

Curriculum Vitæ

Delphine BLANKE

Née le 14-06-71 à Argenteuil (95)

Mél : delphine.blanke@upmc.fr,

Web : <http://www.lsta.upmc.fr/blanke.html>

Coordonnées Professionnelles :

Université UPMC-Paris 6, L.S.T.A.,
8ème étage, Bâtiment A, Bureau 29,
175-179 rue du Chevaleret, 75013 Paris.
Tél : 01.44.27.33.53

Coordonnées Personnelles :

3 sente du Clos d'Arcy,
78300 Poissy.
Tél : 01.30.65.76.96

Cursus Universitaire

Depuis Sept. 2002

Décembre 2004

Maître de Conférences à l'Université Paris 6.

Habilitation à Diriger des Recherches, Université Paris 6.

Spécialité Mathématique/Statistique, soutenue le 6 décembre 2004.

Intitulé : *Contributions à la statistique fonctionnelle des processus.*

Composition du jury :

Yury DAVYDOV (Univ. Lille 1, président),

Peter HALL (Université de Canberra, rapporteur),

Yury KUTOYANTS (Université du Maine, rapporteur),

Wolfgang WEFELMEYER (Université de Cologne, rapporteur),

Denis BOSQ (Université Paris 6, examinateur),

Michel DELECROIX (E.N.S.A.I., examinateur).

Sept. 1998-Août 2002

Sept. 1997-Août 1998

1994-1997

Maître de Conférences à l'Université du Havre.

A.T.E.R., Université Paris 6.

Thèse de Doctorat de l'Université Paris 6,

Spécialité Mathématique/Statistique, sous la direction du Pr. D. BOSQ,
soutenue le 17 novembre 1997, *mention Très Honorable.*

Intitulé de la thèse :

*Estimation non paramétrique pour des processus à temps continu
bruités ou partiellement observés.*

Composition du jury :

Christian GOURIEROUX (prés.), Denis BOSQ (dir. de thèse),

Yury KUTOYANTS (rap.), Elias MASRY (rap., U.S.A.),

Yury DAVYDOV (exam.), Michel DELECROIX (exam.),

Valentine GENON-CATALOT (exam.).

1993-1994

D.E.A. de Statistique, Université Paris 6. *Mention Bien.*

Tâches diverses

- [1] Membre du comité de programme des Journées de la SFdS (Société Française de Statistique) en 2009 (Bordeaux).
- [2] Organisatrice du colloque “Statistique non paramétrique et Statistique des processus”, tenu à Paris en septembre 2007, en l’honneur de D. Bosq (environ 80 participants).
Éditrice du numéro spécial des *Annales de l’ISUP* dédié au colloque.
- [3] Co-Responsable des stages en M2 de Mathématiques appliquées, Option Statistique, depuis Septembre 2006.
- [4] Membre du jury “Industrie et Services” de l’I.S.U.P. (Institut de Statistique de l’Université Pierre et Marie Curie) depuis Septembre 2002.
- [5] Éluë au conseil de l’I.S.U.P. depuis Janvier 2003 (réélue en Janvier 2007).
- [6] Membre de la commission de spécialistes, section 26, de Paris 6 (réélue en 2007 et Vice-Présidente B sur la période 2004-2007).
Suppléante (2004-2007) à la commission 25/26 de l’Université d’Angers.
Membre de la commission 25/26 du Havre de 2000 à 2002.
- [7] Titulaire de la P.E.D.R. de 2000 à 2004 et depuis 2005.
- [8] Referee pour *Annales de l’ISUP*, *Ann. Sci. Math. du Québec*, *C.R. Mathématique*, *Canad. J. Statist.*, *J. Statist. Plann. Inference*, *Statist. Inf. Stoch. Proc.*, *Statist. Probab. Letters*.
- [9] Rapporteur pour les *Mathematical Reviews* depuis décembre 2004.
- [10] Construction du site Web de l’I.S.U.P. et maintenance du site actuel.
Correspondante informatique du L.S.T.A. de 2002 à 2007.
- [11] Responsable de l’organisation du séminaire du L.M.A.H., Université du Havre (2001-2002).
- [12] Présidente de jury de Baccalauréat, Juillet 2000.

Encadrements et Jurys

- ★ *Université Paris 6*. J’encadre (ou co-encadre) 4 étudiants en thèse :
 - F.-X. Lejeune : *Histogramme, régressogramme et polygone de fréquences en temps continu* (thèse soutenue le 14 décembre 2007).
ATER (Université Paris 6).
 - B. Labrador : *Estimation par plus proches voisins en temps continu* (soutenance programmée courant 2008).
ATER (Université Paris-Nord).
 - N. Rekik : *Modèle additif* (début sept. 2004).
 - A. Amiri : *Méthodes récursives en estimation et prévision* (début sept. 2007).

J'ai été membre des jurys de thèse d'A. Bianchi (thèse en co-tutelle entre les universités de Milan et Paris 6) soutenue en juin 2007, avec pour sujet : *Problèmes de statistique inférentielle pour des diffusions multidimensionnelles* et de D. Degras (Paris 6) en décembre 2007.

J'encadre également des mémoires au niveau M2 (DEA) dont les sujets traités ont été, par exemple :

- *Étude des méthodes récursives en estimation.*
- *Étude des estimateurs à ondelettes : cas non bruité et bruité.*
- *Résultats récents portant sur les méthodes de déconvolution.*
- *Résultats récents sur le choix de la fenêtre.*

★ *Université du Havre (1999-2002).* L'université ne disposant pas de 3ème cycle en Mathématiques, mes travaux d'encadrements se sont bornés durant cette période à l'évaluation de mémoires bibliographiques au niveau DEUG (unité d'enseignement "méthodologie" : recherche documentaire) ainsi qu'à la direction de TER (travail d'étude et de recherche donnant lieu à une soutenance) au sein de la maîtrise de Mathématiques.

