ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9413812

SWITCHED HAWKES PROCESSES

Namrata Nadagouda and Mark A. Davenport

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA

ABSTRACT

Hawkes processes are a class of auto-regressive point processes that are commonly used in modeling data in which events tend to cluster and influence the likelihood of future events. Because of their ability to model and explain how events or processes can influence each other, Hawkes processes (and their multivariate extensions) have been applied in a variety of practical applications such as analyzing financial time series, communication networks, and biological networks, to name just a few. In practice, the dynamics of such systems often depend on external factors that may change over time and that may drive different kinds of behavior. In this paper, we consider a switched Hawkes process which can be used to model systems in which the parameters of the process dynamically change depending on some (known) external state. We propose a simple maximum likelihood estimation approach which we validate using synthetic simulations. We then apply our model to a real-world traffic sensor dataset to study traffic patterns during different configurations of the traffic lights at an intersection.

***Index Terms*—** Hawkes processes, switched state models, maximum likelihood estimation, traffic data modeling

1. INTRODUCTION

Hawkes processes [1] are an important tool for analyzing datasets consisting of event times that exhibit auto-regressive behavior. For example, in many settings, the occurrence of an event is often an indicator that more events are likely in the near future. For example, earthquakes [2, 3, 4] are often preceded and/or followed by an elevated level of seismic activity (aftershocks), activity in financial markets [5, 6, 7] can be highly clustered, and activity within many kinds of networks (e.g., communication, social, transportation, etc. [8, 9, 10]), is often clustered/correlated across nodes. In all of these examples as well as many others, Hawkes processes (and multivariate extensions) have been used to aid in understanding the structure of the underlying signals/networks. Specifically, in a Hawkes process the underlying intensity function(s) that determine the rate at which events occur varies as a function of the event history according to a parametric model. This model reveals how much an event increases the probability of future events, as in an earthquake increasing the likelihood of aftershocks or a social media post by a particular user increasing the likelihood of subsequent posts by other users.

In many applications, however, the underlying dynamics may change over time. For instance, the background seismicity and decay rates of earthquakes can change depenging on seismic state (e.g., foreshocks, main shocks or aftershocks [11]). Similarly, activity in a stock market is impacted by price-predictive signals such as the bid-ask spread and queue imbalance, which vary over time [12]. Finally, in a transportation network, the dynamics of a traffic flow can change depending on external factors such as the state of various traffic lights and other control mechanisms. In such scenarios, there are observable external factors that define an underlying system state and that can have a significant impact on the underlying system behavior.

In this paper, we consider a switched version of the Hawkes process that explicitly accounts for changes in behavior depending on the underlying state. Within a given state, the system is described by a set of parameters that characterize the system's behavior. In our model, we allow these parameters to change depending on the underlying state history. We propose and evaluate a general modeling and inference strategy. We then apply this model to a real-world traffic sensor dataset, showing that our approach can effectively learn the underlying traffic flows in this network using only event timing information at each sensor.

2. BACKGROUND

2.1. Hawkes processes

Hawkes processes – first introduced in [13] – are a class of linear, auto-regressive point processes. In a Hawkes process, the occurrence of an event influences the rate at which events will occur in the future. Specifically, the distribution of events is determined by the *conditional intensity function* (CIF). Using the notation in [9], we define the CIF as

$$\lambda(t|\mathcal{H}_t) = \mu + a \sum_{\tau_k \in \mathcal{H}_t} \gamma(t - \tau_k), \tag{1}$$

where $\mathcal{H}_t = \{\tau_k : \tau_k < t\}$ denotes the history of the process at time t (i.e., the set of all events occurring before time

t). The term μ corresponds to the base intensity, a is the coefficient of excitation, and $\gamma(t-\tau_k)$ is the kernel modeling the influence of an event at time τ_k. Typically, we observe a sequence of events $\{\tau_1, \ldots, \tau_K\}$ and will be interested in estimating the parameters μ and a based on these observations.

