

3-2.3 Groupe « Probabilités et Statistiques »

Productions scientifiques du groupe

Emmanuel Bacry a travaillé autour des processus ponctuels et leurs applications (santé, finance haute fréquence, réseaux sociaux). A partir de l'observation d'un système d'agents (patient, le trader, le compte tweeter, etc.) potentiellement en interaction qui engendre une grande quantité d'événements, on cherche à comprendre la géométrie des interactions, la propagation de l'information, la stabilité du système. L'article [A743] introduit une méthode permettant de caractériser les contributions de chaque agent à la volatilité global d'un marché par des processus de Hawkes. Dans [A826] est proposé un formalisme théorique permettant pour la première fois d'intégrer dans un même cadre deux classes de modèles de volatilité historique très populaires : les modèles multifractals (introduits initialement par Mandelbrot) et les modèles plus récents de volatilité rugueuse (introduits par Gatheral et Rosenbaum). Dans le cadre d'un partenariat d'importance avec la caisse nationale d'assurance maladie, Emmanuel et son équipe ont travaillé sur une des plus grosses bases de données médico-administrative. Ils ont pu mettre au point un nouveau modèle de survie [A709] permettant de détecter les effets secondaires d'un grand ensemble de médicaments en une seule passe. Il s'agit d'une méthode très novatrice qui a été illustrée dans le cadre des risques de chutes associées à un ensemble de plus de 80 médicaments [P173]. Enfin Emmanuel a obtenu une nouvelle inégalité de concentration pour les martingales matricielles en temps continu avec une variance observable [A49], dont l'utilisation a été illustrée dans le cadre d'une application sur des données de Twitter.

Patrice Bertrand, en collaboration avec Jean Diatta (université de La Réunion) a poursuivi la ligne de recherche qui consiste à caractériser les modèles de clustering multi-niveaux comme étant des structures de convexité qui sont des convexités d'intervalle. Il a proposé des caractérisations différentes de ses travaux antérieurs [A88]. Ces nouvelles caractérisations permettent d'identifier la hiérarchie d'Apresjan, la hiérarchie faible de Bandelt et Dress, et la hiérarchie du lien simple, comme étant des convexités d'intervalle directement déduites de la dissimilarité d définie sur les données. En outre, il a introduit une suite de dissimilarités définies à partir des chemins valués par d , qui est associée à une filtration de la hiérarchie du lien simple de d , ainsi qu'à une filtration de la hiérarchie faible de Bandelt et Dress de d . Il en résulte une méthode d'affinement de ces structures de clustering bien connues dont on peut contrôler le taux de distortion [P30].

Pierre Brugière s'intéresse à l'apprentissage statistique appliqué à la finance, surtout dans le cadre de la gestion d'actifs, avec un intérêt particulier pour les méthodes de renforcement learning. Il a réalisé une étude de cas avec Gabriel Turinici [P50].

Djalil Chafaï s'est intéressé au spectre de matrices aléatoires en grande dimension, à la statique des gaz de Riesz/Coulomb, ainsi qu'à des dynamiques markoviennes singulières en temps long. L'étude du rayon spectral en grande dimension est menée avec succès dans [A125] via le polynôme caractéristique, après une tentative non-optimale via la formule de Gelfand dans [A126]. Une étude de la norme pour un modèle log-concave est menée dans [A806]. Concernant les gaz de Riesz/Coulomb, Djalil a étudié les mesures d'équilibre dans [A268, A271], des algorithmes de simulation dans [A269, A270], des inégalités fonctionnelles dans [Ch8], des modèles de jellium de Wigner dans [A265, A266], et des inégalités de concentration dans [A267]. Enfin, le temps long pour processus de type Dyson est abordé dans [A112, A173]. Les méthodes qu'il utilise sont variées : couplages, fonctions de Lyapunov, inégalités fonctionnelles, calcul stochastique, analyse gaussienne, analyse spectrale, principe du maximum, analyse complexe, fonctions spéciales, polynômes orthogonaux.

