Limit theorems for topological and geometric functionals

Raphaël Lachièze-Rey (Inria Paris)
Mini-course for the 2025 Aarhus TDA summer school
4-8 August 2025

1. Geometric models
2. Geometric functionals
3. Abnormal fluctuations
4. Conditions for standard statistical behaviour

Geometric models

Random sets

- The favoured probabilistic framework to deal with random geometric structures is that of **Random closed sets**:
  - $\mathscr{F} = \{F \subset \mathbb{R}^d \text{ closed}\}$
  - A set $F \in \mathscr{F}$ is analysed through its intersection with other sets: for $K \subset \mathbb{R}^d$,

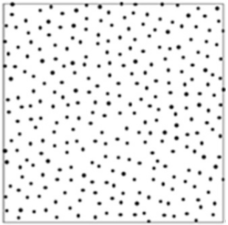
$$\psi_K(F) := \mathbf{1}\{K \cap F \neq \emptyset\}$$

  - A random set F is mathematically a family of "hit-and-miss" informations $\mathbf{1}\{\mathsf{F} \cap K \neq \emptyset\}, K$ compact.
  - The "Effros", or "hit-and-miss" σ-algebra on $\mathscr{F}$ is generated by the $\{\psi_K; K \text{ compact}\}$. Hence the law of a random (closed) set F is characterised by the **capacity functional**

$$T_{\mathsf{F}}(K) := \mathbf{E}\psi_K(\mathsf{F}) = \mathbf{P}(\mathsf{F} \cap K \neq \emptyset), K \text{ compact}.$$

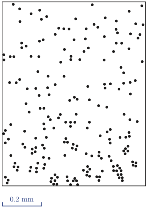
Point patterns I

If $F = \{x_i; i \geqslant 1\}$ is a.s. a set of isolated points, we preferentially note $F = P$ and say P is a **point process**.

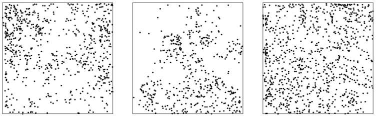


Photoreceptor in a chicken's eye Jiao et al. [2014]

Point patterns II



Cell nuclei in a specimen of joint cartilage Stoyan et al. [1995]



Hickories, maples, oaks Gelfand et al. [2010]

Point patterns III

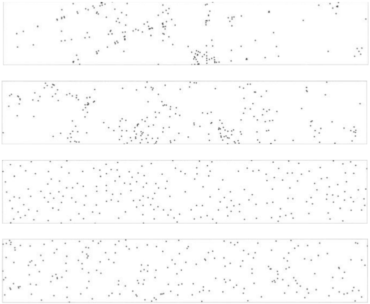


Fig. 11.1. A few examples of point processes. From top to bottom, intersections of random lines, a cluster process, a repulsion process and purely random points.

Some models Lantuéjoul [2002]

Random sets I: Point processes

- **Binomial processes**: for μ a locally finite measure on $\mathbb{R}^d$ without atoms, let $X_1, \ldots, X_n$ i.i.d uniform in B_n the ball centred in 0 with volume n, and $\mathsf{P}_n = \{X_1, \ldots, X_n\}$
- **Exercise:** we have the limit for $A \subset \mathbb{R}^d$ (bounded or not)

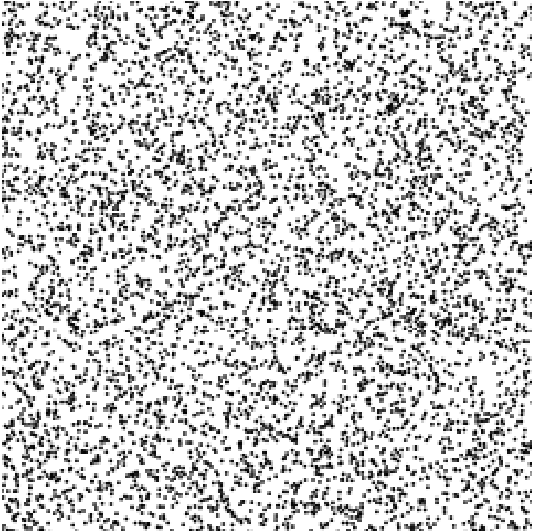
$$\#\mathsf{P}_n \cap A \to \mathrm{Poiss}(\mu(A)) \text{ (Poisson distribution)}$$

- This implies convergence to a limit point process P called **Poisson point process** with intensity μ, characterised by

$$\#\mathsf{P} \cap A \overset{(d)}{=} \mathrm{Poiss}(\mu(A)), A \subset \mathbb{R}^d$$

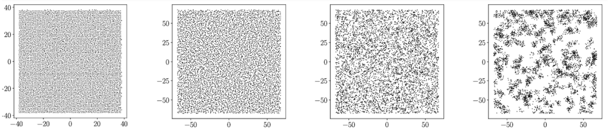
- This is **convergence in the vague topology** (compact)
- stationary Poisson point process: $\mu = \lambda \mathbf{Leb}^d$ $(\lambda > 0)$

stationary Poisson point process



Other point process models

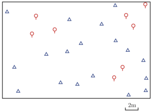
- **Perturbed lattices:** $\mathsf{P} = \{k + U_k; k \in \mathbb{Z}^d\}$ where the $U_k \in \mathbb{R}^d$ are i.i.d.
- **Cluster processes** $\{\tau_x \mathsf{P}_x; x \in \mathsf{P}\}$ where P is a "base" point process, and the P_x are i.i.d. realisations of a "small" point process.
  - **Thomas process**: P is Poisson with intensity $\lambda \textbf{Leb}^d$, P_x is Poisson with Gaussian intensity $\mathcal{N}(0, \sigma^2)$
- Zeros of a random function $F : \mathbb{R}^d \to \mathbb{R}^d$, eigenvalues of a random matrix, system of particles, ...



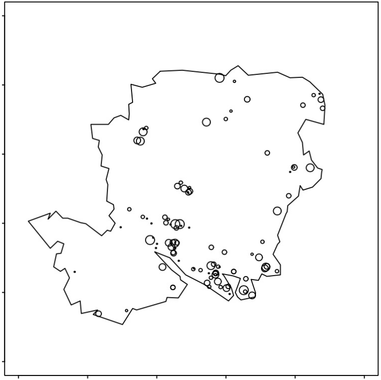
Which is which?

Marked point processes I

One can attach i.i.d. marks M_i of any sort to points of a point process: age, color, shape, etc...



Spruces, birch Stoyan et al. [1995]



gastrointestinal disease in UK
Gelfand et al. [2010]

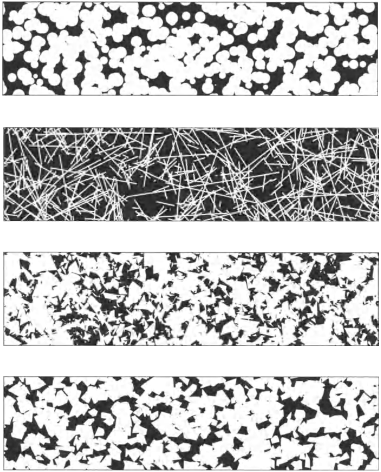
Marked point processes II

Exercise

Let $\mathsf{P} = \{x_i; i \geqslant 1\}$ *a Poisson point process with some intensity* μ *and* M_i *i.i.d. with law* ν *on some space* M. *Let* $\bar{\mathsf{P}} = \{(x_i, M_i); i \geqslant 1\}$. *Show that* $\bar{\mathsf{P}}$ *is Poisson with intensity* $\mu \times \nu$ *on* $\mathbb{R}^d \times \mathsf{M}$, *i.e.*

$$\#\bar{\mathsf{P}} \cap (A \times B) \overset{(d)}{=} \mathrm{Poiss}(\mu(A)\nu(B)).$$

Germ grain I



Some germ grain models Lantuéjoul [2002]

Germ grain II

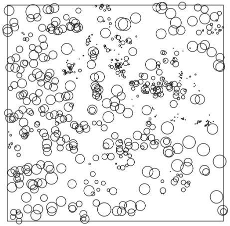


Figure: pines in North America Stoyan et al. [1995]

Germ grain III

Definition

A germ-grain process F is formed by a point process $\mathsf{P} = \sum_i \delta_{x_i}$ marked by a probability measure ν on the space of compact sets (endowed with the trace topology):

$$\mathsf{F} = \bigcup_{x_i \in \mathsf{P}} \tau_{x_i} K_i$$

where the K_i are i.i.d. compact sets with law ν.

Remark: ν need not be a probability measure, one can just define $\{(x_i, K_i)\}$ as a Poisson process with intensity $\mu \times \nu$.
Let P a stationary Poisson process with unit intensity ($\lambda = 1$).

Germ grain IV

- Assume $K_i = B(0, R_i)$ with R_i with some law ν on $\mathbb{R}_+$. Then

$$T_{\mathsf{F}}(K) = 1 - \exp(-\mathbf{E}_\nu K^{\oplus R_1})$$

where $K^{\oplus r} = \{y : d(y, K) \leqslant r\}$

- Deduce the fraction volume (and show these identities)

$$\begin{aligned}\kappa_{\mathsf{F}} :&= \mathbf{E}\mathbf{Leb}^d(\mathsf{F} \cap [0,1]^d)\\ &= \mathbf{Leb}^d(A)^{-1}\mathbf{E}\mathbf{Leb}^d(\mathsf{F} \cap A) = \mathbf{P}(x \in \mathsf{F})\end{aligned}$$

for any non-negligible A, or $x \in \mathbb{R}^d$.

- Generalise to non-circular shapes

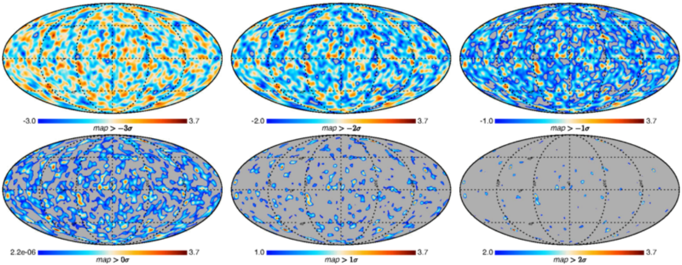
Stationarity

- We study here a set of data in the asymptotics of a large homogeneous random medium:
  Mathematically speaking, assume the random set F satisfies

$$\tau_x \mathsf{F} := \{y + x; y \in \mathsf{F}\} \stackrel{(d)}{=} \mathsf{F}$$

- Quite often, the restriction $\mathsf{P} \cap [0, n]^d$ of a Poisson process is used to approximate n^d i.i.d. uniform variables on $[0, n]^d$.

