Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 147 (2022) 456–480

stochastic processes and their applications

www.elsevier.com/locate/spa

Cluster point processes and Poisson thinning INARMA

Zezhun Chen*, Angelos Dassios

Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

Received 2 March 2021; received in revised form 13 September 2021; accepted 3 February 2022
Available online 10 February 2022

Abstract

In this paper, we consider Poisson thinning Integer-valued time series models, namely integer-valued moving average model (INMA) and Integer-valued Autoregressive Moving Average model (INARMA), and their relationship with cluster point processes, the Cox point process and the dynamic contagion process. We derive the probability generating functionals of INARMA models and compare to that of cluster point processes. The main aim of this paper is to prove that, under a specific parametric setting, INMA and INARMA models are just discrete versions of continuous cluster point processes and hence converge weakly when the length of subintervals goes to zero.

Keywords: Stochastic intensity model; Dynamic contagion process; Integer-valued time series; Poisson thinning

1. Introduction

The Hawkes process, which was first introduced in Hawkes [18,19], is a self-exciting point process such that its intensity depends on the past of the point process itself. Due to its simplicity and flexibility, the Hawkes process can be viewed as a contagion process and applied in different areas, for example seismology in Ogata [33], epidemiology in Kim [26], and sociology in Mohler et al. [32]. It has gained in popularity in recent years. Finance in particular, is a very popular area to apply Hawkes processes, see Bowsher [9], Large [29], Embrechts et al. [17], Bacry et al. [5,6,7,8], Aït-Sahalia et al. [1], and Dassios and Zhao [15,16]. However, in some context such as modelling the credit contagion in Jarrow and Yu [24], the clustering of defaults is consistent with the Hawkes process, but the default intensity could be impacted exogenously by other factors, which means the distribution of cluster centres may not act as a homogeneous Poisson process in the real financial data. In order to address this, Dassios and

* Corresponding author.
E-mail address: z.chen58@lse.ac.uk (Z. Chen).

https://doi.org/10.1016/j.spa.2022.02.002
0304-4149/

Zhao [14] introduced the dynamic contagion process by generalizing the Hawkes process (with exponential decay kernel) and the Cox process with shot noise intensity (exponential decay kernel) used in Dassios and Jang [13], which allows the cluster centres act as a stochastic process.

The standard time series models (AR, MA, ARMA, etc.), on the other hand, are used for sequences of real-valued data. A natural question would be whether we can use time series models for count data. An early contribution has been done by Jacobs and Lewis [21,22,23], who introduced the discrete Autoregressive and Moving average model (DARMA) for stationary discrete time series. However, the correlation structure of DARMA is quite different from the standard time series model. Later, a new model called Integer-valued autoregressive (INAR) time series was defined and examined by McKenzie [30] and Al-Osh and Alzaid [2]. The idea here is to manipulate the operation between coefficients and variables as well as the innovation terms in a way that the values are always integer. The properties of the INAR model are explored by Al-Osh and Alzaid [3], Jin-Guan and Yuan [25], and McKenzie [31]. The Integer-valued Moving Average model (INMA) was introduced and developed by Al-Osh and Alzaid [4], Brännäs and Hall [10], and Brännäs et al. [11]. They apply the similar idea of the INAR model to a standard MA model.

It seems that no one had studied the connection between point processes and integer-valued time series until Kirchner [27], who showed that Hawkes point processes are continuous-time versions of Poisson thinning INAR time series with infinite order and vice versa. The author also mentioned that one can introduce the INARMA model by adding the moving average part into the INAR model and hence make a connection to the dynamic contagion process, which is the main motivation of this paper. Basically, we formally define the INMA model in a similar way to Kirchner and prove that the INMA model with infinite order is actually a discrete version of a Cox point process. We then define the INARMA and prove that it is also a discrete version of the dynamic contagion process, as Kirchner expected.

The paper is organized as follows: Section 2 specifies the terminology and reviews the definitions of three cluster point processes, namely the dynamic contagion process, the Cox process and the Hawkes process, and their probability generating functionals. Section 3 reviews the definition of INAR model, defines the INMA model and INARMA model, and derives their probability generating functionals. Section 4 provides further details on the convergence of probability generating functionals between the INARMA models and the cluster point processes. Section 5 establishes the weak convergence result from the INARMA models to their corresponding cluster point processes. Section 6 verifies the convergence theorem by calculating the joint probability generating functions numerically through simulation. A few concluding remarks are in the final section.

2. Cluster point processes

In this section, we will first define the space we are working on and provide some terminology and notation concerning the integer-valued random measure. Then, we recall the definitions of three cluster point processes, namely the dynamic contagion process, the Cox process and the Hawkes process. Finally, we derive their probability generating functionals by taking advantage of their cluster representation.

2.1. Preliminaries

We will use most of the notation and terminology from Daley and Vere-Jones [12]. Throughout this paper, we work on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $\mathcal{F}$ is the $\sigma-$algebra generating by Ω. A measure μ on the half-line $\mathbb{R}_+$, a complete separable metric space, is boundedly finite if $\mu(A) < \infty$ for every bounded Borel set $A \in \mathcal{B}(\mathbb{R}_+)$. Hence denote $\mathcal{M}^{\#}_{\mathbb{R}_+}$ as the space of all boundedly finite measures and $\mathcal{B}(\mathcal{M}^{\#}_{\mathbb{R}_+})$ as its $\sigma-$algebra.

Definition 2.1. A point process N on the state space $\mathbb{R}_+$ is a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ into $(\mathcal{N}^{\#}_{\mathbb{R}_+}, \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}_+}))$, $N : \Omega \mapsto \mathcal{N}^{\#}_{\mathbb{R}_+}$, such that $N(A)$ is an integer-valued random variable for each bounded $A \in \mathcal{B}(R_+)$. $\mathcal{N}^{\#}_{\mathbb{R}_+}$ is the family of all boundedly finite integer-valued measure $\mu \in \mathcal{M}^{\#}_{\mathbb{R}_+}$

For a point process (random measure) $N \in \mathcal{N}^{\#}_{\mathbb{R}_+}$, they are well-defined only on some bounded area. Consequently, the distribution of a point process is completely determined by the finite dimensional distributions, see Proposition 9.2 II in Daley and Vere-Jones [12]

Definition 2.2. The finite dimensional distributions of a random measure N are the joint distributions for all finite families of bounded Borel sets $A_1, \ldots, A_k$ of $N(A_1), \ldots, N(A_k)$

$$F_k(A_1, \ldots, A_k; x_1, \ldots, x_k) = \mathcal{P}\{N(A_i) \le x_i (i = 1, \ldots, k)\}. \tag{1}$$

Usually, for a non-negative random measure, one would use the Laplace functional to describe the joint distribution of the random measure. As we work on the space $(\mathcal{N}^{\#}_{\mathbb{R}_+}, \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}_+}))$, there are advantages in moving from the Laplace functional to the probability generating functional (p.g.fl)

Definition 2.3. The probability generating functional (p.g.fl) of a point process N on the complete separable metric space $\mathbb{R}_+$ is defined by

$$G[h] = \mathbb{E}\left[\exp\left\{\int_{\mathbb{R}_+} \log h(x) N(dx)\right\}\right], \quad h \in \mathcal{V}(\mathbb{R}_+), \tag{2}$$

where $\mathcal{V}(\mathbb{R}_+)$ is the class of all real-valued Borel functions h defined on $\mathbb{R}_+$ with $1-h$ vanishing outside some bounded set and satisfying $0 \le h(x) \le 1, \forall x \in \mathbb{R}_+$. Later, we will use $\mathcal{V}_0(\mathbb{R}_+)$, the subset of $\mathcal{V}(\mathbb{R}_+)$ satisfying $\inf_{x \in \mathbb{R}_+} h(x) > 0$

One can always use $G[h]$ to describe F_k by setting $h(x) = h_i, x \in A_i$, where h_i is a constant. Then the $G[h]$ will reduce to the joint probability generating function (joint p.g.f).

$$\begin{aligned} G[h] &= \mathbb{E}\left[\exp\left\{\int_{\mathbb{R}_+} \log h(x) N(dx)\right\}\right] \\ &= \mathbb{E}\left[\exp\left(\int_{\cup_{i=1,\ldots,k} A_i} \log h(x) N(dx)\right)\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{k} h_i^{N(A_i)}\right] \end{aligned}$$

In other words, the p.g.fl $G[h]$ is the limit version of the joint p.g.f where the set A_i has the length $dx \to 0$ and $k \to \infty$. When describing the finite dimensional distributions F_k, the p.g.fl and the joint p.g.f are therefore equivalent. For convenience, we will also use the term 'p.g.fl' for those INARMA models to describe their joint p.g.f in Section 3.

2.2. The dynamic contagion process

We first define a generalized version of the dynamic contagion process as in [14]

Definition 2.4. The generalized dynamic contagion process is a cluster point process $N^{(DCP)}$, with stochastic intensity $\lambda^{(DCP)}$ such that

$$\lambda_t^{(DCP)} = \sum_{i:c_i<t}^{N_t^*} \Upsilon_i f(t-c_i) + \sum_{i:\tau_i<t}^{N_t^{(DCP)}} \chi_i \eta(t-\tau_i), \tag{3}$$

where

- $N_t^* \equiv \{c_i\}_{i=1,2,\ldots}$ are the arrival times of the Poisson process with the constant rate $\rho > 0$
- $N_t^{(DCP)} \equiv \{\tau_i\}_{i=1,2,\ldots}$ are the arrival times of the generalized dynamic contagion process
- $\{\Upsilon_i\}$ are i.i.d externally excited jump sizes, realized at times $\{c_i\}$, with distribution $H(x)$, mean μ_Υ and Laplace transform $\hat{h}(u)$
- $\{\chi_i\}$ are i.i.d self-exciting jump sizes, realized at times $\{\tau_i\}$, with distribution $G(y)$, mean μ_χ and Laplace transform $\hat{g}(u)$. They are independent of $\{\Upsilon_i\}$
- $f(u)$ is an Riemann integrable function for any bounded interval in $\mathbb{R}_+$
- $\eta(u)$ is another Riemann integrable function for any bounded interval in $\mathbb{R}_+$

Note that the stationary condition for this point process would be $\int_0^\infty f(u)du < \infty$ and $\mu_\chi \int_0^\infty \eta(u)du < 1$. Following from this definition, we define the other two cluster point processes — the Cox process and the Hawkes process as special cases.

