

Probabilités et Statistique

Mots clés

Modélisation et méthodes statistiques & Stochastiques : Modèles de survie, statistique multidimensionnelle, classification, estimation et contrôle de processus, algorithmes stochastiques, matrices aléatoires.

Composition de l'équipe :

PR et DR : Bernard Bercu (PR, Bx1 et équipe INRIA CQFD), Pierre Del Moral (DR INRIA, responsable de l'équipe probabilités et statistique, ainsi que de l'équipe INRIA ALEA), François Dufour (PR, IPB, responsable de l'équipe INRIA CQFD), Mikhail Nikulin (PR, Bx2), Jérôme Saracco (PR, Bx4 et équipe INRIA CQFD)

MCF et CR : François Caron (CR, INRIA équipe ALEA), Marie Chavent (MCF, Bx2 et équipe INRIA CQFD), Vincent Couallier (MCF, Bx2), Anne Gégout-Petit (MCF, Bx2 et équipe INRIA CQFD), Pierrick Legrand (MCF, Bx2 et équipe ALEA), Léo Gerville-Réache (MCF, Bx2), Brigitte Patouille (MCF, Bx2), Jérôme Poix (MCF, Bx2), Benoîte de Saporta (MC, Bx4 et équipe CQFD), Delphine Féral (MCF, Bx1), Huilong Zhang (MCF, Bx1 et équipe CQFD).

Doctorants : Vanessa Kuentz (thèse MENR, équipe CQFD), Jade Nguyen (thèse financée par le CNRS), Karen Gonzalez (thèse MENR, équipe CQFD), Aurélie Le Cain (thèse financée par le CEA, équipe ALEA), Michele Pace (thèse financée par l'INRIA, équipe ALEA), Karen Gonzalez (thèse MENR, équipe CQFD), Hu Peng (Allocation de recherche Ministère, équipe ALEA), François Giraud (thèse financée par le CEA, équipe ALEA), Arijit Das (thèse co-financée par un contrat région, l'INRIA et un contrat EDF), Victor Vazquez (financement CONACYT), Noureddine Saaidia (financement Alger, S. Nikoulin et N. Seddik), Genia Babykina (thèse financée par le CEMAGREF).

Post-doctorants : Lei Ming (Post-doc INRIA, équipe ALEA, et chercheur invité ANR PREVASSEMBLE avec Météo-France).

Table des matières

1	Présentation de l'équipe	3
2	Activités de recherches	4
2.1	Recherche fondamentale et appliquée	4
2.2	Partenaires industriels et projets ANR	4
3	Production scientifique (2005-2009)	5
4	Collaborations académiques	7
5	Activités d'encadrement de la recherche	8
5.1	Enseignement et encadrement de thèses	8
5.2	Expertises et activités d'édition	10
5.3	Organisation de manifestations scientifiques	10
6	Points forts par thématique	11
6.1	Modélisation et méthodes statistiques	12
6.1.1	Analyse de survie et la fiabilité	12
6.1.2	Statistique multidimensionnelle	14
6.1.3	Statistique appliquée	17
6.2	Modélisation et méthodes stochastiques	19
6.2.1	Contrôle stochastique	19
6.2.2	Algorithmes stochastiques	21
6.2.3	Matrices aléatoires	24
7	Projets de recherches	24
7.1	Modèles et méthodes statistiques	24
7.2	Modèles et méthodes stochastiques	25
8	Points faibles	27
9	Fiches individuelles AERES	28

1 Présentation de l'équipe

Les activités de recherches de l'équipe de Probabilités et de Statistique de l'Institut de Mathématiques de Bordeaux s'articulent autour des *modèles et méthodes statistiques et stochastiques* avec une déclinaison de thèmes marquée autour des axes de recherches complémentaires suivants :

1. Modèles et méthodes statistiques : *les modèles de survie et de vie accélérée, la statistique multidimensionnelle, la classification et l'apprentissage non supervisé.*
2. Modèles et méthodes stochastiques : *estimation et contrôle de processus stochastiques les algorithmes stochastiques, et l'analyse spectrale de matrices aléatoires.*

Cette déclinaison en deux thèmes liés à la statistique et au stochastique offre une meilleure visibilité académique permettant la mise en place de nombreux réseaux de collaborations internationales. Cette visibilité est couplée d'une meilleure exposition nationale permettant la mise en place de nombreux partenariats industriels.

La première thématique centrée autour de la modélisation statistique est soutenue par diverses activités de collaborations et groupes de travail entre des enseignants chercheurs des Universités de Bordeaux 1, 2, et 4. Ces axes de recherches concernent les techniques statistique classique telles la modélisation semi-paramétrique, et non-paramétrique, les techniques de réduction de dimension, mais aussi des techniques statistiques pour la fiabilité et l'analyse des durées de vie, et l'analyse statistique de modèles de vie accélérée. Cette thématique concerne une dizaine d'enseignants chercheurs et sept doctorants, cinq d'entre eux sont aussi reliés aux deux équipes INRIA décrites ci-dessous.

Les deux aspects théoriques et appliqués sont pleinement développés au sein de l'équipe tant sur les aspects des probabilités pures et appliquées que sur les aspects de la statistique mathématique et appliquée. Ces recherches interdisciplinaires s'expriment le plus souvent sous forme de diverses et fructueuses collaborations entre chercheurs ayant des compétences et expertises complémentaires sur la base d'intérêts scientifiques communs. Cette synergie s'est renforcée au fil des années avec la mise en place d'un séminaire bi-mensuel de probabilités et de statistique en 2007 et la récente création de deux équipes INRIA travaillant à l'unisson à l'Institut de Mathématiques de Bordeaux sur des problématiques complémentaires issues de la statistique et des sciences de l'ingénieur :

L'équipe INRIA ALEA (responsable P. Del Moral), créée en 2009, est centrée autour des algorithmes stochastiques et leur applications en filtrage de signaux, en classification non supervisée, en analyse d'événements rares et en optimisation stochastique. La confection et l'analyse de nouvelles classes d'algorithmes stochastiques et de méthodes de simulation en interaction est l'une des spécificité de cette équipe. L'équipe ALEA est une petite équipe constituée uniquement de deux chercheurs INRIA de l'équipe de probabilités et statistique, ainsi que de Pierrick Legrand (MCF Bx2), Josselin Garnier (PR, P7) et Arnaud Doucet (UBC Vancouver et PR Université de Tokyo), et deux doctorants.

L'équipe INRIA CQFD (responsable F. Dufour), créée en 2007, est centrée sur les problèmes de contrôle stochastique et de fiabilité dynamique avec des applications variées en analyse de sûreté de systèmes complexes, en sélection de modèles, en filtrage de signaux ainsi qu'en mathématiques financières. La mise en oeuvre d'algorithmes d'estimation associés à différentes modélisations stochastiques ainsi que l'évaluation des indices de fiabilité de systèmes est l'un des principaux axes de recherches de cette équipe constituée de six enseignants chercheurs de l'équipe de Probabilités et Statistique et de quatre doctorants.

Une dernière thématique plus récente, un peu en marge des précédentes, concerne l'analyse spectrale de matrices aléatoires. Ces thèmes de recherches concernent l'analyse fine de mesures spectrales de matrices aléatoires et le comportement du haut du spectre. Cet axe de recherche est représenté par deux enseignants chercheurs : B. Bercu et D. Féral. Cette thématique émergente offre de nouvelles perspectives de collaboration et d'extension avec l'équipe d'analyse de l'IMB (J.F. Bony, V. Bruneau, S. Kupin).

2 Activités de recherches

2.1 Recherche fondamentale et appliquée

Sur le plan purement théorique, nos recherches s'expriment dans différents domaines des probabilités pures et appliquées, et de la statistique mathématique :

Les recherches en *analyse stochastique* sont diverses et variées. Elles portent notamment sur l'étude fine de processus markoviens par morceaux, les problèmes de contrôle stochastique, les chaînes de Markov non linéaires, l'analyse de processus en interaction, de systèmes de particules ou de processus de branchements, les modèles d'arbres généalogiques, les principes de grandes déviations et les inégalités de concentration exponentielles, les propriétés de propagations du chaos, l'analyse fine de la convergence d'algorithmes stochastiques pour l'estimation et l'optimisation globale, les techniques de martingales, et l'analyse spectrale de matrices aléatoires.

Les recherches en *statistique mathématique* ont un spectre tout aussi large. Elles concernent l'analyse mathématique de techniques de réduction de dimension, les méthodes de classification, l'étude de modèles de renouvellement et de vie accélérée, la convergence de modèles de régression, la confection et l'analyse d'approches géométriques de données décrites par des intervalles.

Sur le plan des domaines d'applications, nos recherches se déclinent suivant six principaux axes : Les méthodes statistiques pour l'analyse de survie et la fiabilité, la modélisation semi- et non-paramétrique, la réduction de dimension, la classification, le contrôle stochastique et l'analyse de risques, et enfin les algorithmes stochastiques et leurs applications en physique, en statistique bayésienne, en analyse d'événements rares, en optimisation combinatoire, et en traitement du signal.

Une description détaillée des points forts et des résultats marquants obtenus dans l'équipe est fournie dans la section 6.

2.2 Partenaires industriels et projets ANR

La plupart des recherches fondamentales et appliquées développées dans l'équipe sont issues ou s'accompagnent de mise en situation pratique sur des problèmes très concrets issus des sciences de l'ingénieur et de la statistique appliquée. Les multiples domaines d'applications traités dans l'équipe sont liés à l'agro-industrie, la biostatistique, l'économie, la génomique, la statistique industrielle, ainsi qu'à la fiabilité et l'électronique, la fiabilité et mécanique, la sûreté de fonctionnement et l'analyse de risques, l'économie, les sondages, mais aussi l'environnement et la pollution de l'air, l'épidémiologie, la cancérologie, le traitement du signal, la poursuite de cibles, et l'analyse d'événements rares.

La majorité de nos recherches appliquées s'inscrivent ainsi dans de nombreux contrats avec nos partenaires industriels :

EDF (la classification de courbes de charges et les méthodes statistiques pour l'estimation de la consommation individuelle, 140 KE), DCNS (filtrage et la poursuite multi cibles, 2 contrats de 20KE), CEA (optimisation stochastique sous contraintes 10KE), (étude de problème de propagation de fissure 15KE), IFREMER (fiabilité de structures plates-formes pétrolières 50KE), VNF (Voies Navigables de France) pour la définition de la méthodologie statistique d'une enquête par sondage des navigants plaisanciers sur le canal des deux mers, 6KE), DANONE (application de méthodes NWAY aux données cliniques en nutrition-santé, 6KE), CHANEL (estimation de courbes de tolérance, 10KE), ALSTOM Transport (7KE), ALSTOM Power (5KE), Striker Spine (biomédical, 7KE) THALES (aéronautique, 12KE), RENAULT, Le Ministère de l'Écologie, du Développement et de l'aménagement et l'Agence de l'Environnement et de la Maîtrise de l'Énergie (Programme PRIMEQUAL 2 sur la qualité de l'air, pollution atmosphérique, projet terminé en 2006, 12KE), Institut Pétrole et Gaz de Moscou.

