Asymptotic Analysis of Statistics of Random Geometric Structures

Joe Yukich

Lehigh University

Geometry and Statistics in Data Sciences, Institut Henri Poincaré, September 21, 2022

Collaborators

- Bartek Błaszczyszyn (Paris)
- Yuliy Baryshnikov (Urbana)
- Pierre Calka (Rouen)
- Raphael Lachièze-Rey (Paris)
- Mathew Penrose (Bath)
- Tomasz Schreiber (Torun)
- Matthias Schulte (Hamburg)
- Aihua Xia (Melbourne)
- D. Yogeshwaran (Bangalore)

I Examples, Goals

· Given a point cloud $\mathcal{X}$ sampled from a distribution having a density κ on a manifold $\mathcal{M}$, how can we understand the topology of $\mathcal{M}$ and the properties of κ?

· Can we use interpoint distances between elements of $\mathcal{X}$ to estimate $\dim(\mathcal{M})$?

· Build a simplicial complex over the point cloud. How many holes and connected components does the complex have?

· What is the entropy of κ?

· **Ref.** Chazal and Michel (2021) An introduction to topological data analysis.

I Examples, Goals

· Questions pertaining to geometric structures on finite 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 structure in terms of **local** contributions $\xi(x, \mathcal{X}),\ x \in \mathcal{X}$.

Example: Vietoris-Rips complex

· $\mathcal{X} \subset \mathbb{R}^d$ a finite point cloud, $r \in (0, \infty)$.

· $\mathrm{Rips}^{(r)}(\mathcal{X})$: 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 $\sigma_k^{(r)}(x, \mathcal{X}) := \frac{\text{number of } k\text{-simplices containing } x}{k+1}$

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

· **Goal** Establish limit theory for $\sum_{x \in \Xi \cap W_\lambda} \sigma_k^{(r)}(x, \Xi \cap W_\lambda)$ when Ξ is infinite point process, W_λ is volume λ window, $\lambda \to \infty$.

Example: Rips complex

· $\mathrm{Cech}^{(r)}(\mathcal{X})$: the intersection of the balls of radius r centered at each element of the $(k+1)$-tuple must be non-empty.

$$\mathrm{Rips}^{(r)}(\mathcal{X}) \subseteq \mathrm{Cech}^{(r)}(\mathcal{X}) \subseteq \mathrm{Rips}^{(2r)}(\mathcal{X}).$$

· Rips and $\check{C}$ech complexes are supposed to capture topological information about the point cloud $\mathcal{X}$.

· Chatterjee, Bobrowski + Kahle, Decreusefond et al., Krebs + Polonik, Lachièze-Rey + Peccati, Reitzner + Schulte.

Example: 'r-offset' models

· Boolean model ('r-offset'): $\mathcal{U}(\mathcal{X}, r) = \bigcup_{x \in \mathcal{X}} B_r(x)$.

· How many components and holes are in this model when r ranges over positive reals? (two of the several Betti numbers of the 'r-offset')

$$\xi^{(r)}_{\text{comp}}(x, \mathcal{X}) := \frac{1}{\text{size of component of } \mathcal{U}(\mathcal{X}, r) \text{ containing } x}.$$

· Component count in 'r-offset': $\sum_{x \in \mathcal{X}} \xi^{(r)}_{\text{comp}}(x, \mathcal{X})$.

· **Goal.** Establish limit theory for $\sum_{x \in \Xi \cap W_\lambda} \xi^{(r)}_{\text{comp}}(x, \Xi \cap W_\lambda)$ when Ξ is a pt process on $\mathbb{R}^d$, W_λ is a volume λ window, $\lambda \to \infty$.

Example: Entropy estimators

· The Shannon entropy of the rv X with density κ is given by

$$H(\kappa) := -\int_{\mathcal{M}} \kappa(x) \log(\kappa(x)) dx.$$

·The entropy $H(\kappa)$ is an information theoretic measure of how the data $\mathcal{X}$ is 'spread out'.

· low entropy $\rightarrow$ data is confined to a small volume

· high entropy $\rightarrow$ data is widely dispersed.

Example: Entropy estimators

· Let $\{X_i\}_{i=1}^n$ be i.i.d. on manifold $\mathcal{M}$ with unknown density κ.

· **Goal.** Estimate Shannon entropy of κ given only interpoint distances between points in $\mathcal{X}_n := \{X_i\}_{i=1}^n$.

· It turns out that we may do this via asymptotic analysis of

$$\sum_{i \leq n} \xi_{\text{entropy}}(X_i, \mathcal{X}_n)$$

where ξ_{entropy} is a score function defined in terms of interpoint distances.

· We may similarly find good estimators of the Rényi ρ-entropy of κ, namely

$$\frac{1}{1-\rho} \log \int (\kappa(x))^\rho dx, \ \rho \neq 1.$$

Example: Critical Points

· $\mathcal{X}$ a point cloud in $\mathbb{R}^d$.

· We define the distance function from $\mathcal{X}$ as

$$d_{\mathcal{X}}(y) := \min_{x \in \mathcal{X}} |y - x|, \; y \in \mathbb{R}^d$$

· Let $f : \mathbb{R}^d \to \mathbb{R}$ be a C^2 function. A point $c \in \mathbb{R}^d$ is called a critical point of f if $\nabla f(c) = 0$, and the scalar $f(c)$ is called a critical value of f .

· **Goal.** Study the number of critical points for $d_{\mathcal{X}}$ (but the distance function may not be C^2). Theory of critical points may be extended for such f.

Example: critical points

· Let $\mathcal{X}_n := \{X_i\}_{i=1}^n$ be i.i.d. on $\mathbb{R}^d$. Express the total number of *critical points* in $\mathcal{X}_n$ as a sum of scores

$$\sum_{x \in \mathcal{X}_n} \xi_{\text{crit}}(x, \mathcal{X}_n)$$

through an appropriate choice of ξ_{crit} which depends on local data.

· **Goal.** Establish limit theory for $\sum_{x \in \mathcal{X}_n} \xi_{\text{crit}}(x, \mathcal{X}_n)$, $n \to \infty$.

Example: Dimension estimators

· $\mathcal{X}$: a point cloud on a manifold $\mathcal{M} \subset \mathbb{R}^d$; $m := \dim(\mathcal{M})$ **unknown**, called the **intrinsic dimension** ($m \leq d$).

· Interpoint distances are **known**.

· $D_j := D_j(x, \mathcal{X}) :=$ dist. between x and its jth nearest neighbor in $\mathcal{X}$.

· **Problem.** Estimate intrinsic dimension m using only the information about interpoint distances.

Example: Dimension estimators

· **Problem.** Estimate intrinsic dimension m using only the information about interpoint distances. Fix $k \geq 3$. Define 'score' at x wrt $\mathcal{X}$, by

$$\xi_k(x, \mathcal{X}) := (k-2) \left(\sum_{j=1}^{k-1} \log \frac{D_k(x, \mathcal{X})}{D_j(x, \mathcal{X})} \right)^{-1} .$$

· **Goal** (Bickel and Levina). Let point cloud be $\mathcal{X}_n := \{X_i\}_{i=1}^n$, i.i.d. sample on manifold $\mathcal{M} \subset \mathbb{R}^d$. Fix $k \geq 3$. When does

$$\lim_{n \to \infty} \mathbb{E}\left[\xi_k(X_1, \mathcal{X}_n)\right] = \dim \mathcal{M}?$$

· Find the distribution (for large n) of the sums $\sum_{i \leq n} \xi_k(X_i, \mathcal{X}_n)$ (after centering and scaling).

Other Examples

· Statistics in stochastic geometry expressible as a sum of scores of local data:

· Surface area estimators: use a point cloud to estimate the surface area of a target object when one only knows whether a given pt in the cloud is inside or outside the target.

· Statistics of random convex hulls: how closely does the convex hull of a point cloud inside a convex set approximate the convex set?

· Statistics of interacting particle systems.

· Covariograms, Ripley's K-function.

Outline

- **I Examples**
- **II Stabilization**
- **III Laws of large numbers**
- **IV Gaussian fluctuations**
- **V Variance asymptotics**
- **VI Corollaries**
- **VII Statistics of general input**
- **VIII Statistics of Poisson functionals**

I Examples: Summary

· All examples involve sums of the form

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

with ξ a score function, $\mathcal{X}$ a finite point cloud.

· When $\mathcal{X} \subset \mathbb{R}^d$ is a random pt configuration, the sums $\sum_{x \in \mathcal{X}} \xi(x, \mathcal{X})$ typically describe a **global** feature of a certain spatial random structure in terms of **local** interactions.

· What is the distribution of these sums for large pt clouds $\mathcal{X}$?

· Laws of large numbers?

· Central limit theorems?

II Stabilization - Terminology

· **Input space** $\mathbf{N}$: space of inputs, these are simple point processes on $\mathbb{R}^d$, random point clouds.

· **Element of input space**: Ξ, random point cloud in $\mathbb{R}^d$, a simple pt process $\Xi := \sum_{x \in \Xi} \delta_x$.

Ex. Homogenous Poisson point process with intensity 1: the number of points in disjoint sets are indep. r.v and the number of points in B is Poisson with mean $\mathrm{Vol}(B)$.

· **Windows**: $W_\lambda := [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

· **Score functions**: $\xi : \mathbb{R}^d \times \mathbf{N} \to \mathbb{R}$

II Stabilization - Terminology

· **Goal.** Given a score function ξ and a pt process $\Xi \in \mathbf{N}$, we seek the limit theory for the total score

$$H_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda), \quad \lambda \to \infty$$

and total measure

$$\mu_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda)\delta_{\lambda^{-1/d}x}, \quad \to \infty.$$

· 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 $\xi(x, \Xi \cap W_\lambda)$. What does 'local' mean?

II Stabilization

· Write $\xi(x, \mathcal{X})$ instead of $\xi(x, \mathcal{X} \cup \{x\})$.

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

· $W_\lambda := [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

· **Key Definition**. ξ is *stabilizing* w.r.t. pt process Ξ on $\mathbb{R}^d$ if for all $x \in \Xi$ there is $R := R^\xi(x, \Xi) < \infty$ a.s. (a 'radius of stabilization') such that

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

for any locally finite $\mathcal{A} \subset \mathbb{R}^d$. ξ is *exponentially stabilizing* w.r.t. Ξ if

$$\sup_{\lambda \geq 1} \sup_{x \in W_\lambda} \mathbb{P}(R^\xi(x, \Xi \cap W_\lambda) \geq r) \leq c \exp(-\frac{r}{c}), \quad r \in [1, \infty).$$

II Stabilization: Examples of stabilizing scores

· Given a point set $\mathcal{X} \subset \mathbb{R}^d$, $x \in \mathcal{X}$, $k \in \mathbb{N}$, let $V_k(x, \mathcal{X})$ be the set of vertices connected to x in (undirected) k nearest neighbors graph on $\mathcal{X}$.