Random fields I



Fantaye, Hansen, Maino, and Marinucci [2014]

Random fields II

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 44, NUMBER 11 NOVEMBER, 1954

Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter

CHARLES COX AND WALTER MUNK
Scripps Institution of Oceanography, La Jolla, California*

(Received April 28, 1954)

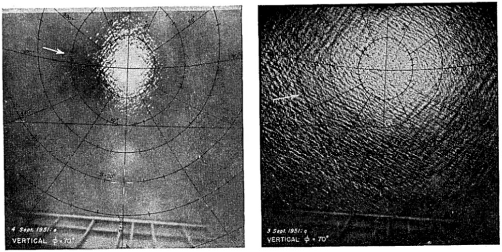
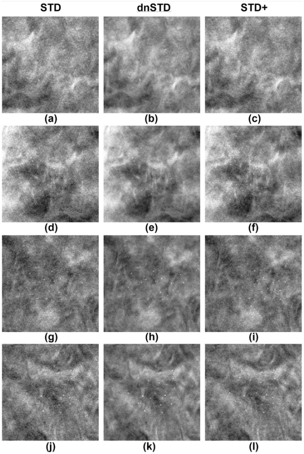


FIG. 1. Glitter patterns photographed by aerial camera pointing vertically downward at solar elevation of $\phi=70°$. The superimposed grids consist of lines of constant slope azimuth α (radial) drawn for every 30°, and of constant tilt β (closed) for every 5°. Grids have been translated and rotated to allow for roll, pitch, and yaw of plane. Shadow of plane can barely be seen along $\alpha=180°$ within white cross. White arrow shows wind direction. *Left:* water surface covered by natural slick, wind 1.8 m sec^{-1}, rms tilt $\sigma=0.0022$. *Right:* clean surface, wind 8.6 m sec^{-1}, $\sigma=0.045$. The vessel *Reverie* is within white circle.

Random fields III



Breast tissue Chan, Helvie, Gao, Hadjilski, 2023

Gaussian fields I

- Given a SDP "covariance" matrix $C = (C_{i,j})_{1 \leqslant i,j \leqslant n}$, there is a unique centred Gaussian vector $\mathsf{X} = (\mathsf{X}_i)_{1 \leqslant i \leqslant n}$ such that

$$\mathbf{E}\mathsf{X}_i\mathsf{X}_j = C_{i,j}.$$

- Given a continuous SDP function $\mathscr{C}(x, y), x \in E$, i.e. such that each finite submatrix $(\mathscr{C}(x_i, x_j)_{1 \leqslant i,j \leqslant n})$ is SDP, there is a unique centred Gaussian process X with

$$\mathbf{E}\mathsf{X}(x)\mathsf{X}(y) = \mathscr{C}(x, y)$$

- We can also prescribe the expectation $m(x)$ by performing the addition $\mathsf{X}(x) \to \mathsf{X}(x) + m(x)$. We only consider centred fields here

Gaussian fields II

- The centred Gaussian field X is stationary if $\mathscr{C}$ is invariant under shifts:

$$\mathscr{C}(x,y) = \tau_z \mathscr{C}(x,y) = \mathscr{C}(x+z, y+z),$$

with $z = -x$ we see that it only depends on $y - x$. In this case we use the abuse of notation $\mathscr{C}(y-x) = \mathscr{C}(x-y)$ instead of $\mathscr{C}(0, y-x)$.

Theorem (Bochner)

SDP yields $\hat{\mathscr{C}} \geqslant 0$ and conversely given any finite measure $\mathscr{S}$, there is a unique SDP $\mathscr{C}$ such that $\mathscr{S} = \hat{\mathscr{C}}$ (in the sense of distributions)

Gaussian fields III

188 15. Gaussian random function

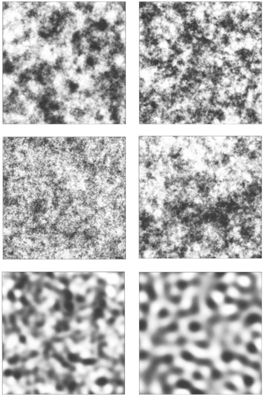


Fig. 15.2. Realizations of gaussian random functions with different covariance functions. From top to bottom and left to right, spherical, exponential, stable, hyperbolic, gaussian and cardinal sine covariances

Lantuéjoul [2002]

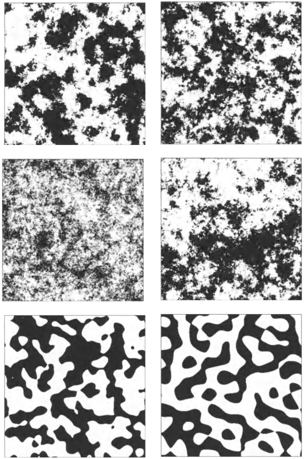
Gaussian excursions I

- For each $\ell \in \mathbb{R}$, we can form random sets with

$$\mathsf{F}_\ell = \{x : \mathsf{X}(x) \geqslant \ell\} \text{ or } \mathsf{L}_\ell = \{x : \mathsf{X}(x) = \ell\}.$$

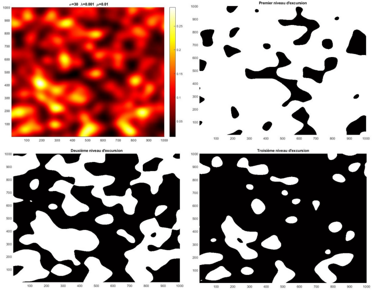
  called (Gaussian) **excursion sets** and **level sets**.
- For $\ell = 0$, F_ℓ is called **nodal domain**, L_ℓ is called **nodal set**.

Gaussian excursions II



Nodal excursions Lantuéjoul [2002]

Gaussian excursions III



Excursions at levels $-1, 0, 1$ Lerbet [2022]

Gaussian excursions IV

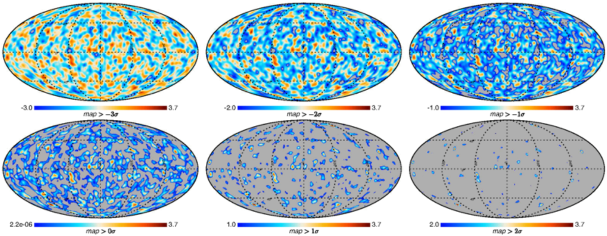


Figure: Fantaye et al. [2014]

Gaussian fields as white noise convolutions

- Let i.i.d. Gaussian variables $W_k \sim \mathcal{N}(0,1); k \in \mathbb{Z}^d$
- Let $\psi : \mathbb{Z}^d \to \mathbb{R}$ summable symmetric and

$$\mathsf{X}(k) = \psi \star W(k) = \sum_i W_{k-i}\psi(i).$$

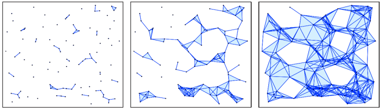
- Covariance $C(k) = \mathbf{E}\mathsf{X}(0)\mathsf{X}(k) = \psi \star \psi(k)$, Fourier $S = \hat{C} = \hat{\psi}^2$
- For any $S \geqslant 0$ on $[0, 2\pi]^d$, one can generate a Gaussian field with spectrum S with $\varphi = \widehat{\sqrt{S}}$
- Idem in the continuous space for a spectral density $s : \mathbb{R}^d \to \mathbb{R}_+$ and a "white noise" W (random measure limit of small i.i.d. Gaussians, e.g. increments of Brownian motion)
- $\mathsf{X}(x) = \int \widetilde{\sqrt{s}}(x-y)dW(y) = W \star \widehat{\sqrt{s}}$ has spectrum $s\mathbf{Leb}^d$

Graphs / Tessellations / partitions I

Random Geometric graphs

Let P a point process,

- a radius $r > 0$: x Connect $x, y \in$ P if $\|x - y\| \leqslant r$.
- **Germ grain model**: Connect two points if their grains K_x, K_y touch, e.g. if $\|x - y\| \leqslant R_x, R_y$ when points are attached to i.i.d. balls $B(x, R_x), x \in$ P.



(a) Sparse regime (b) Thermodynamic regime (c) Dense regime

Geometric graphs for different values for r

Graphs / Tessellations / partitions II

- **Random connection model:** each pair of points x, y is connected independently with probability

$$\varphi(x-y, U_x, U_y), \quad \text{e.g.} \quad \mathbf{1}\{\|x-y\| \leqslant U_x^{-\gamma} U_y^{-\gamma}\}$$

for some i.i.d. marks $U_x, U_y \in [0,1], \gamma \in (0, 1/d)$.

Exercise

- *The number of neighbours $N_{(x,u)}$ of a point at x with a mark u is a Poisson variable with parameter $c_{d,\gamma} u^{-\gamma d}$.*
- *The variance of N_x the number of neighbours of a point at x (with random mark) is finite iff $\gamma < \frac{1}{2d}$.*

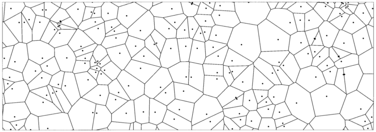
Graphs / Tessellations / partitions III

The Voronoi tessellation

Given a set of points $P \in \mathcal{N}$, for $x \in P$, let the *Voronoi cell* of x

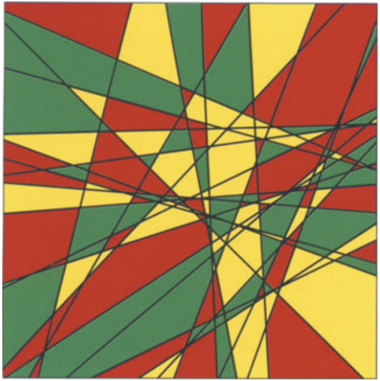
$$\mathrm{Voronoi}(x, P) := \{y \in \mathbb{R}^d : \|y - x\| \leqslant \|y - x'\|, \forall x' \in P\}.$$

The edges form a graph structure/1D lines network; one can also consider the dual *Delaunay* model, where two points x, y are connected if their Voronoi cells touch: $\mathrm{Voronoi}(x, P) \cap \mathrm{Voronoi}(y, P) \neq \emptyset$.



