



HAL
open science

Estimation de la fonction de structure d'un processus ponctuel pour l'étude d'hyperuniformité

Diala Hawat, Guillaume Gautier, Rémi Bardenet, Raphaël Lachièze-Rey

► **To cite this version:**

Diala Hawat, Guillaume Gautier, Rémi Bardenet, Raphaël Lachièze-Rey. Estimation de la fonction de structure d'un processus ponctuel pour l'étude d'hyperuniformité. XXVIIIème Colloque Francophone de Traitement du Signal et des Images GRETSI 2022, Sep 2022, Nancy, France. hal-03747149

HAL Id: hal-03747149

<https://hal.science/hal-03747149v1>

Submitted on 12 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation de la fonction de structure d'un processus ponctuel pour l'étude d'hyperuniformité

Diala HAWAT^{1,2}, Guillaume GAUTIER¹, Rémi BARDENET¹, Raphaël LACHIEZE-REY²

¹Université de Lille, CNRS, Centrale Lille
UMR 9189 – CRISTAL, F-59000 Lille, France

²Université de Paris cité, MAP5, Paris, France

diala.hawat@univ-lille.fr, guillaume.gautier@univ-lille.fr,
remi.bardenet@univ-lille.fr, raphael.lachieze-rey@parisdescartes.fr

Résumé – Un processus ponctuel hyperuniforme est un nuage de points aléatoires réparti *très* régulièrement. Cette propriété a des applications en intégration numérique comme en physique des matériaux. Nous proposons une courte revue des estimateurs de la fonction de structure, une quantité spectrale qui caractérise l'hyperuniformité. Ces estimateurs sont disponibles dans un paquet Python appelé [structure-factor](#).

Abstract – A hyperuniform point process is a random cloud of points that is spread *very* regularly, a property that finds applications in Monte Carlo integration as well as materials science. We give a short survey of estimators of the structure factor, a spectral quantity that characterizes hyperuniformity. All estimators are implemented in a companion Python package called [structure-factor](#).

1 Introduction

Un processus ponctuel est un ensemble aléatoire de points dont l'intersection avec toute fenêtre bornée est finie. Il est dit *hyperuniforme* lorsque la variance du nombre de points observés dans une grande fenêtre est négligeable devant le volume de cette fenêtre. Pour comparer, des points répartis selon un processus de Poisson, donc tirés uniformément et indépendamment conditionnellement à une fenêtre fixée, donnent une variance qui grandit comme le volume de la fenêtre. L'étude de l'hyperuniformité est motivée par de nombreuses applications en traitement du signal, pour contrôler la propagation d'ondes dans des matériaux de synthèse [1], ou pour diminuer l'*aliasing* en répartissant régulièrement des capteurs optiques [2].

L'étude mathématique des processus hyperuniformes étant difficile, des diagnostics empiriques d'hyperuniformité sont souvent utilisés. La pratique courante est d'estimer la *fonction de structure* S du processus ponctuel à partir d'un ou plusieurs échantillons, et d'étudier son comportement au voisinage de zéro, qui caractérise l'hyperuniformité. Plusieurs estimateurs de S existent dans la littérature et les codes informatiques disponibles diffèrent souvent d'un domaine applicatif à l'autre. Il nous a paru important d'en proposer une revue [3] accompagnée d'une implémentation unifiée dans un paquet Python appelé [structure-factor](#)¹, pour tenter de systématiser et garantir la reproductibilité des études numériques d'hyperuniformité.

Dans le présent article, nous proposons une revue rapide de ces estimateurs, insistant sur les plus reliés à l'étude spectrale des séries temporelles, bien connue des traiteurs de signaux.

