

## Liste des sessions

### Groupe 1

1. Équations aux dérivées partielles stochastiques. Laurent Denis (Univ. Évry).
2. Gènes, individus, populations. Amaury Lambert (ENS Ulm).
3. Modèles à données manquantes ou mal observées. Marie-Luce Taupin (Univ. Paris 5).
4. Valeurs extrêmes et applications. Philippe Soulier (Univ. Paris 10).
5. Fiabilité et durée de vie. Christian Paroissin (Univ. Pau).

### Groupe 2

1. Processus autosimilaires. Céline Lacaux (Ecole des Mines de Nancy).
2. Nouvelles directions en mathématiques financières. Peter Tankov (Paris VII).
3. Génétique statistique. Olivier François (IMAG).
4. Statistique spatiale et spatio-temporelle. Cécile Hardouin (Univ. Paris 10).
5. Méthodes régénératives en statistique. Eva Löcherbach (Univ. Paris-Est Créteil).

### Groupe 3

1. Processus de branchement continu et applications. Romain Abraham (Univ. Orléans).
2. Processus autorégressifs. Benoîte de Saporta (Univ. Bordeaux 4).
3. Statistique médicale. Jean-François Dupuy (Univ. Toulouse 3).
4. Selection de modèles. Béatrice Laurent (INSA de Toulouse).
5. Inférence et structure cachée dans les réseaux biologiques. Christophe Ambroise (Genopôle Evry).

### Groupe 4

1. Probabilités non-commutatives et algèbres de Von Neumann de groupes. Dimitri Petritis (IRMAR, Rennes).
2. Simulation de processus de diffusion. Benjamin Jourdain (E.N. Ponts-et-Chaussées).
3. Traitement du signal et statistique. Erwan Le Pennec (Univ. Paris 7).
4. Statistiques bayésiennes : théorie et méthodologie. Judith Rousseau (Univ. Paris 9).
5. Apprentissage séquentiel. Gilles Stoltz (ENS Ulm).

### Groupe 5

1. Méthodes de Monte Carlo adaptatives. François Le Gland (INRIA Rennes), Éric Moulines (ENST Paris).
2. Probabilités et géométrie. Marc Arnaudon (Univ. Poitiers).
3. Machines à vecteurs de support et autres méthodes à noyaux. Fabrice Rossi (INRIA).
4. Tests d'adéquation non paramétriques. Valentin Patilea (INSA Rennes).



## Equations aux Dérivées Partielles Stochastiques

Session organisée par **Laurent Denis**

La théorie des EDPS connaît actuellement un développement important, faisant intervenir des concepts et résultats nouveaux aussi bien en probabilité qu'en analyse.

Les EDPS servent à modéliser de nombreux phénomènes dans des domaines variés tels que : les phénomènes de propagation d'ondes en milieux aléatoires, la dispersion en milieux poreux, la turbulence, les études de populations en biologie, les modèles de taux en finance etc....

En schématisant, on peut dire qu'une EDPS est constituée de deux termes : une équation aux dérivées partielles (terme déterministe) et un bruit aléatoire. Cette session sera l'occasion de présenter les résultats nouveaux pour différents types d'équations.

*Adresse de l'organisateur :*

Laurent DENIS  
Laboratoire Analyse et Probabilités  
Université d'EVRY  
91025 EVRY France  
E-mail : [ldenis@univ-evry.fr](mailto:ldenis@univ-evry.fr)

### Ordre faible pour la discrétisation d'EDP stochastiques

par **Arnaud DEBUSSCHE**

On étudie la discrétisation d'une EDP stochastique de type parabolique écrite sous forme abstraite dans un espace de Hilbert  $H$  :

$$\begin{cases} dX = (AX + f(X))dt + \sigma(X)dW, \\ X(0) = x. \end{cases}$$

Dans le cas de l'équation de la chaleur non linéaire,  $A$  est le Laplacien sur un ouvert borné et muni de conditions aux limites de type Dirichlet, Neumann ou périodiques,  $H$  est l'espace  $L^2$ ,  $f$  est un terme non linéaire et  $dW$  est un terme de bruit de type bruit blanc.

Dans le cas d'une semi discrétisation temporelle, le schéma d'Euler implicite pour cette équation a la forme

$$\begin{cases} X_{k+1} - X_k = \Delta t AX_{k+1} + \Delta t f(X_k) + W((k+1)\Delta t) - W(k\Delta t) \\ X_0 = x. \end{cases}$$

De nombreux travaux ont étudié *l'ordre fort* de ce schéma. Dans le cas classique de l'équation de la chaleur non linéaire en dimension 1 perturbée par un bruit blanc en espace et en temps, ces travaux conduisent à des estimations du type

$$\mathbb{E}(\sup_k |X_k - X(k\Delta t)|_H) \leq c\Delta t^\alpha,$$

pour  $\alpha < 1/4$ . Cet ordre  $1/4$  est raisonnable car il correspond à la régularité temporelle de la solution.

Une autre façon de mesurer la précision d'un schéma est d'étudier son *ordre faible*. On dit que le schéma est d'ordre faible  $\beta$  si pour toute fonction  $\varphi$  suffisamment régulière sur  $H$  on a

$$|\mathbb{E}(\varphi(X_k)) - \mathbb{E}(\varphi(X(k\Delta t)))| \leq c(\varphi)\Delta t^\beta.$$

Dans de nombreuses applications, on ne s'intéresse qu'à l'approximation de telles quantités et c'est donc l'ordre faible qui est important.

Il est connu que dans le cas de la discrétisation d'une équation différentielle stochastique en dimension finie, l'ordre fort du schéma d'Euler est  $1/2$  alors que l'ordre faible est 1. La preuve de ce résultat repose sur l'équation de Kolmogorov associée à l'équation différentielle stochastique.

L'équation de Kolmogorov associée à une EDP stochastique est un objet assez complexe et la méthode se généralise très mal. Après avoir rappelé les résultats en dimension finie, nous montrerons les difficultés qui surviennent en dimension infinie et montrerons comment les surmonter, entre autre grace au calcul de Malliavin. Dans le cas de l'équation de la chaleur non linéaire en dimension un perturbée par un bruit blanc en espace et en temps, on montre que l'ordre faible est le double de l'ordre fort.

*Adresse :*

Arnaud DEBUSSCHE  
ENS Cachan, antenne de Bretagne  
ENS Cachan, Antenne de Bretagne Avenue Robert Schumann  
35170 BRUZ , France  
E-mail : [arnaud.debussche@bretagne.ens-cachan.fr](mailto:arnaud.debussche@bretagne.ens-cachan.fr)  
<<http://www.math.bretagne.ens-cachan.fr/people/arnaud.debussche/>>

## **Temps de sortie et persistance des solitons pour des equations de KdV stochastiques**

par Anne De Bouard et **Eric Gautier**

L'équation de Korteweg de Vries admet des solutions solitons qui sont des ondes progressives localisées se propageant à vitesse et forme constante. Nous considérons le cas de perturbations aléatoires par un bruit additif de petite amplitude. Il est courant en physique d'approcher la solution, correspondant à une donnée initiale générant un soliton en l'absence de bruit, par un soliton modulé aléatoirement (les paramètres du soliton fluctuent aléatoirement). A. de Bouard et A Debussche ont montré la validité d'une telle approximation. Nous faisons ici une étude plus approfondie des temps de sortie de voisinages du soliton et d'un soliton et obtenons leur ordre de grandeur en fonction de l'amplitude du bruit. Ceci permet de quantifier le gain d'une approximation de type soliton modulé dans la description de la persistance des solitons.

*Adresses :*

Anne DE BOUARD  
CMAP  
CMAP, Ecole Polytechnique, Route de Saclay,  
91128 Palaiseau cedex  
E-mail : [debouard@cmappx.polytechnique.fr](mailto:debouard@cmappx.polytechnique.fr)  
<<http://www.cmap.polytechnique.fr/spip.php?article204>>

Eric GAUTIER  
CREST  
ENSAE 3, Avenue Pierre Larousse  
92 245 Malakoff Cedex France  
E-mail : [Eric.Gautier@ensae.fr](mailto:Eric.Gautier@ensae.fr)  
<<http://www.crest.fr/pageperso/eric.gautier/eric.gautier.htm>>

## **Équations aux dérivées partielles stochastiques et processus de Lévy**

par **Eulalia Nualart**

Dans cet exposé nous traitons deux problèmes différents qui ont la particularité d'étudier des résultats liés à des équations aux dérivées partielles stochastiques et des processus de Lévy.

Dans le premier problème nous considérons d'e.d.p.s. paraboliques et hyperboliques linéaires perturbées par un bruit blanc espace-temps, où l'opérateur spatial est le  $L^2$ -générateur d'un processus de Lévy  $X$ . Nous montrons que ces équations ont une solution si et seulement si le processus symétrisé de  $X$  possède des temps locaux. Ce résultat donne un argument probabiliste de la non existence de solutions en dimension strictement plus grande que un. De plus, nous montrons que la solution de l'e.d.p.s est Hölder continue dans sa variable spatiale si et seulement si le temps local mentionné l'est.

Dans le deuxième problème nous considérons un système d'équations des ondes linéaire perturbées par un bruit espace-temps de Lévy. Nous étudierons des propriétés géométriques de l'ensemble espace-temps dans lequel la solution du système visite le niveau zéro.

(Ce travail fut réalisé en collaboration avec D. Khoshnevisan et M. Foondun de l'Université de Utah, EUA).

*Adresse :*

Eulalia NUALART  
Université Paris 13  
Institut Galilée, Université Paris 13,  
93430 Villetaneuse, France.  
E-mail : [eulalia@nualart.es](mailto:eulalia@nualart.es)  
<<http://nualart.es>>

## **Maximum principle for Quasilinear Stochastic PDE's driven by a space-time white noise**

par **Anis Matoussi**

We study the existence and uniqueness of solutions of quasilinear parabolic stochastic PDEs driven by a white noise. Moreover we prove a maximum principle for this solution.

*This is a joint work in progress with Laurent Denis.*

*Adresse :*

Anis MATOUSSI  
Université du Maine  
Département de Mathématiques Équipe Statistiques et Processus Université du Maine Avenue Olivier Messiaen  
72085 Le Mans Cedex 9 France.  
E-mail : [anis.matoussi@univ-lemans.fr](mailto:anis.matoussi@univ-lemans.fr)  
<<http://www.univ-lemans.fr/~amatou/>>



## Gènes, Individus, Populations

Session organisée par **Amaury Lambert**

En biologie des populations, les phénomènes de nature aléatoire se produisent à toutes les échelles : *génétique* (mutations, recombinaisons, réarrangements,...), *écologique* (naissance et mort, prédation, dispersion,...), et *évolutive* (extinction, spéciation).

De plus, des processus stochastiques identiques peuvent servir à modéliser des phénomènes différents : les *processus de fragmentation* et de *branchement* (réplication, division cellulaire, spéciation), les *coalescents* (généalogies de gènes, duels compétitifs, fusions de colonies), les *diffusions* (dynamiques de fréquences alléliques, comportements individuels de dispersion, démographie), etc.

Néanmoins, la complexité et la diversité des systèmes vivants confronte notre communauté à des questions difficiles concernant des objets mathématiques connus, ainsi qu'à l'élaboration d'approches alternatives, voire à la définition de nouveaux objets mathématiques. Cette session regroupe quatre jeunes chercheurs actifs sur ce front (et dont trois sont membres du projet « Modèles Aléatoires de l'Évolution du Vivant » de l'Agence Nationale pour la Recherche).

De manière générale, notre objectif est de faire des prédictions sur des *variables macroscopiques* (compositions alléliques, distributions d'abondances, tailles de population, temps de fixation, taux de spéciation,...), à partir de *modèles individu-centrés* (dits aussi microscopiques), sous des *hypothèses minimales* (échangeabilité, haploïdie, structure spatiale complète,...) qui ont vocation à être relaxées si les patrons observés contredisent nos prédictions.

### Références :

- [1] Bertoin, J. (2006) *Random Fragmentation and Coagulation Processes*. Cambridge U. Press.
- [2] Durrett, R. (2002) *Probability Models for DNA Sequence Evolution*. Springer-Verlag, Berlin.
- [3] Lambert, A. (2008) *Population Dynamics and Random Genealogies*. In *Ninth symposium on probability and stochastic processes held at Guanajuato, Mexico, nov 2006*. Edited by M.E. Caballero, V. Rivero, J. Ruiz de Chávez. À paraître.

### Adresse de l'organisateur :

Amaury LAMBERT  
Laboratoire d'Écologie et Évolution  
CNRS UMR 7625 et Université « Pierre et Marie Curie » Paris 6  
École Normale Supérieure  
46, rue d'Ulm  
75005 Paris  
E-mail : [amaury.lambert@ens.fr](mailto:amaury.lambert@ens.fr)  
<<http://ecologie.snv.jussieu.fr/amaury/>>

### Affiliation Jason :

## Partition allélique et généalogie pour des populations à forte variance du succès reproductif

par **Julien Berestycki**, Nathanaël Berestycki et Jason Schweinsberg

Le coalescent de Kingman est un objet mathématique qui permet de modéliser la façon dont les lignées ancestrales d'individus coalescent lorsque l'on remonte dans le passé. Introduit au début des années 80, il joue depuis un rôle clé en génétique des populations où il est devenu le modèle « standard » de généalogie. L'une des questions qui se pose dans ce domaine est de savoir détecter dans le génome des individus la trace de l'action de la sélection naturelle. Or le coalescent de Kingman permet de calculer le polymorphisme génétique attendu dans une population stable sur laquelle la sélection n'agit pas (il fournit donc l'hypothèse nulle pour des tests statistiques qui cherchent justement à détecter l'action de la sélection). Il apparaît cependant que pour de nombreuses populations (en particulier en

milieu marin) qui présentent une forte variance du succès reproductif, d'autres coalescents sont en fait mieux adaptés pour décrire les généalogies neutres. Nous montrons comment on peut, dans ce cadre, obtenir un analogue asymptotique de la célèbre formule de Ewens qui décrit complètement la diversité attendue.

Dans cet exposé nous introduirons donc quelques uns des modèles les plus utilisés en génétique des populations en nous attachant à souligner les liens profonds qui les relient à l'étude des processus de branchement et de coalescence.

*Adresses :*

Julien BERESTYCKI  
Laboratoire de Probabilités et Modèles Aléatoires  
CNRS UMR 7599 et Université « Pierre et Marie Curie » Paris 6  
175, rue du Chevaleret  
75013 Paris  
E-mail : [jberest@gmail.com](mailto:jberest@gmail.com)

Nathanaël BERESTYCKI  
Statistical Laboratory  
Cambridge University  
DPMMS - 3, Wilberforce Rd.  
Cambridge CB3 0WB  
Royaume-Uni  
E-mail : [nberestycki@gmail.com](mailto:nberestycki@gmail.com)  
<<http://www.statslab.cam.ac.uk/~beresty/>>

Jason SCHWEINSBERG  
University of California at San Diego  
9500 Gilman Drive  
La Jolla CA 92093-0112  
USA  
E-mail : [schwein@math.ucsd.edu](mailto:schwein@math.ucsd.edu)  
<<http://math.ucsd.edu/~jschwein/>>

## Loi de la longueur de l'haplotype à l'issue d'un balayage sélectif

par **Stéphanie Léocard** et Étienne Pardoux

On considère les répercussions d'un balayage sélectif (apparition d'un allèle avantageux qui se fixe rapidement dans la population) sur les gènes partiellement liés au site sous sélection. Ce travail reprend le modèle proposé par Etheridge *et al.* (An approximate sampling formula under genetic hitchhiking, *Ann. Appl. Prob.*, 2006) et par Pfaffelhuber et Studeny (Approximating genealogies for partially linked neutral loci under a selective sweep, *J. Math. Biol.*, 2007). Dans ce cadre, on cherche une loi approchée pour la longueur de l'haplotype.

*Adresses :*

Stéphanie LÉOCARD  
LATP CNRS UMR 6632 et Université de Provence  
CMI, Technopôle Château-Gombert  
39, rue F. Joliot Curie  
13453 Marseille Cedex 13  
E-mail : [leocard@cmi.univ-mrs.fr](mailto:leocard@cmi.univ-mrs.fr)



Étienne PARDOUX  
LATP CNRS UMR 6632 et Université de Provence  
CMI, Technopôle Château-Gombert  
39, rue F. Joliot Curie  
13453 Marseille Cedex 13  
E-mail : [pardoux@cmi.univ-mrs.fr](mailto:pardoux@cmi.univ-mrs.fr)  
<<http://www.latp.univ-mrs.fr/~pardoux>>

## Évolution de la coopération, jeu évolutionnaire et probabilité de fixation en population finie subdivisée

par **Véronique Ladret** et Sabin Lessard

Le modèle de l'île de Wright avec un nombre fini de colonies est considéré dans le cadre d'un jeu linéaire. Une approximation au premier ordre de la probabilité de fixation d'un unique mutant par rapport à l'intensité de sélection est déduite en utilisant la théorie de la coalescence. Comme application, dans le cadre du jeu du prisonnier itéré avec les stratégies 'tit-for-tat' (TFT), qui consiste à coopérer au premier round et à imiter son opposant par la suite, et 'always defect' (AllD), qui consiste à ne jamais coopérer, nous spécifions la condition requise pour que la fixation d'un unique individu coopératif utilisant la stratégie TFT dans une population de AllD, soit favorisée par la sélection et nous étendons la loi du un-tiers de manière à prendre en compte la subdivision de la population. La condition pour l'évolution de la coopération se trouve être moins restrictive que dans une population panmictique.

### *Adresses :*

Véronique LADRET  
Département de mathématiques et de statistique  
Université de Montréal  
C.P. 6128 Succursale Centre-ville  
Montréal, Qué.  
Canada H3C 3J7  
E-mail : [ladretve@dms.umontreal.ca](mailto:ladretve@dms.umontreal.ca)

Sabin LESSARD  
Département de mathématiques et de statistique  
Université de Montréal  
C.P. 6128 Succursale Centre-ville  
Montréal, Qué.  
Canada H3C 3J7  
E-mail : [lessards@dms.umontreal.ca](mailto:lessards@dms.umontreal.ca)

## Théorèmes limites dépendant de la population initiale pour des processus de branchement en environnement aléatoire sous-critiques

par **Vincent Bansaye**

On considère un processus de branchement en environnement aléatoire. Par exemple, chaque jour il fait beau temps ou mauvais temps avec probabilité  $1/2$ . S'il fait beau, chaque fleur se reproduit indépendamment et avec une même loi de reproduction. De même s'il fait mauvais, mais avec une autre loi de reproduction donnant moins d'enfants. Notre processus est sous-critique, c'est-à-dire que les fleurs disparaissent p.s. en temps fini. Est-ce-que planter initialement  $k$  fleurs multiplie par  $k$  la probabilité d'en avoir au moins une en vie au bout d'un temps long? La réponse est oui sans environnement aléatoire, c'est-à-dire pour un processus de Galton-Watson. Sachant qu'il y a au moins une fleur en vie en temps long, y a-t-il une chance pour que les descendances de plusieurs fleurs plantées initialement aient survécu? La réponse est non pour un processus de Galton-Watson. Dans le cas des

environnements aléatoires, on verra que les réponses peuvent différer. On peut alors aussi se poser la question suivante : sachant qu'il y a au moins une fleur en vie en temps long, dois je plutôt l'attribuer au fait qu'il a fait exceptionnellement beau, ou au fait que mes fleurs se sont particulièrement multipliées (avec un temps normal en moyenne) ? Ces différences soulèvent des questions à propos de la limite quasi-stationnaire et du  $Q$ -processus...

*Adresse :*

Vincent BANSAYE

Laboratoire de Probabilités et Modèles Aléatoires

CNRS UMR 7599 et Université « Pierre et Marie Curie » Paris 6

175, rue du Chevaleret

75013 Paris

E-mail : [vbansaye@yahoo.com](mailto:vbansaye@yahoo.com)

<<http://www.proba.jussieu.fr/~vbansaye/>>

## Modèles à données manquantes ou mal observées

Session organisée par Marie-Luce Taupin

Les modèles avec des données manquantes ou partiellement observées sont classiquement considérés dans de nombreux domaines d'application, notamment dans le domaine médical, biostatistique, en économétrie *etc...* Parmi les modèles à données manquantes ou partiellement observées on peut citer les modèles avec données censurées, les modèles avec erreurs sur les variables, le modèle de convolution, la tomographie, les problèmes inverses, les modèles de chaînes de Markov cachées, le filtrage, les modèles de mélange et les modèles à variables latentes. L'objectif de cette session est de présenter des résultats récents dans divers modèles à données manquantes ou partiellement observées, et de mettre en évidence les points communs et les différences de traitement statistique dans ces modèles.

### Références :

- [1] Andersen, P. K. and Borgan, O. and Gill, R. D. and Keiding, N. (1993) *Statistical models based on counting processes*. Springer Series in Statistics.
- [2] Bickel, P. J. and Klaassen, A. J. C. and Ritov, Y. and Wellner, J. A. (1993) *Efficient and adaptive estimation for semiparametric model*. Johns Hopkins Series.
- [3] Carroll, R. J. and Ruppert, D. and Stefanski, L. A. (2006) *Measurement error in nonlinear models. A modern perspective* Monographs on Statistics and Applied Probability. London : Chapman & Hall.
- [4] Fan, J. and Truong, Y. K. (1993) Nonparametric regression with errors in variables. *Ann. Statist.* (21)
- [5] van der Vaart, A. W. *Semiparametric statistics* (2002). Lectures on probability theory and statistics (Saint-Flour, 1999)

### Adresse de l'organisateur :

Marie-Luce TAUPIN  
Laboratoire MAP5, UMR 8145  
Université Paris-René Descartes  
45, rue des Saint-Pères,  
75270 Paris Cedex 06, France  
E-mail : Marie-Luce.Taupin@univ-paris5.fr  
<<http://www.math-info.univ-paris5.fr/~taupin/>>

## Méthodes non paramétriques dans le modèle de convolution

par Fabienne Comte

On considère le modèle de convolution  $Z_i = X_i + \varepsilon_i$  où les  $\varepsilon_i$ ,  $i = 1, \dots, n$  sont des bruits i.i.d. de loi connue. En général, les  $X_i$  sont également i.i.d. et les suites  $(X_i)_{1 \leq i \leq n}$  et  $(\varepsilon_i)_{1 \leq i \leq n}$  sont indépendantes. On observe les  $(Z_i)_{1 \leq i \leq n}$  et on s'intéresse à la loi commune  $f_X$  des  $X_i$ .

Cet exposé se propose de décrire les méthodes non paramétriques (noyaux et projection principalement) mises au point afin d'estimer  $f_X$ , les résultats de vitesse, d'optimalité et d'adaptation obtenus au cours des 20 dernières années, dans des contextes de risques (souvent quadratiques) ponctuels ou intégrés.

Plusieurs pistes récentes de généralisation seront présentées :

- contextes de dépendance dans lesquels les vitesses sont préservées,
- estimation simultanée de la loi du bruit  $\varepsilon$  (observations répétées, observation préalable du bruit seul),
- estimation de fonctionnelles linéaires ou quadratiques de  $f_X$ .

Adresse :

Fabienne COMTE  
Laboratoire MAP5, UMR 8145  
Université Paris-René-Descartes  
45, rue des Saint-Pères  
75270 Paris Cedex 06, France  
E-mail : [fabienne.comte@univ-paris5.fr](mailto:fabienne.comte@univ-paris5.fr)

## Estimation dans un modèle de Cox stratifié partiellement observé

par Jean-François Dupuy

Dans cet exposé, nous nous intéressons au problème de l'estimation d'un modèle de Cox stratifié partiellement observé.

Si  $T^0$  désigne une durée de vie aléatoire,  $C$  une variable aléatoire de censure,  $X$  un  $p$ -vecteur de variables explicatives et  $S$  une variable catégorielle indiquant la strate (sous-population) d'appartenance, l'observation consiste, pour chaque individu, en un quadruplet  $(T, \Delta, X, S)$ , où  $T = \min(T^0, C)$  est la durée observée et  $\Delta = 1(T^0 \leq C)$  est l'indicatrice de censure. Le modèle de Cox stratifié suppose que l'intensité du processus de comptage  $N(t) = 1(T \leq t, \Delta = 1)$  est de la forme  $1(T \geq t)\lambda_{0l}(t) \exp(\beta_0'X)$  si  $S = l$ , où  $\beta_0 \in \mathbb{R}^p$  est le paramètre de régression d'intérêt et  $\{\lambda_{0l}(t) : t \geq 0, l = 1, \dots, L\}$  sont des fonctions positives inconnues considérées comme des paramètres de nuisance.

Nous considérons le problème de l'estimation de  $\beta_0$  à partir d'un échantillon de  $n$  individus indépendants lorsque la valeur de  $S$  n'est pas observée pour certains de ces  $n$  individus. Nous proposons un estimateur de  $\beta_0$  dans ce contexte de données manquantes, et nous étudions ses propriétés asymptotiques à l'aide d'outils issus de la théorie des Z-estimateurs.

Adresse :

Jean-François DUPUY  
Institut de Mathématiques de Toulouse  
Laboratoire de Statistique et Probabilités  
Université Paul Sabatier, 118 route de Narbonne  
31062 Toulouse cedex 4, France  
E-mail : [dupuy@math.ups-tlse.fr](mailto:dupuy@math.ups-tlse.fr)  
<<http://www.lsp.ups-tlse.fr/Fp/Dupuy/>>

## Estimation de paramètres d'équation différentielles stochastiques partiellement observées

par Benjamin Favetto, Valentine Genon-Catalot, Yves Rozenholc et Adeline Samson

Dans le cadre de la lutte contre le cancer, on cherche à estimer "in vivo" les paramètres de microcirculation des tissus (cancéreux ou non) à l'aide d'un examen de type CT-scan avec injection d'un agent de contraste. Nous disposons d'observations bruitées de la micro-circulation que l'on modélise comme suit

$$y_i = (1 \quad 0) U(t_i) + \sigma \varepsilon_i \text{ avec } \varepsilon_i \sim \mathcal{N}(0, 1), i = 0, \dots, n$$

où  $U_t$  est un processus de  $\mathbb{R}^2$  défini par le système bi-dimensionnel d'équations différentielles stochastiques du type Ornstein Uhlenbeck suivant :

$$\begin{aligned} dU_t &= (A(\theta)U_t + F(\theta)B(t))dt + \Sigma dW_t \\ U(t_0) &= 0 \end{aligned}$$

Dans notre modèle,  $A(\theta)$  et  $F(\theta)$  sont des matrices  $(2 \times 2)$  inconnues dépendant du vecteur de paramètres  $\theta$ . Chaque composante du vecteur  $\theta$  a une interprétation biologique.  $B(t)$  est un vecteur

connu dépendant du temps représentant la cinétique de l'agent de contraste dans l'artère. Enfin,  $\Sigma$  est la matrice de volatilité de dimension  $2 \times 2$  et  $W_t$  est un mouvement brownien bidimensionnel. Le problème statistique revient à estimer les paramètres  $\theta, \Sigma$  et  $\sigma$  par maximum de vraisemblance.

