Computational Statistics and Data Analysis 123 (2018) 131–145

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Likelihood based inference for the multivariate renewal Hawkes process☆

Tom Stindl *, Feng Chen

Department of Statistics, UNSW Sydney, Australia

ARTICLE INFO

Article history:
Received 7 October 2017
Received in revised form 9 January 2018
Accepted 30 January 2018
Available online 13 February 2018

Keywords:
Finance
Maximum likelihood
Model assessment
Point process
Prediction
Seismology

ABSTRACT

The recent introduction of the renewal Hawkes (RHawkes) process has extended the modeling capabilities of the classical Hawkes self-exciting process by allowing the immigrant arrival times to follow a general renewal process rather than a homogeneous Poisson process. A multivariate extension to the RHawkes process will be proposed, which allows different event types to interact with self- and cross-excitation effects, termed the multivariate renewal Hawkes (MRHawkes) process model. A recursive algorithm is developed to directly compute the likelihood of the model, which forms the basis of statistical inference. A modified algorithm for likelihood evaluation is also proposed which reduces computational time. The likelihood evaluation algorithm also implies a procedure to assess the goodness-of-fit of the temporal patterns of the events and distribution of the event types by computing independent and uniform residuals. The plug-in predictive density function for the next event time and methods to make future predictions using simulations are presented. Simulation studies will show that the likelihood evaluation algorithms and the prediction procedures are performing as expected. To illustrate the proposed methodology, data on earthquakes arising in two Pacific island countries Fiji and Vanuatu and trade-through data for the stock BNP Paribas on the Euronext Paris stock exchange are analyzed.

1. Introduction

The class of self-exciting point processes provides a flexible framework to model and describe a wide range of events that occur over time whose intensity or rate of arrival of events is influenced by both exogenous and endogenous effects. An attractive feature for this class of models is their ease of interpretation and as such have appeared in a wide variety of application domains. The most commonly applied model of this type is the Hawkes process (Hawkes, 1971), where exogenous events termed immigrants arrive according to a homogeneous Poisson process. This model allows for straightforward calculation of the likelihood and therefore makes likelihood based inference easy to implement.

However, in many applications the Hawkes model fails to provide an adequate fit to the data. There are many works in the literature devoted to extending the Hawkes process. One approach is to allow the model parameters to vary over time while still assuming the immigrants arrive according to a (possible inhomogeneous) Poisson process (Mohler et al., 2011; Chen and Hall, 2013; Roueff et al., 2016; Godoy et al., 2016). Another approach relies on the branching Poisson process interpretation of the Hawkes process (Hawkes and Oakes, 1974) and generalizes the immigrant arrival process. An example is the renewal

☆ This research includes computations using the Linux computational cluster Katana supported by the Faculty of Science, UNSW Sydney.

* Corresponding author.
E-mail addresses: t.stindl@unsw.edu.au (T. Stindl), feng.chen@unsw.edu.au (F. Chen).

https://doi.org/10.1016/j.csda.2018.01.021

Hawkes (RHawkes) process (Wheatley et al., 2016; Chen and Stindl, 2017) where the immigrant arrival process is allowed to be a general renewal process rather than the homogeneous Poisson process. The background event intensity of the RHawkes process is no longer a deterministic constant or function as in the classical Hawkes process or its time-varying extensions, and thus the likelihood function of the RHawkes process is nontrivial to evaluate.

Convinced that the likelihood for the RHawkes process takes exponential time to evaluate, Wheatley et al. (2016) proposed two Expectation–Maximization (EM) type algorithms (Dempster et al., 1977) to calculate the maximum likelihood estimator (MLE) without having to evaluate the likelihood, using two different choices of the sets of missing data. They applied the RHawkes process to model the mid-quote price changes of the E-mini S&P500 future contracts and found the RHawkes model provided a better fit to the data than the Hawkes model. Chen and Stindl (2017) proposed a recursive method to directly evaluate the likelihood of the RHawkes process, in quadratic time, and showed that the likelihood can be directly optimized to obtain the MLEs of the model parameters and their standard errors. They also proposed computationally efficient methods for goodness-of-fit assessment and prediction, and applied their methodologies to earthquake occurrence modeling and to foreign exchange data analysis. They also found that the RHawkes process provided better fit to the earthquake data than the classical Hawkes process and was able to give reasonably accurate predictions of future earthquakes.