D'autres études sont plutôt liées à des projets ANR avec des organismes de recherches nationaux et des partenaires industriels :

Projet région (sur les méthodes de Monte Carlo par chaînes de Markov en interaction, équipe INRIA ALEA, 75KE), ANR PREVASSEMBLE avec Météo France Toulouse sur le filtrage particulaire en assimilation de données (depuis 2009), ANR VIROSCOPY avec l'ENST Paris sur l'analyse de risques en épidémiologie (depuis 2009), ANR NEBBIANO sur la fiabilité des techniques de tatouage d'images avec l'IRISA de Rennes (2006-2009).

De nouveaux projets d'ANR ont été déposés en 2009 : Safe Tee, Arpège (avec l'école polytechnique, le LMP, EADS Astrium et le LATP et l'équipe CQFD), et Propagation (responsable : F. Caron ; sur le pistage radar et optronique passif pour la sauvegarde et la protection d'infrastructure côtière, avec pour partenaires la DCNS, THALES, ECOMER, EXAVISION et l'équipe ALEA).

3 Production scientifique (2005-2009)

Les recherches menées dans l'équipe couvrent différents aspects théoriques et appliqués de la théorie des probabilités et de la statistique mathématique et appliquée. Au niveau des résultats scientifiques, ces recherches interdisciplinaires ont donné lieu à autant de publications dans des journaux de mathématiques que dans des revues de sciences plus appliquées.

Sur le plan qualitatif, les travaux de recherches de l'équipe ont été publiés dans de nombreuses revues internationales de probabilités appliquées et de statistique de tout premier plan :

The Annals of Applied Probability, SIAM Journal of Control and Optimization, Stochastics, Journal of Applied Probability, Stochastic Processes and their Applications, Probability Theory and its Applications, Stochastic Analysis and Applications, Electronic Communications in Probability, Statistica Sinica, Journal of Multivariate Analysis, Journal of the Royal Statistical Society - Series B, Computational Statistics and Data Analysis, Communications in Statistics - Theory and Methods, Computational Statistics, Journal of Statistical Planning and Inference, etc.

Les aspects plus appliqués ont aussi été publiés dans des revues renommées des sciences physiques ou de l'ingénieur, telles :

Optics Communications, IEEE Transactions on Automatic Control, Communications in Statistics - Simulation and Computation, Statistical Modelling, Environmetrics, Case Studies in Business, Industry and Government Statistics, Journal of Statistical Simulation and Computation, International Journal of Product Development, Communications in Dependability and Quality Management, etc.

D'un point de vue plus quantitatif, en quelques chiffres significatifs sur la période 2005-2009, l'équipe a publié autour de 140 articles dans des journaux internationaux, 3 ouvrages, une vingtaine de proceedings de conférences internationales, et une vingtaine de prépublications. Des informations plus fines sont données dans les fiches individuelles des membres de l'équipe fournies dans la section 9.

Il est bien sûr extrêmement difficile de comparer les différents impacts de chacune de ces publications. Ces comparaisons sont souvent liées à des indices de citations de l'article, de la revue, parfois des conférences ou souvent à la renommée de journaux. Ces critères discutables sont très différents dans les communautés probabilité, statistique, et sciences de l'ingénieur. Il est aussi impossible de comparer les impacts d'articles dans ces trois domaines en toute objectivité.

Sur la période considérée, les dix publications qui représentent "le mieux" les deux thématiques développées dans l'équipe sont, dans l'ordre alphabétique des auteurs, décrites ci-dessous.

Concernant les aspects statistiques :

1. Chavent, M., Liquet, B., Saracco, J. A semiparametric approach for multivariate sample selection model. To appear in *Statistica Sinica* (2009).
2. Chavent, M., Guégan, H., Kuentz, V., Patouille, B., Saracco, J. (PCA- and PMF-based methodology for air pollution sources identification and apportionment. To appear in *Environmetrics* 2009).
3. Couallier, V., Denis, L., Bagdonavicius, V., Nikulin, M. and Gerville-Réache, L. Statistical inference through AFT model for biotechnical systems. *Journal of Statistical Planning and Inference* , bf 139, n°139, 1649-1656 (2009).
4. Nikulin, M., Gerville-Réache, L. and Couallier, V. *Statistique des essais accélérés*. Collection Méthodes Stochastiques Appliquées. Paris : Hermes Science ; Paris : Lavoisier. 295 p (2007).
5. A. Sasco, M.Nikulin. Flexible regression models for dynamic prediction of survival in elderly in presence of cancer or other chronic diseases, *Advances in Gerontology*,v.21, #1, 41-48 (2008).

Concernant les aspects stochastiques :

1. B. Bercu and A. Touati, Exponential inequalities for self-normalized martingales with applications, *The Annals of Applied Probability*. Vol. 18, pp 1848-1869 (2008).
2. M. Capitaine, C. Donati-Martin and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices : convergence and non-universality of the fluctuations. *The Annals of Applied Probability*. **37**, 1-47 (2009).
3. O. Costa, F. Dufour. Stability and Ergodicity of Piecewise Deterministic Markov Processes. *SIAM Journal of Control and Optimization*, **47**(2), 1053-1077 (2008).

4. P. Del Moral, F. Patras and S. Rubenthaler Coalescent tree based functional representations for some Feynman-Kac particle models *The Annals of Applied Probability*. Vol. 19, No. 2, 778–825 (2009).
5. P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *The Annals of Applied Probability*. 15(4) :2496-2534, (2005).

4 Collaborations académiques

L'équipe de Probabilités et Statistique possède un très fort réseau de collaborations avec les laboratoires et les organismes de recherches nationaux ainsi qu'avec de nombreuses universités internationales de tout premier plan.

Sur la thématique *Algorithmes stochastiques et méthodes de simulation*, l'équipe ALEA travaille notamment en collaboration avec N. Champagnat de l'équipe INRIA TOSCA sur les aspects dynamiques de population de type génétiques avec mutations rares, et en étroite collaboration avec F. Cérou et A. Guyader de l'équipe INRIA ASPI sur la simulation et l'analyse d'événements rares. Diverses collaborations sont aussi en cours avec F. Patras et S. Rubenthaler de l'université de Nice-Sophia Antipolis sur les propriétés fines de propagations du chaos des systèmes de particules en interaction. Depuis quelques années, l'équipe a initialisé divers projets de recherches sur le thème des polymères dirigés et les fluctuations de systèmes en interaction avec S. Tindel de l'Institut Elie Cartan de Nancy. D'autres projets de collaborations sur la chimie moléculaire et les méthodes de Monte Carlo quantiques sont en cours avec M. Rousset de l'équipe SYMPAF de Lille et B. Jourdain, G. Stoltz, et T. Lelièvre de l'École Nationale des Ponts et Chaussées. L'équipe possède des collaborations actives bien avancées, dont une ANR en cours, avec Ch. Baehr de MétéoFrance Toulouse. Enfin, l'équipe ALEA dirige une ARC INRIA EPS avec l'INRA de Montpellier (J.P. Vila) et les écoles vétérinaires d'Alfort et le LUBEM de l'université de Quimper.

Sur le plan international, l'équipe ALEA travaille depuis quelques années avec les universités suivantes : L'université de Stanford (P. Diaconis et S. Holmes), l'Imperial College of London (D. Crisan et A. Jasra), HEC Montréal (B. Rémillard), l'UBC à Vancouver (A. Doucet), L'Institut de mathématique de l'université d'Oxford (P. Tarres), et l'université d'Erlangen (A. Greven). Enfin, l'équipe ALEA est associée à une équipe INRIA 2AS avec l'université de Wuhan, Chine (Pr Wu, Li Ming).

La plupart de ces collaborations internationales s'articulent autour des méthodes de Monte Carlo, les algorithmes stochastiques, les dynamiques de populations, l'analyse de systèmes de particules en interaction et de modèles de chaînes de Markov en auto-interaction. Ces collaborations s'illustrent à travers de nombreux échanges et visites de ces chercheurs sur des postes de professeurs invités sur l'université de Bordeaux, tels P. Diaconis et S. Holmes, A. Doucet, A. Greven, B. Rémillard, et L.M Wu.

Sur la thématique *contrôle stochastique et des processus markoviens par morceaux*, l'équipe CQFD travaille localement avec le LMP CNRS UMR 5469, l'équipe MC2 EPI INRIA de Bx1, et collabore étroitement avec l'équipe TOSCA EPI INRIA, l'IMT CNRS UMR 5219, l'INSA de Rouen, et les universités Paris V, Paris VI, Paris XI. Sur le plan international, l'équipe CQFD travaille avec les universités suivantes : Wayne State University, Boston University, University of Cincinnati, Universidade de Sao Paulo, Melbourne University, Universidad Mexico. Ces collaborations se sont concrétisées par des échanges et

visites de chercheurs sur des postes de professeurs invités tels que O.L.V. Costa, R. Elliott, B. Miller, A. Piunovskiy. De 2003 à 2006, F. Dufour a été le responsable du projet CAPES/COFECUB de coopération scientifique avec le Brésil. Ce programme scientifique avait pour but de couvrir certains domaines du contrôle optimal. En particulier, il s'agissait d'étudier des problèmes de commande de systèmes non-linéaires déterministes par l'approche de la platitude et d'étudier des problèmes de contrôle optimal pour les systèmes déterministes par morceaux.

Sur la thématique *Modélisation et méthodes statistiques*, l'équipe a aussi de multiples collaborations locales avec le laboratoire d'économie GREThA (UMR CNRS, Bx4) [J. Berlin], en santé publique avec l'ISPED (équipe INSERM, Bx2) [B. Liquet, D. Commenges, P.Barbergé-Gateau, A. Sasco], , en sciences du sport (UFR STAPS, Bx2) [S.Oragio, N.Paris]. EPOC (UMR CNRS, Bx1) [G. Durrieu], le CEMAGREF [F. Vernier, Y. Legat, ...], l'INRA, l'ISVV (Institut des sciences de la vigne et du vin), et Imagen.

La liste précédente se complète sur le plan des collaborations nationales avec le CNAM Paris (A. Gannoun), l'Institut Elie Cartan de Nancy (S. Ferrigno), les Institut de Mathématiques de Bourgogne et de Toulouse (H. Cardot, M. Chaouch, C. Goga), l'école d'ingénieurs de l'Université d'Angers, le Laboratoire Jean Kuntzmann de Grenoble, l'INRIA Rocquencourt (Y. Lechevallier).

Enfin sur le plan international, l'équipe possède d'étroites collaborations avec les universités suivantes : Brunel University (UK), Cantabria Universidad (Espagne), Università di Napoli (Italie), Universidade de Recife (Brésil), Universität Postdam (Allemagne), Steklov Institute (Russie). Enfin, Bernard Bercu est actuellement responsable d'une action ECOS entre la France et le Mexique. Ce programme de coopération scientifique internationale porte sur les problèmes de rupture et de contrôle adaptatif pour des modèles de régression. Il regroupe une vingtaine d'enseignants chercheurs et de doctorants des universités de Bordeaux, Toulouse, Mexico et de Puebla.