Definition 2.5. The (Marked) Cox process with shot-noise intensity, also called doubly stochastic process, is a cluster point process $N^{(C)}$ with stochastic intensity $\lambda^{(C)}$ such that

$$\lambda_t^{(C)} = \sum_{i:c_i<t}^{N_t^*} \Upsilon_i f(t-c_i). \tag{4}$$

It is clear that this is a special case of the dynamic contagion process by letting $\eta(u) = 0, \forall u \in \mathbb{R}_+$. On a bounded area $[0, T]$ where $T > 0$, the process can be considered as a cluster process in which the cluster centres c_i arrive as a homogeneous Poisson process $N^* \sim Pois(\rho)$. Conditional on the arrival of c_i, we then have a cluster whose size follows $N_t^1 \sim Pois(\Upsilon_i f(t-c_i))$ with $c_i \le t \le T$. These clusters are mutually independent and cluster centres are not included in $N^{(C)}$. In order words, the arrivals of cluster centres are indicators that some events will happen around them.

Definition 2.6. The (Marked) Hawkes process is a self-exciting point process $N^{(H)}$ with stochastic intensity $\lambda^{(H)}$ such that

$$\lambda_t^{(H)} = \nu + \sum_{i:\tau_i<t}^{N_t^{(H)}} \chi_i \eta(t-\tau_i), \tag{5}$$

where ν is a positive constant.

Similarly, this is another special case of the dynamic contagion process by replacing the ‘Cox component’ in $\lambda_t^{(DCP)}$ by a positive constant ν. From Hawkes and Oakes [20], the Hawkes process can also be interpreted as a cluster point process. The immigrants (cluster centres) arrive as a homogeneous Poisson process $Pois(\nu)$. Each immigrant generates a Galton–Watson type branching process with expected branching ratio $\mu_\chi \int_0^\infty \eta(u)du$. A cluster is then formed by including all the generations (include the immigrant) from the branching process.

Back to the dynamic contagion process, it is actually a Hawkes process with immigrants arriving as a Cox process rather than a homogeneous Poisson process. Here are the probability generating functionals for these cluster point processes.

Proposition 2.1. *Let $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$ such that $1 - z(.)$ vanishes outside $[0, T]$, where $T > 0$. The probability generating functional (p.g.fl) of the Cox process $N^{(C)}$ on $[0, T]$ is given by*

$$\begin{aligned} G^{(C)}(z(.)) &= \exp\left\{\rho \int_0^T (F^{(C)}(z(.)|c) - 1)dc\right\} \\ F^{(C)}(z(.)|c) &= \hat{h}\left(-\int_0^{T-c} f(u)(z(c+u) - 1)du\right). \end{aligned} \tag{6}$$

Proof. See Appendix A.1. □

Proposition 2.2. *Let $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$ such that $1 - z(.)$ vanishes outside $[0, T]$, where $T > 0$. The probability generating functional (p.g.fl) of the generalized dynamic contagion process $N^{(DCP)}$ on $[0, T]$ is given by*

$$\begin{aligned} G^{(DCP)}(z(.)) &= \exp\left\{\rho \int_0^T \left(\hat{h}\left(-\int_0^{T-u} (F^{(H)}(z(.)|u+v) - 1)f(v)dv\right) - 1\right) du\right\} \\ F^{(H)}(z(.)|u) &= z(u)\hat{g}\left(-\int_0^{T-u} (F^{(H)}(z(.)|u+v) - 1)\eta(v)dv\right), \end{aligned} \tag{7}$$

where $F^{(H)}(z(.)|u)$ is the p.g.fl of a cluster generated by an immigrant (cluster centre) arriving at time u, and including that immigrant. While $F(z(.)|u) = F(z_u(.))$ and $z_u(.) = z(u + .)$ is simply the translation of $z()$.

Proof. See Appendix A.2 □

Corollary 2.1. *Let $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$ such that $1 - z(.)$ vanishes outside $[0, T]$, where $T > 0$. The probability generating functional (p.g.fl) of the Hawkes process $N^{(H)}$ on $[0, T]$ is given by*

$$\begin{aligned} G^{(H)}(z(.)) &= \exp\left\{\nu \int_0^T \left(F^{(H)}(z(.)|u) - 1\right) du\right\} \\ F^{(H)}(z(.)|u) &= z(u)\hat{g}\left(-\int_0^{T-u} (F(z(.)|u+v) - 1)\eta(v)dv\right). \end{aligned} \tag{8}$$

Proof. This result generally follows from Theorem 2 in [20]. We can also derive it from Proposition 2.2 by simply letting $\lambda_u^{(C)} = \nu$ in Eq. (34). □

3. Poisson thinning Integer-valued time series model

In this section, we will review the Poisson thinning INAR model from Kirchner [27]. Then we will define the INMA and INARMA models in a similar way to the INAR model, and derive their probability generating functionals. To conform with the preliminaries of point process, the integer-valued models will be defined on the positive state space $\mathbb{R}_+$.

3.1. Integer-valued Autoregressive model - INAR(∞)

The following results (Definition 3.1, Propositions 3.1, 3.2) are mainly from Kirchner [27]. However, the INAR model in our case is defined on the positive state space $\mathbb{R}_+$ and so the results are slightly different from [27].

Definition 3.1. The INAR(∞) is defined as

$$\begin{aligned} X_n &= \sum_{k=1}^{\infty} \alpha_k \circ X_{n-k} + \varepsilon_n \\ &= \alpha_1 \circ X_{n-1} + \cdots + \alpha_{n-1} \circ X_1 + \varepsilon_n \end{aligned} \tag{9}$$

where

- $\{X_k\}_{k=\dots,-2,-1,0} \equiv 0$ as the process is defined on positive state space $\mathbb{R}_+$
- $\alpha_k \geq 0$ (reproduction coefficients)
- $\varepsilon_n \overset{i.i.d}{\sim} Pois(\alpha_0)$, with $\alpha_0 > 0$ (immigration parameter)
- The thinning operator $\circ$ is defined as

$$\alpha_k \circ X_{n-k} = \sum_{i=1}^{X_{n-k}} \epsilon_i^{(n,k)} \quad \epsilon_i^{(n,k)} \overset{i.i.d}{\sim} Pois(\alpha_k),$$

where $\epsilon_l^{(n,k)}$ are independent over $n \in \mathbb{Z},\ k \in \mathbb{N},\ i \in \mathbb{N}$

Note that the stationary condition for this model is $\sum_{k=1}^{\infty} \alpha_k < 1$. In the early study of the integer-valued time series models, the operator $\circ$ is defined as a binomial thinning operator, which means ϵ_i are Bernoulli random variables. However, Kirchner defines it as a Poisson operator, which will lead to the simpler formulas of probability generating functional. In addition, the p.g.fl derived later can be compared directly to that of the Hawkes process. The following proposition gives the branching representation of the INAR model.

Proposition 3.1. *The INAR(∞) process X_n has the following representation*

$$X_n \overset{d}{=} \sum_{i\in\mathbb{Z}} \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)}, \tag{10}$$

where $F_{n-i}^{(i,j)}$ are independent over i, j and they are the copies of a branching process F_n which is defined by

$$F_n = \sum_{g=0}^{\infty} G_n^{(g)}, \quad n \in \mathbb{Z}. \tag{11}$$

The generation G_n are constructed recursively by

$$G_n^{(0)} = 1_{\{n=0\}} \quad G_n^{(g)} = \sum_{k=1}^{n} \alpha_k \circ G_{n-k}^{(g-1)} = \sum_{k=1}^{n} \sum_{m=1}^{G_{n-k}^{(g-1)}} \epsilon_m^{(n,k,g)}, \quad n \in \mathbb{Z},\ g \in \mathbb{N}, \tag{12}$$

with $\xi_m^{(n,k,g)}$ are independent over n, k, g, m and also independent of $\varepsilon_i,\ i \in \mathbb{Z}$. Furthermore, we have the following distributional equality for the generic family-process (F_n)

$$(F_n)_{n\in\mathbb{Z}} \overset{d}{=} \left(1_{\{n=0\}} + \sum_{i=1}^{n} \sum_{j=1}^{G_i^{(1)}} F_{n-1}^{(i,j)} \right). \tag{13}$$

Proposition 3.2. *Let $z_{.} = \{z_i\}_{i=1,\ldots,n}$ be a sequence of constants such that $0 < z_i \leq 1$. The probability generating functional (p.g.fl) of the INAR sequence $\{X_t\}_{t=1,\ldots,n}$ is given by*

$$\begin{aligned} G^{(X_n)}(z_{.}) &= \exp\left\{ \sum_{i=1}^{n} \alpha_0 (F(z_{.}|i) - 1) \right\} \\ F^{(X_n)}(z_{.}|i) &= z_i \exp\left\{ \sum_{k=1}^{n-i} \alpha_k \left(F(z_{.}|i+k) - 1 \right) \right\}, \end{aligned} \tag{14}$$

where $F^{(X_n)}(z_{.}|t) = F^{(X_n)}(z_{t+.})$ is the p.g.fl of the cluster generated by an immigrant (cluster centre) arriving at time t.

Proof. The (discrete) p.g.fl is given by

$$\begin{aligned} G^{(X_n)}(z_{.}) = \mathbb{E}[\prod_{t=1}^{n} z_t^{X_t}] &= \mathbb{E}\exp\left\{ \sum_{t=1}^{n} \log z_t \sum_{i=1}^{t} \sum_{j=1}^{\varepsilon_i} F_{t-i}^{(i,j)} \right\} \\ &= \mathbb{E}\exp\left\{ \sum_{i=1}^{n} \sum_{j=1}^{\varepsilon_i} \sum_{t=i}^{n} \log z_t F_{t-i}^{(i,j)} \right\}. \end{aligned}$$

The sum $\sum_{t=i}^{n} F_{t-i}^{(i,j)}$ can be interpreted as the cluster, which includes all the generation from time i to time n, generated by one of the immigrants in ε_i. Conditionally on the immigration sequence ε_i and exploiting its independence from the family process $F_n^{(i,j)}$, we have

$$\begin{aligned} G^{(X_n)}(z_{.}) &= \prod_{i=1}^{n} \mathbb{E}\left[\prod_{j=1}^{\varepsilon_i} \mathbb{E}\exp\left\{ \sum_{t=i}^{n} \log z_t F_{t-i}^{(i,j)} \right\} \right] \\ &= \prod_{i=1}^{n} \mathbb{E}\left[F^{(X_n)}(z_{.}|i)^{\varepsilon_i} \right] \\ &= \exp\left\{ \sum_{i=1}^{n} \alpha_0 (F^{(X_n)}(z_{.}|i) - 1) \right\}, \end{aligned}$$

where the p.g.fl of the cluster $F^{(X_n)}(z_{.}|i)$ satisfies the following recursive equation

$$\begin{aligned} F^{(X_n)}(z_{.}|i) &= \mathbb{E}\exp\left\{ \sum_{t=i}^{n} \log z_t F_{t-i} \right\} \\ &= \mathbb{E}\exp\left\{ \sum_{t=0}^{n-i} \log z_{i+t} \left(1_{\{t=0\}} + \sum_{k=1}^{t} \sum_{j=1}^{G_k^{(1)}} F_{t-k}^{(k,j)} \right) \right\} \end{aligned}$$

$$
\begin{aligned}
&= z_i \prod_{k=1}^{n-i} \mathbb{E}\left[\prod_{j=1}^{G_k^{(1)}} \mathbb{E}\exp\left\{\sum_{t=k}^{n-i} \log z_{i+t} F_{t-k}^{(k,j)}\right\}\right] \\
&= z_i \exp\left\{\sum_{k=1}^{n-i} \alpha_k \left(F^{(X_n)}(z_.|i+k) - 1\right)\right\}. \quad \square
\end{aligned}
$$

Since the sequence $\{X_t\}_{t=1,\dots,n}$ takes only integer values, if we fix a bounded area $[0, T]$ and let X_t count the number of points for the equal-length area $((t-1)\Delta, t\Delta]$ where $\Delta = \frac{T}{n}$, the p.g.fl of $\{X_t\}_{t=1,\dots,n}$ will look like the discrete version of the p.g.fl of the Hawkes process.