Nous comparons sur nos données plusieurs approches. La première consiste en une maximisation directe de la vraisemblance des données observées à l'aide d'un recuit simulé. Le calcul de la vraisemblance peut être réalisé : 1/ par un calcul direct,  $y_{0:n} = (y_0, \dots, y_n)$  étant un vecteur gaussien, 2/ par un calcul récursif utilisant la reconstruction du vecteur  $U_{t_i}|y_{0:i}$  par le filtre de Kalman. La deuxième approche repose sur l'utilisation d'un algorithme EM : elle consiste à calculer l'espérance conditionnelle de la vraisemblance des données complètes :

$$Q(\theta, \theta') = E(\log p(y_{0:n}, U; \theta) | y_{0:n}, \theta')$$

en utilisant la loi du vecteur  $U_{t_i}|y_{0:n}$  obtenue par lissage.

*Adresses :*

Benjamin FAVETTO

Laboratoire MAP5, UMR 8145

Université Paris-René Descartes

45, rue des Saint-Pères,

75270 Paris Cedex 06, France

E-mail : [benjamin.favetto@math-info.univ-paris5.fr](mailto:benjamin.favetto@math-info.univ-paris5.fr)

<<http://www.math-info.univ-paris5.fr/map5/>>

Valentine GENON-CATALOT

Laboratoire MAP5, UMR 8145

Université Paris-René Descartes

45, rue des Saint-Pères,

75270 Paris Cedex 06, France

E-mail : [valentine.genon-catalot@math-info.univ-paris5.fr](mailto:valentine.genon-catalot@math-info.univ-paris5.fr)

<[http://www.math-info.univ-paris5.fr/\\_genon/](http://www.math-info.univ-paris5.fr/_genon/)>

Yves ROZENHOLC

Laboratoire MAP5, UMR 8145

Université Paris-René Descartes

45, rue des Saint-Pères,

75270 Paris Cedex 06, France

E-mail : [yves.rozenholc@univ-paris5.fr](mailto:yves.rozenholc@univ-paris5.fr)

<[http://www.math-info.univ-paris5.fr/\\_rozen/](http://www.math-info.univ-paris5.fr/_rozen/)>

Adeline SAMSON

Laboratoire MAP5, UMR 8145

Université Paris-René Descartes

45, rue des Saint-Pères

75270 Paris Cedex 06, France

E-mail : [adeline.samson@univ-paris5.fr](mailto:adeline.samson@univ-paris5.fr)

<<http://adeline.e-samson.org/en/>>

## Mélange de modèles pour le taux de risque conditionnel sous censure

par Stéphane Gaïffas, **Agathe Guilloux** et Vivian Viallon

Soient  $Y$  une v.a. positive et  $X$  un vecteur de covariables dans  $\mathbb{R}^d$ . On voudrait estimer le taux de risque conditionnel de  $Y$  sachant  $X = x$  défini, pour tout  $z > 0$ , par :

$$\alpha_{Y|X}(y, x) = \frac{f_{Y|X}(y, x)}{1 - F_{Y|X}(y, x)},$$

où  $f_{Y|X}$  et  $F_{Y|X}$  sont la densité et la fonction de répartition conditionnelle de  $Y$  sachant  $X$ .

On suppose que  $Y$  peut être censurée aléatoirement par la v.a.  $C$  avec l'hypothèse que  $Y$  et  $C$  sont indépendants conditionnellement à  $X$ . Les observations ont donc la forme :  $(X_i, Z_i = \min(Y_i, C_i), \delta_i = \mathbb{I}(Y_i \leq C_i))$ , pour  $i = 1, \dots, n$ .

Nous considérons les modèles semi-paramétriques suivants : le modèle de Cox  $\alpha_{Y|X}(z, x) = \alpha_0(z)r(<\theta, x >)$ , le modèle des risques additifs d'Aalen  $\alpha_{Y|X}(z, x) = \alpha_0(z) + (<\theta, x >)$  et le modèle single-index  $\alpha_{Y|X}(z, x) = \alpha_0(z, <\theta, x >)$ . On peut également estimer  $\alpha_{Y|X}(y, x)$  de façon entièrement non-para-métrique, c'est-à-dire comme une fonction de  $\mathbb{R}_+ \times \mathbb{R}^d$  dans  $\mathbb{R}_+$ .

En pratique on ne sait pas forcément quel modèle explique bien les données. Une stratégie consiste alors à mélanger les estimateurs associés à chaque modèle et à définir un estimateur agrégé. On calcule une famille  $A(\Lambda) := \{\bar{\alpha}_\lambda ; \lambda \in \Lambda\}$  d'estimateurs, où  $\Lambda =$  ensemble de paramètres des estimateurs. On calcule l'estimateur agrégé :

$$\hat{\alpha} := \sum_{\lambda \in \Lambda} \theta(\bar{\alpha}_\lambda) \bar{\alpha}_\lambda.$$

Le paramètre  $\theta(\bar{\alpha}_\lambda)$  est inversement lié à l'erreur d'estimation empirique de  $\bar{\alpha}_\lambda$ . Pour l'estimateur agrégé, on obtient un résultat du type :

$$E\|\hat{\alpha} - \alpha_{Y|X}\|^2 \leq \inf_{\lambda \in \Lambda} E\|\bar{\alpha}_\lambda - \alpha_{Y|X}\|^2 + \frac{\log |\Lambda|}{Tn}.$$

*Adresses :*

Stéphane GAÏFFAS

Laboratoire de Statistique Théorique et Appliquée (LSTA)

Université Pierre et Marie Curie Paris 6

175, rue du Chevaleret

75013 PARIS France

E-mail : [gaiffas at ccr.jussieu.fr](mailto:gaiffas@ccr.jussieu.fr)

<<http://www.lsta.upmc.fr/gaiffas.php>>

Agathe GUILLOUX

Laboratoire de Statistique Théorique et Appliquée (LSTA)

Université Pierre et Marie Curie Paris 6

175, rue du Chevaleret

75013 PARIS France

E-mail : [aguillou@ccr.jussieu.fr](mailto:aguillou@ccr.jussieu.fr)

<<http://www.lsta.upmc.fr/guilloux.html>>

Vivian VIALLOUX

Laboratoire de Biostatistique Hôpital Cochin

Université Paris Descartes

27, rue du Faubourg Saint Jacques

75014 Paris, France

E-mail : [vivian.viallon@univ-paris5.fr](mailto:vivian.viallon@univ-paris5.fr)

<>

## Théorie des valeurs extrêmes et applications

Session organisée par **Philippe Soulier**

L'objet de cette session est de présenter des développements et des applications récents de la théorie des valeurs extrêmes et de la théorie voisine des franchissements de niveaux de processus stochastiques. Ces domaines ont été particulièrement actifs ces dernières années, comme en témoigne par exemple les récentes monographies [1, 2, 3]. Les exposés de cette session présenteront notamment des résultats nouveaux sur les franchissements de niveaux, sur l'estimation de l'indice des valeurs extrêmes, en théorie des extrêmes multivariés et un travail d'application en climatologie qui introduit des modèles de processus extrémaux.

### Références :

- [1] *Robert J. Adler et Jonathan E. Taylor. Random fields and geometry. Springer, New York. (2007)*
- [2] *Laurens de Haan et Ana Ferreira. Extreme value theory. An introduction. Springer, New York. (2006)*
- [3] *Sydney I. Resnick. Heavy-tail phenomena. Probabilistic and statistical modeling. Springer, New York. (2007)*

### Adresse de l'organisateur :

Philippe SOULIER  
Equipe MODAL'X  
Université Paris-Ouest Nanterre  
92 000 Nanterre France  
E-mail : [philippe.soulier@u-paris10.fr](mailto:philippe.soulier@u-paris10.fr)  
<<http://www.u-paris10.fr/~modalx/>>

## Franchissement de courbe de niveau, formules de Rice et extremum

par **Marie Kratz**

L'étude des extrema est fortement liée à celle des ensembles de niveau, l'approche par les ensembles de niveau permettant d'obtenir des résultats sur les comportements d'extrema de processus ou champs aléatoires. Le domaine des valeurs extrêmes constitue l'une des applications possibles parmi de nombreuses autres (géométrie stochastique, physique, optique, océanographie...) et la littérature sur les ensembles de niveau s'est sensiblement développée depuis le milieu du siècle dernier, en témoignent, suite aux travaux pionniers de Rice, les livres de Cramér et Leadbetter (1967), Adler (1981), Wschebor (1985) et plus récemment Adler et Taylor (2007) [1] et Azaïs et Wschebor (à venir), pour ne citer que les ouvrages parus (ou à paraître) sur le sujet. C'est dans ce cadre d'ensembles de niveau que nous nous placerons, et plus spécifiquement en dimension 1, en s'intéressant au nombre de franchissements d'une courbe de niveau par un processus Gaussien stationnaire. Nous étudierons son comportement asymptotique selon la nature de la courbe de niveau (Kratz & León, soumis). Il s'agit de la généralisation d'un résultat que nous avons obtenu pour un niveau fixé, rendue possible grâce à l'utilisation des formules généralisées de Rice, de notre méthode générale donnant la représentation dans le chaos de Itô-Wiener et des théorèmes de limite centrale pour des fonctionnelles de niveau de processus Gaussiens (Kratz et León, 2001), ainsi que d'outils techniques développés récemment. Nous terminerons la discussion avec des applications, en particulier en référence aux extrema.

### Adresse :

Marie KRATZ  
ESSEC et MAP5, Université René Descartes, Paris  
Université Paris-Ouest Nanterre  
92 000 Nanterre France  
E-mail : [kratz@essec.fr](mailto:kratz@essec.fr)

## **An Auto-Regressive model for maxima : Application to atmospheric chemistry**

par Armelle Guillou, Philippe Naveau, **Gwladys Toulemonde** et Mathieu Vrac

With the increasing number of extreme value theory applications particularly in meteorological problems, it is important to take into account the complexity of certain non independent and temporal observations in statistical models.

In time series analysis, Markov Auto Regressive (AR) processes are convenient because they consider the dependence between the observed value at time  $t$  and values at earlier neighboring time points. By this way, AR processes describe persistence among observations. However, there is no linear AR model dealing with the generalized extreme value distribution.

In this talk we propose an Auto Regressive Markov model of order 1 adapted to the Gumbel distribution. We derive theoretical properties of such a process and we give general properties in the multivariate case. Then, we fit the model on daily and weekly maxima of methane (CH<sub>4</sub>) and carbon dioxide (CO<sub>2</sub>) in Gif-sur-Yvette (France). Finally, simulation results are presented in order to assess the quality of the parameter estimations and of our model on finite samples.

### *Adresses :*

Armelle GUILLOU

Université de Strasbourg I

Université Paris-Ouest Nanterre

92 000 Nanterre France

E-mail : [guillou@math.u-strasbg.fr](mailto:guillou@math.u-strasbg.fr)

Philippe NAVEAU

Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL/CNRS

E-mail : [philippe.naveau@lsce.ipsl.fr](mailto:philippe.naveau@lsce.ipsl.fr)

Gwladys TOULEMONDE

Université Paris VI

E-mail : [gwladys.toulemonde@upmc.fr](mailto:gwladys.toulemonde@upmc.fr)

Mathieu VRAC

Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL/CNRS

E-mail : [mathieu.vrac@lsce.ipsl.fr](mailto:mathieu.vrac@lsce.ipsl.fr)

## **Estimation de l'indice des valeurs extrêmes par la méthode des moindres carrés pondérés**

par **Cécile Mercadier**

L'estimation de l'indice des valeurs extrêmes (noté  $\gamma$ ) est un des problèmes majeurs en Statistique des extrêmes univariés. Pour cette étude, nous nous plaçons dans le domaine d'attraction de la loi de Fréchet ( $\gamma > 0$ ). Des estimateurs de  $\gamma$  (contraints ou pas) construits par la méthode des moindres carrés pondérés sont présentés et utilisés pour l'estimation des quantiles extrêmes.

La « qualité » de chacune de ces procédures dépend fortement du nombre de statistiques d'ordre utilisées. Comme le seuil optimal (défini pour minimiser l'erreur en moyenne quadratique asymptotique) dépend d'un paramètre  $\rho$  dit de second ordre, l'estimation de  $\rho$  est également présentée dans cet exposé.



*Adresse :*

Cécile MERCADIER  
Institut Camille Jordan, Université Claude Bernard Lyon 1  
Université Paris-Ouest Nanterre  
92 000 Nanterre France  
E-mail : [mercadier@math.univ-lyon1.fr](mailto:mercadier@math.univ-lyon1.fr)

## **Modélisation et estimation d'excès conditionnels**

par **Anne-Laure Fougères** et Philippe Soulier

The aim of this talk is the estimation of the conditional limit distribution of a random variable  $Y$  given that another random variable  $X$  exceeds an threshold that tends to infinity. We focus on the framework proposed by Balkema and Embrechts (2007), who assume that the level lines of the density of the pair  $(X, Y)$  are asymptotically locally elliptical. We propose estimators of the normalizing sequences and of the limiting distribution.

*Adresses :*

Anne-Laure FOUGÈRES  
Equipe MODAL'X  
Université Paris-Ouest Nanterre  
92 000 Nanterre France  
E-mail : [Anne-laure.fougeres@u-paris10.fr](mailto:Anne-laure.fougeres@u-paris10.fr)  
<<http://www.u-paris10.fr/~modalx/>>

Philippe SOULIER  
Equipe MODAL'X  
Université Paris-Ouest Nanterre  
92 000 Nanterre France  
E-mail : [philippe.soulier@u-paris10.fr](mailto:philippe.soulier@u-paris10.fr)  
<<http://www.u-paris10.fr/~modalx/>>



## Fiabilité et durée de vie

Session organisée par Laurent Bordes et Christian Paroissin

Les phénomènes faisant intervenir des durées sont présents dans de nombreuses applications comme l'ingénierie (durée de fonctionnement d'un matériel), le biomédical (durée de vie d'un patient) ou encore la finance et l'assurance (temps de cotisation d'un client pour un cabinet d'assurance), par exemple. Ces phénomènes de natures différentes ont en commun d'être des durées aléatoires gouvernées par des lois.

L'estimation de ces lois et de la manière dont elles sont influencées par l'environnement, à partir de données statistiques (durées observées et variables d'environnement), constitue une question fondamentale. On est souvent alors confronté à des problèmes avec données manquantes, incomplètes ou en faible nombre, dans des modèles semi-paramétriques ou non-paramétriques.

Un second problème concerne l'étude de durées liées à des phénomènes de dégradation, celle-ci étant, en général, une quantité mesurée dans le temps considérée comme aléatoire. La durée associée à cette dégradation peut être définie par le temps qu'il faut pour que la dégradation atteigne un seuil donné. L'observation statistique est alors un historique de la dégradation auquel peuvent venir s'ajouter des variables d'environnement.

La modélisation et l'analyse des durées de vies n'est pas un thème nouveau des probabilités et de la statistique. En revanche c'est un thème en constante évolution. Cette évolution tient au fait que les données statistiques prises en compte dans les modèles sont de plus en plus complexes et les modèles eux-mêmes sont de plus en plus généraux.

Les quatre exposés de cette session traitent de ces deux problématiques : la première dans les exposés par J.-Y. Dauxois (exposé long) et par V. Patilea et la seconde dans les exposés par A. Barros & A. Grall et par L. Doyen. Cette session propose donc un panorama diversifié des recherches en cours sur cette thématique.

### *Adresses des organisateurs :*

Laurent BORDES  
Laboratoire de Mathématiques et de leurs Applications  
Université de Pau et des Pays de l'Adour  
64 013 Pau cedex France  
E-mail : lbordes@univ-pau.fr

Christian PAROISSIN  
Laboratoire de Mathématiques et de leurs Applications  
Université de Pau et des Pays de l'Adour  
64 013 Pau cedex France  
E-mail : cparoiss@univ-pau.fr  
<<http://web.univ-pau.fr/~cparoiss>>

## Evénements récurrents, risques concurrents et causes manquantes

par **Jean-Yves Dauxois**

En statistique des durées de vie, les modèles à risques concurrents permettent de prendre en compte les différentes causes de décès ou de panne et de comparer leurs incidences. Nous présentons dans cet exposé des développements récents obtenus dans ce domaine. Nous considérons en particulier le cas des événements récurrents où un modèle de risques concurrents est introduit afin de décrire l'occurrence d'infections nosocomiales sur des patients hospitalisés dans un service de réanimation en France. Une inférence non paramétrique, estimations et tests, est menée dans ce cadre. Ce travail possède naturellement des applications en fiabilité. Nous présentons également dans cet exposé un problème

d'estimation semi-paramétrique (modèle de risques additifs de Aalen) pour risques concurrents quand la cause de mort ou de panne est parfois manquante. Les travaux présentés lors de cet exposé sont le fruit de collaborations différentes avec S. Sencey d'une part et de L. Bordes et P. Joly d'autre part.

*Adresse :*

Jean-Yves DAUXOIS  
Laboratoire de Mathématiques  
Université de Franche-Comté  
16 route de Gray  
25030 Besançon cedex, France  
E-mail : [jean-yves.dauxois@univ-fcomte.fr](mailto:jean-yves.dauxois@univ-fcomte.fr)

## Estimation des paramètres à partir des moments conditionnels en présence de censure

par **Valentin Patilea**

Consider independent copies of a random vector  $Z = (Y, X)'$ . We aim at estimating a parameter  $\theta$  defined through the conditional moment restrictions  $\mathbb{E}[g(Z; \theta_0 | X)] = 0$  a.s. In the framework under study, the random variable  $Y$  is randomly right-censored. Nonlinear regressions under random censoring represent a particular case of the present setup. We propose a new estimation method for models defined by conditional moment restrictions under random censorship that optimizes a criterion based on smoothing techniques. Depending on whether the smoothing parameter is fixed or decreases to zero with the sample size, our approach defines a whole class of estimators. We develop a theory for estimation and inference that focuses on robustness with respect to the bandwidth parameter. Specifically we establish root-n-consistency and asymptotic normality of our estimator as a process depending on the bandwidth within a wide range including fixed bandwidths. We show that this result extends to misspecified models. Our new method apply to non-smooth problems (like quantile regressions) and is simple to implement.

*Adresse :*

Valentin PATILEA  
Centre des Mathématiques - INSA  
Institut National des Sciences Appliquées de Rennes  
20 avenue des Buttes de Coësmes  
CS 14315  
35043 Rennes cedex, France  
E-mail : [valentin.patilea@insa-rennes.fr](mailto:valentin.patilea@insa-rennes.fr)

## Modélisation de propagation de fissures (corrosion sous contrainte) par processus stochastiques

par **Anne Barros** et Antoine Grall

Les phénomènes de corrosion sous contrainte constatés sur différents composants des centrales nucléaires ont été étudiés depuis plusieurs années et les enjeux liés à leur modélisation sont particulièrement importants du point de vue de la sécurité et de l'estimation des durées de vie. Dans cette présentation, nous proposons une approche de modélisation par processus stochastique de phénomènes de fissuration présentant une phase d'incubation suivie d'une phase de propagation. La conséquence principale de l'existence de la phase d'incubation est la non connaissance de l'instant d'apparition des différents défauts. Une fissure est détectée et mesurée pour la première fois lorsque sa taille dépasse un seuil critique caractéristique du système de mesure. En cohérence avec les avis des experts EDF, le cas d'étude considéré a orienté le choix de modélisation vers un processus de Poisson homogène

pour décrire l'occurrence des instants d'initiation des défauts (et donc leur période d'incubation) puis vers un processus Gamma pour la phase de propagation. L'objectif principal de la présentation est de s'intéresser à l'estimation des paramètres du processus composé, sur la base de données de retour d'expérience réalistes. Après avoir déterminé la densité du processus composé qui décrit les dates des instants de détection, on donnera la fonction de vraisemblance et les résultats numériques obtenus à partir des données de corrosion sous contrainte. On évoquera en conclusion le problème de l'estimation de la durée de vie et de l'évaluation d'une politique de maintenance adaptée.

*Adresses :*

Anne BARROS

Laboratoire de Modélisation et de Sécurité des Systèmes  
Université de Technologie de Troyes  
12 rue Marie Curie  
BP 2060  
10010 Troyes, France  
E-mail : [anne.barros@utt.fr](mailto:anne.barros@utt.fr)

Antoine GRALL

Laboratoire de Modélisation et de Sécurité des Systèmes  
Université de Technologie de Troyes  
12 rue Marie Curie  
BP 2060  
10010 Troyes, France  
E-mail : [antoine.grall@utt.fr](mailto:antoine.grall@utt.fr)

## Evaluation de l'efficacité individuelle des maintenances préventives à l'aide d'un modèle de Brown-Proschan

par **Laurent Doyen**

Pour un système réparable, les hypothèses basiques sur l'effet des réparations sont connues sous le nom de réparation minimale ou As Bad As Old (ABAO) et réparation parfaite ou As Good As New (AGAN). Dans le cas ABAO chaque réparation remet le système dans le même état où il était juste avant la défaillance. Dans le cas AGAN, chaque réparation remet à neuf le système. Bien entendu la réalité est entre ces deux cas extrêmes, la réparation est dite imparfaite.

Un des modèles de réparation imparfaite les plus connus est le modèle de Brown-Proschan [1] (BP) pour lequel le système est parfaitement réparé avec une probabilité  $p$  et réparé de façon minimale avec une probabilité  $(1 - p)$ . Cet effet aléatoire de la réparation peut être représenté par des variables aléatoires  $\{B_i\}_{i \geq 1}$  :  $B_i = 1$  si la  $i$ ème réparation est AGAN, et  $B_i = 0$  si la  $i$ ème réparation est ABAO. Le paramètre  $p$  représente alors l'efficacité des réparations.

Dans la littérature les auteurs ont généralement supposés que l'ensemble des effets des réparations  $B_i$ , était connu. Mais en pratique ce n'est pas le cas. Nous nous intéressons donc à l'évaluation de l'effet des réparations, et plus généralement des maintenances, ainsi que de l'usure du système dans le cas où les variables aléatoires  $\{B_i\}_{i \geq 1}$  ne sont pas connues. A notre connaissance seul trois papiers traitent d'estimation pour des modèles BP avec des effets de réparation non connus. Lim [2] a proposé d'utiliser un algorithme EM (Expectation-Maximization), Lim, Lu and Park [3] ont proposé une méthode basée sur une analyse Bayésienne. Langseth and Lindqvist [4] ont généralisé le modèle BP au cas de maintenances préventives et ont proposé d'utiliser une méthode du maximum de vraisemblance. Tous ces papiers proposent uniquement une estimation du paramètre  $p$  qui représente l'efficacité moyenne de l'ensemble des réparations ou actions de maintenances. Ainsi l'inconvénient de toutes ces méthodes et plus généralement de toutes les méthodes d'estimation de l'efficacité des maintenances (Shin Lim Lie (1996), Jack (1998), Yun Choung (1999), Karminskiy Krivstov (2000), Doyen Gaudoin (2004)), est que l'effet des maintenances n'est jamais évalué de façon individuelle.

Or dans un certain nombre de cas pratiques il est reconnu que les effets des maintenances sont très différents d'une maintenance à l'autre.

Dans cette présentation nous proposons un modèle pour un système soumis à des maintenances préventives (MP) et correctives (MC). Les dates de MP sont supposées connues depuis la mise en route du système et sont déterministes. Les durées de maintenance ne sont pas prises en compte, les dates de MC correspondent aux dates de défaillances et sont donc aléatoires. On suppose de plus que les dates de MC ne sont pas connues depuis la mise en route du système. Le processus des MC est donc censuré à gauche. Les MC sont supposées ABAO, et les MP sont supposées de type BP. Le paramètre  $p$  du modèle BP caractérise donc l'efficacité moyenne des MP. L'efficacité individuelle de la  $i$ ème MP est caractérisée par la probabilité que la  $i$ ème MP soit AGAN sachant les instants de défaillance survenus sur la totalité de la période d'observation du système. Ces efficacités de MP ainsi que les paramètres de la première durée de défaillance (qui caractérisent l'usure intrinsèque du modèle) sont estimés grâce à une méthode du type maximum de vraisemblance. L'ensemble de cette méthode a été appliqué et validé sur un jeu de données réelles issues d'une centrale de production d'électricité EDF.

### **Références :**

- [1] BROWN, M. AND PROSCHAN, F. (1983). Imperfect repair. *Journal of Applied Probability*, 20, 851-859.
- [2] LIM, T.J. (1998). Estimating system reliability with fully masked data under Brown-Prochan imperfect repair model. *Reliability Engineering and System Safety*, 59, 277-289.
- [3] LIM, J.H. AND LU, K.L. AND PARK, D.H. (1998). Bayesian imperfect repair model. *Communications in statistics-Theory and methods*, 27, 965-984.
- [4] LANGSETH, H. AND LINDQVIST, B.H. (2004). A maintenance model for components exposed to several failure mechanisms and imperfect repair. *Mathematical and Statistical Methods in Reliability*, World Scientific Publishing Co.

### *Adresse :*

Laurent DOYEN  
SMS-LJK  
Laboratoire Jean Kuntzmann - Université de Grenoble  
IUT 2, département techniques de commercialisation  
2 place Doyen-Gosse  
38031 Grenoble Cedex, France  
E-mail : laurent.doyen@iut2.upmf-grenoble.fr

## Processus autosimilaires

Session organisée par **Céline Lacaux**

Au cours de ces vingt dernières années, de nombreux modèles autosimilaires ont été introduits afin de modéliser des phénomènes naturels. Le mouvement brownien fractionnaire ([2],[3]) est sans doute le modèle le plus utilisé. Cependant, la propriété d'autosimilarité qu'il vérifie est une propriété globale, ce qui restreint son domaine d'application. Dès lors de nombreux champs presque autosimilaires ont été étudiés. La plupart satisfont la propriété d'autosimilarité locale et asymptotique définie dans [1] et [4]. Dans cette session, différentes notions d'autosimilarité seront définies. D'autre part, dans le but de modéliser des phénomènes naturels, estimer les paramètres des modèles "presque" autosimilaires est essentiel. En particulier, un exemple d'application sur des données réelles sera présenté.

