Journées MAS 2016

Phénomènes complexes et hétérogènes

Grenoble

29 - 31 août 2016

Présentation

Depuis 1996, les journées MAS du groupe Modélisation Aléatoire et Statistique de la Société de Mathématiques Appliquées et Industrielles (SMAI) offrent à la communauté des probabilistes et statisticiens français l'opportunité de se rassembler tous les deux ans pour présenter leurs travaux récents. Les journées MAS 2016 seront centrées sur le thème Phénomènes complexes et Hétérogènes.

Ces rencontres comportent plusieurs volets :

- Six conférences plénières centrées sur le thème affiché pour cette édition.

  Vingt sessions parallèles donnant un aperçu des recherches actuelles en France dans le domaine de la modélisation aléatoire et de la statistique sur les thématiques porteuses de probabilités et statistiques, en lien avec les applications et le monde industriel (optimisation stochastique, statistique en grande dimension, modèles probabilistes en biologie, industrie et énergies renouvelables,...).
- Présentation et exposés des lauréats 2014 et 2015 du prix de thèse Jacques Neveu.
- Session posters le premier jour de la conférence. Il n'y a pas de sélection des posters en amont. Un prix "Springer" sera décerné au(x) meilleur(s) poster(s).

Enfin, les journées MAS 2016 sont l'occasion de rendre hommage à Jacques Neveu, entre autres à l'initiative de la création du groupe MAS, récemment décédé.

Nous sommes très heureux de vous accueillir à Grenoble pour cette nouvelle édition et au nom de l'ensemble des membres du comité d'organisation, nous vous souhaitons une excellente conférence.

Jean-François Coeurjolly,
pour le comité d'organisation.

Comité scientifique

Nous sommes très redevables aux membres du comité scientifique pour leur travail de proposition, et la sélection réalisée pour ces journées :

- Bernard Bercu, Université de Bordeaux - Bordeaux INP.
- Francis Comets (Président), Université Paris Diderot.
- Anne Estrade, Université Paris Descartes.
- Anne-Laure Fougères, Université Lyon 1.
- Sylvie Méléard, Ecole Polytechnique.
- Didier Piau, Université de Grenoble Alpes.
- Jean-Michel Poggi, Université Paris-Sud et Université Paris Descartes
- Patricia Reynaud-Bouret, CNRS, Université de Nice Sophia Antipolis.

Comité d'organisation

- Pierre-Olivier Amblard, Gipsa-lab,
- Frédéric Audra, LJK (soutien info),
- Fanny Bastien, IF (infographiste),
- Jean-François Coeurjolly, LJK (Président),
- Juana Dos Santos, Assistante (de choc), LJK,
- Stéphane Girard, LJK,
- Sophie Lambert-Lacroix, TIMC,
- Adeline Leclercq-Samson, LJK,
- Jérôme Lelong, LJK (webmestre),
- Clémentine Prieur, LJK,
- Raphaël Rossignol, IF,
- Laurent Zwald, LJK,

et le soutien de la SMAI

- Nadia Atek (secrétaire SMAI),
- Emmanuel Gobet, Ecole Polytechnique (trésorier SMAI).

Sponsors et partenaires

Nous remercions chaleureusement les organismes et institutions suivantes ayant supporté financièrement la préparation des journées MAS :

Accès, hébergement et informations pratiques

Accès à Grenoble :
par train : La gare de Grenoble se trouve à quelques minutes du centre-ville et est desservie par plusieurs lignes de tram. Si vous arrivez depuis Chambéry par le train, il est préférable de descendre à l'arrêt Gières-Universités qui se trouve à deux stations de tram du campus universitaire (et en particulier de l'arrêt "Condillac-Universités").
par avion : L'aéroport le plus proche est celui de Lyon Saint-Exupéry. Pour se rendre à Grenoble, le plus simple est ensuite d'utiliser la navette bus (34.5 euros aller/retour) et descendre au terminus (gare routière de Grenoble, 100m de la gare ferroviaire).

Accès au campus universitaire :
Depuis le centre ville de Grenoble, des gares ferroviaire ou routière, le plus simple pour se rendre sur le site des journées MAS est de prendre le tram B Direction Gières plaine des sports, et de descendre à l'arrêt "Condillac-Universités" (depuis la gare, il faut compter environ 25 min.). Ensuite, il faut revenir sur ses pas, et traverser la route. Les amphis du hall sud de Stendhal sont juste en face.

Hébergement :
Une suggestion d'hébergements est en lien sur le site de la conférence `http://mas2016.sciencesconf.org/resource/acces`. Quelques hôtels sont situés sur le campus universitaire. Il est néanmoins conseillé, afin de profiter pleinement de la ville, de trouver un hôtel au centre, l'accès à la conférence par le tram rendant la chose aisée.

Lieu de la conférence :
Les journées MAS se dérouleront dans le hall sud de Stendhal. Quatre amphis situés au même endroit sont dédiés à ces journées. Les conférences plénières, exposés des prix Neveu ainsi que l'assemblée générale se dérouleront dans l'amphi 11. Les sessions parallèles auront lieu dans les amphis 7, 8, 9 et 11. L'accueil, les inscriptions, les pause-cafés ainsi que le cocktail de bienvenue et sa session poster, se dérouleront également dans ce hall sud. Les repas de midi se feront au restaurant universitaire Diderot (à 500m du hall sud). Pensez à vous munir des contremarques délivrées lors de votre enregistrement. Ci-après une carte plus précise, positionnant ces différents éléments.

Connexion Wifi :
Le réseau `eduroam` est accessible sur tout le campus universitaire. Lors de votre enregistrement vous avez également reçu un identifiant et mot de passe individuel pour vous connecter en tant que "visiteur" sur le réseau `wifi-campus`.

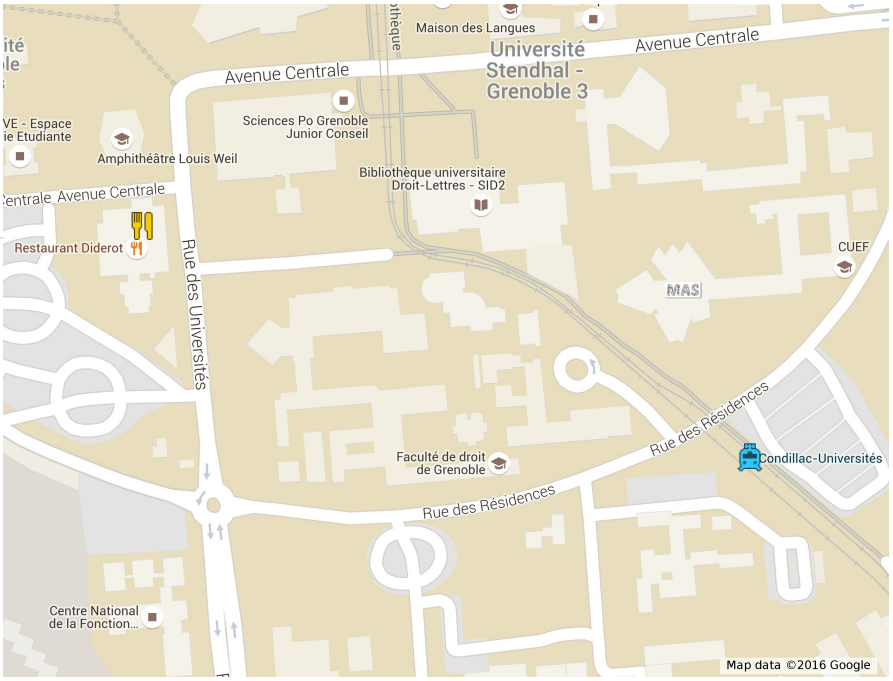
Dîner de conférences :
Le dîner de conférences aura lieu au restaurant le "Téléférique" au sommet des bulles de Grenoble, à la Bastille. Le rendez-vous est 20:00 au restaurant. Le ticket délivré lors de votre enregistrement vous donne accès aux bulles (voir l'image ci-après). Pour les plus courageux, il faut monter 220m de dénivelé par un sympathique chemin accessible depuis la porte Saint-Laurent, succession d'escaliers et chemins agréables (record à 8'30"!!). Le retour se fera en bulles également, vers 22:30 - 23:00 ou à pied selon les possibilités.

Lieu des journées MAS

Hall Sud - Stendhal, conférence

Restaurant Diderot, J. Mas 2016

Condillac-Universités

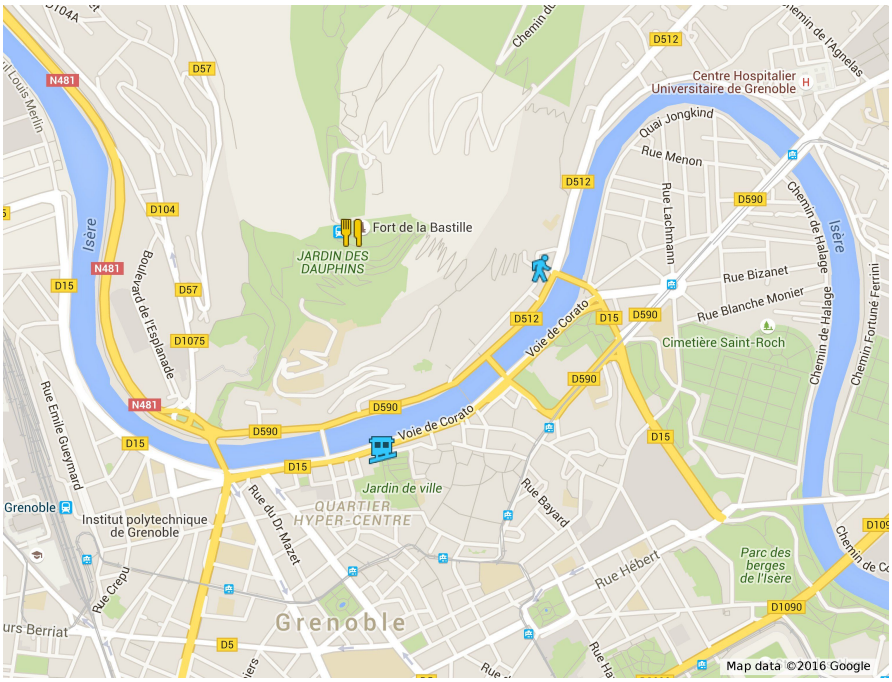


Dîner de conférence

Téléphérique Grenoble-Bastille

Porte Saint-Laurent

Restaurant le Téléférique



Résumés des conférences plénières (par ordre de programmation)

1. Aurélien Ribes (Centre National de Recherches Météorologiques)
   Peut-on prévoir le climat du futur avec des statistiques ?
2. François Delarue (Université Nice Sophia Antipolis)
   Jeux à champ moyen.
3. Amaury Lambert (Université Pierre et Marie Curie)
   A non-exchangeable coalescent arising in phylogenetics.
4. Sylvie Huet (MaIAGE INRA-Jouy-en-Josas)
   Méthodes pénalisées pour la méta-modélisation et l'analyse de sensibilité.
5. Charles Bordenave (CNRS, Université de Toulouse)
   Marche aléatoire sur un graphe aléatoire.
6. Eva Löcherbach (Université de Cergy Pontoise)
   Spiking neurons : interacting processes with memory of variable or infinite length.

1 – Peut-on prévoir le climat du futur avec des statistiques ?

Aurélien Ribes, Centre National de Recherches Météorologiques

En 2016, le changement climatique induit par les activités humaines, au premier rang desquelles les émissions anthropiques de gaz à effet serre, est un phénomène physique de mieux en mieux compris et décrit. C'est également un sujet de préoccupation qui s'est invité dans l'agenda politique. Cependant, malgré des modèles numériques de plus en plus précis et gourmands en moyens de calcul, de larges incertitudes persistent dans l'estimation de la sensibilité du climat planétaire à une augmentation de l'effet de serre. Une certaine compétition est ainsi en train d'émerger entre la modélisation physique, d'une part, et l'étude statistique des quelques 150 ans de données disponibiles, d'autre part, pour réduire nos incertitudes sur le climat de demain.

Après un rapide état des lieux sur les connaissances actuelles sur le changement climatique, je présenterai les modèles et méthodes statistiques principalement utilisées dans la communauté afin de traiter cette question - techniques généralement regroupées sous le nom de détection et attribution du changement climatique. Parmi eux se trouvent notamment des modèles de régression sur variables entachées d'erreur (error in variables) originaux. J'aborderai deux problématiques liées à l'inférence dans ces modèles : l'estimation de matrices de variance de grande dimension, et l'estimation - délicate car reposant sur des échantillons non indépendants - des incertitudes des modèles physiques de climat. Enfin, je discuterai brièvement des synergies possibles entre statistiques et sciences du climat.

2 – Jeux à champ moyen

François Delarue, Université Nice Sophia Antipolis

Les jeux à champ moyen ont été introduits il y a une dizaine d'années par Lasry et Lions pour décrire, asymptotiquement, les états d'équilibre au sein de grandes populations de particules ou d'agents économiques en interaction, chacune ou chacun cherchant à minimiser une fonctionnelle d'énergie ou de coût propre ; ici, le mot "asymptotique" renvoie à la limite prise sur le nombre de particules ou d'agents. L'analyse mathématique comporte plusieurs enjeux, parmi lesquels l'existence d'équilibres asymptotiques, compris comme des lois de probabilité

décrivant la distribution de la population à l'équilibre, l'unicité de tels équilibres et, enfin, le lien entre modèles asymptotiques et modèles non-asymptotiques. Dans cette perspective, j'insisterai sur la notion d'"équation maîtresse" introduite par Lions. L'équation maîtresse est une équation aux dérivées partielles posée sur l'espace des probabilités, dont la solution caractérise l'état de la population à l'équilibre. Je me focaliserai sur les propriétés de cette équation et sur l'utilisation qui peut en être faite pour passer du cadre non-asymptotique au cadre asymptotique.

3 – A non-exchangeable coalescent arising in phylogenetics

Amaury Lambert, Université Pierre et Marie Curie

A popular line of research in evolutionary biology is to use time-calibrated phylogenies in order to infer the underlying process of species diversification. Most models of diversification assume that species are exchangeable and lead to phylogenetic trees whose shape is the same in distribution as that of a Yule pure-birth tree. Here, we propose a non-exchangeable, individual-based, point mutation model of diversification where interspecific pairwise competition (rate d) is always weaker than intraspecific pairwise competition (rate c), and is only felt from the part of individuals belonging to younger species. The only important parameter in this model is d/c, which can also be seen as a selection coefficient.