Invitations

Outre une quinzaine d'invitations à des congrès ou séminaires (voir la liste détaillée, pages 6-7), je suis invitée à l'université de Fudan (Shanghai, Chine) pour 10 jours, courant 2008.

Travaux et Publications

Livre

- [28] *Inference and Prediction in large dimensions.*
(Coll. D. Bosq), 316 pages.
Paru en septembre 2007 dans la collection John Wiley & Sons - Dunod.

Articles soumis ou publiés

- [27] *Prediction and nonparametric regression estimation in continuous time,*
(Coll. D. Bosq). En révision.
- [26] *Assessing the number of mean-square derivatives of a Gaussian process.*
(Coll. C. Vial). À paraître dans *Stoch. Process. Applic.*
- [25] *Adaptive sampling schemes for density estimation.*
J. Statist. Plann. Inference, 2006, 136, 9, 2898-2917.
- [24] *Sample paths adaptive density estimator.*
Math. Methods Statist., 2004, 13, 2, 123-152.
- [23] *Local superefficiency of data-driven projection density estimators in continuous time.* (Coll. D. Bosq).
Statist. Oper. Research Trans., 2004, 28, 1, p. 37-54.
- [22] *Local Hölder exponent estimation for multivariate continuous time processes.*
J. Nonparam. Statist., 2003, 16, 1-2, p. 227-244.
- [21] *Modelization and nonparametric estimation for a dynamical system with noise.* (Coll. D. Bosq, D. Guégan).
Statist. Inf. Stoch. Proc., 2003, 6, p. 267-290.
- [20] *Optimal sampling for density estimation in continuous time.*
(Coll. B. Pumo). *J. Time Ser. Anal.*, 2003, 24, 1, p.1-23.
- [19] *Estimation of local smoothness coefficients for continuous time processes.*
Statist. Inf. Stoch. Proc., 2002, 5, 1, p. 65-93.
- [18] *Estimation of the asymptotic variance of local time density estimators for continuous time processes.* (Coll. F. Merlevède).
Math. Methods Statist., 2000, 9, 3, p. 270-296.
- [17] *A family of minimax rates for density estimators in continuous time.*
(Coll. D. Bosq). *Stochastic Anal. Appl.*, 2000, 18, 6, p. 871-900.
- [16] *Accurate rates of density estimators for continuous time processes.*
(Coll. D. Bosq). *Stat. and Probab. Letters*, 1997, 33, p. 185-191.
- [15] *Estimation de la densité pour des trajectoires non directement observables.*
Publ. Inst. Stat. Univ. Paris, 1996, XXXX, 2, p. 21-36.

Comptes-Rendus de l'Académie des Sciences - Paris

- [14] *Estimation du nombre de dérivées d'un processus Gaussien.*
(Coll. C. Vial). C. R. Math. Acad. Sci. Paris (2006) 343, 10, 661-664.
- [13] *Estimation adaptative de la densité avec données échantillonnées.*
C. R. Math. Acad. Sci. Paris, 2003, 337, 10, p. 675-678.
- [12] *Estimation du coefficient de régularité locale d'une trajectoire de processus.*
C. R. Math. Acad. Sci. Paris, 2002, 334, 2, p. 145-148.
- [11] *Estimation de la densité invariante d'un système dynamique bruité.*
C. R. Acad. Sci. Paris Sér. I Math., 1999, 329, p. 221-224.
- [10] *Estimation non paramétrique de la densité pour des processus à temps continu : vitesses minimax.*
C. R. Acad. Sci. Paris Sér. I Math., 1997, 325, p. 527-530.
- [9] *Estimation de la densité pour des processus bruités à temps continu.*
C. R. Acad. Sci. Paris Sér. I Math., 1995, 321, p. 623-626.

Travaux en cours, Pré-publications et Manuscrits

- [8] *Nonparametric regression in the sampled case with missing data*
en préparation.
- [7] *Estimation of the regularity of a Gaussian process by using quadratic variations,*
manuscrit (21 pages).
- [6] *Assessing the number of mean-square derivatives of a Gaussian process.*
(Coll. C. Vial). Prépublications MODALX, 2006-4, 28 pages.
- [5] Participation à l'article d'océanographie :
Lagrangian analysis of Indian Atlantic interocean exchange in a regional model.
Geophys. Res. Lett., L14611, 33, 2006.
- [4] Mémoire d'habilitation : *Contributions à la statistique fonctionnelle des processus,*
53 pages, décembre 2004.
- [3] *Optimal sampling for density estimation in continuous time.*
(Coll. B. Pumo). Prépublications L.S.T.A., 2002-2, 25 pages.
- [2] *Modelization and nonparametric estimation for a dynamical system with noise.*
(Coll. D. Bosq et D. Guégan).
Document de travail du C.R.E.S.T., 9857, décembre 1998, 27 pages.
- [1] Thèse : *Estimation non paramétrique pour des processus à temps continu bruités ou partiellement observés,* 144 pages, 1997.