### Références :

- [1] A. Benassi, S. Jaffard and D. Roux, *Gaussian processes and Pseudodifferential Elliptic operators*, Revista Mathematica Iberoamericana (13)1 :19–89
- [2] A.N. Kolmogorov, *Wiensche Spiralen und einige andere interessante Kurven in Hilbertsche Raum.*, (1940), C. R. (Dokl.) Acad. Sci. URSS (26) :115–118.
- [3] B.B. Mandelbrot and J. Van Ness, *Fractional Brownian motion, fractionnal noises and applications*, (1968), Siam Review (10) :422–437.
- [4] R.F. Peltier and J. Lévy Véhel, *Multifractional Brownian motion : definition and preliminary results*, (1996), disponible sur <http://www-syntim.inria.fr/fractales/>

### Adresse de l'organisateur :

Céline LACAUX  
Institut Elie Cartan Nancy  
Nancy-Université, CNRS, INRIA  
Boulevard des Aiguillettes  
B.P. 239  
F-54506 Vandœuvre lès Nancy  
E-mail : [Celine.Lacaux@iecn.u-nancy.fr](mailto:Celine.Lacaux@iecn.u-nancy.fr)  
<<http://www.iecn.u-nancy.fr/~lacaux/>>

## Auto-similarité et browniens fractionnaires

par **Jacques Istas**

Auto-similarité et mouvement brownien fractionnaires (mbf) sont intimement liés. En effet, les mbf sont les seuls processus gaussiens auto-similaires à accroissements stationnaires. Nous rappellerons les principales propriétés des mbf : régularité hölderienne, dimension de Hausdorff, estimation... Ces propriétés s'étendent aux mbf euclidiens. Définir la notion d'auto-similarité dans un cadre non-euclidien, par exemple pour des variétés, est délicat puisque les dilations n'ont pas de généralisations naturelles. Dans cet exposé, nous présenterons des mbf indexés par des variétés et en étudierons les principales propriétés.

### Adresse :

Jacques ISTAS  
Laboratoire Jean Kuntzmann  
Université de Grenoble et CNRS  
F-38041 Grenoble Cedex 9  
E-mail : [Jacques.Istas@imag.fr](mailto:Jacques.Istas@imag.fr)  
<<http://ljk.imag.fr/membres/Jacques.Istas/>>

## Modèles germe-grain, changements d'échelles et autosimilarité de tout ordre

par Anne Estrade, **Hermine Biermé** et Ingemar Kaj

**Mots-clés** : autosimilarité, champs généralisés, Processus de poisson ponctuel, champs fractionnaires, Mouvement brownien fractionnaire.

Nous étudions le comportement asymptotique de modèles germe-grain obtenus par aggrégation de boules lancées aléatoirement, dont le volume moyen tend vers 0 ou vers  $+\infty$ . En supposant que l'intensité du rayon obéit à un comportement en loi de puissance, ces modèles convergent, après renormalisation. En particulier nous pouvons obtenir à la limite des champs gaussiens autosimilaires pour tout ordre d'autosimilarité. Nous décrirons précisément ces champs qui généralisent le champ brownien fractionnaire et préciserons leur fonction de covariance.

This work was supported by ANR grant "mipomodim" 05-BLAN-017.

### Références :

- [1] Biermé, H. and Estrade A. : *Poisson random balls : self-similarity and X-ray images*, Adv. Appl. Prob., 38, 1–20, 2006.
- [2] Biermé, H., Estrade A. and Kaj I. : *About scaling behavior of random balls models*, S4G 6th Int. Conference, published by Union of Czech mathematicians and physicists, 63–68, 2006.
- [3] Dobrushin, R. L. : *Gaussian and their subordinated self-similar random generalized fields*, Ann. Probab., 7(1), 1–28, 1979.
- [4] Kaj, I., Leskelä, L., Norros, I. and Schmidt, V. : *Scaling limits for random fields with long-range dependence*, Ann. Probab. 35, 528–550, 2007.

### Adresses :

Anne ESTRADE

MAP5

Université Paris Descartes,

MAP5,

45, rue des Saints-Pères

75006 PARIS

E-mail : [anne.estrade@univ-paris5.fr](mailto:anne.estrade@univ-paris5.fr)

<<http://www.math-info.univ-paris5.fr/~estradea/>>

Hermine BIERMÉ

MAP5

Université Paris Descartes,

MAP5,

45, rue des Saints-Pères

75006 PARIS

E-mail : [hermine.bierme@math-info.univ-paris5.fr](mailto:hermine.bierme@math-info.univ-paris5.fr)

<<http://www.math-info.univ-paris5.fr/~bierme/>>

Ingemar KAJ

Department of mathematics Uppsala University

P.O. Box 480 S-751 06,

Uppsala, Sweden

E-mail : [ikaj@math.uu.se](mailto:ikaj@math.uu.se)

<<http://www.math.uu.se/~ikaj/>>

## Champs gaussiens autosimilaires par rapport à un groupe à m paramètres

par Béatrice Vedel et Marianne Clausel

Si  $\mathcal{A}$  est un sous groupe de  $GL_n(\mathbb{R})$ , le champ Gaussien  $(X(x; \omega))_{x \in \mathbb{R}}$  est dit  $\mathcal{A}$  autosimilaire si, pour un certain homomorphisme  $\chi$  de  $\mathcal{A}$  dans  $\mathbb{R}_+^*$  :

$$X(Ax, \omega) \stackrel{=}{=} \chi(A)X(x, \omega) \text{ pour toute matrice } A \text{ de } \mathcal{A}.$$



Si  $E_1, \dots, E_m$  sont  $m$  matrices commutant deux à deux, et  $\mathcal{A}$  est le sous-groupe à  $m$ -paramètres suivant de  $Gl_n(\mathbb{R})$

$$\mathcal{A} = \left\{ a_1^{E_1} \dots a_m^{E_m}, (a_1, \dots, a_m) \in (\mathbb{R}_+^*)^m \right\}$$

on construit à l'aide d'une représentation harmonisable, des champs gaussiens à accroissements rectangulaires stationnaires,  $\mathcal{A}$ -autosimilaires. On s'intéressera ensuite à la régularité des trajectoires des champs ainsi définis dans les espaces de Besov anisotropes, naturellement associés à ces processus.

*Références :*

- [1] Ayache, A., Léger, S. and Pontier, M., *Drap Brownien Fractionnaire*, Pot. Anal., 17, 31–43, 2002.
- [2] Biermé, H. and Lacaux, C., *Holder Regularity for Operator Scaling Stable Random Fields* submitted, 2007.
- [3] Biermé, H., Meerschaert, M. and H.P. Scheffler, *Operator Scaling Stable Random Fields*, Stochastic Processes and Applications 117(3), 312–332, 2007.
- [4] Bownik, M., *Atomic and molecular decomposition of anisotropic Besov spaces*, Mathematische Zeitschrift 250, 539–571 (2005).
- [5] Kamont, A., *On the Fractional Anisotropic Wiener Field*, Prob. and Math. Stat. 16(1), 85–98, 1996.
- [6] Y.Xiao, *Sample Path Properties of Anisotropic Gaussian Random Fields*, submitted 2007.

*Adresses :*

Béatrice VEDEL

Laboratoire d'analyse et de Mathématiques Appliquées

Université Paris XII

61 avenue du général de Gaulle

94010 Creteil Cedex

France

E-mail : [beatrice.vedel@univ-paris12.fr](mailto:beatrice.vedel@univ-paris12.fr)

<<http://perso-math.univ-mlv.fr/users/vedel.beatrice/home.html>>

Marianne CLAUSEL

Laboratoire d'analyse et de Mathématiques Appliquées

Université Paris XII

61 avenue du général de Gaulle

94010 Creteil Cedex

France

E-mail : [clausel@univ-paris12.fr](mailto:clausel@univ-paris12.fr)

## Estimation du paramètre de fractalité locale et application aux données de fréquences cardiaques

par Jean-Marc Bardet, Vedel Béatrice et Béatrice Vedel

L'autosimilarité relie les comportements locaux et globaux d'un processus, ce qui peut être un frein à la modélisation. Nous construisons un processus gaussien à accroissements stationnaires présentant une forme d'autosimilarité locale (dirigée par le paramètre de fractalité locale) dans une bande de fréquences données. Ceci permet entre autre d'expliquer que dans de nombreuses applications des estimateurs du paramètre d'autosimilarité ont des valeurs supérieures à 1. Une application aux données de fréquences cardiaques mesurées durant un marathon montrent que ce paramètre de fractalité locale pourrait être un témoin essentiel de l'apparition d'une fatigue du coeur.

*Adresses :*

Jean-Marc BARDET

SAMOS-MATISSE

Université Paris I

90 rue Tolbiac

75013 Paris

E-mail : [Jean-Marc.Bardet@univ-paris1.fr](mailto:Jean-Marc.Bardet@univ-paris1.fr)

<<http://matisse.univ-paris1.fr/bardet/>>

Journées MAS 2008, Rennes

Vedel BÉATRICE

Laboratoire d'analyse et de Mathématiques Appliquées

Université Paris XII

61 avenue du général de Gaulle

94010 Creteil Cedex

France

E-mail : [beatrice.vedel@univ-paris12.fr](mailto:beatrice.vedel@univ-paris12.fr)

<<http://perso-math.univ-mlv.fr/users/vedel.beatrice/home.html>>

Béatrice VEDEL

Laboratoire d'analyse et de Mathématiques Appliquées

Université Paris XII

61 avenue du général de Gaulle

94010 Creteil Cedex

France

E-mail : [beatrice.vedel@univ-paris12.fr](mailto:beatrice.vedel@univ-paris12.fr)

<<http://perso-math.univ-mlv.fr/users/vedel.beatrice/home.html>>

## Nouvelles directions en mathématiques financières

Session organisée par **Peter Tankov**

L'objectif de cette session est de présenter en quatre exposés les nouvelles directions de recherche en mathématiques financières autour du grand thème de risque de liquidité et d'imperfections du marché. On parle du risque de liquidité, lorsqu'un investisseur désirant acheter ou vendre un actif, ne trouve pas immédiatement une contrepartie pour cette transaction. Plus généralement, le marché est illiquide ou imparfait, lorsqu'il y a des contraintes sur les dates ou volumes des transactions, ou lorsque les transactions engendrent un coût non-négligeable pour les investisseurs.

Le travail de A. Alfonsi concerne l'optimisation des stratégies d'exécution des transactions (c'est-à-dire, les dates et volumes de transactions) dans un marché avec risque de liquidité via l'étude des carnets d'ordres. B. Bruder étudie le problème d'investissement optimal en présence d'un délai d'exécution (délai entre le passage d'ordre et la transaction effective, qui peut atteindre plusieurs mois pour les actifs comme fonds de placement). E. Voltchkova analyse l'impact de contraintes sur les dates de trading sur la couverture d'options. Finalement, le travail de L. Campi donne un exemple d'un marché imparfait "par excellence" : le marché d'électricité. Dans ce marché, à cause de la nature non stockable d'électricité, la formation des prix est très différente de ce qui est observé dans les cadres plus traditionnels.

*Adresse de l'organisateur :*

Peter TANKOV

Laboratoire de Probabilités et Modèles Aléatoires

Université Paris Diderot – Paris VII

Case 7012, 2 Place Jussieu, 75251 Paris Cedex 05

E-mail : [tankov@math.jussieu.fr](mailto:tankov@math.jussieu.fr)

<<http://www.math.jussieu.fr/~tankov/>>

### Optimal execution strategies in limit order books with general shape functions

par **Aurelien Alfonsi**, Alexander Schied et Antje Schulz

We consider optimal execution strategies for block market orders placed in a limit order book (LOB). We build on the resilience model proposed by Obizhaeva and Wang (2005) but allow for a general shape of the LOB defined via a given density function. Thus, we can allow for empirically observed LOB shapes and obtain a nonlinear price impact of market orders. We distinguish two possibilities for modeling the resilience of the LOB after a large market order : the exponential recovery of the number of limit orders, i.e., of the volume of the LOB, or the exponential recovery of the bid-ask spread. We consider both of these resilience modes and, in each case, derive explicit optimal execution strategies in discrete time. Applying our results to a block-shaped LOB, we obtain a new closed-form representation for the optimal strategy, which explicitly solves the recursive scheme given in Obizhaeva and Wang (2005). We also provide some evidence for the robustness of optimal strategies with respect to the choice of the shape function and the resilience-type.

*Adresses :*

Aurelien ALFONSI

CERMICS

Ecole Nationale des Ponts et Chaussées

6 et 8 avenue Blaise Pascal

Cité Descartes - Champs sur Marne

77455 Marne la Vallée Cedex 2

E-mail : [alfonsi@cermics.enpc.fr](mailto:alfonsi@cermics.enpc.fr)

<<http://cermics.enpc.fr/~alfonsi/home.html>>

Alexander SCHIED

Department of Mathematics, TU Berlin  
Strasse des 17. Juni 136  
10623 Berlin, Germany  
E-mail : [schied@math.tu-berlin.de](mailto:schied@math.tu-berlin.de)

Antje SCHULZ

Quantitative Products Laboratory  
Alexanderstrasse 5  
10178 Berlin, Germany  
E-mail : [aschulz@math.tu-berlin.de](mailto:aschulz@math.tu-berlin.de)

## Impulse control problem on finite horizon with execution delay.

par **Benjamin Bruder** et Huy en Pham

We consider impulse control problems in finite horizon for diffusions with decision lag and execution delay. The new feature is that our general framework deals with the important case when several consecutive orders may be decided before the effective execution of the first one. This is motivated by financial applications in the trading of illiquid assets such as hedge funds. We show that the value functions for such control problems satisfy a suitable version of dynamic programming principle in finite dimension, which takes into account the past dependence of state process through the pending orders. The corresponding Bellman partial differential equations (PDE) system is derived, and exhibit some peculiarities on the coupled equations, domains and boundary conditions. We prove a unique characterization of the value functions to this nonstandard PDE system by means of viscosity solutions. We then provide an algorithm to find the value functions and the optimal control. This easily implementable algorithm involves backward and forward iterations on the domains and the value functions, which appear in turn as original arguments in the proofs for the boundary conditions and uniqueness results.

*Adresses :*

Benjamin BRUDER

Laboratoire de Probabilit es et Mod es Al atoires et SGAM, Soci t  G n rale  
Universit  Paris Diderot – Paris VII  
Case 7012, 2 Place Jussieu, 75251 Paris Cedex 05  
E-mail : [bruder@math.jussieu.fr](mailto:bruder@math.jussieu.fr)  
<<http://www.proba.jussieu.fr/perso.php?id=29>>

Huy en PHAM

Laboratoire de Probabilit es et Mod es Al atoires et Soci t  G n rale et Institut Universitaire de France  
Universit  Paris Diderot – Paris VII  
Case 7012, 2 Place Jussieu, 75251 Paris Cedex 05  
E-mail : [pham@math.jussieu.fr](mailto:pham@math.jussieu.fr)  
<<http://www.proba.jussieu.fr/pageperso/pham/>>

## A joint model of electricity and combustibles prices.

par **Luciano Campi**, Jean-Michel Marin et Nizar Touzi

We propose a joint model for combustibles, electricity spot prices and demand. Combustibles prices follow a multidimensional diffusion process and the electricity spot price is defined via a production function involving combustibles and electricity demand. An explicit relation between combustibles forward prices and electricity forward price is obtained. Moreover, we provide an explicit expression for electricity forward price under the assumption that the electricity demand follows a Gaussian mean-reverting process. Finally, we compare the features our model to the main properties of electricity market data.

*Adresses :*

Luciano CAMPI  
CEREMADE  
Université Paris Dauphine  
Place du Marechal De Lattre De Tassigny  
75775 Paris Cedex 16  
E-mail : [campi@ceremade.dauphine.fr](mailto:campi@ceremade.dauphine.fr)  
<<http://www.geocities.com/lucianocampi/>>

Jean-Michel MARIN  
INRIA  
Saclay  
E-mail : [Jean-Michel.Marin@inria.fr](mailto:Jean-Michel.Marin@inria.fr)  
<<http://www.math.u-psud.fr/~marin/>>

Nizar TOUZI  
Centre de Mathématiques Appliquées  
Ecole Polytechnique  
91128 Palaiseau Cedex  
E-mail : [touzi@cmap.polytechnique.fr](mailto:touzi@cmap.polytechnique.fr)  
<<http://www.cmap.polytechnique.fr/~touzi/>>

## Asymptotic analysis of hedging errors in models with jumps

par Peter Tankov et **Ekaterina Voltchkova**

Most authors who studied the problem of option hedging in incomplete markets, and, in particular, in models with jumps, focused on finding the strategies that minimize the residual hedging error. However, the resulting strategies are usually unrealistic because they require a continuously rebalanced portfolio, which is impossible to achieve in practice due to transaction costs. In reality, the portfolios are rebalanced discretely, which leads to a 'hedging error of the second type', due to the difference between the optimal portfolio and its discretely rebalanced version. In this paper, we analyze this second hedging error and establish a limit theorem for the renormalized error, when the discretization step tends to zero, in the framework of general Itô processes with jumps. The results are applied to the problem of hedging an option with a discontinuous payoff in a jump-diffusion model.

*Adresses :*

Peter TANKOV  
Laboratoire de Probabilités et Modèles Aléatoires  
Université Paris Diderot – Paris VII  
Case 7012, 2 Place Jussieu, 75251 Paris Cedex 05  
E-mail : [tankov@math.jussieu.fr](mailto:tankov@math.jussieu.fr)  
<<http://www.math.jussieu.fr/~tankov/>>

Ekaterina VOLTCHKOVA  
GREMAQ  
Université Toulouse 1 Sciences Sociales,  
21, allée de Brienne,  
31000 Toulouse  
E-mail : [ekaterina.voltchkova@univ-tlse1.fr](mailto:ekaterina.voltchkova@univ-tlse1.fr)



## Génétique statistique

Session organisée par **Olivier François**

La génétique statistique cherche à développer des méthodes pour exploiter l'explosion de données génétiques découlant des projets d'exploration de génomes, en particulier humains. Elle s'intéresse à l'inférence des processus évolutifs, agissant aussi bien à l'échelle moléculaire qu'à l'échelle des individus ou des populations, contribuant à façonner le paysage génomique intra-spécifique.

Depuis l'œuvre monumentale de Ronald Fisher, la statistique joue un rôle fondamental dans la compréhension des mécanismes évolutifs à partir de l'observation du polymorphisme génétique. Plus récemment, de nouvelles approches, souvent computationnelles ou bayésiennes, permettent la cartographie des aires génomiques soumises à la sélection, des points chauds de mutation et de recombinaison, la détection de gènes associés à des maladies ou à des phénotypes d'intérêt, et la compréhension de la structure génétique des populations.

Cette session couvre plusieurs des aspects mentionnés ci-dessus. L'accent est porté sur les approches de criblages génomiques, les études d'association et la détection de la sélection naturelle, mettant en perspectives les aspects d'analyse statistique et les nouveaux challenges pour cette discipline.

### Références :

- [1] Balding DJ, Bishop M and Cannings C. (2007) *Handbook of Statistical Genetics (3th Ed.)*, John Wiley and Sons.
- [2] Akey, J. M., G. Zhang, K. Zhang, L. Jin and M. D. Shriver (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12 : 1805–1814.
- [3] Barreiro, L. B., G. Laval, H. Quach, E. Patin and L. Quintana-Murci (2008) Natural selection has driven population differentiation in modern humans. *Nature Genetics*, in press.
- [4] Sabeti, P. C., et al. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913 –918.

### Adresse de l'organisateur :

Olivier FRANÇOIS  
Grenoble INP, Laboratoire TIMC-IMAG  
Faculté de Médecine, 38706 La Tronche  
E-mail : [olivier.francois@imag.fr](mailto:olivier.francois@imag.fr)

## Détection de la sélection naturelle à l'échelle globale du génome humain

par **Guillaume Laval**

Un des buts de la biologie évolutive est de détecter les gènes soumis à sélection naturelle afin de préciser les bases génétiques de l'adaptation des espèces à leur environnement (ex. détecter les gènes de résistance aux maladies infectieuses, détecter les mutations responsables d'une meilleure digestion du lactose chez l'humain, etc). Cependant de nombreux travaux ont montré que sélection naturelle, histoire évolutive et histoire démographique des populations peuvent affecter les génomes de manière parfois quasi similaire. Par exemple l'augmentation en fréquence d'une mutation favorable (modalité de sélection naturelle dite sélection positive) entraîne les mêmes effets sur les statistiques de détection de la sélection naturelle (ex.  $D$  de Tajima) qu'un simple accroissement de taille des populations. Par conséquent il est primordial de connaître les distributions de ces statistiques attendues sous hypothèse de neutralité évolutive (c'est-à-dire d'absence de sélection naturelle). Actuellement, nous ne disposons pas d'un cadre statistique permettant de prédire ces distributions nulles de manière satisfaisante, c'est à dire en prenant en compte des histoires évolutives et démographiques suffisamment complexes et réalistes pour refléter ce que nous savons de l'évolution humaine. Cependant l'avènement des données de génotypage à grande échelle nous permet de mesurer la variabilité génétique "neutre" à l'échelle globale du génome humain. Le projet HapMap rassemble une collection de 1 million (phase 1) et 3 millions (phase 2) de SNPs (Single Nucleotide Polymorphism ou mutations ponctuelles) autosomaux génotypés dans 4

populations humaines, Afrique, Europe, Chine et Japon. Le projet Perlegen rassemble une collection de 1.5 millions de SNPs autosomaux génotypés dans 3 populations, Afro-Américains, Euro-Américains, Sino-Américains. Ces données nous donnent donc expérimentalement accès aux distributions attendues en absence de sélection naturelle de toute une gamme de statistiques (déséquilibre de liaison, spectre des fréquences alléliques, distances génétiques, etc), distributions qui sont les bases de tous tests statistiques de détection de la sélection naturelle.

Dans les approches statistiques classiques dites “outliers”, une manière de détecter les gènes soumis à sélection est de scanner l’ensemble du génome (données HapMap, Perlegen) afin de déclarer comme soumis à sélection les marqueurs dont les indices de diversité génétique,  $F_{ST}$  (distance génétique entre populations) par exemple, présentent des valeurs extrêmes. En effet le  $F_{ST}$  est une statistique sensible aux effets de la sélection naturelle : la sélection négative (élimination des mutations faiblement délétères) et la sélection balancée (sélection favorisant le maintien de la diversité génétique des populations, ex. avantage sélectif des hétérozygotes) diminuent les valeurs de  $F_{ST}$  par rapport à ce qui est attendu sous hypothèse de neutralité. A l’inverse, la sélection positive (sélection des allèles favorables dans une population donnée) augmente les valeurs de  $F_{ST}$ . Ces approches outliers appliquées aux données des projets internationaux de génotypage « genome entier » (HapMap et Perlegen) ont déjà permis de détecter des dizaines de gènes humains potentiellement soumis à sélection naturelle (Akey et al. 2002, Sabeti et al. 2007 et bien d’autres). Cependant ces approches classiques ne peuvent qu’indiquer de potentielles cibles pour la sélection, elles ne démontrent pas si la sélection naturelle est (ou a été) effectivement à l’œuvre. Une autre manière d’utiliser les données HapMap et Perlegen (Barreiro et Laval et al. 2008), qui est basée sur des comparaisons entre classes de SNPs (établies en fonction leur position génomique ou de leur effet sur les protéines), a notamment permis de mettre en évidence l’existence de la sélection positive, ou sélection naturelle adaptative, en tant que force évolutive agissant au cours de notre histoire récente. Cette étude nous a donc permis de démontrer le rôle joué par différentes modalités de sélection naturelle dans la différenciation génétique des populations humaines, et ce particulièrement pour des marqueurs SNPs synonymes (mutations modifiant les protéines) et non-synonymes (mutations silencieuses).

*Adresse :*

Guillaume LAVAL  
Institut Pasteur, Paris  
E-mail : [glaval@pasteur.fr](mailto:glaval@pasteur.fr)

## Détection de balayages sélectifs à l’aide de chaînes de Markov cachées

par **Simon Boitard**

On parle de balayage sélectif (selective sweep) quand un allèle, en raison de l’avantage sélectif important qu’il représente, voit sa fréquence augmenter très rapidement dans une population donnée jusqu’à être complètement fixé dans cette population. Ce type d’évènement laisse des traces dans le génome d’une population et un des enjeux importants de la génétique des populations est de pouvoir détecter de telles traces de sélection. Nous présentons ici un modèle de chaîne de Markov cachée permettant de détecter des balayages sélectifs à partir d’un échantillon de séquences génomiques. Des simulations indiquent que cette méthode permet d’estimer la position des balayages sélectifs avec plus de précision que les méthodes actuellement existantes.

*Adresse :*

Simon BOITARD  
INRA Toulouse  
E-mail : [simon.boitard@toulouse.inra.fr](mailto:simon.boitard@toulouse.inra.fr)



## **A genome scan method to identify selected loci appropriate for both dominant and codominant markers : A Bayesian perspective**

par **Matthieu Foll**

Identifier des loci influencés par la sélection naturelle présente un enjeu très important dans différents domaines. Ces loci ont une plus grande probabilité d'avoir un rôle dans une fonction biologique importante, et sont également de bons candidats pour être impliqués dans des maladies ou des dysfonctionnements génétiques. Une approche courante pour séparer les effets neutres des effets adaptatifs consiste à identifier les marqueurs ayant un coefficient de différenciation génétique atypique. Nous proposons d'étendre une telle méthode pour estimer directement la probabilité que chaque locus soit soumis à la sélection en utilisant un modèle bayésien. Nous généralisons aussi la méthode aux marqueurs dominants de type AFLP. Nous présentons enfin les résultats d'une analyse de sensibilité portant sur le type de marqueurs utilisé (SNP, SSR, AFLP) et sur le modèle démographique.

*Adresse :*

Matthieu FOLL  
CMPG, Bern  
E-mail : [matthieu.foll@ujf-grenoble.fr](mailto:matthieu.foll@ujf-grenoble.fr)

## **Reseaux d'interaction de protéines et études d'association à l'échelle du génome**

par **Mathieu Emily**

Les études d'association ont pour objectif de détecter des régions du génome statistiquement associées à des maladies complexes (cancer de la prostate, cancer du sein par exemple). La première vague de telles études à l'échelle du génome a démontré la puissance de cette approche. Bien que de nombreux travaux biologiques ont mis en évidence le rôle des interactions génétiques, les premières études d'association se sont focalisés sur des effets marginaux. Dans ce travail, nous proposons une approche, incorporant des réseaux d'interaction de protéines, qui permet de détecter des interactions de paires de zones du genome.

*Adresse :*

Mathieu EMILY  
Bioinformatics Research Center, Aarhus  
E-mail : [memily@birc.au.dk](mailto:memily@birc.au.dk)  
<>



## Statistique spatiale et spatio-temporelle

Session organisée par **Cécile Hardouin**

La statistique spatiale est un domaine dans lequel les recherches s'intensifient. Elle trouve des applications dans des secteurs très divers comme l'épidémiologie, la météorologie, l'environnement, le traitement d'images... Nous avons choisi dans cette session de présenter des exemples relevant aussi bien de problèmes théoriques que de méthodologie.