Proposition 3.3. *Consider the following parametric setting.*

- *Fix the bounded area* $[0, T]$, $T < \infty$
- *Choose* $n > 0$, *the number of all subintervals over* $[0, T]$
- *Set the length of subintervals* $\Delta = \frac{T}{n}$, *the immigrant parameter* $\alpha_0 = \nu\Delta$ *and the reproduction coefficient* $\alpha_k = \chi_i \eta(k\Delta)\Delta, k > 0$
- χ_i *are i.i.d random variables corresponding to the cluster centre* $X_{i\Delta}$ *arriving at* $i\Delta$, *with Laplace transform* $\hat{g}(u) = \mathbb{E}[e^{-u\chi_i}]$
- *Let* $z_i = z(i\Delta)$, *where* $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$

Then the probability generating functional of $\{X_t\}_{t=1,\dots,n}$ *becomes*

$$
\begin{aligned}
&G_{(\Delta)}^{(X_n)}(z(.)) = \exp\left\{\nu \sum_{i=1}^{n} (F(z(.)|i\Delta) - 1)\Delta\right\} \\
&F^{(X_n)}(z(.)|i\Delta) = z(i\Delta)\hat{g}\left(-\sum_{k=1}^{n-i} (F(z(.)|(i+k)\Delta) - 1)\eta(k\Delta)\Delta\right).
\end{aligned}
\tag{15}
$$

Proof. By substituting $\alpha_k = \chi_i \eta(k\Delta)\Delta, k > 0$ into Proposition 3.2, the p.g.fl of the cluster $F^{(X_n)}(z_.|i) = F^{(X_n)}(z(.)|i\Delta)$ becomes

$$
\begin{aligned}
F^{(X_n)}(z(.)|i\Delta) &= z(i\Delta) \prod_{k=1}^{n-i} \mathbb{E}\left[\prod_{j=1}^{G_k^{(1)}} \mathbb{E}\exp\left\{\sum_{t=k}^{n-i} \log z((i+t)\Delta) F_{t-k}^{(k,j)}\right\}\right] \\
&= z(i\Delta)\mathbb{E}\left[\exp\left\{\sum_{k=1}^{n-i} \chi_i \eta(k\Delta)\Delta \left(F^{(X_n)}(z(.)|(i+k)\Delta) - 1\right)\right\}\right] \\
&= z(i\Delta)\hat{g}\left(-\sum_{k=1}^{n-i} (F^{(X_n)}(z(.)|(i+k)\Delta) - 1)\eta(k\Delta)\Delta\right).
\end{aligned}
$$

By substituting $\alpha_0 = \nu\Delta$, the whole p.g.fl of the INAR sequence $\{X_t\}_{t=1,\dots,n}$ becomes

$$
\begin{aligned}
G_{(\Delta)}^{(X_n)}(z(.)) &= \exp\left\{\sum_{i=1}^{n} \alpha_0 (F^{(X_n)}(z_.|i) - 1)\right\} \\
&= \exp\left\{\nu \sum_{i=1}^{n} (F^{(X_n)}(z(.)|i\Delta) - 1)\Delta\right\}. \quad \square
\end{aligned}
$$

3.2. Integer-valued Moving Average model

Definition 3.2. The stationary Poisson thinning INMA(∞) model is defined as

$$\begin{aligned} Y_n &= \sum_{k=0}^{\infty} \beta_k \circ \xi_{n-k} \\ &= \beta_0 \circ \xi_n + \beta_1 \circ \xi_{n-1} + \cdots + \beta_{n-1} \circ \xi_1, \end{aligned} \tag{16}$$

where

- $\beta_k \geq 0$ are some non-negative coefficients
- ξ_k are i.i.d and follow $Pois(\mu)$ with $\mu > 0$
- $\{\xi_k\}_{k=\ldots,-2,-1,0} \equiv 0$ as the process is defined on positive state space $\mathbb{R}_+$
- The thinning operator $\circ$ is defined as

$$\beta_k \circ \xi_{n-k} = \sum_{i=1}^{\xi_{n-k}} u_i^{(n,k)}, \quad u_i^{(n,k)} \overset{i.i.d}{\sim} Pois(\beta_k),$$

Note that the stationary condition for this model is $\sum_{k=0}^{\infty} \beta_k < \infty$. The parameters β_k and μ have a similar interpretation to those in the INAR model. β_k are reproduction coefficients while μ is the arrival intensity of cluster centre. From this model point of view, we can regard ξ_n as the cluster centres. They enter the system starting at time n and trigger other events at each time period ($n \to \sum_{i=1}^{\xi_n} u_i^{(n,0)}, n+1 \to \sum_{i=1}^{\xi_n} u_i^{(n+1,1)}, \ldots$). Y_n is then a counting variable to report the total number of the triggered events from $\xi_n, \xi_{n-1}, \ldots, \xi_1$ over the current time period n. Here are two assumptions we need before proceeding to its probability generating functional.

Assumption 1. $u_i^{(n,k)}$ are mutually independent of each other for $n \in \mathbb{N}$, $k \in \mathbb{N}$, $i \in \mathbb{N}$.

This assumption is made to simply the calculation of probability generating functional and that be compared to the point process counterpart much easier. This means that the number of events $u_i^{(t,k)}$, triggered by one of the cluster centre in ξ_{t-k} and counted by the system Y_t, will not affect the number of events $u_i^{(t+j,k+j)}$, triggered by the same cluster centre and counted by the system Y_{t+j} of any future time $j > 0$.

Proposition 3.4. *Let $z_. = \{z_i\}_{i=1,\ldots,n}$ be a sequence of constants such that $0 < z_i \leq 1$. The probability generating functional (p.g.fl) of the INMA sequence $\{Y_t\}_{t=1,\ldots,n}$ is given by*

$$\begin{aligned} G^{(Y_n)}(z_.) &= \exp\left\{\mu \sum_{t=1}^{n} (F^{(Y_n)}(z_.|t) - 1)\right\} \\ F^{(Y_n)}(z_.|t) &= \exp\left\{\sum_{k=1}^{n-t+1} \beta_{k-1}(z_{t+k-1} - 1)\right\}. \end{aligned} \tag{17}$$

Proof. The aggregated process $S_n = \sum_{t=1}^{n} Y_t$ is actually a cluster point process such that

$$\begin{aligned} S_n &= \sum_{t=1}^{n} Y_t = \sum_{t=1}^{n} \sum_{k=0}^{t-1} \beta_k \circ \xi_{t-k} \\ &= \sum_{t=1}^{n} \sum_{k=0}^{n-t} \beta_k \circ \xi_t \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \sum_{i=1}^{\xi_t} (u_i^{(t,0)} + u_i^{(t+1,1)} + \cdots + u_i^{(n,n-t)}), \quad u_i^{(t,k)} \sim Pois(\beta_k) \\
&\stackrel{d}{=} \sum_{t=1}^{n} \sum_{i=1}^{\xi_t} u_i^t, \quad u_i^t \sim Pois(\sum_{k=0}^{n-t} \beta_k).
\end{aligned} \tag{18}
$$

The last equality follows from the independence of the Poisson random variables. It is now clear that the aggregated process S_n is a cluster process such that

- ξ_t generates the cluster centres independently.
- u_i^t is a cluster generated by one of the cluster centre from ξ_t, with the size of cluster (exclude the cluster centre) following $Pois(\sum_{k=0}^{n-t} \beta_k)$

The (discrete) p.g.fl of Y_t is defined as

$$
G(z_.) = \mathbb{E}[\prod_{j=0}^{n} z_j^{Y_j}].
$$

Now we can derive the p.g.fl of this process by following the similar argument in Proposition 2.1. Conditionally on the arrivals of cluster centres generated by ξ_t, the p.g.fl of cluster u_i^t is

$$
F^{(Y_n)}(z_.|t) = \mathbb{E}[\prod_{k=0}^{n-t} z_{t+k}^{u_i^{(t+k,k)}}] = \exp\left\{\sum_{k=0}^{n-t} \beta_k (z_{t+k} - 1)\right\}.
$$

The cluster centres generated by ξ_t are mutually independent. Then the p.g.fl of $\sum_{i=1}^{\xi_t} u_i^t$ is

$$
G_t(z_.) = \mathbb{E}[F^{(Y_n)}(z_.|t)^{\xi_t}] = \exp\left\{\mu(F^{(Y_n)}(z_.|t) - 1)\right\}.
$$

Clusters centres generated by $\{\xi_t\}_{t=1,\ldots,n}$ are also mutually independent. Finally the p.g.fl of Y_t is

$$
G^{(Y_n)}(z_.) = \prod_{t=1}^{n} G_t(z) = \exp\left\{\mu \sum_{t=1}^{n} (F^{(Y_n)}(z_.|t) - 1)\right\}. \quad \square
$$

Similar to the INAR model, due to the integer-valued nature of the INMA model, if we fix an bounded area $[0, T]$ and let Y_t count the number of points for the equal-length area $((t-1)\Delta, t\Delta]$ with $\Delta = \frac{T}{n}$, the p.g.fl of the INMA sequence $\{Y_t\}_{t=1,\ldots,n}$ will look like the discrete version of the p.g.fl of the Cox process under some specific parametric setting.