· Fix $\phi : \mathbb{R}^+ \to \mathbb{R}^+$. Put

$$\xi(x, \mathcal{X}) = \sum_{y \in V_k(x, \mathcal{X})} \phi(|x - y|).$$

· **Fact**. ξ is stabilizing w.r.t. a rate one homogenous Poisson point process $\mathcal{P}$ on $\mathbb{R}^d$.

· *Proof.* Given $x \in \mathcal{P}$, need to find $R := R^\xi(x, \mathcal{P}) < \infty$ a.s. 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 $\mathcal{A} \subset \mathbb{R}^d$.

II Stabilization: Examples of stabilizing scores

· kNN graph structure at x w.r.t. $\mathcal{P} \cap B_R(x)$ should not change if points are inserted outside $B_R(x)$.

· $d = 2$. For each $t > 0$, construct six disjoint equilateral triangles $T_j(t), j = 1, ..., 6$, having apex at x, edge length t.

· Then

$$R = \min\{t > 0 : \text{card}(T_j(t) \cap \mathcal{P}) \geq k; j = 1, ..., 6\}$$

is a stabilization radius.

II Stabilization: Examples of exponentially stabilizing scores

· Put $R = \min\{t > 0 : \mathrm{card}(T_j(t) \cap \mathcal{P}) \geq k; j = 1, ..., 6\}$

· ξ is exponentially stabilizing w.r.t. a rate one homogenous Poisson point process $\mathcal{P}$ on $\mathbb{R}^d$:

$$\mathbb{P}(R^{\xi}(x, \mathcal{P}) \geq r) \leq \sum_{j=1}^{6} \mathbb{P}(\mathrm{card}(\mathcal{P} \cap T_j(r)) < k).$$

· The number of points from $\mathcal{P}$ in $T_j(r)$ is Poisson r.v. with mean

$$\mathrm{Vol}_2(T_j(r)) = \frac{\sqrt{3}}{4} r^2.$$

· By Chernoff bound for Poisson distribution

$$\mathbb{P}(\mathrm{Poiss}(\frac{\sqrt{3}}{4} r^2) < k) \leq ck \exp(-cr^2), \quad r > 0.$$

· **General rule:** scores defined in terms of 'proximity graphs' are exponentially stabilizing with respect to point processes having void probabilities decaying exponentially fast with the void radius.

· Ξ: a pt process on $\mathbb{R}^d$; $\Xi \cap W_\lambda := \Xi \cap [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

· **Definition (Moment condition).** ξ satisfies the p moment condition w.r.t. Ξ if

$$\sup_{\lambda \geq 1} \sup_{x,y \in W_\lambda} \mathbb{E}\, |\xi(x, (\Xi \cap W_\lambda) \cup \{y\})|^p \leq M_p^\xi := M_p < \infty.$$

III Laws of large numbers

· Ξ: a pt process on $\mathbb{R}^d$; $\Xi \cap W_\lambda := \Xi \cap [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

· **Definition (Moment condition).** ξ satisfies the p moment condition w.r.t. Ξ if

$$\sup_{\lambda \geq 1} \sup_{x,y \in W_\lambda} \mathbb{E}\,|\xi(x, (\Xi \cap W_\lambda) \cup \{y\})|^p \leq M_p^\xi := M_p < \infty.$$

· Palm expectations

· **Definition (Moment condition).** ξ satisfies the p moment condition w.r.t. Ξ if

$$\sup_{\lambda \geq 1} \sup_{x,y \in W_\lambda} \mathbb{E}_x|\xi(x, (\Xi \cap W_\lambda) \cup \{y\})|^p \leq M_p^\xi := M_p < \infty.$$

III Laws of large numbers

$\cdot$ Ξ: a pt process on $\mathbb{R}^d$. Say that Ξ has intensity $\rho(\cdot)$ if for all Borel $B \subset \mathbb{R}^d$ we have

$$\mathbb{E}\,|\Xi \cap B| = \int_B \rho(x)dx.$$

III Laws of large numbers

· **Input**: a stationary point process Ξ on $\mathbb{R}^d$ with constant intensity ρ.
· $\Xi \cap W_\lambda := \Xi \cap [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

$$\mu_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda)\delta_{\lambda^{-1/d}x}.$$

· **Thm (WLLN)**: If ξ is trans. invariant, stabilizing wrt Ξ, and satisfies the p moment condition for some $p \in (1, \infty)$, then for all $f \in \mathbb{B}([-\frac{1}{2}, \frac{1}{2}]^d)$

$$|\lambda^{-1}\mathbb{E}\,\langle \mu_\lambda^\xi, f\rangle - \rho \cdot \mathbb{E}_{\mathbf{0}}\xi(\mathbf{0}, \Xi \cup \{\mathbf{0}\}) \cdot \int_{[-\frac{1}{2},\frac{1}{2}]^d} f(x)dx| \leq \epsilon_\lambda.$$

· Penrose and Y (2003): $\epsilon_\lambda = o(1)$.

· Lachièze-Rey, Schulte, + Y (2017): $\epsilon_\lambda = O(\lambda^{-1/d})$ if ξ is exponentially stabilizing wrt Ξ.

· **Campbell Formula for Poisson point processes** $\mathcal{P}$. Let $\mathcal{P}$ have intensity $\rho : \mathbb{R}^d \to \mathbb{R}^+$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable. The random sum

$$\sum_{x \in \mathcal{P}} f(x)$$

has expected value

$$\mathbb{E} \sum_{x \in \mathcal{P}} f(x) = \int_{\mathbb{R}^d} f(x)\rho(x)dx.$$

· **Mecke Formula**. Let $f : \mathbb{R}^d \times \mathbf{N} \to \mathbb{R}$ be measurable. The random sum

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

has expected value

$$\mathbb{E} \sum_{x \in \mathcal{P}} f(x, \mathcal{P}) = \int_{\mathbb{R}^d} \mathbb{E}\, f(x, \mathcal{P} \cup \{x\})\rho(x)dx.$$

III Laws of large numbers

· **Campbell-Mecke Formula for General Ξ.** Let Ξ be a point process on $\mathbb{R}^d$ with intensity $\rho : \mathbb{R}^d \to \mathbb{R}^+$. Then Ξ admits a family of Palm distributions $(\mathbb{P}_x)_{x \in \mathbb{R}^d}$ on the space $\mathbf{N}$ of simple measures on $\mathbb{R}^d$ such that for integrable $f : \mathbb{R}^d \times \mathbf{N} \to \mathbb{R}$

$$\mathbb{E} \sum_{x \in \Xi} f(x, \Xi) = \int_{\mathbb{R}^d} \mathbb{E}_x f(x, \Xi \cup \{x\}) \rho(x) dx$$

where

$$\mathbb{E}_x f(x, \Xi \cup \{x\}) = \int_{\mathbf{N}} f(x, \mu) \mathbb{P}_x(du), \; x \in \mathbb{R}^d.$$

III LLN: Proof for $f \equiv 1$, Ξ is PPP $\mathcal{P} := \mathcal{P}_\rho$ w. intensity ρ

$$
\begin{aligned}
&|\frac{1}{\lambda}\mathbb{E} \sum_{x \in \mathcal{P} \cap W_\lambda} \xi(x, \mathcal{P} \cap W_\lambda) - \rho \mathbb{E}\, \xi(\mathbf{0}, \mathcal{P})| \\
&= |\frac{1}{\lambda} \int_{W_\lambda} \mathbb{E}\, \xi(x, \mathcal{P} \cap W_\lambda) \rho dx - \rho \mathbb{E}\, \xi(x, \mathcal{P})| dx \\
&\leq \frac{\rho}{\lambda} \int_{W_\lambda} \mathbb{E}\, |\xi(x, \mathcal{P} \cap W_\lambda) - \xi(x, \mathcal{P})| dx
\end{aligned}
$$

III LLN: Proof for $f \equiv 1$, Ξ is PPP $\mathcal{P} := \mathcal{P}_\rho$ w. intensity ρ

$$|\frac{1}{\lambda}\mathbb{E}\sum_{x \in \mathcal{P}\cap W_\lambda} \xi(x, \mathcal{P}\cap W_\lambda) - \rho \mathbb{E}\,\xi(\mathbf{0}, \mathcal{P})|$$

$$= |\frac{1}{\lambda}\int_{W_\lambda} \mathbb{E}\,\xi(x, \mathcal{P}\cap W_\lambda)\rho dx - \rho\mathbb{E}\,\xi(x, \mathcal{P})|dx$$

$$\leq \frac{\rho}{\lambda}\int_{W_\lambda} \mathbb{E}\,|\xi(x, \mathcal{P}\cap W_\lambda) - \xi(x, \mathcal{P})|dx$$

$$= \frac{\rho}{\lambda}\int_{W_\lambda} \mathbb{E}\,|\xi(x, \mathcal{P}\cap W_\lambda) - \xi(x, \mathcal{P})|\mathbf{1}(R^\xi(x, \mathcal{P}\cap W_\lambda) \geq d(x, \partial W_\lambda))dx$$

$$\leq \frac{\rho}{\lambda}\int_{W_\lambda} (\mathbb{E}\,|\xi(x, \mathcal{P}\cap W_\lambda) - \xi(x, \mathcal{P})|^q)^{\frac{1}{q}}$$

$$\cdot\, \mathbb{P}(R^\xi(x, \mathcal{P}\cap W_\lambda) \geq d(x, \partial W_\lambda))^{\frac{q-1}{q}} dx$$

$$\leq 2M_q^{\frac{1}{q}}\frac{\rho}{\lambda}\int_{W_\lambda} \exp(-cd(x, \partial W_\lambda)^{\frac{q-1}{q}})dx = O(\lambda^{-\frac{1}{d}}). \quad \square$$

III LLN: Proof sketch for general f and Ξ w. intensity ρ

· Write also

$$\int_{W_1} f(x)dx = \frac{1}{\lambda}\int_{W_\lambda} f(\lambda^{-1/d}x)dx.$$

·Thus

$$|\frac{1}{\lambda}\mathbb{E}\sum_{x\in\Xi\cap W_\lambda}\xi(x,\Xi\cap W_\lambda)f(\lambda^{-1/d}x) - \rho\mathbb{E}_{\mathbf{0}}\xi(\mathbf{0},\Xi)\int_{W_1} f(x)dx|$$

$$|\frac{\rho}{\lambda}\int_{W_\lambda} f(\lambda^{-1/d}x)[\mathbb{E}_x\xi(x,\Xi\cap W_\lambda) - \mathbb{E}_x\xi(x,\Xi)]dx|$$

$$\leq 2M_q^{\frac{1}{q}}\frac{\rho||f||_\infty}{\lambda}\int_{W_\lambda}\exp(-cd(x,\partial W_\lambda)^{\frac{q-1}{q}})dx$$

$$= O(\lambda^{-\frac{1}{d}}\rho||f||_\infty). \quad \square$$

III Laws of large numbers on *non-uniform* input

· For all $\lambda > 0$ define the λ *re-scaled version* of ξ by

$$\xi_\lambda(x, \mathcal{X}) := \xi(\lambda^{1/d} x, \lambda^{1/d} \mathcal{X}).$$

· Re-scaling is natural when considering Poisson input $\mathcal{P}_\lambda$ having average cardinality λ in a compact set K; dilation by $\lambda^{1/d}$ means that unit volume subsets of $\lambda^{1/d} K$ host on the average one point of $\lambda^{1/d} \mathcal{P}_\lambda$.