In the multidimensional case we may have an ensemble of N point processes that influence each other. In this setting, the CIF for each subprocess can be defined as

$$\lambda_i(t|\mathcal{H}_t) = \mu_i + \sum_{j=1}^{N} a_{i,j} \sum_{\tau_k \in \mathcal{H}_{t,j}} \gamma(t-\tau_k), \qquad (2)$$

where $\boldsymbol{\mu} = [\mu_i]$ with μ_i being the base intensity of the i^{th} subprocess, $\boldsymbol{A} = [a_{i,j}]$ is the "infectivity matrix" where $a_{i,j}$ is a positive quantity describing the influence of the j^{th} subprocess on the i^{th} subprocess and $\mathcal{H}_{t,j}$ is the subset of the history $\mathcal{H}_t$ containing events belonging to the j^{th} subprocess. In this case every event is associated not only with a time of occurrence (τ_k) but also a "mark" $\theta_k \in \{1, \ldots, N\}$ that indicates which subprocess the event is associated with. Using this notation, we can write $\mathcal{H}_{t,j} = \{\tau_k : \tau_k < t, \theta_k = j\}$.

Note that these models assume that the underlying parameterization of the system is constant. We are chiefly interested in the case where the dynamics of the system – corresponding to the model parameters $\boldsymbol{\mu}$ and $\boldsymbol{A}$ – are changing with time. In such settings we must consider more sophisticated models.

2.2. Related work

While state-dependent parameterizations have been studied for traditional Poisson processes (e.g., see [14]), the vast majority of prior work in the field of Hawkes processes assumes a static parameterization. A notable exception is the work of [11] which considers Hawkes processes that change parameterizations according to an underlying Markov model characterizing the state transitions. The state information is assumed to be unknown and not directly observable, and thus [11] aims to estimate both the state transition probabilities and process intensity parameters using a hidden Markov model (HMM) to model the underlying states. To simplify the inference process, [11] considers only univariate Hawkes processes with a kernel that is piecewise constant between events. This approach was recently extended to more general kernels in [15].

The more recent work of [12] is perhaps the most similar to ours. The authors consider a state-dependent Hawkes process consisting of a multivariate Hawkes process that is coupled with an observable Markov chain governed state process. The assumption in this work is that state transitions are triggered by events in the Hawkes process according to a Markov transition matrix that depends on which subprocess the event belongs to. This structure is motivated by the application to limit order books and ensures an efficient parameterization.

3. SWITCHED HAWKES PROCESSES

We now introduce our switched Hawkes process model in which the process switches among a known finite number of states, where each state is characterized by a different multidimensional Hawkes process. Specifically, let an N-dimensional process of order M denote a process with N subprocesses and M states. In this setting, the CIF for the i^{th} subprocess in the s^{th} state can be defined as

$$\lambda_{i,s}(t|\mathcal{H}_t) = \mu_{i,s} + \sum_{j=1}^{N} \sum_{s'=1}^{M} (a_{i,s})_{j,s'} \sum_{\tau_k \in \mathcal{H}_{t,j,s'}} \gamma(t-\tau_k), \quad (3)$$

where $\boldsymbol{\mu} = [\mu_{i,s}]$ with $\mu_{i,s}$ being the base intensity of the i^{th} subprocess, $\boldsymbol{A} = [(a_{i,s})_{j,s'}]$ is the infectivity matrix where $(a_{i,s})_{j,s'}$ represents the influence of events belonging to the j^{th} subprocess that occurred when the system was in state s' on the i^{th} subprocess when the system is in state s. Here, $\mathcal{H}_{t,j,s'}$ now denotes the subset of $\mathcal{H}_t$ defined as

$$\mathcal{H}_{t,j,s'} = \{\tau_k : \tau_k < t, \theta_k = j, s_k = s'\} \qquad (4)$$

where we assume that each event τ_k is accompanied by an additional mark s_k that indicates the corresponding state of the system.