Poisson Voronoi tessellation Lantuéjoul [2002]

Graphs / Tessellations / partitions IV



Poisson line tessellation

Graphs / Tessellations / partitions V

For $\lambda > 0$, let μ the unique measure on the space $\mathscr{L}$ of lines of $\mathbb{R}^2$ that is invariant under the action of Euclideand shifts and such that

$$\mu(\{L : L \cap B(0,1) \neq \emptyset\}) = \lambda.$$

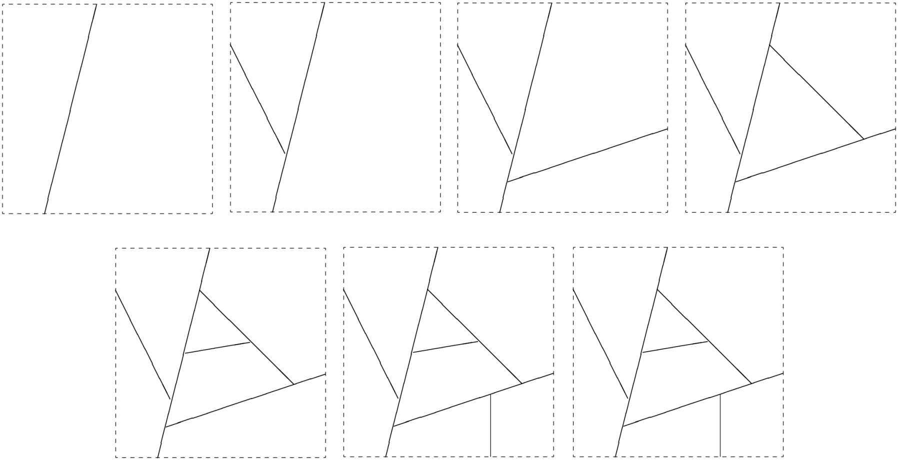
Let $\mathscr{P}$ a Poisson process of intensity μ. The random closed set

$$\mathsf{F} = \cup_{L \in \mathscr{P}} L$$

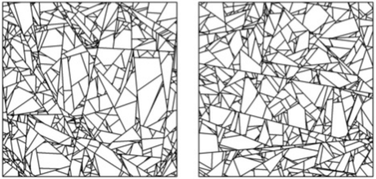
is the *Poisson line intersection* process.

Graphs / Tessellations / partitions VI

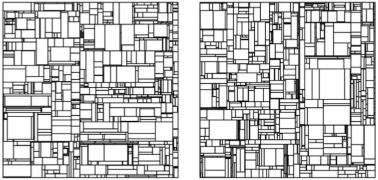
STIT tessellations



Graphs / Tessellations / partitions VII

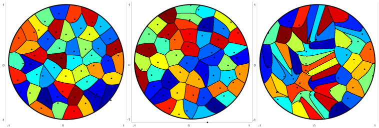


Stit tessellations

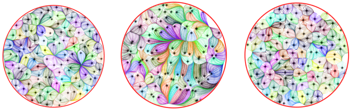


Stit-Mondrian tessellations

Graphs / Tessellations / partitions VIII



Jalowy [2023]



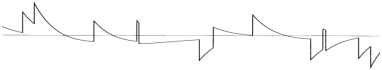
The allocation of the disk to 100 Ginibre points.

The allocation of the disk to 100 Uniform points.

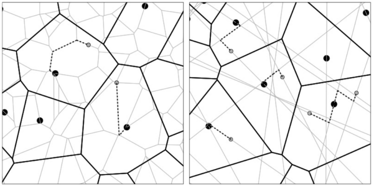
The allocation of the disk to 100 zeros of the GAF.

D. Hawat

Mixed models



Lan02



- Shot noise fields
- Cox processes
- Hawkes processes
- Gibbs processes
- ...

Geometric functionals

Stationarity and LLN

- We study a random set F under the assumption of stationarity

$$\tau_x \mathsf{F} \overset{(d)}{=} \mathsf{F}, x \in \mathbb{R}^d$$

- We wish to estimate parameters of the model based on a spatial average over a window W_n with volume n (ball B_n or cube C_n).

$$\varphi_n(\mathsf{F}) = \frac{1}{n} \int_{W_n} d\mathsf{M}_{\mathsf{F}}(x)$$

- Typical example: Estimate the density of a point process

$$\hat{\lambda} = \frac{1}{n} \# \mathsf{P} \cap W_n$$

- Or estimate the fraction volume of a random set F

$$\hat{\kappa} = \frac{1}{n} \mathbf{Leb}^d(\mathsf{F} \cap W_n).$$

Total volume / mass

Statistics of interest $\varphi_n(F)$:

- Number of points of a point process $\mathsf{F} = \mathsf{P}$:

$$\varphi_n(\mathsf{P}) = \#\mathsf{P} \cap W_n$$

- Volume of a random set F:

$$\varphi_n(\mathsf{F}) = \mathbf{Leb}^d(\mathsf{F} \cap W_n)$$

- Length / $d-1$-dimensional Hausdorff measure of a graph / tessellation / nodal lines set $\mathsf{F} = \mathsf{L}$:

$$\varphi_n(\mathsf{L}) = \mathcal{H}^{d-1}(\mathsf{L} \cap W_n)$$

Standard behaviour:

stationary Poisson process P:

$N_n := \#\mathsf{P} \cap W_n \sim \mathrm{Poiss}(\underbrace{\mathbf{Leb}^d(W_n)}_{=n}) \overset{(d)}{=} \sum_{i=1}^n P_i$ with $P_i \sim \mathrm{Poiss}(1)$ i.i.d.

- Law of Large Numbers (LLN): a.s.

$$\frac{1}{n} N_n = \frac{1}{n} \sum_i P_i \to \mathbf{E} P_1 = 1$$

- Extensive variance:

$$\mathrm{Var}\,(N_n) = \sum_i \mathrm{Var}\,(P_i) = n$$

- CLT:

$$\frac{1}{\sqrt{n}} (\sum_i P_i - n) \to \mathcal{N}(0,1).$$

Main questions:

- **Law of large numbers**:

$$\frac{1}{\mathbf{Leb}^d(W_n)}\varphi_n \to \kappa \in \mathbb{R}?$$

- **Linear variance:** do we have

$$\mathrm{Var}\,(\varphi_n) \asymp n?$$

- **CLT** / Statistic stability / confidence intervals :

$$\tilde{\varphi}_n := \frac{\varphi_n - \mathbf{E}\varphi_n}{\sqrt{\mathrm{Var}\,(\varphi_n(\mathsf{F}))}} \xrightarrow[n\to\infty]{\text{Law}} \mathcal{N}(0,\sigma^2)?$$

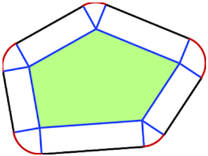
- **Functional CLT** / Uniform tightness for parametric model $\mathsf{F}_\ell, \ell \in \mathbb{R}$:

$$\tilde{\varphi}_n(\mathsf{F}_\ell) \xrightarrow[n\to\infty]{\text{Skohokod space}} G(\ell) \text{ a continuous Gaussian process?}$$

Intrinsic volumes / Curvature measures I

- Let a bounded convex polygon $C \subset \mathbb{R}^d$
- **Steiner formula**: with $C^{\oplus r} = \{y : d(y,C) \leqslant r\}$, $\textbf{Leb}^d(C^{\oplus r})$ is polynomial in r : $\textbf{Leb}^d(C^{\oplus r}) = \ldots$

$$= \underbrace{V^d(C)}_{\textbf{Leb}^d(C^{(d)})} + r \underbrace{V^{d-1}(C)}_{:=\mathcal{H}^{d-1}(\partial C)} + r^2 \underbrace{V^{d-2}(C)}_{=?} + \cdots + \kappa_d r^d \times \underbrace{1}_{\text{Euler}(C)=V^0(C)}$$



- Steiner's formula is still true for any convex bodies!

Intrinsic volumes / Curvature measures II

- $V^k(C)$:="k-th intrinsic volume", can be generalised to any convex set
  - $V^d(C) := \mathbf{Leb}^d(C)$
  - $V^{d-1}(C) :=$ "Perimeter" = "Boundary measure"
  - $V^{d-2}(C) =$ "mean width" $\neq \mathcal{H}^{d-2}(C^{(d-2)}) \ldots$
  - $V^0(C) = 1$
- The intrinsic volumes can be extended to unions of convex sets with:

$$V^k(A \cup B) = V^k(A) + V^k(B) - V^k(A \cap B)$$

  - Ok for volume
  - Less obvious for perimeter
  - It gives a way to extend the intrinsic volume to "polyconvex sets" (finite union of convex bodies). Same interpretation for V^d, V^{d-1},

$$V^0(C) = \mathrm{Euler}(C)$$

Intrinsic volumes / Curvature measures III

Euler characteristic:

- $d = 2$: Number of connected components - Number of bounded holes
- $d = 3$: Number of CC - Number of handles + Number of cavities: bounded CCs of C^c

- Other V^k can be interpreted as "the average k-dimensional cross-section" (Crofton formula)

- Intrinsic volumes can be further extended to smooth sets
- Gives an alternative way to define Euler characteristic
- Computing (probabilistically or not) an intrinsic volume is easy because it is a sum of unordered local contributions
- **Hadwiger's characterisation theorem** Any continuous additive functional φ is a linear combination

$$\varphi(C) = \sum_{k=0}^{d} \lambda_k V^k(C)$$

Intrinsic volumes / Curvature measures IV

- One can see intrinsic volumes as measures over the set C:

$$V^d(C) = \int_C 1\mathbf{Leb}^d(dx)$$

$$V^{d-1}(C) = \int_C \mathbf{1}\{x \in \partial C\}\mathcal{H}^{d-1}(dx)$$

$$V^k(C) = \int_C \underbrace{M^k(C, dx)}_{k-\text{th curvature measure}}$$

- Gauss-Bonnet theorem: for C smooth, e.g. $C = \mathsf{F}_\ell$ excursion set of $\mathcal{C}^3$ Gaussian field,

$$V^0(C) = \int_{\partial C} \underbrace{M^0(C, dx)}_{\text{Gauss curvature at } x} dx$$

Local / Linear functionals I

- Linear functionals in the model are: the number of points, the volume, perimeter, Euler characteristic, etc...
- Often, there is an input P, point process or Gaussian field, and a random set F defined locally from P, and the functional is linear "in F". Examples:
  - P is a Poisson process,

$$\text{Germ grain model: } \mathsf{F} = \bigcup_{x \in \mathsf{P}} B(x,1), \quad \varphi_n(\mathsf{P}) = \mathrm{Vol}(\mathsf{F} \cap W_n)$$

$$\text{NN Graph: } \mathsf{L} = \bigcup_{\{x \text{ nearest neighbour of } y\}} [x,y], \quad \varphi_n(\mathsf{P}) = \mathrm{Euler}(\mathsf{L} \cap W_n)$$

  - Sum of perimeters of Voronoi cells based on a point process
  - X is a Gaussian field, $\varphi_n = \mathrm{Perimeter}(\mathsf{F}_\ell \cap \partial W_n)$

Semi-local functionals

- **Subgraph count**: count the number of triangles / cliques / other abstract finite graph occuring in a random graph; can be seen as "polynomial" in P :

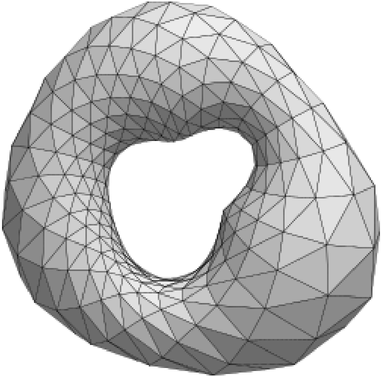
$$\#\text{triangles in geometric graph on } \mathsf{P} = \sum_{x_1,x_2,x_3\in\mathsf{P}} \mathbf{1}\{x_1,x_2,x_3 \text{ pairwise connected}\}$$

