf{1}_{z \in \overline{H}_\lambda} dx_1 dudv.\end{aligned}$$

Using the inclusions

$$\begin{aligned}\check{W}^2 &\subseteq \varphi^{(0)}(W^2) \subseteq \hat{W}^2, \\ \check{W}^3 &\subseteq \varphi^{(1)}(W^3) \subseteq \hat{W}^3, \\ \check{W}^5 &\subseteq \varphi^{(2)}(W^5) \subseteq \hat{W}^5, \\ \varphi^{(3)}(W^4) &\subseteq \hat{W}^4, \\ \check{W}^5 &\subseteq \varphi^{(4)}(W^5) \subseteq \hat{W}^5,\end{aligned}$$

and

$$\mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} \mathbf{1}_{w \in \overline{H}_\lambda} \mathbf{1}_{u+v+w \in \overline{H}_\lambda} \leq \mathbf{1}_{u \in \overline{H}_\lambda} \mathbf{1}_{v \in \overline{H}_\lambda} \mathbf{1}_{w \in \overline{H}_\lambda}, \quad u, v, w \in \mathbb{R}^d,$$

we have

$$\begin{aligned} \ell^2(\check{W} \times (\overline{H}_\lambda \cap \check{W})) &= 2^{-d}\check{\psi}(\lambda) \leq \lambda^{-2}V_{2,\lambda}^2 \leq \ell^2(\hat{W} \times (\overline{H}_\lambda \cap \hat{W})) = 2^d\hat{\psi}(\lambda) \\ \ell^3(\check{W} \times (\overline{H}_\lambda \cap \check{W})^2) &= 2^{-d}\check{\psi}(\lambda)^2 \leq \lambda^{-3}C_\lambda^2 \leq \ell^3(\hat{W} \times (\overline{H}_\lambda \cap \hat{W})^2) = 2^d\hat{\psi}(\lambda)^2 \\ \ell^5(\check{W} \times (\overline{H}_\lambda \cap \check{W})^4) &= 2^{-d}\check{\psi}(\lambda)^4 \leq \lambda^{-5}A_\lambda^2 \leq \ell^5(\hat{W} \times (\overline{H}_\lambda \cap \hat{W})^4) = 2^d\hat{\psi}(\lambda)^4 \\ &\lambda^{-4}B_\lambda^2 \leq \ell^4(\hat{W} \times (\overline{H}_\lambda \cap \hat{W})^3) = 2^d\hat{\psi}(\lambda)^3 \\ \ell^5(\check{W} \times (\overline{H}_\lambda \cap \check{W})^4) &= 2^{-d}\check{\psi}(\lambda)^4 \leq \lambda^{-5}E_\lambda^2 \leq \ell^5(\hat{W} \times (\overline{H}_\lambda \cap \hat{W})^4) = 2^d\hat{\psi}(\lambda)^4 \end{aligned}$$

and the result follows. □

The next statement provides one of the main results of this section: it gives an exhaustive characterization of the asymptotic behavior of  $F_\lambda$ , whenever (4.17) is in order. In order to allow for a comparison with the existing literature, we classify the asymptotic behavior of  $F_\lambda$  according to four regimes, denoted by **(R1)**–**(R4)**. Such a classification is based on the proportion  $\psi(\lambda)$  of space occupied by  $\overline{H}_\lambda$  in the observation window, determining the influence area of a given point of the Poisson measure. This coefficient has to be compared with  $\lambda^{-1}$ , which corresponds to the total window measure divided by the mean number of points. The four regimes are the following:

- (R1)**  $\lambda\psi(\lambda) \rightarrow \infty$ ;
- (R2)** (*Thermodynamic regime*)  $\lambda\psi(\lambda) \asymp 1$ ;
- (R3)**  $\lambda\psi(\lambda) \rightarrow 0$  and  $\lambda\sqrt{\psi(\lambda)} \rightarrow \infty$ ;
- (R4)** The mapping  $\lambda \mapsto \lambda\sqrt{\psi(\lambda)}$  is bounded.

The thermodynamic regime corresponds (after rescaling) to the usual models where the geometry of the interactions does not change as the window of observation grows to the whole space (see Remark 4.5). We will see in Section 4.3.1 that, when specialized to Poissonized disk graphs, our asymptotic approximations and variance estimates concur with those obtained in [28, Chapter 3]. Under regimes **(R1)** and **(R2)**, there is asymptotic normality with convergence at speed  $\lambda^{-1/2}$  in the Wasserstein distance. Under **(R3)** the convergence to the normal law is slower, and under **(R4)** the asymptotic normality is lost: for any converging subsequence the limit is either Poisson or zero.

**Remark 4.8.** One interesting contribution of Theorem 4.9 appears at the end of Point (ii), where it is stated that, under the thermodynamic regime, both chaotic projections of the random variable  $\tilde{F}_\lambda$  contribute to the limit and satisfy a joint CLT. This kind of phenomenon is an example of the “fine Gaussian fluctuations” appearing in the title of the paper.

**Theorem 4.9.** *Let  $\{\overline{H}_\lambda : \lambda > 0\}$  be a family of subsets of  $\mathbb{R}^d$  satisfying (4.9) and let  $\check{\psi}, \hat{\psi}$  be defined according to (4.15)–(4.16). Assume in addition that (4.17) is satisfied, and consider a random variable  $N \sim \mathcal{N}(0, 1)$ . The quantities introduced in Section 4.1 satisfy the following relations: there exist constants  $0 < k < K < \infty$ , independent of  $\lambda$ , such that*

$$\begin{aligned} \frac{V_{1,\lambda}^2}{V_{2,\lambda}^2} &\asymp \lambda\psi(\lambda), & \frac{1}{V_{1,\lambda}^2}A_\lambda &\asymp \lambda^{-1/2}, \quad \text{and} \\ k \max(\lambda^{-1/2}, (\lambda^2\psi(\lambda))^{-1/2}, \lambda^{1/2}\psi(\lambda)) &\leq \frac{1}{V_{2,\lambda}^2} \max(B_\lambda, C_\lambda, D_\lambda, E_\lambda) \\ &\leq K \max(\sqrt{\psi(\lambda)}, \lambda^{-1/2}, (\lambda^2\psi(\lambda))^{-1/2}, \lambda^{1/2}\psi(\lambda)). \end{aligned}$$

Furthermore, one can choose  $K$  in such a way that the following properties (i)–(iii) are verified.