· Equivalent models ($\mathcal{P}_\lambda$ a Poisson pt process w. intensity λ):

$$\sum_{x \in \mathcal{P}_\lambda \cap W_1} \xi_\lambda(x, \mathcal{P}_\lambda \cap W_1) \stackrel{\mathcal{D}}{=} \sum_{x \in \mathcal{P}_1 \cap W_\lambda} \xi(x, \mathcal{P}_1 \cap W_\lambda).$$

III Laws of large numbers on *non-uniform* input

· For all $\lambda > 0$ define the λ *re-scaled version* of ξ by

$$\xi_\lambda(x, \mathcal{X}) := \xi(\lambda^{1/d} x, \lambda^{1/d} \mathcal{X}).$$

· We will replace $\mathcal{X}$ by $\mathcal{P}_{\lambda\kappa}$, a PPP on $\mathbb{R}^d$ with intensity density $\lambda\kappa(x)dx$, namely the number of points of $\mathcal{P}_{\lambda\kappa}$ in B is a Poisson r. v. with mean

$$\int_B \lambda\kappa(x)dx.$$

III Laws of large numbers on *non-uniform* input

· Note: $\mathcal{P}_{\lambda\kappa}$ versus $\mathcal{P}_{\kappa(x)}$. The latter is a homogenous Poisson point process of constant intensity.

· What is asymptotic behavior of $\lambda^{1/d}(\mathcal{P}_{\lambda\kappa} - x_0)$ as a pt process?

III LLN: Convergence of re-scaled pt processes

· Recall: almost every $x \in \mathbb{R}^d$ is a *Lebesgue point* of κ, i.e.,

$$\lim_{\epsilon \to 0} \epsilon^{-d} \int_{B_\epsilon(x)} |\kappa(y) - \kappa(x)| \, dy = 0.$$

· The average number of pts of $\lambda^{1/d}\mathcal{P}_{\lambda\kappa}$ in $B_r(0)$ equals the average number of pts of $\mathcal{P}_{\lambda\kappa}$ in $B_{r\lambda^{-1/d}}(0)$, which equals

$$\int_{B_{r\lambda^{-1/d}}(0)} \kappa(y)\lambda \, dy \ \sim \int_{B_{r\lambda^{-1/d}}(0)} \kappa(0)\lambda \, dy = \kappa(0) \int_{B_r(0)} dy$$

which equals the average number of points of $\mathcal{P}_{\kappa(0)}$ in $B_r(0)$. So

$$\lambda^{1/d}(\mathcal{P}_{\lambda\kappa} - x_0) \xrightarrow{\mathcal{D}} \mathcal{P}_{\kappa(x_0)} \text{ as } \lambda \to \infty,$$

where convergence is in the sense of weak convergence of point processes.

III Laws of large numbers on *non-uniform* input

· $\mathcal{P}_{\lambda\kappa}$: PPP on $\mathbb{R}^d$ with intensity density $\lambda\kappa(x)dx$.

· $x_0 \in \mathbb{R}^d$ a Lebesgue pt for κ:

$$\lambda^{1/d}(\mathcal{P}_{\lambda\kappa} - x_0) \xrightarrow{\mathcal{D}} \mathcal{P}_{\kappa(x_0)} \text{ as } \lambda \to \infty,$$

where convergence is in the sense of weak convergence of point processes.

· If $\xi(\cdot,\cdot)$ is a score defined on $\mathbb{R}^d \times \mathbf{N}$, where $\mathbf{N}$ is the space of locally finite point sets in $\mathbb{R}^d$, one might hope that ξ is *continuous* on the pairs $(\mathbf{0}, \lambda^{1/d}(\mathcal{P}_{\lambda\kappa} - x_0))$ in the sense that

$$\xi(\mathbf{0}, \lambda^{1/d}(\mathcal{P}_{\lambda\kappa} - x_0)) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}_{\kappa(x_0)}), \quad \lambda \to \infty.$$

· This turns out to be the case whenever ξ is stabilizing w.r.t. to $\mathcal{P}_{\kappa(x_0)}$.

· Stabilization is a surrogate for continuity.

III Laws of large numbers on *non-uniform* input

· Put $\mathcal{X}_n := \{X_i\}_{i=1}^n$, where X_i are i.i.d. with density κ on $\mathbb{R}^d$.

· **Lemma (convergence of re-scaled binomial pt process)** Let $x \in \mathbb{R}^d$ be a Lebesgue point for κ. Then

$$n^{1/d}(\mathcal{X}_n - x) \xrightarrow{\mathcal{D}} \mathcal{P}_{\kappa(x)}, \quad n \to \infty.$$

· **'Continuity' lemma for stabilizing scores**. Let $x \in \mathbb{R}^d$ be a Lebesgue point for κ and assume that $R^\xi(x, \mathcal{P}_{\kappa(x)}) < \infty$ a.s. (where $R^\xi(x, \mathcal{P}_{\kappa(x)})$ is the radius of stabilization for ξ at x wrt $\mathcal{P}_{\kappa(x)}$). Then

$$\xi_\lambda(x, \mathcal{P}_{\lambda\kappa}) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)}), \quad \lambda \to \infty,$$

$$\xi_n(x, \mathcal{X}_n) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)}), \quad n \to \infty.$$

· Stabilization is a surrogate for continuity.

III Laws of large numbers

· **Campbell Formula**. Let Ξ be a point process on $\mathbb{R}^d$ with intensity $\rho(x)$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable. The random sum

$$\sum_{x \in \Xi} f(x)$$

has expected value

$$\mathbb{E} \sum_{x \in \Xi} f(x) = \int_{\mathbb{R}^d} f(x) \rho(x) dx.$$

· **Mecke Formula**. Let $f : \mathbb{R}^d \times \mathbf{N} \to \mathbb{R}$ be measurable. The random sum

$$\sum_{x \in \Xi} f(x, \Xi)$$

has expected value

$$\mathbb{E} \sum_{x \in \Xi} f(x, \Xi) = \int_{\mathbb{R}^d} \mathbb{E}_x f(x, \Xi \cup \{x\}) \rho(x) dx.$$

III Laws of large numbers on *non-uniform* input

· For a measure μ on $\mathbb{R}^d$ let $\langle \mu, f \rangle := \int f d\mu$.

·$\mathcal{P}_{\lambda\kappa}$ is PPP with intensity $\lambda\kappa$. Put

$$\mu_\lambda := \sum_{x \in \mathcal{P}_{\lambda\kappa}} \xi_\lambda(x, \mathcal{P}_{\lambda\kappa}) \delta_x.$$

·For all $f \in \mathbb{B}(\mathbb{R}^d)$, the Mecke formula gives

$$\mathbb{E}\left[\langle \mu_\lambda, f \rangle\right] = \lambda \int_{\mathbb{R}^d} f(x) \mathbb{E}\left[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa})\right] \kappa(x)\, dx.$$

· Examine $\mathbb{E}\left[...\right]$. If there is some $p > 1$ such that

$$\sup_\lambda \sup_{x \in \mathbb{R}^d} \mathbb{E}\left|\xi_\lambda(x, \mathcal{P}_{\lambda\kappa})\right|^p < \infty,$$

then uniform integrability, the 'continuity' Lemma, and trans. inv. of ξ show that for all Lebesgue points x of κ

$$\mathbb{E}\, \xi_\lambda(x, \mathcal{P}_{\lambda\kappa}) \to \mathbb{E}\, \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)}) \quad \text{as } \lambda \to \infty.$$

III Laws of large numbers on *non-uniform* input

· The set of points failing to be Lebesgue points of κ has measure zero.

· The bounded convergence theorem gives all $f \in \mathbb{B}(\mathbb{R}^d)$ the WLLN

$$\lim_{\lambda \to \infty} \lambda^{-1} \mathbb{E}\left[\langle \mu_\lambda, f \rangle\right] = \int_{\mathbb{R}^d} f(x) \mathbb{E}\left[\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)})\right] \kappa(x) \, dx.$$

Extensions.

· Convergence of means $\frac{1}{\lambda}\mathbb{E}\left[\langle \mu_\lambda, f\rangle\right]$ may be upgraded to convergence in L^q, $q \in \{1,2\}$, and a.s. convergence.

· Methods yield LLN for measures arising from binomial input:

$$\mu_n := \sum_{i=1}^{n} \xi_n(X_i, \mathcal{X}_n)\delta_{X_i},$$

where $\mathcal{X}_n := \{X_i\}_{i=1}^n$, and $X_i, i \geq 1$, i.i.d. with density κ.

III Laws of large numbers on non-uniform binomial input

Theorem Let $q = 1$ or $q = 2$. Assume:

(i) ξ is stabilizing w.r.t. Poisson input $\mathcal{P}_\tau$ of intensity τ, $\forall \tau > 0$.

(ii) $\sup_n \mathbb{E}\,|\xi_n(X_1, \mathcal{X}_n)|^p < \infty$ for some $p \in (q, \infty)$.

Then for all $f \in \mathbb{B}(\mathbb{R}^d)$

$$\lim_{n\to\infty} n^{-1}\langle \mu_n, f\rangle = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \xi_n(X_i, \mathcal{X}_n) f(X_i)$$

$$= \int f(x)\mathbb{E}\,[\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)})]\kappa(x)dx \text{ in } L^q.$$

If $\sup_\lambda \mathbb{E}\,|\xi_\lambda(\mathbf{0}, \mathcal{P}_{\lambda\kappa})|^p < \infty$ for some $p \in (q, \infty)$, then for all $f \in \mathbb{B}(\mathbb{R}^d)$

$$\lim_{\lambda\to\infty} \lambda^{-1}\langle \mu_\lambda, f\rangle = \int f(x)\mathbb{E}\,[\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)})]\kappa(x)dx \text{ in } L^q.$$

III Laws of large numbers

Corollaries: We can deduce a weak law of large numbers for the following statistics:

· clique counts in the random geometric graph on $\mathcal{P}_{\lambda\kappa}$
· component count in r-offset of $\mathcal{P}_{\lambda\kappa}$
· the number of critical points for $\mathcal{P}_{\lambda\kappa}$
· dimension estimators for input $\mathcal{P}_{\lambda\kappa}$ on manifolds
· entropy estimators for κ.

$\mathcal{P} := \mathcal{P}_1$ a PPP on $\mathbb{R}^d$ with intensity 1. Put
$\mu_\lambda^\xi := \sum_{x \in \mathcal{P} \cap W_\lambda} \xi(x, \mathcal{P} \cap W_\lambda) \delta_{\lambda^{-1/d} x}$.

