Estimation et détection de sauts. Applications aux processus linéaires à valeurs dans $D[0,1]$.

Denis Bosq

LSTA, Université Pierre et Marie Curie - Paris 6, 4 place Jussieu, 75005 Paris, France, denis.bosq@upmc.fr

Résumé. Cet exposé est consacré à l'estimation de l'intensité et de la densité des sauts pour des variables aléatoires à valeurs dans $D[0,1]$, et à la construction de détecteurs pour des sauts constants ou aléatoires. Des théorèmes limites sont obtenus dans le cadre d'observations continues ou à haute fréquence. On en déduit des applications aux sauts de moyennes mobiles et processus autorégressifs à valeurs dans $D[0,1]$. Nous étudions également le cas particulier où il y a une infinité de sauts. Notons que notre approche est différente de celle qui consiste à étudier les sauts des semi-martingales.

Mots-clés. Sauts, processus linéaires fonctionnels, càdlàg, théorèmes limites, estimation, détection.

Abstract. This talk is devoted to the estimation of the intensity and the density of jumps for $D[0,1]$-valued random variables and the construction of detectors for constant or random jumps. Limit theorems are obtained in the context of continuous observations or high frequency data. Applications to jumps for $D[0,1]$-valued moving average and autoregressive processes are considered. We also study the special case where there is an infinity of jumps. Thus, our approach is somewhat different from that which consists in studying jumps in semimartingales.

Keywords. Jumps, functional linear processes, cadlag, limit theorems, estimation, detection

Résumé long

La prévision des processus autorégressifs fonctionnels dans les espaces de Banach et de Hilbert a été beaucoup étudiée, voir Bosq and Blanke (2007), Ferraty and Romain (2011), Horváth and Kokoszka (2012) et Bosq (2014) pour des résultats et des références. En général, on considère des fonctions aléatoires continues avec des applications à la consommation d'électricité, les électrocardiogrammes, les variations de température, ... Les intervalles de temps associés étant un jour, une semaine, un mois, ...

Cet exposé est consacré au cas où des sauts apparaissent dans des applications en finance, économie, variations climatiques, modélisation du prix de l'électricité, ... On peut également remarquer que des grandes variations continues peuvent être considérées comme des sauts puisque dans cette situation, la qualité de la prévision n'est pas bonne.

L'espace naturel pour étudier les sauts est $D = D\,[0,1]$, c'est-à-dire l'espace des fonctions réelles définies sur [0,1] et càdlàg (continues à droite et avec une limite à gauche). Il est bien connu que si D est muni de la norme uniforme, il devient un espace non séparable, et que cela crée des problèmes de mesurabilité. Par conséquent, il vaut mieux remplacer cette norme par la distance de Skorohod, cf Billingsley (1999).

Maintenant, on observe des copies indépendantes ou dépendantes d'une variable aléatoire X à valeurs dans D et on veut estimer l'intensité et la densité des sauts, et si les données sont en temps discret, détecter la position des sauts. Dans la première partie de l'exposé, nous supposons qu'il existe un saut en $t_0 \in \,]\,0,1[$, où t_0 est constant et où les observations sont en temps continu. Alors, il est facile d'obtenir des théorèmes limites pour les estimateurs empiriques de $E(X(t_0) - X(t_0-))$ et $E\,|X(t_0) - X(t_0-)|$, plus précisément la loi forte des grands nombres (LFGN) et le théorème central limite (TCL).

Ensuite, on étudie des données haute-fréquence (HF). La situation est plus compliquée puisqu'on doit d'abord détecter t_0 avant d'estimer l'intensité du saut. Si les trajectoires de X satisfont une condition de Hölder sur $[0, t_0[$ et sur $[t_0, 1]$, on obtient un détecteur consistant de t_0 avec une vitesse de convergence et l'utilisation de ce détecteur permet d'obtenir la LFGN et le TCL.

Dans les parties suivantes, on étudie le cas où la position du saut, soit T, est aléatoire. On obtient à nouveau des théorèmes limites en temps continu et en temps discret. En particulier, on s'intéresse au comportement asymptotique de l'estimateur à noyau de la densité de T, à partir de données discrètes. Puis, on considère le cas où il y a deux sauts aléatoires.

Nous appliquons les résultats précédents aux moyennes mobiles et aux processus autorégressifs à valeurs dans $D([0,1])$. Enfin, nous étudions le cas particulier où il y a une infinité de sauts.

Bibliographie

P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

D. Bosq. Computing the best linear predictor in a Hilbert space. Applications to general ARMAH processes. *J. Multivariate Anal.*, 124:436–450, 2014.

D. Bosq. Estimating and detecting jumps. Applications to $D\,[0,1]$-valued linear processes *Festschrift for Paul Deheuvels*, to appear, Springer, 2014.

D. Bosq and D. Blanke. *Prediction and inference in large dimensions.* Wiley series in probability and statistics. Wiley-Dunod, 2007.

F. Ferraty and Y. Romain, editors. *The Oxford handbook of functional data analysis.* Oxford University Press, Oxford, 2011.

L. Horváth and P. Kokoszka. *Inference for functional data with applications.* Springer Series in Statistics. Springer, New York, 2012.