4 – Méthodes pénalisées pour la méta-modélisation et l'analyse de sensibilité

Sylvie Huet, MaIAGE INRA-Jouy-en-Josas

La modélisation en biologie repose sur des modèles complexes au sens où ils présentent des composantes fortement non linéaires ou font intervenir un grand nombre de variables d'état et de paramètres au travers d'interactions d'ordre élevé. Il est souvent nécessaire de les simplifier soit pour en comprendre le fonctionnement soit pour optimiser des calculs numériques. La construction d'un modèle simplifié, ou métamodèle, d'un modèle complexe à l'aide d'une décomposition de type ANOVA obtenue par projection sur des espaces de Hilbert auto-reproduisant a été proposée récemment. Le métamodèle doit approcher au mieux le modèle initial tout en restant parcimonieux, et en possédant de bonnes qualités prédictives pour relier les variables de sortie aux variables d'entrée. En utilisant les outils de l'estimation fonctionnelle sparse par minimisation de critères convexes, les propriétés prédictives des métamodèles ainsi construits peuvent être établies. Par ailleurs, grâce à la décomposition de type ANOVA, et à la sélection des termes intervenant dans cette décomposition par les méthodes de régularisation de type lasso, les indices de sensibilité du métamodèle approchent les indices de sensibilité du modèle complexe initial.
Ce travail a fait l'objet d'une collaboration avec M.L. Taupin (Université d'Evry, Val d'Essonne).

5 – Marche au hasard sur un digraphe aléatoire

Charles Bordenave, CNRS, Université de Toulouse

Nous nous intéresserons aux propriétés d'une marche au hasard sur un graphe fini dirigé aléatoire. Nous essaierons de décrire la mesure invariante de cette chaîne de Markov irréversible et d'étudier son temps de mélange. Nous montrerons que sur un graphe typique, cette chaîne satisfait le phénomène de "cut-off". Ce phénomène, proposé par Aldous et Diaconis en 1986, quantifie que la chaîne atteint brusquement un état proche de l'équilibre. Nous discuterons enfin des résultats très préliminaires sur le trou spectral de cette chaîne de Markov.

L'essentiel de l'exposé se basera sur un travail commun avec Pietro Caputo et Justin Salez disponible sur http ://arxiv.org/abs/1508.06600.

6 – Spiking neurons : interacting processes with memory of variable or infinite length

Eva Löcherbach, Université de Cergy Pontoise

We give an overview of recent models proposed to describe large or infinite networks of spiking neurons. These models are constituted by systems of non-Markovian processes with a countable number of interacting components, either in discrete or in continuous time. Each component is represented by a point process indicating the presence of a spike or not at a given time. Each particle's rate (in continuous time) or probability (in discrete time) of having a spike depends on the entire time evolution of the system since the last spike time of the particle. We discuss these models and several theoretical results that have been obtained (existence of a stationary solution, perfect simulation). In a second step, we adopt an equivalent description by means of an associated interacting particle system which is Markovian. We discuss its longtime behavior and study propagation of chaos in mean field systems.

Résumés des exposés des lauréats du prix de thèse Jacque Neveu (par ordre de programmation)

1. Emilie Kaufmann (CNRS, CRIStAL)
   Stratégies bayésiennes et fréquentistes dans des modèles de bandits.
2. Julien Reygner (CERMICS, École des Ponts ParisTech)
   Autour des modèles d'Atlas.
3. Erwan Scornet (Laboratoire de Statistique Théorique et Appliquée, UPMC)
   Promenade en forêts aléatoires

7 – Stratégies bayésiennes et fréquentistes dans des modèles de bandits

Emilie Kaufmann, CNRS, CRIStAL

Dans un modèle de bandit stochastique à plusieurs bras, un agent choisit séquentiellement des actions conduisant à des récompenses aléatoires, de façon à réaliser un compromis entre exploration et exploitation de son environnement. Ces modèles, initialement introduits dans le contexte des essais cliniques, sont aujourd'hui beaucoup étudiés pour diverses applications liées à l'optimisation de contenu web. Dans cet exposé, nous présenterons des stratégies optimales pour certains problèmes de bandit, et nous verrons en particulier que l'utilisation d'outils bayésiens est pertinent pour résoudre un problème de nature fréquentiste.

8 – Autour des modèles d'Atlas

Julien Reygner, CERMICS, École des Ponts ParisTech

Les modèles d'Atlas décrivent des systèmes de particules évoluant sur la droite réelle, au sein desquels la dynamique d'une particule ne dépend que de son rang dans le système. De tels modèles apparaissent naturellement en mécanique statistique, en finance ou en théorie des files d'attente, et leur étude repose sur divers outils : mouvement brownien réfléchi, propagation du chaos, théorie ergodique des processus de Markov. Dans cet exposé, nous présenterons quelques aspects du comportement en temps long et à grande échelle de ces modèles.

9 – Promenade en forêts aléatoires

Erwan Scornet, Laboratoire de Statistique Théorique et Appliquée, UPMC

Les forêts aléatoires, inventées par Breiman en 2001, comptent parmi les algorithmes d'apprentissage supervisé les plus utilisés, notamment en grande dimension. Elles possèdent en pratique de bonnes capacités prédictives et sont faciles à utiliser puisqu'elles ne nécessitent pas la calibration de multiples paramètres. Cependant, l'algorithme est difficile à analyser théoriquement car l'estimateur dépend de manière complexe des données. En effet, à chaque étape de l'algorithme, un critère empirique est minimisé sur un sous-ensemble des données initiales. Récemment, des travaux ont permis d'améliorer notre compréhension des forêts aléatoires, en s'intéressant notamment à leur normalité asymptotique ou à leur convergence. Dans cet exposé, je présenterai un premier résultat de convergence pour les forêts aléatoires de Breiman. Je montrerai également en quoi cette procédure possède de bonnes propriétés dans des contextes parcimonieux.

Résumé des sessions parallèles par ordre de programmation

Groupe 1 :

- PDMP pour la biologie (Romain Yvinnec et Florent Malrieu).
- Tests multiples (Magalie Fromont).
- Processus Gaussiens pour les expériences numériques (François Bachoc).
- Percolation (Marie Théret).

Groupe 2 :

- Grande dimension et génomique (Laurent Jacob).
- Statistiques pour les codes numériques (Merlin Keller).
- Géométrie stochastique (David Coupier).
- Probabilités et algorithmique (Irène Marcovici).

Groupe 3 :

- Optimisation stochastique : théorie et méthodes (Gersende Fort).
- Modèles statistiques autour de l'énergie (Jairo Cugliari).
- Applications statistiques des processus ponctuels déterminantaux (Rémi Bardenet).
- Modèles de polymères (Quentin Berger).

Groupe 4 :

- Optimisation stochastique et applications (Bruno Gaujal).
- Modèles et inférence pour données écologiques (Marie-Pierre Etienne).
- Inférence géométrique et topologique (Bertrand Michel et Clément Levrard).
- Matrices et graphes aléatoires (Camille Male).

Groupe 5 :

- Système de particules en interaction (Emmanuel Jacob).
- Parcimonie dans l'apprentissage statistique (Farida Enikeeva).
- Changement climatique (Anne-Catherine Favre).
- Inégalités de concentration (Pierre Youssef).

Groupe 1
Lundi 13:30-15:25, Amphi 7

Session : Modèles probabilistes et déterministes par morceaux pour la biologie

Organisée par
Romain Yvinec, INRA Tours
Florent Malrieu, LMPT, Université François-Rabelais de Tours

Résumé. Les récents progrès expérimentaux en biologie moléculaire et cellulaire motivent l'utilisation de modèles probabilistes de type non diffusif (chaîne de Markov en temps continu, hybride déterministe/stochastique). Une grande classe de processus, introduite dans la littérature par [4], fait intervenir des mouvements déterministes (type EDO) interrompus par des sauts aléatoires. Ces processus sont parfois très simples à définir mais posent de nombreuses questions théoriques intéressantes [3]. Au cours de cette session, nous verrons une revue de résultats récents sur les processus de Markov déterministes par morceaux, des limites d'échelles associées à ces processus, et deux applications en biologie moléculaire et en dynamique des populations.

1 – Quelques PDMP pour la biologie

Bertrand Cloez, Inra Mistea Montpellier

Un processus de Markov déterministe par morceaux (PDMP en anglais) est un processus évoluant selon une dynamique déterministe entre des sauts aléatoires. Après avoir rapidement rappelé la définition et les propriétés élémentaires de ces processus, nous ferons un panorama de quelques applications en biologie (division cellulaire, dynamique de population, neuroscience...). Des questions de comportement en temps long (convergence, extinction...) et de passage à l'échelle (liens EDP/EDO) seront abordées.

2 – Stochastic models of protein production with cell division and gene replication

Renaud Dessalles, INRA Unité MaIAGE - INRIA Paris

Protein production is the fundamental process by which the genetic information of a biological cell is synthesised into a functional product, the proteins. For prokaryotic cells (like bacteria), this is a highly stochastic process and results from the realisation of a very large number of elementary stochastic processes of different nature. This talk confronts experimental measures of protein production with classical stochastic models of gene expression (like those of Rigney and Schieve (1977) and Berg (1978)) ; it shows that these models cannot properly explain experimental measures of protein variance, especially for highly expressed proteins. We propose extended versions of these classical models, that take into account other

possible sources of variability such as the effect of cell division or gene replication on the number of proteins. Mathematical analysis of these models are presented as well as their biological interpretation with respect to experimental measures.

Joint work with Vincent Fromion (INRA Unité MaIAGE) and Philippe Robert (INRIA Paris).

3 – Échantillonnage dans une population structurée branchante

Aline Marguet, CMAP, École Polytechnique

On s'intéresse à l'évolution d'une population de cellules. Chaque individu dans la population est caractérisé par un trait (son âge, sa taille, le nombre de parasites, ...) qui évolue au cours du temps et qui détermine la dynamique de la cellule (sa durée de vie, son nombre de descendants, ...). Lorsqu'on échantillonne un individu uniformément au temps t, on cherche à connaître son trait et l'histoire de son trait le long de sa lignée ancestrale. Dans le prolongement des travaux [1] et [2], nous introduirons un processus auxiliaire qui permet de caractériser l'évolution du trait d'un individu typique. Ce processus auxiliaire apparaît dans une formule dite "Many-To-One" permettant de séparer la dynamique de l'ensemble de la population en deux phénomènes : le développement de la population en terme de nombre d'individus et l'évolution du trait des individus.

4 – Méthode de pénalisation et moyennisation pour des processus markovien déterministes par morceaux

Alexandre Genadot, Institut de Mathématiques de Bordeaux, Inria Bordeaux-Sud-Ouest

Considérons un processus markovien déterministe par morceaux, avec frontière, dont la composante modale est d'évolution plus rapide que la composante euclidienne. Supposons que l'on caricature ce fait en accélérant infiniment la dynamique du mode... que devient alors la composante euclidienne ? Peut-on décrire, à la limite, son comportement ? Que se passe-t-il à la frontière ? Nous verrons comment aborder ces questions en s'appuyant sur une méthode de pénalisation.

Références

[1] Bansaye, V. and Delmas, J.-F. and Marsalle, L. and Tran, V. C., *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees*, The Annals of Applied Probability, p. 2263–2314, 2011..

[2] Cloez, B., *Limit theorems for some branching measure-valued processes*, arXiv preprint arXiv :1106.0660, 2011..

[3] Malrieu, F., *Some simple but challenging Markov processes*, Ann. Fac. Sci. Toulouse Math. (6) vol. 24(4), 857–883 2015..

[4] Davis, M. H. A., *Piecewise-deterministic Markov processes : a general class of nondiffusion stochastic models*, J. Roy. Statist. Soc. Ser. B 46, no. 3, 353–388, With discussion. MR MR790622 (87g :60062), 1984..

Groupe 1
Lundi 13:30-15:25, Amphi 8

Session : Processus gaussiens pour les expériences numériques

Organisée par
François Bachoc, Université Toulouse III

Résumé. La place des expériences numériques est devenue prépondérante dans de nombreux domaines scientifiques, tels quel l'ingénierie, la physique, la biologie ou l'économie. Effectuer une expérience numérique s'apparente à obtenir une évaluation (coûteuse !) d'une fonction déterministe et inconnue. Pour pallier à ce coût, un paradigme répandu est de modéliser cette fonction inconnue comme une réalisation de processus gaussien. On peut alors utiliser la loi du processus gaussien, conditionnelle aux évaluations de la fonction, pour limiter le nombre de ces évaluations. Dans cette session, nous exposerons plusieurs avancées récentes, théoriques et pratiques, concernant les processus gaussiens pour les expériences numériques.

1 – Quelques défis actuels autour des méthodes par processus gaussiens

David Ginsbourger, Idiap Research Institute et Université de Berne, Suisse

Suite à l'essor des méthodes par processus gaussiens en planification d'expériences numériques et en apprentissage statistique, leur étude a récemment connu un regain d'intêret non-seulement en statistiques appliquées mais aussi dans d'autres domaines, et notamment en optimisation numérique. Après quelques rappels sur les bases de ces méthodes, nous porterons notre attention sur une sélection de défis contemporains en lien avec les exposés suivants. En particulier : comment et à quel point peut-on incorporer des hypothèses structurelles dans des modèles de type processus gaussien ? Quelles sont les limites des processus gaussiens concernant la taille du jeu d'apprentissage et comment aller au delà du modèle classique pour faire face à des jeux de donnés de grande taille ? Enfin, comment peut-on tirer partie de modèles de processus gaussiens pour générer des algorithmes d'optimisation globale parcimonieux en nombre d'évaluations, et quelles garanties peuvent être apportées pour de tels algorithmes ?

2 – Estimation de la courbe d'actualisation par krigeage sous contraintes

Hassan Maatouk, Institut National des Sciences Appliquées Lyon, hassan.maatouk@insa-lyon.fr

La construction de structures par terme est au coeur de l'évaluation financière et de la gestion du risque voir e.g. [1] et [2]. Une structure par terme est une courbe qui d'écrit l'évolution d'une grandeur économique ou financière comme une fonction de la maturité ou horizon de temps. Des exemples typiques sont la structure par terme des taux d'intérêt sans risque, la structure par terme d'obligations, la structure par terme de probabilités de défaut et la structure par terme de volatilités implicites de rendements d'actifs financiers. En pratique, les cotations des marchés des produits financiers sous-jacents sont utilisées et

fournissent une information partielle sur les structures par terme considérées. De plus, cette information est plus au moins fiable en fonction de la liquidité de la maturité des marchés en question. Le problème est d'obtenir une courbe continue en la maturité à partir de ces informations.

Dans [1], une courbe d'actualisation ainsi que des probabilités de défaut sont estimés en utilisant la Régression par Processus Gaussien sous contraintes de type inégalité. L'estimation des paramètres de la fonction de covariance du processus gaussien initial a été possible en adaptant également la méthode de validation croisée. Pour étudier le modèle proposé, quelques exemples numériques en dimension 1 et 2 sont inclus. Les résultats numériques montrent que les simulations du processus gaussien conditionnel respectent à la fois les contraintes linéaires de type égalité et la monotonie décroissante par rapport à la maturité.

