ON MEHLER'S FORMULA

Giovanni Peccati
(Luxembourg University)

Conférence Géométrie Stochastique

Nantes — April 7, 2016

OVERVIEW

- ⋆ I will discuss two joint works: Last, Peccati and Schulte (PTRF, 2016+), and Bachmann and Peccati (EJP, 2016) .
- ⋆ Common thread: a connection with **Mehler's formula**, providing a mixture-type representation of the Ornstein-Uhlenbeck semi-group on the Poisson space.
- ⋆ Use of **Malliavin/Stein** techniques.
- ⋆ Connections with the theory of **geometric stabilization** – Penrose and Yukich (2001), see also Kesten and Lee (1996).
- ⋆ Our concentration bounds complete e.g. results by Wu (2000), Houdré and Privault (2002), and Breton, Houdré and Privault (2007).

Overview

- ⋆ I will discuss two joint works: Last, Peccati and Schulte (PTRF, 2016+), and Bachmann and Peccati (EJP, 2016) .
- ⋆ Common thread: a connection with **Mehler's formula**, providing a mixture-type representation of the Ornstein-Uhlenbeck semi-group on the Poisson space.
- ⋆ Use of **Malliavin/Stein** techniques.
- ⋆ Connections with the theory of **geometric stabilization** – Penrose and Yukich (2001), see also Kesten and Lee (1996).
- ⋆ Our concentration bounds complete e.g. results by Wu (2000), Houdré and Privault (2002), and Breton, Houdré and Privault (2007).

OVERVIEW

- ⋆ I will discuss two joint works: Last, Peccati and Schulte (PTRF, 2016+), and Bachmann and Peccati (EJP, 2016) .
- ⋆ Common thread: a connection with **Mehler's formula**, providing a mixture-type representation of the Ornstein-Uhlenbeck semi-group on the Poisson space.
- ⋆ Use of **Malliavin/Stein** techniques.
- ⋆ Connections with the theory of **geometric stabilization** – Penrose and Yukich (2001), see also Kesten and Lee (1996).
- ⋆ Our concentration bounds complete e.g. results by Wu (2000), Houdré and Privault (2002), and Breton, Houdré and Privault (2007).

OVERVIEW

- ⋆ I will discuss two joint works: Last, Peccati and Schulte (PTRF, 2016+), and Bachmann and Peccati (EJP, 2016) .
- ⋆ Common thread: a connection with **Mehler's formula**, providing a mixture-type representation of the Ornstein-Uhlenbeck semi-group on the Poisson space.
- ⋆ Use of **Malliavin/Stein** techniques.
- ⋆ Connections with the theory of **geometric stabilization** – Penrose and Yukich (2001), see also Kesten and Lee (1996).
- ⋆ Our concentration bounds complete e.g. results by Wu (2000), Houdré and Privault (2002), and Breton, Houdré and Privault (2007).

Overview

- ⋆ I will discuss two joint works: Last, Peccati and Schulte (PTRF, 2016+), and Bachmann and Peccati (EJP, 2016) .
- ⋆ Common thread: a connection with **Mehler's formula**, providing a mixture-type representation of the Ornstein-Uhlenbeck semigroup on the Poisson space.
- ⋆ Use of **Malliavin/Stein** techniques.
- ⋆ Connections with the theory of **geometric stabilization** – Penrose and Yukich (2001), see also Kesten and Lee (1996).
- ⋆ Our concentration bounds complete e.g. results by Wu (2000), Houdré and Privault (2002), and Breton, Houdré and Privault (2007).

FRAMEWORK

- ⋆ For every $t \geq 1$, η_t is a Poisson measure on $\mathbb{R}^d$ ($d \geq 1$), with intensity $t \times$ Lebesgue.
- ⋆ We denote by $F_t = F_t(\eta_t)$ a generic square-integrable functional of η_t, write $m(t) = \mathbb{E}[F_t]$, $v(t) = \mathbf{Var}\, F_t$, and

$$\widetilde{F}_t = \frac{F_t - m(t)}{v(t)^{1/2}}.$$

- ⋆ We shall write $\eta = \eta_1$, $F = F_1$, $m = m(1)$, ... and so on.

