Geometric statistics of point processes: limit theory and (some) statistical learning

Bartłomiej Błaszczyszyn
[Inria/ENS Paris]

4th Workshop on Cumulants, Concentration and Superconcentration
University of Münster, January 10–11 2019

Geometric marks of point processes

Consider $\Phi = \sum_i \delta_{X_i}$ simple ($X_i \neq X_j$ for $i \neq j$) , stationary point process in $\mathbb{R}^d$. Recall, Φ is a random object on the space $\mathbb{M}$ of locally finite counting measures on $\mathbb{R}^d$ with a suitable σ-field.

Consider $\Phi = \sum_i \delta_{X_i}$ simple ($X_i \neq X_j$ for $i \neq j$) , stationary point process in $\mathbb{R}^d$. Recall, Φ is a random object on the space $\mathbb{M}$ of locally finite counting measures on $\mathbb{R}^d$ with a suitable σ-field.

Consider a measurable, real-valued marking (score) function $\xi = \xi(x, \phi)$ defined on $\mathbb{R}^d \times \mathbb{M}$, with $x \in \phi$. Assume ξ is translation invariant; i.e., $\xi(x + a, \phi + a) = \xi(x, \phi)$ for all $a \in \mathbb{R}^d$ and $\phi \in \mathbb{M}$.

Consider $\Phi = \sum_i \delta_{X_i}$ simple ($X_i \neq X_j$ for $i \neq j$) , stationary point process in $\mathbb{R}^d$. Recall, Φ is a random object on the space $\mathbb{M}$ of locally finite counting measures on $\mathbb{R}^d$ with a suitable σ-field.

Consider a measurable, real-valued marking (score) function $\xi = \xi(x, \phi)$ defined on $\mathbb{R}^d \times \mathbb{M}$, with $x \in \phi$. Assume ξ is translation invariant; i.e., $\xi(x + a, \phi + a) = \xi(x, \phi)$ for all $a \in \mathbb{R}^d$ and $\phi \in \mathbb{M}$.

We will call

$$M_i := \xi(X_i, \Phi)$$

geometric mark of the point X_i of Φ (produced by the marking function ξ).

Consider $\Phi = \sum_i \delta_{X_i}$ simple ($X_i \neq X_j$ for $i \neq j$) , stationary point process in $\mathbb{R}^d$. Recall, Φ is a random object on the space $\mathbb{M}$ of locally finite counting measures on $\mathbb{R}^d$ with a suitable σ-field.

Consider a measurable, real-valued marking (score) function $\xi = \xi(x, \phi)$ defined on $\mathbb{R}^d \times \mathbb{M}$, with $x \in \phi$. Assume ξ is translation invariant; i.e., $\xi(x + a, \phi + a) = \xi(x, \phi)$ for all $a \in \mathbb{R}^d$ and $\phi \in \mathbb{M}$.

We will call

$$M_i := \xi(X_i, \Phi)$$

geometric mark of the point X_i of Φ (produced by the marking function ξ).

Note that $\tilde{\Phi} := \sum_i \delta_{(X_i, M_i)}$ is a stationary marked point process.

Consider $\Phi = \sum_i \delta_{X_i}$ simple ($X_i \neq X_j$ for $i \neq j$) , stationary point process in $\mathbb{R}^d$. Recall, Φ is a random object on the space $\mathbb{M}$ of locally finite counting measures on $\mathbb{R}^d$ with a suitable σ-field.

Consider a measurable, real-valued marking (score) function $\xi = \xi(x, \phi)$ defined on $\mathbb{R}^d \times \mathbb{M}$, with $x \in \phi$. Assume ξ is translation invariant; i.e., $\xi(x + a, \phi + a) = \xi(x, \phi)$ for all $a \in \mathbb{R}^d$ and $\phi \in \mathbb{M}$.

We will call

$$M_i := \xi(X_i, \Phi)$$

geometric mark of the point X_i of Φ (produced by the marking function ξ).

Note that $\tilde{\Phi} := \sum_i \delta_{(X_i, M_i)}$ is a stationary marked point process.

We will sometimes (ab)use the same notation for the weighted point measure $\tilde{\Phi} := \sum_i M_i \delta_{X_i}$, where atoms of Φ are weighted by the values of their marks (possibly signed measure).

Simple examples of geometric marks

- Distance to the nearest neighbour:
  $M_i = R_i := \min\{|X_i - X_j| : x_j \in \Phi, X_i \neq X_i\}$.
- Volume of the Voronoi cell:
  $M_i = |A_i| := |\{y \in \mathbb{R}^d : |y - X_i| \leq \min_{X_j \in \Phi} |y - X_j|\}|$.
- Shot-noise:
  $m_i = S_i := \sum_{i \neq j} \ell(|X_j - X_i|)$ with some response function $\ell(\cdot)$.
- Number of neighbors within distance R:
  shot-noise with response function $\ell(r) = 1(r \leq R)$.

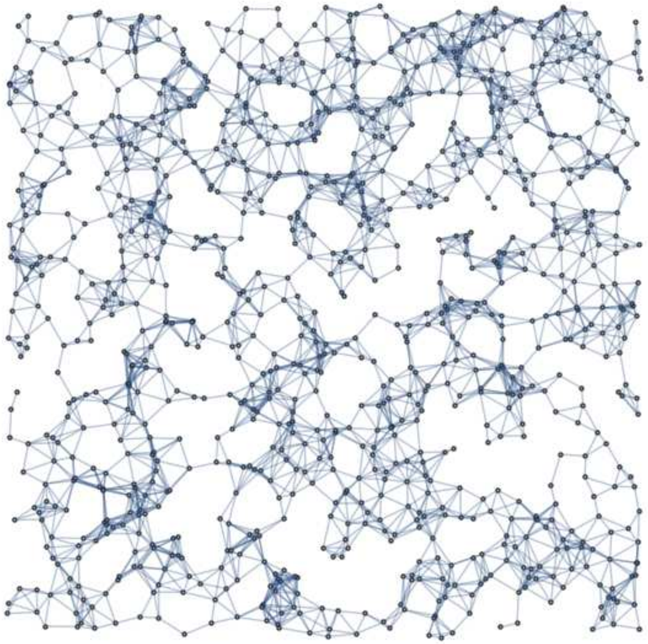
- Distance to the nearest neighbour:
  $M_i = R_i := \min\{|X_i - X_j| : x_j \in \Phi, X_i \neq X_i\}$.
- Volume of the Voronoi cell:
  $M_i = |A_i| := |\{y \in \mathbb{R}^d : |y - X_i| \leq \min_{X_j \in \Phi} |y - X_j|\}|$.
- Shot-noise:
  $m_i = S_i := \sum_{i \neq j} \ell(|X_j - X_i|)$ with some response function $\ell(\cdot)$.
- Number of neighbors within distance R:
  shot-noise with response function $\ell(r) = 1(r \leq R)$.

Many interesting statistics of more complex geometric models can be represented as functions (e.g. sums) of such geometric marks.
For example: ...

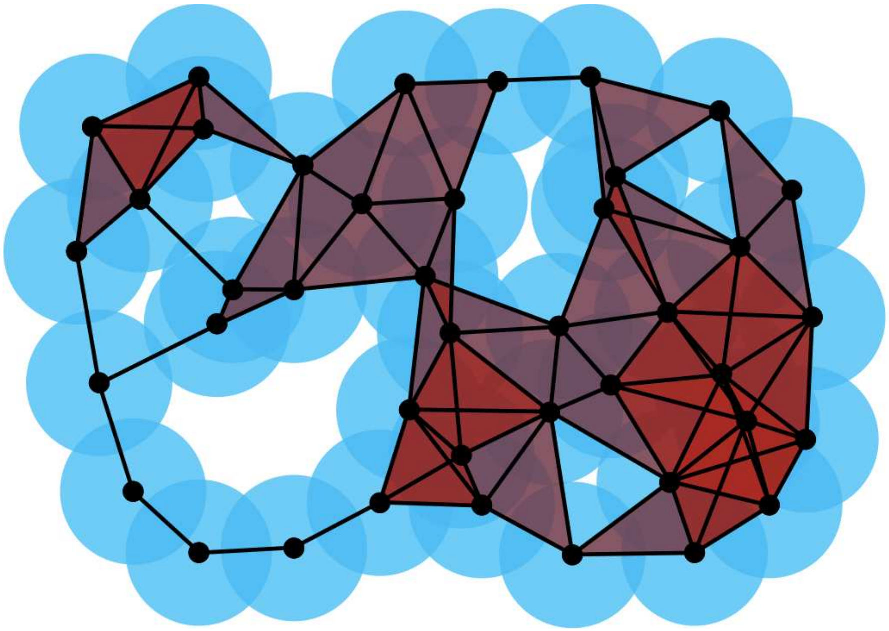
Characteristics of Gilbert graph

For $R > 0$ there exist an edges between any two points $X_i, X_j \in \Phi$ iff $|X_i - X_j| \leq R$. The sum of marks M_i of points in a given bounded window may represent for example:

- Total number of edges:
  $\xi(x, \Phi) = \frac{1}{2}(\Phi(B_x(r)) - 1)$, where $B_x(r)$ is the ball of radius r centered at x
- Total edge length
  $\xi(x, \Phi) = \frac{1}{2} \sum_{y \in \Phi \cap B_x(r)} |x - y|$.
- Sub-graph Γ count
  $\xi(x, \Phi) = \frac{1}{k} \sum_{\substack{y_1, \ldots, y_k \in \Phi \cap B_x(kr) \\ \text{distinct}}} 1(G(\{y_1, \ldots, y_k\}, r) \cong \Gamma)$,
  where Γ is an abstract graph with k vertexes.



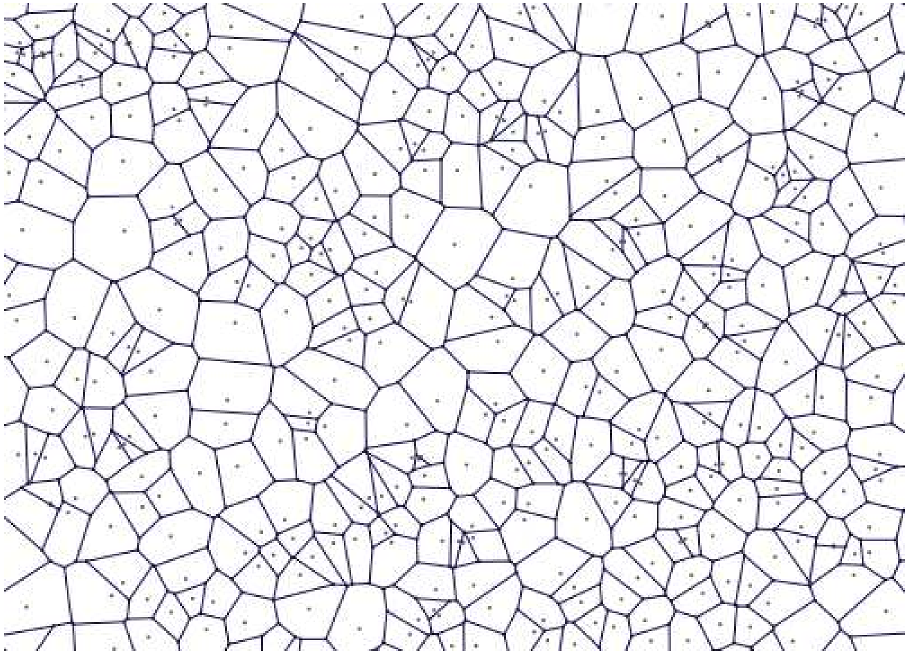
- More involved geometric and topological properties of data encored in Čech complex that is an extension of the Gilbert graph allowing for higher-dimensional edges called facets.



- $\boldsymbol{k}$-covered region volumes in the Boolean model.
- Morse critical points.

Further examples

- Properties of k-nearest neighbor graphs.
- In particular intrinsic volumes of faces of Voronoi tessellations



Further examples

Many engineering characteristics of communication networks, for example cell loads in wireless communication networks capturing in a concise way the quality of service offered by the base stations, whose locations are modeled by the points of $\mathbf{\Phi}$.

Further examples

Many engineering characteristics of communication networks, for example cell loads in wireless communication networks capturing in a concise way the quality of service offered by the base stations, whose locations are modeled by the points of $\mathbf{\Phi}$.

In what follows we shall present two subjects related to geometric marks.

We shall study the asymptotic of the weighted point measure $\tilde{\Phi}$ "compressed" to the unit volume window:

$$\mu_n := \sum_{X_i \in W_n} M_i \delta_{X_i/n^{1/d}},$$

where $W_n = [0, n^{1/d}]^d$.

We shall study the asymptotic of the weighted point measure $\tilde{\Phi}$ "compressed" to the unit volume window:

$$\mu_n := \sum_{X_i \in W_n} M_i \delta_{X_i/n^{1/d}},$$

where $W_n = [0, n^{1/d}]^d$.

More precisely, expectation asymptotics, variance asymptotics, and central limit theorems (CLT) for the integrals

$$\mu_n(f) := \int_{W_1} f(x)\, \mu_n(\mathrm{d}x)$$

of μ_n with respect to test functions f.

We shall study the asymptotic of the weighted point measure $\tilde{\Phi}$ "compressed" to the unit volume window:

$$\mu_n := \sum_{X_i \in W_n} M_i \delta_{X_i/n^{1/d}},$$

where $W_n = [0, n^{1/d}]^d$.
More precisely, expectation asymptotics, variance asymptotics, and central limit theorems (CLT) for the integrals

$$\mu_n(f) := \int_{W_1} f(x)\, \mu_n(\mathrm{d}x)$$

of μ_n with respect to test functions f.

$\Rightarrow$ We shall define and use some (new?) mixing condition for point processes expressed in terms of the correlations functions and related to the cumulant measures.

Subject II: Statistical learning of geometric marks

We shall address (in a more computer-science way) the following "practical" problem:
Suppose the marking (score) function ξ is not know. One aims at learning this function from the examples of marked point patterns, in order to predict the marks of new point patterns.

We shall address (in a more computer-science way) the following "practical" problem:
Suppose the marking (score) function ξ is not know. One aims at learning this function from the examples of marked point patterns, in order to predict the marks of new point patterns.

$\Rightarrow$ We shall use some new "representation" of the weighted measure $\tilde{\mathbf{\Phi}}$ via its scattering moments. These are new operators based on wavelet transforms computed at different scales. They have many interesting properties studied up to now mostly in empirical way.

We shall address (in a more computer-science way) the following "practical" problem:
Suppose the marking (score) function ξ is not know. One aims at learning this function from the examples of marked point patterns, in order to predict the marks of new point patterns.

