Central limit theorems for point processes with focus on Gibbsian functionals

Christian Hirsch

joint work with M. Otto & A. M. Svane

DSTS spring meeting, Aalborg, May 09, 2023

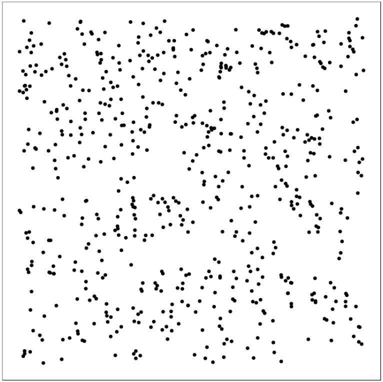
1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 Main results

- Dataset consisting of ***634 neurons***
- ***Minicolumn Hypothesis.*** Arrangement in vertical columns



Christoffersen, Møller & Christensen. *Modelling columnarity of pyramidal cells in the human cortex*

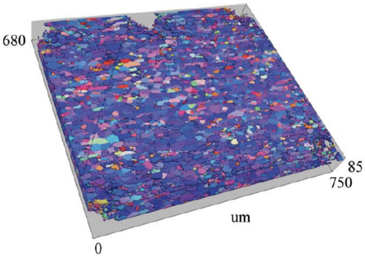
Rafati & et. al. *Detection and spatial characterization of minicolumnarity in the human cortex*

Virtual material design

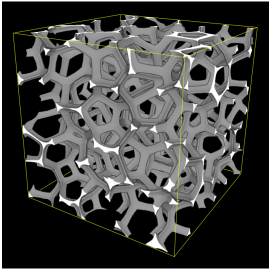
▷ Stochastic-geometry models for heterogeneous materials

⇝ Computation of material characteristics through simulations

Statistical hypothesis tests?



Extra low carbon strip steel



Aluminium alloy foam

1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 Main results

1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 Main results

Null model

$\mathcal{P} = \{X_1, \ldots, X_N\}$ = Poisson point process

- $X_1, X_2, \cdots =$ iid random vectors in sampling window Q
- $N =$ Poisson random variable

Test statistics

- $H := \sum_{X_i \in Q} g(X_i, \mathcal{P})$
- $g(X_i, \mathcal{P}) :=$ Score function

Examples

- $g(X_i, \mathcal{P}) = \#\{X_j : |X_i - X_j| < r\} \rightsquigarrow$ Ripley's K-function
- $g(X_i, \mathcal{P}) =$ Area/Perimeter of Voronoi cell at X_i
- $H =$ Persistent Betti numbers $\rightsquigarrow$ Christophe & Anne Marie

Issue. Distribution of test statistic not known

$\rightsquigarrow$ Asymptotically exact tests in large windows

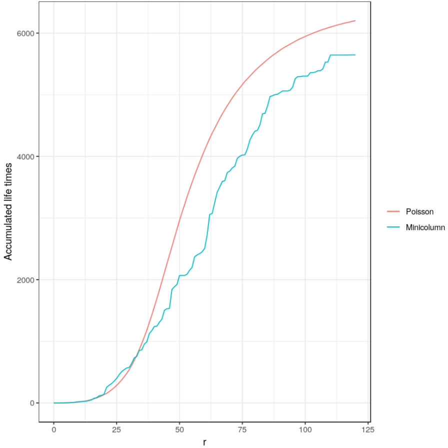
- $H_n := H(\mathcal{P}_n) := H(\mathcal{P} \cap Q_n)$

Asymptotic normality

$$\frac{H_n - \mathbb{E}[H_n]}{\sqrt{|Q_n|}} \Rightarrow \mathcal{N}(0; \sigma^2)$$

