

# Journées MAS 2020

## Programme détaillé

### Table des Matières

<b>1</b>	<b>Conférences plénières</b>	<b>1</b>
<b>2</b>	<b>Prix Neveu</b>	<b>3</b>
<b>3</b>	<b>Sessions parallèles</b>	<b>4</b>
3.1	Sessions 1 . . . . .	4
3.2	Sessions 2 . . . . .	7
3.3	Sessions 3 . . . . .	9
3.4	Sessions 4 . . . . .	12
3.5	Sessions 5 . . . . .	14
3.6	Sessions 6 . . . . .	17

## 1 Conférences plénières

### Béatrice de Tilière (Université Paris-Dauphine)

Titre : Modèle de dimères sur les graphes minimaux : le cas elliptique et au-delà.

Résumé : Le modèle de dimères représente la répartition de molécules di-atomiques à la surface d'un cristal. Ceci est modélisé par des couplages parfaits d'un graphe planaire choisis selon la mesure de Boltzmann. Lorsque le graphe est périodique Kenyon, Okounkov et Sheffield montrent que le diagramme de phase est donné par la courbe spectrale qui a la propriété remarquable d'être de Harnack; ils établissent de plus une correspondance entre ces courbes et les modèles de dimères sur les graphes bipartis. Un autre résultat important est l'expression locale obtenue par Kenyon pour les corrélations lorsque le graphe sous-jacent est isoradial et le modèle est critique. Dans une série de travaux en collaboration avec Cédric Boutillier (Sorbonne Université) et David Cimasoni (Université de Genève), nous généralisons ces résultats dans un cadre unifié. Nous considérons le modèle sur des graphes minimaux (une famille légèrement plus grande que les isoradiaux) et établissons une correspondance explicite avec les courbes de Harnack; nous obtenons également une formule locale pour les corrélations pour la famille à deux paramètres de mesures de Gibbs.

### Davide Faranda (Université Paris-Saclay)

Titre : Physique Statistique des événements extrêmes

Résumé: Comment modéliser le comportement moyen, la variabilité et les événements extrêmes de systèmes complexes ? Complexe est ce que nous percevons comme spatialement, temporellement et dynamiquement riche et esthétiquement beau. Cette richesse se décline sous la forme de tourbillons turbulents, de la rage d'un orage ou de la diffusion exponentiellement rapide d'un virus ou d'une crise économique. Dans cet exposé je combine le point de vue de la physique statistique et de la théorie des systèmes dynamiques pour concevoir des

outils mathématiques qui agissent comme des loupes sur les systèmes complexes. La première question spécifique que j'ai abordée est de déterminer le nombre de variables, d'équations ou de données dont nous avons besoin pour décrire un événement spécifique, que cela se révèle ou non extrême pour le système que nous examinons. L'un des résultats de cette recherche est la possibilité d'étudier certains des systèmes les plus complexes comme par exemple un ensemble de vortex turbulents dans une géométrie confinée (écoulement de von Karman) avec trois équations dynamiques simples. Ces équations ne décrivent pas seulement l'état moyen du système, mais elles permettent une prévisibilité à court terme de son mouvement. Nous avons ainsi découvert que les événements extrêmes dans la circulation atmosphérique tels que les cyclones montrent souvent une structure plus simple que le comportement moyen de l'atmosphère : ils agissent comme des condensats où quasiment tous les degrés de liberté de la dynamique sont gelés. Cela implique qu'ils correspondent à des briques spécifiques de la géométrie mathématique sous-jacente et parfois inconnue de ces systèmes: les points fixes instables de la dynamique. équipés de ces outils statistiques, nous pouvons donc rechercher l'empreinte de points fixes instables dans les phénomènes naturels et découvrir leur correspondance avec les événements extrêmes rencontrés dans la vie quotidienne: jusqu'à présent nous avons découvert qu'ils apparaissent comme des tourbillons turbulents extrêmes dans l'atmosphère, comme des tempêtes, des ouragans, des tremblements de terre ou même comme des états extrêmes d'autres systèmes complexes.

### **Josselin Garnier (Ecole Polytechnique)**

Titre : Propagation des ondes et imagerie en milieux aléatoires

Résumé : Dans cet exposé, nous considérons différentes configurations de propagation des ondes en milieux aléatoires, pour des applications à l'imagerie dans des milieux complexes. Nous montrons qu'une analyse asymptotique basée sur une technique de séparation d'échelles permet de dériver des équations différentielles stochastiques effectives satisfaites par le champ d'ondes et de calculer les moments du champ d'ondes. Nous discutons d'une conjecture physique qui affirme que le champ d'ondes acquiert une statistique gaussienne lorsque l'onde se propage sur une grande distance dans un milieu aléatoire. Nous présentons une configuration où cette conjecture est juste et une autre où elle est fautive.

### **Amandine Marrel (CEA)**

Titre : Analyse de sensibilité basée sur les mesures de dépendance HSIC – Application au traitement des incertitudes en simulation numérique

Résumé : Dans le cadre du traitement des incertitudes en simulation numérique, l'analyse de sensibilité globale vise à étudier l'impact des incertitudes de chaque paramètre d'entrée  $(X_i)_{i=1\dots d}$  sur la (ou les) sortie(s) du simulateur  $Y$ . Dans le cadre d'une formalisation probabiliste, des mesures de dissimilarité sont utilisées comme mesures de dépendance pour comparer la loi de probabilité jointe  $(X_i, Y)$  avec le produit des lois marginales. On peut citer par exemple la distance de covariance basée sur l'utilisation des fonctions caractéristiques. Plus généralement, ces mesures font partie d'une classe plus large de mesures de dépendance basées sur la « redescription » des variables dans des espaces de Hilbert à noyaux reproduisants (RKHS). Tirant parti de « l'astuce des noyaux », ces mesures basées sur les RKHS ont connu un engouement croissant ces dernières années pour l'analyse de sensibilité des simulateurs numériques. Parmi elles, le critère d'indépendance de Hilbert–Schmidt noté HSIC pour Hilbert Schmidt Independence Criterion généralise la notion de covariance entre deux variables aléatoires et permet ainsi de capturer un très large spectre de formes de dépendance entre les variables. Cette présentation introduira les principes théoriques et l'inférence statistiques autour des mesures HSIC. Les tests statistiques d'indépendance construits sur ces

mesures seront plus précisément décrits. Leur utilisation pratique dans le cadre de l'analyse de sensibilité pour le screening et la hiérarchisation des variables d'entrée sera décrite et illustrée sur des applications industrielles. Certains travaux récents seront également présentés, afin d'étendre leur utilisation (sorties fonctionnelles, plans d'expériences stratifiés notamment) et d'accroître la puissance des tests statistiques associés.

### **Cristina Toninelli (Université Paris Dauphine)**

Titre : Modèle de Fredrickson-Andersen 2-spin facilité

Résumé: Le modèle de Fredrickson-Andersen 2-spin facilité (FA-2f) est un système de particules en interaction avec contraintes cinétiques (KCM). Son évolution suit un processus de Markov à temps continu de type Glauber soumis à la contrainte suivante: une particule ne peut être créée/tuée sur un site que si au moins 2 de ses plus proches voisins sont vides. Cette contrainte induit une dynamique coopérative et vitreuse caractérisée par des échelles de temps qui divergent rapidement lorsque la densité de sites vides,  $q$ , tend vers 0. Si on note  $\tau$  le premier temps auquel l'origine est vidée, notre résultat pour le processus FA-2f stationnaire sur  $\mathbb{Z}^d$  s'écrit de la façon suivante:  $\tau \sim \exp\left(\frac{(d\lambda(d,2)+o(1))}{q^{1/(d-1)}}\right)$  w.h.p. pour toute dimension  $d \geq 2$ , où  $\lambda(d,2)$  est la constante de seuil pour la percolation bootstrap à 2-voisins. Nous expliquerons le mécanisme de relaxation conduisant à ce résultat qui résout une controverse dans la littérature physique, nous donnerons un aperçu des techniques de preuve et enfin nous mentionnerons d'autres résultats pouvant être obtenus avec nos techniques pour des KCM plus généraux, notamment des résultats d'universalité pour  $d = 2$ .

[En collaboration avec I.Hartarsky and F.Martinelli]

### **Rémi Rhodes (Université Aix-Marseille)**

Title : Solving the Liouville model

Abstract: we will review recent progress in the understanding of some specific conformal field theory called the Liouville model.

## **2 Prix Neveu**

### **Elsa Cazelles (IRIT), Prix Neveu 2018**

Titre: Apprentissage de données et transport optimal : outils statistiques et algorithmiques.

