

OFFRE DE THÈSE EN STATISTIQUE DES MODÈLES DYNAMIQUES ALÉATOIRES NON-MARKOVIENS

PIERRE ALQUIER* AND NICOLAS MARIE†

SUJET

Depuis la fin des années 1990, les probabilistes et statisticiens s'intéressent à la modélisation des séries temporelles issues de processus non-markoviens. Cela pose des difficultés à la fois en temps continu et en temps discret. L'évolution des outils d'analyse stochastique et de la théorie des matrices aléatoires ont permis d'envisager de nouvelles approches pour étudier de telles séries temporelles ces dernières années.

La première partie de ce sujet de thèse porte sur la complétion et la classification de séries temporelles multi-variées en grande dimension. Le doctorant devra se familiariser avec les matrices aléatoires et la signature des fonctions continues et à variation bornée. La seconde partie porte sur la sélection de fenêtre pour un estimateur non-paramétrique de la tendance d'une équation différentielle dirigée par le mouvement brownien fractionnaire (EDf). Cette seconde partie s'inscrit dans le cadre du projet PRC (CNRS) France-Russie DIRaH dont Nicolas Marie est membre.

Dans la mesure où les deux parties de ce sujet de thèse intègrent une composante sélection de modèle, des applications en Statistical Learning sont envisageables.

Partie I - Complétion et classification de séries temporelles multi-variées en grande dimension. D'une part, soit le modèle

$$\mathbf{X} = \mathbf{M}_{r,\tau} + \varepsilon,$$

où \mathbf{X} est une matrice $d \times T$ dont les lignes sont des séries temporelles, ε est une matrice aléatoire dont les lignes sont indépendantes et sous-gaussiennes, $\mathbf{M}_{r,\tau}$ est la matrice dont les lignes sont les tendances des séries temporelles considérées, r est le rang de $\mathbf{M}_{r,\tau}$ supposé faible et τ est un paramètre caractérisant une propriété spécifique à la tendance d'une série temporelle comme la τ -périodicité.

Dans Alquier et Marie [1], un contrôle en $O(r(d + \tau))$ et non $O(r(d + T))$ du risque quadratique de l'estimateur des moindres carrés de $\mathbf{M}_{r,\tau}$ a été établi en tirant parti de la structure de série temporelle des lignes de \mathbf{X} . Il en découle un contrôle du risque quadratique d'un estimateur adaptatif pour (r, τ) sélectionné par pénalisation.

Via une démarche analogue, il est envisageable d'exploiter la structure de série temporelle des lignes de la matrice \mathbf{X} pour améliorer la borne de risque connue (cf. Koltchinskii et al. [3]) d'un estimateur de ses valeurs manquantes (complétion).

D'autre part, la signature est un opérateur qui à une trajectoire multidimensionnelle, continue et à variation bornée associe un vecteur dont les composantes ont une signification géométrique. La signature avait initialement été introduite en analyse stochastique par Terry Lyons pour donner un sens aux équations différentielles dirigées par des processus à trajectoires très irrégulières comme le mouvement brownien fractionnaire. Depuis 2014, T. Lyons explore l'utilisation de la signature en Machine Learning et notamment pour classer des séries temporelles (cf. Levin

et Lyons [4]).

Les travaux existants se bornent à des considérations algorithmiques. Il serait intéressant d'établir des inégalités d'oracle pour ces classificateurs et traiter la question de la sélection de l'ordre de troncature de la signature.

Partie II - Estimation de la tendance d'une Edf réfléchie et sélection de fenêtre. La seconde partie du travail de thèse portera sur l'estimation de la tendance du processus X_ε , solution du problème de réflexion de Skorokhod

$$(1) \quad \begin{cases} X_\varepsilon(t) = \int_0^t b(X_\varepsilon(u))du + \varepsilon B(t) + Y_\varepsilon(t) \\ -\dot{Y}_\varepsilon(t) \in \mathcal{N}_{[m(t), M(t)]}(X_\varepsilon(t)) \text{ } |DY_\varepsilon| \text{-p.p. avec } Y_\varepsilon(0) = x_0 \end{cases} ; t \in [0, T],$$

où $b : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction lipschitzienne, $\varepsilon > 0$, B est un mouvement brownien fractionnaire de paramètre de Hurst $H \in [1/2, 1[$ et $m, M : [0, T] \rightarrow \mathbb{R}$ sont des fonctions lipschitziennes telles que $m(t) \leq M(t)$ pour tout $t \in [0, T]$.

