

Proposition de communication (journée doctorant)

Thibault Espinasse

23 mars 2011

Actuellement en 3eme année de thèse, sous la direction de Fabrice Gamboa et Jean-Michel Loubes à l'Institut de Mathématiques de Toulouse, je travaille en amont sur des problèmes posés par l'entreprise Médiamobile (spécialisée dans la prédiction de trafic routier). L'objet de cette communication est un travail effectué avec Fabrice Gamboa et Jean-Michel Loubes sur le filtrage de données manquantes à partir de données de vitesse obtenues sur le réseau parisien. Ce travail est consitué d'un premier papier théorique proposant une modélisation de la stationarité d'un processus indexé par un graphe, et d'un travail appliqué en cours d'élaboration qui concerne la prédiction de données manquantes sur le réseau.

1 Stationarité de processus Gaussiens indexés par un graphe, approximation de Whittle

Dans le premier travail, on propose une construction générale de processus indexés par des graphes. La motivation initiale est de modéliser des processus indexés par un graphe (par exemple de vitesses sur le réseau routier), et dont les corrélations ne dépendent que de la structure du réseau. Pour cela, on cherche à généraliser les définitions existantes pour les graphes distance-transitifs [7]. En particulier, cette construction doit être compatible avec le cas de \mathbb{Z}^d (voir par exemple [6]) et le cas des arbres homogènes (voir [1], [2]).

On s'intéresse ensuite au cas particulier des *ARMA*, afin de mettre en place une approximation de Whittle (voir par exemple [3],[5]), classique pour les séries chronologiques, pour les processus ainsi définis. Tout comme pour les séries chronologiques, cette représentation fait apparaître la mesure spectrale du graphe concerné. La convergence peut-être obtenue dans les cas où la densité est assez régulière. Tout comme dans $\mathbb{Z}^d, d \geq 2$ [6], on peut prouver la normalité asymptotique en débiaisant cet estimateur.

2 Mesure spectrale du réseau routier et prédiction de données manquantes

Pour mettre en pratique l'approximation de Whittle, la mesure spectrale du graphe routier doit-être estimée. Ce travail est en cours. Les premières simulations laissent supposer que la mesure spectrale de la matrice d'adjacence

converge vers un mélange de lois. Il s'agit alors de proposer un modèle adéquat puis d'estimer les paramètres concernés.

Tout comme pour les séries chronologiques, le filtrage des valeurs manquantes consiste en une estimation de l'opérateur de projection des valeurs manquantes sur les variables observées. Bickel a prouvé dans [4] que l'inverse de la matrice de covariance peut-être estimé efficacement en inversant la covariance empirique, lorsque le processus est assez régulier. Dans un cadre non paramétrique, on utilisera l'estimation empirique de l'opérateur de covariance pour estimer ce projecteur.

References

- [1] Jean-Pierre Arnaud. Fonctions sphériques et fonctions définies positives sur l'arbre homogène. *C. R. Acad. Sci. Paris Sér. A-B*, 290(2):A99–A101, 1980.
- [2] Jean-Pierre Arnaud. Stationary processes indexed by a homogeneous tree. *Ann. Probab.*, 22(1):195–218, 1994.
- [3] Robert Azencott and Didier Dacunha-Castelle. *Series of irregular observations*. Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1986. Forecasting and model building.
- [4] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008.
- [5] Ulf Grenander and Gábor Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.
- [6] Xavier Guyon. Parameter estimation for a stationary process on a d -dimensional lattice. *Biometrika*, 69(1):95–105, 1982.
- [7] Herbert Heyer. Stationary random fields over graphs and related structures. In *Wavelets, multiscale systems and hypercomplex analysis*, volume 167 of *Oper. Theory Adv. Appl.*, pages 157–171. Birkhäuser, Basel, 2006.