# Stabilization via semigroup interpolations

Giovanni Peccati (Luxembourg University)

UT Austin: May 5, 2014

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

 Mainly based on joint work with G. Last and M. Schulte (Karlsruhe).

- Mainly based on joint work with G. Last and M. Schulte (Karlsruhe).
- One of the latest instalments in a rich line of research, focussing on probabilistic approximations via the use of infinite-dimensional integration by parts formulae. Landmark contributions: Peccati, Solé, Utzet and Taqqu (2010), Reitzner and Schulte (2012), Hug, Last and Schulte (2013), Eichelsbacher and Thäle (2013).

- Mainly based on joint work with G. Last and M. Schulte (Karlsruhe).
- One of the latest instalments in a rich line of research, focussing on probabilistic approximations via the use of infinite-dimensional integration by parts formulae. Landmark contributions: Peccati, Solé, Utzet and Taqqu (2010), Reitzner and Schulte (2012), Hug, Last and Schulte (2013), Eichelsbacher and Thäle (2013).
- A parallel (and richer) theory exists on Gaussian spaces see the monograph by Nourdin and Peccati (2012).

For every  $t \ge 1$ ,  $\eta_t$  is a Poisson measure on  $\mathbb{R}^d$  ( $d \ge 1$ ), with intensity  $t \times$  Lebesgue.

- For every t ≥ 1, η<sub>t</sub> is a Poisson measure on ℝ<sup>d</sup> (d ≥ 1), with intensity t × Lebesgue.
- We denote by  $F_t = F_t(\eta_t)$  a generic centered and square-integrable functional of  $\eta_t$ , write  $v(t) = \text{Var } F_t$ , and

$$\widetilde{F}_t = v(t)^{-1/2} F_t, \quad t > 0.$$

- For every  $t \ge 1$ ,  $\eta_t$  is a Poisson measure on  $\mathbb{R}^d$  ( $d \ge 1$ ), with intensity  $t \times$  Lebesgue.
- We denote by  $F_t = F_t(\eta_t)$  a generic centered and square-integrable functional of  $\eta_t$ , write  $v(t) = \text{Var } F_t$ , and

$$\widetilde{F}_t = v(t)^{-1/2}F_t, \quad t > 0.$$

Assuming that  $v(t) \ge \sigma t$ , as  $t \to \infty$ , we want to deduce "optimal" bounds of the type

$$d_{Kol}(\widetilde{F}_t, N) := \sup_{z \in \mathbb{R}} \left| \mathbb{P}(\widetilde{F}_t \leq z) - \mathbb{P}(N \leq z) \right| \leq C t^{-1/2},$$

where  $N \sim \mathcal{N}(0, 1)$ .

- For every  $t \ge 1$ ,  $\eta_t$  is a Poisson measure on  $\mathbb{R}^d$  ( $d \ge 1$ ), with intensity  $t \times$  Lebesgue.
- We denote by  $F_t = F_t(\eta_t)$  a generic centered and square-integrable functional of  $\eta_t$ , write  $v(t) = \text{Var } F_t$ , and

$$\widetilde{F}_t = v(t)^{-1/2}F_t, \quad t > 0.$$

Assuming that  $v(t) \ge \sigma t$ , as  $t \to \infty$ , we want to deduce "optimal" bounds of the type

$$d_{\mathcal{K}ol}(\widetilde{F}_t, \mathcal{N}) := \sup_{z \in \mathbb{R}} \left| \mathbb{P}(\widetilde{F}_t \leq z) - \mathbb{P}(\mathcal{N} \leq z) \right| \leq C t^{-1/2},$$

where  $N \sim \mathcal{N}(0, 1)$ .

In a non-dynamic setting, we shall write  $\eta = \eta_1$ ,  $F = F_1$ , ... and so on. Chaos

■ Recall that every *F* ∈ *L*<sup>2</sup>(σ(η)) admits a unique chaotic decomposition of the type

$$F = \mathbb{E}(F) + \sum_{n \geq 1} I_n(f_n),$$

where

$$I_n(f_n) = \int \cdots \int f_n(x_1, ..., x_n) \mathbf{1}_{\{no \ diagonals\}} \hat{\eta}(dx_1) \cdots \hat{\eta}(dx_n)$$

stands for a **multiple Wiener** -Itô integral with respect to the compensation  $\hat{\eta} = \eta$  – Leb.