Proposition 3.5. *Consider the following parametric setting.*

- *Fix the bounded area* $[0, T]$, $T < \infty$
- *Choose* $n > 0$, *the number of all subintervals over* $[0, T]$
- *Set the length of subintervals* $\Delta = \frac{T}{n}$, $\mu = \rho\Delta$ *and* $\beta_k = \Upsilon_t f(k\Delta)\Delta, k \geq 0$
- Υ_t *are i.i.d random variables corresponding to the cluster centre* $\xi_{t\Delta}$ *arriving at* $t\Delta$, *with the Laplace transform* $\hat{h}(u) = \mathbb{E}[e^{-u\Upsilon_t}]$
- *Let* $z_k = z(k\Delta)$ *where* $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$

Then the probability generating functional of the sequence $\{Y_t\}_{t=1,\ldots,n}$ *becomes*

$$G^{(Y_n)}_{(\Delta)}(z(.)) = \exp\left\{\rho \sum_{t=1}^{n} (F^{(Y_n)}(z(.)|t\Delta) - 1)\Delta\right\}$$
$$F^{(Y_n)}(z(.)|t\Delta) = \hat{h}\left(-\sum_{k=1}^{n-t+1} f(k\Delta)(z((t+k-1)\Delta) - 1)\Delta\right). \tag{19}$$

Proof. By substituting $\beta_k = \Upsilon_t f(k\Delta)\Delta, k \geq 0$ into Proposition 3.4, the p.g.fl of the cluster part $F^{(Y_n)}(z_.|t) = F^{(Y_n)}(z(.)|t\Delta)$ becomes

$$\begin{aligned} F^{(Y_n)}(z(.)|t\Delta) &= \mathbb{E}[\prod_{k=0}^{n-t} z_{t+k}^{u_i^{(t+k,k)}}] \\ &= \mathbb{E}\left[\exp\left\{\sum_{k=0}^{n-t} \Upsilon_t f(k\Delta)\Delta(z_{t+k} - 1)\right\}\right] \\ &= \hat{h}\left(-\sum_{k=1}^{n-t+1} f(k\Delta t)(z((t+k-1)\Delta) - 1)\Delta\right). \end{aligned}$$

Then substituting $\mu = \rho\Delta$, the p.g.fl of the INMA sequence $\{Y_t\}_{t=1,\ldots,n}$ becomes

$$\begin{aligned} G^{(Y_n)}_{(\Delta)}(z(.)) &= \exp\left\{\mu \sum_{t=1}^{n} (F^{(Y_n)}(z(.)|t\Delta) - 1)\right\} \\ &= \exp\left\{\rho \sum_{t=1}^{n} (F^{(Y_n)}(z(.)|t\Delta) - 1)\Delta\right\}. \quad \square \end{aligned}$$

3.3. Integer-valued Autoregressive Moving Average model

Definition 3.3. The stationary Poisson thinning INARMA(∞, ∞) model is defined as

$$\begin{aligned} Z_n &= \sum_{k=1}^{\infty} \alpha_k \circ Z_{n-k} + Y_n \\ &= \sum_{k=1}^{\infty} \alpha_k \circ Z_{n-k} + \sum_{j=0}^{\infty} \beta_j \circ \xi_{n-j} \\ &= \alpha_1 \circ Z_{n-1} + \cdots + \alpha_{n-1} \circ Z_1 + \beta_0 \circ \xi_n + \cdots + \beta_{n-1} \circ \xi_1, \end{aligned} \tag{20}$$

where

- ξ_j are i.i.d and follow $Pois(\mu)$ with $\mu > 0$
- $\{\xi_j\}_{j=\ldots,-2,-1,0} \equiv 0$ and $\{Z_k\}_{k=\ldots,-2,-1,0} \equiv 0$ as the process is defined on $\mathbb{R}_+$
- α_i and β_i are positive coefficients
- The thinning operator $\circ$ is defined as

$$\alpha_k \circ Z_{n-k} = \sum_{i=1}^{Z_{n-k}} \epsilon_i^{(n,k)} \quad \epsilon_i^{(n,k)} \overset{i.i.d}{\sim} Pois(\alpha_k)$$

$$\beta_k \circ \xi_{n-k} = \sum_{i=1}^{\xi_{n-k}} u_i^{(n,k)}, \quad u_i^{(n,k)} \overset{i.i.d}{\sim} Pois(\beta_k).$$

- $\epsilon_i^{(n,k)}$ and $u_i^{(n,k)}$ are mutually independent over $n \in \mathbb{N}, i \in \mathbb{N}, \ k \in \mathbb{N}$.

Note that the stationary condition for this process is $\sum_{i=1}^{\infty} \alpha_i < 1$, $\sum_{i=0}^{\infty} \beta_i < \infty$. The INARMA model simply combines the INAR components and the INMA components from previous sections. It is a generalized INAR model whose immigrants process ε_n is replaced by the INMA model Y_n.

Proposition 3.6. *Let $z_{\cdot} = \{z_i\}_{i=1,\dots,n}$ be a sequence of constants such that $0 < z_i \le 1$. The probability generating functional (p.g.fl) of the INARMA sequence $\{Z_t\}_{t=1,\dots,n}$ is given by*

$$\begin{aligned} G^{(Z_n)}(z_{\cdot}) &= \exp\left\{\mu \sum_{i=1}^{n}\left(\exp\left\{\sum_{k=0}^{n-i} \beta_k (F^{(X_n)}(z_{\cdot}|i+k)-1)\right\}-1\right)\right\} \\ F^{(X_n)}(z_{\cdot}|i) &= z_i \exp\left\{\sum_{k=1}^{n-i} \alpha_k \left(F^{(X_n)}(z_{\cdot}|i+k)-1\right)\right\}, \end{aligned} \tag{21}$$

where $F^{(X_n)}(z_{\cdot}|t)$ is the p.g.fl of the cluster generated by an immigrant (cluster centre) arriving at time t, and including that immigrant. While $F^{(X_n)}(z_{\cdot}|t) = F^{(X_n)}(z_{t+\cdot})$ is simply the translation of $z_{\cdot}$.

Proof. The $F^{(X_n)}(z_{\cdot}|i)$ is exactly the same as the one in INAR model, because this is the cluster generated by the autoregressive structure in the INARMA model and it is irrelevant to Y_i. Hence we can apply the result directly from Proposition 3.2

$$F^{(X_n)}(z_{\cdot}|i) = z_i \exp\left\{\sum_{k=1}^{n-i} \alpha_k \left(F^{(X_n)}(z_{\cdot}|i+k)-1\right)\right\}.$$

Then we can apply a similar argument to the INAR model such that

$$\begin{aligned} G^{(Z_n)}(z_{\cdot}) &= \prod_{i=1}^{n} \mathbb{E}\left[\prod_{j=1}^{Y_i} \mathbb{E}\exp\left\{\sum_{t=i}^{n} \log z_t F_{t-i}^{(i,j)}\right\}\right] \\ &= \prod_{i=1}^{n} \mathbb{E}\left[F^{(X_n)}(z_{\cdot}|i)^{Y_i}\right]. \end{aligned}$$

Now apply the p.g.fl of the INMA model from Proposition 3.4

$$\begin{aligned} G^{(Z_n)}(z_{\cdot}) &= \prod_{i=1}^{n} \mathbb{E}\left[F^{(X_n)}(z_{\cdot}|i)^{Y_i}\right] \\ &= \exp\left\{\mu \sum_{i=1}^{n}\left(\exp\left\{\sum_{k=0}^{n-i} \beta_k (F^{(X_n)}(z_{\cdot}|i+k)-1)\right\}-1\right)\right\}. \quad \square \end{aligned}$$

Similar to the INAR model, due to the integer-valued nature of INARMA model, if we fix a bounded area $[0, T]$ and let Z_t count the number of points for the equal-length area $((t-1)\Delta, t\Delta]$ with $\Delta = \frac{T}{n}$, the p.g.fl of the INARMA sequence $\{Z_t\}_{t=1,\dots,n}$ will look like the discrete version of the p.g.fl of the generalized dynamic contagion process.

Theorem 3.1. *Consider the following parametric setting.*

- *Fixed the terminal time T, $T < \infty$*
- *Choose $n > 0$, the number of all subintervals over $[0, T]$*
- *Set the length of subintervals $\Delta = \frac{T}{n}$, the parameters of INAR part $\alpha_k = \chi_i \eta(k\Delta)\Delta$, $k > 0$ and the parameters of INMA part $\mu = \rho\Delta t$, $\beta_j = \Upsilon_i f(j\Delta)\Delta$, $j \geq 0$*
- *Υ_i are i.i.d random variables corresponding to each cluster centre $\xi_{i\Delta}$ arriving at $i\Delta$, with the Laplace transform $\hat{h}(u) = \mathbb{E}[e^{-u\Upsilon_i}]$*
- *χ_i are i.i.d random variables corresponding to each INAR cluster centre $Z_{i\Delta}$ arriving at $i\Delta$, with the Laplace transform $\hat{g}(u) = \mathbb{E}[e^{-u\chi_i}]$*
- *Let $z_t = z(t\Delta)$ with $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$.*

The probability generating functional of the INARMA sequence $\{Z_t\}_{t=1,\dots,n}$ becomes

$$\begin{aligned} G^{(Z_n)}_{(\Delta)}(z(.)) &= \exp\left\{\rho \sum_{i=1}^{n} \left(\hat{h}\left(-\sum_{k=0}^{n-i}(F^{(X_n)}(z(.)|(i+k)\Delta) - 1) f(k\Delta)\Delta\right) - 1\right)\Delta\right\} \\ F^{(X_n)}(z(.)|i\Delta) &= z(i\Delta)\hat{g}\left(-\sum_{k=1}^{n-i}\left(F^{(X_n)}(z(.)|(i+k)\Delta) - 1\right)\eta(k\Delta)\Delta\right). \end{aligned} \tag{22}$$

Proof. By substituting $\alpha_k = \chi_i \eta(k\Delta)\Delta$ into Proposition 3.6, the p.g.fl $F^{(X_n)}(z_.|i) = F(z(.)|i\Delta)$ is exactly the same as the one in Proposition 3.3

$$\begin{aligned} F^{(X_n)}(z(.)|i\Delta) &= z(i\Delta)\mathbb{E}\left[\exp\left\{\sum_{k=1}^{n-i}\chi_i\eta(k\Delta)\Delta\left(F^{(X_n)}(z(.)|(i+k)\Delta) - 1\right)\right\}\right] \\ &= z(i\Delta)\hat{g}\left(-\sum_{k=1}^{n-i}(F^{(X_n)}(z(.)|(i+k)\Delta) - 1)\eta(k\Delta)\Delta\right). \end{aligned}$$

By substituting $\beta_k = \Upsilon_i f(k\Delta)\Delta$, the p.g.fl of the whole INARMA sequence $\{Z_t\}_{t=1,\dots,n}$ becomes the p.g.fl of INMA sequence $\{Y_t\}_{t=1,\dots,n}$. Then we can apply Proposition 3.5

$$\begin{aligned} G^{(Z_n)}_{(\Delta)}(z(.)) &= G^{(Y_n)}_{(\Delta)}(F^{(X_n)}(z(.)|i\Delta)) \\ &= \exp\left\{\rho \sum_{i=1}^{n} \left(\hat{h}\left(-\sum_{k=0}^{n-i}(F^{(X_n)}(z(.)|(i+k)\Delta) - 1) f(k\Delta)\Delta\right) - 1\right)\Delta\right\}, \end{aligned}$$

where the $z(.)$ is replaced by $F^{(X_n)}(z(.)|i\Delta)$ in $G^{(Y_n)}_{(\Delta)}$. □

4. Convergence of probability generating functionals

In this section, we will prove the convergence results of the p.g.fl.s between the INARMA models and the cluster point processes.