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

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

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

Lachièze-Rey, Schulte + Y (2017): $\epsilon_\lambda = O(\lambda^{-1/2})$ (Stein's method)

IV Gaussian fluctuations

· Simplify: study sum of scores (not measures). Put

$$H_\lambda^\xi := \sum_{x \in \mathcal{P} \cap W_\lambda} \xi(x, \mathcal{P} \cap W_\lambda).$$

· **Thm (CLT)**: Assume ξ is exponentially stabilizing wrt $\mathcal{P}$ and satisfies the p moment condition for some $p \in (5, \infty)$. If $\mathrm{Var} H_\lambda^\xi = \Omega(\lambda)$, then

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left[\frac{H_\lambda^\xi - \mathbb{E}\, H_\lambda^\xi}{\sqrt{\mathrm{Var} H_\lambda^\xi}} \leq t \right] - \mathbb{P}[N(0,1) \leq t] \right| \leq \epsilon_\lambda = o(1).$$

IV Gaussian fluctuations: Proofs

· Three approaches:

(i) Blocking methods: intuitive, need only $2+\epsilon$ moments on scores; suboptimal rates.

(ii) Malliavin-Stein methods: optimal rates for Poisson and binomial input; $5+\epsilon$ moments.

(iii) Cumulants: works for input having factorizing correlations; requires all moments, suboptimal rates.

IV Gaussian fluctuations: Blocking Method

· Recalling that ξ_λ dilates by factor of $\lambda^{\frac{1}{d}}$, we write equivalently

$$H_\lambda^\xi := \sum_{x \in \mathcal{P} \cap W_\lambda} \xi(x, \mathcal{P} \cap W_\lambda) = \sum_{x \in \mathcal{P}_\lambda \cap W_1} \xi_\lambda(x, \mathcal{P}_\lambda \cap W_1).$$

· Subdivide $W_1 := [-\frac{1}{2}, \frac{1}{2}]^d$ into sub-cubes $\{Q_i\}_{i \geq 1}$ of edge length $c \frac{\log \lambda}{\lambda^{\frac{1}{d}}}$.

· There is an event E_λ, $\mathbb{P}(E_\lambda^c) = 1 - o(1)$, such that on E_λ the sums

$$\sum_{x \in \mathcal{P}_\lambda \cap Q_i} \xi_\lambda(x, \mathcal{P}_\lambda) \text{ and } \sum_{x \in \mathcal{P}_\lambda \cap Q_j} \xi_\lambda(x, \mathcal{P}_\lambda)$$

are independent if Q_i and Q_j are separated by at least one subcube.

· Conditional on event E_λ we obtain Gaussian fluctuations by M-dependent methods.

· Exponential stabilization is surrogate for independence.

IV Gaussian fluctuations: Malliavin-Stein Method

· $F = f(\mathcal{P})$, a measurable function of Poisson pt process $\mathcal{P}$ on $\mathbb{R}^d$. Called a Poisson functional.

· Difference operators (discrete gradient). For x, y in $\mathbb{R}^d$

$$D_x F = f(\mathcal{P} \cup \{x\}) - f(\mathcal{P}).$$

$$D_{x,y} F = f(\mathcal{P} \cup \{x, y\}) - f(\mathcal{P} \cup \{x\}) - f(\mathcal{P} \cup \{y\}) + f(\mathcal{P}).$$

IV Gaussian fluctuations: Malliavin-Stein Method

· $F = f(\mathcal{P})$, a measurable function of Poisson pt process $\mathcal{P}$ on $\mathbb{R}^d$.
· Difference operators (discrete gradient). For x in $\mathbb{R}^d$

$$D_x F = f(\mathcal{P} \cup \{x\}) - f(\mathcal{P}).$$

· **Poincaré Inequality**: $\mathrm{Var} F \leq \mathbb{E} \int_{\mathbb{R}^d} (D_x F)^2 dx.$

· This is Poisson space counterpart to Chernoff-Nash-Poincaré Inequality:

$$\mathrm{Var} f(X) \leq \mathbb{E} \, ||\nabla f(X)||^2.$$

· ∇ is gradient, f is smooth, $X = (X_1, ..., X_d)$ i.i.d. standard Gaussian vector.

IV Gaussian fluctuations: Malliavin-Stein Method

$$H(\mathcal{P}_\lambda) := \sum_{x \in \mathcal{P}_\lambda \cap W_1} \xi_\lambda(x, \mathcal{P}_\lambda \cap W_1).$$

· Difference operators (discrete gradient). For x, y in $W_1 := [-\frac{1}{2}, \frac{1}{2}]^d$

$$D_x H(\mathcal{P}_\lambda) = H(\mathcal{P}_\lambda \cup \{x\}) - H(\mathcal{P}_\lambda)$$

$$D_{x,y} H(\mathcal{P}_\lambda) = H(\mathcal{P}_\lambda \cup \{x, y\}) - H(\mathcal{P}_\lambda \cup \{x\}) - H(\mathcal{P}_\lambda \cup \{y\}) + H(\mathcal{P}_\lambda)$$

· **Thm** (Last, Peccati, Schulte) Assume $\mathbb{E}\, H(\mathcal{P}_\lambda)^2 < \infty$ and that there are positive constants c, p s.t for all $x, y \in W_1$ we have $\mathbb{E}\, |D_x H(\mathcal{P}_\lambda \cup \{y\})|^{4+p} \leq c$. Then

$$d_K \left(\frac{H(\mathcal{P}_\lambda) - \mathbb{E}\, H(\mathcal{P}_\lambda)}{\sqrt{\mathrm{Var} H(\mathcal{P}_\lambda)}}, N(0,1) \right) \leq C(S_1 + S_2 + S_3).$$

IV Gaussian fluctuations: Malliavin-Stein Method

$$S_1 = \frac{\lambda}{\mathrm{Var} H(\mathcal{P}_\lambda)} \sqrt{\int_{W_1^2} \mathbb{P}\left(D_{x,y} H(\mathcal{P}_\lambda) \neq 0\right)^{\frac{p}{8+2p}} dx dy}$$

$$S_2 = \frac{\lambda^{\frac{3}{2}}}{\mathrm{Var} H(\mathcal{P}_\lambda)} \sqrt{\int_{W_1} \left(\int_{W_1} \mathbb{P}\left(D_{x,y} H(\mathcal{P}_\lambda) \neq 0\right)^{\frac{p}{16+4p}} dy \right)^2 dx}$$

$$S_3 = \frac{\sqrt{\lambda}}{\mathrm{Var} H(\mathcal{P}_\lambda)} + \frac{\lambda}{(\mathrm{Var} H(\mathcal{P}_\lambda))^{\frac{3}{2}}} + \frac{\lambda^{\frac{5}{4}} + \lambda^{\frac{3}{2}}}{(\mathrm{Var} H(\mathcal{P}_\lambda))^2}$$

Claim: ξ exponentially stabilizes w.r.t. $\mathcal{P}_\lambda$ and $\mathrm{Var} H(\mathcal{P}_\lambda) = \Theta(\lambda)$

$$\Longrightarrow (S_1 + S_2 + S_3) = O\left(\frac{1}{\sqrt{\lambda}}\right).$$

· Suffices to show $\mathbb{P}\left(D_{x,y} H(\mathcal{P}_\lambda) \neq 0\right) \leq C \cdot \mathbf{1}(|x-y| \leq c\lambda^{\frac{-1}{d}})$.

· We show this in the special case that the stabilization radius is deterministic, i.e. is bounded by $c\lambda^{\frac{-1}{d}}$.

$$H(\mathcal{P}_\lambda) := \sum_{x \in \mathcal{P}_\lambda \cap W_1} \xi_\lambda(x, \mathcal{P}_\lambda \cap W_1).$$

· We show $\mathbb{P}\left(D_{x,y} H(\mathcal{P}_\lambda) \neq 0\right) \leq C \cdot \mathbf{1}(|x-y| \leq c\lambda^{\frac{-1}{d}})$:

$$\begin{aligned}
&\mathbb{P}\left(D_{x,y} H(\mathcal{P}_\lambda) \neq 0\right) \\
&\leq \mathbb{P}(D_x \xi_\lambda(y, \mathcal{P}_\lambda) \neq 0) + \mathbb{P}(D_y \xi_\lambda(x, \mathcal{P}_\lambda) \neq 0) \\
&\qquad + \mathbb{E} \sum_{z \in \mathcal{P}_\lambda} \mathbf{1}(D_{x,y} \xi_\lambda(z, \mathcal{P}_\lambda) \neq 0) \\
&\leq \mathbf{1}(|x-y| \leq c\lambda^{\frac{-1}{d}}) + \mathbf{1}(|x-y| \leq c\lambda^{\frac{-1}{d}}) \\
&\qquad + \lambda \int_{W_1} \mathbf{1}(|x-z| \leq c\lambda^{\frac{-1}{d}}) \mathbf{1}(|y-z| \leq c\lambda^{\frac{-1}{d}}) dz \\
&\leq C \cdot \mathbf{1}(|x-y| \leq 2c\lambda^{\frac{-1}{d}}). \qquad \square
\end{aligned}$$

IV Gaussian fluctuations: Cumulant Method

· **Method of moments.** If all the moments of a sequence of random variables $(X_n)_{n \geq 1}$ converge to those of a normal N, then

$$X_n \xrightarrow{\mathcal{D}} N.$$

· **Cumulants.** For a random variable Y with all finite moments, expanding the logarithm of the Laplace transform (mgf) in a formal power series gives cumulant generating function

$$\kappa_Y(t) := \log \mathbb{E}\,(e^{tY}) = \log(1 + \sum_{k=1}^{\infty} \frac{M_k t^k}{k!}) = \sum_{k=1}^{\infty} \frac{c^{(k)}(Y) t^k}{k!},$$

where $M_k = \mathbb{E}\,(Y^k)$ is the k th moment of Y and $c^{(k)} := c^{(k)}(Y)$ denotes the k th cumulant of Y.

· $c^{(k)} = i^{-k} \, \frac{d^k \log \mathbb{E}\,(e^{itY})}{dt^k} \, |_{t=0}$.

Cumulants.

$$\kappa_Y(t) := \log \mathbb{E}\,(e^{tY}) = \log(1 + \sum_{k=1}^{\infty} \frac{M_k t^k}{k!}) = \sum_{k=1}^{\infty} \frac{c^{(k)}(Y) t^k}{k!}.$$

· First cumulant is the mean. Second and third cumulants are the second and third central moments of Y. But this is not true for higher order cumulants.

· Additivity of log $\Rightarrow$ kth order cumulant of a sum of independent r.v. is the sum of the cumulants.

· The cumulant is 'cumulative', hence the terminology.

· Cumulant generating function for $N(\mu, \sigma^2)$ is $\kappa(t) = \mu t + \frac{\sigma^2}{2} t^2$.

· Cumulants of order three or more vanish.

