

# Hawkes and INAR( $\infty$ ) processes

Matthias Kirchner

*Department of Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland*

Received 8 July 2015; received in revised form 3 January 2016; accepted 23 February 2016

Available online 2 March 2016

## Abstract

In this paper, we discuss integer-valued autoregressive time series (INAR), Hawkes point processes, and their interrelationship. Besides presenting structural analogies, we derive a convergence theorem. More specifically, we generalize the well-known INAR( $p$ ),  $p \in \mathbb{N}$ , time series model to a corresponding model of infinite order: the INAR( $\infty$ ) model. We establish existence, uniqueness, finiteness of moments, and give formulas for the autocovariance function as well as for the joint moment-generating function. Furthermore, we derive a branching-process – as well as an AR( $\infty$ ) – and an MA( $\infty$ ) representation for the model. We compare Hawkes process properties with their INAR( $\infty$ ) counterparts. Given a Hawkes process  $N$ , in the main theorem of the paper we construct an INAR( $\infty$ )-based family of point processes and prove its convergence to  $N$ . This connection between INAR and Hawkes models will be relevant in applications.

© 2016 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

MSC: 60G55; 60F99; 37M10

Keywords: Hawkes process; Integer-valued time series; Weak convergence of point processes; Branching process

## 0. Introduction

In this paper, we show that Hawkes point processes are continuous-time versions of integer-valued autoregressive time series and – vice versa – that integer-valued autoregressive time series are discrete-time versions of Hawkes point processes; see [Theorem 2](#) for the main result of the paper. To start with, we outline the history of the concepts involved.

Standard time series theory for sequences of real-valued data points has been developed in seminal works like Whittle [25] and Box and Jenkins [3]. This theory led to the natural question

---

E-mail address: [matthias.kirchner@math.ethz.ch](mailto:matthias.kirchner@math.ethz.ch).

<http://dx.doi.org/10.1016/j.spa.2016.02.008>

0304-4149/© 2016 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of time series models for count data. In the count-data context, the starting point is also a defining system of difference equations of the form “ $X_n - \sum \alpha_k X_{n-k} = \varepsilon_n + \sum \beta_k \varepsilon_{n-k}$ ,  $n \in \mathbb{Z}$ ”. The main idea of the construction is to manipulate these equations in such a way that their solutions are integer-valued. This can be achieved by giving the error terms “ $(\varepsilon_n)$ ” a distribution supported on  $\mathbb{N}_0$  and by substituting all multiplications with thinning operations. In the above spirit, autoregressive integer-valued (INAR) time series were defined and examined by McKenzie [19] and Al-Osh and Alzaid [1]. The modern definition of the INAR model comes from [17]. Latour [16] generalizes the model to the multivariate case. For an exhaustive collection of properties of the INAR model; see [8]. For a textbook reference; see [9].

The Hawkes process was introduced in [11,12] as a model for contagious processes such as measles infections or hijackings. As a point process in continuous time, the Hawkes process allows for the modeling of intensities which depend on the past of the process itself. Its alternative name, “self-exciting point process”, stems from the fact that, given the occurrence of an event, intensity jumps upwards and then decays gradually. Theoretical cornerstones for the model are Hawkes [13] which establishes the representation as a cluster process, Ogata [21] which covers calibration issues and propagates a recursive method for likelihood calculations, Brémaud and Massoulié [4] which extends the original model by generalizing the affine dependence on the past to Lipschitz dependence, Brémaud and Massoulié [5] which proves the existence of a specific borderline case of the model, and Liniger [18] which puts the subtleties of the definition and the construction on a solid mathematical foundation – especially for the marked multivariate case. For a textbook reference that covers many aspects of the Hawkes process; see [7].

To the best of our knowledge, the close connection between INAR and Hawkes processes has not been studied before. The correspondence between the model classes becomes even more direct if one applies infinite autoregression instead of finite autoregression for the time series model. This was our main motivation for generalizing the existing  $\text{INAR}(p)$  framework with  $p < \infty$  to the case  $p = \infty$ . For the new  $\text{INAR}(\infty)$  model, we give an explicit construction and show uniqueness; see Theorem 1. Then we derive three alternative descriptions of the process, namely a branching-process, an autoregressive, and a moving-average representation. Furthermore, we calculate basic quantities such as the joint moment-generating function and the autocovariance function. These are mainly presented for comparison with their Hawkes process counterparts. We observe that the equivalent branching-construction of INAR and Hawkes models form the core of the connection. This equivalence yields corresponding equations for generating functions, similar moment structures, and analogous stability criteria. Theorem 2 establishes a convergence result. In this theorem, for a given Hawkes process  $N$ , we construct a specific family of  $\text{INAR}(\infty)$  sequences  $\left\{ \left( X_n^{(\Delta)} \right)_{n \in \mathbb{Z}} \right\}_{\Delta > 0}$ . From each member of this family, we derive a point process  $N^{(\Delta)}$  by setting

$$N^{(\Delta)}((a, b]) := \sum_{n: \Delta n \in (a, b]} X_n^{(\Delta)}, \quad a < b.$$

The theorem states that  $N^{(\Delta)}$  converges weakly to the Hawkes process  $N$  when  $\Delta$  goes to zero. This result is relevant for applications of INAR and Hawkes processes. In particular, the convergence theorem yields an estimation method for the Hawkes process by estimating the more tractable approximating INAR model instead. We work out this estimation method in [15]. Moreover, from a purely theoretical point of view, the presented line of thought is useful: the time series perspective on point processes as well as the point process perspective on (integer-valued) time series can be fertile for constructing and understanding event-data models; see Section 4.

The paper is organized as follows: Section 1 introduces the INAR( $\infty$ ) model. Section 2 presents the Hawkes process. Section 3 establishes the convergence theorem. Furthermore, it collects structural analogies between the two model classes. In the final section, we discuss the broader interpretation of the INAR–Hawkes relation.

## 1. The INAR( $\infty$ ) model

Throughout the paper, we consider a basic probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  carrying all random variables involved.

### 1.1. Definition and existence

**Definition 1.** For an  $\mathbb{N}_0$ -valued random variable  $Y$  and a constant  $\alpha \geq 0$ , the *reproduction operator*  $\circ$  is defined by

$$\alpha \circ Y := \sum_{n=1}^Y \xi_n^{(\alpha)},$$

where  $\xi_n^{(\alpha)} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\alpha)$ ,  $n \in \mathbb{N}$ , independently of  $Y$ . We refer to  $\xi_n^{(\alpha)}$ ,  $n \in \mathbb{N}$ , as *offspring variables* and to  $(\xi_n^{(\alpha)})$  as *offspring sequence*.

Here and throughout the paper, we use the convention that  $\sum_{n=p}^q a_n := 0$ ,  $q < p$ , for any sequence  $(a_n)_{n \in \mathbb{Z}} \subset \mathbb{R}$ .