4.1. Dynamic contagion process and INARMA model

The p.g.fl of the generalized dynamic contagion process is given by

$$\begin{aligned} G^{(DCP)}(z(.)) &= \exp\left\{\rho\int_0^T\left(\hat{h}\left(-\int_0^{T-u}(F^{(H)}(z(.)|u+v)-1)f(v)dv\right)-1\right)du\right\} \\ F^{(H)}(z(.)|u) &= z(u)\hat{g}\left(-\int_0^{T-u}(F^{(H)}(z(.)|u+v)-1)\eta(v)dv\right). \end{aligned} \tag{23}$$

The p.g.fl of the INARMA model with specific parametric setting in Theorem 3.1 is given by

$$\begin{aligned} G^{(Z_n)}_{(\Delta)}(z(.)) &= \exp\left\{\rho\sum_{i=1}^{n}\left(\hat{h}\left(-\sum_{k=0}^{n-i}(F^{(X_n)}(z(.)|(i+k)\Delta)-1)f(k\Delta)\Delta\right)-1\right)\Delta\right\} \\ F^{(X_n)}(z(.)|i\Delta) &= z(i\Delta)\hat{g}\left(-\sum_{k=1}^{n-i}(F^{(X_n)}(z(.)|(i+k)\Delta)-1)\eta(k\Delta)\Delta\right). \end{aligned} \tag{24}$$

Lemma 4.1. *If $r(u)$ is an Riemann integrable function over an interval $[a,b]$ such that $r(u)$ is bounded and the set D, the discontinuities of $r(u)$, has Lebesgue measure 0, then there exist positive constants M and k that satisfy the following inequality*

$$\left|\int_a^b r(u)du - R_n\right| \le M\Delta^k \sim O(\Delta^k), \quad n>0, \tag{25}$$

where

- *n - number of subintervals over $[a,b]$ which has the partition $\{x_0, x_1, \ldots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$*
- *$R_n = \sum_{i=1}^n r(t_i)\Delta_i$, where $x_i \in [x_{i-1}, x_i]$, $\Delta_i = x_i - x_{i-1}$ and $\Delta = \max_{i=1,\ldots,n}\Delta_i$*

Proof. From the definition of Riemann integral, for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left|\int_a^b r(u)du - R_n\right| < \epsilon, \quad for\ \Delta < \delta.$$

Then conversely, for every choice of δ, there exists ϵ such that the above inequality holds and it converges to 0 when $\delta \to 0$ from which we can infer that the ϵ is the function of Δ with a positive power. Then we let $\delta = \Delta$ and let $\epsilon = M\Delta^k > 0$ for some positive M and k such that the above inequality also holds for the case of equality . □

Proposition 4.1. *Let Θ be the parameter space to specify the generalized dynamic contagion process and the INARMA model and $z() \in \mathcal{V}_0(\mathbb{R}_+)$. There exists a positive constant k such that the rate of convergence for the absolute difference of the log p.g.fl.s between the generalized dynamic contagion process and the INARMA model is given by*

$$\begin{aligned} &D^{(DCP)}(z(.), \Delta|\Theta) = \left|\log G^{(DCP)}(z(.)) - \log G^{(Z_n)}_{(\Delta)}(z(.))\right| \sim O(\Delta^k) \\ &\lim_{n\to\infty} D^{(DCP)}(z(.), \Delta|\Theta) = 0. \end{aligned} \tag{26}$$

Proof. See Appendix A.3. □

Corollary 4.1. *Let Θ be the parameter space to specify the Cox process and the INMA model and $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$. There exists a positive constant k such that the rate of convergence for the absolute difference of the log p.g.fl.s between the Cox process and the INMA model is given by*

$$\begin{aligned} &D^{(C)}(z(.), \Delta|\Theta) = \left|\log G^{(C)}(z(.)) - \log G^{(Y_n)}_{(\Delta)}(z(.))\right| \sim O(\Delta^k) \\ &\lim_{n\to\infty} D^{(C)}(z(.), \Delta|\Theta) = 0. \end{aligned} \tag{27}$$

Proof. See Appendix A.4. □

Corollary 4.2. *Let Θ be the parameter space to specify the Hawkes process and the INAR model and $z() \in \mathcal{V}_0(\mathbb{R}_+)$. There exists a positive constant k such that the rate of convergence for the absolute difference of the log p.g.fl.s between the Hawkes process and the INAR model is given by*

$$\begin{aligned} &D^{(H)}(z(.), \Delta|\Theta) = \left|\log G^{(H)}(z(.)) - \log G^{(X_n)}_{(\Delta)}(z(.))\right| \sim O(\Delta^k) \\ &\lim_{n\to\infty} D^{(H)}(z(.), \Delta|\Theta) = 0. \end{aligned} \tag{28}$$

Proof. See Appendix A.5. □

5. Links between the INARMA models and the cluster point processes

In this section, we will construct a family of random measures $\{N_n\}_{n=1,2,...}$ on $\mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}_+})$ by aggregating the integer-valued time series and explain how the discrete time models can mimic the behaviour of those continuous time cluster point processes N. We prove that, under the weak convergence theorem, N_n will converge weakly to N as $n \to \infty$.

5.1. Preliminaries and definition

As discussed in the previous section, we can always fix a bounded area $[0, T]$ and choose a number $n > 0$, large enough. Then a continuous point process $N((0, T])$ can be treated as the sum of the bin-size count $\{N((t-1)\Delta, t\Delta)\}_{t=1,...,n}$ with $\Delta = \frac{T}{n}$. Conversely, for example the INAR model, we let the sequence $\{X_t\}_{t=1,...,n}$ be the measures for the bin-size count $\{N((t-1)\Delta, t\Delta)\}_{t=1,...,n}$. Hence if we specify the parameters in integer-valued time series models carefully and if n is large enough, we would expect the aggregation of the integer-valued time series can approximate the continuous cluster point process.

Definition 5.1. For $n > 0$, let $\{X_t\}_{t=1,...,n}$, $\{Y_t\}_{t=1,...,n}$ and $\{Z_t\}_{t=1,...,n}$ be the INAR sequence, the INMA sequence and the INARMA sequence defined in Section 3 with the parametric setting $\Delta = \frac{T}{n}$, $\alpha_0 = \nu\Delta$, $\alpha_k = \chi_i\eta(k\Delta)\Delta$ for $k > 0$, $\beta_j = \Upsilon_i f(j\Delta)\Delta$ for $j \geq 0$ and $\mu = \rho\Delta$. Define the following three families of point processes,

$$\begin{aligned} N_n^{(H)}(A) &= \sum_{t:t\Delta\in A} X_t \\ N_n^{(C)}(A) &= \sum_{t:t\Delta\in A} Y_t \\ N_n^{(DCP)}(A) &= \sum_{t:t\Delta\in A} Z_t \end{aligned} \tag{29}$$

where A is a bounded set in $\mathcal{B}(\mathbb{R}_+)$ and T is a constant such that $T \geq \sup A$. The joint distribution of these point processes is uniquely determined by their p.g.fl.s derived in Section 3.

The idea here is basically followed from Kirchner [27]. To prove the weak convergence, he defined the INAR model and constructed a family of point processes $N^{(\Delta)}$ by aggregating the INAR sequence over $A \in \mathcal{B}(\mathbb{R})$, the Borel σ-algebra on $\mathbb{R}$. Then he proved the weak convergence of $N^{(\Delta)}$ to the Hawkes process N from the definition point of view, see definition 5 and Theorem 2 in Kirchner [27]. He also mentioned this can be proved in a different way by showing the convergence of the Laplace functional of $N^{(\Delta)}$. In our case, we will use probability generating functionals.

5.2. Weak convergence

From Definition 2.1 in section 2 and Proposition 9.2.II in Daley and Vere-Jones [12], we can say that the distribution of a random measure (point process) ζ on $(\mathcal{N}^{\#}_{\mathbb{R}_+}, \mathcal{B}(\mathcal{N}^{\#}_{\mathbb{R}_+}))$ is completely determined by its finite-dimensional distributions. Then for the weak convergence of random measure on $\mathcal{N}^{\#}_{\mathbb{R}_+}$, it is sufficient to prove the convergence of finite dimensional distributions, which is established by Theorem 11.1. VII in [12].

Proposition 5.1. *Let $\mathcal{X}$ be a complete separable metric space and let $\mathcal{P}$, $\{\mathcal{P}_n\}$ be distributions on $(\mathcal{M}^{\#}_{\mathcal{X}}, \mathcal{B}(\mathcal{M}^{\#}_{\mathcal{X}}))$. Then $\mathcal{P}_n \to \mathcal{P}$ weakly if and only if the finite-dimensional distributions of $\mathcal{P}_n$ converge weakly to those of $\mathcal{P}$.*

In our case, the state space is $\mathcal{X} = \mathbb{R}_+$. Also, there is one-to-one mapping from finite dimensional distributions to its probability generating functional. Hence it is sufficient to prove the convergence of the p.g.fl.s between point processes. This is confirmed by another Proposition 11.1.VIII in [12]. We only write down part of it here.

Proposition 5.2. *Each of the following conditions is equivalent to the weak convergence $\mathcal{P}_n \to \mathcal{P}$, assuming the function f ranges over the space of continuous functions vanishing outside a bounded set.*

- *The distribution of $\int_{\mathcal{X}} f d\zeta$ under $\mathcal{P}_n$ converges weakly to its distribution under $\mathcal{P}$*
- *For point process, the p.g.fl.s $G_n[z]$ converge to $G[z]$ for each continuous $z \in \mathcal{V}_0(\mathcal{X})$*

Before establishing the convergence theorem, we need to first show the probability measures of those point processes defined in 5.1 are uniformly tight. Here we refer and combine the results of Lemma 1 and 2 in Kirchner [27]. We also derive a similar one for $N^{(C)}_n$ and $N^{(DCP)}_n$.