Maximum likelihood estimation has been commonly used for parameter estimation in Hawkes processes [16, 17]. We develop an extension of this procedure to estimate $\boldsymbol{\mu}$ and $\boldsymbol{A}$ in a switched process. Suppose that we observe the sequence of events $\{(\tau_1, \theta_1, s_1), \ldots, (\tau_K, \theta_K, s_K)\}$ on $[0, T)$. The likelihood of these events is given by

$$L(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_T) = \left(\prod_{k=1}^{K} \lambda_{\theta_k, s_k}(\tau_k)\right) \exp\left(-\int_0^T \lambda(t)dt\right) \qquad (5)$$

where

$$\lambda(t) = \sum_{i=1}^{N} \sum_{s=1}^{M} \lambda_{i,s}(t). \qquad (6)$$

(Here $\lambda_{i,s}(t)$ is used to denote the CIF instead of $\lambda_{i,s}(t|\mathcal{H}_t)$ and, for brevity, we will follow this notation henceforth.)

Considering the negative logarithm of the likelihood leads us to a convenient separable form as below.

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_T) = \sum_{i=1}^{N} \sum_{s=1}^{M} \mathcal{L}_{i,s}(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_T) \qquad (7)$$

where

$$\mathcal{L}_{i,s}(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_T) = \int_0^T \lambda_{i,s}(t)dt - \sum_{\tau_k \in \mathcal{H}_{T,i,s}} \log \lambda_{i,s}(\tau_k). \qquad (8)$$

Since every $\lambda_{i,s}(t)$ depends only on $\mu_{i,s}$ and the sub-matrix $\boldsymbol{A}_{i,s} = [(a_{i,s})_{j,s'}$ for $j = 1 : N$ and $s' = 1 : M]$, each of

$\mathcal{L}_{i,s}(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_{\mathcal{T}})$ can be optimized independently for $MN + 1$ parameters instead of the composite $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{A}|\mathcal{H}_{\mathcal{T}})$ involving $MN(MN + 1)$ parameters. The parameters are estimated via (8) which is a convex optimization problem that can be solved using a quasi-Newton method as described in [9].

Unlike the prior works discussed above, our model makes no assumptions on the nature of the state process or any relation between the state process and the event process. We instead assume that we have implicit knowledge of the states and state transition times (which could be arbitrary). While some previous works consider the joint estimation of the state and event processes under specific modeling assumptions on the state transitions, this often involves elaborate computations. We instead provide a simple method for estimating the parameters when given knowledge of the state information. In cases where the dynamics are known or are relatively simpler to estimate via some form of side information, this can lead to a much simpler inference process. As an example, in the case of data from traffic sensors, information about the configuration of the traffic lights might be available. Similarly, if we consider the situation of estimating the firing rates of neurons, the underlying brain states (sleep vs awake) might be easy to estimate. In the latter case, our method can be modified to estimate switched variants of Poisson processes (which are popularly used to model neuron firings) [18]. We also note that our approach extends very naturally to multivariate Hawkes processes and can accommodate arbitrary kernels.

4. SIMULATIONS

To illustrate the concept of a switched Hawkes system, we simulate a simple model of a two-dimensional Hawkes process of order 2. The event occurrence times are generated according to the thinning method [19], considering the CIFs corresponding to the underlying state. The marks indicating which subprocess the events belong to, are generated according to the state-dependent mark distribution function. We use an exponential decaying kernel for $\gamma(t)$. The data generated by this model is illustrated in Figure 1.

We are interested in understanding the differences between the inference drawn by the multivariate model and the switched model on such data. Towards this end, we estimate parameters for both the models on a data simulated using the switched model. This model is of a similar structure to the system in Figure 1. The parameters used are

$$\boldsymbol{\mu}_1 = \begin{bmatrix}1\\0\end{bmatrix} \quad \boldsymbol{\mu}_2 = \begin{bmatrix}0\\1\end{bmatrix} \quad (\boldsymbol{A}_1)_1 = \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} \quad (\boldsymbol{A}_2)_2 = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$$

with $(\boldsymbol{A}_1)_2$ and $(\boldsymbol{A}_2)_1$ set to 0. Here $\boldsymbol{\mu}_s = [\mu_{i,s}$ for $i = 1 : N]$ and $(\boldsymbol{A}_s)_{s'} = [(a_{i,s})_{j,s'}$ for $i = 1 : N$ and $j = 1 : N]$.

