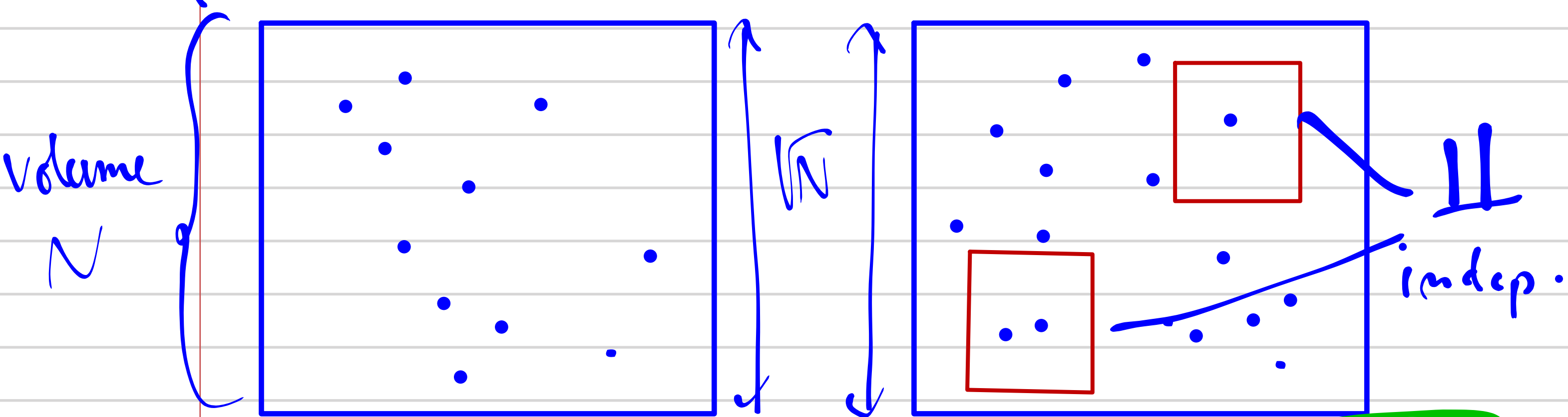
Stabilizing Functionals & 2^d order Poincaré Inequalities

I - Some examples

W_N : Window of $\mathbb{R}^d$ of volume N

X : Point process $\subset W_N$

$(X_N = X \cap W_N)$

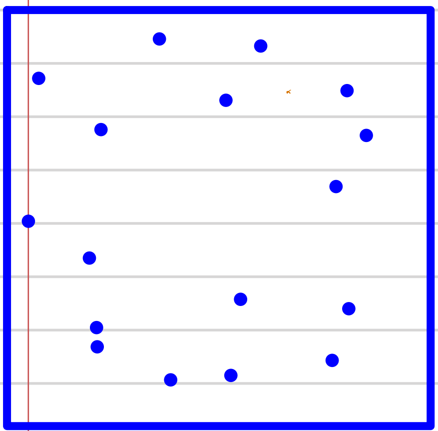


Binomial
N points iid uniform

Poisson
$\mathcal{P}(N)$ points iid uniform

$F = F(X, W_N)$ Random Functional, L^2, $\in \mathbb{R}$

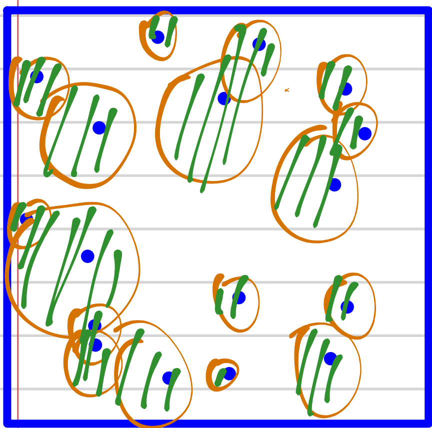
$$F(X,W_N) = \sum_{x\in X_N} \underbrace{\xi(x, X, W_N)}_{\text{"score function"}}$$



$F(X,W_N) = \# X \cap W_N$

Number of points

$\xi(x, X, W_N) = 1$

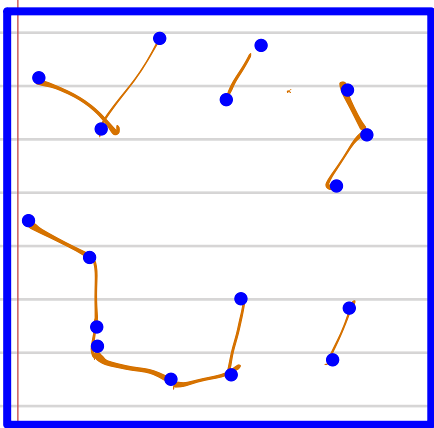


$$F(X,W_N) = \mathcal{L}^d\Big(\bigcup_{x\in X} B(x, R_x)\Big) \quad R_x \text{ iid}$$

Boolean Model (Volume, ...)