Résumé : La thématique principale de mes travaux de recherche est l'analyse de données pouvant être décrites par des distributions de probabilité supportées sur un espace Euclidien, au moyen du transport optimal. Des analyses statistiques d'ordre un et deux sont une première approche pour comprendre les tendances générales d'un ensemble d'observations. Dans le contexte du transport optimal, ces études correspondent au barycentre et à la décomposition en composantes géodésiques principales dans l'espace de Wasserstein. Notamment, je m'intéresse à des estimateurs régularisés du barycentre, afin de gérer le bruit provenant des observations. Par ailleurs, dans le cas d'étude de données sous forme de séries temporelles, le transport optimal est utilisé afin de comparer l'information spectrale de ces objets.

### **Léo Miolane (NYU), Prix Neveu 2019**

Titre: Estimation statistique et "modèles de spins"

Résumé: Nous allons nous intéresser aux liens entre certains problèmes d'estimation statistique et des modèles de "spins" issus de la physique statistique. Nous nous concentrerons principalement sur l'estimation d'une matrice de rang 1 bruitée, dans un cadre bayésien.

Dans cet exemple, la distribution a posteriori de la matrice de rang 1 est similaire à des modèles de spins à champ moyen. Comme pour les modèles "physiques", nous verrons que ces modèles présentent des transitions de phases qui ont d'importantes implications en termes d'estimation statistique.

### 3 Sessions parallèles

#### 3.1 Sessions 1

• **Matrices et graphes aléatoires**, organisateur Pascal Maillard (Université Toulouse 3)

1) Mireille Capitaine

Titre : Fluctuations de la transformée de Stieltjes de la mesure spectrale empirique de polynômes autoadjoints en une matrice de Wigner et une matrice diagonale déterministe  
Résumé: Nous présenterons le résultat suivant. Lorsque la dimension tend vers l'infini, le processus analytique sur  $\mathbb{C} \setminus \mathbb{R}$  des traces non normalisées des résolvantes d'un polynôme autoadjoint en une matrice de Wigner et une matrice diagonale déterministe, recentrées, converge vers un processus Gaussien complexe centré dont la covariance s'exprime en termes de fonctions de subordination à valeurs opérateurs définies en probabilités libres. Il s'agit d'un travail en collaboration avec Serban Belinschi, Sandrine Dallaporta et Maxime Février

2) Simon Coste

Titre : Convergence of the characteristic polynomial of sparse matrices

Abstract : The reverse characteristic polynomial of a matrix  $A$  is  $q(z) = \det(I - zA)$ . I will show that when  $A$  is a (non-Hermitian)  $n \times n$  random matrix with iid Bernoulli( $d/n$ ) entries, with  $d$  fixed, then when  $n$  goes to infinity, the polynomial  $q$  converges towards a random infinite holomorphic function, which is the exponential of a random series with Poisson coefficients. This entails a short proof of previous results on the asymptotics of the eigenvalues of  $A$ , a strategy inspired by a recent paper by Bordenave, Chafai and Garcia-Zelada.

3) Cécile Mailler

Titre : Cellules de Voronoi dans les arbres de fragmentation de Devroye

Résumé: Prenons  $k$  nœuds d'un graphe à  $n$  nœuds, et faisons partir de ces  $k$  nœuds  $k$  épidémies concurrentes qui se propagent de manière déterministe à la même vitesse le long des arêtes de l'arbre. Un nœud ne peut être infecté que par une seule épidémie, la première qui l'atteint. Nous nous intéressons à la taille des territoires finaux de ces  $k$  épidémies et appelons ces territoires les «cellules de Voronoi» des  $k$  points initiaux. Dans ce travail en collaboration avec Markus Heydenreich (Munich) et Alex Drewitz (Cologne), nous étudions la taille des cellules de Voronoi de  $k$  nœuds tirés uniformément au hasard parmi les  $n$  nœuds d'un arbre aléatoire de fragmentation.

4) Franck Gabriel

Titre : Tweaking algorithms in finite-size random matrix approximations.

Abstract : For a large family of random matrices, as the dimensions goes to infinity, observables, e.g. normalized traces, converge to a finite limit which define a natural limiting mathematical object. As a result, operations on these limiting objects can be approximated by similar operations on finite-size random matrices. This is, for instance, the underlying idea behind the Gaussian Random Features approximation of Kernel Ridge Regression in which the kernel Gram matrix is approximated by random covariant matrix. Although the finite-size random matrices are good approximations of the limiting objects, applying the same operations on the finite size random matrices as on the limiting object might be probably not the best approximation one could consider in order to estimate the result of the operations on the limiting object. Interestingly, the "natural" algorithm can in fact be tweaked to get more effective implementations. In this talk, we propose two innovative applications of

such tweaking approach. In the first example, more theoretical, we show that two random matrices  $A_N, B_N$  which are invariant by conjugation by permutations and independent are asymptotically “free conditionally on their diagonal” (the so-called free with amalgamation property): using a fixed point algorithm, this allows to get beyond the naive approximation of the eigenvalue distribution of the limit of  $A_N + B_N$  obtained by simply considering the eigenvalue distribution of  $A_N + B_N$ . In the second one, more practical and which finds its motivation in Supervised Machine Learning, we consider Kernel Ridge Regression (KRR) and the afore-mentioned Gaussian Random Features (GRF) approximation with non-zero ridge: in this case, we show that by finely tuning the ridge in the GRF model in a non-trivial way, one obtains a better finite-size approximation of the KRR estimator.

- **Apprentissage statistique robuste**, organisateur Adrien Saumard (ENSAI)

- 1) Claire Bréchet (Univ. Rennes 2)

Titre : Approximation de données par des unions d’ellipsoïdes et partitionnement.

Résumé : Je vais introduire des substituts pour la fonction distance au support d’une mesure, dont les sous-niveaux sont des unions de boules ou des unions d’ellipsoïdes. J’énoncerai plusieurs résultats. En particulier, je parlerai des vitesses d’approximation de ces substituts par leurs versions empiriques, construites à partir d’échantillons de points. J’expliquerai aussi comment mettre à profit de tels estimateurs pour partitionner des données qui ont une structure géométrique particulière. Les résultats présentés sont issus des papiers [1, 2], et de travaux en cours. [1] C. Bréchet. Robust anisotropic power-functions-based filtrations for clustering. In 36th International Symposium on Computational Geometry (SoCG 2020), vol. 164, 2020. [2] C. Bréchet, C. Levrard. A k-points-based distance for robust geometric inference. Bernoulli, 26(4), 2020.

- 2) Timothée Matthieu (LMO, Orsay)

Titre : étude robuste de M-estimateurs consistants via une distance de transport optimal

Résumé : Dans cette présentation, je vais exposer des résultats de consistance et de stabilité des M-estimateurs pour l’estimation robuste de la moyenne multivariée quand les données sont corrompues. En particulier, je vais introduire une famille de distances venant de la théorie du transport optimal qui paraît particulièrement adaptée au problème de l’estimation robuste, et grâce à cette famille de distance je vais donner les conditions sur les données anormales (en terme de quantité de données anormales et/ou en terme de moments de la distribution des données anormales) pour qu’un M-estimateur converge.

- 3) Jules Dépersion (CREST, ENSAE)

Titre : Robustesse de l’estimateur de Stahel-Donoho.

Résumé : Nous étudions des versions median-of-mean (MOM) de la Stahel-Donoho outlyingness (SDO) et des fonctions d’écart absolu médian (MAD) pour construire des estimateurs robustes aux données contaminées ou « heavy-tailed ». Grâce aux propriétés presque isométriques des versions MOM de SDO et MAD, on donne les premières bornes non asymptotiques de la « médiane de Stahel-Donoho », un estimateur original qui permet de donner des bornes en « norme de prédiction » et pas seulement en norme L2, et qui complètent la consistance [Maronna 1995] et la normalité asymptotique [Zuo, Cui, He, 2004] des estimateurs de Stahel-Donoho. Nous montrons également que la version MOM de MAD peut être utilisée pour construire un estimateur de la matrice de covariance sous une seule hypothèse de moment L2 ou d’un paramètre d’échelle si un second moment n’existe pas.

- 4) Edouard Genetay (CREST, ENSAI)

Titre : Construction d’un algorithme de clustering robuste via bootstrap MOM.

Résumé : Nous présenterons un nouvel algorithme de clustering robuste: K-bMOM. L’estimateur à la base de cet algorithme hérite ses propriétés robustes d’une variante bootstrap de la stratégie Median-Of-Means (bMOM). Nous commencerons par présenter la robustesse de bMOM et de K-bMOM à la présence d’outliers via leur "breakdown points", puis nous

étayerons ces résultats par des simulations numériques dans différents contextes.