2 Définitions

Soit un processus ponctuel $\mathcal{X} \subset \mathbb{R}^d$ stationnaire, i.e., de loi invariante par translation. Dans ce cas, le nombre de points $\mathcal{X}(A)$ qui tombe dans une fenêtre A est une variable aléatoire, de moyenne proportionnelle au volume de A ; la constante de proportionnalité $\rho > 0$ est appelée *intensité* du processus. Quant à la variance de $\mathcal{X}(A)$, on peut s'attendre à ce qu'elle grandisse comme $|A|$, le volume de A ; c'est le cas pour le processus de Poisson, un processus de référence sans aucune interaction entre points, en particulier au sens où, pour $A \cap B = \emptyset$, $\mathcal{X}(A)$ et $\mathcal{X}(B)$ sont des variables aléatoires indépendantes. Lorsque, contrairement à Poisson, les interactions entre paires de points largement espacés ne sont pas nulles, la variance de $\mathcal{X}(A)$ peut être négligeable devant le volume de A ; on parle alors de processus *hyperuniforme*.

Pour caractériser les interactions entre les points de \mathcal{X} , on utilise la *fonction de corrélation par paire* $g : \mathbb{R}^d \rightarrow [0, +\infty[$, où $g(\mathbf{r})d\mathbf{r}$ s'interprète comme la probabilité d'avoir un point de \mathcal{X} en \mathbf{r} , sachant qu'il y en a un en $\mathbf{0}$. On peut estimer g directement [4], ou encore s'intéresser à la *fonction de structure*

$$S(\mathbf{k}) \triangleq 1 + \rho \mathcal{F}(g - 1)(\mathbf{k}), \quad (1)$$

où pour $\mathbf{k} \in \mathbb{R}^d$, $\mathcal{F}(f)(\mathbf{k}) \triangleq \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\langle \mathbf{k}, \mathbf{x} \rangle}$. En analyse spectrale des signaux aléatoires, la quantité la plus proche de S est sans doute la transformée de Fourier de l'autocovariance d'un processus stationnaire [5]. Notons enfin que lorsque \mathcal{X} est également supposé isotrope, i.e., de loi invariante par rotation, g et S sont des fonctions radiales. En notant $r = \|\mathbf{r}\|_2$ et $k =$

1. <https://github.com/For-a-few-DPPs-more/structure-factor>

$\|\mathbf{k}\|_2$, on peut alors écrire, avec un abus de notation,

$$S(k) = 1 + \rho(2\pi)^{d/2} \int_0^\infty r^{d/2} (g(r) - 1) \frac{J_{d/2-1}(kr)}{k^{d/2-1}} dr, \quad (2)$$

où J_ν est la fonction de Bessel du premier type d'ordre ν .

La fonction de structure caractérise les processus hyperuniformes. En effet, sous de faibles hypothèses, un processus ponctuel stationnaire est hyperuniforme ssi $S(\mathbf{0}) = 0$ [6]. De plus, lorsque la fonction de structure satisfait $|S(\mathbf{k})| \sim c\|\mathbf{k}\|_2^\alpha$ au voisinage du zéro, l'exposant α permet de quantifier la variance du nombre de points qui tombent dans une boule de rayon R , cf. Table 1.

TABLE 1 – Classes d'hyperuniformité, selon Torquato [1].

Classe	I	II	III
α	> 1	$= 1$	$\in]0, 1[$
$\text{Var}[\mathcal{X}(B(0, R))]$	$\mathcal{O}(R^{d-1})$	$\mathcal{O}(R^{d-1} \log(R))$	$\mathcal{O}(R^{d-\alpha})$

Comme exemple fondamental, un processus de Poisson stationnaire est isotrope mais pas hyperuniforme : ses fonctions de corrélation et de structure sont constantes et égales à 1. Sur l'échantillon de la Figure 1, on ne décèle effectivement pas de structure à l'œil nu. En revanche, le processus de Ginibre, qui correspond à la limite, quand $N \rightarrow \infty$, du processus des valeurs propres (complexes) d'une matrice aléatoire $N \times N$ remplie de gaussiennes complexes standard i.i.d., est hyperuniforme de classe I, avec $\alpha = 2$. En effet, Ginibre est stationnaire et isotrope, avec $g(r) = 1 - \exp(-r^2)$ et $S(k) = 1 - \exp(-k^2/4)$. On décèle sa régularité, même à petite échelle de distance, sur l'échantillon de la Figure 1.