### Références :

- [1] HAINING R. *Spatial data analysis in the social and environmental sciences*, (2000) Cambridge University press.
- [2] GUYON X. *Random fields on a network*, (1995) Modeling, statistics and applications Springer-Verlag.
- [3] LAWSON A.B. *Statistical methods in spatial epidemiology*, (2001) Wiley.
- [4] MOORE M. (Ed.) *Spatial statistics : methodological aspects and applications*, (2001) LNS 159 Springer.

### Adresse de l'organisateur :

Cécile HARDOUIN  
SAMOS-MATISSE, Centre d'Economie de la Sorbonne  
Université Paris 1  
90 rue de Tolbiac 75634 Paris Cedex 13 France  
E-mail : [hardouin@univ-paris1.fr](mailto:hardouin@univ-paris1.fr)  
<<http://samos.univ-paris1.fr/spip/HARDOUIN-Cecile>>

## Simulation conditionnelle de modèles stochastiques spatiaux

par **Christian Lantuéjoul**

Dans de nombreuses disciplines (sciences de la terre, météorologie, physique des matériaux ...), il est souvent avantageux d'interpréter le phénomène spatial étudié en tant que réalisation particulière d'un modèle stochastique, car cela permet son investigation au moyen des outils et des techniques fournis par le calcul des probabilités. C'est à ce niveau qu'interviennent les simulations conditionnelles. Conçues au départ à des fins de prédiction, elles peuvent aussi servir à quantifier la connaissance que les données apportent au phénomène et aussi à tester la sensibilité des paramètres du modèle. Bien entendu, tous les résultats obtenus sont totalement tributaires du modèle spatial retenu. Ces résultats ne valent que ce que vaut l'adéquation du modèle au phénomène étudié. Dans cette présentation, nous proposons quelques principes simples pour concevoir des algorithmes de simulation conditionnelle de fonctions et d'ensembles aléatoires. De nos jours, ces algorithmes doivent être de plus en plus raffinés pour faire face à l'évolution radicale du type et du nombre de données disponibles.

### Adresse :

Christian LANTUÉJOUL  
Ecole des Mines de Paris  
35 rue Saint-Honoré  
77305 Fontainebleau, France  
E-mail : [Christian.Lantuejoul@ensmp.fr](mailto:Christian.Lantuejoul@ensmp.fr)

## Composite likelihood methods for space and space-time covariance models

par **Carlo Gaetan**

In the last years there has been a growing interest in the construction of space and space-time covariance functions for Gaussian random fields. However exact computation of the likelihood requires

calculation of the inverse and determinant of the covariance matrix, and this evaluation is slow when the number of observations is large. This fact motivates the search for approximations to the likelihood function that require smaller computational burdens. In this talk we propose a composite likelihood approach to the estimation of the covariance function. An identification criterion based on the composite likelihood is also introduced.

*Adresse :*

Carlo GAETAN  
Dipartimento Di Statistica ,  
Università Ca' Foscari  
Venezia San Giobbe, Canaregio 873,  
30121 Venezia, Italia  
E-mail : [gaetan@unive.it](mailto:gaetan@unive.it)  
<<http://www.dst.unive.it/~gaetan>>

## Modélisation spatialement explicite d'une dynamique forestière et inférence

par **Fabien Campillo**, Nicolas Desassis et Vivien Rossi

Comprendre comment évolue le peuplement forestier tropical est un enjeu crucial pour mettre au point des stratégies d'exploitation qui maintiennent la biodiversité. La prise en compte de la spatialisation des arbres est essentielle à une bonne compréhension des mécanismes écologiques.

On modélise le peuplement d'une espèce et sa dynamique par un processus markovien ponctuel de branchement. En plus de la position d'un individu, la variable d'état comprend par exemple sa taille. Ce type de modèle permet de rendre compte des phénomènes de compétition pour l'accès aux ressources.

Le générateur du processus est explicité. Une procédure de simulation de Monte Carlo est proposée. Cette approche permet de faire appel à des méthodes statistiques de type MCMC afin de caler le modèle à des données, en l'occurrence à un jeu de données spatio-temporelles du dispositif de Paracou en Guyane.

*Adresses :*

Fabien CAMPILLO  
INRA/INRIA - UMR ASB  
2 place Viala  
34060 Montpellier cedex 06, France  
E-mail : [Fabien.Campillo@inria.fr](mailto:Fabien.Campillo@inria.fr)  
<<http://www-sop.inria.fr/mere/personnel/campillo/>>

Nicolas DESASSIS  
INRA/INRIA - UMR ASB  
2 place Viala  
34060 Montpellier cedex 06, France  
E-mail : [Nicolas.Desassis@supagro.inra.fr](mailto:Nicolas.Desassis@supagro.inra.fr)

Vivien ROSSI  
Unité de recherche des Dynamique des forêts naturelles CIRAD  
Campus International de Baillarguet,  
34398 Montpellier cedex 5, France  
E-mail : [viven.rossi@cirad.fr](mailto:viven.rossi@cirad.fr)  
<<http://vrossi.free.fr/>>

## Prédiction d'un champ aléatoire spatial multivarié composé de variables de différente nature

par Jean-Noël Bacro, **Pierrette Chagneau**, Frédéric Mortier et Nicolas Picard

Dans la plupart des problèmes de géostatistique, les données collectées sont multivariées et les variables étudiées n'ont pas toutes la même nature (variables gaussiennes, variables de comptage, variables ordinales ou nominales). Prédire le champ aléatoire spatial constitué par ces variables nécessite d'avoir recours à des méthodes capables de traiter simultanément des variables de nature différente. Ces méthodes sont peu nombreuses dans la littérature. L'objectif de ce travail est de proposer un modèle permettant de prédire un champ aléatoire spatial multivarié composé à la fois de variables gaussiennes, de variables de Poisson, de variables ordinales et nominales. Le modèle proposé est basé sur la construction de processus spatiaux dépendants  $Z_k$ , où  $Z_k(s)$  peut être une variable aléatoire discrète ou continue. Dans le cas non spatial, les variables ordinales et nominales sont traitées grâce au modèle probit ordinal et au modèle probit multinomial. Le principe de ces modèles est de redéfinir le problème en termes de variables latentes gaussiennes. Nous généralisons ces modèles au cas spatial en introduisant des processus gaussiens sous-jacents. Le modèle repose sur une structuration hiérarchique comparable à celle proposée par [1]. La dépendance spatiale entre les processus est modélisée par une approche moyenne mobile ([2]). La structure de dépendance obtenue est souple. Le modèle est appliqué à la prédiction de caractéristiques du sol à partir de données recueillies en Guyane française.

[1] Wolpert R., Ickstadt K. (1998) Poisson/gamma random field models for spatial statistics. *Biometrika* 85 , no. 2, 251–267.

[2] Ver Hoef J., Barry, R. (1998) Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plann. Inference* 69 , no. 2, 275–294.

### Adresses :

Jean-Noël BACRO  
I3M  
Université de Montpellier 2  
CC 051, Place Eugène Bataillon,  
34095 Montpellier cedex 5, France  
E-mail : [bacro@math.univ-montp2.fr](mailto:bacro@math.univ-montp2.fr)

Pierrette CHAGNEAU  
CIRAD, UR 37  
Campus international de Baillarguet  
TA C-37|D, 34398 Montpellier Cedex 5, France  
E-mail : [pierrette.chagneau@cirad.fr](mailto:pierrette.chagneau@cirad.fr)

Frédéric MORTIER  
CIRAD, UR 39  
Campus International de Baillarguet  
TA A-39|C, 34398 Montpellier cedex 5, France  
E-mail : [frederic.mortier@cirad.fr](mailto:frederic.mortier@cirad.fr)

Nicolas PICARD  
CIRAD, UR 37  
Campus international de Baillarguet  
TA C-37|D, 34398 Montpellier cedex 5, France  
E-mail : [nicolas.picard@cirad.fr](mailto:nicolas.picard@cirad.fr)

## Autour d'une technique de rééchantillonnage non paramétrique pour les champs aléatoires

par **Lionel Truquet**

Dans un article récent de Bickel et Levina [1], les auteurs montrent la consistance d'une méthode de rééchantillonnage appliquée à la synthèse de textures. Cet algorithme utilisé pour des champs de Markov unilatéraux est basée sur l'estimation de la spécification conditionnelle en un site par une autre obtenue par rééchantillonnage. La méthode d'estimation est basée sur l'utilisation d'un lissage par noyau. Contrairement à Bickel et Levina, nous obtenons des résultats de nature globale. Nous montrons que ce cadre se généralise pour des champs non causaux et nous étudions le problème de calculs de vitesse de convergence pour les lois jointes pour un tel estimateur à l'aide de la méthode de contraction de Dobrushin [2].

[1] Bickel, P., Levina, E. (2006) Texture synthesis and nonparametric resampling of random fields. *The Annals of Statistics*, Vol. 34, No 4, 1751-1773.

[2] Dobrushin, R.L. (1970) Prescribing a system of random variables by conditional distributions. *Theor. Probab. Appl.* 15, 458-486.

[3] Föllmer, H. (1988) *Random Fields and Diffusion Processes*. Ecole d'Eté de Probabilités de St. Flour XVI. Lecture Notes in Mathematics 1362, Springer, 101-203

*Adresse :*

Lionel TRUQUET  
SAMOS-MATISSE, CES et CREST  
Samos, Université Paris 1  
90, rue de Tolbiac,  
75634 Paris Cedex 13, France  
E-mail : [truquet@gmail.com](mailto:truquet@gmail.com)

## Méthodes régénératives en statistiques

Session organisée par **Eva Löcherbach**

Dans un problème de statistique des processus lié à l'observation d'un processus de Markov fort  $X = (X_t)_{t \geq 0}$ , supposons qu'il existe un point  $x_0$  récurrent. Dans ce cas, nous pouvons décomposer la trajectoire du processus en des *cycles de vie*  $R_n, n \geq 1$  :

1.  $R_n < \infty$  p.s.,  $R_{n+1} = R_n + R_1 \circ \theta_{R_n}$ .
2.  $X_{R_n} = x_0$ .
3.  $(X_{R_n+t})_{t \geq 0}$  est indépendant de  $\sigma\{X_s, s \leq R_n\}$ .

Dans ce cas, la trajectoire du processus peut être décomposée en des cycles de vie  $[R_n, R_{n+1}[$ ,  $n = 1, 2, \dots$ , et des théorèmes limites, par exemple le théorème ergodique pour des fonctionnelles additives, se démontrent comme une simple conséquence de la loi des grands nombres. Des théorèmes limites pour des martingales associées à  $X$  peuvent également être démontrés sous une condition sur les queues de la distribution de la longueur d'un cycle de vie  $R_2 - R_1$ . Or, en général, un point récurrent  $x_0$  n'existe pas. Athreya et Ney, [1] et Nummelin, [6], donnent une construction qui permet d'introduire un atôme récurrent pour une chaîne de Markov. Cette méthode a été utilisée dans [4] pour obtenir des théorèmes limites pour des martingales fonctionnelles additives, et ensuite développé dans [5] au cadre des processus de Markov récurrent en temps continu. Ce dernier travail a donné lieu à une entière caractérisation des vitesses de convergence dans un problème de statistique lié à l'observation de  $X$  – également dans le cas récurrent-nul.

Dans le cadre des chaînes de Markov récurrentes, Stéphan Cléménçon a notamment utilisé la structure régénérative pour construire des procédures d'estimation explicites, par exemple le Bootstrap, dont la vitesse de convergence dépasse celle des méthodes classiques, voir par exemple [2].

Finalement, indiquons un champs d'applications nouveau : les chaînes de Markov d'ordre variable. "Ordre variable" veut dire que la longueur du passé dont on a besoin pour prédire l'état suivant du processus dépend du passé observé. De tels processus ont été introduits dans la théorie de l'information par Rissanen, [7]. On les a ensuite utilisés dans d'autres domaines des sciences, par exemple en biologie pour modéliser des données de protéines, voir Leonardi et Galves, [3]. L'existence de temps de régénération est un outil principal pour l'étude de tels processus, pour l'existence, la simulation exacte de la mesure invariante et l'étude de la vitesse de convergence vers l'équilibre.

### Références :

- [1] Athreya, K.B., Ney, P. A new approach to the limit theory of recurrent Markov chains. *Trans. Am. Math. Soc.* 245 (1978), 493-501.
- [2] Bertail, P., Cléménçon, S. : *Regeneration-based statistics for Harris Markov chains*. In : *Dependance in Probability and Statistics*. Eds. P. Bertail, P. Soulier, P. Doukhan, *Lecture Notes in Statistics*, No 187, 1–53, Springer (2006).
- [3] Leonardi, F., Galves, A.. Sequence Motif identification and protein classification using probabilistic trees. In *Lecture Notes in Computer Science 3594*, 190–193, 2005.
- [4] Höpfner, R., Löcherbach, E. : *Limit theorems for null recurrent Markov processes*. *Memoirs AMS* **161**, Number 768, 2003.
- [5] Löcherbach, E., Loukianova, D.. On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. To appear in SPA 2007.
- [6] Nummelin, E. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 43 (1978), 309–318.
- [7] Rissanen, J.. A universal data compression system. *IEEE Trans. Inform. Theory* 29, no. 5, 656–664, 1983.

### Adresse de l'organisateur :

Eva LÖCHERBACH

Laboratoire d'Analyse et de Mathématiques Appliquées UMR CNRS 8050

Université Paris 12 - Val de Marne

94000 Créteil CEDEX France

E-mail : locherbach@univ-paris12.fr

<<http://perso-math.univ-mlv.fr/users/locherbach.eva/>>

## Nummelin splitting for continuous time Harris recurrent Markov processes and applications

par Dasha Loukianova et Eva Löcherbach

Let  $X$  be a Harris recurrent strong Markov process in continuous time. We introduce a sequence of stopping times that allow to study an analogue of a life-cycle decomposition for  $X$ , which is an extension of the well-known splitting technique of Nummelin to the time-continuous case.

Due to the time-continuous setting specificity we can not get a real independence between the life cycles. However we get a weak form of independence, which implies a strong mixing property for increments of functionals of  $X$  during a life cycle.

As a consequence we are able to give deterministic equivalents of additive functionals of  $X$  and to generalise a law of iterated logarithm for additive functionals and martingale additive functionals obtained by Chen to the time-continuous setting.

We apply our results to the problem of non-parametric kernel estimation of the drift of multi-dimensional recurrent, but not necessarily ergodic, diffusion processes.

*Adresses :*

Dasha LOUKIANOVA  
Université d'Evry  
Bd François Mitterrand, 91025 Evry Cedex, France  
E-mail : [dasha.loukianova@mathuniv-evry.fr](mailto:dasha.loukianova@mathuniv-evry.fr) web

Eva LÖCHERBACH  
LAMA  
Université Paris 12 Val de Marne 94220 Créteil Cedex, France  
E-mail : [locherbach@univ-paris12.fr](mailto:locherbach@univ-paris12.fr)  
<<http://perso-math.univ-mlv.fr/users/locherbach.eva/>>

## Bootstrap (pseudo-) régénératif pour les chaînes de Markov

par Stéphane Cléménçon

Cet exposé s'attachera à décrire les principes du *bootstrap régénératif* pour le rééchantillonnage des chaînes de Markov. L'idée consiste à rééchantillonner un nombre aléatoire de blocs de données définis par les instants de (pseudo-) régénération, puis, en les mettant "bout à bout", à reconstruire une "trajectoire bootstrap" de taille comparable à la séquence initiale. On montrera en particulier que, dans le cas régénératif, la vitesse d'approximation de cette méthode est exactement celle du bootstrap i.i.d.,  $O_P(n^{-1})$  pour la distribution de moyennes empiriques du type  $n^{-1} \sum_{i \leq n} f(X_i)$ .

*Références :*

- [1] Sharp probability inequalities for Markov chains. (2006), with P. Bertail. To appear in **Theory of Probability and its Applications**.
- [2] Approximate Regenerative Block Bootstrap : some simulation studies. (2006), with P. Bertail. To appear in **Comp. Stat. and Data Analysis**.
- [3] Regeneration-based statistics for Harris Markov chains. (2006), with P. Bertail. In **Dependence in Probability and Statistics**, LNS, No 187, 1-53. Springer.
- [4] Regenerative Block Bootstrap for Markov Chains. (2005), with P. Bertail. In **Bernoulli**.
- [5] Note on the regeneration-based bootstrap for atomic Markov chains. (2005), with P. Bertail. In **Test**.
- [6] Edgeworth expansions for suitably normalized sample mean statistics of atomic Markov chains. (2004), with P. Bertail. In **PTRF**.
- [7] Approximate Regenerative Block-Bootstrap for Markov Chains : second-order properties. (2004), with P. Bertail. In **Compstat 2004 Proc.** Physica Verlag.
- [8] Extreme Values Statistics for Markov Chains via the (Pseudo-) Regenerative Method. (2007), with P. Bertail and J. Tressou. To appear in **Extremes**.



*Adresse :*

Stéphan CLÉMENÇON  
ENST Télécom Paris  
46, rue Barrault, 75634 Paris CEDEX 13, France  
E-mail : [stephan.clemencon@enst.fr](mailto:stephan.clemencon@enst.fr)  
<<http://metarisk.inapg.inra.fr/content/view/full/331>>

## **Uniqueness without regeneration scheme for chains with unbounded variable length memory**

par **Gallo Alessandro**

We consider a certain class of unbounded context trees which have been introduced by D. Maia in her PhD Thesis (March, 2007). D. Maia found a condition which implies the existence of a regeneration scheme for the stationary process consistent with the tree. Considering the same class of trees, we give a condition which allows us to keep the uniqueness of the stationary consistent process, without the regeneration scheme.

*Adresse :*

Gallo ALEXSANDRO  
Universidade de São Paulo  
Instituto de Matemática e Estatística, Universidade de São Paulo, BP 66281, 05315-970 São Paulo, Brasil  
E-mail : [gsandro@ime.usp.br](mailto:gsandro@ime.usp.br)

## **On a problem of statistical inference in null recurrent diffusions**

par **Reinhard Höpfner**

We consider a prototype example for statistical inference in null recurrent one dimensional diffusions. First, in a parametric model with one-dimensional parameter  $\vartheta$ , we prove local asymptotic mixed normality (LAMN) and efficiency of the sequence of maximum likelihood estimates (MLE). Their rate of convergence is  $n^{\alpha/2}$  where  $\alpha$  is ranging over  $(0, 1)$ , depends on  $\vartheta$ , and is related to the tail index (in the sense of regular variation at  $\infty$ ) of the invariant measure under  $\vartheta$ . Second, in a semiparametric model (we add an unknown nuisance function with compact support to the drift of the parametric model), we prove a local asymptotic minimax bound and specify asymptotically efficient estimates for the unknown parameter.

Höpfner, R., Löcherbach, E. : Limit theorems for null recurrent Markov processes. *Memoirs AMS, American Mathematical Society* 2003.

Höpfner, R., Kutoyants, Yu. : : On a problem in statistical inference in null recurrent diffusions. *Stat. Inference Stoch. Processes* **6**, 25-42 (2003).

Khasminskii, R. : Stochastic stability of differential equations. *Sijthoff and Noordhoff*, 1980.

Khasminskii, R. : Asymptotic behaviour of parabolic equations arising from one-dimensional null recurrent diffusions. *J. Diff. Equations* **161**, 154-173 (2000).

Touati, A. : Théorèmes limites pour les processus de Markov récurrents. *C. R. Acad. Sci. Paris Série I* **305**, 841-844 (1988).

Journées MAS 2008, Rennes

*Adresse :*

Reinhard HÖPFNER

Johannes-Gutenberg Universität Mainz

Institut für Mathematik, Johannes-Gutenberg Universität Mainz, Staudinger Weg 9, 55099 Mainz, Allemagne

E-mail : [hoepfner@mathematik.uni-mainz.de](mailto:hoepfner@mathematik.uni-mainz.de)

<<http://joguinf.informatik.uni-mainz.de/~hoepfner/>>

## Processus de branchements continus et applications

Session organisée par **Romain Abraham**

Les processus de branchements continus sont les seules limites possibles de processus de Galton-Watson renormalisés. Ils sont donc très adaptés pour modéliser l'évolution de grandes populations, sur de longs intervalles de temps. Une structure généalogique (i.e. un arbre continu) associée à ces processus de branchement n'a été introduite qu'avec les travaux de Le Gall-Le Jan [3] puis de Duquesne-Le Gall [2], généralisant ainsi le CRT (Continuum Random Tree) d'Aldous. Cette structure généalogique est extrêmement riche. Elle a permis d'obtenir de nombreux résultats particulièrement intéressants sur ces processus de branchement continu et a conduit à de nombreuses applications : processus de fragmentation/coalescence [4], applications en biologie (arbres philogéniques, mutations dans une population, populations multi-types ...). Nous proposons ici de présenter des aspects théoriques concernant ces processus de branchement ainsi que des applications récentes utilisant ces objets.

### Références :

- [1] J. Berestycky - N. Berestycky - J. Schweinsberg. Beta-coalescents and continuous stable random trees. *Ann. Probab.* 35 :1835-1887, 2007.
- [2] T. Duquesne - J.F. Le Gall. *Random trees, Lévy processes and spatial branching processes*. Vol. 281, Astérisque, 2002.
- [3] J.F. Le Gall - Y. Le Jan. Branching processes in Lévy processes : The exploration process. *Ann. Probab.* 26 :213-252, 1998.
- [4] G. Miermont. Self-similar fragmentations derived from the stable tree : splitting at nodes. *Probab. Th. rel. Fields* 131 :341-375, 2005.

### Adresse de l'organisateur :

Romain ABRAHAM  
Laboratoire MAPMO  
Université d'Orléans  
45 067 Orléans cedex 2  
E-mail : [Romain.Abraham@univ-orleans.fr](mailto:Romain.Abraham@univ-orleans.fr)  
<<http://www.univ-orleans.fr/mapmo/membres/abraham/>>

## Arbres continus stables

par **Thomas Duquesne**

Les arbres continus stables qui sont des espaces métriques aléatoires constituant les analogues continus des arbres de Galton-Watson. Dans cet exposé nous donnons une construction élémentaire et nous détaillons certaines propriétés géométriques et fractales (mesures de Hausdorff et de packing) ainsi que le lien avec le super mouvement Brownien.

### Adresse :

Thomas DUQUESNE  
Laboratoire PMA  
Université Paris 6  
75252 PARIS Cedex  
E-mail : [thomas.duquesne@math.u-psud.fr](mailto:thomas.duquesne@math.u-psud.fr)

## Processus de Dawson Watanabe et diffusion de Feller multitypes conditionnés à la non-extinction

par **Nicolas Champagnat**

Nous nous intéressons à des processus de Dawson-Watanabe multitypes critiques ou sous-critiques conditionnés à ne jamais s'éteindre. Sur tout intervalle de temps borné, la loi du processus conditionné est une  $h$ -transformée de la loi du processus non conditionné. Nous obtenons également la transformée de Laplace de ses transitions. Ces caractérisations de la loi du processus sont ensuite utilisées pour prouver plusieurs résultats sur le comportement en temps long de la masse du processus — une diffusion de Feller multitype, avec différents types de conditionnement. Cette étude est menée dans le cas général des matrices de mutation irréductibles, où l'on examine également le problème de l'interversion des limites, ainsi que pour plusieurs exemples de matrices de mutation réductibles dans le cas bi-type, où différents comportements sont observés en fonction du (ou des) type(s) conditionnés à ne pas s'éteindre.

*Adresse :*

Nicolas CHAMPAGNAT  
INRIA Sophia Antipolis  
06902 Sophia Antipolis Cedex  
E-mail : [Nicolas.Champagnat@sophia.inria.fr](mailto:Nicolas.Champagnat@sophia.inria.fr)  
<<http://www-sop.inria.fr/tosca/personnel/Nicolas.Champagnat/>>

## Elagage d'arbres continus

par **Guillaume Voisin**

Etant donnés deux mécanismes de branchement  $\psi$  et  $\psi_0$ , nous cherchons à construire simultanément deux processus de branchement continus  $X$  et  $X^0$  tels que, pour tout temps  $t$ ,  $X_t^0 \leq X_t$  (autrement dit,  $X^0$  est une sous population de la population  $X$ ). Sous certaines conditions liant  $\psi_0$  et  $\psi$ ,  $\psi$  étant un mécanisme de branchement critique ou sous-critique très général, une telle construction est possible en considérant l'arbre généalogique associé au processus total  $X$  et en élaguant ce dernier afin d'obtenir un arbre généalogique décrivant l'évolution de la sous-population  $X^0$ .

*Adresse :*

Guillaume VOISIN  
Laboratoire MAPMO  
45067 Orléans cedex 2  
E-mail : [guillaume.voisin@univ-orleans.fr](mailto:guillaume.voisin@univ-orleans.fr)

## Arbres de fragmentation invariants par changement de racine

par **Bénédicte Haas**

Un processus de fragmentation décrit l'évolution d'un système de masses se scindant au fur et à mesure que le temps avance. On considère ici des processus de fragmentation ayant une propriété de *branchement* (différentes masses se fragmentent indépendamment les unes des autres) et une propriété d'auto-similarité (le taux de fragmentation d'une masse  $m$  est proportionnel à  $m^\alpha$ ,  $\alpha$  étant un indice *négligé* fixé).

Un arbre de fragmentation est un arbre continu aléatoire enraciné décrivant la généalogie d'un tel processus. On dit qu'il est *invariant par changement de racine uniforme* si en ré-enracinant cet arbre à une feuille prise "uniformément" au hasard, on obtient un arbre ayant la même loi que l'arbre initial avec sa racine initiale.

On caractérisera dans cet exposé les arbres de fragmentation possédant cette propriété d'invariance par changement de racine. Ceci fait partie d'un travail fait en collaboration avec Jim Pitman (Berkeley) et Matthias Winkel (Oxford).

Journées MAS 2008, Rennes

*Adresse :*

Bénédicte HAAS

Ceremade

Université Paris-Dauphine

75016 Paris

E-mail : `haas.ceremade.dauphine.fr`



## Processus autorégressifs

Session organisée par **Benoîte de Saporta**

Les processus autorégressifs sont utilisés pour modéliser des séries chronologiques dans de nombreux domaines : finance, biologie, climatologie, ... A partir du modèle autorégressif linéaire simple d'ordre 1 AR(1) :

$$X_n = aX_{n-1} + b + \varepsilon_n$$

de nombreux modèles plus sophistiqués sont étudiés : modèles à coefficient aléatoires (indépendants [6], markoviens [8]) éventuellement multidimensionnels, modèles ARCH [2], GARCH, modèles à bifurcation [9, 5], modèles non linéaires. Les recherches actuelles se concentrent autour de deux thèmes : l'étude des propriétés théoriques de ces processus (existence, moments et queue d'une loi stationnaire, comportement extrême [7]) dont la connaissance permet de mieux choisir les modèles en fonction des données ; et d'autre part l'estimation des paramètres des modèles [1] pour pouvoir répondre aux problèmes pratiques et faire des prévisions.