5 Activités d'encadrement de la recherche

L'équipe est investie dans de nombreuses activités d'encadrements de la recherches, tant sur le plan de l'enseignement et l'encadrement de thèses, que sur son implication dans différents comités d'expertise d'organismes de recherches ou dans l'édition de journaux scientifiques internationaux. Ces activités d'encadrement s'illustrent aussi avec l'organisation de nombreuses conférences, ateliers de travail, et autres manifestations scientifiques sur l'université de Bordeaux. Ces trois aspects sont décrits en détail dans les sections suivantes.

5.1 Enseignement et encadrement de thèses

En premier lieu, l'équipe joue un rôle essentiel dans l'organisation d'un master lié aux probabilités et à la statistique sur l'université de Bordeaux. Depuis 2009, M. Chavent et B. Bercu sont responsables de la spécialité *Statistique et Fiabilité* du Master MIMSE (Ingénierie Mathématique, Statistique et Economique) des universités Bordeaux 1, 2 et 4, V. Couallier ayant la responsabilité du M2 professionnel dans cette spécialité. De 2007 à 2009, F. Dufour et M. Nikulin étaient les responsables de cette spécialité. J. Saracco a eu quant à lui la responsabilité du M2 professionnel *Ingénierie des risques économiques et financiers* de la

spécialité 4 *Ingénierie économique* de ce master. Depuis 2007, F. Dufour est responsable de la deuxième année de l'école d'ingénieurs MATMECA de l'Université Bordeaux 1.

Concernant l'encadrement des thèses sur la période 2005-2009, l'équipe offre une variété de sujets de thèse sur divers sujets émergents des probabilités numériques et de la statistique avec un soucis constant sur le devenir des étudiants. Les thèses en cours sont décrites ci-dessous avec les sujets, les financements, et les différents enseignants chercheurs qui les encadrent :

- A partir de septembre ou octobre 2009, Hu Peng. Sujet : Simulation particulière d'événements rares. Applications en mathématiques financières (financement Allocations ministère, P. Del Moral).
- A partir de septembre ou octobre 2009, François Giraud : Inversion stochastique pour l'imagerie radar SAR/ISAR (financement CEA CESTA, P. Del Moral, P. Minvielle).
- Depuis septembre 2008, Michele Pace. Sujet : analyse numérique et aspects informatiques d'une nouvelle classe d'algorithmes de recherches aléatoires en interaction (financement INRIA ; F. Caron et P. Del Moral).
- Depuis septembre 2008, Aurélie Le Cain. Sujet : Modélisation statistique de champs électromagnétiques associés à la superposition de faisceaux laser (financement CEA CESTA ; B. Bercu, P. Del Moral et A. Bourgeade).
- Depuis 2008, Noureddine Saaidia. Sujet : études de la famille des lois d'Inverse Gaussiennes et ses applications en analyse de survie et fiabilité. (S. Nikoulin et N. Seddik).
- Depuis novembre 2007, Genia Babykina. Sujet : Prise en compte de facteurs de risque dépendant du temps dans un modèle de prévision des casses de canalisations sous pression (financement Cemagref, J. Saracco et V. Couallier).
- Depuis octobre 2007, Jade Nguyen. Sujet : Sur les méthodes d'estimation récursive pour les modèles semiparamétriques (financement CNRS & projet région Aquitaine, B. Bercu, J. Saracco).
- Depuis septembre 2007, Victor Vazquez. Sujet : Résultats asymptotiques pour les modèles ARX en poursuite adaptative (financement CONACYT, B. Bercu et R. Montes de Oca)
- Depuis septembre 2006, Karen Gonzalez. Sujet : Processus markoviens déterministes par morceaux et fiabilité. (financement MENR ; B. de Saporta, F. Dufour).
- Depuis septembre 2006, Vanessa Kuentz. Sujet : Contributions à la réduction de dimension. (financement MENR ; M. Chavent, J. Saracco).

Les thèses qui se sont achevées dans la période considérée sont décrites ci-dessous avec les dates précises, les encadrants, et les postes actuels des doctorants :

- M. Rousset 04-07 (P. Del Moral et L. Miclo). [CR INRIA à Lille].
- Ch. Baehr 04-08 (D. Bakry et P. Del Moral). [ingénieur à Météo France].
- S. Orazio 04-08 (L. Gerville-Réache et S. Nikoulin).
- M. Chaouch 05-08 (J. Saracco) [ATER à l'Université de Bourgogne].
- S. Juillard, 04-07 (F. Dufour). [ingénieur financier à la CNP].
- S. Marret, 04-08 (F. Dufour). [ingénieur à DCNS].
- L. Sanae, 04-08 (F. Dufour). [ingénieur financier à NATAXIS].
- R. Lopez 03-06 (B. Bercu et J.C. Fort). [ingénieur à ASTRIUM].
- P. Cenac 03-06 (B. Bercu et G. Fayolle). [MCF à l'université de Bourgogne].

5.2 Expertises et activités d'édition

La plupart des membres de l'équipe participent ou ont participé à des jury de concours ou des comités d'expertise universitaires ou d'organismes de recherches nationaux :

M. Chavent était membre expert du jury du concours externe CNRS n°BAP E 154 d'accès au corps d'ingénieur d'étude (profil statisticien) en 2008.

Depuis 2007, F. Dufour est responsable de la deuxième année de l'école d'ingénieurs MATMECA de l'Université Bordeaux 1.

P. Del Moral a été membre expert du comité d'évaluation CNRS de l'UMR 6620, Lab. de Math. de l'Université B. Pascal de Clermont Ferrand en 2007. Il a aussi été membre des commission spécialistes, sections 25 et 26, du laboratoire J.A. Dieudonné de 2004 à 2007, et sections 61 du lab. IS3-CNRS de Sophia-Antipolis en 2005.

Depuis 2008, F. Dufour et P. Del Moral sont membres de la commission Jeunes Chercheurs du Centre de recherche INRIA Bordeaux - Sud-Ouest. Depuis 2008, F. Dufour est membre du Jury CR2 INRIA du Centre de recherche INRIA Bordeaux - Sud-Ouest.

En 2009, B. Bercu a été membre des commissions de sélection de Pau, l'INSA et l'Université Paul Sabatier de Toulouse.

Sur le plan de l'édition, l'équipe est investie dans l'édition d'une dizaine de journaux internationaux de probabilités et de statistique :

The Case Studies in Business, Industry and Government Statistics (CSBIGS) journal (J. Saracco, éditeur associé depuis 2006), Statistical Inference for Stochastic Processes (M. Nikulin, éditeur associé depuis 2002), Siam Journal on Control and Optimization (F. Dufour, éditeur associé depuis 2009), Revista de Matemática : Teoría y Aplicaciones (P. Del Moral, éditeur associé depuis 2009), ESAIM : Proceedings (P. Del Moral, éditeur en chef avec E. Cancès, et J.-F. Gerbeau, depuis 2006), Applied Mathematics and Optimization (P. Del Moral, éditeur associé depuis 2008), Stoch. Proc. and their Applications (P. Del Moral, éditeur associé de 2006 à 2009), Stochastic Analysis and Applications (P. Del Moral, éditeur associé depuis 2001), M2AN (P. Del Moral, éditeur invité avec Nicolas Hadjiconstantinou (MIT University) pour un volume special en 2010 sur les méthodes probabilistes.

5.3 Organisation de manifestations scientifiques

Les membres de l'équipe ont organisé de nombreuses manifestations scientifiques sur différents thèmes de la statistique et des probabilités appliquées sur l'Université de Bordeaux ou sur d'autres universités nationales. Certaines manifestations permettent de mettre en place, ou s'inscrivent dans des collaborations inter universitaires ou européennes. D'autres manifestations sont étroitement liées à des projets ANR ou aux activités de recherches de l'équipe en collaborations avec ses partenaires industriels. Elles permettent un transfert bilatéral entre nos recherches et l'industrie. Enfin, l'équipe s'investit dans l'organisation de manifestations scientifiques d'associations nationales de chercheurs en probabilités et statistique sur l'université de Bordeaux. Ces dernières sont décrites en détail ci-dessous :

- Les prochaines **Journées MAS de la SMAI** auront lieu à Bordeaux en 2010 sur le thème Algorithmes Stochastiques et Combinatoire.
- Le Workshop en l'honneur de W.Q. Meeker, "Advances on Degradation and Accelerated Life-Testing Models with Applications to Reliability and Survival Analysis", aura lieu en novembre 2009 (M. Nikulin & N. Limnios).

- Cinquième Rencontre de Statistiques Mathématiques entre les universités de Bordeaux, Santander, Toulouse et Valladolid, Juin 2009 (B. Bercu).
- The Sixth International Conference on Mathematical Methods in Reliability”, 22-27 Juin, 2009, Moscow, Russie, organisée par N.Balakrishnan, M.Nikulin et V.Rykov.
- 41-ièmes Journées de Statistique, du 25 au 29 Mai 2009, Bordeaux (A. Gégout-Petit, présidente du comité d’organisation, et équipe de probabilités et statistique de Bordeaux), 450 participants.
- Journée SMAI-IMdR Mathématiques Appliquées et Sûreté de Fonctionnement, 6 février 2009, Pau (B. de Saporta et C. Paroissin), 45 participants.
- Workshop in Honor of W.Q. Meeker, Advances on Degradation and Accelerated Life-Testing Models with Applications to Reliability and Survival Analysis, Novembre 2009 (M. Nikulin & N. Limnios).
- Workshop on Multitarget Tracking, Mai 2009, Bordeaux (F. Caron).
- Workshop on Stochastic Algorithms, Juin 2009, Bordeaux (P. Del Moral & L.M. Wu, équipe associée INRIA 2AS).
- International conference on accelerated life testing in reliability and quality control, Juin 2008, Bordeaux organisée par M. Nikulin, et F. Guerin (100 personnes).
- International conference on accelerated life testing in reliability and quality control, Juin 2008, Bordeaux (M. Nikulin, F. Guerin).
- Journées de Probabilités et Statistique de Bordeaux with the CEA CESTA Octobre 2008 (P. Del Moral).
- Journées Toeplitz, Probabilités et Matrices Aléatoires de Bordeaux. Juin 2007, Bordeaux (B. Bercu).
- Workshop ”Asymptotical Statistics and Its Applications in Survival Analysis and Reliability”, en l’honneur du Professor Ildar Ibragimov”, 6 et 7 Septembre 2007, organisé by M.Nikulin and D.Commenges (80 personnes).
- Workshop on Polymers models and related topics, Nice, Février 2007 (P. Del Moral) dans le cadre du projet ANR Chaire d’Excellence de P. Diaconis.
- International Workshop on Rare Event Simulation (P. Del Moral, G. Rubino, B. Tuffin), Nice Avril 2007 dans le cadre du projet ANR Chaire d’Excellence de P. Diaconis.
- International Conference ”Statistical Models for Biomedical and Technical Systems”, Cyprus, 29-31 mai, 2006, organisée par Ilia Vonta et Mikhail Nikulin (180 personnes)
- Conférence MAS Lille 2006, Session : *Méthodes particulières en physique et en ingénierie stochastique* (P. Del Moral).
- Depuis 1998 les professeurs C.Huber, N.Limnios, M.Mesbah and M.Nikuline forment le Comité Scientifique du Séminaire Européen, Mathematical Methods for Survival Analysis, Reliability and Quality of Life.