We use the constraint of no self-excitation while estimating the arameters in both the models. The multivariate model

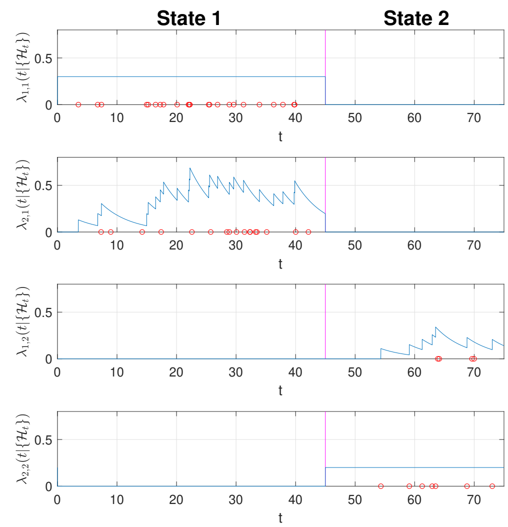

Fig. 1. Simulation of a two dimensional switched Hawkes process of order two. This figure represents the CIF of the two subprocesses in the two states. The vertical line at t = 45s represents a state transition. The red circles indicate the event occurrence times. State 1 is characterized by innovations at node 1 and excitations at node 2. We can observe that every event of the first subprocess leads to an increase in the CIF of the other subprocess, and vice versa in state 2.

learns a single set of parameters of the system with

$$\boldsymbol{\mu} = \begin{bmatrix}0.68\\0.53\end{bmatrix} \quad \boldsymbol{A} = \begin{bmatrix}0 & 0.28\\0.46 & 0\end{bmatrix}.$$

This is equivalent to learning a model that captures the average performance of the system in both the states. The switched model more accurately estimates parameters that clearly indicate the differing dynamics of the two states of the system. In this case, the parameters learnt are

$$\boldsymbol{\mu}_1 = \begin{bmatrix}1.00\\0\end{bmatrix} \quad \boldsymbol{\mu}_2 = \begin{bmatrix}0\\1.07\end{bmatrix}$$

$$(\boldsymbol{A}_1)_1 = \begin{bmatrix}0 & 0\\0.96 & 0\end{bmatrix} \quad (\boldsymbol{A})_{2,2} = \begin{bmatrix}0 & 0.92\\0 & 0\end{bmatrix}$$

with $(\boldsymbol{A}_1)_2$ and $(\boldsymbol{A}_2)_1$ being equal to 0.

5. APPLICATION TO TRAFFIC SENSOR DATASET

In this section we consider a real-world traffic sensor dataset which consists of time-stamped events representing the passage of a vehicle over a sensor. In the subset of the dataset that we consider here, there are 16 sensors placed at different lanes in the 4 legs of an intersection as illustrated in Figure 2. The traffic lights at the intersection switch among 6 different configurations. The section of the data used for our experiments consists of sensor observations collected over a duration of two hours, consisting of approximately 8000 events in total. Accordingly we use a 16 dimensional