- Number of connected components Betti_0 / Number of "holes" Betti_{d-1} $(d \geqslant 2)$
- Other Betti numbers $\text{Betti}_k, 0 \leqslant k \leqslant d-1$. Interestingly

$$\underbrace{\text{Euler}(\mathsf{F})}_{\text{Local}} = \sum_k (-1)^k \underbrace{\text{Betti}_k(\mathsf{F})}_{\text{Non-local}}$$

Random set / Simplicial complex

- To study homology on a random set F, one can either:
  - Study equivalence classes in terms of homotopy (Algebraic Topology)
  - Study random sets based on complexes, such as the Čech complex on a point set, or a triangulation of F

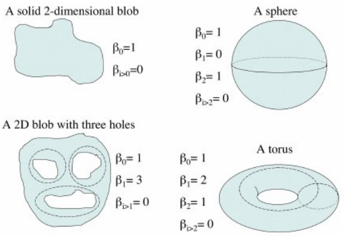


Betti numbers

Betti numbers of $F \subset \mathbb{R}^d$ smooth manifold

- $\mathrm{Betti}_k(F)$: Number of equivalence classes of kD loops for $0 \leqslant k < d$
  - $\mathrm{Betti}_0(F)$: Number of 0D loops ($\approx$ points) $= \#CC(E)$
  - $\mathrm{Betti}_1(F)$: Number of 1D loops ($\approx$ circles)
  - $\mathrm{Betti}_2(F)$: Number of 2D loops ($\approx$ bounded holes)

Euler characteristic: $\mathrm{Euler}(E) := \sum_k (-1)^k \mathrm{Betti}_k(F) = \int_{\partial E} \kappa_F(x)dx$



Walker 2016

Another representation

- A reason why subgraph counts is relevant for topological analysis:
  - Let $N_k(F)$ the number of k-dimensional simplices
  - Then

$$\mathrm{Euler}(F) = \sum_{k=0}^{\infty} (-1)^k N_k(F)$$

Uses of Euler characteristic

- The Euler characteristic can be used to estimate the percolation threshold of large stationary models
- The mean Euler characteristic is useful to estimate rare crossing events for random fields

Highly non-linear: optimisation functionals

- Lenght of minimal spanning tree
- Optimal transport cost
- Length of minimal spanning path (Traveling Salesman Problem)

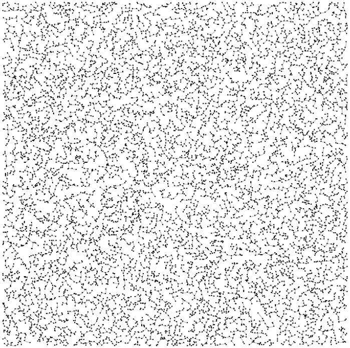


Figure: Minimal spanning tree on Poisson points

Abnormal fluctuations

Random Geometric Graph and Random Connection Model I

- Let $n \to \infty$ some parameter
- Let P a stationary Poisson process
- For any $x, y \in \mathsf{P} \cap W_n$, draw an edge between x and y with probability

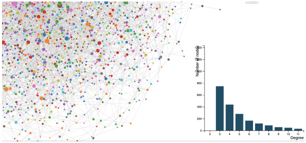
$$\psi\left(\|x-y\|, U_x, U_y\right) \in [0,1]$$

  - ψ interaction profile
  - $U_x, x \in \mathsf{P}_n$: i.i.d. marks with some law ν (marked point process).
    - If $\nu = \delta_0$ (no randomness), we have the **random geometric graph**, and the number of neighbours follows a Poisson distribution with parameter

$$\#\text{neighbours}(x) \overset{(d)}{=} \text{Poiss}\left(\int \psi(\|x-y\|)\mu(dy)\right).$$

Random Geometric Graph and Random Connection Model II

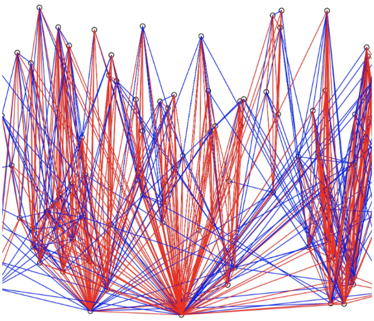
- This does not reflect real social networks where it is known that the number of neighbours experiences a power law



https://helios2.mi.parisdescartes.fr/~rlachiez/social/index.html

- More generally, ν determines the law of the neighbour profile. To reflect social network, we choose ν, ψ such that the number of neighbours has a "heavy" power law tail

Random Geometric Graph and Random Connection Model III



Age dependent 1D random connection model (Credit. P. Grac

Back to RCM model I

- **Age-dependant random connection model:** Number of neighbours of typical point x with mark $U_x \sim \mathscr{U}_{[0,1]}$ on $\mathbb{R}$:

$$\#\text{Neighbours}(x, U_x) \overset{(d)}{=} \text{Poiss}(U_x^{-\gamma})$$

- $\gamma < 1/2$: finite second order moment,
- $\gamma \geqslant 1/2$: infinite second order moment

Back to RCM model II

Call m-clique of the graph a group of m points which are all mutually connected, consider the number of m-cliques in $[-n, n]$, approximable by

$$\varphi_n(\mathsf{P}) = \sum_{x \in \mathsf{P} \cap [-n,n]} \frac{1}{m} \#\{m\text{-cliques} \ni x\}$$

Theorem

Clique counts

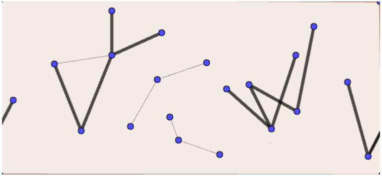
- *[Hirsch and Owada [2024]]: Stable limit for $\gamma > 1/2$ (with the right renormalisation)*
- *[Hirsch, Lr, and Owada [2025]]: CLT for $\gamma < 1/2$*

For more general subgraph counts, the result depends on the shape of the graph: for instance, for subtree counts, it depends on the number ℓ of leaves:

Theorem

ℓ-subtree counts:

- *[Hirsch and Owada [2024]]: Stable limit for* $\gamma > 1/2\ell$
- *[Hirsch, Lr, and Owada [2025]]: CLT for* $\gamma < 1/2\ell$



7 wedge trees in this example, with $\ell = 2$

Perturbed lattices

- Let $U_k, k \in \mathbb{Z}^d$ i.i.d. variables in $\mathbb{R}^d$
- Let $\mathsf{P}_0 = \{k + U_k\}$ the perturbed lattice
- Let $U \sim \mathscr{U}_{[0,1]^d}$. Then $\mathsf{P} = \tau_U \mathsf{P}_0$ is a stationary point process
- Consider

$$\varphi_n(\mathsf{P}) = \#\mathsf{P} \cap C_n.$$

Theorem (Mastrilli '25)

- *If* $d \geqslant 3$, $\tilde{\varphi}_n(\mathsf{P}) \to \mathcal{N}(0,1)$
- *If* $d = 2$ *and* $\mathbf{E}\|U_k\|^\nu < \infty$ *for some* $\nu > 0$, $\tilde{\varphi}_n(F) \to \mathcal{N}(0,1)$
- *If* $d = 1$ *and the characteristic function* φ *of the* U_k *satisfies* $\varphi(x) \sim 1 - c|x|^\alpha$ *for some* $c > 0, \alpha \in [1,2)$, *there is no CLT (the limit is stable for* $\alpha > 1$*)*

Hyperuniformity I

Let $\mathsf{P}_m \subset \mathbb{C}$ be either

- **(Ginibre process)** the set of eigenvalues of $M = (M_{i,j})_{1\leqslant i,j\leqslant m}$ where $M_{i,j} = R_{i,j} + iI_{i,j}$ i.i.d. $\mathcal{N}(0,1)$
- **(GAF zeros)** the zeros of the random "Gaussian Analytic Function" (GAF)

$$F(z) = \sum_{k=1}^{m}(R_k + iI_k)\frac{z^k}{k!}$$

where the R_k, I_k are i.i.d. $\mathcal{N}(0,1)$ variables.

Hyperuniformity II

Theorem

$\mathsf{P}_m \to \mathsf{P}$ *where* P *is stationary and hyperuniform:*

$$\frac{Var(\#\mathsf{P} \cap B_n)}{\mathbf{Leb}^d(B_n)} \to 0$$

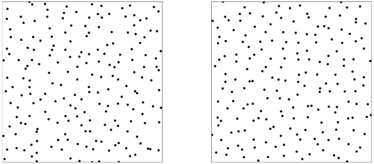


Figure: *Left:* Ginibre, *Right:* GAF zeros

References: Hough et al. [2009], Lr [2025+]

Linear statistics for hyperuniform point processes

Let f smooth with fast decay with $\int f \neq 0$. For an ergodic process with unit intensity $\lambda = 1$:

$$\mathbf{E} \underbrace{\sum_{x \in \mathsf{P}} f(x/R)}_{e.g.\ \#\mathsf{P} \cap B_R} = R^d \int f$$

For a unit intensity Poisson process:

$$\mathrm{Var}\left(\mathsf{P}(f_R)\right) = R^d \int f^2$$

For a hyperuniform process

$$\mathrm{Var}\left(\mathsf{P}(f_R)\right) \sim R^{d-\alpha} c_f$$

where $\alpha > 0$ is the hyperuniformity index (can be arbitrarily large).

Planar Gaussian random waves I

- Let $\mathsf{X}(x), x \in \mathbb{R}^2$ the random Gaussian field with covariance measure the Bessel function of 1t order defined by

$$\mathscr{C}(x) := J_1(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix \cdot e^{i\theta}} d\theta,$$

- it means that $\mathscr{S} := \hat{\mathscr{C}}$ is the uniform distribution on the circle $\partial B(0,1)$, and X does not have a white noise decomposition.
- This system does not have strong mixing properties because the correlation C decreases "slowly" at ∞, namely

$$\mathscr{C}(x) \sim_{\|x\| \to \infty} \sin(\|x\| + c) \|x\|^{-1/2}.$$

Planar Gaussian random waves II

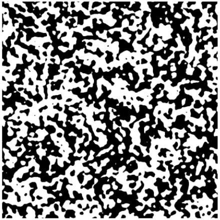
- By Maruyama's theorem, X is mixing. This long range interactions still yields non-standard behaviour.
- We study the excursion and level sets

$$\mathsf{F}_\ell = \{x : \mathsf{X}(x) \geqslant \ell\}, \mathsf{L}_\ell = \{x : \mathsf{X}(x) = \ell\}.$$

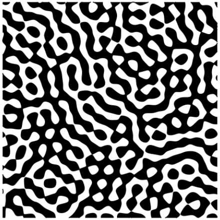
- When the covariance decays fast, e.g. Bargmann-Fock field

$$\mathscr{C}(x-y) = \exp(-\|x-y\|^2),$$

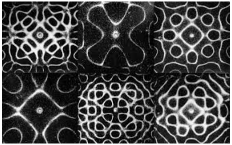
geometric functionals have a standard behaviour.



