Limit Theory for Statistics of Random Geometric Structures

Joe Yukich

Lehigh University

Stochastic Geometry and Its Applications, Nantes, April 4-8, 2016

Introduction

Talk is based on joint work with B. Błaszczyszyn and D. Yogeshwaran

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

The sums describe some global feature of the random system in terms of local contributions $\xi(x, \mathcal{X}),\ x \in \mathcal{X}$.

Introduction

Questions pertaining to geometric structures on random input $\mathcal{X} \subset \mathbb{R}^d$ often involve analyzing sums of spatially correlated terms

$$\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X}),$$

where the $\mathbb{R}$-valued score function ξ, defined on pairs $(x, \mathcal{X})$, represents the interaction of x with respect to $\mathcal{X}$.

The sums describe some global feature of the random system in terms of local contributions $\xi(x, \mathcal{X}),\ x \in \mathcal{X}$.

Ex. 1: Statistics of random graphs

Clique counts. $\mathcal{X} \subset \mathbb{R}^d$ finite, $r \in (0, \infty)$.

$\cdot$ Join two points of $\mathcal{X}$ iff they are at distance at most r. Vietoris-Rips complex (with parameter r) is simplicial complex whose k-simplices correspond to unordered $(k+1)$-tuples of points in $\mathcal{X}$ all pairwise within r of each other.

$\cdot$ For $k \in \mathbb{N}$ and $x \in \mathcal{X}$, put $\sigma_k(x, \mathcal{X}) := \frac{\text{number of } k\text{-simplices containing } x}{k+1}$

Ex. 1: Statistics of random graphs

Clique counts. $\mathcal{X} \subset \mathbb{R}^d$ finite, $r \in (0, \infty)$.

· Join two points of $\mathcal{X}$ iff they are at distance at most r. Vietoris-Rips complex (with parameter r) is simplicial complex whose k-simplices correspond to unordered $(k+1)$-tuples of points in $\mathcal{X}$ all pairwise within r of each other.

· For $k \in \mathbb{N}$ and $x \in \mathcal{X}$, put $\sigma_k(x, \mathcal{X}) := \frac{\text{number of } k\text{-simplices containing } x}{k+1}$

· Total number of k-simplices in Vietoris-Rips complex: $\sum_{x \in \mathcal{X}} \sigma_k(x, \mathcal{X})$.

Chatterjee, Decreusefond et al., Eichelsbacher, Lachièze-Rey + Peccati, Reitzner + Schulte, Penrose + Y

Ex. 1: Statistics of random graphs

Total edge length of graphs. $\mathcal{X} \subset \mathbb{R}^d$ finite. Given $x \in \mathcal{X}$, let x_{NN} be the nearest neighbor of x.

· Undirected nearest neighbor graph on $\mathcal{X}$: include an edge $\{x, y\}$ if $y = x_{NN}$ and/or $x = y_{NN}$.

· For $x \in \mathcal{X}$, put

$$\xi(x, \mathcal{X}) := \begin{cases} \frac{1}{2}||x - x_{NN}|| & \text{if } \ x, x_{NN} \ \text{ are mutual n.n.} \\ ||x - x_{NN}|| & \text{otherwise.} \end{cases}$$

Ex. 1: Statistics of random graphs

Total edge length of graphs. $\mathcal{X} \subset \mathbb{R}^d$ finite. Given $x \in \mathcal{X}$, let x_{NN} be the nearest neighbor of x.

· Undirected nearest neighbor graph on $\mathcal{X}$: include an edge $\{x, y\}$ if $y = x_{NN}$ and/or $x = y_{NN}$.

· For $x \in \mathcal{X}$, put

$$\xi(x, \mathcal{X}) := \begin{cases} \frac{1}{2}||x - x_{NN}|| & \text{if} \quad x, x_{NN} \quad \text{are mutual n.n.} \\ ||x - x_{NN}|| & \text{otherwise.} \end{cases}$$

· Total edge length of n.n. graph on $\mathcal{X}$: $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$.

Chatterjee; Last, Peccati, + Schulte; Steele; Penrose + Y

Ex. 2: Germ-grain models

· $\mathcal{X} \subset \mathbb{R}^d$ a collection of 'germs'.

· $S_x, x \in \mathcal{X}$, a collection of 'grains' (closed bounded sets).

· Germ-grain model: $\bigcup_{x \in \mathcal{X}} (x \oplus S_x)$.

