Recent Berry-Esseen bounds obtained with Stein's method and Poincare inequalities, with Geometric applications

Raphaël Lachièze-Rey, Univ. Paris 5 René Descartes,

Georgia Tech.

Minimal spanning tree

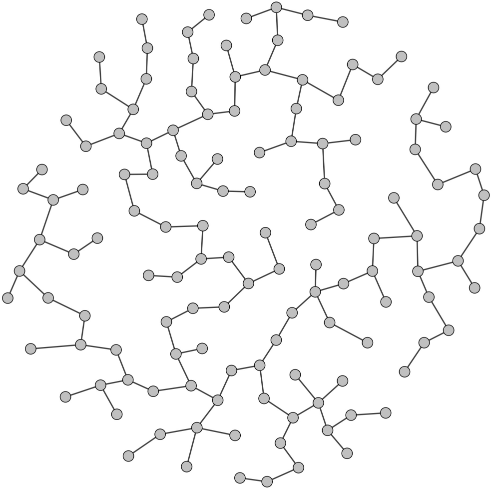
- X : Finite set in $\mathbb{R}^d$
- $M(X)$: Connected graph on X minimizing

$$\sum_{\{x,y\} \text{ edge}} \|x - y\|.$$

- Unique if the points of X are "in general position" (for interesting random point processes, happens a.s.) M(X) : **Minimal Spanning Tree**
- No loops
- We are interested in the functional

$$\varphi(X) = \sum_{\{x,y\} \text{ edge of } M(X)} \|x - y\|$$

Example



Random input

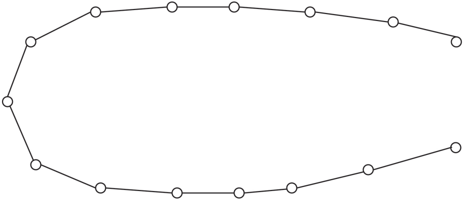
The random input X_n will tipically be,

- either a Poisson process with intensity 1 on the window $\mathbb{X}_n := [0, n^{1/d}]^d$ **"Poisson input"**
- Or a set of n uniform iid points on $\mathbb{X}_n$ **"Binomial input"**,

and we study the law of $\varphi(X_n)$ in the asymptotics $n \to \infty$.

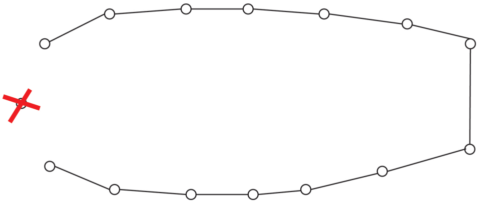
What happens when you remove a point

- If you remove a point, it might not make a big difference, but it might also change the structure far away. With high probability ?



What happens when you remove a point

- If you remove a point, it might not make a big difference, but it might also change the structure far away. With high probability ?



First and second order derivatives

- For establishing **limit theorems**, we will quantify this dependency through **discrete derivatives**.
- Introduce the **first order derivative**, for $x \in \mathbb{R}^d$:

$$D_x\varphi(X) = \varphi(X \cup \{x\}) - \varphi(X)$$

  Related to a classical notion of **influence**

- Say that a point y has no interaction with a point x if

$$D_x\varphi(X \cup \{y\}) = D_x\varphi(X)$$

  i.e.

$$D_y(D_x\varphi(X)) = 0.$$

- This is termed the **second order derivative** and is symmetric in x, y :

$$D^2_{y,x}\varphi(X) = \varphi(X \cup \{x, y\}) - \varphi(X \cup \{x\}) - \varphi(Y \cup \{y\}) + \varphi(X).$$

- x and y **"don't interact"** if $D^2_{y,x}\varphi(X) = 0$.

Stabilization

- N : Gaussian standard variable
- $\tilde{\varphi}(X_n) = \mathbf{Var}(\varphi(X_n))^{-1/2}(\varphi(X_n) - \mathbf{E}\varphi(X_n))$

We already know since the 90's (Kesten & Lee) that

$$\tilde{\varphi}(X_n) \to N$$

in law, as $n \to \infty$. They introduced the idea of **stabilization radius** : Given a point $x \in \mathbb{R}^d$, there is a.s. a radius $R_x > 0$ independent of n such that for $y \notin B(x, R)$,

$$D^2\varphi_{x,y}\varphi(X_n) = 0$$

The question is : At what speed does the convergence occur ?