Séminaires et Congrès

2005-2009

- [27] Invitation au *Colloque SSC & SFdS* à Ottawa, Canada. Mai 2008.
- [26] *Séminaire du L.M.A.C.*
Université de Technologie, Compiègne. Février 2008.
- [25] *Groupe de travail "Statistique et applications"*.
Université de Caen. Décembre 2007.
- [24] *Séminaire ModalX, "Probabilités et Statistique"*.
Université de Nanterre. Mai 2007.
- [23] *Groupe de Travail "Applications des mathématiques"*.
ENS Cachan, Antenne de Bretagne. Novembre 2006.
- [22] Invitation aux *4èmes Rencontres Statistique Fonctionnelle et Opérationnelle*. Université de Grenoble. Juin 2006.
- [21] *Séminaire parisien de Statistiques*.
Institut Henri Poincaré, Paris. Avril 2006.
- [20] *Séminaire du Laboratoire Angevin de REcherche en Mathématiques*. Université d'Angers. Novembre 2005, Angers.
- [19] Invitation au colloque *Statistique asymptotique des processus stochastiques V*. Université du Maine. Janvier 2005, Le Mans.

2000-2004

- [18] Séminaire de Statistique non paramétrique et semi-paramétrique.
C.R.E.S.T. et Paris 6. Novembre 2003, Paris.
- [17] Congrès de la SFDS. (Communication). Juin 2003, Lyon.
- [16] Invitation au colloque *Statistique asymptotique des processus stochastiques IV*. Université du Maine. Décembre 2002, Le Mans.
- [15] Invitation *International conference on current advances and trends in nonparametric statistics*. Juillet 2002, Crète (Grèce).
- [14] Séminaire de Statistique non paramétrique et semi-paramétrique.
C.R.E.S.T. et Paris 6, février 2002, Paris.
- [13] Séminaire de l'IRMAR : *Processus stochastiques et applications*. Université de Rennes 1, janvier 2002, Rennes.
- [12] Invitation au colloque *Statistique asymptotique des processus stochastiques III*. Université du Maine, décembre 2000, Le Mans.
- [11] Séminaire de Statistique non paramétrique et semi-paramétrique.
C.R.E.S.T. et Paris 6, janvier 2000, Paris.

1995-1999

- [10] Séminaire du *Laboratoire de Mathématiques Raphaël Salem*. Université de Rouen, mars 1999.
- [9] Séminaire de Statistique de l'*E.N.S.A.I.*
Campus de l'*E.N.S.A.I.*, décembre 1998, Bruz.

- [8] Communication au Congrès de l'A.S.U., mai 1998, Rennes.
- [7] Communication au Colloque des Jeunes Probabilistes et Statisticiens, avril 1998, Aussois.
- [6] Séminaire de Statistique non paramétrique et semi-paramétrique. C.R.E.S.T. et Paris 6, février 1998, Paris.
- [5] Communication au Congrès de l'A.S.U., mai 1997, Carcassonne.
- [4] Séminaire de Statistique non paramétrique et semi-paramétrique. C.R.E.S.T. et Paris 6, mars 1997, Paris.
- [3] Séminaire Jeunes du L.S.T.A., octobre 1996, Jussieu, Paris.
- [2] Communication au Congrès de l'A.S.U., mai 1996, Université Laval, Canada.
- [1] Séminaire de Statistique non paramétrique et semi-paramétrique. C.R.E.S.T. et Paris 6, décembre 1995, Paris.

Divers

Classée **2ème** sur le poste PR, profil *biostatistique*, du CNAM Paris (2005).

Classée **3ème** sur le poste PR, profil *modèle, calcul scientifique et statistique*, INSA de Rouen (2006).

Classée **3ème** sur le poste PR, profil *mathématiques appliquées, statistiques, probabilités*, Université d'Angers (2006).

Classée **4ème** sur le poste PR, profil *probabilités et statistiques*, Université de Lille 1 (2007).

Classée **4ème** sur le poste PR, profil *statistiques appliquées*, Université de Dijon (2007).

- Anglais* : Lu, écrit, parlé.
- Espagnol* : Lu, écrit.
- Langages* : Bases de html, LISP et de Turbo Pascal.
- Logiciels* : R, Splus, Statgraphics, Eviews, Dreamweaver.

Enseignements

- Depuis 2002 **Maître de Conférences, Université Paris 6.**
Cours (compléments) en M2 (ex-DEA de Statistique) : Statistique des processus en temps continu.
Rappels sur les processus à temps continu, estimation de la densité et de la régression (temps discret et continu), processus ARH, ...
Cours de Simulation : I.S.U.P. 1ère année (2003-2004),
Générateurs de nombre aléatoires, simulation de v.a. discrètes et continues, ...
T.D. de Statistique Mathématique, I.S.U.P. 1ère année.
Estimation ponctuelle, modèle exponentiel, exhaustivité et information, estimation bayésienne, intervalles de confiance, tests ...
T.D. de Statistique des Processus, M1 de Mathématiques appliquées
Processus stationnaires, Représentation de Wold, densité spectrale, processus de Poisson et de Wiener, intégrale stochastique, équations différentielles stochastiques, ...
T.D. d'Intégration et Mesure, I.S.U.P. 1ère année.
Tribu et mesure, construction de l'intégrale, intégrale de Lebesgue, théorèmes de convergence, mesures produit...
- 1998-2002 **Maître de Conférences, Université du Havre.**
Cours de Mathématiques Appliquées, M.S.T. T.A.M., 1ère année.
Cours de Mathématiques, DEUG SV, 1ère année.
Cours/TD de Statistique, M.S.T. C.I.C., 1ère année.
T.D. de Statistique Inférentielle, Maîtrise de Mathématiques.
T.D. de Processus Stochastiques, Maîtrise de Mathématiques.
T.D. de Probabilités, DEUG M.I.A.S., 2ème année.
T.D. de Statistique, DEUG SNV, 2ème année.
T.D. de Statistique, M.S.T. T.A.M., 1ère année.
T.D. d'Analyse, DEUG M.I.A.S., 1ère année.
- 1997-1998 **A.T.E.R., Université Paris 6.**
T.D. de Probabilités à l'I.S.U.P., 1ère année.
T.D. de Statistique Mathématique à l'I.S.U.P., 1ère année.
T.D. d'Intégration et Mesure à l'I.S.U.P., 1ère année.
T.D. de Probabilités et Statistique, DEUG SCVT, 1ère année.
Encadrement de projets d'Econométrie, I.S.U.P., 2ème année.
(sous les logiciels SAS et E-views).
- 1996-1997 *T.D. de Statistique Inférentielle à l'E.N.S.A.I., Rennes.*
1995-1996 *T.D. de Statistique Inférentielle à l'E.N.S.A.I., Paris.*
1994-1995 *T.D. de Statistique Univariée à l'E.N.S.A.I., Paris.*