In many areas researchers also encounter multi-type event sequence data. For example, in earthquake modeling, the data may contain earthquakes from several neighboring regions. In finance, the tick history data on a specific stock typically records both the times of trades and the times of quotes; and order book data records the arrival times and other features of limit and market orders such as the side of the trade. In these applications, it is of interest to study not only the interactions within events of the same type, but also the interactions between events of different types. Therefore, multivariate point processes, where different components are allowed to interact with each other, are needed. A multivariate point process model for this purpose is the multivariate Hawkes process (Hawkes, 1971). Bowsher (2007) modeled the timing of trades and mid-quote price changes for a NYSE stock using a generalized bivariate Hawkes process that allows the baseline event rate to vary with time. Embrechts et al. (2011) fit the bivariate Hawkes process to daily data on the negative and positive exceedances of certain threshold levels for the Dow Jones Industrial Average index. Bacry et al. (2013) showed that the multivariate Hawkes process can demonstrate the Epps effect and lead–lag effect observed in financial data. When interpreted as branching Poisson processes, both the multivariate Hawkes process and the generalization considered by Bowsher (2007) assume the arrival processes of immigrants to be Poisson and therefore do not allow over- or under-dispersion of the numbers of immigrants, or serial correlation of the numbers of immigrants in non-overlapping time intervals, even events of the same type. Such assumptions restrict the modeling capabilities of the multivariate Hawkes process unnecessarily.

In this paper we consider a point process model which extends the renewal Hawkes process by allowing the events of the process to be of different types and, in addition to the self-excitation effect among events of the same type, allowing events of each type to affect the future occurrence rates of events of other types through the mutual excitation mechanism adopted in the multivariate Hawkes processes. The model also extends the multivariate Hawkes process in that the immigrant events of different types can arrive according to general renewal processes, rather than Poisson processes in the classical multivariate Hawkes process. This implies that the numbers of immigrant events of the same type in non-overlapping time intervals are allowed to have serial correlation and to be over- or under-dispersed relative to the Poisson distribution. We naturally call this model the multivariate renewal Hawkes process, or MRHawkes process for short.

Similar to the RHawkes process, the MRHawkes process can be efficiently simulated by utilizing the branching process interpretation. Also similar to the RHawkes process, the MRHawkes process does not have an easy to evaluate likelihood function. We derive an algorithm to efficiently evaluate the likelihood of the MRHawkes process model, using an approach analogous to that of Chen and Stindl (2017). We demonstrate the feasibility of fitting the MRHawkes process model to data by likelihood maximization, on simulated data and on real life data. The time and space complexities of the algorithm for MRHawkes process likelihood evaluation are both polynomial in the number of events observed, and therefore the algorithm can be fairly slow on large data sets. To overcome this issue, we shall propose a simple modification to the algorithm, which can yield a good approximation of the likelihood in quadratic time and linear storage space. We will also provide an approach to assess two aspects of the goodness-of-fit of the MRHawkes model, the temporal patterns of the events and the event type distribution using the Rosenblatt residuals (Rosenblatt, 1952) and the Universal residuals (Brockwell, 2007) respectively. A simulation based approach to predict future event occurrences will also be proposed.

The rest of the paper is structured as follows. Section 2 introduces the MRHawkes process model. Section 3 derives an algorithm to evaluate the likelihood of the MRHawkes process. A method to evaluate the goodness-of-fit is presented in Section 4 and methods for future events prediction in Section 5. Results of our simulation studies are presented in Section 6 together with methods to simulate the process and the assessment of the predictive performance of the model. Applications in seismology and finance are presented in Section 7 with an analysis of earthquakes arising in two Pacific island countries Fiji and Vanuatu and a data set of trade-throughs for the stock BNP Paribas, traded on the Euronext Paris stock exchange.

2. Model and notation

Let $\{(\tau_i, z_i),\ i = 1, 2, \ldots\}$ be a realization of a multivariate point process where $\tau_1 < \tau_2 < \cdots$ are distinct and interpretable as the occurrence time of the ith event and $z_i \in \{1, \ldots, M\}$ indicates the ith event type. Let the associated