**Definition 2.** For  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}_0$ , let  $\varepsilon_n \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\alpha_0)$ ,  $n \in \mathbb{Z}$ , and  $\xi_l^{(n,k)} \sim \text{Pois}(\alpha_k)$ , independently over  $n \in \mathbb{Z}$ ,  $k \in \mathbb{N}$ ,  $l \in \mathbb{N}$ , and also independent of  $(\varepsilon_n)$ . An *integer-valued autoregressive time series of infinite order* (INAR( $\infty$ )) is a sequence of random variables  $(X_n)_{n \in \mathbb{Z}}$  which is a solution to the system of stochastic difference equations

$$\varepsilon_n = X_n - \sum_{k=1}^{\infty} \alpha_k \circ X_{n-k} \quad (1)$$

$$:= X_n - \sum_{k=1}^{\infty} \sum_{l=1}^{X_{n-k}} \xi_l^{(n,k)}, \quad n \in \mathbb{Z}. \quad (2)$$

We call  $\alpha_0$  *immigration parameter*,  $(\varepsilon_n)$  *immigration sequence*,  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}$ , *reproduction coefficients*, and  $K := \sum_{k=1}^{\infty} \alpha_k$  *reproduction mean*.

In most situations, it is enough to use the reproduction notation from (1) in Definition 2 without explicitly writing out the offspring sequences as in (2) – keeping in mind that each “ $\circ$ ” operates independently over  $k \in \mathbb{N}$  and  $n \in \mathbb{Z}$ . Clearly, Definitions 1 and 2 depend on the choice of the distribution of the offspring variables. A more obvious option would have been sequences of Bernoulli variables. This would yield the binomial thinning operator from [24] which has in fact been the choice in the cited INAR( $p$ ) literature. The Poisson choice for the offspring sequences, however, again yields a Poisson distribution for “ $X_n | X_{n-1}, X_{n-2}, \dots$ ”. This in turn leads to formulas that are simpler and that can be compared with their Hawkes counterparts more directly. We will address this issue in Sections 3.4 and 4. For the following existence theorem, any  $\mathbb{N}_0$ -valued distribution with finite first moments would do for the offspring variables. Throughout our paper, “stationary” is understood as “strictly stationary”.

**Theorem 1.** Let  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}_0$ , with reproduction mean  $K := \sum_{k=1}^{\infty} \alpha_k < 1$ . Then (2) has an almost surely unique stationary solution  $(X_n)_{n \in \mathbb{Z}}$ , where  $X_n \in \mathbb{N}_0$ ,  $n \in \mathbb{Z}$ , and  $\mathbb{E}X_n \equiv \alpha_0/(1 - K)$ .

**Proof.** See Appendix A.1.  $\square$

## 1.2. Branching structure

We highlight the branching nature of the solution to (2). As we will see, one can interpret this solution as a model for the size of a population, where each individual is alive exactly during one time-step. Each individual is either an immigrant or stems from a prior individual. This is similar to a Galton–Watson framework with immigration; see Section 5 in [23]. In contrast to the Galton–Watson setup, each INAR( $\infty$ ) individual does not only have offspring at the next time-step but (potentially) at any future time. The proof of Theorem 1 formalizes this structure and then establishes that the construction indeed yields a process with the desired properties. The next proposition summarizes the branching representation of the INAR( $\infty$ ) process. We emphasize the branching intuition of family processes consisting of generation processes by the unusual but suggestive notation for stochastic processes  $(F_n)$  and  $(G_n)$ . The branching formulation will be useful for the derivation of the moment-generating function. It furthermore summarizes the most elegant and, at the same time, most efficient way, for simulating from the INAR( $\infty$ ) model:

**Proposition 1.** Let  $(X_n)$  be an INAR( $\infty$ ) sequence with respect to an immigration parameter  $\alpha_0 > 0$  and reproduction coefficients  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}$ , so that  $\sum_{k=1}^{\infty} \alpha_k < 1$ ; see Definition 2. Then

$$X_n \stackrel{d}{=} \sum_{i \in \mathbb{Z}} \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)}, \quad n \in \mathbb{Z}, \quad (3)$$

where  $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\alpha_0)$ ,  $i \in \mathbb{Z}$ , and  $(F_n^{(i,j)})$  are independent (over  $i \in \mathbb{Z}$  and  $j \in \mathbb{N}$ ) copies of a branching process  $(F_n)$  defined by

$$F_n := \sum_{g=0}^{\infty} G_n^{(g)}, \quad n \in \mathbb{Z}. \quad (4)$$

The generations  $(G_n)$  in (4) are constructed recursively by

$$G_n^{(0)} := 1_{\{n=0\}} \quad \text{and} \quad G_n^{(g)} := \sum_{k=1}^n \alpha_k \circ G_{n-k}^{(g-1)} := \sum_{k=1}^n \sum_{m=1}^{G_{n-k}^{(g-1)}} \xi_m^{(n,k,g)}, \quad n \in \mathbb{Z}, \quad g \in \mathbb{N}, \quad (5)$$

with  $\xi_m^{(n,k,g)} \sim \text{Pois}(\alpha_k)$  independently over  $m, n, k, g$ —and also independent of  $(\varepsilon_i)_{i \in \mathbb{Z}}$ . Furthermore, we have the following distributional equality for the generic family-process  $(F_n)$  from (4):

$$(F_n)_{n \in \mathbb{Z}} \stackrel{d}{=} \left( 1_{\{n=0\}} + \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} F_{n-i}^{(i,j)} \right)_{n \in \mathbb{Z}}. \quad (6)$$

**Proof.** See Appendix A.2.  $\square$

### 1.3. Moment generating function

From the representation of the  $\text{INAR}(\infty)$  sequence as a superposition of shifted i.i.d. family processes in [Proposition 1](#), we derive equations for the joint moment-generating function of the model. First, we fix some notation:

**Definition 3.** For any sequence  $(t_n)_{n \in \mathbb{N}_0} \subset \mathbb{R}$ , let  $\text{supp}((t_n)) := \sup\{n \in \mathbb{N}_0 : t_n \neq 0, t_k = 0, k > n\}$  be the support of the sequence  $(t_n)$ . Furthermore, for  $A \subset \mathbb{R}$ , let  $c_{00}(A) := \{(t_n)_{n \in \mathbb{N}_0} \subset A : \text{supp}((t_n)) < \infty\}$ , the space of sequences in  $A$  with a finite number of nonzero values. For any time series  $(Y_n)_{n \in \mathbb{N}_0}$ , we define the joint moment-generating function

$$M_{(Y_n)}((t_n)_{n \in \mathbb{N}_0}) := \mathbb{E} \exp \left\{ \sum_{n=0}^{\infty} t_n Y_n \right\}, \quad (t_n) \in c_{00}(\mathbb{R}). \quad (7)$$

The somewhat unusual definitions above have been chosen for most direct comparison between the  $\text{INAR}(\infty)$  joint moment-generating function and the Laplace functional of a Hawkes process; see [Propositions 7](#) and [9](#).