Lemma 5.1. *For any bounded interval $[a, b]$ on $\mathbb{R}_+$, we can always find a constant $T > b$ and define $\Delta = \frac{T}{n} \in (0, \delta)$ for some constant $\delta > 0$ as long as $n > \left[\frac{T}{\delta}\right]$. Let $N^{(H)}_n$ be the point process defined in 5.1 Then there exists a constant $B^{(H)}$ such that*

$$\mathbb{E}[N^{(H)}_n([a, b])] < (b - a + 2\delta)\nu B^{(H)}$$
$$B^{(H)} = \begin{cases} (1-K)^{-1}, & \text{if } K < 1 \\ (1 + K + K^2 + \cdots + K^m), & \text{otherwise} \end{cases} \tag{30}$$
$$K = \mu_\chi \sum_{k=1}^{\infty} \eta(k\Delta)\Delta$$

Proof. The coefficients $(b-a+2\delta)\nu$ denote the upper bound of the expected number of immigrants over the fixed time interval $[a,b]$, whose derivation is given in Kirchner [27]. In the stationary case where the branching ratio $K<1$. The expected size of a cluster for INAR(∞) over a long time horizon is evaluated as $(1+K+K^2+\cdots)=(1-K)^{-1}$. In the non-stationary case, since the offspring is produced by Poisson distribution, there is a positive waiting time before a new generation is produced. So over the bounded interval $[a,b]$, there exists a constant $m>0$ and the size of a cluster is the sum of m generations $(1+K+K^2+\cdots+K^m)$ □

Lemma 5.2. *For any bounded interval $[a,b]$ on $\mathbb{R}_+$, we can always find a constant $T>b$ and define $\Delta=\frac{T}{n}\in(0,\delta)$ for some constant $\delta>0$ as long as $n>\left[\frac{T}{\delta}\right]$. Let $N_n^{(C)}$ and $N_n^{(DCP)}$ be the point processes defined in 5.1. Then there exist constants $B^{(H)}$ and $L(T)$ such that*

$$\begin{aligned}&\mathbb{E}[N_n^{(C)}([a,b])]<(b-a+2\delta)\rho L(T)\\&\mathbb{E}[N_n^{(DCP)}([a,b])]<(b-a+2\delta)\rho L(T)B^{(H)},\end{aligned}\tag{31}$$

where $L(T)=\mu_\Upsilon(\int_0^T f(t)dt+c)$. The constant c is defined as $c=\left|\int_0^T f(t)dt-\sum_{k=0}^{n-1}f(k\Delta)\Delta\right|$

Proof. From the definition of INMA model, the expectation is

$$\begin{aligned}\mathbb{E}[Y_t]&=\mathbb{E}[\beta_0\circ\xi_t+\cdots+\beta_{t-1}\circ\xi_1]\\&=\mathbb{E}[\xi_i]\sum_{k=0}^{n-1}\mathbb{E}[\beta_k]\\&=\rho\Delta\sum_{k=0}^{n-1}\mu_\Upsilon f(k\Delta)\Delta\\&\le\rho\Delta\mu_\Upsilon\left(\int_0^T f(t)dt+c\right)\\&\le\rho L(T)\Delta.\end{aligned}$$

The number of subintervals over $[a,b]$ is $[\frac{b-a}{\Delta}]+1<\frac{b-a}{\Delta}+2$. Finally we have

$$\begin{aligned}\mathbb{E}[N_n^{(C)}([a,b])]&=\sum_{t,t\Delta\in[a,b]}\mathbb{E}[Y_t]\\&\le\left(\left[\frac{b-a}{\Delta}\right]+1\right)\rho L(T)\Delta\\&<\left(\frac{b-a}{\Delta}+2\right)\rho L(T)\Delta\\&<(b-a+2\delta)\rho L(T).\end{aligned}$$

The upper bound for $\mathbb{E}[N_n^{(DCP)}([a,b])]$ can be derived similarly as that of $\mathbb{E}[N_n^{(H)}([a,b])]$. We need to replace ν by $\rho L(T)$ □

Lemma 5.3. *The families of the probability measures $\mathcal{P}_n^{(C)},\mathcal{P}_n^{(H)},\mathcal{P}_n^{(DCP)}$ on $\left(\mathcal{N}^\#_{\mathbb{R}_+},\mathcal{B}(\mathcal{N}^\#_{\mathbb{R}_+})\right)$ corresponding to the point processes $N_n^{(C)},N_n^{(H)},N_n^{(DCP)}$ respectively are uniformly tight.*

Proof. For any bounded interval $[a, b]$ on $\mathbb{R}_+$, we can always find a constant $T > b$ and define $\Delta = \frac{T}{n}$, such that $\Delta \in (0, \delta)$ for some constant $\delta > 0$ as long as $n > \left[\frac{T}{\delta}\right]$. To show the tightness, for every $\epsilon > 0$, we can let $M_\epsilon^{(H)} = (b - a + 2\delta)\frac{\nu B^{(H)}}{\epsilon}$, $M_\epsilon^{(C)} = (b - a + 2\delta)\rho L(T)\frac{1}{\epsilon}$ and $M_\epsilon^{(DCP)} = (b - a + 2\delta)\frac{\rho L(T)B^{(H)}}{\epsilon}$ such that

$$P(N_n^{(H)}([a,b]) > M_\epsilon^{(H)}) \leq \frac{\mathbb{E}[N_n^{(H)}([a,b])]}{M_\epsilon^{(H)}} < (b-a+2\delta)\frac{\nu B^{(H)}}{M_\epsilon^{(H)}} = \epsilon$$
$$P(N_n^{(C)}([a,b]) > M_\epsilon^{(C)}) \leq \frac{\mathbb{E}[N_n^{(C)}([a,b])]}{M_\epsilon^{(H)}} < \frac{(b-a+2\delta)\rho L(T)}{M_\epsilon^{(H)}} = \epsilon$$
$$P(N_n^{(DCP)}([a,b]) > M_\epsilon^{(DCP)}) \leq \frac{\mathbb{E}[N_n^{(DCP)}([a,b])]}{M_\epsilon^{(H)}} < (b-a+2\delta)\frac{\rho L(T)B^{(H)}}{M_\epsilon^{(DCP)}} = \epsilon$$

Here we apply the Markov inequality. □

Theorem 5.1. *Let $N^{(H)}, N^{(C)}, N^{(DCP)}$ be the Hawkes process, the Cox process and the generalized dynamic contagion process defined in Section 2. For $n > 0$, let $N_n^{(H)}$, $N_n^{(C)}$ and $N_n^{(DCP)}$ be the point processes defined in 5.1. Then we have the following weak convergence results*

$$\begin{aligned} &N_n^{(H)} \xrightarrow{w} N^{(H)} \\ &N_n^{(C)} \xrightarrow{w} N^{(C)} \\ &N_n^{(DCP)} \xrightarrow{w} N^{(DCP)} \quad \text{as } n \to \infty. \end{aligned} \tag{32}$$

Proof. Uniform tightness of the three families of point processes is followed by Lemma 5.3. From the preliminaries in Section 2, the distribution of a random measure N on $\mathcal{N}^\#_{\mathbb{R}_+}$ is completely determined by the finite dimensional distributions see Proposition 9.2.III in [12], i.e. the joint distribution for all finite families of bounded Borel sets $A_1, \ldots, A_k$ on $\mathbb{R}_+$ of the random variable $N(A_1), \ldots, N(A_k)$. From the tightness lemma, it is clear that all finite dimensional distribution for the point processes $N_n^{(.)}$ restricted to $[a, b]$ are uniformly tight. Consequently, there always exists a constant $T > b$ such that we can uniquely describe the finite dimensional distributions by its probability generating functional on the bounded area $[0, T]$. Combining the convergence results in Proposition 4.1, Corollaries 4.1 and 4.2 in Section 4, i.e. the absolute difference of the log p.g.fl.s between the point processes $N_n^{(.)}$ and $N^{(.)}$ goes to 0 as $\Delta \to 0$, equivalently $n \to \infty$

$$\lim_{n\to\infty} \left|\log G^{(H)}(z(.)) - \log G^{(X_n)}_{(\Delta)}(z(.))\right| = 0$$
$$\lim_{n\to\infty} \left|\log G^{(C)}(z(.)) - \log G^{(Y_n)}_{(\Delta)}(z(.))\right| = 0$$
$$\lim_{n\to\infty} \left|\log G^{(DCP)}(z(.)) - \log G^{(Z_n)}_{(\Delta)}(z(.))\right| = 0,$$

we can now apply Proposition 5.2 and state that the families of point processes $N_n^{(H)}$, $N_n^{(C)}$ and $N_n^{(DCP)}$ converge weakly to $N^{(H)}$, $N^{(C)}$ and $N^{(DCP)}$ respectively as $n \to \infty$. □

6. Concluding remarks

In this paper, we review the continuous cluster point process in a general parametric setting. Then we review the Poisson thinning INAR model and introduce the Poisson thinning INMA and the INARMA models. We prove that these integer-valued time series models, under some specific parametric setting, are actually the discrete versions of the cluster point processes $N_t^{(.)}$ with continuous stochastic intensity $\lambda_t^{(.)}$. We confirm Kirchner's thought in [27] on the relationship between the INARMA model and the dynamic contagion process. If there is a simple and effective estimation procedure for the INARMA model, for example the one Kirchner did in [28] for the INAR model, then the dynamic contagion process can be applied to those Hawkes-based processes. However, there are some potential issues left to be addressed. For example, can we make use of the structure standard ARMA model to perform estimation for the integer-valued version? How can we deal with random variables in the coefficients of time series models (random coefficients)? These are all proposed as topics for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the anonymous referee for their very helpful comments and suggestions which have significantly improved this article.

Appendix. Proofs

A.1. Proof of Proposition 2.1

The Cox process is basically a cluster point process such that,

- The arrivals of cluster centres c_i follow $N^* \sim Pois(\rho)$ a homogeneous Poisson process
- Conditionally on c_i, each cluster centre will generate a cluster, the size of which follows $N_T^1 \sim Pois(\Upsilon_i f(T-c_i))$.