3 – Nested estimation of kriging models for large data-sets

Didier Rullière, ISFA

Gaussian random fields are widely studied from a statistical point of view, and have found applications in many areas, including geostatistics, climate science and computer experiments. One limitation of Gaussian process models is that classical statistical procedures (such as Gaussian conditioning and Maximum likelihood) entail computationally intensive calculations when the data-set size n is large. Hence, there is a lively research activity for designing and analyzing approximate procedures which are computationally cheaper. We propose a new procedure based on the aggregation of several Gaussian process models, each based on a different subset of the total data set. We support this procedure with asymptotic results. Furthermore the effectiveness of the method is illustrated on an industrial case study.

4 – Optimisation séquentielle par processus Gaussiens

Emile Contal, CMLA, ENS Cachan

L'optimisation globale de fonctions "boites noires" est une étape cruciale de nombreuses applications, du design industriel à la calibration de modèles. Cette tâche est particulièrement complexe lorsque chaque évaluation de la fonction est coûteuse. Le but de l'optimisation séquentielle est de simultanément explorer la fonction inconnue et converger vers son optimum global, en utilisant le moins d'évaluations possible. Le modèle des processus Gaussiens apporte un cadre efficace et rigoureux pour étudier et construire divers algorithmes. Dans cet exposé nous supposerons que la fonction inconnue est tirée suivant un processus Gaussien. Nous étudierons des propriétés du processus qui sont fondamentales pour décrire la complexité du problème d'optimisation. Nous en déduirons un algorithme avec des garanties théoriques, puis nous analyserons les performances empiriques dans des cas réels ou simulés. Nous discuterons enfin des limitations pratiques de tels algorithmes dans des cadres complexes.

Références

[1] Cousin, A. and Maatouk, H. and Rullière, D., *Kriging of financial term-structures*, in revision, ArXiv, 2015.

[2] Kenyon, C. and Stamm, R., *Discounting, Libor, CVA and Funding : Interest Rate and Credit Pricing*, Palgrave Macmillan, 2012.

Groupe 1
Lundi 13:30-15:25, Amphi 9

Session : Percolation

Organisée par
Marie Théret, LPMA, Université Paris Diderot

Résumé. Dans cette session nous passerons en revue quelques avancées récentes sur les modèles de type percolation (percolation, percolation orientée, percolation continue, percolation paire).

1 – Percolation et modèles de dimères

Vincent Beffara, Institut Fourier, Université Grenoble Alpes.

Pour un pavage par dimères de loi uniforme dans le "diamant aztèque", il apparaît une interface entre une phase ordonnée ("solide") au bord et une zone désordonnée ("liquide") au centre du domaine ; les fluctuations autour de cette interface convergent vers le processus d'Airy, et sont bien comprises. Si on remplace la mesure uniforme en donnant à chaque dimère un poids de manière périodique, il peut apparaître une troisième phase qualitativement différente ("gazeuse"), et l'interface entre gaz et liquide est plus difficile à étudier. Le même processus d'Airy apparaît à la limite, et j'expliquerai comment pour étendre les preuves précédentes à ce cas il est utile de voir la phase gazeuse comme un modèle de percolation sous-critique.

2 – Le nombre de chemins en percolation orientée.

Olivier Garet, IECL, Université de Lorraine - Nancy.

Si N_n désigne le nombre de chemins de hauteur n dans la percolation de Bernoulli orientée, la quantité $N_n^{1/n}$ converge presque sûrement. Il s'agit là d'une conjecture bien naturelle, que j'ai démontrée il y a peu avec Jean-Baptiste Gouéré et Régine Marchand. J'essaierai de donner quelques idées des ingrédients de la preuve, en particulier du lien avec certaines techniques sous-additives récentes.

3 – Percolation et percolation de premier passage dans le modèle booléen

Jean-Baptiste Gouéré, LMPT, Université François Rabelais - Tours

Jetons de manière indépendante et homogène une infinité de boules de rayons aléatoires dans un espace euclidien. Notons S la réunion de ces boules. Voici deux questions :

1) Toutes les composantes connexes de S sont-elles bornées ?
2) Un marcheur se déplace à vitesse 1 en dehors de S et à vitesse infinie dans S. S'il optimise son trajet, le temps nécessaire pour aller de l'origine à un point éloigné x sera-t-il asymptotiquement proportionnel à la norme euclidienne de x ?

On vérifie facilement que si la réponse à la première question (qui porte sur un modèle de percolation) est non, alors la réponse à la deuxième question (qui porte sur un modèle de

percolation de premier passage) est également non. Dans cet exposé nous précisons les liens entre ces deux modèles. Travail en collaboration avec Marie Théret.

4 – Percolation paire sur $\mathbb{Z}^2$

Irène Marcovici, IECL, Université de Lorraine - Nancy.

Dans le modèle classique de percolation par arête sur $\mathbb{Z}^2$, on fixe un paramètre $p \in (0,1)$, et pour chaque arête, on décide de la garder avec probabilité p et de l'effacer avec probabilité $1-p$, de manière indépendante pour différentes arêtes. On s'intéresse alors à l'existence d'une composante connexe infinie dans le sous-graphe aléatoire obtenu.

Ici, nous voulons conditionner cette percolation à être paire, c'est-à-dire à ce que dans le sous-graphe obtenu, chaque sommet soit de degré pair. Nous montrons tout d'abord, en utilisant le formalisme des mesures de Gibbs, l'existence et l'unicité de la mesure de percolation paire de paramètre p. Cette construction permet de faire le lien avec les contours du modèle d'Ising sur $\mathbb{Z}^2$, pour une certaine température dépendant de p.

Nous nous intéressons ensuite à l'existence d'une composante connexe infinie. La difficulté principale découle du fait que le conditionnement crée des dépendances à longue portée, et qu'il n'y a plus de propriété de monotonie aussi simple que celle obtenue dans le modèle classique avec le couplage croissant.

Il s'agit d'un travail commun avec Olivier Garet et Régine Marchand.

Groupe 1
Lundi 13:30-15:25, Amphi 11

Session : Tests multiples

Organisée par
Magalie Fromont, Université Rennes 2, IRMAR

Résumé. Dans le contexte actuel, favorisant la multiplicité des sources d'information et de questionnement, les tests multiples viennent répondre à de plus en plus de problèmes de Statistique complexes. La question des tests multiples est ici abordée sous un angle essentiellement théorique, des éléments classiques ou historiques jusqu'aux développements les plus récents. La dimension pratique, motivant ce type de question et en posant les contraintes, est également traitée au travers de quelques applications.

1 – Évaluation théorique des tests multiples

Patricia-Reynaud Bouret, Université de Nice Sophia-Antipolis, LJAD

Travail en collaboration avec Magalie Fromont (IRMAR, Université Rennes 2) et Matthieu Lerasle (LJAD, Université de Nice Sophia-Antipolis).
Après un rappel des définitions usuelles relatives aux tests multiples, notamment celles des différents critères d'évaluation liés à l'erreur de première espèce comme le Family-Wise Error Rate (FWER), nous décrirons quelques procédures de tests multiples bien connues ainsi que leurs propriétés. Nous nous interrogerons ensuite sur l'évaluation de ces procédures du point de vue de l'erreur de deuxième espèce, et nous verrons qu'il n'existe dans la littérature des tests multiples que très peu de critères liés à cette erreur. Partant d'un parallèle entre les tests multiples et les tests agrégés développés en théorie des tests d'hypothèses simples, adaptatifs au sens du minimax, nous introduirons la notion de vitesse de séparation par famille ou Family-Wise Separation Rate (FWSR), comme point de départ d'une théorie minimax pour les tests multiples contrôlant le FWER. Nous étudierons alors les propriétés d'adaptation (ou non) au sens du minimax de différentes procédures de tests multiples dans des cadres gaussiens classiques.

2 – Continuous Testing for Poisson process intensities

Franck Picard, CNRS, Univ. Lyon 1

Next Generation Sequencing technologies now allow the genome-wide mapping of binding events along genomes, like the binding of transcription factors for instance. More generally, the field of epigenetics is interested in the regulation of the genome by features that are spatially organized. One open question that remains is the comparison of spatially ordered features along the genome, between biological conditions. An example would be to compare the location of transcription factors between disease and healthy individuals. We propose here to model the spatial occurences of genomic features in each condition by a Poisson process

with a heterogeneous intensity on [0,1], and we restate the problem as the comparison of Poisson process intensites in continuous time. Contrary to global testing approaches that consist in testing whether the two intensities are equal on [0,1], we focus on a local testing strategy using scanning windows. Our method is based on kernel to build the test statistics, and on monte-carlo simulations to compute the p-value process. By using the continuous testing framework, we provide a procedure that controls the Family Wise Error Rate as well as the False Discovery Rate in continous time. We illustrate our method on experimental data, and discuss its extensions in the general framework of testing for Poisson process intensities. Joint work with Etienne Roquain (Paris 6), Anne-Laure Fougères (Lyon 1), Patricia Reynaud-Bouret (Nice).

3 – Usage de poids optimaux data-driven dans la procédure de Benjamini et Hochberg

Guillermo Durand, Université Pierre et Marie Curie, LPMA

On cherche à améliorer la puissance de Benjamini et Hochberg en mettant des poids sur les p-values. La question se pose alors de choisir les poids les plus pertinents. Dans un contexte où les p-values sont réparties dans un nombre fini de groupes, on propose des poids calculés à partir des p-values, qui satisfont un résultat de maximalité en puissance quand le nombre de tests tend vers l'infini.

4 – Inférence post hoc en test multiple

Pierre Neuvial, CNRS, Université d'Évry Val D'Essone

Travail en collaboration avec Gilles Blanchard et Etienne Roquain.

Lorsque l'on teste simultanément un grand nombre d'hypothèses nulles, une pratique courante dans les applications (notamment en génomique ou en neuro-imagerie) consiste à (i) sélectionner un sous-ensemble d'hypothèses candidates, puis (ii) raffiner cette sélection à l'aide de connaissances a priori. Le contrôle de mesures de risque classiques en test multiple comme le False Discovery Rate ne fournit aucune garantie statistique sur les ensembles d'hypothèses obtenus par ce processus.

Ce fossé entre les besoins des applications et les garanties fournies par les méthodes actuelles motive le développement de procédures dites post hoc, c'est-à-dire pour lesquelles les ensembles d'hypothèses sélectionnés peuvent être définis par l'utilisateur de la procédure, après avoir "vu les données". Goeman et Solari (Stat. Science, 2011) ont proposé des procédures post hoc reposant sur la notion de "closed testing".

Nous introduisons une construction alternative de procédures post hoc. Celle-ci repose sur le contrôle d'une nouvelle mesure de risque, appelée joint Family-Wise Error Rate (JFWER). Nous proposons des procédures de contrôle du JFWER adaptées notamment au cas où la loi jointe des statistiques de test sous l'hypothèse nulle est connue, et discutons leur performance ainsi que les liens avec la procédure proposée par Goeman et Solari.

Groupe 2
Lundi 17:00-18:55, Amphi 7

Session : Grande dimension et génomique

Organisée par
Laurent Jacob, CNRS/Université Lyon 1, UMR 5558 LBBE

Résumé. Les technologies à haut débit en biologie moléculaire permettent de mesurer un grand nombre de facteurs dans des échantillons. L'accès simultanée à l'expression de milliers de gènes ou la présence de millions de SNPs a le potentiel d'améliorer notre compréhension des phénomènes biologiques et notre capacité à les prédire. Le nombre d'échantillons disponible demeure toutefois réduit, donnant lieu à des problèmes d'inférence en grande dimension ("grand p, petit n"). La session a pour objectif d'exposer des techniques développées pour répondre aux défis statistiques et computationnels rencontrés dans ce contexte.

1 – Régression en grande dimension et épistasie par blocs pour les études d'association

Christophe Ambroise, Université d'Evry Val d'Essonne - INRA, UMR 8071 Laboratoire de Mathématiques et Modélisation d'Évry

Dans le domaine des études d'association pan-génome (GWAS) une partie de la littérature est consacrée à la détection des interactions existant entre deux ou plusieurs parties du génome (épistasie).

La plupart des approches considèrent les interactions entre loci déjà connus pour être associés au phénotype étudié. Dans cette présentation nous explorons des approches statistiques multi-variées permettant de détecter des épistasies au niveau de groupes de SNP sans filtrage préalable.

2 – Une approche de sélection de variables pour améliorer l'estimation d'héritabilité dans les modèles linéaires mixtes parcimonieux

Anna Bonnet, AgroParisTech - INRA, UMR 518 MIA

L'héritabilité d'un caractère biologique est définie comme la part de sa variation au sein d'une population qui est causée par des facteurs génétiques. Nous proposons dans un premier temps un estimateur de l'héritabilité dans les modèles linéaires mixtes parcimonieux, dont nous avons étudié les propriétés théoriques. Nous mettons en évidence que lorsque la taille des effets aléatoires est trop grande par rapport au nombre d'observations, nous ne pouvons fournir une estimation précise pour l'héritabilité. Nous avons ensuite proposé une méthode de sélection de variables afin de réduire la taille des effets aléatoires, dans le but d'améliorer la précision de l'estimation de l'héritabilité. Néanmoins, nous montrons que ce type d'approche fonctionne uniquement lorsque le nombre composantes non nulles dans les effets aléatoires, c'est à dire le nombre de variants génétiques qui influencent la variation phénotypique, est

assez faible. Nous avons finalement établi un critère empirique pour déterminer les cas où il était possible de faire de la sélection de variables.

3 – Détection d'outliers en grande dimension : application à la génomique des populations

Mickael Blum, CNRS, Université Grenoble Alpes, UMR 5525 Labo TIMC-IMAG

Notre objectif est de détecter quelles sont les variables outliers dans des jeux de données de grande dimension. Les méthodes de détection d'outliers sont utilisées en génomique pour détecter quels sont les gènes qui permettent aux individus de s'adapter à leur environnement. Nous proposons une approche rapide basée sur l'analyse en composantes principales. Le principe est de considérer comme gènes candidats ceux qui sont excessivement corrélés avec les composantes principales. Pour ce faire, nous calculons pour chaque marqueur génétique un vecteur qui mesure l'association entre un marqueur génétique et les composantes principales. Nous utilisons ensuite la distance de Mahalanobis pour trouver quels sont les vecteurs atypiques. En utilisant un jeu de données humains comprenant un peu plus d'un millier d'individus et des centaines de milliers de marqueurs génétiques, nous montrons que cette approche permet de détecter des exemples d'adaptation biologique chez l'homme.

4 – Deciphering splicing from high-throughput RNA sequencing with sparse regression techniques.