IV Gaussian fluctuations: Proof idea for cumulant method

· **Theorem.** (Marcinkiewicz) Let $(X_n)_{n \geq 1}$ be mean zero random variables, $\mathrm{Var} X_n = 1$. If $\lim_{n \to \infty} c^{(k)}(X_n) = 0$ for all k large, then

$$X_n \xrightarrow{\mathcal{D}} N(0,1) \text{ as } n \to \infty.$$

· **Corollary**. (linear growth of cumulants implies CLT) If $(Y_n)_{n \geq 1}$ are mean zero random variables with $c^{(k)}(Y_n) = O(n)$ for all k large, if $\mathrm{Var} Y_n = \Omega(n^\alpha)$ for some $\alpha \in (0, \infty)$, then

$$\frac{Y_n}{\sqrt{\mathrm{Var} Y_n}} \xrightarrow{\mathcal{D}} N(0,1) \text{ as } n \to \infty.$$

· **Proof.** $c^{(k)}\left(\frac{Y_n}{\sqrt{\mathrm{Var} Y_n}}\right) = \frac{c^{(k)}(Y_n)}{(\mathrm{Var} Y_n)^{k/2}} = O\left(\frac{n}{(n^\alpha)^{k/2}}\right) \to 0$ for k large. $\square$

IV Gaussian fluctuations: Cumulant Method

Proof idea for CLT $H_\lambda^\xi := \sum_{x \in \mathcal{P}_\lambda} \xi_\lambda(x, \mathcal{P} \cap W_\lambda)$.

· To show $\frac{H_\lambda^\xi - \mathbb{E}\, H_\lambda^\xi}{\sqrt{\mathrm{Var} H_\lambda^\xi}} \xrightarrow{\mathcal{D}} N(0,1)$, it suffices to show that kth order cumulant for H_λ^ξ has linear growth $O(\lambda)$.

· Combinatorial formula for cumulants in terms of moments:

$$c^{(k)}(Y) = \sum_{\pi \in \Pi[k]} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{i=1}^{|\pi|} M_{|\pi(i)|}\,.$$

· Here $\Pi[k]$ is the set of all unordered partitions of $\{1, ..., k\}$, and for a partition $\pi = \{\pi(1), \ldots, \pi(l)\} \in \Pi[k]$, $|\pi|$ denotes the number of its elements (in this case $|\pi| = l$), while $|\pi(i)|$ the number of elements of subset $\pi(i)$.

· **Fact:** kth cumulant for H_λ^ξ can be expressed as a linear combination of terms of the form

$$|\mathbb{E}\,\Pi_{i=1}^{p+q}\xi(x_i,.) - \mathbb{E}\,\Pi_{i=1}^{p}\xi(x_i,.)\mathbb{E}\,\Pi_{i=p+1}^{p+q}\xi(x_i,.)|\mathbb{E}\,|\xi(x_j,.)|^r\mathbb{E}\,|\xi(x_l,.)|^u,$$

with $p+q+r+u=k$, with second argument of ξ standing for $\mathcal{P}\cap W_\lambda$.

IV Gaussian fluctuations: Cumulant Method

· If ξ stabilizes exponentially fast then mixed moments 'factorize': for all $p, q \in \mathbb{N}$ there are constants c_{p+q} and C_{p+q} s.t. for all $x_1, ..., x_{p+q} \in \mathbb{R}^d$,

$$|\mathbb{E}\,\Pi_{i=1}^{p+q}\xi(x_i, \mathcal{P} \cap W_\lambda) - \mathbb{E}\,\Pi_{i=1}^{p}\xi(x_i, \mathcal{P} \cap W_\lambda)\mathbb{E}\,\Pi_{i=p+1}^{p+q}\xi(x_i, \mathcal{P} \cap W_\lambda)|$$
$$\leq C_{p+q}\exp(-c_{p+q}s),$$

where

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

· This is main ingredient for showing $c^{(k)}(H_\lambda^\xi) = O(\lambda)$ for all $k \geq 2$. □

IV Gaussian fluctuations: mixed moments factorize.

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

· Let $R = \max_{i \le p+q} R^{\xi}(x_i, \mathcal{P})$. Hölder's inequality and the existence of all moments for ξ imply

$$
\begin{aligned}
&\mathbb{E}\,\Pi_{i=1}^{p+q} \xi(x_i, \mathcal{P}) \\
&= \mathbb{E}\,\Pi_{i=1}^{p+q} \xi(x_i, \mathcal{P}) \mathbf{1}(R \le \frac{s}{3}) + \mathbb{E}\,\Pi_{i=1}^{p+q} \xi(x_i, \mathcal{P}) \mathbf{1}(R > \frac{s}{3}) \\
&= \mathbb{E}\,\Pi_{i=1}^{p+q} \xi(x_i, \mathcal{P} \cap B_{\frac{s}{3}}(x_i)) \mathbf{1}(R \le \frac{s}{3}) + C_{p+q} \exp(-c_{p+q} s) \\
&= \mathbb{E}\,\Pi_{i=1}^{p+q} \xi(x_i, \mathcal{P} \cap B_{\frac{s}{3}}(x_i)) + 2C_{p+q} \exp(-c_{p+q} s) \\
&= \mathbb{E}\,\Pi_{i=1}^{p} \xi(x_i, \mathcal{P} \cap B_{\frac{s}{3}}(x_i)) \cdot \mathbb{E}\,\Pi_{i=p+1}^{p+q} \xi(x_i, \mathcal{P} \cap B_{\frac{s}{3}}(x_i)) \\
&\qquad\qquad + 2C_{p+q} \exp(-c_{p+q} s) \\
&= \mathbb{E}\,\Pi_{i=1}^{p} \xi(x_i, \mathcal{P}) \cdot \mathbb{E}\,\Pi_{i=p+1}^{p+q} \xi(x_i, \mathcal{P}) + 3C_{p+q} \exp(-c_{p+q} s). \qquad \square
\end{aligned}
$$

IV Gaussian fluctuations

We will use factorizing in variance asymptotics, specifically to show that the limit variance exists and is finite.

IV Gaussian fluctuations: Extensions

Gaussian fluctuations hold in these set-ups:

· the score function ξ need not be translation invariant

· stabilizing functionals on Poisson and binomial input on general metric spaces, including manifolds

· stabilizing functionals of general input on Euclidean spaces. Input must have factorizing correlations.

· multi-dimensional CLT for a vector of stabilizing functionals

· rates of multivariate normal convergence

· ξ can be polynomially stabilizing.

V Variance asymptotics; Poisson input

· Given homogenous rate τ Poisson input $\mathcal{P}_\tau$ on $\mathbb{R}^d$, and a score ξ, put

$$V^\xi(\tau) := \int_{\mathbb{R}^d} [\mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_\tau \cup \{x\})\xi(x, \mathcal{P}_\tau \cup \{\mathbf{0}\}) - \mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_\tau)\mathbb{E}\,\xi(x, \mathcal{P}_\tau)]dx.$$

Put

$$H^\xi_{\lambda\kappa} := \sum_{x \in \mathcal{P}_{\lambda\kappa}} \xi_\lambda(x, \mathcal{P}_{\lambda\kappa}), \quad \mathrm{supp}(\kappa) = \mathrm{K},$$

Thm (variance asymptotics): If ξ is exponentially stabilizing wrt $\mathcal{P}_\tau, \tau > 0$, and satisfies the p moment condition for some $p \in (2, \infty)$, then

$$\lim_{\lambda \to \infty} \lambda^{-1} \mathrm{Var} H^\xi_{\lambda\kappa} = \int_K \mathbb{E}\,\xi^2(\mathbf{0}, \mathcal{P}_{\kappa(x)})\kappa(x)dx + \int_K V^\xi(\kappa(x))\kappa^2(x)dx \in [0, \infty)$$

V Variance asymptotics

· Recall: if ξ stabilizes, then $\xi_\lambda(x, \mathcal{P}_{\lambda\kappa} \cup \{x\}) \xrightarrow{\mathcal{D}} \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)} \cup \{\mathbf{0}\})$.

· **Lemma** (continuity lemma for pairs of scores) Let x be a Lebesgue point for κ. If ξ is stabilizing w.r.t. $\mathcal{P}_{\kappa(x)}$, then for all $z \in \mathbb{R}^d$, we have as $\lambda \to \infty$

$$\left(\xi_\lambda(x, \mathcal{P}_{\lambda\kappa} \cup \{x + \lambda^{-1/d} z\}),\ \xi_\lambda(x + \lambda^{-1/d} z, \mathcal{P}_{\lambda\kappa} \cup \{x\})\right)$$

$$\left(\xi_\lambda(\mathbf{0}, (\mathcal{P}_{\lambda\kappa} - x) \cup \{\lambda^{-1/d} z\}),\ \xi_\lambda(\lambda^{-1/d} z, (\mathcal{P}_{\lambda\kappa} - x) \cup \{\mathbf{0}\})\right)$$

$$\xrightarrow{\mathcal{D}} \left(\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)} \cup \{z\}), \xi(z, \mathcal{P}_{\kappa(x)} \cup \{\mathbf{0}\})\right).$$

· We use this lemma to prove variance asymptotics. (Remember it for the next slide.)

V Variance asymptotics

- Mecke's Formula for the Poisson process $\mathcal{P}_{\lambda\kappa}$ says

$$\lambda^{-1}\mathrm{Var} H_\lambda^\xi = \lambda^{-1}[\mathbb{E}\,(H_\lambda^\xi)^2 - (\mathbb{E}\,H_\lambda^\xi)^2]$$
$$= \lambda \int_K \int_K \{\mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa} \cup \{y\})\xi_\lambda(y, \mathcal{P}_{\lambda\kappa} \cup \{x\})]$$
$$- \mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa})]\mathbb{E}\,[\xi_\lambda(y, \mathcal{P}_{\lambda\kappa})]\}\kappa(x)\kappa(y)\,dy\,dx$$
$$+ \int_K \mathbb{E}\,[\xi_\lambda^2(x, \mathcal{P}_{\lambda\kappa})]\kappa(x)\,dx.$$

V Variance asymptotics

· Mecke's Formula for the Poisson process $\mathcal{P}_{\lambda\kappa}$ says

$$\begin{aligned}\lambda^{-1}\mathrm{Var} H^{\xi}_{\lambda} &= \lambda^{-1}[\mathbb{E}\,(H^{\xi}_{\lambda})^2 - (\mathbb{E}\,H^{\xi}_{\lambda})^2] \\ &= \lambda \int_K \int_K \{\mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa} \cup \{y\})\xi_\lambda(y, \mathcal{P}_{\lambda\kappa} \cup \{x\})] \\ &\quad - \mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa})]\mathbb{E}\,[\xi_\lambda(y, \mathcal{P}_{\lambda\kappa})]\}\kappa(x)\kappa(y)\,dy\,dx \\ &\quad + \int_K \mathbb{E}\,[\xi^2_\lambda(x, \mathcal{P}_{\lambda\kappa})]\kappa(x)\,dx.\end{aligned}$$

· Put $y = x + \lambda^{-1/d}z$ in the right-hand side of the above (so $\lambda dy = dz, z \in \lambda^{1/d}(K - x)$). The two point correlation function $\{...\}$ becomes

$$\begin{aligned}\{...\} := \mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa} \cup \{x + \lambda^{-1/d}z\})\xi_\lambda(x + \lambda^{-1/d}z, \mathcal{P}_{\lambda\kappa} \cup \{x\})] \\ -\mathbb{E}\,[\xi_\lambda(x, \mathcal{P}_{\lambda\kappa})]\mathbb{E}\,[\xi_\lambda(x + \lambda^{-1/d}z, \mathcal{P}_{\lambda\kappa})].\end{aligned}$$

· Continuity lemma for pairs implies $\{...\}$ converges to

$$\mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)} \cup \{z\})\xi(z, \mathcal{P}_{\kappa(x)} \cup \{\mathbf{0}\}) - \mathbb{E}\,\xi(\mathbf{0}, \mathcal{P}_{\kappa(x)})\mathbb{E}\,\xi(z, \mathcal{P}_{\kappa(x)}).$$

V Variance asymptotics; Poisson input

· Assuming exponential stabilization, the integrand in the above is dominated by an integrable function of z over $\mathbb{R}^d$.