Rapport de Recherche

(D. Blanke - Mars 2008)

L'essentiel de mes activités de recherche porte sur l'estimation fonctionnelle ou paramétrique pour des processus. L'ensemble de ces travaux peut se diviser en quatre parties : estimation en temps continu, estimation pour des observations discrétisées, estimation pour des observations bruitées et estimation de paramètres structurels. Les numéros entre crochets réfèrent au Curriculum Vitæ p. 4-5 (resp. p. 6-7 pour les exposés), les manuscrits [15] à [26] pouvant être consultés à l'adresse :

http://www.lsta.upmc.fr/blanke/trav_public.html

Ce rapport se conclut par mes perspectives de recherche pour l'avenir.

1 Estimation fonctionnelle en temps continu

Travaux [1], [4], [10], [16]-[17], [23]-[24], [27]-[28].

Exposés [4], [6]-[8], [10], [16], [20].

1.1 Vitesses de convergence pour l'estimateur à noyau

En premier lieu, on cherche à estimer la densité marginale $f(x_1, \dots, x_d)$ associée à un processus à temps continu $\{X_t, t \in \mathbb{R}\}$ à valeurs dans \mathbb{R}^d (sous l'hypothèse que les X_t aient tous la même loi). Si on dispose de l'observation d'une réalisation de ce processus sur un intervalle de temps $[0, T]$, l'estimation de f peut se faire à l'aide d'un estimateur à noyau usuel, adapté au temps continu. Dans [17] nous avons donné, pour l'étude de l'erreur quadratique de cet estimateur, une *nouvelle famille* de vitesses de convergence qui sont propres au temps continu. Ces vitesses sont calées entre la vitesse "paramétrique" $\frac{1}{T}$ (initialement obtenue par Castellana et Leadbetter, 1986) et la vitesse "optimale" (c'est-à-dire la vitesse du cas i.i.d. obtenue pour une régularité fixée de la densité). Elles sont fonction d'un paramètre, γ_0 , lié à la dimension et à la nature des trajectoires observées ; on remarque qu'elles sont d'autant meilleures quand les composantes du processus ont des trajectoires irrégulières.

Ce phénomène peut être interprété de la manière suivante : plus les trajectoires sont irrégulières, plus on se rapproche du processus (non mesurable)

constitué par un continuum de v.a. i.i.d., et ainsi, plus les vitesses d'estimation peuvent être bonnes. Pour modéliser cette propriété, on fait des hypothèses sur le comportement local de la densité jointe $f_{(X_0, X_u)}$ au voisinage de $u = 0$ et de la diagonale de \mathbb{R}^{2d} . En particulier, dans le cas réel ($d = 1$), on a une alternative simple concernant le comportement de la variance de l'estimateur (voir [16]). En effet, la vitesse obtenue est de l'ordre de $1/T$ pour des processus stationnaires dont la densité du couple $(X_0, \frac{X_u - X_0}{u^{\gamma_0}})$ existe pour un paramètre $\gamma_0 \in]0, 1[$ (par exemple $\gamma_0 = 1/2$ pour certaines classes de diffusions homogènes ergodiques). D'autre part, si $\gamma_0 = 1$ (par exemple pour un processus Gaussien dérivable en moyenne quadratique) la variance est alors dégradée et de l'ordre de $\ln(1/h_T)/T$ où h_T dénote la fenêtre associée à l'estimateur à noyau.

Plus généralement, pour des processus d -dimensionnels, la vitesse d'estimation va dépendre du paramètre $\gamma_0 = \sum_{i=1}^d \gamma_i$ où $\gamma_i \in]0, 1[$ correspond à la régularité de la i -ème composante $\frac{X_u^{(i)} - X_0^{(i)}}{u^{\gamma_i}}$. On peut obtenir ainsi toute une *famille de vitesses intermédiaires* atteintes par l'estimateur à noyau en temps continu et qui n'ont pas d'équivalent en temps discret. Ces vitesses sont également *minimax* au sens où si l'on considère l'ensemble des processus pour lesquels l'estimateur à noyau a une vitesse de convergence donnée, alors *il n'existe pas* d'estimateur ayant une meilleure vitesse de convergence sur cet ensemble. J'ai repris ces travaux dans le livre [28] (avec l'article [27] actuellement en révision), en les complétant avec l'étude de l'estimateur à noyau de la régression pour lequel on observe des résultats similaires. Ces résultats permettent alors d'obtenir des prédicteurs convergeant à des vitesses variées, suivant la nature du processus sous-jacent.