$\Rightarrow$ We shall use some new "representation" of the weighted measure $\tilde{\mathbf{\Phi}}$ via its scattering moments. These are new operators based on wavelet transforms computed at different scales. They have many interesting properties studied up to now mostly in empirical way.

Mathematically, we shall study the asymptotic of the scattering moments of $\tilde{\mathbf{\Phi}}$ as the scale grows infinitely small or large.
The CLT (established in the first part) is useful for large scales, and suggests how to estimate the variance asymptotic of $\tilde{\mathbf{\Phi}}$, in particular to detect the hyperuniformity or hyperfluctuations (to be exaplained).

Limit theory for geometric statistics of point processes having fast decay of correlations

based on a joint work with

D. Yogeshwaran [ISI Bangalore,]
Joe Yukich [Lehigh University, Bethlehem]

LLN and CLT for iid rv's — classic results

$Y_1, Y_2, \ldots$ independent, identically distributed rv's, $S_n = \sum_{i=1}^{n} Y_i$.

□ (Weak LLN) If $\mu := \mathbf{E}(Y_1)$ finite then

$$\frac{S_n}{n} \xrightarrow{\mathbf{P}} \mu \qquad n \to \infty.$$

□ (CLT) If $\sigma^2 := \mathrm{Var}\,(Y_1) < \infty$ then

$$\frac{S_n - n\mu}{\sqrt{\mathrm{Var}\,(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \Rightarrow N(0,1)\,.$$

Looking for LLN and LCT for dependent rv's.

Dependent sums in geometric context

$\mathcal{P} = \{X_i\} \subset \mathbb{R}^d$ — point process, geometric input data.
Consider sums

$$S_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i, \mathcal{P}_n)\,,$$

where

- $\mathcal{P}_n = \mathcal{P} \cap W_n$ data truncated to the observation window $W_n = [0, n^{1/d}]^d$ of volume n.,
- $\xi(X_i, \mathcal{P}_n)$ score function of the relative position of point X_i in $\mathcal{P}_n$.

Dependent sums in geometric context

$\mathcal{P} = \{X_i\} \subset \mathbb{R}^d$ — geometric input data

$$S_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i, \mathcal{P}_n) \, .$$

Two-fold dependence of the summands in S_n:

- via the score function ξ,
- via possibly correlated data $\mathcal{P}$.

Dependent sums in geometric context

$\mathcal{P} = \{X_i\} \subset \mathbb{R}^d$ — geometric input data

$$S_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i, \mathcal{P}_n)\,.$$

Two-fold dependence of the summands in S_n:

- via the score function ξ,
- via possibly correlated data $\mathcal{P}$.

Special cases:

- $\xi(x, \mathcal{P}) = \xi(x)$ — linear score function, no dependence via ξ,
- Poisson process $\mathcal{P}$ — independent data.

Poisson process — independent data

Assume $\mathcal{P} = \{X_i\}$ form homogeneous Poisson point process on $\mathbb{R}^d$ of intensity λ.
Remind:

- □ $\mathcal{P}(B)$ — number of points in B is Poisson $(\lambda|B|)$ rv,
- □ $\mathcal{P}(B_1), \ldots \mathcal{P}(B_k)$ are independent rv's.

Poisson process — independent data

Assume $\mathcal{P} = \{X_i\}$ form homogeneous Poisson point process on $\mathbb{R}^d$ of intensity λ.

Remind:

- □ $\mathcal{P}(B)$ — number of points in B is Poisson $(\lambda|B|)$ rv,
- □ $\mathcal{P}(B_1), \ldots \mathcal{P}(B_k)$ are independent rv's.

Fact: given $\mathcal{P}(W_n) = n$, points of $\mathcal{P}$ in W_n, are independent, identically distributed (uniform on W_n).

Poisson process — independent data

Assume $\mathcal{P} = \{X_i\}$ form homogeneous Poisson point process on $\mathbb{R}^d$ of intensity λ.
Remind:

- □ $\mathcal{P}(B)$ — number of points in B is Poisson $(\lambda|B|)$ rv,
- □ $\mathcal{P}(B_1), \ldots \mathcal{P}(B_k)$ are independent rv's.

Fact: given $\mathcal{P}(W_n) = n$, points of $\mathcal{P}$ in W_n, are independent, identically distributed (uniform on W_n).

In case of linear score functions of Poisson process $S_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i)$ are Poisson-randomized sums of iid rv's hence LLN and CLT reduce to (almost) classic setting.

Stabilizing score functions

To control the dependence via the score function ξ define:

- $\mathcal{R} = \mathcal{R}(x, \mathcal{X})$ is a stabilization radius of ξ on data $\mathcal{X}$ at $x \in \mathcal{X}$ if any modification of data $\mathcal{X}$ outside the ball $B_x(\mathcal{R})$ of radius $\mathcal{R}$ centered at x does not change the value of $\xi(x, \mathcal{X})$

$$\xi\left(x, \mathcal{X} \cap B_x(\mathcal{R})\right) = \xi\left(x, \left(\mathcal{X} \cap B_x(\mathcal{R})\right) \cup \left(\mathcal{X}' \cap B_x^c(\mathcal{R})\right)\right)$$

for any data set $\mathcal{X}'$, with B_x^c denoting the complement of the ball B_x.

Exponential stabilization on Poisson data

Consider translation invariant score function $\xi(x, \mathcal{X}) = \xi(x + a, \mathcal{X} + a)$ for all $a \in \mathbb{R}^d$.

- ξ is exponentially stabilizing on Poisson input if

$$\sup_{1 \le n \le \infty} \mathbf{P}\left(\mathcal{R}(0, \mathcal{P}_n \cup \{0\}) > r \right) \le c_1 e^{-c_2 r}$$

for some constants $c_1 < \infty$, $c_2 > 0$ and all $r \ge 0$.

Theorem ...[Penrose & Yukich (2003)]...

Assume ξ is exponentially stabilizing on Poisson process with intensity λ and consider $S_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i, \mathcal{P}_n)$.

□ (Mean) If $\mathbf{E}(\xi^p(0, \mathcal{P} \cup \{0\})) < \infty$ for some $p > 1$ then

$$\frac{\mathbf{E}(S_n)}{n} \to \lambda \mathbf{E}(\xi(0, \mathcal{P} \cup \{0\})) \qquad n \to \infty.$$

...

Theorem ...[Baryshnikov &Yukich (2005)]...

□ (Variance) If $\mathbf{E}(\xi^p(0, \mathcal{P} \cup \{0\})) < \infty$ for some $p > 2$ then

$$\frac{\mathrm{Var}\,(S_n)}{n} \to \sigma^2(\xi) < \infty \qquad n \to \infty\,,$$

where

$$\sigma^2(\xi) = \lambda \mathbf{E}\big(\xi^2(0, \mathcal{P} \cup \{0\})\big) + \lambda^2 \int_{\mathbb{R}^d} \mathbf{E}(\xi(0, \mathcal{P} \cup \{0, x\})\xi(x, \mathcal{P} \cup \{0, x\})) - \Big(\mathbf{E}(\xi(0, \mathcal{P} \cup \{0\}))\Big)^2 \,\mathrm{d}x\,.$$

...

Limit theory for Poisson data cont'd

Theorem ...[Baryshnikov-Yukich (2005)]...

- (Mean) & (Variance) imply weak LLN for S_n.
- If moreover $\sigma^2(\xi) > 0$ then the CLT for S_n holds.

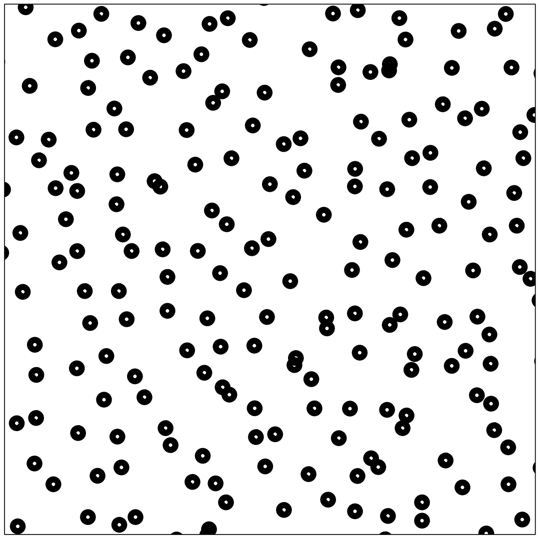
Limit theory for Poisson data cont'd

Theorem ...[Baryshnikov-Yukich (2005)]...

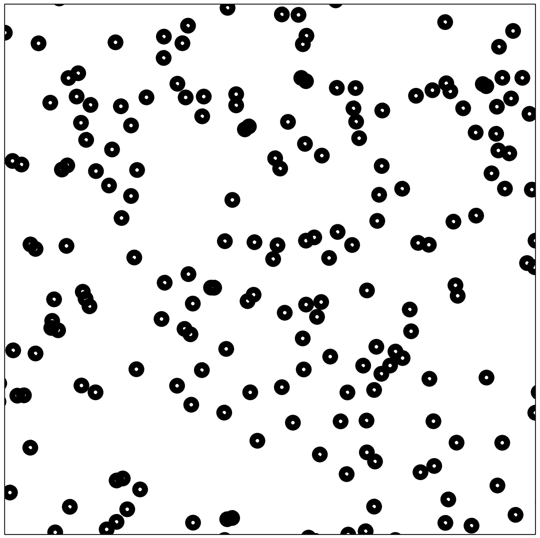
- (Mean) & (Variance) imply weak LLN for S_n.
- If moreover $\sigma^2(\xi) > 0$ then the CLT for S_n holds.

Goal: extend the theory to correlated data.

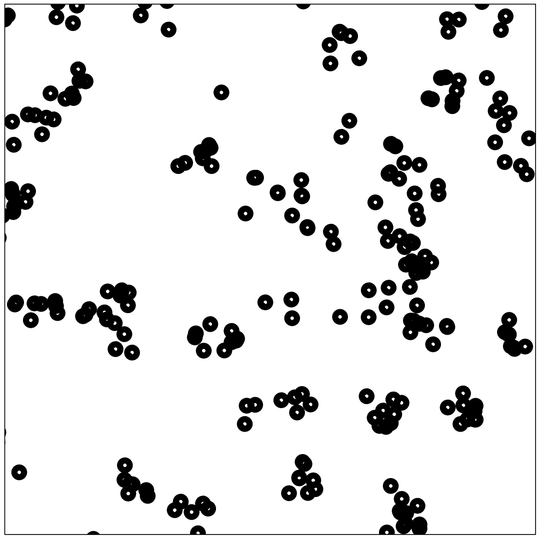
Sample data realizations and their models



negative correlation
(hyper-)regulararity
rigidity



independence



positive correlation
(hyper-)fluctuation
clustering

- determinantal,
- α-determinantal $\alpha < 0$,
- Gibbsian and other hard-hore,

- Poisson
- Binomial

- permanental ,
- α-permanental $\alpha > 0$,
- cluster processes,
- Cox processes

Outline of the remaining part of the talk

- controlling correlations of points
- examples of point processes with fast decaying of correlations
- main results
- comments on previous results
- proof idea

Mixing point processes

- (Usual) mixing of a point process roughly says that configurations of points in distant regions U, V are asymptotically independent

$$\mathbf{E}(\phi(\mathcal{P} \cap U) \cdot \psi(\mathcal{P} \cap V)) - \mathbf{E}(\phi(\mathcal{P} \cap U))\, \mathbf{E}(\psi(\mathcal{P} \cap V)) \to 0$$

as distance$(U, V) \to \infty$,

where ϕ, ψ are two functions of the configuration of points.

Mixing point processes

- (Usual) mixing of a point process roughly says that configurations of points in distant regions U, V are asymptotically independent

$$\mathbf{E}(\phi(\mathcal{P} \cap U) \cdot \psi(\mathcal{P} \cap V)) - \mathbf{E}(\phi(\mathcal{P} \cap U))\, \mathbf{E}(\psi(\mathcal{P} \cap V)) \to 0$$

as distance$(U, V) \to \infty$,

where ϕ, ψ are two functions of the configuration of points.

Mixing implies ergodicity, hence LLN for sums without boundary effects $\tilde{S}_n = \sum_{X_i \in \mathcal{P}_n} \xi(X_i, \mathcal{P})$.
Too weak for LLN with boundary effects and CLT.

Stronger mixing properties

- alpha-mixing Total variation convergence of $(\phi(\mathcal{P} \cap U), \psi(\mathcal{P} \cap V))$ to the independence. Still not enough for CLT.
- Brillinger mixing Reduced cumulant (singed) measures having finite total variation. Directly usable in proofs of CLT for point counts. More difficult to verify for examples of point processes.
- mixing of correlation functions — our approach.

Consider simple point process $\mathcal{P}$ (no multiple points) on $\mathbb{R}^d$;
k-point correlation function $\rho^{(k)}(x_1, ..., x_k)$ of $\mathcal{P}$. Informally

$$\mathsf{P}\left(\mathcal{P}(\mathrm{d}x_1) \geq 1, \ldots, \mathcal{P}(\mathrm{d}x_k) \geq 1\right) = \rho^{(k)}(x_1, ..., x_k)\mathrm{d}x_1 \ldots \mathrm{d}x_k.$$

Consider simple point process $\mathcal{P}$ (no multiple points) on $\mathbb{R}^d$;
k-point correlation function $\rho^{(k)}(x_1, ..., x_k)$ of $\mathcal{P}$. Informally

$$\mathsf{P}\left(\mathcal{P}(\mathrm{d}x_1) \geq 1, \ldots, \mathcal{P}(\mathrm{d}x_k) \geq 1 \right) = \rho^{(k)}(x_1, ..., x_k)\mathrm{d}x_1 \ldots \mathrm{d}x_k.$$

Formally $\rho^{(k)}$ is the density the corresponding factorial moment measure $\alpha^{(k)}(B_1 \times \cdots \times B_k) = \mathsf{E}\Big(\prod_{1 \leq i \leq k} \mathcal{P}(B_i)\Big) = \int_{B_1 \times \cdots \times B_k} \rho^{(k)}(x_1, ..., x_k)\, \mathrm{d}x_1 \ldots \mathrm{d}x_k$, where $B_1, ..., B_k$ are mutually disjoint bounded Borel sets in $\mathbb{R}^d$.

$\{\rho^{(k)}, k \geq 1\}$ characterize the distribution of simple point process having finite some exponential moments.