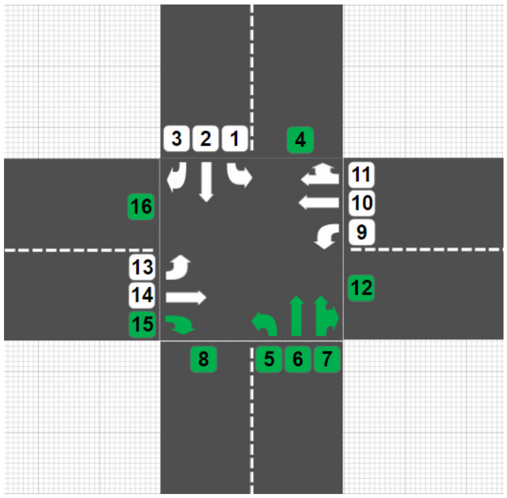

Fig. 2. Depiction of one of the states in the traffic dataset. Each of the small squares indicates a sensor placed in the corresponding lane with the ones having arrows in front of them being inbound sensors and the rest outbound sensors. The arrows represent the direction of traffic flow through the corresponding lanes. The shaded sensors along with the shaded arrow marks indicate the sensors that are active and the existing connections.

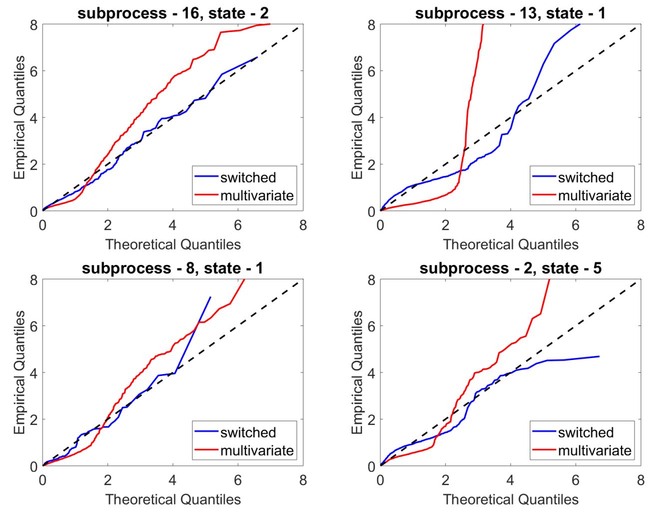

Fig. 3. Representative QQ-plots for the parameters learnt on the traffic dataset.

switched Hawkes process of order 6 to model the data. By applying our model to this dataset, we aim to learn the connections between the events at inbound sensors and those at outbound sensors, thereby obtaining estimates of which are the dominant traffic flows through the intersection.

An example state is as represented in Figure 2. In our inference, we use the constraint of no self-excitation since a vehicle passing over a sensor does not pass the same sensor again. For this state, the 7 most significant values in the corresponding infectivity matrix includes the 5 connections one would expect to observe in this state, i.e., 5-16, 6-4, 7-4, 7-12, and 15-8.

We also estimate a multivariate Hawkes process for this data and investigate which of the two models better explains the data. According to the theorem of random time change [20], for a point process with intensity $\lambda(t)$ and event times $\{\tau_k\}_{k=1}^{n}$, the rescaled inter-arrival times given by $\{\int_{\tau_{k-1}}^{\tau_k} \lambda(t)dt\}_{k=1}^{n}$ form a sequence of exponential random variables with unit rate. Accordingly, for an N-dimensional Hawkes process, the integrated intensities for each of the subprocesses form N independent sequences of i.i.d. exponential random variables with unit rate. Similarly, for an N-dimensional switched Hawkes process of order M, they form MN independent sequences [12]. In Figure 3, we compare the empirical quantiles for both the models with the theoretical quantiles for the unit rate exponential distribution, known as a *QQ-plot*. The closer the QQ-plot is to the reference line (slope = 1), the better the model matches the true distribution of the data. We observe that the switched model is a far better fit compared to the multivariate model.

We note that this data is relatively noisy with the observed sensor events not being strictly in accordance with traffic signal information. Nevertheless, our model is robust enough to pick the strong connections and resolve the causality between the inbound and outbound sensors. In some cases the model also estimates spurious influences between adjacent sensors. To avoid such false connections, we can use the constraint of no influence among sensors in adjacent lanes and the knowledge of inbound-outbound orientations in addition to the constraint of no self-excitation. The results obtained with these additional constraints often exhibit substantial improvements.