3 Estimer la fonction de structure

On suppose disposer d'un échantillon $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ de \mathcal{X} restreint à une fenêtre $W \subset \mathbb{R}^d$, de volume $|W| < \infty$.

3.1 Pour un processus stationnaire

On suppose ici que $W = \prod_{j=1}^d [-L_j/2, L_j/2]$. L'estimateur de (1) le plus courant est la *scattering intensity*

$$\hat{S}_{\text{SI}}(\mathbf{k}) \triangleq \frac{1}{\rho|W|} \left| \sum_{j=1}^N e^{-i\langle \mathbf{k}, \mathbf{x}_j \rangle} \right|^2, \quad \mathbf{k} \in \mathbb{R}^d \setminus \{\mathbf{0}\}. \quad (3)$$

Comme son nom l'indique, la *scattering intensity* mime l'intensité mesurée dans une expérience de diffraction sur un assemblage de N particules dont les positions sont décrites par $\mathcal{X} \cap W$. D'un point de vue signal, (3) contient les notions de transformée de Fourier discrète et de périodogramme [7].

On peut montrer que $\hat{S}_{\text{SI}}(\mathbf{k})$ est un estimateur asymptotiquement non biaisé de la fonction de structure $S(\mathbf{k})$ pour des vecteurs d'onde $\mathbf{k} = (k_1, \dots, k_d)$ dont au moins une composante vaut $k_j = 2\pi n/L_j$ avec $n \in \mathbb{Z} \setminus \{0\}$. L'asymptotique est à comprendre au sens où la fenêtre d'observation croît jusqu'à couvrir tout l'espace \mathbb{R}^d . Pour une fenêtre W donnée, de tels vecteurs sont qualifiés de « vecteurs d'onde autorisés » [8], car ils permettent d'annuler une partie du biais de l'estimateur.

Une façon naturelle de généraliser l'estimateur (3) est d'introduire une *fonction de pondération* (ou *taper* en anglais) $t(\cdot, W)$, supportée sur W . On obtient alors un nouvel estimateur, dit pondéré ou *tapered*, sous la forme

$$\hat{S}_{\text{T}}(t, \mathbf{k}) \triangleq \frac{1}{\rho} \left| \sum_{j=1}^N t(\mathbf{x}_j, W) e^{-i\langle \mathbf{k}, \mathbf{x}_j \rangle} \right|^2, \quad \mathbf{k} \in \mathbb{R}^d. \quad (4)$$

On retrouve (3) lorsque $t(\mathbf{x}, W) = |W|^{-1/2} \mathbb{1}_W(\mathbf{x})$.

Pour chaque estimateur pondéré ; on peut définir l'analogue des « vecteurs d'onde autorisés » pour la *scattering intensity*, sur lesquels l'estimateur associé est asymptotiquement non biaisé. Ils correspondent en fait aux zéros de la transformée de Fourier de la fonction pondération t . Pour des vecteurs d'onde plus généraux, [5] proposent deux façons de corriger une partie du biais associé à la géométrie de W et au choix du taper t . La première, dite indirecte, corrige le terme associé au périodogramme, via l'estimateur

$$\hat{S}_{\text{UDT}}(t, \mathbf{k}) \triangleq \frac{1}{\rho} \left| \sum_{j=1}^N t(\mathbf{x}_j, W) e^{-i\langle \mathbf{k}, \mathbf{x}_j \rangle} \right|^2 - \rho |\mathcal{F}(t)(\mathbf{k}, W)|^2. \quad (5)$$

La seconde, dite directe, corrige le terme relatif à la transformée de Fourier discrète, via l'estimateur

$$\hat{S}_{\text{DDT}}(t, \mathbf{k}) \triangleq \frac{1}{\rho} \left| \sum_{j=1}^N t(\mathbf{x}_j, W) e^{-i\langle \mathbf{k}, \mathbf{x}_j \rangle} - \rho \mathcal{F}(t)(\mathbf{k}, W) \right|^2. \quad (6)$$

L'avantage du débiaisage direct (6) est que l'estimateur est presque sûrement positif, contrairement à (5).