$\xi(x) = \mathcal{L}^d(\quad)$

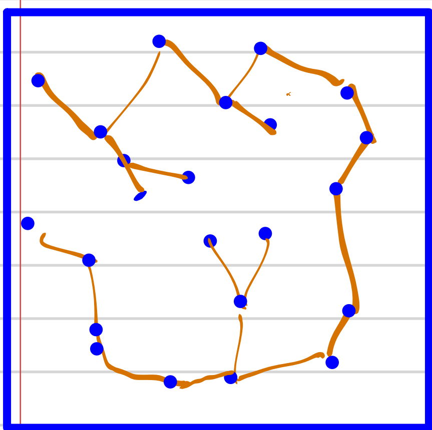
closest points



Nearest neighbour graph

$F(X,W_N)$ = # edges, Length, ...

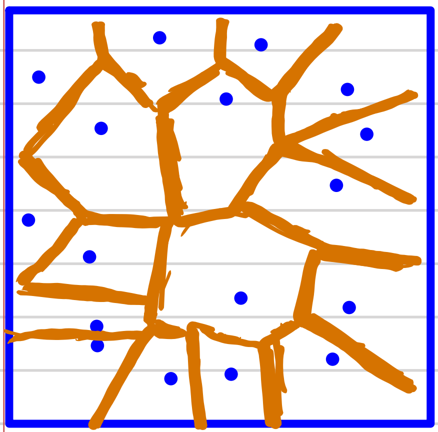
$\xi(x) = \frac{1}{2} \#\,\text{edges}\,(x)$



Minimal Spanning Tree

$F(X)$ = length, Weighted length

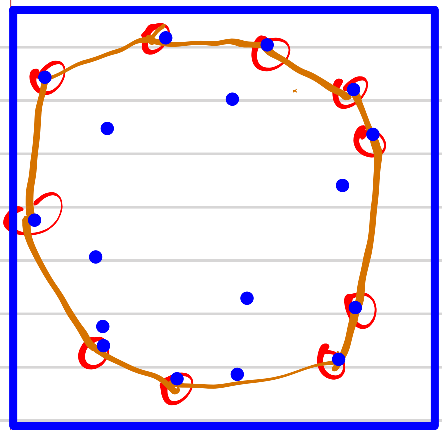
$$\xi(x) = \sum_{e\in \text{edges}(x)} \frac{w(e)}{2}, \ \ldots$$



Voronoi Tesselation

$F(X)$ = length, # vertices, ...

$\xi(x) = \frac{1}{2}\,\text{length}(\text{cell}(x))$, ...

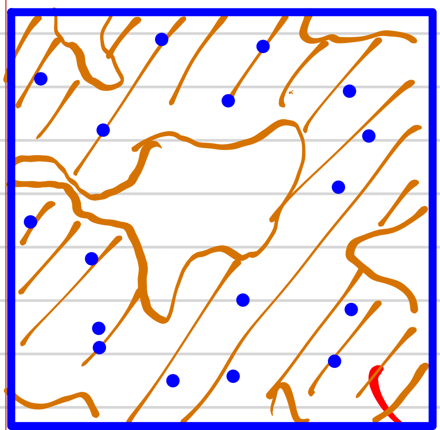


Convex hull

$F(X)$ = Volume, perimeter, ...

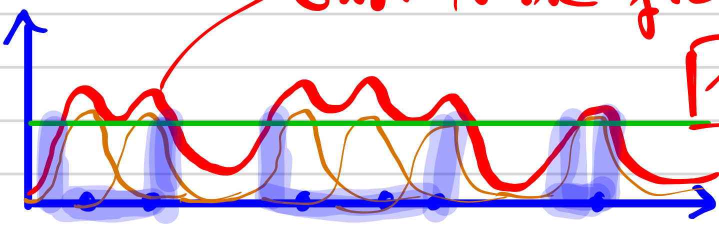
$\xi(x) = 1(x \text{ "extremal"})\,W(x)$

"Surface order scaling"



Shot Noise field

1D



excursion set

- Volume
- Perimeter
- # connected components

2D

$\xi(x) = ?$

II - Stabilization

II - Stabilization

The score function ξ stabilizes on input X if $\forall x \in X, \exists R_x > 0$ such that $\xi(x, X) = \xi(x, X \cap B(x, R_x))$

