Estimating the hyperuniformity exponent of point processes

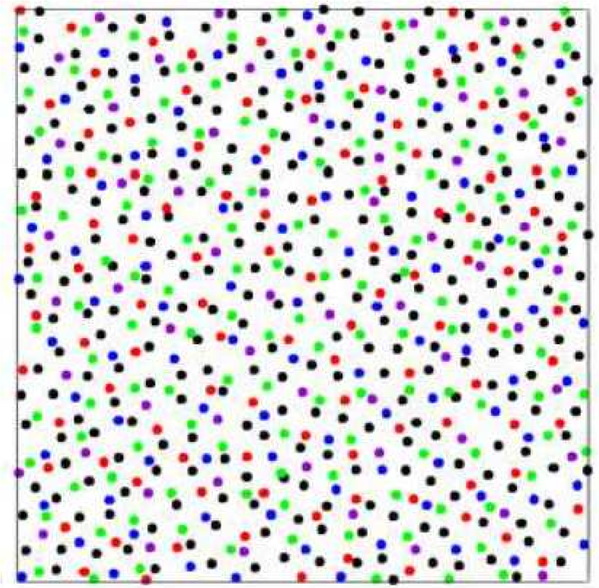
Gabriel Mastrilli
ENSAI Rennes

Bartek Błaszczyszyn
Inria/ENS Paris

Frederic Lavancier
ENSAI Rennes

Anzère, École d'été CUSO, September 1-4, 2024

A striking feature of nature?

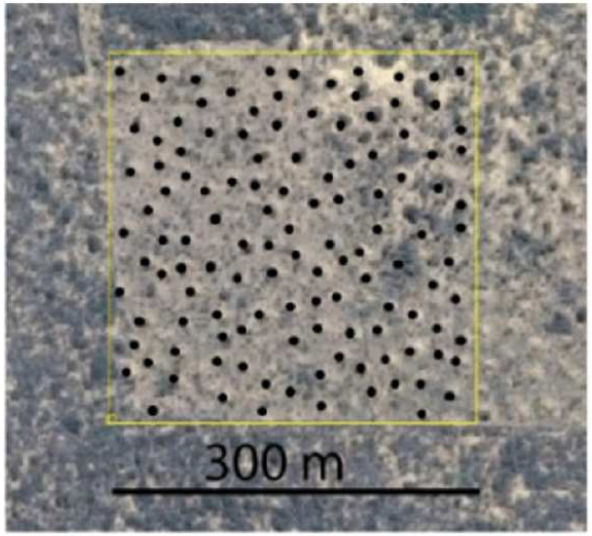


(Jiao et al. 2014)

Avian photoreceptors

(Huang et al. 2021)

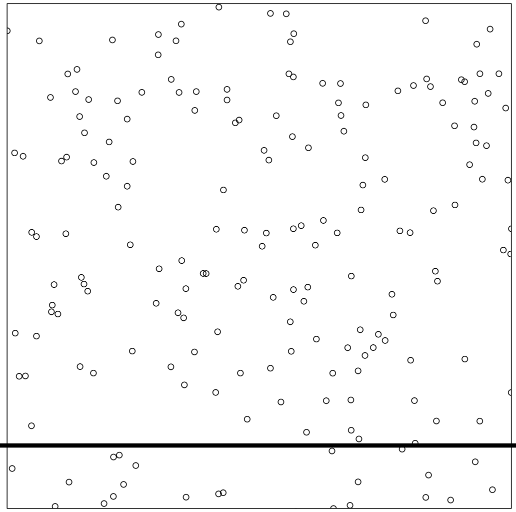
Swimming algae



...

Termite mounds

Patterns more "regular" than complete independence (Poisson model)



Examples in physics

- Cristals (Torquato and Stillinger 2003),
- Plasmas Plasmas (Jancovici 1981),
- Gas (Torquato, Scardicchio, and Zachary 2008),
- Fuilds (Lei and Ni 2019),
- Ices (Martelli, Torquato, Giovambattista, and Car 2017),
- Engineering/materials (Gorsky et al. 2019)
- ...

MODELS
All crystals [27], many quasicrystals [32, 33], stealthy and other hyperuniform disordered ground states [62, 63, 65, 68, 143], perturbed lattices [134, 137–139, 145], g_2-invariant disordered point processes [27], one-component plasmas [35, 146], hard-sphere plasmas [147, 148], random organization models [56], perfect glasses [68], and Weyl-Heisenberg ensembles [136].
Some quasicrystals [33], classical disordered ground states [68, 143], zeros of the Riemann zeta function [34, 71], eigenvalues of random matrices [14], fermionic point processes [34], superfluid helium [61, 144], maximally random jammed packings [36, 38, 39, 41, 43], perturbed lattices [137]. density fluctuations in early Universe [17, 18, 145], and perfect glasses [68].
Classical disordered ground states [135], random organization models [52, 54], perfect glasses [68], and perturbed lattices [139].

Problem formulation

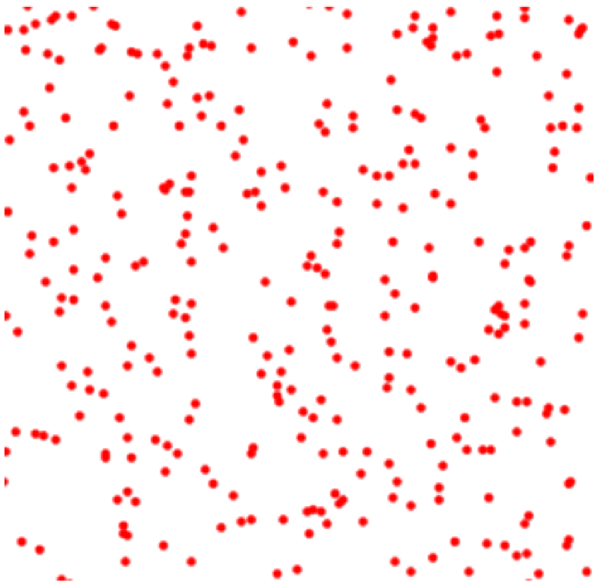
Hyperuniform point processes

- Point process Φ — random, locally finite configuration of points in $\mathbb{R}^d$. Considered as an atomic measure. Assume stationary (translation invariant distribution) and ergodicity.
- Φ is said hyperuniform if

