Given the points of regime change, estimation of parameters of the current regime by filtering of the PDMP, properties of the estimators.

ANR-09-SEGI-004

Partial report of team CQFD on task 1.1.2

L'objectif de cette tâche est de proposer des méthodes d'estimation des paramètres d'un Processus Markovien Déterministe par Morceaux (PDMP en abrégé). Le partenaire concerné est l'INRIA CQFD. Les personnes sont R. Azaïs, F. Dufour et A. Gégout-Petit.

1 Contexte

L'un des objectifs du projet ANR Fautocoes est d'utiliser les PDMP pour modéliser les systèmes physiques complexes et les phénomènes tels que la propagation de fissures dans les structures mécaniques. La modélisation par les PDMP est intéressante dans ce contexte car elle permet de s'appuyer sur les lois de propagation de fissures issues de la mécanique ; ces lois intevenant dans le flot déterministe du PDMP. Il reste donc à modéliser les changements de régime de propagation en des temps aléatoires par des changements stochastiques de loi mécanique ou des paramètres de ces lois en cours de propagation. La variable physique (comme la longueur de la fissure) est alors unPDMP. Une étape importante de la modélisation est donc l'estimation des paramètres de la dynamique d'un PDMP. Le flot étant déterministe, cette estimation concerne essentiellement l'estimation de l'intensité de saut λ et celle du noyau de transition q.

Le premier problème relève de l'analyse de survie. L'analyse de survie est un domaine de recherche historique et particulièrement développé dans les études de fiabilité [8] et de biostatistique [1] . D'un point de vue mathématique, un PDMP est un processus ponctuel marqué et les techniques d'estimation de ces processus ont déjà été développée par de nombreux auteurs (voir [1], [2], [4] et les références incluses). Cependant l'estimation de l'intensité d'un PDMP pose des problèmes spécifiques dans la mesure ou l'intensité est fonction de la variable physique.

2 Approche

Nous nous proposons d'estimer l'intensité λ dans le contexte spécifique des PDMP. La définition de l'intensité donnée par [3] est la suivante :

$$P(T_{n+1} > t) = exp(-\int_{T_n}^{T_n+t} \lambda(\Phi(X_s))ds)\mathbb{I}_{t<t^*(X_{T_n})}.$$

Mais de part les propriétés du flot Φ, on peut réécrire :

$$\int_{T_n}^{T_n+t} \lambda(\Phi(X_s))ds = \int_0^t \tilde{\lambda}(X_{T_n}, s))ds.$$

Nous supposons que l'ensemble Z des valeurs possibles de X aux instants de saut, c'est-à-dire les valeurs des X_{T_n} est fini, soit $Z = \{z_1, z_2, \dots, z_l\}$. Si pour un z_i donné, $(X_{T_n} = z_i)$,l'intensité de $S_{n+1} = T_{n+1} - T_n$ est $\tilde{\lambda}(z_i, s)$. De part le caractère markovien d'un PDMP, pour tous les n tels que $(X_{T_n} = z_i)$, l'intensité de saut de S_{n+1} est la même. Si z_i est récurrent, il est donc intéressant de regarder le processus X_t sur un temps long pour faire des statistiques sur $\tilde{\lambda}(z_i, s)$.

L'estimation de l'intensité s'obtient en général à partir de celle de l'intensité cumulée qui est par essence discontinue puisque constante par morceaux. Il convient donc de lisser l'intensité cumulée pour obtenir l'intensité. Ce lissage peut-être obtenu par des méthodes à noyaux [9]. D'autre méthode de lissage sont disponibles pour la densité d'une variable aléatoire sous l'hypothèse qu'elle soit décomposable sur une base orthonormale de splines, par exemple Koo & al [6] et de Kooperberg [7] estiment les coefficients de la décomposition d'une densité sur une base de fonctions splines.

3 Résultats

Nous avons étudié le problème de l'estimation de $\tilde{\lambda}(z_k, s)$ dans le cadre présenté ci-dessus. Sous l'hypothèse que le noyau de transition q ne dépend par du temps passé dans le mode sortant, on peut proposer un estimateur de type Nelson Aalen car processus de comptage

$$N_t = \sum_{i=1}^{n} \mathbb{I}_{\{S_{i+1} \leq t\}} \mathbb{I}_{\{X_{T_i = z_k}\}}$$

admet une intensité de type multiplicative dans la filtration

$$\mathcal{F}_t = \sigma\{Z_i, \mathbb{I}_{S_{i+1} \leq s}, 0 \leq s \leq t; 0 \leq i \leq n\}$$

cette filtration n'est pas du tout la filtration historique du PDMP qui elle voit apparaître les temps $(T_{i+1} = T_i + S_{i+1}, Z_{i+1})$ les après les autres alors que dans $\mathcal{F}_t$ les Z_i sont tous connus à l'origine et les S_i démarrent tous à l'origine comme si on étudiait des temps de saut indépendants. Le caractère markovien du processus et l'hypothèse sur le noyau implique que l'intensité de $\mathbb{I}_{\{S_{i+1} \leq t\}} \mathbb{I}_{\{X_{T_i} = z_k\}}$ dans $\mathcal{F}_t$ est $\mathbb{I}_{\{X_{T_i} = z_k\}} \tilde{\lambda}(z_k, s)$. Les estimateurs de type Nelson-Aalen sont donc adéquats pour cette estimation et nous obtenons des résultats de normalité asymptotique quand n tend vers l'infini (ou de manière équivalente le temps t).

D'autre part nous avons adapté les méthode de lissage de Koo & al et de [6] et de Kooperberg [7] au problèmes de l'estimation de l'intensité par l'intermédiaire de l'estimateur de Nelson Aalen de notre intensité. Nous les avons aussi étendus aux plus général de décomposition sur une base orthonormale de fonctions régulières (splines, Fourier, polynômes, etc...)

4 Dissémination des résultats

Les résultats qui font l'objet de la thèse de Romain Azaïs, sont en cours de rédaction et devrait être soumis pour publication dans le courant de l'année 2011.

Références

[1] Andersen P.K., Borgan O., Gill R.D., Keiding N. (1993) : *Statistical models based on counting processes.* Springer-Verlag, Springer Series in Statistics.

[2] Brémaud P. (1981) : *Point processes and queues : Martingale dynamics.* Springer-Verlag, Springer Series in Statistics.

[3] Davis M.H.A. (1993) : *Markov models and optimzation.* Monographs on Statistics and Applied Probability, vol.49, Chapman & Hall, London.

[4] Elliot R.J., Aggoun L., Moore J.B. (1995) : *Hidden Markov Models : Estimation and control.* Springer, Applications of Mathematics, 29.

[5] Jacobsen M. (1982) : *Statistical Analysis of Counting Processes.* Lecture Notes in Statistics 12. Springer-Verlag, New-York.

[6] Koo J.-Y., Kooperberg C., Park J. (1999) : *Logspline density estimation under censoring and truncation.* Board of the Foundation of the Scandinavian Journal of Statistic Vol 26 : 87-105.

[7] Kooperberg C., Stone C.J. (1991) : *A study of logspline density estimation.* Computational Statistics & Data Analysis 12, 327-347 North-Holland.

[8] Meeker W.Q. and Escobar L.A. (1998) : *Statistical Methods for Reliability Data.* John Wiley & Sons.

[9] Ramlau-Hansen H. (1983) : *Smoothing Counting Process Intensities by Means of Kernel Functions.* Ann. Statist. Volume 11, Number 2, 453-466, 1983.