- d_W : **Wasserstein distance**, defined by

$$d_W(U,V) = \sup_{h \text{ 1-Lipschitz}} |\mathbf{E}[h(U) - h(V)]|.$$

- d_K : **Kolmogorov distance**, defined by

$$d_K(U,V) = \sup_{t\in\mathbb{R}} |\mathbf{P}(U \leqslant t) - \mathbf{P}(V \leqslant t)|.$$

The aim of a "**2d-order Poincaré inequality**" in the Poisson framework is to bound $d_W(\tilde{\varphi}(X), N)$ (or $d_K(\tilde{\varphi}(X), N)$) in terms of $\mathbf{P}(D^2_{x,y}\varphi(X) \neq 0)$.

Stein's method and Berry-Essèen bounds

We have $\mathbf{E}[Nf(N) - f'(N)] = 0$ for f smooth enough. **Stein's method** gives, for any variable U,

$$d_W(U,N) \leqslant c \sup_{f:\underbrace{\|f\|_\infty \leqslant 1, \|f'\|_\infty \leqslant 1}_{(*)}, \|f''\|_\infty \leqslant 1} |\mathbf{E}Uf(U) - f'(U)|$$

$$d_K(U,N) \leqslant c \sup_{t\in\mathbb{R}, f \text{ satisfies } (**)} |\mathbf{E}Uf(U) - f'(U)|.$$

where f satisfies $(**)$ if it satisfies $(*)$ and some **second order Taylor inequality** depending on t :

$$|\underbrace{f(s+h) - f(s) - f'(s)h}_{\text{2d order difference}}| \leqslant \underbrace{h^2(|s|+1)}_{\text{2d order term}} + \underbrace{h(\mathbf{1}_{\{x\leqslant t\leqslant x+h\}} - \mathbf{1}_{\{x+h\leqslant t\leqslant x\}})}_{\text{has to be dealt with specifically}}.$$

In the case of a random input X and a functional $\varphi(X)$, the challenge is then to express

$$\mathbf{E}[\varphi(X)f(\varphi(X)) - f'(\varphi(X))]$$

in terms of the derivatives $D_x\varphi(X), D^2_{x,y}\varphi(X)$. This is where **Stein's method** has to be combined with other analytic methods

- **Malliavin calculus** for Poisson input Peccati, Nourdin, Last, Reitzner, Schulte, LR, ... Based on an **orthogonal chaotic decomposition**
- **Another specific decomposition** for binomial input Chatterjee, Peccati & LR

In some sense, **Stein's method** deals with the **target law**, and the decomposition deals with the random **input process**.

A "2d-order Poincaré"-like inequality LR, Schulte, Yukich

We need

$$\sup_{x\in X_n} \sup_{A\subset X_n, |A|\leqslant 1} \mathbf{E}[D_x\varphi(X_{n-1-|A|}\cup A)^7] \leqslant \text{constant},$$

$$\psi_n(x,y) = \sup_{A\subset X_n, |A|\leqslant 1} \mathbf{P}(D^2_{x,y}\varphi(X_{n-2-|A|}\cup A)\neq 0)^{1/6}, x,y\in\mathbb{E}$$

small when x, y are far away

Then, with $\sigma^2 = \mathbf{Var}(\varphi(X_n))$, **typically $\sigma^2 \sim n$**

$$d_K(\tilde{\varphi}(X_n), N) \leqslant \frac{n}{\sigma^3} + \frac{1}{\sigma^2}\left[\sqrt{n} + n\sqrt{\int_{X_n^2}\psi_n(x,y)dxdy} + n^{3/2}\sqrt{\int_{X_n}\left(\int_{X_n}\psi_n(x,y)dy\right)^2 dx}\right].$$

Comments

- Based on previous works of Chatterjee 2008, and LR&Peccati 2015
- A similar result exists with **Poisson input** Last, Peccati, Schulte 2014
- Already used to give optimal **Berry-Essèen bounds** for more simple functionals, or combinatorial functionals
  - Boolean model LR, Peccati
  - Nearest neighbour graph Last, Peccati, Schulte
  - Voronoi tessellation (Voronoi set approximation) LR, Peccati
  - Proximity graphs (work in progress) Goldstein, Johnson, LR
  - Longest increasing subsequences ? (with C. Houdré)

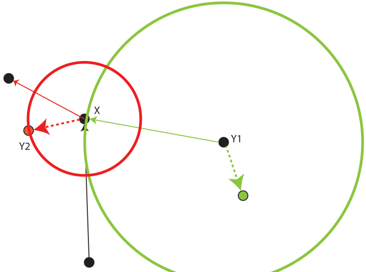
  All these examples are **exponentially stabilizing**. This is not the case for
  - Minimal spanning tree
  - Random sequential packing
  - Travelling salesman problem
  - Matching problems
  - ...