• **Géostatistiques**, organisateurs Thomas Romary (Mines Paris Tech) et Mike Pereira (Chalmers University of Technology)

1) Cédric Travelletti, Universität Bern

Title : Uncertainty Quantification and Experimental Design for Large-Scale Linear Inverse Problems under Gaussian Process Priors, with Applications to Geophysics

Abstract : When using Gaussian process (GP) priors for solving inverse problems in a Bayesian framework, the computational complexity scales cubically in the number of data-points. In the context of inverse problems involving integral operators, one faces additional difficulties that hinder inversion on large grids since covariance matrices can become too large to be stored. In this talk, we introduce an implicit representation of the posterior covariance matrix that reduces the memory footprint by only storing low rank intermediate matrices, while allowing individual elements to be accessed on-the-fly without needing to build the full posterior covariance. Moreover, this representation allows for fast sequential inclusion of new observations. These faculties are crucial when considering optimal experimental design tasks and we here demonstrate our approach by computing optimal data collection plans for excursion set recovery for a gravimetric inverse problem, where the goal is to provide fine resolution estimates of high density regions inside the Stromboli volcano, Italy. Optimal data collection plans are computed by extending the weighted integrated variance reduction (wIVR) criterion to inverse problems. Our results show that this criterion is able to significantly reduce the uncertainty on the excursion volume, reaching close to minimal levels of residual uncertainty. Overall, our techniques allow the advantages of probabilistic models to be brought to bear on large-scale inverse problems arising in the natural sciences. Particularly, applying the latest developments in Bayesian optimal design on realistic large-scale problems opens new venues of research at the interface of machine learning and traditional experimental science.

2) Chantal de Fouquet, Mines ParisTech

Titre : Estimation géostatistique à base physique : application à la caractérisation de panaches de contaminants

Résumé : La géométrie des panaches de pollution dans l'atmosphère, les sols ou les nappes phréatiques est liée à la variabilité du milieu, à la localisation des sources, mais surtout au mode de transport, à l'évolution des polluants et à leurs interactions avec l'environnement. Lorsque les données sont peu nombreuses, le résultat des techniques géostatistiques usuelles (krigeage) peut apparaître physiquement peu cohérent avec la connaissance a priori sur les mécanismes physique en jeux. Sur un cas synthétique bidimensionnel d'un panache de contamination radioactive en milieu non saturé, deux estimateurs géostatistiques tenant compte explicitement de la modélisation physique sont introduits (Pannecoucke et al., 2019) : ils s'appuient sur la modélisation physique de l'écoulement et de transport en considérant un déversement accidentel et des conditions aux limites connues (la chronique des précipitations et la charge hydraulique aux limites du domaine) et des paramètres hydrauliques hétérogènes, simulés par méthodes géostatistiques. La grandeur à prédire est ainsi modélisée comme la réalisation d'une fonction aléatoire non stationnaire dont les deux premiers moments, moyenne et variogramme, sont estimés à partir des résultats de la modélisation numérique. Le krigeage avec dérive externe utilise la moyenne des résultats des simulations d'écoulement en dérive. La seconde méthode, originale, est un krigeage avec un variogramme non stationnaire empirique, calculé sur les résultats des simulations d'écoulement. Ces deux méthodes non stationnaires sont comparées au krigeage usuel. Très général, ce couplage des simulations numériques et de la géostatistique s'avère particulièrement utile dans le cas de données peu nombreuses. Le choix de l'une de ces méthodes repose sur le nombre de simulations disponibles, la qualité de la modélisation physique, ainsi que sur le problème d'estimation (linéaire ou non linéaire) à résoudre.

3) Mike Pereira, Chalmers University

Titre : Champs gaussiens et variétés riemanniennes: quand la géostatistique se passe de matrices de covariance

Résumé : De nombreuses applications en statistique spatiale nécessitent que les données soient modélisées par des processus gaussiens définis sur des domaines non euclidiens, ou avec des propriétés non stationnaires. L'utilisation de tels modèles se fait généralement au prix d'une augmentation drastique des coûts opérationnels (en termes de calcul et de stockage), ce qui les rend difficiles à appliquer à de grands jeux de données. Dans cet exposé, nous proposons une solution à ce problème qui repose sur la définition d'une classe de champs aléatoires sur des variétés riemanniennes. La discrétisation de ces champs aléatoires, entreprise à l'aide d'une approche par éléments finis, fournit alors une caractérisation explicite des champs, qui est exploitée afin de fournir des algorithmes efficaces de simulation, prédiction et filtrage de données. Nous illustrons cette approche par un cas d'étude de filtrage de données sismiques.

4) Thomas Romary, Mines ParisTech

Titre : Simulations géostatistiques par réseaux antagonistes génératifs

Résumé : Une des applications historiques de la géostatistique est l'évaluation des ressources du sous-sol, qu'elles soient minières, pétrolières ou hydriques. Cette évaluation doit se faire sur la base de données parcellaires et est par conséquent incertaine. Un grand nombre de modélisations probabilistes ont été proposées pour représenter les milieux d'intérêt ainsi que des algorithmes dédiés à leur conditionnement aux données disponibles. Ces différents modèles ne sont cependant pas capables de représenter des configurations géologiques complexes. Depuis une vingtaine d'années, le centre de géosciences de Mines ParisTech développe le logiciel FLUMY, basé sur un modèle génétique d'environnements de dépôts sédimentaires qui permet de générer des simulations très réalistes. Dans une optique de quantification d'incertitudes, ce modèle est en revanche difficile à conditionner aux données et relativement lent à générer un grand nombre de réalisations. Dans cette présentation, on montrera qu'il est possible de générer des simulations réalistes à l'aide des réseaux antagonistes génératifs tout en les conditionnant aux données disponibles au travers d'une approche bayésienne variationnelle.

## 3.2 Sessions 2

• **Trajectoires rugueuses et EDPS**, organisateurs Ismaël Bailleul (Université Rennes 1) et Nicolas Marie (Université Paris-Nanterre)

1) Yvain Bruned

Title: Ramification of Volterra-type Rough Paths

Abstract: In this talk, we will present an extension of the new approach proposed by Harang and Tindel for dealing with stochastic Volterra equation using the ideas of Rough Path theory. The main idea of this construction is to consider iterated integral convolutions with the Volterra kernel. This leads to the corresponding abstract objects called Volterra-type Rough Paths as well as the notion of the convolution product, an extension of the natural tensor product used in Rough Path theory. It is joint work with Foivos Katsetsiadis.

2) William Salkeld

Title: Probabilistic Rough Paths

Abstract: In this talk, I will explain some of the foundation results for a new regularity structure developed to study interactive systems of equations and their mean-field limits. At the heart of this solution theory is a Taylor expansion using the so called Lions measure derivative. This quantifies infinitesimal perturbations of probability measures induced by infinitesimal variations in a linear space of random variable. I will detail some of the distinctions between this regularity structure and previous examples and motivate a quotient operations on branched rough paths which elucidates the link between probabilistic rough paths with

an empirical law and their mean-field limits. This talk is based on ArXiv:2106.09801 and ongoing work with my supervisor Francois Delarue at Universite Cote d'Azur.

3) Adeline Fermanian

Title: Framing RNN as a kernel method: A neural ODE approach

Abstract: Building on the interpretation of a recurrent neural network (RNN) as a continuous-time neural differential equation, we show, under appropriate conditions, that the solution of a RNN can be viewed as a linear function of a specific feature set of the input sequence, known as the signature. This connection allows us to frame a RNN as a kernel method in a suitable reproducing kernel Hilbert space. As a consequence, we obtain theoretical guarantees on generalization and stability for a large class of recurrent networks. Our results are illustrated on simulated datasets.

4) Carlo Bellingeri :

Title : Beyond geometric rough paths

Abstract : In the classical theory of rough paths, we restrict the analysis to so-called geometric rough paths, when the driving noise has very low Hölder regularity. This assumption naturally has many interesting analytical properties, but it fits poorly in probabilistic contexts involving stochastic calculus and non-commutative probability. In this talk, we will present two new families of rough objects more adapted for these situations: the quasi-geometric rough paths and the non-commutative rough paths. Even if they have different definitions, we will show how these new notions share deep connections with geometric rough paths, allowing us to explore richer algebraic structures.

• **Probabilités et algorithmique**, organisateur Nicolas Broutin (Sorbonne Université)

1) Luis Fredes (Orsay)

Title: Aldous-Broder theorem: the non reversible case and new combinatorial proof

Abstract: The Aldous–Broder algorithm is a famous algorithm used to sample a uniform spanning tree of any finite connected graph  $G$ , but it is more general: it states that given a reversible  $M$  Markov chain on  $G$  started at  $r$  and up to the cover time, the tree rooted at  $r$  formed by the steps of successive first entrance in each node (different from the root) has a probability proportional to the product of these edges according to  $M$ , where the edges are directed toward  $r$ . In this talk I will present an extension to the non-reversible case and a new combinatorial proof of this theorem. (Based on a joint work with Jean-François Marckert.)