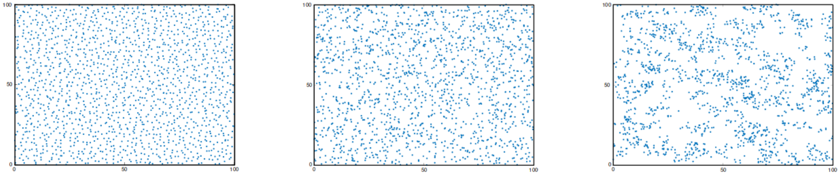
- Penrose & Yukich; (2003). $\rightsquigarrow$ Foundational general result
- Biscio, Chenavier, H. & Svane; (2020). $\rightsquigarrow$ 2D M-bounded persistent Betti numbers

Application to minicolumn dataset

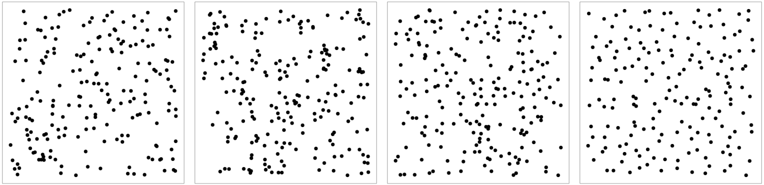


***p*-value = 0.012**

Beyond complete spatial randomness?



Y. Hiraoka, T. Shirai. *Limit theorems for persistence diagrams*



1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 Main results

1 Motivation

2 Complete spatial randomness

3 **Gibbs point processes**

4 Main results

Idea. Interactions via density wrt. Poisson process to reflect

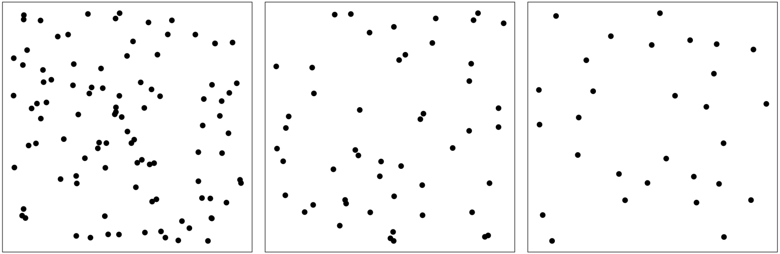
$\rightsquigarrow$ ***Gibbs point processes $\mathcal{X}_n$.***

$$\mathbb{P}(\mathcal{X}_n \in A) = \mathbb{E}[f(\mathcal{P}_n)\mathbb{1}\{\mathcal{P}_n \in A\}]$$

Strauss process.

$$f(\varphi) = \frac{1}{Z}\gamma^{s_R(\varphi)}$$

- $s_R(\varphi) = \#\{\{x, y\} \subseteq \varphi \colon |x - y| < R\} =$ number of R-close pairs



Strauss process with $\gamma = 1$ (left), $\gamma = 0.5$ (middle), $\gamma = 0$ (right) by `spatstat`

Advantages.

- Very flexible interactions
- Simulation methods available $\rightsquigarrow$ (Møller & Waagepetersen, 2004)
  - MCMC methods
  - Perfect simulation
- Parameter estimation via maximum pseudolikelihood
- Interaction quantification via ***Papangelou intensity***

$$\kappa(x, \varphi) := \frac{f(\{x\} \cup \varphi)}{f(\varphi)}$$

- Strauss process: $\kappa(x, \varphi) = \gamma^{t_R(x,\varphi)}$; $t_R(x, \varphi) := \#(\varphi \cap B_R(y))$

Disadvantages.

- Mainly repulsive point patterns
- Simulation may be slow
- No closed-form moments $\rightsquigarrow$ determinantal point processes
- Problems in modelling densely packed hard sphere systems

CLTs in growing domains $\rightsquigarrow$ ***infinite-volume Gibbs processes***

Georgii-Nguyen-Zessin (GNZ) equation

$$\mathbb{E}\Big[\sum_{X_i \in \mathcal{X}} h(X_i, \mathcal{X})\Big] = \lambda \mathbb{E}\Big[\int h(x, \mathcal{X} \cup \{x\})\kappa(x, \mathcal{X})\mathrm{d}x\Big]$$