(i) (Regime **(R1)**) If  $\lambda\psi(\lambda) \rightarrow \infty$ , the first chaos projection  $F_{1,\lambda}$  dominates and

$$\text{Var}(F_{1,\lambda}) \sim \text{Var}(F_\lambda) \asymp \lambda^3\psi(\lambda)^2 \rightarrow \infty,$$

and

$$d_W(\tilde{F}_\lambda, N) \leq K \lambda^{-1/2}.$$

(ii) (Regime **(R2)**) If  $\lambda\psi(\lambda) \asymp 1$ ,

$$\text{Var}(F_\lambda) \asymp \text{Var}(F_{1,\lambda}) \asymp \text{Var}(F_{2,\lambda}) \asymp \lambda,$$

and

$$d_W(\tilde{F}_\lambda, N) \leq K \lambda^{-1/2}.$$

In this case one has also that, as  $\lambda \rightarrow \infty$ , the pair

$$\left( \frac{F_{1,\lambda}}{\text{Var}(F_{1,\lambda})^{1/2}}, \frac{F_{2,\lambda}}{\text{Var}(F_{2,\lambda})^{1/2}} \right),$$

converges in distribution to a two-dimensional Gaussian vector  $(N_1, N_2)$ , such that  $N_i \sim \mathcal{N}(0, 1)$  and  $N_1, N_2$  are independent.

(iii) (Regimes **(R3)** and **(R4)**) If  $\lambda\psi(\lambda) \rightarrow 0$ , then the second chaos projection  $F_{2,\lambda}$  dominates,

$$\text{Var}(F_\lambda) \sim \text{Var}(F_{2,\lambda}) \asymp \lambda^2\psi(\lambda),$$

and

$$d_W(\tilde{F}_\lambda, N) \leq \frac{K}{\lambda\sqrt{\psi(\lambda)}}.$$

Moreover, one has that  $\tilde{F}_\lambda$  converges in distribution to  $N$  if and only if  $\lambda^2\psi(\lambda) \rightarrow \infty$  (or, equivalently,  $\text{Var}(F_\lambda) \rightarrow \infty$ ).

*Proof of Theorem 4.9.* Theorem 4.2 and assumption (4.17) yield

$$\begin{aligned} V_{2,\lambda}^2 &= 2D_\lambda^2 \asymp \lambda^2\psi(\lambda) \\ V_{1,\lambda}^2 &= C_\lambda^2 \asymp \lambda^3\psi(\lambda)^2 \\ A_\lambda^2 &\asymp \lambda^5\psi(\lambda)^4 \\ B_\lambda^2 &\leq 2^d \lambda^4 \psi(\lambda)^3 \\ E_\lambda^2 &\asymp \lambda^5\psi(\lambda)^4. \end{aligned}$$

(i) We assume  $\lambda\psi(\lambda) \rightarrow \infty$ . Applying (4.1), one deduces that

$$\begin{aligned} d_W(\tilde{F}, N) &\leq \frac{1}{V_{1,\lambda}^2} \max(A_\lambda, B_\lambda, C_\lambda, D_\lambda, E_\lambda) \\ &\leq \frac{K\sqrt{\lambda}}{\lambda^3\psi(\lambda)^2} \max(\lambda^2\psi(\lambda)^2, \lambda^{3/2}\psi(\lambda)^{3/2}, \lambda\psi(\lambda), \lambda^{1/2}\psi(\lambda)^{1/2}) \\ &\leq \frac{K}{\sqrt{\lambda}} \end{aligned}$$

for  $\lambda$  large enough. Notice that (4.5) does not yield a better bound because  $\psi(\lambda) \leq 1$ .

(ii) Applying again (4.1), the conclusion is deduced from Point (i), because  $\lambda\psi(\lambda) \geq c$  for some constant  $c > 0$  and for  $\lambda$  large enough. The last statement at Point (ii) follows from an application of Proposition 3.14.

(iii) Using (4.1) again yields

$$\begin{aligned} d_W(\tilde{F}, N) &\leq \frac{1}{\lambda^2\psi(\lambda)} \max(\lambda^{5/2}\psi(\lambda)^2, \lambda^2\psi(\lambda)^{3/2}, \lambda^{3/2}\psi, \lambda\psi(\lambda)^{1/2}) \\ &\leq \frac{K}{\lambda\sqrt{\psi(\lambda)}} \end{aligned}$$

for  $\lambda$  large enough, because  $\lambda\psi(\lambda) \rightarrow 0$ . To conclude the proof, we have to show that, if  $\lambda^2\psi(\lambda)$  does not diverge to infinity, then  $\tilde{F}_\lambda$  does not converge in distribution to  $N$ . To prove this negative result, one could apply the product formula (2.4) to prove that, whenever  $\lambda^2\psi(\lambda)$  is not diverging to infinity and is bounded away from zero, there exists a sequence  $\lambda_n, n \geq 1$ , such that  $\lambda_n \rightarrow \infty$  and  $\sup_n E[\tilde{F}_{\lambda_n}^6] < \infty$ , so that the desired conclusion is deduced from the last part of Theorem 4.2 (the case when  $\lambda^2\psi(\lambda)$  is not bounded away from zero can be dealt with by a direct argument). However, the statement of the forthcoming Theorem 4.12 is much stronger, and it is therefore not necessary to spell out the details of these computations.  $\square$

**Corollary 4.10.** *Assume that the geometric rule defined by  $\overline{G}_\lambda$  (see (4.12)) does not depend on  $\lambda$ , in such a way that  $\overline{G}_\lambda = \overline{G}$  for some fixed measurable set  $\overline{G}$ . Assuming (4.17) (see Remark 4.6), one has that  $\tilde{F}_\lambda$  converges in distribution to  $N \sim \mathcal{N}(0, 1)$ , with a rate at most of the order  $\lambda^{-1/2}$  with respect to  $d_W$ .*