Pour aller plus loin, il est possible de moyennner les estimateurs venant de plusieurs *tapers*, selon une approche dite de *multitapering* bien établie pour les séries temporelles [7]. L'estimateur obtenu est nommé \hat{S}_{MT} . En débiaisant indirectement ou directement, on obtient \hat{S}_{UDMT} et \hat{S}_{DDMT} , respectivement. L'espoir est de diminuer la variance de l'estimateur sans en augmenter le biais.

3.2 Pour un processus stationnaire et isotrope

On suppose ici que $W = B(\mathbf{0}, R)$ est la boule centrée de rayon $R > 0$. L'équivalent isotrope de la *scattering intensity* peut être construit à partir de la formulation radiale (2), de la fonction de structure, sous la forme

$$\hat{S}_{\text{BI}}(k) \triangleq 1 + \frac{(2\pi)^{d/2}}{\rho|W|\omega_{d-1}} \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{J_{d/2-1}(k\|\mathbf{x}_i - \mathbf{x}_j\|_2)}{(k\|\mathbf{x}_i - \mathbf{x}_j\|_2)^{d/2-1}}, \quad (7)$$

où ω_{d-1} est la surface de la sphère unité en dimension d . Cet estimateur, dit de Bartlett isotrope, est une généralisation multidimensionnelle de l'estimateur de [9]. A l'instar de la *scattering intensity*, les « nombres d'onde autorisés » sont ici les zéros de la fonction $k \mapsto J_{d/2}(Rk)$. Sur ces valeurs, l'estimateur de Bartlett est asymptotiquement non biaisé.

De façon alternative, il est possible d'adopter une stratégie en deux étapes pour approximer la fonction de structure radiale.

Il suffit de (i) construire un estimateur \hat{g} de la fonction de corrélation par paire g [4], (ii) estimer l'intégrale unidimensionnelle (2), en remplaçant g par \hat{g} et en utilisant un schéma de quadrature. Deux quadratures traitent spécifiquement d'intégrales de la forme (2), l'une impliquant directement les fonctions de Bessel [10], l'autre une discrétisation « à la Fourier » de la transformée de Hankel [11]. On nomme respectivement les deux estimateurs obtenus \hat{S}_{HO} et \hat{S}_{HBC} .

4 Expériences

Les estimateurs présentés dans la Section 3 sont regroupés dans un paquet Python, [structure-factor](#),¹ accompagné d'une [documentation](#) et d'un [tutoriel](#).

Démonstration des estimateurs. Nous considérons deux réalisations d'approximativement 5800 points, l'une du processus de Poisson d'intensité $1/\pi$, l'autre de Ginibre. La Figure 1 illustre les estimateurs de la Section 3.1. La première colonne correspond à Ginibre, la seconde à Poisson. La première ligne montre les deux échantillons. Le processus de Ginibre montre plus de régularité à grande échelle, et nous allons voir que l'estimation de S confirme son hyperuniformité. La deuxième ligne montre en effet les résultats de \hat{S}_{SI} , évalué sur des vecteurs d'ondes parmi les vecteurs d'ondes autorisés. Les trois dernières lignes correspondent respectivement aux estimateurs \hat{S}_{MT} , \hat{S}_{DDMT} et \hat{S}_{UDMT} , appliqués à quatre tapers sinusoidaux [3, Section 3.1.4]. Sur ces figures, les nuages de points gris correspondent aux valeurs estimées de $\hat{S}(\mathbf{k})$ en fonction de $\|\mathbf{k}\|_2$. Notons que les deux processus étant isotropes, les fonctions de structure sont radiales. Comme les estimateurs que nous comparons ici sont évalués sur une grille de \mathbb{R}^2 , nous moyennons les estimées de S sur des régions annulaires. La moyenne plus ou moins trois écarts-type empiriques de la moyenne est montrée en rouge pour chaque bin. Finalement, les droites vertes représentent la valeur théorique de S pour Ginibre, et les lignes noires pointillées celle de Poisson, qui sert de référence générale sur les graphes de structure.