BF excursion



Random wave excursion



Related to Chladni figures

Theorem

- *The variance of the nodal volume is "linear", i.e. in* n

$$Var\left(\mathbf{Leb}^d(\mathsf{F}_\ell \cap B_n)\right) \asymp n$$

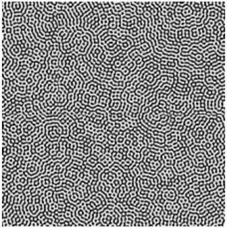
- *The nodal length is hyperfluctuating and depends on* ℓ *(Berry's conjecture Berry [2002]):*

$$Var\left(\mathcal{H}^{d-1}(\mathsf{L}_\ell \cap B_n)\right) \asymp \begin{cases} n^{3/2} \text{ if } \ell \neq 0 \\ n\ln(n) \text{ if } \ell = 0. \end{cases}$$

- *The variance of the number of connected components is in* $n^{3/2}$ *for* $\ell \neq 0$ *Beliaev et al. [2019]*

$$Var(\mathrm{Betti}_0(\mathsf{F}_0 \cap B_n)) \geqslant cn^{3/2}$$

Random waves nodal length



Nodal set of the excursion random wave

The nodal length does not follow a CLT Marinucci et al. [2016]: with

$$\varphi_n = \mathcal{H}^{d-1}(\mathsf{L}_\ell \cap B_n)$$

we have for some $a, b \geqslant 0$

$$\tilde{\varphi}_n \to aX_1^2 + bX_2^2$$

where X_1, X_2 are independent Gaussian variables.

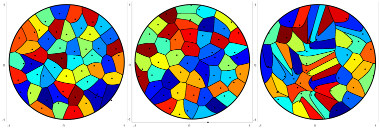
2D optimal transport I

Call allocation of C_n to $\mathsf{P} \cap C_n$ a partition of C_n into disjoint cells $\mathscr{B} = \{B_x, x \in \mathsf{P} \cap C_n\}$, such that each cell has equal volume

$$\mathbf{Leb}^d(B_x) = \frac{n}{\#\mathsf{P} \cap C_n}.$$

Call allocation 2-cost the quantity

$$C(\mathscr{B}) = \inf_{\mathscr{B}} \sum_x \int_{B_x} \|x-y\|^2 dy.$$



Jalowy [2023]

2D optimal transport II

Call optimal cost the minimal value of the cost

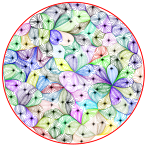
$$\varphi_n^p := \mathsf{W}_2^2(\mathsf{P} \cap C_n) = \inf_{\mathscr{B}} C(\mathscr{B}).$$

This is the optimal transport cost for the Wasserstein distance.

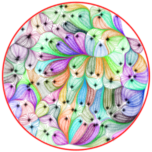
Theorem (Ajtai, Komlos, and Tusnady [1984])

- *If* $d \geqslant 3$, $\varphi_n \asymp n$
- *If* $d = 2$, $\varphi_n \asymp n \ln(n)$
- *If* $d = 1$, $\varphi_n \asymp n^2$

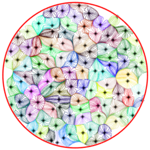
Allocations to Poisson points



The allocation of the disk to 100 Ginibre points.



The allocation of the disk to 100 Uniform points.



The allocation of the disk to 100 zeros of the GAF.

Credit: D. Haway

Conditions for standard statistical behaviour

Formulation

In the greater generality, we can represent φ_n as a stationary random signed measure M, i.e. $\varphi_n(\mathsf{F}) = \mathsf{M}(W_n)$. **Examples**:

- Number of points of P : $\mathsf{M}(A) = \#\mathsf{P} \cap A$
- Volume of a random set F : $\mathsf{M}(A) = \mathbf{Leb}^d(\mathsf{F} \cap A)$
- Number of triangles of a graph on a point process P :

$$\mathsf{M}(A) = \sum_{x \in \mathsf{P} \cap A} \frac{1}{3} \#\{ \text{ triangles } \ni x\}$$

- Number of bounded connected components of a “full” random set F:

$$\mathsf{M}(A) = \int_{\mathsf{F} \cap A} \frac{1}{\mathbf{Leb}^d(\text{connected component containing } x)} dx$$

We will assume Local Square Integrability (LSI): for A bounded

$$\mathbf{E}\mathsf{M}(A)^2 < \infty.$$

Example of a useful result

- Sometimes $\mathsf{M}(dx)$ is called the **score function**, when randomness comes from a point process P we use the notation $\mathsf{M}(dx) = \xi(x, \mathsf{P})$ and

$$\varphi_n = \sum_{x \in \mathsf{P} \cap W_n} \xi(x, \mathsf{P})$$

Theorem (Beck [1987])

For M LSI (wide sense) stationary random measure,

$$\limsup_{n \to \infty} \frac{Var(\mathsf{M}(B_n))}{n^{1-\frac{1}{d}}} > 0.$$

- Not true in all generality for a rectangle instead of a ball
- The $\liminf$ can be zero
- True for a rectangle if M is a "disordered" point process

Ergodicity and LLN I

- The basic assumption is that of **ergodicity**. It roughly means that a quantity can be estimated by doing spatial averages.

Definition

A stationary random measure M is **ergodic** if for all bounded function f,

$$\frac{1}{n}\int_{B_n} f(\tau_x \mathsf{M})dx \to \mathbf{E} f(\mathsf{M})$$

If for instance one wishes to estimate the fraction volume / intensity with $f(P) = \#P \cap B_1$,

$$\lambda = \mathbf{E}\#\mathsf{P} \cap B(0,1),$$

$$\hat{\lambda}_n := \frac{1}{n}\int_{B_n} f(\tau_x \mathsf{P})dx = \frac{1}{n}\#\mathsf{P} \cap B_n + o(1) \to \lambda \text{ a.s.}$$

Ergodicity and LLN II

- **What can go wrong?** Let P be a stationary Poisson process with random intensity Λ . The actual intensity is

$$\lambda = \mathbf{E}\mathsf{P}([0,1]^d) = \mathbf{E}\Lambda$$

but we have a.s.

$$\hat{\lambda} := \frac{1}{n}\#\mathsf{P} \cap W_n \to \Lambda$$

which in general does not give λ.

- Non integrable functional for 2D optimal transport:

$$\sum_{x \in \mathsf{P} \cap W_n} \mathrm{Cost}_1(\mathrm{Cell}(x, \mathsf{P})) \asymp n \ln(n)$$

Ergodicity and LLN III

- Discrete version (see [Meester and Roy, 1996, Sec. 2.2]): for f integrable

$$\frac{1}{n^d}\sum_{k_1=1}^{n}\cdots\sum_{k_d=1}^{n} f(\tau_{(k_1,\ldots,k_d)}\mathsf{F}) \to \mathbf{E}f(\mathsf{F})$$

- This will typically apply to functionals $\varphi_n = \varphi(\mathsf{P}\cap C_n)$ which are approximatively additive, i.e. for all rectangle $\mathcal{R}$ union of two rectangles $\mathcal{R}_1, \mathcal{R}_2$,

$$\varphi(P\cap\mathcal{R}) = \varphi(P\cap\mathcal{R}_1) + \varphi(P\cap\mathcal{R}_2) + \underbrace{o(\mathrm{diam}(\mathcal{R}_i)^d)}_{\text{Boundary terms}}$$

Ergodicity and LLN IV

because then

$$\varphi_n(\mathsf{P} \cap C_n) \approx \sum_{k_1=-n}^{n} \cdots \sum_{k_d=-n}^{n} \varphi_n(\mathsf{P} \cap \tau_k C_1)$$

- See for instance Yukich [1998] for general definitions and results
- Quite often, one can also have more easily a weak LLN using Bienaymé-Chebyshev inequality with

$$\mathrm{Var}\left(\frac{1}{n}\varphi_n(\mathsf{F})\right) \to 0 \Rightarrow \frac{1}{n}\varphi_n(F) - \frac{1}{n}\mathbf{E}\varphi_n(F) \xrightarrow[n\to\infty]{\mathbf{P}} 0.$$

LLN for Betti numbers I

Definition

Čech k-simplex of P: set of k points $x_1, \dots, x_k \in$ P such that

$$\bigcap_{i=1}^{k} B(x_i, r) \neq \emptyset.$$

Theorem (Yogeshwaran et al. [2017])

Let P *an ergodic stationary point process with finite moments on each compact. We have for some* $b_k > 0$

$$\frac{1}{n}\mathrm{Betti}_k(\mathsf{P} \cap C_n) \xrightarrow[n\to\infty]{a.s.} b_k.$$

LLN for Betti numbers II

Remember that $\mathrm{Betti}_k(\mathcal{K})$ is the number of equivalence classes of k-dimensional loops in a set of complexes $\mathcal{K}$.

Lemma (Topological lemma)

For two complexes $\mathcal{K} \subset \mathcal{K}_1$,

$$|\mathrm{Betti}_k(\mathcal{K}_1) - \mathrm{Betti}_k(\mathcal{K})| \leqslant \sum_{j=k}^{k+1} \#\{\text{j-simplex in } \mathcal{K}_1 \setminus K\}$$

Approximation: Decompose C_n in cubes of size t: $C_j^t, j = 0, \ldots, N$ with $Nt^d = n$

$$\mathrm{Betti}_k(\mathsf{P} \cap C_n) \approx \sum_j \mathrm{Betti}_k(\mathsf{P} \cap C_j^t)$$

LLN for Betti numbers III

The difference consists in simplexes touching the boundary, negligible for t large enough.
With the previous inequality, the difference is bounded by

$$\#\{k-\text{simplices touching } \partial C_j^t\}$$
$$+\#\{(k+1)-\text{simplices touching } \partial C_j^t\} \approx N \times 2drt \asymp \frac{n}{t^{d-1}}$$

then apply the ergodic theorem to the translates

$$\text{Betti}_k(\mathsf{P} \cap C_j), j = 0, \dots, N$$

for t arbitrarily large.