D'une part, l'existence, l'unicité et l'approximation de la solution du Problème (1) ont été étudiées dans Castaing et al. [2]. D'autre part, la consistance, une vitesse de convergence et la loi limite de l'estimateur

$$\hat{\tau}_\varepsilon(t) := \frac{1}{h_\varepsilon} \int_0^t \int_0^T K\left(\frac{s-u}{h_\varepsilon}\right) dX_\varepsilon(s)du ; t \in [0, T]$$

de la tendance du Problème (1) ont été établies dans Marie [5], où $K : \mathbb{R} \rightarrow \mathbb{R}_+$ désigne un noyau.

La question de la sélection de la fenêtre h_ε n'a pas encore été traitée. La cross-validation et une méthode type Goldenshluger-Lepski sont envisageables. Pour la seconde méthode, un contrôle du risque quadratique de l'estimateur adaptatif devra être obtenu au moins dans le cas $H = 1/2$ et idéalement dans le cas $H > 1/2$.

ENCADREMENT

La thèse sera encadrée par :

- **Pierre ALQUIER**
Chercheur au RIKEN AIP
Ancien professeur de l'ENSAE Paris
- **Nicolas MARIE**
Enseignant-chercheur à l'ESME Sudria (Cti)
Chercheur associé de l'université Paris Nanterre

DÉROULEMENT DE LA THÈSE ET MODALITÉS DE CANDIDATURE

Le candidat retenu intégrera l'équipe de recherche Modélisation Mathématique et Applications de l'ESME Sudria, une école d'ingénieurs accréditée par la Cti, et le laboratoire Modal'X de l'université Paris Nanterre à partir de septembre 2020 et pour une durée de 3 ans.

Le doctorant sera basé sur le site d'Ivry-sur-Seine de l'ESME Sudria. Il effectuera son service d'enseignements en tronc commun. Par ailleurs, le doctorant pourra effectuer de courtes visites, ou éventuellement un stage de 6 mois, au RIKEN AIP à Tokyo pour travailler avec Pierre Alquier.

Le dossier de candidature est à adresser à Nicolas Marie (nicolas.marie@esme.fr) et doit comprendre :

- (1) Un curriculum vitae détaillé.

- (2) Un relevé des notes de M2.
- (3) Une lettre de motivation.
- (4) Une lettre de recommandation.

REFERENCES

- [1] P. Alquier et N. Marie. *Matrix Factorization for Multivariate Time Series Analysis*. Electronic Journal of Statistics 13, 2, 4346-4366, 2019.
- [2] C. Castaing, N. Marie et P. Raynaud de Fitte. *Sweeping Processes Perturbed by Rough Signals*. Soumis, arXiv: 1702.06495.
- [3] V. Koltchinskii, K. Lounici et A. Tsybakov. *Nuclear-Norm Penalization and Optimal Rates for Noisy Low-Rank Matrix Completion*. The Annals of Statistics 39, 5, 2302-2329, 2011.
- [4] D. Levin et T. Lyons. *Learning from the Past, Predicting the Statistics fo the Future, Learning an Evolving System*. arXiv: 1309.0260.
- [5] N. Marie. *Nonparametric Estimation of the Trend in Reflected Fractional SDE*. Statistics and Probability Letters, DOI: 10.1016/j.spl.2019.108659, 2019.

*RIKEN CENTER FOR ADVANCED INTELLIGENCE PROJECT, TOKYO, JAPON
Email address: pierre.alquier.stat@gmail.com

†ESME SUDRIA, PARIS, FRANCE
Email address: nicolas.marie@esme.fr

†LABORATOIRE MODAL'X, UNIVERSITÉ PARIS NANTERRE, NANTERRE, FRANCE
Email address: nmarie@parisnanterre.fr