2) Philippe Chassaing (Institut Elie Cartan, Université de Lorraine)

Titre: Champs de vecteurs et grands objets combinatoires

Résumé: On montrera que l'énumération asymptotique des automates complets déterministes accessibles, qui sont l'input de l'algorithme de Moore, peut être menée à bien via l'analyse des trajectoires d'un champ de vecteurs étroitement lié aux nombres de Stirling de deuxième espèce, et on examinera de ce point de vue quelques autres triangles combinatoires. (En collaboration avec Jules Flin et Alexis Zevio)

3) Nicolas Chenavier (Université du Littoral Côte d'Opale)

Titre: Deux problèmes d'extrêmes sur un graphe de Delaunay.

Résumé: Le graphe de Delaunay d'un ensemble de points du plan est une triangulation telle qu'aucun point de l'ensemble n'est à l'intérieur du disque circonscrit d'un des triangles. Cet objet fournit un algorithme pour calculer l'arbre couvrant minimal et possède un grand nombre d'applications, notamment en segmentation d'image ou en méthodes des éléments finis pour construire des maillages. Dans cet exposé, on s'intéresse à deux problèmes d'extrêmes sur un graphe de Delaunay lorsque celui-ci est généré par un processus ponctuel de Poisson. Le premier porte sur la longueur du chemin le plus court entre deux sommets pris au hasard dans le graphe. En re-normalisant par la distance euclidienne entre ces deux points, on fournit un encadrement de l'espérance de la distance. Le second problème porte sur le maximum des degrés des sommets du graphe se situant dans la fenêtre  $W_n = [-n, n]^2$ . Lorsque  $n$  tend vers



l'infini, on montre que celui-ci se concentre, avec grande probabilité, sur au plus deux entiers consécutifs. On fournit également l'ordre exact du maximum. Le premier problème est basé sur un travail avec O. Devillers (INRIA Nancy) et le second avec G. Bonnet (Université de Bochum).

### 3.3 Sessions 3

• **Théorèmes limites pour les produits de matrices i.i.d.**, organisateur Christophe Cuny (Université de Brest)

1) Hui Xiao (Hildesheim's University) Title: Large deviation expansions for the coefficients of random walks on the general linear group

Abstract: Consider a sequence of independent and identically distributed random elements  $(g_n)_{n \geq 1}$  on the general linear group  $GL(d, \mathbb{R})$  and the left random walk  $G_n := g_n \dots g_1$ . Under suitable conditions, we establish Bahadur-Rao-Petrov type large deviation expansions for the coefficients  $\langle f, G_n v \rangle$  of the product  $G_n$ , where  $v \in \mathbb{R}^d$  and  $f \in (\mathbb{R}^d)^*$ . In particular, we obtain an explicit rate function in the large deviation principle, thus improving significantly the known large deviation bounds. Moreover, we prove local limit theorems with large deviations for the coefficients, and large deviation expansions under Cramér's change of probability measure. For the proofs we establish the Hölder regularity of the invariant measure of the Markov chain  $(G_n v / |G_n v|)$  under the changed probability, which is of independent interest. Joint work with I. Grama and Q. Liu.

2) Richard Aoun (American University of Beirut)

Title: Law of large numbers of the spectral radius of random matrix products and related results.

Abstract: The goal of the talk is to expose recent results in random matrix products theory without irreducibility assumptions. The main focus will be to show a law of large numbers for the spectral radius of an i.i.d random walk on the general linear group (joint work with Cagri Sert). The talk will shed light to the relation of such a problem with stationary measures on the projective space. Open questions will be stated.

3) Cagri Sert (University of Zürich)

Title: Counting limit theorems for representations of Gromov-hyperbolic groups.

Abstract: Given a representation  $\rho : \Gamma \rightarrow \mathrm{SL}_d(\mathbb{R})$  of a finitely generated hyperbolic group  $\Gamma$  endowed with a generating set  $S$ , under various assumptions, we obtain some of the classical limit theorems (law of large numbers, central limit theorem, large deviation principles) of probability theory for norms and spectral radii of matrices with respect to counting measures on spheres (or balls) with respect to word-metric induced by  $S$ . The automatic structure of Gromov-hyperbolic groups allows one to turn this seemingly deterministic problem into a problem of establishing limit theorems for additive functionals of a (or rather several) Markov chain(s). We will give a brief overview of the notions, results and the techniques involved. Joint work with I. Cipriano, R. Dougall and S. Cantrell.

• **Développements récents en Bayésien computationnel**, organisateur Nicolas Chopin (ENSAE)

1) Julien Stoeher (Université Paris Dauphine)

Titre: Gibbs-ABC: Component-wise approximate Bayesian computation via Gibbs-like steps (joint work with G. Clarté, C.P. Robert and R. Ryder)

Abstract: Approximate Bayesian computation methods are useful for generative models with intractable likelihoods. These methods are however sensitive to the dimension of the parameter space, requiring exponentially increasing resources as this dimension grows. To tackle this difficulty, we explore a Gibbs version of the ABC approach that runs component-wise approximate Bayesian computation steps aimed at the corresponding conditional posterior

distributions, and based on summary statistics of reduced dimensions. While lacking the standard justifications for the Gibbs sampler, the resulting Markov chain is shown to converge in distribution under some partial independence conditions. The associated stationary distribution can further be shown to be close to the true posterior distribution and some hierarchical versions of the proposed mechanism enjoy a closed form limiting distribution. Experiments also demonstrate the gain in efficiency brought by the Gibbs version over the standard solution

.2) Gabriel Ducrocq (ENSAE, IPP)

Titre: Cube thinning: une méthode de thinning sous contraintes (joint work with Nicolas Chopin)

Résumé: Il est parfois nécessaire de sous-échantillonner les échantillons obtenus par méthodes de Monte-Carlo Markov chain. Cette pratique est appelée thinning. Nous présentons une nouvelle méthode de thinning en nous appuyant sur deux idées: l'utilisation de control variates bien choisies et une méthode d'échantillonnage d'une population sous contrainte, tirée de la littérature sur les sondages. Nous comparons notre méthode aux Kernel Stein thinning développé dans "Optimal Thinning of MCMC Output" (Riabziz et al. 2020) ainsi qu'à la méthode de thinning standard.

3) Avetik Karagulyan (transition entre ENSAE et KAUST)

Titre: Penalized Langevin Dynamics

Abstract: We study the problem of sampling from a probability distribution on  $\mathbb{R}^p$  defined via a convex and smooth potential function. We first consider two continuous-time diffusion-type processes, termed Penalized Langevin dynamics (PLD) and Penalized Kinetic Langevin dynamics (PKLD), the drift of which is the negative gradient of the potential plus a linear penalty that vanishes when time goes to infinity. An upper bound on the Wasserstein-2 distance between the distribution of the PLD at time  $t$  and the target is established. This upper bound highlights the influence of the speed of decay of the penalty on the accuracy of approximation. As a consequence, considering the low-temperature limit we infer a new non-asymptotic guarantee of convergence of the penalized gradient flow for the optimization problem.

4) Julyan Arbel (Inria Grenoble)

Titre: Improving MCMC convergence diagnostic with a local version of  $\hat{R}$

Abstract : Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis. Among the most popular methods, the potential scale reduction factor (commonly named  $\hat{R}$ ) is an indicator that monitors the convergence of all chains to the stationary distribution, based on a comparison of the between- and within-variances of the chains. Several improvements have been suggested since its introduction by Gelman and Rubin (1992). Here, we analyse some properties of the theoretical value  $R$  associated to  $\hat{R}$  in the case of a localized version that focuses on quantiles of the distribution. This leads to proposing a new indicator, which is shown to allow both for localizing the MCMC convergence in different quantiles of the distribution, and at the same time for handling some convergence issues not detected by other  $\hat{R}$  versions.

• **Géométrie stochastique**, organisateur Raphaël Lachièze-Rey (Université de Paris)

1) Pierre Calka (Université de Rouen)

Titre. Simplexes aléatoires et fluctuations d'interfaces convexes

Résumé. L'exposé débutera par une introduction générale au domaine de la géométrie stochastique et sera ensuite focalisé sur des modèles de polytopes aléatoires construits comme enveloppes convexes de nuages de points aléatoires dans l'espace euclidien. Les résultats principaux portent sur les fluctuations maximales dans le sens radial et dans le sens longitudinal de la frontière d'un tel polytope aléatoire, quand la taille du nuage tend vers l'infini. On verra

que cette question est notamment liée à un problème auxiliaire portant sur le volume d'un simplexe aléatoire dans la boule. Il s'agit d'un travail en collaboration avec J. E. Yukich.