**Proposition 2.** Let  $(X_n)$  be an  $\text{INAR}(\infty)$  sequence with respect to immigration parameter  $\alpha_0 \geq 0$  and reproduction coefficients  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}$ , such that  $K = \sum_{k=1}^{\infty} \alpha_k < 1$ ; see [Definition 2](#). Then there exists a constant  $\delta > 0$  such that

$$M_{(X_n)}((t_n)) \leq \exp \left\{ d \alpha_0 \frac{1+K}{2K} \right\} < \infty, \quad (t_n) \in c_{00}((-\infty, \delta]), \quad (8)$$

where  $d := \text{supp}((t_n)) + 1$  is the maximal number of nonzero values of  $(t_n)$ . Furthermore, we have that

$$M_{(X_n)}((t_n)) = \exp \left\{ \alpha_0 \sum_{i \in \mathbb{Z}} \left( M_{(F_n)}((t_{i+n})_{n \in \mathbb{N}_0}) - 1 \right) \right\}, \quad (9)$$

where we set  $t_m := 0$  for  $m < 0$ . Here,  $(F_n)$  denotes the generic family-process from [Proposition 1](#). Its joint moment-generating function  $M_{(F_n)}$  is the unique solution to

$$M_{(F_n)}((s_n)) = e^{s_0} \exp \left\{ \sum_{k=1}^{\infty} \alpha_k \left( M_{(F_n)}((s_{n+k})_{n \in \mathbb{N}}) - 1 \right) \right\}, \quad (s_n) \in c_{00}((-\infty, \delta]). \quad (10)$$

**Proof.** See [Appendix A.3](#).  $\square$

As a matter of fact, every moment-generating function that is finite in a neighborhood of zero has a Taylor series about zero. The coefficients of this series are the joint moments. Consequently, from [Proposition 2](#), we obtain

**Corollary 1.** Let  $(X_n)$  be an  $\text{INAR}(\infty)$  sequence as in [Theorem 1](#). For  $m \in \mathbb{N}$  and  $k_1, \dots, k_m \in \mathbb{N}_0$ , we have that

$$\mathbb{E} \left[ X_1^{k_1} X_2^{k_2} \dots X_m^{k_m} \right] < \infty.$$

#### 1.4. ARMA representations

We represent the INAR( $\infty$ ) model as both, an autoregressive as well as a moving-average time series. These explicit representations allow the application of standard linear time-series theory on the INAR( $\infty$ ) model.

**Proposition 3.** Let  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}_0$ , with  $K = \sum_{k=1}^{\infty} \alpha_k < 1$ , and let  $(X_n)$  be the corresponding INAR( $\infty$ ) process; see [Definition 2](#). Then

$$u_n := X_n - \sum_{k=1}^{\infty} \alpha_k X_{n-k} - \alpha_0, \quad n \in \mathbb{Z}, \quad (11)$$

defines a stationary sequence  $(u_n)$  with  $\mathbb{E}u_n \equiv 0$ ,  $n \in \mathbb{Z}$ , and

$$\mathbb{E}[u_n u_{n'}] = \begin{cases} 0, & n \neq n', \\ \frac{\alpha_0}{1-K}, & n = n'. \end{cases} \quad (12)$$

Furthermore, we have that

$$(X_n - \mu_X) - \sum_{k=1}^{\infty} \alpha_k (X_{n-k} - \mu_X) = u_n, \quad n \in \mathbb{Z}, \quad m < 0, \quad (13)$$

where  $\mu_X := \mathbb{E}X_0 = \alpha_0/(1-K)$ . In other words,  $(u_n)$  is a (dependent) white-noise sequence and the time series  $(X_n - \mu_X)_{n \in \mathbb{Z}}$  can be described in terms of a solution to an ordinary standard AR( $\infty$ ) system of difference equations.

**Proof.** See [Appendix A.4](#).  $\square$

From [Proposition 3](#), we find the Wold decomposition of the INAR( $\infty$ ), that is, a representation as a standard MA( $\infty$ ) time series. It will be most helpful for establishing the second-order properties of the process.

**Proposition 4.** Let  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}_0$ , with  $\sum_{k=1}^{\infty} \alpha_k < 1$ . Then the corresponding INAR( $\infty$ ) process from [Definition 2](#) is a solution to the family of equations

$$X_n - \mu_X = \sum_{k=0}^{\infty} \beta_k u_{n-k}, \quad n \in \mathbb{Z}, \quad (14)$$

where  $(u_n)$  is the white-noise sequence from [Proposition 3](#),  $\mu_X := \mathbb{E}X_0 = \alpha_0/(1-K)$ ,  $\beta_0 := 1$ , and  $\beta_k := \sum_{i=1}^k \alpha_i \beta_{k-i}$ ,  $k \in \mathbb{N}$ . Furthermore,  $\beta_k \geq 0$ ,  $k \in \mathbb{N}_0$ , and  $\sum_{k=0}^{\infty} \beta_k = 1/(1-K) < \infty$ .

**Proof.** See [Appendix A.5](#).  $\square$

Note that the moving-average coefficients  $(\beta_k)_{k \in \mathbb{N}_0}$  defined in [Proposition 4](#) correspond to  $(\mathbb{E}F_k)_{k \in \mathbb{N}_0}$ , the expected values of the family-process (4) after  $k = 0, 1, 2, \dots$  time steps. Also note that [Proposition 4](#) is not a corollary of the standard result stating that ARMA( $p, q$ ) processes ( $p, q < \infty$ ) are MA( $\infty$ ) processes; for example, see Theorem 3.1.1 of [6]. The argumentation for the case  $p = \infty$  has to be more subtle: we have to prove that the (in general infinitely many) zeros of the involved power series can be bounded away from the unit circle.

### 1.5. Second order properties

The  $\text{MA}(\infty)$  representation makes the second moments of the  $\text{INAR}(\infty)$  sequence particularly tractable:

**Proposition 5.** *Let  $(X_n)$  be an  $\text{INAR}(\infty)$  process with reproduction coefficients  $\alpha_k \geq 0$ ,  $k \in \mathbb{N}$ , such that  $K := \sum_{k=1}^{\infty} \alpha_k < 1$ , and immigration parameter  $\alpha_0 \geq 0$ . Furthermore, let  $R(j) := \text{Cov}(X_n, X_{n+j})$ ,  $j \in \mathbb{Z}$ , be the autocovariance function of the (stationary) sequence. Then*

$$R(j) = \frac{\alpha_0}{1-K} \sum_{k=0}^{\infty} \beta_k \beta_{k+|j|} \geq 0, \quad j \in \mathbb{Z}, k \in \mathbb{N} \quad (15)$$

where  $\beta_0 := 1$  and  $\beta_k := \sum_{i=1}^k \alpha_i \beta_{k-i}$ . In addition, we have that

$$\sum_{j=0}^{\infty} R(j) \leq \frac{\alpha_0}{(1-K)^3} < \infty. \quad (16)$$

**Proof.** See [Appendix A.6](#).  $\square$

## 2. The Hawkes process

After the first section on the new  $\text{INAR}(\infty)$  model, the following shorter section formally presents the well-known Hawkes point process. We treat point processes as random counting-measures and only consider point processes on  $\mathbb{R}$ . First, we fix some general notation and terminology. Then we recall the definition, the existence theorem, and selected properties of the (univariate) Hawkes process. For the general theory, we mainly follow Resnick [22, Chapter 3]. For the Hawkes part, our main references are the seminal papers Hawkes [11,12] and Hawkes [13].