Elsa Bernard, Mines ParisTech - Institut Curie, U900 Cancer et génome

Detecting and quantifying alternatively spliced isoforms from RNA-seq data is an important but challenging task. Several state-of-the-art methods are based on sparse probabilistic models. However, explicitly listing the –possibly exponentially– large set of candidate transcripts is intractable for genes with many exons. We develop a technique based on network flow optimization which can efficiently tackle the sparse estimation problem on the full set of candidate transcripts. We show that the penalized likelihood maximization can be reformulated as a convex cost flow problem over a network, which can be solved with polynomial-time algorithms. We also propose to infer the transcripts jointly across several related samples. We formulate a convex optimization problem that allows to share information between samples and that we solve efficiently. Finally we extend our techniques to RNA-seq data from targeted experiments on human disease gene such as the breast cancer susceptibility gene BRCA1, and show results of importance in a clinical diagnostic context.

Groupe 2
Lundi 17:00-18:55, Amphi 8

Session : Méthodes statistiques pour les codes de calcul

Organisée par
Merlin Keller, EDF R&D

Résumé. L'utilisation croissante de codes de simulation numériques pour la compréhension et la prévision des systèmes physiques mène à des problématiques spécifiques, qui peuvent être dus à leur caractère coûteux (temps de calcul important), irrégulier, ou encore à la dimension importante des variables d'entrée et/ou de sortie. Nous présentons ici plusieurs approches récentes pour y répondre, que ce soit dans le cadre d'une analyse de sensibilité (étude de l'influence relative des entrées sur les sorties), de l'optimisation robuste, de la vérification d'un code, ou du calage de ses paramètres incertains.

1 – Analyse de sensibilité globale pour modèles à entrées et/ou sorties multivariées

Clémentine Prieur, Université Grenoble Alpes.

De nombreux codes numériques prennent en entrée des variables vectorielles et/ou fonctionnelles pour produire en sortie une ou plusieurs quantités d'intérêt. Aussi bien les entrées que les quantités d'intérêt peuvent présenter des corrélations. Nous présenterons dans ce contexte différents outils statistiques pour la visualisation des incertitudes d'une part, pour l'analyse de sensibilité globale d'autre part. Les méthodologies proposées seront illustrées sur quelques applications récentes [2, 1].

2 – Optimisation bayésienne multi-objectifs sous contraintes

Emmanuel Vazquez, Laboratoire des Signaux et Systèmes, CentraleSupélec

Nous proposons un nouvel algorithme en optimisation bayésienne, appelé BMOO (Bayesian Multi-Objective Optimization). Cet algorithme permet d'aborder des problèmes d'optimisation sans dérivées, mono- ou multi-objectifs, sous contraintes d'inégalité non-linéaires. Les objectifs et les contraintes peuvent potentiellement être coûteux à évaluer (ce qui implique que le nombre d'évaluations pouvant être utilisées pour mener à bien l'optimisation est très limité). L'optimisation bayésienne consiste à modéliser les fonctions évaluées par des processus aléatoires et à adopter une stratégie de décision séquentielle pour choisir les évaluations. Notre contribution principale consiste à proposer une règle de domination étendue pour gérer simultanément les objectifs et les contraintes. Puis, nous utilisons un critère d'échantillonnage traduisant l'amélioration du front de Pareto (de type hyper-volume dominé). L'algorithme proposé peut démarrer sans connaissance d'un point admissible. Le calcul et l'optimisation du critère d'échantillonnage sont effectués en utilisant des techniques de type Monte-Carlo séquentiel. Nous avons comparé notre algorithme aux algorithmes de l'état de l'art pour des

problèmes d'optimisation sous contrainte mono- et multi-objectifs et les résultats obtenus sont très satisfaisants.

3 – Processus gaussiens pour l'utilisation massive et automatique de simulateurs numériques

François Bachoc, Institut Mathématique de Toulouse

De nombreuses études en sciences nucléaires se fondent sur l'utilisation intensive de simulateurs numériques. Ces simulateurs sont alors utilisés automatiquement dans potentiellement des milliers de conditions de calculs, sans contrôle individuel sur l'aspect physique et numérique de chacun des calculs. Cette présentation présente une étude de cas de cette problématique lors de l'utilisation du code de thermo-mécanique Germinal. Nous montrons comment le modèle par processus gaussiens permet d'obtenir une approximation peu couteuse du code Germinal. Nous comparons alors cette approximation avec celles fournies par d'autres techniques classiques : les réseaux de neurones et les méthodes à noyaux. Dans un second temps, nous montrons comment les processus gaussiens peuvent aider à détecter les problèmes numériques liées à l'utilisation massive de simulateurs. On distingue alors la détection de calculs complètement faux (pour lesquels la détection se fait bien par plusieurs techniques d'approximation de codes) et la détection d'inconsistance entre calculs (qui est plus délicate et pour laquelle les processus gaussiens sont particulièrement intéressants).

4 – Sélection bayésienne de modèle pour l'identification des paramètres d'un code de calcul

Guillaume Damblin, CEA DEN/DANS/DM2S/STMF/LGLS

Nous présentons une approche bayésienne pour la validation d'un code de calcul simulant une quantité physique d'intérêt. La validation est appréhendée comme un problème de test statistique qui confronte l'hypothèse nulle selon laquelle le code de calcul prédit parfaitement la quantité physique d'intérêt, avec l'hypothèse alternative selon laquelle une erreur systématique subsiste entre le système physique et les prédictions du code. Lorsque le code dépend d'un paramètre inconnu, l'hypothèse nulle correspond à l'existence d'une valeur du paramètre permettant un ajustement exact du code au système physique, tandis que l'hypothèse alternative correspond à la situation pour laquelle chaque valeur du paramètre définit une fonction d'erreur non nulle entre la réponse du code et la quantité physique d'intérêt. En supposant la linéarité du code par rapport au paramètre et en représentant *a priori* l'erreur éventuelle par un processus gaussien, le facteur de Bayes est calculé à partir des mesures physiques disponibles de façon à discriminer laquelle de ces deux hypothèses statistiques est la plus probable. Une attention particulière est portée au choix de lois *a priori* objectives necessitant le calcul d'un facteur de Bayes intrinsèque [3], qui dans un cas particulier est égal au facteur de Bayes.

Références

[1] L. Gilquin, T. Capelle, E. Arnaud and C. Prieur, *Sensitivity Analysis and Optimisation of a Land Use and Transport Integrated Model*, hal-01291774.

[2] S. Nanty, C. Helbert, A. Marrel, N. Pérot and C. Prieur, *Sampling, Metamodeling, and Sensitivity Analysis of Numerical Simulators with Functional Stochastic Inputs*, SIAM/ASA Journal on Uncertainty Quantification, 2016, 4(1).

[3] Berger, J.O. and Pericchi, L.R., *The intrinsic Bayes factor for model selection and prediction*, Journal of the American Statistical Association, 1996, 433(91).

Groupe 2
Lundi 17:00-18:55, Amphi 9

Session : Probabilités et algorithmique

Organisée par
Irène Marcovici, Institut Élie Cartan de Lorraine, Université de Lorraine

Résumé. Les probabilités et l'algorithmique sont des domaines qui se nourrissent mutuellement. D'un côté, l'analyse d'algorithmes fait naturellement intervenir l'étude d'objets discrets aléatoires. De l'autre côté, pour étudier et engendrer aléatoirement des structures combinatoires complexes, il est souvent nécessaire de développer des algorithmes sophistiqués. Dans ce contexte, une question d'actualité est aussi la classification des distributions de probabilités qui peuvent être décrites à l'aide de différents modèles de calcul. Cette session présentera différents résultats récents de nature théorique illustrant les interactions entre probabilités et algorithmique. Ils auront en commun de faire intervenir des dynamiques discrètes sur des structures aléatoires complexes.

1 – Mesures atteignables asymptotiquement par itération d'une mesure par un automate cellulaire

Mathieu Sablik, Institut de Mathématiques de Marseille, Aix Marseille Université

Les automates cellulaires modélisent des phénomènes à interactions locales où chaque cellule est actualisée en parallèle. Leur étude empirique est basée sur l'observation d'un diagramme espace-temps initié par une configuration tirée au hasard. Il est donc naturel de s'intéresser aux valeurs d'adhérence de la suite des itérés d'une mesure par un automate cellulaire, cela correspond aux comportements asymptotiques typiques observés.

Une question naturelle est donc de caractériser les ensembles de mesures limites atteignables. En plus des contraintes topologiques classiques, il apparaît naturellement des contraintes liées à la calculabilité. Ce sont en fait les seules : pour un ensemble de mesures données, il existe un automate cellulaire qui réalise cet ensemble comme mesures limites. Ce type de construction permet de réaliser du calcul sur l'espace des mesures de probabilité à l'aide d'un automate cellulaire.

2 – Bestiaire de chaînes de Markov à mémoire variable et de marches aléatoires persistantes.

Peggy Cénac, Institut Mathématique de Bourgogne, Université de Bourgogne

Cet exposé présentera une petite zoologie de chaînes de Markov à mémoire variable, avec des conditions d'existence et unicité de mesure invariante. Il sera ensuite question de marches aléatoires persistantes, construites à partir de chaînes de Markov à mémoire non bornée, où les longueurs de sauts de la marche ne sont pas forcément intégrables. Un critère de récurrence/transience s'exprimant en fonction des paramètres du modèle sera énoncé. Suivront

plusieurs exemples illustrant le caractère instable du type de la marche lorsqu'on perturbe légèrement les paramètres. Les travaux décrits dans cet exposé sont le fruit de plusieurs collaboration avec B. Chauvin, F. Paccaut et N. Pouyanne ou B. de Loynes, A. Le Ny et Y. Offret.

3 – Simulation parfaite de réseaux fermés de files d'attentes

Christelle Rovetta, Inria Paris (DYOGENE) et ENS

L'algorithme de simulation parfaite fut introduit en 1996 par Propp et Wilson [1]. Il permet l'échantillonnage non biaisé de la distribution stationnaire d'une chaîne de Markov ergodique. Le coût à payer est la simulation en parallèle de tous les états possibles de la chaîne. Dans cet exposé, nous nous intéressons à la simulation parfaite de réseaux fermés de files d'attentes. L'évolution de tels réseaux est modélisée par une chaîne de Markov ergodique dont l'espace d'états est de cardinalité exponentielle en le nombre de files, rendant ainsi l'algorithme de simulation parfaite inutilisable en pratique. Nous présentons ici une représentation compacte de l'espace des états que l'on appelle diagramme. Cette dernière a une complexité polynomiale. Nous montrerons comment simuler en parallèle tous les états à l'aide des diagrammes et réaliserons ainsi la simulation parfaite de réseaux fermés de files d'attentes.

4 – Couplage par le passé : accélérer l'échantillonnage pour le modèle des sphères dures

Rémi Varloot, Microsoft Research - Inria Paris (DYOGENE) et ENS

Le modèle des sphères dures est une représentation simple de l'état d'un gaz, utilisé en physique statistique pour obtenir une équation d'état. On le retrouve aussi en théorie des graphes, où l'on parle d'ensembles indépendants. La complexité combinatorielle du modèle est telle que l'échantillonnage d'une configuration admissible se fait le plus souvent à l'aide de l'algorithme de Monte-Carlo ou de ses variantes. Parmi celles-ci, on pourra citer l'algorithme de couplage par le passé (CFTP), introduit par Propp et Wilson en 1996 [1]. Celui-ci est particulièrement adapté du fait de la propriété de monotonie que présente le modèle. Après avoir présenté le modèle et détaillé les particularités de l'algorithme, je m'intéresserai à une amélioration possible de celui-ci, visant à accélérer l'échantillonnage.

Références

[1] J.G. Propp et D.B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random structures and Algorithms, 1996, 1-2(9).

Groupe 2
Lundi 17:00-18:55, Amphi 11

Session : Géométrie Stochastique

Organisée par
David Coupier, Université Lille 1.

Résumé. P. Calka présentera dans l'exposé long de cette session un thème historique de la géométrie stochastique, à savoir l'étude de la frontière de l'enveloppe convexe d'un grand nombre de points jetés aléatoirement dans un polytope. Puis, trois exposés sur des thématiques bien différentes illustreront le large spectre couvert par la géométrie stochastique : l'apparition du Brownian net comme limite d'échelle d'une famille de marches aléatoires coalescentes-branchantes (E. Schertzer), un problème d'optimisation en géométrie algorithmique (X. Goaoc) et l'absence de composante connexe infinie dans des graphes aléatoires géométriques (S. Le Stum).

1 – Polytopes aléatoires : étude asymptotique

Pierre Calka, Université de Rouen.

L'exposé débutera par une brève introduction à la géométrie stochastique et à quelques modèles classiques. Nous nous intéressons en particulier aux polytopes aléatoires qui sont construits comme enveloppes convexes de nuages de points aléatoires de l'espace euclidien. Lorsque la taille du nuage tend vers l'infini, nous étudions le comportement de certaines variables telles que le nombre de points extrémaux ou le volume. Nous proposons une méthode inédite qui fournit le calcul explicite de variances limites en même temps que la convergence du processus frontière du polytope. Celle-ci repose sur la notion de stabilisation pour des processus ponctuels de Poisson et l'utilisation de graphes de dépendance. Nous discutons en détail le cas de points uniformes dans un polytope simple fixé. Cet exposé est issu de plusieurs travaux en collaboration avec T. Schreiber et J. Yukich.

2 – Propriétés géométriques du filet Brownien

Emmanuel Schertzer, Université Paris 6.

Dans cet exposé, j'introduirai un modèle de percolation dirigée en dimension 1+1, tel que l'ensemble des trajectoires partant de l'origine se décrit naturellement comme un bouquet de marches aléatoires coalescentes-branchantes avec mort. Dans une certaine limite d'échelle, ce modèle converge vers le filet Brownien avec mort (Brownian net with killing). Je présenterai certaines propriétés géométriques remarquables de cet objet limite, et je discuterai du lien entre ces propriétés géométriques et certains modèles de mécanique statistique. Ce travail est en collaboration avec C. Newman et K. Ravishankar.

3 – Complexité lissée d'enveloppe convexe

Xavier Goaoc, Université Paris-Est Marne-la-Vallée

Spielman et Teng ont apporté une réponse éclairante à une question importante en analyse d'algorithmes (comment l'algorithme du simplexe peut-il être si efficace en pratique et si inefficace en théorie ?) au travers d'un modèle probabiliste original, qu'ils ont appelé "complexité lissée". J'introduirai ce modèle et certaines des questions qu'il soulève en géométrie algorithmique via l'exemple de l'enveloppe convexe. Aucune connaissance en algorithmique ne sera nécessaire (voire utile).

4 – Absence de percolation dans les modèles germes-grains arrêtés

Simon Le Stum, Université Lille 1.

Dans cet exposé, nous étudierons la question de la percolation pour des graphes orientés obtenus comme étant l'état final d'une dynamique germes-grains. Nous présenterons deux dynamiques germes-grains dans le plan. Chacune d'entre elles, exercée sur des configurations poissoniennes de points, génère un graphe orienté "outdegree one", c'est-à-dire que de chacun des sommets du graphe est issu une unique arête sortante. Nous généraliserons la structure de ces deux exemples en donnant une construction formelle des modèles germes-grains dans R^d. Le point central de l'exposé concerne l'introduction de deux hypothèses portant sur ces modèles qui, étant vérifiées, garantissent l'absence de percolation dans le modèle. Nous détaillerons chacune de ces deux hypothèses et vérifierons que les deux exemples de dynamiques données dans la première partie de l'exposé les satisfont.