6 Points forts par thématique

L’équipe est essentiellement organisée autour d’un thème interdisciplinaire fédérateur : *l’inférence statistique et l’ingénierie stochastique* en parfaite adéquation avec le tissu scientifique universitaire des universités, des écoles d’ingénieur de Bordeaux et du Centre INRIA de Bordeaux Sud ouest. Les membres de l’équipe ont un spectre de recherche assez large des

aspects fondamentaux vers les aspects plus appliqués des probabilités, de la statistique, et des sciences de l'ingénieur. Cette structuration interdisciplinaire est couplée avec un investissement très dynamique des membres de l'équipe avec de nombreux partenaires industriels et divers organismes ou instituts de recherches nationaux, tels l'INRIA, le CNRS, le CEA, l'INSERM, et l'INRA. Cette synergie offre à l'équipe de nombreux contrats industriels variés avec une forte implication dans la recherche fondamentale tant au niveau des publications dans des revues de tout premier plan qu'au niveau des projets ANR nationaux.

Les points forts de l'équipe sur le plan de la recherche fondamentale et appliquée sur la période 2005-2009 sont brièvement développés ci-après autour des deux axes principaux de l'équipe : *La modélisation et méthodes statistiques & la modélisation et méthodes stochastiques*. Les premiers résultats marquants de l'équipe sur le nouveau thème de recherche lié aux *matrices aléatoires* sont développés dans la section 6.2.3.

6.1 Modélisation et méthodes statistiques

Cet axe de recherche de l'équipe est tout d'abord associé à des études fondamentales sur une classe riche de modèles et méthodes de la statistique mathématique : étude de modèles de régression semi-paramétriques, et non-paramétriques, étude de quantiles conditionnels uni-variés et multi-variés, méthode de réduction de dimension.

Les deux principales activités de recherches appliquées concernent : *les modèles et méthodes statistiques pour l'analyse de survie et la fiabilité, et la statistique multidimensionnelle en incluant les méthodes et modèles semi et non paramétriques et les méthodes statistiques de classification et d'analyse des données*

Ces deux aspects seront développés en détails dans la section 6.1.1 et la section 6.1.2. L'équipe est aussi investie dans de nombreuses projets de recherches de statistique appliquée. Ces projets sont décrits dans la section 6.1.3

6.1.1 Analyse de survie et la fiabilité

Cette thématique est l'un des principaux axes de recherche des statisticiens de l'université Bordeaux 2 depuis 1994. Elle concerne essentiellement les enseignants chercheurs suivants :

Vincent Couallier (MCF, Bx2), Jérôme Poix (MCF, Bx2), Léo Gerville-Réache (MCF, Bx2), Mikhail Nikulin (PR, Bx2).

Les études menées dans cette thématique portent sur la modélisation et l'inférence statistique de problèmes de survie et de fiabilité basées sur les modèles de dégradation et de vies accélérées. Ces nouvelles méthodes élaborées au cours des quinze dernières années par l'équipe sont adaptées à l'analyse de survie pour les systèmes complexes évoluant dans un environnement dynamique (exprimé en terme de vecteur de covariables dépendant du temps). Ces résultats ont un fort impact dans le milieu médical, en anthropologie, en biologie, en cancérologie, en démographie ainsi que dans l'analyse de la durée de vie de composants industriels et de la fiabilité de systèmes complexes issus des sciences de l'ingénieur.

La qualité de ces recherches peut se mesurer au nombre de publications et au nombres de projets de collaborations des chercheurs investis dans cette thématique avec des organismes de recherches nationaux et internationaux tels l'ISPED, de l'Université Bordeaux2 (M. Nikulin, D. Commenges, P.Barbergé-Gateau et A. Sasco), l'institut Bergonié de Bordeaux (V. Couallier et J. Bonnet), l'Institut Mathématique de Steklov Saint Petersburg (M. Nikulin

et V. Solev), le Département de statistiques de l'Université de Vilnius (M. Nikulin, V. Bagdonavicius), La *School of Public Health*, China University of Taiwan (M. Nikulin, H. Wu), L'institut ISTIA d'Angers (avec F. Guerin), le laboratoire de mathématiques appliquées de l'UTC, Compiègne (avec N. Limnios), L'université René Descartes Paris V (avec C. Huber).

Les résultats de modélisation et d'inférence statistique développés dans cette thématique sont divers et variés. Les premiers portent sur la modélisation et l'estimation en analyse de survie sous environnement variable, prenant en compte des processus de stress dépendants du temps. Dans ce contexte, l'équipe a développé des extensions de modèles et de méthodes d'estimation associées pour des covariables qui ne sont plus des vecteurs fixes réels mais des trajectoires de processus. Il en a découlé une généralisation naturelle des modèles de Cox, les modèles GPH, GLPH, avec un développement systématique des tests associés. Ces travaux théoriques ont été appliqués en biologie sur des données de cancer du poumon (avec V. Bagdonavicius et M. Hafdi).

Dans la même orientation, l'équipe a développé de nouveaux modèles permettant des croisements de fonctions de survie (cross-effects), effets non pris en compte dans les modèles classiques et qui pourtant sont envisageables dans de nombreuses applications biomédicales. L'impact de la chimiothérapie couplée à la radiothérapie a été ainsi analysée (publications et collaborations avec H. Wu, Taiwan, et V. Bagdonavicius, Lituanie).

Une autre avancée concerne la prise en compte de processus de dégradation partiellement observés et la modélisation des durées de vie correspondantes. Différents modèles ont été proposés qui généralisent les modèles de temps d'atteinte d'un processus pour un seuil fixé. Dans le cadre des modèles à risques compétitifs, l'influence des covariables dépendant du temps (vues comme des trajectoires partiellement observées de processus stochastiques) peuvent être classées en deux classes : les variables endogènes, modélisant par exemple un processus de dégradation ou de perte de performance, et les covariables exogènes, modélisant la dynamique de variation des conditions expérimentales.

Parmi d'autres, les applications en gérontologie (M. Nikulin, P. Barberger-Gateau) sur la démence, en génétique (V. Couallier, J. Bonnet) sur la dégradation de l'ADN, en industrie automobile (L. Gerville-Réache dans le cadre d'une thèse CIFRE), sur la dégradation de pièces moteurs témoignent du fort impact de ces recherches.

Les modèles de vie accélérées ont aussi été appliqués à l'économétrie pour les systèmes de crédits et l'analyse du risque de ruine. M. Nikulin et J. Poix travaillent actuellement sur ce thème très porteur.

D'autres études reposent sur le conditionnement de l'intensité de défaillance de systèmes en fonction de la trajectoire observée de dégradation, ou sur le développement de nouvelles méthodes d'estimation paramétrique par maximum de vraisemblance pour des processus de comptage modélisant des systèmes réparables avec maintenance corrective imparfaite. Dans le cadre d'un co-encadrement de thèse (V. Couallier) avec le Cemagref, un nouveau modèle stochastique pour la maintenance imparfaite de systèmes réparables avec covariables dépendant du temps a été développé en collaboration avec Yves Le Gat.

Enfin, les développements les plus récents concernent la mise en oeuvre des modèles de vie accélérée dans le cadre des systèmes réparables et redondants. Dans une optique de maintenabilité maximale, les unités en stand-by attendent dans un état intermédiaire entre le fonctionnement nominal et l'attente à *froid*. L'attente en redondance est alors assimilée par le principe de Sedyakin à un niveau abaissé de contrainte. Les études portent actuellement sur l'analyse du comportement asymptotique des estimateurs dans un modèles

semi-paramétrique. Les premiers résultats ont été publiés en 2008 dans les proceedings du séminaire européen (*Mathematical Methods in Survival Analysis and Quality of Life*, www.dma.utc.fr/nlimnios/SEMINAIRE).

6.1.2 Statistique multidimensionnelle

Cette thématique est très active et en plein développement depuis ces trois dernières années au sein de l'équipe. Elle concerne les enseignants chercheurs suivants :

Bernard Bercu (PR, Bx1 et équipe INRIA CQFD), François Caron (CR, INRIA équipe ALEA), Marie Chavent (MCF, Bx2 et équipe INRIA CQFD), Anne Gégout-Petit (MCF, Bx2 et équipe INRIA CQFD), Pierrick Legrand (MCF, Bx2 et équipe ALEA), Brigitte Patouille (MCF, Bx2).

On peut faire ressortir trois grands axes qui recouvrent le large spectre des compétences de l'équipe en Statistique multidimensionnelle tant sur le plan théorique qu'appliqué.

- Le premier axe concerne les méthodes semi et non paramétrique dans le cadre de modèles de régression. Une compétence majeure de l'équipe concerne plus particulièrement les méthodes de type SIR (Sliced Inverse Regression), permettant l'estimation d'un modèle semiparamétrique très général de régression. Un autre compétence porte sur l'estimation nonparamétrique des quantiles conditionnels ou non, uni-variés ou multivariés (spatiaux) pour des données complètes ou censurées.
- Le second axe concerne les méthodes statistiques de classification et d'analyse des données. L'équipe possède des compétences en classification supervisée ou non supervisée avec des approches géométriques et stochastiques, certaines méthodes permettant d'obtenir une classification d'individus (parfois modélisés par des intervalles), d'autres de variables. Un autre point fort de cet axe est le développement de nouvelles méthodes d'analyse de données qualitatives.
- Enfin, un troisième axe peut être mis en exergue dans le cadre de cette thématique. Il s'agit de la statistique appliquée : le point de départ des approches statistiques développées est une problématique réelle issue généralement de collaborations interdisciplinaires telles que celles concernant la pollution de l'air par exemple.

Méthodes et modèles semi et non paramétriques. Les modèles semiparamétriques de régression incluant une réduction de dimension (de la partie explicative du modèle) sont apparus dans la littérature statistique au début des années 1990 et sont en pleine extension depuis les années 2000. Parmi ces méthodes, les méthodes de type SIR ont montré qu'elles possèdent de nombreux avantages par rapport aux méthodes concurrentes tant sur le plan théorique que sur le plan computationnel et appliqué. Un des objectifs de l'équipe a été d'étendre certains résultats théoriques de ces méthodes et de développer des outils afin de rendre l'application tout à fait opérationnelle.