### Références :

- [1] *An autoregressive model with time-varying coefficients for wind fields* (2006) P. Ailliot, V. Monbet and M. Prevosto, *Environmetrics* 2, pp107–117
- [2] *The tail of the stationary distribution of an autoregressive process with ARCH(1) errors* (2001) M. Borkovec and C. Klüppelberg, *Ann. Appl. Probab.* 4, pp1220–1241
- [3] *Data assimilation in the presence of forecast bias* (1998) Dee, D., and da Silva, A.M., *Quart. J. Roy. Meteor. Soc.*, 124, pp269–295
- [4] *A one dimensional Kalman filter for the correction of 2 m-temperature forecasts* (2002) Galanis, G., and Anadranistakis, M., *Meteorological Applications*, 9, pp437–441
- [5] *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging* (2007) *Ann. Appl. Probab.* 17, pp1538–1569
- [6] *Random difference equations and renewal theory for products of random matrices* (1973) H. Kesten, *Acta Math* 131, pp 207–248
- [7] *Extremal behaviour of models with multivariate random recurrence representation* (2007) C. Klüppelberg and S. Pergamenchtchikov, *Stochastic Process. Appl.* 117, pp432–456
- [8] *Tail of the stationary solution of the stochastic equation  $Y_{n+1} = a_n Y_n + b_n$  with Markovian coefficients* (2005) B. de Saporta, *Stochastic Process. Appl.* 12, pp1954–1978
- [9] *Least-squares for bifurcation autoregressive processes* (2005) *Statistics and Probability Letters* 74, pp77–88

### Adresse de l'organisateur :

Benoîte DE SAPORTA  
GREThA et INRIA Bordeaux Sud Ouest  
Université Montesquieu Bordeaux IV  
Avenue Léon Duguit  
33608 Pessac Cedex, France  
E-mail : [Benoite.deSaporta@math.u-bordeaux1.fr](mailto:Benoite.deSaporta@math.u-bordeaux1.fr)  
<<http://www.math.u-bordeaux.fr/~saporta/>>

## Comportement extrême des processus autorégressifs multivariés

par **Serge Pergamenchtchikov**

Cet exposé est consacré aux processus autorégressifs d'ordre  $q \geq 1$  à coefficients aléatoires. Pour souligner l'importance de ces modèles, en particulier, on montre que les processus de type de ARCH autorégressifs sont équivalents en distribution aux processus autoregressifs multivariés.

Pour les valeurs extrêmes des composantes de ces processus on montre des théorèmes limites et on calcule les indices extrêmes. De plus, pour étudier le comportement des "clusters" dans ces modèles on obtient la limite en distribution pour le processus ponctuel d'excès d'un grand seuil. Il se trouve que cette limite est un processus de comptage de Poisson.

*Adresse :*

Serge PERGAMENCHTCHIKOV  
Laboratoire de Mathématiques Raphaël Salem  
Université de Rouen  
Avenue de l'Université, BP.12  
Technopôle du Madrillet  
76801 Saint-Etienne-du-Rouvray France  
E-mail : [Serge.Pergamenchtkhikov@univ-rouen.fr](mailto:Serge.Pergamenchtkhikov@univ-rouen.fr) web

## Un modèle à espace d'état pour la correction de prévisions météorologiques

par Pierre Ailliot et **Valérie Monbet**

La qualité des prévisions météorologiques est déterminante pour de nombreuses applications, et en particulier pour la gestion des énergies renouvelables. Or si les prévisions fournies par les modèles numériques sont généralement bonnes à grande échelle, on observe fréquemment des erreurs importantes localement.

Une méthode couramment utilisée consiste alors à corriger les sorties des modèles numériques par des modèles stochastiques. Plusieurs auteurs ([3], [4]) ont proposé d'utiliser un modèle à espace d'état linéaire dans lequel le processus caché représente "l'erreur" du modèle numérique. Ensuite, la prédiction obtenue par le modèle numérique est corrigée en temps réel à partir des observations disponibles et en utilisant le filtre de Kalman. Un des principaux intérêts de ce modèle est de permettre un apprentissage adaptatif de l'erreur de prévision.

Dans cette présentation, nous étendons ce modèle. Nous supposons que l'erreur peut être décomposée en deux termes : une erreur de niveau et une erreur de phase. Par exemple, si on s'intéresse au passage d'une tempête, l'erreur de niveau représente l'erreur commise sur l'intensité de la tempête et l'erreur de phase l'erreur commise sur la date d'arrivée de la tempête. Ces deux termes n'étant pas directement observables, ils sont inclus dans l'état caché du modèle à espace d'état qui devient non-linéaire. Nous supposons également que la dynamique de l'état caché est décrite par un processus d'Ornstein-Uhlenbeck, ce qui permet de prendre en compte directement l'irrégularité des dates d'observations.

Après avoir introduit le modèle, nous comparerons les résultats obtenus avec les autres modèles existants dans la littérature sur des données de vent en Bretagne. Nous parlerons également des problèmes liés à l'inférence statistique.

*Adresses :*

Pierre AILLIOT  
Laboratoire de Mathématiques  
Université de Brest  
6, Avenue Victor Le Gorgeu  
B.P. 809  
29285 BREST France  
E-mail : [pierre.ailliot@univ-brest.fr](mailto:pierre.ailliot@univ-brest.fr)  
<<http://pagesperso.univ-brest.fr/~ailliot/index.html>>

Valérie MONBET  
laboratoire Lab-STICC  
Université de Bretagne Sud  
Centre de Recherche Yves Coppens  
Campus de Tohannic, BP 573  
56017 Vannes cedex France  
E-mail : [valerie.monbet@univ-ubs.fr](mailto:valerie.monbet@univ-ubs.fr)  
<<http://web.univ-ubs.fr/sabres/Monbet4/monbet.html>>



## Limit Theorems for Bifurcating Markov Chains. Application to the Detection of Cellular Aging

par **Julien Guyon**

We propose a general method to study dependent data in a binary tree, where an individual in one generation gives rise to two different offspring, one of type 0 and one of type 1, in the next generation. For any specific characteristic of these individuals, we assume that the characteristic is stochastic and depends on its ancestors' only through the mother's characteristic. The dependency structure may be described by a transition probability  $P(x, dydz)$  which gives the probability that the pair of daughters' characteristics is around  $(y, z)$  given that the mother's characteristic is  $x$ . Note that  $y$ , the characteristic of the daughter of type 0, and  $z$ , that of the daughter of type 1, may be conditionally dependent given  $x$ , and their respective conditional distributions may differ. We then speak of bifurcating Markov chains.

We derive laws of large numbers and central limit theorems for such stochastic processes. We then apply these results to detect cellular aging in *Escherichia Coli*, using the data of E. J. STEWART *et al.* and a bifurcating autoregressive model.

*Adresse :*

Julien GUYON  
Société Générale  
Tour Société Générale  
17 cours Valmy  
92987 Paris-La Défense, France  
E-mail : [julien.guyon@sgcib.com](mailto:julien.guyon@sgcib.com)  
<<http://cermics.enpc.fr/~guyon/home.html>>

## Analyse asymptotique des processus autoregressif à bifurcation par une approche martingale

par Bernard Bercu, Benoîte de Saporta et **Anne Gégout-Petit**

A l'aide de méthodes martingales, nous étudions précisément le comportement asymptotique de l'estimateur des moindres carrés des coefficients d'un processus autorégressif de bifurcation. Sous des hypothèses assez faibles sur le bruit, nous obtenons une vitesse de convergence dans la loi des grands nombres ainsi qu'une loi forte quadratique.

*Adresses :*

Bernard BERCU  
Institut de Mathématiques de Bordeaux et INRIA Bordeaux Sud Ouest  
Université Bordeaux 1  
351 Cours de la Libération  
33405 Talence France  
E-mail : [bernard.bercu@math.u-bordeaux1.fr](mailto:bernard.bercu@math.u-bordeaux1.fr)  
<<http://www.math.u-bordeaux.fr/~bercu/>>

Benoîte DE SAPORTA  
GREThA et INRIA Bordeaux Sud Ouest  
Université Montesquieu Bordeaux IV  
Avenue Léon Duguit  
33608 Pessac Cedex, France  
E-mail : [Benoite.deSaporta@math.u-bordeaux1.fr](mailto:Benoite.deSaporta@math.u-bordeaux1.fr)  
<<http://www.math.u-bordeaux.fr/~saporta/>>

Journées MAS 2008, Rennes

Anne GÉGOUT-PETIT

Institut de Mathématiques de Bordeaux et INRIA Bordeaux Sud Ouest

Université Victor Segalen Bordeaux 2,

UFR Sciences et Modélisation

146, Rue Léo Saignat

33076 Bordeaux France

E-mail : [anne.gegout-petit@sm.u-bordeaux2.fr](mailto:anne.gegout-petit@sm.u-bordeaux2.fr)

<<http://www.sm.u-bordeaux2.fr/~agp/>>

## Statistique médicale

Session organisée par **Jean-François Dupuy**

Les données recueillies en recherche biomédicale sont de plus en plus complexes. Leur analyse nécessite le développement de nouveaux outils statistiques.

Par exemple, les modèles non-linéaires à effets mixtes sont très utilisés dans l'analyse des essais thérapeutiques et en pharmacologie, pour étudier les cinétiques virales et la relation dose-réponse. Ces modèles suscitent de nombreuses questions où sont requises les compétences du statisticien : développement de méthodes et d'algorithmes d'estimation, sélection de modèles...

De même, l'étude des modèles de régression pour durées censurées est souvent motivée par des applications médicales. Les développements récents de ces modèles ont pour but de prendre en compte la complexité croissante des données recueillies dans les études cliniques et épidémiologiques : risques concurrents, événements récurrents,...

Cette session propose quatre exposés de travaux récents s'inscrivant dans ces thèmes.

*Adresse de l'organisateur :*

Jean-François DUPUY  
Institut de Mathématiques de Toulouse,  
Laboratoire de Statistique et Probabilités,  
Université Paul Sabatier  
31062 Toulouse cedex 9, France  
E-mail : dupuy@cict.fr  
<<http://www.lsp.ups-tlse.fr/Fp/Dupuy/>>

### **Modélisation de dynamiques virales. Application aux dynamiques du VIH et de l'Hépatite C**

par **Marc Lavielle**

L'efficacité des traitements antorétroviraux, pour des pathologies telles que le VIH ou l'hépatite C, est quantifiée par la décroissance de la charge virale chez le malade. Différents modèles définis par des systèmes d'équations différentielles ordinaires ont été développés pour décrire la dynamique de cette charge virale. Ces modèles prennent en compte la physiologie de la réplication des virus ainsi que les mécanismes des différentes options thérapeutiques.

Il existe une grande variabilité inter-patients dans ces pathologies : l'utilisation de modèles à effets mixtes pour l'étude de ces cinétiques virales permet de mieux comprendre les différences de réponse au traitement d'un patient à l'autre.

L'analyse de données réelles est difficile pour différentes raisons : les modèles sont ici hautement non linéaires, il y a peu de données par sujet, de nombreuses données sont censurées à gauche (inférieures à une limite de quantification). Nous utilisons alors l'algorithme SAEM implémenté dans le logiciel MONOLIX pour l'estimation des paramètres du modèle par maximum de vraisemblance ainsi que pour la sélection de modèle (modèle de covariables, modèle de covariance, modèle d'erreur,...).

*Adresse :*

Marc LAVIELLE  
Département de Mathématiques  
Université Paris-Sud  
Bat 425, Equipe de Probabilités, Statistique et Modélisation  
91405 Orsay Cedex, France  
E-mail : [Marc.Lavielle@math.u-psud.fr](mailto:Marc.Lavielle@math.u-psud.fr)

## Nonparametric estimators in nonlinear mixed effects models for population pharmacokinetics

par **Julie Antic**

During the drug development, nonlinear mixed effects models are routinely used to study the drug's pharmacokinetics. The distribution of random effects is of special interest because it allows to describe the heterogeneity of the drug's kinetics in the population of patients. Parametric estimators generally rely on normally assumptions which are biologically unrealistic and difficult to check. In this respect, nonparametric estimators are very attractive. A naive estimator is nearly exclusively used in the drug companies but this choice is more motivated by the habits than by statistical arguments. We compare several nonparametric estimators among the most used or documented. We find out that the currently used naive estimator is not consistent when the data are sparse like in phase II or III of clinical trials. A maximum likelihood estimator, with mild regularity assumptions, appears more satisfactory both in theory and in practice. However, it is difficult to compute : the likelihood is not explicit and has several local maxima : a stochastic EM algorithm is proposed to perform this optimization.

*Adresse :*

Julie ANTIC

Ecole Nationale Vétérinaire de Toulouse

23 Chemin des Capelles,

B.P. 87614, 31076 Toulouse cedex 3, France

E-mail : [j.antic@envt.fr](mailto:j.antic@envt.fr)

<<http://www.biostat.envt.fr/spip/spip.php?rubrique30>>

## Inference non-paramétrique dans un modèle d'évènements récurrents en présence d'un évènement terminal et de censure indépendante

par **Segolen Geffray**

Dans les études cliniques et épidémiologiques, des évènements récurrents peuvent survenir chez les individus. Les évènements ischémiques chez les patients présentant un risque vasculaire ou les récurrences de tumeurs chez les patients cancéreux sont des exemples classiques. Souvent, la survenue de ces évènements est associée à un risque élevé de décès de sorte que certains individus décèdent au cours de l'étude. Le décès est ici traité comme un risque concurrent dépendant qui s'exerce à chaque récurrence de la maladie. Dans ce contexte, nous nous intéressons aux fonctions de répartition spécifiques à une cause donnée correspondant à la durée entre deux évènements successifs conditionnellement au fait d'avoir survécu aux évènements antérieurs. Nous considérons également les fonctions de répartition multivariées correspondant aux durées entre plusieurs évènements successifs. Des estimateurs non-paramétriques pour toutes ces fonctions sont considérés en présence de censure droite indépendante. Des résultats de consistance forte et de convergence faible sont alors proposés. Les propriétés des estimateurs introduits sont évaluées à distance finie par simulations. La méthodologie proposée est enfin appliquée à des données réelles provenant d'un essai clinique comparant l'effet de deux médicaments dans la prévention des évènements ischémiques majeurs.

*Adresse :*

Segolen GEFFRAY

Faculté de Médecine, Université de Nantes

1 rue Gaston Veil,

BP 53508 44035 Nantes Cedex 1 France

E-mail : [Segolen.Geffray@univ-nantes.fr](mailto:Segolen.Geffray@univ-nantes.fr)

<<http://www.medecine.univ-nantes.fr>>

## **Modèle de régression pour la probabilité conditionnelle d'un risque concurrent**

par **Aurélien Latouche**, A. Allignol, J.P. Fine et J. Yan

En présence d'événements concurrents, nous présentons un modèle de régression pour la probabilité conditionnelle pour un événement de type  $k$  (Pepe et Mori 1993, *Statistics in Medicine*). Cette probabilité représente le probabilité d'avoir expérimenté l'événement de type  $k$  avant l'instant  $t$  sachant que l'on a expérimenté aucun des autres événements en compétition. Le modèle considéré est un modèle à odds proportionnel avec des effets dépendants du temps. L'estimation repose sur le modèle "temporal process regression" (Fine, Yan et Kosorok, *Biometrika* 2004). Nous présentons également un paquet R permettant l'estimation de l'effet de covariables et la représentation graphique des probabilités conditionnelles.

### *Adresses :*

Aurélien LATOUCHE

Faculté de Médecine, Université Versailles Saint-Quentin, EA 2506

E-mail : [aurelien.latouche@uvsq.fr](mailto:aurelien.latouche@uvsq.fr)

<http://www.uvsq.fr/>

A. ALIGNOL

J.P. FINE

J. YAN



## Sélection de Modèles

Session organisée par **Béatrice LAURENT**

L'idée de sélectionner un modèle à l'aide d'un critère de type maximum de vraisemblance pénalisé s'est développée dans les années 70 avec les travaux de Mallows et Akaike. Depuis une quinzaine d'années, Barron, Birgé et Massart [1], Birgé et Massart (voir par exemple [2], [3], et [6]) ont développé une théorie non asymptotique pour la sélection de modèles, qui s'appuie sur des résultats de concentration. Il s'agit de déterminer de manière précise le terme de pénalisation, en fonction du nombre de paramètres qui intervient dans chaque modèle, mais aussi de la complexité de la taille de la collection de modèles considérée, afin d'établir des inégalités de type "oracle". Ces inégalités prouvent que le modèle sélectionné se comporte presque aussi bien que le "meilleur" modèle de la collection (en terme de risque par exemple). On peut déduire des inégalités de type "oracle" que les estimateurs construits à partir du modèle sélectionné sont adaptatifs, c'est-à-dire atteignent les vitesses minimax d'estimation sur de nombreuses classes fonctionnelles simultanément. Parallèlement, les méthodes de seuillage d'ondelettes développées par Donoho et Johnstone [4], Donoho, Johnstone, Kerkycharian et Picard [5] fournissent également des procédures d'estimation adaptatives, qui peuvent dans certains cas s'interpréter comme des procédures de sélection de modèles via un critère pénalisé.

L'objectif de la session est de présenter quelques travaux récents sur la sélection de modèles et les méthodes de seuillage. Les exposés couvrent un large spectre allant de la théorie aux applications. La sélection de variables en régression gaussienne à variance inconnue, la sélection de graphe pour des réseaux biologiques, l'estimation dans des modèles gaussiens hétéroscédastiques et l'estimation de l'intensité d'un processus de Poisson de support inconnu seront les thèmes abordés au cours de la session.

### Références :

- [1] Barron, A.R., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Th. Rel. Fields.* **113**, 301-415
- [2] Birgé, L. and Massart, P. (1997) From model selection to adaptive estimation. *Festschrift for Lucien Lecam : Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds.) 55-87, Springer-Verlag, New-York.
- [3] Birgé, L. and Massart, P. (2001) Gaussian Model selection. *Journal of European Mathematical Society*, **3**, 203-268.
- [4] Donoho, D. and Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage. *JASA*, **90**, 1200-1224.
- [5] Donoho, D., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508-539.
- [6] Massart, P. (2003) Concentration Inequalities and Model Selection. *Cours de l'Ecole d'été de Saint-Flour*, Springer.

### Adresse de l'organisateur :

Béatrice LAURENT  
INSA de Toulouse  
Département de Génie Mathématique  
135 Avenue de Rangueil  
31077 Toulouse Cedex 4 (France)  
E-mail : [beatrice.laurent@insa-toulouse.fr](mailto:beatrice.laurent@insa-toulouse.fr)  
<<http://www-gmm.insa-toulouse.fr/~laurent/>>

## Sélection de modèle et sélection de variables en régression gaussienne à variance inconnue

par **Yannick BARAUD**, Christophe GIRAUD et Sylvie HUET

Nous observons un vecteur gaussien dont les composantes sont indépendantes et de même variance et cherchons à estimer son espérance par sélection de modèle. Nous ne supposons pas que la variance

des composantes est connue (ou convenablement estimée par un estimateur ad-hoc). Ainsi, nous nous intéresserons aux critères de choix de modèle de type FPE, AIC ou BIC, introduits de manière heuristique dans les années 70 et couramment utilisés depuis, qui ont l'avantage de ne pas dépendre de cette variance inconnue. Le but de cette exposé est double. D'abord présenter une analyse non-asymptotique de ces critères, puis en proposer de nouveaux pour le problème de la sélection de variables. Des simulations viendront illustrer les performances de ces nouveaux critères par rapport aux anciens et par rapport au LASSO adaptatif proposé par Zou en 2006.

*Adresses :*

Yannick BARAUD  
Laboratoire J-A Dieudonné  
Universite de Nice Sophia-Antipolis  
Parc Valrose  
06108 Nice cedex 02, France  
E-mail : Yannick.BARAUD@unice.fr  
<[http://math1.unice.fr/\\_baraud/](http://math1.unice.fr/_baraud/)>

Christophe GIRAUD  
Laboratoire J-A Dieudonné  
Universite de Nice Sophia-Antipolis  
Parc Valrose  
06108 Nice cedex 02, France  
E-mail : cgiraud@math.unice.fr  
<[http://math1.unice.fr/\\_cgiraud/](http://math1.unice.fr/_cgiraud/)>

Sylvie HUET  
Unité MIA INRA  
78352 Jouy-en-Josas Cedex,  
France  
E-mail : sylvie.huet@jouy.inra.fr

## **Estimation simultanée de la moyenne et de la variance en régression gaussienne hétéroscédastique**

par **Xavier GENDRE**

Nous nous intéressons à l'estimation simultanée de la moyenne et de la variance d'un échantillon gaussien. Pour cela, nous construisons une procédure de choix de modèle basée sur un critère de vraisemblance pénalisée. Nous présenterons une inégalité de type oracle et, sous quelques hypothèses, des résultats adaptatifs pour le risque minimax.

*Adresse :*

Xavier GENDRE  
Laboratoire J-A Dieudonné  
Universite de Nice Sophia-Antipolis  
Parc Valrose  
06108 Nice cedex 02, France  
E-mail : Xavier.GENDRE@unice.fr  
<[http://math1.unice.fr/\\_gendre/](http://math1.unice.fr/_gendre/)>

## **Estimation par seuillage de l'intensité d'un processus de Poisson à support inconnu ou infini**

par Patricia REYNAUD et Vincent RIVOIRARD



Le but de cet exposé est d'estimer l'intensité  $f$  d'un processus de Poisson à l'aide des réalisations de ce processus. En utilisant une décomposition par ondelettes de  $f$ , nous proposons une procédure par seuillage aléatoire qui ne suppose pas connu le support de  $f$ . En particulier, ce dernier peut être infini. Nous montrons l'optimalité de notre procédure dans le cadre des approches oracle et minimax. Ce faisant, nous établissons que les vitesses minimax sur les espaces de Besov  $B_{p,q}^\alpha$  ( $p \leq 2$ ) pour les fonctions à support infini sont les mêmes que pour les fonctions à support compact (à un éventuel terme logarithmique près). La deuxième partie de notre exposé porte sur la calibration optimale du seuil de notre procédure. Ce travail s'appuie sur des inégalités oracle établies pour des classes fonctionnelles particulières. Il est prolongé par des simulations réalisées sur une large gamme de signaux.

*Adresses :*

Patricia REYNAUD

DMA

Ecole Normale Supérieure

45 Rue d'Ulm,

75230 Paris Cedex 05, France

E-mail : [reynaud@dma.ens.fr](mailto:reynaud@dma.ens.fr)

<<http://www.dma.ens.fr/~reynaud/>>

Vincent RIVOIRARD

Laboratoire de Mathématiques d'Orsay et DMA

Université Paris-Sud 11

Bâtiment 425

91405 Orsay Cedex, France

E-mail : [Vincent.Rivoirard@math.u-psud.fr](mailto:Vincent.Rivoirard@math.u-psud.fr)

<<http://www.math.u-psud.fr/~rivoirar>>

## Sélection de graphe pour des réseaux bayésiens

par Nicolas VERZELEN

Récemment, les données provenant de la génomique ont fait naître un intérêt grandissant pour les problèmes d'estimation en grande dimension. Dans cette optique, on s'intéresse ici à l'estimation de la covariance d'un vecteur Gaussien de moyenne nulle. On considère  $X \sim \mathcal{N}(0_p, \Sigma)$  un vecteur gaussien de taille  $p$ . Notre objectif est d'estimer la matrice  $\Sigma$  en utilisant  $n$  réalisations indépendantes du vecteur  $X$ . On parle de problème en grande dimension lorsque  $n$  est supposé plus petit que  $p$ .

Une approche classique dans ce cadre est de supposer que la matrice de covariance  $\Sigma$  ou une fonctionnelle de  $\Sigma$  appartient ou *est proche* d'un bon espace vectoriel de faible dimension. Ainsi, selon le contexte il est commode de supposer (voir [3]) que la matrice  $\Sigma$  est creuse, que son inverse  $\Sigma^{-1}$  est creuse ou alors que  $K$  la décomposition de Cholevski de  $\Sigma^{-1}$  est creuse. Ces deux dernières hypothèses correspondent respectivement à la sparsité d'un modèle graphique gaussien associé à  $X$  et à la sparsité d'un réseau bayésien associé à  $X$  (voir [2]) pour plus de détails).

Après avoir rappelé les liens qui unissent le graphe d'un réseau bayésien associé à  $X$  et la matrice  $K$ , on s'appuiera sur des méthodes maintenant classiques de pénalisation (voir par exemple [1]) pour sélectionner le graphe du réseau bayésien et construire un estimateur  $\hat{K}$  de maximum de vraisemblance pénalisé. Plus précisément, on montrera sous quelle condition sur la pénalité, l'estimateur  $\hat{K}$  satisfait une inégalité oracle non asymptotique en distance de Kullback. On pourra ainsi en déduire les performances de critères classiques comme AIC ou BIC.

## Références

- [1] Y. Baraud, C. Giraud, S. Huet (2007), *Gaussian model Selection with unknown variance*, Annals of Statistics, à paraître.

- [2] D. M. Edwards (2000), *Introduction to graphical models*, Springer, New-York.
- [3] C. Lam and J. Fan (2007) *Sparsistency and Rates of Convergence in Large Covariance Matrices Estimation*, arXiv :math0711.3933.v1.

*Adresse :*

Nicolas VERZELEN  
Laboratoire de Mathématiques d'Orsay  
Université Paris-Sud 11  
Bâtiment 425  
91405 Orsay Cedex, France  
E-mail : [Nicolas.Verzelen@math.u-psud.fr](mailto:Nicolas.Verzelen@math.u-psud.fr)

## Inférence et structure cachée dans les réseaux biologiques

Session organisée par **Christophe Ambroise**

Les réseaux sont aujourd'hui au cœur des préoccupations des biologistes, qu'il s'agisse de réseaux d'interactions (par exemple entre protéines), de réseaux de régulation (quels gènes, quand il sont exprimés incitent ou inhibent quels autres gènes?), ou de voies métaboliques (quelles sont les cascades faisant intervenir successivement des composés chimiques).

A partir de données indiquant pour chaque composant (pour chaque nœud) une valeur numérique au cours de plusieurs expériences, indépendantes ou recueillies au cours d'une évolution temporelle, est-il possible d'inférer un graphe, dont les arêtes, simples ou valuées, décriront l'interaction entre ces composants? Des approches d'inférence statistique des réseaux fondées sur les modèles graphiques gaussiens [1] sont présentés dans cette session.