· The double integral in the above thus converges to

$$\int_K \int_{\mathbb{R}^d} \{\mathbb{E}\, \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)} \cup z) \xi(z, \mathcal{P}_{\kappa(x)} \cup \mathbf{0}) - \mathbb{E}\, \xi(\mathbf{0}, \mathcal{P}_{\kappa(x)}) \mathbb{E}\, \xi(z, \mathcal{P}_{\kappa(x)})\} \kappa(x)^2 dz dx$$

by dominated convergence and the assumed moment bounds.

· The integral is finite because of the factorizing property of moments of ξ.
□

VI Corollaries

This approach leads to central limit theorems and variance asymptotics for:

· Dimension estimators

· Simplex counts in Rips complex

· Entropy estimators

· Empirical covariogram

VI Corollary: Dimension estimators

· Let $\mathcal{X}_n := \{X_i\}_{i=1}^n$ be i.i.d. on manifold $\mathcal{M} \subset \mathbb{R}^d$; $m := \text{dim}(\mathcal{M})$ **unknown**, called the intrinsic dimension. Interpoint distances are **known**.

· $D_j(x, \mathcal{X}_n)$:= dist. between x and its jth nearest neighbor in $\mathcal{X}_n$.

· **Goal.** Estimate intrinsic dimension m using only the information about interpoint distances. Fix $k \geq 3$. Define 'score' at X_1 wrt $\mathcal{X}_n$ by

$$\xi_k(X_1, \mathcal{X}_n) := (k-2) \left(\sum_{j=1}^{k-1} \log \frac{D_k(X_1, \mathcal{X}_n)}{D_j(X_1, \mathcal{X}_n)} \right)^{-1}.$$

· **LLN, CLT** Assume X_i have a.e. continuous density bounded away from zero and infinity. Fix $k \geq 3$.

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i \leq n} \xi_k(X_i, \mathcal{X}_n) = m = \dim \mathcal{M} \quad \text{in prob.}$$

and ……

VI Corollary: Dimension estimators

.....and (CLT)

$$\frac{\sum_{i \le n} \xi_k(X_i, \mathcal{X}_n) - \mathbb{E} \sum_{i \le n} \xi_k(X_i, \mathcal{X}_n)}{\sqrt{\mathrm{Var} \sum_{i \le n} \xi_k(X_i, \mathcal{X}_n)}} \xrightarrow{\mathcal{D}} N(0, \sigma^2_{\xi_k}).$$

Why does this work?

VI Corollary: Dimension estimators

- $\mathcal{P} :=$ homogeneous rate one Poisson pt process on $\mathbb{R}^m$, $m \geq 2$.
- Fix an integer $k \geq 3$.
- $D_j := D_j(x, \mathcal{P}) :=$ dist. between x and its jth nearest neighbor in $\mathcal{P}$.
- Conditional on D_k, for all $1 \leq j \leq k-1$

$$\frac{D_j^m}{D_k^m} \stackrel{\mathcal{D}}{=} U_{(j)} \Rightarrow -\log \frac{D_j^m}{D_k^m} \stackrel{\mathcal{D}}{=} X_{(j)}, \qquad X \stackrel{\mathcal{D}}{=} \exp(1).$$

VI Corollary: Dimension estimators

· $\mathcal{P} :=$ homogeneous rate one Poisson pt process on $\mathbb{R}^m$, $m \geq 2$.

· Fix an integer $k \geq 3$.

· $D_j := D_j(x, \mathcal{P}) :=$ dist. between x and its jth nearest neighbor in $\mathcal{P}$.

· Conditional on D_k, for all $1 \leq j \leq k-1$

$$\frac{D_j^m}{D_k^m} \stackrel{\mathcal{D}}{=} U_{(j)} \Rightarrow -\log \frac{D_j^m}{D_k^m} \stackrel{\mathcal{D}}{=} X_{(j)}, \qquad X \stackrel{\mathcal{D}}{=} \exp(1).$$

· The sum over $j = 1, ..., k-1$ has a gamma$(k-1, 1)$ distribution: $m \sum_{j=1}^{k-1} \log \frac{D_k}{D_j} \stackrel{\mathcal{D}}{=} \Gamma_{k-1}$, where Γ_{k-1} is gamma$(k-1, 1)$. Reciprocating:

$$\Rightarrow (k-2) \left(\sum_{j=1}^{k-1} \log \frac{D_k}{D_j} \right)^{-1} \stackrel{\mathcal{D}}{=} m(k-2)(\Gamma_{k-1})^{-1}.$$

· Expectation of LHS is m. In other words the LHS is an unbiased estimator of dimension for any $k \geq 3$ (Bickel + Levina '05).

VI Corollary: Simplex counts in Rips complex

· Rips complex 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 $\sigma_k^{(r)}(x, \mathcal{X}) := \frac{\text{number of } k\text{-simplices containing } x}{k+1}$

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

· **WLLN; CLT for simplex counts; (critical regime)**. Let Ξ be a stationary point process on $\mathbb{R}^d$ with intensity ρ. For $r \in (0, \infty)$ fixed

$$\lambda^{-1}\mathbb{E} \sum_{x \in \Xi \cap W_\lambda} \sigma_k^{(r)}(x, \Xi \cap W_\lambda) \to \rho \mathbb{E}_{\mathbf{0}} \xi(\mathbf{0}, \Xi).$$

If Ξ has 'factorizing correlations' then $\sum_{x \in \Xi \cap W_\lambda} \sigma_k^{(r)}(x, \Xi \cap W_\lambda)$ has Gaussian fluctuations after centering and scaling.

· **Extensions:**

· $\mathcal{P}_{\lambda\kappa}$ is Poisson input on a C^1 submanifold with intensity $\lambda\kappa$, κ continuous with bounded support; rates of conv. in multivariate CLT.

VI Corollary: Shannon entropy estimators

· Shannon differential entropy of a r.v. X with density κ on the manifold $\mathcal{M}$:

$$H(\kappa) := -\int_{\mathcal{M}} \kappa(x) \log(\kappa(x)) dx.$$

· We seek estimators of $H(\kappa)$ in terms of nearest neighbor distances.

· dim($\mathcal{M}$) = m. Put

$$\xi(x, \mathcal{X}) := \log(e^{\gamma} \omega_m D_1^m(x, \mathcal{X}))$$

where $\omega_m := \pi^{m/2} [\Gamma(1 + \frac{m}{2})]^{-1}$ is volume of the unit radius m-dimensional ball, $\gamma := 0.57721...$ is Euler's constant.

· Let $\mathcal{X}$ be $\mathcal{P}_a$, a homog. PPP of intensity a on $\mathbb{R}^m$. Then

$$\mathbb{E}\left[\xi(\mathbf{0}, \mathcal{P}_a)\right] = a \int_0^{\infty} \log(e^{\gamma} u) e^{-au} du$$

since the probability that the volume of the nearest neighbors ball around $\mathbf{0}$ exceeds u is e^{-au}.

VI Corollary: Shannon entropy estimators

· Relation between expected score and intensity:

$$\mathbb{E}\,[\xi(\mathbf{0},\mathcal{P}_a)] = \gamma + a\int_0^\infty (\log u)e^{-au}du = -\log a$$

since $\int_0^\infty \log(\frac{w}{a})e^{-w}dw = -\log a - \gamma$, γ is Euler's constant.

· Entropy estimators $S(n^{1/m}\mathcal{X}_n) := \frac{1}{n}\sum_{i\le n}\xi(n^{1/m}X_i, n^{1/m}\mathcal{X}_n)$. LLN for stabilizing ξ yields

$$\begin{aligned}\mathbb{E}\,S(n^{1/m}\mathcal{X}_n) &= \mathbb{E}\,\xi(n^{1/m}X_1, n^{1/m}\mathcal{X}_n)\\ &= \mathbb{E}\,\xi(\mathbf{0}, n^{1/m}(\mathcal{X}_n - X_1))\\ &\to -\int_K \mathbb{E}\,[\xi(\mathbf{0},\mathcal{P}_{\kappa(x)})]\kappa(x)dx\\ &= -\int_K \log(\kappa(x))\kappa(x)dx.\end{aligned}$$

VI Corollary: Shannon entropy estimators

· Entropy estimators $S(n^{1/m}\mathcal{X}_n) := \frac{1}{n}\sum_{i\leq n}\xi(n^{1/m}X_i, n^{1/m}\mathcal{X}_n)$.

LLN, CLT for Shannon entropy estimators. Let κ be a density function with compact support on C^1 manifold $\mathcal{M}$, κ bounded away from zero and infinity, $m = \dim(\mathcal{M})$. The Shannon entropy estimators $S(n^{1/m}\mathcal{X}_n)$ converge in L^2 to the Shannon entropy $-\int_{\mathcal{M}}\kappa(x)\log(\kappa(x))dx$. Moreover

$$\frac{S(n^{1/m}\mathcal{X}_n) - \mathbb{E}\,S(n^{1/m}\mathcal{X}_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,\sigma^2),$$

with a variance σ^2 having a closed form.

VI Corollary: Covariograms

· Consider the random field $\{\zeta(x, \mathcal{P})\}_{x \in \mathbb{R}^d}$, $\mathcal{P}$ a pt process.

· W_λ: window of volume λ.

· The covariogram $\gamma(h), h > 0$, of the random field $\{\zeta(x, \mathcal{P})\}_{x \in W_\lambda}$: the average squared difference between the values of ζ at points separated by distance which is 'close' to h, namely

$$\gamma(h) := \frac{1}{\delta^{d-1} \cdot |W_\lambda|} \int_{x,y \in W_\lambda :\ h-\delta \le |x-y| \le h+\delta} \mathbb{E}\left(\zeta(x, \mathcal{P}) - \zeta(y, \mathcal{P})\right)^2 dx dy.$$

· For a fixed h and x, given a small 'tolerance' δ, there is a set of measure proportional to δ^{d-1} on which points y in this set have separation distance from x in the interval $(h - \delta, h + \delta)$, whence the scaling by δ^{d-1}.

· Covariogram: a measure of spatial dependence of the field. It may be estimated by the so-called empirical covariogram, obtained by sampling the random field at point locations given by an independent stationary pt process Ξ.