- Pour les aspects théoriques : diverses extensions des méthodes SIR au cas où la variable à expliquer est multidimensionnelle ont été développées [J. Saracco en collaboration L. Barreda et A. Gannoun (CNAM, Paris)]. Pour une de ces méthodes nommée PMS_α , des résultats de convergence asymptotique ont été obtenus [J. Saracco avec A. Gannoun]. Une autre extension a été développée lorsque les variables explicatives ne sont plus uniquement quantitatives, c'est à dire dans le cas où une covariable est qualitative. Les simulations ont confirmé les différents résultats de convergence obtenus pour

ce modèle [J. Saracco, en collaboration avec B. Liquet (ISPED, Bx 2)]. Dans le cadre de modèles de sélection (très utilisés en économie), un modèle semiparamétrique de sélection (avec des variables dépendantes multidimensionnelles) a été proposé et une méthode d'estimation en deux étapes (une étape de type SIR multivarié, suivie d'une étape d'analyse canonique) a été développée pour estimer la partie paramétrique du modèle, la partie fonctionnelle étant estimée nonparamétriquement. La théorie asymptotique et l'implémentation numérique ont montré tout le potentiel de cette nouvelle approche vis à vis des méthodes paramétriques existantes [M. Chavent, J. Saracco, en collaboration avec B. Liquet].

- Pour les aspects plus *méthodologiques* : pour que l'application sur données réelles des méthodes de SIR soit effective, cela nécessite de connaître la dimension du sous-espace de réduction de dimension. Nous avons donc développé une approche par bootstrap du choix simultané de la dimension K et du paramètre α pour la méthode SIR_α , approche qui a l'avantage d'avoir une visualisation graphique 3-D du critère permettant de choisir en pratique le couple (K, α) optimal pour un jeu de données réelles [J. Saracco, en collaboration avec B. Liquet]. Un autre travail permettant d'assouplir une des étapes d'estimation des méthodes SIR (qui repose sur un *tranchage* du support de la variable à expliquer) a été de proposer à la place de ce tranchage (qui peut être vu comme une classification *dure*) une étape de classification floue. Cette extension a montré son efficacité en simulations numériques [J. Saracco, en collaboration avec S. Ferrigno (Univ. Nancy) et A. Gannoun]. Tout récemment, une approche de type *Bagging* (plusieurs échantillonnages obtenus par bootstrap permettent d'obtenir des estimateurs des matrices d'intérêt qui sont ensuite agrégées) a été développée, tant sur le plan théorique que pratique, pour les méthodes de type SIR afin d'estimer plus efficacement l'espace de réduction de dimension lorsque les échantillons sont de petites tailles (faible nombre d'observations) [J. Saracco, V. Kuentz, en collaboration avec B. Liquet].

Dans la modélisation nonparamétrique, l'essentiel des compétences de l'équipe concerne l'estimation des quantiles conditionnels, univariés ou bien multivariés. Pour ce qui est des quantiles conditionnels univariés, les apports principaux concernent l'étude des estimateurs non paramétriques de type noyau ou polynômes locaux dans le cadre de données censurées que l'on rencontre souvent dans les problèmes de biostatistique. Des résultats théoriques de convergence ont ainsi été obtenus et la mise en oeuvre sur données simulées et réelles a montré la validité des approches développées [J. Saracco, en collaboration avec A. Gannoun et K. Yu (Brunel Univ.)]. Une large étude de comparaison (pratique et théorique) avec les différents estimateurs non paramétriques concurrents de la fonction de répartition conditionnels et des quantiles conditionnels a aussi été faite dans ce cadre de données censurées [J. Saracco, en collaboration avec A. Gannoun et K. Yu]. Lorsque la variable à expliquer est multidimensionnelle, on parle de quantiles multivariés ou spatiaux ou encore géométriques. L'extension de la notion de quantiles dans des espaces de dimension supérieure ou égale à 2 n'est pas directe, ni tout à fait naturelle du fait qu'il n'y a pas de relation d'ordre total dans \mathcal{R}^d . Nous avons dans ce cadre étudié la notion de quantiles géométriques (introduits par Chaudhuri en 1996) conditionnels ou non conditionnels et nous avons proposé des estimateurs non paramétriques de ces quantités ainsi que les algorithmes itératifs correspondants. Nous avons aussi précisé un certain nombre de propriétés asymptotiques pour ces estimateurs [J. Saracco, en collaboration avec M. Chaouch (Univ. de Bourgogne) et A. Gannoun].

Pour poursuivre sur cet axe de recherche, on peut aussi mettre en avant qu'une approche semiparamétrique d'estimation de courbes de référence couplant l'estimation de quantiles conditionnels unidimensionnels ou multivariés et la réduction de dimension via les méthodes SIR a été développée tant sur les aspects théoriques qu'appliqués dans le cadre d'un contrat de recherche avec le CERIES (centre de recherche de sur la peau humaine financé par CHANNEL) portant sur la construction de courbes de référence pour des propriétés biophysiques de la peau humaine [J. Saracco en collaboration avec C. Guinot (CERIES), A. Gannoun et B. Liquet].

Nous nous sommes également intéressés à des problèmes de sélection de modèles pour des données de survie : nous proposons de choisir entre un modèle à risques proportionnels et un modèle stratifié en utilisant le critère théorique ELL (Expected Log-Likelihood), proche du critère optimal KL (Kullback-Leibler), que nous estimons par le critère LCV (Likelihood Cross Validation). Nous avons illustré les performances de cette approche sur simulations et nous l'avons mise en oeuvre sur des données réelles [J. Saracco, en collaboration avec B. Liquet]. Dans le cadre de l'analyse des niveaux d'expression en génomique (domaine en grand développement actuellement), nous avons développé une nouvelle méthodologie purement non-paramétrique [J. Saracco, en collaboration avec A. Gannoun et B. Liquet].

Enfin, nous avons aussi proposé récemment un estimateur semiparamétrique du mode conditionnel dans le cadre d'un modèle de régression. Cet estimateur repose sur un combinaison paramétrique des estimateurs nonparamétriques de mode, de la médiane et de l'espérance conditionnelle, il en découle un algorithme itératif d'estimation. Les propriétés de convergence des estimateurs des parties paramétrique et fonctionnelle ont été étudiées [J. Saracco, en collaboration avec A. Gannoun et K. Yu].

Méthodes statistiques de classification et d'analyse des données.

L'axe de recherche en classification est en pleine expansion depuis quelques années. Il se décline sous les deux aspects statistique et stochastique.

Sur le plan statistique, les méthodes de classification développées s'inscrivent dans le cadre de la réduction de dimension et plus particulièrement autour de la classification automatique d'un tableau de données multidimensionnelles. L'approche adoptée est géométrique dans le sens où les problèmes de classification sont vus comme des problèmes d'optimisation d'un critère d'homogénéité, ce critère étant basé sur des distances pour la classification d'individus et sur des mesures d'association pour la classification de variables.

Dans ce cadre, nous avons développé la méthode DIVCLUS-T qui est une méthode descendante de classification hiérarchique basée sur la même approche que les arbres de décision mais d'un point de vue non supervisé [M. Chavent en collaboration avec Y. Lechevallier (DR INRIA) et O. Briant (MdC, INPG)]. Une variante intégrant des contraintes de contiguïté sur les individus a été envisagée et appliquée à la classification d'unités hydrographiques du bassin versant de la charente [M. Chavent en collaboration avec Y. Lechevallier et F. Vernier (CEMAGREF)]. Nous avons également mis en oeuvre cette méthode pour répondre à une problématique en comptabilité financière [M. Chavent en collaboration avec H. Stolowy (Pr HEC)].

D'autres résultats concernent les méthodes de classification de données décrites par des intervalles. Ces résultats sont associées à des études géométriques pour la définition d'un hyper-rectangle central optimisant une mesure d'adéquation permettant pour la construction d'un algorithme de classification des nuées dynamiques convergent [M. Chavent et J.

Saracco]. Les résultats obtenus avec la distance de Hausdorff ont été appliqués pour la définition d'une version normalisée permettant de gérer les problèmes d'échelles de mesures différentes, ou encore pour la définition d'une version avec distances adaptatives [M. Chavent, en collaboration avec R. Verde (univ. de Naples), et F. De Carvalho (univ. de Récif)].

Dans certaines applications, on peut s'intéresser spécifiquement à la classification de variables plutôt qu'à la classification des observations. C'est le cas par exemple en analyse sensorielle (mise en place de groupes de descripteurs), en biochimie (classification de gènes), en marketing (segmentation d'un panel de consommateurs), etc. La classification de variables peut également être l'une des solutions possibles pour réduire la dimension d'un tableau de données, tout en conservant le plus d'information possible. Concernant les variables qualitatives, nous avons développé et étudié différents algorithmes de classification (k-means, classification hiérarchique ascendante et descendante) pour la classification de variables qualitatives. Les prototypes des classes de variables qualitatives sont des variables latentes quantitatives maximisant le critère d'homogénéité égal à la somme des rapports de corrélations des variables de la classe à la variable latente [M. Chavent, V. Kuentz, J. Saracco].

Il existe un lien entre l'analyse factorielle et la classification de variable. En effet l'utilisation de la rotation en analyse factorielle permet d'obtenir des facteurs plus clairement corrélés aux variables et donc implicitement des groupes de variables. Cette question de la rotation étant peu abordée dans le cas qualitatif, nous avons proposé un critère de rotation en Analyse des Correspondances Multiples basé sur des rapports de corrélations. Nous avons obtenu une solution analytique en dimension deux. Cette solution analytique est en particulier utilisée pour améliorer les méthodes de classification de variables qualitatives [M. Chavent, V. Kuentz, J. Saracco].

Une approche alternative pour la classification non supervisée a été développée, basée sur des modèles bayésiens non/semi paramétriques. Ces méthodes font l'hypothèse que les données sont distribuées selon un mélange infini de distributions et permettent l'estimation du nombre de classes à partir des données. Ces modèles sont estimés à l'aide de méthodes de Monte Carlo par Chaîne de Markov et séquentielles [F. Caron]. Un des points forts concerne le développement d'une nouvelle classe de processus de Dirichlet non homogènes préservant la nature de Dirichlet sur chaque marginale temporelle, pour des applications de classification dynamique non supervisée ([F. Caron, en collaboration avec M. Davy et A. Doucet]).

D'autres méthodes classification ont été développées dans le cadre du projet HEVEA (Handicap : Etude et Valorisation de l'Ecologie Auditive), qui a pour but de contribuer au réglage des implants cochléaires (prothèses auditives) afin d'optimiser la communication des personnes présentant des déficiences auditives. Des méthodes ondelette et des analyses de régularité ont été réalisées pour aboutir à une classification des environnements sonores du patient. D'autre part, la sélection automatique de paramétrage se fait à l'aide d'un algorithme évolutionnaire interactif [P. Legrand].