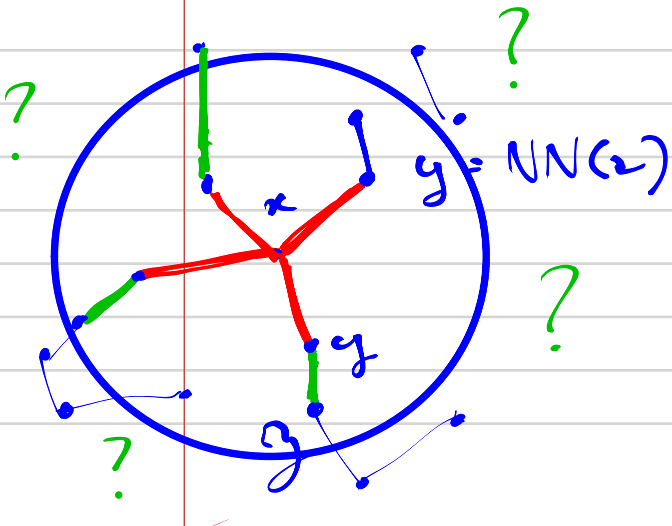
$$= \xi(x, X \cap B(x, R_x) \cup A)$$

for $A \subset B(x, R_x)^c$

$\boxed{\text{The contribution of each } x \in X \text{ depends on its neighbourhood}}$

Examples

- $F(X, W_N) = \# X \cap W_N$, $R_x = 0$
- $F(x, W_N) = \#$ Nearest neighbour graph



$R_x = \mathrm{Max}\ \|x - z\|$ such that $z \in B_{NN}(x, z)$

N.N. distance

Exponential stabilization :

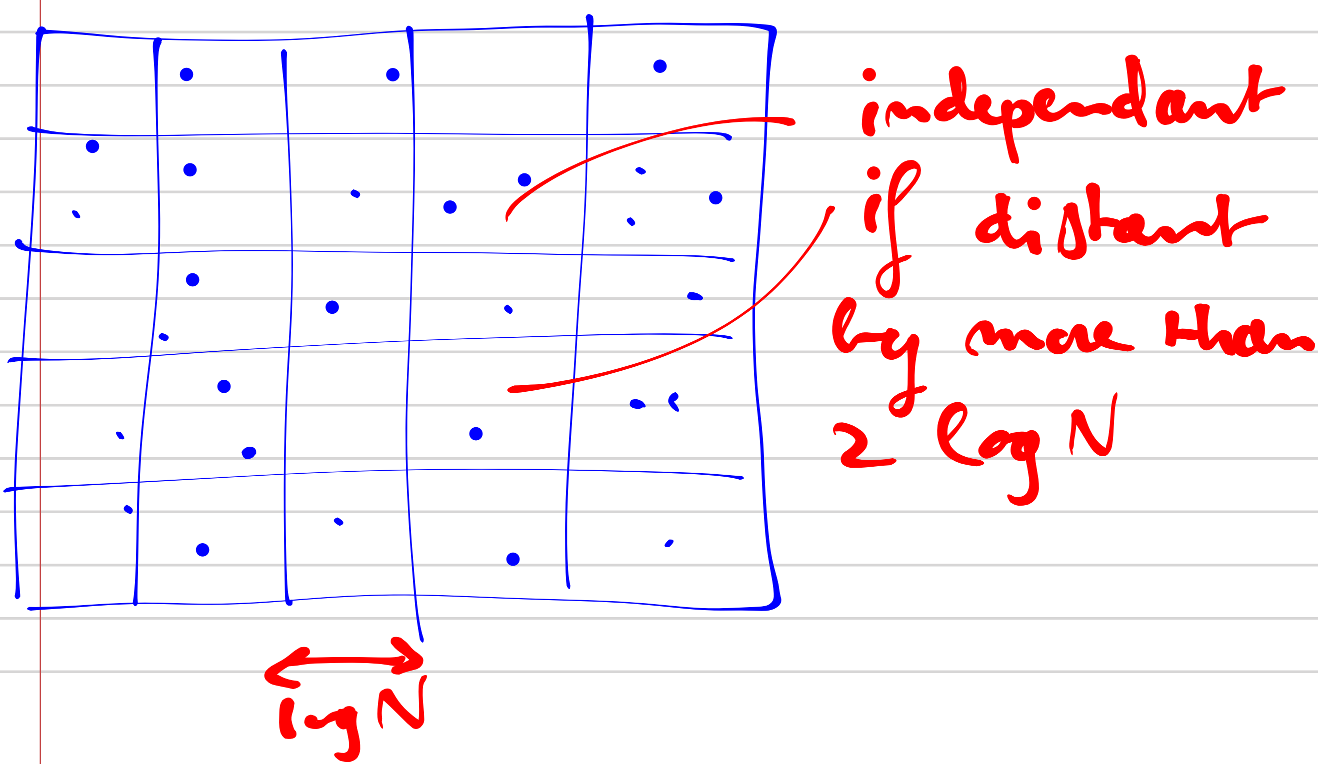
$$\sup_N \sup_{x \in W_N} \mathbb{P}(R_x(x, W_N) \geq t) \leq C e^{-ct^{\gamma}}$$