2) Thomas Opitz (BioSP, INRAE)

Titre : La géométrie stochastique des ensembles d'excursions des mélanges de processus gaussiens et de leurs limites extrêmes

Résumé : L'étude des ensembles d'excursions d'un processus stochastique spatial au-dessus d'un seuil élevé permet de caractériser les épisodes d'événements extrêmes. Les lois de probabilité de certaines caractéristiques géométriques de ces ensembles, telles que leur surface, leur périmètre ou leur caractéristique d'Euler, nous fournissent des informations utiles sur la structure des agrégats spatiaux de valeurs extrêmes. L'estimation de ces caractéristiques est une approche prometteuse pour acquérir de nouvelles connaissances dans les domaines des applications climatiques, environnementales et écologiques. Dans ce travail, nous considérons des séries de champs gaussiens stationnaires, indépendants et identiquement distribués, avec une variable aléatoire pour la moyenne ou la variance du champ. Ainsi, nous obtenons les champs de type "Gaussian location mixture" ou "Gaussian scale mixture", respectivement. Les membres de cette classe de processus ont été utilisés pour la modélisation flexible de données spatiales répliquées. De plus, les modèles standard en théorie des valeurs extrêmes pour les dépassements de seuil au-dessus d'un seuil élevé, connus sous le nom de processus de Pareto généralisés, sont obtenus dans la limite lorsque le seuil tend vers l'infini.

En exploitant les nombreux résultats théoriques disponibles pour les ensembles d'excursions des champs gaussiens, nous proposons d'étendre ces résultats aux mélanges de processus gaussiens et à leur processus de Pareto limites. En particulier, nous dérivons les formules pour les espérances de certaines caractéristiques géométriques. Finalement, l'estimation de telles caractéristiques sera illustrée sur des données climatiques.

3) Gilles Bonnet (Ruhr-Universität Bochum)

Titre: Convergence faible du processus de points formé par l'intersection d'hyperplans Poissonien

Résumé: Je présenterai les résultats de l'article intitulé comme cet exposé (arXiv:2007.06398, travail conjoint avec A. Baci et C. Thäle). Cet article traite du processus de points d'intersection d'un processus d'hyperplans de Poisson stationnaire et isotrope dans  $\mathbb{R}^d$  d'intensité  $t > 0$ , où seuls les hyperplans qui intersectent une boule centrée de rayon  $R > 0$  sont considérés. En prenant  $R = t^{-\frac{d}{d+1}}$ , on montre que ce processus ponctuel converge en distribution, lorsque  $t \rightarrow \infty$ , vers un processus ponctuel de Poisson sur  $\mathbb{R}^d \setminus \{0\}$  dont la mesure d'intensité a une densité de loi proportionnelle à  $\|x\|^{-(d+1)}$  par rapport à la mesure de Lebesgue. Une borne sur la vitesse de convergence en termes de la mesure de distance de Kantorovich-Rubinstein est également fournie. Pour obtenir notre résultat nous appliquons un théorème général d'approximation fonctionnelle de Poisson sur des espaces abstraits. Nous présentons également des implications sur la convergence faible de l'enveloppe convexe du processus du point d'intersection et la convergence de son  $f$ -vecteur. Nous réfutons et corrigeons ainsi une conjecture de Devroye et Toussaint [J. Algorithms 14.3 (1993), 381–394] en géométrie computationnelle.

4) Elie Calie (Orange)

Titre : Modélisation des communications device-to-device

Résumé : La communication directe d'un device à un autre, appelée « device-to-device », sera une fonctionnalité importante de la téléphonie mobile de cinquième génération (5G). Elle permettra à la fois d'augmenter la couverture des antennes, et de gagner en capacité en laissant certains flux de données transiter directement entre les devices sans passer par les stations de base. Orange utilise la géométrie stochastique pour modéliser ce nouveau mode de communication. Les modèles utilisés tiennent compte non seulement de la voirie, mais aussi du shadowing, de la mobilité et de la bufferisation.

### 3.4 Sessions 4

- **Modèles bayésiens**, organisateur Zacharie Naulet (Université Paris-Saclay)

- 1) Maxime Vono (Lagrange Mathematics and Computing Research Center Huawei, Paris)

Titre: Augmentation des données asymptotiquement exacte: modèles, algorithmes et théorie

Résumé: Dans cet exposé, je vais présenter un cadre statistique général qui permet d'effectuer naturellement de l'inférence Bayésienne distribuée. Je présenterai un résultat théorique de convergence non-asymptotique et illustrerai les bénéfices de la méthode proposée sur des problèmes de traitement d'images et de Machine Learning.

- 2) Hong-Phuong Dang (ENSAI, Rennes)

Titre: Interaction entre les méthodes bayésiennes et optimisation en utilisant Small-Variance Asymptotics dans le cadre de l'apprentissage de dictionnaire.

Résumé: Dans les modèles probabilistes, les méthodes d'échantillonnage et les approximations variationnelles sont souvent utilisées pour l'inférence. Mais ces méthodes passent difficilement à l'échelle. Une alternative consiste à relâcher le modèle probabiliste dans une formulation non probabiliste et utiliser un algorithme évolutif pour résoudre le problème de minimisation associé. Nous présentons une analyse dite de Small-Variance Asymptotics (SVA) du modèle bayésien qui utilise dans le cadre de l'apprentissage de dictionnaire. Cette approche s'obtient en faisant tendre la variance de la vraisemblance du modèle vers 0. L'analyse montre une interaction entre les méthodes bayésiennes et optimisation. Les résultats illustrent la pertinence de la méthode proposée.

- 3) Thibault Randrianarisoa (Sorbonne Université, LPSM, Paris)

Titre: Optional Pólya trees: vitesses de contraction de la loi a posteriori et quantification de l'erreur

Résumé: En estimation de densité, nous étudions une classe de lois a priori par arbres, les Optional Pólya trees. Nous obtenons des vitesses de contraction de la loi a posteriori qui sont quasi-optimales en norme infinie et s'adaptent à la régularité höldérienne de la densité, pour des valeurs inférieures à 1. Du point de vue de la quantification de l'incertitude, nous obtenons et étudions des bandes de confiance dérivées à partir de bandes de crédibilité et d'autres aspects de la postérieure. Finalement, nous illustrons ces résultats avec des simulations numériques.

- 4) Mariia Vladimirova (Inria Grenoble Rhone-Alpes, Mistis, Grenoble)

Titre: Impact précis de la profondeur sur les lois à priori des unités dans les réseaux de neurones bayésiens

Résumé: Les connaissances théoriques et l'élaboration d'une théorie complète sont souvent le moteur du développement de méthodes nouvelles et améliorées. Les réseaux de neurones sont des modèles puissants, mais ils manquent toujours d'un support théorique complet. L'approche bayésienne appliquée aux réseaux de neurones est considérée comme l'un des cadres les mieux adaptés pour obtenir des explications théoriques et améliorer les modèles. Notre résultat principal est une description précise des queues des lois à priori des unités cachées. Cette description montre que les queues des unités deviennent plus lourdes lorsque l'on va plus en profondeur. Cette découverte éclaire le comportement des unités cachées des réseaux de neurones bayésiens finis. De plus, nous étudions les propriétés de dépendance entre les unités cachées et généralisons les caractérisations des queues aux poids dépendants, montrant une nouvelle perspective sur les distinctions entre les réseaux de neurones bayésiens finis et infinis.

- **Statistiques fonctionnelles**, organisatrices Sophie Dabo (Université de Lille) et Anne-Françoise Yao (Université Clermont-Auvergne)

- 1) Julien Ah Pine Laboratoire ERIC, Université de Lyon.

Title : Multiple kernel SVM for classifying functional data in Sobolev spaces

Abstract: In this contribution we are interested in classifying smooth functional data that belong to Sobolev Spaces. We propose to apply the multiple kernel SVM approach. Our motivation is twofold. Firstly, kernel functions allow us to deal with non linearly separable cases by projecting the data on more appropriate RKHS. Secondly, the successive derivative functions provide different views of the functional data and can help to discriminate distinct classes. In our framework, each set of derivative functions of the same order leads to a kernel matrix and a linear combination of these kernel matrices is estimated during the learning process. We illustrate the application of our proposal on real world datasets.

2) Pamela Llop, Universidad Nacional del Litoral -CONICET

Title: An input/output-based aggregation rule for functional data classification.