6.1.3 Statistique appliquée

Une étude sur la pollution atmosphérique a été réalisée en réponse à l'APR "Particules" du programme PRIMEQUAL/PREDIT piloté par le MEDAD et l'ADEME. La question

étudiée a été l'identification et la quantification des contributions de sources de poussières fines à un environnement. Ce projet, fruit d'une collaboration avec un organisme agréé de surveillance de la qualité de l'air (AIRAQ) et un laboratoire de physique nucléaire spécialisé dans l'analyse élémentaire par faisceaux d'ions de haute énergie (CENBG), a été réalisé selon un protocole en trois phases : le prélèvement de particules fines sur le site d'intérêt, l'analyse chimique de ces prélèvements à l'aide de la technique multi-élémentaire PIXE et enfin le traitement statistique des données. Dans cette dernière phase, nous avons étudié les limites des outils classiques d'analyse factorielle et nous avons proposé d'utiliser des modèles spécifiques qui s'apparentent à l'approximation d'une matrice par le produit de deux matrices de rang inférieur, sous certaines contraintes physiques dont la positivité des coefficients. L'algorithme PMF suivi d'une normalisation des résultats en fonction des contraintes physiques du modèle, a ainsi permis d'identifier les sources et surtout de quantifier leurs contributions à l'empoussièrement sur la période de l'étude [M. Chavent, B. Patouille, V. Kuentz, J. Saracco, en collaboration avec H. Guégan (CENBG)].

Nous venons de répondre au nouvel APR 2009 *Pollution atmosphérique longue distance* du programme PRIMEQUAL, toujours en collaboration avec la cellule ARCANE du CENBG et l'organisme agréé de surveillance de la qualité de l'air ATMO Poitou-Charentes. Le contexte de cette proposition est l'étude expérimentale des phénomènes de transport de particules à longue distance. L'originalité de cette proposition réside dans la nouvelle méthode d'analyse multi-élémentaire envisagée (technique PESA) permettant de doser de façon non destructive certains éléments chimiques non mesurés avec la méthode PIXE, dans le protocole de prélèvement sur deux sites simultanément, et sur les méthodes statistiques qui seront développées pour tenir compte de l'aspect multi-bloc des données recueillies.

Dans un autre contexte, une étude sur l'analyse de durée de vie de prothèses lombaires, sujet proposé par la société Styker Spine, a fait l'objet d'un contrat industriel et d'une publication dans *Journal of Statistical Planning and Inference*. Il s'agissait de modéliser et d'estimer la durée de vie d'arthrodèses lombaires à partir de données d'essais effectuées sur plusieurs configurations géométriques (plusieurs tailles, plusieurs longueur de vis, présence/absence de connecteur) pour définir les effets de vieillissement sous contraintes nominales ou accélérées. Ce travail a nécessité des ressources théoriques et pratiques sur les modèles de vie accélérée.

Avec l'institut Bergonié et la société Imagene, une collaboration à long terme existe et porte sur l'analyse statistique et le calcul de probabilité pour la modélisation de la dégradation de l'ADN extrait à l'état sec. L'équipe intervient comme support mathématique et statistique pour le développement de la R&D d'Imagene dans la recherche de nouveau procédé de conservation de l'ADN. Les résultats les plus importants ont fait l'objet d'une première publication scientifique dans *Analytical Biochemistry* (un autre article est en révision pour *Nucleic Acid Research*).

Enfin, une collaboration avec le laboratoire d'Anthropologie des Populations du Passé, UMR 5199 - PACEA, Université Bordeaux 1 a également donné lieu à de nombreuses applications de la Statistique, portant le plus souvent sur les modèles linéaires, linéaires généralisés, avec estimation classique (par maximum de vraisemblance) ou par GEE. La première étude a porté sur une série de squelettes portugais (fin du 19^{ème}/début du 20^{ème} siècle). Son objectif était d'étudier les relations entre âge, perte osseuse, arthrose et ostéoporose (article publié dans *HOMO - Journal of Comparative Human Biology*, 2007). Une seconde étude sur des collections de squelettes a porté sur l'analyse statistique de données multidimension-

nelles dichotomiques codant la présence d'entéropathies et sur un modèle permettant de prédire ces atteintes en fonction de variables telles que l'activité physique, le genre, l'âge ou la collection. Un article a été soumis à *American Journal of Physical Anthropology* en 2009.

6.2 Modélisation et méthodes stochastiques

Les recherches menées dans cette thématique sont diverses et variées.

Les premières portent sur *la stabilité de processus stochastiques et plus particulièrement sur le contrôle optimal*. Ces deux axes de recherche se sont fortement développés ces dernières années tant du point de vue théorique que des applications en analyse de risque et en sûreté de fonctionnement. Les principaux résultats liés au contrôle stochastique seront développés dans la section 6.2.1.

Dans un autre registre, l'équipe étudie aussi l'analyse de systèmes de particules en interaction ou branchements de type champ moyen et les modèles d'arbres généalogiques associés, les principes de grandes déviations et les inégalités de concentration exponentielles, l'analyse fine de la convergence d'algorithmes stochastiques pour l'estimation et l'optimisation globale, les techniques de martingales, et l'analyse spectrale de matrices aléatoires. Les principaux résultats liés *aux algorithmes stochastiques et à l'étude des matrices aléatoires* seront développés respectivement dans la section 6.2.2, et dans la section 6.2.3.

Deux autres thèmes émergent sont abordés dans l'équipe :

Le premier concerne l'estimation de processus autorégressifs. Il s'agit d'utiliser des résultats de convergence de martingales pour l'étude asymptotique de l'estimateur des moindres carrés du vecteur des paramètres d'un processus auto-régressifs de bifurcation. Les principaux résultats sur ce thème concernent la consistance et l'analyse des fluctuations des estimateurs par des techniques de martingales.

D'autres études concernent l'utilisation de processus stochastiques pour l'étude des modèles statistiques de causalité. Il est le résultat d'une collaboration avec l'équipe biostatistique de l'Unité INSERM U 897 de Bordeaux. Les principaux résultats sur ce thème concernent l'étude des propriétés d'indépendance locale faible conditionnelle entre processus stochastiques dans une large classe de semi-martingales et une définition originale liée à la notion de vraisemblance sur la base de certains processus, en utilisant le théorème de Girsanov. Ces travaux permettent d'associer un modèle graphique à cette notion d'indépendance pour une classe de processus stochastiques. Ce thème de recherche est transverse par rapport aux thèmes centraux de l'équipe proba-stat de l'IMB.

6.2.1 Contrôle stochastique

Cette thématique est au coeur des recherches de l'équipe INRIA CQFD. Elle concerne les enseignants chercheurs suivants :

Bernard Bercu (PR, Bx1 et équipe INRIA CQFD), François Dufour (PR, IPB, responsable de l'équipe INRIA CQFD), Anne Gégout-Petit (MCF, Bx2 et équipe INRIA CQFD), Benoîte de Saporta (MC, Bx4 et équipe CQFD), Huilong Zhang (MCF, Bx1 et équipe CQFD).

Cet axe de recherche s'est pleinement développé ces dernières années tant sur le plan théorique que sur le plan des applications en analyse de risque et en sûreté de fonctionnement

Concernant les aspects théoriques, les premiers résultats marquants de l'équipe portent sur des problèmes de stabilité stochastique et de contrôle optimal de processus markoviens.

Un premier axe de recherche concerne le contrôle optimal singulier d'équations différentielles stochastiques. Il a été dégagé de nouveaux résultats d'existence en utilisant des techniques de changement de temps contrôlées. Il s'agit à notre connaissance des premiers résultats de la littérature obtenus dans un cadre aussi général : forme du coût quelconque et gain non-linéaire. De plus, il a été dégagé une nouvelle forme du principe du maximum où les variables adjointes sont solutions d'équations différentielles stochastiques rétrogrades singulières. Ces travaux se sont fait en collaboration avec B. Miller de l'Université de Monash, Melbourne, Australie.

Un second axe de recherche traite des problèmes de contrôle pour les processus markoviens déterministes par morceaux. Il existe très peu de résultats dans la littérature concernant les problèmes de contrôle dits *continus* pour des coûts moyennés à long terme contrairement au contrôle impulsionnel. En collaboration avec O.L.V. Costa (Université de Sao Paulo, Brésil), nous avons obtenus des résultats d'existence pour différentes classes de contrôle admissible.

Dans le domaine de la stabilité stochastique, nous nous sommes intéressés aux problèmes d'existence de lois invariantes et d'ergodicité pour différentes classes de processus markoviens. Il a été montré que lorsque l'ensemble test intervenant dans la condition de drift est un ensemble '*petite*', le critère de Foster-Lyapunov assure l'existence de probabilités invariantes pour une chaîne de Markov quelconque. De plus, il a été identifié une catégorie de chaînes de Markov possédant des propriétés très intéressantes, englobant les chaînes irréductibles ainsi que les chaînes qui ont une composante continue (T-chain suivant la terminologie anglosaxonne de S. Meyn et R. Tweedie). Il a été notamment montré que pour de telles chaînes, le critère de Foster Lyapunov fournit une condition nécessaire et suffisante assurant l'existence d'une loi invariante lorsque l'ensemble test est un ensemble *petite*. Ces résultats généralisent de nombreux travaux de la littérature. La décomposition de Doeblin est fondamentale dans la théorie des chaînes de Markov qui ne sont pas nécessairement irréductibles. Il a été mis en évidence que cette nouvelle classe de chaînes de Markov admet la décomposition de Doeblin. Des résultats similaires ont été analysés pour la décomposition de Yoshida. Les propriétés ergodiques (récurrence, Harris récurrence, existence de mesures invariantes non nécessairement finies, ergodicité) des processus markoviens déterministes par morceaux ont été complètement caractérisées à l'aide d'un noyau markovien qui peut être explicitement calculé en fonctions des paramètres du processus. Il s'agit des premiers résultats de la littérature concernant cette très grande classe processus. Ces travaux se sont fait en collaboration avec O.L.V. Costa (Université de Sao Paulo, Brésil).

Concernant les aspects appliqués, les résultats marquants de l'équipe portent essentiellement sur l'application des processus markoviens déterministes par morceaux sur des problématiques liées à la fiabilité et à la sûreté de fonctionnement de systèmes complexes issus de l'industrie (analyse de pannes ou de dégradations physiques). Cet axe de recherche s'est pleinement développé dans l'équipe CQFD en collaborations avec des partenaires industriels comme EADS-Astrium sur des problématiques de propagation de fissure et de maintenance préventive. En particulier, il a été montré à partir de l'analyse de cas tests de la littérature que les processus markoviens déterministes par morceaux offrent un cadre de modélisation très général et particulièrement pertinent pour traiter des problèmes de fiabilité dynamique.

6.2.2 Algorithmes stochastiques

Cette thématique, au coeur des recherches de l'équipe ALEA, est en pleine expansion depuis ces deux dernières années. Elle concerne principalement les enseignants chercheurs suivants :

Bernard Bercu (PR, Bx1 et équipe INRIA CQFD), François Caron (CR, INRIA équipe ALEA), Pierre Del Moral (DR INRIA, équipe INRIA ALEA), Pierrick Legrand (MCF, Bx2 et équipe ALEA).

Les point forts de l'équipe sur ce thème émergent s'articulent principalement autour de deux axes de recherches :

1) La confection et analyse de nouvelles classes d'algorithmes stochastiques en interaction : échantillonneur de Gibbs et recuits simulés en interaction, algorithmes d'apprentissage et d'exploration aléatoire de type génétiques, chaînes de Markov en auto-interaction avec leurs mesures d'occupation.