· Surface area, Euler characteristic, clump count,... may be expressed as $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$ for appropriate ξ. For example, for $x \in \mathcal{X}$ we put

$$\xi_{\text{clump}}(x, \mathcal{X}) := (\text{size of clump of germ-grain model containing } x)^{-1}.$$

Ex. 2: Germ-grain models

· $\mathcal{X} \subset \mathbb{R}^d$ a collection of 'germs'.

· $S_x, x \in \mathcal{X}$, a collection of 'grains' (closed bounded sets).

· Germ-grain model: $\bigcup_{x \in \mathcal{X}} (x \oplus S_x)$.

· Surface area, Euler characteristic, clump count,... may be expressed as $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$ for appropriate ξ. For example, for $x \in \mathcal{X}$ we put

$$\xi_{\text{clump}}(x, \mathcal{X}) := (\text{size of clump of germ-grain model containing } x)^{-1}.$$

· Clump count in germ-grain model equals $\sum_{x \in \mathcal{X}} \xi_{\text{clump}}(x, \mathcal{X})$.

· Baddeley; Hall; Hug, Last + Schulte; Molchanov; Penrose + Y; Schneider + Weil; Stoyan;...

Ex. 3: Random packing (Random sequential adsorption)

· $\mathcal{X} \subset \mathbb{R}^d$ finite. Assign elements $x \in \mathcal{X}$ time marks $\tau_x \in [0, 1]$.

· Let $B_1, B_2, ...$ be a sequence of unit volume d-dimensional Euclidean balls with centers arriving sequentially at points $x \in \mathcal{X}$ and at arrival times τ_x.

· The first ball B_1 to arrive is packed. Recursively, for $i = 2, 3, ...$, the ith ball is packed if it does not overlap any ball in $B_1, B_2, ..., B_{i-1}$ which has already been packed.

Ex. 3: Random packing (Random sequential adsorption)

· $\mathcal{X} \subset \mathbb{R}^d$ finite. Assign elements $x \in \mathcal{X}$ time marks $\tau_x \in [0, 1]$.

· Let $B_1, B_2, \ldots$ be a sequence of unit volume d-dimensional Euclidean balls with centers arriving sequentially at points $x \in \mathcal{X}$ and at arrival times τ_x.

· The first ball B_1 to arrive is packed. Recursively, for $i = 2, 3, \ldots$, the ith ball is packed if it does not overlap any ball in $B_1, B_2, \ldots, B_{i-1}$ which has already been packed.

· For $x \in \mathcal{X}$ define packing functional

$$\rho(x, \mathcal{X}) := \begin{cases} 1 & \text{if ball arriving at x is packed} \\ 0 & \text{otherwise} \end{cases},$$

Then total number of packed balls equals $\sum_{x \in \mathcal{X}} \rho(x, \mathcal{X})$.

· Rényi, Coffman, Dvoretzky + Robbins; Flory, Itoh + Shepp; Torquato,...

Ex. 4: Statistics of random convex hulls

· $\mathcal{X} \subset \mathbb{R}^d$ finite. Let $\mathrm{co}(\mathcal{X})$ denote the convex hull of $\mathcal{X}$.

· For $x \in \mathcal{X}$, $k \in \{0, 1, ..., d-1\}$, we put

$f_k(x, \mathcal{X}) := \frac{1}{k+1}(\text{number of k} - \text{dimensional faces containing x}).$

Ex. 4: Statistics of random convex hulls

· $\mathcal{X} \subset \mathbb{R}^d$ finite. Let $\mathrm{co}(\mathcal{X})$ denote the convex hull of $\mathcal{X}$.

· For $x \in \mathcal{X}$, $k \in \{0, 1, ..., d-1\}$, we put

$f_k(x, \mathcal{X}) := \frac{1}{k+1}(\mathrm{number\ of\ k-dimensional\ faces\ containing\ x}).$

· Total number of k-dimensional faces of $\mathrm{co}(\mathcal{X})$ equals $\sum_{x \in \mathcal{X}} f_k(x, \mathcal{X})$.

· Rényi + Sulanke; Bárány; Buchta; Calka, Schreiber + Y; Groeneboom, Reitzner, Vu,...

General questions

· When $\mathcal{X} \subset \mathbb{R}^d$ is a random pt configuration, the sums $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$ describe a global feature of some spatial random system.

General questions

· When $\mathcal{X} \subset \mathbb{R}^d$ is a random pt configuration, the sums $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$ describe a global feature of some spatial random system.

· **Question.** What is the distribution of these sums for large pt configurations $\mathcal{X}$? LLN? CLT?

Goals

$\mathcal{P}$: stationary pt process on $\mathbb{R}^d$

Restrict to windows: $\mathcal{P}_n := \mathcal{P} \cap [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$.