VI Corollary: Covariograms

This gives rise to the empirical covariogram ($h > 0$):

$$\gamma_\lambda(h \pm \delta) = \frac{1}{\delta^{d-1}|W_\lambda|} \sum_{x_i, x_j \in \Xi \cap W_\lambda:\ |x_j - x_i| \in (h-\delta, h+\delta)} (\zeta(x_i, \mathcal{P}) - \zeta(x_j, \mathcal{P}))^2$$

$$= \frac{1}{\delta^{d-1}|W_\lambda|} \sum_{x_i \in \Xi \cap W_\lambda} \xi(x_i, \Xi_\lambda) \quad (*)$$

with

$$\xi(x_i, \Xi_\lambda) := \sum_{x_j \in \Xi \cap W_\lambda:\ |x_j - x_i| \in (h-\delta, h+\delta)} (\zeta(x_i, \mathcal{P}) - \zeta(x_j, \mathcal{P}))^2.$$

· **LLN, CLT** May deduce the limit theory for the sums (*) as $\lambda \to \infty$.

VII General input

· **Question.** If the input Ξ is neither Poisson nor binomial, when do we get results which are qualitatively similar?

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

$$\sum_{x \in \Xi \cap W_\lambda} \delta_{\lambda^{-1/d} x}$$

where Ξ is determinantal pt process, $\Xi \cap W_\lambda := \Xi \cap [-\frac{1}{2}\lambda^{\frac{1}{d}}, \frac{1}{2}\lambda^{\frac{1}{d}}]^d$.

· Nazarov and Sodin (2012): establish asymptotic normality of

$$\sum_{x \in \Xi \cap W_\lambda} \delta_{\lambda^{-1/d} x}$$

where Ξ is zero set of Gaussian analytic function.

· **Goal**. Extend these results to ξ-weighted statistics

$$\mu_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda) \delta_{\lambda^{-1/d} x}.$$

VII General input

· How to extend these results (LLN, CLT, Variance asymptotics) to non-Poisson input?

· Can we establish analogs of the main results for input having spatial dependencies?

VII General input: correlation functions

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

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

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

Rks.

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

$\rho^{(k)}(x_1, \ldots, x_k) \geq \Pi_{i=1}^k \rho^{(1)}(x_i)$ implies Ξ is attractive

$\rho^{(k)}(x_1, \ldots, x_k) \leq \Pi_{i=1}^k \rho^{(1)}(x_i)$ implies Ξ is repulsive

VII General input: factorizing correlations

· **Definition (factorizing correlations).** A pt process Ξ *has factorizing correlations* if there is a fast decreasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ such that for all $p, q \in \mathbb{N}$ there are decay 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)

· Poisson input has factorizing correlations ($\phi = 0$)

VII General input

· **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 non-negative definite kernel of locally trace class integral operator from $L^2(\mathbb{R}^d)$ to itself.

· DPP is repulsive.

· **Fact** If $|K(x, y)| \leq \phi(|x - y|)$, with ϕ fast decreasing, then the DPP has factorizing correlations.

· **Ex.** Infinite Ginibre ensemble on complex plane with kernel

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

has factorizing correlations.

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 trans. invariant (in the class of Gaussian power series, it is the only one which is trans. inv.).

· Z_F exhibits local repulsivity.

· Z_F has factorizing correlations (Nazarov and Sodin (2012)).

VII General input

- Further examples of pt processes with factorizing correlations:
  - Permanental pt processes with fast decreasing kernel
  - Certain Gibbs pt processes.

VII General input

· Ξ: pt process on $\mathbb{R}^d$ with intensity ρ.

· $\Xi \cap W_\lambda := \Xi \cap [\frac{-\lambda^{\frac{1}{d}}}{2}, \frac{\lambda^{\frac{1}{d}}}{2}]^d$. Put

$$\mu_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda) \delta_{\lambda^{-1/d} x}.$$

Recall:
Thm - LLN (Błaszczyszyn, Yogeshwaran, Y (2019)): If ξ is trans. inv., stabilizing wrt Ξ and satisfies the p moment condition for some $p \in (1, \infty)$, then for all $f \in \mathbb{B}([-\frac{1}{2}, \frac{1}{2}]^d)$

$$\lim_{\lambda \to \infty} \lambda^{-1} \mathbb{E} \langle \mu_\lambda^\xi, f \rangle = \rho \cdot \mathbb{E}_{\mathbf{0}} \xi(\mathbf{0}, \Xi) \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f(x) dx.$$

VII General input: Gaussian fluctuations

Thm- CLT(Błaszczyszyn, Yogeshwaran, Y (2019, 2022))
$\mu_\lambda^\xi := \sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda) \delta_{\lambda^{-1/d} x}$. Assume:

· Ξ has factorizing correlations.

· ξ is exponentially stabilizing wrt Ξ,

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

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

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

· **Remarks.** Extends Soshnikov (2002) and Shirai + Takahashi (2003) who restrict to $\xi \equiv 1$, determinantal input.

· Extends Nazarov and Sodin (2012), who restrict to $\xi \equiv 1$, Ξ the zero set of Gaussian entire function.

VII General input: Variance asymptotics

- Given input Ξ with factorizing correlations and a score ξ, put

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

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

- **Thm (BYY '19)**: If ξ is exponentially stabilizing wrt Ξ, if ξ satisfies the p moment condition for some $p \in (2,\infty)$, then for all $f \in \mathbb{B}([-\frac{1}{2},\frac{1}{2}]^d)$

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

- **Rk.** Extends Soshnikov (2002), who assumes $\xi \equiv 1$, determinantal input.

· Recall: To show

$$\frac{\langle \mu_\lambda^\xi, f \rangle}{\sqrt{\mathrm{Var}\langle \mu_\lambda^\xi, f \rangle}} \xrightarrow{\mathcal{D}} N(0,1),$$

it suffices to show that cumulants for $\langle \mu_\lambda^\xi, f \rangle$ have linear growth $O(\lambda)$.

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

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

· Need to show that the mixed moments factorize, 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 φ is fast decreasing and $s := \inf_{i \in \{1,...,p\},\ j \in \{p+1,...,p+q\}} |x_i - x_j|$.

VII General input: Proof of CLT via cumulants

· **Fact**. Ξ has factorizing correlations and ξ exponentially stabilizing $\Rightarrow$ mixed moments factorize.

VII General input: Applications

· These general results immediately yield limit theory (WLLN, Gaussian fluctuations, variance asymptotics) for statistics of geometric structures on pt processes with factorizing correlations (FC). This includes:

· Rips clique count on FC point processes, including DPP with fast decreasing kernel, zero set of Gaussian entire function.

· number of critical points for FC point processes.

Example: k-critical points

· Let $\mathcal{X}$ be a point cloud in $\mathbb{R}^d$.
· We define the distance function from $\mathcal{X}$ as

$$d_{\mathcal{X}}(y) := \min_{x \in \mathcal{X}} |y - x|, \; y \in \mathbb{R}^d.$$

· Let $f : \mathbb{R}^d \to \mathbb{R}$ be a C^2 function. A point $c \in \mathbb{R}^d$ is called a critical point of f if $\nabla f(c) = 0$, and the real number $f(c)$ is called a critical value of f .

· We would like to study the number of critical points for $d_{\mathcal{X}}$ (but the distance function may not be C^2). Theory of critical points may be extended for such f.

Example: k-critical points

· $\mathcal{X} \subset \mathbb{R}^d$ a finite pt set; $\mathcal{Y} \subset \mathcal{X}$ a set of $k+1$ points, with k fixed, $k \leq d$.

· Let $c_{\mathcal{Y}}$ and $r_{\mathcal{Y}}$ denote the center and radius of the unique $(k-1)$-dimensional sphere containing $\mathcal{Y}$.

· The subset $\mathcal{Y} \subset \mathcal{X}$ generates an index k critical point of $\mathcal{X}$ iff

$$c_{\mathcal{Y}} \in \operatorname{int}(\operatorname{conv}(\mathcal{Y})) \text{ and } \mathcal{X} \cap B_{r_{\mathcal{Y}}}(c_{\mathcal{Y}}) = \emptyset.$$

· Let Ξ_λ be the restriction of Ξ to volume λ window W_λ. We express the total number of *index* k *critical points* in Ξ_λ as a sum of scores

$$\sum_{x \in \Xi_\lambda} \xi_{\mathrm{crit}}^{(k)}(x, \Xi_\lambda)$$

through an appropriate choice of $\xi_{\mathrm{crit}}^{(k)}$ which depends on local data.

Example: k-critical points

· Let Ξ_λ be the restriction of Ξ to volume λ window W_λ. We may express the total number of *index k critical points* in Ξ_λ as a sum of scores

$$\sum_{x \in \Xi_\lambda} \xi_{\text{crit}}^{(k)}(x, \Xi_\lambda).$$

· **LLN, CLT for critical pts** We may deduce from our general results the limit theory for

$$\sum_{x \in \Xi_\lambda} \xi_{\text{crit}}^{(k)}(x, \Xi_\lambda), \quad \lambda \to \infty.$$

When the input is Poisson we may obtain rates of convergence in (multivariate) CLT.

· Re-scale to obtain limit results for input on compact sets, as intensity of input tends to infinity.

VIII Stabilization of General Poisson Functionals

· Most models of physical systems involve particles which interact 'locally', inducing long-range interactions.

· We let $H(\mathcal{X})$ be a generic functional of a point cloud $\mathcal{X}$. We do not assume that it may be represented as a sum of stabilizing scores.

· Still, many models of physical systems involve particles which interact 'locally', inducing long-range interactions.

· Can this be exploited to yield limit theory for $H(\mathcal{X})$, when the size of the point cloud $\mathcal{X}$ tends to infinity?

VIII Stabilization of General Poisson Functionals

· We take our particles to be points, usually the realization of an i.i.d. collection of r.v. or a homogeneous Poisson point process $\mathcal{P}_1$ on $\mathbb{R}^d$.

· For ease of exposition, we consider $W_\lambda := [-\frac{1}{2}\lambda^{1/d}, \frac{1}{2}\lambda^{1/d}]^d$.

· H: a generic functional defined on finite point sets.

· We are interested in the behavior of the **Poisson** functional $H(\mathcal{P}_1 \cap W_\lambda)$.

VIII Stabilization of Poisson Functionals

Natural questions:

1. (LLN) When do we have $\lim_{\lambda\to\infty} \frac{H(\mathcal{P}_1 \cap W_\lambda)}{\lambda} = \text{constant} \quad \text{a.s.?}$
2. (CLT) When do we have

$$\frac{H(\mathcal{P}_1 \cap W_\lambda) - \mathbb{E}\, H(\mathcal{P}_1 \cap W_\lambda)}{\sqrt{\mathrm{Var} H(\mathcal{P}_1 \cap W_\lambda)}} \xrightarrow{\mathcal{D}} N(0,1)?$$

a. Geometric Graph (Gilbert Graph)

· **Def.** Given a finite point set $\mathcal{X}$, $r \in (0, \infty)$, define the r-offset

$$\mathcal{U}(\mathcal{X}, r) := \bigcup_{x \in \mathcal{X}} B_r(x).$$

· When $\mathcal{X}$ is PPP we get the Boolean model. It gives rise to the geometric graph $G_r(\mathcal{X})$:

$$x \leftrightarrow y \text{ iff } B_{r/2}(x) \cap B_{r/2}(y) \neq \emptyset.$$

· **Def.** Let $H(\mathcal{X})$ be the number of edges in geometric graph $G_r(\mathcal{X})$.