Asymptotic analysis

Bad event

$$\Omega_N = \{ \sup_{x \in X} R_x > \log N \}$$

If Ω_N^c is satisfied, $F(X, W_N)$ behaves like a sum of m-dependant variables



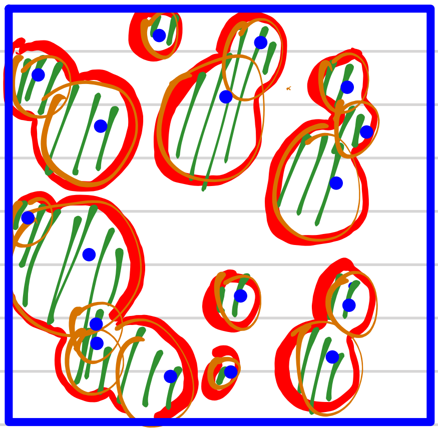
Polynomial stabilisation

$$\sup_{x,w} \mathbb{P}(R_x > t) \leq c(1+t)^{-\alpha} \quad (\text{large } \alpha)$$

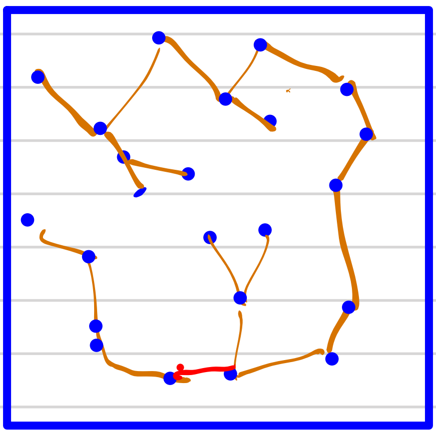
Weak stabilisation:

almost surely,

$$F(X \cup \{0\}, W_N) - F(X, W_N) \xrightarrow[N\to\infty]{} : \Delta_0$$



Betti numbers of Boolean Model



Length of Minimal Spanning Tree

First use of Stabilization by Kesten & Lee '96 to prove a CLT for the MST

Results (Penrose, Yukich, ...)
2003+

Assumptions

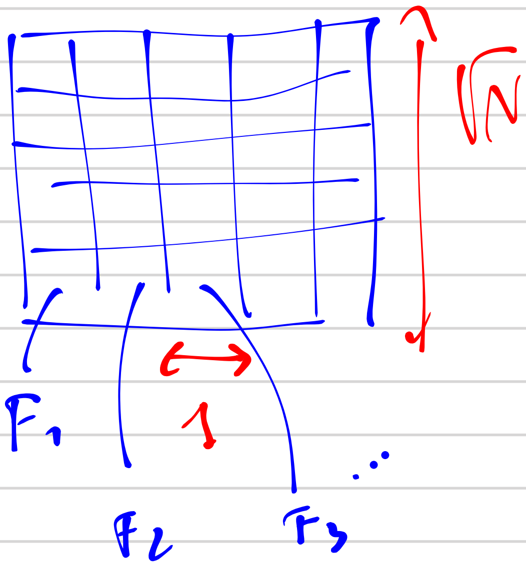
- (Moment $p>4$) $\sup_{x,w} \mathbb{E}\left[\xi(x, \mathcal{X}, w)^p\right] < \infty$
- Stabilization
- Non-degenerate variance:
  $\mathrm{Var}(f_0) > 0 \Rightarrow \mathrm{Var}(F_N) \geq cN$

Then:

- LLN
- Linear Variance
- CLT (Good speed if exp. stab.)

- NN graphs
- Voronoi tessellations
- RSA
- ...

Dependency graphs



Conditionally on Ω_N, each F_i is independant of all but $O(\log N)$ other F_j

Then use Stein's Method . . .