Abstract: In [1], the authors introduce a new nonlinear aggregation rule for functional data by aggregating a collection of basic estimators previously calibrated using a training data set. In that work, the authors also introduce a more flexible form of the classifier which discards a small percentage of those preliminary experts (outputs) that are bad initial estimators and behaves differently from the rest. In [2], for the finite dimensional setting, the authors propose an alternative procedure that adds to the classifier information about the inputs, in order to improve the results by discarding data that are not “similar” to the point to be classified. The consistency of that classifier is proved by combining versions of consistency results for kernel classification rule. In this work we combine and extend the ideas from [1] and [2] by creating an aggregation rule for functional data which takes into account input and output information. As previously, the classifier is built from a collection of  $M$  arbitrary training classifiers, taking advantage of the abilities of each expert. Although the method allows to combine classifiers of very different nature, it also performs very well when aggregating experts of the same nature, for instance, when combining different nearest neighbors classifiers. We prove the consistency of the aggregation rule and we show its good performance via simulations studies and also via real data examples.

[1] Cholaquidis, A., Fraiman, R., Kalemkerian, J., Llop, P. (2016). A nonlinear aggregation type classifier. *Journal of Multivariate Analysis*, 146, 269-281.

[2] Fischer, A., Mougeot, M. (2019), Aggregation using input–output trade-off, *Journal of Statistical Planning and Inference*, 200, 1-19.

3) Anne-Françoise Yao Laboratoire LMBP, Université Clermont-Auvergne

Title: Some asymptotic results on kernel regression method for functional data with values in finite dimensional manifolds.

Abstract : This work deals with the problem of kernel regression where the response is a real-valued variable and the regressor is a functional variable. Namely, we investigate the case where the regressor lies in a finite dimensional submanifold of a Hilbert space. We discuss the choice the kernel which is a central issue in this setting and give some asymptotic results of the estimator when the observations are independent identically distributed. We motivate the practical interest of such an estimator through some applications.

• **Percolation**, organisateur Sébastien Martineau (Sorbonne Université)

1) Irène Marcovici, Institut Elie Cartan de Lorraine, Université de Lorraine

Titre : Automates cellulaires et perturbations aléatoires

Résumé : Les automates cellulaires sont des systèmes dynamiques pour lesquels le temps et l'espace sont discrets. Ils permettent de modéliser l'évolution d'un ensemble de « cellules » interagissant entre elles de manière locale : au cours du temps, chacune actualise son état en fonction de ce qu'elle perçoit de son voisinage. La règle locale qui définit le système peut mener à une grande complexité de comportements macroscopiques, qui dépendent souvent fortement de l'état initial. Cependant, quand on perturbe un automate cellulaire par un bruit aléatoire, on s'attend généralement à ce que les erreurs « percolent » et effacent progres-

sivement toute information contenue dans la configuration initiale, entraînant l'ergodicité du système. Lorsque le bruit est suffisamment élevé, un couplage avec un processus de percolation orientée dans le diagramme espace-temps permet effectivement de montrer l'ergodicité. Mais lorsque le bruit est faible, l'ergodicité est souvent difficile à prouver. Je montrerai comment d'autres comparaisons avec la percolation permettent néanmoins de démontrer l'ergodicité en présence d'une petite perturbation aléatoire, pour certaines familles spécifiques d'automates cellulaires. Nous nous intéresserons aussi à la capacité de certains automates cellulaires de « réparer » des erreurs présentes seulement dans la configuration initiale. L'exposé sera basé sur des travaux effectués en collaboration avec Nazim Fatès, Mathieu Sablik, Siamak Taati.

2) Mendes Oulamara (IHÉS & Université Paris-Saclay)

Titre : Universalité en percolation

Résumé : La transformation de Yang-Baxter induit des couplages locaux entre les configurations de percolation sur deux graphes isoradiaux différents. Nous verrons comment on peut déduire des résultats d'universalité et d'invariance à partir de ces couplages.

3) Ivailo Hartarsky (CEREMADE, Université Paris-Dauphine, PSL)

Titre : Weakly constrained-degree percolation

Abstract: In constrained-degree percolation we consider the  $d$ -dimensional hypercubic lattice and supply each edge with a clock. Each clock rings only once at an independent and identically distributed time. When the clock of an edge rings, it becomes open, provided that there are currently at most  $k - 1$  open edges incident with each of its vertices, where  $2 \leq k \leq 2d$  is a parameter of the model. The study is made difficult by the lack of most key properties of more conventional percolation models such as monotonicity in the underlying graph, finite energy, FKG inequality, finite dependence range, etc. We provide the first non-perturbative argument for this model. It allows us to determine the asymptotic value of the critical time as  $d$  and  $k$  diverge, but also to give quantitative upper bounds for any  $d \geq 3$  and most values of  $k$ . This establishes the conjectured non-triviality of the phase transition of emergence of infinite open clusters for a wide range of parameters. The talk is based on joint work with Bernardo de Lima available at <https://arxiv.org/abs/2010.08955>.

4) Jérôme Casse (LMO, Université Paris-Saclay)

Titre : "Percolation de dernier passage généralisée : étude sur le cylindre" Résumé : La percolation de dernier passage dirigée est, classiquement, un modèle de croissance dans le quart de plan discret. Pour croître de la case  $(i, j)$ , il faut que les cases  $(i - 1, j)$  et  $(i, j - 1)$  soient présentes dans notre amas de croissance, puis attendre un temps aléatoire  $\tau(i, j)$ . Ce modèle est notamment intéressant pour modéliser le temps d'assèchement d'un terrain. Dans cet exposé, je présente une généralisation de la percolation de dernier passage dirigée dans le cas où le temps à attendre  $\tau(i, j)$  dépend des temps d'arrivée des cases  $(i - 1, j)$  et  $(i, j - 1)$  dans l'amas et je présente ce modèle non pas comme un modèle de croissance dans le quart de plan, mais dans un cylindre de taille  $L$ . Dans le cylindre, il apparaît ainsi une ligne de front pour notre amas. L'objet de cet exposé va être d'étudier deux propriétés asymptotiques (en temps) de cette ligne de front: sa vitesse et sa forme. Nous verrons que, dans des cas particuliers dits solubles ou intégrables, cette vitesse et cette forme ont une forme explicite en fonction des paramètres du modèle. Puis, j'expliquerai par quelle magie ces cas sont solubles, alors que les autres ne le sont a priori pas.

### 3.5 Sessions 5

- **Statistique et Géométrie**, organisateur Clément Levrard (Université de Paris)

1) Théo Lacombe:

Title: Averaging Radon measures: an Optimal Transport perspective.

Abstract: This talk will start by introducing an important metric over spaces of probability measures: the so-called Wasserstein distance. This metric, which arises from Optimal Transport theory, leads to consider interesting yet challenging problems, both of theoretical and computational nature. We will focus on the one of computing Fréchet means (a.k.a. barycenters) of samples of probability measures, and will discuss how it can be extended to general Radon measures (i.e. non-negative measures with total mass non necessarily equal to 1) through a variant of Optimal Transport due to Figalli and Gigli (2010).

2) Vincent Divol

Title: Summarizing the topology of complex datasets with (expected) persistence diagrams

Abstract: The goal of the talk is twofold. First, we will take some time to explain what are persistence diagrams: these are summaries describing in a multi-scale fashion the topology of some object (e.g. a graph, a point cloud or a time series), that are routinely used in the field of Topological Data Analysis. They consist in finite point sets in the plane. In particular, they do not naturally belong to a vector space, a fact that may hinder their use in standard machine learning pipelines. We will present two techniques to bypass this problem: the use of feature maps on the space of persistence diagrams, and quantization algorithms.

3) Alice Le Brigant

Title: Classifying histograms using the Fisher-Rao geometry of beta distributions

Abstract : These last few decades, Riemannian geometry has proved to be a powerful tool to deal with non linear data. When the data consists of probability distributions, the Fisher-Rao metric provides such a Riemannian framework. In this talk, we will see how to compare, average and classify beta distributions using the Fisher-Rao metric. We show that this geometry is negatively curved, which makes it a suitable framework to classify distributions through the popular K-means algorithm. We illustrate the use of these geometric tools for classification of histograms of medical data.

• **Probabilités et Statistiques pour la biologie**, organisatrice Camille Coron (Université Paris-Saclay)

1) Arnaud Personne

Titre : Comment modéliser l'évolution d'une forêt ? Le hasard peut-il à lui seul expliquer la répartition des individus sur une parcelle de forêt ?

