J. Appl. Prob. **,***-*** (2019)
*Doi:**.****/jpr.2019.***
Applied Probability Trusts 2019

AN ELEMENTARY DERIVATION OF MOMENTS OF HAWKES PROCESSES

LIRONG CUI,* ** *Beijing Institute of Technology*

ALAN HAWKES,*** *Swansea University*

HE YI,**** *Beijing Institute of Technology*

Abstract

Hawkes processes have been widely used in many areas, but their probability properties can be quite difficult. In the paper an elementary approach is presented to obtain moments of Hawkes processes and/or the intensity of a number of marked Hawkes processes, in which the detailed outline is given step-by-step, and it works not only for all Markovian Hawkes processes, but also for some non-Markovian Hawkes processes. The approach is simpler and more convenient than usual methods such as the Dynkin formula and martingale methods. The method is applied for one-dimensional Hawkes processes and other related processes such as Cox processes, dynamic contagion processes, non-homogenous Poisson processes and non-Markovian cases. Several results are obtained which may be useful in studying Hawkes processes and other counting processes. Our proposed method is an extension of Dynkin formula, which is simple and easy to use.

Keywords: Hawkes process; Dynkin formula; moments; counting process; Gamma decay function

2010 Mathematics Subject Classification: Primary 60G55
Secondary 60G57; 60G44; 60J80

Received 13 November, 2017; Revision received 23, 2019.

* Postal address: School of Management & Economics, Beijing Institute of Technology, Beijing 100081, China
** Email: Lirongcui@bit.edu.cn
*** Postal address:School of Management, Swansea University, Fabian Way, Swansea SA1 8EN, UK
**** Postal address: School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China

1. Introduction

Hawkes [12]-[14] introduced a family of models for stochastic point processes called 'self-exciting and mutually-exciting point processes', the essential property of which was that the occurrence of any event increased the probability of further events occurring. The main theoretical properties derived were the Bartlett spectrum and the corresponding covariance density function, useful tools for discriminating between models, means and variances of event counts. These processes have since become known as Hawkes processes and, despite a low uptake of applications for the first thirty years or so, they have recently become widely used in many different fields of study. For example, see [20], [21], [18], [22], [5, 6], [3] and [16].

This paper considers simple point processes on $\mathbb{R}_{\geq 0}$. We denote such a process by $N(t) := \sum_{r\in\mathcal{N}} I_{T_r\leq t}$, $t \geq 0$, where $\{T_r\}$ denote the (disjoint) points of $N(t)$. We interpret $\mathbb{R}_{\geq 0}$ as 'time' and the points T_r as 'events' in time. I_A is an indicator function for event A, i.e., $I_A = 1$ if event A occurs, otherwise, $I_A = 0$. An important property of simple point processes is the (conditional) intensity function, often denoted by $\lambda(t)$, such that

$$\begin{cases} \mathbb{P}\{N(t+\Delta) - N(t) = 1|\mathcal{F}_t\} = \lambda(t)\Delta + o(\Delta), \\ \mathbb{P}\{N(t+\Delta) - N(t) \geq 2|\mathcal{F}_t\} = o(\Delta), \end{cases}$$

where $(\mathcal{F}_t)_{t\geq 0}$, representing the history of the process up to time t is the filtration $\mathcal{F}_t = \sigma(N(s), 0 \leq s < t)$: so that the history includes knowledge of both intensity and the events prior to time t. Note that the above property means that the point process is orderly: i.e., events cannot occur simultaneously.

A simple but fairly general marked Hawkes self-exciting process can be defined by its intensity in the form

$$\lambda(t) = v(t) + \sum_{T_r<t} Z_r\gamma(t - T_r), \quad t > 0.$$

The function $\gamma(u) \geq 0$, $u > 0$ controls the effect that an event has on the intensity of events at time u later: this may be called the exciting kernel. Z_r is a mark associated with the event occurring at time T_r: these are usually supposed to be i.i.d. random variables. Note that in this simple class of processes the mark multiplies the exciting kernel so that an event with a large mark will increase the future intensity much more than an event with a small mark will. For example, an earthquake with large magnitude will increase the probability of further earthquakes much more than a small earthquake will. Similarly, a large jump in the price of a financial asset is more likely to cause more future jumps than a small jump will. Marks can enter the model in more general ways but this simple version will suffice in this paper. $v(t)$ is the base intensity: we will treat this as a deterministic function of time, although it may be partly a function of some exogenous process, representing economic conditions, weather etc., depending on the process that is being modeled.

It is important to realize that the intensity should always be non-negative, and this will be assured if $v(t)$, Z_r and $\gamma(u)$ are all non-negative. In this paper we will mostly consider exponential kernels $\gamma(u) = e^{-\beta u}$, $u > 0$, so that our basic class of self-exciting processes will be defined by

$$\lambda(t) = v(t) + \sum_{T_r < t} Z_r e^{-\beta(t-T_r)}, \quad t > 0. \tag{1}$$

Sometimes it is useful to suppose that a process has been running for some time, effectively taking the start time as minus infinity for the stationary process. In this paper we will always take the process as starting at time zero, so that the event times, T_r, are always positive.

A similarly simple class of mutually-exciting models for marked multivariate point processes, $N_k(t)$, $k = 1, 2, ..., K$, can be defined by the intensity functions

$$\lambda_k(t) = v_k(t) + \sum_{j=1}^{K} \sum_{T_{j,r} < t} Z_{j,r} \alpha_{kj} e^{-\beta_{kj}(t-T_{j,r})}, \tag{2}$$

where $\lambda_k(t)$ and $v_k(t)$ are the intensity function and baseline intensity, respectively, for the process of events $N_k(t)$. $\{T_{j,r}\}_{r=1,2,...}$ denotes the series of times at which events of type j occur and $\{Z_{j,r}\}_{r=1,2,...}$, the marks associated with those events, are typically assumed to be i.i.d. with distribution function $G_{Z_j}(z)$. α_{kj} and β_{kj} are parameters that determine the exciting effect that an event of type j has on the future intensity of events of type k. If $j = k$ we have a self-exciting component; $j \neq k$ we have a cross-exciting component.

Hawkes processes have been intensively studied and there have been generalizations to the models and advances in understanding their mathematical properties. Distributions of counts and interval properties of Hawkes processes can be quite difficult. Some results have been obtained, often using relatively advanced methods such as infinitesimal generators and Dynkin's formula (for example, see [10]). In this paper we consider a class of relatively simple Hawkes processes and introduce an elementary method to obtain expectations of the general form $\mathbb{E}[g(N(t), \lambda(t), t)]$, for example $\mathbb{E}[N^m(t)]$, $\mathbb{E}[\lambda^k(t)]$ and $\mathbb{C}\text{ov}[N(t), \lambda(t)]$. The method is elementary in principle, although the algebra can sometimes be complicated or we need to resort to numerical integration.

First we briefly describe methods that have been used in the literature to find such expectations or, better still, to find distributions of $N(t)$ and/or $\lambda(t)$. Oakes [19] studied the simple self-exciting model ($v(t) \equiv v$, $Z_r \equiv \alpha$ in Equation (1), which is called as a simple Hawkes process). Using the branching process representation of the simple Hawkes process, developed in [15], he obtained a set of equations for the probability generating function (p.g.f) of $N(t)$, and hence found the mean

$$\mathbb{E}[N(t)] = \frac{vt}{1-m} - \frac{m}{(1-m)^2}\frac{v}{\beta}\left\{1 - e^{-\beta(1-m)t}\right\}, \tag{3}$$

where $m = \alpha/\beta$ is the branching ratio (the expected number of immediate descendants of any individual in the process). Adamopoulos [1] generalized Oakes [15] cluster process representation of the univariate self-exciting process to a mutually-exciting process. He derived a multivariate probability functional (p.g.fl) of the process from which a variety of probability generating functions, and therefore moments, can be obtained. For example, he obtained the expected number of type j descendants of a type i event at the origin that occur in the interval $(0, t)$. He also obtained some forward recurrence time distributions and distributions of time intervals between events. Numerical results were given for a simple bivariate earthquake model as specified by Equation (2) with all marks $Z_{j,r}$ equal to 1 and parameters $v_1 = \alpha_{11} = \alpha_{12} = 0$ but each of the parameters $v_2, \alpha_{21}, \alpha_{22} > 0$. Thus, type 2 events were self-exciting and also excited type 1 events; type 1 events were entirely generated by the occurrence of type 2 events, even the baseline being zero. Errais et al. [11] modeled the occurrence of defaults in a portfolio of assets using the simple marked self-exciting process of Equation (2), but with different notation and with time-varying baseline intensity

$$\lambda(t) = c + e^{-\kappa t}(\lambda_0 - c) + \sum_{T_r < t} Z_r e^{-\kappa(t - T_r)}.$$

The marks, Z_r, were proportional to the loss of the defaulting asset, so large losses were likely to lead to further defaults by causing large increases in the intensity.

As the joint process $(N(t), \lambda(t))$ is Markovian, they were able to use the infinitesimal generator and associated Dynkin formula to obtain closed expressions for the moments of the Hawkes intensity. In particular

$$\mathbb{E}[\lambda(t)] = \left(\frac{\kappa c}{\mu} + \lambda_0\right) e^{\mu t} - \frac{\kappa c}{\mu}, \tag{4}$$

where $\mu = \mathbb{E}(Z_r) - \kappa$. A similar method was used to find $\mathbb{E}[\lambda^2(t)]$ and, therefore, the variance. They also show that a conditional transform of $J = (L, N)$, where $L(t)$ is the cumulative loss up to time t, satisfies

$$\mathbb{E}[\exp\left(uJ_T|\mathcal{F}_t\right)] = \exp\left(a(t) + b(t)\lambda(t) + uJ_t\right),$$

where the functions $a(t)$ and $b(t)$ satisfy certain differential equations. Differentiating with respect to u leads to differential equations that can be solved, algebraically or numerically, to obtain various moments. For example,

$$\mathbb{E}[N(t)] = \frac{\kappa c + \mu\lambda_0}{\mu^2}\left(e^{\mu t} - 1\right) - \frac{\kappa c}{\mu} t. \tag{5}$$

Dassios and Zhao [8] introduced the dynamic Contagion model, denoted as $(N_D(t), \lambda_D(t))$, where $N_D(t)$ and $\lambda_D(t)$ are corresponding counting process and its intensity function. They could use piecewise-deterministic Markov process theory and the work of [9] to show that the infinitesimal generator operating $\mathcal{A}$ on a function $g(n, \lambda, t)$ to obtain a conditional joint Laplace transform-probability generation function $\mathbb{E}\left[\theta^{N_D(t) - N_D(u)} e^{-\upsilon\lambda_D(t)} | \mathcal{F}_u\right]$, $t > u \geq 0$, for the distributions of $N_D(t)$ and $\lambda_D(t)$ in terms of a function that satisfies a certain ODE. Marginal Laplace transforms, p.g.fs of $N_D(t)$ and $\lambda_D(t)$ are obtained by setting $\theta = 1$ and $\upsilon = 0$, respectively. Moments can be obtained in the usual way by taking appropriate derivatives with respect to θ and υ. However, instead they apply the infinitesimal generator to functions of the form $g(n, \lambda, t) = n^m \lambda^k$ to obtain a differential equation whose solution yields the expectation $\mathbb{E}\left[N_D^m(t)\lambda_D^k(t)\right]$. In this way they obtain, for example, $\mathbb{E}[\lambda_D(t)]$, $\mathbb{E}[\lambda_D^2(t)]$, $\mathbb{E}[N_D(t)]$,

$\mathbb{E}[N_D^2(t)]$, $\mathbb{E}[N_D(t)\lambda_D(t)]$ and hence also variances and covariance. In each case there is a different differential equation to solve which, using their methods, includes using martingale properties of the generator.

Chen et al. [3] gave perhaps the easiest solution for the expectation of intensity of the univariate Hawkes process with exponential exciting kernel. They wrote its intensity function in the integral form $\lambda(t) = v + \int_0^t \alpha e^{-\beta(t-u)} \mathrm{d}N(u)$. On taking expectations we get $\mathbb{E}\left[\lambda(t)\right] = v + \int_0^t \alpha e^{-\beta(t-u)} \mathbb{E}\left[\lambda(u)\right] \mathrm{d}u$ and the Laplace transform of this is $\varphi^*(s) = \int_0^\infty \mathbb{E}\left[\lambda(t)\right] e^{-st} \mathrm{d}t = \frac{v}{s} + \frac{\alpha}{s+\beta}\varphi^*(s)$. So $\varphi^*(s) = \frac{v}{\beta-\alpha}(\frac{\beta}{s} - \frac{\alpha}{s+\beta-\alpha})$, which has inverse transform

$$\mathbb{E}\left[\lambda(t)\right] = \frac{v}{\beta - \alpha}\left(\beta - \alpha e^{-(\beta-\alpha)t}\right), \qquad t > 0, \alpha \neq \beta. \tag{6}$$

This is equivalent to Equation (4) if, in that equation, $\lambda_0 = c = v$, so that the baseline intensity is constant, and $\kappa = \beta$, $\mathbb{E}[Z_r] = \alpha$.

Summarizing the methods have been used so far for obtaining moments and count distributions of Hawkes processes, we know that the methods are: (i) infinitesimal generator and Dynkin's formula, (ii) probability generating function, (iii) Martingale method and (iv) Laplace transform, but all these methods are only for Hawkes processes with exponential kernel intensity functions.

The rest of the paper is organized as follows. In Section 2 provides our main results in the paper, in which a procedure of obtaining moments for Hawkes processes is developed and illustrated by applying it to a simple Hawkes process, and the relation of our method and Dynkin's formula is discussed as well. In Section 3 various results are obtained for a simple one-dimensional marked Hawkes processes. In Section 4 the method is applied to a variety of processes such as general mutually-exciting processes, Cox processes, dynamic contagion processes, non-homogenous Poisson processes, a one-dimensional self-exciting process with multiple marks and a non-Markovian case. Finally, Section 5 contains conclusions and discussion.

2. Main Result

Methods of finding moments that make use of infinitesimal generator, Dynkin formula and martingale may be considered a bit advanced by many people. In this section we propose a new approach for finding moments that use probabilistic arguments that are fairly elementary in principle, although the complicated algebraic manipulations are needed sometimes for this approach. Somewhat like [8], they lead to differential equations that may be solved algebraically or by numerical integration.

The aim of the method proposed in this paper is to calculate the expectation value of any arbitrary function of point process $N(t)$ and its intensity function $\lambda(t)$, denoted as $f(t) := \mathbb{E}[g(N(t), \lambda(t), t)]$: for example, the special cases of $\mathbb{E}[N^m(t)]$, $\mathbb{E}[\lambda^n(t)]$ and $\mathbb{C}\mathrm{ov}[N(t), \lambda(t)]$. The procedure for this method is described below in subsection 2.1, taking the simple Hawkes process as an example, and $f_{m,n}(t) = \mathbb{E}[N^m(t)\lambda^n(t)]$ as an objective function. This process is relatively simple but illustrates most of the ideas used in the general method. Subsection 2.2 provides some comparisons with a method using infinitesimal generators and Dynkin's formula.

2.1. Procedure of the elementary approach

The main idea of our elementary approach for obtaining the moments of Hawkes processes is as follows. Given the history of Hawkes process $\mathcal{F}_t = \sigma(N(s), 0 \le s < t)$ and a small positive $\Delta > 0$, we know that $N(t+\Delta) - N(t)$ is Bernoulli distributed approximately, thus $\lambda(t+\Delta | N(t+\Delta) - N(t), \mathcal{F}_t)$ are two-point distributed random variables approximately, then it is easily to obtain $\mathbb{E}[g(N(t+\Delta), \lambda(t+\Delta), t+\Delta)|\mathcal{F}_t]$ based on these Bernoulli and two-point distributions approximately. Furthermore, some differential or partial differential equations on $\mathbb{E}[g(N(t), \lambda(t), t)]$ can be given via letting $\Delta \to 0$ and using double expectation theorem, these differential or partial differential equations can be solved analytically or numerically to obtain $\mathbb{E}[g(N(t), \lambda(t), t)]$. In the following, we present a procedure exemplified by application for a simple Hawkes process to give the detailed steps to show how our elementary approach works. Consider a simple Hawkes process $N(t)$, $t \ge 0$ with an intensity function given by

$$\lambda(t) = v + \int_0^t \alpha e^{-\beta(t-u)} \mathrm{d}N(u) = v + \sum_{T_r < t} \alpha e^{-\beta(t-T_r)},$$

where T_r are positive event times, $r = 1, 2, ...,$ $v, \alpha, \beta > 0$.