Laëtitia Comminges s'est spécialisée sur le modèle de régression linéaire, sur la parcimonie en général, la détection de signal, l'estimation minimax, l'estimation de fonctionnelles non lisses, et l'estimation de norme et de variance, l'estimation robuste et adaptative. Elle a considéré des problèmes d'estimation de la norme f_2 (et f_2 au carré) dans le modèle de régression linéaire sparse avec variance inconnue, ainsi que le problème de test d'hypothèse de nullité du paramètre de régression sous des alternatives locale sparse avec mesure de séparation en f_2 . Laëtitia a pu établir les taux minimax d'estimation et de test pour ces trois problèmes [A253].

Clément Cosco a concentré son attention sur le modèle des polymères dirigés en milieu aléatoire et l'équation KPZ, qui est une EDPS Non-Linéaire introduite par les physiciens Kardar, Parisi et Zhang afin de modéliser certaines croissances d'interfaces dans un espace inhomogène (par exemple la frontière d'un feu de forêt,...). En grande dimension ($d \geq 3$) et dans le régime de faible désordre, Clément a contribué à établir que le comportement de l'équation KPZ est intimement lié au théorème local limite du modèle de polymère dirigés. Cela lui a permis, dans un article avec Shuta Nakajima et Makoto Nakashima [A333], de prouver que l'équation KPZ régularisée se comporte à l'échelle macroscopique comme l'équation de la chaleur stochastique avec bruit additif. Les résultats précédents sur le sujet étaient restreints à une sous-partie de cette région. En dimension $d = 2$, Clément a conjecturé que les propriétés du point d'arrivée favori des trajectoires du polymère sont caractérisées par celles du maximum d'un champ gaussien log-corrélé. Dans le but de montrer ce comportement, il a obtenu avec Ofer Zeitouni [P84] une borne supérieure sur les moments du champ en accord avec ce phénomène. Pour cela, il a généralisé à un très grand nombre de marches une méthode de Caravenna-Sun-Zygouras sur des diagrammes d'intersection de 3 marches aléatoires. Il a très récemment obtenu une borne inférieure correspondante.

Béatrice de Tilière s'intéresse aux modèles exactement solubles tels que le modèle de dimères, le modèle d'Ising et les arbres/forêts couvrantes. Le premier ensemble de résultats principaux concerne l'étude de divers modèles Z -invariants en dehors du point critique : forêts couvrantes/Laplacien massique et modèle d'Ising genre 1 [A177, A178], dimères genre 1 [A174] et genre général [A176]. Avant ces travaux, seul le genre 0 était connu. Béatrice montre que ces modèles ont une propriété d'intégrabilité très forte, comme partiellement prédit par Baxter; elle paramètre ces modèles en utilisant une courbe de Harnack associée, et étudie les plongements du graphe sous-jacent [A175]. Les techniques utilisées mélangent probabilité, combinatoire, analyse complexe sur les surface de Riemann et géométrie algébrique. Dans son autre résultat principal [A807], Béatrice démontre un lien fort entre le modèle d'Ising d'une part et les forêts couvrantes et la marche aléatoire d'autre part. Ceci explique en particulier l'importance de l'analyse complexe discrète pour la compréhension du modèle d'Ising.