2) L'étude des différents domaines d'applications liés à la physique, l'inférence bayésienne, l'optimisation stochastique et au traitement du signal : *classification non supervisée de données fonctionnelles, chaînes de Markov cachées, estimation du haut du spectre d'opérateurs de Schrödinger, filtrage de trajectoires de signaux non linéaires et non gaussien en terme d'arbres généalogiques, techniques de type recuits simulés en interaction pour l'optimisation globale de fonction non convexes et l'exploration d'espaces combinatoires complexes.*

Concernant les aspects théoriques, les résultats marquants portent sur différents résultats mathématiques sur l'analyse fine d'algorithmes stochastiques :

Le premier concerne l'étude des propriétés de concentration exponentielle, les fluctuations autour des lois limites, les inégalités de concentration pour les martingales [B. Bercu, en collaboration avec A. Touati], ainsi que la concentration exponentielle et les déviations modérées de systèmes de particules en interaction de type champ moyen [P. Del Moral, en collaboration avec E. Rio (Pr. Univ. Versailles, en détachement INRIA de 2008 à 2009 dans l'équipe ALEA), B. Bercu et L.M. Wu (Pr. Wuhan Univ. et Clermont Ferrand)]. Ces travaux développent des techniques de semigroupes permettant d'obtenir des théorèmes de la limite centrale fonctionnels pour une classe générale de systèmes de particules en interaction, incluant entre autre les filtres particulaires, le méthodes de Monte Carlo séquentielles, les modèles de McKean Vlasov issus de la mécanique des fluides, et les algorithmes stochastiques de type génétiques. Outre ces résultats de fluctuations, ces recherches ont permis d'étendre des inégalités exponentielles de type Bennett, Bernstein et Hoeffding classiques pour les variables indépendantes ou faiblement dépendantes aux systèmes de particules en interaction. Ces résultats de concentration exponentielle sont les premiers de ce type pour des processus en interaction de type champ moyen.

Le second concerne l'analyse des propriétés de propagation du chaos pour des interprétations particulières de formules de champ moyen (P. Del Moral, en collaboration avec F. Patras [DR CNRS Nice], L. Miclo [DR CNRS Marseille], S. Rubenthaler [MdC Nice]). Ces travaux offrent des décompositions exactes des lois de blocs de particules explicites en fonction du paramètre de précision. Ces études sont les premiers résultats de ce type pour des systèmes de particules en interaction de type champ moyen. Elles offrent aussi de nouvelles connections entre la combinatoire des arbres de coalescence et les processus en interaction.

Le dernier porte sur le comportement en temps long d'algorithmes génétiques neutres avec formules explicites des lois invariantes de population finies. (P. Del Moral, en collaboration avec F. Patras [DR CNRS Nice], L. Miclo [DR CNRS Marseille], S. Rubenthaler [MdC Nice]). Ces travaux offrent tout d'abord une description explicite de la mesure invariante d'une population génétique de taille fixée. Ces résultats complètent des études existantes, la plupart fondées sur des techniques de renormalisation temporelle pour des tailles de populations infinies. Le second fait marquant de ces études concerne l'analyse de la convergence de ces populations finies vers leur état d'équilibre en fonction de la convergence à l'équilibre de l'ancêtre commun et du temps de coalescence des lignes ancestrales associées à l'arbre généalogique de la population.

Concernant les aspects plus appliqués, les résultats marquants de l'équipe portent essentiellement sur l'analyse stochastique de problèmes d'inférence statistique, d'analyse de risques ou d'optimisation globale issus de la physique et des sciences de l'ingénieur. Plus précisément, un axe majeur de recherche de l'équipe porte sur l'analyse de problèmes de filtrage de signaux ou de chaînes de Markov cachées par des techniques d'inférence bayésienne et la confection d'algorithmes stochastiques performants permettant de simuler des lois conditionnelles complexes.

Les premières études concernent le développement d'outils combinant des techniques d'analyse fractales et des algorithmes stochastiques de type génétiques pour la classification et l'optimisation stochastique. L'utilisation d'ondelettes et d'outils d'analyse fractale est appropriée à l'analyse des signaux irréguliers. La caractérisation de la régularité locale est importante dans la description de ces signaux. Pour étudier la régularité d'un signal, on utilise l'exposant de Hölder. On dispose de plusieurs méthodes permettant de l'estimer. Certaines méthodes, basées sur des décompositions en ondelettes, fournissent un environnement mathématique favorable au calcul de la vitesse de convergence d'un estimateur basé sur une conservation ou un contrôle de la régularité locale (par exemple dans une application sur le débruitage ou sur l'interpolation de signaux ou d'images). D'autre part les algorithmes évolutionnaires constituent un outil performant dans le domaine de l'optimisation. Il s'agit de techniques d'optimisation stochastique de plus en plus utilisées pour résoudre des problèmes d'optimisation complexe présentant un espace de recherche de grande dimension. La combinaison de ces outils fournit un environnement complet pour la résolution de problèmes concrets dans le domaine du vivant [P. Legrand].

D'autres résultats marquants de l'équipe concernent la confection d'algorithmes efficaces de type Monte Carlo séquentiels pour des problèmes d'inférence bayésienne associés à des processus à sauts ou à des processus de Dirichlet. Ces techniques émergentes permettent de résoudre des problèmes d'estimation de densité et la classification non supervisée. Elles ont donné lieu à diverses applications : analyse de données synchrones/asynchrones et le cas binaire de capteur valide/défaillant dans des modèles de chaînes de Markov cachées, déconvolution aveugle et détection de rupture dans des modèles de filtrage de signaux, apprentissage et classification non supervisée de données fonctionnelles [F. Caron, en collaboration avec A. Doucet].

D'autres études concernent l'application des méthodes particulières pour la simulation et l'analyse d'événements rares (F. Cérou, P. Del Moral, A. Guyader). Ces nouveaux outils de simulation modernes ont permis des avancées significatives sur ce sujet et complètent les techniques d'échantillonnage plus classiques, développées dans la communauté statistique durant ces vingt dernières années.

De nouvelles études en cours avec un fort impact portent sur l'approximation d'enveloppes de Snell pour les problèmes d'arrêt optimal par rapport à des critères de type moyenne (P. Del Moral et Hu Peng). Ce sujet fondamental en optimisation stochastique est aussi au coeur d'un projet de collaboration avec l'équipe OSIRIS de l'EDF de Clamart. D'autres aspects théoriques et numériques fondés sur des techniques de grilles déterministes sont en cours de développement dans l'équipe [F. Dufour et B. De Saporta].

L'équipe a aussi mis en place de nouvelles approximations particulières de distributions de branchements pour les problèmes de filtrage multi-cibles (F. Caron, P. Del Moral, A. Doucet, et M. Pace). Ce sujet en cours de développement est aussi au coeur d'un contrat entre l'équipe ALEA et la DCNS de Toulon, et il est au centre d'un projet ANR déposé cette année.

Un dernier point fort de l'équipe concerne la mise en lumière d'une nouvelle classe d'algorithmes stochastique de type recuits simulés ou algorithmes de Gibbs en interaction permettant de résoudre des problèmes d'optimisation globale sous contraintes (P. Del Moral, et F. Giraud, stagiaire ENS Cachan, ce sujet est aussi au coeur d'un projet de collaboration avec P. Minvielle du CEA CESTA). L'étude des applications de cette nouvelle classe de modèles stochastiques à l'analyse combinatoire, et plus particulièrement au calcul de cardinaux d'espaces de solutions complexes est en cours de développement (P. Del Moral, A. Doucet, et A. Jasra).

Dans un autre registre, l'équipe est aussi investie dans divers projets de recherches biomédicales.

L'objectif du projet HEVEA (Handicap : Etude et Valorisation de l'Ecologie Auditive) est de contribuer au réglage des implants cochléaires (prothèses auditives) afin d'optimiser la communication des personnes présentant des déficiences auditives. Ce projet est le produit d'une collaboration entre les organismes suivants : IMB Bordeaux2, Innotech, Université du Littoral, ex-équipe COMPLEX Inria Rocquencourt, Hôpital Avicenne, équipe ARP-UFRL Université Paris 7 [P. Legrand].

L'écologie auditive consiste en des méthodes d'auto-adaptativité des prothèses auditives. Le projet HEVEA se décompose en quatre sous-projets, "Echantillonnage des scènes sonores", "caractérisation des scènes sonores", "Paramétrage en fonction des scènes sonores", "Sélection automatique de paramétrage". Pour mener à bien ces quatre sous-projets, des méthodes ondelette et des analyses de régularité ont été réalisées pour aboutir à une classification des environnements sonores du patient. D'autre part, la sélection automatique de paramétrage se fait grâce à un algorithme évolutionnaire interactif. Au cours de ce projet le premier algorithme évolutionnaire fonctionnant sur pocket PC a été développé et ce logiciel est régulièrement actualisé en fonction des nouvelles versions d'OS sur pocket PC [P. Legrand].

Un nouveau projet de recherche concerne la rétroaction stimuli-activité cérébrale [P. Legrand]. L'objectif de ce projet est de 1) Créer une base de données de signaux EEG obtenus par acquisition sur des sujets dans un état psychophysiologique donné. 2) Développer une méthode de classification de signaux EEG selon l'état psychophysiologique du sujet. Définition d'une distance sur les classes obtenues. 3) Mettre en oeuvre un système d'écoute avec génération paramétrée de musique. 4) Développer un prototype permettant une rétroaction entre la musique générée et l'état psychophysiologique d'un sujet afin de l'amener vers un état cible.

6.2.3 Matrices aléatoires

Une toute nouvelle thématique concerne le comportement asymptotique du spectre de matrices aléatoires. C'est une thématique très récente de l'équipe puisqu'elle est représentée par *B. Bercu (PR, Bx1 et équipe INRIA CQFD)*, et *D. Féral (MCF, Bx1) recrutés respectivement en 2006 et 2008*.

Un premier axe de recherche porte sur les modèles dits Gaz de Coulomb discrets et continus. Les gaz de Coulomb discrets sont les analogues discrets des modèles plus classiques de gaz de Coulomb continus rencontrés en théorie des matrices aléatoires. Des résultats de grandes déviations pour la mesure spectrale et la λ -valeur propre maximale^a associées ont été obtenus par D. Féral. Un second axe de recherche est consacré au comportement asymptotique du spectre de matrices aléatoires structurées comme les matrices de Toeplitz, de Hankel, ou de Jacobi. Le comportement asymptotique de la mesure spectrale de matrices aléatoires circulantes a récemment été établi par B. Bercu et W. Bryc.