ω-mixing of correlation functions

Definition. The correlation functions are ω-mixing, with function $\omega = \omega(k, s)$ of $k = 1, 2, \ldots$, $s \geq 0$ if for all $p, q \geq 1$

$$|\rho^{(p+q)}(x_1, \ldots, x_{p+q}) - \rho^{(p)}(x_1, \ldots, x_p)\rho^{(q)}(x_{p+1}, \ldots, x_{p+q})| \leq \omega(p+q, s),$$

where $s := d(\{x_1, \ldots, x_p\}, \{x_{p+1}, \ldots, x_{p+q}\})$ separation distance between $\{x_1, \ldots, x_p\}$ and $\{x_{p+1}, \ldots, x_{p+q}\}$.

ω-mixing of correlation functions

Definition. The correlation functions are ω-mixing, with function $\omega = \omega(k, s)$ of $k = 1, 2, \ldots$, $s \geq 0$ if for all $p, q \geq 1$

$$|\rho^{(p+q)}(x_1, \ldots, x_{p+q}) - \rho^{(p)}(x_1, \ldots, x_p)\rho^{(q)}(x_{p+1}, \ldots, x_{p+q})| \leq \omega(p+q, s),$$

where $s := d(\{x_1, \ldots, x_p\}, \{x_{p+1}, \ldots, x_{p+q}\})$ separation distance between $\{x_1, \ldots, x_p\}$ and $\{x_{p+1}, \ldots, x_{p+q}\}$.

We assume $\omega(k, s) \searrow 0$ when $s \to \infty$ meaning

correlation functions asymptotically factorize (some physicists say "cluster") for large separation distance indicating asymptotic independence of $\mathcal{P}$.

Relations to other mixing properties

Depend on the function ω. For example:

- $\sum_k \frac{\epsilon^k}{k!}\omega(k,s) < \infty$ for some $\epsilon > 0$ and $\to 0$ as $s \to \infty$ implies α-mixing.
- $\omega(k,s) = C_k e^{-s}$ (called fast decay of correlations) implies Brillinger mixing.

Relations to other mixing properties

Some Examples of point processes having fast decay of correlations:

- determinantal and permanental processes with fast decaying kernel $K(x,y) \leq Ce^{-c|x-y|}$, for $C < \infty$, $c > 0$.
- α-permanental and determinantal processes with with fast decaying kernel.
- Zero set of Gaussian entire function,
- some Gibbs point processes,
- processes with finite range dependence (e.g. Matern hard core).

Stabilizing score functions revisited

Definition. The score function ξ is exponentially stabilizing on correlated input $\mathcal{P}$ if

$$\sup_{1 \le n \le \infty} \sup_{x_1, \ldots, x_l \in W_n} \mathbf{P}_{x_1, \ldots, x_l}\big(R^{\xi}(x_1, \mathcal{P}_n) > t\big) \le C e^{-c_l t}$$

for some constants $C < \infty$, $c_l > 0$ and all $t \ge 0$,
where $\mathbf{P}_{x_1, \ldots, x_l}$ are Palm probabilities of $\mathcal{P}$; play the role of conditional probabilities given $\mathcal{P}$ has atoms at $x_1, \ldots, x_l$.

Moment conditions revisited

Definition. Given $p \in [1, \infty)$, say that the pair $(\xi, \mathcal{P})$ satisfies the p-moment condition if

$$\sup_{1 \le n \le \infty} \sup_{1 \le p' \le \lfloor p \rfloor} \sup_{x_1, \ldots, x_{p'} \in W_n} \mathbf{E}_{x_1, \ldots, x_{p'}} \max\{|\xi(x_1, \mathcal{P}_n)|, 1\}^p \le M_p < \infty$$

for some constant $M_p := M_p^{\xi}$, where $\mathbf{E}_{x_1, \ldots, x_{p'}}$ are Palm expectations.

Measure valued sums

Charge points of $\mathcal{P}_n = \mathcal{P} \cap W_n$ in $W_n = [0, n^{1/d}]^d$, by the values of their score functions, contract the space by $n^{-1/d}$ as $n \to \infty$ to obtain weighted (signed) point measure on $W_1 = [0, 1]^d$

$$\mu_n^{\xi} := \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n) \delta_{n^{-1/d} x}.$$

Charge points of $\mathcal{P}_n = \mathcal{P} \cap W_n$ in $W_n = [0, n^{1/d}]^d$, by the values of their score functions, contract the space by $n^{-1/d}$ as $n \to \infty$ to obtain weighted (signed) point measure on $W_1 = [0,1]^d$

$$\mu_n^\xi := \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n) \delta_{n^{-1/d}x}.$$

Total mass is the previously considered sum

$$\mu_n^\xi(\mathbb{R}^d) = \mu_n^\xi(W_n) = S_n\,.$$

Integrals of test functions f are denoted by

$$\mu_n^\xi(f) = \int_{W_1} f(x)\, \mu_n^\xi(\mathrm{d}x).$$

Theorem [BB, Yogeshwaran, Yukich (2018+)]

Assume ξ is exponentially stabilizing on point process $\mathcal{P}$, with intensity λ, having fast decay of correlations.

- (Mean) If $(\xi, \mathcal{P})$ satisfies p-moment condition for some $p > 1$ then for all bounded functions f on $W_1 = [0,1]^d$

$$\left| \frac{\mathsf{E}\big(\mu_n^{\xi}(f)\big)}{n} - \lambda \mathsf{E}_0\big(\xi(0,\mathcal{P})\big) \int_{W_1} f(x)\,\mathrm{d}x \right| = O(n^{-1/d}).$$

- (Covariance) If $(\xi, \mathcal{P})$ satisfies p-moment condition for some $p > 2$ then for all bounded functions f, g on $W_1 = [0,1]^d$

$$\lim_{n\to\infty} \frac{\mathrm{Cov}\Big(\mu_n^{\xi}(f)\mu_n^{\xi}(g)\Big)}{n} = \sigma^2(\xi) \int_{W_1} f(x)g(x)\,\mathrm{d}x \in [0,\infty),$$

where

$$\begin{aligned}\sigma^2(\xi) = \lambda \mathsf{E}_0\big(\xi^2(0,\mathcal{P}\cup\{0\})\big) \\ + \lambda^2 \int_{\mathbb{R}^d} \mathsf{E}_{0,x}\Big(\xi(0,\mathcal{P}\cup\{0,x\})\xi(x,\mathcal{P}\cup\{0,x\})\Big) \\ - \Big(\mathsf{E}_0\big(\xi(0,\mathcal{P}\cup\{0\})\big)\Big)^2\,\mathrm{d}x\,.\end{aligned}$$

...

Main result cnt'd

- (Mean) & (Variance) imply the weak LLN for $\mu_n^\xi(f)$.
- CLT for $\mu_n^\xi(f)$ holds for some natural "admissible" subclass of stabilizing score functions on input processes with fast decay of correlations, provided $\sigma^2(\xi) > 0$.
- Multivariate CLT holds for $\left(\mu_n^\xi(f_1), \ldots, \mu_n^\xi(f_k)\right)$ by the Cramér-Wold device.
- Extensions for surface and smaller-order variance scaling (when $\sigma^2(\xi) = 0$, which is the case for some "very regular" (called hyperuniform) processes as some determinantal point processes having projection kernel (e.g. Ginibre).

Previous results

Presented approach extends some previous CLT results for point counts (constant marking function) of correlated point processes

- determinantal processes Soshnikov (2002)
- determinantal and permanental processes Shirai & Takahashi (2003),
- Gaussian entire functions Nazarov & Sodin (2012)

Main steps of the proof

We use the method of cumulants with the following intermediate steps:

- □ Fast decay of correlations of $\mathcal{P} = \mathbf{\Phi}$ and exponential stabilization of $\boldsymbol{\xi}$ on $\mathcal{P}$ implies fast decay of correlations of the weighted point measure $\tilde{\mathbf{\Phi}}$; result of an independent interest. Proof based on factorial moment expansions for point processes BB (1995).

Main steps of the proof

We use the method of cumulants with the following intermediate steps:

- Fast decay of correlations of $\mathcal{P} = \mathbf{\Phi}$ and exponential stabilization of ξ on $\mathcal{P}$ implies fast decay of correlations of the weighted point measure $\tilde{\mathbf{\Phi}}$; result of an independent interest. Proof based on factorial moment expansions for point processes BB (1995).
- Fast decay of correlations of $\tilde{\mathbf{\Phi}}$ implies Brillinger mixing of $\tilde{\mathbf{\Phi}}$; proof inspired by Nazarov & Sodin (2012) considering point counts.

Main steps of the proof

We use the method of cumulants with the following intermediate steps:

- □ Fast decay of correlations of $\mathcal{P} = \Phi$ and exponential stabilization of ξ on $\mathcal{P}$ implies fast decay of correlations of the weighted point measure $\tilde{\Phi}$; result of an independent interest. Proof based on factorial moment expansions for point processes BB (1995).
- □ Fast decay of correlations of $\tilde{\Phi}$ implies Brillinger mixing of $\tilde{\Phi}$; proof inspired by Nazarov & Sodin (2012) considering point counts.
- □ Brillinger mixing with enough fast increase of variance implies vanishing of cumulants order $k \geq 3$, of

$$\frac{\mu_n^{(\xi)}(f) - \mathsf{E}[\mu_n^{(\xi)}(f)]}{(\mathsf{Var}\left(\mu_n^{(\xi)}(f)\right))^{1/2}}$$

implying normal convergence (a classical result of Marcinkiewicz).

More details: fast decay of correlations with marks

Correlation functions of the weighted point measure $\tilde{\Phi}$: for $p \geq 1$, $k_1, \ldots, k_p \geq 1$,

$$m^{(k_1,\ldots,k_p)}(x_1, \ldots, x_p) := \mathsf{E}_{x_1,\ldots,x_p}(\xi(x_1, \mathcal{P})^{k_1} \ldots \xi(x_p, \mathcal{P})^{k_p}) \times \rho^{(p)}(x_1, \ldots, x_p).$$

More details: fast decay of correlations with marks

Correlation functions of the weighted point measure $\tilde{\Phi}$: for $p \geq 1$, $k_1, \ldots, k_p \geq 1$,

$$m^{(k_1,\ldots,k_p)}(x_1, \ldots, x_p) := \mathsf{E}_{x_1,\ldots,x_p}(\xi(x_1, \mathcal{P})^{k_1} \ldots \xi(x_p, \mathcal{P})^{k_p}) \times \rho^{(p)}(x_1, \ldots, x_p).$$

Fast decay of correlations if

$$m^{(k_1,\ldots,k_{p+q})}(x_1, \ldots, x_{p+q}) \approx m^{(k_1,\ldots,k_p)}(x_1, \ldots, x_p) \times m^{(k_{p+1},\ldots,k_{p+q})}(x_{p+1}, \ldots, x_{p+q})$$

up to an additive error exponentially decaying in the separation distance of the two groups of arguments $\{x_1, \ldots, x_p\}$ and $\{x_{p+1}, \ldots, x_{p+q}\}$.

Correlation functions of the weighted point measure $\tilde{\Phi}$: for $p \geq 1$, $k_1, \ldots, k_p \geq 1$,

$$m^{(k_1,\ldots,k_p)}(x_1,\ldots,x_p) := \mathsf{E}_{x_1,\ldots,x_p}(\xi(x_1,\mathcal{P})^{k_1} \ldots \xi(x_p,\mathcal{P})^{k_p}) \times \rho^{(p)}(x_1,\ldots,x_p).$$

Fast decay of correlations if

$$m^{(k_1,\ldots,k_{p+q})}(x_1,\ldots,x_{p+q}) \approx m^{(k_1,\ldots,k_p)}(x_1,\ldots,x_p) \times m^{(k_{p+1},\ldots,k_{p+q})}(x_{p+1},\ldots,x_{p+q})$$

up to an additive error exponentially decaying in the separation distance of the two groups of arguments $\{x_1,\ldots,x_p\}$ and $\{x_{p+1},\ldots,x_{p+q}\}$.

Fact: Fast decay of correlations of the input process $\mathcal{P} = \tilde{\Phi}$ and exponential stabilizing of the score function implies fast decay of correlations of the weighted point measure $\tilde{\Phi}$.

More details: fast decay of correlations with marks

Correlation functions of the weighted point measure $\tilde{\Phi}$: for $p \geq 1$, $k_1, \dots, k_p \geq 1$,

$$m^{(k_1,\dots,k_p)}(x_1,\dots,x_p) := \mathsf{E}_{x_1,\dots,x_p}(\xi(x_1,\mathcal{P})^{k_1}\dots\xi(x_p,\mathcal{P})^{k_p}) \times \rho^{(p)}(x_1,\dots,x_p).$$

Fast decay of correlations if

$$m^{(k_1,\dots,k_{p+q})}(x_1,\dots,x_{p+q}) \approx m^{(k_1,\dots,k_p)}(x_1,\dots,x_p) \times m^{(k_{p+1},\dots,k_{p+q})}(x_{p+1},\dots,x_{p+q})$$

up to an additive error exponentially decaying in the separation distance of the two groups of arguments $\{x_1,\dots,x_p\}$ and $\{x_{p+1},\dots,x_{p+q}\}$.

Fact: Fast decay of correlations of the input process $\mathcal{P} = \tilde{\Phi}$ and exponential stabilizing of the score function implies fast decay of correlations of the weighted point measure $\tilde{\Phi}$.

Proof using factorial moment expansion (an ersatz of the chaos expansion for non-Poisson inputs) of $m^{(k_1,\dots,k_{p+q})}$ with respect to factorial moments of Φ.

More details: fast decay of correlations and Brillinger mixing

Fast decay of correlations of $\mathbf{\Phi}$ is equivalent to the the Ursell functions $\boldsymbol{m_\top(x_1,\ldots,x_p)}$ (densities of cumulant measures) being absolutely bounded by some function $\boldsymbol{\phi_\top(\cdot)}$ exponentially decaying in the arguments diameter

$$|\boldsymbol{m_\top(x_1,\ldots,x_k)}| \le \boldsymbol{C_k^\top \phi_\top\big(c_k^\top \mathbf{diam}(x_1,\ldots,x_k)\big)}.$$

More details: fast decay of correlations and Brillinger mixing

Fast decay of correlations of Φ is equivalent to the the Ursell functions $m_\top(x_1,\dots,x_p)$ (densities of cumulant measures) being absolutely bounded by some function $\phi_\top(\cdot)$ exponentially decaying in the arguments diameter

$$|m_\top(x_1,\dots,x_k)| \le C_k^\top \phi_\top\big(c_k^\top \mathbf{diam}(x_1,\dots,x_k)\big).$$

This imples Brillinger mixing

$$\sup_{x_1\in\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} |m_\top(x_1,\dots,x_k)|\, \mathrm{d}x_2\cdots \mathrm{d}x_k < \infty.$$

More details: fast decay of correlations and Brillinger mixing

Fast decay of correlations of $\boldsymbol{\Phi}$ is equivalent to the the Ursell functions $m_\top(x_1,\ldots,x_p)$ (densities of cumulant measures) being absolutely bounded by some function $\phi_\top(\cdot)$ exponentially decaying in the arguments diameter

$$|m_\top(x_1,\ldots,x_k)| \leq C_k^\top \phi_\top\big(c_k^\top \mathbf{diam}(x_1,\ldots,x_k)\big).$$

This imples Brillinger mixing

$$\sup_{x_1\in\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} |m_\top(x_1,\ldots,x_k)|\,\mathrm{d}x_2\cdots\mathrm{d}x_k < \infty.$$

Similarly for the weighted point measure $\tilde{\boldsymbol{\Phi}}$.