Goals

$\mathcal{P}$: stationary pt process on $\mathbb{R}^d$

Restrict to windows: $\mathcal{P}_n := \mathcal{P} \cap [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$.

Goal. Given a score function $\xi(\cdot,\cdot)$ defined on pairs $(x, \mathcal{X})$, given a pt process $\mathcal{P}$, we seek the limit theory (LLN, CLT, variance asymptotics) for the total score

$$\sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n)$$

and total measure

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

Tractable problems must be *local* in the sense that points far away from x should not play a role in the evaluation of the score $\xi(x, \mathcal{P}_n)$.

Stabilization

We assume translation invariant scores: $\xi(x, \mathcal{X}) = \xi(\mathbf{0}, \mathcal{X} - x)$.

Recall $\mathcal{P}_n := \mathcal{P} \cap [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$

Stabilization

We assume translation invariant scores: $\xi(x, \mathcal{X}) = \xi(\mathbf{0}, \mathcal{X} - x)$.

Recall $\mathcal{P}_n := \mathcal{P} \cap [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$

Key Definition. ξ is *stabilizing* wrt pt process $\mathcal{P}$ on $\mathbb{R}^d$ if for all $x \in \mathcal{P}$ there is $R := R^{\xi}(x, \mathcal{P}) < \infty$ a.s. (a 'radius of stabilization') such that

$$\xi(x, \mathcal{P} \cap B_R(x)) = \xi(x, \mathcal{P} \cap B_R(x) \cup (\mathcal{A} \cap B_R^c(x))).$$

for any locally finite $\mathcal{A} \subset \mathbb{R}^d$. ξ is *exponentially stabilizing* wrt $\mathcal{P}$ if there is a constant c such that

$$\sup_{x \in \mathbb{R}^d} \sup_n P[R^{\xi}(x, \mathcal{P}_n) \geq r] \leq c \exp(-\frac{r}{c}), \quad r \in [1, \infty).$$

Moment condition

$\mathcal{P}$: a pt process on $\mathbb{R}^d$; $\mathcal{P}_n := \mathcal{P} \cap [-\frac{n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$.

Definition. ξ satisfies the p moment condition wrt $\mathcal{P}$ if

$$\sup_n \sup_{x,y \in \mathbb{R}^d} \mathbb{E}\,|\xi(x, \mathcal{P}_n \cup \{y\})|^p < \infty.$$

Weak law of large numbers for Poisson input $\mathcal{H}$

Let $\mathcal{H}$ be a rate 1 Poisson pt process on $\mathbb{R}^d$; $\mathcal{H}_n := \mathcal{H} \cap [\frac{-n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$.

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

Thm (WLLN): If ξ is stabilizing wrt $\mathcal{H}$, if ξ satisfies the p moment condition for some $p \in (1, \infty)$, then for all $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ we have

$$|n^{-1}\mathbb{E}\,\langle \mu_n^\xi, f\rangle - \mathbb{E}\,\xi(\mathbf{0}, \mathcal{H} \cup \{\mathbf{0}\}) \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f(x)dx| \le \epsilon_n.$$

Weak law of large numbers for Poisson input $\mathcal{H}$

Let $\mathcal{H}$ be a rate 1 Poisson pt process on $\mathbb{R}^d$; $\mathcal{H}_n := \mathcal{H} \cap [\frac{-n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$.

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

Thm (WLLN): If ξ is stabilizing wrt $\mathcal{H}$, if ξ satisfies the p moment condition for some $p \in (1, \infty)$, then for all $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ we have

$$|n^{-1} \mathbb{E} \langle \mu_n^\xi, f \rangle - \mathbb{E}\, \xi(\mathbf{0}, \mathcal{H} \cup \{\mathbf{0}\}) \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f(x) dx| \leq \epsilon_n.$$

Penrose and Y (2003): $\epsilon_n = o(1)$.

Schulte + Y (2016): $\epsilon_n = O(n^{-1/d})$ if ξ is exponentially stabilizing wrt $\mathcal{H}$.

Gaussian fluctuations for Poisson input $\mathcal{H}$ on $\mathbb{R}^d$

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

Thm (CLT): Assume ξ is exponentially stabilizing wrt $\mathcal{H}$ and that ξ satisfies the p moment condition for some $p \in (5, \infty)$. If $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ satisfies $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n)$, then

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E}\langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \leq t \right] - P[N(0,1) \leq t] \right| \leq \epsilon_n.$$

Gaussian fluctuations for Poisson input $\mathcal{H}$ on $\mathbb{R}^d$

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

Thm (CLT): Assume ξ is exponentially stabilizing wrt $\mathcal{H}$ and that ξ satisfies the p moment condition for some $p \in (5, \infty)$. If $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ satisfies $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n)$, then

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E}\,\langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \le t \right] - P[N(0,1) \le t] \right| \le \epsilon_n.$$

Penrose + Y (2005), Penrose (2007): $\epsilon_n = O((\log n)^{3d} n^{-1/2})$.