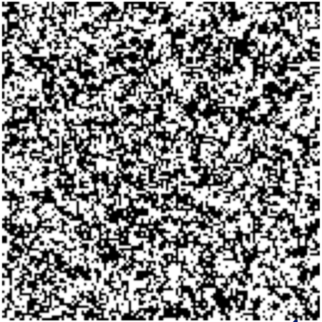
- Close relation to ***statistical physics***

Existence

- Achievable by ***tightness arguments***

Uniqueness

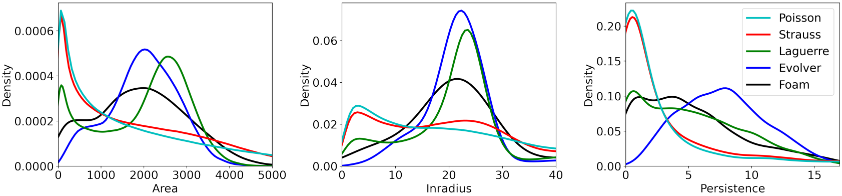
- Highly non-trivial
- Often possible for ***small intensities***
- Non-uniqueness related with ***phase transition***



Ising model

Null models. Poisson and Strauss ($\gamma = 0.01 \rightsquigarrow$ highly repulsive)

Exploratory analysis. KDE for the face areas, face-inradii and total persistences.



z-scores

H_0\Statistic	T_{Area}	T_{I}	T_{Pers}
Poisson	12.8	14.94	35.94
Strauss	11.84	14.36	32.48
Laguerre	6.92	1.85	1.33
Evolver	12.42	4.03	20.47

H., Krebs & Redenbach. *Persistent homology based goodness-of-fit tests for tessellations*

1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 **Main results**

Consider functionals of the form

- $H_n := H(\mathcal{X}_n)$

H is ***weakly stabilizing*** if

- $H((\varphi \cap A) \cup \{y\}) - H(\varphi \cap A) \to \Delta(y, \varphi, \infty) < \infty$ for $A \uparrow \mathbb{R}^d$

Theorem A. CLT for weakly stabilizing functionals

Let H be translation-invariant. Assume that H together with $\mathcal{X}_n$ satisfy conditions on moments and ***weak stabilization***. Then, for some $\sigma \geqslant 0$,

$$|Q_n|^{-1/2}(H_n - \mathbb{E}[H_n]) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Example. Persistent Betti numbers

Consider sums of ***stabilizing score functions:***

- $H_n = \sum_{X_i \in \mathcal{X}_n} g\big(X_i, \mathcal{X}_n\big)$
- $g(x, \varphi) = g(x, \varphi \cap B_{R(x,\varphi)}(x))$
- $\{\varphi \colon R(x, \varphi) \leqslant r\} = \{\varphi \colon R(x, \varphi \cap B_r(x)) \leqslant r\}$

Exponential stabilization

$$\limsup_{r\uparrow\infty} \sup_n \sup_{x \in Q_n} \frac{\log \mathbb{P}(R(x, \mathcal{X}_n \cup \{x\}) > r)}{r^\alpha} < 0$$

Theorem B. Quantitative normal approximation

Let g be translation covariant. Assume that g together with $\mathcal{X}_n$ satisfies conditions on moments, variance lower bound & ***exponential stabilization***. Then,

$$d_{\mathsf{K}}\Big(\frac{H_n - \mathbb{E}[H_n]}{\sqrt{\mathsf{Var}(H_n)}}, \mathcal{N}(0,1)\Big) \leqslant O\Big(\frac{(\log |Q_n|)^a}{\sqrt{|Q_n|}}\Big),$$

where $d_{\mathsf{K}}(X, Y) := \sup_{u \in \mathbb{R}} |\mathbb{P}(X \leqslant u) - \mathbb{P}(Y \leqslant u)| =$ Kolmogorov distance.

Example. Total edge length in Voronoi tessellation

1 Motivation

2 Complete spatial randomness

3 Gibbs point processes

4 Main results

Poisson point processes

Theorem A.