Procedure of the method: the start

Step 1. Set an objective function $f(t) = \mathbb{E}[g(N(t), \lambda(t), t)]$ that is calculated: for example, $\mathbb{E}[N^m(t)\lambda^n(t)]$. (The required conditions on $g(n, \lambda, t)$ will be given in subsection 2.2.)

Step 2. Find the probabilities of the following related point process events.

For a simple (or orderly) point process, we need conditional probabilities

$$\mathbb{P}\{N(t+\Delta)-N(t)=0|\mathcal{F}_t\} \text{ and } \mathbb{P}\{N(t+\Delta)-N(t)=1|\mathcal{F}_t\},$$

where the filtration $\mathcal{F}_t = \sigma(N(s), 0 \leq s < t)$ represents the history of the process up to time t.

For the simple Hawkes process, we have

$$\mathbb{P}\{N(t+\Delta)-N(t)=0|\mathcal{F}_t\} = 1-\lambda(t)\Delta+o(\Delta),$$

$$\mathbb{P}\{N(t+\Delta)-N(t)=1|\mathcal{F}_t\} = \lambda(t)\Delta+o(\Delta).$$

Step 3. Calculate the intensity function values $\lambda(t+\Delta)$ given the following two events:

$$\{N(t+\Delta)-N(t)=0\} \text{ and } \{N(t+\Delta)-N(t)=1\}.$$

For the simple Hawkes process, we have,

(i) when $\{N(t+\Delta)-N(t)=0\}$ occurs,

$$\lambda_0(t+\Delta) = v+\alpha \sum_{T_i<t+\Delta} e^{-\beta(t-T_i+\Delta)} = (1-\beta\Delta)\lambda(t)+\beta v\Delta+o(\Delta);$$

(ii) when $\{N(t+\Delta)-N(t)=1\}$ occurs,

$$\begin{aligned}\lambda_1(t+\Delta) &= v+\alpha\sum_{T_i<t} e^{-\beta(t-T_i+\Delta)}+\alpha e^{-\beta(t-T_{N(t)+1}+\Delta)} \\ &= \lambda_0(t+\Delta)+\alpha(1-\beta\Delta_1)+o(\Delta_1), \quad 0<\Delta_1<\Delta.\end{aligned}$$

Step 4. Calculate and simplify $\mathbb{E}[g(N(t+\Delta),\lambda(t+\Delta),t+\Delta)|\mathcal{F}_t]$ by using the results obtained in ***Step 2*** and ***Step 3***.

For the simple Hawkes process, we have

$$\begin{aligned}&\mathbb{E}[N^m(t+\Delta)\lambda^n(t+\Delta)|\mathcal{F}_t] \\ &= N^m(t)\lambda_0^n(t+\Delta)[1-\lambda(t)\Delta]+[N(t)+1]^m\lambda_1^n(t+\Delta)\lambda(t)\Delta+o(\Delta) \\ &= N^m(t)\lambda^n(t)+N^m(t)\sum_{i=0}^{n-1}\binom{n}{i}(-\nabla_1)^{n-i}\lambda^i(t)-N^m(t)\sum_{i=0}^{n}\binom{n}{i}(-\nabla_1)^{n-i}\lambda^{i+1}(t)\Delta \\ &\quad+\sum_{j=0}^{m}\binom{m}{j}N^j(t)\sum_{i=0}^{n}\binom{n}{i}\alpha^{n-i}\lambda^{i+1}(t)\Delta+o(\Delta),\end{aligned}$$

where $\nabla_1 = \beta[\lambda(t)-v]\Delta$, $0<\Delta_1<\Delta$.

Step 5. Take expectations on both sides on the formula of $\mathbb{E}[g(N(t+\Delta),\lambda(t+\Delta),t+\Delta)|\mathcal{F}_t]$ which was gotten in ***Step 4***, then let $\Delta \to 0$ to get the differential or partial differential equation on $\mathbb{E}[g(N(t),\lambda(t),t)]$.

For the simple Hawkes process, we have

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[N^m(t)\lambda^n(t)] &= \lim_{\Delta\downarrow 0}\frac{\mathbb{E}[N^m(t+\Delta)\lambda^n(t+\Delta)]-\mathbb{E}[N^m(t)\lambda^n(t)]}{\Delta}\\ &= n\beta v\mathbb{E}[N^m(t)\lambda^{n-1}(t)]-n\beta\mathbb{E}[N^m(t)\lambda^n(t)]\\ &\quad+\sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N^j(t)\lambda^{n+1}(t)],\\ &\quad+\sum_{j=0}^{m}\sum_{i=0}^{n-1}\binom{m}{j}\binom{n}{i}\alpha^{n-i}\mathbb{E}[N^j(t)\lambda^{i+1}(t)]. \qquad (7)\end{aligned}$$

Note that where $\sum_{i=0}^{w}[\bullet]=0$, when $w<0$.

Step 6. Solve the established differential equations based on the boundary conditions.

For the simple Hawkes process, this differential equation (obtained in ***Step 5***) can be solved, subject to initial conditions $\mathbb{E}[N^i(0)]=0$, $\mathbb{E}[\lambda^i(0)]=v^i$, $(i=1,2,\ldots)$.

Procedure of the method: the end.

This procedure provides the main result for our elementary method, the quantities $\lambda_0(t+\Delta)$ and $\lambda_1(t+\Delta)$ are required, and the double expectation theorem is used, then differential equations are developed, which can provide solution for $\mathbb{E}[g(N(t),\lambda(t),t)]$

Remark 1. In ***Steps 3*** and ***4***, the symbol $o(\Delta)$ is used, which is for $X=o(\Delta)$ such that

$$\lim_{\Delta\downarrow 0}\frac{X}{\Delta}=0, \qquad a.e.,$$

where X is a random variable. In the paper the notation $o(\Delta)$ is used in many places, some of them are for the deterministic sense, some for the random case like just mentioned in ***Step 3***. However, which one does not matter for the calculations, the reason for this can be seen in *Remark* 2.

Remark 2. We have used the conclusion $\lim_{\Delta\downarrow 0}\mathbb{E}[X/\Delta]=0$ such that $X=o(\Delta)$ (of course, it is not a general infinitesimal, which is raised in our paper), this operation means that

$$\lim_{\Delta\downarrow 0}\mathbb{E}[X/\Delta]=\mathbb{E}[\lim_{\Delta\downarrow 0}X/\Delta]=0.$$

The holding of above equality needs, in general, the sequence $\{\frac{o(\Delta)}{\Delta}\}_{\Delta>0} = \{\frac{X}{\Delta}\}_{\Delta>0}$ being uniformly integrable. In fact, we have

$$\sup_{0<\Delta<\delta} \mathbb{E}[|X/\Delta|] < \infty,$$

where δ is a given very small positive value, it is well-known that this condition can guarantee the the sequence $\{\frac{X}{\Delta}\}_{\Delta>0}$ being uniformly integrable. Thus the interchange of expectation and limit is a reasonable operation in our procedure. In the following, we shall show some facts appeared in our paper. Note we assume that

$$\mathbb{E}[|g(N(t), \lambda(t), t)|] < \infty.$$

Case 1. In ***Step 3*** in Section 2.1, for

$$\lambda_0(t+\Delta) = v + \alpha \sum_{T_i<t+\Delta} e^{-\beta(t-T_i+\Delta)} = (1-\beta\Delta)\lambda(t) + \beta v\Delta + o(\Delta),$$

where $o(\Delta) = [\frac{(\beta\Delta)^2}{2!} - \frac{(\beta\Delta)^3}{3!} + \frac{(\beta\Delta)^4}{4!} - \cdots][\lambda(t) - v] = [e^{-\beta\Delta} - 1 + \beta\Delta][\lambda(t) - v]$. On the other hand, we can know that

$$\mathbb{E}[|\frac{o(\Delta)}{\Delta}|] \leq |\frac{e^{-\beta\Delta} - 1 + \beta\Delta}{\Delta}|\mathbb{E}[|\lambda(t) - v|],$$

and $|\frac{e^{-\beta\Delta}-1+\beta\Delta}{\Delta}| \leq \frac{1}{\Delta}\sum_{n=2}^{\infty}(-1)^n\frac{(\beta\Delta)^n}{n!} \leq \frac{1}{\Delta}\sum_{n=2}^{\infty}\frac{(\beta\Delta)^n}{n!} \leq \beta\sum_{n=1}^{\infty}\frac{(\beta\Delta)^n}{n!} = \beta(e^{\beta\Delta} - 1)$. If we choose $\Delta\beta < 1$, i.e., $\Delta < 1/\beta = \delta$, then we have

$$|\frac{e^{-\beta\Delta} - 1 + \beta\Delta}{\Delta}| \leq \beta(e-1).$$

For any given time t , we know that $\mathbb{E}[|\lambda(t) - v|] \leq C$, i.e., it has a bound (where C does not related to Δ). Thus we can obtain that

$$\mathbb{E}[|\frac{o(\Delta)}{\Delta}|] \leq \beta(e-1)C$$

provided that $\Delta < 1/\beta = \delta$. It proves $\sup_{0<\Delta<\delta} \mathbb{E}[|X/\Delta|] < \infty$.

Case 2. In ***Step 3*** in Section 2.1, for

$$\begin{aligned}\lambda_1(t+\Delta) &= v + \alpha\sum_{T_i<t} e^{-\beta(t-T_i+\Delta)} + \alpha e^{-\beta(t-T_{N(t)+1}+\Delta)} \\ &= \lambda_0(t+\Delta) + \alpha(1-\beta\Delta_1) + o(\Delta_1), \;\; 0<\Delta_1<\Delta,\end{aligned}$$

where Δ_1 is a random variable, but it has the lower and upper bounds: 0 and Δ , respectively. Similarly, we can prove that $\sup_{0<\Delta<\delta} \mathbb{E}[|X/\Delta|] < \infty, \Delta < 1/\beta = \delta$.

Case 3. In ***Step 3*** in Section 2.2, in $\mathbb{E}[g(N(t+\Delta), \lambda(t+\Delta), t+\Delta)|\mathcal{F}_t]$, we used that

$$o(\Delta) = g(N(t+\Delta), \lambda_i(t+\Delta), t+\Delta)P\{N(t+\Delta) - N(t) \geq 2\}, (i \geq 2),$$

it has assumed that

$$P\{N(t+\Delta) - N(t) \geq 2\} = o_D(\Delta)$$

which is a deterministic case (using the letter "D" to denote it). On the other hand, based on the boundness of $\mathbb{E}[|g(N(t), \lambda(t), t)|]$ for any time t , we have

$$\mathbb{E}[o(\Delta)/\Delta] = [o_D(\Delta)/\Delta]\mathbb{E}[g(N(t+\Delta), \lambda_i(t+\Delta), t+\Delta)] = o_D(\Delta)/\Delta,$$

which proves that $\sup\limits_{0<\Delta<\delta} \mathbb{E}[|X/\Delta|] < \infty$.

Remark 3. The double expectation theorem is used in our elementary method, i.e.,

$$\mathbb{E}\left[g(N(t+\Delta), \lambda(t+\Delta), t+\Delta)\right] = \mathbb{E}\left[\mathbb{E}[g(N(t+\Delta), \lambda(t+\Delta), t+\Delta)|\mathcal{F}_t]\right].$$

Remark 4. We have not mentioned the marks in Hawkes processes in this procedure, but, in fact, the i.i.d. marks do not affect our method in essential, they just make the computations complicated. Thus in the following context, we consider them not always.

2.2. Relation of our method to use of infinitesimal generators and Dynkin's formula

From the procedure provided above, we know that the key issue is to get the differential equation. To embody this key point, we summarize and provide the detailed steps for a general function $g(N(t), \lambda(t), t)$ which and its related expectations must satisfy the following conditions:

(i) $\frac{\partial}{\partial t}g(N, \lambda, t)$ is uniformly continuous in t;

(ii) $\frac{\partial}{\partial \lambda}g(N, \lambda, t)$ and $\frac{\partial}{\partial \lambda}g(N, \lambda, t)\frac{\mathrm{d}\lambda}{\mathrm{d}t}$ are uniformly continuous in λ;

(iii) All expectations exist and are finite.

Given the filtration $\mathcal{F}_t = \sigma(N(s), 0 \leq s < t)$, and noting that $\lambda(t) = \lambda(t, N(t), Z)$, where the random variable Z is a mark in the Hawkes process $N(t)$ with distribution function $G_Z(z)$.

$$\begin{aligned}
&\mathbb{E}[g(N(t+\Delta),\lambda(t+\Delta),t+\Delta)|\mathcal{F}_t]\\
&\quad= g(N(t),\lambda_0(t+\Delta),t+\Delta)[1-\lambda(t)\Delta]\\
&\qquad+ g(N(t)+1,\lambda_1(t+\Delta),t+\Delta)[\lambda(t)\Delta]+o(\Delta)\\
&\quad= g(N(t),\lambda_0(t+\Delta),t+\Delta) - g(N(t),\lambda_0(t+\Delta),t+\Delta)\lambda(t)\Delta\\
&\qquad+ g(N(t)+1,\lambda_1(t+\Delta),t+\Delta)\lambda(t)\Delta+o(\Delta),
\end{aligned}$$

where $\lambda_0(t+\Delta)=\lambda(t+\Delta,N(t),Z_1,Z_2,\ldots,Z_{N(t)})$,

$\lambda_1(t+\Delta)=\lambda(t+\Delta,N(t)+1,Z_1,Z_2,\ldots,Z_{N(t)+1})$,

$\lim_{\Delta\downarrow 0}\lambda_0(t+\Delta)=\lambda(t,N(t),Z_1,Z_2,\ldots,Z_{N(t)})$ and,

$\lim_{\Delta\downarrow 0}\lambda_1(t+\Delta)=\lambda(t,N(t)+1,Z_1,Z_2,\ldots,Z_{N(t)+1})$.

We have

$$\begin{aligned}
&\lim_{\Delta\downarrow 0}\frac{\mathbb{E}\{\mathbb{E}[g(N(t+\Delta),\lambda(t+\Delta),t+\Delta)|\mathcal{F}_t]-g(N(t),\lambda(t),t)\}}{\Delta}\\
&\quad=\lim_{\Delta\downarrow 0}\frac{\mathbb{E}[g(N(t),\lambda_0(t+\Delta),t+\Delta)]-\mathbb{E}[g(N(t),\lambda_0(t+\Delta),t)]}{\Delta}\\
&\qquad+\lim_{\Delta\downarrow 0}\frac{\mathbb{E}[g(N(t),\lambda_0(t+\Delta),t)]-\mathbb{E}[g(N(t),\lambda(t),t)]}{\Delta}\\
&\qquad+\lim_{\Delta\downarrow 0}\{\mathbb{E}[\lambda(t)g(N(t)+1,\lambda_1(t+\Delta),t+\Delta)]\\
&\qquad-\mathbb{E}[\lambda(t)g(N(t),\lambda_0(t+\Delta),t+\Delta)]\}.
\end{aligned}$$

We get

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[g(N(t),\lambda(t),t)]=\mathbb{E}[\frac{\partial}{\partial t}g(N(t),\lambda(t),t)]+\mathbb{E}[\frac{\partial}{\partial\lambda}g(N(t),\lambda(t),t)\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)]\\
&+\mathbb{E}[\lambda(t)g(N(t)+1,\lambda(t)+Z,t)]-\mathbb{E}[\lambda(t)g(N(t),\lambda(t),t)]
\end{aligned}\tag{8}$$

provided that

$$\lim_{\Delta\downarrow 0}\frac{\mathbb{E}[g(N(t),\lambda_0(t+\Delta),t)]-\mathbb{E}[g(N(t),\lambda(t),t)]}{\Delta}=\mathbb{E}[\frac{\partial}{\partial\lambda}g(N(t),\lambda(t),t)\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)]$$

and

$$|\mathbb{E}[\lambda(t)g(N(t)+1,\lambda(t)+Z,t)]-\mathbb{E}[\lambda(t)g(N(t),\lambda(t),t)]|<\infty$$

hold.