Résumé : En 2001, Hubbell présente un modèle de dynamique des populations dit « neutre » visant à décrire l'évolution d'une forêt dans le temps. Dans ce modèle toutes les espèces sont d'égales compétitrices et la composition de la communauté est déterminée uniquement par la dispersion stochastique des individus. Nous nous intéresserons à une extension de ce modèle dans lequel chaque espèce possède un avantage sélectif caractérisé par un paramètre aléatoire (processus de saut, diffusion...). Ce modèle, largement utilisé en génétique des populations et connu sous le nom de Modèle de Moran en environnement aléatoire. Il s'agit d'un modèle Markovien dont la dynamique peut être approchée quand la taille de la communauté étudiée est assez grande par des processus continus en temps (diffusion de Wright-Fisher ou pdmp). Nous donnerons un ordre de grandeur de l'erreur commise en effectuant cette approximation diffusion pour une taille de population fixée. Ensuite nous nous servirons de ce résultat pour obtenir des informations sur le processus discret à partir du processus continu: approximation des moments, comportement en temps long, persistance...

2) Aline Marguet

Titre : Comportement asymptotique d'une population de cellules infectée par des parasites  
Résumé : On considère une classe générale de processus de Markov branchants, pour la modélisation de la prolifération d'une infection par des parasites d'une population de cellules. Chaque cellule contient une quantité de parasites évoluant suivant une diffusion avec sauts positifs, dont les différents taux (croissance, diffusion et saut) sont des fonctions de la quantité de parasites. Le taux de division d'une cellule dépend également de la quantité de parasites qu'elle

contient. À la division, une cellule donne naissance à deux cellules filles et partage ses parasites entre ses descendants. Une cellule peut également mourir à un taux dépendant de la quantité de parasites qu'elle contient. Nous présenterons le comportement asymptotique de l'infection dans la population de cellules. En particulier, nous nous intéresserons à la quantité de parasites d'une cellule "typique", ainsi qu'à la question de la survie de la population de cellules. Nous étudierons l'influence de la loi de répartition des parasites à la division ainsi que celle de la forme des taux de division et de mort des cellules comme fonction de la quantité de parasites qu'elles contiennent.

3) Veronica Miro Pina

Title: Xi-coalescents arising in population models with bottlenecks.

Abstract: In this work, we study a family of Xi-coalescents that arise from a class of Wright-Fisher models with recurrent demographic bottlenecks. This family of Xi-coalescents constructed from i.i.d mass, is an extension of the symmetric coalescent (Gonzalez Casanova et al. 2021). We study the process which counts the frequency of blocks of each size in a n-coalescent, when n tends to infinity. We show that the multivariate Lamperti transform of this process is a Markov Additive Process (MAP). This allows us to provide some asymptotics for the length of order r in a n-coalescent, which is defined as the sum of the lengths of all the branches that carry a subtree with r leaves. Joint work with: Adrián González Casanova, Arno Siri-Jégousse and Emmanuel Schertzer.

4) Marie Perrot-Dockes Title: Post hoc false discovery proportion inference under a Hidden Markov Model

Abstract: We address the multiple testing problem under the assumption that the true/false hypotheses are driven by a Hidden Markov Model (HMM), which is recognized as a fundamental setting to model multiple testing under dependence since the seminal work of Sun and Cai (2009). While previous work has concentrated on deriving specific procedures with a controlled False Discovery Rate (FDR) under this model, following a recent trend in selective inference, we consider the problem of establishing confidence bounds on the false discovery proportion (FDP), for a user-selected set of hypotheses that can depend on the observed data in an arbitrary way. We develop a methodology to construct such confidence bounds first when the HMM model is known, then when its parameters are unknown and estimated, including the data distribution under the null and the alternative, using a nonparametric approach. In the latter case, we propose a bootstrap-based methodology to take into account the effect of parameter estimation error. We show that taking advantage of the assumed HMM structure allows for a substantial improvement of confidence bound sharpness over existing agnostic (structure-free) methods, as witnessed both via numerical experiments and real data examples.

• **Apprentissage et réseaux de neurones**, organisateur Edouard Oyallon (Sorbonne Université)

1) Nicolas Keriven, CNRS, GIPSA-lab,

Title : Graph Neural Networks on Large Random Graphs: Convergence, Stability, Universality

2) Michael Arbel, INRIA Grenoble

Title: “Annealed Flow Transport Monte Carlo”

Abstract: “Annealed Importance Sampling (AIS) and its Sequential Monte Carlo (SMC) extensions are state-of-the-art methods for estimating normalizing constants of probability distributions. We propose here a novel Monte Carlo algorithm, Annealed Flow Transport (AFT), that builds upon AIS and SMC and combines them with normalizing flows (NFs) for improved performance. This method transports a set of particles using not only importance sampling (IS), Markov chain Monte Carlo (MCMC) and resampling steps - as in SMC, but also relies on NFs which are learned sequentially to push particles towards the successive



annealed targets. We provide limit theorems for the resulting Monte Carlo estimates of the normalizing constant and expectations with respect to the target distribution. Additionally, we show that a continuous-time scaling limit of the population version of AFT is given by a Feynman–Kac measure which simplifies to the law of a controlled diffusion for expressive NFs. We demonstrate experimentally the benefits and limitations of our methodology on a variety of applications."

3) Joakim Andén, KTH

Title: Interpretable neural networks for audio classification using joint time–frequency scattering

Abstract: Convolutional networks have enjoyed great success extracting relevant structural information for a range of data, including audio signals. However, since their filters are learned from optimizing over a training set, analyzing their properties and predicting their behavior poses difficulties. We consider the joint time-frequency scattering network, a convolutional network with fixed wavelet filters that performs comparably to fully learned networks for several audio classification tasks. The fixed nature of the network lets us analyze how it captures different time-frequency structures, including amplitude modulation, harmonic structure, and frequency modulation. These results illustrate the expressive power of the scattering network in representing commonly occurring structures.

4) Julien Rabin, ENSICAEN

Title: Une approche hiérarchique pour le transport optimal semi-discret appliqué à la synthèse de texture

Abstract : Dans cet exposé nous nous intéressons à l'estimation des paramètres du transport optimal dans le cadre semi-discret, c'est-à-dire d'une mesure à densité vers une distribution discrète. L'application principalement visée est la synthèse d'images où, à la manière d'un réseau de neurones, des images aléatoires sont itérativement transformées localement par des plans de transport, dont les paramètres sont appris sur une base de données, garantissant ainsi les propriétés statistiques des images générées. Afin d'obtenir des résultats visuellement satisfaisant, il est nécessaire de considérer des voisinages (patches) de grandes tailles, ce qui augmente la dimension du problème étudié et rend difficile l'utilisation des algorithmes conventionnels. Pour pallier ce problème, nous proposons une approche hiérarchique multi-niveaux qui, au contraire des approches multi-échelle proposées précédemment, permet d'obtenir de manière plus rapide une estimation approchée de ces plans de transport. Ce travail est en collaboration avec Arthur Leclaire (IMB, Bordeaux) et Bruno Galerne (Institut Denis Poisson).

### 3.6 Sessions 6

• **Statistique et image**, organisateur Alasdair Newson (Télécom Paris)

1) Claire Launay :

Titre: Processus pixelliques déterminantaux : Synthèse de textures et inférence.

Résumé: Les processus ponctuels déterminantaux (DPP) sont des modèles probabilistes qui modélisent les corrélations négatives ou la répulsion à l'intérieur d'un ensemble d'éléments. Ils ont tendance à générer des sous-ensembles d'éléments diversifiés ou éloignés les uns des autres. Cette notion de similarité ou de proximité entre les points de l'ensemble est définie et conservée dans le noyau associé à chaque DPP. Ces processus ont récemment gagné en influence dans la communauté de Machine Learning, principalement grâce à leur capacité à sous-échantillonner de grands ensembles de données de manière élégante et efficace. Lors de cette présentation, je m'intéresserai à l'application des DPP à des questions de traitement d'images, lorsque l'ensemble des points de départ correspond aux pixels d'une image. Dans ce cadre discret et fini, nous avons adapté la définition et les propriétés des DPPs que nous appelons processus pixelliques déterminantaux (DPixP). Ce nouveau cadre implique aussi de

prendre des hypothèses de périodicité et de stationnarité qui ont des conséquences sur le noyau du processus. Nous nous sommes intéressés aux propriétés de répulsions de tels processus et à leur application à la synthèse de textures gaussiennes, en utilisant les modèles shot noise. Nous avons également étudié l'estimation du noyau d'un DPixP à partir d'un ou de plusieurs échantillons.