Variance upper bound with Poincaré inequality I

- For a functional $\varphi(x_1, \dots, x_n)$, and i.i.d. variables $X'_1, \dots, X'_n$, $X_1, \dots, X_n$, we have the Efron-Stein variance upper bound

$$\mathrm{Var}\left(\varphi(X_1, \dots, X_n)\right) \leqslant \sum_i \Big(\mathbf{E}\varphi(X_1, \dots, X_i, X'_{i+1}, \dots, X'_n) - \mathbf{E}\varphi(X_1, \dots, X'_i, X'_{i+1}, \dots, X'_n)\Big)^2 .$$

- Does the functional vary a lot when one point is resampled?
- For a stationary Poisson process P,

$$\mathrm{Var}\left(\varphi(\mathsf{P} \cap W_n)\right) \leqslant \int_{W_n} \mathbf{E}[\varphi(\mathsf{P} \cup \{x\}) - \varphi(\mathsf{P})]^2 \lambda dx$$

It quite often matches the lower bound (Lr and Peccati [2017], Schulte and Trapp [2024])

Variance upper bound with Poincaré inequality II

- **Example:** Number of connected components of the r-Germ grain model / r-geometric graph:
- Adding a ball can create at most one CC, but it cannot merge more than κ_d CCs for any existing point around:

$$\begin{aligned} |\varphi(\mathsf{P} \cup \{x\}) - \varphi(\mathsf{P})| &\leqslant \kappa_d \\ \mathbf{E}\, |\varphi(\mathsf{P} \cup \{x\}) - \varphi(\mathsf{P})|^q &\leqslant \kappa_d^q \text{ (with } q = 2) \end{aligned}$$

- Works similarly for other Betti numbers
- With the representation $\mathrm{Euler} = \sum_{k=0}^{d-1} (-1)^k \mathrm{Betti}_k$, it gives also an upper bound for the Euler characteristic.

Variance lower bounds

More tricky , often requires conditionnal variance formula - Example of Volume of Voronoi cells :

$$\varphi_n = \sum_{x \in \mathsf{P} \cap W_n} \mathrm{Vol}(\mathrm{Cell}(x, \mathsf{P})) \Rightarrow \mathrm{Var}\,(\varphi_n) \asymp n^{1-1/d}$$

$$\varphi_n = \sum_{x \in \mathsf{P}} \mathrm{Per}(\mathrm{Cell}(x, \mathsf{P}) \cap W_n) \Rightarrow \mathrm{Var}\,(\varphi_n) \asymp n.$$

Stronger assumption: mixing or "asymptotic independence"

- Say that F (stationary) is mixing if for bounded $A, B \subset \mathbb{R}^d$,

$$\mathbf{P}(A \cap \mathsf{F} = \emptyset, \tau_x B \cap \mathsf{F} = \emptyset) \to \mathbf{P}(A \cap \mathsf{F} = \emptyset)\mathbf{P}(B \cap \mathsf{F} = \emptyset)$$

Theorem (Maruyama [1949])

A stationary centred Gaussian field $\mathsf{X}(x), x \in \mathbb{R}^d$ *with covariance function* $\mathscr{C}$ *and spectral measure* $S = \widehat{\mathscr{C}}$ *is*

- **ergodic** *iff* S *has no atoms*
- **mixing** *iff* $\mathscr{C}(x) \to 0$ *as* $\|x\| \to \infty$.

- **Exercise:** The Poisson line intersection process is mixing
- Mixing is a convenient concept of ergodic theory but we sometimes need a stronger assumption.

A CLT with Brillinger mixing in dimension $d \geqslant 2$ I

- **(Refresher on) the method of cumulants:** for a variable X with exponential moments ($\mathbf{E}e^{t|X|} < \infty$ for some $t > 0$), call $\kappa_m(X)$ the m-th order cumulant

$$\kappa_m(X) = \frac{d^m}{dt^m} \ln \mathbf{E}e^{tX}|_{t=0}$$

- It retrieves some familiar quantities

$$\begin{aligned}\kappa_1(X) =& \mathbf{E}X\\ \kappa_2(X) =& \mathrm{Var}\,(X)\\ \kappa_3(X) =& \mathbf{E}(X - \mathbf{E}X)^3\\ & \dots\end{aligned}$$

A CLT with Brillinger mixing in dimension $d \geqslant 2$ II

- The m-th cumulant is a linear combination of moments of order $j \leqslant m$.
- For the Gaussian variable, $\mathbf{E}e^{tX} = e^{t^2}, \kappa_m(X) = 0$ for $m \geqslant 3$. **Marcinkiewicz theorem:** Gaussian laws are the only laws on $\mathbb{R}$ for which only finitely many cumulants do not vanish (i.e. $\ln(\mathbf{E}e^{tX})$ is polynomial).

Theorem

Let X_n centred with unit variance. $X_n \to \mathcal{N}(0,1)$ if and only if for all but finitely many $m, \kappa_m(X_n) \to 0$.

- See also the **4th moment theorems** (in some frameworks, $\kappa_m(X) \to 0$ for $m = 3, 4$ is sufficient)

A CLT with Brillinger mixing in dimension $d \geqslant 2$ III

- With Beck's lemma for a random measure M, we have a CLT for

$$\widetilde{\mathsf{M}(B_n)} := \mathrm{Var}\left(\mathsf{M}(B_n)\right)^{-1/2}\left(\mathsf{M}(B_n - \mathbf{E}\mathsf{M}(B_n)\right)$$

  if for sufficiently high $m \geqslant 3$

$$\kappa_m(\widetilde{\mathsf{M}(B_n)}) = \frac{\kappa_m(\mathsf{M}(B_n))}{\mathrm{Var}\left(\mathsf{M}(B_n)\right)^{m/2}} \to 0$$

  Hence we have a CLT if e.g. for all m, $\kappa_m(\mathsf{M}(B_n)) = O(n)$ (except maybe for strongly hyperuniform processes in $d = 1$).
- **Example:** Poisson process: $\kappa_m(\mathrm{Poiss}(n)) = n$ hence we have a CLT
- The "Poisson-like" condition $\kappa_m(B_n) = O(n)$ is strongly related to "Brillinger mixing" (equivalent under some regularity assumptions), and to fast decay of correlations at every order:
  - Recall $\kappa = \mathbf{P}(x \in \mathsf{F})$.

A CLT with Brillinger mixing in dimension $d \geqslant 2$ IV

- Recall that **plain mixing** is related to

$$\mathbf{P}(0, y \in \mathsf{F}) - \mathbf{P}(0 \in \mathsf{F})\mathbf{P}(y \in \mathsf{F}) = \kappa(\mathbf{P}(y \in \mathsf{F} | 0 \in \mathsf{F}) - \kappa) \to 0.$$

- **Fast decay of correlations** : for all $p, q \geqslant 0$

$$|\mathbf{P}(x_1, \ldots, x_{p+q} \in \mathsf{F}) - \mathbf{P}(x_1, \ldots, x_p \in \mathsf{F})\mathbf{P}(x_{p+1}, \ldots, x_{p+q} \in \mathsf{F})| \leqslant \Phi_{p,q}(\mathrm{dist}(\{x_1, \ldots, x_p\}, \{x_{p+1}, \ldots, x_{p+q}\}))$$

where $\Phi_{p,q}$ has fast decay at ∞ (faster than any power function).

Theorem (Błaszczyszyn, Yogeshwaran, and Yukich [2019])

Let F a random measure with fast decay of correlations, then we have Brillinger mixing and CLT

Some applications

Semi-local (non-linear) statistics over non-Poissonian input:

- Random geometric graphs on Determinantal processes
- Voronoi tessellations on zeros of Gaussian Analytic functions
- Betti numbers (far from percolation thresholds)

Geometric stabilisation and weak dependency

We introduce the concept of **stabilisation**, essential to prove that a geometric functional is statistically stable.

- Let P a stationary point process on $\mathbb{R}^d$ and a functional

$$\varphi_n(\mathsf{P}) = \sum_{x \in \mathsf{P} \cap W_n} \xi(x, \mathsf{P})$$

- Concept: the contribution of each point $x \in \mathsf{P}$ depends on a ball around x with radius R_x :

$$\xi(x, \mathsf{P} \cup A) = \xi(x, \mathsf{P} \cap B(x, R_x) \cup A) \text{ for all finite } A$$

- By stationarity, the law of R_x does not depend on x
- We should have some control over R_x's tail.

Example: NN graphs and subgraph counts

- Let $\varphi_n(\mathsf{P})$ the length of the NN graph on $\mathsf{P} \cap W_n$, we have stabilisation with

$$R_x = \sup\{2\|y\| : y \sim x\}$$
$$\mathbf{P}(R_x > t) = \mathbf{P}(R_0 > t) \leqslant ce^{-c' t^d}$$

- Still works (with different constants) with the k-NN graph, or for weighted edge lengths $\sum_{x,y \in \mathsf{P}} \|x - y\|^\alpha \mathbf{1}\{x \sim y\}$
- Also works for the number of triangles in W_n, for the geometric graph, ...

Quantifying stabilisation

- **Stretched exponential stabilisation** for some $\alpha > 0$

$$\sup_{x,n} \mathbf{P}(R_x > t) \leqslant C \exp(-c\|t\|^{\alpha})$$

- **Polynomial stabilisation**

$$\sup_{x,n} \mathbf{P}(R_x > t) \leqslant C(1+t)^{-d-\alpha}$$

- **Weak stabilisation:** There exists a random variable D^{∞} such that

$$D_0\varphi_n(\mathsf{P}) := \varphi_n(\mathsf{P}) - \varphi_n(\mathsf{P} \cup \{0\}) \xrightarrow[n\to\infty]{\mathbf{P}} D^{\infty}$$

$$\text{Strong implies Weak: } \mathbf{E}\sum_x \mathbf{1}\{0 \in B(x, R_x)\} < \infty$$

hence a.s. for some R_0 suff. large, 0 does not influence $\xi(x, \mathsf{P})$ further than R_0, and conversely.

CLT and 2d order variance with stabilisation on Poisson input

Theorem (Penrose Yukich 2001, Th. 3.1)

Assume we have

- *Weak stabilisation*
- *the 4th moment is uniformly bounded: for finite* $A \subset \mathbb{R}^d$

$$\sup_{n,A} \mathbf{E} \left| \varphi_n(\mathsf{P} \cup \{0\} \cup A) - \varphi_n(\mathsf{P} \cup A) \right|^4 < \infty$$

Then $Var(\varphi_n) = O(n)$ *and*

$$n^{-1/2}(\varphi_n - \mathbf{E}\varphi_n) \to \mathcal{N}(0, \underbrace{\sigma^2}_{>0?})$$

The 4th moment assumption is not always easy to check (see RCMs)

Stabilisation of topological functionals

Consider the germ grain F based on a Poisson process and a radius r

$$\mathsf{F} = \cup_{x \in \mathsf{P}} B(x, r)$$

$$\varphi_n = \mathrm{Betti}_k(\mathsf{F} \cap W_n).$$

Theorem (Yogeshwaran, Subag, and Adler [2017])

$$\mathrm{LLN} : \frac{1}{n}\varphi_n \xrightarrow[n\to\infty]{a.s.} \hat{\beta}_k$$

$$\text{Variance for } k = 0 \; / \; k = d-1 : \frac{1}{n} Var(\varphi_n) \asymp \sigma^2 > 0$$

$$\mathrm{CLT} : \frac{1}{\sqrt{n}}(\varphi_n(\mathsf{P}) - \mathbf{E}\varphi_n(\mathsf{P})) \to \mathcal{N}(0, \sigma^2)$$

CLT proof with weak stabilisation

Let us prove that the number of connected components stabilises.