III - Stein's Method

III - Stein's Method

Idea If $N \sim \mathcal{N}(0,1)$, for f test function, $\mathbb{E}[N f(N)] = \mathbb{E}[f'(N)]$

(Integration By Parts)

$\Rightarrow$ For X close to $\mathcal{N}(0,1)$,
$$\left| \mathbb{E}\left[X f(X)\right] - \mathbb{E}[f'(X)] \right| \simeq 0$$

Adapted to Wasserstein Distance

$$d_W(X,N) := \sup_{h \text{ 1-Lip.}} \left| \mathbb{E}[h(X)] - \mathbb{E}[h(N)] \right|$$

Magic: The solution f_h of the ODE

$$h(x) - \mathbb{E}[h(N)] = x h(x) - h'(x)$$

satisfies: $\sup_{h \text{ 1-Lip.}} \|f_h'\|_\infty, \|f_h''\|_\infty < \infty$

$$\to d_W(X,N) \leq \sup_{\|f'\|_\infty, \|f''\|_\infty < C} \left| \mathbb{E}[X f(X) - f'(X)] \right|$$

Idea #2 : Find T such that

- $\mathbb{E}\, T = 1$
- $\mathrm{Var}(T)$ Small
- $\mathbb{E}[X f(X)] \simeq \mathbb{E}[T f'(X)]$

$\Rightarrow$ · Dependency graphs
· Size-Bias couplings
· etc...

Malliavin Calculus

$$\bullet\ T = \text{"} \langle DF, -DL^{-1}F \rangle_{L^2(W)} \text{"}$$

$$= \int_W D_x F \left(-D_x (L^{-1} F)\right) dx$$

Nourdin, Peccati, Nualart, ...
Gaussian World

Peccati, Solé, Utzet, Taqqu, Last, ...
Poisson World

Malliavin Operators

X: Poisson input, $F(X)$: L^2 functional s

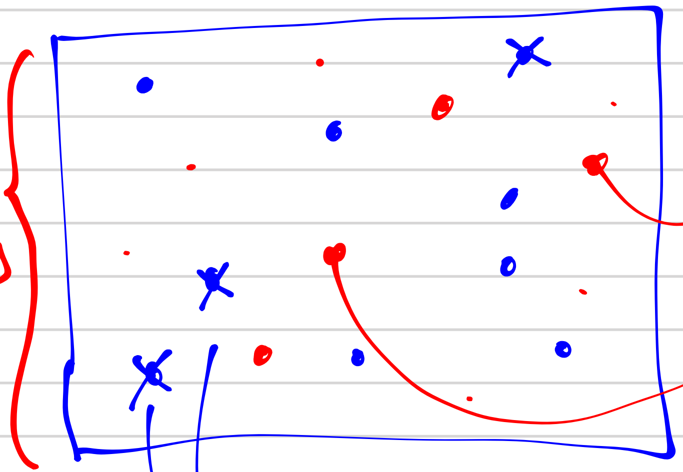
- $DF = x \mapsto F(X \cup \{x\}) - F(X)$
- Semi-group (X': indep. copy)

$$P_t F = \mathbb{E}[F(X^t \cup X'_{1-t}) | X]$$

Remove each point indep^ly with proba. t

Add each point with proba $1-t$

Glauber dynamics



$\cdot X'$ added (Proba $= 1-t$)

$\cdot X$ Removed (Proba $= t$)

$$P_0 F(X) = F(X), \quad P_\infty F(X) = \mathbb{E} F(X)$$

Generator:

$$LF(X) := \frac{d}{dt} P_t F(X)$$

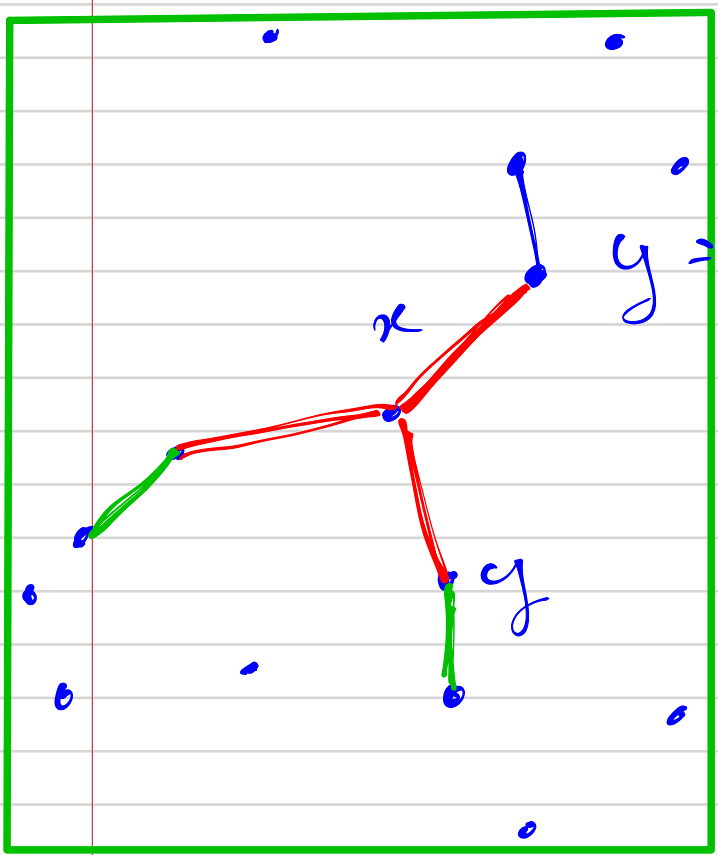
Inverse: L^{-1} (defined if $\mathbb{E}F(X) = 0$)