In many applications, it is easy to get a good estimate on the second order derivative. **Example** : **Nearest neighbours graph length** :

$$\varphi(X) = \sum_{x \in X} \|x - NN(x, X)\|$$

where $NN(x, X)$ is the nearest neighbour of x in X. We have

$D^2_{x,y}\varphi(X) \neq 0$ implies that **some ball with diameter $\|x - y\|$ contains at most one point of X.**



Therefore, with Poisson or binomial input,

$$\mathbf{P}(D^2_{x,y}\varphi(X) \neq 0) \leqslant c\|x-y\|^{-d}\exp(-cn\|x-y\|^d)$$

for some $c > 0$. This is enough to get $d(\tilde{\varphi}(X), N) \leqslant Cn^{-1/2}$ for some $C > 0$, with either **Poisson** or **binomial input**, and **Wasserstein** or **Kolmogorov distance**.

Derivatives estimates for the MST

Getting a bound for the **MST** is harder. Recall that

$$\varphi(X) = \sum_{\{x,y\} \text{ edge of the MST}} \|x - y\|.$$

It is easy to see that

$$|D_x\varphi(X)| \leqslant \|x - NN(x, X)\| + \|x - \underbrace{NN(x, X \setminus NN(x, X))}_{\text{Second nearest neighbour}}\|,$$

which gives a constant $C > 0$ such that, for all $n \geqslant 1, x \in X_n$

$$\mathbf{E}|D_x\varphi(X_n)|^7 dx \leqslant C$$

Second-order derivative

- Getting a good estimate on

$$\mathbf{P}(D^2_{x,y}\varphi(X) \neq 0)$$

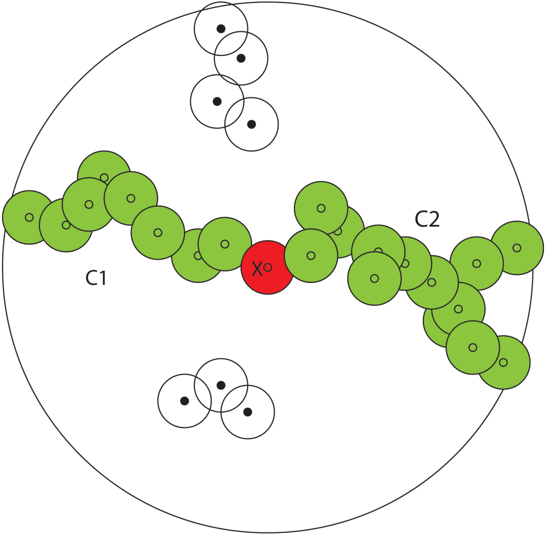
is the key for obtaining a good bound on $d_W(\tilde{\varphi}(X), N)$.

- Chatterjee & Sen 2013 obtained a bound directly without using such estimates. They obtained that in dimension 2, for some $\gamma > 0$,

$$d_W(\tilde{\varphi}(X), N) \leqslant Cn^{-\gamma},$$

and γ is related to the **2-arm exponent** β, that we define below.

Two-arm event in x among $B(x, R)$ at level $\ell > 0$



Two-arm event

Given a point set X, a distance $\ell > 0$, define

$$X^{\oplus \ell} = \bigcup_{x \in X} B(x, \ell).$$

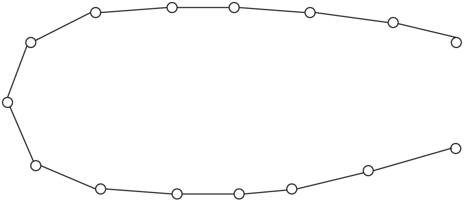
For $x \in X$ and $R > 0$, a **two-arm event** with these parameters is realized if

- $(X \setminus x)^{\oplus \ell} \cap B(x, R)$ has at least **two connected components** C_1, C_2
- $C_1 \cup C_2 \cup B(x, \ell)$ is connected
- C_1 and C_2 both touch $\partial B(x, R)$.