Un troisième axe de recherche concerne le comportement asymptotique des valeurs propres extrémales de certaines matrices aléatoires. Les modèles d'étude sont des déformations de modèles classiques et bien connus, à savoir les matrices de Wigner et les matrices de Wishart. Il s'agit en particulier d'étudier les effets de la perturbation sur la loi de la plus grande valeur propre pour des matrices aléatoires de grande dimension. A l'aide d'arguments combinatoires, des résultats d'universalité des fluctuations ont ainsi été établis par D. Féral et S. Péché pour certaines matrices de Wigner déformées puis pour des matrices de Wishart non blanches. Une autre approche, basée sur les transformées de Stieltjes, a permis d'étudier des classes assez générales de matrices de Wigner déformées et de mettre en évidence un phénomène remarquable de non-universalité des fluctuations des valeurs propres extrémales, voir D. Féral en collaboration avec M. Capitaine et C. Donati-Martin.

Cette thématique émergente offre de nouvelles perspectives de collaboration et d'extension avec les équipes d'Analyse et d'EDP de l'IMB, en particulier avec J.F. Bony, V. Bruneau, E.M. Ouhabaz et S. Kupin recruté cette année à l'IMB.

7 Projets de recherches

7.1 Modèles et méthodes statistiques

Sur le plan des méthodes de classification stochastique, les projets de recherches de l'équipe concernent le développement mathématique et informatique de nouvelles techniques algorithmiques d'apprentissage et de classification bayésienne. Ces travaux en cours de développement sont au coeur d'un contrat avec l'EDF pour la classification et la prévision de courbes de charges. En ce qui concerne les méthodes de classification statistique, nous développons actuellement un algorithme de partitionnement utilisant les quantiles multivariés (dits géométriques). Des problèmes théoriques restent encore non résolus pour la classification de variables qualitatives (monotonie de l'indice hiérarchique par exemple) ainsi que pour la rotation en ACM (solution analytique en dimension quelconque par exemple).

Un sujet de recherche sur lequel l'équipe a déjà établi des contacts avec d'autres laboratoires universitaires (Univ. Montpellier 2, Agrocampus Rennes, Univ. Rennes 2) ainsi

que DANONE Research, porte sur le développement de méthodes d'analyse de cubes de données, en particulier en classification et en traitement des données manquantes.

Enfin pour conclure sur les méthodes de classification, des projets de collaborations des approches stochastiques et statistiques portent le traitement de données d'intervalles et l'extraction non supervisée et incrémentale de classes sur un flot de données (encore appelé *data stream clustering*).

L'intérêt des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les estimations. L'idée est d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calcul peut être très intéressant et les applications d'une telle approche sont nombreuses. En nous focalisant sur la méthode SIR, nous allons proposer un estimateur récursif de la partie paramétrique du modèle semiapramétrique sous-jacent, et nous étudierons les propriétés asymptotiques de cet estimateur qui sera implémenté afin de valider son comportement numérique.

Enfin, il est important de souligner que de récents contacts académiques avec le département de génie mécanique et productique de l'université Bordeaux 1, le Cemagref de Bordeaux-Cestas, l'Institut de la Vigne et du Vin ouvrent de très bonnes perspectives de recherches sur le thème de l'analyse de survie et la fiabilité. Le récent projet ANR Safe Tee déposé dans la thématique Véhicule de Transport Terrestre associe l'IMB dans le contexte de la démonstration de fiabilité et les modèles de vies accélérées (avec ALSTOM TIS, le LAAS, l'ENIT, entre autres).

7.2 Modèles et méthodes stochastiques

Sur le thème des algorithmes stochastiques, les projets de recherches de l'équipe portent sur le développement mathématique et informatique de nouvelles classes de modèles stochastiques pour *l'optimisation combinatoire, la poursuite multi-cible, la classification non supervisée, la recherche de temps d'arrêt optimaux et les approximations d'enveloppes de Snell*. Ces projets à court terme s'inscrivent le plus souvent comme la suite naturelle des nos récentes études décrites dans les points forts de l'équipe dans la thématique des algorithmes stochastiques et les méthodes de simulation. Il est important de souligner que ces projets en cours de développement sont liés à des projets précis avec des partenaires industriels et des organismes de recherches nationaux.

Un premier projet de recherche concerne le filtrage non linéaire et la poursuite multi-cible. Ce projet s'inscrit dans un récent dépôt d'ANR, ANR Propagation (responsable : F. Caron ; sur le pistage radar et optronique passif pour la sauvegarde et la protection d'infrastructure côtière, avec pour partenaires la DCNS, les sociétés THALES, ECOMER, et EXAVISION et l'équipe INRIA ALEA).

Un second projet s'inscrit dans de fructueuses collaborations avec le CEA CESTA sur les aspects optimisation stochastique [P. Del Moral et P. Minvielle].

Un troisième avec l'équipe OSIRIS de l'EDF de Clamart porte sur les aspects numériques des mathématiques financières et le calculs d'enveloppes de Snell [P. Del Moral et N. Oudjane, équipe INRIA ALEA en collaboration avec l'équipe ASPI de l'IRISA de Rennes].

Deux autres projets en cours de développement dans le cadre de deux récentes ANRs (ANR PREVASSEMBLE et ANR VIROSCOPY). Ces projets concernent le développement

des méthodes particulières en mécanique des fluides et en prédiction météo [P. Del Moral, en collaboration avec Ch. Baehr de Météo France Toulouse], et l'analyse de risque et la simulation d'événements rares en épidémiologie [P. Del Moral et A. Doucet et A. Jasra].

Dans un autre registre, d'autres projets en cours portent sur le développement d'approches stochastiques et numériques originales pour l'analyse de dynamiques d'éco-systèmes bactériens, notamment de type alimentaire, dans un but de modélisation, de prévision, de détection, de prévention et de contrôle de risques de prolifération critique (dans le cas de pathogènes). Ces projets s'inscrivent dans le cadre d'une action de recherche collaborative INRIA (ARC INRIA EPS) dirigée par P. Del Moral en collaboration avec J.P. Vila (INRA Montpellier), J. P. Gauchy (INRA Paris), L. Coroller (LUBEM de Quimper), et J.C. Augustin de l'ENV d'Alfort.

Sur le thème du contrôle stochastique, nous souhaitons développer la thématiques contrôle stochastique sur un plan théorique mais aussi dans le cadre d'applications aux systèmes complexes allant du monde industriel jusqu'au monde du vivant. Ceci permettra d'accroître des synergies déjà existantes avec les entreprises : EADS, EDF, la DCN, le CEA et Thales. Il est à noter que la demande d'ANR FauToCoES du programme Arpège (avec l'école polytechnique, le LMP, EADS Astrium, le LATP, l'équipe CQFD et l'équipe MC2) sur le contrôle tolérant aux fautes pour les systèmes embarqués vient d'être acceptée.

Nos objectifs sont d'aborder les thèmes suivants : 1) contrôle optimal de processus en temps discret ou continu : contrôle singulier de processus de diffusion, contrôle ergodique de processus déterministes par morceaux, analyse de stabilité stochastique. 2) les processus markoviens décisionnels : conditions suffisantes d'existence de contrôle optimal, analyse d'algorithmes stochastiques d'itération. 3) la programmation dynamique : analyse de problème de contrôle avec modèle incertain, partiellement observé, avec contrainte. 4) la programmation linéaire en dimension infinie : analyse de problème de contrôle avec contrainte, développement de procédure numérique pour calculer la valeur optimale de la fonction coût. 5) les aspects numériques du contrôle : étude et analyse d'algorithmes d'approximation de stratégies optimales, étude de convergence, vitesse de convergence

En lien avec l'étude des processus markoviens déterministes par morceaux, qui est un thème majeur de notre équipe, nous souhaitons développer des techniques d'estimation des paramètres de ces processus. A notre connaissance ce champs de recherche n'a été que très partiellement étudié. Ce thème de recherche a des liens avec l'étude des modèles de survie pour l'estimation de l'intensité des changements de régime. Mais le caractère déterministe des portions de trajectoire et de certains temps de changements de régime, pose des problèmes très spécifiques. Ces résultats auront des applications immédiates en industrie notamment dans des problèmes de propagation de fissures.

Autour de l'estimation des processus autorégressifs de bifurcation, nous souhaitons d'une part étendre nos résultats existants à une classe de processus autorégressifs de bifurcation d'ordre p , avec $p > 1$. D'autre part, nous souhaitons étudier des problèmes d'estimation lorsque les coefficients d'autorégression sont aléatoires.

Nous souhaitons également poursuivre l'étude des propriétés fines des processus autorégressifs à changement de régime Markovien (moments, queue,...) en vue d'applications en économie et économétrie.

8 Points faibles

Comme en témoigne le compte rendu d'activité décrit précédemment, l'équipe de probabilités et de statistique est en surchauffe permanente. La gestion des contrats industriels, des projets d'ANR ou ARC INRIA, la gestion des équipes INRIA associées, la mise en place de manifestations scientifiques de toute sorte, et son implication forte dans l'enseignement et les diverses charges administratives au sein des universités, mais aussi dans la communauté scientifique nationale et la recherche internationale de haut niveau sont chaque jour de plus en plus difficile à gérer par les membres de l'équipe.

Les universités de Bordeaux ont permis de mettre en place une équipe de probabilités et de statistique avec une série de recrutements d'enseignants chercheurs en parfaite adéquation avec le tissu scientifique local. La composition actuelle de l'équipe est la suivante :

- 1 PR Bx 1, 1 PR Bx2 et 1 PR Bx4, 1PR IPB & 1 DR INRIA.
- 2 MCF de Bx1, 7 MCF de Bx2, 1 MCF de Bx4 & 1 CR INRIA.

A travers l'INRIA, nous avons aussi bénéficié de structures administratives supplémentaires (assistante d'équipe projets, services de ressources humaines, de documentation et de relations industrielles performants), de délégations assez régulières (une année pour F. Dufour et pour E. Rio), et d'un poste de chercheur invité (A. Doucet sur une période de trois mois). L'équipe a aussi été appuyée par deux récents recrutements de chercheurs INRIA, effectués en 2007 et 2008 (P. Del Moral et F. Caron). De plus, une offre de poste d'ingénieur de développement INRIA est en cours.

Néanmoins, le fort dynamisme et les synergies mises en place par cette équipe s'accompagnent d'un manque évident de moyens humains pour poursuivre ses projets de recherches émergents, son investissement dans les contrats avec ses partenaires industriels, et dans ses activités d'encadrement de la recherche y compris le travail d'édition dans des journaux scientifiques et l'encadrement des étudiants. Ce manque de moyen humain empêche aussi la mise en place de projets de collaborations naturels entre l'équipe de probabilités et statistique et d'autres équipes INRIA ou universitaires travaillant sur des thèmes connexes, comme *l'analyse fonctionnelle, la combinatoire, la théorie des opérateurs, le traitement du signal et de l'image, la biostatistique ou encore les équipes travaillant sur les dynamiques de population ou sur la biologie.*

Sur ce plan, il y a donc une nécessité absolue de recruter au plus vite du personnel MdC, Pr ou chercheur INRIA ou CNRS. Rester sans recrutement rapide étoufferait la plupart de nos projets de recherches et anéantirait les synergies et les perspectives de collaborations mise en place depuis des cinq dernières années.