Laure Dumaz s'intéresse au spectre des opérateurs de Schrödinger, qui s'écrivent $-\Delta + V$ où Δ est le laplacien et V un potentiel aléatoire. Depuis les travaux d'Anderson dans les années 50, ce domaine de recherche a été l'objet de nombreux travaux, notamment dans les années 80. Un des enjeux majeurs est de prouver la délocalisation en dimension 3. Afin de mieux comprendre comment les vecteurs propres se localisent lorsque l'énergie est finie et se délocalisent pour les hautes énergies, Laure a détaillé avec Cyril Labbé l'étude en dimension 1 et dans le cas où le potentiel V est un bruit blanc [P98, A389, P99]. Ce potentiel avait été peu regardé dans la littérature en raison de son irrégularité. Il apparaît comme la limite de modèles discrets avec des potentiels suffisamment décorrélés. Dans tous les régimes d'énergie, une description détaillée de la forme des vecteurs propres est obtenue, ainsi que la limite microscopique des valeurs propres. Dans le cas haute énergie, grâce à une transformation unitaire bien choisie, un opérateur limite de la forme Dirac apparaît, qui était conjecturé par Edelman et Sutton comme limite du bulk de certaines matrices aléatoires symétriques et qui est relié aux opérateurs de manège hyperbolique de Valkó et Virág. Les techniques utilisées sont celles du calcul stochastique, ce qui est novateur dans ce domaine et permet d'obtenir des résultats très précis. Laure a également étudié d'autres opérateurs liés aux matrices aléatoires dans les papiers [A390] avec Cyril Labbé et [A391] avec Benedek Valkó et Yun Li.

Marc Hoffmann s'intéresse à plusieurs aspects de la statistique mathématique : avec Gilles Blanchard et Markus Reiß, il a pu établir de nouvelles bornes d'estimation pour la régularisation des problèmes inverses par arrêt optimal [A107, A108]. En biologie mathématique des populations, il a poursuivi ses travaux avec Aline Marguet et Marie Doumic sur les modèles structurés [A386, A534] ainsi qu'avec Frédéric Pack Shoenberg pour la modélisation de la propagation d'épidémie par processus de Hawkes modifiés [A795]. En inférence géométrique, avec Clément Berenfeld, John Harvey et Krishnan Shankar, il a établi les vitesses optimales pour l'estimation d'une densité à support sur une sous-variété inconnue ainsi que pour l'estimation du reach (au sens de Federer) via une nouvelle utilisation de la fonction de défaut de convexité [A73, A74]. Avec Mathias Trabs [P137], il a poursuivi son programme d'inférence à travers les échelles, établissant rigoureusement la non-monotonie en échelle des taux minimax dans les modèles dits de dispersion, résultat un peu inattendu. Avec Laetitia Della Maestra, il a réalisé l'estimation non-paramétrique pour des systèmes de particules en interaction en champ moyen de type McKean-Vlasov [A353, A354], obtenant la propriété LAN ainsi que des bornes minimax et des inégalités oracles pour l'estimation anisotrope de la solution de l'équation de Fokker-Planck associée et de son terme de transport. Ceci permet en particulier de tester la présence d'interactions. Les inégalités de type Bernstein nécessaires pour réaliser un tel programme évitent l'écueil des méthodes de couplage ou des inégalités géométriques via un argument simple de changement de mesure, dans l'esprit des travaux de Lacker. Finalement, avec Kolyan Ray [P136], Marc a obtenu les premiers résultats de contraction optimale de lois a posteriori pour l'estimation du coefficient de diffusion dans un modèle multidimensionnel à partir d'observations haute fréquence.

François Huveneers a approfondi l'étude des propriétés de transport dans les systèmes hamiltoniens quantiques et classiques. Dans [A802], une description du passage de la phase localisée à la phase ergodique est proposée pour des chaînes quantiques désordonnées. Dans [A545], l'existence d'un plateau pré-thermal est mise en évidence pour une chaîne d'oscillateurs classique, et la durée de vie de ce régime transient est quantifiée. Dans [A779], l'annulation des coefficients de transport est démontrée pour une chaîne anharmonique dans laquelle les interactions sont suffisamment diluées, rendant rigoureuse une approche heuristique basée sur l'existence d'effets de Griffiths. D'autre part, François a étudié l'émergence d'une limite hydrodynamique dans une limite d'échelle hyperbolique (système d'équations d'Euler) pour un système admettant un ensemble complet d'intégrales du mouvement et qui n'est donc pas ergodique [A85, A550]. Cela montre que l'ergodicité n'est pas une condition nécessaire pour l'obtention d'équations macroscopiques dans l'échelle spatio-temporelle hyperbolique. Enfin, avec François Simenhaus, il a progressé dans la compréhension des fluctuations d'une particule évoluant dans un champ diffusif à l'aide d'arguments théoriques et de simulations numériques [A553]. Ce travail fournit un cadre cohérent à plusieurs observations antérieures, parfois d'apparence contradictoire.

