UNIVERSITÉ DES ANTILLES ET DE LA GUYANE

THÈSE

pour obtenir le grade de
Docteur de l'Université des Antilles et de la Guyane

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

présentée par

Florencia CHIMARD
(florencia.chimard@univ-ag.fr)

Mélanges de Processus Ponctuels spatio-temporels et approche bayésienne semi-paramétrique

soutenue publiquement le jeudi 16 Décembre 2010

Composition du jury :

Président :	Professeur Jeffrey ROSENTHAL	University of Toronto, Canada
Rapporteur :	Professeur Lance WALLER	Emory University, Atlanta USA
Rapporteur :	Professeur Patrick BROWN	University of Toronto, Canada
Directeur :	Professeur Jean VAILLANT	Université des Antilles et de la Guyane
Co-directeur :	Professeur Richard EMILION	Université d'Orléans, France
Examinateur :	Professeur Alex MERIL	Université des Antilles et de la Guyane

A mes parents,
A mon frère Thony qui nous a quitté prématurément.

Résumé

Mots-clés : Processus ponctuel, processus de Dirichlet, loi a priori stick-breaking, modèle bayésien hiérarchique, méthodes MCMC

Les processus ponctuels sont souvent utilisés comme modèles de répartitions spatiales ou spatio-temporelles d'occurrences. Pour notre part, dans cette thèse, nous nous intéressons à l'analyse statistique bayésienne de ces processus dans le cas où nous disposons de k cartes exhaustives à des dates d'observations échelonnées dans le temps. Nous proposons deux contextes d'étude.

Tout d'abord, nous considérons des occurrences constituant la réalisation d'un processus de Cox spatio-temporel dont l'intensité est associée à un processus shot noise généralisé. Le modèle correspond à une mesure d'intensité liée à des contributions générées par un processus caché de Poisson et qui suivent un processus de Dirichlet centré sur la loi Gamma. A partir des positions spatiales des occurrences observées entre plusieurs paires de dates d'observations consécutives, nous proposons d'inférer sur les paramètres d'intérêt à l'aide de méthodes MCMC dans le cadre d'un modèle bayésien hiérarchique. Un algorithme avec augmentation des données est proposé et testé sur des jeux de données artificielles.

D'autre part, nous analysons la situation où l'ensemble d'étude est un ensemble discret de positions possibles pour chaque occurrence du phénomène. La présence/absence d'une unique occurrence en une position donnée implique que nous aurons des données binaires. Par conséquent, nous développons un modèle de mélange de lois de Bernoulli avec un paramètre d'intensité d'arrière plan suivant un processus autorégressif d'ordre 1 log-gaussien. Nous utilisons une approche bayésienne hiérarchique pour mener à bien l'inférence statistique de notre modèle. Nous développons un algorithme Metropolis-within-Gibbs pour calculer la loi a posteriori des paramètres d'intérêt. Des tests sont effectués sur des données artificielles et sur des données concernant le virus de la feuille jaune de la canne à sucre.

Abstract

Keywords : Point process, Dirichlet process, stick-breaking prior, hierarchical bayesian model, MCMC methods

Point processes are often used as tools for describing spatial or spatio-temporal point patterns. In this Phd dissertation, we give an overview of bayesian statistical analysis for point processes and recent tools like the Dirichlet process and its diverse extensions. We focus on situations where the available data are maps of the studied point process at different observations dates. Two contexts are considered.

Firstly, we consider occurrences of events in a studied area forming the realization of a spatio-temporal Cox process directed by a generalized shot noise intensity measure. A hidden Poisson process generates contributions to the intensity measure which are distributed according to a Dirichlet process centered on the Gamma distribution. For data consisting of spatial locations of occurrences between several pairs of consecutive observation dates, we develop statistical inference about the parameters of interest by means of MCMC methods within the framework of hierarchical bayesian modeling. A data augmentation algorithm is introduced and tested on artificial data.

Secondly, we analyse the case where the point process support is discrete with at most one occurrence for a given element of the support. For such binary data, we present and discuss models based on Bernoulli distribution mixture with a background intensity following a log-gaussian. The statistical inference for these models is developped by using a hierarchical bayesian approach. Tests are carried out on artificial data and data from Yellow Leaf Sugarcane Virus observations.

Remerciements

Je voudrais en premier lieu exprimer toute ma gratitude au professeur Jean VAILLANT, pour sa disponibilité, son soutien, ses encouragements, au cours de ces trois années, mais aussi pour avoir fait preuve, à plusieurs reprises, de sa confiance en mon travail de recherche.

Je remercie également le professeur Richard EMILION, co-directeur de thèse, pour ses précieuses remarques et les conseils qu'il m'a prodigués.

Je voudrais exprimer toute ma reconnaissance au professeur Jeffrey ROSENTHAL qui m'a fait l'honneur de participer et présider ce jury de thèse. Je le remercie de m'avoir accueilli au sein de son département à l'Université de Toronto pour un séjour de recherche auquel cette thèse doit beaucoup.

Je remercie sincèrement les professeurs Patrick BROWN et Lance WALLER, ainsi que le professeur Alex MERIL, pour les remarques et suggestions qui m'ont permis d'améliorer cette thèse.

Je tiens à remercier toute ma famille et tous mes amis pour avoir contribué à un environnement favorable à mes activités de recherche.

Pour finir, je remercie toutes les personnes qui ont contribué de près ou de loin à la concrétisation de cette thèse.

Table des matières