Framework

- ⋆ For every $t \geq 1$, η_t is a Poisson measure on $\mathbb{R}^d$ ($d \geq 1$), with intensity $t \times$ Lebesgue.
- ⋆ We denote by $F_t = F_t(\eta_t)$ a generic square-integrable functional of η_t, write $m(t) = \mathbb{E}[F_t]$, $v(t) = \mathbf{Var}\, F_t$, and

$$\widetilde{F}_t = \frac{F_t - m(t)}{v(t)^{1/2}}.$$

- ⋆ We shall write $\eta = \eta_1$, $F = F_1$, $m = m(1)$, ... and so on.

FRAMEWORK

- ⋆ For every $t \geq 1$, η_t is a Poisson measure on $\mathbb{R}^d$ ($d \geq 1$), with intensity $t \times$ Lebesgue.
- ⋆ We denote by $F_t = F_t(\eta_t)$ a generic square-integrable functional of η_t, write $m(t) = \mathbb{E}[F_t]$, $v(t) = \mathbf{Var}\, F_t$, and

$$\widetilde{F}_t = \frac{F_t - m(t)}{v(t)^{1/2}}.$$

- ⋆ We shall write $\eta = \eta_1$, $F = F_1$, $m = m(1)$, ... and so on.

GOALS

- ⋆ **Goal 1**: If $v(t) \geq \sigma t$, as $t \to \infty$, we look for "optimal" Berry-Esseen 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 $N \sim \mathscr{N}(0,1)$. Tool: **second order Poincaré inequalities**.

- ⋆ **Goal 2**: In a non dynamic setting, we look for (exponential and Gaussian) concentration bounds on **upper and lower tails**

$$\mathbb{P}\Big(F - m \geq r\Big), \quad \mathbb{P}\Big(F - m \leq -r\Big), \quad r > 0.$$

Tool: **logarithmic Sobolev inequalities**.

GOALS

- $\star$ **Goal 1**: If $v(t) \geq \sigma t$, as $t \to \infty$, we look for "optimal" Berry-Esseen bounds of the type

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

  where $N \sim \mathscr{N}(0,1)$. Tool: **second order Poincaré inequalities**.
- $\star$ **Goal 2**: In a non dynamic setting, we look for (exponential and Gaussian) concentration bounds on **upper and lower tails**

$$\mathbb{P}\Big(F - m \geq r\Big), \quad \mathbb{P}\Big(F - m \leq -r\Big), \quad r > 0.$$

  Tool: **logarithmic Sobolev inequalities**.

Chaos and Malliavin Operators

- $\star$ Recall that every $F \in L^2(\sigma(\eta))$ admits a unique **chaotic decomposition** of the type

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

- $\star$ Define the **Ornstein-Uhlenbeck semigroup** $\{T_s : s \geq 0\}$ by

$$T_s F = \mathbb{E}(F) + \sum_{n \geq 1} e^{-ns} I_n(f_n).$$

- $\star$ For a functional F of η and $x \in \mathbb{R}^d$, define

$$D_x F(\eta) = F(\eta + \delta_x) - F(\eta)$$

(**add-one cost operator**).

CHAOS AND MALLIAVIN OPERATORS

- ⋆ Recall that every $F \in L^2(\sigma(\eta))$ admits a unique **chaotic decomposition** of the type

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

- ⋆ Define the **Ornstein-Uhlenbeck semigroup** $\{T_s : s \geq 0\}$ by

$$T_s F = \mathbb{E}(F) + \sum_{n \geq 1} e^{-ns} I_n(f_n).$$

- ⋆ For a functional F of η and $x \in \mathbb{R}^d$, define

$$D_x F(\eta) = F(\eta + \delta_x) - F(\eta)$$

(**add-one cost operator**).

Chaos and Malliavin Operators

- ⋆ Recall that every $F \in L^2(\sigma(\eta))$ admits a unique **chaotic decomposition** of the type

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

- ⋆ Define the **Ornstein-Uhlenbeck semigroup** $\{T_s : s \geq 0\}$ by

$$T_s F = \mathbb{E}(F) + \sum_{n\geq 1} e^{-ns} I_n(f_n).$$

- ⋆ For a functional F of η and $x \in \mathbb{R}^d$, define

$$D_x F(\eta) = F(\eta + \delta_x) - F(\eta)$$

(**add-one cost operator**).