*Proof.* We are in one of the following situations:

1. If  $\overline{G}$  has finite positive Lebesgue measure,  $\tilde{\psi}(\lambda) \asymp \hat{\psi}(\lambda) \asymp \ell(\overline{G})/\ell(Q_\lambda) \asymp \lambda^{-1}$ . It corresponds to the case (ii) in Theorem 4.9, meaning the two chaoses codominate. It follows that  $\tilde{F}_\lambda$  converges to the normal law with a rate at most of the order of  $\lambda^{-1/2}$  with respect to  $d_W$ .
2. If  $\overline{G}$  does not have finite measure,  $\lambda\psi(\lambda) \asymp \ell(\overline{G} \cap Q_\lambda) \rightarrow \infty$  and we are in the situation of Point (i) of Theorem 4.9, that is: the first chaos dominates. We therefore deduce that

$$d_W(\tilde{F}_\lambda, N) \leq K\lambda^{-1/2},$$

for some  $K > 0$ , and the conclusion follows.  $\square$

As announced, we shall now deal more thoroughly with the case where  $\lambda^2\psi(\lambda)$  does not diverge to infinity. In the proof of the next statement we shall use the following notation: if  $X$  is a random variable with finite moments of every order, then we write  $\{\chi_m(X) : m \geq 1\}$  to indicate the sequence of its cumulants (see [23, Chapter 3] for an introduction to this concept). For instance,  $\chi_1(X) = E[X]$ ,  $\chi_2(X) = \text{Var}(X)$ , and so on.

**Remark 4.11.** The proof of Theorem 4.12 provided below is based on diagram formulae and the method of moments and cumulants. An alternate proof could be deduced from the classic results by Silverman and Brown [35], combined with a Poissonization argument. Another proof of this result, complete with explicit bounds in the total variation distance, appears in [24]. The proof provided below has the merit of illustrating an application of diagram formulae (that are typically used to deduce CLTs) to a non-central result.

**Theorem 4.12** (Poisson approximations). *Let the assumptions and notation of Theorem 4.9 prevail, and assume in addition that  $\lambda\psi(\lambda) \rightarrow 0$  and the mapping  $\lambda \mapsto \lambda^2\psi(\lambda)$*

does not diverge to infinity, as  $\lambda \rightarrow \infty$ . Then, there exists a sequence  $\{\lambda_n\}$  such that  $\lim_{n \rightarrow \infty} \lambda_n = \infty$  and the following properties (i)-(iii) hold.

- (i)  $\lim_{n \rightarrow \infty} \text{Var}(F_{\lambda_n}) = \lim_{n \rightarrow \infty} \text{Var}(F_{2,\lambda_n}) = 2c$ , where  $c$  is a nonnegative constant possibly depending on the sequence  $\{\lambda_n\}$ , and we have used the notation (4.2).
- (ii) If  $c = 0$ , then  $E[|\tilde{F}_{\lambda_n}|] \rightarrow 0$ , as  $n \rightarrow \infty$ .
- (iii) If  $c > 0$ , then  $F_{2,\lambda_n}$ , and therefore  $F_{\lambda_n} - E[F_{\lambda_n}]$ , converges in distribution to  $X = 2P(c/2)$ , as  $n \rightarrow \infty$ , where  $P(c/2)$  indicates a centered Poisson random variable with parameter  $c/2$ .

In particular, the family  $\{\tilde{F}_\lambda\}$  does not verify a CLT as  $\lambda \rightarrow \infty$ .

*Proof.* Since  $\text{Var}(F_\lambda) \sim \text{Var}(F_{2,\lambda}) \asymp \lambda^2 \psi(\lambda)$ , Point (i) is immediately deduced from the Bolzano-Weierstrass theorem. Point (ii) follows from a direct application of Slivnyak-Mecke formula (see [31, Corollary 3.2.3]), yielding that, as  $n \rightarrow \infty$ ,

$$\frac{E[F_{\lambda_n}]}{\text{Var}(F_{\lambda_n})^{1/2}} \asymp \lambda_n \sqrt{\psi(\lambda_n)} \rightarrow 0.$$

We shall prove Point (iii) by using the method of cumulants. First of all, we observe that since  $\lambda_n^2 \psi(\lambda_n)$  is bounded and bounded away from zero, one has that  $V_{1,\lambda_n}^2 \asymp \lambda_n^3 \psi(\lambda_n)^2 \rightarrow 0$ , that is: as  $n \rightarrow \infty$ , the limits of  $F_{2,\lambda_n}$  and  $F_{\lambda_n} - E[F_{\lambda_n}]$  coincide. We recall that the law of the random variable  $X = 2P(c/2)$  is determined by its moments or, equivalently, by its cumulants (see e.g. [23, pp. 42-43]). Standard computations imply that  $\chi_1(X) = 0$  and, for every  $m \geq 2$ ,  $\chi_m(X) = 2^{m-1}c$ . We are therefore left to show that, for every  $m \geq 3$ ,

$$\chi_m(I_2(f_{2,\lambda_n})) \rightarrow 2^{m-1}c, \tag{4.18}$$

where  $f_{2,\lambda}$  is the kernel appearing in formula (4.2). Our proof of (4.18) is based on the use of the so-called ‘diagram formulae’ for multiple Poisson integrals as stated in [23, Corollary 7.4.1] (observe that this statement only deals with simple kernels: however, one can immediately extend this result to our framework by using e.g. Surgailis [36, Theorem 3.1]). Fix  $m \geq 3$ , and introduce the following notation and definitions. We shall write  $[2m] = \{1, 2, \dots, 2m\}$  and indicate by  $\pi_0$  the partition of  $[2m]$  given by  $\pi_0 = \{\{1, 2\}, \{3, 4\}, \dots, \{2m-3, 2m-2\}, \{2m-1, 2m\}\}$ . Given a partition  $\pi$  of  $[2m]$ , we write  $|\pi|$  to indicate the number of blocks of  $\pi$ . The function  $(f_{2,\lambda})_\pi$ , in  $|\pi|$  variables, is obtained as follows: (1) consider the function