1.2 Estimation adaptative

Le choix de la fenêtre optimale pour l'estimateur à noyau dépendant fortement de la valeur prise par γ_0 , il est intéressant de considérer un estimateur adaptatif relativement à ce paramètre. Dans une première étape, on sélectionne un candidat pour γ_0 dans une grille de valeurs possibles, en suivant un algorithme initié par Lepskii. La valeur ainsi retenue est alors "injectée" dans l'estimateur à noyau usuel. Dans l'article [24], j'ai en premier lieu recherché les vitesses intermédiaires pour la *convergence presque sûre (simple et uniforme)* de l'estimateur à noyau pour γ_0 connu. La deuxième partie de l'article est alors consacrée à l'étude de la convergence presque sûre de l'estimateur *adaptatif*. On montre alors que les vitesses sont du même ordre que celles obtenues dans le cas où γ_0 est supposé connu.

Dans un cadre différent, il est encore possible d'obtenir des estimateurs

avec une vitesse paramétrique, sans hypothèses locales sur la nature des trajectoires. Ainsi, dans [23], nous considérons une extension au temps continu d’un estimateur par projection modifiée. Cet estimateur atteint une vitesse suroptimale sur une sous-classe de densités \mathcal{F}_0 (choisie par le statisticien) tout en conservant une vitesse “raisonnable” en dehors de \mathcal{F}_0 . Plus précisément, soit (E, \mathcal{B}, μ) un espace mesuré avec μ σ -finie, et telle que $L^2(\mu)$ soit de dimension infinie. Soit également $X = (X_t, t \in \mathbb{R})$ un processus stochastique défini sur l’espace de probabilité (Ω, \mathcal{A}, P) et à valeurs dans (E, \mathcal{B}) , les X_t ayant même densité f par rapport à μ . On note \mathcal{F} la famille de ces densités f telles que $f = \sum_{j=0}^{\infty} a_j e_j$, $\sum_{j=0}^{\infty} a_j^2 < \infty$ où $(e_j, j \geq 0)$ est un système orthonormal de $L^2(\mu)$. L’estimateur par projection modifiée se définit alors comme un estimateur par projection, obtenu en tronquant la série dès que les coefficients de Fourier empiriques restants sont *tous* inférieurs à un certain seuil (contrairement aux estimateurs à seuil classiques où seuls les coefficients supérieurs à un seuil donné sont conservés). On s’intéresse maintenant à la classe $\mathcal{F}_0 = \bigcup_{K=0}^{\infty} \mathcal{F}_0(K)$, constituée par les densités f admettant un développement de Fourier fini par rapport à la base choisie : $f = \sum_{j=0}^K a_j e_j$ avec $a_K \neq 0$. Notons que le choix du système orthonormal peut s’effectuer en privilégiant une suite de densités $(\varphi_j, j \geq 0)$, linéairement indépendantes et qui engendreront (e_j) par orthonormalisation. On démontre alors que, pour le risque L^2 intégré et *sans hypothèse d’irrégularité sur les trajectoires*, la vitesse de convergence est à la fois $1/T$ sur \mathcal{F}_0 et quasi-optimale sur $\mathcal{F} \setminus \mathcal{F}_0$. Le même type de résultat est obtenu pour la convergence uniforme presque sûre. Notons également que dans les cas usuels, l’erreur quadratique intégrée asymptotique de l’estimateur par projection modifiée est strictement plus petite sur \mathcal{F}_0 que celle de l’estimateur temps local.

2 Estimation dans le cas discrétisé et applications numériques

Travaux [3]–[4], [8], [13], [20], [25], [28]. Exposés [13], [17]–[20], [27].

Une fois les vitesses d’estimation connues dans le cadre idéal où l’ensemble de la trajectoire est observée sur $[0, T]$, il est naturel de se pencher sur le problème des données discrétisées et en particulier sur l’existence d’un *plan d’échantillonnage optimal*, c’est-à-dire tenant compte des propriétés locales de régularité des trajectoires sous-jacentes. De manière plus spécifique, on se place dans le cadre où le statisticien dispose d’un grand nombre d’observations fréquentes dans le temps (échantillonnage à haute fréquence). Ceci se traduit par le modèle théorique $X_{t_{1,n}}, \dots, X_{t_{n,n}}$ où $t_0 = 0 < t_{1,n} < \dots < t_{n,n}$,

$t_{i+1,n} - t_{i,n} = \delta_n$ avec $\delta_n \rightarrow 0$ et $n\delta_n \rightarrow +\infty$. Le choix de δ_n peut s'avérer crucial dans le cas où par exemple on cherche, afin de minimiser les coûts, à réduire au maximum la durée totale d'observation du processus $T_n = n\delta_n$. Dans l'article [3],[20] écrit en collaboration avec B. Pumo, nous donnons les *choix optimaux* de δ_n pour l'étude de l'erreur quadratique de l'estimateur à noyau de la densité. En particulier dans le cadre réel, on observe que, pour une trajectoire régulière, les observations doivent être plus éloignées les unes des autres que dans le cas d'une trajectoire irrégulière. L'explication naturelle est que la corrélation locale entre les données est bien plus importante pour des phénomènes "lisses" et les *simulations* effectuées confirment également ces résultats théoriques. J'ai étendu ces résultats dans le livre [28] au problème de l'estimation de la fonction de régression, pour laquelle des plans d'échantillonnage similaires sont obtenus. Dans ce cadre et en présence de données manquantes, un article avec *simulations* est en cours de préparation ([8]).