Lorsque des réseaux biologiques sont disponibles (après inférence statistique par exemple), leur description statistique permet d'approfondir la compréhension des processus biologiques sous-jacents. Cette étude statistique des réseaux permet par exemple de dégager des structures cachées expliquant des relations privilégiées entre nœuds du réseaux ou de mettre en évidence des formes de relation particulière (telles les boucles de rétroaction). L'emploi des modèles de mélange pour cette tâche descriptive semble prometteur [2].

### Références :

- [1] Meinshausen, N. and Bühlmann, P. (2006), High-dimensional graphs and variable selection with the lasso, *The Annals of Statistics* **34**, 1436–1462.
- [2] Daudin, J., Picard, F. and Robin, S. (2008). A mixture model for random graph. *Statistics and computing* **18**, 1–36.

### Adresse de l'organisateur :

Christophe AMBROISE  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
La genopole  
Tour Evry 2  
523 place des Terrasses  
91000 Evry, France  
E-mail : [Christophe.Ambroise@genopole.cnrs.fr](mailto:Christophe.Ambroise@genopole.cnrs.fr)  
<<http://stat.genopole.cnrs.fr/~cambroise/>>

## Maximum de vraisemblance pénalisé pour l'inférence de réseaux creux par modèles graphiques Gaussien à structure cachée

par **Christophe Ambroise**, Julien Chiquet et Catherine Matias

Les modèles graphiques gaussiens se sont avérés être un modèle bien adapté à l'inférence des graphes d'interaction biologiques. L'absence d'arête entre deux gènes est traduite par l'indépendance conditionnelle des expressions de ceux-ci conditionnellement à l'ensemble des autres. L'estimation de ces graphes de dépendances représente un défi pour le statisticien, qui dispose dans ce contexte, comme souvent en biologie, de moins d'observations que de variables.

Nous proposons un modèle Bayésien qui force la parcimonie du modèle (i.e. l'inférence d'un faible nombre d'arêtes). Ces procédures sont inspirées de la régularisation par norme  $L_1$  très populaire en régression linéaire.

En particulier, le modèle suppose l'existence d'une structure particulière sur le réseau. Cette structure fait l'hypothèse de l'existence de groupes fonctionnels, sans que ceux-ci soient connus au préalable. Ces deux a priori, graphe creux et structuré permettent d'obtenir une estimation des graphes de dépendance biologiquement plausible.

*Adresses :*

Christophe AMBROISE  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
La genopole  
Tour Evry 2  
523 place des Terrasses  
91000 Evry, France  
E-mail : [Christophe.Ambroise@genopole.cnrs.fr](mailto:Christophe.Ambroise@genopole.cnrs.fr)  
<<http://stat.genopole.cnrs.fr/~cambroise/>>

Julien CHIQUET  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
E-mail : [Julien.Chiquet@genopole.cnrs.fr](mailto:Julien.Chiquet@genopole.cnrs.fr)  
<<http://stat.genopole.cnrs.fr/~jchiquet/>>

Catherine MATIAS  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
E-mail : [Catherine.Matias@genopole.cnrs.fr](mailto:Catherine.Matias@genopole.cnrs.fr)  
<<http://stat.genopole.cnrs.fr/~cmatias/>>

## Test de validation de graphe

par **Fanny Villers** et Nicolas Verzelen

Les modèles graphiques gaussiens constituent des outils prometteurs pour l'analyse des réseaux biologiques. Dans de nombreuses applications, les biologistes ont une connaissance préalable du réseau et désirent examiner la qualité de leur modèle en utilisant des données d'expression. C'est pourquoi nous proposons un test de validation de graphe. Il s'agit à partir d'un jeu de données et d'un graphe minimal proposé par le biologiste, c'est à dire un graphe dans lequel toutes les arêtes correspondent à des interactions biologiques pertinentes, de tester s'il ne manque pas d'arêtes entre des sommets. Notre procédure est basée sur le test d'une hypothèse linéaire dans un modèle de régression multivariée, dont les variables explicatives sont aléatoires.

La procédure présentée permet de traiter des données en grande dimension. Nous contrôlons le niveau exact du test et montrons des résultats non asymptotiques sur la puissance du test.

*Adresses :*

Fanny VILLERS  
INRA  
INRA, Unité MIA  
78352 Jouy-en-Josas Cedex  
E-mail : [fanny.villers@jouy.inra.fr](mailto:fanny.villers@jouy.inra.fr)

Nicolas VERZELEN  
E-mail : [nicolas.verzelen@math.u-psud.fr](mailto:nicolas.verzelen@math.u-psud.fr)

## Méthodes Bayésiennes pour l'analyse de structures dans les réseaux

par **Pierre Latouche** et Etienne Birmelé

Les réseaux sont très largement utilisés en biologie, en sciences sociales, et dans les nouvelles technologies de l'information. Ils ont vocation à modéliser, à l'aide d'arêtes, les interactions entre objets d'intérêt. La recherche de classes, également appelées communautés ou modules, d'objets ayant des profils d'interactions similaires, est de plus en plus utilisée afin d'analyser la structure des réseaux.

Récemment, le modèle MixNet pour la détection de communautés a été proposé par Daudin et al. (2006) avec une méthode associée d'estimation par maximum de vraisemblance, basée sur des techniques variationnelles. L'algorithme permet l'estimation du nombre de classes dans un réseau et génère des probabilités d'appartenance des sommets aux différentes classes.

En nous appuyant sur les travaux d'Hofman et Wiggins (2007), nous montrons d'abord comment MixNet peut être décrit dans un contexte Bayésien. Nous présentons ensuite les équations, obtenues à l'aide d'une approche de type Variational Bayes, permettant l'optimisation des hyperparamètres du modèle. Enfin, en utilisant des données simulées et réelles, nous analysons la le nombre et la qualité des classes obtenues.

## Références

- Daudin, J. and Picard, F. and Robin, S. (2006) : A Mixture Model for Random Graphs. *Tech. rep, INRIA*.
- Hofman, J.M. and Wiggins, C.H. (2007) : A Bayesian Approach to Network Modularity. *ArXiv e-prints*.
- Jordan, M. and Ghahramani, Z. and Jaakkola, T. (1998) : An introduction to variational methods for graphical models. In : Jordan, M. : *Learning in Graphical Models*. MIT Press.

### Adresses :

Pierre LATOUCHE  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
La genopole  
Tour Evry 2  
523 place des Terrasses  
91000 Evry, France  
E-mail : [Pierre.Latouche@genopole.cnrs.fr](mailto:Pierre.Latouche@genopole.cnrs.fr)

Etienne BIRMELE  
Laboratoire Statistique et Génome (UMR CNRS 8071, INRA 1152)  
E-mail : [Etienne.Birmele@genopole.cnrs.fr](mailto:Etienne.Birmele@genopole.cnrs.fr)  
<<http://stat.genopole.cnrs.fr/~ebirmele/>>

## Grphe aléatoire à structure latente et covariables

par **Hugo Zanghi**

Dans le contexte des graphes aléatoires, le modèle de mélange MixNet suppose un mélange de distribution sur les arêtes et une structure cachée sur les nœuds et semble une bonne alternative aux modèles existant. En particulier, la distribution des degrés résultante s'avère très proche des distributions réelles observées. La structure cachée reflète l'existence de modules, ou de nœuds possédant des comportement similaires en terme de connections. L'estimation des paramètres du modèle permet donc de trouver ces modules et de synthétiser les informations sur les types d'interactions.

Dans de nombreuses situations, des informations additionnelles sur les nœuds et arêtes sont disponibles : position du gène sur le chromosome, séquence des bases, expression du gène dans certaines conditions, co-expression entre deux gènes, fonction de la protéine...

Nous proposons une extension du modèle MixNet, prenant en compte ces informations additionnelles. Les informations disponibles aux nœud sont considérés comme des covariables et participer à l'estimation de la structure cachée.

Journées MAS 2008, Rennes

*Adresse :*

Hugo ZANGHI

Exalead

Exalead

10 place de la Madeleine

75008 Paris, France

E-mail : [Hugo.Zanghi@exalead.com](mailto:Hugo.Zanghi@exalead.com)

## Probabilités non-commutatives et algèbres de von Neumann de groupes

Session organisée par **Dimitri Petritis**

La mécanique quantique, introduite dans les années 30 pour décrire le comportement de systèmes (sub-) atomiques, peut être vue mathématiquement comme une généralisation non-commutative de la théorie des probabilités où les variables aléatoires réelles sont remplacées par des opérateurs auto-adjoints sur un espace de Hilbert adéquat, les probabilités par des états sur l'algèbre des opérateurs, la loi des variables aléatoires (loi image d'un état) par la mesure obtenue par l'application de l'état sur la mesure spectrale de l'opérateur. Ce formalisme, introduit par von Neumann [2], est aussi décrit dans les textes plus récents [4, 3, 1].

En faisant abstraction de la structure hilbertienne sous-jacente, on peut formuler abstraitement une probabilité non-commutative comme un état tracial sur une algèbre d'opérateurs  $\mathcal{A}$ . Même en se limitant au cas des algèbres de von Neumann, l'algèbre d'opérateurs est un objet très compliqué. Il y a cependant des situations où l'algèbre découle d'un objet plus simple, ce dernier permettant de déterminer une partie significative des propriétés de l'algèbre. Un exemple de cette situation simplificatrice est le cas où l'algèbre  $\mathcal{A}$  est engendrée par la représentation d'un groupe (localement compact). Or physiquement les groupes correspondent à des transformations agissant sur le système; leurs actions (egodiques) fournissent donc des exemples d'espaces de probabilité non-commutatifs. L'exposé de Bachir BEKKA, après une introduction générale au sujet, montre la puissance de la description algébrique et son apport à l'étude des groupes et des leurs actions; l'exposé de Jean-Romain HEU complète l'exposé précédent en présentant des résultats spécifiques associés à la dynamique sur les nilvariétés.

En se souvenant que les états sur les algèbres jouent le rôle des probabilités, les transformations physiques que les états peuvent subir doivent préserver leur nature probabiliste. La notion classique d'opérateur de Markov est remplacé ici par la notion de transformation complètement positive; ces transformations correspondent aux modifications les plus générales (canaux quantiques) que l'on peut effectuer sur les états d'un système en préservant sa positivité, même lorsque le système est couplé à un environnement arbitraire. Les exposés de Nadine GUILLOTIN et de Bunrith-Jacques LIM étudient le comportement asymptotique des états lorsqu'ils subissent l'action répétée d'un canal quantique.

### Références :

- [1] Meyer, Paul-André, *Quantum probability for probabilists*, Lecture Notes in Mathematics 1538, Springer-Verlag, Berlin 1993.
- [2] von Neumann, Johann, *Mathematische Grundlagen der Quantenmechanik*, Unveränderter Nachdruck der ersten Auflage von 1932. Die Grundlehren der mathematischen Wissenschaften, Band 38, Springer-Verlag, Berlin, 1968.
- [3] Parthasarathy, K. R., *An introduction to quantum stochastic calculus*, Monographs in Mathematics, 85, Birkhäuser Verlag, Basel, 1992.
- [4] Varadarajan, V. S., *Geometry of quantum theory*, Second edition, Springer-Verlag, New York, 1985.

### Adresse de l'organisateur :

Dimitri PETRITIS  
Institut de recherche mathématique de Rennes  
Université de Rennes 1  
35042 Rennes Cedex, France  
E-mail : [dimitri.petritis@univ-rennes1.fr](mailto:dimitri.petritis@univ-rennes1.fr)  
<<http://perso.univ-rennes1.fr/dimitri.petritis/>>

## Probabilités non commutatives et algèbres de von Neumann de groupes

par **Bachir Bekka**

En mécanique quantique, les fonctions sur un espace  $\Omega$  sont remplacés par des opérateurs sur un espace de Hilbert. Le rôle de l'espace de probabilités fondamental  $(\Omega, \mathbf{P})$  est alors joué par une algèbre d'opérateurs (algèbre de von Neumann)  $\mathcal{A}$  munie d'une trace  $\tau$ . Le couple  $(\mathcal{A}, \tau)$  est appelé espace de probabilités non commutatif. Dans les vingt dernières années, de nombreuses notions et résultats classiques (indépendance de variables aléatoires, TCL, ...) ont été généralisés au cas non commutatif. D'autre part, les groupes (non commutatifs) ou leurs actions ergodiques fournissent des exemples significatifs d'espaces de probabilités non commutatifs. Le but de l'exposé sera de présenter le sujet et de montrer son apport à l'étude des groupes et actions de groupes ainsi qu'à celui du comportement de grandes matrices aléatoires.

*Adresse :*

Bachir BEKKA  
Institut de recherche mathématique de Rennes  
Université de Rennes 1  
35042 Rennes Cedex, France  
E-mail : [bachir.bekka@univ-rennes1.fr](mailto:bachir.bekka@univ-rennes1.fr)  
<<http://perso.univ-rennes1.fr/bachir.bekka/>>

## Propriétés statistiques des matrices de Pauli traversant des canaux bruités

par **Nadine Guillotin**

Nous étudions les propriétés statistiques du triplet  $(\sigma_x, \sigma_y, \sigma_z)$  des matrices de Pauli traversant une suite de canaux bruités, modélisés par la répétition d'une application complètement positive et préservant la trace. Plus précisément, un théorème central limite non-commutatif pour la loi du triplet ainsi qu'un principe de grandes déviations sont présentés.

*Adresse :*

Nadine GUILLOTIN  
Institut Camille Jordan  
Université de Lyon 1  
69000 Lyon, France  
E-mail : [Nadine.Guillotin@univ-lyon1.fr](mailto:Nadine.Guillotin@univ-lyon1.fr)  
<<http://igd.univ-lyon1.fr/~guillotin/>>

## Dynamique dans les nilvariétés

par **Jean-Romain Heu**

En nous appuyant sur des résultats relatifs aux tores, nous aborderons certaines questions de dynamique dans les nilvariétés, et plus particulièrement dans les nilvariétés associées aux groupes de Heisenberg.

*Adresse :*

Jean-Romain HEU  
Institut de recherche mathématique de Rennes  
Université de Rennes 1  
35042 Rennes Cedex, France  
E-mail : [jean-romain.heu@univ-rennes1.fr](mailto:jean-romain.heu@univ-rennes1.fr) web



## **Canaux quantiques agissant sur des matrices densité**

par **Bunrith-Jacques Lim**

Nous étudierons le lien entre les mesures répétées sur un système quantique et l'action des canaux quantiques sur des opérateurs positifs de classe trace.

*Adresse :*

Bunrith-Jacques LIM  
Institut de recherche mathématique de Rennes  
Université de Rennes 1  
35042 Rennes Cedex, France  
E-mail : [bunrith.lim@univ-rennes1.fr](mailto:bunrith.lim@univ-rennes1.fr)



## Simulation de processus de diffusion

Session organisée par **Benjamin Jourdain**

Les processus de diffusion sont utilisés comme modèles dans un grand nombre de domaines d'applications (finance, géophysique,...). C'est pourquoi l'obtention de procédures de simulation efficaces de ces processus est un enjeu très important. Ce domaine de recherche très actif a connu des développements récents très intéressants parmi lesquels :

- la simulation d'équations différentielles stochastiques à coefficients discontinus [4, 8, 9, 11],
- le développement de procédures de simulation exacte en loi [1, 2, 3, 5],
- le développement de schémas d'ordre de convergence faible élevé : une première possibilité consiste à remplacer les intégrales browniennes itérées que l'on ne sait pas simuler par des variables aléatoires prenant un nombre fini de valeurs et partageant les mêmes moments jusqu'à un certain ordre [6, 7, 10, 12, 13]. Une seconde alternative repose sur l'intégration d'équations différentielles ordinaires sur des horizons aléatoires bien choisis [14, 15].

Ces développements seront abordés dans les exposés de la session.

### Références :

- [1] Beskos, A., Papaspiliopoulos, O. and Roberts, G. : Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* 12(6) :1077-1098, 2006.
- [2] Beskos, A., Papaspiliopoulos, O., Roberts, G. and Fearnhead, P. : Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68(3)333-382, 2006.
- [3] Beskos, A. and Roberts, G. : Exact simulation of diffusions. *Ann. Appl. Probab.* 15(4) :2422-2444, 2005.
- [4] Étoré, P. and Lejay, A. : A Donsker theorem to simulate one-dimensional processes with measurable coefficients. *ESAIM Probab. Stat.* 11 :301-326, 2007.
- [5] Jourdain, B. and Sbairi, M. : Exact retrospective Monte Carlo computation of arithmetic average Asian options, *Monte Carlo Methods Appl.* 13(2) :135-171, 2007.
- [6] Kusuoka, S. : Approximation of expectation of diffusion process and mathematical finance. *Taniguchi Conference on Mathematics Nara '98*, 147-165, *Adv. Stud. Pure Math.*, 31, Math. Soc. Japan, Tokyo, 2001.
- [7] Kusuoka, S. : Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. *Advances in mathematical economics*. Vol. 6, 69-83, 2004.
- [8] Lejay, A. : Stochastic differential equations driven by processes generated by divergence form operators. I. A Wong-Zakai theorem. *ESAIM Probab. Stat.* 10 :356-379, 2006.
- [9] Lejay, A. and Martinez, M. : A scheme for simulating one-dimensional diffusion processes with discontinuous coefficients. *Ann. Appl. Probab.* 16(1) :107-139, 2006.
- [10] Lyons, T. and Victoir, N. : Cubature on Wiener space. *Stochastic analysis with applications to mathematical finance*. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 460 :169-198, 2004.
- [11] Martinez, M. and Talay, D. : Discrétisation d'équations différentielles stochastiques unidimensionnelles à générateur sous forme divergence avec coefficient discontinu. *C. R. Math. Acad. Sci. Paris* 342(1) :51-56, 2006.
- [12] Ninomiya, S. : A new simulation scheme of diffusion processes : application of the Kusuoka approximation to finance problems. *Math. Comput. Simulation* 62 :479-486, 2003.
- [13] Ninomiya, S. : A partial sampling method applied to the Kusuoka approximation. *Monte Carlo Methods Appl.* 9(1) :27-38, 2003.
- [14] Ninomiya, S. and Victoir, N. : Weak approximation of stochastic differential equations and application to derivative pricing, *Appl. Math. Finance*, 15(2) :107-121, 2008. 2005.
- [15] Ninomiya, M. and Ninomiya, S. : A new weak approximation scheme of stochastic differential equations by using the Runge-Kutta method, preprint, 2008.

*Adresse de l'organisateur :*

Benjamin JOURDAIN  
CERMICS  
École des Ponts  
6 et 8 avenue Blaise Pascal  
77420 Champs sur Marne. France  
E-mail : [jourdain@cermics.enpc.fr](mailto:jourdain@cermics.enpc.fr)  
<<http://cermics.enpc.fr/~jourdain/>>

## Processus à coefficients discontinus en dimension un : simulations et applications

par Pierre Étoré, **Antoine Lejay** et Miguel Martinez

Les processus de diffusion à coefficients discontinus, avec un générateur infinitésimal sous forme divergence ou non-divergence, ont de nombreuses applications potentielles en écologie, physique, géophysique, médecine, finance ... car ils peuvent servir à modéliser des phénomènes de « barrières perméables » et des interfaces entre des milieux de caractéristiques très différentes.

En dimension un, il est possible de représenter ces processus comme solution d'équations différentielles stochastiques (EDS) faisant intervenir le temps local du processus aux discontinuités.

Nous présenterons donc divers modèles ainsi que divers algorithmes permettant de simuler ces processus. Ces algorithmes reposent sur une analyse des propriétés de ces EDS, et nous mettrons en évidence les différences et difficultés par rapport au cas où les coefficients sont réguliers.

### *Adresses :*

Pierre ÉTORÉ  
CMAP  
Ecole Polytechnique  
Route de Saclay  
91128 Palaiseau Cedex  
E-mail : [etore@cmmap.polytechnique.fr](mailto:etore@cmmap.polytechnique.fr)  
<<http://www.cmapx.polytechnique.fr/~etore>>

Antoine LEJAY  
INRIA Nancy Grand-Est  
IECN  
Campus scientifique  
BP 239  
54506 Vandoeuvre-lès-Nancy CEDEX  
E-mail : [Antoine.Lejay@iecn.u-nancy.fr](mailto:Antoine.Lejay@iecn.u-nancy.fr)  
<<http://www.iecn.u-nancy.fr/~lejay/>>

Miguel MARTINEZ  
LAMA  
Université Paris-Est Marne-la-Vallée  
5 Bd Descartes  
Cité Descartes, Champs-sur Marne  
77454 Marne-la-Vallée Cedex 2  
E-mail : [miguel.martinez@univ-mlv.fr](mailto:miguel.martinez@univ-mlv.fr)

## Pratique de l'extrapolation dans les schémas de discrétisation des équations différentielles stochastiques

par **Julien Guyon**

Dans les salles de marchés, on estime le prix et la couverture de produits dérivés. La première étape, cruciale, consiste à identifier les facteurs de risque liés à un produit et à modéliser leur dynamique jointe. Dans la seconde étape, on vient évaluer numériquement le prix, la stratégie de couverture et leurs sensibilités aux paramètres du modèle. Pour ce faire, on est parfois ramené au problème suivant : estimer par méthode de Monte Carlo la valeur moyenne  $E[f(X_{t_1}, \dots, X_{t_N})]$  où  $X$  est un processus diffusif multidimensionnel. L'estimation doit être rapide et robuste. Des schémas de discrétisation d'ordre élevé (deux ou plus) tendent à l'accélérer, en réduisant la taille de la grille de discrétisation temporelle qui, à nombre de trajectoires simulées donné, suffit à plonger l'estimateur au coeur de l'erreur Monte Carlo. Une extrapolation de schémas d'ordre un a, théoriquement, les mêmes effets et, pratiquement, la vertu de la simplicité. On présente ici des modalités d'extrapolation et les résultats de leur implémentation.

*Adresse :*

Julien GUYON  
Société générale, Equity Derivatives Quantitative Research, Paris  
GEDS/DAI/GSD/QTR  
17, cours Valmy  
92 987 Paris-La Défense, France  
E-mail : [julien.guyon@sgcib.com](mailto:julien.guyon@sgcib.com)

## Estimation de l'erreur de localisation pour les équations de HJB

par **Junbo Huang**

De nombreux problèmes de contrôle stochastique en finance, notamment les problèmes d'optimisation de portefeuille, sont liés à un type d'équation aux dérivées partielles non linéaires du second ordre, appelées EDP de Hamilton-Jacobi-Bellman (HJB). Malheureusement, ces équations ne peuvent être résolues explicitement que dans peu de cas. Donc le recours aux méthodes numériques est inévitable. Mais afin d'approcher numériquement la solution d'une équation de HJB, cette équation doit être localisée dans un domaine borné et il faut pour cela comprendre le comportement au bord de cette équation. Comme de nombreuses équations sont obtenues sur un domaine non borné, deux questions se posent naturellement : quelles sont les conditions au bord pertinentes pour l'équation de HJB localisée ? Et quelle est l'erreur de localisation si les conditions au bord ne sont pas données précisément ?

Pour l'équation de HJB obtenue à partir du problème de maximisation d'utilité de portefeuille en horizon fini, une approximation des conditions au bord peut être obtenue en utilisant la méthode de martingale et l'algorithme de Robbins-Monro. Quant à la deuxième question, nous montrons que grâce à l'interprétation probabiliste des équations de HJB, l'erreur de localisation est contrôlée par l'erreur au bord et la probabilité maximale pour que la valeur du portefeuille atteigne le bord.

*Adresse :*

Junbo HUANG  
Projet TOSCA  
INRIA Sophia Antipolis-Méditerranée  
2004, Route des Lucioles  
B.P. 93  
06902 Sophia Antipolis Cedex  
France  
E-mail : [junbo.huang@sophia.inria.fr](mailto:junbo.huang@sophia.inria.fr)

## Méthodes de simulation exacte. Application au calcul du prix d'options asiatiques.

par Benjamin Jourdain et **Mohamed Sbai**

Nous présentons l'algorithme de Beskos, Papaspiliopoulos et Roberts [1] qui permet de simuler de manière exacte les trajectoires de la solution d'une équation différentielle stochastique unidimensionnelle. Dans plusieurs situations, comme en finance, on est seulement intéressé par le calcul de l'espérance d'une fonction de la solution en un instant donné. Dans cette optique, nous considérons également une méthode de calcul exact d'espérances dont les idées remontent au travail de Wagner sur les estimateurs Monte Carlo sans biais de certaines intégrales fonctionnelles (Journal of Computational Physics 71, 1987). En s'inspirant de ces deux méthodes, nous proposons une méthode de simulation exacte pour le calcul du prix d'options asiatiques dans le cadre du modèle Black and Scholes. En comparaison avec les autres méthodes de Monte Carlo, nous évitons ainsi le biais provenant de la discrétisation en temps. Nous présentons aussi quelques résultats numériques et nous considérons au passage des techniques de réduction de variance afin d'améliorer notre estimation du prix.

*Adresses :*

Benjamin JOURDAIN  
CERMICS  
École des Ponts  
6 et 8 avenue Blaise Pascal  
77420 Champs sur Marne. France  
E-mail : [jourdain@cermics.enpc.fr](mailto:jourdain@cermics.enpc.fr)  
<<http://cermics.enpc.fr/~jourdain/>>

Mohamed SBAI  
CERMICS  
École des Ponts  
6 et 8 avenue Blaise Pascal  
77420 Champs sur Marne. France  
E-mail : [sbai@cermics.enpc.fr](mailto:sbai@cermics.enpc.fr)  
<<http://cermics.enpc.fr/~sbai/home.html>>

## Simulation exacte des grecques et des prix en dim 1.

par Victor Reutenauer et **Etienne Tanré**

Dans un travail récent, Beskos, Papaspiliopoulos et Roberts [1] décrivent une méthode de simulation exacte des solutions d'EDS unidimensionnelles dont le coefficient de diffusion est constant. Leur algorithme repose sur une procédure de rejet de trajectoires que l'on sait simuler sans erreur de discrétisation (pont brownien, pont de Bessel de dimension 3 partant de 0). Dans ce travail, nous modifions leur méthode de simulation afin de rejeter plus rapidement les trajectoires; nous relâchons un peu leurs hypothèses en proposant une méthode de simulation quasi-exacte sous des hypothèses un peu plus faibles. Enfin, grâce au calcul de Malliavin, nous étendons la méthode au calcul des dérivées par rapport à la condition initiale de prix d'options (calcul du Delta et du Gamma).

Nous présentons les résultats numériques pour le modèle de taux de Cox Ingersoll Ross (CIR).