$$\Phi(x_1, \dots, x_{2m}) = f_{2,\lambda}(x_1, x_2) \times f_{2,\lambda}(x_3, x_4) \times \dots \times f_{2,\lambda}(x_{2m-1}, x_{2m}),$$

given by the juxtaposition of  $m$  copies of  $f_{2,\lambda}$ , and (2) identify two variables  $x_i, x_j$  in the argument of  $\Phi$  if and only if  $i$  and  $j$  are in the same block of  $\pi$ . According to [23, Corollary 7.4.1], one has therefore that

$$\chi_m(I_2(f_{2,\lambda_n})) = \sum_{\pi \in \mathcal{M}_m} \lambda_n^{|\pi|} \int_{(\mathbb{R}^d)^{|\pi|}} (f_{2,\lambda_n})_\pi d\ell^{|\pi|},$$

where the symbol  $\mathcal{M}_m$  stands for the class of those partitions  $\pi$  of  $[2m]$  satisfying the following properties: (a) every block of  $\pi$  contains at least two elements, (b) given any two blocks  $b_0 \in \pi_0$  and  $b_1 \in \pi$ , the intersection  $b_0 \cap b_1$  contains at most one element, and (c) the diagram  $\Gamma(\pi_0, \pi)$ , as defined in [23, Section 4.1], is connected in the sense of [23, p. 47]. There are exactly  $2^{m-1}$  partitions  $\pi \in \mathcal{M}_m$  such that  $|\pi| = 2$ , and for any such partition one has that

$$\lambda_n^{|\pi|} \int_{(\mathbb{R}^d)^{|\pi|}} (f_{2,\lambda_n})_\pi d\ell^{|\pi|} = \|f_{2,\lambda_n}\|_{L^2((\lambda_n \ell)^2)}^2 \rightarrow c.$$

On the other hand, if  $\pi \in \mathcal{M}_m$  and  $|\pi| \geq 3$ , a change of variables analogous to the ones defined in the proof of Theorem 4.7 yields that, for some constant  $C$  independent of  $n$ ,

$$0 \leq \lambda_n^{|\pi|} \int_{(\mathbb{R}^d)^{|\pi|}} (f_{2,\lambda_n})_\pi d\ell^{|\pi|} \leq C \lambda_n^{|\pi|} \psi(\lambda_n)^{|\pi|-1} = C \lambda_n^2 \psi(\lambda_n) \times (\lambda_n \psi(\lambda_n))^{|\pi|-2} \rightarrow 0,$$

thus concluding the proof. □

### 4.3.1 Two examples

We now present some explicit examples. The notation of Section 4.3 will prevail throughout this section.

**Example 4.13** (Disk graph). Assume

$$\bar{G}_\lambda = B(0, r_\lambda)$$

for some  $r_\lambda > 0$ , meaning that two points of  $\eta$  in  $Q_\lambda$  are connected whenever their distance is smaller than  $r_\lambda$ . It yields  $\psi(\lambda) \asymp r_\lambda^d/\lambda$  (it is easy to verify that (4.17) is satisfied). Then  $\tilde{F}_\lambda$  is asymptotically normal iff  $\lambda r_\lambda^d \rightarrow \infty$ , and

$$d_W(\tilde{F}_\lambda, N) \leq C \lambda^{-1/2} \max(1, r_\lambda^{-d/2}).$$

According to the classification based on the four regimes **(R1)**–**(R4)**, the above result yields the following exhaustive description of the asymptotic behavior of  $F_\lambda$  (note how we are able to distinguish the contribution of each chaotic projection) :

- (R1)** If  $r_\lambda \rightarrow \infty$ , then  $\text{Var}(F_\lambda) \asymp \lambda(r_\lambda^d)^2$ ,  $\tilde{F}_\lambda$  satisfies a CLT with an upper bound of the order of  $\lambda^{-1/2}$  on the Wasserstein distance, and the projection of  $\tilde{F}_\lambda$  on the first Wiener chaos dominates.
- (R2)** If  $r_\lambda \asymp 1$ , then  $\text{Var}(F_\lambda) \asymp \lambda$ ,  $\tilde{F}_\lambda$  satisfies a CLT with an upper bound of the order of  $\lambda^{-1/2}$  on the Wasserstein distance, and the projections of  $\tilde{F}_\lambda$  on the first and second Wiener chaos both contribute to the limit and satisfy a joint CLT.
- (R3)** If  $r_\lambda \rightarrow 0$  and  $\lambda r_\lambda^d \rightarrow \infty$ , then  $\text{Var}(F_\lambda) \asymp \lambda r_\lambda^d$ ,  $\tilde{F}_\lambda$  satisfies a CLT with an upper bound of the order of  $(\lambda r_\lambda^d)^{-1/2}$  on the Wasserstein distance, and the projection of  $\tilde{F}_\lambda$  on the second Wiener chaos dominates in the limit.
- (R4)** If  $\lambda r_\lambda^d$  converges to a finite constant  $c \geq 0$ , then  $F_\lambda$  converges either to zero (if  $c = 0$ ) or to a multiple of a Poisson random variable (if  $c > 0$ ).