Dans [4],[13],[25] la convergence presque sûre de l'estimateur de la densité est tout d'abord étudiée : on retrouve ainsi les plans d'échantillonnage précédents. Dans un deuxième temps, le problème est abordé sous un angle adaptatif. On construit en particulier un estimateur doublement adaptatif relativement à la régularité r_0 de la densité et à celle de la trajectoire, γ_0 . Cet estimateur se construit à l'aide d'une double grille permettant de sélectionner un couple de valeurs candidates pour (r_0, γ_0) que l'on injecte alors dans l'estimateur. On montre que les vitesses de convergence de cet estimateur doublement adaptatif sont les mêmes que celles obtenues pour (r_0, γ_0) connu. Cette méthode est relativement bien adaptée dans le cas où l'on dispose d'une machine que l'on cherche à calibrer, en testant différents pas de discrétisation durant une période d'apprentissage donnée.

3 Estimation avec erreurs sur les variables

Travaux [1]-[4], [9], [11], [15], [20]-[21].

Exposés [1]-[3], [5], [9], [11], [13], [20].

3.1 Temps continu et discrétisé

Il s'agissait du point de départ de ma thèse : le but étant d'estimer la densité marginale d'un processus $\{X_t^0, t \in \mathbb{R}\}$ à valeurs dans \mathbb{R}^d , sous des hypothèses de stationnarité adéquates, et avec pour base l'observation d'une trajectoire bruitée $X_t = X_t^0 + \varepsilon_t, t \in [0, T]$. Pour n observations ponctuelles

et des estimateurs adaptés (comme l'estimateur à noyau de déconvolution), les résultats suivants étaient déjà connus :

1. les vitesses d'estimation dépendent de la densité du bruit (supposée connue),
2. ces vitesses peuvent décroître de $n^{-4/7}$ (bruit de loi exponentielle) à $(\ln n)^{-2}$ (bruit Gaussien) ;
3. ces vitesses sont *minimax* (au sens où il n'existe pas de meilleur estimateur pour ce type de problème) lorsque les observations sont supposées indépendantes.

Sur ces bases il était donc intéressant de se pencher sur l'extension de tels résultats dans le cas où le statisticien dispose de l'observation globale d'une trajectoire de processus. J'ai ainsi étudié le comportement asymptotique d'une version "temps continu" de l'estimateur à noyau de déconvolution.

Les différentes propriétés de cet estimateur ont été publiées dans une note [9] où l'*erreur quadratique* puis la *normalité asymptotique* sont traitées. On retrouve en particulier, pour une large classe de processus, les mêmes ordres de vitesses que pour le temps discret, avec la même discussion suivant la loi du bruit. Cependant (et de la même manière que dans le cas non bruité) le passage en temps continu permet d'améliorer ces vitesses lorsque les trajectoires sont suffisamment irrégulières et apportent ainsi plus d'information. Par exemple pour un bruit suivant une loi exponentielle, l'estimateur peut atteindre la vitesse $T^{-2/3}$ ("meilleure" que la vitesse optimale $n^{-4/7}$ du cas discret). Les démonstrations de l'ensemble de ces résultats ainsi que la normalité asymptotique multidimensionnelle de l'estimateur ont fait l'objet d'un article [15]. La *convergence presque sûre* (simple et uniforme) est également traitée dans la thèse ([1], chapitres 2 et 3). Ces résultats ont enfin été généralisés à l'estimation de la fonction de régression (avec en particulier la normalité asymptotique multidimensionnelle au chapitre 3 de la thèse) afin de pouvoir traiter les problèmes liés à la prédiction.

En ce qui concerne le bruit Gaussien la vitesse logarithmique $(\ln T)^{-2}$ ne semble malheureusement pas être améliorable. Cependant on montre ([1], chapitre 4) que pour des bruits de niveau asymptotique *faible*, les *vitesses usuelles* d'estimation non paramétrique (et même paramétriques) peuvent être atteintes par les estimateurs usuels (on ne suppose plus alors leur loi connue). Nous avons repris ces résultats dans [3],[20] pour le problème des observations discrétisées et nous les avons également illustrés par des *simulations*.

Finalement dans [1], chapitre 4, le problème d'estimation de la densité d'un processus quand le processus initial est observé via un filtre est

également abordé. Pour le cas particulier des “petites moyennes de trajectoires”, nous donnons des conditions sous lesquelles l’estimateur de la densité atteint encore les vitesses de convergence optimales.

3.2 Temps discret

Je me suis également intéressée à l’estimation de la densité de la mesure invariante d’un *système dynamique* en temps discret et perturbé par un bruit. Soit l’espace de probabilité $(E, \mathcal{B}(E), \mu)$ où E est un fermé de \mathbb{R}^d et $\mathcal{B}(E)$ est la tribu borélienne associée. Rappelons qu’un système dynamique en temps discret est en général défini par une application mesurable $\varphi : E \rightarrow E$ où E est un fermé de \mathbb{R}^d . L’état du système à l’instant t est donné par $X_t = \varphi(X_{t-1})$, $t \geq 1$ où X_0 est un vecteur aléatoire donné de E , représentant l’état initial. Néanmoins un tel système purement théorique n’est pas très réaliste car les observations X_t sont en général corrompues par un bruit. Supposons donc que l’on observe n observations bruitées Y_1, Y_2, \dots, Y_n menant au modèle plus naturel $Y_t = \psi(Y_{t-1}, \Delta_t)$, $t = 1, \dots, n$ où ψ est une fonction mesurable : $E \times F \rightarrow E$, ($F \in \mathcal{B}_{\mathbb{R}^d}$) et (Δ_t) représente le bruit.