Groupe 3
Mardi 10:00-11:55, Amphi 7

Session : Modèles statistiques autour de l'énergie

Organisée par
Jairo Cugliari, Univ Lyon

Résumé. Les interactions entre le milieu academique et le milieu professionnel lié aux energies ont donné des riches collaborations dont cette session en temoigne : des domaines divers de la mathématique appliquée pour la modélisation, prévision et optimisation des energies.

1 – Méthodes d'ensembles pour la prévision dans le domaine de l'énergie

Yannig Goude, EDF-Université Paris-Sud

Les méthodes d'ensemble permettent d'obtenir relativement aisément de bons résultats pour un grand nombre d'applications et en particulier dans le domaine de l'analyse prédictive. Elles sont génériques et nécessitent peu d'intervention humaine pour leur calibration. De plus, le boom des data sciences et des outils open source associés (`R`, `Python` par exemple) permettent au data scientist de disposer d'un nombre croissant de méthodes de prévision (par exemple le package `caret` en `R` recense plus de 200 modèles de régression/classification). Ainsi, de nombreux challenges de prévision ont récemment été remportés par des approches ensemblistes, la démarche classique des challengers étant de développer un certains nombre de modèles différents, de les confronter, puis finalement les agréger. Dans cet exposé nous présentons une étude comparative de ces approches d'ensemble pour la prévision dans l'énergie (consommation d'électricité, prix, photo-voltaïque, éolien). Nous comparons des méthodes ensemblistes classiquement de nature "off-line" telle que le bagging, le boosting ou le stacking à des méthodes d'agrégation "on-line" (les données sont observées instant après instant, de manière séquentielle) présentées dans [1].

2 – Prévision à court terme de production électrique éolienne

Lucie Montuelle, Université Paris Diderot, ANR Forewer

La prévision de production électrique éolienne est une problématique touchant des enjeux économiques, industriels et environnementaux. Dans le cadre du projet ANR Forewer, nous nous sommes intéressées à la prévision en temps réel de la production électrique éolienne sur un parc à partir des données de capteurs météorologiques. Les méthodes d'apprentissage statistique ont été éprouvées et comparées à des modèles paramétriques, proches du modèle physique. Sur les données considérées, la supériorité de l'apprentissage et en particulier des forêts aléatoires, bien calibrées, a été montrée. De plus, cette procédure semble assez robuste à l'erreur de mesure de la vitesse du vent. Ce travail a été effectué en collaboration avec A. Fischer, M. Mougeot et D. Picard [2].

3 – Classification de consommateurs électriques à l'aide de modèles de mélanges en régression en grande dimension.

Emilie Devijver, Department of Mathematics and Leuven Statistics Research Centre (LStat), KU Leuven, Leuven, Belgium

De nombreuses informations sur les consommateurs individuels sont désormais disponibles grâce, par exemple, aux nouvelles techniques de *smart grid*. L'exploitation de ces résultats passe par la modélisation à différentes échelles et l'exploitation du profil de charge. La segmentation des consommateurs basée sur la classification de charge de consommation est une approche naturelle dans cette direction. Dans cet exposé, on illustre sur des données réelles l'utilisation d'une méthode basée sur les modèles de régression en grande dimension qui effectue simultanément la sélection de modèles (pour réduire la dimension) et la classification. On insistera sur les avantages de la méthode par rapport à ce jeu de données électriques individuelles irlandaises. Ce travail est en collaboration avec Yannig Goude et Jean-Michel Poggi, et actuellement soumis [3].

4 – Deux contributions pour l'optimisation de système électrique fortement renouvelables

Robin Girard, MINES ParisTech, PSL - Research University, PERSEE

Comparer plusieurs options pour le mix électriques du futur afin de considérer notamment la possibilité d'une production majoritairement d'origine renouvelable est une tâche dans laquelle un grand nombre de problèmes de statistique et d'optimisation émergent (voir e.g. [4]). Nous présenterons ici deux problèmes d'optimisation particuliers auxquels nous avons récemment contribué et qui visent à mieux modéliser et optimiser un système électrique fortement renouvelable. Dans le premier [5] il s'agit d'optimiser un ensemble de moyens de production et de stockage dans le réseau de distribution sous des contraintes liées au réseau. La technique proposée repose sur une relaxation assez classique des contraintes pour obtenir un cône du second ordre et notre contribution est une procédure itérative de coupes dont nous démontrons qu'elles sont exactes et nous rapprochent de la solution du problème non relaxé. Dans le deuxième problème [6] il s'agit de dimensionner un stockage, la technique proposée s'inspire du principe de Bellman et d'algorithmes de transformée de Legendre rapide, elle permet une résolution du problème de gestion du stockage en un temps quasi linéaire.

Références

[1] Gaillard, P. & Goude, Y., *Forecasting Electricity Consumption by Aggregating Experts ; How to Design a Good Set of Experts*, Antoniadis, A. ; Poggi, J.-M. ; Brossat, X. (Eds.) Forecasting Electricity Consumption by Aggregating Experts ; How to Design a Good Set of Experts Modeling and Stochastic Learning for Forecasting in High Dimensions, Springer International Publishing, 2015, 217, 95-115.

[2] Fischer, A., Montuelle, L., Mougeot, M. & Picard, D., *Real-time wind power forecast*, Preprint, 2016.

[3] Devijver, E., Goude, Y. & Poggi, J.-M., *Clustering electricity consumers using high-dimensional regression mixture models*, arXiv :1507.00167.

[4] Ademe, *«Vers un mix 100% renouvelable»*, `http://mixenr.ademe.fr/`.

[5] S.Y. Abdelouadoud, R. Girard, F.P. Neirac, T. Guiot, *Optimal power flow of a distribution system based on increasingly tight cutting*, planes added to a second order cone relaxation, Inter. J. of Electrical Power & Energy Systems, vol. 69, 2015, pp. 9–17.

[6] R. Girard, V. Barbesant, F. Foucault and G. Kariniotakis, *Fast dynamic programming with application to storage planning*, 2014 IEEE PES T&D Conference and Exposition, Chicago, IL, USA, 2014, pp. 1-5.

Groupe 3
Mardi 10:00-11:55, Amphi 8

Session : Applications statistiques des processus ponctuels déterminantaux

Organisée par
Rémi Bardenet, CNRS & CRIStAL, Université de Lille

Résumé. Les processus ponctuels déterminantaux (DPP) ont récemment fait leur apparition en statistiques. Ils encodent de la répulsivité entre les points d'un échantillon, tout en possédant des propriétés statistiques et computationnelles attractives que n'ont pas les processus de Gibbs généraux. Dans cette section, nous proposons quelques applications choisies des DPP en statistique, comme outil de modélisation et d'analyse.

1 – Modélisation et statistique des processus ponctuels déterminantaux

Frédéric Lavancier, Laboratoire Jean Leray, Université de Nantes & Inria Rennes.

Les processus ponctuels déterminantaux (DPPs) sont des objets connus depuis longtemps en probabilité, notamment pour leur rôle joué dans l'étude des valeurs propres de matrices aléatoires. De façon générale, les DPPs forment une classe de processus présentant de la dépendance négative. Leur utilisation en statistique est relativement récente. Définis sur un ensemble discret, les DPPs sont utilisés en machine learning ou en théorie des sondages. Définis sur un ensemble continu, les DPPs permettent l'échantillonnage de points régulièrement espacés (plus que dans le cas indépendant) et sont bien adaptés à la modélisation des processus ponctuels spatiaux répulsifs. Je montrerai dans cet exposé que les DPPs présentent en effet des propriétés remarquables pour leur utilisation en statistique : ils peuvent être simulés de façon exacte, des modèles paramétriques relativement flexibles peuvent être construits facilement, tous leur moments sont connus et l'inférence peut être conduite de manière standard (maximum de vraisemblance, minimum de contraste,...). J'insisterai particulièrement sur leur utilisation en statistique spatiale, domaine dans lequel les processus ponctuels répulsifs sont généralement modélisés par des processus de Gibbs. En comparaison, la simulation des modèles de Gibbs repose sur des algorithmes de type MCMC et leurs moments ainsi que la vraisemblance sont inconnus. En revanche, la flexibilité de modélisation offerte par les DPPs est moindre que celle offerte par les modèles de Gibbs, ce que je montrerai en déterminant le DPP "le plus répulsif". Cet exposé est basé sur des travaux effectués en collaboration avec Christophe Biscio , Jesper Møller et Ege Rubak de l'université d'Aalborg.

2 – Plans de sondages déterminantaux

Xavier Mary et Vincent Loonis, Modal'X, Université Paris Ouest et INSEE.

Nous proposons l'étude d'une nouvelle classe de plans de sondages, basée sur les processus déterminantaux. Les spécificités liées aux sondages (plans de taille fixe, probabilités d'inclusion simple fixées) seront examinées. Dans un deuxième temps, les propriétés des estimateurs

basés sur ces plans de sondages seront étudiées. Enfin, la question de plans “optimaux” sera abordée.

3 – Grandes matrices de Wishart corrélées : fluctuations et indépendance asymptotique aux bords

Jamal Najim, CNRS & Université Paris-Est Marne-la-Vallée.

On étudie le comportement asymptotique des valeurs propres de grandes matrices (complexes) de Wishart corrélées aux bords du spectre limite et autour des points intérieurs, de densité nulle (cusp points).

Pour ce modèle matriciel, le support de la distribution spectrale limite peut présenter plusieurs composantes connexes. Sous des hypothèses raisonnables, nous montrons que les valeurs propres extrémales, i.e. celles qui convergent presque sûrement vers les bords du spectre limite fluctuent selon la loi de Tracy-Widom, à la vitesse $N^{2/3}$; nous montrons également que les valeurs propres autour d'un « cusp point » fluctuent selon une distribution décrite par le noyau de Pearcey. De plus, étant donnés plusieurs bords (strictement positifs) du support, nous montrons que les valeurs propres extrémales associées sont asymptotiquement indépendantes. Enfin, nous étudions le cas où le bord gauche est l'origine. Dans ce cas, nous montrons que la plus petite valeur propre fluctue selon la distribution décrite par le noyau de Bessel (« hard edge Tracy-Widom »). En guise d'application, nous étudions le comportement asymptotique du nombre de conditionnement (rapport entre la plus grande et plus petite valeur propre).

Il s'agit d'un travail conjoint avec Walid Hachem (Télécom Paristech) et Adrien Hardy (Université de Lille), fondé sur les articles [1, 2].

4 – Processus ponctuel de β-Ginibre et déploiement d'un réseau cellulaire

Aurélien Vasseur, Télécom ParisTech

On souhaite valider le processus ponctuel de β-Ginibre comme modèle réaliste pour la distribution des locations de stations de base dans un réseau cellulaire. Le processus de β-Ginibre est un processus déterminantal dans lequel la répulsion est controlée par le paramètre β : en particulier, lorsque β tend vers zéro, le processus converge vers un processus de Poisson. Les simulations sur des données réelles collectées à Paris montrent que les locations des stations de base peuvent être “fittées” avec un processus de β-Ginibre, ceci à l'aide d'un outil statistique spécifique aux processus ponctuels : la fonction J. L'observation des résultats permet en outre de constater que la superposition indépendante de processus ponctuels de β-Ginibre tend vers un processus de Poisson. Afin de démontrer ce résultat de façon théorique, on introduit une topologie relativement fine sur les processus ponctuels et associée à une distance : la distance de Kantorovich-Rubinstein. La convergence d'une superposition indépendante de processus ponctuels déterminantaux, puis d'une suite de processus de β-Ginibre vers un processus de Poisson sont ainsi établies.

Références

[1] Hachem, Hardy, Najim, *Large Complex Correlated Wishart Matrices : Fluctuations and Asymptotic Independence at the Edges.*, Annals of Probability, 2016, 44(13).

[2] Hachem, Hardy, Najim, *Large Complex Correlated Wishart Matrices : The Pearcey Kernel and Expansion at the Hard Edge.*, Electron. J. Probab., 2016, 1(36).

Groupe 3
Mardi 10:00-11:55, Amphi 9

Session : Modèles de Polymères

Organisée par
Quentin Berger, Université Pierre et Marie Curie

Résumé. Cette session porte sur différents modèles de polymères (et interfaces), et l'étude de leurs interactions soit avec leur milieu – polymères dirigés en environnement aléatoire, modèles d'accrochage –, soit avec eux-mêmes – polymères chargés ou partiellement dirigés en auto-interaction. Le but est de comprendre comment, en fonction des paramètres du modèle, les interactions modifient les configurations typiques du polymère, et d'étudier les caractéristiques de la transition de phase le cas échéant.

1 – Désordre et phénomènes critiques : les modèles d'accrochage

Giambattista Giacomin, Université Paris Diderot

Les modèles exactement solubles possèdent un rôle central dans le progrès de la la mécanique statistique, en particulier pour comprendre les transitions de phase et les phénomènes critiques. Mais ces modèles sont très particuliers, et souvent peu réalistes : montrer la stabilité des résultats lorsque l'on introduit des « impuretés » – ou « désordre » – est vite devenu un enjeu très important. Je présenterai quelques prédictions remarquables, basées sur l'approche du « groupe de renormalisation », dans le contexte des modèles d'accrochage où l'on observe une transition dite de localisation. Les modèles d'accrochage forment une vaste classe, qui contient notamment des modèles de polymères et d'interfaces. C'est précisément pour ces modèles que l'effet du désordre sur la transition a pu être appréhendé, et les résultats obtenus confirment et vont parfois au delà des prédictions physiques.

2 – Polymères chargés

Julien Poisat, Université Paris Dauphine

On s'intéresse à un polymère non-dirigé représenté par la trajectoire d'une marche simple sur le réseau $\mathbb{Z}$ et portant des charges aléatoires et i.i.d. Chaque auto-intersection du polymère donne une contribution au hamiltonien égale au produit des charges qui se rencontrent. La loi jointe du polymère et des charges est donnée par la mesure de Gibbs associée à ce hamiltonien. Les paramètres sont le biais des charges et la température du système. À l'aide d'une formule de Ray-Knight pour les temps locaux de la marche simple, nous obtenons une représentation spectrale pour l'énergie libre annealed du polymère, ainsi que l'existence d'une courbe critique dans le diagramme de phase, séparant une phase balistique d'une phase sous-balistique. La transition de phase est du premier ordre. Nous obtenons également une loi des grands nombres, une théorème central limite et un principe de grande déviation, ainsi que le comportement de l'énergie libre à faible couplage. (basé sur un article en collaboration avec Francesco Caravenna, Frank den Hollander et Nicolas Pétrélis [1].) Si le temps le permet,

j'évoquerai des conjectures sur le cas annealed multi-dimensionnel. (travail en préparation avec Quentin Berger et Frank den Hollander.)