Some Remarkable Relations

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

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

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

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

⋆ Let L be the the **generator** of $\{T_s\}$, then

$$L = -\delta D.$$

⋆ The **pseudo-inverse** of L is written L^{-1}, and we have

$$-DL^{-1}F = \int_0^\infty e^{-s} T_s \, DF \, ds.$$

Some Remarkable Relations

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

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

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

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

- Let L be the the **generator** of $\{T_s\}$, then

$$L = -\delta D.$$

- The **pseudo-inverse** of L is written L^{-1}, and we have

$$-DL^{-1}F = \int_0^\infty e^{-s} T_s \, DF \, ds.$$

Some Remarkable Relations

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

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

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

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

- Let L be the the **generator** of $\{T_s\}$, then

$$L = -\delta D.$$

- The **pseudo-inverse** of L is written L^{-1}, and we have

$$-DL^{-1}F = \int_0^\infty e^{-s} T_s \, DF \, ds.$$

MEHLER'S FORMULA

Theorem (Mehler's formula)

For every $s \geq 0$ *, define* $\eta^{(s)}$ *to be a* e^{-s}–**thinning** *of* η*, and let* $\hat{\eta}^{(s)}$ *be an independent Poisson measure with intensity* $(1 - e^{-s})\times$ *Lebesgue. The Ornstein-Uhlenbeck semigroup* $\{T_s : s \geq 0\}$ *can be represented as follows:*

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

Berry-Esseen Bounds: 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 $f : \mathbb{R}^d \to \mathbb{R}$,

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

- ⋆ The first example of a **second order Poincaré estimate** appears in Chatterjee (2007):

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

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

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

where D stands for the **Malliavin derivative**.

Berry-Esseen Bounds: 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 $f : \mathbb{R}^d \to \mathbb{R}$,

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

- ⋆ The first example of a **second order Poincaré estimate** appears in Chatterjee (2007):

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

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

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

where D stands for the **Malliavin derivative**.

Berry-Esseen Bounds: 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 $f : \mathbb{R}^d \to \mathbb{R}$,

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

- ⋆ The first example of a **second order Poincaré estimate** appears in Chatterjee (2007):

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

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

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

where D stands for the **Malliavin derivative**.

Towards the Poisson Framework

- ⋆ We shall build on the following Poincaré inequality: for every $F \in L^2(\sigma(\eta))$,

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

- ⋆ 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 relation $d_{Kol} \le 2\sqrt{d_W}$ 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.

Towards the Poisson Framework

- ⋆ We shall build on the following Poincaré inequality: for every $F \in L^2(\sigma(\eta))$,

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

- ⋆ 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 relation $d_{Kol} \le 2\sqrt{d_W}$ 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.

TOWARDS THE POISSON FRAMEWORK

- ⋆ We shall build on the following Poincaré inequality: for every $F \in L^2(\sigma(\eta))$,

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

- ⋆ 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 relation $d_{Kol} \leq 2\sqrt{d_W}$ 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.

TOWARDS THE POISSON FRAMEWORK

- ⋆ We shall build on the following Poincaré inequality: for every $F \in L^2(\sigma(\eta))$,

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

- ⋆ 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 relation $d_{Kol} \leq 2\sqrt{d_W}$ 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.

A Bound Based on Stein's Method

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: for every $F \in L^2(\sigma(\eta))$ with mean zero and variance 1,

$$
\begin{aligned}
d_{Kol}(F,N) \le \mathbb{E}\Big|1 - \int (D_xF)(-D_xL^{-1}F)\,dx\Big| \\
+ \frac{\sqrt{2\pi}}{8}\mathbb{E}\int (D_xF)^2|D_xL^{-1}F|\,dx \\
+ \frac{1}{2}\mathbb{E}\int (D_xF)^2|F||D_xL^{-1}F|\,dx \\
+ \sup_t \mathbb{E}\int (D_x\mathbf{1}\{F>t\})(D_xF)|D_xL^{-1}F|\,dx.
\end{aligned}
$$

Similar bounds in the **binomial setting**: Lachièze-Rey and Peccati (Ann. Appl. Probab., 2016+).

A Bound Based on Stein's Method

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: for every $F \in L^2(\sigma(\eta))$ with mean zero and variance 1,

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

Similar bounds in the **binomial setting**: Lachièze-Rey and Peccati (Ann. Appl. Probab., 2016+).