**Remark 4.14.** Explicit estimates on the variances and on the rates of convergence to normal in the Wasserstein distance for edge counting statistics are also provided in [30, Theorem 6.3]. Their variance estimates are of the order

$$\max(\lambda^3 \delta_\lambda^{2d}, \lambda^2 \delta_\lambda^d) = \max(\lambda^3 \psi(\lambda)^2, \lambda^2 \psi(\lambda)),$$

corresponding exactly to our findings ( $\delta_\lambda = \lambda^{-1/d} r_\lambda = \psi(\lambda)^{1/d}$  in their notation). They obtained convergence to the normal law if  $\delta_\lambda \rightarrow 0$  and  $\lambda^{4/3} \delta_\lambda^d \rightarrow \infty$ , whereas Example 4.13 yields a CLT whenever  $\lambda^2 \psi(\lambda) \rightarrow \infty$ , dealing with the cases  $\limsup_\lambda \delta_\lambda > 0$  and  $\lambda^{-2/d} = o(\delta_\lambda)$ ,  $\delta_\lambda \leq \lambda^{-4/(3d)}$ , and gives a negative answer if  $\delta_\lambda \leq C \lambda^{-2/d}$  for some  $C > 0$ .

**Remark 4.15.** For every fixed  $\lambda$ , the  $U$ -statistic  $\frac{1}{2} F_\lambda$  has the same law as the random variable counting the number of edges in a disk graph, with radius  $\delta_\lambda = \lambda^{-1/d} r_\lambda$ , based on random points of the form  $\{Y_1, \dots, Y_{N(\lambda)}\}$ , where  $\{Y_i\}$  indicates a collection of i.i.d. random variables uniformly distributed on  $Q_1 = [-\frac{1}{2}, \frac{1}{2}]^d$ , and  $N(\lambda)$  is an independent Poisson random variable with parameter  $\lambda$ . As such, each  $\frac{1}{2} F_\lambda$  is just a subgraph

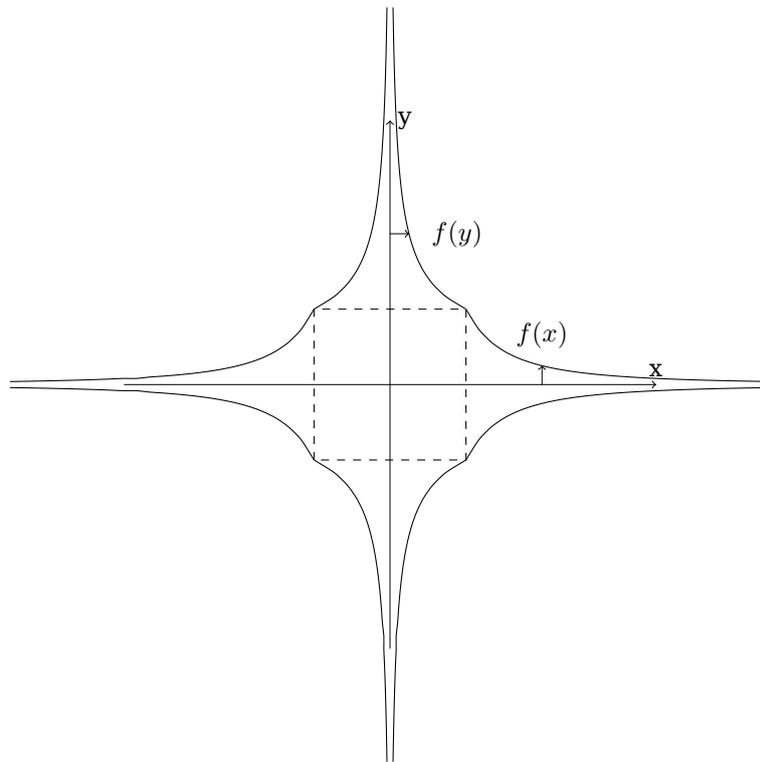


Figure 1: General form of  $G_\lambda = \{(x, y) : |y| < |f(x)| \text{ and } |x| < |f(y)|\}$ .

counting statistic based on a Poissonized random geometric graph, and enters the general framework of [28, Section 3.4], where general  $m$ -dimensional CLTs are obtained for these objects. It is immediately checked that our variance estimates coincide with those stated in [28, p. 56] (for the case  $k = 2$ ), whereas our estimates in the Wasserstein distance refine the findings of [28, Theorems 3.9 and 3.10] (in the case  $k = 2$  and  $m = 1$ ), where no information on the rate of convergence is given. Previous references for CLTs for Poissonized disk graphs are [2, 9], where no explicit rates of convergence are provided either. A generalization of the previously described findings to general subgraph counting in a disk graph model can be found in [10, Section 3].

**Example 4.16.** We present here examples of geometric stationary graphs with arbitrarily long connections and behaviors spanning the whole spectrum of possibilities indicated in Theorem 4.9. In all cases, condition (4.17) is easily checked. We define  $\bar{G} \subseteq \mathbb{R}^2$  as in Fig. 1, symmetric with respect to all axes, and consider different choices for  $f$ , a monotone function  $\mathbb{R}_+^* \mapsto \mathbb{R}_+^*$ . We set  $G_\lambda = \alpha_\lambda \bar{G}$  with different values for  $\alpha_\lambda > 0$ . We refer the reader to Fig. 2 for results of simulations, where different values for  $\lambda$  have been used.

- (a) If  $f(x) = 1/x$ ,  $\alpha_\lambda = 1$ , then  $\lambda\psi(\lambda) \asymp \log(\lambda)$ , whence there is a CLT with an upper bound of the order of  $\lambda^{-1/2}$  (Situation (i) of Theorem 4.9), see Fig. 2a.
- (b) If  $f(x) = 1/x^2$ ,  $\alpha_\lambda = 1$ , then  $\lambda\psi(\lambda) \asymp 1$ , whence there is a CLT with an upper bound of the order of  $\lambda^{-1/2}$  (Situation (ii) of Theorem 4.9), see Fig. 2b.
- (c) If  $f(x) = 1/x^2$ ,  $\alpha_\lambda = \lambda^{-1/4}$ , then  $\lambda\psi(\lambda) \asymp \lambda^{-1/2}$ , whence there is a CLT with an upper bound of the order of  $\lambda^{-1/4}$  (Situation (iii) of Theorem 4.9), see Fig. 2c.