Equation (8) can be used for any function $g(N,\lambda,t)$ for which the expectations of $\mathbb{E}[\frac{\partial}{\partial\lambda}g(N(t),\lambda(t),t)\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)]$ and $\mathbb{E}[\lambda(t)g(N(t)+1,\lambda(t)+Z,t)]$ can be calculated or can be expressed as functions of $\mathbb{E}[g(N(t),\lambda(t),t)]$ or Equation (8) can be solved analytically or numerically.

For completeness, we recall that the definition of an infinitesimal generator $\mathcal{A}$ of process $\boldsymbol{X}(t)$ is given as for any function $g : \Re^n \to \Re$,

$$\mathcal{A}g(\boldsymbol{x}) = \lim_{t\downarrow 0} \frac{\mathbb{E}[g(\boldsymbol{X}(t))] - g(\boldsymbol{x})}{t}, \ \boldsymbol{x} \in \Re^n, \ \boldsymbol{X}(0) = \boldsymbol{x}, \ \lim_{t\downarrow 0} g(\boldsymbol{X}(t)) = g(\boldsymbol{x}).$$

Furthermore, we can take an integral operation from time t to T $(t < T)$ on both sides of Equation (8) getting

$$\begin{aligned}
&\mathbb{E}[g(N(T), \lambda(T), T)|\mathcal{F}_t] - g(N(t), \lambda(t), t) \\
&= \int_t^T \mathbb{E}\{\frac{\partial}{\partial u} g(N(u), \lambda(u), u)|\mathcal{F}_t\}\mathrm{d}u + \int_t^T \mathbb{E}\{\frac{\partial}{\partial \lambda} g(N(u), \lambda(u), u)\frac{\mathrm{d}}{\mathrm{d}u}\lambda(u)|\mathcal{F}_t\}\mathrm{d}u \\
&\quad + \int_t^T \mathbb{E}[\lambda(u) g(N(u)+1, \lambda(u)+Z, u)|\mathcal{F}_t]\mathrm{d}u - \mathbb{E}[\lambda(t) g(N(t), \lambda(t), t)] \\
&= \mathbb{E}\{\int_t^T [\lambda(u) \int_0^\infty [g(N(u)+1, \lambda(u)+z, u) - g(N(u), \lambda(u), u)]\mathrm{d}G_Z(z) \\
&\qquad + \frac{\partial g}{\partial u} + \frac{\partial g}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}u}|\mathcal{F}_t]\mathrm{d}u\}.
\end{aligned}$$

That is to say, we have

$$\mathbb{E}[g(N(T), \lambda(T), T)|\mathcal{F}_t] - g(N(t), \lambda(t), t) = \mathbb{E}[\int_t^T \mathcal{A}g(N(u), \lambda(u), u)\mathrm{d}u|\mathcal{F}_t], \tag{9}$$

which is the Dynkin formula. The infinitesimal generator $\mathcal{A}$ is given for a function $g: \ \boldsymbol{N} \times \Re^+ \times \Re^+ \to \Re$ with continuous partial derivatives by

$$\mathcal{A}g(N, \lambda, t) = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \lambda \int_0^\infty [g(N+1, \lambda+z, t) - g(N, \lambda, t)]\mathrm{d}G_Z(z). \tag{10}$$

Based on the above analysis, we see that, if the process is Markovian, the results obtained by using the Dynkin formula can also be obtained by using our method. Furthermore, our method is very naïve and simple. However, our method is more general because, as the following example shows, it can also be applied in non-Markovian cases for which the Dynkin formula is not applicable. In fact, another non-Markovian case example will be presented in Section 4 as well.

Remark 5. We presented our elementary method by using a procedure rather than formula like Equation (8). Because (i) our method does not need any advanced theory and works for more situations rather than limited to the Markovian cases which the Dynkin formula needs; (ii) Our method can be operated simply without any conditions to be checked if it works. In fact, In subsection 3.2.3, we shall show this point via the comparison between our method and Dynkin formula.

We consider any regular counting process such that the probability of multiple events in an interval of length Δ is of order $o(\Delta)$ as $\Delta \to 0$. Then we have the following Theorem:

Theorem 1. *For a counting process such that* $\mathbb{P}\{N(t+\Delta) - N(t) \geq 2\} = o(\Delta)$*, then we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N^m(t)] = \sum_{j=0}^{m-1} \binom{m}{j} \mathbb{E}[N^j(t)\lambda(t)], \quad (m \geq 1). \tag{11}$$

Proof. The proof is similar to that in ***Steps 4*** and ***5*** in the procedure for the simple Hawkes process. □

Remark 6. Theorem 1 does not require the Markov property, it holds for all counting processes that do not have multiple events occurring simultaneously. This fact tells us that our method is an extension of Dynkin formula.

3. Applications on Self-Exciting Hawkes Processes

In this section, the applications on several self-exciting Hawkes processes for our elementary method given in Section 2 are shown, in which the detailed steps or numerical examples are presented.

3.1. Applications on a simple Hawkes process

Consider a simple Hawkes process $N(t), t \geq 0$ with an intensity function given in Section 2. In the following steps, we demonstrate our method for $\mathbb{E}[\lambda^n(t)]$, $t \geq 0$, and derive all general equations (i.e., equations (12), (14) and (15)). Then consider special cases like $n = m = 2$.

Step 1. Set our target as $\mathbb{E}[\lambda^n(t)]$ for any positive integer n $(n \geq 1)$.

Step 2. Find two probabilities such as

$$\mathbb{P}\{N(t+\Delta) - N(t) = 0|\mathcal{F}_t\} = 1 - \lambda(t)\Delta + o(\Delta),$$
$$\mathbb{P}\{N(t+\Delta) - N(t) = 1|\mathcal{F}_t\} = \lambda(t)\Delta + o(\Delta),$$

where the filtration $\mathcal{F}_t = \sigma(N(s), 0 \leq s < t)$ represents the history of the simple Hawkes process up to time t, and $\Delta > 0$. And we also know that $\mathbb{P}\{N(t+\Delta) - N(t) \geq$

$2|\mathcal{F}_t\} = o(\Delta)$ due to the assumption.

Step 3. Calculate the intensity function $\lambda(t+\Delta)$ given $\mathcal{F}_t$ and two events

$$\{N(t+\Delta) - N(t) = 0\} \text{ and } \{N(t+\Delta) - N(t) = 1\}$$

(i) When $\{N(t+\Delta) - N(t) = 0\}$ occurs,

$$\lambda_0(t+\Delta) = v + \sum_{T_i < t+\Delta} \alpha e^{-\beta(t-T_i+\Delta)} = (1-\beta\Delta)\lambda(t) + \beta v\Delta + o(\Delta);$$

(ii) When $\{N(t+\Delta) - N(t) = 1\}$ occurs,

$$\begin{aligned}\lambda_1(t+\Delta) &= v + \sum_{T_i<t} \alpha e^{-\beta(t-T_i+\Delta)} + \alpha e^{-\beta\Delta_1} \\ &= \lambda_0(t+\Delta) + \alpha(1-\beta\Delta_1) + o(\Delta_1),\ 0 < \Delta_1 < \Delta.\end{aligned}$$

Step 4. Calculate $\mathbb{E}[\lambda^n(t+\Delta)|\mathcal{F}_t]$.

Based on values of $\lambda_0(t+\Delta)$ and $\lambda_1(t+\Delta)$ given in **Step 3**, we have

$$\begin{aligned}\mathbb{E}[\lambda^n(t+\Delta)|\mathcal{F}_t] &= \lambda_0^n(t+\Delta)[1-\lambda(t)\Delta] + \lambda_1^n(t+\Delta)\lambda(t)\Delta + o(\Delta) \\ &= \lambda_0^n(t+\Delta) + \lambda(t)\Delta\sum_{i=0}^{n-1}\binom{n}{i}\lambda_0^i(t+\Delta)(\alpha-\beta\Delta_1)^{n-i} + o(\Delta).\end{aligned}$$

$\mathbb{E}[\lambda^m(t+\Delta)|\mathcal{F}_t]$ is gotten because when $\{N(t+\Delta)-N(t) = 0\}$ occurs, then $\lambda(t+\Delta) = \lambda_0(t+\Delta)$. When $\{N(t+\Delta) - N(t) = 1\}$ occurs, then $\lambda(t+\Delta) = \lambda_1(t+\Delta)$.

Step 5. Get a differential equation on $\mathbb{E}[\lambda^n(t)]$. Let $\Delta \to 0$, and using the result obtained in **Step 4** and the following equality

$$\mathbb{E}[\mathbb{E}[\lambda^n(t+\Delta)|\mathcal{F}_t]] = \mathbb{E}[\lambda^n(t+\Delta)],$$

then we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^n(t)] = n\beta v\mathbb{E}[\lambda^{n-1}(t)] - n\beta\mathbb{E}[\lambda^n(t)] + \sum_{i=0}^{n-1}\binom{n}{i}\alpha^{n-i}\mathbb{E}[\lambda^{i+1}(t)],\ (n \geq 1), \quad (12)$$

Step 6. Solve the established differential equation for getting $\mathbb{E}[\lambda^n(t)]$. Now we consider the case of $\mathbb{E}[N^2(t)]$.

We have from Equation (11),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[N^2(t)\right] = \mathbb{E}\left[\lambda(t)\right] + 2\mathbb{E}\left[N(t)\lambda(t)\right], \quad (13)$$

thus we need two terms: $\mathbb{E}[\lambda(t)]$ and $\mathbb{E}[N(t)\lambda(t)]$, for getting $\mathbb{E}[N(t)^2]$.

Similarly, for $\mathbb{E}[N(t)\lambda(t)]$, following **Steps 1** to **5**, we can get differential equations as follows.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N(t)\lambda(t)] = \beta v\mathbb{E}[N(t)] + \mathbb{E}[\lambda(t)] + (\alpha-\beta)\mathbb{E}[N(t)\lambda(t)] + \mathbb{E}[\lambda^2(t)]. \tag{14}$$

We thus can set up and solve a chain of differential equations.

$$\mathbb{E}[\lambda(t)] \to \mathbb{E}[\lambda^2(t)] \to \cdots \to \mathbb{E}[\lambda^n(t)],\ (n \geq 1).$$

When $n = 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)] = (\alpha-\beta)\mathbb{E}[\lambda(t)] + \beta v \tag{15}$$

with initial condition $\mathbb{E}[\lambda(0)] = v$. This differential equation is easily solved as

$$\mathbb{E}[\lambda(t)] = \begin{cases} \dfrac{-\beta v}{\alpha-\beta} + \dfrac{\alpha v}{\alpha-\beta}e^{(\alpha-\beta)t}, & \text{when } \alpha \neq \beta, \\ v + v\alpha t, & \text{when } \alpha = \beta. \end{cases} \tag{16}$$

Taking $m = 1$, in Equation (11) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[N(t)\right] = \mathbb{E}\left[\lambda(t)\right] \Rightarrow$$

$$\mathbb{E}[N(t)] = \int_0^t \mathbb{E}\left[\lambda(u)\right]\mathrm{d}u = \begin{cases} \dfrac{-\beta vt}{\alpha-\beta} + \dfrac{\alpha v}{(\alpha-\beta)^2}[e^{(\alpha-\beta)t} - 1], & \text{when } \alpha \neq \beta, \\ vt + \dfrac{1}{2}v\alpha t^2, & \text{when } \alpha = \beta. \end{cases} \tag{17}$$

Furthermore, we have from Equation (12),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^2(t)] = (2\beta v + \alpha^2)\mathbb{E}[\lambda(t)] + (2\alpha - 2\beta)\mathbb{E}[\lambda^2(t)],\ (n = 2). \tag{18}$$

Solving Equation (18), we have, with initial condition $\mathbb{E}[\lambda^2(0)] = v^2$,

$$\mathbb{E}[\lambda^2(t)] = \begin{cases} \left[\dfrac{\Delta_{11}}{\alpha-\beta} + v^2 - \dfrac{\Delta_{12}}{\alpha-\beta}\right]e^{-2(\alpha-\beta)t}, & \text{when } \alpha \neq \beta, \\ v(\alpha^2 + 2\beta v)(\dfrac{1}{2}t^2 + t) + v^2, & \text{when } \alpha = \beta, \end{cases}$$

where $\Delta_{11} = (\alpha^2+2\beta v)v\left[\dfrac{\alpha e^{3(\alpha-\beta)t}}{3(\alpha-\beta)} - \dfrac{\beta e^{2(\alpha-\beta)t}}{2(\alpha-\beta)}\right], \Delta_{12} = (\alpha^2+2\beta v)v\left[\dfrac{\alpha}{3(\alpha-\beta)} - \dfrac{\beta}{2(\alpha-\beta)}\right]$.

Similarly, solving Equation (14), we get, with initial condition $\mathbb{E}[N(0)\lambda(0)] = 0$,

$$\mathbb{E}[N(t)\lambda(t)] = \begin{cases} e^{(\alpha-\beta)t}\int_0^t [\beta v\mathbb{E}[N(u)] + \mathbb{E}[\lambda(u)] + \mathbb{E}[\lambda^2(u)]]\mathrm{d}u, & \text{when } \alpha \neq \beta, \\ \int_0^t [\beta v\mathbb{E}[N(u)] + \mathbb{E}[\lambda(u)] + \mathbb{E}[\lambda^2(u)]]\mathrm{d}u, & \text{when } \alpha = \beta. \end{cases} \tag{19}$$

Based on Equation (13), we have, with initial condition $\mathbb{E}\left[N^2(0)\right] = 0$,

$$\mathbb{E}\left[N^2(t)\right] = \int_0^t \left[\mathbb{E}\left[\lambda(u)\right] + 2\mathbb{E}\left[N(u)\lambda(u)\right]\mathrm{d}u, \tag{20}$$

which is a closed-form for the second order moment of a simple Hawkes process. In fact, these are not so easy to describe in general but results for low order moments are easy to obtain. Results for all first and second order moments can be found from Equations (16) and (17) for first order then Equations (14), (18) to (20), for second order moments - see example 1 below. All expectations of order m can be found as follows: calculate all expectations of order less than m; find expectation $\mathbb{E}[\lambda^m(t)]$ using the recursion of powers of $\lambda(t)$ from Equation (12); finally, use appropriate versions of corresponding Equation (21) (to be given in subsection 3.2) to obtain successively all expectations of order m,

$$\mathbb{E}[\lambda^m(t)] \to \mathbb{E}[N_H(t)\lambda^{m-1}(t)] \to \cdots \to \mathbb{E}[N_H^{m-1}(t)\lambda(t)] \to \mathbb{E}[N_H^m(t)],\ (m \geq 1).$$

The whole recursion can be represented as

order 1: $\mathbb{E}[\lambda(t)] \to \mathbb{E}[N(t)] \to$ order 2

order 2: $\mathbb{E}[\lambda^2(t)] \to \mathbb{E}[N(t)\lambda(t)] \to \mathbb{E}[N^2(t)] \to$ order 3

...

order $m-1$: $\cdots \to$ order m

order m: $\mathbb{E}[\lambda^m(t)] \to \mathbb{E}[N(t)\lambda^{m-1}(t)] \to \cdots \to \mathbb{E}[N^{m-1}(t)\lambda(t)] \to \mathbb{E}[N^m(t)],\ (m \geq 1)$

In the following, a numerical example is given for a simple Hawkes process.