Gaussian fluctuations for Poisson input $\mathcal{H}$ on $\mathbb{R}^d$

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

Thm (CLT): Assume ξ is exponentially stabilizing wrt $\mathcal{H}$ and that ξ satisfies the p moment condition for some $p \in (5, \infty)$. If $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ satisfies $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n)$, then

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E}\langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \leq t \right] - P[N(0,1) \leq t] \right| \leq \epsilon_n.$$

Penrose + Y (2005), Penrose (2007): $\epsilon_n = O((\log n)^{3d} n^{-1/2})$.

Last, Peccati + Schulte (2016): $\epsilon_n = \gamma_1 + \gamma_2 + ... + \gamma_5$.

Gaussian fluctuations for Poisson input $\mathcal{H}$ on $\mathbb{R}^d$

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

Thm (CLT): Assume ξ is exponentially stabilizing wrt $\mathcal{H}$ and that ξ satisfies the p moment condition for some $p \in (5, \infty)$. If $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ satisfies $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n)$, then

$$\sup_{t \in \mathbb{R}} \left| P\left[\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E} \langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \leq t \right] - P[N(0,1) \leq t] \right| \leq \epsilon_n.$$

Penrose + Y (2005), Penrose (2007): $\epsilon_n = O((\log n)^{3d} n^{-1/2})$.

Last, Peccati + Schulte (2016): $\epsilon_n = \gamma_1 + \gamma_2 + ... + \gamma_5$.

Lachièze-Rey, Schulte, + Y (2016): $\epsilon_n = O(n^{-1/2})$.

Variance asymptotics for Poisson input; volume order fluctuations

Given homogenous rate 1 Poisson input $\mathcal{H}$ on $\mathbb{R}^d$, and a score ξ, put

$$\sigma^2(\xi) := \mathbb{E}\,\xi^2(\mathbf{0}, \mathcal{H}) + \int_{\mathbb{R}^d} \mathbb{E}\,\xi(\mathbf{0}, \mathcal{H} \cup \{x\})\xi(x, \mathcal{H} \cup \{\mathbf{0}\}) - \mathbb{E}\,\xi(\mathbf{0}, \mathcal{H})\mathbb{E}\,\xi(x, \mathcal{H})dx$$

Variance asymptotics for Poisson input; volume order fluctuations

Given homogenous rate 1 Poisson input $\mathcal{H}$ on $\mathbb{R}^d$, and a score ξ, put

$$\sigma^2(\xi) := \mathbb{E}\,\xi^2(\mathbf{0},\mathcal{H}) + \int_{\mathbb{R}^d} \mathbb{E}\,\xi(\mathbf{0},\mathcal{H}\cup\{x\})\xi(x,\mathcal{H}\cup\{\mathbf{0}\}) - \mathbb{E}\,\xi(\mathbf{0},\mathcal{H})\mathbb{E}\,\xi(x,\mathcal{H})dx$$

Thm (variance asymptotics): If ξ is exponentially stabilizing wrt $\mathcal{H}$, if ξ satisfies the p moment condition for some $p \in (2,\infty)$, then for all $f \in B([-\frac{1}{2},\frac{1}{2}]^d)$ we have

$$\lim_{n\to\infty} n^{-1}\mathrm{Var}\langle \mu_n^\xi, f\rangle = \sigma^2(\xi)\int_{[-\frac{1}{2},\frac{1}{2}]^d} f^2(x)dx \in [0,\infty).$$

Baryshnikov + Y. (2005); Penrose (2007)

· **Question.** If the input pt process is not Poisson, when do we get results which are qualitatively similar?

· Soshnikov (2002): establishes asymptotic normality of the *linear* statistics

$$\sum_{x \in \mathcal{P}_n} \delta_{n^{-1/d}x}$$

where $\mathcal{P}$ is determinantal pt process, $\mathcal{P}_n := \mathcal{P} \cap W_n$.

· **Question.** If the input pt process is not Poisson, when do we get results which are qualitatively similar?

· Soshnikov (2002): establishes asymptotic normality of the *linear* statistics

$$\sum_{x \in \mathcal{P}_n} \delta_{n^{-1/d}x}$$

where $\mathcal{P}$ is determinantal pt process, $\mathcal{P}_n := \mathcal{P} \cap W_n$.