On remarque d'abord que l'estimateur *multitaper* débiaisé indirectement donne quelques valeurs négatives, conduisant à de grandes imprécisions en échelle log-log. De plus, le biais de l'estimateur \hat{S}_{MT} est remarquablement grand sur les petites valeurs de $\|\mathbf{k}\|_2$ non-autorisées. L'estimateur \hat{S}_{DDMT} semble uniformément plus proche de la valeur théorique que \hat{S}_{SI} , ce que nous allons vérifier en estimant leurs erreurs quadratiques moyennes intégrées en \mathbf{k} . Auparavant, nous montrons simplement les résultats des estimateurs isotropes.

La Figure 2 illustre les résultats des estimateurs de la Section 3.2. La deuxième ligne montre les résultats de l'estimateur de Bartlett \hat{S}_{BI} , évalué sur des nombres d'ondes parmi les nombres d'ondes autorisés. Les deux dernières lignes présentent respectivement les résultats de \hat{S}_{HO} , \hat{S}_{HBC} , moyennés sur des bins en $k = \|\mathbf{k}\|_2$ comme pour la *scattering intensity*. Les légendes des Figures 1-2 sont identiques.

Notons que \hat{S}_{BI} semble le plus précis, et que parmi les estimateurs basés sur une quadrature numérique, \hat{S}_{HBC} semble plus stable que \hat{S}_{HO} sur les petites valeurs de k . Pour plus

d'expériences, nous renvoyons le lecteur vers [3, Section 4].

Comparaison quantitative des estimateurs. Concernant le temps de calcul, \hat{S}_{BI} est le plus coûteux, étant une somme d'un nombre quadratique d'évaluations, tandis que \hat{S}_T est la somme de N exponentielles. L'utilisation des filtres *multitaper* dans \hat{S}_{MT} multiplie le coût de \hat{S}_T par le nombre de tapers, mais ce dernier est typiquement petit [5] et l'on peut facilement paralléliser les évaluations, un cœur par *taper*.

La Table 2 montre les estimations \widehat{iMSE} de l'erreur quadratique moyenne de $\hat{S}(k)$, intégrée sur un intervalle de valeurs de k proches de 0. Une petite $iMSE$ est donc désirable pour les diagnostics graphiques d'hyperuniformité. La contribution de la variance intégrée \widehat{iVar} à \widehat{iMSE} , pour les deux estimateurs supposés les plus précis, \hat{S}_{SI} et \hat{S}_{DDMT} . L'estimation de la $iMSE$ se fait en moyennant sur 50 échantillons indépendants. On constate un facteur de 4 entre la \widehat{iMSE} de \hat{S}_{DDMT} et celle de \hat{S}_{SI} , et que ce facteur est largement significatif. Dans ce cas, nous pouvons donc être confiants que le *multitapering* réduit bien l'erreur quadratique moyenne. On constate que la variance domine le biais dans la $iMSE$.

La Table 3 présente la même étude pour les estimateurs isotropes \hat{S}_{BI} , \hat{S}_{HO} et \hat{S}_{HBC} . L'erreur \widehat{iMSE} de \hat{S}_{BI} est de deux ordres de grandeur plus petite que \hat{S}_{HO} et \hat{S}_{HBC} , significative, et encore dominée par la variance, cf. aussi [3, Section 5].

Diagnostics d'hyperuniformité. Pour effectuer un diagnostic empirique d'hyperuniformité, [1] propose de régresser linéairement $\log \hat{S}$ sur $\log k$, et de comparer l'ordonnée à l'origine H de la droite obtenue à un seuil proche de 0. Pour inférer la classe d'hyperuniformité cf. Table 1, on peut également régresser un monôme k^α à \hat{S} au voisinage de 0. La Figure 3 montre la distribution empirique des indices H et α obtenus via l'estimateur \hat{S}_{BI} , cf. [3, Section 4.4] pour plus de détails.