Example 1. For a simple Hawkes process, if $\alpha = 2$, $\beta = 3$, $v = 1$, we have

$$\mathbb{E}[\lambda(t)] = 3 - 2e^{-t}$$

from Equation (16), then from Equation (17)

$$\mathbb{E}[N(t)] = \int_0^t \mathbb{E}[\lambda(u)]\mathrm{d}u = -2 + 3t + 2e^{-t};$$

$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^2(t)] + 2\mathbb{E}[\lambda^2(t)] = 10\mathbb{E}\left[\lambda(t)\right]$ from Equation (12) $\Rightarrow \mathbb{E}[\lambda^2(t)] = 15 - 20e^{-t} + 6e^{-2t}$;

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N(t)\lambda(t)] + \mathbb{E}[N(t)\lambda(t)] = 3\mathbb{E}\left[N(t)\right] + 2\mathbb{E}[\lambda(t)] + \mathbb{E}[\lambda^2(t)] \text{ from Equation (14)}$$

$$\Rightarrow \mathbb{E}[N(t)\lambda(t)] = 6 + 9t - 18te^{-t} - 6e^{-2t};$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N^2(t)] = \mathbb{E}\left[\lambda(t)\right] + 2\mathbb{E}\left[N(t)\lambda(t)\right] \text{ from Equation (21)}$$

$$\Rightarrow \mathbb{E}[N^2(t)] = \int_0^t \mathbb{E}\left[\lambda(u)\right] + 2\mathbb{E}\left[N(u)\lambda(u)\right]\mathrm{d}u$$

$$\Rightarrow \mathbb{E}[N^2(t)] = -44 + 15t + 9t^2 + 36te^{-t} + 38e^{-t} + 6e^{-2t}.$$

We further obtain in the usual way

$$\mathbb{V}\mathrm{ar}[\lambda(t)] = 6 - 8e^{-t} + 2e^{-2t},$$

$$\mathbb{V}\mathrm{ar}[N(t)] = -48 + 27t + 24te^{-t} + 46e^{-t} + 2e^{-2t},$$

$$\mathbb{C}\mathrm{ov}[N(t), \lambda(t)] = \mathbb{E}[N(t)\lambda(t)] - \mathbb{E}[N(t)]\mathbb{E}[\lambda(t)] = 12 - 12te^{-t} - 10e^{-t} - 2e^{-2t}.$$

As shown in FIGURE 1, we see that correlation coefficient $\rho(t)$ of $N(t)$ and $\lambda(t)$ is a decreasing function of time t and $\lim_{t\downarrow 0}\rho(t) = 1$, $\lim_{t\uparrow\infty}\rho(t) = 0$, as we intuitively expect.

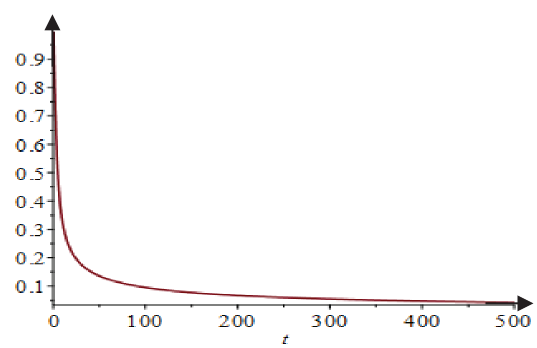

FIGURE 1: The correlation coefficient $\rho(t)$ of $N(t)$ and $\lambda(t)$ for a simple Hawke process..

3.2. Applications on a simple self-exciting marked Hawkes process

In this subsection, we shall use our method to obtain some results for the simple self-exciting marked Hawkes process defined as follows (see Equation (21)). A simple self-exciting marked Hawkes process $N_H(t)$ has intensity

$$\lambda(t) = v + \sum_{T_r < t} Z_i e^{-\beta(t - T_r)}, \; t > 0, \tag{21}$$

with $\{Z_r\}$ a sequence of i.i.d. random variables with distribution function $G_Z(z)$. An even simpler model is obtained if $G_Z(z)$ is a degenerate distribution with all values equal to a constant α, which is a simple Hawkes process we have discussed previously. It is easy to see that the pair $(N_H(t), \lambda(t))$ forms a Markov process. Within this subsection we simplify the notation by writing $\mathbb{E}[Z^r] = \mu_r$, $r > 0$. Note that throughout this paper we always use μ_i as the ith moments for random variables, their distribution functions can be known from the context.

3.2.1. *A result for* $\mathbb{E}[N_H^m(t)\lambda^n(t)]$.

Similar to that in subsection 3.1, we have that if there is no event occuring in $(t, t+\Delta)$, with probability $1 - \lambda(t)\Delta + o(\Delta)$, then $N_H(t+\Delta) = N_H(t)$ and,

$$\lambda_0(t+\Delta) = v + \sum_{T_i<t+\Delta} Z_i e^{-\beta(t+\Delta-T_i)} = v + e^{-\beta\Delta} \sum_{T_i<t} Z_i e^{-\beta(t-T_i)} = \lambda(t) - \nabla_1 + o(\Delta).$$

If there is an event occuring in $(t, t+\Delta)$, with probability $\lambda(t)\Delta + o(\Delta)$, then $N_H(t+\Delta) = N_H(t) + 1$ and

$$\begin{aligned} \lambda_1(t+\Delta) &= v + \sum_{T_i<t+\Delta} Z_i e^{-\beta(t+\Delta-T_i)} \\ &= \lambda(t) - \nabla_1 + Z_{N_H(t)+1}\nabla_2 + o(\Delta), \end{aligned}$$

where ∇_1 was given in ***Step 4*** of subsection 2.1, and $\nabla_2 = 1 - \beta\Delta_1$. Thus we can get Theorem 2 given below, which is a differential equation that can be solved recursively.

Theorem 2. *For any non-negative integers m and n, suppose the random marks Z_i, $(i = 1, 2, \ldots)$ with distribution function $G_Z(z)$ are independent of previous values of Hawkes process $N_H(t)$ and its intensity $\lambda(t)$; and its kernel function is $\gamma(t) = e^{-\beta t}$. Then we have, using the notation $f_{m,n}(t) = \mathbb{E}[N_H^m(t)\lambda^n(t)]$,*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f_{m,n}(t) &= n\beta v f_{m,n-1}(t) - n\beta f_{m,n}(t) + \sum_{j=0}^{m-1} \binom{m}{j} f_{j,n+1}(t) \\ &\quad + \sum_{j=0}^{m} \sum_{i=0}^{n-1} \binom{m}{j} \binom{n}{i} \mu_{n-i} f_{j,i+1}(t), \end{aligned} \tag{22}$$

where $\sum_{i=0}^{w} [\bullet] = 0$, when $w < 0$, and $\mu_r = \mathbb{E}[Z^r] = \int_0^\infty u^r \mathrm{d}G_Z(u) < \infty$.

We see that $f_{m,n}(t)$ satisfied a first-order differential difference equation, which can be solved easily. Furthermore, we can know that getting $f_{m,n}(t)$ needs all values for $f_{j,i}(t)$, $(j,i) \in \{(0,1,\ldots,m-1),(0,1,\ldots,n+1)\}$.

In addition, Theorem 2 can be obtained by using the Dynkin formula approach too. We now show that the same general result can be derived from Dynkin's formula. Because of the Markov property for stochastic process $(N_H(t),\lambda(t))$, and using the results in [9] and [8], we consider the function $g(N,\lambda,t)=N_H^m(t)\lambda^n(t)$, then we have

$$\begin{aligned}\mathcal{A}g(N,\lambda,t) &= \beta(v-\lambda)\frac{\partial g}{\partial\lambda}+\lambda[\int_0^\infty g(N+1,\lambda+x,t)\mathrm{d}G(x)-g(N,\lambda,t)]\\ &= n\beta(v-\lambda)N^m\lambda^{n-1}+(N+1)^m\lambda\sum_{i=0}^{n}\binom{n}{i}\lambda^i\mu_{n-i}-N^m\lambda^{n+1}\end{aligned}$$

provided that

$$|(N+1)^m\sum_{i=0}^{n}\binom{n}{i}\lambda^i\mu_{n-i}-N^m\lambda^n|<\infty,$$

where the operator $\mathcal{A}$ is the infinitesimal generator of the Hawkes process with intensity function $\lambda(t)=v+\sum_{T_i<t}Z_ie^{-\beta(t-T_i)}$, $\mu_i=\int_0^\infty z^i\mathrm{d}G_Z(z)$, $i=0,1,\ldots$. Based on the Dynkin formula, we have, for $t<T$,

$$\mathbb{E}[N_H^m(T)\lambda^n(T)|\mathcal{F}_t]=N_H^m(t)\lambda^n(t)+\mathbb{E}[\int_t^T\mathcal{A}g(N_H(u),\lambda(u),u)\mathrm{d}u|\mathcal{F}_t].$$

Take expectation operation on both sides of the above equation and set $t=0$: we have

$$\mathbb{E}[N_H^m(T)\lambda^n(T)]=\int_0^T\mathbb{E}[\mathcal{A}g(N_H(u),\lambda(u),u)\mathrm{d}u].$$

Differentiating the above equation, we get

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}f_{m,n}(t)=\mathbb{E}[\mathcal{A}g(N_H(t),\lambda(t),t)]\\ &=n\beta vf_{m,n-1}(t)-n\beta f_{m,n}(t)\\ &\quad+\sum_{j=0}^{m-1}\binom{m}{j}f_{j,n+1}(t)+\sum_{j=0}^{m}\sum_{i=0}^{n-1}\binom{m}{j}\binom{n}{i}\mu_{n-i}f_{j,i+1}(t),\end{aligned}$$

which is coincident with our result presented in Theorem 2. This again shows that our method gets the same result as the Dynkin formula in the cases where the latter method is valid.

3.2.2. *Auto-covariance and cross-covariance*

Here we consider the calculations of auto-covariance and cross-covariance functions for the simple self-exciting marked Hawkes process $N_H(t)$ and its intensity $\lambda(t)$ as follows.

Because auto-covariances are symmetric we need consider only $u < t$ when finding $\mathbb{E}[\lambda(t)\lambda(u)]$ and $\mathbb{E}[N_H(t)N_H(u)]$. Then

$$\begin{aligned}&\mathbb{E}[\lambda(t+\Delta)\lambda(u)|\mathcal{F}_t]\\&=\lambda(t)\lambda(u)-\nabla_1\lambda(u)+\mathbb{E}[Z_{N_H(t)+1}\nabla_2-\nabla_1]\lambda(u)\lambda(t)\Delta+o(\Delta),\end{aligned}$$

and so $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)\lambda(u)] = \beta v\mathbb{E}[\lambda(u)] + (\mu_1 - \beta)\mathbb{E}[\lambda(t)\lambda(u)]$, with solution

$$\mathbb{E}[\lambda(t)\lambda(u)] = \beta v\mathbb{E}[\lambda(u)]\frac{e^{(\mu_1-\beta)(t-u)}-1}{(\mu_1-\beta)} + e^{(\mu_1-\beta)(t-u)}\mathbb{E}[\lambda^2(u)], \qquad u<t, \tag{23}$$

in terms of expectations that we already know how to find.

We define the cross-covariance function of $N_H(t)$ and $\lambda(t)$ as

$$\mathbb{C}\mathrm{ov}[N_N(t),\lambda(t)] = \mathbb{E}[N_H(t)\lambda(u)] - \mathbb{E}[N_H(t)]\mathbb{E}[\lambda(u)].$$

To find $\mathbb{E}[N_H(t)\lambda(u)]$, when $u > t$,

$$\begin{aligned}&\mathbb{E}[N_H(t)\lambda(u+\Delta)|\mathcal{F}_u]\\&=N_H(t)\lambda(u)-N_H(t)\nabla_1(u)+N_H(t)\mathbb{E}[Z\nabla_2-\nabla_1(u)]\lambda(u)\Delta+o(\Delta).\end{aligned}$$

Then $\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}[N_H(t)\lambda(u)] = \beta v\mathbb{E}[N_H(t)] + (\mu_1-\beta)\mathbb{E}[N_H(t)\lambda(u)]$, which has solution

$$\mathbb{E}[N_H(t)\lambda(u)] = \beta v\mathbb{E}[N_H(t)]\frac{e^{(\mu_1-\beta)(u-t)}-1}{(\mu_1-\beta)} + e^{(\mu_1-\beta)(u-t)}\mathbb{E}[N_H(t)\lambda(t)]. \tag{24}$$

When $u < t$,

$$\mathbb{E}[N_H(t+\Delta)\lambda(u)|\mathcal{F}_t] = N_H(t)\lambda(u)[1-\lambda(t)\Delta] + [N_H(t)+1]\lambda(u)\lambda(t)\Delta + o(\Delta),$$

then

$$\mathbb{E}[N_H(t+\Delta)\lambda(u)] = \mathbb{E}[N_H(t)\lambda(u)] + \mathbb{E}[\lambda(u)\lambda(t)]\Delta + o(\Delta),$$

which means $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_H(t)\lambda(u)] = \mathbb{E}[\lambda(u)\lambda(t)]$ which has solution

$$\mathbb{E}[N_H(t)\lambda(u)] = \mathbb{E}[N_H(u)\lambda(u)] + \int_u^t \mathbb{E}[\lambda(u)\lambda(v)]\mathrm{d}v, \qquad u<t, \tag{25}$$

in terms of expectations that we know how to calculate. Finally,

$$\mathbb{E}[N_H(t+\Delta)N_H(u)|\mathcal{F}_t] = N_H(t)N_H(u)[1-\lambda(t)\Delta] + (N_H(t)+1)N_H(u)\lambda(t)\Delta + o(\Delta)$$

gives us the differential equation

$$\frac{d}{dt}\mathbb{E}[N_H(t)N_H(u)] = \mathbb{E}[N_H(u)\lambda(t)]$$

with solution in terms of expectations that we already know how to find

$$\mathbb{E}[N_H(t)N_H(u)] = \mathbb{E}[N_H^2(u)] + \int_u^t \mathbb{E}[N_H(u)\lambda(v)]\mathrm{d}v, \qquad u < t. \tag{26}$$

3.2.3. *Correlation between* $N_H(t)$ *and* $\int_0^t \lambda(u)\mathrm{d}u$

Similar to subsection 3.2.2, the related discussions are presented as follows.