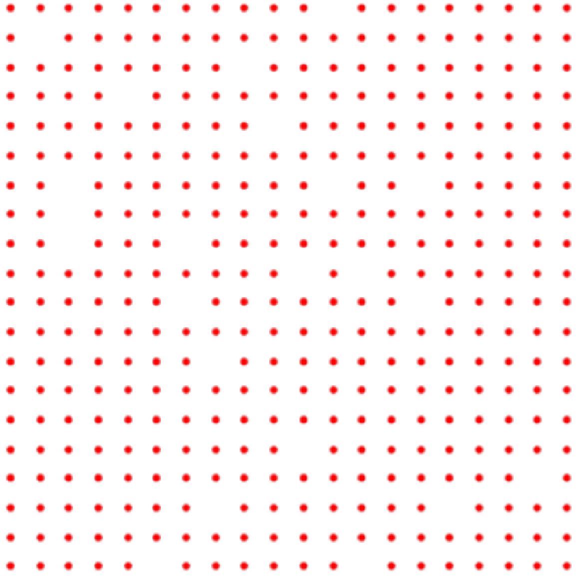
$$\mathrm{Var}[\Phi(B_0(R))] \underset{r\to\infty}{=} o(R^d),$$

where $B_0(R)$ is a ball of radius R in $\mathbb{R}^d$.
- Remember, for Poisson point process Φ (complete independent configuration of points) $\mathrm{Var}[\Phi(B_0(R))] \sim R^d$.
- Hyperuniformity $\equiv$ sub-Poissonian growth in number variance.

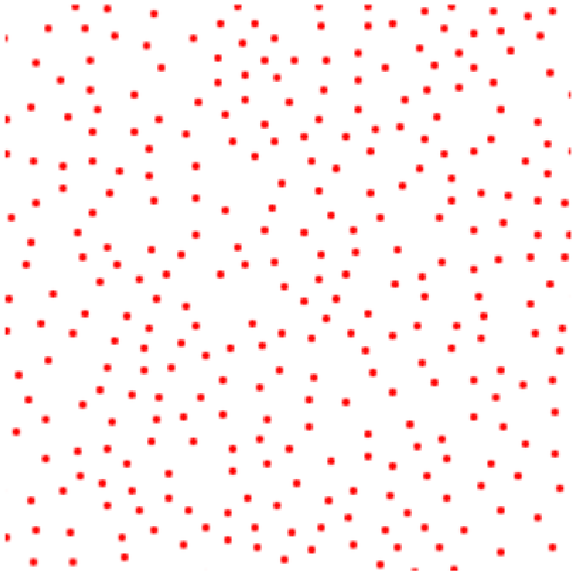
Can you recognize hyperuniformity?



(a) Perturbed Ginibre: hyperuniform.



(b) Thinned URL: not hyperuniform.



(c) Matérn-III (RSA): not hyperuniform.

Hyperuniformity Cases

- Asymptotic behavior for different hyperuniformity exponents:

$$\mathrm{Var}[\Phi(B_0(R))] = \begin{cases} O(R^{d-\alpha}) & \text{for } 0 < \alpha < 1, \text{ (weak hyperuniformity)} \\ O(R^{d-1}\log(R)) & \\ O(R^{d-1}) & \text{(strong hyperuniformity),} \end{cases}$$

where α is called the degree (or strenght) of the hyperuniformity.

- Are there any point processes exhibiting degree $\alpha > 1$?
- No, when counting the points! We need finer tools to capture large-scale fluctuations.
- The reason lies in the indicator function $\mathbf{1}(x \in B_0(R))$ used in

$$\mathrm{Var}[\Phi(B_0(R))] = \mathrm{Var}\Big[\sum_{x\in\Phi} \mathbf{1}(x \in B_0(R))\Big] = \mathrm{Var}\Big[\sum_{x\in\Phi} \mathbf{1}\left(\frac{x}{R} \in B_0(1)\right)\Big]$$

which introduces an unavoidable boundary effect of the order of the "surface volume", of all orders R^{d-1}.

- By using sufficiently smooth functions $f(x)$ instead of $\mathbf{1}(x \in B_0(R))$, we obtain the variance rate

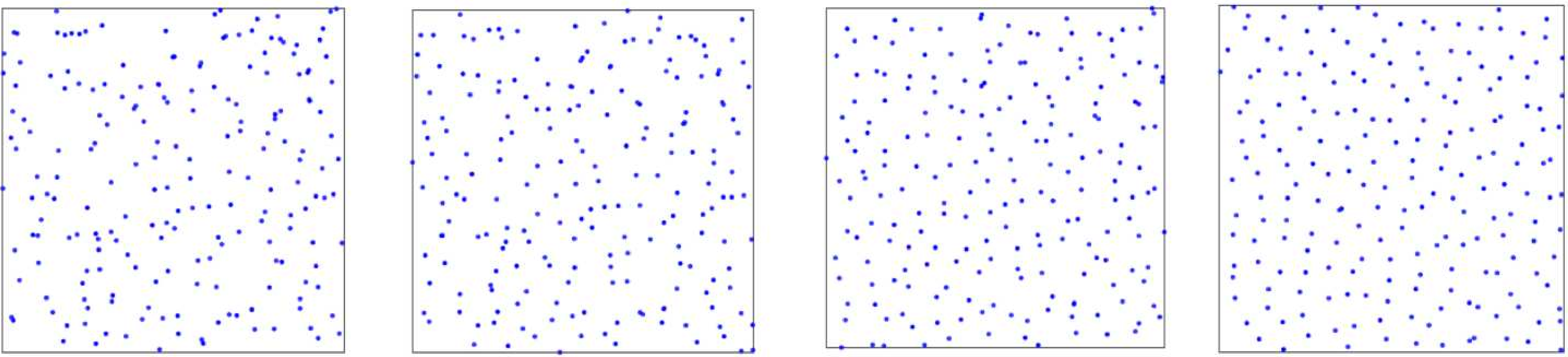
$$\mathrm{Var}\left[\sum_{x \in \Phi} f\left(\frac{x}{R}\right)\right] = O(R^{d-\alpha})$$

for hyperuniform point processes of degree $\alpha \geq 0$.