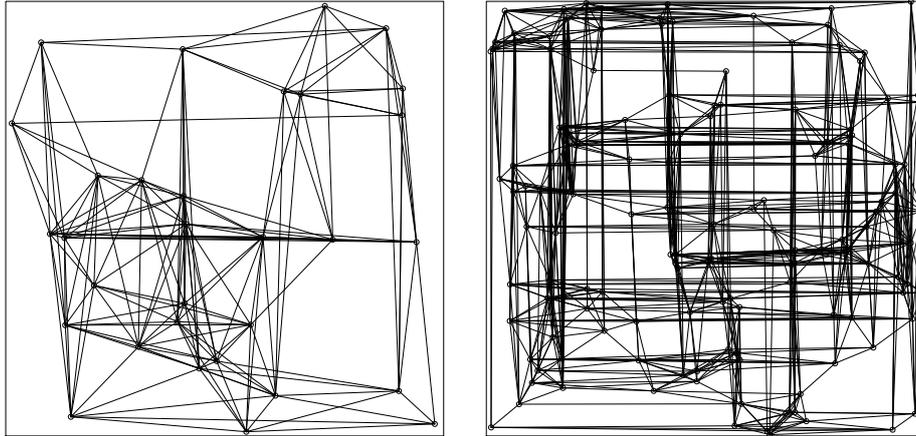


Figure 2a:  $\lambda = 25, \lambda = 150$

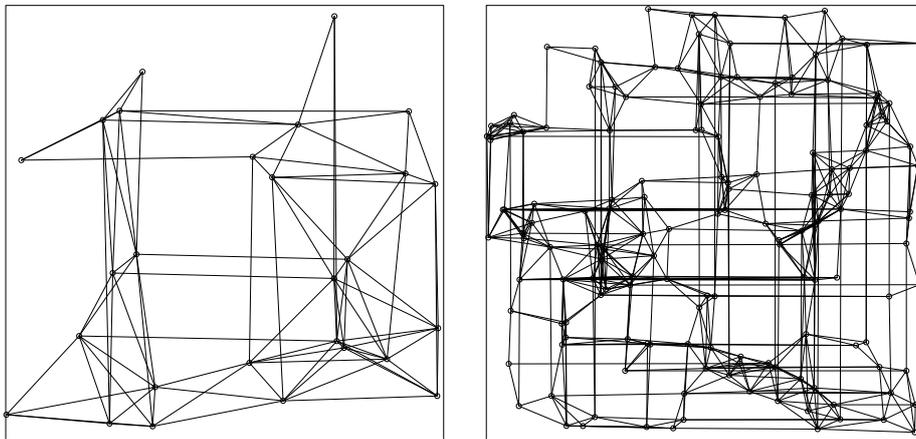


Figure 2b:  $\lambda = 25, \lambda = 150$

- (d)** If  $f(x) = 1/x$ ,  $\alpha_\lambda = \lambda^{-1/2}$ , then  $\lambda\psi(\lambda) \asymp \log(\lambda)\lambda^{-1}$ , whence there is a CLT with an upper bound of the order of  $\log(\lambda)^{-1}$  (Situation (iii) of Theorem 4.9), see Fig. 2d.
- (e)**  $f(x) = 1/x^2, \alpha_\lambda = \lambda^{-1/2}$ . According to Theorem 4.12,  $\tilde{F}_\lambda$  converges in distribution to a multiple of a Poisson random variable, see Fig. 2e.

In all cases, the convergence to a normal law goes hand in hand with the almost sure convergence of the number of connections to infinity, and with the convergence of the variance to infinity. In the case **(d)**, the convergence is very slow, the number of connections behaves asymptotically like a Poisson law with parameter  $\log(\lambda)$ , due to the long-range connections within the point process. In the case **(e)**, the asymptotic properties of  $G_\lambda$  do not yield long range connections and the number of connections converges towards a Poisson-type limit.