Alessandra Iacobucci a pour rôle principal d'apporter aux chercheurs et chercheuses du Ceremade son expertise dans l'utilisation des méthodes mathématiques pour la modélisation et la résolution numérique de problèmes théoriques. Son travail pendant la période d'évaluation s'est organisé autour de trois axes de recherche : les propriétés de transport de chaînes d'atomes en interaction soumises à des forçages thermiques et mécaniques aux bords, les propriétés spectrales du générateur de la dynamique de Langevin, et les propriétés de transport du système au niveau macroscopique. Alessandra a implémenté des algorithmes de dynamique moléculaire avec échantillonnage de mesures stationnaires, et a calculé les coefficients de transport, aussi bien par les formules de Green-Kubo via la simulation de la dynamique d'équilibre, que par la simulation de la réponse linéaire de la dynamique stationnaire hors-équilibre [Th38]. Pour analyser les propriétés spectrales du générateur, elle a implémenté un algorithme basé sur la méthode de Fourier-Galerkin [Th38, A555]. L'analyse macroscopique des propriétés de transport nécessite l'intégration d'un système d'équations aux dérivées partielles d'advection-diffusion non-linéaires, qu'elle a implémenté numériquement par une discrétisation aux différences finies et par un algorithme de point fixe permettant de déterminer l'état stationnaire du système [A556].

Cyril Labbé a étudié le spectre d'opérateurs de Schrödinger aléatoires dans le cas particulier où le potentiel est un bruit blanc. Ses travaux dans ce domaine ont été obtenus avec Laure Dumaz, et ont donc déjà été cités plus haut. D'autre part, Cyril s'est intéressé au phénomène de cutoff pour les processus de Markov. En collaboration avec Hubert Lacoïn [A585, A586], il a établi ce phénomène pour le processus d'exclusion simple asymétrique et faiblement asymétrique en dimension 1. Enfin, Cyril a contribué à l'étude d'EDP stochastiques singulières, domaine qui connaît un essor particulier depuis l'introduction de théories nouvelles il y a une dizaine d'années. En collaboration avec Martin Hairer, il a étendu la partie analytique de la théorie des structures de régularité aux espaces de Besov [A503]. Cet article fournit un cadre technique assez large permettant la construction et l'étude d'EDP stochastiques singulières.

Katia Meziani étudie la tomographie homodyne quantique, et les performances théoriques d'un estimateur à noyau de la fonction de Wigner avec borne minimax pour la norme L^∞ et borne inférieure pour la norme L^2 et estimation adaptative. Dans le cadre de la régression linéaire, Katia a établi un nouveau critère Muddling Labels Regularization (MLR) et a montré au travers de résultats théoriques [A662] qu'il permet des performances de généralisation bien supérieures aux critères usuelles tels que AIC, BIC, SURE ou CV. Avec grande probabilité le critère est différentiable et ne nécessite pas de connaître le niveau de bruit. De ce critère découlent plusieurs procédures qui surpassent les performances du Lasso, Ridge et elasticNet. L'avantage premier de MLR est son adaptation comme perte pour entraîner des réseaux de neurones. Les résultats empiriques surpassent dans le cadre de la régression, des réseaux de neurones classiques et pour de nombreux jeux de données de cas les méthodes usuelles en machine Learning comme Catboost, Xgboost et RF.