*Adresses :*

Victor REUTENAUER  
Calyon, FIM, Interest Rates and Hybrid Quantitative Research  
9, quai du Président Paul Doumer.  
92920 Paris-La Défense cedex  
E-mail : [Victor.Reutenauer@calyon.com](mailto:Victor.Reutenauer@calyon.com)

Journées MAS 2008, Rennes

Etienne TANRÉ  
INRIA, EPI Tosca  
2004, route des Lucioles, BP 93  
06902 Sophia-Antipolis Cedex  
E-mail : [Etienne.Tanre@sophia.inria.fr](mailto:Etienne.Tanre@sophia.inria.fr)  
<<http://www-sop.inria.fr/members/Etienne.Tanre/>>





## Traitement du signal et Statistique

Session organisée par **Erwan Le Pennec**

Les interactions entre le traitement du signal et les statistiques sont naturelles. L'un a besoin de méthodes statistiques efficaces dès lors que la modélisation est probabiliste tandis que l'autre y trouve des applications naturelles. L'objet de cette session est de présenter des travaux récents à l'interface entre ces deux disciplines.

Il sera en particulier question du problème de la représentation des données. L'efficacité des traitements statistiques (estimation, détection,...) est en effet conditionné par l'utilisation d'une représentation (description) des signaux adaptés à ceux-ci mais également au problème étudié. Cette problématique sera étudiée à travers les exemples proposés par les orateurs : finance, images, courbes et signaux sur la sphère.

*Adresse de l'organisateur :*

Erwan LE PENNEC  
LPMA / Université Paris Diderot  
Chevaleret  
75013 PARIS  
E-mail : [lepenec@math.jussieu.fr](mailto:lepenec@math.jussieu.fr)  
<<http://www.math.jussieu.fr/~lepenec>>

### Pas reçu

par **Emmanuel Bacry**

Pas reçu.

*Adresse :*

Emmanuel BACRY  
CMAP / Ecole Polytechnique  
E-mail : [bacry@cmap.polytechnique.fr](mailto:bacry@cmap.polytechnique.fr)  
<<http://www.cmap.polytechnique.fr/~bacry>>

### Statistical M-Estimation and Consistency in large deformable models for Image Warping

par Jérémie Bigot, **Sébastien Gadat** et Jean-Michel Loubes

The problem of defining appropriate distances between shapes or images and modelling the variability of natural images by group transformations is at the heart of modern image analysis. Results about the estimation of template images when the data are observed in a random framework are scarce and most of the existing results are stated in a deterministic setting. In this paper, we consider a set of images randomly warped from a mean template which has to be recovered. For this, we define an appropriate statistical parametric model to generate random diffeomorphic deformations in two-dimensions. Then, we focus on the problem of estimating the mean pattern when the images are observed with noise. This problem is challenging both from a theoretical and a practical point of view. M-estimation theory enables us to build an estimator defined as a minimizer of a well-tailored empirical criterion. We prove the convergence of this estimator and propose an iterative algorithm to compute this M-estimator in practice. Simulations of template extraction and an application to image classification and clustering are also provided.

*Adresses :*

Jérémie BIGOT

Institut de Mathématique de Toulouse / UPS

Université Paul Sabatier

E-mail : [bigot@mat.ups-tlse.fr](mailto:bigot@mat.ups-tlse.fr)

<<http://www.lsp.ups-tlse.fr/Fp/Bigot/>>

Sébastien GADAT

Institut de Mathématique de Toulouse / UPS

Université Paul Sabatier

E-mail : [gadat@mat.ups-tlse.fr](mailto:gadat@mat.ups-tlse.fr)

<<http://www.lsp.ups-tlse.fr/Fp/Gadat/>>

Jean-Michel LOUBES

Institut de Mathématique de Toulouse / UPS

Université Paul Sabatier

E-mail : [loubes@mat.ups-tlse.fr](mailto:loubes@mat.ups-tlse.fr)

<<http://www.lsp.ups-tlse.fr/Fp/Loubes/>>

## Applications des estimateurs d'excès de masse

par Mathilde Mougeot

*Adresse :*

Mathilde MOUGEOT

LPMA / Université Paris X

Université Paris X

E-mail : [mathilde.mougeot@u-paris10.fr](mailto:mathilde.mougeot@u-paris10.fr)

## Estimation spectrale sur la sphère avec des ondelettes

par Frédéric Guillaux et Gilles Faÿ

On s'intéresse à un processus aléatoire indexé par la sphère. S'il est stationnaire, l'analyse harmonique sur la sphère permet de définir son spectre de puissance. On cherche à estimer ce spectre à partir d'une unique observation du processus. L'exemple qui motive ce travail est l'étude du Fond Diffus Cosmologique, qui est l'un des grands thèmes de la physique actuelle [1], comme en témoigne l'attribution de deux prix Nobel en 30 ans.

Une classe d'ondelettes de seconde génération, les needlets, a été développée sur la sphère et ses propriétés statistiques étudiées [2,3,4]. Son utilisation permet une estimation efficace du spectre de puissance, dans un contexte réaliste comprenant des données manquantes, un bruit hétéroscédastique, et la fusion d'observations indépendantes [5].

Références :

1. Three-year WMAP observations : Temperature analysis (2007) Hinshaw et al., *Astrophys. J.*
2. Localized tight frames on spheres (2006) Narcowich, Petrushev, Ward, *SIAM J. Math. Anal.*
3. Asymptotics for spherical needlets (2008) Baldi, Kerkycharian, Marinucci, Picard, *Ann. Statist.* (à paraître)
4. Practical wavelet design on the sphere (2008) Guillaux, Faÿ, Cardoso, *Appl. Comp. Harmon. Anal.* (à paraître)
5. CMB Power spectrum estimation using wavelets (2008) Faÿ, Guillaux et al. (preprint)

Journées MAS 2008, Rennes

*Adresses :*

Frédéric GUILLOUX

Laboratoire AstroParticule et Cosmologie / Université Paris Diderot adresse

Université Paris X

E-mail : [guilloux@math.jussieu.fr](mailto:guilloux@math.jussieu.fr)

<<http://www.math.jussieu.fr/~guilloux>>

Gilles FAY

Laboratoire AstroParticule et Cosmologie / Université Paris Diderot adresse

Université Paul Sabatier

E-mail : [Gilles.Fay@univ-lille1.fr](mailto:Gilles.Fay@univ-lille1.fr)

<<http://math.univ-lille1.fr/~fay>>



## Statistiques bayésiennes : théorie et méthodologie

Session organisée par **Judith Rousseau**

Dans cette session nous proposons de parcourir certains aspects théoriques des statistiques bayésiennes et plus particulièrement des statistiques bayésiennes non paramétriques : Les théorèmes de Bernstein-Von Mises sont utiles en particulier pour comparer les régions de confiance fréquentistes avec les régions de crédibilité bayésiennes, les résultats théoriques dans les modèles mal spécifiés sont mal connus. D'un point de vue méthodologique J.M. Marin aborde un des grands thèmes porteurs des statistiques : les choix de modèles et S. Donnet présente un travail dans un domaine moins connu de la statistique bayésienne en France bien qu'extrêmement important en pratique, à savoir la construction de lois a priori à l'aide de dires d'experts.

*Adresse de l'organisateur :*

Judith ROUSSEAU  
CEREMADE, UMR 7534  
Université Paris Dauphine  
Place du Maréchal deLattre de Tassigny,  
75016 Paris, France  
E-mail : [rousseau@ceremade.dauphine.fr](mailto:rousseau@ceremade.dauphine.fr)  
<<http://www.ceremade.dauphine.fr/~rousseau/>>

### Choix bayésiens de modèles dans les modèles graphiques

par **Jean Michel Marin**

Soit  $f(y|\theta)$  la vraisemblance du modèle paramétrique considéré,  $\theta \in \Theta$  étant un paramètre inconnu. Dans le paradigme bayésien,  $\theta$  est considéré comme une quantité aléatoire de densité  $\pi(\theta)$ , dite loi a priori sur  $\theta$ . L'estimation bayésienne est basée sur le calcul de la loi a posteriori de  $\theta$ , la loi conditionnelle de  $\theta$  sachant  $y$  dont la densité est notée  $\pi(\theta|y)$ . D'après le théorème de Bayes, nous avons

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta).$$

Considérons deux modèles bayésiens paramétriques  $\mathcal{M}_1$  et  $\mathcal{M}_2$  où

$$\mathcal{M}_i : x \sim f_i(y|\theta_i), \theta_i \in \Theta_i, \theta_i \sim \pi_i(\theta_i)$$

et munissons l'espace des modèles d'une loi de probabilité a priori. Soient  $\mathbb{P}(\mathcal{M}_1)$  et  $\mathbb{P}(\mathcal{M}_2)$  les probabilités a priori des deux modèles, un choix de modèle bayésien est basé sur la loi a posteriori des différents modèles, ie sur les probabilités  $\mathbb{P}(\mathcal{M}_1|y)$  et  $\mathbb{P}(\mathcal{M}_2|y)$  où

$$\mathbb{P}(\mathcal{M}_i|y) \propto \int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i\mathbb{P}(\mathcal{M}_i).$$

Typiquement, si  $\mathbb{P}(\mathcal{M}_1|y) > \mathbb{P}(\mathcal{M}_2|y)$ , on choisira le modèle 1. Mettons en lumière quelques difficultés associées à cette méthodologie.

- La distribution a posteriori des modèles est très sensible au choix des lois a priori des paramètres des modèles ( $\pi_1(\theta_1)$  et  $\pi_2(\theta_2)$ ). Aussi, dans le cas où nous disposons d'informations a priori, il est important que ces lois a priori soient équitables. C'est un problème difficile très peu étudié.
- Il n'est pas possible d'utiliser des lois a priori impropres, distributions dont la densité est d'intégrale infinie. Dans ce cas, les probabilités a posteriori des modèles ne sont définies qu'à une constante arbitraire près. Ainsi, lorsque l'on ne dispose pas d'information a priori, il est difficile de mettre la méthodologie bayésienne de choix de modèles. C'est un problème qui a été beaucoup étudié mais, dans de nombreux cas, les réponses apportées ne sont pas satisfaisantes.
- Pour des modèles complexes, nous ne pouvons pas calculer explicitement  $\int_{\Theta_i} f_i(y|\theta_i)\pi_i(\theta_i)d\theta_i$ . Il convient alors de déterminer une méthode d'approximation raisonnable de cette quantité.

- Enfin, lorsque le nombre de modèles en compétition est très important, il n'est pas possible de calculer explicitement la loi a posteriori des modèles. L'exploration de l'espace des modèles peut alors s'avérer très difficile.

Dans cette présentation, nous abordons toutes ces difficultés et montrons comment elles peuvent être surmontées pour différents types de modèles.

*Adresse :*

Jean Michel MARIN  
INRIA, Université Paris - Sud  
Université Paris-Sud,  
91 940 Orsay, France  
E-mail : [jean-michel.marin@inria.fr](mailto:jean-michel.marin@inria.fr)

## Combinaison de dires de différents experts dans l'élicitation d'une loi a priori

par Isabelle Albert, **Sophie Donnet**, Chantal Guihenneuc, Samantha Low-Choi, Kerrie Mengersen et Judith Rousseau

Dans ce travail nous nous intéressons à l'élicitation de lois a priori, dans une modélisation bayésienne à partir des connaissances que peuvent fournir un ou plusieurs experts. L'élicitation de lois a priori par dires d'experts a été largement étudiée dans le passé, tant d'un point de vue psychologique que d'un point de vue statistique. Il s'agit d'un problème, qui s'il est conceptuellement simple est difficile à mettre en oeuvre. Ici nous nous intéressons plus particulièrement à la construction de lois a priori en utilisant les connaissances de plusieurs experts. Il s'agit alors de combiner ces différentes connaissances de manière cohérente et permettant de prendre en compte les interactions possibles entre les experts. Nous proposons une modélisation hiérarchique. Nous appliquons notre approche dans deux problèmes différents. Le premier concerne une modélisation dose-réponse pour la listéria, sur les souris. Le second modélise la durée d'une thèse en mathématiques appliqués dans une université australienne.

*Adresses :*

Isabelle ALBERT  
INRA, Unité Mét@risk, AgroParisTech adresse  
Université Paris-Sud,  
91 940 Orsay, France  
E-mail : [albert@inapg.inra.fr](mailto:albert@inapg.inra.fr)

Sophie DONNET  
CEREMADE, UMR 7534  
Université Paris Dauphine  
Place du Maréchal de Lattre de Tassigny,  
75016 Paris, France  
E-mail : [Donnet@ceremade.dauphine.fr](mailto:Donnet@ceremade.dauphine.fr)

Chantal GUIHENNEUC  
Laboratoire MAP5, UMR 8145 et INSERM  
Université Paris-René Descartes  
45, rue des Saint-Pères,  
75270 Paris Cedex 06, France  
E-mail : [chantal.guihenneuc@univ-paris5.fr](mailto:chantal.guihenneuc@univ-paris5.fr)

Samantha LOW-CHOI  
School of Mathematical Sciences, QUT  
Queensland University of Technology,  
Brisbane, Queensland  
E-mail : [s.lowchoy@qut.edu.au](mailto:s.lowchoy@qut.edu.au)

Kerrie MENGERSEN  
School of Mathematical Sciences, QUT  
Queensland University of Technology,  
Brisbane, Queensland  
E-mail : [k.mengersen@qut.edu.au](mailto:k.mengersen@qut.edu.au)

Judith ROUSSEAU  
CEREMADE, UMR 7534  
Université Paris Dauphine  
Place du Maréchal deLattre de Tassigny,  
75016 Paris, France  
E-mail : [rousseau@ceremade.dauphine.fr](mailto:rousseau@ceremade.dauphine.fr)

## Un théorème de Bernstein-von Mises semi-paramétrique avec a priori Gaussiens

par **Ismael Castillo**

Dans un modèle semi-paramétrique d'inconnues  $\theta, f$ , attribuons une mesure de probabilité a priori au paramètre  $\theta$  ainsi qu'à la fonction  $f$ . Le conditionnement de cette probabilité a priori en  $(\theta, f)$  par les données définit une mesure aléatoire, l'a posteriori. Le théorème de Bernstein-von Mises s'intéresse à la convergence de la marginale en  $\theta$  de cet a posteriori vers une loi normale, modulo renormalisation. Dans cet exposé nous discuterons des hypothèses et d'un cadre dans lesquels cette conclusion est vérifiée, principalement pour des a priori gaussiens. Nous étudierons en détails l'exemple du modèle de translation en bruit blanc gaussien,  $dX(t) = f(t - \theta)dt + (1/\sqrt{n})dW(t)$ ,  $t \in [0, 1]$ , avec  $f$  une fonction symétrique 1-périodique et  $W$  le mouvement Brownien.

*Adresse :*

Ismael CASTILLO  
Vrije Universiteit  
Vrije Universiteit  
De Boelelaan,  
Amsterdam, the Netherlands  
E-mail : [i.castillo@few.vu.nl](mailto:i.castillo@few.vu.nl)

## Mélange de modèles pour le taux de risque conditionnel sous censure

par **Willem Kruijer**

In statistical modeling it is usually assumed that the unknown distribution of the data is contained in the model. Often however there is no strong evidence for this assumption. If the model does not contain the distribution that generated the data, it is said to be misspecified. In practice many models still perform well under misspecification, and it is of interest to formulate conditions under which this happens. In this talk we do this for some nonparametric (Bayesian) regression problems. Under certain conditions on the true error distribution and on the space of regression functions, the posterior concentrates, at the optimal rate, around the point in the model that is closest to  $p_0$  in Kullback-Leibler sense.

Kleijn and van der Vaart [?] give asymptotic results for misspecified i.i.d models in Bayesian nonparametrics. For non-i.i.d., but correctly specified models, convergence rates were obtained by Ghosal and van der Vaart [?]. Although the results of these papers could possibly be combined, we take an alternative information theoretic approach, based on recent work by Zhang ([?],[?]). He showed that general results regarding posterior convergence rates can be derived from the information inequality. We extend his results to misspecified models with independent non-identically distributed observations, and study the consequences for nonparametric regression.

*Bibliographie :*

- [1] Ghosal, S. and van der Vaart, A.W. (2007) Convergence rates of posterior distributions for noniid observations. *Annals of Statistics* **35**(1), 192–223.
- [2] Kleijn, B.J.K. and van der Vaart, A.W. (2006) Misspecification in infinite-dimensional Bayesian statistics. *Annals of Statistics* **34**(2), 837–877.
- [3] Zhang, T. (2006) Information-theoretic upper and lower bounds for statistical estimation. *IEEE Trans. Inform. Theory* **52** 1307–1321.
- [4] Zhang, T. (2006) From  $\epsilon$ -entropy to KL-entropy : analysis of minimum information complexity density estimation. *Annals of Statistics* **34**(5), 2180–2210.

*Adresse :*

Willem KRUIJER  
Vrije Universiteit  
Vrije Universiteit  
De Boelelaan,  
Amsterdam, the Netherlands  
E-mail : [kruijer@few.vu.nl](mailto:kruijer@few.vu.nl)



## Apprentissage séquentiel

Session organisée par **Gilles Stoltz**

La statistique classique part d'un jeu d'observations complet, modélise et estime la loi de ces dernières, afin d'en déduire certaines propriétés. Lors de l'application pratique de ces méthodes, on est souvent confronté au fait que l'on ne dispose pas de toutes les observations immédiatement, mais qu'elles sont révélées pas à pas. Par exemple, en météorologie, on accède à une nouvelle observation (ou série d'observations) chaque jour, après avoir formé nos prédictions la veille. Le jeu de données s'enrichit peu à peu, et il s'agit de construire des stratégies de prévision qui s'adaptent et utilisent efficacement cet accroissement des données. Les mêmes problèmes se posent, par exemple, dans le cadre de l'investissement boursier ou de la prévision de consommation électrique. Différents paradigmes, stochastiques ou dits de suites individuelles, sont considérés ; les seconds ont l'avantage et les défauts de la robustesse, ils s'appliquent à toutes les situations mais sont parfois, pas toujours, trop précautionneux pour améliorer réellement les procédures existantes (sur lesquelles leurs algorithmes de prédiction se fondent). On se reportera aux exposés de Yannig Goude et Vivien Mallet, où l'on dispose de plusieurs procédures de prédiction de base, que l'on combine séquentiellement plutôt que d'en sélectionner une. Les premiers exploitent quant à eux le caractère stochastique pour, par exemple, sélectionner les variables les plus significatives dans le jeu de données (voir l'exposé de Mohamed Hebiri) et/ou considérer des observations imparfaites (voir l'exposé introductif de Rémi Munos).

*Adresse de l'organisateur :*

Gilles STOLTZ

Ecole normale supérieure & HEC Paris

45 rue d'Ulm, 75005 Paris & 1 rue de la libération, 78351 Jouy-en-Josas, France

E-mail : [gilles.stoltz@ens.fr](mailto:gilles.stoltz@ens.fr)

<<http://www.dma.ens.fr/~stoltz>>

### **Bandit algorithms for tree search with applications to games, optimization, and control**

par **Rémi Munos**

Bandit-based methods for tree search have recently gained popularity when applied to huge trees, e.g., in the game of go [2]. Their efficient exploration policy of the tree enables to return rapidly a good value and improve precision if more time is provided. The UCT algorithm [3], a tree search method based on Upper Confidence Bounds (UCB, see [1]), is believed to adapt locally to the effective smoothness of the tree. However, it is known that UCT may be over-optimistic for some problems, leading to high regret. I will describe alternative bandit-based approaches and analyze a Bandit Algorithm for Smooth Trees (BAST) which takes into account actual smoothness of the rewards for performing efficient cuts of sub-optimal branches with high confidence. I will then illustrate several fields of applications, such as minimax search in games, global optimization of Lipschitz functions, and discounted optimal control.

The first part of the talk will be based on the paper

<http://hal.inria.fr/inria-00150207/en/>

*Références :*

- [1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine Learning Journal*, 47(2-3) :235–256, 2002.
- [2] S. Gelly, Y. Wang, R. Munos, and O. Teytaud. Modication of UCT with patterns in Monte-Carlo go. Technical Report INRIA RR-6062, 2006.
- [3] L. Kocsis and C. Szepesvari. Bandit based Monte-Carlo planning. *Proceedings of the European Conference on Machine Learning*, pages 282-293, 2006.

*Adresse :*

Rémi MUNOS  
INRIA Lille, projet SequeL  
40 avenue Halley  
59650 Villeneuve d'Ascq, France  
E-mail : [remi.munos@inria.fr](mailto:remi.munos@inria.fr)  
<<http://sequel.futurs.inria.fr/munos>>

## Sparse conformal predictors

par **Mohamed Hebiri**

Conformal predictors exploit the similarity of new data points with previous data to build prediction intervals. This approach was introduced in [3] and was illustrated on the example of ridge regression. In this talk, we will explore the properties of conformal predictors based on sparse solutions for linear regression problems. Sparse representations can be obtained with a LASSO procedure [2] and we will explain how to adapt the LARS algorithm [1] to efficiently design sparse conformal predictors.

*Références :*

- [1] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2) :407-499, 2004.
- [2] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1) :267-288, 1996.
- [3] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

*Adresse :*

Mohamed HEBIRI  
LPMA, Université Paris Diderot (Paris-VII)  
175 rue du Chevaleret  
75013 Paris, France  
E-mail : [hebiri@math.jussieu.fr](mailto:hebiri@math.jussieu.fr)

## Prévision de consommation d'électricité

par **Yannig Goude**

Cet exposé traite du problème de mélange séquentiel de prédicteurs et son application à la prévision de séries temporelles, en particulier la consommation d'électricité au périmètre EDF. Le but est la prédiction d'observations que l'on acquiert séquentiellement et pour lesquelles nous disposons d'un nombre fini  $M$  de prévisions individuelles. Nous cherchons à construire une prévision agrégée, appelée prévision mélange, qui, sous des hypothèses assez générales, prévoit presque aussi bien qu'un meilleur prédicteur inconnu a priori, appelé oracle. Nous supposons que le meilleur prédicteur (en un sens que nous définissons) parmi les  $M$  considérés peut changer au cours du temps, cette suite de prédicteurs optimaux formant un nouvel oracle. Nous introduisons de nouvelles stratégies de mélange exploitant des techniques de détection de rupture de manière à suivre séquentiellement cet oracle.

Le problème de la prévision de consommation d'électricité est abordé dans ce cadre. Nous utilisons comme prédicteurs individuels des modifications du modèle utilisé à l'opérationnel par EDF et appliquons ces nouveaux algorithmes sur des cas concrets de prévisions difficiles. Nous montrons qu'il est possible d'améliorer sensiblement les prévisions du modèle opérationnel lors de périodes réputées atypiques.

*Adresse :*

Yannig GOUDE  
EDF R&D, département OSIRIS  
1 avenue du général de Gaulle  
92170 Clamart, France  
E-mail : [yannig.goude@edf.fr](mailto:yannig.goude@edf.fr)

## **Agrégation séquentielle pour la prévision de la qualité de l'air**

par **Vivien Mallet**

La prévision, éventuellement opérationnelle, de la qualité de l'air s'effectue grâce à des modèles de transport réactif. Ces modèles intègrent une équation d'advection-diffusion-réaction dont les coefficients sont estimés par des paramétrisations physiques et de nombreuses données (champs météorologiques et émissions, par exemple). Une grande limitation des modèles provient des fortes incertitudes dans leurs formulations physiques et leurs données d'entrée. Pour tenir compte de ces incertitudes, plusieurs modèles alternatifs sont construits et utilisés simultanément en prévision – on parle de prévision d'ensemble.

L'objectif de nos travaux est de combiner linéairement les résultats d'un ensemble de modèles, pour produire une prévision meilleure que toute prévision issue ou dérivée de l'ensemble. Plusieurs méthodes d'apprentissage statistique sont appliquées, souvent avec succès, et comparées; on peut citer notamment l'algorithme de pondération par poids exponentiels à la Gibbs, la "ridge regression" et divers autres issus de la communauté des suites individuelles, voir [1, 2]. La présentation abordera les difficultés et les spécificités de l'application, et notamment, observations spatialisées (stations de mesure sur l'Europe), faible pertinence du passé lointain pour la prévision, importance des événements extrêmes.

*Références :*

- [1] N. Cesa-Bianchi et Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- [2] V. Mallet, G. Stoltz et B. Mauricette. Description of Sequential Aggregation of Methods and their Performances for Ozone Ensemble Forecasting, rapport technique DMA-07-08, Ecole normale supérieure, 2007.

*Adresse :*

Vivien MALLET  
INRIA Rocquencourt, projet CLIME  
BP 105  
78153 Le Chesnay Cedex, France  
E-mail : [vivien.mallet@inria.fr](mailto:vivien.mallet@inria.fr) web



## Méthodes de Monte Carlo adaptatives

Session organisée par François Le Gland et Éric Moulines

Les méthodes de Monte Carlo avec interaction, également appelés méthodes particulières ou méthodes de Monte Carlo séquentielles, consistent à exploiter les poids attribués aux différentes simulations en cours, pour dupliquer et autoriser à se poursuivre les simulations de plus fort poids, ou au contraire arrêter les simulations de plus faible poids. De nombreuses stratégies d'adaptation sont proposées et étudiées, par exemple pour apprendre la distribution d'importance optimale, adapter la taille de l'échantillon ou les instants de redistributions, etc., avec le but commun de réduire la variance des estimateurs proposés.

Les méthodes particulières connaissent un très grand succès dans le domaine générique de la navigation, de la localisation et de la poursuite : l'exposé d'Etienne Mémin présente un certain nombre de techniques de filtrage particulière pour le suivi visuel de points ou de courbes, ou pour l'estimation d'un mouvement fluide, à partir d'une modélisation de faible dimension, éventuellement apprise sur la séquence d'images. Deux autres domaines où les méthodes particulières connaissent actuellement un fort développement sont ceux de la simulation d'évènements rares et de la simulation moléculaire : l'exposé d'Arnaud Guyader présente une stratégie adaptative de branchement multi-niveaux pour estimer la queue d'une distribution, avec une application en protection des données multimedia, l'exposé de Nadia Oudjane présente une stratégie d'apprentissage de la distribution d'importance optimale, avec une application à l'évaluation de value-at-risk (VaR) pour un portefeuille complexe, et l'exposé de Raphaël Roux présente un algorithme de calcul d'énergie libre mettant en jeu un système de particules en interaction et permettant d'échapper aux états métastables. Enfin, l'exposé de Jimmy Olsson présente une méthodologie pour l'adaptation de la distribution d'importance, basée sur le contrôle de la taille effective de l'échantillon, estimé à l'aide du coefficient de variation des poids.

### Références :

- [1] Olivier Cappé, Éric Moulines, and Tobias Rydén. *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer-Verlag, New York, 2005.
- [2] Pierre Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications. Springer-Verlag, New York, 2004.
- [3] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.