Examples: perturbed lattices

$$\Phi_{\alpha} = \{y + U + U_y + V_y | y \in \mathbb{Z}^2\}$$

where U, $(U_y)_{y\in\mathbb{Z}^2}$ are i.i.d. uniform on $[-1/2, 1/2]^2$, and $(V_y)_{y\in\mathbb{Z}^2}$ are i.i.d. with characteristic function φ s.t. $1 - |\varphi(k)|^2 \sim_0 t|k|^{\alpha}$.
(for $V_y \equiv 1$ — cloaked lattice (Klatt, Kim, and Torquato 2020)).



(a) $\alpha = 0.5$ (b) $\alpha = 1$ (c) $\alpha = 2$ (d) $\alpha = 4$

Figure: Different degrees α of hyperuniformity (Torquato 2018).

Hyperuniformity in frequency domain

- Bartlett spectrum (structure factor) $\boldsymbol{S}$ of point process $\boldsymbol{\Phi}$ is a complex-valued function on $\mathbb{R}^{\boldsymbol{d}}$

$$\boldsymbol{S(k) := 1 + \lambda \mathcal{F}[g-1](k),}$$

  where

  - $\boldsymbol{\lambda} := \mathbb{E}[\boldsymbol{\Phi}([0,1]^{d})]$ intensity of $\boldsymbol{\Phi}$,
  - $\boldsymbol{\mathcal{F}}$ denotes the Fourier transform on $\mathbb{R}^{\boldsymbol{d}}$,
  - $\boldsymbol{g}$ is pair-correlation function of $\boldsymbol{\Phi}$ (assumed $\boldsymbol{g - 1 \in L^1}(\mathbb{R}^{\boldsymbol{d}})$), defined via second correlation function $\boldsymbol{\rho^{(2)}(dx, dy) = \mathbb{E}[\Phi(dx)\Phi(dy)] = \lambda^2 g(x-y)dxdy}$, $\boldsymbol{x \neq y}$.
  - Equivalently, $\boldsymbol{g}$ represents (if it exists) the density of the mean measure under Palm probability

$$\mathbb{E}^{\boldsymbol{0}}[\boldsymbol{\Phi(B)}] = \int_{\boldsymbol{B}} \boldsymbol{g(x)\, dx.}$$

Hyperuniformity in frequency domain

- Fourier-Campbell formula: For all $f_1, f_2 \in L^2(\mathbb{R}^d)$:

$$\mathbf{Cov}\left[\sum_{x\in\Phi} f_1(x), \sum_{x\in\Phi} f_2(x)\right] = \lambda \int_{\mathbb{R}^d} \mathcal{F}[f_1](k)\overline{\mathcal{F}[f_2]}(k)S(k)dk.$$

- In particular, for all $f \in L^2(\mathbb{R}^d)$:

$$\mathbf{Var}\left[\sum_{x\in\Phi} f(x)\right] = \lambda \int_{\mathbb{R}^d} |\mathcal{F}[f](k)|^2 S(k)dk.$$

- Consequently,

$$\mathbf{Var}\left[\sum_{x\in\Phi} f\left(\frac{x}{R}\right)\right] = R^d \times \lambda \int_{\mathbb{R}^d} |\mathcal{F}[f](k)|^2 S(k/R)dk.$$

Hyperuniformity in frequency domain

$$\mathrm{Var}\left[\sum_{x\in\Phi} f\left(\frac{x}{R}\right)\right] = R^d \times \lambda \int_{\mathbb{R}^d} |\mathcal{F}[f](k)|^2 S(k/R) dk.$$

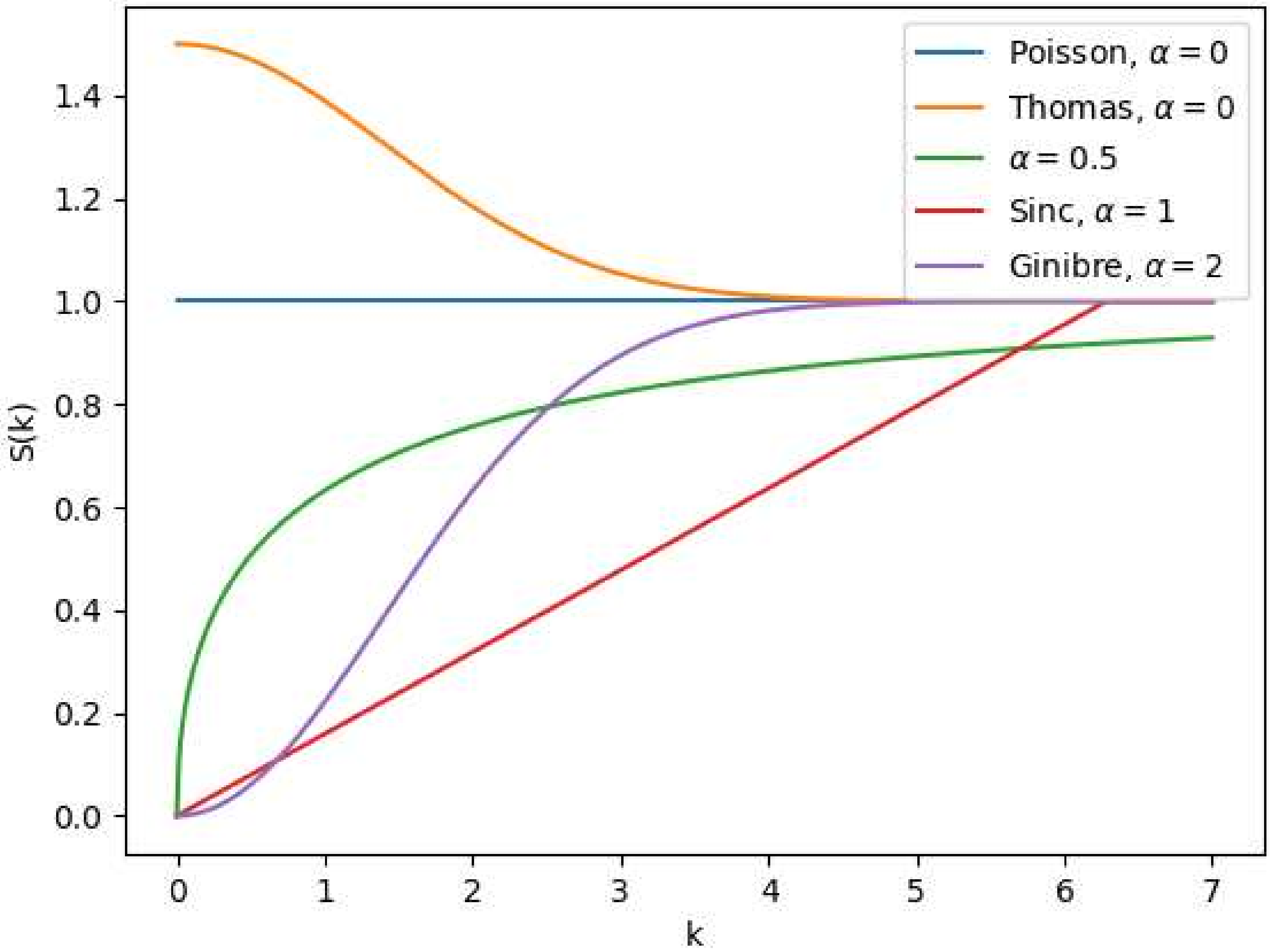
- If $S(0) > 0$ then the RHS is $\sim R^d$, hence Φ is not hyperuniform.
- If $S(0) = 0$ then RHS is $\ll R^d$, hence Φ is hyperuniform (low frequencies of point process disappear).
- Assume:

$$S(k) \underset{|k|\to 0}{\sim} c|k|^\alpha,$$

where $c > 0$ and $\alpha \geq 0$ are constants.

- If moreover f is sufficiently smooth then the RHS is $\sim R^{d-\alpha}$.

Structure function for theoretical point process models



Estimation of α

(on one realization)

Estimation of the degree α of hyperuniformity?

- State-of-the-art:

  1. Estimation of S with $\widehat{S}_R$. Example:

$$\widehat{S}_R(k) = \frac{1}{\#\{\Phi \cap [-R,R]^d\}} \left| \sum_{x \in \Phi \cap [-R,R]^d} e^{-ik.x} \right|^2 .$$

For large window R : $\widehat{S}_R(k) \simeq S(k)$.

  2. Estimation of the behavior of $\widehat{S}_R$ at 0.

For small frequencies k_R : $\widehat{S}_R(k_R) \simeq c|k_R|^\alpha$.

- Idea: combine the two asymptotic regimes...

- PROPOSITION: Assume: $S(k) \underset{|k|\to 0}{\sim} c|k|^{\alpha}$ ($\alpha \geq 0$, $c > 0$). Let f be a Schwartz function and $j \in (0,1)$ then

$$\mathbf{Var}\left[\sum_{x\in\Phi\cap[-R,R]^d} f(x/R^j)\right] \underset{R\to\infty}{\sim} R^{j(d-\alpha)}\lambda \int_{\mathbb{R}^d} |\mathcal{F}[f](k)|^2 c|k|^{\alpha} dk.$$

- If $\int f = 0$, one can expect:

$$\left(\sum_{x\in\Phi\cap[-R,R]^d} f(x/R^j)\right)^2 \simeq R^{j(d-\alpha)} \text{ cst.}$$

Multi-scale linear regression to estimate α

- If $\int f = 0$, one can expect:

$$\log\left\{\left(\sum_{x\in\Phi\cap[-R,R]^d} f(x/R^j)\right)^2\right\} \simeq \log(R)(d-\alpha)j + \text{cst.}$$

- One considers not only several "scales" j to reduce the variance of the estimator but also several "tapers"...

Multi-scale, multi-tapers estimator

- For several scales $j \in J$, $0 < j < 1$
  and several smooth (Schwartz) function (tapers) f_i, $i \in I$ with $\int f_i = 0$,

Least-square estimator of α:

$$\widehat{\alpha} = d - \sum_{j \in J} \frac{\hat{w}_j}{\log(R)} \log\left(\sum_{i \in I}\left(\sum_{x \in \Phi \cap [-R,R]^d} f_i(x/R^j)\right)^2\right),$$

with weights:

$$\forall j \in J,\ \hat{w}_j = \frac{|J| j - \sum_{j' \in J} j'}{|J|\left(\sum_{j' \in J} j'^2\right) - \left(\sum_{j' \in J} j'\right)^2}.$$

Two properties: $\sum_{j \in J} \hat{w}_j = 0$ and $\sum_{j \in J} j \hat{w}_j = 1$.

Consistency

- Observe:

$$\widehat{\alpha}(I,J,R)-\alpha=\sum_{j\in J}\frac{\hat{w}_j}{\log(R)}\log\left(\sum_{i\in I}\left(R^{\frac{\alpha-d}{2}j}\sum_{x\in\Phi\cap[-R,R]^d}f_i(x/R^j)\right)^2\right).$$

- PROPOSITION: Assume:
  - $S(k)\sim c|k|^\alpha$ as $|k|\to 0$, where $\alpha\geq 0$ and $c>0$.
  - for each $j\in J$, there exists $i_j\in I$ such that:

$$R^{\frac{\alpha-d}{2}j}\sum_{x\in\Phi\cap[-R,R]^d}f_{i_j}(x/R^j)\xrightarrow[R\to\infty]{Law}X_j,$$

  - $\mathbb{P}[X_j=0]=0$.

Then $\widehat{\alpha}(I,J,R)\to\alpha$ in probability as $R\to\infty$.