If there is no event occuring in $(t, t+\Delta)$, with probability $1-\lambda(t)\Delta + o(\Delta)$, then $N_H(t+\Delta) = N_H(t)$ and

$$\begin{aligned}
\int_0^{t+\Delta} \lambda(u)\mathrm{d}u &= \int_0^t \lambda(u)\mathrm{d}u + \int_t^{t+\Delta} [v + \sum_{T_i<u} Z_i e^{-\beta(u-T_i)}]\mathrm{d}u \\
&= \int_0^t \lambda(u)\mathrm{d}u + \int_t^{t+\Delta} [v + \sum_{T_i<t} Z_i e^{-\beta(u-T_i)}]\mathrm{d}u \\
&= \int_0^t \lambda(u)\mathrm{d}u + \lambda(t)\Delta + o(\Delta).
\end{aligned}$$

If there is an event occuring in $(t, t+\Delta)$, with probability $\lambda(t)\Delta + o(\Delta)$, then $N_H(t+\Delta) = N_H(t) + 1$ and we get

$$\begin{aligned}
&\int_0^{t+\Delta} \lambda(u)\mathrm{d}u \\
&= \int_0^t \lambda(u)\mathrm{d}u + \int_t^{T_{N_H(t)+1}} [v + \sum_{T_i<u} Z_i e^{-\beta(u-T_i)}]\mathrm{d}u + \int_{T_{N_H(t)+1}}^{t+\Delta} [v + \sum_{T_i<u} Z_i e^{-\beta(u-T_i)}]\mathrm{d}u \\
&= \int_0^t \lambda(u)\mathrm{d}u + \int_t^{t+\Delta} [v + \sum_{T_i<t} Z_i e^{-\beta(u-T_i)}]\mathrm{d}u + Z_{N_H(t)+1} \int_{T_{N_H(t)+1}}^{t+\Delta} e^{-\beta(u-T_{N_H(t)+1})}\mathrm{d}u \\
&= \int_0^t \lambda(u)\mathrm{d}u + \lambda(t)\Delta + Z_{N_H(t)+1}\beta^{-1}[1 - e^{-\beta(t+\Delta-T_{N_H(t)+1})}] + o(\Delta) \\
&= \int_0^t \lambda(u)\mathrm{d}u + \lambda(t)\Delta + \nabla_3 + o(\Delta),
\end{aligned}$$

where $\nabla_3 = Z_{N_H(t)+1}\Delta'$, $0 < \Delta' < \Delta$. Then we have

$$\begin{aligned}
&\mathbb{E}\{N_H^m(t+\Delta)\lambda^n(t+\Delta)[\int_0^{t+\Delta}\lambda(u)\mathrm{d}u]^l|\mathcal{F}_t\}\\
&= N_H^m(t)[\lambda(t)-\Delta_1]^n[\int_0^t\lambda(u)\mathrm{d}u+\lambda(t)\Delta]^l[1-\lambda(t)\Delta]+[N_H(t)+1]^m\\
&\quad\times[\lambda(t)-\nabla_1+\nabla_2]^n[\int_0^t\lambda(u)\mathrm{d}u+\lambda(t)\Delta+\nabla_3]^l\lambda(t)\Delta+o(\Delta)\\
&=\{N_H^m(t)\lambda^n(t)+N_H^m(t)\sum_{i=0}^{n-1}\binom{n}{i}(-\nabla_1)^{n-i}\lambda^i(t)-N_H^m(t)\sum_{i=0}^{n}\binom{n}{i}\\
&\quad\times(-\nabla_1)^{n-i}\lambda^{i+1}(t)\Delta+\sum_{j=0}^{m}\binom{m}{j}N_H^j(t)\sum_{i=0}^{n}\binom{n}{i}(\nabla_2-\nabla_1)^{n-i}\lambda^{i+1}(t)\Delta\}\\
&\quad\times[\int_0^t\lambda(u)\mathrm{d}u]^l+lN_H^m(t)\lambda^{n+1}(t)[\int_0^t\lambda(u)\mathrm{d}u]^{l-1}\Delta+o(\Delta).
\end{aligned}$$

Thus similar to Theorem 2, we can get

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{N_H^m(t)\lambda^n(t)[\int_0^t\lambda(u)\mathrm{d}u]^l\}\\
&=n\beta v\mathbb{E}\{N_H^m(t)\lambda^{n-1}(t)[\int_0^t\lambda(u)\mathrm{d}u]^l\}-n\beta\mathbb{E}\{N_H^m(t)\lambda^n(t)[\int_0^t\lambda(u)\mathrm{d}u]^l\}\\
&\quad+\sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}\{N_H^j(t)\lambda^{n+1}(t)[\int_0^t\lambda(u)\mathrm{d}u]^l\}+\sum_{j=0}^{m}\sum_{i=0}^{n-1}\binom{m}{j}\binom{n}{i}\mu_{n-i}\mathbb{E}\{N_H^j(t)\\
&\quad\times\lambda^{i+1}(t)[\int_0^t\lambda(u)\mathrm{d}u]^l\}+l\mathbb{E}\{N_H^m(t)\lambda^{n+1}(t)[\int_0^t\lambda(u)\mathrm{d}u]^{l-1}\}.
\end{aligned}\tag{27}$$

Similar to that in subsection 3.1, we thus have the following chain recursions for getting $\mathbb{E}[N_H^m(t)\int_0^t\lambda(u)\mathrm{d}u]$,

order 0: $\mathbb{E}[\int_0^t\lambda(u)\mathrm{d}u] \to$ order 1

order 1: $\mathbb{E}[\lambda(t)\int_0^t\lambda(u)\mathrm{d}u] \to \mathbb{E}[N_H(t)\int_0^t\lambda(u)\mathrm{d}u] \to$ order 2

...

order $m-1$: $\cdots \to$ order m

order m: $\mathbb{E}[\lambda^m(t)\int_0^t\lambda(u)\mathrm{d}u] \to \mathbb{E}[N_H(t)\lambda^{m-1}(t)\int_0^t\lambda(u)\mathrm{d}u] \to$

$\cdots \to \mathbb{E}[N_H^m(t)\int_0^t\lambda(u)\mathrm{d}u]$.

Similar recursions can be given for $l = 2$, $l = 3$ and so on.

For example, for $\mathbb{E}\{N_H(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\}$, we have

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\int\limits_0^t \lambda(u)\mathrm{d}u] = \mathbb{E}[\lambda(t)], \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{\lambda(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} = \beta v \mathbb{E}[\int\limits_0^t \lambda(u)\mathrm{d}u] + (\mu_1 - \beta)\mathbb{E}\{\lambda(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}[\lambda^2(t)], \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{N_H(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} = \mathbb{E}\{\lambda(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}[N_H(t)\lambda(t)]. \end{cases}$$

Remark 7. The first equation above holds always, the reason of listing it here is that it will be used below.

In fact, we can use the Dynkin formula to give the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = \beta v \mathbb{E}[\int_0^t \lambda(u)\mathrm{d}u] + (\mu_1 - \beta)\mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}[\lambda^2(t)].$$

Let $g(N_H(t), \lambda(t), t) = \lambda(t)\int\limits_0^t \lambda(u)\mathrm{d}u$, then we have

$$\begin{aligned} \mathcal{A}g(N_H(t), \lambda(t), t) &= \frac{\partial g}{\partial t} + \frac{\partial g}{\partial \lambda}\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} + \lambda(t)\{[\lambda(t) + Z]\int_0^t \lambda(u)\mathrm{d}u - \lambda(t)\int_0^t \lambda(u)\mathrm{d}u\} \\ &= \lambda^2(t) + \frac{\mathrm{d}\lambda(t)}{\mathrm{d}t}\int_0^t \lambda(u)\mathrm{d}u + \lambda(t)Z\int_0^t \lambda(u)\mathrm{d}u \\ &= \lambda^2(t) + [\beta v - \beta\lambda(t)]\int_0^t \lambda(u)\mathrm{d}u + \lambda(t)Z\int_0^t \lambda(u)\mathrm{d}u, \end{aligned}$$

then we have

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{\lambda(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} = \mathbb{E}[\mathcal{A}g(N_H(t), \lambda(t), t)] \\ &= \beta v \mathbb{E}[\int\limits_0^t \lambda(u)\mathrm{d}u] + (\mu_1 - \beta)\mathbb{E}\{\lambda(t)[\int\limits_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}[\lambda^2(t)], \end{aligned}$$

which is the equation we obtained by using our elementary approach.

Remark 8. When we see the procedure of using the Dynkin formula above, we can know it may sometimes have some confuse on derivation process for $\frac{\partial g}{\partial \lambda}$, because there is a $\lambda(t)$ in the integral, which tells us that our method is more simple and direct.

Example 2. For the simple self-exciting Hawkes processes in Example 1, we have

$$\mathbb{E}[\int_0^t \lambda(u)\mathrm{d}u] = \int_0^t \mathbb{E}[\lambda(u)]\mathrm{d}u = -2 + 3t + 2e^{-t},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = 3\mathbb{E}[\int_0^t \lambda(u)\mathrm{d}u] + \mathbb{E}[\lambda^2(t)] = 9 + 9t - 14e^{-t} + 6e^{-2t}$$
$$\Rightarrow \mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = e^{-t}\int_0^t e^u(9 + 9u - 14e^{-u} + 6e^{-2u})\mathrm{d}u = 9t - 14te^{-t} + 6e^{-t} - 6e^{-2t},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{N_H(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = \mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} + \mathbb{E}[N_H(t)\lambda(t)] = 6 + 18t - 32te^{-t} + 6e^{-t} - 12e^{-2t}$$
$$\Rightarrow \mathbb{E}\{N_H(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = -32 + 6t + 9t^2 + 32te^{-t} + 26e^{-t} + 6e^{-2t},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\{[\int_0^t \lambda(u)\mathrm{d}u]^2\} = 2\mathbb{E}\{\lambda(t)[\int_0^t \lambda(u)\mathrm{d}u]\} = 18t - 28te^{-t} + 12e^{-t} - 12e^{-2t}$$

$$\Rightarrow \mathbb{E}\{[\int_0^t \lambda(u)\mathrm{d}u]^2\} = -22 + 9t^2 + 28te^{-t} + 16e^{-t} + 6e^{-2t}.$$

Then we have

$$\mathbb{V}\mathrm{ar}[N_H(t)] = 2e^{-2t} + 46e^{-t} + 24te^{-t} + 27t - 48,$$
$$\mathbb{V}\mathrm{ar}[\int_0^t \lambda(u)\mathrm{d}u] = 2e^{-2t} + 24e^{-t} + 16te^{-t} + 12t - 26,$$
$$\mathbb{C}\mathrm{ov}[N_H(t), \int_0^t \lambda(u)\mathrm{d}u] = \mathbb{E}\{N_H(t)[\int_0^t \lambda(u)\mathrm{d}u]\} - \mathbb{E}[N_H(t)]\mathbb{E}[\int_0^t \lambda(u)\mathrm{d}u]$$
$$= 2e^{-2t} + 34e^{-t} + 20te^{-t} + 18t - 36.$$

Thus we have the correlation coefficient

$$\tilde{\rho}(t) = \frac{2e^{-2t} + 34e^{-t} + 20te^{-t} + 18t - 36}{\sqrt{2e^{-2t} + 46e^{-t} + 24te^{-t} + 27t - 48}\sqrt{2e^{-2t} + 24e^{-t} + 16te^{-t} + 12t - 26}}.$$

As shown in FIGURE 2, we see that correlation coefficient $\tilde{\rho}(t)$ of $N_H(t)$ and $\int_0^t \lambda(u)\mathrm{d}u$ is an increasing function of time t and $\lim_{t\downarrow 0}\tilde{\rho}(t) \approx 0.8662$, $\lim_{t\uparrow\infty}\tilde{\rho}(t) = 1$, which is interesting compared with FIGURE 1.

3.2.4. *Generating functions*

This subsection contains results for probability generating functions for the distribution of $N_H(t)$, Laplace transforms for the distribution of $\lambda(t)$ and the joint distribution of the two. As usual, we are here concerned with obtaining the appropriate partial differential equation. We start with the joint distribution:

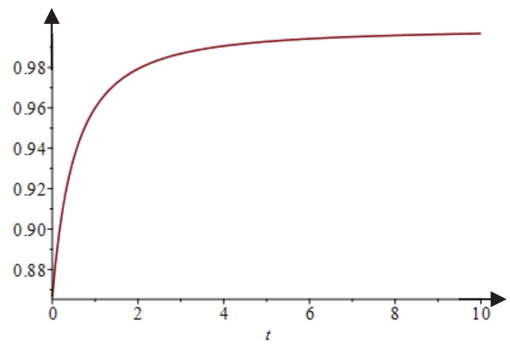

FIGURE 2: The curve of correlation coefficient $\tilde{\rho}(t)$.

Theorem 3. *For the Hawkes process with intensity function* $\lambda(t) = v + \sum_{T_r<t} Z_r e^{-\beta(t-T_r)}$, *we have*

$$\frac{\partial}{\partial t}G(\theta, s, t) = [1 - \beta s - \theta\phi(s)]\frac{\partial}{\partial s}G(\theta, s, t) - \beta v s G(\theta, s, t), \tag{28}$$

$G(\theta, s, t) := \mathbb{E}[\theta^{N_H(t)}e^{-s\lambda(t)}]$ *is the p.g.f/Laplace transform of the joint distribution of* $(N_H(t), \lambda(t))$*;* $\phi(s) = \mathbb{E}[e^{-sZ}] = \int_0^\infty e^{-sz}\mathrm{d}G_Z(z)$*, is the Laplace transform of the density function of the marked random variables. The initial condition is* $G(\theta, s, 0) = e^{-sv}$.

Proof.

$$\begin{aligned}
&G(\theta, s, t+\Delta|\mathcal{F}_t) = \mathbb{E}[\theta^{N_H(t+\Delta)}e^{-s\lambda(t+\Delta)}|\mathcal{F}_t] \\
&= \theta^{N_H(t)}e^{-s[\lambda(t)-\Delta(\beta\lambda(t)-\beta v)]}[1-\lambda(t)\Delta] \\
&\quad + \theta\mathbb{E}\left[\theta^{N_H(t)}e^{-s[\lambda(t)-\Delta(\beta\lambda(t)-\beta v)+Z_{N_H(t)+1}(1-\beta\Delta')]}\lambda(t)\right]\Delta + o(\Delta) \\
&= \theta^{N_H(t)}e^{-s\lambda(t)}\{1+s[\beta\lambda(t)-\beta v]\Delta\}[1-\lambda(t)\Delta] \\
&\quad + \theta\mathbb{E}\left[\theta^{N_H(t)}e^{-s[\lambda(t)-\Delta(\beta\lambda(t)-\beta v)]}\lambda(t)e^{-sZ_{N_H(t)+1}(1-\beta\Delta')}\right]\Delta + o(\Delta),\ 0<\Delta'<\Delta.
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{\partial}{\partial t}G(\theta, s, t) &= [\beta s - 1 + \theta\phi(s)]\mathbb{E}[\theta^{N_H(t)}e^{-s\lambda(t)}\lambda(t)] - \beta v s\mathbb{E}[\theta^{N_H(t)}e^{-s\lambda(t)}] \\
&= [1 - \beta s - \theta\phi(s)]\frac{\partial}{\partial s}G(\theta, s, t) - \beta v s G(\theta, s, t),
\end{aligned}$$

the proof is completed. □

Corollary 1. *Under the same conditions as above, the Laplace transform of the distribution of* $\lambda(t)$*,* $g(s, t) := \mathbb{E}\left[e^{-s\lambda(t)}\right] = G(1, s, t)$*, and satisfies the partial differential* *equation*

$$\frac{\partial}{\partial t}g(s,t) = [1-\beta s-\phi(s)]\frac{\partial}{\partial s}g(s,t) - \beta v s g(s,t), \tag{29}$$

with boundary/initial conditions $g(0,t)=1$, $g(s,0)=e^{-sv}$.

The proof follows simply from the theorem by taking $g(s,t)=G(1,s,t)$.

Corollary 2. *Under the same conditions as above, the p.g.f. of the distribution of* $N_H(t)$ *is given by* $G(\theta,t)=G(\theta,0,t)$ *and satisfies the differential equation*

$$\frac{\partial}{\partial t}G(\theta,t) = (\theta-1)\mathbb{E}\left[\theta^{N_H(t)}\lambda(t)\right], \tag{30}$$

subject to conditions $G(1,t)=1$; $G(\theta,0)=1$.

The joint probability generating function for $(N_H(t),\lambda(t))$ can be given by solving a partial differential equation, which provides, at least, a way to find the moments and count distribution for Hawkes processes.

4. Other Applications

Previous sections have dealt with results for the simple self-exciting Hawkes. In this section we consider briefly results for some other models such as mutually-exciting process, Cox process, dynamic contagion process, inhomogenous Poisson process and non-Markovian process.