· Nazarov and Sodin (2012): establish asymptotic normality of the *linear* statistics

$$\sum_{x \in \mathcal{P}_n} \delta_{n^{-1/d}x}$$

where $\mathcal{P}$ is zero set of Gaussian analytic function, $\mathcal{P}_n := \mathcal{P} \cap W_n$.

· We want to extend these results to non-linear statistics

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

Clustering pt processes

Def. Given a simple pt process $\mathcal{P}$ on $\mathbb{R}^d$, the k pt correlation function $\rho^{(k)} : (\mathbb{R}^d)^k \to [0, \infty)$ is defined via

$$\mathbb{E}\left[\Pi_{i=1}^k \mathrm{card}(\mathcal{P} \cap B_i)\right] = \int_{B_1} ... \int_{B_k} \rho^{(k)}(x_1, ..., x_k) dx_1 ... dx_k,$$

where $B_1, ..., B_k$ are disjoint subsets of $\mathbb{R}^d$.

Rk. $\rho^{(k)}(x_1, ..., x_k) = \Pi_{i=1}^k \rho^{(1)}(x_i)$ characterizes the Poisson pt process

Clustering pt processes

Def. Given a simple pt process $\mathcal{P}$ on $\mathbb{R}^d$, the k pt correlation function $\rho^{(k)} : (\mathbb{R}^d)^k \to [0, \infty)$ is defined via

$$\mathbb{E}\left[\Pi_{i=1}^k \text{card}(\mathcal{P} \cap B_i)\right] = \int_{B_1} ... \int_{B_k} \rho^{(k)}(x_1, ..., x_k) dx_1 ... dx_k,$$

where $B_1, ..., B_k$ are disjoint subsets of $\mathbb{R}^d$.

Rk. $\rho^{(k)}(x_1, ..., x_k) = \Pi_{i=1}^k \rho^{(1)}(x_i)$ characterizes the Poisson pt process

Key Definition A pt process $\mathcal{P}$ *clusters* if there is a fast decreasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $p, q \in \mathbb{N}$ there are constants $c_{p,q}$ and $C_{p,q}$ such that for all $x_1, ..., x_{p+q} \in \mathbb{R}^d$,

$$|\rho^{(p+q)}(x_1, ..., x_{p+q}) - \rho^{(p)}(x_1, ..., x_p)\rho^{(q)}(x_{p+1}, ..., x_{p+q})| \leq C_{p,q}\phi(-c_{p,q}s),$$

where $s := \inf_{i \in \{1,...,p\},\ j \in \{p+1,...,p+q\}} |x_i - x_j|$.

(ϕ 'fast decreasing' means ϕ decaying faster than any power)

Ex. 1: Determinantal pt process

A pt process is determinantal (DPP) if its correlation functions satisfy

$$\rho^{(k)}(x_1, ..., x_k) = \det(K(x_i, x_j))_{1 \leq i \leq j \leq k},$$

where $K(\cdot, \cdot)$ is Hermitian kernel of integral operator from $L^2(\mathbb{R}^d)$ to itself.

Ex. 1: Determinantal pt process

A pt process is determinantal (DPP) if its correlation functions satisfy

$$\rho^{(k)}(x_1, ..., x_k) = \det(K(x_i, x_j))_{1 \leq i \leq j \leq k},$$

where $K(\cdot, \cdot)$ is Hermitian kernel of integral operator from $L^2(\mathbb{R}^d)$ to itself.

Fact (Błaszczyszyn, Yogeshwaran, + Y (2016)). If $|K(x, y)| \leq \phi(||x - y||)$, with ϕ fast decreasing, then the DPP clusters.

Ex. Infinite Ginibre ensemble on complex plane clusters with kernel

$$K(z_1, z_2) = \exp(iIm(z_1 \bar{z}_2) - \frac{1}{2}|z_1 - z_2|^2).$$

Ex. 2: Zero set of Gaussian entire function

· Let $X_j, j \geq 1$, be i.i.d. standard complex Gaussians. Consider the Gaussian entire function

$$F(z) := \sum_{j=1}^{\infty} \frac{X_j}{\sqrt{j!}} z^j.$$

· Zero set $Z_F := F^{-1}(\{0\})$ is stationary.

Ex. 2: Zero set of Gaussian entire function

· Let $X_j, j \geq 1$, be i.i.d. standard complex Gaussians. Consider the Gaussian entire function

$$F(z) := \sum_{j=1}^{\infty} \frac{X_j}{\sqrt{j!}} z^j.$$

· Zero set $Z_F := F^{-1}(\{0\})$ is stationary.