Examples

- $F(X) = \# X \cap W$

  $D_x F(X) = 1$ for $x \in W$

- $F(X) = \text{length}\,(NN \text{ graph})$



$$D_x F(X) = \| x - NN(x) \| + \sum_{y:\, x = NN(y)} \left(\| y - x \| - \| y - NN^2(y) \| \right)$$

$$\mathbb{E}\left[D_x F^2 \right] \sim C!$$

1st order Poincaré inequality

$$F(X) - \mathbb{E} F(X) = -\int_0^\infty \frac{d}{dt} P_t F(X)\, dt$$

$P_0 F(X)$, $P_\infty F(X)$

$$= \int_0^\infty -L P_t F(X)\, dt$$

$$\operatorname{Var} F(X) = \mathbb{E}\, F(X)\big(F(X) - \mathbb{E}(F(X))\big)$$

$$= -\int_0^\infty \mathbb{E}\left[F(X) L P_t F(X)\right] dt$$

$$(IPP) = \mathbb{E}\left[\int_0^\infty e^{-t} \langle D_x F(X), P_t D_x F(X)\rangle\, dt\right]$$

$$\boxed{\operatorname{Var} F(X) \leq \int_{W_N} \mathbb{E}\, |D_x F|^2\, dx}$$

<u>Contractivity</u>: $\mathbb{E}\left[|P_t F|^p\right] \leq \mathbb{E}\left[|F|^p\right]$

- <u>Examples</u>
  - $F(X) = \# X \cap W_N$, $D_x F = 1$, $\operatorname{Var}(F(X)) = N$
  - $\xi(x)$ is exp. stabilizing
    - Nearest neighbour graph
    - Boolean Model
    - Voronoi tesselation

$\operatorname{Var} F(X) \sim N$

"linear variance"

$$\mathbb{E}\, |D_x F|^2 \sim \mathbb{E}\left[\xi(x)\right]^2 \sim C^t$$

Distance with $N \sim \mathcal{N}(0,1)$

STEIN

$$\bullet\ d_W(F,N) = \sup_{h\ 1\text{-Lip.}} \left| \mathbb{E}[h(F)] - \mathbb{E}[h(N)] \right|$$

$$= \sup_{f : \|f'\|, \|f''\| < c} \left| \mathbb{E}[F f(F) - f'(F)] \right|$$

CHAIN RULE

$$\bullet\ D_x f(F) = f(F(X \cup \{x\})) - f(F(X))$$

$$= f(F(X) + D_x F) - f(F(X))$$

$$= f'(F) D_x F + O(D_x F^2)$$

IPP

$$\bullet\ \mathbb{E}[F f(F)] = \mathbb{E}[\langle D f(F), -DL^{-1} F \rangle]$$

$$\bullet\ \left| \mathbb{E}[f'(F) - F f(F)] \right|$$

$$\leq \Big| \mathbb{E}\Big[\underbrace{f'(F)}_{\leq c} \big(\overbrace{1 - \langle DF, -DL^{-1} F \rangle}^{I}\big)\Big]\Big|$$

$$+ \mathbb{E}\left[O(D_x F^2) \| D_x L^{-1} F \|_\infty \right]$$

Fueled by Wiener-Ito decomposition

Peccati, Solé, Utzet, Taqqu 2009
Leads to

$$d_W(F,N) \le c\left[\mathrm{Var}(T) + \int_W \mathbb{E}\, D_x F^2 \, |D_x L^{-1} F| \, dx\right]$$

→ Can deal with "U-Statistics"
↳ LR + Peccati '12, Reitzner, Schulte

Problem: Geometric control over $L^{-1}F$

IV - 2d Order Poincaré inequality

→ Bound in terms of the 2d order derivative (cf. Chatterjee '08 for Gaussian vectors)

$$D^2_{x,y} F(x) = D_x(D_y F(x)) = D_y(D_x F(x))$$

How much is x's contribution affected by the presence of y?