3 – Localisation pour le polymère log-gamma

Vu-Lan Nguyen, Université Paris Diderot

De manière générale, les polymères dirigés en environnement aléatoire sont « localisés » dans la phase dite de fort désordre. Dans cet exposé, basé sur un travail en collaboration avec F. Comets [2], nous considérerons le modèle de polymère dirigé en environnement log-gamma introduit récemment par Seppalainen, qui a la particularité d'être exactement soluble. Pour la version point-à-ligne du modèle stationnaire, la localisation peut être exprimée comme le piégeage du point final dans un potentiel donné par un marche aléatoire indépendante.

4 – Limites d'échelle de la marche partiellement dirigée en auto-interaction

Nicolas Pétrélis, Université de Nantes

La marche aléatoire partiellement dirigée en auto-interaction a été introduite en 1979 par Zwanzig et Lauritzen pour étudier la transition d'effondrement d'un homopolymère plongé dans une solvant répulsif. Jusqu'à récemment son étude était réalisée à l'aide de méthodes purement combinatoires. En dimension 2, une représentation probabiliste appropriée permet néammoins d'aller plus loin dans la compréhension de la structure géométrique adoptée par la marche dans chacun de ses trois régimes (étendu, critique et effondré). Nous présenterons dans cet exposée les limites d'échelles obtenues dans chaque régime et nous considérerons un modèle alternatif non dirigé.

Références

[1] F. Caravenna, F. den Hollander, N. Pétrélis et J. Poisat, *Annealed scaling for a charged polymer*, Mathematical Physics, Analysis and Geometry, à paraître.

[2] F. Comets et Vu-Lan Nguyen, *Localization in log-gamma polymers with boundaries*, Probab. Th. Rel. Fields., à paraître.

Groupe 3
Mardi 10:00-11:55, Amphi 11

Session : Optimisation stochastique : théorie et méthodes

Organisée par
Gersende Fort, LTCI, CNRS et Télécom ParisTech

1 – On Complexity of Convex Stochastic Optimization

Anatoli Juditsky, LJK, Univ. Grenoble Alpes

We present an overview of complexity results of convex stochastic optimization. We discuss lower information bounds and compare known complexity estimates for Sample Average Approximation and Stochastic Approximation. We also consider problems with functional constraints and briefly discuss the problem of estimating the optimal value of the stochastic problem.

2 – Sur la complexité de l'identification du meilleur bras sous contrainte de risque dans un modèle de bandits

Aurélien Garivier, IMT, Univ. Toulouse 3

Nous considérons un modèle d'optimisation discrète où, à chaque instant, le choix d'une option donne accès à une observation bruitée de la valeur associée. Nous donnons une estimation précise du nombre de tirages nécessaires pour identifier avec un risque δ donné l'option ayant la plus grande valeur moyenne associée : nous donnons une borne inférieure (qui implique elle-même un problème d'optimisation), et nous décrivons un algorithme (appelé "Track-and-Stop") qui atteint asymptotiquement cette borne inférieure quand le risque δ tend vers 0.
Travail joint avec Emilie Kaufmann, présenté à la conférence COLT 2016 (http ://arxiv.org/abs/1602.04589)

3 – On the Online Frank-Wolfe Algorithms

Jean Lafond, LTCI, Télécom ParisTech

In this paper, the online variants of the classical Frank-Wolfe algorithm are considered. We consider minimizing the regret with a stochastic cost. Our online algorithms only require *simple* iterative updates and a *non-adaptive* step size rule, in contrast to the *hybrid* schemes commonly considered in the literature. The proofs rely on bounding the duality gaps of the online algorithms. Several novel results are derived. The regret bound and anytime optimality for a strongly convex stochastic cost are shown to be as fast as $\mathcal{O}(\log^3 T/T)$ and $\mathcal{O}(\log^2 T/T)$, respectively, where T is the number of rounds played. Moreover, the online projection-free algorithms are shown to converge even when the loss is *non-convex*, i.e., the algorithms find a stationary point to the stochastic cost as $T \to \infty$. Numerical experiments on realistic data sets are presented to support our theoretical claims.

4 – De l'arrêt optimal à l'optimisation stochastique

Jérôme Lelong, LJK, Université Grenoble Alpes

Nous considérons le problème d'arrêt optimal

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}[Z_\tau] \tag{1}$$

où $(Z_t)_{0 \leq t \leq T}$ est un processus stochastique càdlàg adapté à une filtration brownienne sous–jacente et $\mathcal{T}$ l'enemble des temps d'arrêt relatifs à cette filtration à valeurs dans $[0, T]$. Ce problème est souvent abordé en discrétisant Z sur une grille de temps et en résolvant l'équation de programmation dynamique associée à l'enveloppe de Snell de ce processus à temps discret.

En utilisant la formulation duale de ce problème introduite par [1] et [2], il est possible de réécrire (1) comme la solution d'un problème d'optimisation stochastique en dimension infinie. Dans ce travail, nous proposons une approximation en dimension finie et étudions la convergence de la solution approchée vers la vraie solution. L'algorithme obtenu peut être implémenté très efficacement sur une architecture multi–processeur comme le montrent les mesures de performance que nous avons effectuées. Nous appliquons ce résultat au calcul du prix des options américaines.

Références

[1] M. B. Haugh and L. Kogan, *Pricing american options : a duality approach*, *Operations Research*, 2004, 52(2).

[2] L. C. G. Rogers, *Monte Carlo valuation of American options*, *Mathematical Finance*, 2002, 12(3).

Groupe 4
Mardi 13:30-15:25, Amphi 7

Session : Statistique et Ecologie

Organisée par
Marie-Pierre Etienne, AgroParisTech

Résumé. Cette session vise à offrir un aperçu des questions actuelles en écologie qui nécessitent des développements méthodologiques et font ainsi émerger de nouvelles questions statistiques.

1 – Enjeux et difficultés des sciences participatives

Camille Coron, Laboratoire de Mathématiques d'Orsay

Au cours des vingt dernières années, la récolte de données faisant intervenir des citoyens s'est beaucoup développée, notamment en écologie, transport, météorologie, consommation d'énergie,... Les données qui en sont issues sont généralement très nombreuses (en particulier beaucoup plus nombreuses que les données standardisées obtenues par les scientifiques ou par les professionnels du domaine concerné), mais manquent souvent d'exactitude, de précision, et d'homogénéité. L'utilisation et la valorisation de ces données participatives est donc actuellement un enjeu très important. Je présenterai notamment dans cet exposé un travail de recherche mené avec Benjamin Auder (Université Paris Sud), Christophe Giraud (Université Paris Sud), Romain Julliard (Museum National d'Histoire Naturelle) et Clément Calenge (Office National de la Chasse et de la Faune Sauvage), dans lequel nous combinons données standardisées et données participatives, pour estimer des abondances relatives d'espèces.

2 – Impact du réseau social dans un modèle d'échange de graines dans une population finie

Pierre Barbillon, AgroParisTech

Les échanges de graines entre paysans est un sujet d'étude important dans la mesure où ils ont une influence sur l'évolution de la diversité des variétés cultivées. Ces échanges sont structurés par un réseau social entre paysans. Afin de mieux comprendre ce processus dynamique d'échange et son importance, nous proposons d'étudier un modèle stochastique d'extinction-colonisation prenant en compte la complexité du réseau social : les échanges ne sont supposés possibles qu'au travers un réseau social fixé tandis qu'un phénomène d'extinction peut intervenir aléatoirement pour chaque paysan à chaque génération. Il est alors possible d'explorer l'influence des propriétés topologiques du réseau sur la persistance d'une variété donnée dans le réseau au bout d'un nombre de génération fixé. Nous nous concentrons principalement sur quatre types de réseaux sociaux afin de décrire des systèmes d'organisation différents. Nous prenons en compte le fait que le nombre de fermes est fini et donc responsable de variabilité dans les résultats. Ceux-ci sont obtenus par calcul exact lorsque le nombre de fermes est petit et par simulation autrement. La précision des résultats de simulation est améliorée par un filtre particulaire ou par des techniques de splitting.

3 – Geostatistics for point processes : predicting the intensity of ecological point process data

Edith Gabriel, Laboratoire de Mathématique, Université d'Avignon

In many ecological studies, mapping the intensity of animal or plant species is cumbersome as soon as these objects are not accessible by automated methods, as image analysis for example. The knowledge at large scale of the underlying process variability can then only be obtained through sampling and spatial prediction. Here, we aim to estimate the local intensity of a point process at locations where it has not been observed using the best linear unbiased combination of the point process realization on sampled windows. We show that the weight function associated to the estimator is the solution of a Fredholm equation of second kind. Both the kernel and the source term of the Fredholm equation are related to the second order characteristics of the point process. We propose here to restrict the solution space of the Fredholm equation to that generated by linear combinations of (i) step functions, which lead to a direct solution as we then get adapted kriging equations (Gabriel et al, 2016) ; and (ii) basis splines, which provide a continuous approximation. Results will be illustrated to estimate and predict the intensity of Montagu's Harriers' nest locations in a region of France.

4 – Estimation de paramètres de mouvement en écologie : Application aux modèles à base d'équations diférentielles stochastiques

Pierre Gloaguen, AgroParisTech

L'étude et la compréhension du mouvement des individus en écologie a pour but de 1) Comprendre la structuration des territoires des individus, 2) Aider à la mise en place de mesures adéquates de gestion pour la préservation de ces territoires.
Le déplacement des individus est étudié au travers de mesures de mouvement (balises GPS), afin de mettre en évidence les mécanismes structurant leurs déplacements.
Nous nous intéressons ici aux modèles où l'ensemble des positions d'un individu au cours du temps $(X_t)_{t\geq 0}$ est modélisé par une diffusion solution d'une équation différentielle stochastique (EDS) du type :

$$dX_t = b(X_t, \theta)dt + \sigma dW_t$$

où $b(\cdot)$ est la fonction de dérive dépendant de paramètres θ et σ un paramètre réglant le caractère diffusif du mouvement. L'objectif est alors d'inférer θ à partir d'observations discrètes de positions au cours du temps.
Plusieurs algorithmes d'inférence ont été proposés dans le cadre d'EDS observées à temps discret, le plus souvent pour de hautes fréquences d'échantillonage.
Nous proposons ici d'étudier la performance de ces algorithmes pour des processus d'échantillonage typique de données GPS pour l'analyse du mouvement, à savoir irrégulier et à moyenne, voire basse fréquence.

Groupe 4
Mardi 13:30-15:25, Amphi 8

Session : Inférence géométrique et topologique, estimation de formes

Organisée par
Clément Levrard, Université Paris Diderot
Bertrand Michel, Université Pierre et Marie Curie

Résumé. Cette session traite d'analyse de données et d'inférence pour des données comportant de l'information géométrique ou topologique. Il y sera notamment question d'analyse topologique des données, de reconstruction, d'estimation de support et d'estimation de surfaces.

1 – Une introduction à l'algorithme Mapper

Steve Oudot, INRIA Saclay

Mapper est sans doute l'outil de l'analyse topologique de données le plus largement utilisé dans les sciences appliquées et dans l'industrie. Son utilisation principale se trouve en analyse exploratoire, où elle permet d'obtenir de nouvelles représentations des données donnant une vision plus haut-niveau de leur structure qu'un simple clustering. La sortie de Mapper prend la forme d'un graphe dont les noeuds représentent des sous-populations homogènes des données et dont les arcs représentent certaines relations de similarité. Pour autant, l'instabilité intinsèque de cette sortie et la difficulté du réglage des paramètres qui en decoule sont un réel frein à l'utilisation de la méthode. En pratique les utilisateurs en sont réduits à une utilisation purement heuristique, avec un réglage des paramètres à l'aveugle. Cet exposé se concentrera sur l'étude de la structure des graphes fournis par Mapper et tentera d'éclairer d'une lumière nouvelle leurs propriétés de stabilité, avec à la clé de nouveaux outils théoriques pour le réglage des paramètres.

2 – Reconstruction simpliciale de variétés par estimation d'espaces tangents

Eddie Aamari, INRIA Saclay

On considère le problème de reconstruction de variété dans un cadre semi-asymptotique. Sous des hypothèses de régularité géométrique, on proposera un estimateur calculable $\hat{M}$ du support M d'une distribution inconnue dont on observe un n-échantillon i.i.d. L'estimateur $\hat{M}$ possède la même topologie que M et l'approxime, pour la perte donnée par la distance de Hausdorff, à la vitesse optimale au sens minimax. La méthode développée se base sur le complexe de Delaunay tangentiel. Après avoir réduit la question à l'estimation des espaces tangents de M, le problème est résolu avec des ACP locales. On examinera la robustesse de la méthode dans le cadre d'un modèle de mélange, où une technique de débruitage reposant sur l'ACP locale sera présentée.

3 – Un test d'existence du bord du support

Catherine Aaron, Université Blaise Pascal, Clermont-Ferrand

Lorsque l'on observe des données tirées sur $\mathbb{R}^d$ avec une densité absolument continue par rapport a la mesure de Lebesgue et à support compact, les méthodes classiques d'estimation du support fournissent un estimateur naturel de la frontière du support. En revanche, lorsque l'on observe des données tirées sur une sous-variété compacte de dimension $d' < d$, d'une part les estimateurs du support ne fournissent pas, le plus souvent, un estimateur naturel de la frontière du support et de plus, la question même de l'existence de la frontière se pose. On présentera dans cet exposé un test d'existence de la frontière. La statistique de test utilisée est basée sur la distance maximum entre un point de l'échantillon et le barycentre de ses k-plus proches voisins. On montrera que sous certaines conditions de régularité qui seront discutées, la p-value peut être estimée via une loi du chi2. Un critère de choix du paramètre k sera proposé. Les performances seront illustrées par des simulations.

4 – Estimations de surfaces moyennes via les métamorphose de formes fonctionnelles

Benjamin Charlier, Université de Montpellier

Une forme fonctionnelle `fshape` est une surface sur laquelle est définie une fonction à valeurs réelles. Ce type de données, très courant en imagerie médicale, reste complexe à analyser d'un point de vue statistique. Pour analyser un jeu de données composé de fshapes, nous proposons de modéliser et de quantifier les variations géométriques et fonctionnelles de manières jointes. Dans cet exposé, nous décrivons un cadre théorique et numérique pour calculer une moyenne de surfaces fonctionnelles à la manière des modèles statistiques de déformations. Le cadre mathématique permet de montrer que les formulations variationnelles proposées pour résoudre ce problème possèdent bien des solutions. Une méthode de résolution algorithmique est implémentée dans le logiciel `fshapesTk` qui est disponible en ligne.