### 2.1. Preliminaries

Let  $\mathcal{B} := \mathcal{B}(\mathbb{R})$  be the Borel-sets in  $\mathbb{R}$  and  $\mathcal{B}_b := \{B \in \mathcal{B}(\mathbb{R}) : B \text{ bounded}\}$ . A measure  $m$  on  $\mathbb{R}$  is a *point measure* if  $m(B) \in \mathbb{N}_0$ ,  $B \in \mathcal{B}_b$ . We denote the space of point measures on  $\mathbb{R}$  by  $M_p := M_p(\mathbb{R})$ . Let  $C_K^+ := C_K^+(\mathbb{R})$  be the space of nonnegative continuous functions on  $\mathbb{R}$  with compact support. Point measures  $(m_n)$  converge vaguely to a point measure  $m$  if  $\lim_{n \rightarrow \infty} \int f(t) m_n(dt) \rightarrow \int f(t) m(dt)$ ,  $f \in C_K^+(\mathbb{R})$ ; we write  $m_n \xrightarrow{v} m$ . Vague convergence yields the *vague topology* on  $M_p$ . The Borel  $\sigma$ -algebra generated by this topology,  $\mathcal{M}_p := \mathcal{B}(M_p)$ , coincides with the  $\sigma$ -algebra generated by the sets  $\{m \in M_p : m(A) = k\}$ ,  $A \in \mathcal{B}_b$ ,  $k \in \mathbb{N}_0$ ; see Lemma 1.4. in [14]. Any measurable mapping  $\Phi : (M_p, \mathcal{M}_p) \rightarrow (\mathbb{R}, \mathcal{B})$  such that  $\lim_{n \rightarrow \infty} \Phi(m_n) = \Phi(m)$  whenever  $m_n \xrightarrow{v} m$ , is *continuous with respect to the vague topology*. Our basic underlying probability space is  $(\Omega, \mathcal{F}, \mathbb{P})$ . A measurable mapping  $N : (\Omega, \mathcal{F}) \rightarrow (M_p, \mathcal{M}_p)$ ,  $\omega \mapsto N(\omega)$  is called *point process*. The *history* of a point process  $N$  is the filtration  $(\mathcal{H}_t^N)$ , where, for  $t \in \mathbb{R}$ ,

$$\mathcal{H}_t^N := \sigma \left( \left\{ \omega \in \Omega : N(\omega)((a, b]) \in B \right\} : -\infty < a < b \leq t, B \subset \mathbb{N}_0 \right). \quad (17)$$

We assume that  $\mathcal{H}_t^N \subset \mathcal{F}$ ,  $t \in \mathbb{R}$ . Note that our definition of a point process allows multiple points, i.e., we may have that “ $\mathbb{P}[N(\{t\}) > 1 | N(\{t\}) > 0] > 0$ ”. Also note that, for  $t \in \mathbb{R}$ , the  $\sigma$ -algebra  $\mathcal{H}_t^N$  includes all sets of the form

$$\left\{ \omega : N(\omega)(\{t_n\}) = k_n, N(\omega)((t_{n-1}, t_n)) = 0, n = 0, -1, -2, \dots \right\}$$

with  $t =: t_0 \geq t_{-1} \geq \dots$  and  $k_0, k_{-1}, \dots \in \mathbb{N}$ . (18)

## 2.2. Definition and existence

**Definition 4.** For any point measure  $m \in M_p$ , define the *Hawkes intensity*

$$\lambda(t|m) := \eta + \int_{\mathbb{R}} h(t-s)m(ds), \quad t \in \mathbb{R},$$

where  $\eta > 0$  is a constant and  $h : \mathbb{R} \rightarrow \mathbb{R}_0^+$  is a nonnegative measurable function with  $h(t) = 0$ ,  $t \leq 0$ . We refer to  $\eta$  as *immigration intensity* and to  $h$  as *reproduction intensity*.

The immigration intensity is often called *baseline intensity* and the reproduction intensity is often called *excitement function*. However, our objective is to highlight the similarity between Hawkes and INAR processes. Consequently, we make use of a joint branching-process terminology; see [Definition 2](#).

**Definition 5.** Let  $\lambda$  be a Hawkes intensity as in [Definition 4](#). A *Hawkes process* is a point process  $N$  that is a solution to the family of equations

$$\mathbb{E} \left[ 1_A N((a, b]) \right] = \mathbb{E} \left[ 1_A \int_a^b \lambda(s|N) ds \right], \quad a < b, A \in \mathcal{H}_a^N. \quad (19)$$

A priori, it is not clear whether this family of equations has a solution, whether any possible solution would be unique (in a distributional sense), and whether the distribution of a solution would be stationary. These questions are answered by the following proposition; see [\[13\]](#):

**Proposition 6.** Let  $\lambda$  be a Hawkes intensity with immigration intensity  $\eta > 0$  and reproduction intensity  $h$  such that  $\int_0^\infty h(t)dt < 1$ . Then there is precisely one stationary process that satisfies [\(19\)](#).

The existence and uniqueness result above is established by the observation that the solution to [\(19\)](#) must be a cluster process or – more specifically – a branching process with immigration: the points are interpreted as individuals that are either immigrants or offspring. The immigrants (or cluster centers) stem from a homogeneous Poisson process with intensity  $\eta$ . These immigrants form generation zero of the following branching procedure: an immigrant at time  $s \in \mathbb{R}$  triggers an inhomogeneous Poisson process with intensity  $h(\cdot - s)$  where  $h$  is the reproduction intensity of the process as in [Definition 4](#). These offspring individuals form generation one. Each of these first-generation individuals again triggers an inhomogeneous Poisson process in a similar way, etc., so that the families (or clusters) are generated by cascades of inhomogeneous Poisson processes.

## 2.3. The Laplace functional

The cluster and branching process point of view is also fertile beyond the results of [Proposition 6](#). For example, it leads to equations for the *Laplace functional* of a Hawkes



process. The Laplace functional  $\Psi_N$  of a point process  $N$  is a functional defined on the space of nonnegative measurable functions with compact support by

$$\Psi_N[f] := \mathbb{E} \exp \left\{ - \int_{-\infty}^{\infty} f(t) N(dt) \right\}.$$

The next proposition is Theorem 2 in [13] – with a slight modification as the original statement refers to the probability generating functional whereas we prefer the nowadays more common Laplace functional notion:

**Proposition 7.** *Let  $N$  be a Hawkes process with immigration intensity  $\eta > 0$  and reproduction intensity  $h$  as in Proposition 6. Then the Laplace functional  $\Psi_N$  of  $N$  is*

$$\Psi_N[f] = \exp \left\{ \eta \int_{-\infty}^{\infty} \left( \Psi_F[f(t + \cdot)] - 1 \right) dt \right\},$$

where  $\Psi_F$  is a functional that is the unique solution to

$$\Psi_F[f] = e^{-f(0)} \exp \left\{ \int_0^{\infty} h(t) \left( \Psi_F[f(t + \cdot)] - 1 \right) dt \right\}.$$

In both equalities,  $f$  denotes an arbitrary measurable, nonnegative function with compact support.

### 3. Links between INAR( $\infty$ ) and Hawkes processes

In the following section, we first explain how discrete-time INAR( $\infty$ ) processes can approximate a continuous-time Hawkes process. After the convergence theorem, we establish a number of properties of the approximating sequence. Finally, we collect some structural analogies of the two models.