Minimax property of the MST

- Given a finite set X in general position and $x, y \in X$, x and y are connected in X iff there is no path $x_0 = x, x_1 \in X, \ldots, x_{q-1} \in X, x_q = y$ such that $\|x_i - x_{i+1}\| < \|x - y\|$.
- In other words, x and y are connected in the MST by the path γ minimizing

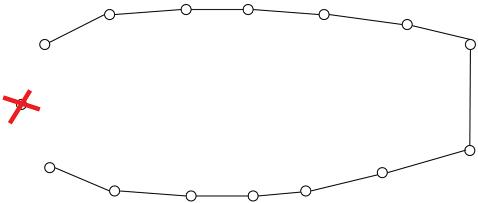
$$\max_{\{a,b\} \text{ edge of } \gamma} \|a - b\|.$$



Minimax property

- Given a finite set X in general position and $x, y \in X$, x and y are connected in X iff there is no path $x_0 = x, x_1 \in X, \dots, x_{q-1} \in X, x_q = y$ such that $\|x_i - x_{i+1}\| < \|x - y\|$.
- In other words, x and y are connected in the MST by the path γ minimizing

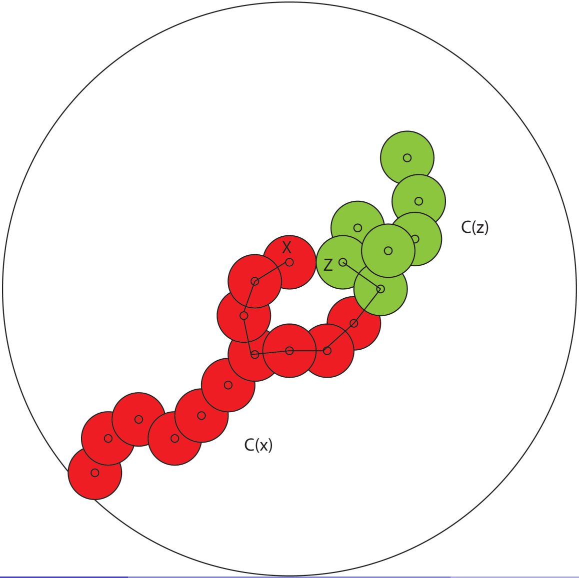
$$\max_{\{a,b\} \text{ edge of } \gamma} \|a - b\|.$$



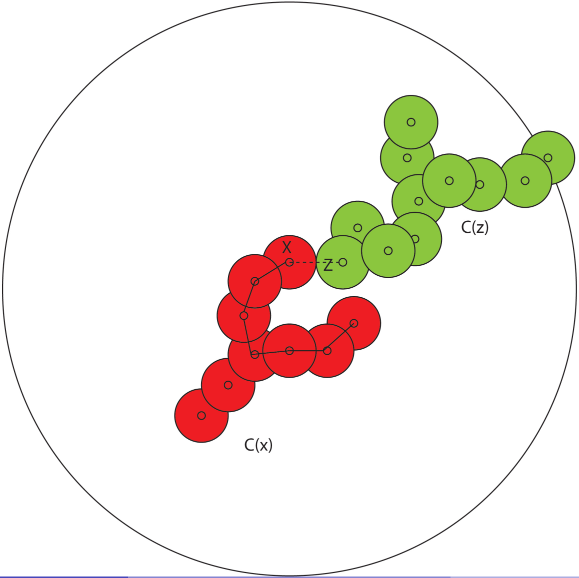
Stabilization radius

- Given $x \in X$, we are looking for some $R > 0$ such that for y outside $B(x, R)$, $D^2_{x,y}M(X) = 0$.
- Such a number is called a **stabilization radius**. This notion is fundamental for understanding the asymptotics of geometric functionals.
- To estimate $R = R(x, X)$, we introduce z "close to" x, and study if the removal/addition of a point y outside $B(x, R)$ can affect the presence of the edge $\{x, z\}$ in the MST.