- We let $\mathscr{C}_r$ the class of disjoint connected components touching $B(0,r)$, hence that can be modified by adding/removing 0
- $\mathscr{C}_r$ is a.s. bounded by κ_d
- If they are disconnected globally they are a fortiori disconnected inside $B(0,3r)$
- Two components C, C' disconnected close to 0 might be connected far from 0, but they reach their "connected status" at some finite radius $R(C,C')$ (they cannot be connected "at ∞")
- Once we are over all such radii, "it stabilises".
- More generally, use **Meyer-Vietoris sequence**

Stabilisation and TDA

- In the context of TDA/persistent homology, one studies the evolution of Betti numbers when r evolves:

$$\mathrm{Betti}_k^{r,s}(\mathsf{P}\cap W_n) := \#\{\text{(equiv. classes of) cycles born before } r \text{ and merged after } s\}$$

- $\mathrm{Betti}_k^{r,r}(\mathsf{P}\cap W_n) = \mathrm{Betti}_k(\mathsf{F}_r\cap W_n)$ germ grain model / geometric graph with parameter r

Lemma (Hiraoka et al, Lm 5.3)

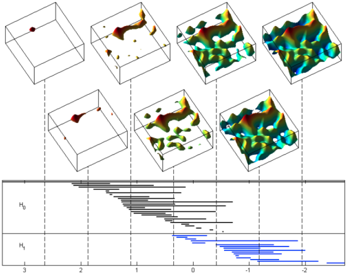
"Deterministic stabilisation:" For any configuration P, and $0\leqslant r\leqslant s$, there is $D_{r,s}(P)$ and $R=R_{r,s}(P)$ such that for $n\geqslant R$

$$\mathrm{Betti}_k^{r,s}((P\cap B_n)\cup\{0\}) - \mathrm{Betti}_k^{r,s}(P\cap B_n) = D_{r,s}(P)$$

Implies CLT. Krebs et al. [2021], Trinh [2019] : binomial case, multidimensional version, inhomogeneous intensity

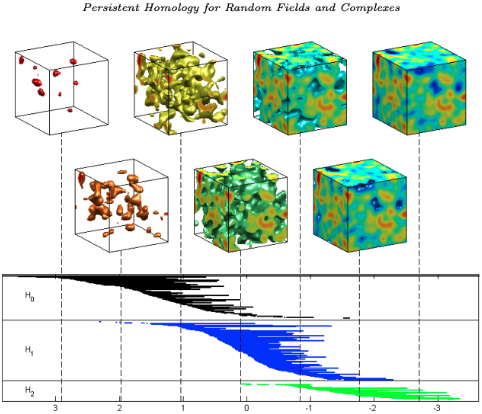
Limit theorems for excursion / random fields functionals I

- TDA is about studying a parametric "functional" model
- Lots of natural data can be modeled through a random field, ℓ is the varying parameter



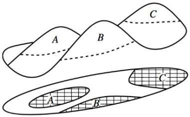
Adler, Bobrowski, Borman, Subag, and Weinberger [2010]

Limit theorems for excursion / random fields functionals II



Adler et al. [2010]

Limit theorems for excursion / random fields functionals III



Adler and Taylor [2007]

- X centred stationary Gaussian field on $\mathbb{R}^d$
- $\ell \geqslant 0, \mathsf{F}_\ell = \{x : \mathsf{X}(x) \geqslant \ell\}$
- $0 \leqslant k \leqslant d-1$, $\varphi_n = \mathrm{Betti}_k(\mathsf{F}_\ell \cap W_n)$
- Further assume white noise convolution for some L^2 symmetric ψ :

$$\mathsf{X}(x) = \int \psi(x-y)dW(y)$$

i.e. $\mathscr{C} = \psi \star \psi$.

Limit theorems for excursion / random fields functionals IV

Theorem (Beliaev, McAuley, and Muirhead [2024])

Assume $\int \psi \neq 0$ *and*

- *Smoothness* $\psi \in \mathscr{C}^5(\mathbb{R}^d)$
- *Fast decay*

$$|\psi(x)|, |\partial_i \psi(x)|, |\partial_{i,j}\psi(x)| \leqslant C(1+\|x\|)^{-9d-1}, x \to \infty$$

Let $\varphi_n = Var(\mathrm{Betti}_0(\mathsf{F}_\ell \cap W_n))$. *Then*

$$Var(\varphi_n) \sim \sigma_\ell n \text{ with } \sigma_\ell > 0$$

$$\frac{\varphi_n - \mathbf{E}\varphi_n}{\sqrt{n}} \to \mathcal{N}(0, \sigma_\ell^2).$$

Theorem (Hirsch and Lr [2025])

Assume still $\int \psi \neq 0$ *and*

- *More smoothness:* ψ *of class* $\mathcal{C}^{2^{13}}$
- *Faster decay* $|\partial_\alpha \psi(x)| \leqslant c(1 + \|x\|)^{-55d}$
- *Either* F_ℓ *or* F_ℓ^c *does not percolate (and* $\ell \neq \ell_c$ *critical level)*

Then there is a Functional CLT for Betti numbers for $\varphi_{n,\ell} = \mathrm{Betti}_k(\mathsf{F}_\ell \cap W_n)$, *weakly in the Skohokod topology*

$$\tilde{\varphi}_{n,\ell} := \frac{\varphi_{n,\ell} - \mathbf{E}}{\sqrt{n}} \to G_k(\ell)$$

where G_k *is a non-zero centred continuous Gaussian field with* $Var(G_k(\ell)) = \sigma_\ell^2 > 0$.

FCLT proof

- **CLT Proof** Based on stabilisation
- **Multivariate CLT** (Cramér-Wold device): for finitely many coefficients λ_i, levels ℓ_i;

$$\sum_i \lambda_i \varphi_{n,\ell_i} \to \mathcal{N}(0,\sigma_\lambda^2) \Rightarrow \ (\tilde{\varphi}_{n,\ell_1},\dots,\tilde{\varphi}_{n,\ell_q}) \to (G_{k,\ell_1},\dots,G_{k,\ell_q}).$$

- **Tightness**: requires uniform bound on increments of the form

$$\sup_n \mathbf{E}|\tilde{\varphi}_{n,\ell} - \tilde{\varphi}_{n,\ell'}|^4 \leqslant c\|\ell - \ell'\|^\beta$$

with $\beta > 1$. In any case, we always have to control that $\tilde{\varphi}_{n,\ell}$ does not vary too much when ℓ varies

Proof I

Also based on stabilisation ideas but on Gaussian input

- Poisson: "discrete" white noise
- Gaussian field: based on Gaussian white noise $W(dx)$

$\Rightarrow$ What is stabilisation for a Gaussian field?

$$\mathsf{X}(x) = \int \psi(x-y)dW(y) \approx \sum_{y \in \varepsilon\mathbb{Z}^d} \psi(x-y)W_y \text{ with } W_y \sim \mathcal{N}(0, \varepsilon^d)$$

Stability under resampling: let $B \subset \mathbb{R}^d$ and W' an independand copy of W and the resampled field at B

$$\mathsf{X}^B(x) = \int_B \psi(x-y)dW'(y) + \int_{B^c} \psi(x-y)dW(u)$$

Proof II

- $\mathsf{X}^B \overset{(d)}{=} \mathsf{X}$ because $W - W1_B + W'1_B \overset{(d)}{=} W$ for all $B \subset \mathbb{R}^d$
- $\mathsf{X} \approx \mathsf{X}_B$ far from B because $W = W'$ outside B : for y at distance more than R from B

$$\mathbf{E}\|\mathsf{X}(y) - \mathsf{X}^B(y)\|^2 \leqslant c \int_{B(0,R)^c} |\psi|^2 \leqslant c(1+R)^{-109d}$$

- The same holds for the fields derivatives far from B :

$$\partial_i \mathsf{X}(x) \approx \partial_i \mathsf{X}^B(x), \partial_{i,j}\mathsf{X}(x) \approx \partial_{i,j}\mathsf{X}^B(x), \ldots$$

Small variations analysis I

- When a "smooth topological structure" evolves "smoothly", the topology does not change when there is no "topological accident"
- Germ grain model: Merge/appearing/disappearing of ball, such accidents are coded into the barcodes
- The number of such accidents in a small interval $[\ell, \ell + h]$ goes to 0 as $h \to 0$
- Accidents of random functions: $X(x)$ is **Morse function** if and only if each critical point has $X''(x) \neq 0$ and there are no two critical points at the same level
- The topology of a level set changes only when a critical point is crossed

Small variations analysis II

Lemma (Topological lemma)

let $X : W \to \mathbb{R}$ Morse and Δ a smooth perturbation, then (after removing boundary components) there is an isotopy between the level sets $\{X \geqslant \ell\}$ and $\{X + \Delta \geqslant \ell\}$ if no critical point "crosses" level ℓ of $X + t\Delta$ for $t \in [0, 1]$

- When one resamples around 0, the difference stabilises weakly when the window converges to ∞ as after a while the critical points close to 0 are not created/deleted. Beliaev et al. [2024]: a.s.

$$\mathrm{Betti}_k(\mathsf{F}_\ell; W_n) - \mathrm{Betti}_k(\mathsf{F}_\ell; W_n^{B_1}) \xrightarrow[n\to\infty]{a.s.} \Delta_k$$

Moments

- By the topological lemma, the Betti number at some level ℓ is bounded by the number of critical points from $-\infty$ until ℓ, globally

$$\mathrm{Betti}_k(\{X \geqslant \ell\}) \leqslant c_{d,k} \#\text{critical points}(X)$$

- For regular (say $\mathcal{C}^3$) Gaussian fields it is easy to prove that there is a finite moment of order 2 for the number of critical points, hence for the Betti numbers.
- For stabilisation techniques, one requires higher moments, typically 4th order.
- Proving even 3d order moments is complicated (even if the field is $\mathscr{C}^\infty$), Beliaev et al. [2024] proved it for their CLT
- Almost simultaneously, Gass and Stecconi [2024] and Ancona and Letendre [2025] proved that if the field is $\mathscr{C}^{k+1}$, the zeros have moments up to order k (hence all moments if the field is $\mathscr{C}^\infty$)

Stabilisation I

We decompose

$$\varphi_n = \sum_{x \in W_n : \nabla \mathsf{X}(x) = 0} \text{Contribution of } x$$
$$= \sum_{x \in W_n : \nabla \mathsf{X}(x) = 0} \mathrm{Betti}_k(C(x, \mathsf{F}_\ell) \cap W_n) \mathbf{1}\{x \text{ is the reference point}\}$$

where $C(x, \mathsf{F}_\ell)$ is the connected component containing x and

$$\mathbf{1}\{x \text{ is the reference point}\} = \mathbf{1}\{x \text{ is the smallest critical point of } C(x, \mathsf{F}_\ell) \cap W_n \text{ in lexicographical order}\}$$

Stabilisation II

- Do we have stabilisation of

$$\mathrm{Betti}_k(C(x,\mathsf{F}_\ell)\cap W_n) \text{ and } \mathbf{1}\{x \text{ is the reference point}\}?$$

under the continuous perturbation

$$\mathsf{X} \to \tilde{\mathsf{X}}$$

where

- either we resample the white noise far from x continuously

$$\tilde{\mathsf{X}} = \mathsf{X}^{B(x,R)^c}$$

- or we slightly modify the level

$$\tilde{\mathsf{X}} = \mathsf{X}(h+\cdot)$$

Stabilisation III

- We consider the perturbation in a continuous manner

$$\tilde{\mathsf{X}}^{(t)} = \mathsf{X} + t(\tilde{\mathsf{X}} - \mathsf{X}), \;\; \tilde{\mathsf{F}}_\ell^{(t)} = \{\mathsf{X}^{(t)} \geqslant \ell\}, t \in [0,1].$$

- Stabilisation depends a lot on the size of the component $C(x, \mathsf{F}_\ell)$!