More details: Brillinger mixing, variance asymptotic and CLT

Brillinger mixing of $\tilde{\Phi}$ implies all cumulants of $\mu_n^{(\xi)}(f)$ of order $k \geq 1$ grow as n (window volume).

Brillinger mixing of $\tilde{\Phi}$ implies all cumulants of $\mu_n^{(\xi)}(f)$ of order $k \geq 1$ grow as n (window volume).

Consequently, the kth cumulant of

$$\frac{\mu_n^{(\xi)}(f) - \mathsf{E}[\mu_n^{(\xi)}(f)]}{(\mathsf{Var}\left(\mu_n^{(\xi)}(f)\right))^{1/2}}$$

when $n \to \infty$ behaves as

$$\frac{n}{(\mathsf{Var}\left(\mu_n^{(\xi)}(f)\right))^{k/2}}.$$

More details: Brillinger mixing, variance asymptotic and CLT

Brillinger mixing of $\tilde{\Phi}$ implies all cumulants of $\mu_n^{(\xi)}(f)$ of order $k \geq 1$ grow as n (window volume).

Consequently, the kth cumulant of

$$\frac{\mu_n^{(\xi)}(f) - \mathsf{E}[\mu_n^{(\xi)}(f)]}{(\mathsf{Var}\left(\mu_n^{(\xi)}(f)\right))^{1/2}}$$

when $n \to \infty$ behaves as

$$\frac{n}{(\mathsf{Var}\left(\mu_n^{(\xi)}(f)\right))^{k/2}}.$$

Hence, for $k \geq 3$ and large enough these cumulants tend to 0 with $n \to \infty$, provided the variance grows as n^{δ}, with some $\delta > 0$.

References I

- □ Penrose &Yukich (2003) Weak laws of large numbers in geometric probability, *Ann. Appl. Probab.*
- □ Soshnikov (2002) Gaussian limit for determinantal random point fields, *Ann. Probab.*
- □ Baryshnikov & Yukich (2005) Gaussian limits for random measures in geometric probability. *Ann. Appl. Prob.*
- □ Shirai & Takahashi (2003) Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes, *J. Funct. Anal.*,
- □ Nazarov & Sodin (2012) Correlation functions for random complex zeroes: Strong clustering and local universality, *Comm. Math. Phys.*
- ⇒ BB, Yogeshwaran, & Yukich (2018+) Limit theory for geometric statistics of point processes having fast decay of correlations *Ann. Probab.*

On scattering moments of geometrically marked point processes

based on a joint work (in progress) with
Antoine Brochard [PhD student PSL/ENS and Huawei, Paris]
Stephane Mallat [College de France and ENS Paris]
Sixin Zhang [Peking University]

Moments and transforms

We know Fourier and Laplace transforms, moments, factorial moments, ... of random measures and point processes.

Moments and transforms

We know Fourier and Laplace transforms, moments, factorial moments, ... of random measures and point processes.

This talk is about scattering moments — a new class of operators, which in some sense (mostly to be yet explored) capture distribution, but also individual realizations of point patterns, proposed recently by Stephane Mallat (see Mallat (2012)) for signal (image) analysis.

Moments and transforms

We know Fourier and Laplace transforms, moments, factorial moments, ... of random measures and point processes.

This talk is about scattering moments — a new class of operators, which in some sense (mostly to be yet explored) capture distribution, but also individual realizations of point patterns, proposed recently by Stephane Mallat (see Mallat (2012)) for signal (image) analysis.

This is a discrete family of nonlinear and noncommuting operators, computing at different scales the modulus of a wavelet transform of a one- or higher-dimensional signal (e.g. image).

Moments and transforms

We know Fourier and Laplace transforms, moments, factorial moments, ... of random measures and point processes.

This talk is about scattering moments — a new class of operators, which in some sense (mostly to be yet explored) capture distribution, but also individual realizations of point patterns, proposed recently by Stephane Mallat (see Mallat (2012)) for signal (image) analysis.

This is a discrete family of nonlinear and noncommuting operators, computing at different scales the modulus of a wavelet transform of a one- or higher-dimensional signal (e.g. image).

They are Lipschitz-continuous with respect smooth signal diffeomorphisms and this makes them useful in signal processing, in particular in relation to statistical learning, as they allow one to learn intrinsic properties of some class of signals from a smaller number of signal samples.

Wavelet

Following Bruna, Mallat, Bacry, Muzy (2015),
let ψ be a continuous, bounded, complex valued function on $\mathbb{R}^d$ of zero average $\int_{\mathbb{R}^d} \psi(x)\, \mathrm{d}x = 0$
and such that $|\psi(x)| = O(|x|^{-d})$ for $|x| \to \infty$.

Usually ψ is normalized so that $\int_{\mathbb{R}^d} |\psi(x)|\, \mathrm{d}x = 1$.

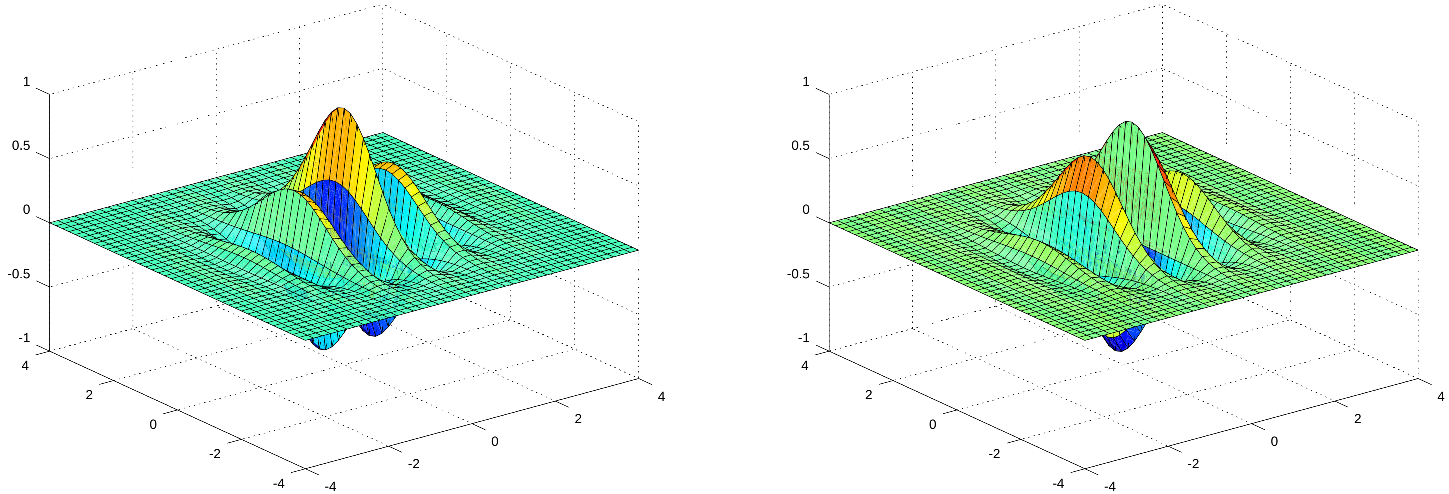
We call ψ (d-dimensional) mother wavelet.

Example: 2D Morlet wavelet

Morlet wavelet on the plane

$$\psi(x) = \exp(i\,\omega\cdot x)\exp(-|x|^2/2),$$

where i is the imaginary unit and $\omega\cdot x$ is the scalar product of some nonzero vector parameter $\omega \in \mathbb{R}^2$, called spatial frequency, with $x \in \mathbb{R}^2$.



Real and Imaginary part of the Morlet wavelet with $\omega = (5.5, 0)$.

Scaling and rotating the mother wavelet

Consider a discrete family of re-scaled and rotated wavelets

$$\psi_{(j,\theta)} = \psi_{(j,\theta)}(x) := 2^{-jd}\psi(2^{-j}r_{-\theta}x),$$

with the scale parameter $j \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ and the rotation parameter $\theta \in [0, 2\pi)$; ($r_\theta x$ denotes the rotation of $x \in \mathbb{R}^2$ by the angle θ with respect to the origin).

Wavelet transforms of a signal in $\mathbb{R}^d$

In general, signal is modeled as a (random, possibly signed) measure $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{x})$ on $\mathbb{R}^d$.

Wavelet transforms of a signal in $\mathbb{R}^d$

In general, signal is modeled as a (random, possibly signed) measure $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{x})$ on $\mathbb{R}^d$.

Wavelet transform of (a realization of) $\boldsymbol{\Lambda}$ at scale $\mathbf{2}^{\boldsymbol{j}}$ and angle $\boldsymbol{\theta}$, is a (random) filed on $\mathbb{R}^d$ defined as a convolution of $\boldsymbol{\Lambda}$ with the wavelet $\boldsymbol{\psi}_{(\boldsymbol{j},\boldsymbol{\theta})}$:

$$(\boldsymbol{\Lambda} \star \boldsymbol{\psi}_{(\boldsymbol{j},\boldsymbol{\theta})})(\boldsymbol{x}) := \int_{\mathbb{R}^d} \boldsymbol{\psi}_{(\boldsymbol{j},\boldsymbol{\theta})}(\boldsymbol{x} - \boldsymbol{y})\, \boldsymbol{\Lambda}(\mathrm{d}\boldsymbol{y})\,.$$

Wavelet transforms of a signal in $\mathbb{R}^d$

In general, signal is modeled as a (random, possibly signed) measure $\Lambda = \Lambda(\mathrm{d}x)$ on $\mathbb{R}^d$.

Wavelet transform of (a realization of) Λ at scale 2^j and angle θ, is a (random) filed on $\mathbb{R}^d$ defined as a convolution of Λ with the wavelet $\psi_{(j,\theta)}$:

$$(\Lambda \star \psi_{(j,\theta)})(x) := \int_{\mathbb{R}^d} \psi_{(j,\theta)}(x - y)\, \Lambda(\mathrm{d}y)\,.$$

Observe: The zero average property of the mother wavelet $\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x = 0$ implies that the wavelet transform $\Lambda \star \psi_{(j,\theta)}(x)$ at the scale j has larger absolute values for x where the Λ is has more variability at this given scale. It (almost) vanishes where Λ is (almost) uniform at this scale.

In this talk we are interested in purely atomic signals, that is weighted point processes

$$\Lambda = \tilde{\Phi} := \sum_i M_i \delta_{X_i},$$

where $\Phi = \sum_i \delta_{X_i}$ is a simple, stationary point process in $\mathbb{R}^d$ and M_i are marks of the points of Φ produced by a real valued, translation invariant score (marking) function m:

$$M_i = m(X_i, \Phi),$$

with $m(x + a, \phi + a) = m(x, \phi)$ for all $x, a \in \mathbb{R}^d$ and signal $\phi \ni x$.

Wavelet transforms of purely atomic signals in $\mathbb{R}^d$

In this talk we are interested in purely atomic signals, that is weighted point processes

$$\Lambda = \tilde{\Phi} := \sum_i M_i \delta_{X_i},$$

where $\Phi = \sum_i \delta_{X_i}$ is a simple, stationary point process in $\mathbb{R}^d$ and M_i are marks of the points of Φ produced by a real valued, translation invariant score (marking) function m:

$$M_i = m(X_i, \Phi),$$

with $m(x + a, \phi + a) = m(x, \phi)$ for all $x, a \in \mathbb{R}^d$ and signal $\phi \ni x$.

In this case the wavelet transforms are just shot-noise fields of $\tilde{\Phi}$ with wavelets as the response function

$$(\tilde{\Phi} \star \psi_{(j,\theta)})(x) = \sum_i M_i \psi_{(j,\theta)}(x - x_i).$$

Scattering moments: introducing non-linearity

Define the scattering fields as the modulus of the (complex valued) wavelet transforms

$$S_{j,\theta}\tilde{\Phi} := |\tilde{\Phi} \star \psi_{(j,\theta)}(x)| = \left|\sum_i M_i \psi_{(j,\theta)}(x - x_i)\right|.$$

Scattering moments: introducing non-linearity

Define the scattering fields as the modulus of the (complex valued) wavelet transforms

$$S_{j,\theta}\tilde{\Phi} := |\tilde{\Phi} \star \psi_{(j,\theta)}(x)| = \left|\sum_i M_i \psi_{(j,\theta)}(x - x_i)\right|.$$

Define the scattering moments of the (stationary) signal $\tilde{\Phi}$ as the expected values of these new fields (without loss of generality) calculated at the origin $\mathbf{0} \in \mathbb{R}^d$

$$\bar{S}\tilde{\Phi}(j,\theta) := \mathsf{E}[S_{j,\theta}\tilde{\Phi}(\mathbf{0})] \quad j \in \mathbb{Z},\ \theta \in [0, 2\pi).$$

Define the scattering fields as the modulus of the (complex valued) wavelet transforms

$$S_{j,\theta}\tilde{\Phi} := |\tilde{\Phi} \star \psi_{(j,\theta)}(x)| = \left|\sum_i M_i \psi_{(j,\theta)}(x - x_i)\right|.$$

Define the scattering moments of the (stationary) signal $\tilde{\Phi}$ as the expected values of these new fields (without loss of generality) calculated at the origin $\mathbf{0} \in \mathbb{R}^d$

$$\bar{S}\tilde{\Phi}(j,\theta) := \mathsf{E}[S_{j,\theta}\tilde{\Phi}(\mathbf{0})] \quad j \in \mathbb{Z},\ \theta \in [0, 2\pi).$$

If $\tilde{\Phi}$ is isotropic $\bar{S}\tilde{\Phi}(j,\theta) = \bar{S}\tilde{\Phi}(j)$ do not depend on θ.