#### 3.1. Preliminaries

Let  $Y_n$ ,  $n \in \mathbb{N}$ , and  $Y$  be random variables with values in some topological space. The sequence  $(Y_n)$  converges weakly to  $Y$  if  $\lim_{n \rightarrow \infty} \mathbb{E} \varphi(Y_n) = \mathbb{E} \varphi(Y)$  for all nonnegative continuous bounded functions  $\varphi$ . We define weak convergence of point processes with respect to the vague topology  $\mathcal{M}_p$  on  $M_p$ ; see Section 2.1. In this case, weak convergence of point processes is equivalent to convergence of their finite-dimensional distributions; see [7, Theorem 11.1.VII]. General weak convergence theory, as developed in the monograph Billingsley [2], considers sequences in metric spaces. Therefore it is important to note that the vague topology is metrizable; see [22, Proposition 3.17]. In other words, we may treat  $(M_p, \mathcal{M}_p)$  as a metric space where necessary. A most helpful theorem in the weak-convergence context is the continuous mapping theorem; see Theorem 5.1 in [2]. We apply it in the following form:

**Proposition 8.** *Let  $(N_n)$  and  $N$  be point processes such that  $N_n \xrightarrow{w} N$ ,  $n \rightarrow \infty$ . Furthermore, let  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  be a bounded, measurable function with compact support and with a set of discontinuities  $D_f \in \mathcal{B}$  such that  $\mathbb{P}[N(D_f) > 0] = 0$ . Then,  $\int f(t) N_n(dt) \xrightarrow{w} \int f(t) N(dt)$  for  $n \rightarrow \infty$ .*

### 3.2. The convergence theorem

Next to the conditions on the reproduction intensity  $h$  from [Definition 5](#) and [Proposition 6](#), we introduce an additional assumption: we want  $h$  piecewise continuous. We say a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is *piecewise continuous* if its set of discontinuities  $D_f \subset \mathbb{R}$  is finite and for all  $t_0 \in D_f$  the limits  $\lim_{t \rightarrow t_0^-} f(t)$  and  $\lim_{t \rightarrow t_0^+} f(t)$  exist and are finite. Combining all assumptions on  $h$  yields the following important technical

**Lemma 1.** *Let  $h : \mathbb{R} \rightarrow \mathbb{R}_0^+$  be a piecewise continuous function with  $h(t) = 0$ ,  $t \leq 0$ , and  $\int h(t) dt < 1$ . Then there exist constants  $\delta > 0$  and  $\tilde{K} < 1$  such that, for any  $\Delta \in (0, \delta)$ ,*

$$K^{(\Delta)} := \Delta \sum_{k=1}^{\infty} h(k\Delta) \leq \tilde{K} < 1. \quad (20)$$

In the sequel of the section, let  $\delta > 0$ ,  $K^{(\Delta)} \leq \tilde{K} < 1$ , and  $\Delta \in (0, \delta)$ , be as in the lemma above. We state the main mathematical result of our paper:

**Theorem 2.** *Let  $N$  be a Hawkes process with immigration intensity  $\eta$  and reproduction intensity  $h$  as in [Lemma 1](#). For  $\Delta \in (0, \delta)$ , let  $(X_n^{(\Delta)})$  be an INAR( $\infty$ ) sequence with immigration parameter  $\Delta\eta$  and reproduction coefficients  $\Delta h(k\Delta)$ ,  $k \in \mathbb{N}$ . From the sequences  $\{(X_n^{(\Delta)})\}_{\Delta \in (0, \delta)}$ , we define a family of point processes by*

$$N^{(\Delta)}(A) := \sum_{k:k\Delta \in A} X_k^{(\Delta)}, \quad A \in \mathcal{B}, \Delta \in (0, \delta). \quad (21)$$

Then, we have that

$$N^{(\Delta)} \xrightarrow{w} N \quad \text{for } \Delta \rightarrow 0.$$

**Proof.** See [Appendix A.11](#).  $\square$

Our proof uses the standard weak-convergence approach – as followed in the Hawkes context, e.g., by Brémaud and Massoulié [\[5\]](#). First, tightness of the approximating family is established. By Prokhorov's theorem, tightness yields weak subsequential limits for all subsequences. Then we show that all those potential weak subsequential limits have the same distribution as the Hawkes process. This will establish the result. An alternative approach would be convergence of Laplace functionals that are given by

**Proposition 9.** *For some  $\Delta \in (0, \delta)$ , let  $N^{(\Delta)}$  be as in [Theorem 2](#). Let  $f$  be a nonnegative measurable function with compact support. Then the Laplace functional of  $N^{(\Delta)}$  evaluated at  $f$  is*

$$\Psi_{N^{(\Delta)}}[f] = \exp \left\{ \Delta \eta \sum_{i \in \mathbb{Z}} \left( \Psi_{F^{(\Delta)}}^{(\Delta)} \left( (f((i+n)\Delta))_{n \in \mathbb{N}_0} \right) - 1 \right) \right\}.$$

Here, the function  $\Psi_{F^{(\Delta)}}^{(\Delta)}$  operating on sequences  $(s_n)_{n \in \mathbb{N}_0} \in c_{00}([0, \infty))$  is a solution to

$$\Psi_{F^{(\Delta)}}^{(\Delta)}((s_n)_{n \in \mathbb{N}_0}) = e^{-s_0} \exp \left\{ \sum_{k=1}^{\infty} \Delta h(k\Delta) \left( \Psi_{F^{(\Delta)}}^{(\Delta)}((s_{k+n})_{n \in \mathbb{N}_0}) - 1 \right) \right\}.$$

**Proof.** See A.7.  $\square$

The similarities between the formulas in Proposition 9 and the corresponding equations for the Hawkes process in Proposition 7 are striking. Still, rather than establishing the convergence result from Theorem 2 via the Laplace functionals, we choose a more direct reasoning on the process level that contains useful information on the approximating point process family as a by-product. The information necessary for the convergence proof is collected in the following lemmas:

**Lemma 2.** For any  $\Delta \in (0, \delta)$ , let  $N^{(\Delta)}$  be a point process as in Theorem 2. Then, for  $A \in \mathcal{B}$ , we have that

$$A \cap \{k\Delta : k \in \mathbb{Z}\} = \emptyset \Rightarrow N^{(\Delta)}(A) = 0, \text{ almost surely.} \quad (22)$$

For the expectation, we find that

$$\mathbb{E}N^{(\Delta)}(\{k\Delta\}) = \Delta \frac{\eta}{1 - K^{(\Delta)}} < \delta \frac{\eta}{1 - \tilde{K}}, \quad k \in \mathbb{Z}, \quad (23)$$

and, for  $a < b$ ,

$$\mathbb{E}N^{(\Delta)}([a, b]) < (b - a + 2\delta) \frac{\eta}{1 - \tilde{K}} < \infty. \quad (24)$$

**Proof.** See Appendix A.8.  $\square$

**Lemma 3.** For  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$ , the approximating family of point processes from Theorem 2, we have that

$$\sup_{\Delta \in (0, \delta)} \text{Var}(N^{(\Delta)}(A)) < \infty, \quad A \in \mathcal{B}_b.$$

**Proof.** See Appendix A.9.  $\square$

A family of random variables  $(Y_i)_{i \in I}$  is uniformly integrable if  $\lim_{M \rightarrow \infty} \sup_{i \in I} \mathbb{E}[1_{|Y_i| > M} |Y_i|] = 0$ . We obtain uniform integrability of the random variables in question as a corollary from Lemma 3:

**Lemma 4.** Let  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$  be the approximating family of point processes from Theorem 2 and  $A \in \mathcal{B}_b$ . Then we have that the family of random variables  $(N^{(\Delta)}(A))_{\Delta \in (0, \delta)}$  is uniformly integrable.