· **CLT (critical regime):** $d \geq 2, r > 0$:

$$\frac{H(\mathcal{P}_1 \cap W_\lambda) - \mathbb{E}\, H(\mathcal{P}_1 \cap W_\lambda)}{\sqrt{\mathrm{Var} H(\mathcal{P}_1 \cap W_\lambda)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Example: r-offset

· $\mathcal{X}_n$ a point cloud of n i.i.d. uniform points in $[-\frac{1}{2}, \frac{1}{2}]^d$, $r > 0$.

· r-offset: $\mathcal{U}(\mathcal{X}_n, r) = \bigcup_{x \in \mathcal{X}_n} B_r(x)$.

· Number of holes and components in $\mathcal{U}(\mathcal{X}_n, r)$ depends on nr^d:

$nr^d \to 0$	subcritical regime (sparse)
$nr^d \to c \in (0, \infty)$	critical regime
$nr^d \to \infty$	supercritical regime

· Different regimes: Hall; Hug, Last + Schulte; Yogeshwaran + Adler, Penrose + Y; Bobrowski + Mukherjee, Bobrowski + Kahle, Krebs + Polonik, Owada + Adler; Hiraoka, Shirai, Trinh.

b. Betti Numbers for Cech Complex on PPP

Def (Cech Complex). Given a finite point set $\mathcal{X}$, $r \in (0, \infty)$, put

$$C^{(r)}(\mathcal{X}) := \mathrm{Cech}^{(r)}(\mathcal{X}) := \{\sigma \subset \mathcal{X} : \bigcap_{x \in \sigma} B_r(x) \neq \emptyset\}.$$

VIII Stabilization of Poisson Functionals

· Topological data analysis: determine Betti numbers of $C^{(r)}(\mathcal{X})$. Zeroth Betti number is the number of connected components, β_1 is the number of one dimensional holes, β_2 is the number of two dimensional holes,...

· **CLT (critical regime)** $d \geq 2, k \in \{0, 1, ..., d-1\}, r > 0$:

$$\frac{\beta_k(C^{(r)}(\mathcal{P}_1 \cap W_\lambda)) - \mathbb{E}\, \beta_k(C^{(r)}(\mathcal{P}_1 \cap W_\lambda))}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

· Yogeshwaran, Subag, Adler, 2017
· Hiraoka; Shirai, Trinh, 2018
· Krebs and Polonik, 2022: multivariate CLT
· Hirsch and Krebs, 2021: FCLT

VIII Stabilization of Poisson Functionals

· Write $W_\lambda := \cup_{i=1}^{\lambda} W_{\lambda,i}$, where $W_{\lambda,i}$ are disjoint sub-cubes of volume 1.

· In general H is NOT additive:

$$H(\mathcal{P}_1 \cap W_\lambda) \neq \sum_{i=1}^{n} H(\mathcal{P}_1 \cap W_{\lambda,i}).$$

VIII Stabilization of Poisson Functionals

- Restriction of H to disjoint sets does not give independence

- $H(\mathcal{P}_1 \cap A)$ and $H(\mathcal{P}_1 \cap B)$ are dependent in general.

VIII Stabilization of Poisson Functionals: Difference operators

· A key idea is to compare $H(\mathcal{P}_1 \cap W_\lambda)$ with $H((\mathcal{P}_1 \cap W_\lambda) \cup \{\mathbf{0}\})$.

· In other words, what happens to H when we insert an extra point at the origin into homogenous rate 1 Poisson input $\mathcal{P}_1$? How does H change?

VIII Stabilization of Poisson Functionals: Difference operators

· Let $\mathcal{P}_1$ be unit intensity PPP on $\mathbb{R}^d$.

·**Def.** Say that H **weakly stabilizes** on $\mathcal{P}_1$ if there is an a.s. finite r.v. Δ such that

$$H((\mathcal{P}_1 \cap W_\lambda) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap W_\lambda) \to \Delta.$$

· This condition says that the 'add-one cost' does not propagate far.

VIII Stabilization of Poisson Functionals: Difference operators

· **Def.** Say that H **weakly stabilizes** on $\mathcal{P}_1$ if there is an a.s. finite r.v. Δ such that

$$H((\mathcal{P}_1 \cap W_\lambda) \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap W_\lambda) \to \Delta.$$

· **Def.** Recall first order difference operator

$$D_{\mathbf{0}} H(\mathcal{P}_1) := H(\mathcal{P}_1 \cup \{\mathbf{0}\}) - H(\mathcal{P}_1).$$

· Weak stabilization says that the first order difference operator has a behavior which is determined by local data (what is outside the cube W_λ plays no role).

VIII Stabilization of Poisson Functionals: Difference operators

· Which functionals H stabilize in the above sense?

· Consider the nearest neighbors graph (put an edge between every point and its nearest neighbor).

VIII Stabilization of Poisson Functionals: CLT and variance asymptotics

· **Theorem**. Let H be a functional on locally finite point sets in $\mathbb{R}^d$. Assume:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) bounded increments:

$$\sup_{Q,\ Q \text{ a cube}} \mathbb{E}\,|H(\mathcal{P}_1 \cap Q \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost" (i.e. stabilizes)

Then

$$\frac{H(\mathcal{P}_1 \cap W_\lambda) - \mathbb{E}\,H(\mathcal{P}_1 \cap W_\lambda)}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap W_\lambda)}{\lambda} \to \sigma^2.$$

VIII Stabilization of Poisson Functionals: Difference operators

Applications

· The following functionals stabilize and satisfy the CLT and variance asymptotics:

a. Betti numbers: critical and sub-critical regime. In the latter regime there is no infinite component, so showing stabilization is less technical (Trinh- 2019):

b. The number of edges in the random geometric graph with parameter r,

VIII Stabilization of Poisson Functionals: CLT

· **Theorem**. Let H be a functional on locally finite point sets in $\mathbb{R}^d$. Assume:

(i) translation invariance, i.e., $H(\mathcal{X} + y) = H(\mathcal{X}),\ y \in \mathbb{R}^d$,

(ii) bounded increments:

$$\sup_Q \mathbb{E}\,|H(\mathcal{P}_1 \cap Q \cup \{\mathbf{0}\}) - H(\mathcal{P}_1 \cap Q)|^{2+\epsilon} < \infty,$$

(iii) stability of 'add-one cost". Then

$$\frac{H(\mathcal{P}_1 \cap W_\lambda) - \mathbb{E}\, H(\mathcal{P}_1 \cap W_\lambda)}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

$$\frac{\mathrm{Var} H(\mathcal{P}_1 \cap W_\lambda)}{\lambda} \to \sigma^2.$$

Remarks:

· no formula for σ^2, no rate of normal convergence,

· Trinh has extension to non-uniform input.

· If the add-one cost $D_{\mathbf{0}} H(\mathcal{P}_1)$ is non-degenerate, then $\sigma^2 > 0$.

Proof of CLT.

1. Express H as a sum of martingale differences and use McLeish CLT
2. Trinh: use Poincaré inequality for $H := H(\mathcal{P})$:

   · **Poincaré Inequality**: $\mathrm{Var} H \leq \mathbb{E} \int_{\mathbb{R}^d} (D_x H)^2 dx.$

Concluding Remarks

- **II Stabilization** (a surrogate for continuity and independence)
- **III Laws of large numbers** (Mecke-Campbell formula)
- **IV Gaussian fluctuations** (three approaches)
- **V Variance asymptotics** (fast decay of mixed moments)
- **VI Corollaries**
- **VII Statistics of general input** (factorizing correlations)
- **VIII Statistics of Poisson functionals** (weak stabilization)

References

Barbour, A. D. and Xia, A. (2006). Normal approximation for random sums. *Adv. in Appl. Probab.* **38** 693–728.

Baryshnikov, Yu. and Yukich, J. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.* **15** 213–253.

Błaszczyszyn, B., Yogeshwaran, D. and J. Yukich (2019). Limit theory for geometric statistics of point processes having fast decay of correlations, Annals of Probability, Vol. 47, Number 2, 835-895.

Błaszczyszyn, B., Yogeshwaran D. and J. Yukich (2022+). Limit theory for asymptotically de-correlated dynamic spatial random models, preprint.

Bobrowski, O. and Mukherjee, S. (2015). The topology of probability distributions on manifolds. *PTRF* **161** 651–686.

Hug, D., Last, G., and Schulte, M. (2016). Second order properties and central limit theorems for geometric functionals of Boolean models. *Ann. Appl. Probab.* **26** 73–135.

References

Lachièze-Rey, R. and Peccati, G. (2017). New Berry-Esseen bounds for functionals of binomial point processes. *Ann. Appl. Probab.* **27** 1992–2031.

Lachièze-Rey, R., Schulte, M., and Yukich, J. (2019). Normal approximation for stabilizing functionals. *Ann. Appl. Probab.* **29** 931–991.

Last, G., Peccati, G., and Schulte, M. (2016). Normal approximations on the Poisson space: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory and Related Fields* **165** 667–723.

Penrose, M. D. (2007). Gaussian limits for random geometric measures. *Electron. J. Probab.* **12** 989–1035.

Penrose, M. D. and Yukich, J. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11** 1005–1041.

References

Penrose, M. D. and Yukich, J. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303.

Penrose, M. D. and Yukich, J. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications*, Lecture Note Series Inst. for Math. Sci. National Univ. Singapore **5** 37–58.

Penrose, M. D. and Yukich, J. (2013). Limit theory for point processes in manifolds. *Ann. Appl. Probab.* **23** 2161–2211.

Schreiber, T. (2010). Limit theorems in stochastic geometry. Oxford University Press, Oxford.

Schulte, M. and Yukich, J. E. (2019). Multivariate second order Poincaré inequalities for Poisson functionals. *EJP* **24**, paper no. 130.

Schulte, M. and Yukich, J. (2022). Rates of multivariate normal approximation for statistics in geometric probability, Ann. Appl. Probab., to appear.

THANK YOU

Example: Ripley's K function

· Ξ : stationary point process on $\mathbb{R}^d$ with intensity ρ.

· $A \subset \mathbb{R}^d$, $|A|$ finite.

· The K-function is defined by

$$K(r) = \frac{1}{|A|\rho^2}\mathbb{E} \sum_{x \in \Xi \cap A} \sum_{y \in \Xi} \mathbf{1}(|x-y| \leq r), \quad r > 0.$$

· Naive estimator of K-function (W_λ a window of volume λ):

$$\hat{K}(r) = \frac{1}{\lambda\rho^2} \sum_{x \in \Xi \cap W_\lambda} \sum_{y \in \Xi \cap W_\lambda} \mathbf{1}(|x-y| \leq r), \; r > 0.$$

· **Goal.** Express this estimator (as well as less naive ones) as a sum of scores

$$\sum_{x \in \Xi \cap W_\lambda} \xi(x, \Xi \cap W_\lambda)$$

through an appropriate choice of a score ξ which should depend only on local data. Find the limit theory as $\lambda \to \infty$.