A Bound Based on Stein's Method

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: for every $F \in L^2(\sigma(\eta))$ with mean zero and variance 1,

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

Similar bounds in the **binomial setting**: Lachièze-Rey and Peccati (Ann. Appl. Probab., 2016+).

General Second Order Poincaré Inequalities

Theorem (Last, Peccati and Schulte, 2016+)
Let $F \in L^2(\sigma(\eta))$ *be centered and such that* $\mathbf{Var}\, F = 1$. *Let* $N \sim \mathscr{N}(0,1)$. *Then,*

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

or, in a dynamic setting,

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

THE BOUNDS

Here,

$$\gamma_1 = 4\sqrt{\iint \left[\mathbb{E}(D_{x_1}F)^2(D_{x_2}F)^2\right]^{1/2}\left[\mathbb{E}(D^2_{x_1,x_3}F)^2(D^2_{x_2,x_3}F)^2\right]^{1/2} dx_1dx_2dx_3},$$

$$\gamma_2 = \left[\int \mathbb{E}(D^2_{x_1,x_3}F)^2(D^2_{x_2,x_3}F)^2\, dx_1dx_2dx_3\right]^{1/2},$$

$$\gamma_3 = \int \mathbb{E}|D_xF|^3\, dx,$$

$$\gamma_4 = \frac{1}{2}\left[\mathbb{E}F^4\right]^{1/4}\int \left[\mathbb{E}(D_xF)^4\right]^{3/4} dx,$$

$$\gamma_5 = \left[\int [\mathbb{E}(D_xF)^4\, dx\right]^{1/2},$$

$$\gamma_6 = \left[\int 6\left[[\mathbb{E}(D_{x_1}F)^4\right]^{1/2}\left[[\mathbb{E}(D^2_{x_1,x_2}F)^4\right]^{1/2} + 3[\mathbb{E}(D^2_{x_1,x_2}F)^4\, dx_1dx_2\right]^{1/2}$$

Application: the Nearest Neighbour Graph

- For every t, we consider the restriction of η_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.

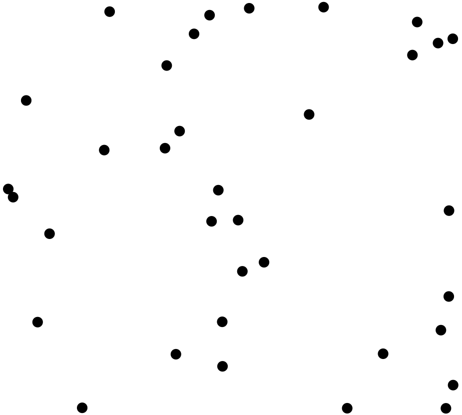
- Here is an example for $k = 1$ (*courtesy of M. Schulte*)

APPLICATION: THE NEAREST NEIGHBOUR GRAPH

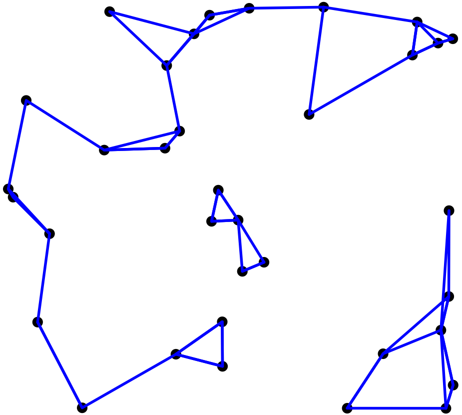
- For every t, we consider the restriction of η_t to a compact window $H \subset \mathbb{R}^d$. We build the associated k-**nearest neighbour graph** as follows: two distinct points x, y in $\eta_t \cap H$ are linked by an edge if and only if x is one of the k-nearest neighbours of y, or y is one of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

$$d_{Kol}\left(\frac{L_t^\alpha - \mathbb{E}(L_t^\alpha)}{\mathbf{Var}^{1/2} L_t^\alpha}, N\right) = d_{Kol}\left(\frac{F_t - \mathbb{E}(F_t)}{\mathbf{Var}^{1/2} F_t}, N\right) \leq a(t),$$

where

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

(in such a way that $\mathbf{Var}\, F_t \geq \sigma_\alpha t$, 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), $a(t) = O((\log t)^{3d} t^{-1/2})$.