Vere-Jones [34] gives the p.g.fl of a cluster process as

$$G(z(.)) = G_0(F(z(.)|t)), \tag{33}$$

where $G_0()$ is the p.g.fl of the process of cluster centres and $F(z(.)|t)$ is the p.g.fl for a cluster given that the cluster centre occurs at time t. Combining the second bullet point, we have

$$\begin{aligned}
F^{(C)}(z(.)|c) &= \mathbb{E}[\exp \int_{\mathbb{R}_+} \log z(s) N_T^1(ds)] \\
&= \mathbb{E} \exp \left\{ \Upsilon_i \int_c^T f(s-c)(z(s)-1)ds \right\} \\
&= \mathbb{E} \exp \left\{ \Upsilon_i \int_0^{T-c} f(u)(z(c+u)-1)du \right\} \\
&= \hat{h} \left(-\int_0^{T-c} f(u)(z(c+u)-1)du \right).
\end{aligned}$$

Hence the p.g.fl of the Cox process is

$$G^{(C)}(z(.)) = \mathbb{E}\left[\exp \int_{\mathbb{R}_+} \log F^{(C)}(z(.)|c)N^*(dc)\right]$$
$$= \exp\left\{\rho \int_0^T (F^{(C)}(z(.)|c) - 1)dc\right\}. \quad \square$$

A.2. *Proof of* Proposition 2.2

The generalized dynamic contagion process is a cluster process,

- The arrivals of immigrants follow the Cox process with intensity $\lambda_t^{(C)}$.
- Each immigrant generates a Galton–Watson type branching process with expected branching ratio $\mu_\chi \int_0^\infty \eta(u)du < 1$. The cluster is formed by including all generations from the branching process.

Let $\mathcal{F}_t^{(C)}$ be the filtration generated by $\lambda_t^{(C)}$. Conditionally on $\mathcal{F}_t^{(C)}$, the p.g.fl of the generalized dynamic contagion process is just the p.g.fl of the Hawkes process with its immigration process being an inhomogeneous Poisson process. Then we can apply Theorem 2 in [20]

$$G(z(.)|\mathcal{F}_t^{(C)}) = \exp\left\{\int_0^T \left(F^{(H)}(z(.)|u) - 1\right)\lambda_u^{(c)}du\right\}$$
$$F^{(H)}(z(.)|u) = z(u)\hat{g}\left(-\int_0^{T-u} (F^{(H)}(z(.)|u+v) - 1)\eta(v)dv\right).$$

The underlying intensity function is $\lambda_t = \nu + \sum_{i:\tau_i<t} \gamma(t-\tau_i)$ in [20]. In our case, we are working on the bounded area $[0,T]$ and $1 - h(u) = 0$ when u lies outside $[0,T]$. By the definition of p.g.fl, $F(z(.)|u) = 1$ when u lies outside $[0,T]$. The ranges of integrals for $G(z(.)|\mathcal{F}_t^{(C)})$ and $F^{(H)}(z(.)|u)$, therefore, reduce to $[0,T]$ and $[0,T-u]$ respectively. Then we substitute $\gamma(t-\tau_i)$ with $\chi_i f(t-\tau_i)$ and take expectation with respect to χ_i. Finally, the unconditional p.g.fl of the generalized dynamic contagion process is $\mathbb{E}[G(z(.)|\mathcal{F}_t^{(C)})]$, which turns out to be the p.g.fl of the Cox process. Then we can apply the results from Proposition 2.1

$$G^{(DCP)}(z(.)) = \mathbb{E}\left[\exp\left\{\int_0^T \left(F^{(H)}(z(.)|u) - 1\right)\lambda_u^{(C)}du\right\}\right]$$
$$= \exp\left\{\rho \int_0^T \left(\hat{h}\left(-\int_0^{T-c} (F^{(H)}(z(.)|u+c) - 1)f(u)du\right) - 1\right)dc\right\}. \quad \square \tag{34}$$

A.3. *Proof of* Proposition 4.1

Let us define the following quantities

$$I_1 = \int_0^T \left(\hat{h}(I_2(u)) - 1\right)du$$
$$I_2(u) = \int_0^{T-u} (1 - F^{(H)}(z(.)|u+v))f(v)dv$$
$$R_1 = \sum_{i=1}^n (\hat{h}(I_2((i-1)\Delta)) - 1)\Delta$$

$$R_2(i) = \sum_{k=1}^{n-i+1} (1 - F^{(H)}(z(.)|(k+i-1)\Delta)) f((k-1)\Delta)\Delta$$
$$R_3(i) = \sum_{k=1}^{n-i+1} (1 - F^{(X_n)}(z(.)|(k+i-1)\Delta)) f((k-1)\Delta)\Delta$$
$$J_i = \mu_\chi \left| \int_0^{T-i\Delta t} (F^{(H)}(z(.)|u+v) - 1)\eta(v)dv - \sum_{k=1}^{n-i} (F^{(H)}(z(.)|u+v)\eta(k\Delta)\Delta \right|.$$

Then $D^{(DCP)}(z(.), \Delta|\Theta)$ can be decomposed as

$$\begin{aligned} D^{(DCP)}(z(.), \Delta|\Theta) &= \left| \log G^{(DCP)}(z(.)) - \log G^{(Z_n)}_{(\Delta)}(z(.)) \right| \\ &= \rho \left| \int_0^T (\hat{h}(I_2(u)) - 1)du - \sum_{i=1}^{n} (\hat{h}(R_3(i)) - 1)\Delta \right| \\ &\le \rho \left| \int_0^T (\hat{h}(I_2(u)) - 1)du - R_1 \right| + \rho \left| R_1 - \sum_{i=1}^{n} (\hat{h}(R_3(i)) - 1)\Delta \right|. \end{aligned} \tag{35}$$

Here we add the inter-median term R_1 which is the Riemann sum of its corresponding integral. Then apply Lemma 4.1 to the first part

$$\left| \int_0^T (\hat{h}(I_2(u)) - 1)du - R_1 \right| \sim O(\Delta^{k_1}).$$

For the second part, we make use of the property of the convex function $\hat{h}(u)$ such that

$$\begin{aligned} \left| R_1 - \sum_{i=1}^{n} (\hat{h}(RS_3(i)) - 1)\Delta \right| &= \left| \sum_{i=1}^{n} (\hat{h}(I_2((i-1)\Delta)) - 1)\Delta - \sum_{i=1}^{n} (\hat{h}(R_3(i)) - 1)\Delta \right| \\ &\le \sum_{i=1}^{n} |I_2((i-1)\Delta) - R_3(i)| \, \mu_\Upsilon \Delta \\ &< \sum_{i=1}^{n} |I_2((i-1)\Delta) - R_2(i)| \, \mu_\Upsilon \Delta \\ &+ \sum_{i=1}^{n} |R_2(i) - R_3(i)| \, \mu_\Upsilon \Delta, \end{aligned} \tag{36}$$

which again separates into two parts. For the first part, apply Lemma 4.1

$$\sum_{i=1}^{n} |I_2((i-1)\Delta t) - R_2(i)| \, \mu_\Upsilon \Delta t \sim O(\Delta^{k_2}).$$

For the second part,

$$
\begin{aligned}
|R_2(i) - R_3(i)| &= \sum_{k=0}^{n-i} \left| F^{(H)}(z(.)|(k+i)\Delta) - F^{(X_n)}(z(.)|(k+i)\Delta) \right| f(k\Delta)\Delta \\
&\leq f_m \Delta \sum_{k=0}^{n-i} \left| F^{(H)}(z(.)|(k+i)\Delta) - F^{(X_n)}(z(.)|(k+i)\Delta) \right| \\
&\leq f_m \Delta \sum_{k=0}^{n} \left| F^{(H)}(z(.)|k\Delta) - F^{(X_n)}(z(.)|k\Delta) \right| \\
f_m &= \max_{k=1,\dots,n} f(k\Delta).
\end{aligned}
$$

For the absolute difference $\left| F^{(H)}(z(.)|k\Delta) - F^{(X_n)}(z(.)|k\Delta) \right|$, we can solve it backwardly. When $i = n$,

$$
|F^{(H)}(z(.)|n\Delta) - F^{(X_n)}(z(.)|n\Delta)| = 0.
$$

When $i = n-1$,

$$
\begin{aligned}
&|F^{(H)}(z(.)|(n-1)\Delta) - F^{(X_n)}(z(.)|(n-1)\Delta)|, \\
&\leq z((n-1)\Delta) \left(|\hat{g}(-I_{n-1}) - \hat{g}(-R_{n-1})| + |\hat{g}(-R_{n-1}) - \hat{g}(-R_{n-1}^b)| \right) \\
&\leq J_{n-1} + \sum_{k=1}^{1} \mu_\chi \eta(k\Delta) |F^{(H)}(z(.)|(n-1+k)\Delta) - F^{(X_n)}(z(.)|(n-1+k)\Delta)|\Delta \\
&= J_{n-1},
\end{aligned}
$$

where we use the condition $z(.) \in \mathcal{V}_0(\mathbb{R}_+)$ and $z(u) \leq 1, u \in (0, T)$. Then when $i = n-2$,

$$
\begin{aligned}
&|F^{(H)}(z(.)|(n-2)\Delta) - F^{(X_n)}(z(.)|(n-2)\Delta)|, \\
&\leq z((n-2)\Delta) \left(|\hat{g}(-I_{n-2}) - \hat{g}(-R_{n-2})| + |\hat{g}(-R_{n-2}) - \hat{g}(-R_{n-2}^b)| \right) \\
&\leq J_{n-2} + \sum_{k=1}^{2} \mu_\chi \eta(k\Delta) |F^{(H)}(z(.)|(n-2+k)\Delta) - F^{(X_n)}(z(.)|(n-2+k)\Delta)|\Delta \\
&\leq J_{n-2} + J_{n-1}\mu_\chi \eta(\Delta)\Delta \\
&= J_{n-2} + O(\Delta^{k_3+1}) \sim O(\Delta^{k_3}).
\end{aligned}
$$

Note that J_i is the absolute difference between the Integral and its Riemann sum, hence we can apply Lemma 4.1

$$
\begin{aligned}
&J_i \leq M_i \Delta^{k_3} \leq M' \Delta^{k_3} \sim O(\Delta^{k_3}) \\
&M' = \max_{i=0,\dots,n} M_i.
\end{aligned}
$$

When $i = n-j,\ j = 1, 2, \dots, n$,

$$
\begin{aligned}
&|F^{(H)}(z(.)|(n-j)\Delta) - F^{(X_n)}(z(.)|(n-j)\Delta)|, \\
&\leq z((n-j)\Delta) \left(|\hat{g}(-I_{n-j}) - \hat{g}(-R_{n-j})| + |\hat{g}(-R_{n-j}) - \hat{g}(-R_{n-j}^b)| \right) \\
&\leq J_{n-j} + \sum_{k=1}^{j} \mu_\chi \eta(k\Delta) |F^{(H)}(z(.)|(n-j+k)\Delta) - F^{(X_n)}(z(.)|(n-j+k)\Delta)|\Delta \\
&= J_{n-j} + jO(\Delta^{k_3+1}) \sim O(\Delta^{k_3}).
\end{aligned}
$$