A family of probability measures  $(\mathbb{P}^{(i)})_{i \in I}$  on  $(M_p, \mathcal{M}_p)$  is uniformly tight if, for all  $\varepsilon > 0$ , there exists a compact set  $K \subset M_p$  such that  $\mathbb{P}^{(i)}[K^c] < \varepsilon$ ,  $i \in I$ .

**Lemma 5.** The family of the probability measures  $(\mathbb{P}^{(\Delta)})_{0 < \Delta < \delta}$  on  $(M_p, \sigma(\mathcal{M}_p))$  corresponding to the point processes  $(N^{(\Delta)})_{0 < \Delta < \delta}$  is uniformly tight.

**Proof.** See Appendix A.10.  $\square$

### 3.3. Structural analogies

Besides the formal convergence result from Theorem 2, we point out a number of structural parallels between the Hawkes and the INAR( $\infty$ ) model. The branching structure of the INAR( $\infty$ )

model described in Section 1.2 is the same as the branching structure of the Hawkes process described after Proposition 6. This similar underlying structure yields analogous equations for the moment-generating function of the INAR( $\infty$ ) and the Laplace functional of the Hawkes process; see Propositions 2 and 7. Consequently, we can expect similar distributional properties. In the following, we compare the models more directly by considering a specific Hawkes process  $N$  together with its approximating family of INAR( $\infty$ ) sequences,  $(X^{(\Delta)})$ ,  $\Delta \in (0, \delta)$ , obtained from Theorem 2.

- (i) The defining Eqs. (2) and (19) have similar interpretations: taking expectations conditional on  $\mathcal{H}_{n-1}^{X^{(\Delta)}} := \sigma(X_k^{(\Delta)} : k \leq n-1)$  on both sides of (2) yields

$$\frac{\mathbb{E}[X_n^{(\Delta)} | \mathcal{H}_{n-1}^{X^{(\Delta)}}]}{\Delta} = \eta + \sum_{k=-\infty}^{n-1} h((n-k)\Delta) X_k^{(\Delta)}, \quad n \in \mathbb{Z},$$

which is similar to the local version of (19)

$$\frac{\mathbb{E}[N(dt) | \mathcal{H}_t^N]}{dt} = \eta + \int_{-\infty}^t h(t-s) N(ds), \quad t \in \mathbb{R}.$$

- (ii) The stability criteria from Theorem 1 and Proposition 6 correspond: for the time series case,  $\sum_{k=1}^{\infty} \Delta h(\Delta k) < 1$  is a sufficient existence condition which in the Hawkes case becomes  $\int_0^{\infty} h(t) dt < 1$ . Brémaud and Massoulié [5] establishes the existence of a nontrivial Hawkes process, where the total weight of the reproduction intensity equals 1 and the immigration intensity is zero. The analogous statement for INAR( $\infty$ ) sequences can be derived in a similar way.
- (iii) From the correspondence of the generating functions, we know that the moments of the INAR( $\infty$ ) and the Hawkes process must be similar. The first moments are

$$\frac{\mathbb{E}X_n^{(\Delta)}}{\Delta} = \frac{\eta}{1 - K^{(\Delta)}}, \quad \text{respectively,} \quad \frac{\mathbb{E}N(dt)}{dt} = \frac{\eta}{1 - K}.$$

For the first equality see Theorem 1; for the second equality see [11].

- (iv) For the autocovariances of both models, i.e., for

$$R^{(\Delta)}(n) := \frac{\mathbb{E}[X_0^{(\Delta)} X_n^{(\Delta)}]}{\Delta^2} - \left( \frac{\mathbb{E}X_0^{(\Delta)}}{\Delta} \right)^2, \quad n \in \mathbb{Z},$$

respectively, for

$$r(t) := \frac{\mathbb{E}[dN_0 dN_t]}{(dt)^2} - \left( \frac{\mathbb{E}N(dt)}{dt} \right)^2, \quad t \in \mathbb{R},$$

we find at the origin

$$R^{(\Delta)}(0) = \frac{1}{\Delta} \frac{\eta}{1 - K^{(\Delta)}} + \sum_{k=1}^{\infty} \Delta h(k\Delta) R^{(\Delta)}(k), \quad (25)$$

respectively,

$$r(0) = \frac{1}{dt} \frac{\eta}{1 - K} + \int_{0+}^{\infty} h(s) r(s) ds. \quad (26)$$

Implicit equations of Yule–Walker type are valid in both cases:

$$R^{(\Delta)}(n) = \sum_{k=1}^{\infty} h(\Delta k) R^{(\Delta)}(|n| - k) \Delta, \quad n \neq 0, \quad (27)$$

respectively,

$$r(t) = \int_0^{\infty} h(s) r(|t| - s) ds, \quad t \neq 0. \quad (28)$$

Eqs. (25) and (27) are standard facts given the representation of the INAR( $\infty$ ) sequence as a standard AR( $\infty$ ) process in Proposition 3; see, e.g., [6]. Eq. (28) can be derived as in [12], where  $r$  is called “complete autocovariance density”. Eq. (26) corresponds to (10) in [12] evaluated at zero.

### 3.4. The choice of the offspring distribution

As a last remark, we again refer to the choice of the offspring distribution in Definition 1. An obvious alternative to the Poisson distribution would have been the Bernoulli distribution; see the discussion after Definition 2. With the Bernoulli choice, each individual would have not more than one offspring at each future point in time instead of potentially unboundedly many. We want to indicate that in the limit (in the sense of Theorem 2 where all reproduction coefficients go to zero) this option would yield the same result: for  $\Delta \in (0, 1)$  and  $(\alpha_k) \subset [0, 1]$  such that  $\sum \alpha_k < 1$ , let  $\xi_k^{(\Delta)} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\Delta \alpha_k)$ ,  $k \in \mathbb{N}$ , and  $\tilde{\xi}_k^{(\Delta)} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\Delta \alpha_k)$ ,  $k \in \mathbb{N}$ . Then one can easily show that

$$\lim_{\Delta \rightarrow 0^+} \frac{\mathbb{P} \left[ \xi_k^{(\Delta)} = 0 \right]}{\mathbb{P} \left[ \tilde{\xi}_k^{(\Delta)} = 0 \right]} = \lim_{\Delta \rightarrow 0^+} \frac{\mathbb{P} \left[ \xi_k^{(\Delta)} = 1 \right]}{\mathbb{P} \left[ \tilde{\xi}_k^{(\Delta)} = 1 \right]} = 1, \quad k \in \mathbb{N}.$$

So when  $\Delta$  is very small, the offspring distribution candidates, Poisson and Bernoulli, become very similar. Roughly speaking, the limiting procedure in Theorem 2 is nothing else than a (complicated) superposition of limits of the form  $\sum_{k=1}^{\infty} \xi_k^{(\Delta)}$ . For these kinds of sums, we have that

$$\lim_{\Delta \rightarrow 0^+} \frac{\mathbb{P} \left[ \sum_{k=1}^{\infty} \xi_k^{(\Delta)} = n \right]}{\mathbb{P} \left[ \sum_{k=1}^{\infty} \tilde{\xi}_k^{(\Delta)} = n \right]} = 1, \quad n \in \mathbb{N}_0.$$

If  $\alpha_k > 0$  infinitely often, then – by the Poisson limit theorem – the two considered probabilities are even equal for all  $\Delta \in [0, 1]$ . In view of the above, one can expect that Theorem 2 with Bernoulli offspring would yield the same limit as Poisson offspring, namely the Hawkes process.