Groupe 4
Mardi 13:30-15:25, Amphi 9

Session : Matrices et Graphes Aléatoires

Organisée par
Camille MALE, Université Paris Descartes, CNRS

Résumé. Lors de cette session dédiée à la théorie spectrale des matrices aléatoires, deux thèmes importants actuels seront déclinés. D'une part, des problèmes concernant la répartition des valeurs propres de matrices d'adjacence de graphes aléatoires. Le modèle du graphe d'Erdös-Rényi est formellement très proche du modèle classique des matrices des matrices de Wigner (matrices symétriques à entrées indépendantes). Cependant, de nouveaux phénomènes apparaissent lorsque le nombre moyen de voisins d'un sommet pris au hasard n'est pas de l'ordre du nombre total de sommet. D'autre part, des questions relatives au lien entre grandes matrices aléatoires et la théorie des probabilités libres. Cette dernière permet de décrire la répartition des valeurs propres de polynômes en des matrices aléatoires indépendantes.

1 – Spectre de graphe aléatoires

Laurent Ménard, Paris Ouest Nanterre La Defense.

Nous ferons un survol des résultats connus (et de questions ouvertes) sur le spectre des matrices d'adjacence de graphes aléatoires dilués. On se concentrera sur le graphe d'Erdös-Rényi (n sommets, et on met une arête entre deux sommets avec probabilité $\frac{c}{n}$ où c est un paramètre fixé), qui malgré sa simplicité se révèle déjà difficile à étudier. L'exposé sera peu technique et accessible sans prérequis sur les matrices aléatoires ou les graphes aléatoires.

2 – On eigenvalue distribution of random matrix ensembles related with the Ihara zeta function of large random graphs

Oleksiy Khorunzhiy, Laboratoire de Mathématiques de Versailles

We consider real symmetric $n \times n$ random matrices of the form

$$H^{(n,\rho)}(v) = \frac{v^2}{\rho} B^{(n,\rho)} - \frac{v}{\sqrt{\rho}} A^{(n,\rho)},$$

where $A^{(n,\rho)}$ is the adjacency matrix of the ensemble of the Erdös-Rényi random graphs $\{\Gamma^{(n,\rho)}\}$, $(A^{(n,\rho)})_{ij} = a_{ij}$ equal to 1 with probability ρ/n and 0 otherwise, and $B^{(n,\rho)} = diag\,(\sum_{k=1}^{n} a_{ik})$. Matrix $H^{(n,\rho)}$ generalizes the discrete analog of the Laplace operator of $\Gamma^{(n,\rho)}$. The eigenvalue distribution of $\{H^{(n,\rho)}\}$ determines the average value of the Ihara zeta function of $\Gamma^{(n,\rho)}$.

By using the classical method of moments, we show that the limiting eigenvalue distribution of $H^{(n,\rho)}$ exists as $n, \rho \to \infty$ and is given by a shifted semicircle law. Similar statement is proved in the case when the graphs Γ are given by the random graphs of the long-range percolation model. This model can be considered as a generalization of the Erdös-Rényi ensemble of random graphs $\{\Gamma^{(n,\rho)}\}$.

3 – Grandes déviations de traces de matrices aléatoires

Fanny Augeri, Institut de Mathématiques de Toulouse.

Partant du théorème de Wigner qui donne la convergence presque sûre des traces des matrices de Wigner vers les nombres de Catalan, on s'intéressera aux grandes déviations des traces les modèles suivants : les matrices de Wigner à entrées Gaussiennes, les matrices de Wigner dites "sans queues Gaussiennes", et les β-ensembles associés à un potentiel convexe et à croissance polynomiale.

4 – Comportement en temps long de l'équation de Fokker-Planck libre pour certains potentiels non convexes

Mylène Maïda, Laboratoire Paul Painlevé, Lilles

On expliquera dans cet exposé comment des idées et des techniques venues des probabilités libres ou de l'étude des polynômes orthogonaux permettent d'aborder la question de la convergence vers l'équilibre de l'équation dite des milieux granulaires avec interaction logarithmique (dite aussi équation de Fokker-Planck ou MacKean-Vlasov non linéaire) pour certains potentiels non convexes. Il s'agit d'un travail en commun avec Catherine Donati-Martin et Benjamin Groux.

Groupe 4
Mardi 13:30-15:25, Amphi 11

Session : Optimisation stochastique en action

Organisée par
Bruno Gaujal, INRIA

Résumé. Cette session montre comment une grande variété de techniques d'optimisation stochastique (respectivement l'analyse non-lisse, les jeux en champ moyen, les politiques d'indice et les méthodes de décomposition) peuvent être les bons outils pour résoudre les problèmes liés à la production d'énergie, les marchés de l'énergie, l'exploration des réseaux sociaux et de gestion distribuée des barrages.

1 – Nonsmooth analysis for stochastic optimization : theory, algorithms and applications in energy

Jérôme Malick, CNRS, Univ. Grenoble Alpes

Prendre des décisions optimales dans un contexte incertain est complexe : une modélisation fine de l'aléatoire peut donner naissance à des problèmes d'optimisation difficiles, potentiellement non-convexes, hors de porté d'approches frontales. C'est le cas par exemple pour l'optimisation de la production d'électricité incluant les énergies renouvelables intermittentes sujettes à de l'aléa météo. Dans cet exposé, je vais présenter des outils mathématiques et algorithmiques issus de l'analyse non-lisse pour résoudre des problèmes d'optimisation stochastique hétérogènes et de grande taille. J'illustrerai l'intérêt de ces méthodes sur des problèmes de production d'EDF.

2 – Optimization in Energy Markets : the Good, the Bad and the Ugly

Nicolas Gast, INRIA

Electricity networks are evolving. On the one hand, the generation becomes more volatile and less dispatchable because of renewable energies, like solar or wind. On the other hand, communication capabilities makes possible a real-time control of the consumption, via real-time prices or congestion signal mechanisms.

In this talk, I will review some results that concern real-time electricity markets. I will question the assumption that a free and honest market would lead to an optimal use of the generation and storage. I will show that under quite general conditions, the market leads to an efficient use of the resources. This result relies on tools from stochastic optimization and lagrangian decomposition. I will also show that these markets have several issues : prices are very volatile and might lead to under-investment.

3 – Whittle Index Policy for Crawling Ephemeral Content

Konstantin Avrachenkov, INRIA Sophia Antipolis

We consider the task of scheduling a crawler to retrieve from several sites their ephemeral content. This is content, such as news or posts at social network groups, for which a user typically loses interest after some days or hours. Thus development of a timely crawling policy for ephemeral information sources is very important. We first formulate this problem as an optimal control problem with average reward. The reward can be measured in terms of the number of clicks or relevant search requests. The problem in its exact formulation suffers from the curse of dimensionality and quickly becomes intractable even with moderate number of information sources. Fortunately, this problem admits a Whittle index, a celebrated heuristics which leads to problem decomposition and to a very simple and efficient crawling policy. We derive the Whittle index for a simple deterministic model and provide its theoretical justification. We also outline an extension to a fully stochastic model. This is a joint work with V.S. Borkar from IIT Bombay.

4 – Decomposition methods in stochastic optimization, with applications to Energy

Vincent Leclere, ENPC

Multistage stochastic optimization problem are hard to solve. Indeed, the extensive formulation (i.e. a deterministic equivalent problem) has a huge size. We claim that decomposition methods which allow to see this huge problem as a collection of coordinated smaller problems are an efficient way to address multistage stochastic optimization. Decomposition methods can be done in at least three ways : by scenarios (e.g. Progressive Hedging approaches), by time (e.g. dynamic programming approaches) or spatially by disconnecting subparts of the system. The Dual Approximate Dynamic Programming (DADP) algorithm we present fall in the last category and is motivated by the study of the optimal management of an hydroelectric valley composed of N linked dams. Dualizing the coupling constraint, and fixing a multiplier allow to solve the problem as N independant dams. Unfortunately in a stochastic setting this approach fails. The DADP algorithm rely on an approximation of the multiplier through a conditional expectation, allowing to efficiently solve the subproblems by dynamic programming. We present theoretical results and interpretation of the DADP algorithm as well as numerical results.

Groupe 5
Mercredi 10:30-12:25, Amphi 7

Session : Système de particules en interaction

Organisée par
Emmanuel Jacob, Ecole Normale Supérieure de Lyon

Résumé. La modélisation de nombreux phénomènes aléatoires, principalement inspirés de physique statistique, met en jeu un système de particules en interaction : chaque particule a un comportement stochastique et interagit avec les autres particules, également stochastiques.
Cette session entend donner un aperçu de la diversité des modèles que l'on est amenés à considérer, selon que l'accent est mis sur l'aspect d'un système d'équations différentielles stochastiques liées ou contraintes, sur l'aspect champ moyen où les interactions entre particules ont l'effet d'un milieu aléatoire, ou sur l'aspect d'un graphe particulier d'interaction, et de l'influence de cette structure de graphe sur l'évolution du système.

1 – Kinetically constrained spin models : some results and open issues

Cristina Toninelli, Université Pierre et Marie Curie

Kinetically constrained spin models (KCSM) are a class of interacting particle systems which have been introduced in physics literature to model liquid-glass and more general jamming transitions. The evolution of KCSM is given by a continuous time Markov process whose elementary moves are birth and death of particles. The key feature is that a move occurs only if the configuration satisfies a local constraint. We will show that the constraints induce the existence of clusters of blocked particles, the occurrence of several invariant measures, non-attractiveness, ergodicity breaking transitions and the failure of classic coercive inequalities to analyse relaxation to equilibrium. We will present some techniques which have been developed to establish the scaling of spectral gap and mixing times. We will conclude by presenting some open problems.

2 – Traveling waves pour le modèle de Kuramoto quenched.

Eric Luçon, Université Paris Descartes

Je parlerai du comportement en temps long du modèle de Kuramoto en milieu aléatoire. Il s'agit d'un système de N diffusions sur le cercle en interaction de type champ-moyen. Le désordre consiste en la donnée d'une fréquence aléatoire pour chacune des particules. Sur un intervalle de temps borné [0, T], il y a propagation du chaos (presque-sûrement par rapport au désordre) : la mesure empirique du système converge pour N grand vers la solution d'une équation de Fokker-Planck nonlinéaire, dont il est possible de calculer explicitement les solutions stationnaires. Le but de cet exposé est de montrer que sur une échelle de temps d'ordre $\sqrt{N}$, le système n'est plus auto-moyennant : les fluctuations du désordre induisent

des traveling waves, dont la direction et la vitesse dépendent du tirage du désordre.
Travail en commun avec Christophe Poquet (Lyon 1)

3 – Processus de contact et percolation par groupements cumulatifs

Laurent Ménard, Paris ouest

Nous présenterons le processus de contact (aussi appelé Susceptible-Infected-Susceptible) qui est un modèle classique de système de particules en interactions utilisé pour modéliser la propagation d'une infection sur un graphe. Ce processus admet une transition de phase sur un graphe infini : en fonction du taux d'infection, une épidémie disparaît ou non au cours du temps. Nous essaierons de voir comment détecter cette transition à l'aide d'un processus de percolation original. L'exposé sera peu technique et accessible sans prérequis sur le processus de contact ou la percolation.

4 – Marche renforcée, milieu aléatoire et système de particules

Xiaolin Zeng, Université Lyon 1 et ENS Lyon.

Nous nous intéressons au processus de sauts renforcé par site. Il s'agit d'une marche aléatoire en milieu aléatoire, où l'environnement est un potentiel aléatoire généralisant la loi inverse-gaussienne. Cette loi inverse-gaussienne peut être caractérisée comme la loi du temps auquel un mouvement brownien avec dérive touche 1, et nous proposons une caractérisation similaire concernant la marche renforcée.

Groupe 5
Mercredi 10:30-12:25, Amphi 8

Session : Méthodes parcimonieuses en apprentissage statistique

Organisée par
Farida Enikeeva, Université de Poitiers, Laboratoire de Mathématiques et Applications

Résumé. En apprentissage statistique, on est souvent confronté à des situations où la dimension du modèle p est plus grand que le nombre d'observations. Dans ce cas si on ne fait pas d'hypothèses additionnelles sur la structure des données, il est impossible de définir des procédures d'inférence statistique qui sont consistantes. Néanmoins, dans bien des applications, l'information a priori que l'on a sur les données nous permet de faire une hypothèse dite de parcimonie, qui consiste à supposer que la dimension effective du modèle est beaucoup plus petite que p. Une approche fructueuse consiste alors à reformuler ce problème d'apprentissage sparse sous forme d'un problème d'optimisation, en faisant intervenir une pénalité permettant de traduire cette hypothèse de parcimonie que l'on a faite sur le modèle. Ainsi les méthodes qui seront présentées dans cette session font tout à la fois intervenir de la statistique et de l'optimisation. Il s'agira à chaque fois d'obtenir un bon compromis entre le contrôle du risque statistique et des algorithmes pouvant être exécutés avec une faible complexité de calcul.

1 – GAP Safe screening rules for sparse multi-task and multi-class models

Alexandre Gramfort, Telecom ParisTech

High dimensional regression benefits from sparsity promoting regularizations. Screening rules leverage the known sparsity of the solution by ignoring some variables in the optimization, hence speeding up solvers. When the procedure is proven not to discard features wrongly the rules are said to be *safe*. In this paper we derive new safe rules for generalized linear models regularized with ℓ_1 and ℓ_1/ℓ_2 norms. The rules are based on duality gap computations and spherical safe regions whose diameters converge to zero. This allows to discard safely more variables, in particular for low regularization parameters. The GAP Safe rule can cope with any iterative solver and we illustrate its performance on coordinate descent for multi-task Lasso, binary and multinomial logistic regression, demonstrating significant speed ups on all tested datasets with respect to previous safe rules. Joint work with Eugene Ndiaye, Olivier Fercoq and Joseph Salmon.

2 – Apprentissage de la structure pour la parcimonie structurée

Nino Shervashidze, INRIA–SIERRA, Laboratoire d'Informatique de l'ENS, Paris

La parcimonie structurée a récemment émergé au croisement des statistiques, de l'apprentissage et du traitement du signal comme une approche prometteuse pour l'inférence en grande dimension. Les méthodes actuelles d'apprentissage sous l'hypothèse de parcimonie structurée

supposent que l'on sait a priori comment pondérer (ou pénaliser) chaque sous-ensemble de variables pendant le processus de sélection de sous-ensembles, ce qui n'est pas le cas en général. Nous proposons une approche bayésienne pour inférer des poids de sous-ensembles à partir de données. Nous modélisons les poids des groupes avec des hyperparamètres de lois a priori à queue lourde sur des groupes de variables et nous dérivons une approximation bayésienne variationnelle afin d'inférer ces hyperparamètres à partir de données. Nos expériences permettent de vérifier que notre approximation est capable de retrouver les hyperparamètres quand les données sont générées par le modèle, et montrent l'intérêt de l'apprentissage des poids dans le cadre de problèmes de débruitage synthétiques et réels.