Stefano Olla s'intéresse à l'origine microscopique des conditions aux limites des lois de conservation macroscopiques. Dans [A680], il a établi la limite hydrodynamique quasi-statique avec le profil macroscopique satisfaisant les conditions d'entropie pour l'équation statique de Burgers avec les conditions aux limites correspondantes. Dans [A577], il considère la limite d'une équation cinétique linéaire avec réflexion-transmission-absorption à une interface et avec un noyau de diffusion dégénéré. L'équation découle d'une chaîne microscopique d'oscillateurs en contact avec un bain de chaleur. Dans [A578], une chaîne infinie d'oscillateurs harmoniques couplés avec un thermostat Langevin à l'origine est étudiée. Dans la limite haute fréquence, les coefficients de réflexion-transmission de l'énergie des ondes sont obtenus. De manière surprenante, la diffusion ne couple pas l'énergie des ondes à différentes fréquences. Dans [A504], Stefano introduit une perturbation stochastique conservatrice de masse de l'équation de Schrödinger Non-Linéaire discrète, qui modélise l'action d'un bain de chaleur à une température donnée. La distribution canonique de Gibbs correspondante est l'unique mesure invariante. Dans le cas de la focalisation cubique unidimensionnelle et dans la limite du temps long et de basse température, la solution converge vers l'onde stationnaire de l'équation continue qui minimise l'énergie pour une masse donnée.

Madalina Olteanu a travaillé sur la question de l'importance et de la sélection des variables, dans le cadre non-supervisé (clustering type k -means) et supervisé (forêts aléatoires). En utilisant des approches pénalisées inspirées par le cadre de la régression, elle a pu proposer des méthodes de clustering sparse et group-sparse pour données mixtes, [C18, C21]. Un package R a été également développé et publié. Dans le cadre de l'apprentissage supervisé et des forêts aléatoires une nouvelle méthode d'importance, utilisant les corrélations entre les variables combinées à une procédure de clustering [C19] a été proposée. Madalina a également travaillé en collaboration avec des collègues en physique statistique et géographie quantitative, sur le développement d'une nouvelle approche pour quantifier la ségrégation résidentielle. Il s'agit de la construction d'un nouvel indice de ségrégation, qui tient compte à la fois du caractère multi-scalaire et individuel du phénomène [A95, A730, A732, A745]. Dans le cadre des mêmes travaux, elle a pu également s'intéresser à l'apport des réseaux de neurones non-supervisés pour l'étude de la ségrégation spatiale [A731]. Enfin, Madalina a abordé la question de la détection de ruptures et d'anomalies, pour des données complexes, temporelles ou spatio-temporelles, et pour une partie du travail censurées et échantillonnées de manière irrégulière. Les méthodologies proposées combinent des étapes de clustering avec de la détection de rupture off-line [Ch13, Ch14, A599].

Julien Poisat s'intéresse à différents modèles de polymères. Dans [A75], il étudie un polymère chargé et montre l'existence de deux phases, localisée et délocalisée. Pour estimer la position du point critique, il analyse l'énergie libre du modèle, celui de la marche faiblement auto-évitante et la fonction de taux pour les déviations vers le bas du temps d'auto-intersection de la marche. Dans [A289], il étudie un polymère interagissant avec une ligne présentant des impuretés (modèle d'accrochage désordonné) réparties suivant un renouvellement. Quand les corrélations de ce désordre sont sommables, le critère de Harris est semblable au cas 'bruit blanc'. Dans le cas plus difficile des corrélations non sommables, une réponse partielle est fournie. La méthode la plus originale de ce papier est la combinaison entre un développement de type 'chaos' de la fonction de partition et des inégalités de découplage sur le désordre, permettant de se ramener au cas 'bruit blanc'. Dans [A739], Julien détermine l'asymptotique en temps long de la probabilité de survie d'une marche aléatoire en dimension un parmi des obstacles répartis suivant un renouvellement. Les résultats diffèrent fortement du cas 'bruit blanc'. Dans [P180] (soumis), il utilise des techniques de renouvellement, héritées des modèles de polymères, pour donner le comportement de la marche conditionnée à survivre.