Fine Gaussian fluctuations on the Poisson space, I

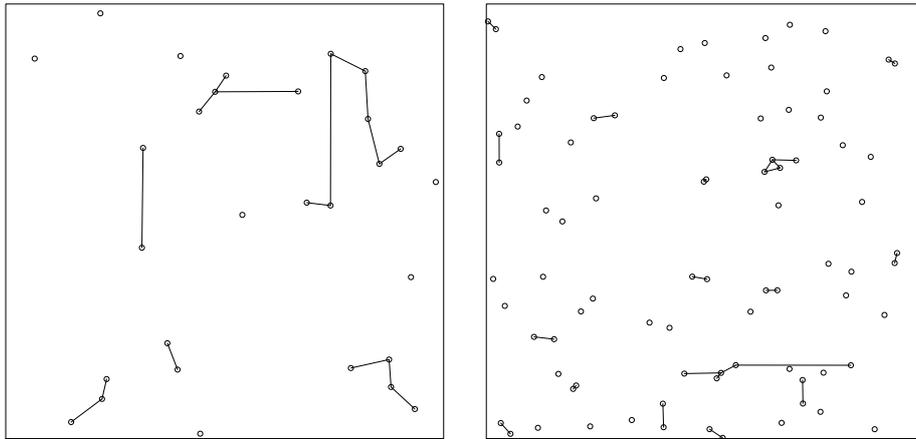


Figure 2c:  $\lambda = 25, \lambda = 75$

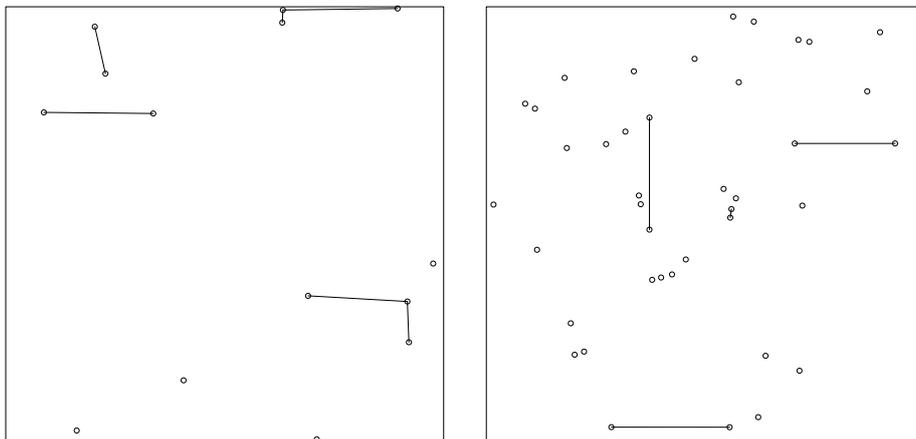


Figure 2d:  $\lambda = 10, \lambda = 50$

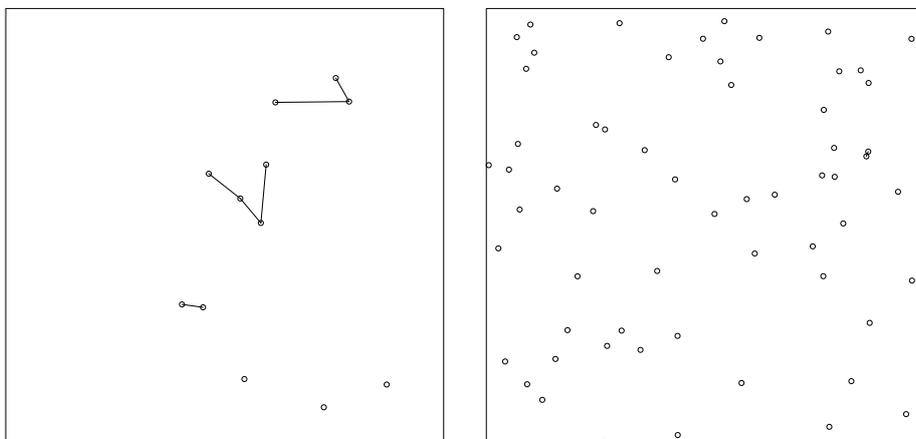


Figure 2e:  $\lambda = 10, \lambda = 50$

## References

- [1] Y. Barishnikov and M. Penrose (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Prob.* **15** (1A), 213-253. MR-2115042
- [2] R. N. Bhattacharya and J. K. Ghosh (1992). A class of  $U$ -statistics and asymptotic normality of the number of  $k$ -clusters. *J. Multiv. Analysis* **43**, 300-330. MR-1193616
- [3] N.H. Bingham, C.M. Goldie and J.L. Teugels (1989). *Regular variation*. Encyclopedia of mathematics and its applications, Vol. 27, *Regular variation*, Cambridge. MR-1015093
- [4] R. Blei and S. Janson (2004). Rademacher chaos: tail estimates versus limit theorems. *Ark. Mat.* **42**(1), 13-29. MR-2056543
- [5] B. N. Clark, C. J. Colbourn and D.S. Johnson (1990). Unit disk graphs. *Discrete Mathematics*, **86**(1-3), 165-177. MR-1088569
- [6] L. Decreusefond, E. Ferraz and H. Randriam (2011). Simplicial Homology of Random Configurations. Preprint. MR-2842939
- [7] E. Ferraz and A. Vergne (2011). Statistics of geometric random simplicial complexes. Preprint.
- [8] L. Heinrich, H. Schmidt and V. Schmidt (2006). Central limit theorems for Poisson hyperplane tessellations. *Ann. App. Probab.*, **16**(2), 919-950. MR-2244437
- [9] S.R. Jammalamadaka and S. Janson (1986). Limit theorems for a triangular scheme of  $U$ -statistics with applications to inter-point distances. *Ann. Probab.* **14**(4), 1347-1358. MR-0866355
- [10] R. Lachièze-Rey and G. Peccati (2012). Fine Gaussian fluctuations on the Poisson space II: rescaled kernels, marked processes and geometric  $U$ -statistics. Preprint available at the address arXiv:1205.0632v2. MR-2914431
- [11] G. Last and M. Penrose (2010). Poisson process Fock space representation, chaos expansion and covariance inequalities. *Probab. Theory Rel. Fields.* **150** (3-4), 663-690. MR-2824870
- [12] G. Last, M. D. Penrose, M. Schulte and C. Thaele (2012). Moments and central limit theorems for some multivariate Poisson functionals. Preprint.
- [13] W. S. Kendall and I. Molchanov (2009). *New perspectives in stochastic geometry*, Oxford University Press. MR-2668353
- [14] C. McDiarmid (2003). Random channel assignment in the plane. *Random Structure and Algorithms*, **22**, 187-212. MR-1954610
- [15] N.T. Minh (2011). Malliavin-Stein method for multi-dimensional  $U$ -statistics of Poisson point processes. Preprint.
- [16] T. Müller (2008). Two-point concentration in random geometric graphs. *Combinatorics* **28**, 529-545. MR-2501248
- [17] I. Nourdin and G. Peccati (2009). Stein's method on Wiener chaos. *Probab. Theory Relat. Fields* **145**(1), 75-118. MR-2520122
- [18] I. Nourdin and G. Peccati (2012). *Normal approximations using Malliavin calculus: from Stein's method to universality*. Cambridge University Press. MR-2962301
- [19] I. Nourdin, G. Peccati and G. Reinert (2010). Stein's method and stochastic analysis of Rademacher sequences. *The Electronic Journal of Probability* **15**, 1703-1742 (Electronic). MR-2735379
- [20] D. Nualart and G. Peccati (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, **33**(1), 177-193. MR-2118863
- [21] D. Nualart and J. Vives (1990). Anticipative calculus for the Poisson process based on the Fock space. In *Sem. de Proba. WWIV, LNM 1426*, pp. 154-165. Springer-Verlag. MR-1071538
- [22] G. Peccati and M.S. Taqqu (2008). Central limit theorems for double Poisson integrals. *Bernoulli* **14**(3), 791-821. MR-2537812
- [23] G. Peccati and M.S. Taqqu (2010). *Wiener chaos: moments, cumulants and diagrams*. Springer-Verlag. MR-2791919
- [24] G. Peccati (2011). The Chen-Stein method for Poisson functionals. Preprint.