Define the scattering fields as the modulus of the (complex valued) wavelet transforms

$$S_{j,\theta}\tilde{\Phi} := |\tilde{\Phi} \star \psi_{(j,\theta)}(x)| = \left|\sum_i M_i \psi_{(j,\theta)}(x - x_i)\right|.$$

Define the scattering moments of the (stationary) signal $\tilde{\Phi}$ as the expected values of these new fields (without loss of generality) calculated at the origin $0 \in \mathbb{R}^d$

$$\bar{S}\tilde{\Phi}(j,\theta) := \mathsf{E}[S_{j,\theta}\tilde{\Phi}(0)] \quad j \in \mathbb{Z},\ \theta \in [0, 2\pi).$$

If $\tilde{\Phi}$ is isotropic $\bar{S}\tilde{\Phi}(j,\theta) = \bar{S}\tilde{\Phi}(j)$ do not depend on θ.

Second order scattering moments are defined by considering the first scattering field $S_{j,\theta}\tilde{\Phi}$ as the signal density

$$\bar{S}\tilde{\Phi}(j_1,\theta_1,j_2,\theta_2) := \mathsf{E}[||\tilde{\Phi} \star \psi_{(j_1,\theta_1)}| \star \psi_{(j_2,\theta_2)}(0))|].$$

Remarks and questions on scattering moments

- The non-linearity produced by the modulus $|\cdot|$ in $\mathsf{E}\Big[\Big|\sum_i M_i \psi_j(0 - X_i)\Big|\Big]$ makes the scattering moments (a priori) depend on all correlation functions of $\tilde{\Phi}$, (which would not be the case if the square $|\cdot|^2$ of the norm is taken).

Remarks and questions on scattering moments

- The non-linearity produced by the modulus $|\cdot|$ in $\mathsf{E}\left[\left|\sum_i M_i \psi_j(0 - X_i)\right|\right]$ makes the scattering moments (a priori) depend on all correlation functions of $\tilde{\Phi}$, (which would not be the case if the square $|\cdot|^2$ of the norm is taken).
- To what extend the scattering moments characterize the correlation functions of $\tilde{\Phi}$ (and its distribution)? We do not know.

Remarks and questions on scattering moments

- The non-linearity produced by the modulus $|\cdot|$ in $\mathsf{E}\left[\left|\sum_i M_i\psi_j(0 - X_i)\right|\right]$ makes the scattering moments (a priori) depend on all correlation functions of $\tilde{\Phi}$, (which would not be the case if the square $|\cdot|^2$ of the norm is taken).
- To what extend the scattering moments characterize the correlation functions of $\tilde{\Phi}$ (and its distribution)? We do not know.
- We study their asymptotic when $j \to -\infty$ and $j \to \infty$ (at small and large scales). This is inspired by results for non-marked, 1D-Poisson process obtained in Bruna, Mallat, Bacry, Muzy (2015).

Empirical scattering moments

Empirical scattering moments are calculated when replacing the expectations $\mathsf{E}[\cdots]$ by the empirical averaging over $\boldsymbol{x}$ in a given observation window $\boldsymbol{W}$

$$\hat{S}\tilde{\Phi}(j,\theta) := \frac{1}{|W|}\int_W S_{j,\theta}\tilde{\Phi}(x)\,\mathrm{d}x \quad j \in \mathbb{Z},\ \theta \in [0, 2\pi).$$

Empirical scattering moments

Empirical scattering moments are calculated when replacing the expectations $\mathsf{E}[\cdots]$ by the empirical averaging over $\boldsymbol{x}$ in a given observation window $\boldsymbol{W}$

$$\hat{S}\tilde{\Phi}(j,\theta) := \frac{1}{|W|}\int_W S_{j,\theta}\tilde{\Phi}(x)\,\mathrm{d}x \quad j \in \mathbb{Z},\ \theta \in [0,2\pi).$$

- When the scale $\boldsymbol{j}$ is small with respect to the window $\boldsymbol{W}$ the empirical scattering moments are good estimators of the theoretical ones, hence they are invariant with respect to realizations of $\tilde{\Phi}$ (rather characterize the distribution of $\tilde{\Phi}$).

Empirical scattering moments

Empirical scattering moments are calculated when replacing the expectations $\mathsf{E}[\cdots]$ by the empirical averaging over $\boldsymbol{x}$ in a given observation window $\boldsymbol{W}$

$$\hat{S}\tilde{\boldsymbol{\Phi}}(j,\boldsymbol{\theta}) := \frac{1}{|\boldsymbol{W}|}\int_{\boldsymbol{W}} S_{j,\boldsymbol{\theta}}\tilde{\boldsymbol{\Phi}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad j\in\mathbb{Z},\ \boldsymbol{\theta}\in[0,2\pi).$$

- When the scale j is small with respect to the window $\boldsymbol{W}$ the empirical scattering moments are good estimators of the theoretical ones, hence they are invariant with respect to realizations of $\tilde{\boldsymbol{\Phi}}$ (rather characterize the distribution of $\tilde{\boldsymbol{\Phi}}$).
- For larger scales the empirical scattering moments carry additional information about the given realization. Lipschitz-continuous with respect smooth signal diffeomorphisms, can be used to statistical learning and/or classification of signal patterns.
  Some application example will be given in the second part of this talk.

Outline of the remaining part

- Limit results for scattering moments of weighted point processes,
- Statistical learning of score functions using scattering moments. (Briefly)

Limit results for scattering moments

Notation, assumptions (recap)

- $\Phi = \sum_i \delta_{X_i}$ simple, stationary point process on $\mathbb{R}^d$ of intensity λ.
- $M_i := m(X_i, \Phi) \in \mathbb{R}$ real marks defined by some translation-invariant marking (score) function.
- $\tilde{\Phi} = \sum_i M_i \delta_{X_i}$ random, purely atomic, possible signed measure, where marks are point weights.

Notation, assumptions (recap)

- $\Phi = \sum_i \delta_{X_i}$ simple, stationary point process on $\mathbb{R}^d$ of intensity λ.
- $M_i := m(X_i, \Phi) \in \mathbb{R}$ real marks defined by some translation-invariant marking (score) function.
- $\tilde{\Phi} = \sum_i M_i \delta_{X_i}$ random, purely atomic, possible signed measure, where marks are point weights.
- ψ (mother wavelet) continuous, bounded, real (for simplicity of the presentation) function on $\mathbb{R}^d$, zero average $\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x = 0$, normalized $||\psi||_1 = \int_{\mathbb{R}^d} |\psi(x)|\,\mathrm{d}x = 1$, with compact support $\mathrm{supp}(\psi) = B$.
- $\psi_j(x) := 2^{-jd}\psi(2^{-j}(x)$ wavelet at scale $j \in \{\ldots, -1, 0, 1, \ldots\}$. For simplicity no rotation considered, irrelevant if Φ is isotropic.

Notation, assumptions (recap)

- $\Phi = \sum_i \delta_{X_i}$ simple, stationary point process on $\mathbb{R}^d$ of intensity λ.
- $M_i := m(X_i, \Phi) \in \mathbb{R}$ real marks defined by some translation-invariant marking (score) function.
- $\tilde{\Phi} = \sum_i M_i \delta_{X_i}$ random, purely atomic, possible signed measure, where marks are point weights.
- ψ (mother wavelet) continuous, bounded, real (for simplicity of the presentation) function on $\mathbb{R}^d$, zero average $\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}x = 0$, normalized $||\psi||_1 = \int_{\mathbb{R}^d} |\psi(x)|\,\mathrm{d}x = 1$, with compact support $\mathrm{supp}(\psi) = B$.
- $\psi_j(x) := 2^{-jd}\psi(2^{-j}(x))$ wavelet at scale $j \in \{\dots, -1, 0, 1, \dots\}$. For simplicity no rotation considered, irrelevant if Φ is isotropic.
- $S_j(x) := S_j\tilde{\Phi}(x) = \left|\sum_i M_i \psi_j(x - X_i)\right|$ scattering filed of $\tilde{\Phi}$ at scale j; stationary random field on $\mathbb{R}^d$.
- $\bar{S}(j) := \bar{S}\tilde{\Phi}(j) = \mathsf{E}[S_j(0)]$ scattering moment at scale $j \in \mathbb{Z}$.

Small scale limit

Theorem. Assume the second order correlation function $\rho^2(x, y)$ of Φ exists and denote by $\kappa(\cdot)$ the reduced second order correlation (i.e., $\rho^2(x, y) = \lambda\kappa(x - y)$). Assume $\int_B |\mathbb{E}^{0,u}(m(0, \Phi))\kappa(u)| du < \infty$, where $\mathsf{E}^{0,u}$ is two-point Palm expectation of Φ. Then, as $j \to -\infty$

$$\bar{S}(j) = \lambda \mathsf{E}^0[m(0, \Phi)] + O(2^{jd} \int_B \mathsf{E}^{0,2^j u}[m(0, \Phi)]\kappa(2^j u) \mathrm{d}u).$$

The scattering moments at small scales converge to the intensity $\lambda \mathsf{E}^0[m(0, \Phi)]$ of the random measure $\tilde{\Phi}$. The speed of convergence depends on the reduced second order correlation (more repulsion faster convergence).

Small scale limit — proof idea

At small scales the wavelet functions "see" the points separately and hence

$$\mathsf{E}\Big[\Big|\sum_i \boldsymbol{M_i \psi_j(0 - X_i)}\Big|\Big] \approx \mathsf{E}\Big[\sum_i \Big|\boldsymbol{M_i \psi_j(0 - X_i)}\Big|\Big]. \qquad (*)$$

Small scale limit — proof idea

At small scales the wavelet functions “see” the points separately and hence

$$\mathsf{E}\Big[\Big|\sum_i \boldsymbol{M_i \psi_j (0 - X_i)}\Big|\Big] \approx \mathsf{E}\Big[\sum_i \Big|\boldsymbol{M_i \psi_j (0 - X_i)}\Big|\Big]. \qquad (*)$$

Use Campbell’s formula to calculate the expression in the right-hand-side of (*).

Small scale limit — proof idea

At small scales the wavelet functions “see” the points separately and hence

$$\mathsf{E}\Big[\Big|\sum_i M_i \psi_j(0 - X_i)\Big|\Big] \approx \mathsf{E}\Big[\sum_i \Big|M_i \psi_j(0 - X_i)\Big|\Big]. \qquad (*)$$

Use Campbell’s formula to calculate the expression in the right-hand-side of (*).

Remark: Factorial moment expansion can give higher order approximations in (*). TODO

Denote $\boldsymbol{Y_j := \sum_{x\in\Phi} \psi(2^{-j}x)m(x,\Phi)}$. Recall $\boldsymbol{\bar{S}(j) = 2^{-jd}}\mathsf{E}[|\boldsymbol{Y_j}|]$.)

Theorem Assume $\boldsymbol{Y_j}$ satisfy the CLT as $\boldsymbol{j \to \infty}$ and some moment conditions usually required for the CLT. Then

$$\lim_{j\to\infty} \frac{\bar{S}(j)}{2^{-jd}\sqrt{\mathrm{Var}(Y_j)}} = \mathsf{E}[|\mathcal{N}(0,1)|] = \sqrt{2/\pi}.$$

The scattering moments at large scale reveal the variance asymptotic of $\boldsymbol{Y_j}$. It is related but in general not the same as the asymptotic of the variance of the total mass of $\boldsymbol{\tilde{\Phi}}$. To be explained.

Large scale limit — CLT assumption and proof idea

The CLT for $\boldsymbol{Y_j := \sum_{x\in\Phi}\psi(2^{-j}x)m(x,\Phi)}$ can be established for a large class of score functions and point processes including Poisson one. (There exists a reach literature, e.g. BB, Yogeshwaran, Yukich (2019) concerning point processes with fast decay of correlations.)

Large scale limit — CLT assumption and proof idea

The CLT for $\boldsymbol{Y_j := \sum_{x\in\Phi} \psi(2^{-j}x)m(x,\Phi)}$ can be established for a large class of score functions and point processes including Poisson one. (There exists a reach literature, e.g. BB, Yogeshwaran, Yukich (2019) concerning point processes with fast decay of correlations.)

By the CLT, with $\mathsf{E}[\boldsymbol{Y_j}] = \boldsymbol{0}$ (since the wavelet function is centered), and by the continuity of $|\cdot|$ we have

$$\frac{|\boldsymbol{Y_j}|}{\sqrt{\mathbf{Var}(\boldsymbol{Y_j})}} \xrightarrow[j\to\infty]{d} |\mathcal{N}(\mathbf{0},\mathbf{1})|.$$

Large scale limit — CLT assumption and proof idea

The CLT for $Y_j := \sum_{x\in\Phi}\psi(2^{-j}x)m(x,\Phi)$ can be established for a large class of score functions and point processes including Poisson one. (There exists a reach literature, e.g. BB, Yogeshwaran, Yukich (2019) concerning point processes with fast decay of correlations.)

By the CLT, with $\mathsf{E}[Y_j] = 0$ (since the wavelet function is centered), and by the continuity of $|\cdot|$ we have

$$\frac{|Y_j|}{\sqrt{\mathrm{Var}(Y_j)}} \xrightarrow[j\to\infty]{d} |\mathcal{N}(0,1)|.$$

Under some suitable moment assumption the convergence holds also in L1.

Hence

$$\frac{\mathsf{E}[|Y_j|]}{\sqrt{\mathrm{Var}(Y_j)}} \to_{j\to\infty} \mathsf{E}[|\mathcal{N}(0,1)|] = \sqrt{2/\pi}.$$

Suppose the power spectral measure (Bartlett spectrum) of the random measure $\tilde{\Phi}$ admits density $\mu_{\tilde{\Phi}}(\nu)$ (sometimes called "structure" or "scattering" function). Denote by $\widehat{\psi}$ the Fourier transform of the wavelet function ψ. Then

$$\begin{aligned}\mathrm{Var}(Y_j) &= \mathrm{Var}\Big(\int_{\mathbb{R}^d} \psi(2^{-j}x)\, \tilde{\Phi}(\mathrm{d}x)\Big) \\ &= 2^{jd} \int_{\mathbb{R}^d} |\widehat{\psi}(\nu)|^2 \mu_{\tilde{\Phi}}(2^{-j}\nu)\, \mathrm{d}\nu \qquad (*) \\ (\text{as } j \to \infty) &\sim 2^{jd} \quad \text{(volume scaling)}\end{aligned}$$

provided $0 < \mu_{\tilde{\Phi}}(0) < \infty \iff$ volume scaling of the variance of $\tilde{\Phi}$.