Chaos

■ Recall that every F ∈ L<sup>2</sup>(σ(η)) admits a unique chaotic decomposition of the type

$$F = \mathbb{E}(F) + \sum_{n \geq 1} I_n(f_n),$$

where

$$I_n(f_n) = \int \cdots \int f_n(x_1, ..., x_n) \mathbf{1}_{\{no \ diagonals\}} \hat{\eta}(dx_1) \cdots \hat{\eta}(dx_n)$$

stands for a **multiple Wiener** -Itô integral with respect to the compensation  $\hat{\eta} = \eta$  – Leb.

This decomposition will play an important role, mostly in the background. See Günter's talk and my mini course for more details.

# Type of geometric variables

One has two kind of variables: the techniques may differ very much when passing from one class to the other.

 Random variables having a finite chaotic expansion. By virtue of a result by Reitzner and Schulte (2012), these variables are basically finite linear combinations of *U*-statistics; examples include subgraph counting or total length statistics in the Gilbert graph.

# Type of geometric variables

One has two kind of variables: the techniques may differ very much when passing from one class to the other.

- Random variables having a finite chaotic expansion. By virtue of a result by Reitzner and Schulte (2012), these variables are basically finite linear combinations of *U*-statistics; examples include subgraph counting or total length statistics in the Gilbert graph.
- (2) Random variables having an infinite expansion. Examples include subgraph counting and total length statistics in the *k*-nearest neighbour graph, intrinsic volumes of Poisson-Voronoi tessellations and Boolean models (see Günter's talk).

# Type of geometric variables

One has two kind of variables: the techniques may differ very much when passing from one class to the other.

- Random variables having a finite chaotic expansion. By virtue of a result by Reitzner and Schulte (2012), these variables are basically finite linear combinations of *U*-statistics; examples include subgraph counting or total length statistics in the Gilbert graph.
- (2) Random variables having an infinite expansion. Examples include subgraph counting and total length statistics in the *k*-nearest neighbour graph, intrinsic volumes of Poisson-Voronoi tessellations and Boolean models (see Günter's talk).

In this talk, we are interested in random variables of the type (2) for which the chaotic decomposition is not easily to amenable to analysis. Our idea for dealing with this situation is to suitably extend the concept of a **second-order Poincaré inequality**.

#### Gaussian framework

■ Recall the usual **Poincaré-Chernoff-Nash inequality**: for a *d*-dimensional standard Gaussian vector  $X = (X_1, ..., X_d)$  and for every smooth mapping  $f : \mathbb{R}^d \to \mathbb{R}$ ,

 $\operatorname{Var} f(X) \leq \mathbb{E}[\|\nabla f(X)\|_{\mathbb{R}^d}^2].$ 

#### Gaussian framework

■ Recall the usual **Poincaré-Chernoff-Nash inequality**: for a *d*-dimensional standard Gaussian vector  $X = (X_1, ..., X_d)$  and for every smooth mapping  $f : \mathbb{R}^d \to \mathbb{R}$ ,

 $\operatorname{Var} f(X) \leq \mathbb{E}[\|\nabla f(X)\|_{\mathbb{R}^d}^2].$ 

The first example of a second order Poincaré estimate appears in Chatterjee (2007): for *f* and *X* as above,

 $d_{TV}(f(X), N) \leq C \mathbb{E}[\|\operatorname{Hess} f(X)\|_{op}^4]^{1/4} \times \mathbb{E}[\|\nabla f(X)\|_{\mathbb{R}^d}^4]^{1/4}$ 

## Gaussian framework

■ Recall the usual **Poincaré-Chernoff-Nash inequality**: for a *d*-dimensional standard Gaussian vector  $X = (X_1, ..., X_d)$  and for every smooth mapping  $f : \mathbb{R}^d \to \mathbb{R}$ ,

 $\operatorname{Var} f(X) \leq \mathbb{E}[\|\nabla f(X)\|_{\mathbb{R}^d}^2].$ 

The first example of a second order Poincaré estimate appears in Chatterjee (2007): for *f* and *X* as above,  $d_{TV}(f(X), N) \leq C \mathbb{E}[\|\text{Hess}f(X)\|_{oo}^{4}]^{1/4} \times \mathbb{E}[\|\nabla f(X)\|_{wd}^{4}]^{1/4}$ 

In Nourdin, Peccati and Reinert (2010): extension to functionals F of a general Gaussian field X,

 $d_{TV}(F, N) \leq C \mathbb{E}[\|D^2 F\|_{op}^4]^{1/4} \times \mathbb{E}[\|DF\|^4]^{1/4},$ 

where *D* stands for the Malliavin derivative.