Pour le processus de Poisson, les valeurs de H obtenues sont toutes plus grandes que 0.4, ce qui confirme empiriquement la non-hyperuniformité. Au contraire, les valeurs de H pour l'ensemble de Ginibre sont toujours négatives, ce qui est révélateur d'un phénomène d'hyperuniformité. Finalement, la concentration des valeurs de α autour de 2 pour Ginibre confirme les résultats théoriques connus.

5 Discussion

L'analyse spectrale des processus ponctuels gagne à s'inspirer des méthodes développées pour les séries temporelles, comme le *multitapering* [5]. Le choix des *tapers* reste une question largement ouverte, que ce soit pour l'estimateur *multitaper* directement débiaisé \hat{S}_{DDMT} , ou l'estimateur de Bartlett \hat{S}_{BI} , qui, d'après notre expérience, sont les plus précis.

La comparaison de ces deux estimateurs est une voie naturelle pour poursuivre les travaux, mais est plus subtile qu'il n'y paraît. Par exemple, l'estimateur *multi-taper* est moins coûteux, trivialement parallélisable, mais nécessite plus d'évaluations que l'estimateur de Bartlett pour une précision similaire. Nous espérons que le présent article, qui résume notre revue [3], et notre paquet logiciel seront une étape vers

la systématisation de l'étude empirique de l'hyperuniformité.

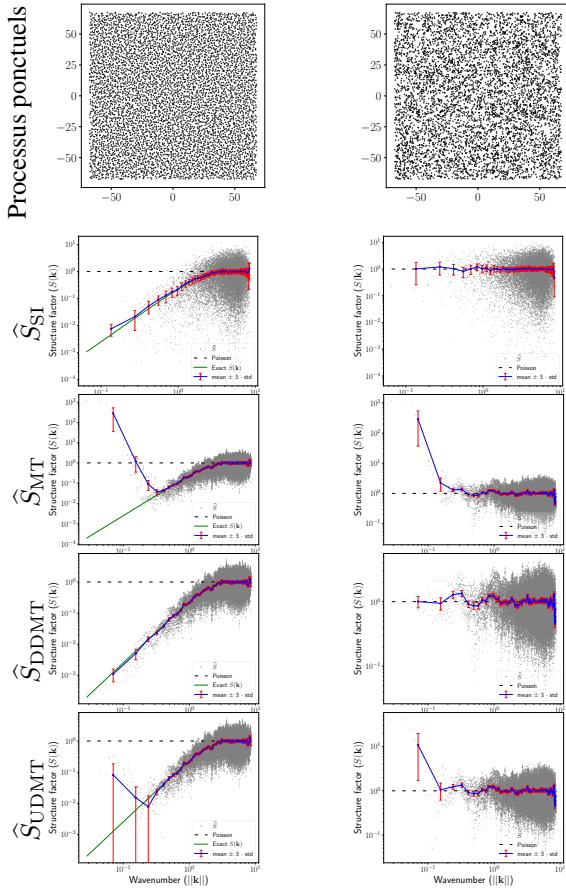


FIGURE 1 – Estimateurs de la Section 3.1 appliqués à l'ensemble de Ginibre (gauche) et le processus de Poisson (droite).

TABLE 2 – $\widehat{\text{iMSE}}$ des estimateurs stationnaires.

Estimateurs	$\widehat{\text{iVar}}$	$\text{CI}[\widehat{\text{iMSE}}]$	$\widehat{\text{iVar}}$	$\text{CI}[\widehat{\text{iMSE}}]$
\widehat{S}_{SI}	0.32	0.32 ± 0.02	1.31	1.34 ± 0.06
$\widehat{S}_{\text{DDMT}}$	0.08	0.08 ± 0.007	0.37	0.38 ± 0.02
	Ginibre		Poisson	

TABLE 3 – $\widehat{\text{iMSE}}$ des estimateurs stationnaires isotropes.