Theorem (Vanneuville, Muirhead, Ribera, Severo, Beffara, Gayet, ...)

For a standard Gaussian field in the supercritical regime, the typical bounded component's diameter decreases exponentially fast:

- *In dimension $d \geqslant 2$ at level $\ell > \ell_c = 0$,*
- *In dimension $d \geqslant 3$ at level $\ell > \ell_c > 0$ with positive association $\mathscr{C}(x) \geqslant 0$ or if $\mathscr{C}$ decays sufficiently fast*

Stabilisation IV

- We assume therefore that ℓ evolves in (ℓ_c, ∞), so that on a small perturbation, $C(x, \mathsf{F}_\ell^{(t)})$ stays inside $C(x, \mathsf{F}_{\ell'})$ for some $\ell' \in (\ell_c, \ell)$ satisfying

$$\mathbf{P}(\mathrm{diam}(C(x, \mathsf{F}_{\ell'})) > R) \leqslant \exp(-cR^\alpha)$$

- Quantifying the variation of x's contribution hence amounts to quantify if
  - Another critical point y becomes the smallest critical point of $C(x, \mathsf{F}_\ell^{(t)})$ in the LG order
  - Some critical points appear / disappear /merge during the perturbation
  - ...

Stabilisation V

- There is one useful formula when studying Gaussian fields, called **Kac-Rice** formula, that gives the number of zeros of a random field
- To estimate the probability that topological accidents happen, we bound by counting the number of zeros of a specific function:
- **Example 1:** Critical points crossing the level

$$\mathbf{P}(\text{a critical point crosses level } \ell) = \mathbf{P}((\tilde{\mathsf{X}}^{(t)} - \ell, \nabla \mathsf{X}^{(t)}) \text{ vanishes}) \leqslant \mathbf{E}\#\mathrm{Zeros}((t,y) \mapsto (\mathsf{X}^{(t)}(y) - \ell, \nabla \tilde{\mathsf{X}}^{(t)}(y)))$$

- **Example 2:** (Dis)appearing critical points

$$\begin{aligned}&\mathbf{P}(\text{a critical point appears in } B(x,R))\\ &\leqslant \mathbf{P}(\det \mathrm{Hess}\mathsf{F}^{(t)}(y) = 0, y \in B(x,R), t \in [0,1])\\ &\leqslant \mathbf{E}\#\mathrm{Zeros}((t,y) \mapsto \det \mathrm{Hess}\mathsf{F}^{(t)}(y))\end{aligned}$$

Kac-Rice formula

- **Kac-Rice** formula for a deterministic function $\psi : \mathbb{R} \to \mathbb{R}$

$$\#\mathrm{Zeros}(\psi) = \lim_{\varepsilon} \int |\psi'(x)| \frac{\mathbf{1}\{\psi(x) \in [-\varepsilon, \varepsilon]\}}{2\varepsilon} dx.$$

Then take the expectation for a random Gaussian function. Using

$$\frac{\mathbf{P}(\psi(x) \in [-\varepsilon, \varepsilon])}{2\varepsilon} \to \frac{1}{\sqrt{2\pi \mathrm{Var}\,(\psi(x))}}$$

We have

$$\mathbf{E}\#\mathrm{Zeros}(\psi) = \int \mathbf{E}[|\psi'(x)| | \psi(x) = 0] \frac{1}{\sqrt{2\pi \mathrm{Var}\,(\psi(x))}} dx.$$

In higher dimensions

$$\mathbf{E}\#\mathrm{Zeros}(\psi) = \int \mathbf{E}[\|\nabla \psi(x)\| | \psi(x) = 0] \frac{1}{(2\pi)^{d/2} \mathrm{Var}\,(\psi(x))} dx.$$

Alexander duality

- Final remark: If the components of $C(x, \mathsf{F}_\ell)$ are too large, we can work in the connected components of the complement using **Alexander duality**:

$$\mathrm{Betti}_k(\mathsf{F}_\ell \cap W_n) = \mathrm{Betti}_{d-1-k}(W_n \setminus \mathsf{F}_\ell)$$

we use also the self-duality of the Gaussian field: $\mathsf{X} \overset{(d)}{=} -\mathsf{X}$, hence

$$\mathbb{R}^d \setminus \mathsf{F}_\ell \overset{(d)}{=} \mathsf{F}_{-\ell}.$$

All in all, we can work at level ℓ either

- in the subcritical regime (F_ℓ percolates)
- in the supercritical regime ($\mathsf{F}_{-\ell}$ percolates)

Thank you!

References I

R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer, 2007.

R.J. Adler, O. Bobrowski, M.S. Borman, E. Subag, and S. Weinberger. Persistent homology for random fields and complexes. *IMS Coll.*, 6: 124–143, 2010.

M. Ajtai, J. Komlos, and G. Tusnady. On optimal matchings. *Combinatorica*, 4:259–264, 1984.

M. Ancona and T. Letendre. Multijet bundles and application to the finiteness of moments for zeros of gaussian fields. *Analysis & PDEs*, 18 (6):1433–1476, 2025.

J. Beck. Irregularities of distribution. I. *Acta Math.*, 159:1–49, 1987.

D. Beliaev, M. McAuley, and S. Muirhead. Fluctuations of the number of excursion sets of planar Gaussian fields. arXiv:1908.10708, 2019.

References II

D. Beliaev, M. McAuley, and S. Muirhead. A central limit theorem for the number of excursion set a central limit theorem for the number of excursion set components of gaussian fields. *Ann. Prob.*, 52(3): 882–=922, 2024.

M. V. Berry. Statistics of nodal lines and points in chaotic quantum billiards: perimeter corrections, fluctuations, curvature. *Journal of Physics A: Mathematical and General*, 35(13):3025, 2002.

B. Błaszczyszyn, D. Yogeshwaran, and J. Yukich. Limit theory for geometric statistics of point processes having fast decay of correlations. *Ann. Prob.*, 47:835–895, 2019.

Y. Fantaye, F. Hansen, D. Maino, and D. Marinucci. Cosmological applications of the gaussian kinematic formula. *Physical Review D*, 91 (6), 2014.

References III

L. Gass and M. Stecconi. The number of critical points of a Gaussian field: finiteness of moments. *Prob. Th. Rel. Fields*, https://doi.org/10.1007/s00440-024-01273-5, 2024.

A. E. Gelfand, P. Diggle, M. Fuentes, and R. Webb. *Handbook of spatial statistics*. Chapman & Hall, 2010.

C. Hirsch and R. Lr. Functional central limit theorem for topological functionals of gaussian critical points. arXiv:2411.11429, 2025.

C. Hirsch and T. Owada. Limit theorems under heavy-tailed scenario in the age dependent random connection models. *arXiv preprint arXiv:2409.05226*, 2024.

C. Hirsch, R. Lr, and T. Owada. Normal approximation for subgraph counts in age-dependent random connection models. https://arxiv.org/abs/2505.09318, 2025.

References IV

J. Ben Hough, M. Krishnapur, Y. Peres, and B. Viràg. *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. University Lecture Series. Institute of Mathematical Statistics, 2009.

Jonas Jalowy. The Wasserstein distance to the circular law. *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques*, 59(4): 2285–2307, 2023.

Y. Jiao, T. Lau, H. Hatzikirou, M. Meyer-Hermann, J. C. Corbo, and S. Torquato. Avian photoreceptor patterns represent a disordered hyperuniform solution to a multiscale packing problem. *Phys. Rev. E*, 89 (022721), 2014.

J. Krebs, B. Roycraft, and W. Polonik. On approximation theorems for the Euler characteristic with applications to the bootstrap. *Elec. J. Stat.*, 15 (2):4462–4509, 2021.

R. Lr. Hyperuniformity of random stationary measures, transport and rigidity. 2025+.

References V

R. Lr and G. Peccati. New Berry-Esseen bounds for functionals of binomial point processes. *Annals of Applied Probability*, 27:1992–2031, 2017.

C. Lantuéjoul. *Geostatistical simulation: models and algorithms*. Springer, Berlin, 2002.

A. Lerbet. *Études statistiques et géométriques de champs aléatoires shot noise. Similarités et différences avec les modèles Gaussiens.* PhD thesis, Université de Poitiers, https://tel.archives-ouvertes.fr/tel-03774818/, 2022.

D. Marinucci, G. Peccati, M. Rossi, and I. Wigman. NON-UNIVERSALITY OF NODAL LENGTH DISTRIBUTION FOR ARITHMETIC RANDOM WAVES. *Geom. Funct. Anal*, 26:926–960, 2016.

G. Maruyama. The harmonic analysis of stationary stochastic processes. *Mem. Fac. Sci. Kyushu Univ.*, Ser. A4:45–106, 1949.

References VI

R. Meester and R. Roy. *Continuum percolation*. Cambridge University Press, 1996.

M. Schulte and V. Trapp. Lower bounds for variances of poisson functionals. *Electron. J. Probab*, 29:1–43, 2024.

D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. Wiley, Chichester, second edition, 1995.

K. D. Trinh. On central limit theorems in stochastic geometry for add- one cost stabilizing functionals. *Elec. Comm. Prob.*, 24, 2019.

D Yogeshwaran, E Subag, and R. J. Adler. Random geometric complexes in the thermodynamic regime. *Prob. Th. Rel. Fields*, 167:107–142, 2017.

J. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*. Number 1675 in Lecture notes in Mathematics. Springer, Berlin, 1998.