4.1. Mutually-exciting process

We now consider a fairly general model comprising K mutually-exciting marked point processes with exponential exciting kernels. Suppose the intensity of $N_l(t)$ is given by

$$\lambda_l(t) = v_l + \sum_{j=1}^{K}\lambda_{lj}(t), \quad l=1,\ldots,K, \tag{31}$$

where

$$\lambda_{lj}(t) = \sum_{T_{j,r}<t} Z_{j,r}\alpha_{lj}e^{-\beta_{lj}(t-T_{j,r})}, \tag{32}$$

and $\{T_{j,r}\}_{r=1,2,\ldots}$ is a sequence of occurrence times of type-j events and $\{Z_{j,r}\}_{r=1,2,\ldots}$ the marks associated with them. Note that there is an extra factor α_{lj} that contributes to the effect of a type-j event, and its mark, on the future intensity of type-l events:

this is because the same event and same mark may have different effects on various other types of event. Then

$$\begin{aligned}\mathbb{E}\big[N_k^m(t+\Delta)\lambda_l^n(t+\Delta)|\mathcal{F}_t\big] &= N_k^m(t)\big\{\lambda_l(t)-\Delta\sum_{j=1}^{K}\lambda_{lj}(t)\beta_{lj}\big\}^n\big\{1-\Delta\sum_{j=1}^{K}\lambda_j(t)\big\}\\ &+\mathbb{E}\big[N_k^m(t)\sum_{i=1,i\neq k}^{K}\Delta\lambda_i(t)\big\{\lambda_l(t)-\Delta\sum_{j=1}^{K}\lambda_{lj}(t)\beta_{lj}+Z_{i,N_i(t)+1}\alpha_{li}(1-\Delta'\beta_{li})\big\}^n\big]\\ &+\mathbb{E}\big[\big(N_k(t)+1\big)^m\Delta\lambda_k(t)\big\{\lambda_l(t)-\Delta\sum_{j=1}^{K}\lambda_{lj}(t)\beta_{lj}+Z_{k,N_k(t)+1}\alpha_{lk}(1-\Delta'\beta_{lk})\big\}^n\big]+o(\Delta),\end{aligned}$$

where $0<\Delta'<\Delta$. Taking expectations and appropriate limits as $\Delta\to 0$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[N_k^m(t)\lambda_l^n(t)\right] &= -n\mathbb{E}\left[N_k^m(t)\lambda_l^{n-1}(t)\sum_{j=1}^{K}\lambda_{lj}(t)\beta_{lj}\right]-\mathbb{E}\left[\sum_{j=1}^{K}N_k^m(t)\lambda_l^n(t)\lambda_j(t)\right]\\ &+\mathbb{E}\left[N_k^m(t)\sum_{j=1,j\neq k}^{K}\lambda_j(t)\sum_{i=0}^{n}\binom{n}{i}\lambda_l^i(t)\mu_{n-i}^{(q)}\alpha_{lj}^{n-i}\right]\\ &+\mathbb{E}\left[(N_k(t)+1)^m\lambda_k(t)\sum_{i=0}^{n}\binom{n}{i}\lambda_l^i(t)\mu_{n-i}^{(k)}\alpha_{lk}^{n-i}\right],\end{aligned}$$

where $\mu_{n-i}^{(j)}=\mathbb{E}\left[Z_{j,r}^{n-i}\right]$ is the $(n-i)$th moment of $Z_{j,r}$ for all $r=1,2,\ldots$.

This can be written as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[N_k^m(t)\lambda_l^n(t)\right] &= -n\sum_{j=1}^{K}\beta_{lj}\mathbb{E}\left[N_k^m(t)\lambda_{lj}(t)\lambda_l^{n-1}(t)\right]+\sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}\left[N_k^j(t)\lambda_k(t)\lambda_l^n(t)\right]\\ &+\sum_{i=0}^{n-1}\binom{n}{i}\sum_{j=1,j\neq k}^{K}\mu_{n-i}^{(j)}\alpha_{lj}^{n-i}\mathbb{E}\left[N_k^m(t)\lambda_l^i(t)\lambda_j(t)\right]\\ &+\sum_{j=0}^{m}\sum_{i=0}^{n-1}\binom{m}{j}\binom{n}{i}\mu_{n-i}^{(k)}\alpha_{lk}^{n-i}\mathbb{E}\left[N_k^j(t)\lambda_l^i(t)\lambda_k(t)\right].\end{aligned} \tag{33}$$

4.2. Cox process

Cox [4] introduced a doubly stochastic model of point processes, now usually called Cox processes. More recent treatments of the properties of general classes of Cox processes are given, for example, by [2] and [17]. For this paper, we consider a simple version used by [7] in the context of catastrophe insurance. It has intensity

$$\lambda(t)=v+\sum_{T_r<t}Y_re^{-\beta(t-T_r)},\quad t>0, \tag{34}$$

$\{Y_r\}_{r=1,2,\ldots}$ is a sequence of i.i.d. random variables with distribution function $G_Y(y)$, $y>0$. This looks much like Equation (1) for the Hawkes process when $v(t)=v$. But there is one very big difference: the events $\{T_r\}$ on the right of Equation (34) arise from a time homogeneous Poisson process with rate η, and the $\{Y_r\}$ are marks associated with events of the Poisson process; in contrast, the events on the right of Equation (1) belong to the same process that is generated by the intensity $\lambda(t)$ on the left — it is this that drives the contagious property of the Hawkes process, with the occurrence of past events increasing the intensity that governs future events.

Note that the above Cox model involves two point processes, the Cox process $N_C(t)$ and the Poisson process, $N_P(t)$, that drives it. These can be considered as forming a bivariate mutually-exciting pair. In this case the general form of Equation (2) can be written as the special case

$$\begin{cases} \lambda_C(t) = v + \sum_{T_{P,r}<t} Y_i e^{-\beta(t-T_{P;r})}, \\ \lambda_P(t) = \eta. \end{cases} \tag{35}$$

In general, each intensity of a bivariate mutually-exciting Hawkes process consists of a baseline intensity and two exciting terms: a self-exciting term and a cross-exciting term. In this case we see that the Cox intensity contains a term that is excited by the Poisson process, but no self-exciting term; the Poisson intensity has no exciting terms at all. So they are not really mutually-exciting, because the Cox process is not excited at all!

Similarly, we know that the triple $(N_C(t), N_P(t), \lambda_C(t))$ forms a Markov process.

For the Cox process we have two point processes, $N_C(t)$ and the Poisson process, $N_P(t)$, that drives it. Therefore, for small time increment Δ, we need four probabilities

$$\mathbb{P}\{\{N_C(t+\Delta) - N_C(t) = 0\} \cap \{N_P(t+\Delta) - N_P(t) = 0\} | \mathcal{F}_t\}$$
$$= [1 - \lambda(t)\Delta][1 - \eta\Delta] + o(\Delta) = 1 - (\eta + \lambda(t))\Delta + o(\Delta);$$

$$\mathbb{P}\{\{N_C(t+\Delta) - N_C(t) = 0\} \cap \{N_P(t+\Delta) - N_P(t) = 1\} | \mathcal{F}_t\} = \eta\Delta + o(\Delta);$$

$$\mathbb{P}\{\{N_C(t+\Delta) - N_C(t) = 1\} \cap \{N_P(t+\Delta) - N_P(t) = 0\} | \mathcal{F}_t\} = \lambda(t)\Delta + o(\Delta);$$

$$\mathbb{P}\{\{N_C(t+\Delta) - N_C(t) = 1\} \cap \{N_P(t+\Delta) - N_P(t) = 1\} | \mathcal{F}_t\} = o(\Delta).$$

Note that the reasons are, for example, for the above last equality as follows.

Because, we can consider two cases: (i) $\{N_C(t+\Delta) - N_C(t) = 1\}$ occurs not later than occurring of $\{N_P(t+\Delta) - N_P(t) = 1\}$; (ii) $\{N_C(t+\Delta) - N_C(t) = 1\}$ occurs after occurring of $\{N_P(t+\Delta) - N_P(t) = 1\}$.

For case (i), we have

$$
\begin{aligned}
&\mathbb{P}\{\{N_C(t+\Delta)-N_C(t)=1\}\cap\{N_P(t+\Delta)-N_P(t)=1\}|\mathcal{F}_t\}\\
&=\mathbb{P}\{\{N_P(t+\Delta)-N_P(t)=1\}|\{N_C(t+\Delta)-N_C(t)=1\},\mathcal{F}_t\}\\
&\quad\times\mathbb{P}\{\{N_C(t+\Delta)-N_C(t)=1\}|\mathcal{F}_t\}\\
&=\mathbb{P}\{\{N_P(t+\Delta)-N_P(t)=1\}|\mathcal{F}_t\}\mathbb{P}\{\{N_C(t+\Delta)-N_C(t)=1\}|\mathcal{F}_t\}\\
&\quad\text{(due to } N_C(t) \text{ does not affect } N_P(t), \text{ and } \mathbb{P}\{N_C(t+\Delta)-N_C(t)\geq 2\}=o(\Delta) \text{ if}\\
&\quad\{N_P(t+\Delta)-N_P(t)=1\} \text{ affects } N_C(t))\\
&=\eta\Delta\lambda(t)\Delta+o(\Delta)=o(\Delta).
\end{aligned}
$$

For case (ii), we have,

$$
\begin{aligned}
&\mathbb{P}\{\{N_C(t+\Delta)-N_C(t)=1\}\cap\{N_P(t+\Delta)-N_P(t)=1\}|\mathcal{F}_t\}\\
&=\mathbb{P}\{\{N_C(t+\Delta)-N_C(t)=1\}|\{N_P(t+\Delta)-N_P(t)=1\},\mathcal{F}_t\}\mathbb{P}\{\{N_P(t+\Delta)-N_P(t)=1\}|\mathcal{F}_t\}\\
&=\lambda(t+\Delta)\lambda(t)\Delta\mathbb{P}\{\{N_P(t+\Delta)-N_P(t)=1\}|\mathcal{F}_t\}+o(\Delta)\\
&=\lambda(t+\Delta)\lambda(t)\Delta\eta\Delta+o(\Delta)=o(\Delta),
\end{aligned}
$$

where

$$
\begin{aligned}
\lambda(t+\Delta) &= v+\sum_{T_{P,i}<t+\Delta} Y_i e^{-\beta(t-T_{P,i}+\Delta)}\\
&= v+e^{-\beta\Delta}[\lambda(t)-v]+Y^*e^{-\beta\Delta_1}, \qquad (0<\Delta_1<\Delta)\\
&= (1-\beta\Delta)\lambda(t)+\beta v\Delta+Y^*e^{-\beta\Delta_1}+o(\Delta).
\end{aligned}
$$

For the Cox process we have

$$
\lambda(t+\Delta)=\begin{cases}\lambda(t)-\Delta[\beta\lambda(t)-\beta v]+o(\Delta), & \text{given } \{N_P(t+\Delta)-N_P(t)=0\},\\ \lambda(t)-\Delta[\beta\lambda(t)-\beta v]+Y_{N_P(t)+1}+o(\Delta), & \text{given } \{N_P(t+\Delta)-N_P(t)=1\}.\end{cases}
$$

Thus, apart from the constant term, the intensity decays exponentially with an added jump, Y, when a new Poisson event occurs at a time that precedes the end of the time increment by Δ', $(\Delta'<\Delta)$.

Let $\nabla_4=Y_{N_P(t)+1}(1-\beta\Delta')$, so $\lim_{\Delta\downarrow 0}(\nabla_4-\nabla_1)=Y_{N_P(t)+1}$. Then we have

$$
\begin{aligned}
&\mathbb{E}[N_C^m(t+\Delta)\lambda^n(t+\Delta)|\mathcal{F}_t]\\
&=N_C^m(t)(\lambda(t)-\nabla_1)^n(1-(\lambda(t)+\eta)\Delta)+N_C^m(t)(\lambda(t)-\nabla_1+\nabla_4)^n\eta\Delta\\
&\quad+(N_C(t)+1)^m(\lambda(t)-\nabla_1)^n\lambda(t)\Delta+o(\Delta),
\end{aligned}
$$

and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_C^m(t)\lambda^n(t)] &= \lim_{\Delta\downarrow 0}\frac{\mathbb{E}[N_C^m(t+\Delta)\lambda^n(t+\Delta)]-\mathbb{E}\left[N_C^m(t)\lambda^n(t)\right]}{\Delta}\\ &= n\beta v\mathbb{E}[N_C^m(t)\lambda^{n-1}(t)]-n\beta\mathbb{E}[N_C^m(t)\lambda^n(t)]\\ &\quad+\eta\sum_{i=0}^{n-1}\binom{n}{i}\mu_{n-i}\mathbb{E}[N_C^m(t)\lambda^i(t)]\\ &\quad+\sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N_C^j(t)\lambda^{n+1}(t)], \end{aligned} \tag{36}$$

where $\mu_r=\mathbb{E}[Y^r]$ is the rth moment of the mark Y. In fact, because of the Markov property of $(N_C(t),\lambda(t))$, we have the same result for Cox process by using the Dynkin formula for the objective function $\mathbb{E}[N_C^m(t)\lambda^n(t)]$.

Differential Equation (36) can be solved, subject to initial conditions $\mathbb{E}[N_C^i(0)]=0$, $\mathbb{E}[\lambda^i(0)]=v^i$, $(i=1,2,\ldots)$, we see that this involves similar functions with various powers, so that such equations must be solved recursively. Start with $m=0$, $n=1$ and we have

$$\frac{\mathrm{d}\mathbb{E}[\lambda(t)]}{\mathrm{d}t}+\beta\mathbb{E}[\lambda(t)]=\beta v+\eta\mu_1,$$

which has solution

$$\mathbb{E}[\lambda(t)]=v+\frac{\eta\mu_1}{\beta}\left(1-e^{-\beta t}\right). \tag{37}$$

Alternatively, this is also easily obtained by taking expectations of Equation (34), because it is easily to obtain the expectation of a function of Poisson random variables.

If we let $m=1$, $n=0$ then Equation (36) becomes

$$\frac{\mathrm{d}\mathbb{E}\left[N_C(t)\right]}{\mathrm{d}t}=\mathbb{E}[\lambda(t)]$$

or, rather obviously,

$$\mathbb{E}\left[N_C(t)\right]=\int_0^t\mathbb{E}[\lambda(u)]\mathrm{d}u=\left(\frac{\eta\mu_1}{\beta}+v\right)t-\frac{\eta\mu_1}{\beta^2}\left(1-e^{-\beta t}\right). \tag{38}$$

The general equation using $m=0$ is

$$\frac{\mathrm{d}\mathbb{E}[\lambda^n(t)]}{\mathrm{d}t}+n\beta\mathbb{E}[\lambda^n(t)]=n\beta v\mathbb{E}[\lambda^{n-1}(t)]+\eta\sum_{i=0}^{n-1}\binom{n}{i}\mu_{n-i}\mathbb{E}[\lambda^i(t)],$$

so that we have a simple recursion for $\mathbb{E}[\lambda^n(t)]$ in terms of lower powers of n. Recursion is not so simple for $m>0$, but we give the first two equations that are obviously soluble in terms of powers that we already know.

$$\begin{cases} \dfrac{\mathrm{d}\mathbb{E}[N_C(t)\lambda(t)]}{\mathrm{d}t} + \beta\mathbb{E}[N_C(t)\lambda(t)] = \beta v\mathbb{E}[N_C(t)] + \eta\mu_1\mathbb{E}[N_C(t)] + \mathbb{E}[\lambda^2(t)], \\ \dfrac{\mathrm{d}\mathbb{E}[N_C^2(t)]}{\mathrm{d}t} = \mathbb{E}[\lambda(t)] + 2\mathbb{E}[N_C(t)\lambda(t)]. \end{cases}$$

4.3. Dynamic contagion processes

Dassios and Zhao [8] introduced a dynamic contagion process, $N_D(t)$, with intensity function

$$\lambda_D(t) = v + (\lambda_0 - v)e^{-\beta t} + \sum_{S_r<t} Y_r e^{-\beta(t-S_r)} + \sum_{T_r<t} Z_r e^{-\beta(t-T_r)}, \tag{39}$$

where $\{S_r\}_{r=1,2,\ldots}$ are event occurring times of a homogenous Poisson process with rate η and $\{Y_r\}_{r=1,2,\ldots}$ are their associated marks, assumed to be i.i.d. random variables with distribution function $G_Y(y)$, $y > 0$. $\{T_r\}_{r=1,2,\ldots}$ are event occurring times of a dynamic contagion process with corresponding marks, $\{Z_r\}_{r=1,2,\ldots}$, assumed to be i.i.d. random variables with distribution function $G_Z(z)$, $z > 0$. It is assumed that the random variables Y_r, Z_r are independent of the history of the process prior to their occurrences. In this case the baseline intensity is not constant.