Estimateurs	$\widehat{\text{iVar}}$	$\text{CI}[\widehat{\text{iMSE}}]$	$\widehat{\text{iVar}}$	$\text{CI}[\widehat{\text{iMSE}}]$
\widehat{S}_{BI}	0.003	0.004 ± 0.0003	0.05	0.05 ± 0.009
\widehat{S}_{HO}	0.37	0.38 ± 0.09	2.12	2.14 ± 0.93
\widehat{S}_{HBC}	0.03	0.03 ± 0.01	2.44	2.45 ± 1.33
	Ginibre		Poisson	

Références

- [1] Salvatore Torquato. Hyperuniform states of matter. *Physics Reports*, 2018.
- [2] Ming-Jie Sun, Xin-Yu Zhao, and Li-Jing Li. Imaging using hyperuniform

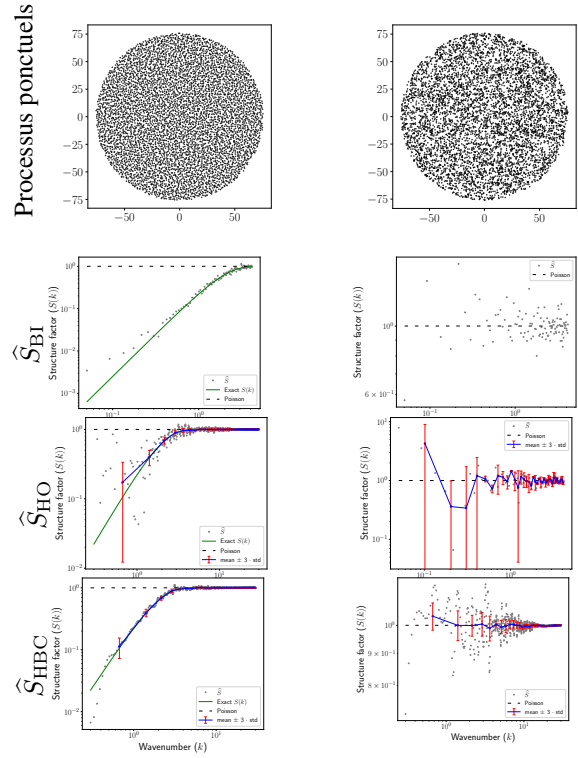
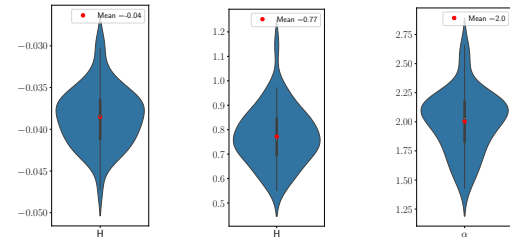


FIGURE 2 – Estimateurs de la Section 3.2 appliqués à l'ensemble de Ginibre (gauche) et le processus de Poisson (droite).



(a) H : Ginibre (b) H : Poisson (c) α : Ginibre

FIGURE 3 – Diagnostics d'hyperuniformité

- sampling with a single-pixel camera. *Opt. Lett.*, 2018.
- [3] Diala Hawat, Guillaume Gautier, Rémi Bardenet, and Raphaël Lachièze-Rey. On estimating the structure factor of a point process, with applications to hyperuniformity. 2022. arXiv : 2203.08749.
- [4] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns : Methodology and Applications with R*. Chapman and Hall/CRC, 2015.
- [5] Tuomas A. Rajala, Sofia C. Olhede, and David John Murrell. Spectral estimation for spatial point patterns. 2020. arXiv : 2009.01474.
- [6] Simon Coste. Order, fluctuations, rigidities. 2021.
- [7] Donald B. Percival and Andrew T. Walden. *Spectral Analysis for Univariate Time Series*. Cambridge University Press, 2020.
- [8] Michael Andreas Klatt, Günter Last, and D. Yogeshwaran. Hyperuniform and rigid stable matchings. *Random Structures & Algorithms*, 2020.
- [9] M. S. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 1964.
- [10] Hidenori Ogata. A numerical integration formula based on the Bessel functions. *Publications of the Research Institute for Mathematical Sciences*, 2005.
- [11] Natalie Baddour and Ugo Chouinard. Theory and operational rules for the discrete Hankel transform. *Journal of the Optical Society of America A*, 2015.