

1-3.3 Groupe thématique « Probabilités et Statistiques » (PS)

Thèmes de recherche

Les thèmes principaux du groupe « Probabilités et Statistiques » sont : d'une part, la physique statistique à l'équilibre, la physique statistique hors équilibre et les inégalités fonctionnelles en grande dimension concernant les probabilités ; d'autre part, l'analyse bayésienne et la statistique computationnelle, la grande dimension et l'inférence géométrique, le clustering et l'apprentissage statistique et enfin la statistique des processus concernant les statistiques.

• **Physique statistique à l'équilibre.** De nombreux membres du groupe (Chafaï, Cosco, de Tilière, Dumaz, Labbé, Poisat, Simenhaus) étudient les propriétés statiques des grands modèles de mécanique statistique à l'équilibre, qu'ils soient discrets ou continus, notamment :

- Les systèmes de spins exactement solubles tels que le modèle de dimères, le modèle d'Ising et les forêts couvrantes (de Tilière). Il s'agit d'un domaine très actif au niveau international, avec l'attribution de 3 médailles Fields (Werner 2006, Smirnov 2010, Duminił-Copin 2022).
- Les polymères aléatoires, qui modélisent le positionnement spatial d'une longue chaîne de monomères en interaction avec elle-même et son environnement (Cosco, Poisat, Simenhaus). Le but ici est de comprendre le comportement macroscopique de la marche (diffusif, sur-diffusif ou localisé) en fonction du réseau sous-jacent et de l'interaction considérée. On peut faire varier les paramètres du modèle et chercher à analyser les éventuelles transitions de phase qui en résultent.
- Les systèmes de particules en interaction, tels que les gaz de Riesz et de Coulomb ou les modèles de valeurs propres de matrices/opérateurs aléatoires (Chafaï, Dumaz, Labbé). L'enjeu est ici de décrire le positionnement macroscopique (distribution empirique, fluctuations) et microscopique (rigidité, localisation ou délocalisation) d'un grand nombre de particules (ou de valeurs propres) ayant tendance à se repousser entre elles.

• **Physique statistique hors-équilibre.** Dès lors que les propriétés statiques d'un modèle de physique statistique à l'équilibre sont bien comprises, on peut chercher à étudier son comportement dynamique, lorsque le système est initialisé ou maintenu en dehors de l'équilibre. Cette question fondamentale intéresse plusieurs membres du groupe (Chafaï, Cosco, Huvneers, Iacobucci, Labbé, Olla, Salez, Simenhaus), et se décline en une variété de problèmes, notamment :

- L'étude des propriétés de diffusion et de transport (par exemple via l'établissement de limites hydro-dynamiques) pour des systèmes hamiltoniens classiques ou quantiques soumis à des forçages thermiques et mécaniques aux bords (Huvneers, Iacobucci, Olla).
- La classification des modèles cinétiquement contraints, qui constituent une classe de modèles stochastiques pour la transition liquide/verre (Toninelli). Éclaircir les mécanismes qui sous-tendent cette célèbre transition de phase est un problème physique majeur, qui a motivé de nombreux efforts aussi bien théoriques qu'expérimentaux (voir par exemple sc-glass.uchicago.edu).
- L'étude des temps de mélange et l'établissement du phénomène de cutoff, qui est une transition de phase mystérieuse et remarquable dans la convergence à l'équilibre de certains processus de Markov ergodiques (Chafaï, Labbé, Salez, Simenhaus).
- L'analyse du comportement en temps long des marches aléatoires en environnement aléatoire et des diffusions non-linéaires de type équation de la chaleur stochastique ou KPZ qui en sont les limites d'échelles (Cosco, Labbé, Simenhaus).

• **Inégalités fonctionnelles en grande dimension.** Une petite partie des membres du groupe (Chafaï, Labbé, Lehec, Salez) s'intéresse également aux inégalités fonctionnelles discrètes ou continues pour les semi-groupes de Markov, et à leurs liens avec le phénomène de concentration de la mesure, notamment :

- Les inégalités de Poincaré sur des espaces discrets ou continus, qui permettent de contrôler l'isopérimétrie via l'inégalité de Cheeger, et qui fournissent de la concentration sous la mesure invariante.
- Les inégalités « entropiques » de type log-Sobolev ou log-Sobolev modifiées, qui caractérisent la décroissance exponentielle de l'entropie et quantifient l'hyper-contractivité du semi-groupe de Markov considéré.
- Les inégalités de courbure au sens de Bakry-Emery ou d'Ollivier, qui sont inspirées de la célèbre notion de courbure de Ricci en géométrie Riemannienne et qui fournissent notamment des inégalités de concentration 'locales en temps', capables de se propager le long du semi-groupe de Markov au lieu d'être seulement valables sous la mesure stationnaire.