3 – Complexity analysis of the Lasso regularization path

Julien Mairal, INRIA–LEAR, Grenoble

We will study an intriguing phenomenon related to the regularization path of the Lasso estimator. The regularization path of the Lasso can be shown to be piecewise linear, making it possible to "follow" and explicitly compute the entire path. We analyze this popular strategy, and prove that its worst case complexity is exponential in the number of variables. We then oppose this pessimistic result to an (optimistic) approximate analysis : We show that an approximate path with at most $O(1/\sqrt{\varepsilon})$ linear segments can always be obtained, where every point on the path is guaranteed to be optimal up to a relative ε-duality gap.

4 – Low Rank Matrix Completion with Exponential Family Noise

Jean Lafond, Telecom ParisTech

The matrix completion problem consists in reconstructing a matrix from a sample of entries, possibly observed with noise. A popular class of estimator, known as nuclear norm penalized estimators, are based on minimizing the sum of a data fitting term and a nuclear norm penalization. Here, we investigate the case where the noise distribution belongs to the exponential family and is sub-exponential. Our framework allows for a general sampling scheme. We first consider an estimator defined as the minimizer of the sum of a log-likelihood term and a nuclear norm penalization and prove an upper bound on the Frobenius prediction risk. The rate obtained improves on previous works on matrix completion for exponential family. When the sampling distribution is known, we propose another estimator and prove an oracle inequality w.r.t. the Kullback–Leibler prediction risk, which translates immediately into an upper bound on the Frobenius prediction risk. Finally, we show that all the rates obtained are minimax optimal up to a logarithmic factor.

Groupe 5
Mercredi 10:30-12:25, Amphi 9

Session : Inégalités de concentration

Organisée par
Pierre Youssef, Univ. Paris Diderot

Résumé. Les inégalités de concentration et le phénomène de concentration de la mesure constituent un outil puissant qui est utilisé dans différents domaines. L'exemple de base de ce phénomène est illustré par la loi des grands nombres qui affirme que la somme de variables aléatoires indépendantes se concentre autour de leur moyenne avec grande probabilité. Plus généralement, il s'agit souvent d'étudier la concentration d'une fonction d'un processus aléatoire autour de sa moyenne. Dans cette session, nous verrons certains types de ces inégalités et leurs applications en traitement du signal, géométrie des convexes en grande dimension, matrices aléatoires et graphes aléatoires.

1 – On the gap between RIP-properties and sparse recovery conditions

Guillaume Lecué, CNRS - ENSAE

We prove that iid random vectors that satisfy a rather weak moment assumption can be used as measurement vectors in Compressed Sensing, and the number of measurements required for exact reconstruction is the same as the best possible estimate exhibited by a random Gaussian matrix. We then show that this moment condition is necessary, up to a log log factor.
In addition, we explore the noisy setup and consider the problem of recovering sparse vectors from undetermined linear measurements via ℓ_p-constrained basis pursuit. Previous analyses of this problem based on generalized restricted isometry properties have suggested that two phenomena occur if $p \neq 2$. First, one may need substantially more than $s \log(en/s)$ measurements (optimal for $p = 2$) for uniform recovery of all s-sparse vectors. Second, the matrix that achieves recovery with the optimal number of measurements may not be Gaussian (as for $p = 2$). We present a new, direct analysis which shows that in fact neither of these phenomena occur. Via a suitable version of the null space property we show that a standard Gaussian matrix provides ℓ_q/ℓ_1-recovery guarantees for ℓ_p-constrained basis pursuit in the optimal measurement regime. Our result extends to several heavier-tailed measurement matrices. As an application, we show that one can obtain a consistent reconstruction from uniform scalar quantized measurements in the optimal measurement regime.
Joint work with Sjoerd Dirksen, Shahar Mendelson and Holger Rauhut. This talk is based on the two publications [1, 2].

2 – Small ball estimates for quasi-norms
Omer Friedland, Univ. Paris 6

We study two types of small ball estimates for random vectors in finite dimensional spaces equipped with a quasi-norm. In the first part, we obtain bounds for the small ball probability of random vactors under some smoothness assumptions on their density functions. In the second part, we obtain Littlewood-Offord type estimates for quasi-norms. This is a joint work with Ohad Giladi and Olivier Guédon.

3 – Inégalités de concentrations non commutatives dans un cadre de dépendance
Marwa Banna, Télécom ParisTech

On s'intéresse à des inégalités de déviation pour la plus grande valeur propre d'une somme de matrices aléatoires auto-adjointes. Plus précisément, on établit une inégalité de type Bernstein pour la plus grande valeur propre de la somme de matrices auto-ajointes, centrées et géométriquement β-mélangeantes dont la plus grande valeur propre est bornée. Ceci étend d'une part le résultat de Merlevède et al. (2009) à un cadre matriciel et généralise d'autre part le résultat de Tropp (2012) pour des sommes de matrices indépendantes. Travail en collaboration avec F. Merlevède et P. Youssef.

4 – La concentration de tirages pondérés sans remise
Anna Ben-Hamou, Université Paris Diderot, LPMA

Dans son article de 1963, Hoeffding a montré que la somme induite par des tirages uniformes sans remise dans une population finie était plus petite, en ordre convexe, que celle induite par des tirages avec remise. En particulier, les tirages sans remise concentrent "au moins aussi bien" que les tirages avec remise. Lorsque la taille de l'échantillon grandit, on peut même supposer que les tirages sans remise concentrent "largement mieux", la variance étant alors de l'ordre du nombre d'éléments non-tirés. Cela a été confirmé par Serfling en 1967. Dans cet exposé, nous tenterons de généraliser cette comparaison au cas de tirages pondérés, et notamment aux tirages dit biaisés par la taille, qui interviennent dans de nombreuses situations, comme par exemple dans le modèle de configuration pour les graphes aléatoires. Il s'agit d'un travail en commun avec Justin Salez et Yuval Peres.

Références

[1] Guillaume Lecué and Shahar Mendelson, *Sparse recovery under weak moment property*, To appear in Journal of the European mathematical society, 2015.

[2] Sjoerd Dirksen, Guillaume Lecué and Holger Rauhut, *On the gap between RIP-properties and sparse recovery conditions*, To appear in IEEE Transactions on Information Theory, 2015..

Groupe 5
Mercredi 10:30-12:25, Amphi 11

Session : Changement climatique

Organisée par
Anne-Catherine Favre, Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE)

Résumé. Il est important d'être en mesure de déterminer l'impact du changement climatique, par exemple sur les pluies extrêmes, les crues et l'écologie. Dans ce contexte des méthodes statistiques innovantes ont été devéloppées pour la détection et l'attribution du changement climatique. Cette session exposera ces méthodes et montrera des applications aux pluies extrêmes en Afrique de l'Ouest et en Méditerranée ainsi qu'à l'écologie en milieu boréal.

1 – How to revise return periods for record events in a changing climate

Philippe Naveau, Laboratoire des Sciences du Climat et de l'Environnement (LSCE)

Breaking a record simply means that the current observation exceeds all past measurements. Such a type of an event is regularly followed to a media frenzy and, in such instances, climatologists are often asked if the frequency of this record is different from previous ones.

This leads to the question of Detection and Attribution (D&A) ("Detection" is the process of demonstrating that climate has changed in some defined statistical sense, without providing a reason for that change and "Attribution" is the process of establishing the most likely causes for the detected change with some defined level of confidence, see the IPCC definition).

The field of statistics has become one of the mathematical foundations in D&A studies because computing uncertainties represent difficult inferential challenges when analyzing small probabilities.

In this context, we will give a brief overview on the main statistical concepts underpinning the D&A and proposes new methodological approaches to revise return periods for record events in a changing climate. We will show the advantages of our method throughout theoretical results and simulation studies.

2 – Tendances et contrastes régionaux du régime des pluies extrêmes en Afrique de l'Ouest depuis les années 1950

Théo Vischel, Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE)

L'étude de l'évolution des pluies extrêmes est au cœur d'enjeux à l'articulation entre la compréhension du climat de notre planète et ses impacts sociétaux qui nécessitent de définir des stratégies de gestion des risques hydrologiques. On aborde cette problématique en Afrique de l'Ouest, région caractérisée par de forts gradients pluviométriques et marquée par une évolution des pluies très contrastée au cours du dernier siècle. La détection de non-stationnarités

dans les séries d'extrêmes est rendue difficile par d'importants effets d'échantillonnage liés à la forte variabilité naturelle de la pluie et la faible quantité de données disponibles pour la documenter. On verra comment la théorie de valeurs extrêmes permet de s'accommoder de ces contraintes pour étudier l'évolution des précipitations en prenant en compte l'information locale (séries pluviométriques ponctuelles), tout en respectant une certaine cohérence régionale. On s'intéressera ici à documenter les contrastes régionaux de l'évolution de l'occurrence, l'intensité et la dépendance spatiale des précipitations extrêmes. Les résultats montrent que la tournure prise par le régime des précipitations au cours des 15 dernières années dans la région est typique d'un climat plus extrême : les quantités annuelles en relative hausse se sont accompagnées d'une diminution persistante du nombre de jours pluvieux et d'une progression des événements extrêmes. Cela pose de nombreuses questions aux climatologues pour comprendre les processus atmosphériques associés à cette évolution et aux hydrologues qui ont longtemps attribué l'augmentation récente des inondations en Afrique de l'Ouest à un changement d'usage des sols.

3 – Copules de valeurs extrêmes extra-paramétrisées : extension au cadre spatial

Julie Carreau, HydroSciences Montpellier (HSM)

Les copules de valeurs extrêmes sont motivées par la théorie des valeurs extrêmes multivariées. Cependant, la plupart des copules de haute dimension sont trop simplistes pour les applications. Récemment, une classe flexible de copules de valeurs extrêmes a été proposée en combinant deux copules de valeurs extrêmes avec une pondération prenant ses valeurs dans l'hyper-cube unitaire.

Dans une étude multi-sites, la dimension de la copule (et donc de l'hyper-cube) est le nombre de sites et cette approche extra-paramétrisée devient rapidement sur-paramétrisée. De plus, l'interpolation spatiale n'est pas évidente. Le but de ce travail consiste à étendre cette approche au cadre spatial. En considérant la pondération comme une fonction de covariables, la complexité du modèle est réduite. Par ailleurs, le modèle est défini en tout point de l'espace et peut être interprété en termes de distances.

Nous nous concentrons sur l'extension spatiale basée sur les copules de Gumbel et décrivons les structures de dépendance possibles du modèle. Nous appliquons le modèle d'abord sur des données synthétiques puis sur des données de précipitations dans la région Méditerranéenne. .

4 – Extreme cold winters and ecological impacts on northern forests

Anthony Davison, Chaire de statistique, EPFL

Global change is expected to have severe impacts on northern climates, opening impassable seaways and damaging infrastructure building on permafrost. It will also affect the life-cycles of insects, such as the moth Epirrita autumnata, whose larvae are kept in check by winters with minimum temperatures below $-36°$C. Since these larvae can devastate boreal forests, spatial prediction of future minimum temperatures is an important ingredient in understanding the potential ecological impact of warming. In this paper a Bayesian hierarchical approach is used to fit a spatial max-stable process to winter temperature data from around 20 stations in northern Finland, in which times of occurrences are included to enable efficient inference and prediction of the probabilities of major outbreaks in future years. The work is joint with Emeric Thibaud, Dan Cooley, Juha Aalto and Juha Heikkinen.

Lundi 29 Août

9:30 - 10:20 : Inscription autour d'un café
10:20 - 10:30 : Accueil et informations pratiques, *Amphi 11*
10:30 - 11:00 : Hommage à Jacques Neveu (Francis Comets, Etienne Pardoux)
11:00 - 12:00 : Conférence plénière, **Aurélien Ribes**
12:00 - 13:30 : Déjeûner, restaurant Diderot
13:30 - 15:25 : Sessions parallèles (groupe 1)

- PDMP pour la biologie (Romain Yvinnec et Florent Malrieu), *Amphi 7*
- Processus Gaussiens pour les expériences numériques (François Bachoc), *Amphi 8*
- Percolation (Marie Théret), *Amphi 9*
- Tests multiples (Magalie Fromont), *Amphi 11*

15:30 - 16:30 : Conférence pléniere, **François Delarue**
16:30 - 17:00 : Pause-Café
17:00 - 18:55 : Sessions parallèles (groupe 2)

- Grande dimension et génomique (Laurent Jacob), *Amphi 7*
- Statistiques pour les codes numériques (Merlin Keller), *Amphi 8*
- Probabilités et algorithmique (Irène Marcovici), *Amphi 9*
- Géométrie stochastique (David Coupier), *Amphi 11*

19:00 - 20:30 : Cocktail de bienvenue et session poster (hall sud)

Mardi 30 Août

8:30 - 9:30 : Conférence plénière, **Amaury Lambert**
9:30 - 10:00 : Pause-Café
10:00 - 11:55 : Sessions parallèles (groupe 3)

- Modèles statistiques autour de l'énergie (Jairo Cugliari), *Amphi 7*
- Applications statistiques des DPP (Rémi Bardenet), *Amphi 8*
- Modèles de polymères (Quentin Berger), *Amphi 9*
- Optimisation stochastique : théorie et méthodes (Gersende Fort), *Amphi 11*

12:00 - 13:30 : Déjeûner, restaurant Diderot
13:30 - 15:25 : Sessions parallèles (groupe 4)

- Modèles et inférence pour données écologiques (Marie-Pierre Etienne), *Amphi 7*
- Inférence géométrique et topologique (Bertrand Michel et Clément Levrard), *Amphi 8*
- Matrices et graphes aléatoires (Camille Male), *Amphi 9*
- Optimisation stochastique et applications (Bruno Gaujal), *Amphi 11*

15:30 - 16:00 : Pause-Café
16:00 - 16:30 : Prix Neveu, **Emilie Kaufmann**
16:30 - 17:30 : Conférence plénière, **Sylvie Huet**
17:30 - 18:30 : Assemblée Générale du groupe SMAI-MAS
19:30 - . . . : Dîner de conférence à la Bastille !

Mercredi 31 Août

9:00 - 10:00 : Conférence plénière, **Charles Bordenave**
10:00- 10:30 : Pause-Café
10:30 - 12:25 : Sessions parallèles (groupe 5)

- Système de particules en interaction (Emmanuel Jacob), *Amphi 7*
- Parcimonie dans l'apprentissage statistique (Farida Enikeeva), *Amphi 8*
- Inégalités de concentration (Pierre Youssef), *Amphi 9*
- Changement climatique (Anne-Catherine Favre), *Amphi 11*

12:25 - 14:00 : Déjeûner, restaurant Diderot
14:00 - 15:00 : Prix Neveu, **Julien Reygner et Erwan Scornet**
15:00 - 16:00 : Conférence plénière, **Eva Löcherbach**
16:00 - : **Clôture de la conférence**