For a functional *F* of η and x ∈ ℝ<sup>d</sup>, define D<sub>x</sub>F(η) = F(η + δ<sub>x</sub>) − F(η) (add-one cost operator). We shall build on the following Poincaré inequality: for every F ∈ L<sup>2</sup>(σ(η)),

$$\mathbf{Var} \mathbf{F} \leq \mathbb{E} \left\{ \int_{\mathbb{R}^d} (D_x \mathbf{F})^2 dx \right\}.$$

For a functional *F* of η and x ∈ ℝ<sup>d</sup>, define
 D<sub>x</sub>F(η) = F(η + δ<sub>x</sub>) − F(η) (add-one cost operator). We shall build on the following Poincaré inequality: for every F ∈ L<sup>2</sup>(σ(η)),

$$\operatorname{Var} F \leq \mathbb{E} \left\{ \int_{\mathbb{R}^d} (D_x F)^2 dx 
ight\}.$$

Note that we are looking for optimal rates, and that the estimates on the Gaussian space typically yield suboptimal results.

For a functional *F* of η and x ∈ ℝ<sup>d</sup>, define
 D<sub>x</sub>F(η) = F(η + δ<sub>x</sub>) − F(η) (add-one cost operator). We shall build on the following Poincaré inequality: for every F ∈ L<sup>2</sup>(σ(η)),

$$\operatorname{Var} F \leq \mathbb{E} \left\{ \int_{\mathbb{R}^d} (D_x F)^2 dx 
ight\}.$$

- Note that we are looking for optimal rates, and that the estimates on the Gaussian space typically yield suboptimal results.
- In the Poisson framework, it is much easier to work with the Wasserstein distance  $d_W$ ; however, the usual bound  $d_{Kol} \leq 2\sqrt{d_W}$  would yield suboptimal bounds.

For a functional *F* of η and x ∈ ℝ<sup>d</sup>, define
 D<sub>x</sub>F(η) = F(η + δ<sub>x</sub>) − F(η) (add-one cost operator). We shall build on the following Poincaré inequality: for every F ∈ L<sup>2</sup>(σ(η)),

$$\operatorname{Var} F \leq \mathbb{E} \left\{ \int_{\mathbb{R}^d} (D_x F)^2 dx 
ight\}.$$

- Note that we are looking for optimal rates, and that the estimates on the Gaussian space typically yield suboptimal results.
- In the Poisson framework, it is much easier to work with the Wasserstein distance  $d_W$ ; however, the usual bound  $d_{Kol} \leq 2\sqrt{d_W}$  would yield suboptimal bounds.
- One additional difficulty in the Poisson setting is that linear functionals of a Poisson measure are in general very far from being Gaussian.

# Main ingredient: The Ornstein-Uhlenbeck semigroup in Mehler's form

For every  $s \ge 0$ , define  $\eta^{(s)}$  to be a  $e^{-s}$ -thinning of  $\eta$ , and let  $\hat{\eta}^{(s)}$  be an independent Poisson measure with intensity  $(1 - e^{-s}) \times$  Lebesgue. The collection of operators  $\{T_s : s \ge 0\}$  given by

$$T_{s}F(\eta) := \mathbb{E}\left[F(\eta^{(s)} + \hat{\eta}^{(s)}) \,|\, \eta\right]$$

is the Ornstein-Uhlenbeck semigroup.

# Main ingredient: The Ornstein-Uhlenbeck semigroup in Mehler's form

For every  $s \ge 0$ , define  $\eta^{(s)}$  to be a  $e^{-s}$ -thinning of  $\eta$ , and let  $\hat{\eta}^{(s)}$  be an independent Poisson measure with intensity  $(1 - e^{-s}) \times$  Lebesgue. The collection of operators  $\{T_s : s \ge 0\}$  given by

$$T_{s}F(\eta) := \mathbb{E}\left[F(\eta^{(s)} + \hat{\eta}^{(s)}) \,|\, \eta\right]$$

is the Ornstein-Uhlenbeck semigroup.

It is sometimes convenient to work with  $P_t := T_{\log 1/t}$ ,  $t \in [0, 1]$ , so that  $P_0F = \mathbb{E}(F)$  and  $P_1F = F$ .