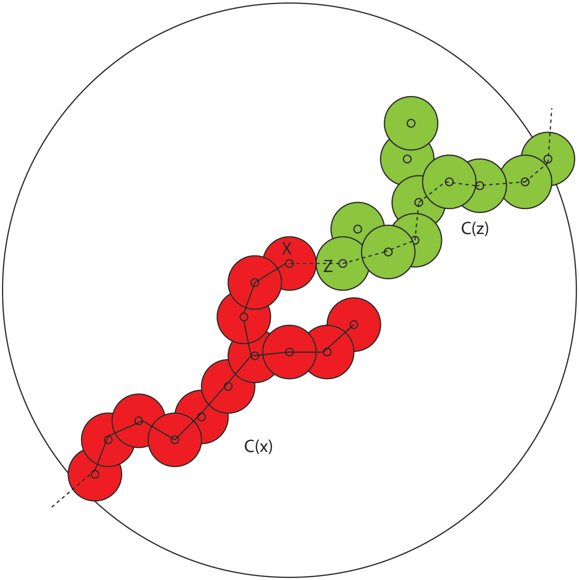
Case 1 : x and z are not connected no matter what is X outside $B(x, R)$



Case 2 : x and z are connected no matter what is X outside $B(x, R)$



Case 3 : Depends on $X \cap B(x, R)^c$



Several cases occur. Call $\ell = \|x - z\|$, $C(x)$ the connected component of $(X \setminus \{z\})^{\oplus \ell}$ containing x, $C(z)$ the component of z in $(X \setminus \{x\})^{\oplus \ell}$.

- If $C(x)$ and $C(z)$ meet inside $B(x, R)$, by the minimax property, $\{x, z\}$ is not an edge of the MST, no matter what is X outside $B(x, R)$
- If $C(x)$ is contained in $B(x, R)$ and disjoint from $C(z)$, then $\{x, z\}$ is an edge no matter what.
- If $C(x)$ and $C(y)$ do not meet inside $B(x, R)$, but both touch the boundary, they might be connected outside $B(x, R)$, or not. This is a two arm-event. Therefore

$$\begin{aligned}&\mathbf{P}(\{z, x\} \text{ affected by } X \setminus B(x, R)) \\ &\qquad \leqslant \mathbf{P}(\text{two-arm event in } B(x, R) \text{ at level } \ell = \|z - x\|).\end{aligned}$$

We need to estimate this probability.

Critical radius

It turns out that this problem is easily solved in some cases :

- **If ℓ is small**, the component $C(x)$ quickly "extincts", and **the radius R is very small with high probability**.
- **If ℓ is large**, the components $C(x)$ and $C(z)$ are unlikely to stay disconnected for very long, here again **R is small**.
- There is a **critical value ℓ^***, which is also the **continuum percolation threshold**, around which a good uniform estimate cannot be obtained.

Unfortunately, several (random) z, and therefore several (random) ℓ, have to be tested. A "two-arm exponent β" is such that

$$\mathbf{P}(\text{ two-arm event in } B(x,R) \text{ at level } \ell) \leqslant cR^{-d\beta},$$

for ℓ uniformly in some interval $[\ell^* - \varepsilon, \ell^* + \varepsilon]$.

Berry-Essèen bounds

- In dimension 2, Chatterjee manages to exhibit such a positive $\beta > 0$. He then obtains **Berry-Essen bounds** in $n^{-\frac{\beta}{\beta+p}}$, where $p > 1$ is arbitrary (with an ad-hoc method).
- In dimension $d \geqslant 3$, he obtains

$$\mathbf{P}(\text{ two-arm event in } B(x,R) \text{ at level } \ell) \leqslant C \log(n)^{-d/2},$$

which gives a **Berry-Esseen bound** in $\log(n)^{-d/8p}$.

Work in progress : We use the general bounds obtained with second order derivatives to generalise his results to **binomial input** and **Kolmogorov distance**.

Number of connected components

- X : random point process
- F : **union of balls** centred in X with random radii/critical radius

$$\varphi(X) = \#\{\text{connected components of } F\}.$$

- Then

$$D^2_{x,y}\varphi(X) \neq 0$$

if x and y are two "**breaking points**" of a connected component of F.

- Let $x \in X, R > 0$. A **two-arm event** is realized in $B(x,R)$ if removing x cuts its connected component in 2 components that touch the boundary. If such an event is not realized, $D^2_{x,y}\varphi(X) = 0$ for any y outside $B(x,R)$.