Dans l’article ([2],[21]) écrit en collaboration avec D. Bosq et D.Guégan, nous avons introduit et étudié l’influence de différents bruits typiques. Nous avons alors estimé la densité f (lorsqu’elle existe) de la mesure invariante μ associée au modèle non bruité $Y_t = \varphi(Y_{t-1})$ pour des observations Y_t données comme ci-dessus. Deux cas particuliers ont été considérés dans l’article : le premier est le *modèle de déconvolution* classique (où la loi du bruit est supposée connue) et le second correspond à la *propagation de petites erreurs*. Nous montrons que sous des hypothèses adéquates les *vitesse optimales* du cas indépendant sont encore *atteintes* par les estimateurs. Finalement des *simulations* sont proposées dans [21], confortant ces résultats théoriques.

4 Estimation de paramètres structurels

Travaux [4], [6], [7], [12], [14], [18]–[19], [22], [26].

Exposés [12], [14]–[15], [20]–[26].

4.1 Estimation d’une variance asymptotique

Avec F. Merlevède nous avons étudié, dans [18], l’estimation du terme de variance asymptotique obtenue sous les conditions de Castellana et Leadbetter (cf section 1.1). Nous utilisons un estimateur basé sur le temps local

car il présente en effet l'intérêt d'être *sans biais* et d'atteindre, avec des hypothèses usuelles, la vitesse paramétrique T^{-1} (Bosq and Davydov, 1999). Rappelons tout d'abord que si X est observé sur $[0, T]$, sa mesure d'occupation sur $[0, T]$ est définie par $\nu_T(B) = \int_0^T \mathbf{1}_B(X_t) dt$, $B \in \mathcal{B}_{\mathbb{R}}$, où $\mathcal{B}_{\mathbb{R}}$ est la tribu des ensembles Boréliens dans \mathbb{R} , et la mesure empirique μ_T sur $[0, T]$ est alors donnée par $\nu_T T^{-1}$. Si de plus ν_T est p.s. absolument continue par rapport à la mesure de Lebesgue λ sur \mathbb{R} , alors un temps local (ou densité d'occupation) pour X est défini comme une fonction mesurable $l_T(x, \omega)$ telle que $l_T(\cdot, \omega)$ soit une version de $\frac{d\nu_T}{d\lambda}$ pour presque tout ω dans Ω . Dans [18] et en supposant en particulier le processus (X_t) strictement stationnaire, nous avons commencé par établir un *théorème central limite* pour cet estimateur temps local. La variance asymptotique étant inconnue (ce qui empêche alors la construction pratique d'intervalles de confiance), nous avons construit et étudié un estimateur adapté (dérivé d'un estimateur de la densité spectrale) pour cette variance. Nous donnons un éventail de vitesses de convergence possibles pour l'erreur quadratique, dépendant essentiellement de l'existence de moments d'ordre p (avec $p \geq 2$) du temps local.

4.2 Estimation de l'index de régularité d'un processus

Soit $\{X_t, t \in \mathbb{R}\}$ un processus réel, séparable et mesurable dont les trajectoires satisfont localement une *condition de type Hölder* :

$$\limsup_{\delta \searrow 0} \sup_{t_1 \leq t \leq t_2 - \delta} \sup_{0 \leq s \leq \delta} |X_{t+s} - X_t| / s^\gamma L(s) \leq 1, \quad 0 \leq t_1 < t_2$$

où $0 < \gamma \leq 1$ et $L(\cdot)$ est une fonction à variation lente. Notre but est alors d'estimer γ_0 avec $\gamma_0 = \sup \{\gamma, \gamma \in \Gamma\}$ où Γ représente l'ensemble des γ vérifiant la relation précédente. Évidemment, le premier intérêt consiste à rechercher de l'information sur la *régularité locale* des trajectoires du processus observé. Notons cependant qu'en tenant compte des travaux exposés ci-dessus, un estimateur préliminaire pour γ_0 pourra servir, dans un deuxième temps, à la construction d'estimateurs de type "plug-in" pour la densité.

Dans le cas spécifique des processus Gaussiens stationnaires, il existe une relation simple et connue entre la dimension fractale des trajectoires et l'index fractal des covariances. Ceci a mené à un grand nombre de travaux d'estimation dans le cas Gaussien. Pour un futur contexte d'estimation non paramétrique, je me suis placée dans un cadre plus général en considérant des processus non nécessairement Gaussiens et en supposant d'autre part :

- soit que la trajectoire est observée globalement sur $[0, T_N]$ où (T_N) est une suite de réels tels que $T_N \nearrow \infty$, $T_{N+1} - T_N \geq a > 0$, $N \geq 1$ avec a

constante,

- soit que l'on dispose de n observations échantillonnées $X_0, \dots, X_{(n-1)\delta_n}$ où δ_n est le pas d'échantillonnage tel que $\delta_n \rightarrow 0$, $n\delta_n \rightarrow \infty$ quand $n \rightarrow \infty$ (c'est-à-dire que les observations sont fréquentes et disponibles sur une longue durée).