□ THEOREM:(Multivariate central limit theorem) Assume that

- $S(k) \sim c|k|^{\alpha}$, as $|k| \to 0$ where $c > 0$ and $0 < \alpha < d$,
- Φ is Brillinger mixing.[1]

Then:

$$\left(R^{\frac{\alpha-d}{2}j} \sum_{x \in \Phi \cap [-R,R]^d} f_i(x/R^j) \right)_{i \in I, j \in J} \xrightarrow[R \to \infty]{Law} (\sqrt{c} N(i,j,\alpha))_{i \in I, j \in J},$$

where $(N(i,j,\alpha))_{i \in I, j \in J}$ is a Gaussian vector with zero mean and covariance matrix:

$$\Sigma(\alpha) := \left(\mathbf{1}_{j_1=j_2} \int_{\mathbb{R}^d} \mathcal{F}[f_{i_1}](k) \overline{\mathcal{F}[f_{i_2}]}(k) |k|^{\alpha} dk \right)_{(j_1,j_2) \in J^2, (i_1,i_2) \in I^2}.$$

[1] Remember mixing $\mathbb{P}_{\Phi \cap (B_1 \cup (x+B_2))} \xrightarrow{x \to \infty} \mathbb{P}_{\Phi \cap B_1} \times \mathbb{P}_{\Phi \cap B_2}$.

Brillinger mixing concerns the rate of convergence in the mixing process.

Asymptotic confidence intervals

- Under the assumptions of the CLT, let:
  - $a \in (0, 1)$,
  - for all $\beta \geq 0$ and $q \in (0, 1)$, let $F^{-1}(q; \beta)$ be the quantile of order q of

$$\sum_{j \in J} w_j \log \left(\sum_{i \in I} N(i, j, \beta)^2 \right).$$

Then

$$\left[\widehat{\alpha} - \frac{F^{-1}(1 - a/2; (\widehat{\alpha})_+)}{\log(R)},\ \widehat{\alpha} - \frac{F^{-1}(a/2; (\widehat{\alpha})_+)}{\log(R)} \right]$$

is an asymptotic confidence interval of order $1 - a$.

- Assume:
  - $S(k) \sim c|k|^{\alpha}| + c_1|k|^{\beta}$, with $\beta > \alpha \geq 0$ and $c, c_1 > 0$ constants.
  - Φ is Brillinger mixing,
  - $f_i = \psi_i(x) = e^{-\frac{1}{2}|x|^2} \prod_{l=1}^{d} H_{i_l}(x_l)$ where $H_n(y)$ are the Hermite polynomials and $I = \{i \in \mathbb{N}^d |\ |i|_\infty \leq N_I,\ \int \psi_i = 0\}$.

Then, there exists $R_0 > 0$ and $0 < C(\epsilon, J) < \infty$ such that for all $R \geq R_0$:

$$\mathbb{P}\left(\log(R)\left|\widehat{\alpha}(I,J,R) - \alpha\right| \geq \epsilon\right) \leq C(\epsilon, J)\left(\left(\frac{|I|}{R^{2j}}\right)^{\beta-\alpha} + \frac{1}{|I|}\right).$$

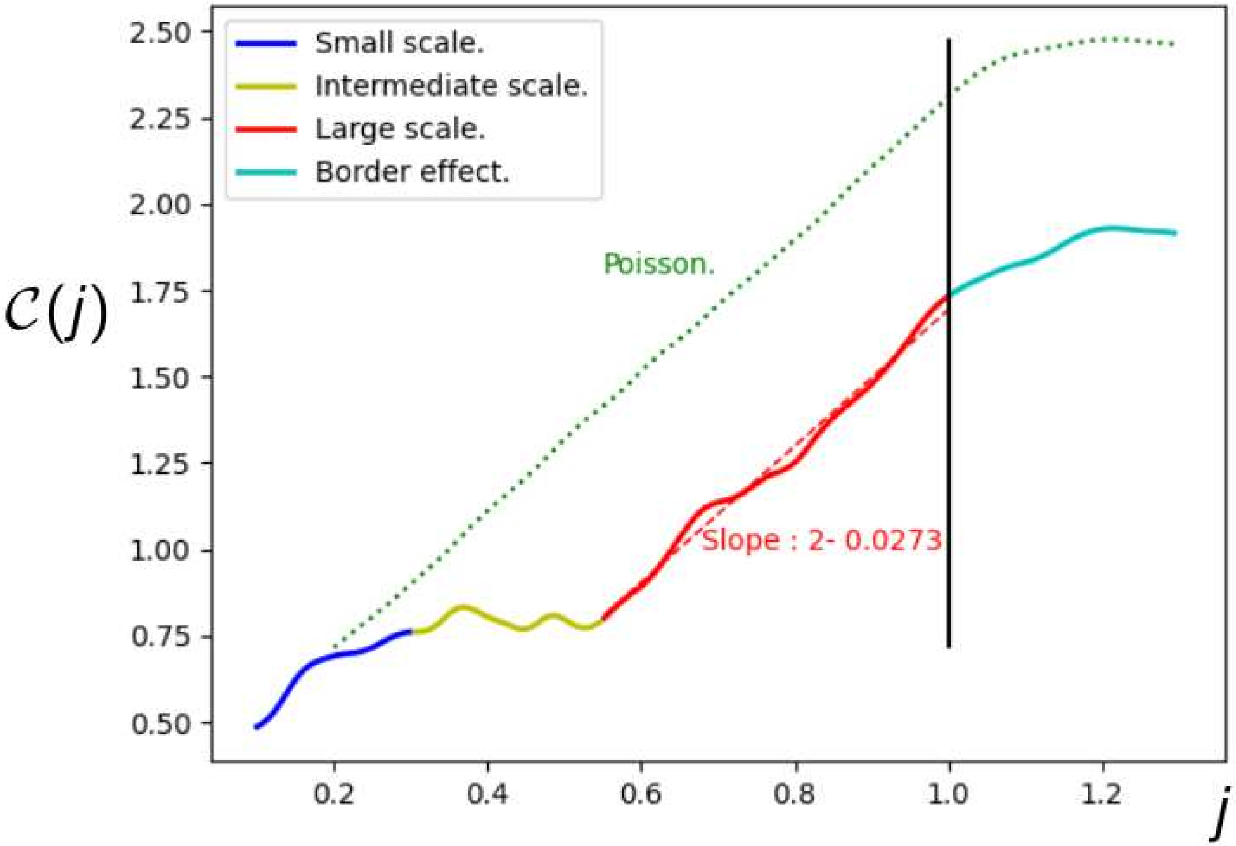
- Variance scales as $|I|^{-1}$, Bias can be high if $|I|$ is large for fixed R.