LENGTH OF THE NEAREST NEIGHBOUR GRAPH

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

$$d_{Kol}\left(\frac{L_t^\alpha - \mathbb{E}(L_t^\alpha)}{\mathbf{Var}^{1/2} L_t^\alpha}, N\right) = d_{Kol}\left(\frac{F_t - \mathbb{E}(F_t)}{\mathbf{Var}^{1/2} F_t}, N\right) \leq a(t),$$

where

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

(in such a way that $\mathbf{Var}\, F_t \geq \sigma_\alpha t$, 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), $a(t) = O((\log t)^{3d} t^{-1/2})$.

Proposition (Last, Peccati and Schulte, 2016+)
Let $F_t \in L^2(\sigma(\eta_t))$, $t \geq 1$, and assume there are finite constants $p_1, p_2, c > 0$ such that

$$\mathbb{E}|D_x F_t|^{4+p_1} \leq c, \quad \mathbb{E}|D^2_{x_1,x_2} F_t|^{4+p_2} \leq c,$$

Moreover, assume that $\mathbf{Var} F_t / t > v$, $t \geq 1$, *with* $v > 0$ *and that*

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

Let $N \sim \mathcal{N}(0,1)$. *Then,*

$$d_{Kol}\left(\frac{F_t - \mathbb{E}(F_t)}{\sqrt{\mathbf{Var} F_t}}, N \right) \leq \frac{C}{\sqrt{t}}, \quad t \geq 1.$$

Connections with Stabilization Theory

- Our result requires to bound a quantity of the type

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

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

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

- Then, it suffices to show that

$$\sup_{x,t} \int t\,\mathbb{P}\left\{R_t(x,\eta_t) \leq \|x-y\| \text{ or } R_t(x,\eta_t+\delta_y) \neq R_t(x,\eta_t)\right\}^\beta dy < \infty.$$

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

CONNECTIONS WITH STABILIZATION THEORY

- Our result requires to bound a quantity of the type

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

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

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

- Then, it suffices to show that

$$\sup_{x,t} \int t \, \mathbb{P}\left\{R_t(x, \eta_t) \leq \|x - y\| \text{ or } R_t(x, \eta_t + \delta_y) \neq R_t(x, \eta_t)\right\}^\beta dy < \infty.$$

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

CONNECTIONS WITH STABILIZATION THEORY

- ⋆ Our result requires to bound a quantity of the type

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

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

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

- ⋆ Then, it suffices to show that

$$\sup_{x,t} \int t \, \mathbb{P}\,\{R_t(x, \eta_t) \leq \|x - y\| \text{ or } R_t(x, \eta_t + \delta_y) \neq R_t(x, \eta_t)\}^\beta \, dy < \infty.$$

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

CONNECTIONS WITH STABILIZATION THEORY

- ⋆ Our result requires to bound a quantity of the type

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

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

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

- ⋆ Then, it suffices to show that

$$\sup_{x,t} \int t\,\mathbb{P}\,\{R_t(x,\eta_t) \leq \|x-y\| \text{ or } R_t(x,\eta_t+\delta_y) \neq R_t(x,\eta_t)\}^\beta \, dy < \infty.$$

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 following estimate:

Proposition (Last, Peccati and Schulte, 2016+)

Let $N \sim \mathscr{N}(0,1)$. *There exists a finite constant* C_α *such that*

$$d_{Kol}\left(\frac{L_t^\alpha - \mathbb{E}(L_t^\alpha)}{\mathbf{Var}^{1/2}\, L_t^\alpha}, N\right) \leq \frac{C_\alpha}{\sqrt{t}}.$$

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

Proposition (Last, Peccati and Schulte, 2016+)

Let $N \sim \mathscr{N}(0,1)$. There exists a finite constant C_α such that

$$d_{Kol}\left(\frac{L_t^\alpha - \mathbb{E}(L_t^\alpha)}{\mathbf{Var}^{1/2}\, L_t^\alpha}, N\right) \leq \frac{C_\alpha}{\sqrt{t}}.$$

Write

$$\Phi(x) = x \log x$$

By approximation, and exploiting the fact that the mapping

$$(x, y) \mapsto y(\Phi'(x+y) - \Phi'(x))$$

is convex, Mehler's formula can be used to provide an intrinsic proof of the following fundamental inequality