- [25] G. Peccati, J.-L. Solé, M.S. Taqqu and F. Utzet (2010). Stein’s method and normal approximation of Poisson functionals. *Ann. Probab.* **38**(2), 443-478. MR-2642882
- [26] G. Peccati and C. Zheng (2010). Multi-dimensional Gaussian fluctuations on the Poisson space. *Elect. J. Probab.* **15**(48), 1487-1527. MR-2727319
- [27] G. Peccati and C. Zheng (2011). Universal Gaussian fluctuations on the discrete Poisson chaos. To appear in *Bernoulli*.
- [28] M. Penrose (2003). *Random Geometric Graphs*. Oxford University Press. MR-1986198
- [29] M. Penrose and J. Yukich (2005). Normal approximation in geometric probability. In: *Stein’s Method and Applications*, eds. A.D. Barbour and L.H.Y. Chen, World Scientific, pp. 37-58. MR-2201885
- [30] M. Reitzner and M. Schulte (2011). Central Limit Theorems for U-Statistics of Poisson Point Processes. To appear in: *Ann. Probab.*
- [31] R. Schneider and W. Weil (2008). *Stochastic and integral geometry*. Springer-Verlag. MR-2455326
- [32] M. Schulte (2011). A Central Limit Theorem for the Poisson-Voronoi Approximation. *Advances in Applied Mathematics* **49**, 285-306.
- [33] M. Schulte and C. Thaele (2010). Exact and asymptotic results for intrinsic volumes of Poisson  $k$ -flat processes. Preprint.
- [34] M. Schulte and C. Thaele (2012). The scaling limit of Poisson-driven order statistics with applications in geometric probability. *Stochastic Processes and their Applications* **122**, 4096-4120. MR-2971726
- [35] B. Silverman and T. Brown (1978). Short Distances, Flat Triangles and Poisson Limits. *Journal of Applied Probability* **15**(4), 815-825. MR-0511059
- [36] D. Surgailis (1984). On multiple Poisson stochastic integrals and associated Markov semi-groups. *Probab. Math. Statist.* **3**(2), 217-239. MR-0764148

## 5 Appendix: Malliavin operators on the Poisson space

We now define some Malliavin-type operators associated with a Poisson measure  $\eta$ , on the Borel space  $(Z, \mathcal{Z})$ , with non-atomic control measure  $\mu$ . We follow the work by Nualart and Vives [21].

### The derivative operator $D$ .

For every  $F \in L^2(P)$ , the derivative of  $F$ ,  $DF$  is defined as an element of  $L^2(P; L^2(\mu))$ , that is, of the space of the jointly measurable random functions  $u : \Omega \times Z \mapsto \mathbb{R}$  such that  $E \left[ \int_Z u_z^2 \mu(dz) \right] < \infty$ .

**Definition 5.1.** 1. The domain of the derivative operator  $D$ , written  $\text{dom}D$ , is the set of all random variables  $F \in L^2(P)$  admitting a chaotic decomposition (2.1) such that

$$\sum_{k \geq 1} k k! \|f_k\|_{L^2(\mu^k)}^2 < \infty,$$

2. For any  $F \in \text{dom}D$ , the random function  $z \mapsto D_z F$  is defined by

$$D_z F = \sum_{k \geq 1} k I_{k-1}(f_k(z, \cdot)).$$

### The divergence operator $\delta$ .

Thanks to the chaotic representation property of  $\eta$ , every random function  $u \in L^2(P, L^2(\mu))$  admits a unique representation of the type

$$u_z = \sum_{k \geq 0} I_k(f_k(z, \cdot)), \quad z \in Z, \tag{5.1}$$

where the kernel  $f_k$  is a function of  $k + 1$  variables, and  $f_k(z, \cdot)$  is an element of  $L^2_s(\mu^k)$ . The divergence operator  $\delta(u)$  maps a random function  $u$  in its domain to an element of  $L^2(P)$ .

**Definition 5.2.** 1. The domain of the divergence operator, denoted by  $\text{dom}\delta$ , is the collection of all  $u \in L^2(P, L^2(\mu))$  having the above chaotic expansion (5.1) satisfied the condition:

$$\sum_{k \geq 0} (k + 1)! \|f_k\|_{L^2(\mu^{(k+1)})}^2 < \infty.$$

2. For  $u \in \text{dom}\delta$ , the random variable  $\delta(u)$  is given by

$$\delta(u) = \sum_{k \geq 0} I_{k+1}(\tilde{f}_k),$$

where  $\tilde{f}_k$  is the canonical symmetrization of the  $k + 1$  variables function  $f_k$ .

As made clear in the following statement, the operator  $\delta$  is indeed the adjoint operator of  $D$ .

**Lemma 5.3** (Integration by parts). For every  $G \in \text{dom}D$  and  $u \in \text{dom}\delta$ , one has that

$$E[G\delta(u)] = E[\langle DG, u \rangle_{L^2(\mu)}].$$

The proof of Lemma 5.3 is detailed e.g. in [21].

### The Ornstein-Uhlenbeck generator $L$ .

**Definition 5.4.** 1. The domain of the Ornstein-Uhlenbeck generator, denoted by  $\text{dom}L$ , is the collection of all  $F \in L^2(P)$  whose chaotic representation verifies the condition:

$$\sum_{k \geq 1} k^2 k! \|f_k\|_{L^2(\mu^k)}^2 < \infty$$

2. The Ornstein-Uhlenbeck generator  $L$  acts on random variable  $F \in \text{dom}L$  as follows:

$$LF = - \sum_{k \geq 1} k I_k(f_k).$$

### The pseudo-inverse of $L$ .

**Definition 5.5.** 1. The domain of the pseudo-inverse of the Ornstein-Uhlenbeck generator, denoted by  $L^{-1}$ , is the space  $L^2_0(P)$  of centered random variables in  $L^2(P)$ .

2. For  $F = \sum_{k \geq 1} I_k(f_k) \in L^2_0(P)$ , we set

$$L^{-1}F = - \sum_{k \geq 1} \frac{1}{k} I_k(f_k).$$