Vincent Rivoirard a établi les vitesses de concentration bayésiennes non-paramétriques pour différents processus ponctuels, comme les processus d'Aalen [A383, A384] et surtout les processus de Hawkes multivariés linéaires [A382] et non-linéaires en établissant pour ces derniers les conditions qui permettent d'inférer le graphe de connectivité fonctionnelle en présence d'inhibition. Ses travaux sur les processus de Hawkes se sont déclinés dans le cadre applicatif pour la modélisation de l'activité neuronale [A596] ou celle de la mortalité quotidienne due au COVID 19 [A190]. Dans le cadre minimax adaptatif, il a établi de nouveaux résultats pour des problématiques de déconvolution qui ont mis en jeu de nouvelles techniques fondées sur la transformée de Mellin pour la déconvolution multiplicative [A533], la transformée de Cauchy pour la déconvolution libre [A668] ou des stratégies de déformation pour des données circulaires [P176]. Les approches par pénalisation ont été revisitées en proposant de nouveaux contrastes pénalisés qui fournissent des procédures combinant une calibration optimale des hyperparamètres et un faible coût computationnel [A299, A588, A718, A817]. Des extensions de procédures classiques ont enfin été envisagées pour spécifiquement prendre en compte la nature des données dans le cadre fonctionnel [P25] ou lorsque celles-ci sont discrètes [A123, A541].

Christian Robert a introduit dans [O8] un état de l'art des techniques d'inférence sur ces objets, avec des perspectives originales sur l'inférence sur le nombre de composantes, avec quelques dizaines de spécialistes sur les modèles de mélanges de distributions. Dans [A701], Christian défend une perspective radicalement différente sur l'objet central de la statistique, à savoir conduire des tests d'hypothèses et d'adéquation de modèles, et donc l'abandon de la notion de significativité dans les tests et de l'emploi associé de p-values. Cet article d'opinion fait suite à l'avertissement de plusieurs sociétés de statistique sur les dérives résultant de l'emploi aveugle, voire biaisé, de ces objets. Il a aussi donné lieu à une tribune dans *Nature*, restreinte aux trois principaux auteurs. Dans [A87], une approche originale des méthodes ABC (Approximate Bayesian Computation) est proposée, laquelle remplace l'utilisation de statistiques résumées (subjectives) par des mesures de distance intrinsèques comme la distance de Wasserstein. Cet article a donné lieu à de nombreuses citations et généralisations. Dans [A748], une autre approche des méthodes ABC est explorée, adossant la sélection des statistiques résumées à l'emploi de forêts aléatoires pondérant leur importance relative dans l'explication des données. Dans [A443], Christian pose les caractérisations théoriques de l'emploi des méthodes ABC dans un contexte où le modèle supposé est incorrect (ou mal spécifié), obtenant des résultats surprenants sur la robustesse des approches originelles (contre le manque de robustesse des approches plus sophistiquées). Dans [A309], Christian et ses collaborateurs introduisent une méthode permettant de combattre la malédiction de la dimension dont souffrent les méthodes ABC (en termes du nombre de paramètres du modèle).

Angelina Roche s'est focalisée sur deux problèmes : d'une part, le modèle de régression à sortie réelle avec plusieurs covariables fonctionnelles [P187] : l'objectif est d'étudier l'estimateur Lasso dans ce contexte infini-dimensionnel. Ce modèle a été étudié dans la communauté Machine Learning, sous le nom de multiple kernel learning mais avec des hypothèses plus restreintes sur les covariables et les résultats théoriques ne s'étendaient pas à la dimension infinie. L'originalité de l'article consiste à démontrer des inégalités-oracles de sparsité sharp sous des hypothèses assez faibles, ce qui a conduit à revisiter les hypothèses de type valeurs propres restreintes usuellement admises dans la communauté. D'autre part, la question de l'estimation des fonctions propres de l'opérateur de covariance lorsque les données fonctionnelles ne sont pas observées entièrement. Cette situation intervient dans la plupart des travaux théoriques jusqu'à présent, mais en certains points avec potentiellement présence de bruit, ce qui est le cas dans la majorité des cas pratiques [A776]. Il s'agit d'un article de review mettant en perspective les résultats de la thèse de Ryad Belhakem avec la littérature existant sur le sujet.