## 4. Conclusion

The mathematical formulation of the correspondence between INAR and Hawkes processes in Theorem 2 has the following heuristic interpretation relevant for practical applications: let  $(N^{(\Delta)})$  be the approximating family of INAR( $\infty$ )-based point processes for a Hawkes process  $N$  as in Theorem 2. Then for  $\Delta > 0$  (small), the finite-dimensional distributions of  $N^{(\Delta)}$

are approximately equal to the finite-dimensional distributions of  $N$ . In particular, we find, for  $n \in \mathbb{N}$ ,

$$\begin{aligned} & \left( N\left((0, \Delta]\right), N\left((\Delta, 2\Delta]\right), \dots, N\left((n-1)\Delta, n\Delta]\right) \right) \\ & \stackrel{d}{\approx} \left( N^{(\Delta)}\left((0, \Delta]\right), N^{(\Delta)}\left((\Delta, 2\Delta]\right), \dots, N^{(\Delta)}\left((n-1)\Delta, n\Delta]\right) \right) \\ & = \left( X_1^{(\Delta)}, X_2^{(\Delta)}, \dots, X_n^{(\Delta)} \right), \end{aligned} \quad (29)$$

where  $(X_n^{(\Delta)})$  is the INAR( $\infty$ ) sequence that the point process  $N^{(\Delta)}$  is based on; see [Theorem 2](#). So  $(X_n^{(\Delta)})$  is an approximative model for the bin-count sequences of the considered Hawkes process  $N$ . This point of view can be very fertile. For example, it leads to a nonparametric estimation procedure for the Hawkes process. Instead of fitting a Hawkes process directly, one fits the corresponding INAR( $\infty$ ) model from [Theorem 2](#) on the bin-counts for some small  $\Delta > 0$ . Kirchner [15] gives a detailed discussion of this estimation procedure including asymptotic properties, bias issues and techniques for an optimal bin-size choice  $\Delta$ . Also note that the Hawkes bin-count sequence view in (29) on the INAR( $\infty$ ) model is another argument in favor of choosing the Poisson instead of the Bernoulli distribution for the offspring sequences: clearly a Hawkes event can have potentially more than one direct offspring event in a future time-interval.

For any  $\mathbb{N}_0$ -valued time series model, one can construct a point process model the way it is done in [Theorem 2](#). So, studying integer-valued time series can be inspiring for developing and understanding point process models. For example, one might want to consider the corresponding point process of an integer-valued autoregressive moving-average (INARMA) time series. For INARMA time series, a moving-average part is added to the autoregressive part in the defining difference Eqs. (2); see [9]. In fact, the corresponding point process is nothing else but the “dynamic contagion process” as proposed in [27]. Also for integer-valued time series theory, it is inspiring to translate point process models into the discrete-time setup. For example, one might want to translate the generalizing results on selfexciting point processes in [4] to the INAR context. Here, the selfexcitement of the point process is modeled not as an affine but as a general Lipschitz function of the past of the process. The analogous generalization in time series theory is “nonlinear Poisson autoregression” as studied in [10] for the case  $p = 1$ . In the latter paper, the authors also find a Lipschitz condition for the transfer function. Yet another idea might be to consider marked INAR sequences in analogy to marked Hawkes processes; see [18].

One can expect that the results of our paper also hold in a multivariate setup – with the obvious modifications. However, the notation would become even more tedious, so that we have decided to concentrate on the univariate case. In view of the many INAR/Hawkes-correspondences presented in this paper we conclude: INAR( $\infty$ ) sequences are discrete-time versions of Hawkes processes and, vice versa, Hawkes processes are continuous-time versions of INAR( $\infty$ ) sequences.

## Acknowledgments

M.K. takes pleasure in thanking Thomas Mikosch for his most valuable comments on an earlier version of the paper and Paul Embrechts for guiding him to the topic of Hawkes processes

and its applications to quantitative risk management. He also thanks an anonymous referee whose report improved some of the proofs as well as the final overall structure of the paper considerably. Furthermore, M.K. thanks Isabel Marquez da Silva for sharing her expertise on integer-valued time series models and Rita Kirchner as well as Anne MacKay for help with the editing. The author acknowledges financial support from RiskLab at the ETH Zurich and the Swiss Finance Institute.

## Appendix. Proofs

### A.1. Proof of [Theorem 1](#)

Let  $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(\alpha_0)$ ,  $i \in \mathbb{Z}$ , be the immigration terms from [Definition 2](#). For the  $j$ th potential immigrant at time  $i \in \mathbb{Z}$ , we define generation processes  $(G_n^{(g,i,j)})$ ,  $g \in \mathbb{N}_0$ , by the following recursive procedure:

$$G_n^{(0,i,j)} := 1_{\{n=0\}}, \quad n \in \mathbb{Z}, i \in \mathbb{Z}, j \in \mathbb{N} \quad (\text{A.1})$$

$$G_n^{(g,i,j)} := \sum_{k=1}^n \alpha_k \circ G_{n-k}^{(g-1,i,j)}. \quad (\text{A.2})$$

For all distributional properties of the construction, it will be enough to apply defining Eq. (A.2) for the generations. We will consider an explicit representation of the involved offspring sequences later. Note that  $G_n^{(g,i,j)} = 0$  whenever  $n < 0$ . A family originating from the  $j$ th individual immigrant at time  $i$  is the superposition of all the corresponding generation-processes:

$$F_n^{(i,j)} := \sum_{g=0}^{\infty} G_n^{(g,i,j)}, \quad n \in \mathbb{Z}, i \in \mathbb{Z}, j \in \mathbb{N}. \quad (\text{A.3})$$

The candidate series  $(\tilde{X}_n)$  for a solution of (2) is the superposition of all these families – modulo an appropriate shift in the time index:

$$\tilde{X}_n := \sum_{i=-\infty}^n \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)}, \quad n \in \mathbb{Z}. \quad (\text{A.4})$$

Note that only family and generation processes indexed by  $(i, j)$  with  $j \in \{1, \dots, \varepsilon_i\}$  come into play. Also note that all infinite series involved in the construction above are well-defined because their partial sums are nondecreasing. As a first step, we remind ourselves that  $K = \sum_{k=1}^{\infty} \alpha_k$  and establish

$$\mathbb{E} \sum_{n=0}^{\infty} G_n^{(g,i,j)} = K^g, \quad g \in \mathbb{N}_0, i \in \mathbb{Z}, j \in \mathbb{N}, \quad (\text{A.5})$$

by induction: for  $g = 0$ , (A.5) is correct because we have  $\mathbb{E} \sum_{n=0}^{\infty} G_n^{(0,i,j)} = \sum_{n=0}^{\infty} 1_{\{n=0\}} = 1 = K^0$ . For  $g > 0$ , one can show that  $\mathbb{E} \sum_{n=0}^{\infty} G_n^{(g,i,j)} = K \cdot \mathbb{E} \sum_{n=0}^{\infty} G_n^{(g-1,i,j)}$  and (A.5) follows. By (A.5),  $\sum_{n=0}^{\infty} F_n^{(i,j)}$  has expectation  $1/(1-K)$ . In particular,  $\sum_{n=0}^{\infty} F_n^{(i,j)}$  is almost surely finite. Since  $F_n^{(i,j)}$  is a sum of integers, we have that  $F_n^{(i,j)} \in \mathbb{N}_0$ ,  $n \in \mathbb{Z}$ , almost surely.