Suppose the power spectral measure (Bartlett spectrum) of the random measure $\tilde{\Phi}$ admits density $\mu_{\tilde{\Phi}}(\nu)$ (sometimes called "structure" or "scattering" function). Denote by $\widehat{\psi}$ the Fourier transform of the wavelet function ψ. Then

$$\begin{aligned}\mathrm{Var}(Y_j) &= \mathrm{Var}\Big(\int_{\mathbb{R}^d} \psi(2^{-j}x)\,\tilde{\Phi}(\mathrm{d}x)\Big)\\ &= 2^{jd}\int_{\mathbb{R}^d} |\widehat{\psi}(\nu)|^2 \mu_{\tilde{\Phi}}(2^{-j}\nu)\,\mathrm{d}\nu \qquad (*)\\ \text{(as } j\to\infty) &\sim 2^{jd} \quad \text{(volume scaling)}\end{aligned}$$

provided $0 < \mu_{\tilde{\Phi}}(0) < \infty \iff$ volume scaling of the variance of $\tilde{\Phi}$.

When $\mu_{\tilde{\Phi}}(0) = 0$ (hyperuniformity) or $\mu_{\tilde{\Phi}}(0) = \infty$ (hyperfluctuation of $\tilde{\Phi}$) the expression in (*) has, respectively, sub-volume or super-volume scaling, but the volume exponents are in general different from these regarding the total mass of $\tilde{\Phi}$.

Proposition: When $\tilde{\Phi}$ exhibits volume order variance scaling (e.g. Poisson case) then $\mathrm{Var}(Y_j) \sim 2^{jd}$ and

$$\lim_{j\to\infty} 2^{jd/2}\bar{S}_j = \sqrt{2/\pi}||\psi||_2^2\sqrt{\sigma^2},$$

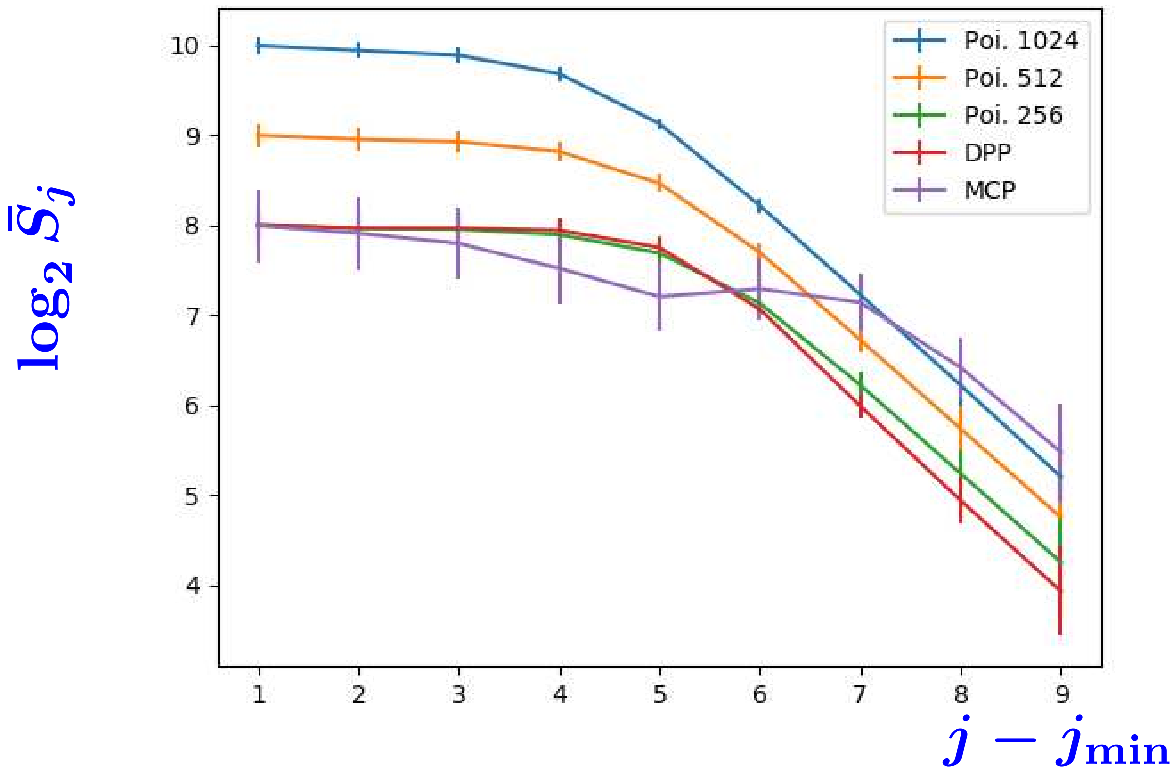
where $\sigma^2 = \lambda\mathbb{E}^0[m^2(0,\Phi)] + \int_{\mathbb{R}^d}\mathbb{E}^{0,x}[m(0,\Phi)m(x,\Phi)]\lambda\kappa(x) - (\lambda\mathbb{E}^0[m(0,\Phi)])^2\mathrm{d}x$.

Large scale limit — variance scaling classification

When $j \to \infty$

- $\log \bar{S}_j \sim -jd/2$ indicates Poisson-like variance asymptotic,
- $\log \bar{S}_j \lesssim -jd/2$ indicates hyperuniformity (e.g. Ginibre),
- $\log \bar{S}_j \gtrsim -jd/2$ indicates hyperfluctuation (e.g. Poisson line intersections).

Examples; $d = 2$: volume order variance scaling



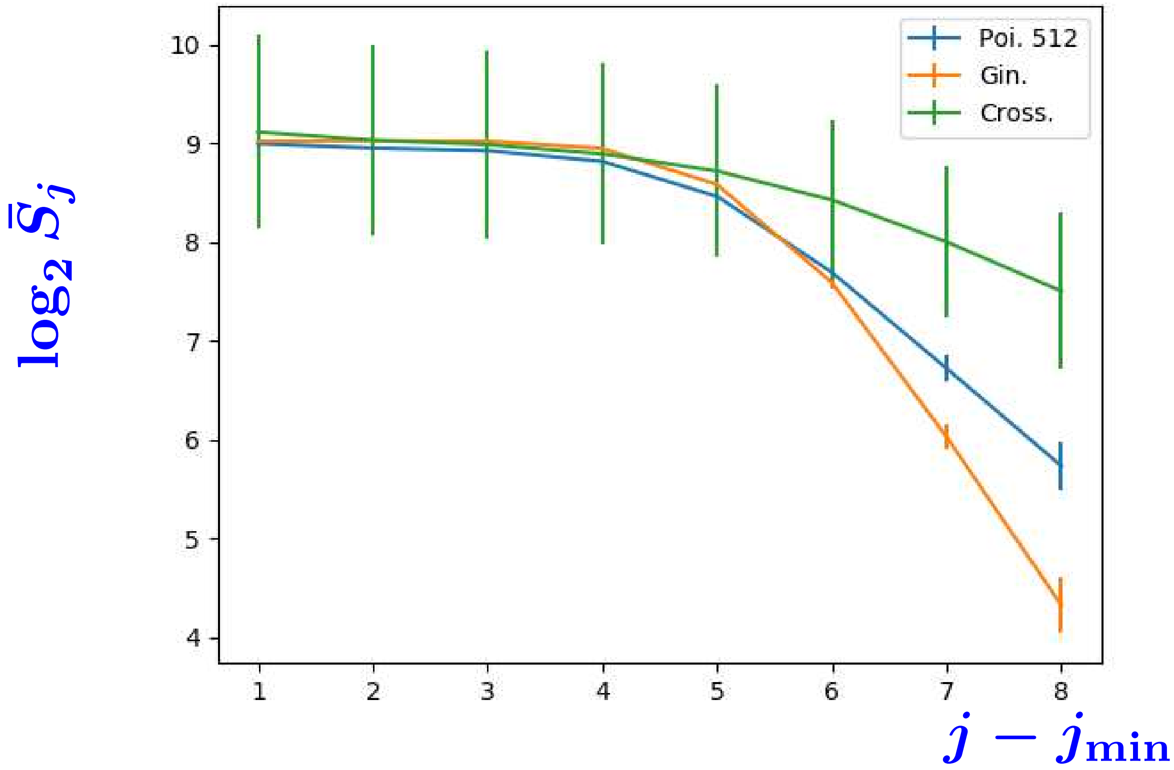
Poi. — Poisson process, DPP — Gaussian determinantal process, MCP — Matérn cluster process; all non-marked $(m(x, \phi) = 1)$.

Note $\log_2(\bar{S}_{j_{\min}}) = \log_2(\lambda)$; e.g. $\log_2(512) = 9$, etc.

For large j, $\log_2(\bar{S}_j) \sim \frac{-j2}{2} = -j$.

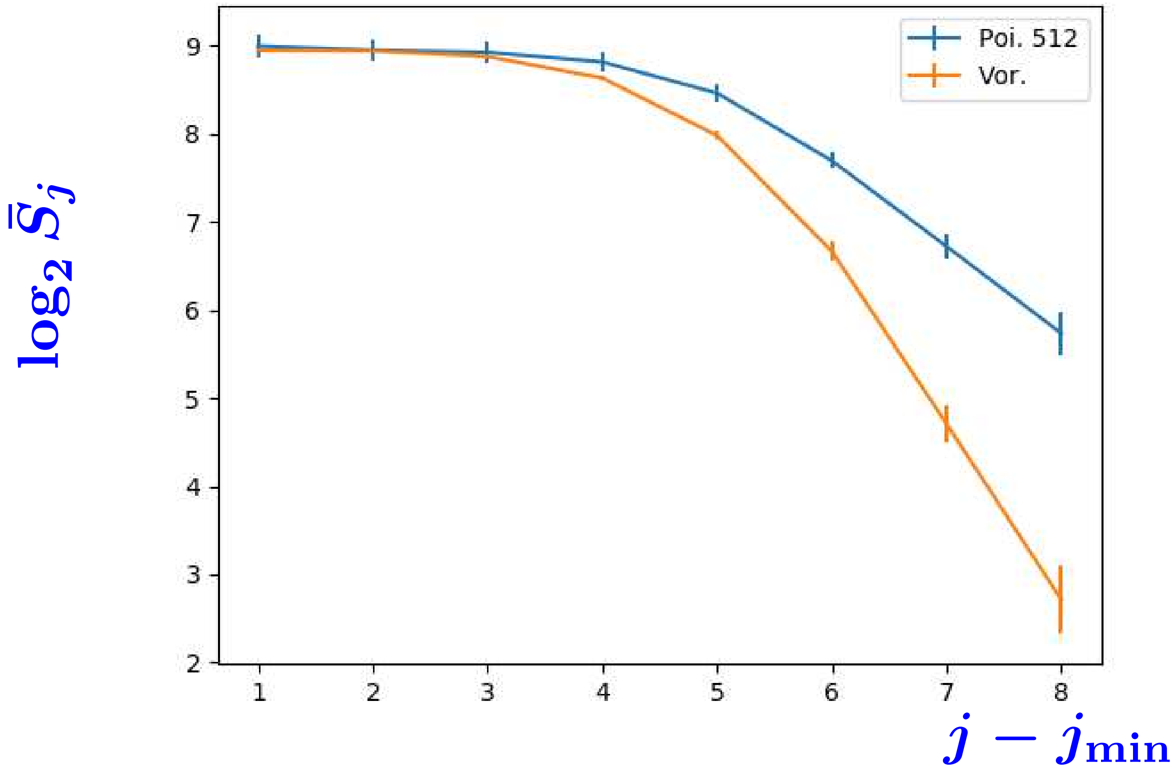
Error bars — 95% confidence intervals calculated on 500 realizations.

Examples: hyperuniformity and hyperfluctuation



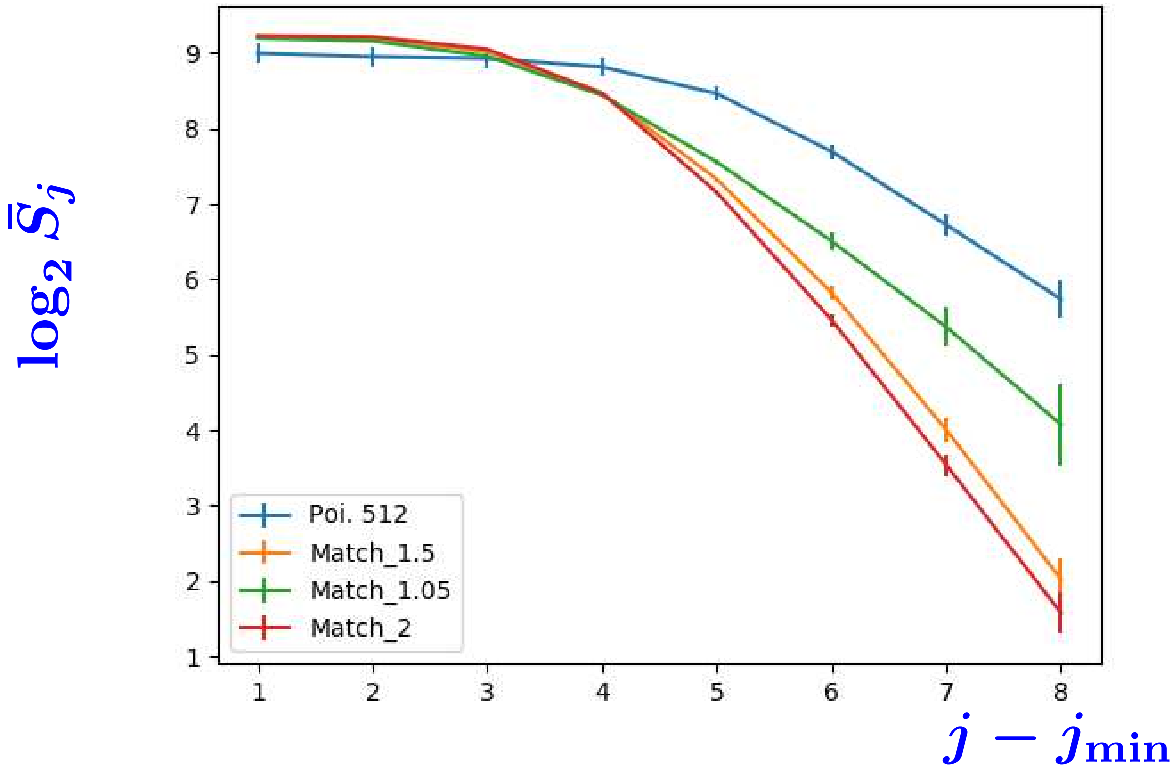
Poi. — Poisson process, $\log_2(\bar{S}_j) \sim_{j\to\infty} -j$,
Cross. — Poisson line crossing, $\log_2(\bar{S}_j) \sim_{j\to\infty} > -j$ (hyperfluct.),
Ginibre — determinantal process; $\log_2(\hat{S}_j) \sim_{j\to\infty} < -j$ (hyperunif.).