J'ai alors étudié pour ces deux cas le comportement *asymptotique* et *presque sûr* d'une famille d'estimateurs basés sur des petits accroissements de processus. De plus, les estimateurs les plus efficaces étant particulièrement sensibles à la valeur du coefficient de *régularité du second ordre*, ces résultats sont complétés en donnant un estimateur préliminaire de ce paramètre avec sa vitesse exacte de convergence. Finalement, des *simulations* sont effectuées dans [19] et permettent de comparer le comportement de l'ensemble des estimateurs proposés. La mise en parallèle des résultats [22] (cas continu) et [12],[19] (cas échantillonné) nous permet de remarquer que, comme attendu, la connaissance globale de la trajectoire sur $[0, T_N]$ apporte plus d'informations, ce qui permet d'améliorer les vitesses de convergence et le comportement général de nos estimateurs.

4.3 Estimation du nombre de dérivées d'un processus Gaussien

Dans le travail joint avec C. Vial (voir [26] et [6],[14]), nous considérons un processus Gaussien réel, X , de régularité inconnue $p_0 \in \mathbb{N}$ dans le cas où la dérivée d'ordre p_0 en moyenne quadratique, notée $X^{(p_0)}$, est supposée satisfaire une condition de type Hölder. Dans un premier temps, à partir des observations discrétisées $X(t_1), \dots, X(t_n)$, on étudie la reconstruction de $X(t)$, $t \in [0, 1]$, par $\tilde{X}_r(t)$. Ici $\tilde{X}_r(t)$ dénote une interpolation polynômiale par morceaux, de degré $r \geq 1$. On montre que l'erreur quadratique d'interpolation $E(X(t) - \tilde{X}_r(t))^2$ décroît quand r augmente, puis se stabilise dès que r dépasse p_0 . Ce point nous permet, dans une deuxième partie, de construire un estimateur \hat{p} du paramètre p_0 grâce à un critère empirique basé sur l'interpolation polynômiale par morceaux. On établit alors, la convergence presque sûre de \hat{p} vers p_0 via une inégalité exponentielle pour $P(\hat{p} \neq p_0)$. Finalement, on montre que $\tilde{X}_{\hat{p}}(t)$ converge presque sûrement vers $X(t)$ avec une vitesse quasi-optimale.

Nous continuons à travailler sur ce sujet, en étudiant actuellement un deuxième estimateur basé sur les variations quadratiques du processus (voir [7]). De plus, par la suite, nous comptons appliquer ces résultats (via des méthodes de type plug-in) à des problèmes d'estimation et de prévision où la régularité p_0 était, jusqu'à présent, supposée connue.

Perspectives de recherche

Outre les travaux en cours de développement cités précédemment, voici quelques thèmes dont j'envisage le développement dans les années à venir.

- **Estimation quand les paramètres structurels sont inconnus.** On dispose d'estimateurs adaptatifs dans le cas où les paramètres γ_0 et/ou r_0 sont inconnus. Une première question concerne l'application de telles méthodes à des données réelles : comment se comportent de tels estimateurs ? Sont-ils robustes devant des données aberrantes ? ... Il serait également intéressant de regarder le comportement d'estimateurs de la densité de type “plug-in” (en se basant sur les résultats obtenus pour l'estimation de γ_0). Une comparaison générale (tant d'un point de vue théorique que numérique) pourra alors être faite avec les estimateurs adaptatifs du type précédent.
- **Modèle additif.** Comme on l'a vu, le comportement des estimateurs dépend de γ_0 où γ_0 peut être interprété comme la somme des irrégularités apportées par chaque composante du processus d -dimensionnel (avec $d \geq 2$). On peut ainsi chercher à étudier ce phénomène pour l'estimation de la régression, tout particulièrement au sein d'un modèle additif qui, par construction, va différencier le rôle joué par chacune de ces composantes.
- **Prévision.** Pour des raisons soit physiques, soit liées à la nature des appareils délivrant les mesures, l'intégralité des courbes n'est pas toujours à la disposition du Statisticien, les données à traiter se présentant plutôt sous la forme de n observations discrétisées X_{t_1}, \dots, X_{t_n} . Nous avons vu ici que la localisation optimale des instants t_i devait se faire en fonction de la nature de la trajectoire sous-jacente (en particulier des instants t_i trop proches les uns des autres peuvent conduire à des estimateurs inconsistants). Dans le cadre de la prévision non paramétrique et en s'appuyant sur les travaux développés dans [28], les axes suivants pourraient être ainsi privilégiés :
 - *Prévision du processus sur un intervalle futur.* La problématique est la suivante : on dispose de l'observation du processus à N instants fréquents t_1, \dots, t_N et on s'intéresse à la prévision du phénomène, non pas en un instant précis, mais sur tout un intervalle. Supposons par exemple données les n observations sur un intervalle $[0, T_n]$, on désire alors effectuer une prévision du phénomène X_t sur l'intervalle $[T_n, T_n + h]$, $h > 0$. Ce travail peut alors se faire en 2 étapes : tout d'abord on envisage la reconstruction de la trajectoire sur $[0, T_n]$ par

des techniques d'interpolation optimales (encore peu développées en dehors du cadre Gaussien) en vue d'utiliser, dans un deuxième temps, les travaux existant sur la prévision des processus à temps continu par des méthodes fonctionnelles.

- *Prévision de franchissement de niveaux.* Nous pouvons également étudier la prévision de franchissement de niveaux grâce à l'utilisation des prédicteurs précédemment cités. En effet jusqu'à présent peu de travaux semblent exister en dehors du cas Gaussien. Un des buts serait ainsi de pouvoir définir des *zones d'alarmes*, dont les applications seraient nombreuses dans ce contexte de prévision.