(Penrose & Yukich, 2001)

- Proof based on ***Martingale CLT***

$$H_n - \mathbb{E}[H_n] = \sum_{i \leqslant k_n} \Delta_{i,n} := \sum_{i \leqslant k_n} \mathbb{E}[H_n | \mathcal{F}_{i,n}] - \mathbb{E}[H_n | \mathcal{F}_{i-1,n}],$$

- Relies on ergodic theorem

Theorem B.

(Lachièze-Rey, Schulte & Yukich, 2019)

- Proof based on ***Malliavin-Stein theory***
- No log-corrections

Gibbs point processes

Theorem A. /

Theorem B.

(Schreiber & Yukich, 2013), (Xia & Yukich, 2015)

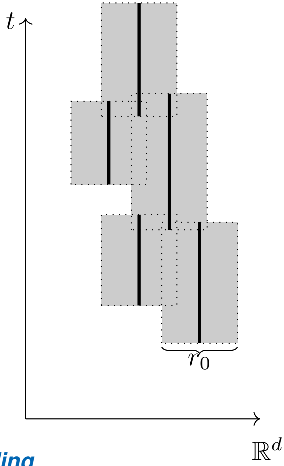
- ▷ ***Graphical construction***
- ▷ Constraint on intensity, e.g., for Strauss $|B_{(r_0+1)}(o)|\tau < 1$

(Chen, Röllin & Xia, 2022)

- ▷ ***Palm coupling***
- ▷ Require fast decay of correlations

(Cong & Xia, 2023+)

- ▷ No Gibbs assumption needed
- ▷ Restricted to Wasserstein distance



~~***Graphical construction***~~ $\rightsquigarrow$ ***Disagreement coupling***

Gibbs processes $\mathcal{X}$

- Finite interaction range $r_0 > 0$
- Papangelou intensity bounded by some $\alpha_0 > 0$
- $\alpha_0 < \alpha_c(r_0)$, the critical intensity for Poisson continuum percolation

Theorem A.

- Weak stabilization
- Growth condition. $|H(\varphi \cup \{y\}) - H(\varphi)| \leqslant \exp\big(c\varphi(B_r(y))\big)$

Theorem B.

- Exponential stabilization
- $\sup_{n\geqslant 1} \sup_{x_1,\dots,x_5 \in Q_n} \mathbb{E}\big[g(x_1, \mathcal{X}_n \cup \{x_1, \dots, x_5\})^5\big] < \infty$
- $\mathsf{Var}(H_n) \in \Omega(|Q_n|)$
- $g(x, \varphi \cup \{y\}) = 0$ if $g(x, \varphi) = 0$ $\rightsquigarrow$ hereditary condition

Summer School on TDA & Spatial Statistics Aalborg, 26/06/23–30/06/23

http://www.dstda.com

- Wojciech Chachólski (KTH)
- Anne Estrade (Université Paris Cité)
- Erika Roldán (MPI Leipzig)
- Rasmus Waagepetersen (Aalborg University)

Literature

[1] C. A. N. Biscio, N. Chenavier, C. Hirsch, and A. M. Svane.
Testing goodness of fit for point processes via topological data analysis.
Electron. J. Stat., 14(1):1024–1074, 2020.

[2] C. A. N. Biscio and A. M. Svane.
A functional central limit theorem for the empirical Ripley's K-function.
Electron. J. Stat., 16(1):3060–3098, 2022.

[3] L. H. Y. Chen, A. Röllin, and A. Xia.
Palm theory, random measures and Stein couplings.
Ann. Appl. Probab., 31(6):2881–2923, 2021.

[4] A. D. Christoffersen, J. Møller, and H. S. Christensen.
Modelling columnarity of pyramidal cells in the human cerebral cortex.
Aust. N. Z. J. Stat., 63(1):33–54, 2021.

[5] T. Cong and A. Xia.
Convergence rate for geometric statistics of point processes with fast decay dependence.
arXiv preprint arXiv:2205.13211, 2022.

[6] C. Hirsch, J. Krebs, and C. Redenbach.
Persistent homology based goodness-of-fit tests for spatial tessellations.
arXiv preprint arXiv:2209.13151, 2022.