Note that, apart from changing subscript C into D, and having a time-varying baseline intensity, Equation (39) differs from the first part of Equation (35) simply by adding a self-exciting term to the cross-exciting term that is already there. Together with the expression for $\lambda_P(t)$ in Equation (35) we once again have a bivariate mutually-exciting process $(N_D(t), N_P(t))$ and, because of the exponential form of the exciting terms, the triple $(N_D(t), N_P(t), \lambda_D(t))$ forms a Markov process: in fact, because $N_P(t)$ is Poisson, $(N_D(t), \lambda_D(t))$ is a Markov process. In introducing the dynamic contagion process, [8] described it as a generalization of both the Hawkes process (with exponential decay) and the Cox process with shot noise intensity (with exponential decay). We see that, in fact, it is a special case of a bivariate mutually-exciting Hawkes process.

The Dynamic contagion process was defined in Equation (39). We restate it here in a slightly different form. The intensity is

$$\lambda(t) = v(t) + \sum_{S_r<t} Y_i e^{-\beta(t-S_r)} + \sum_{T_r<t} Z_r e^{-\beta(t-T_r)},$$

where the baseline intensity is the time-varying function $v(t) = v + (\lambda_0 - v)e^{-\beta t}$.

In this model we have two point processes: the Dynamic contagion process $N_D(t)$, which has intensity $\lambda(t)$, and the Poisson process $N_P(t)$, which has constant intensity η.

Similarly, we seek to find a differential equation for $\mathbb{E}\left[N_D^m(t)N_P^k(t)\lambda^n(t)\right]$. The result is as follows.

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_D^m(t)N_P^k(t)\lambda^n(t)] \\
&= n[v'(t) + \beta v(t)]\mathbb{E}[N_D^m(t)N_P^k(t)\lambda^{n-1}(t)] - n\beta\mathbb{E}[N_D^m(t)N_P^k(t)\lambda^n(t)] \\
&\quad + \eta\sum_{r=0}^{k}\binom{k}{r}\sum_{i=0}^{n-1}\binom{n}{i}\mu_{n-i}^{(Y)}\mathbb{E}[N_D^m(t)N_p^r(t)\lambda^i(t)] + \eta\sum_{r=0}^{k-1}\binom{k}{r}\mathbb{E}[N_D^m(t)N_p^r(t)\lambda^n(t)] \\
&\quad + \sum_{j=0}^{m}\binom{m}{j}\sum_{i=0}^{n-1}\binom{n}{i}\mu_{n-i}^{(Z)}\mathbb{E}[N_D^j(t)N_P^k(t)\lambda^{i+1}(t)] + \sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N_D^j(t)N_P^k(t)\lambda^{n+1}(t)],
\end{aligned} \tag{40}$$

where $v'(t) = \dfrac{\mathrm{d}v(t)}{\mathrm{d}t} = -\beta(\lambda_0 - v)e^{-\beta t}$.

Equation (40), of course, is a special case of Equation (33) when $v(t) \equiv v$ is constant. We can consider special cases when some of the parameters m, k, n are zero. We can also consider other cases such as $\mathbb{E}[N_D^m(t)\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)]$, where

$$\lambda_1(t) = \sum_{S_r<t} Y_r e^{-\beta(t-S_r)}; \quad \lambda_2(t) = \sum_{T_r<t} Z_r e^{-\beta(t-T_r)}.$$

With the given baseline intensity the process is Markovian and so Equation (40) can also be derived by Dynkin's formula. However, our proof is valid for more general baseline intensity functions.

4.4. Inhomogenous Poisson processes

Here we simply take $\lambda(t)$ to be a given deterministic function: there are no exciting kernels of the kind we have so far assumed. Using our method, we can get the following differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[g(N(t))] = \{\mathbb{E}[g(N(t)+1)] - \mathbb{E}[g(N(t))]\}\lambda(t), \tag{41}$$

which can be solved analytically for some functions $g : \Re^+ \times \boldsymbol{N} \to \Re$ under some mild conditions.

Example 3. $g(x) = \theta^x$, $\theta > 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\theta^{N(t)}] = \{\mathbb{E}[\theta^{N(t)+1}] - \mathbb{E}[\theta^{N(t)}]\} = (\theta - 1)\mathbb{E}[\theta^{N(t)}]\lambda(t),$$

so that

$$\mathbb{E}[\theta^{N(t)}] = \exp\left[\int_0^t (\theta-1)\lambda(u)\mathrm{d}u\right], \tag{42}$$

which is of course the very well known expression for the p.g.f. of the inhomogeneous Poisson process.

Example 4. $g(x) = x^n$, $n = 1, 2, \ldots$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N^m(t)] = \{\mathbb{E}[(N(t)+1)^m] - \mathbb{E}[N^m(t)]\}\lambda(t) = \lambda(t)\sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N^j(t)],$$

which can be integrated to give

$$\mathbb{E}[N^m(t)] = \sum_{j=1}^{m-1}\binom{m}{j}\int_0^t \lambda(u)\mathbb{E}[N^j(u)]\mathrm{d}u. \tag{43}$$

In fact, we can have, for an inhomogeneous Poisson process $N(t)$ whose intensity function as $\lambda(t)$, and

$$\mathbb{P}\{N(t)=k\} = \frac{[\Lambda(t)]^k}{k!}e^{-\Lambda(t)}, k = 0, 1, \ldots,$$

where $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N^m(t)] &= \frac{\mathrm{d}}{\mathrm{d}t}\{\sum_{k=0}^{\infty}k^m\frac{[\Lambda(t)]^k}{k!}e^{-\Lambda(t)}\} \\
&= \lambda(t)\sum_{k=0}^{\infty}k^m\frac{[\Lambda(t)]^{k-1}}{(k-1)!}e^{-\Lambda(t)} - \lambda(t)\sum_{k=0}^{\infty}k^m\frac{[\Lambda(t)]^k}{k!}e^{-\Lambda(t)} \\
&= \sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N^j(t)\lambda(t)],
\end{aligned}$$

which is the same as Equation (43) with initial condition $\mathbb{E}[N^m(0)] = 0$.

4.5. A multi-marked Hawkes process

We now consider a new type of Hawkes process whose intensity function is given by

$$\lambda(t) = v + \sum_{j=1}^{K}\sum_{T_r<t} Z_{j,r}e^{-\beta_j(t-T_r)}, \tag{44}$$

where T_r is the occurrence time of the rth event of the point process $N_H(t)$; for each $j = 1$ to K, $\{Z_{j,r} \sim G_{Z_j}(z)\}_{r=1,2,\ldots}$ is a sequence of i.i.d. marks; these series are also independent of each other and independent of previous values of the process $N(t)$. The model looks very similar to the standard mutually-exciting process. However, in this case there is only one point process and each event on the right of this equation is operated on by a kernel that consists of K different exponential functions, each multiplied by a different mark sequence, which combine additively to produce the output intensity. Note that the rates β_j are in general different: if they were all the same then this would become a simpler self-exciting process with marks that are just the sums of the individual marks $\sum_{j=1}^{K} Z_{j,r}$ with a distribution that is the convolution of the various $G_{Z_j}(z)$ distributions. Our method can still work on this more general Hawkes process. In the following, we only consider the case of $K = 2$.

Theorem 4. *For a Hawkes process with intensity function*

$$\lambda(t) = v + \sum_{T_r<t} Z_{1,r} e^{-\beta_1(t-T_r)} + \sum_{T_r<t} Z_{2,r} e^{-\beta_2(t-T_r)}, \tag{45}$$

we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_H^m(t)\lambda^n(t)] = &-n\beta_1\mathbb{E}[N_H^m(t)\lambda^{n-1}(t)\lambda_1(t)] - n\beta_2\mathbb{E}[N_H^m(t)\lambda^{n-1}(t)\lambda_2(t)] \\ &+ \sum_{j=0}^{m-1}\tbinom{m}{j}\mathbb{E}[N_H^j(t)\lambda^{n+1}(t)] + \sum_{j=0}^{m}\sum_{i=0}^{n-1}\tbinom{m}{j}\tbinom{n}{i} w_{n-i}\mathbb{E}[N_H^j(t)\lambda^{i+1}(t)],\end{aligned}$$

where for $(k = 1, 2, \ldots;\ r = 1, 2, \ldots)$

$$\lambda_1(t) = \sum_{T_r<t} Z_{1,r} e^{-\beta_1(t-T_r)};\quad \lambda_2(t) = \sum_{T_r<t} Z_{2,r} e^{-\beta_2(t-T_r)};\quad w_k = \int_0^\infty z^k \mathrm{d}G_{Z_{1,r}+Z_{2,r}}(z).$$

We can also treat the intensity components separately

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_H^m(t)\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)] \\ &= -(n_1\beta_1 + n_2\beta_2)\mathbb{E}[N_H^m(t)\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)] + \sum_{j=0}^{m-1}\binom{m}{j}\mathbb{E}[N_H^j(t)\lambda(t)\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)] \\ &\quad + \sum_{j=0}^{m}\sum_{i=0}^{n_1-1}\binom{m}{j}\binom{n_1}{i}\mu_{n_1-i}^{(1)}\mathbb{E}[N_H^j(t)\lambda(t)\lambda_1^{i}(t)\lambda_2^{n_2}(t)] \\ &\quad + \sum_{j=0}^{m}\sum_{i=0}^{n_2-1}\binom{m}{j}\binom{n_2}{i}\mu_{n_2-i}^{(2)}\mathbb{E}[N_H^j(t)\lambda(t)\lambda_1^{n_1}(t)\lambda_2^{i}(t)] \\ &\quad + \sum_{j=0}^{m}\sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1}\binom{m}{j}\binom{n_1}{i_1}\binom{n_2}{i_2}\mu_{n_1-i_1}^{(1)}\mu_{n_2-i_2}^{(2)}\mathbb{E}[N_H^j(t)\lambda(t)\lambda_1^{i_1}(t)\lambda_2^{i_2}(t)].\end{aligned}$$

where $\mu_k^{(i)} = \int_0^\infty z_i^k \mathrm{d}G_{Z_i}(z_i)$, $(i = 1, 2;\ k = 1, 2, \ldots)$.

Proof. The proof is similar to the previous ones. The details are omitted here. □

Some special cases are considered in the following.

When $n_1 = n_2 = 0$, by using Theorem 4, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_H^m(t)] = \sum_{i=0}^{m-1}\binom{m}{i}\mathbb{E}[N_H^i(t)\lambda(t)],$$

which is coincident with Theorem 1 provided for general counting processes. In particular, $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_H(t)] = \mathbb{E}[\lambda(t)] \Rightarrow \mathbb{E}[N_H(t)] = \int_0^t \mathbb{E}[\lambda(u)]\mathrm{d}u$.

When $m = 0$, by using Theorem 4, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)] = &-(n_1\beta_1 + n_2\beta_2)\mathbb{E}[\lambda_1^{n_1}(t)\lambda_2^{n_2}(t)]\\ &+\sum_{i=0}^{n_1-1}\tbinom{n_1}{i}\mu_{n_1-i}^{(1)}\mathbb{E}[\lambda(t)\lambda_1^{i}(t)\lambda_2^{n_2}(t)]\\ &+\sum_{i=0}^{n_2-1}\tbinom{n_2}{i}\mu_{n_2-i}^{(2)}\mathbb{E}[\lambda(t)\lambda_1^{n_1}(t)\lambda_2^{i}(t)]\\ &+\sum_{i_1=0}^{n_1-1}\sum_{i_2=0}^{n_2-1}\tbinom{n_1}{i_1}\tbinom{n_2}{i_2}\mu_{n_1-i_1}^{(1)}\mu_{n_2-i_2}^{(2)}\mathbb{E}[\lambda(t)\lambda_1^{i_1}(t)\lambda_2^{i_2}(t)].\end{aligned} \tag{46}$$

4.6. Non-Markovian Hawkes Processes

In this subsection, a non-Markovian Hawkes process, denoted as $N_{HG}(t)$, is considered by using our method, whose intensity kernel function is $h(t) = te^{-t}$, $(t \geq 0)$ (Gamma decay function), i.e., its intensity function is

$$\lambda(t) = v + \int_0^t h(t-s)\mathrm{d}N_{HG}(t) = v + \sum_{T_i<t} h(t-T_i) = v + \sum_{T_i<t}[(t-T_i)e^{-(t-T_i)}], \tag{47}$$

where $\{T_i\}_{i=1,2,\dots}$ is the series of times at which events occur for the Hawkes process $N_{HG}(t)$.

We shall give the results for the moments of $N_{HG}(t)$ by using our method, then present a closed-form for $\mathbb{E}[N_{HG}^2(t)]$ as an example. Because the higher order moments can also be done similarly, here they are omitted.

It is easy to know that $h'(t) = e^{-t} - te^{-t} = h_2(t) - h(t)$, where $h_2(t) = e^{-t}$, for $t \geq 0$, $h_2(t) = 0$, for $t < 0$.

First we have, when $\{N_{HG}(t+\Delta) - N_{HG}(t) = 0\}$ occurs,

$$\begin{aligned}\lambda_0(t+\Delta) &= v + \sum_{T_i<t+\Delta} h(t-T_i+\Delta) = v + \sum_{T_i<t} h(t-T_i+\Delta)\\ &= v + \sum_{T_i<t} h(t-T_i) + \Delta \sum_{T_i<t} h'(t-T_i) + o(\Delta)\\ &= \lambda(t) + \Delta[\sum_{T_i<t} h_2(t-T_i) - \sum_{T_i<t} h(t-T_i)] + o(\Delta)\\ &= \lambda(t) + \Delta[\lambda_2(t) - \lambda(t) + v] + o(\Delta),\end{aligned}$$

where $\lambda_2(t)$ is a stochastic process, which is defined as

$$\lambda_2(t) := \int_0^t h_2(t-s)\mathrm{d}N_{HG}(t) = \sum_{T_i<t} h_2(t-T_i) = \sum_{T_i<t} e^{-(t-T_i)}. \tag{48}$$

Note that $\lambda_2(t)$ depends on the point process $N_{HG}(t)$ whose the moments are considering now. $(\lambda_2(0) = 0,\ \lambda_2(t) = 0,\ \text{when } t < T_1)$

On the other hand, we have, when $\{N_{HG}(t+\Delta) - N_{HG}(t) = 1\}$ occurs,

$$\begin{aligned}\lambda_1(t+\Delta) &= v + \sum_{T_i<t+\Delta} h(t-T_i+\Delta) = v + \sum_{T_i<t} h(t-T_i+\Delta) + h(t-T^*+\Delta)\\ &= \lambda_0(t+\Delta) + h(\Delta_1) + o(\Delta), \quad 0<\Delta_1<\Delta, \quad t<T^*<t+\Delta.\end{aligned}$$

Using our elementary method presented in Section 2, then we have the following result.