#### Some remarkable relations

(Integration by parts) Consider the restriction of D to the space

dom 
$$D := \left\{ \varphi : \mathbb{E} \left[ \int \varphi(x)^2 dx \right] < \infty \right\},$$

as well as its adjoint  $\delta$ . Then, for  $\varphi \in \operatorname{dom} \delta$ 

$$\mathbb{E}[\delta(\varphi)F] = \mathbb{E}\int \varphi(x)D_xF\,dx.$$

#### Some remarkable relations

(Integration by parts) Consider the restriction of D to the space

dom 
$$D := \left\{ \varphi : \mathbb{E} \left[ \int \varphi(x)^2 dx \right] < \infty \right\},$$

as well as its adjoint  $\delta$ . Then, for  $\varphi \in \operatorname{dom} \delta$ 

$$\mathbb{E}[\delta(\varphi)F] = \mathbb{E}\int \varphi(x)D_xF\,dx.$$

• Let *L* be the generator of  $\{T_s\}$ , then  $L = -\delta D$ .

#### Some remarkable relations

(Integration by parts) Consider the restriction of D to the space

dom 
$$D := \left\{ \varphi : \mathbb{E} \left[ \int \varphi(x)^2 dx \right] < \infty \right\},$$

as well as its adjoint  $\delta$ . Then, for  $\varphi \in \operatorname{dom} \delta$ 

$$\mathbb{E}[\delta(\varphi)F] = \mathbb{E}\int \varphi(x)D_xF\,dx.$$

Let *L* be the generator of {*T<sub>s</sub>*}, then *L* = -δ*D*.
 The (pseudo)-inverse of *L* admits the representation

$$L^{-1}=-\int_0^\infty T_s\,ds=-\int_0^1 P_t\,\frac{dt}{t},$$

and

$$-DL^{-1}F = -\int_0^1 P_t DF \, dt.$$

$$D_x F = \sum_n n I_{n-1}(f_n(x, \cdot))$$

$$D_x F = \sum_n n I_{n-1}(f_n(x, \cdot))$$
$$P_t F = \sum_n t^n I_n(f_n)$$

■ 
$$D_x F = \sum_n n I_{n-1}(f_n(x, \cdot))$$
  
■  $P_t F = \sum_n t^n I_n(f_n)$   
■  $LF = -\sum_n n I_n(f_n)$ 

$$D_{x}F = \sum_{n} n I_{n-1}(f_{n}(x, \cdot))$$
  

$$P_{t}F = \sum_{n} t^{n} I_{n}(f_{n})$$
  

$$LF = -\sum_{n} n I_{n}(f_{n})$$
  

$$L^{-1}F = -\sum_{n} n^{-1} I_{n}(f_{n}).$$

The following bound is due to Eichelsbacher and Thäle (2013) (building on Schulte (2012)), and is based on a subtle use of Stein's method (see my mini-course):

The following bound is due to Eichelsbacher and Thäle (2013) (building on Schulte (2012)), and is based on a subtle use of Stein's method (see my mini-course): for every  $F \in L^2(\sigma(\eta))$  with mean zero and variance 1,

$$d_{Kol}(F,N) \leq \mathbb{E}|1 - \int (D_x F)(-D_x L^{-1} F) dx| + \frac{\sqrt{2\pi}}{8} \mathbb{E} \int (D_x F)^2 |D_x L^{-1} F| dx + \frac{1}{2} \mathbb{E} \int (D_x F)^2 |F| |D_x L^{-1} F| dx + \sup_t \mathbb{E} \int (D_x \mathbf{1} \{F > t\}) (D_x F) |D_x L^{-1} F| dx.$$

#### Theorem (Last, Peccati and Schulte, 2013)

Let  $F \in L^2(\sigma(\eta))$  be centered and such that Var F = 1. Let  $N \sim \mathcal{N}(0, 1)$ . then,

$$d_{Kol}(F, N) \leq \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + \gamma_6,$$

or, in a dynamic setting,

$$d_{Kol}(F_t, N) \leq t \times (\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + \gamma_6).$$

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

# The bounds

#### Here,

$$\begin{split} \gamma_1 &:= 4 \sqrt{\int \left[\mathbb{E}(D_{x_1}F)^2 (D_{x_2}F)^2\right]^{1/2} \left[\mathbb{E}(D_{x_1,x_3}^2F)^2 (D_{x_2,x_3}^2F)^2\right]^{1/2} dx_1 dx_2 dx_3},\\ \gamma_2 &:= \left[\int \mathbb{E}(D_{x_1,x_3}^2F)^2 (D_{x_2,x_3}^2F)^2 dx_1 dx_2 dx_3\right]^{1/2},\\ \gamma_3 &:= \int \mathbb{E}|D_xF|^3 dx, \end{split}$$