6. CONCLUSION

In this paper we have proposed an extension of the traditional multidimensional Hawkes process to a switched Hawkes process that can adapt based on dynamically changing state information. We consider a simple maximum likelihood estimation approach to fitting the model parameters and demonstrate the effectiveness of this approach on both synthetic simulations and real-world experiments. In the process, we also demonstrate the ineffectiveness of traditional (non-state-aware) Hawkes processes compared to our proposed approach. From our experiments, we conclude that our proposed approach is also a promising direction for obtaining improved stochastic models for multivariate Poisson processes in which the events are correlated and state-dependent. In the future, we hope to further explore the use of these techniques in the context of predicting future traffic behavior in a network and in modeling state-dependent neural activity.

7. ACKNOWLEDGEMENTS

This work was supported, in part, by NSF grant CCF-1350616 and a gift from the Alfred P. Sloan Foundation.

8. REFERENCES

[1] Patrick J Laub, Thomas Taimre, and Philip K Pollett, "Hawkes processes," *arXiv preprint arXiv:1507.02822*, 2015.

[2] Yosihiko Ogata, "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics*, vol. 50, no. 2, pp. 379–402, 1998.

[3] H Pratiwi, I Slamet, DRS Saputro, et al., "Self-exciting point process in modeling earthquake occurrences," in *Journal of Physics: Conference Series*. IOP Publishing, 2017, vol. 855, p. 012033.

[4] Alex Reinhart et al., "A review of self-exciting spatio-temporal point processes and their applications," *Statistical Science*, vol. 33, no. 3, pp. 299–318, 2018.

[5] Graham Simon, *Hawkes Processes in Finance: A Review with Simulations*, Ph.D. thesis, University of Oregon, 2016.

[6] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy, "Hawkes processes in finance," *Market Microstructure and Liquidity*, vol. 1, no. 01, pp. 1550005, 2015.

[7] Alan G Hawkes, "Hawkes processes and their applications to finance: a review," *Quantitative Finance*, vol. 18, no. 2, pp. 193–198, 2018.

[8] Ke Zhou, Hongyuan Zha, and Le Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes," in *Artificial Intelligence and Statistics*, 2013, pp. 641–649.

[9] Michael G Moore and Mark A Davenport, "Analysis of wireless networks using hawkes processes," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2016, pp. 1–5.

[10] Marian-Andrei Rizoiu, Young Lee, Swapnil Mishra, and Lexing Xie, "A tutorial on hawkes processes for events in social media," *arXiv preprint arXiv:1708.06401*, 2017.

[11] Ting Wang, Mark Bebbington, and David Harte, "Markov-modulated hawkes process with stepwise decay," *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 3, pp. 521–544, 2012.

[12] Maxime Morariu-Patrichi and Mikko S Pakkanen, "State-dependent hawkes processes and their application to limit order book modelling," *arXiv preprint arXiv:1809.08060*, 2018.

[13] Alan G Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971.

[14] Shaochuan Lu, "Markov modulated poisson process associated with state-dependent marks and its applications to the deep earthquakes," *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 1, pp. 87–106, 2012.

[15] Tian Zheng, Jing Wu, and James Curley, "Markov-modulated hawkes processes for sporadic and bursty event occurrences," *arXiv preprint arXiv:1903.03223*, 2019.

[16] Tohru Ozaki, "Maximum likelihood estimation of hawkes' self-exciting point processes," *Annals of the Institute of Statistical Mathematics*, vol. 31, no. 1, pp. 145–155, 1979.

[17] Jakob Gulddahl Rasmussen, "Lecture notes: Temporal point processes and the conditional intensity function," *arXiv preprint arXiv:1806.00221*, 2018.

[18] Don H Johnson, "Point process models of single-neuron discharges," *Journal of computational neuroscience*, vol. 3, no. 4, pp. 275–299, 1996.

[19] Yosihiko Ogata, "On lewis' simulation method for point processes," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 23–31, 1981.

[20] David Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, Springer, 2003.