2) Remi Laumont :

Titre: Approche Bayésienne pour la restauration d'images utilisant des a-prioris Plug & Play: quand Langevin rencontre Tweedie

Resumé : Depuis les travaux de Venkatakrishnan et al. [2] en 2013, les méthodes Plug & Play (PnP) sont souvent appliquées pour la restauration d'image dans un contexte Bayésien. Ces méthodes visent à calculer les estimateurs Minimum Mean Square Error (MMSE) ou Maximum A Posteriori (MAP) pour des problèmes inverses en imagerie en combinant une vraisemblance explicite et un a-priori implicite défini par un algorithme de débruitage. Dans la littérature, les méthodes PnP diffèrent principalement par le schéma itératif utilisé que cela soit pour l'échantillonnage ou l'optimisation. Dans le cas des algorithmes d'optimisation, des travaux récents garantissent la convergence vers un point fixe d'un certain opérateur, point fixe qui n'est pas nécessairement le MAP. Dans le cas des algorithmes d'échantillonnage de la littérature, à notre connaissance, il n'existe pas de preuves de convergence. Par ailleurs, il reste d'importantes questions ouvertes portant sur la bonne définition des modèles Bayésiens sous-jacents ou encore des estimateurs calculés, ainsi que leurs propriétés de régularité, nécessaires pour assurer la stabilité du schéma numérique. L'un des algorithmes que nous présentons, PnP-ULA (Plug & Play Unadjusted Langevin Algorithm) a été développé afin de répondre à ces questions. Cet algorithme permet d'échantillonner l'a-posteriori. Exploitant des résultats récents de convergence sur les chaînes de Markov, nous donnons des conditions de convergence pour cet algorithme sous des hypothèses réalistes notamment concernant l'opérateur de débruitage. Enfin, nous montrons l'efficacité de notre méthode sur des problèmes inverses classiques d'imagerie tels que le défloutage, l'interpolation ou le débruitage. En outre, nous fournissons une première étude d'incertitude sur les résultats obtenus. La référence de l'article relatif à cette présentation est [1].

Références [1] Rémi Laumont, Valentin de Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie. <https://arxiv.org/pdf/2103.04715.pdf>, 2021. [2] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. Plug-and-play priors for model based reconstruction. In 2013 IEEE Global Conference on Signal and Information Processing, pages 945–948. IEEE, 2013.

3) Lara Raad

Conditional Gaussian patch models for texture synthesis

Exemplar-based texture synthesis is defined as the process of generating, from an input texture sample, new texture images that are perceptually equivalent to the input. We model texture self-similarities with conditional Gaussian distributions in the patch space. These models are then embedded in a patch-based framework for texture synthesis. The Gaussian distribution for each patch is inferred from the set of its nearest neighbors in the patch space obtained from the input sample. In patch-based approaches one of the difficulties is putting the patches back together. We will compare a quilting technique [1], where every new patch overlaps the previously synthesized one and the overlapped parts are blended together across the minima error path, to conditional Gaussian models [2] constraining the simulated Gaussian patches to respect the values of its overlapping zone. The results show that it is possible to dispose of the quilting step in the case of periodic and pseudo-periodic textures. However, for more complex textures, the conditional models are less performing and are fast limited by the size of the texture input sample.

[1] Efros, Alexei A., and William T. Freeman. "Image quilting for texture synthesis and

transfer." Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001. [2] Raad, Lara, Agnès Desolneux, and Jean-Michel Morel. "A conditional multiscale locally Gaussian texture synthesis algorithm." Journal of Mathematical Imaging and Vision 56.2 (2016): 260-279.

4) Jonathan Vacher

Flexibly Regularized Mixture Models and Application to Image Segmentation

Abstract

Probabilistic finite mixture models are widely used for unsupervised clustering. These models can often be improved by adapting them to the topology of the data. For instance, in order to classify spatially adjacent data points similarly, it is common to introduce a Laplacian constraint on the posterior probability that each data point belongs to a class. Alternatively, the mixing probabilities can be treated as free parameters, while assuming Gauss-Markov or more complex priors to regularize those mixing probabilities. However, these approaches are constrained by the shape of the prior and often lead to complicated or intractable inference. Here, we propose a new parametrization of the Dirichlet distribution to flexibly regularize the mixing probabilities of over-parametrized mixture distributions. Using the Expectation-Maximization algorithm, we show that our approach allows us to define any linear update rule for the mixing probabilities, including spatial smoothing regularization as a special case. We then show that this flexible design can be extended to share class information between multiple mixture models. We apply our algorithm to artificial and natural image segmentation tasks, and we provide quantitative and qualitative comparison of the performance of Gaussian and Student-t mixtures on the Berkeley Segmentation Dataset. We also demonstrate how to propagate class information across the layers of deep convolutional neural networks in a probabilistically optimal way, suggesting a new interpretation for feedback signals in biological visual systems. Our flexible approach can be easily generalized to adapt probabilistic mixture models to arbitrary data topologies.

• **Marches aléatoires/Milieux aléatoires**, organisateur Alexis Devulder (Université Versailles Saint-Quentin)

1) Roland Diel

Titre : Estimation non paramétrique pour une marche aléatoire en milieu aléatoire.

Résumé : On considère une marche aléatoire évoluant dans un environnement i.i.d sur  $\mathbb{Z}$ . Dans cet exposé, nous nous intéresserons à l'estimation non paramétrique de la loi de l'environnement. Nous verrons d'abord comment construire des estimateurs pour les moments de l'environnement puis comment utiliser ces estimateurs pour approcher la fonction de répartition de la loi. Nous verrons ensuite comment la vitesse asymptotique de la marche influe sur la vitesse de convergence de ces estimateurs.

2) Xiaolin Zeng

Titre : Loi de Dirichlet sur un graphe décomposable

Résumé : La loi de Dirichlet provient de l'urne de Dirichlet; de plus la marche aléatoire en environnement de Dirichlet correspond à la marche renforcée dirigée. Nous allons donner une généralisation de la loi de Dirichlet sur un graphe fini et décomposable, de façon que la loi de Dirichlet classique correspond au cas du graphe complet. Si le temps le permet, nous allons donner une interprétation probabiliste de cette loi, via des randonnées aléatoires. Savoir si cette généralisation provient d'une marche renforcée reste une question ouverte. Travail en collaboration avec J. Wesolowski et B. Kolodziejek.

3) Xinxin Chen

Titre : Fixed points for branching Brownian motions. Joint work with C. Garban and A. Shekhar [Université Claude Bernard Lyon 1]

Abstract : We consider a particle system by attaching to each atom of some point process  $\theta$  an independent branching Brownian motions (BBM) with critical drift  $-\sqrt{2}$  and study all

point processes which are left invariant. Under the assumption that  $\theta(\mathbb{R}_+) < \infty$  a.s., we show that all fixed points are distributed as the extremal point process of BBM with some random shift.

• **Arbres aléatoires**, organisateur Jean-François Delmas (ENPC)

1) Igor Kortchemski

Titre : Limites de grands arbres aléatoires.

Résumé: À quoi un grand arbre aléatoire de Bienaymé-Galton-Watson ressemble-t-il ? Il s'agira de dresser un panorama impressionniste de résultats récents qui visent à répondre à cette question.

2) Alice Contat

Titre : Une surprenante symétrie pour le greedy independent set sur les arbres de Cayley.

Résumé : Nous prouvons que la taille  $G_n$  du greedy independent set sur un arbre de Cayley uniforme a (presque) la même loi que la taille de son complémentaire. En particulier, nous montrerons que  $G_n$  a la même loi que le nombre de sommets à hauteur paire dans un arbre de Cayley. Nous montrerons aussi comment explorer  $G_n$  de façon markovienne.

3) Michel Nassif

Titre: Scaling limits of general additive functionals of conditioned Galton-Watson trees.

Abstract: We study the scaling limits of general additive functionals on conditioned Galton-Watson trees of the form

$$F(\mathbf{t}) = \sum_{v \in \mathbf{t}} f(\mathbf{t}_v),$$

where  $\mathbf{t}_v$  is the subtree of  $\mathbf{t}$  above vertex  $v$  and  $f$  is a given toll function. Many usual indices can be written in that form, e.g. the total path length, the Wiener index, and Shao and Sokal's  $B_1$  index. We assume that the offspring distribution is critical with finite or infinite variance. We express the limit as a functional of the stable Lévy tree and study some of its properties. We also describe a phase transition when the toll function depends only on the size and the height of the tree.

4) Jean-Jil Duchamps

Title: The Moran Forest (joint work with F. Bienvenu and F. Foutel-Rodier)

Abstract: Consider the random graph obtained as the stationary distribution of the following Markov chain on graphs labelled with  $1, \dots, n$ : at each step, a uniformly chosen vertex is disconnected from its neighbors and reconnected to another uniformly chosen vertex. This random graph is the forest that records all parent-offspring pairs that are alive in a population evolving according to the Moran process. The Moran forest can be built from two different simple constructions, involving either uniform attachment trees or the uniform tree on  $1, \dots, n$ , that enable us to derive some of its characteristics : distribution of degrees, of a uniformly chosen tree, of the largest tree.