$$\begin{split} \gamma_4 &:= \frac{1}{2} [\mathbb{E}F^4]^{1/4} \int [\mathbb{E}(D_x F)^4]^{3/4} \, dx, \\ \gamma_5 &:= \left[ \int [\mathbb{E}(D_x F)^4 \, dx \right]^{1/2}, \\ \gamma_6 &:= \left[ \int 6 [[\mathbb{E}(D_{x_1} F)^4]^{1/2} [[\mathbb{E}(D_{x_1, x_2}^2 F)^4]^{1/2} + 3 [\mathbb{E}(D_{x_1, x_2}^2 F)^4 \, dx_1 \, dx_2]^{1/2}, \end{split}$$

2

▲口 → ▲圖 → ▲ 臣 → ▲ 臣 →

#### One has also the simpler bound

$$d_W(F, N) \leq \gamma_1 + \gamma_2 + \gamma_3,$$

where

$$d_{W}(F,N) = \sup_{h:|h'| \leq 1} |\mathbb{E}[h(F)] - \mathbb{E}[h(N)]|$$

is the 1-Wasserstein distance.

For every *t*, we consider the restriction of  $\eta_t$  to a compact window  $H \subset \mathbb{R}^d$ . We build the associated *k*-**nearest neighbour graph** as follows: two distinct points *x*, *y* in  $\eta_t \cap H$  are linked by an edge if and only if *x* is one of the *k*-nearest neighbours of *y*, or *y* is one of the *k*-nearest neighbours of *x*.

For every *t*, we consider the restriction of  $\eta_t$  to a compact window  $H \subset \mathbb{R}^d$ . We build the associated *k*-**nearest neighbour graph** as follows: two distinct points *x*, *y* in  $\eta_t \cap H$  are linked by an edge if and only if *x* is one of the *k*-nearest neighbours of *y*, or *y* is one of the *k*-nearest neighbours of *x*.

Here is an example for k = 1 (*courtesy of M. Schulte*)

## Application: the nearest neighbour graph



# Application: the nearest neighbour graph



#### Length of the nearest neighbour graph

We wish to establish an upper bound (for  $\alpha \in [0, 1]$ ) of the type

$$d_{\mathcal{K}ol}\left(\frac{L_t^{\alpha} - \mathbb{E}(L_t^{\alpha})}{\operatorname{Var}^{1/2}L_t^{\alpha}}, N\right) = d_{\mathcal{K}ol}\left(\frac{F_t - \mathbb{E}(F_t)}{\operatorname{Var}^{1/2}F_t}, N\right) \le a(t),$$

where

$$L_t^{\alpha} := \sum_{\boldsymbol{x} \sim \boldsymbol{y}; \boldsymbol{x}, \boldsymbol{y} \in \eta_t \cap H} \| \boldsymbol{x} - \boldsymbol{y} \|^{\alpha}, \quad \boldsymbol{F}_t = t^{\alpha/d} L_t^{\alpha}$$

(in such a way that **Var**  $F_t \ge \sigma_{\alpha} t$ , see Penrose and Yukich (2001)).

## Length of the nearest neighbour graph

We wish to establish an upper bound (for  $\alpha \in [0, 1]$ ) of the type

$$d_{\mathsf{Kol}}\left(\frac{L^{\alpha}_t - \mathbb{E}(L^{\alpha}_t)}{\operatorname{Var}^{1/2} L^{\alpha}_t}, \mathsf{N}\right) = d_{\mathsf{Kol}}\left(\frac{\mathsf{F}_t - \mathbb{E}(\mathsf{F}_t)}{\operatorname{Var}^{1/2} \mathsf{F}_t}, \mathsf{N}\right) \le a(t),$$

where

$$L_t^{\alpha} := \sum_{\boldsymbol{x} \sim \boldsymbol{y}; \boldsymbol{x}, \boldsymbol{y} \in \eta_t \cap H} \| \boldsymbol{x} - \boldsymbol{y} \|^{\alpha}, \quad F_t = t^{\alpha/d} L_t^{\alpha}$$

(in such a way that **Var**  $F_t \ge \sigma_{\alpha} t$ , see Penrose and Yukich (2001)).

Previous findings for  $\alpha = 1$ : Avram and Bertsimas (1993),  $a(t) = O((\log t)^{1+3/4} t^{-1/4})$ Penrose and Yukich (2005),  $a(t) = O((\log t)^{3d} t^{-1/2})$ .

# A general Berry-Esséen bound

Let  $H \subset \mathbb{R}^d$  be a compact set.