· Z_F exhibits local repulsivity.

· Z_F strongly clusters (Nazarov and Sodin (2012)).

Ex. 3: Gibbs pt processes

Consider the class $\mathbf{\Psi}$ of Hamiltonians consisting of:

· pair potentials without negative part,

· area interaction Hamiltonians, and

· hard core Hamiltonians.

· For $\Psi \in \mathbf{\Psi}$, let $\mathcal{P}^{\beta\Psi}$ be the Gibbs pt process having Radon-Nikodym derivative $\exp(-\beta\Psi(\cdot))$ with respect to a reference homogeneous Poisson pt process $\mathcal{H}_\tau$ on $\mathbb{R}^d$ of intensity τ.

· There is a range of inverse temperature and activity parameters (β and τ) such that $\mathcal{P}^{\beta\Psi}$ clusters (Schreiber and Y, 2013).

Weak law of large numbers for clustering input

Let $\mathcal{P}$ be clustering pt process on $\mathbb{R}^d$. Recall $\mathcal{P}_n := \mathcal{P} \cap [\frac{-n^{1/d}}{2}, \frac{n^{1/d}}{2}]^d$ and

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

Thm (BYY '16): If ξ is stabilizing wrt $\mathcal{P}$, if ξ satisfies the p moment condition for some $p \in (1, \infty)$, then for all $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$ we have

$$\lim_{n \to \infty} n^{-1} \mathbb{E} \langle \mu_n^\xi, f \rangle = \mathbb{E}\, \xi(\mathbf{0}, \mathcal{P} \cup \{\mathbf{0}\}) \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f(x) dx \cdot \rho^{(1)}(\mathbf{0}).$$

Gaussian fluctuations for clustering input $\mathcal{P}$

Thm (BYY '16) $\mu_n^\xi := \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n) \delta_{n^{-1/d}x}$. Assume

· $\mathcal{P}$ clusters,

· ξ has deterministic radius of stabilization wrt $\mathcal{P}$,

· ξ satisfies the p moment condition for some $p \in (2, \infty)$, and

· $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n^\alpha)$ for some $\alpha \in (0,1)$, $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$. Then

$$\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E}\,\langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \xrightarrow{\mathcal{D}} N(0,1).$$

Gaussian fluctuations for clustering input $\mathcal{P}$

Thm (BYY '16) $\mu_n^\xi := \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n) \delta_{n^{-1/d}x}$. Assume

· $\mathcal{P}$ clusters,

· ξ has deterministic radius of stabilization wrt $\mathcal{P}$,

· ξ satisfies the p moment condition for some $p \in (2, \infty)$, and

· $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n^\alpha)$ for some $\alpha \in (0,1)$, $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$. Then

$$\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E} \langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \xrightarrow{\mathcal{D}} N(0,1).$$

Remarks. If $\mathcal{P}$ is determinantal with fast decreasing kernel then this extends Soshnikov (2002), who restricts to linear statistics $\sum_{x \in \mathcal{P}_n} \delta_{n^{-1/d}x}$, that is he puts $\xi \equiv 1$.

· If $\mathcal{P}$ is zero set of Gaussian entire function, this extends Nazarov and Sodin (2012), who also restrict to $\sum_{x \in \mathcal{P}_n} \delta_{n^{-1/d}x}$.

Gaussian fluctuations for clustering input $\mathcal{P}$

Thm (BYY '16) $\mu_n^\xi := \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n) \delta_{n^{-1/d}x}$. Assume

· $\mathcal{P}$ clusters and clustering coeff. satisfy mild growth condition

· ξ exponentially stabilizing wrt $\mathcal{P}$,

· ξ satisfies the p moment condition for some $p \in (2, \infty)$, and

· $\mathrm{Var}\langle \mu_n^\xi, f \rangle = \Omega(n^\alpha)$ for some $\alpha \in (0,1)$, $f \in B([-\frac{1}{2}, \frac{1}{2}]^d)$. Then

$$\frac{\langle \mu_n^\xi, f \rangle - \mathbb{E}\, \langle \mu_n^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f \rangle}} \xrightarrow{\mathcal{D}} N(0,1).$$

Remark. If $\mathcal{P}$ is determinantal with fast decreasing kernel (e.g. Ginibre) then $\mathcal{P}$ satisfies stated condition.