Fabrice Rossi a obtenu la mise au point d'un modèle de co-clustering de données mixtes [Ch4] unique dans sa capacité à créer des variables supplémentaires à partir des variables d'origine, afin de regrouper dans un même cluster des variables de nature différente. Cette méthode est basée sur une approche de type *minimum description length*, principe aussi utilisé pour développer un estimateur de densité par histogramme non régulier sans méta-paramètre à calibrer [A830]. Il a pu aussi proposer un nouvel algorithme d'alignement de graphes permettant de comparer des programmes représentés par le graphe d'appels entre leur fonction [C55]. Cette méthode améliore significativement l'état de l'art. Fabrice a développé un nouveau modèle parcimonieux pour le clustering de textes [A782] et un modèle qui combine détection de ruptures et clustering pour des échanges de texte modélisés par un graphe dynamique [A325]. En détection de ruptures et d'anomalies, il a proposé une approche spatio-temporelle pour des données parcimonieuses et relevées de façon irrégulière [A599]. Ses collaborations en archéologie ont donné lieu à trois communications dans les conférences de référence du domaine [C39, C62, C63]. Enfin, il a eu l'opportunité de participer à la rédaction d'un article d'état de l'art sur l'utilisation de la visualisation d'information dans le but de permettre à des utilisateurs d'outils d'apprentissage automatique de savoir s'ils peuvent faire confiance aux résultats de ces outils [A288].

Robin Ryder a obtenu en collaboration avec des linguistes de l'INALCO et du MPI Leipzig la première reconstruction de l'histoire des langues Sino-Tibétaines [A786], en explorant par MCMC la loi a posteriori d'un ensemble d'arbres phylogéniques au vu de données lexicales, ce qui correspond à l'état de l'art en Statistique pour la Linguistique Historique. Avec Grégoire Clarté il a récemment proposé [P79] une nouvelle méthode d'inférence de l'histoire jointe de données lexicales et phonologiques, ce qui permet d'étendre l'ensemble des familles auxquelles ces méthodes peuvent être appliquées, et de prendre en compte un éventail plus large de données ; ils ont pu notamment appliquer ces méthodes à l'histoire des langues des signes. Les arbres obtenus sont maintenant essentiels pour de nombreuses analyses statistiques cross-linguistiques [A784]. Robin a aussi établi des techniques de couplage de MCMC pour contrôler la qualité d'un échantillonneur d'une loi sur un ensemble d'arbres phylogéniques [A566]. Pour d'autres problèmes d'inférence bayésienne computationnelle, la loi a posteriori n'est pas disponible sous forme close et on recourt à des méthodes approchées. Il a pu notamment proposer une méthode générique pour contrôler la qualité de l'approximation [A608], et avons développé une nouvelle classe d'algorithmes sans vraisemblance (dite ABC-Gibbs) pour l'inférence approchée en grande dimension, en approchant des lois a posteriori conditionnelles [A309].

Justin Salez a délaissé ses anciens sujets de recherche autour des propriétés structurelles et spectrales des graphes aléatoires pour s'intéresser aux aspects quantitatifs de la convergence à l'équilibre des processus de Markov. Avec ses collaborateurs, il a notamment démontré le mystérieux phénomène de cutoff pour différents systèmes de particules en interaction tels que le processus d'exclusion avec réservoirs [A791], le processus 'Zero-Range' [A526, A705] ou les 'Activated Random Walks' [P48]. Il a par ailleurs établi diverses inégalités fonctionnelles discrètes, telles que des principes de comparaison de formes de Dirichlet [A525, A528], des inégalités de log-Sobolev [A789, P221], ou des inégalités de log-Sobolev modifiées [A527, P190]. Enfin, il s'est récemment intéressé aux divers analogues discrets de la notion de courbure de Ricci, et à leurs remarquables implications pour les graphes et les chaînes de Markov. Cela lui a permis de résoudre un problème ouvert dû à E. Milman, A. Naor et Y. Ollivier sur l'incompatibilité entre courbure positive et expansion pour les graphes de degrés bornés [P174, A790], et d'apporter un premier critère général pour l'émergence du cutoff pour les marches aléatoires sur des groupes Abéliens [P189].