• **Analyse bayésienne et statistique computationnelle.** Le Ceremade est historiquement un des fers de lance de la statistique bayésienne en France, sous l'impulsion notamment de Christian Robert et ses collaborateurs. Outre les aspects d'analyse bayésienne et leur liens avec la statistique computationnelle (Robert, Ryder, Stoehr), les propriétés fréquentistes des lois a posteriori sont aussi étudiées au sein du laboratoire (Hoffmann, Rivoirard). Plus précisément :

- Méthodes de simulations markoviennes et inférence bayésienne approximative : échantillonnage efficace sur un espace de grande dimension, via des techniques MCMC notamment. Lorsqu'il est difficile, voire impossible, d'évaluer la vraisemblance des données (dans les modèles hiérarchiques notamment), étude des méthodes approchées (ABC) et leurs extensions à des espaces de grande dimension ou de géométrie complexe (Robert, Ryder, Stoehr).
- Inférence bayésienne dans les modèles à variable latente, problèmes de vraisemblance cachée : perspectives théoriques de convergence de ces méthodes d'inférence et de validation de ces techniques, généralement dans le contexte de modèles particuliers comme les modèles de mélanges (Robert, Ryder, Stoehr).
- Vitesses de contraction (en statistique bayésienne non-paramétrique) : étude fréquentiste de la vitesse de concentration optimale de la loi a posteriori autour d'une loi du modèle (Hoffmann, Rivoirard).

• **Grande dimension et inférence géométrique.** Héritière de la statistique non-paramétrique, elle mobilise une partie de la recherche de l'équipe. Il s'agit de comprendre la structure parcimonieuse ou géométrique de basse dimension, dans un contexte de données massives et de combat contre le fléau de la dimension (Bacry, Comminges, Hoffmann, Meziani, Rivoirard, Robert, Roche). On y étudie notamment :

- L'estimation de fonctions, de fonctionnelles dans des modèles de régression avec co-variables en grande dimension, sous des hypothèses de parcimonie. Les algorithmes afférents sont issus de l'estimation adaptative classique (inégalités oracles) et de l'interface avec l'apprentissage statistique (Bacry, Comminges, Meziani, Rivoirard).
- L'estimation non-paramétrique d'une densité ou d'une fonctionnelle de son support lorsque la densité est concentrée sur une structure géométrique simple, possiblement inconnue. Regroupe des techniques de reconstruction de signal dans un cadre de géométrie des sous-variétés (Hoffmann, Ryder).
- L'estimation de données fonctionnelles, lorsque les observations sont des réalisations indépendantes de courbes aléatoires spécifiées seulement par des caractéristiques du second ordre (Rivoirard, Roche).
- Problèmes inverses dans un cadre mal ou sévèrement mal posé, déconvolution, recherche de critères d'arrêt pour minimiser le coût de l'estimation adaptative (Hoffmann, Meziani, Rivoirard).

• **Clustering et apprentissage statistique.** Il s'agit d'un domaine en plein développement à l'interface avec les applications et l'algorithmique. Dans la recherche effectuée au sein du Ceremade (Bertrand, Brugière, Meziani, Olteanu, Rossi), on y retrouve notamment les thèmes suivants :

- Classification non supervisée (clustering), et structure de classification hiérarchique (Bertrand).
- Apprentissage non-supervisé parcimonieux, détection de rupture et d'anomalie dans des données graphiques, relationnelles et textuelles (Olteanu, Rossi).
- Méthodes d'apprentissage par renforcement et apprentissage profond (Brugière, Meziani).

• **Statistique des processus.** Une partie des membres du groupe (Bacry, Hoffmann, Olteanu, Rivoirard, Rossi) étudie l'inférence de modèles via une dimension temporelle, à travers l'étude de processus stochastiques dans un contexte statistique :

- Statistique pour des processus ponctuels, de type Hawkes ou de diffusion de type McKean-Vlasov, notamment en interaction dans une limite champ moyen (Bacry, Hoffmann, Rivoirard).
- Processus stationnaires et détection de ruptures spatio-temporelles (Olteanu, Rossi).