---

Theorem (Log-Sobolev Inequality; Wu, 2000)

Let $G = G(\eta) > 0$ be integrable. Then,

$$\mathbb{E}[\Phi(G)] - \Phi(\mathbb{E}[G]) \leq \mathbb{E} \int (D_x \Phi(G) - \Phi'(G) D_x G)\, dx.$$

CONCENTRATION

⋆ When applied to random variables of the type $G = e^{\lambda F}$, log-Sobolev + 'Herbst argument' yield concentration estimates (typically, exponential) on F, provided F is Lipschitz and

$$\int (D_x F)^2 \, dx \leq c$$

Problem: these assumptions are often *not adapted* to a geometric setting.

⋆ In Bachmann and Peccati (EJP, 2016): combine Log-Sobolev with **Mecke's formula**, in order to deduce concentration estimates involving quantities of the type

$$V = \int (F(\eta) - F(\eta - \delta_x))^2 \, \eta(dx),$$

that are indeed amenable to geometric analysis. Close in spirit to Boucheron, Lugosi and Massart (2003).

⋆ When applied to random variables of the type $G = e^{\lambda F}$, log-Sobolev + 'Herbst argument' yield concentration estimates (typically, exponential) on F, provided F is Lipschitz and

$$\int (D_x F)^2 \, dx \leq c$$

Problem: these assumptions are often *not adapted* to a geometric setting.

⋆ In Bachmann and Peccati (EJP, 2016): combine Log-Sobolev with **Mecke's formula**, in order to deduce concentration estimates involving quantities of the type

$$V = \int (F(\eta) - F(\eta - \delta_x))^2 \, \eta(dx),$$

that are indeed amenable to geometric analysis. Close in spirit to Boucheron, Lugosi and Massart (2003).

CONCENTRATION

- ⋆ When applied to random variables of the type $G = e^{\lambda F}$, log-Sobolev + 'Herbst argument' yield concentration estimates (typically, exponential) on F, provided F is Lipschitz and

$$\int (D_x F)^2 \, dx \leq c$$

Problem: these assumptions are often *not adapted* to a geometric setting.

- ⋆ In Bachmann and Peccati (EJP, 2016): combine Log-Sobolev with **Mecke's formula**, in order to deduce concentration estimates involving quantities of the type

$$V = \int (F(\eta) - F(\eta - \delta_x))^2 \, \eta(dx),$$

that are indeed amenable to geometric analysis. Close in spirit to Boucheron, Lugosi and Massart (2003).

Representative Statement

Theorem (Bachmann and Peccati, 2016)
Assume that $DF \geq 0$ and $V \leq c$, then

$$\mathbb{P}[F \geq m + r] \leq e^{-r^2/c}, \quad r > 0.$$

Assume that $F \geq 0$, $DF \geq 0$ and $V \leq cF^{\alpha}$ ($\alpha \in [0, 2)$), then

$$\mathbb{P}[F \geq m + r] \leq \exp\left\{ -\frac{((r+m)^{1-\alpha/2} - m^{1-\alpha/2})^2}{2c} \right\}, \quad r > 0.$$

Applications to: **length power functionals**, **subgraph counting** (also, Bachmann and Reitzner, 2015), intrinsic proof of the **convex distance inequality for Poisson measures** (Reitzner, 2013), U**-statistics**, **component counts** (Bachmann, 2015).

Representative Statement

Theorem (Bachmann and Peccati, 2016)

Assume that $DF \geq 0$ *and* $V \leq c$*, then*

$$\mathbb{P}[F \geq m + r] \leq e^{-r^2/c}, \quad r > 0.$$

Assume that $F \geq 0$*,* $DF \geq 0$ *and* $V \leq cF^{\alpha}$ *(*$\alpha \in [0, 2)$*), then*

$$\mathbb{P}[F \geq m + r] \leq \exp\left\{-\frac{((r+m)^{1-\alpha/2} - m^{1-\alpha/2})^2}{2c}\right\}, \quad r > 0.$$

Applications to: **length power functionals**, **subgraph counting** (also, Bachmann and Reitzner, 2015), intrinsic proof of the **convex distance inequality for Poisson measures** (Reitzner, 2013), U-**statistics**, **component counts** (Bachmann, 2015).

ADVERTISING