Then the whole sum becomes

$$\begin{aligned}&\sum_{i=0}^{n}|F^{(H)}(z(.)|i\Delta)-F^{(X_n)}(z(.)|i\Delta)|\Delta\\&\leq\Delta\sum_{i=0}^{n-1}\left(J_i+iO(\Delta^{k_3+1})\right)\sim O(\Delta^{k_3}).\end{aligned}\tag{37}$$

Then the second part in Eq. (36) becomes

$$\begin{aligned}\sum_{i=0}^{n}|R_2(i)-R_3(i)|\,\mu_\Upsilon\Delta\leq\sum_{i=1}^{n}&\left(f_m\Delta\sum_{k=0}^{n}\left|F^{(H)}(z(.)|k\Delta)-F^{(X_n)}(z(.)|k\Delta)\right|\right)\\&\times\mu_\Upsilon\Delta\sim O(\Delta^{k_3}).\end{aligned}$$

Finally, let $k=\min\{k_1,k_2,k_3\}$. $\square$

A.4. Proof of Corollary 4.1

The results follow from Proposition 4.1. The p.g.fl of the Cox process $G^{(C)}$ can be derived from the p.g.fl of the generalized dynamic contagion process $G^{(DCP)}$ by letting $\eta(u)=0$ such that $F^{(H)}(z(.))$ becomes

$$F^{(H)}(z(.)|u)=z(u)\hat{g}(0)=z(u).$$

Similarly, $G^{(Y_n)}_{(\Delta)}$ can be derived from $G^{(Z_n)}_{(\Delta)}$ by letting $F^{(X_n)}(z(.)|i\Delta)=z(i\Delta)$. Then $D^{(C)}(z(.),\Delta|\Theta)$ will have the same form as Eqs. (35) and (36) such that

$$\begin{aligned}D^{(C)}(z(.),\Delta|\Theta)&=\left|\log G^{(C)}(z(.))-\log G^{(Y_n)}_{(\Delta))}(z(.))\right|\\&=\rho\left|\int_0^T(\hat{h}(I_2(u))-1)du-\sum_{i=1}^{n}(\hat{h}(R_3(i))-1)\Delta\right|\\&\leq\rho\left|\int_0^T(\hat{h}(I_2(u))-1)du-R_1\right|+\rho\left|R_1-\sum_{i=1}^{n}(\hat{h}(R_3(i))-1)\Delta\right|\\&\leq\rho\left|\int_0^T(\hat{h}(I_2(u))-1)du-R_1\right|+\rho\sum_{i=1}^{n}|I_2((i-1)\Delta t)-R_2(i)|\,\mu_\Upsilon\Delta\\&\quad+\rho\sum_{i=1}^{n}|R_2(i)-R_3(i)|\,\mu_\Upsilon\Delta\\&\sim O(\Delta^{k_1})+O(\Delta^{k_2}),\end{aligned}$$

where $R_2(i)-R_3(i)=0$ since $F^{(H)}(z(.)|i\Delta)=z(i\Delta)=F^{(X_n)}(z(.)|i\Delta)$. Finally, we can take $k=\min\{k_1,k_2\}$. $\square$

A.5. Proof of corollary Corollary 4.2

Similarly, this result follows from Proposition 4.1. From the p.g.fl.s point of view, $G^{(H)}$ can be recovered by replacing ρ and $\hat{h}\left(-\int_0^{T-u}(F^{(H)}(z(.)|u+v)-1)f(v)dv\right)$ by ν and

$F^{(H)}(z(.)|u)$ in $G^{(DCP)}$ respectively. $G_{(\Delta)}^{(X_n)}$ can be derived from $G_{(\Delta)}^{(Z_n)}$ in a similar way. Then $D^{(H)}$ becomes

$$\begin{aligned} D^{(H)}(z(.), \Delta|\Theta) &= \nu \left| \int_0^T (F^{(H)}(z(.)|u) - 1)du - \sum_{i=1}^{n} (F^{(X_n)}(z(.)|i\Delta) - 1)\Delta \right| \\ &\leq \nu \left| \int_0^T (F^{(H)}(z(.)|u) - 1)du - \sum_{i=1}^{n} (F^{(H)}(z(.)|i\Delta) - 1)\Delta \right| \\ &+ \nu \left| \sum_{i=1}^{n} (F^{(H)}(z(.)|i\Delta) - 1)\Delta - \sum_{i=1}^{n} (F^{(X_n)}(z(.)|i\Delta) - 1)\Delta \right|. \end{aligned}$$

Adopting the similar technique as in Proposition 4.1, we add the term $\sum_{i=1}^{n}(F^{(H)}(z(.)|i\Delta)-1)\Delta$ which is the right Riemann sum of the integral. Then we can apply Lemma 4.1

$$\left| \int_0^T (F^{(H)}(z(.)|u) - 1)du - \sum_{i=1}^{n} (F^{(H)}(z(.)|i\Delta) - 1)\Delta \right| \sim O(\Delta^{k_1}), \quad k_1 > 0.$$

The second part is

$$\begin{aligned} &\left| \sum_{i=1}^{n} (F^{(H)}(z(.)|i\Delta) - 1)\Delta - \sum_{i=1}^{n} (F^{(X_n)}(z(.)|i\Delta) - 1)\Delta \right| \\ &\leq \sum_{i=1}^{n} |F^{(H)}(z(.)|i\Delta) - F^{(X_n)}(z(.)|i\Delta)|\Delta \sim O(\Delta^{k_2}). \end{aligned}$$

This result follows from the inequality (37). Finally, we can take $k = \min\{k_1, k_2\}$. □

References

[1] Y. Aït-Sahalia, J. Cacho-Diaz, R.J. Laeven, Modeling financial contagion using mutually exciting jump processes, J. Financ. Econ. 117 (3) (2015) 585–606.

[2] M. Al-Osh, A.A. Alzaid, First-order integer-valued autoregressive (INAR(1)) process, J. Time Series Anal. 8 (3) (1987) 261–275.

[3] M. Al-Osh, A.A. Alzaid, First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties, Stat. Neerl. 42 (1) (1988) 53–61.

[4] M. Al-Osh, A.A. Alzaid, Integer-valued moving average (INMA) process, Statist. Papers 29 (1) (1988) 281–300.

[5] E. Bacry, K. Dayri, J.-F. Muzy, Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data, Eur. Phys. J. B 85 (5) (2012) 1–12.

[6] E. Bacry, S. Delattre, M. Hoffmann, J.-F. Muzy, Modelling microstructure noise with mutually exciting point processes, Quant. Finance 13 (1) (2013) 65–77.

[7] E. Bacry, S. Delattre, M. Hoffmann, J.-F. Muzy, Some limit theorems for Hawkes processes and application to financial statistics, Stochastic Process. Appl. 123 (7) (2013) 2475–2499.

[8] E. Bacry, I. Mastromatteo, J.-F. Muzy, Hawkes processes in finance, Mark. Microstruct. Liq. 1 (01) (2015) 1550005.

[9] C.G. Bowsher, Modelling security market events in continuous time: Intensity based, multivariate point process models, J. Econometrics 141 (2) (2007) 876–912.

[10] K. Brännäs, A. Hall, Estimation in integer-valued moving average models, Appl. Stoch. Models Bus. Ind. 17 (3) (2001) 277–291.

[11] K. Brännäs, J. Hellström, J. Nordström, A new approach to modelling and forecasting monthly guest nights in hotels, Int. J. Forecast. 18 (1) (2002) 19–30.

[12] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure, Springer Science & Business Media, 2007.

[13] A. Dassios, J.-W. Jang, Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity, Finance Stoch. 7 (1) (2003) 73–95.
[14] A. Dassios, H. Zhao, A dynamic contagion process, Adv. Appl. Probab. 43 (3) (2011) 814–846.
[15] A. Dassios, H. Zhao, Efficient simulation of clustering jumps with CIR intensity, Oper. Res. 65 (6) (2017) 1494–1515.
[16] A. Dassios, H. Zhao, A generalized contagion process with an application to credit risk, Int. J. Theor. Appl. Finance 20 (01) (2017) 1750003.
[17] P. Embrechts, T. Liniger, L. Lin, Multivariate Hawkes processes: an application to financial data, J. Appl. Probab. 48 (A) (2011) 367–378.
[18] A.G. Hawkes, Point spectra of some mutually exciting point processes, J. R. Stat. Soc. Ser. B Stat. Methodol. 33 (3) (1971) 438–443.
[19] A.G. Hawkes, Spectra of some self-exciting and mutually exciting point processes, Biometrika 58 (1) (1971) 83–90.
[20] A.G. Hawkes, D. Oakes, A cluster process representation of a self-exciting process, J. Appl. Probab. (1974) 493–503.
[21] P.A. Jacobs, P.A. Lewis, Discrete time series generated by mixtures. I: Correlational and runs properties, J. R. Stat. Soc. Ser. B Stat. Methodol. 40 (1) (1978) 94–105.
[22] P.A. Jacobs, P.A. Lewis, Discrete time series generated by mixtures II: asymptotic properties, J. R. Stat. Soc. Ser. B Stat. Methodol. 40 (2) (1978) 222–228.
[23] P.A. Jacobs, P.A. Lewis, Stationary discrete autoregressive-moving average time series generated by mixtures, J. Time Series Anal. 4 (1) (1983) 19–36.
[24] R.A. Jarrow, F. Yu, Counterparty risk and the pricing of defaultable securities, J. Finance 56 (5) (2001) 1765–1799.
[25] D. Jin-Guan, L. Yuan, The integer-valued autoregressive (INAR(p)) model, J. Time Series Anal. 12 (2) (1991) 129–142.
[26] H. Kim, Spatio-Temporal Point Process Models for the Spread of Avian Influenza Virus (H5N1) (Ph.D. thesis), UC Berkeley, 2011.
[27] M. Kirchner, Hawkes and INAR(∞) processes, Stochastic Process. Appl. 126 (8) (2016) 2494–2525.
[28] M. Kirchner, An estimation procedure for the Hawkes process, Quant. Finance 17 (4) (2017) 571–595.
[29] J. Large, Measuring the resiliency of an electronic limit order book, J. Financial Mark. 10 (1) (2007) 1–25.
[30] E. McKenzie, Some simple models for discrete variate time series 1, JAWRA J. Am. Water Resour. Assoc. 21 (4) (1985) 645–650.
[31] E. McKenzie, Some ARMA models for dependent sequences of Poisson counts, Adv. Appl. Probab. (1988) 822–835.
[32] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, G.E. Tita, Self-exciting point process modeling of crime, J. Amer. Statist. Assoc. 106 (493) (2011) 100–108.
[33] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, J. Amer. Statist. Assoc. 83 (401) (1988) 9–27.
[34] D. Vere-Jones, Stochastic models for earthquake occurrence, J. R. Stat. Soc. Ser. B Stat. Methodol. 32 (1) (1970) 1–45.