[7] J. Møller and R. P. Waagepetersen.
Statistical Inference and Simulation for Spatial Point Processes.
CRC, Boca Raton, 2004.

Normal approximation for functionals on Gibbs processes via disagreement couplings

Moritz Otto (Aarhus)

based on joint work with C. Hirsch & A. M. Svane

DSTS spring meeting
Aalborg
May 09, 2023

Quantitative normal approximation

Let

$$H_n := \sum_{X_i \in \mathcal{X}_n} g(X_i, \mathcal{X}_n).$$

Theorem. Assume that g together with $\mathcal{X}$ satisfy conditions on moments, variance lower bounds and exponential stabilization. Then

$$\mathrm{d_K}\left(\frac{H_n - \mathbb{E}H_n}{\sqrt{\mathrm{Var}(H_n)}}, \mathcal{N}(0,1)\right) \leq O\left(\frac{(\log|Q_n|)^a}{\sqrt{|Q_n|}}\right),$$

where $\mathrm{d_K}(X, Y) := \sup_{u\in\mathbb{R}} |\mathbb{P}(X \leq u) - \mathbb{P}(Y \leq u)|$ Kolmogorov distance.

Proof outline

We use a technique from Chen–Röllin–Xia 2021 that combines

- Palm calculus
- Stein's method

to bound the Kolmogorov distance using a Palm coupling.

To apply this method to Gibbs functionals, we exploit

- Gibbs interpretation of the Palm measure
- Disagreement coupling.

Palm measures

- Ξ point process with intensity measure Λ
- Ξ_x Palm measure of Ξ at $x \in \mathbb{R}^d$ if

$$\mathbb{E} \int f(x, \Xi) \, \Xi(dx) = \mathbb{E} \int f(x, \Xi_x) \, \Lambda(dx), \quad f \geq 0.$$

Stein's method for normal approximation

- $h_u(w)$ smooth approximation of the indicator $1\{w \leq u\}$.
- Let $N \sim \mathcal{N}(0,1)$ and consider the Stein equation

$$f_u'(w) - wf_u(w) = h_u(w) - \mathbb{E}h_u(N), \quad u \in \mathbb{R}.$$

- Let H be such that $\mathbb{E}H = 0$ and $\mathrm{Var}(H) = 1$. Assume that

$$\mathbb{E}[Hf(H)] = \mathbb{E}\int_{-\infty}^{\infty} f'(H+t)\hat{K}(t)\,dt.$$

  for all sufficiently smooth f for some $\hat{K}(t)$.
- Put $K(t) := \mathbb{E}\hat{K}(t)$. For all $u \in \mathbb{R}$ we have

$$\begin{aligned}\mathbb{P}(H \leq u) - \mathbb{P}(N \leq u) &\approx \mathbb{E}h_u(H) - \mathbb{E}h_u(N)\\ &= \mathbb{E}\int_{-\infty}^{\infty} f_u'(H)K(t)\,dt - \mathbb{E}\int_{-\infty}^{\infty} f_u'(H+t)\hat{K}(t)\,dt.\end{aligned}$$

How to choose $\hat{K}(t)$?

- $\Xi := \sum_{X_i \in \mathcal{X}_n} g(X_i, \mathcal{X}_n)\, \delta_{X_i}$ with intensity measure Λ
- Palm versions Ξ_x, $x \in \mathbb{R}^d$, defined on the same prob. space
- Let $\lambda := \mathbb{E}|\Xi|$, $\sigma^2 := \mathrm{Var}(|\Xi|)$ and set

$$H := \frac{|\Xi| - \lambda}{\sigma^2}, \qquad H_x := \frac{|\Xi_x| - \lambda}{\sigma^2}, \qquad \Delta_x := H_x - H.$$