Examples / implementation issues

Implementation issues: case non-hyperuniform

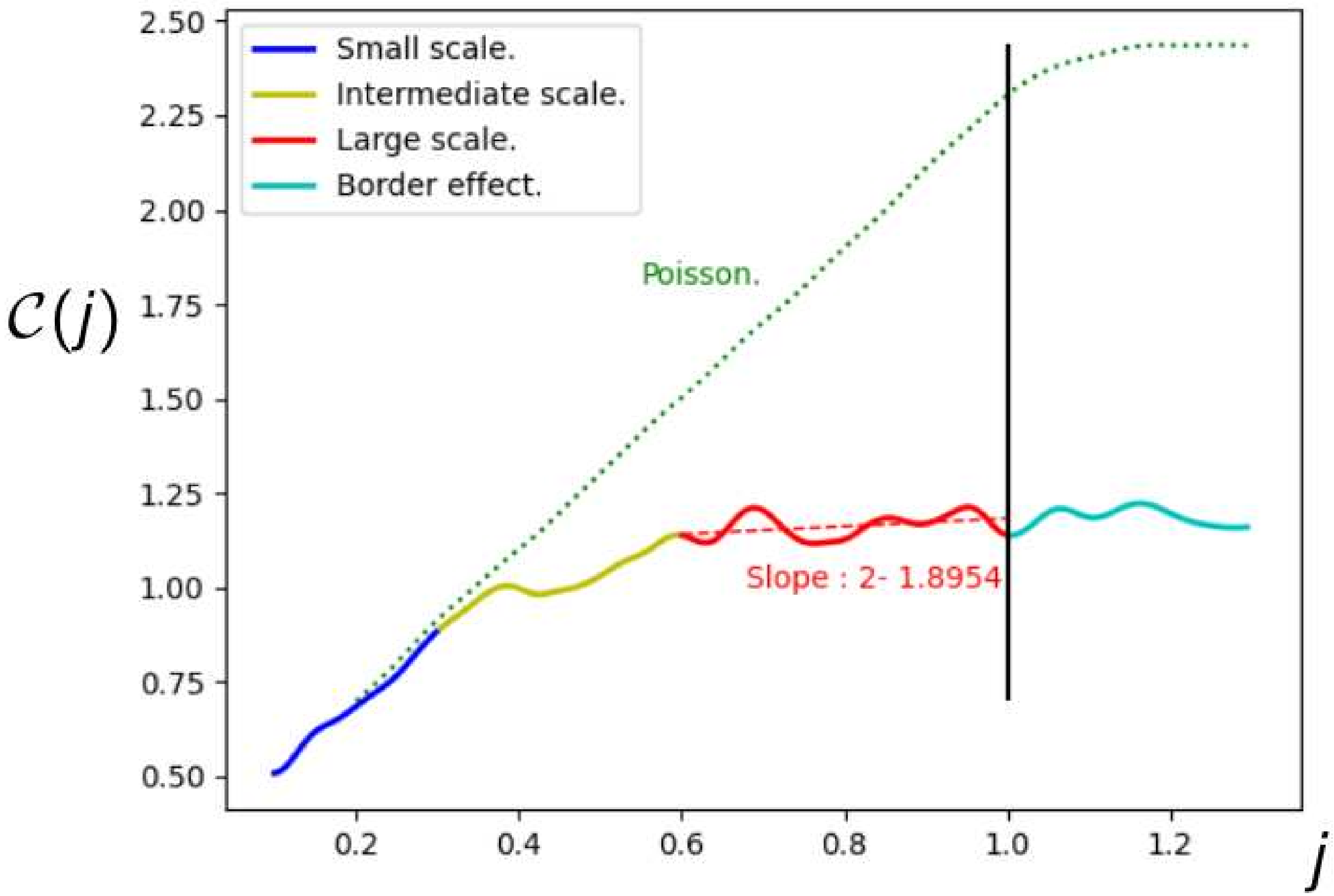
$$\widehat{\alpha} = d - \text{ slope of } \mathcal{C}, \text{ with}$$

$$\mathcal{C} : j \mapsto \frac{1}{\log(R)} \log \left(\sum_{i \in I} \left(\sum_{x \in \Phi \cap [-R,R]^d} f_i \left(x / R^j \right) \right)^2 \right).$$



Matérn-III model: 5000 points, $I = \{75$ Hermite tapers$\}$, R = 35.

Implementation issues: case of strong hyperuniform



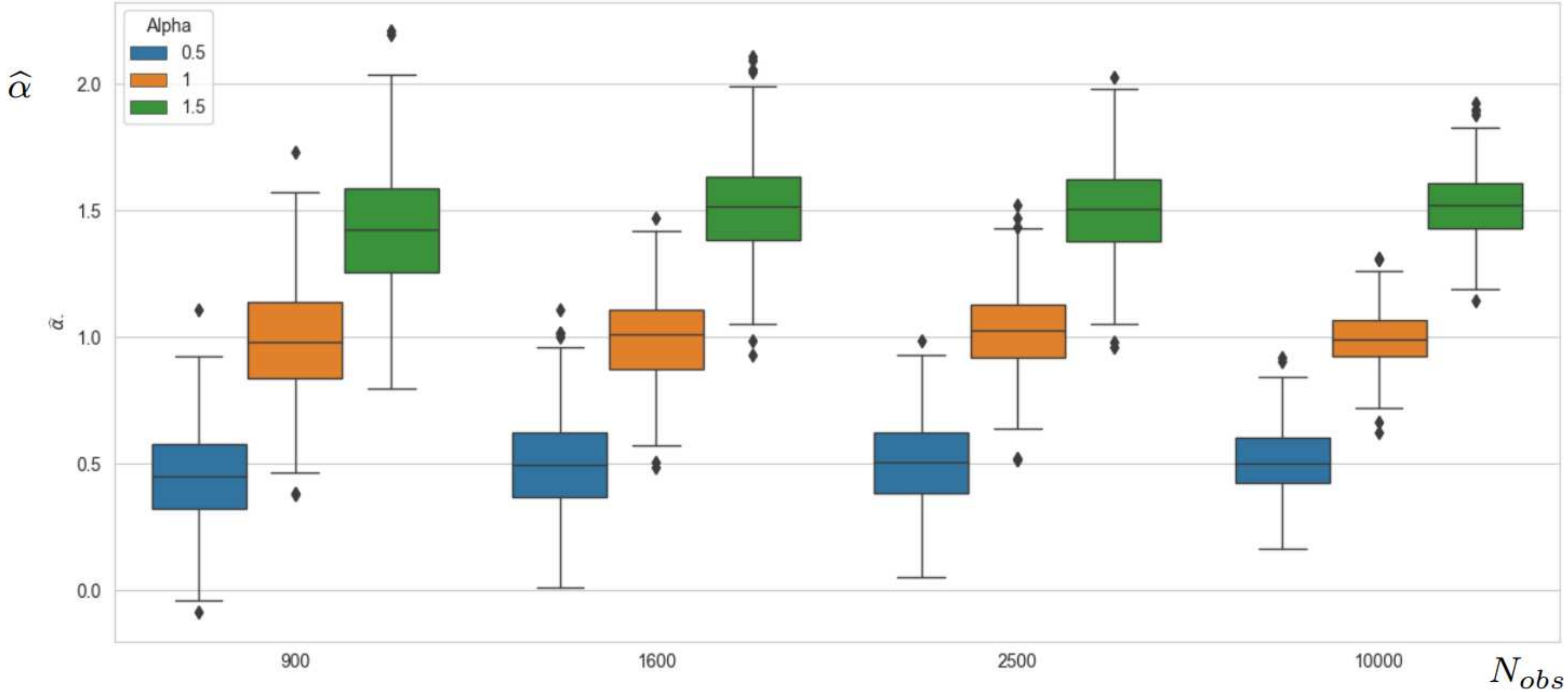
Ginibre model, 1600 points, $I = \{75$ Hermite tapers$\}$, R = 20.