Variance asymptotics for clustering input $\mathcal{P}$

- Given clustering input $\mathcal{P}$ and a score ξ, put

$$\sigma^2(\xi) := \mathbb{E}\,\xi^2(\mathbf{0},\mathcal{P})\rho^{(1)}(\mathbf{0})+$$

$$\int_{\mathbb{R}^d} \mathbb{E}\,\xi(\mathbf{0},\mathcal{P}\cup x)\xi(x,\mathcal{P}\cup\mathbf{0})\rho^{(2)}(\mathbf{0},x)-\mathbb{E}\,\xi(\mathbf{0},\mathcal{P})\rho^{(1)}(\mathbf{0})\mathbb{E}\,\xi(x,\mathcal{P})\rho^{(1)}(\mathbf{x})dx.$$

Variance asymptotics for clustering input $\mathcal{P}$

· Given clustering input $\mathcal{P}$ and a score ξ, put

$$\sigma^2(\xi) := \mathbb{E}\,\xi^2(\mathbf{0},\mathcal{P})\rho^{(1)}(\mathbf{0}) +$$

$$\int_{\mathbb{R}^d} \mathbb{E}\,\xi(\mathbf{0},\mathcal{P}\cup x)\xi(x,\mathcal{P}\cup\mathbf{0})\rho^{(2)}(\mathbf{0},x) - \mathbb{E}\,\xi(\mathbf{0},\mathcal{P})\rho^{(1)}(\mathbf{0})\mathbb{E}\,\xi(x,\mathcal{P})\rho^{(1)}(\mathbf{x})dx.$$

· **Thm (BYY '16)**: If ξ is exponentially stabilizing wrt $\mathcal{P}$, if ξ satisfies the p moment condition for some $p \in (2,\infty)$, then for all $f \in B([-\frac{1}{2},\frac{1}{2}]^d)$ we have

$$\lim_{n\to\infty} n^{-1}\mathrm{Var}\langle \mu_n^{\xi}, f\rangle = \sigma^2(\xi)\int_{[-\frac{1}{2},\frac{1}{2}]^d} f^2(x)dx \in [0,\infty).$$

· **Rk.** When $\mathcal{P}$ is determinantal with fast decreasing kernel this extends Soshnikov (2002), who assumes $\xi \equiv 1$.

Proof idea for CLT

· Given ξ, consider k mixed moment functions $m_{(k)} : (\mathbb{R}^d)^k \to \mathbb{R}$ given by

$$m_{(k)}(x_1, ..., x_k; \mathcal{P}_n) := \mathbb{E}\, \Pi_{i=1}^k \xi(x_i, \mathcal{P}_n) \rho^{(k)}(x_1, ..., x_k).$$

· We show that the mixed moments cluster, that is for all $p, q \in \mathbb{N}$ there are constants $c_{p,q}$ and $C_{p,q}$ s.t. for all $x_1, ..., x_{p+q} \in \mathbb{R}^d$,

$$|m_{(p+q)}(x_1, ..., x_{p+q}) - m_{(p)}(x_1, ..., x_p) m_{(q)}(x_{p+1}, ..., x_{p+q})| \leq C_{p,q} \varphi(-c_{p,q} s),$$

where

$$s := \inf_{i \in \{1,...,p\},\ j \in \{p+1,...,p+q\}} |x_i - x_j|$$

and where φ is fast decreasing.

Proof idea for CLT

· Given ξ, consider k mixed moment functions $m_{(k)} : (\mathbb{R}^d)^k \to \mathbb{R}$ given by

$$m_{(k)}(x_1, ..., x_k; \mathcal{P}_n) := \mathbb{E}\, \Pi_{i=1}^k \xi(x_i, \mathcal{P}_n) \rho^{(k)}(x_1, ..., x_k).$$

· We show that the mixed moments cluster, that is for all $p, q \in \mathbb{N}$ there are constants $c_{p,q}$ and $C_{p,q}$ s.t. for all $x_1, ..., x_{p+q} \in \mathbb{R}^d$,

$$|m_{(p+q)}(x_1, ..., x_{p+q}) - m_{(p)}(x_1, ..., x_p) m_{(q)}(x_{p+1}, ..., x_{p+q})| \le C_{p,q} \varphi(-c_{p,q} s),$$

where

$$s := \inf_{i \in \{1,...,p\},\ j \in \{p+1,...,p+q\}} |x_i - x_j|$$

and where φ is fast decreasing.