Theorem 5. *For a Hawkes process $N_{HG}(t)$ with intensity function*

$$\lambda(t) = v + \sum_{T_i<t} [(t-T_i)e^{-(t-T_i)}],$$

then we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}^m(t)\lambda^n(t)\lambda_2^l(t)] &= n\mathbb{E}[N_{HG}^m(t)\lambda^{n-1}(t)\lambda_2^{l+1}(t)] - (n+l)\mathbb{E}[N_{HG}^m(t)\lambda^n(t)\lambda_2^l(t)]\\ &\quad + nv\mathbb{E}[N_{HG}^m(t)\lambda^{n-1}(t)\lambda_2^l(t)]\\ &\quad + \sum_{i=0}^{m-1}\binom{m}{i}\mathbb{E}[N_{HG}^i(t)\lambda^{n+1}(t)[\lambda_2(t)+1]^l]\\ &\quad + \sum_{k=0}^{l-1}\binom{l}{k}\mathbb{E}[N_{HG}^m(t)\lambda^{n+1}(t)\lambda_2^k(t)],\end{aligned} \tag{49}$$

where $\sum_0^w [*] = 0$, *if* $w<0$ *and* $\lambda^i(t) = 0$, *if* $i<0$.

Proof. The proof is similar to the previous ones based on some complicated algebraic manipulations, here it is omitted. □

Remark 9. Theorem 5 is presented as an example for our elementary approach on a non-Markovian Hawkes process in which the Dynkin formula cannot work. It is again to prove our method is an extension of Dynkin formula.

Special cases: From Equation (49), we have the following special cases.

(i) When $m = 0$, $n = 1$, $l = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)] = \mathbb{E}[\lambda_2(t)] - \mathbb{E}[\lambda(t)] - v.$$

(ii) When $m = 0$, $n = 0$, $l = 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_2(t)] = \mathbb{E}[\lambda(t)] - \mathbb{E}[\lambda_2(t)].$$

To summarize cases (i) and (ii), we have a set of 2 differential equations as follows.

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)] = \mathbb{E}[\lambda_2(t)] - \mathbb{E}[\lambda(t)] - v, \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_2(t)] = \mathbb{E}[\lambda(t)] - \mathbb{E}[\lambda_2(t)]. \end{cases}$$

Numerical results, with $\mathbb{E}[\lambda_2(0)] = 0$, $\mathbb{E}[\lambda(0)] = v$, and set $v = 2$, are,

$$\begin{cases} \mathbb{E}[\lambda(t)] = \frac{1}{2}e^{-2t} + t + \frac{3}{2}, \\ \mathbb{E}[\lambda_2(t)] = -\frac{1}{2}e^{-2t} + t + \frac{1}{2}. \end{cases}$$

(iii) When $m = 0$, $n = 2$, $l = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^2(t)] = 2\mathbb{E}[\lambda(t)\lambda_2(t)] - 2\mathbb{E}[\lambda^2(t)] + 2v\mathbb{E}[\lambda(t)].$$

(iv) When $m = 0$, $n = 1$, $l = 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)\lambda_2(t)] = \mathbb{E}[\lambda^2(t)] + \mathbb{E}[\lambda_2^2(t)] - 2\mathbb{E}[\lambda(t)\lambda_2(t)] + v\mathbb{E}[\lambda_2(t)].$$

(v) When $m = 0$, $n = 0$, $l = 2$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_2^2(t)] = 2\mathbb{E}[\lambda(t)\lambda_2(t)] - 2\mathbb{E}[\lambda_2^2(t)] + \mathbb{E}[\lambda(t)].$$

To summarize cases (iii), (iv) and (v), we have a set of 3 differential equations as follows.

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda^2(t)] = 2\mathbb{E}[\lambda(t)\lambda_2(t)] - 2\mathbb{E}[\lambda^2(t)] + 2v\mathbb{E}[\lambda(t)], \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda(t)\lambda_2(t)] = \mathbb{E}[\lambda^2(t)] + \mathbb{E}[\lambda_2^2(t)] - 2\mathbb{E}[\lambda(t)\lambda_2(t)] + v\mathbb{E}[\lambda_2(t)], \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_2^2(t)] = 2\mathbb{E}[\lambda(t)\lambda_2(t)] - 2\mathbb{E}[\lambda_2^2(t)] + \mathbb{E}[\lambda(t)]. \end{cases}$$

Numerical results, with $\mathbb{E}[\lambda_2^2(0)] = 0$, $\mathbb{E}[\lambda^2(0)] = v^2 = 4$, $\mathbb{E}[\lambda(0)\lambda_2(0)] = 0$, are,

$$\begin{cases} \mathbb{E}[\lambda^2(t)] = \frac{9}{8}t^2 + \frac{51}{16}t + \frac{137}{16} + \frac{7}{4}e^{-2t} + \frac{3}{4}te^{-2t} + \frac{7}{64}e^{-4t}, \\ \mathbb{E}[\lambda_2^2(t)] = \frac{9}{8}t^2 + \frac{27}{16}t + \frac{41}{64} - \frac{3}{4}e^{-2t} - \frac{3}{4}te^{-2t} + \frac{7}{64}e^{-4t}, \\ \mathbb{E}[\lambda(t)\lambda_2(t)] = \frac{9}{8}t^2 + \frac{37}{16}t + \frac{47}{64} - \frac{5}{8}e^{-2t} - \frac{7}{64}e^{-4t}. \end{cases}$$

(vi) When $m = 1$, $n = 1$, $l = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}(t)\lambda(t)] = E[N_{HG}(t)\lambda_2(t)] - \mathbb{E}[N_{HG}(t)\lambda(t)] + v\mathbb{E}[\lambda(t)] + \mathbb{E}[\lambda^2(t)].$$

(vii) When $m = 1$, $n = 0$, $l = 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}(t)\lambda_2(t)] = \mathbb{E}[N_{HG}(t)\lambda(t)] - \mathbb{E}[N_{HG}(t)\lambda_2(t)] + \mathbb{E}[\lambda(t)\lambda_2(t)] + \mathbb{E}[\lambda(t)].$$

To summarize cases (vi) and (vii), we have a set of 2 differential equations as follows.

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}(t)\lambda(t)] = E[N_{HG}(t)\lambda_2(t)] - \mathbb{E}[N_{HG}(t)\lambda(t)] + v\mathbb{E}[\lambda(t)] + \mathbb{E}[\lambda^2(t)], \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}(t)\lambda_2(t)] = \mathbb{E}[N_{HG}(t)\lambda(t)] - \mathbb{E}[N_{HG}(t)\lambda_2(t)] + \mathbb{E}[\lambda(t)\lambda_2(t)] + \mathbb{E}[\lambda(t)]. \end{cases}$$

Numerical results, with $\mathbb{E}[N_{HG}(0)\lambda(0)] = 0$, $\mathbb{E}[N_{HG}(0)\lambda_2(0)] = 0$, are,

$$\begin{cases} \mathbb{E}[N_{HG}(t)\lambda(t)] = \frac{3}{8}t^3 + \frac{17}{8}t^2 + \frac{943}{128}t + \frac{715}{256} - \frac{7}{256}e^{-2t} + \frac{5}{4}te^{-2t} + \frac{3}{16}t^2e^{-2t} - \frac{7}{128}e^{-4t}, \\ \mathbb{E}[N_{HG}(t)\lambda_2(t)] = \frac{3}{8}t^3 + \frac{17}{8}t^2 + \frac{823}{128}t - \frac{345}{256} + \frac{331}{256}e^{-2t} - \frac{13}{8}te^{-2t} - \frac{3}{16}t^2e^{-2t} + \frac{7}{128}e^{-4t}. \end{cases}$$

Based on Theorem 1, we have

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[N_{HG}^2(t)] = \sum_{j=0}^{1}\binom{2}{j}\mathbb{E}[N_{HG}^j(t)\lambda(t)] \\ &= \mathbb{E}[\lambda(t)] + 2\mathbb{E}[N_{HG}(t)\lambda(t)] \\ &= \frac{1}{2}e^{-2t} + t + \frac{3}{2} + 2\left[\frac{3}{8}t^3 + \frac{17}{8}t^2 + \frac{943}{128}t + \frac{715}{256} - \frac{7}{256}e^{-2t} + \frac{5}{4}te^{-2t} + \frac{3}{16}t^2e^{-2t} - \frac{7}{128}e^{-4t}\right] \\ &= \frac{3}{4}t^3 + \frac{17}{4}t^2 + \frac{999}{56}t + \frac{883}{128} + \frac{49}{128}e^{-2t} + \frac{5}{2}te^{-2t} + \frac{3}{8}t^2e^{-2t} - \frac{7}{56}e^{-4t}. \end{aligned}$$

$$\mathbb{E}[N_{HG}^2(t)] = \frac{-1}{5376}(1008t^2 + 7728t + 4893)e^{-2t} + \frac{3}{16}t^4 + \frac{17}{12}t^3 + \frac{999}{112}t^2 + \frac{883}{128}t + \frac{1}{32}e^{-4t} + \frac{225}{256}.$$

On the other hand, we also have

$$\mathbb{E}[N_{HG}(t)] = \int_0^t \mathbb{E}[\lambda(u)]\mathrm{d}u = \frac{1}{4} - \frac{1}{4}e^{-2t} + \frac{1}{2}t^2 + \frac{3}{2}t.$$

The curves for $\mathbb{E}[N^2_{HG}(t)]$, $\mathbb{E}[N_{HG}(t)]$ and a sample path of $\lambda(t)$ are presented in FIGURES 3 and 4, respectively.

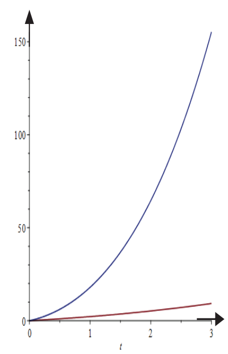

FIGURE 3: The curves of $\mathbb{E}[N^2_{HG}(t)]$ and $\mathbb{E}[N_{HG}(t)]$.

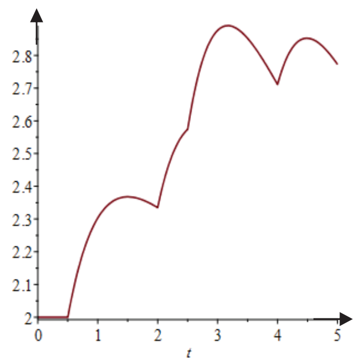

FIGURE 4: A sample path of $\lambda(t)$ with $v = 2$ ($T_i = 0.5, 2, 2.5, 4,$ $(i = 1, ..., 4)$

Remark 10. The key issue for our method to be used successfully is to introduce a stochastic process $\lambda_2(t)$ for this non-Markovian Hawkes process with Gamma decay kernel function. In fact, the process $(N_{HG}(t), \lambda(t), \lambda_2(t))$ has the Markov property. In general, how to introduce some stochastic processes makes the Markov property true for non-Markovian Hawkes processes is a difficult question, which is similar to the method of supplementary variables. When using our method, one does not need to check the Markov property being true or not, if the differential equations can be established successfully, then our method is applicable.

5. Conclusions and Discussion

In the paper an elementary approach is presented to obtain moments of the counting process and/or the intensity of a number of marked Hawkes processes, in which the detailed procedure is given step-by-step, and it works not only for all Markovian Hawkes processes, but also for some non-Markovian Hawkes processes. When the process is Markovian the same results can be obtained by the famous Dynkin formula. Our method starts by finding the expectation of the required moment at time $t + \Delta$ conditional on the history of the process up to time t. Taking a further expectation and finding the limit as $\Delta \to 0$ yields a differential equation for the required moment. It may be possible to solve this explicitly in simple cases, otherwise it may be solved by standard numerical methods for differential equations. Our method does not depend on having a Markov process, so that it is more general for this purpose than using a Dynkin formula.

Results have been found for simple versions of a self-exciting process, mutually-exciting Hawkes processes, Cox processes, dynamic contagion processes and non-homo -geneous Poisson processes. We show that all of these actually belong to the Hawkes process family. Some results are also obtained for a self-exciting process with multiple marks. We also present an application of our elementary method to a non-Markovian Hawkes processes with Gamma decay kernel function, which shows again that our method is an extension of Dynkin formula.

We use exponential exciting-kernel functions. This often results in the process having the Markov property, but the main advantage is the simple form we get, to order $o(\Delta)$, for $\lambda(t+\Delta)$. We also use the Gamma decay kernel function by introducing a new related stochastic process although the marks and exponential parameter are not used, which is for presenting a simple example in non-Markovian case. As mentioned previous, the Marks and parameters do not affect our method, they just bring some complexity but difficulty. The method in introducing a new stochastic process for non-Markovian Hawkes process looks much like the method of supplementary variables in essential. The Gamma decay kernel function in the example makes some difference with exponential kernel function for random intensity functions, the former's intensity function increases gradually, but the latter's intensity has a jump when an event occurs.

An advantage of our method is that it begins by looking at the conditional expectation of the function we are actually trying to derive. For example, to obtain results for $\mathbb{E}[\theta^{N(t)}e^{-s\lambda(t)}]$ we would begin simply with $\mathbb{E}[\theta^{N(t+\Delta)}e^{-s\lambda(t+\Delta)}|\mathcal{F}_t]$ where the filtration $\mathcal{F}_t$ represents the history of the process up to time t. In contrast, for the dynamic contagion process, [8] (Theorem 3.1 in their paper) approach the same problem by beginning with an objective function $g(N, \lambda, t) = e^{c(t)}A^n(t)e^{-B(t)\lambda}$ and then use the infinitesimal generator to obtain equations for the functions $c(t)$, $A(t)$, and $B(t)$. Our approach seems more obvious.

We have given several examples in the paper but the method can be applied to many problems.

Acknowledgements

This work is supported by the NSF of China under grants 71871021 and 71631001. The authors sincerely thank a referee and the editor for their valuable comments and suggestions for improving the paper greatly.

References

[1] Adamopoulos, L. (1975). Some counting and interval properties of the mutually-exciting processes. *J. Appl. Prob.* **12,** 78–86.

[2] Bremaud, P. and Massoulie, L. (2002). Power spectra of general shot noises and Hawkes processes with a random excitation. *Adv. Appl. Prob.* **34,** 205–222.

[3] Chen, J. M., Hawkes, A. G., Scalas, E. and Trinh, M. (2018). Performance of information criteria for selection of Hawkes process models of financial data. *Quant. Finance* **18(2),** 225–235.

[4] Cox, D. R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc. B* **17,** 129–164.

[5] Cui, L. R., Chen, Z. L. and Gao, H. D. (2018). Reliability for systems with self-healing effect under shock models. *Qua. Tech. and Quant. Manag.* **15,** 551–567.

[6] Cui, L. R., Li, Z. P. and Yi, H. (2018). Partial self-exciting point processes and their parameter estimations. *Comm. in Stat.-Simu. and Comput.* online.

[7] Dassios, A. and Jang, J. W. (2003). Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochas.* **7,** 73–95.

[8] Dassios, A. and Zhao, H. B. (2011). A dynamic contagion process. *Adv. Appl. Prob.* **43,** 814–846.

[9] Davis, M. H. A. (1984). Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. B* **46,** 353–388.

[10] Duffie, D., Filipovic, D. and Schachermayer, W. (2003). Affine processes and applications in finance. *Ann. Appl. Prob.* **13,** 984–1053.

[11] Errais, E., Giesecke, K. and Goldberg, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1,** 642–665.

[12] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58,** 83–90.

[13] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc, B* **33(3),** 438–443.

[14] Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. *Chapter in Stochastic Point Processes.* P.A.W. Lewis (ed.), Wiley, New York, 261–271.

[15] Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Prob.* **11,** 493–503.

[16] Li, Z. P., Cui, L. R. and Chen J. H. (2018). Traffic accident modelling via self-exciting point processes. *Reliab. Eng. Syst. Saf.* **180,** 312–320.

[17] Møller, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.* **35,** 614–640.

[18] Mohler, G., Short, M., Brantingham, P., Schoenberg, F. and Tita, G. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106(493),** 100–108.

[19] Oakes, D. (1975). The Markovian self-exciting process. *J. Appl. Prob.* **12,** 69–77.

[20] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83(401),** 9–27.

[21] Veen, A. and Schoenberg, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *J. Amer. Statist. Assoc.* **103,** 614–624.

[22] Zhu, L. (2013). Nonlinear Hawkes processes. Doctoral Thesis, New York University.