### Adresses des organisateurs :

François LE GLAND  
INRIA Rennes — Bretagne Atlantique  
IRISA  
Campus de Beaulieu  
35042 RENNES Cedex, France  
E-mail : [legland@irisa.fr](mailto:legland@irisa.fr)  
<<http://www.irisa.fr/aspi/legland/>>

Éric MOULINES  
ENST Paris  
Département TSI  
37-39 rue Dareau  
75014 PARIS, France  
E-mail : [moulines@tsi.enst.fr](mailto:moulines@tsi.enst.fr)  
<<http://www.tsi.enst.fr/~moulines/>>

## Techniques de filtrage pour le suivi visuel d'entités régies par une dynamique non linéaire

par Etienne Mémin

Dans cet exposé nous passerons en revue un certain nombre de techniques de filtrage proposées pour le suivi d'entités dynamiques à partir de séquences d'images. Nous nous intéresserons en particulier au suivi de variables d'état régies par une dynamique non linéaire estimée directement à partir des données images ou définie par un modèle physique. Plusieurs applications seront décrites. Nous montrerons comment une dynamique non linéaire définie à partir des données et couplée à une vraisemblance gaussienne peut être avantageusement utilisée pour une application de suivi de points dans une séquences d'images. Nous aborderons ensuite le suivi d'écoulements fluides à partir de séquences d'images. Nous décrirons une approche de filtrage définie à partir d'une description paramétrique de dimension réduite du mouvement fluide et permettant de suivre les grandes échelles d'un écoulement. En dernier lieu, nous décrirons une technique de filtrage pour le suivi de courbes dans des séquences. L'approche se caractérise ici par une dynamique décrite au moyen d'une équation différentielle stochastique définie à partir des données et incluant un modèle de bruit de faible dimension.

*Adresse :*

Etienne MÉMIN  
Université de Rennes 1  
IRISA  
Campus de Beaulieu  
35042 RENNES Cedex, France  
E-mail : [memin@irisa.fr](mailto:memin@irisa.fr)  
<<http://www.irisa.fr/prive/memin/>>

## Méthode particulière pour l'estimation de la queue d'une distribution

par Frédéric Cérou, Teddy Furon et **Arnaud Guyader**

On veut estimer la probabilité d'un évènement rare  $\mathcal{R}$ , correspondant au dépassement d'un seuil par une variable aléatoire. Précisément, soit  $X$  un élément aléatoire (variable, vecteur, image, courbe) à valeurs dans un espace  $E$  et  $\Phi : E \rightarrow \mathbb{R}$  une fonction suffisamment régulière. On s'intéresse à la probabilité de l'évènement  $\mathcal{R} = \{\Phi(X) \geq L\}$ , où  $L$  est un scalaire supposé grand par rapport aux valeurs usuelles de la variable aléatoire  $\Phi(X)$ , de sorte que  $\mathcal{R}$  soit en effet très peu fréquent.

On suppose qu'aucun calcul analytique n'est possible : on sait juste simuler des réalisations de  $X$  et évaluer  $\Phi$  en ces points. On est donc amené à utiliser des méthodes de type Monte Carlo. L'évènement  $\mathcal{R}$  étant rare, une méthode de Monte Carlo naïve est vouée à l'échec, c'est pourquoi on propose ici une méthode d'estimation multi-niveaux. D'un point de vue théorique, celle-ci peut s'interpréter et s'analyser grâce aux formules de Feynman-Kac pour un système de particules en interaction. Une application à l'estimation de probabilité de fausse alarme en protection de données multimédia (watermarking) illustrera l'exposé.

*Adresses :*

Frédéric CÉROU  
INRIA Rennes — Bretagne Atlantique  
IRISA  
Campus de Beaulieu  
35042 RENNES Cedex, France  
E-mail : [frederic.cerou@irisa.fr](mailto:frederic.cerou@irisa.fr)  
<<http://www.irisa.fr/aspi/fcerou/>>

Teddy FURON  
INRIA Rennes — Bretagne Atlantique  
IRISA  
Campus de Beaulieu  
35042 RENNES Cedex, France  
E-mail : [teddy.furon@irisa.fr](mailto:teddy.furon@irisa.fr)  
<<http://www.irisa.fr/temics/staff/furon/>>

Arnaud GUYADER  
Université de Haute Bretagne  
Laboratoire de Statistique  
Place du Recteur H. Le Moal  
35043 RENNES Cedex, France  
E-mail : [arnaud.guyader@uhb.fr](mailto:arnaud.guyader@uhb.fr)  
<[http://www.uhb.fr/sc\\_sociales/labstats/AGUYADER/](http://www.uhb.fr/sc_sociales/labstats/AGUYADER/)>

## Variance reduction approach based on particle methods for Value at Risk

par Nadia Oudjane

Computing Value at Risk of large and complex portfolios is known to be one of the most challenging application of numerical methods to finance. The VaR is defined to be an extreme quantile of the distribution of the portfolio losses over a given time horizon. The classical crude Monte Carlo approach requires to simulate a great number of market prices for the given time horizon and to compute for each simulation the portfolio's value. However, only a small fraction of those computations will really contribute to the VaR computation. Typically, for 99% VaR, only 10 percents of the simulations are of interest. Hence, importance sampling technique can be used to artificially bias the simulations into the region of interest and then to reduce the variance of the estimation. The quality of the importance sampling approximation strongly depends on the choice of the proposal distribution. Unfortunately, the minimal variance proposal distribution cannot be used since it depends on the unknown expectation. We propose a new adaptive particle algorithm which approximates progressively the minimal variance proposal distribution required for VaR computation. Our approach achieves important variance reductions especially in the case of extreme VaR.

*Adresse :*

Nadia OUDJANE  
EDF R&D  
Département OSIRIS  
1 avenue du Général de Gaulle  
92141 CLAMART Cedex, France  
E-mail : [nadia.oudjane@edf.fr](mailto:nadia.oudjane@edf.fr)

## Adaptive methods for sequential importance sampling

par Julien Cornebise et Jimmy Olsson

In this talk we discuss new adaptive proposal strategies for sequential Monte Carlo algorithms — also known as particle filters — relying on new criteria evaluating the quality of the proposed particles. The choice of the proposal distribution is a major concern and can dramatically influence the quality of the estimates. Thus, we show how the long-used coefficient of variation of the weights can be used for estimating the  $\chi^2$ -distance between the target and instrumental distributions of the auxiliary particle filter. As a by-product of this analysis we obtain an auxiliary adjustment multiplier weight type for which this  $\chi^2$ -distance is minimal. Moreover, we establish an empirical estimate of linear complexity of the Kullback-Leibler divergence between the involved distributions. Guided by these results, we discuss adaptive designing of the particle filter proposal distribution, e.g. by means of population Monte Carlo techniques, and illustrate the methods on a numerical example.

*Adresses :*

Julien CORNEBISE  
ENST Paris  
Département TSI  
37-39 rue Dareau  
75014 PARIS, France  
E-mail : [cornebise@ccr.jussieu.fr](mailto:cornebise@ccr.jussieu.fr)  
<<http://www.lsta.upmc.fr/doct/cornebise/index.fr.html>>

Jimmy OLSSON  
ENST Paris  
Département TSI  
37-39 rue Dareau  
75014 PARIS, France  
E-mail : [olsson@tsi.enst.fr](mailto:olsson@tsi.enst.fr)

## Calculs d'énergie libre par système de particules en interaction

par Benjamin Jourdain, Tony Lelièvre et **Raphaël Roux**

Un problème important en simulation moléculaire est de calculer des différences d'énergies entre certains états d'un système physique. La difficulté de ce genre de calcul vient de deux problèmes. Premièrement, la dimension du système est en général très grande, ce qui impose un recours à des méthodes de Monte-Carlo. Deuxièmement, le système présente dans la plupart des cas de nombreux états métastables, de sorte qu'une simulation directe serait beaucoup trop coûteuse en temps.

Je présenterai une méthode utilisant une force biaisante adaptative (ABF), qui permet de sortir des métastabilités en un temps raisonnable, faisant intervenir une EDS non linéaire, ainsi qu'une approximation de cette EDS par un système de particules en interaction.

### *Adresses :*

Benjamin JOURDAIN  
CERMICS – ENPC  
6 et 8 avenue Blaise Pascal  
Cité Descartes – Champs sur Marne  
77455 MARNE la VALLÉE Cedex, France  
E-mail : [jourdain@cermics.enpc.fr](mailto:jourdain@cermics.enpc.fr)  
<<http://cermics.enpc.fr/~jourdain/>>

Tony LELIÈVRE  
CERMICS – ENPC  
6 et 8 avenue Blaise Pascal  
Cité Descartes – Champs sur Marne  
77455 MARNE la VALLÉE Cedex, France  
E-mail : [lelievre@cermics.enpc.fr](mailto:lelievre@cermics.enpc.fr)  
<<http://cermics.enpc.fr/~lelievre/>>

Raphaël ROUX  
CERMICS – ENPC  
6 et 8 avenue Blaise Pascal  
Cité Descartes – Champs sur Marne  
77455 MARNE la VALLÉE Cedex, France  
E-mail : [rouxr@cermics.enpc.fr](mailto:rouxr@cermics.enpc.fr)



## Probabilités et géométrie

Session organisée par **Marc Arnaudon**

Cette session aura pour but de réunir un spectre assez large de chercheurs qui utilisent des structures géométriques (variétés, groupes de Lie, fibrés vectoriels, connexions), dans le domaine des probabilités et statistiques, théoriques et appliquées. D'un point de vue théorique, on verra comment les probabilités, la géométrie et le calcul stochastique permettent de construire différents modèles physiques. Du côté des applications seront principalement concernés la comparaison et la discrimination de formes et le traitement des images. Il est espéré que la mise en présence de ces différents thèmes de recherche sera profitable à tous.

*Adresse de l'organisateur :*

Marc ARNAUDON  
Laboratoire de Mathématiques et Applications  
Département de mathématiques  
Université de Poitiers  
Téléport 2, rue Marie et Pierre Curie  
BP 30179  
86962 FUTUROSCOPE CHASSENEUIL Cedex  
E-mail : [Marc.Arnaudon@math.univ-poitiers.fr](mailto:Marc.Arnaudon@math.univ-poitiers.fr)  
<<http://www-math.sp2mi.univ-poitiers.fr/~arnaudon/>>

### Modélisation, comparaison et discrimination de formes : questions géométriques et statistiques

par **Alain Trouvé**

Dans cet exposé, nous aborderons la problématique de la construction d'espaces de formes ( $n$ -uplets labellés de points, nuages de points, sous-variétés, images en niveau de gris) au travers de l'action de groupes de transformations de dimension infinie et la géométrie Riemannienne. Nous essayerons de mettre en lumière les nouveaux enjeux posés pas les problèmes de constructions de modèles probabilistes et l'estimation de modèles génératifs.

*Adresse :*

Alain TROUVÉ  
CMLA ENS-CACHAN  
Bât Laplace  
61, Avenue du Président Wilson  
94235 Cachan Cedex  
Département de mathématiques  
Université de Poitiers  
Téléport 2, rue Marie et Pierre Curie  
BP 30179  
86962 FUTUROSCOPE CHASSENEUIL Cedex  
E-mail : [trouve@cmla.ens-cachan.fr](mailto:trouve@cmla.ens-cachan.fr)

### Statistical Computing on Manifolds for Computational Anatomy

par **Xavier Pennec**

Computational anatomy is an emerging discipline that aim at analyzing and modeling the biological variability of the human anatomy. The goal is not only to model the normal variations among

a population, but also to discover morphological differences between normal and pathological populations. To reach this goal, the method is to identify representative anatomically geometric features (points, tensors, curves, surfaces, volume transformations), and to describe their statistical distribution. Unfortunately, these geometric features belong to manifolds that are not vector spaces. Based on a Riemannian manifold structure, we first developed a consistent statistical framework to compute first order moments and the related statistical (Mahalanobis) distance. We also provided tractable approximations of the generalized normal and chi-square laws for small variances, which allows to perform intrinsic statistical tests. Application examples are provided with the evaluation of the performances of rigid registration algorithms and with a statistical model of the scoliotic spine. Then, we extend the Riemannian computing framework to PDEs for smoothing and interpolation of fields of features with the example of positive definite symmetric matrices (tensors). These covariance matrices are used to in Diffusion Tensor Imaging or to describe the joint anatomical variability at different places (Green function) in shape variability analysis. We show that the choice of a convenient Riemannian metric allows to generalize consistently to tensor fields many important geometric data processing algorithms such as interpolation, filtering, diffusion and restoration of missing data. The methodology is exemplified with Diffusion Tensor MR Images (DTI), and with the modeling of the variability of the brain from a dataset of sulcal lines delineated on the cerebral cortex. In this context, we obtain a dense 3D variability map which proves to be in accordance with previously published results on smaller samples subjects. We also propose innovative methods to analyze the asymmetry of brain variability, and the correlation between different points of the brain.

*Adresse :*

Xavier PENNEC  
INRIA, Sophia Antipolis  
2004 route des Lucioles BP 93  
06902 SOPHIA ANTIPOLIS Cedex - FRANCE  
E-mail : [Xavier.Pennec@sophia.inria.fr](mailto:Xavier.Pennec@sophia.inria.fr)  
<<http://www-sop.inria.fr/asclepios/personnel/Xavier.Pennec/>>

## Mesures de Yang-Mills en deux dimensions et revêtements ramifiés aléatoires

par **Thierry Lévy**

Dans cet exposé, je décrirai une classe de champs d'holonomie aléatoire sur des surfaces qui contient à la fois la mesure de Yang-Mills telle qu'elle est utilisée par les physiciens et des modèles de revêtements ramifiés aléatoires. Par ailleurs, j'expliquerai comment la dualité de Schur-Weyl permet de relier d'une autre façon la mesure de Yang-Mills associée au groupe unitaire à des fibrés principaux singuliers dont le groupe de structure est le groupe symétrique et qui ne sont autres que des revêtements ramifiés.

*Adresse :*

Thierry LÉVY  
DMA - Ecole normale supérieure  
45, rue d'Ulm  
75230 Paris Cedex 05 - France  
E-mail : [levy@dma.ens.fr](mailto:levy@dma.ens.fr)  
<<http://www.dma.ens.fr/~levy/>>

## Machines à vecteurs de support et autres méthodes à noyaux

Session organisée par **Fabrice Rossi**

La session porte sur les méthodes d'apprentissage automatique dites « à noyaux » [1, 2]. Ces méthodes s'appuient sur une fonction noyau  $k$  (symétrique et positive) et surtout sur un espace de Hilbert à noyau reproduisant  $\mathcal{H}$  engendré par  $k$ . Le modèle des données est choisi dans  $\mathcal{H}$  par optimisation d'un compromis entre l'adéquation aux données et la régularité (au sens de la norme de  $\mathcal{H}$ ) du modèle. La méthode la plus connue est celle des machines à vecteurs de support pour lesquelles l'adéquation modèle/données est mesurée par le *hinge loss* (dans le cadre d'une discrimination en deux classes). Plus généralement, ces méthodes permettent d'estimer des modèles de discrimination à  $m$  classes et des modèles de régression scalaire, multiple ou à valeur dans des espaces plus complexes.

### Références :

- [1] Bernhard Schölkopf and Alex Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.
- [2] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

### Adresse de l'organisateur :

Fabrice ROSSI  
Projet AXiS  
INRIA Rocquencourt  
Domaine de Voluceau, Rocquencourt, B.P. 105  
78153 Le Chesnay Cedex  
France  
E-mail : [Fabrice.Rossi@inria.fr](mailto:Fabrice.Rossi@inria.fr)  
<<http://apiacoa.org/>>

## Les machines à noyaux et leur mise en œuvre efficace

par **Stéphane Canu**

Cette présentation vise à introduire les machines à noyaux en se focalisant sur la plus populaire, le séparateur à vaste marge (SVM), en faisant le point sur les différentes facettes de son utilisation. L'accent sera mis sur la notion de parcimonie et sur les considérations pratiques liées à la mise en œuvre de ce type de méthode.

### Adresse :

Stéphane CANU  
Laboratoire LITIS  
INSA de Rouen  
Avenue de l'Université - BP 8  
76801 Saint-Étienne-du-Rouvray Cedex, France  
E-mail : [stephane.canu@insa-rouen.fr](mailto:stephane.canu@insa-rouen.fr)  
<[http://asi.insa-rouen.fr/enseignants/\\_scanu/](http://asi.insa-rouen.fr/enseignants/_scanu/)>

## Apprentissage du noyau par optimisation convexe

par **Francis Bach**

Dans le cadre des méthodes à noyaux pour l'apprentissage supervisé, le choix du noyau est crucial pour la performance. Je présenterai une méthode pour l'apprentissage du noyau à partir de données utilisant une pénalisation convexe par une norme proche de la norme  $l_1$ . Des résultats récents montrant que cette pénalisation donne lieu à des estimateurs consistents du noyau seront présentés.

*Adresse :*

Francis BACH  
INRIA - Willow project  
Département d'Informatique, Ecole Normale Supérieure  
45, rue d'Ulm  
75230 Paris Cedex  
E-mail : [francis.bach@mines.org](mailto:francis.bach@mines.org)  
<<http://www.di.ens.fr/~fbach/>>

## Régression à sortie noyau : application à la complétion de graphes

par **Florence d'Alché-Buc**

L'abondance de données structurées (séquences, arbres, graphes,...) pose aujourd'hui le problème de l'extension des méthodes d'apprentissage statistique à des sorties structurées. Lorsqu'on peut définir un noyau sur le produit cartésien de l'ensemble des sorties par lui-même, l'astuce des noyaux peut être utilisée et il devient possible de travailler dans le cadre de la régression vers un nouvel espace de sortie doté du produit scalaire défini par le noyau choisi. Nous montrons comment des algorithmes et des classes de fonctions habituellement utilisées pour la régression non paramétrique peuvent être étendus à ce cadre : arbres de régression, k-plus-proches voisins, forêts aléatoires, bagging, gradient boosting. Puis nous nous focalisons sur la problématique de complétion de graphes en restant d'abord dans le cadre supervisé puis en montrant des résultats préliminaires dans le cadre semi-supervisé. Nous illustrons ces approches sur la complétion de réseaux d'interactions protéines-protéines chez la levure puis chez l'homme.

*Adresse :*

Florence D'ALCHÉ-BUC  
IBISC 3190 CNRS & Genopole  
IBISC  
523 Place des Terrasses  
91000 Evry  
E-mail : [florence.dalche@ibisc.fr](mailto:florence.dalche@ibisc.fr)  
<<http://www.ibisc.fr/~dalche/>>

## Risques garantis pour les M-SVM

par **Yann Guermeur**

Les machines à vecteurs support (SVM) sont des modèles de l'apprentissage automatique qui font actuellement l'objet de nombreux travaux de recherche, ceci pour deux raisons principales : d'une part, leurs performances constituent l'état de l'art dans de multiples domaines de la reconnaissance des formes, d'autre part, elles possèdent des propriétés statistiques remarquables. Le premier modèle de SVM proposé par Vapnik et ses co-auteurs calcule des dichotomies. Il peut être utilisé pour effectuer des tâches de discrimination à catégories multiples, dans le cadre de l'application de méthodes de décomposition. Des SVM multi-classes ont également été proposées dans la littérature, parmi lesquelles nous distinguons celles qui s'appuient sur un modèle affine multivarié, que nous nommons M-SVM. Cette communication décrit les trois principaux modèles de M-SVM, ainsi que les bornes que nous avons établies sur leur risque.

Journées MAS 2008, Rennes

*Adresse :*

Yann GUERMEUR

LORIA - Equipe ABC

Campus Scientifique

BP 239

54506 Vandoeuvre-lès-Nancy Cedex, France

E-mail : [Yann.Guermeur@loria.fr](mailto:Yann.Guermeur@loria.fr)

<<http://www.loria.fr/~guermeur/>>



## Tests d'adéquation non paramétrique

Session organisée par **Valentin Patilea**

Cette session propose des nouvelles méthodes pour tester l'adéquation d'un modèle des densités ou d'un modèle des régressions contre des alternatives non paramétriques. Les exposés présentent, d'une part, des procédures statistiques pour tester la monotonie d'une régression ou l'adéquation d'un modèle paramétrique de régression en présence d'une censure aléatoire et, d'une autre part, des méthodes pour tester l'adéquation d'un état quantique à un état quantique donné ou pour tester l'homogénéité des différentes composantes d'un mélange des lois pour lequel les pondérations sont connues.

*Adresse de l'organisateur :*

Valentin PATILEA  
I.R.M.A.R.  
INSA de Rennes  
20, Avenue des Buttes de Coësmes, CS 14315  
35043 Rennes cedex, France  
E-mail : [valentin.patilea@insa-rennes.fr](mailto:valentin.patilea@insa-rennes.fr)  
<<http://www.math.univ-rennes1.fr/stat/>>

### Monotone rearrangements and a test for strict monotonicity

par **Holger Dette**

A new test for strict monotonicity of the regression function is proposed which is based on a composition of an estimate of the inverse of the regression function with a common regression estimate. This composition is equal to the identity if and only if the "true" regression function is strictly monotone, and a test based on an  $L^2$ -distance is investigated. The asymptotic normality of the corresponding test statistic is established under the null hypothesis of strict monotonicity.

*Adresse :*

Holger DETTE  
Fakultät für Mathematik  
Mathematik III  
Universitätsstraße 150  
Ruhr-Universität Bochum  
44780 Bochum, Deutschland  
E-mail : [holger.dette@ruhr-uni-bochum.de](mailto:holger.dette@ruhr-uni-bochum.de)  
<<http://www.ruhr-uni-bochum.de/mathematik3/dette.htm>>

### Tests d'adéquation d'un état quantique

par **Cristina Butucea** et **Katia Meziani**

En physique quantique, une mesure d'un état quantique de la matière ne peut produire que des résultats aléatoires. L'état quantique est caractérisé soit par sa matrice de densité  $\rho$  (auto-adjointe, définie positive et de trace 1), soit par sa fonction de Wigner  $W_\rho$  (une fonction bivariée, de masse 1, pas nécessairement positive). La loi des observations est déterminée par  $\rho$  et par la mesure qui est ici la tomographie quantique homodyne.

Nous proposons deux méthodes pour tester en norme  $L_2$  l'adéquation à un état quantique donné,  $\rho_0$  ou  $W_{\rho_0}$ , respectivement. Il s'agit de considérer des estimateurs d'une fonctionnelle quadratique comme statistiques de test et d'étudier le comportement des tests dans ce double problème inverse non paramétrique.

*Adresses :*

Cristina BUTUCEA  
Laboratoire Paul Painlevé - UMR CNRS 8524  
UFR Mathématiques  
Université des Sciences et Technologies de Lille 1  
59655 Villeneuve d'Ascq cedex, France  
E-mail : [cristina.butucea@math.univ-lille1.fr](mailto:cristina.butucea@math.univ-lille1.fr)  
<<http://math.univ-lille1.fr/~butucea/>>

Katia MEZIANI  
LPMA / UMR 7599  
Université Denis Diderot  
Boîte courrier 7012  
75251 Paris cedex 05, France  
E-mail : [nounge@hotmail.com](mailto:nounge@hotmail.com)  
<<http://www.proba.jussieu.fr/>>

## Test d'homogénéité dans un modèle de mélange

par Florent Autin et **Christophe Pouet**

Nous considérons un modèle de mélange dont les pondérations sont connues. Deux échantillons de même taille sont observés et le problème considéré est de tester l'homogénéité des différentes composantes du mélange. L'introduction du mélange permet de considérer des variables non-identiquement distribuées et représente donc un complément intéressant du cas où les variables sont identiquement distribuées au sein de chaque échantillon. Le cadre choisi pour résoudre ce problème est l'approche minimax. Pour les espaces de Besov, nous retrouvons la vitesse de test habituelle ; du point de vue adaptatif, nous prouvons l'existence d'une perte inévitable. Les tests considérés ici reposent sur des techniques d'ondelettes. Un point important à prendre en compte est l'influence des pondérations dans le mélange. En effet, certaines conditions sont imposées sur les pondérations afin de permettre de récupérer les composantes du mélange et jouent donc un rôle au niveau des constantes dans la vitesse de test.

*Adresses :*

Florent AUTIN  
L.A.T.P.  
Université de Provence  
C.M.I. 39 rue F. Joliot-Curie  
13453 Marseille cedex 13, France  
E-mail : [autin@cmi.univ-mrs.fr](mailto:autin@cmi.univ-mrs.fr)  
<<http://www.cmi.univ-mrs.fr/~autin/>>

Christophe POUET  
Université de Provence  
C.M.I. 39 rue F. Joliot-Curie  
13453 Marseille cedex 13, France  
E-mail : [pouet@cmi.univ-mrs.fr](mailto:pouet@cmi.univ-mrs.fr)  
<<http://www.cmi.univ-mrs.fr>>

## Tests non paramétriques d'adéquation pour des modèles de régression en présence de censure

par **Olivier Lopez**

Nous nous intéressons au test non paramétrique de l'hypothèse suivante,

$$H_0 : \exists \theta_0 \in \Theta \subset \mathbb{R}^k \text{ tel que } E[Y|X] = f(\theta_0, X),$$



où  $f$  est une fonction connue,  $X \in \mathbb{R}^d$ , et  $Y \in \mathbb{R}$  une variable aléatoire censurée à droite. Introduisant une variable de censure  $C$ , les observations sont constituées de répliques i.i.d.  $(T_i, \delta_i, X_i)_{1 \leq i \leq n}$  où

$$\begin{aligned} T_i &= \inf(Y_i, C_i), \\ \delta_i &= \mathbf{1}_{Y_i \leq C_i}. \end{aligned}$$

Nous proposons deux procédures de test de l'hypothèse  $H_0$  contre une alternative non paramétrique. Ces procédures sont vues comme des généralisations de statistiques de test proposées, en l'absence de censure, par Zheng (1996), et Horowitz et Spokoiny (2001).

Dans un premier temps, nous considérons le cas où  $C$  est indépendant des variables  $(Y, X)$ . Dans ce contexte, nous démontrons des propriétés de consistance sous des alternatives se rapprochant de l'hypothèse nulle, et nous comparons notre approche à celle de Stute, W., González-Manteiga, W. et Sánchez-Sellero, C. (2000). Dans un second temps, nous modifions notre procédure afin de considérer le cas plus général où  $C$  est indépendant de  $Y$  conditionnellement à  $X$ .

*Références :*

- [1] Horowitz, J.L. & Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.
- [2] Stute, W., González-Manteiga, W. & Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods* **29**, 1611–1629.
- [3] Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.

*Adresse :*

Olivier LOPEZ  
CREST-ENSAI et IRMAR  
ENSAI  
Campus de Ker Lann, rue Blaise Pascal  
BP 37203 – 35172 Bruz Cedex, France  
E-mail : [lopez@ensai.fr](mailto:lopez@ensai.fr)  
<<http://www.ensai.com/>>