Benchmark on perturbed lattices

Assume

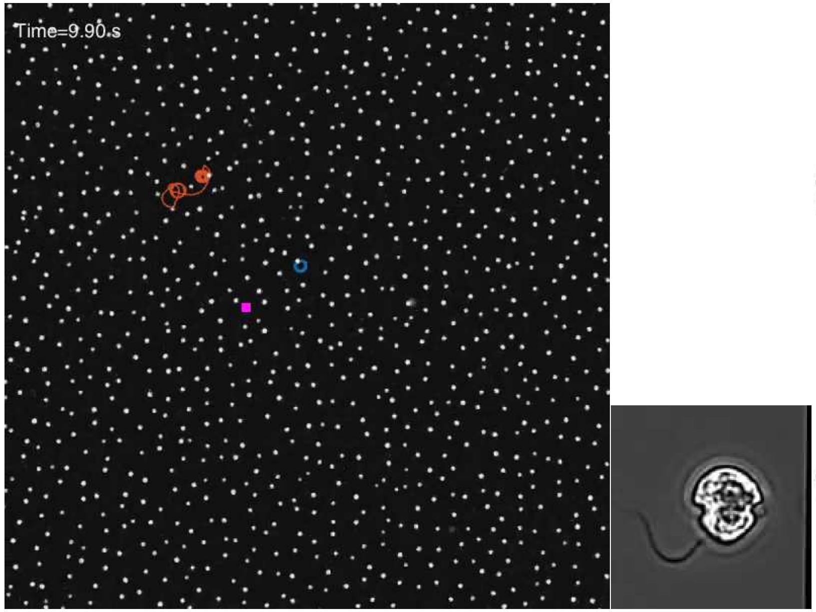
$$\Phi_{\alpha} = \{y + U + U_y + V_y | y \in \mathbb{Z}^2\}$$

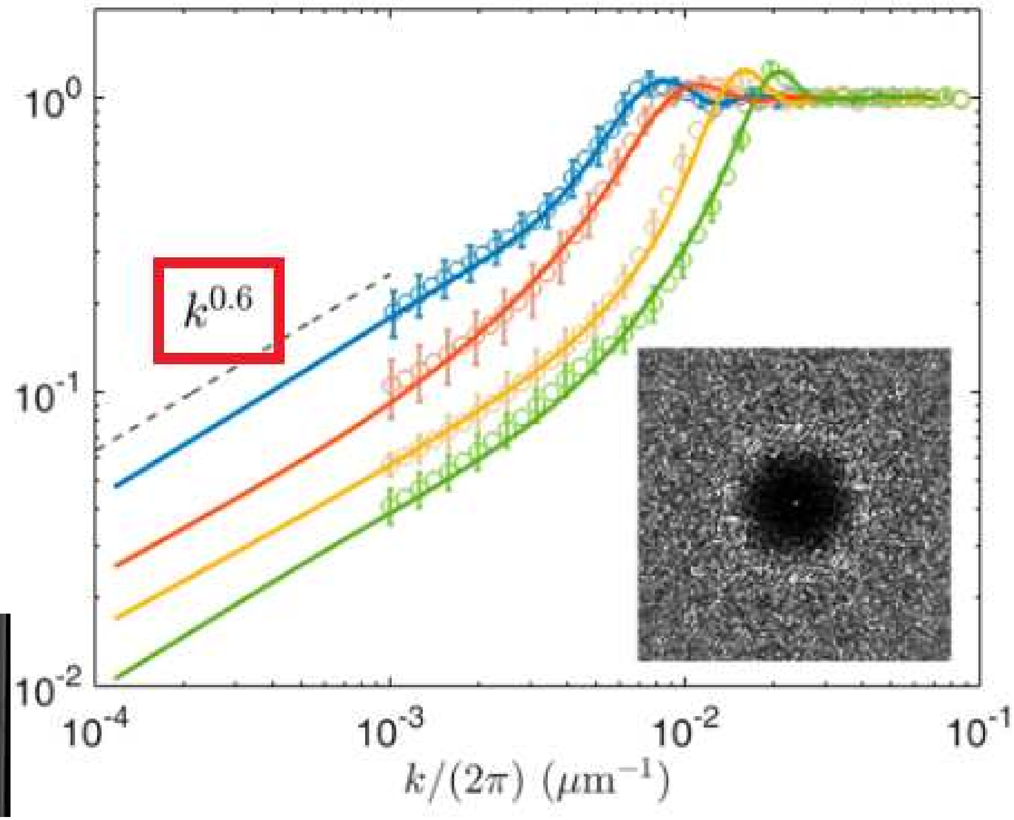
where U, $(U_y)_{y\in\mathbb{Z}^2}$ are i.i.d. uniform on $[-1/2, 1/2]^2$, and $(V_y)_{y\in\mathbb{Z}^2}$ are i.i.d. with characteristic function φ s.t. $1 - |\varphi(k)|^2 \sim_0 t|k|^{\alpha}$.
(for $V_y \equiv 1$ — cloaked lattice (Klatt, Kim, and Torquato 2020).