François Simenhaus s'est intéressé, avec Julien Poisat, à la mesure d'équilibre d'un modèle de polymère interagissant avec des interfaces répulsives distribuées selon un processus de renouvellement à queue lourde. Dans un premier article, une description de la loi limite de la fonction de partition après une renormalisation adaptée a été obtenue [A739]. Dans un second temps, des résultats de localisation très précis pour la trajectoire du polymère ont pu être fournis [P180]. L'étude des marches aléatoires en environnement symétrique et diffusif, et notamment l'ordre des fluctuations, reste aujourd'hui une des questions importantes du domaine. Dans cette perspective, l'article [A553] étudie un processus évoluant dans un environnement simplifié, diffusif et guidé par l'équation de la chaleur afin de faire apparaître deux comportements caractéristiques : en temps court l'évolution est sous-diffusive, du fait de l'évolution rapide de l'environnement, tandis qu'en temps long, l'environnement est décrit par ses grands modes évoluant suffisamment lentement pour que le marcheur ait le temps de rejoindre des minima du potentiel évoluant diffusivement.

Julien Stoehr cherche à échantillonner efficacement une loi de probabilité potentiellement multimodale, définie sur un espace de grande dimension ($d \geq 50$) et à calculer des espérances sous cette loi. Avec Changye Wu et Christian Robert [P220], il a étudié une calibration du schéma d'intégration symplectique utilisé pour résoudre le système de Hamilton. En construisant une mesure empirique fournissant à chaque itération un intervalle aléatoire sur lequel intégrer le système tout en préservant les propriétés de convergence de l'algorithme tout en limitant le coût computationnel de chaque itération. Julien s'intéresse également aux méthodes d'échantillonnage préférentiel qui se comportent mal dès que $d \geq 10$. C'est notamment le cas si l'on veut utiliser ces méthodes en synécologie pour un grand nombre d'espèces et de sites d'observation. Avec Stéphane Robin (Sorbonne Université), il a travaillé sur un projet avec des données en synécologie et construit un algorithme Monte Carlo EM qui combine l'échantillonnage préférentiel, des approches variationnelles et des méthodes de type vraisemblance composite. Dans les modèles bayésiens hiérarchiques, le nombre important de paramètres rend délicate l'inférence autour de la loi *a posteriori*. Dans un contexte où il est difficile d'évaluer la vraisemblance des données et tout ou partie des lois conditionnelles du modèle, il a proposé avec Grégoire Clarté, Christian Robert et Robin Ryder [A309] de traiter le problème à l'aide d'une méthode qui combine l'échantillonneur de Gibbs et les méthodes ABC.

Cristina Toninelli a dédié les cinq années passées aux systèmes de particules avec contraintes cinétiques (KCM) : une classe de modèles stochastiques pour la transition liquide/verre. Avec ses collaborateurs, elle a démontré que les KCM en dimension 2 se répartissent en 5 classes d'universalité [A509, A513, A675]. Le comportement des échelles de temps est obtenu dans chaque classe, et un critère simple est fourni pour déterminer la classe d'un modèle à partir de sa contrainte et d'un raisonnement combinatoire. Pour démontrer ces résultats, Cristina a dû développer une nouvelle boîte à outils, car la dégénérescence des taux de transition (non-monotonie, non-unicité des mesures invariantes, temps de mélange anormaux) empêche l'utilisation des méthodes classiques. Elle a également déterminé le comportement asymptotique, en toute dimension, pour l'un des plus célèbres KCM : le modèle FA2f [A512]. Cela a nécessité d'appréhender et contrôler son mécanisme coopératif de relaxation, dominé par le mouvement de gouttelettes large et rares qui se déforment pour évoluer dans un environnement dynamique aléatoire.