Then

$$\mathbb{E}[Hf(H)] = \mathbb{E} \int \frac{f(H_x) - f(H)}{\sigma^2}\, \Lambda(dx) = \mathbb{E} \iint_0^{\Delta_x} \frac{f'(H+t)}{\sigma^2}\, dt\, \Lambda(dx)$$

$$= \mathbb{E} \iint_{-\infty}^{\infty} \frac{f'(H+t)}{\sigma^2} \big(1\{\Delta_x > t > 0\} - 1\{\Delta_x < t \le 0\}\big)\, dt\, \Lambda(dx).$$

Therefore let

$$\hat{K}(t) := \frac{1}{\sigma^2} \int \big(1\{\Delta_x > t > 0\} - 1\{\Delta_x < t \le 0\}\big)\, \Lambda(dx).$$

How to couple Ξ and Ξ_x, $x \in \mathbb{R}^d$?

- Let $\mathcal{X}_x^{\Xi}$ Palm process of $\mathcal{X}$ at x wrt Ξ
- Then $\Xi_x := \sum_{X_i \in \mathcal{X}_x^{\Xi}} g(X_i, \mathcal{X}_x^{\Xi})\, \delta_{X_i}$ is a Palm version of Ξ

We use the following Gibbs interpretation of $\mathcal{X}_x^{\Xi}$:

Lemma. Let $\mathcal{X}$ be a Gibbs process with PI κ. Then the reduced process $\mathcal{X}_x^{\Xi} \setminus \{x\}$ is a Gibbs process with PI κ^x given by

$$\kappa^x(y, \omega) := \kappa(y, \omega \cup \{x\}) \frac{g(x, \omega \cup \{x, y\})}{g(x, \omega \cup \{x\})}$$

where $0/0 := 0$.

Poisson embedding of finite Gibbs processes

- total ordering $<$ on $\mathbb{R}^d$
- For $B \subset \mathbb{R}^d$ with $|B| < \infty$ define partition function

$$Z_B(\psi) := \int e^{-H(\mu,\psi)}\, \Pi_B(d\mu) \in (0, +\infty], \quad \psi \in \mathsf{N}.$$

- Define $p \colon \mathbb{R}^d \times \mathsf{N} \to [0,1]$ by

$$p(x,\psi) := \kappa(x, \psi_{(-\infty,x)}) \frac{Z_{(x,\infty)}(\psi_{(-\infty,x)} \cup \{x\})}{Z_{(x,\infty)}(\psi_{(-\infty,x)})}$$

where $\infty/\infty := 0$.

- Φ Poisson process on $\mathbb{R}^d \times \mathbb{R}_+$ with intensity measure $\lambda_d \otimes \lambda_1$

Let

$$x_1 := \min\{x \in \mathbb{X} : \exists t \geq 0 : (x,t) \in \Phi \text{ and } t \leq p(x,0)\},$$
$$x_{n+1} := \min\{x > x_n : \exists t \geq 0 : (x,t) \in \Phi,\ t \leq p(x,\{x_1,\dots,x_n\})\}.$$

Let $\tau := \sup\{n \geq 1 : x_n \in \mathbb{R}^d\}$ and define for $\tau < \infty$,

$$T(\Phi) := \{x_1,\dots,x_\tau\}.$$

Theorem. $T(\Phi)$ is Gibbs process with PI κ.

Disagreement coupling

- Iterative Poisson embedding yields

  Theorem. We find Gibbs processes $\mathcal{X}$, $\mathcal{X}_x^{\Xi} \setminus \{x\}$, $x \in \mathbb{R}^d$, such that for every Borel set $P \subset Q_n$ with $\mathrm{dist}(P, \{x\} \cup Q_n^c) > r$,

  $$\mathbb{P}(\mathcal{X} \cap P \neq (\mathcal{X}_x^{\Xi} \setminus \{x\}) \cap P) \leq c_1 e^{-c_2 r}, \quad r \geq r_0.$$

- This allows us to control

  $$\hat{K}(t) := \frac{1}{\sigma^2} \int \big(1\{\Delta_x > t > 0\} - 1\{\Delta_x < t \leq 0\}\big)\, \Lambda(dx).$$

  and, therefore, to bound

  $$d_K(H, \mathcal{N}(0,1)).$$