#### Proposition (Last, Peccati and Schulte, 2014)

Let  $F_t \in L^2(\sigma(\eta_t))$ ,  $t \ge 1$ , and assume there are finite constants  $p_1, p_2, c > 0$  such that

$$\mathbb{E} |D_x \mathcal{F}_t|^{4+
ho_1} \leq c, \quad \mathbb{E} |D_{x_1,x_2}^2 \mathcal{F}_t|^{4+
ho_2} \leq c,$$

Moreover, assume that  $\operatorname{Var} F_t / t > v$ ,  $t \ge 1$ , with v > 0 and that

$$m := \sup_{x \in H, \ t \ge 1} t \int \mathbb{P}(D_{x,y}^2 F_t \neq 0)^{p_2/(16+4p_2)} \, dy < \infty$$

Let N be a standard Gaussian random variable. Then, there exists a finite constant C, depending uniquely on  $c, p_1, p_2, v, m$  and the measure of H, such that

$$d_{Kol}\left(rac{F_t - \mathbb{E}(F_t)}{\sqrt{\operatorname{Var}F_t}}, N
ight) \leq C t^{-1/2}, \quad t \geq 1.$$

クへで 18/20

Our result requires to bound a quantity of the type

$$\sup_{x\in H, t\geq 1} t \int \mathbb{P}(D_{x,y}^2 F_t \neq 0)^\beta \, dy$$

Our result requires to bound a quantity of the type

$$\sup_{x\in H, t\geq 1} t \int \mathbb{P}(D_{x,y}^2 F_t \neq 0)^\beta \, dy$$

Assume that there exist **radii of stabilization**  $\{R_t(x, \eta_t)\}$ , verifying

$$D_x F_t(\eta_t) = D_x F_t(\eta_t \cap B^d(x, R_t(x, \eta_t))).$$

Our result requires to bound a quantity of the type

$$\sup_{x\in H, t\geq 1} t \int \mathbb{P}(D_{x,y}^2 F_t \neq 0)^\beta \, dy$$

Assume that there exist **radii of stabilization**  $\{R_t(x, \eta_t)\}$ , verifying

$$D_x F_t(\eta_t) = D_x F_t(\eta_t \cap B^d(x, R_t(x, \eta_t))).$$

Then, it suffices to show that

$$\sup_{x,t} \int t \mathbb{P}\left(y \in B^{d}(x, R_{t}(x, \eta_{t})) \text{ or } \\ R_{t}(x, \eta_{t} + \delta_{y}) \neq R_{t}(x, \eta_{t})\right)^{\beta} dy < \infty.$$

Our result requires to bound a quantity of the type

$$\sup_{x\in H, t\geq 1} t \int \mathbb{P}(D_{x,y}^2 F_t \neq 0)^\beta \, dy$$

Assume that there exist **radii of stabilization**  $\{R_t(x, \eta_t)\}$ , verifying

$$D_x F_t(\eta_t) = D_x F_t(\eta_t \cap B^d(x, R_t(x, \eta_t))).$$

Then, it suffices to show that

$$\begin{split} \sup_{x,t} \int t \, \mathbb{P} \left( y \in B^d(x, R_t(x, \eta_t)) \text{ or } \right. \\ \left. R_t(x, \eta_t + \delta_y) \neq R_t(x, \eta_t) \right)^\beta \, dy < \infty. \end{split}$$

This is very close to the **add-one cost stabilization** by Penrose and Yukich (2001).

This strategy works very well with the *k*-nng, yielding the estimate

This strategy works very well with the *k*-nng, yielding the estimate

Proposition (Last, Peccati and Schulte, 2014)

There exists a finite constant  $C_{\alpha}$  such that

$$d_{Kol}\left(rac{L_t^lpha-\mathbb{E}(L_t^lpha)}{\mathbf{Var}^{1/2}L_t^lpha},\mathbf{N}
ight)\leq rac{C_lpha}{\sqrt{t}}.$$

This strategy works very well with the *k*-nng, yielding the estimate

Proposition (Last, Peccati and Schulte, 2014)

There exists a finite constant  $C_{\alpha}$  such that

$$d_{Kol}\left(rac{L_t^{lpha}-\mathbb{E}(L_t^{lpha})}{\operatorname{Var}^{1/2}L_t^{lpha}}, \mathsf{N}
ight) \leq rac{C_{lpha}}{\sqrt{t}}.$$

Further applications will appear in Günter's talk, as well as in the mini course.