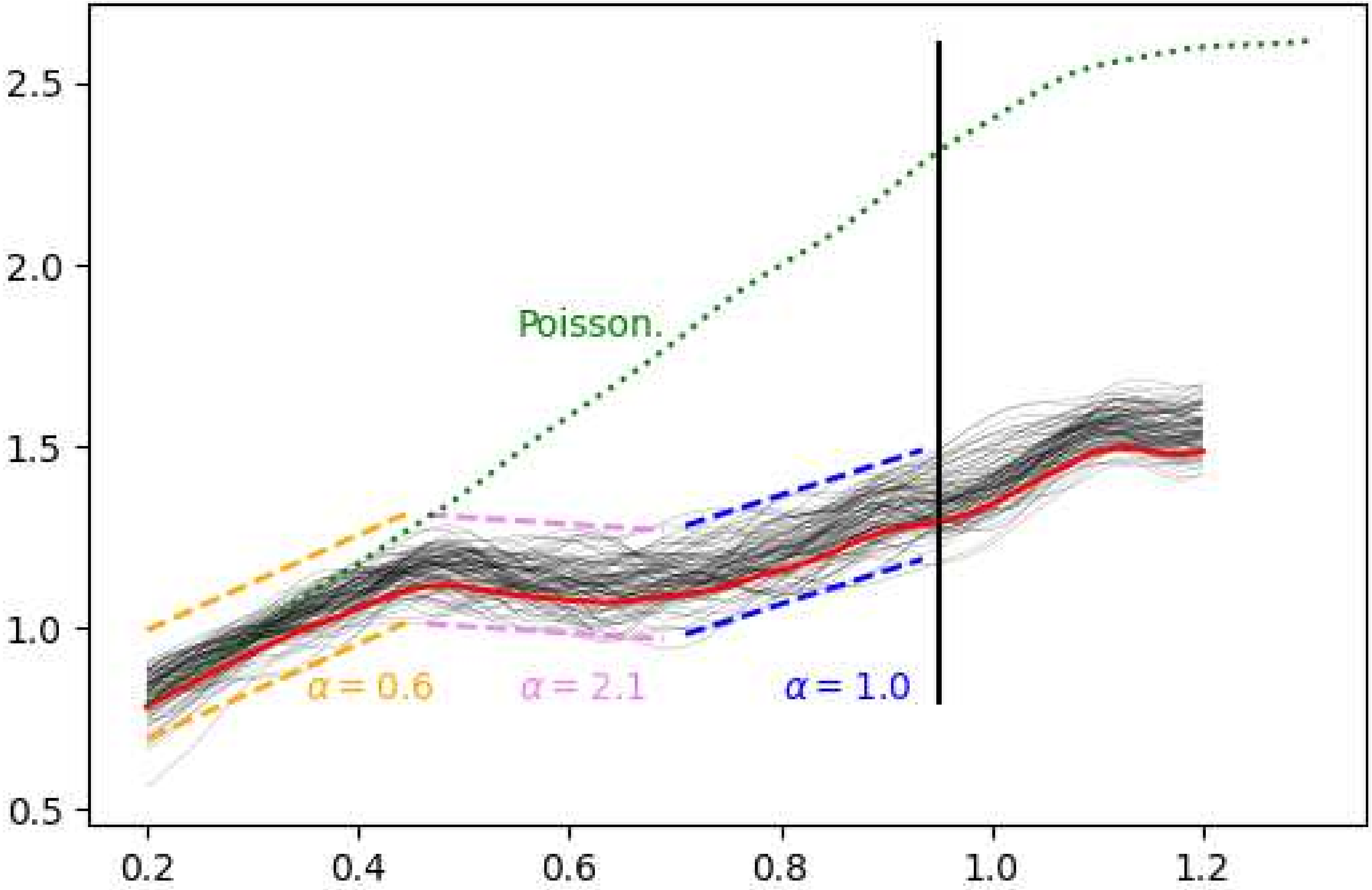
Perturbed lattices, $I = \{75$ Hermite tapers$\}$.

Real data — System of marine algae (Huang et al. 2021)





Marine algae — our estimation of α



Estimating α for an algae system (approximately 900 points).

Conclusions

- Hyperuniformity — the variance of random systems grows slower than the volume of the window $\equiv$ low frequencies disappear.
- Assume point process on $\mathbb{R}^d$ having Bartlett spectrum $S(k) \sim_0 c|k|^\alpha$ with $c > 0$ and $\alpha \geq 0$. Case $\alpha > 0$ indicates hyperuniformity.
- Multi-scale, multi-taper estimators of α applicable on one realization

$$\widehat{\alpha}(I, J, R) := d - \sum_{j \in J} \frac{w_j}{\log(R)} \log\left(\sum_{i \in I}\left(\sum_{x \in \Phi \cap [-R,R]^d} f_i(x/R^j)\right)^2\right).$$

- Brillinger-mixing + $\alpha < d$: CLT + confidence intervals.
- $\alpha \geq d$: consistency criterion.
- Choice of the number of tapers: bias/variance trade-off.

For more details, see:

- Mastrilli, G., BB, Lavancier, F. (2024). Estimating the hyperuniformity exponent of point processes. arXiv:2407.16797
- Klatt, M. A., Last, G. and Henze, N. A genuine test for hyperuniformity. (2022) arXiv:2210.12790
- Hawat, D., Gautier, G., Bardenet, R. and Lachièze-Rey, R. On estimating the structure factor of a point process, with applications to hyperuniformity. (2023) Statistics and Computing
- Torquato, S. Hyperuniform states of matter. (2018)Physics Reports
- Torquato, S. and Stillinger, F. H. Local density fluctuations, hyperuniformity, and order metrics. (2003) Physical Review E.
- ...

Thanks for your attention