· $\mathcal{P}$ clusters and ξ exp. stabilizing $\Rightarrow$ mixed moments cluster

Proof idea for CLT

· Given ξ, consider k mixed moment functions $m_{(k)} : (\mathbb{R}^d)^k \to \mathbb{R}$ given by

$$m_{(k)}(x_1, ..., x_k; \mathcal{P}_n) := \mathbb{E}\, \Pi_{i=1}^k \xi(x_i, \mathcal{P}_n) \rho^{(k)}(x_1, ..., x_k).$$

· We show that the mixed moments cluster, that is for all $p, q \in \mathbb{N}$ there are constants $c_{p,q}$ and $C_{p,q}$ s.t. for all $x_1, ..., x_{p+q} \in \mathbb{R}^d$,

$$|m_{(p+q)}(x_1, ..., x_{p+q}) - m_{(p)}(x_1, ..., x_p) m_{(q)}(x_{p+1}, ..., x_{p+q})| \le C_{p,q} \varphi(-c_{p,q} s),$$

where

$$s := \inf_{i \in \{1,...,p\},\ j \in \{p+1,...,p+q\}} |x_i - x_j|$$

and where φ is fast decreasing.

· $\mathcal{P}$ clusters and ξ exp. stabilizing $\Rightarrow$ mixed moments cluster

· cumulants of $\frac{\langle \mu_n^\xi, f\rangle - \mathbb{E}\,\langle \mu_n^\xi, f\rangle}{\sqrt{\mathrm{Var}\langle \mu_n^\xi, f\rangle}}$ converge to cumulants of $N(0,1)$.

Applications

General results yield WLLN, Gaussian fluctuations, variance asymptotics for statistics of geometric structures on clustering pt processes (CPP):

(i) Vietoris-Rips clique count on any CPP, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

... extends Chatterjee, Lachièze-Rey + Peccati, Reitzner + Schulte, Yogeshwaran + Adler (2015)

Applications

General results yield WLLN, Gaussian fluctuations, variance asymptotics for statistics of geometric structures on clustering pt processes (CPP):

(i) Vietoris-Rips clique count on any CPP, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

... extends Chatterjee, Lachièze-Rey + Peccati, Reitzner + Schulte, Yogeshwaran + Adler (2015)

(ii) total volume and surface area of germ-grain model with germs given by points in CPP, i.i.d. grains with bounded diameter.

... extends Penrose + Y (2005); Hug, Last, Schulte (2015)

Applications

General results yield WLLN, Gaussian fluctuations, variance asymptotics for statistics of geometric structures on clustering pt processes (CPP):

(i) Vietoris-Rips clique count on any CPP, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

... extends Chatterjee, Lachièze-Rey + Peccati, Reitzner + Schulte, Yogeshwaran + Adler (2015)

(ii) total volume and surface area of germ-grain model with germs given by points in CPP, i.i.d. grains with bounded diameter.

... extends Penrose + Y (2005); Hug, Last, Schulte (2015)

(iii) total edge length in knn graph on DPP with fast decreasing kernel.

... extends Steele (1988); Penrose + Y (2005); Hug, Last, Schulte (2015)

Applications

General results yield WLLN, Gaussian fluctuations, variance asymptotics for statistics of geometric structures on clustering pt processes (CPP):

(i) Vietoris-Rips clique count on any CPP, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

... extends Chatterjee, Lachièze-Rey + Peccati, Reitzner + Schulte, Yogeshwaran + Adler (2015)

(ii) total volume and surface area of germ-grain model with germs given by points in CPP, i.i.d. grains with bounded diameter.

... extends Penrose + Y (2005); Hug, Last, Schulte (2015)

(iii) total edge length in knn graph on DPP with fast decreasing kernel.

... extends Steele (1988); Penrose + Y (2005); Hug, Last, Schulte (2015)

(iv) the number of k-faces of convex hull of DPP (with fast decreasing kernel) on a volume n ball ... extends Reitzner (2005); Calka, Schreiber + Y (2008, 2013).

Applications

General results yield WLLN, Gaussian fluctuations, variance asymptotics for statistics of geometric structures on clustering pt processes (CPP):

(i) Vietoris-Rips clique count on any CPP, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

... extends Chatterjee, Lachièze-Rey + Peccati, Reitzner + Schulte, Yogeshwaran + Adler (2015)

(ii) total volume and surface area of germ-grain model with germs given by points in CPP, i.i.d. grains with bounded diameter.

... extends Penrose + Y (2005); Hug, Last, Schulte (2015)

(iii) total edge length in knn graph on DPP with fast decreasing kernel.

... extends Steele (1988); Penrose + Y (2005); Hug, Last, Schulte (2015)

(iv) the number of k-faces of convex hull of DPP (with fast decreasing kernel) on a volume n ball ... extends Reitzner (2005); Calka, Schreiber + Y (2008, 2013).

THANK YOU