Example: Voronoi-surface marking of Poisson points



Poi. — Poisson process, $\log_2(\bar{S}_j) \sim_{j\to\infty} -j$,
Vor. — Voronoi-surface marking of Poisson, $\log_2(\bar{S}_j) \sim_{j\to\infty} < -j$.

Example: stable maching of Poisson to 2D lattice



$\log_2(\bar{S}_j) \sim_{j\to\infty} < -j$

Hyperuniform (dependent) thinning of Poisson of intensity $\lambda > 1$ obtained as its stable maching to the square lattice of intensity 1 on 2D; see Klatt, Last, Yogeshwaran (2018).

Statistical learning of geometric marks

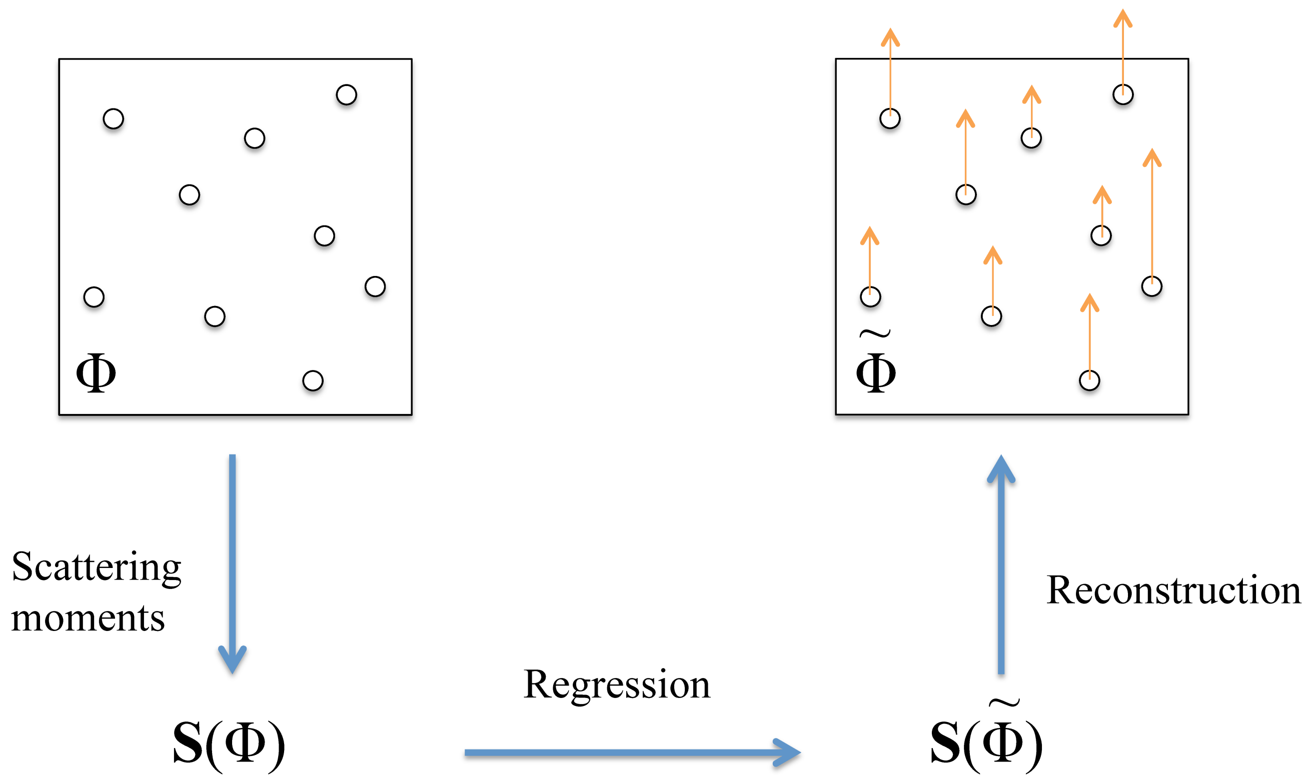
Problem of learning of geometric marks

Suppose the marking function $\boldsymbol{m}$ is not known explicitly.

One observes only some realizations of the marked point process $\tilde{\boldsymbol{\Phi}}$ with points restricted to some finite observation window $\boldsymbol{W}$. Denote these realizations by $\tilde{\phi}_k = \sum_i \delta_{(x_i(k),m_i(k))}$, with $x_i(k) \in W$, $k = 1, \ldots$.

The problem consists in learning the function $\boldsymbol{m}$ so as to be able to calculate approximations of the unobserved marks $m_i = m(x_i, \phi)$ for a new realization $\phi = \sum_i \delta_{x_i}$ of (only points) of the point process $\boldsymbol{\Phi}$.

Learning and reconstructing in a nut-shell



Learning and reconstructing marks with scattering transforms

Recall the problem: the marks $m_i = m(x_i, \phi)$ of the observed marked point patterns $\tilde{\phi} = \sum_i \delta_{x_i, m_i}$ are produced by unknown function $m = m(x, \phi)$, which one wants to learn from data.

Learning and reconstructing marks with scattering transforms

Recall the problem: the marks $m_i = m(x_i, \phi)$ of the observed marked point patterns $\tilde{\phi} = \sum_i \delta_{x_i, m_i}$ are produced by unknown function $m = m(x, \phi)$, which one wants to learn from data.

- Learning: Capture the relation between the marks and the points through the relation between the vector of the (say first order) empirical scattering moments $\hat{\mathbf{S}}\tilde{\phi} := (\hat{S}\tilde{\phi}(j, \theta) : j, \theta)$ of the marked point patterns $\tilde{\phi}$ and the vector of (say first- and second-order) empirical moments $\hat{\mathbf{S2}}\phi := (\hat{S}\phi(j_1, \theta_1, j_2, \theta_2) : j_1, \theta_1, j_2, \theta_2)$ of the non-marked one ϕ. This relation can be established established using some regression model (e.g., linear ridge regression) on the training data set, where points and marks are observed.

Learning and reconstructing marks with scattering transforms

Recall the problem: the marks $m_i = m(x_i, \phi)$ of the observed marked point patterns $\tilde{\phi} = \sum_i \delta_{x_i, m_i}$ are produced by unknown function $m = m(x, \phi)$, which one wants to learn from data.

- Learning: Capture the relation between the marks and the points through the relation between the vector of the (say first order) empirical scattering moments $\hat{\mathbf{S}}\tilde{\phi} := (\hat{S}\tilde{\phi}(j,\theta) : j, \theta)$ of the marked point patterns $\tilde{\phi}$ and the vector of (say first- and second-order) empirical moments $\hat{\mathbf{S2}}\phi := (\hat{S}\phi(j_1, \theta_1, j_2, \theta_2) : j_1, \theta_1, j_2, \theta_2)$ of the non-marked one ϕ. This relation can be established established using some regression model (e.g., linear ridge regression) on the training data set, where points and marks are observed.
- Reconstructing: Estimate marks $m_i = m(x_i, \phi)$ of a new configuration, where only points are observed, (numerically) solving an inverse problem, where marks are reconstructed from the estimated (regressed) scattering moments.

Linear regression — brief reminder of a well known statistical approach

Let $\boldsymbol{X_k} := \hat{\mathbf{S}}\mathbf{2}\phi_{\boldsymbol{k}}$ and $\boldsymbol{Y_k} := \hat{\mathbf{S}}\tilde{\phi}_{\boldsymbol{k}}$, $\boldsymbol{k} = \mathbf{1}, \ldots, \boldsymbol{n}$, be the scattering transforms of $\boldsymbol{n}$ realizations of marked point patterns $\tilde{\phi}_{\boldsymbol{k}}$. (In $\boldsymbol{X_k}$ only points are considered, while in $\boldsymbol{Y_k}$ points with marks).

One is looking for a common, linear relation between $\boldsymbol{X_k}$ and $\boldsymbol{Y_k}$ for all samples $\boldsymbol{k}$, represented by some matrix $\mathbb{B}$ and vector $\boldsymbol{\beta_0}$ such that

$$\mathbb{B}\boldsymbol{X_k} + \boldsymbol{\beta_0} \approx \boldsymbol{Y_k} \quad \text{for all } \boldsymbol{k} = \mathbf{1}, \ldots, \boldsymbol{n}. \tag{1}$$

Denote by $\beta(p)$ the p th line of the matrix $\mathbb{B}$ in (1). For $p = (j, \theta)$ it corresponds to the scattering moment in Y_k at scale 2^j and angle θ. Similarly, let $\beta_0(p)$ be the p th component for the vector β_0. Let $Y_k(p) := S\tilde{\phi}_k(p)$ be the $p = (j, \theta)$-component of $\mathbf{S}_k$. The linear ridge model consists in minimizing the regularized sum of the squared residuals

$$\sum_{k=1}^{n} [\beta(p) X_k + \beta_0(p) - Y_k(p)]^2 + \lambda(p) ||\beta(p)||^2 \,,$$

for some (Tikhonov) regularization parameters $\lambda(p) \geq 0$ (usually needed in high dimensional regression problems). These parameters are chosen (by the cross-validation) to minimize this squared residuals on the validation set, a subset of the training set. where $||\cdot||$ is the Euclidean norm.

Solution of the linear ridge regression — brief reminder

The linear ridge regression problem admits explicit solution

$$[\hat{\beta}_0(p), \hat{\beta}(p)]^\top := (\mathbb{X}^\top \mathbb{X} + \lambda(p)\mathbf{I})^{-1} \mathbb{X}^\top \mathbb{Y}(p), \qquad (2)$$

where $\mathbb{X}$ is the matrix with lines X_k appended with the first column of 1's, $\mathbb{Y}(p)$ is the column vector with elements $Y_k(p)$, $k = 1, \ldots, n$, $\mathbf{I}$ is the appropriate identity matrix and $^\top$ stands for the matrix transpose.

The linear ridge regression problem admits explicit solution

$$[\hat{\beta}_0(p), \hat{\beta}(p)]^\top := (\mathbb{X}^\top \mathbb{X} + \lambda(p)\mathbf{I})^{-1}\mathbb{X}^\top \mathbb{Y}(p), \tag{2}$$

where $\mathbb{X}$ is the matrix with lines X_k appended with the first column of 1's, $\mathbb{Y}(p)$ is the column vector with elements $Y_k(p)$, $k = 1, \ldots, n$, $\mathbf{I}$ is the appropriate identity matrix and $^\top$ stands for the matrix transpose.

Using (2) one can calculate estimates $\hat{\hat{S}}\tilde{\phi}$ of the empirical, (marked) scattering moments of a new configuration $\tilde{\phi} = \sum_i \delta_{(x_i)}$ for which only points are given, by using its empirical, non-marked scattering moments $\hat{\mathbf{S}}2\phi$

$$\hat{\hat{S}}\tilde{\phi}(p) := \hat{\beta}(p)\,\hat{\mathbf{S}}2\phi + \hat{\beta}_0(p)$$

Solution of the linear ridge regression — brief reminder

The linear ridge regression problem admits explicit solution

$$[\hat{\beta}_0(p), \hat{\beta}(p)]^\top := (\mathbb{X}^\top \mathbb{X} + \lambda(p)\mathbf{I})^{-1} \mathbb{X}^\top \mathbb{Y}(p), \qquad (2)$$

where $\mathbb{X}$ is the matrix with lines X_k appended with the first column of 1's, $\mathbb{Y}(p)$ is the column vector with elements $Y_k(p)$, $k = 1, \ldots, n$, $\mathbf{I}$ is the appropriate identity matrix and $^\top$ stands for the matrix transpose.

Using (2) one can calculate estimates $\hat{\hat{S}}\tilde{\phi}$ of the empirical, (marked) scattering moments of a new configuration $\tilde{\phi} = \sum_i \delta_{(x_i)}$ for which only points are given, by using its empirical, non-marked scattering moments $\hat{\mathbf{S2}}\phi$

$$\hat{\hat{S}}\tilde{\phi}(p) := \hat{\beta}(p)\, \hat{\mathbf{S2}}\phi + \hat{\beta}_0(p)$$

Remember, expression (2) requires tuning of the regularization parameters $\lambda(p) \geq 0$ usually needed in high dimensional regression problems when the matrix $\mathbb{X}^\top \mathbb{X}$ is not invertible.

Knowing non-marked configuration $\phi = \sum_i \delta_{x_i}$ and having calculated approximations $\hat{\hat{\mathbf{S}}}\tilde{\phi}$ of its marked scattering moments, we estimate unknown marks $m_i = m(x_i, \phi)$ by looking for a solution to the following minimization problem

$$\arg \min_{\tilde{\phi}' : \phi' = \phi} ||\hat{\mathbf{S}}\tilde{\phi}' - \hat{\hat{\mathbf{S}}}\tilde{\phi}||^2, \qquad (3)$$

where we minimize over all arbitrarily marked configurations $\tilde{\phi}'$ sharing the points with given ϕ (hence over unknown marks) and $\hat{\mathbf{S}}\tilde{\phi}'$ denotes the scattering moment calculated for $\tilde{\phi}'$. It should be noted that (3) is a non convex optimization problem. To solve it we use L-BFGS-B algorithm, see Byrd, Lu, Nocedal (1995). This is a steepest descent algorithm for which it is important to optimize (via cross-validation) the number of iterations.

Examples of geometric marks

Consider $\tilde{\phi} = \sum_i \delta(x_i, m_i)$ with points on the plane $\mathbb{R}^2$ and the following marks

- Voronoi cell surface: $m_i = |A_i| := |\{y \in \mathbb{R}^2 : |y - x_i| \leq \min_{x_j \in \phi} |y - x_j|\}|$.
- Shot-noise: $m_i = S_i := \sum_i \mathbf{1}(x_i \neq x_j)\ell(|x_j - x_i|)$ with the response function $\ell(r) = r^\beta$ for some $\beta > 2$.