Examples

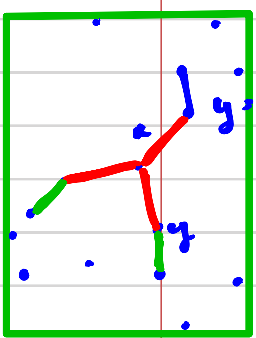
- $F(X) = \# X \cap K$

  $D_x F(X) = 1$ (whether $y \in X$ or not)

  $\to D^2_{xy} F(X) = 0$

- $F(X) = \text{Length (NN graph)}$ 2^d N. N.

$$D_x F(X) = \|x - NN(x)\| + \sum_{y:\, y = NN(x)} \left[\|x-y\| - \|y - NN_2(y)\| \right]$$

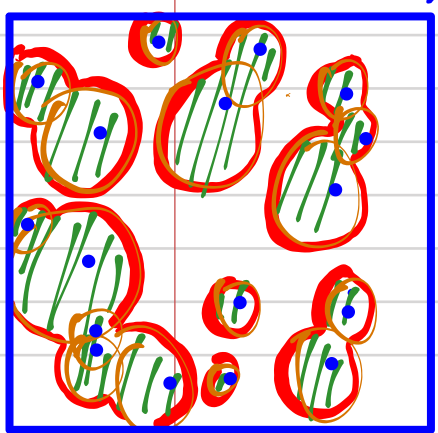


$D^2_{xy} F(X)$ if y is "far from x"

NN distance

$$|D^2_{xy} F(X)| \sim D_x F(X)\, 1_{\{y \in B_{NN}(x,3)\}}$$

- $F(X) = \text{Volume (Boolean Model}(X))$



$D^2_{xy} F = 0$ if

$B(x, R_x) \cap B(y, R_y) = \emptyset$

2^d order PI (Last, Peccati, Schulte 2016)

$$\boxed{d_W\left(\frac{F}{\sigma}, \mathcal{N}\right) \leq \sqrt{\iiint \sqrt{\mathbb{E}(D_{x_1}F)^2(D_{x_2}F)^2} \times \sqrt{\mathbb{E}(D^2_{x_1x_2}F)^2(D^2_{x_2x_3}F)^2}\, dx_1\, dx_2\, dx_3} \times \frac{1}{\sigma^4}}$$

$$\sim \sqrt{\iiint_{W^3} \gamma(x_1-x_2)\,\gamma(x_2-x_3)\,\frac{dx_1\,dx_2\,dx_3}{|W|^2}}$$

$$\boxed{\operatorname{Var} F = \sigma^2} \sim |W|$$

$$\sim \sqrt{\int_W C \cdot dx_1} \sim \sqrt{\frac{|W|}{|W|^2}} \sim \frac{1}{\sqrt{|W|}}$$

$$+ \sqrt{\iiint \mathbb{E}(D^2_{x_1x_3}F)^2(D^2_{x_2x_3}F)^2\, dx_1\, dx_2\, dx_3} \times \frac{1}{\sigma^4}$$

$$+ \int \mathbb{E}\,|D_x F|^3\, dx \times \frac{1}{\sigma^3}$$

Based on Mehler Formula:

$$L^{-1}F = -\int_0^\infty P_{e^{-t}}F\, dt = -\int_0^1 s^{-1} P_s F\, ds$$

Also works for combinatorial problems

Binomial input: LR + Peccati 2016

How to bound $\mathrm{Var}(T)$?

$$\boxed{T = \int_W D_x F \cdot D_x L^{-1} F \cdot dx \ ?}$$

Use 1st order PI!

$$\mathrm{Var}(T) \leq \int \mathbb{E}\left(D_y T\right)^2 dy$$

$$\leq \iiint_{W^3} \mathbb{E}\Big[D_y\left(D_x F \, D_x L^{-1} F\right) \times D_y\left(D_z F \, D_z L^{-1} F\right)\Big] dx\,dy\,dz$$

PRODUCT FORMULA: $D(FG) = (DF)G + F(DG) + (DF)(DG)$

$\rightarrow$ Appearance of terms $D^2_{xy} F$

Use Mehler formula for $DL^{-1}F$

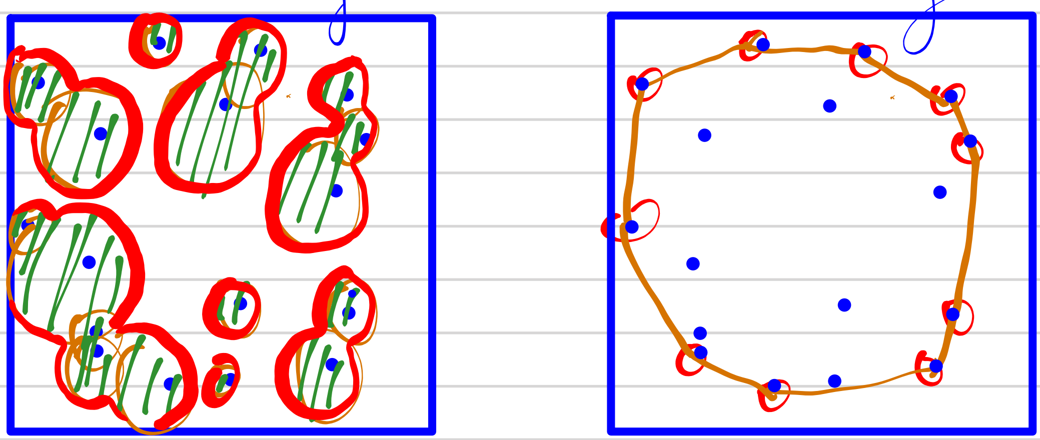
LR, Schulte, Yukich '19

Use of 2^d order Poincaré inequalities over $\begin{cases} \text{Poisson input} \\ \text{Binomial input} \end{cases}$ with marks

On $\begin{cases} \mathbb{R}^d \\ \text{Smooth manifold} \\ \text{Regular Metric space} \end{cases}$

under exponential stabilization

of volume order / surface order
scaling / scaling



→ CLT with optimal speed

for $\begin{cases} \text{Wasserstein distance} \\ d_K(F, N) = \sup_t |\mathbb{P}(F > t) - \mathbb{P}(N > t)| \end{cases}$

Kolmogorov

Schulte, Trapp '22

Lower bound on the variance

If $\int_{W^2} \mathbb{E}\left[(D^2_{x,y} F)^2 \, dx dy \right]$

$$\leq \alpha \int_W \mathbb{E}\left(D^2_x F\right)^2 dx =: M$$
$< \infty$

Then 1st order Poincaré inequality is sharp:

$$\mathrm{Var}(F) \geq \frac{4}{(\alpha+2)^2 \alpha} \times M$$

Thäle, Trauthwein '22

Weaker moment assumption

$$\left(\mathbb{E}(D^2_{xy} F)^4\right)^{\frac{1}{4}} \to \left(\mathbb{E}(D^2_{x,y} F)^{2+\varepsilon}\right)^{\frac{1}{2+\varepsilon}}$$

Improvement on 2d order P.I.

LR, Peccati, Yang '22

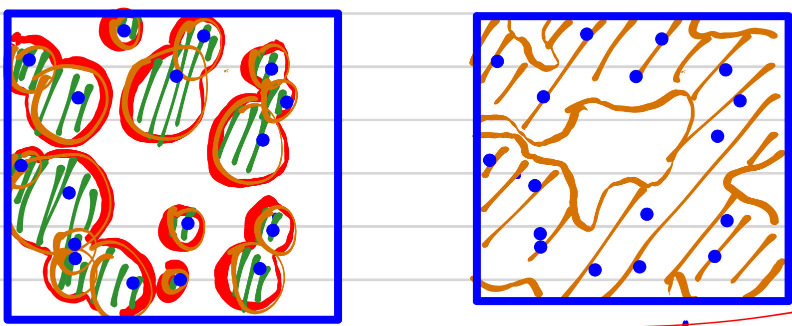
2 scale stabilization

Deal with functionals with long range

(weak stabilization)

- Length of Minimal Spanning Tree
  ↳ Idea of Chatterjee & Sen 2017

- # Connected components (Boolean Model)



- # Connected components of heavy tailed excursions

- Length of "Online NN graph"

<u>Idea</u> Introduce intermediate scale $\ell_N = o(\sqrt{N})$

- Moment assumption $\sup_x \mathbb{E}|D_x F|^p < \infty$
- Non-degenerate variance:

$$\mathrm{Var}(F) \geq aN > 0$$

Then

$$d_W(F,\mathcal{N}) \leq \sqrt{\int \sup_x \mathbb{E}\left[\left|D_x F(X_N) - D_x F(X_N \cap B(x,\ell_N))\right|^{\frac{p-1}{p}}\right] dx} + \frac{\ell_N^{d/2}}{N^{d/2}}$$

- Bound with Kolmogorov distance d_K
- Can include Marks
- Multi-variate versions

Thank you
for your attention!