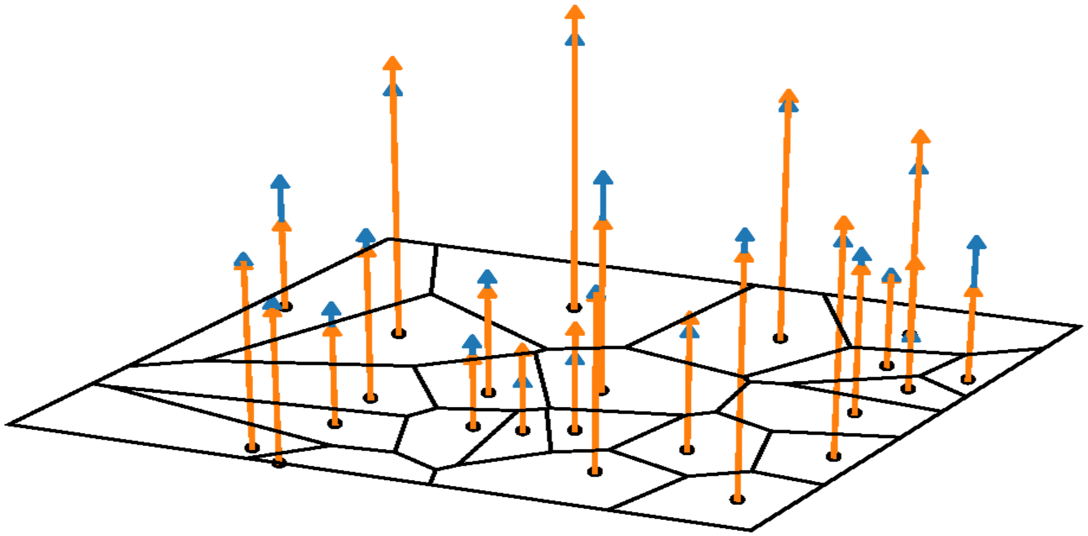
Examples of geometric marks

Consider $\tilde{\phi} = \sum_i \delta(x_i, m_i)$ with points on the plane $\mathbb{R}^2$ and the following marks

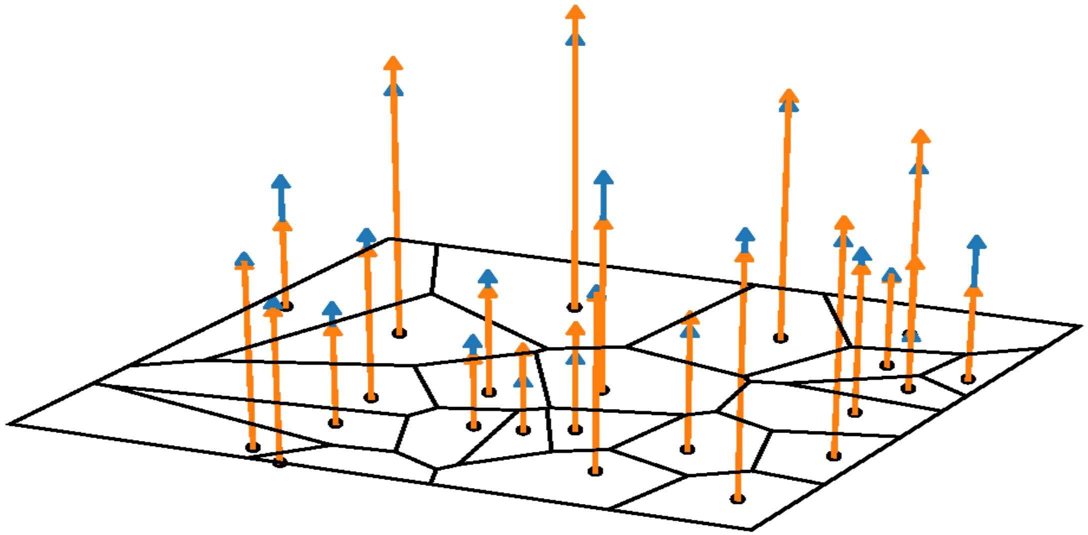
- Voronoi cell surface:
  $m_i = |A_i| := |\{y \in \mathbb{R}^2 : |y - x_i| \leq \min_{x_j \in \phi} |y - x_j|\}|$.
- Shot-noise: $m_i = S_i := \sum_i \mathbf{1}(x_i \neq x_j)\ell(|x_j - x_i|)$ with the response function $\ell(r) = r^\beta$ for some $\beta > 2$.

In what follows we illustrate of the quality of the procedure of learning and prediction of these marks. The training data set — $10\,000$ Poisson point patterns (about 30 points in each) for each model. Regression problem has dimension $1401 \longrightarrow 57$ (number of fist- and second-order scattering moments calculated for non-marked point patterns, and 57 for marked patterns. More details in Brochard, BB, Mallat, Zhang (2019).

Voronoi cell surface area reconstruction example

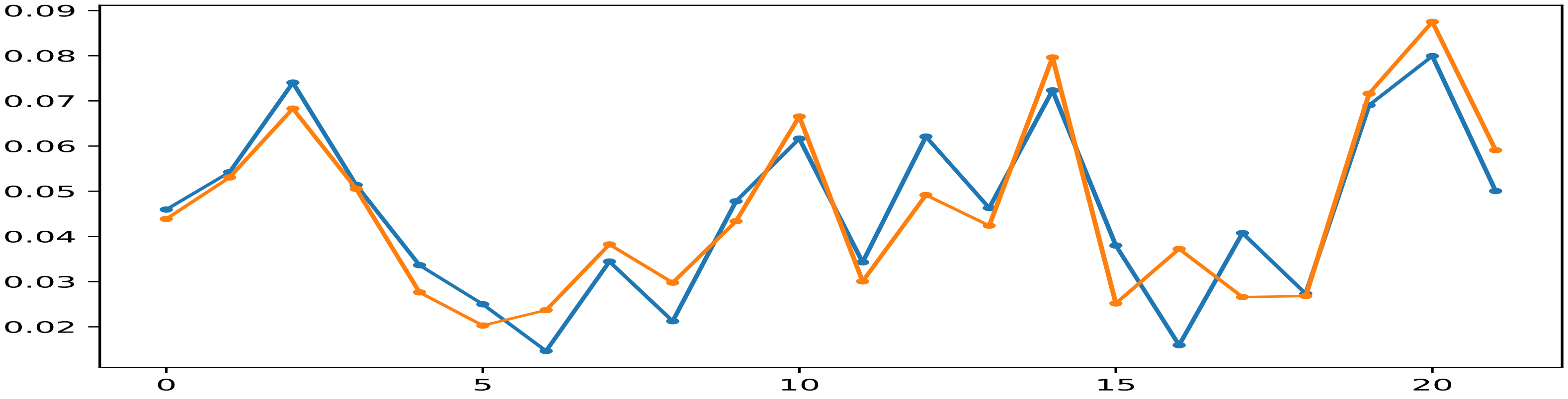


Blue — exact, orange — reconstructed

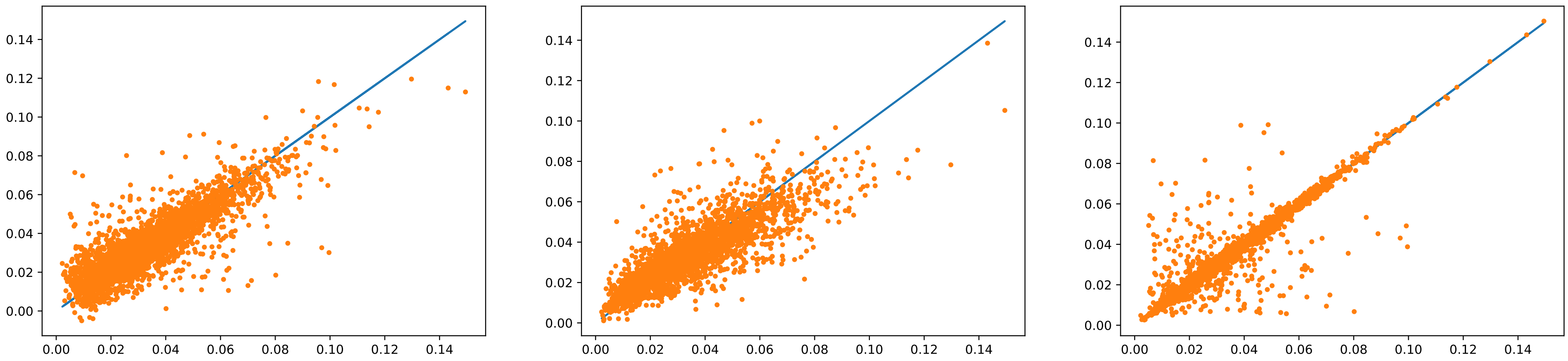


Blue — exact, orange — reconstructed

1D lexicographic representation of points to see better



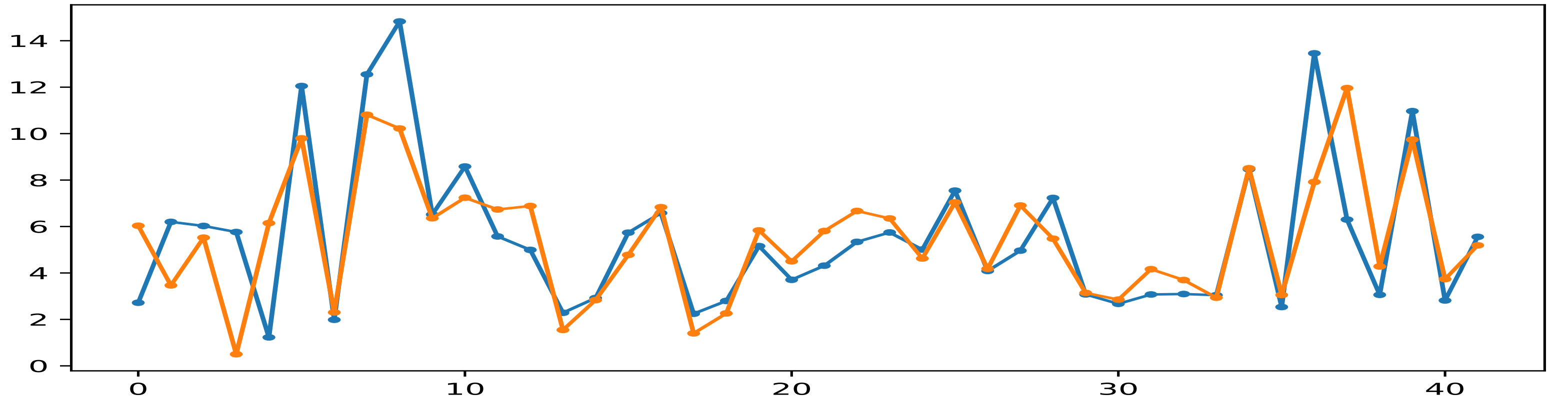
Voronoi cell surface area Q-Q plots



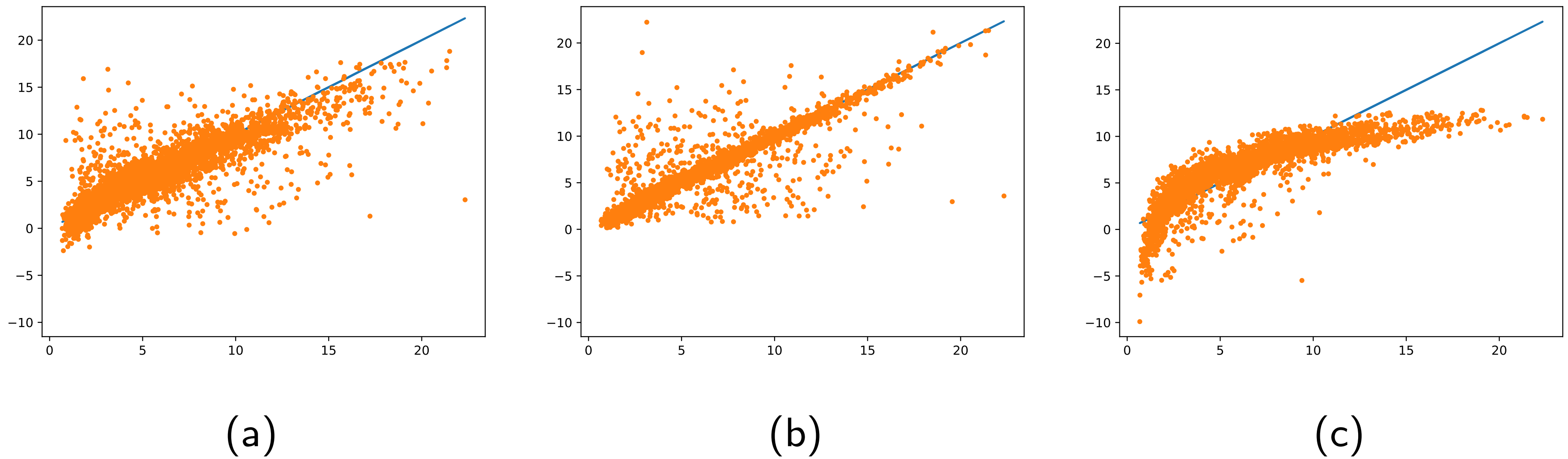
(a) (b) (c)

Reconstructed marks vs true values for all points of 100 test images. Reconstruction from (a) estimated moments, (b) exact moments, (c) benchmark based on distance matrix representation.

Shot-noise



Reconstructed image example.



Q-Q plot: (a) estimated moments, (b) exact moments, (c) benchmark.

Conclusions

- Scattering moments are nonlinear and noncommuting operators, computing at different scales the modulus of a wavelet transform.
- At small scale their capture the intensity of point processes, at large scale their variance scaling.
- Numerical evidences confirm that they can capture (in statistically exploitable way) existing rigidity. For example, they allow one to estimate values of marks given locations of points.
- More work required to understand better these, apparently statistically useful, operators.

References II

- ☐ Byrd, Lu, J. Nocedal (1995) “A limited memory algorithm for bound constrained optimization”, *SIAM J. Sci. Stat. Comput*,
- ☐ S. Mallat (2012) “Group invariant scattering” *Commun Pur Appl Math*,
- ☐ Bruna, Mallat, Bacry, Muzy (2015), “Intermittent process analysis with scattering moments”, *Ann Stat*,
- ☐ Klatt, Last, Yogeshwaran (2018) “Hyperuniform and rigid stable matchings”, arXiv:1810.00265
- ☐ BB, Yogeshwaran, Yukich (2019) “Limit theory for geometric statistics of point processes having fast decay of correlations”, *Ann Probab*,
- ⇒ Brochard, BB, Mallat, Zhang (2019) “Statistical learning of geometric characteristics of wireless networks”, in *Proc. INFOCOM 2019*,
- ⇒ Brochard, BB, Mallat, Zhang, “Limit results for scattering moments”, article in preparation.

Thank you.