For  $\tilde{X}_n$ ,  $n \in \mathbb{Z}$ , we find that

$$\mathbb{E}\tilde{X}_n = \sum_{i=-\infty}^n \mathbb{E} \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)} = \alpha_0 \sum_{i=-\infty}^n \mathbb{E} F_{n-i}^{(i,1)} = \alpha_0 \mathbb{E} \sum_{i=-\infty}^n F_{n-i}^{(1,1)} = \frac{\alpha_0}{1-K}. \quad (\text{A.6})$$

Note that in the present work, nearly all random variables involved are nonnegative. Consequently, in most cases the interchange of summations as in the first equality above is covered by Tonelli's Theorem. Also Wald's Identity as applied in the second equality above will be applied without further notice. From (A.6), we get that  $\tilde{X}_n$  is almost surely finite and therefore almost surely  $\mathbb{N}_0$ -valued. Note that, by construction, the generations  $(G_n^{(i,j)})$  and therefore the families  $(F_n^{(i,j)})$  are independently and identically distributed time series over  $i \in \mathbb{Z}$  and  $j \in \mathbb{N}$ . Stationarity of  $(\tilde{X}_n)$  then follows from the i.i.d. property of the immigration sequence  $(\varepsilon_i)$ .

To show that our candidate sequence  $(\tilde{X}_n)$  indeed solves (2), we have to work with an explicit representation for the reproductions involved in (A.2). To that aim, let

$$\xi_{g,i,j,m}^{(n,k)} \sim \text{Pois}(\alpha_k), \quad \text{independently over } i, n \in \mathbb{Z} \text{ and } k, j, g, m \in \mathbb{N}. \quad (\text{A.7})$$

In our branching terminology,  $\xi_{g,i,j,m}^{(n,k)}$  denotes the number of offspring individuals at time  $n$  whose parent lived at time  $n-k$ . This parent belongs to the  $(g-1)$ -th generation of family  $(i, j)$ . Furthermore, this parent is the  $m$ th such individual. We repeat the defining recursion for the generation processes from above – this time we represent the involved offspring sequences explicitly:

$$G_n^{(0,i,j)} := 1_{\{n=0\}}, \quad n \in \mathbb{Z}, i \in \mathbb{Z}, j \in \mathbb{N} \quad (\text{A.8})$$

$$G_n^{(g,i,j)} := \sum_{k=1}^n \alpha_k \circ G_{n-k}^{(g-1,i,j)} \quad (\text{A.9})$$

$$:= \sum_{k=1}^n \sum_{m=1}^{G_{n-k}^{(g-1,i,j)}} \xi_{g,i,j,m}^{(i+n,k)}, \quad n \in \mathbb{Z}, i \in \mathbb{Z}, j \in \mathbb{N}, g \in \mathbb{N}. \quad (\text{A.10})$$

It is obvious that (A.10) justifies the distributional assumptions on (A.9) (i.e., (A.2)), used in the first part of this proof. For any  $n \in \mathbb{Z}$ , we find

$$\tilde{X}_n = \sum_{i=-\infty}^n \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)} = \sum_{i=-\infty}^{n-1} \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)} + \sum_{j=1}^{\varepsilon_n} F_0^{(n,j)} = \sum_{i=-\infty}^{n-1} \sum_{j=1}^{\varepsilon_i} \sum_{g=0}^{\infty} G_{n-i}^{(g,i,j)} + \varepsilon_n. \quad (\text{A.11})$$

Note that the third summation really starts in  $g=1$  because  $G_{n-i}^{(0,i,j)} = 1_{\{n-i=0\}} = 0$  whenever  $i \leq n-1$ . For the triple sum in (A.11), we obtain

$$\begin{aligned} \sum_{i=-\infty}^{n-1} \sum_{j=1}^{\varepsilon_i} \sum_{g=1}^{\infty} G_{n-i}^{(g,i,j)} &= \sum_{i=-\infty}^{n-1} \sum_{j=1}^{\varepsilon_i} \sum_{g=1}^{\infty} \sum_{k=1}^{n-i} G_{n-k-i}^{(g-1,i,j)} \sum_{m=1}^{G_{n-k-i}^{(g-1,i,j)}} \xi_{g,i,j,m}^{(i+n-k,k)} \\ &= \sum_{k=1}^{\infty} \left( \sum_{i=-\infty}^{n-k} \sum_{j=1}^{\varepsilon_i} \sum_{g=1}^{\infty} \sum_{m=1}^{G_{n-k-i}^{(g-1,i,j)}} \xi_{g,i,j,m}^{(n,k)} \right). \end{aligned} \quad (\text{A.12})$$



For (A.12), we use the fact that it is irrelevant whether we let  $k$  run over  $\{1, 2, \dots, n-i\}$  or over  $\mathbb{N}$  because  $G_{n-i-k}^{(g-1,i,j)} = 0$  whenever  $k > n-i$ ; by the same argument, we may let  $i$  run up to  $n-k$  only. For fixed  $n \in \mathbb{Z}$  and  $k \in \mathbb{N}$ , the term in the bracket is a sum of i.i.d.  $\text{Pois}(\alpha_k)$  random variables  $\xi_{n,j,g,m}^{(n+i,k)}$  over the (stochastic) index set

$$I^{(n,k)} := \left\{ (g, i, j, m) \in \mathbb{Z}^4 : 1 \leq g, i \leq n-k, 1 \leq j \leq \varepsilon_i, 1 \leq m \leq G_{n-k-i}^{(g-1,i,j)} \right\}.$$

For the size of  $I^{(n,k)}$ , we obtain

$$\left| I^{(n,k)} \right| = \sum_{i=-\infty}^{n-k} \sum_{j=1}^{\varepsilon_i} \sum_{g=1}^{\infty} G_{n-i-k}^{(g-1,i,j)} \stackrel{\text{(A.3)}}{=} \sum_{i=-\infty}^{n-k} \sum_{j=1}^{\varepsilon_i} F_{n-i-k}^{(i,j)} \stackrel{\text{(A.4)}}{=} \tilde{X}_{n-k} \quad (< \infty, \text{ a.s.}). \quad (\text{A.13})$$

Let  $(\xi_m^{(n,k)})$  be the offspring sequences from Definition 2. Note that, for  $n \in \mathbb{Z}$  and  $k \in \mathbb{N}$ ,

$$\left( \xi_l^{(n,k)} : l = 1, 2, \dots, \tilde{X}_{n-k} \right) \quad \text{and} \quad \left( \xi_{g,i,j,m}^{(n,k)} : (g, i, j, m) \in I^{(n,k)} \right) \quad (\text{A.14})$$

are equally distributed – no matter which order we choose for the second set. Also note that, for  $(n, k) \neq (n', k')$ , we have that  $I^{(n,k)} \cap I^{(n',k')} = \emptyset$ . Consequently, the independence properties over  $n$  and  $k$  necessary for the reproductions are preserved. We did indeed make correspondence (A.14) explicit. The reordering of the offspring sequences, however, is cumbersome. It involves a function that is recursively defined on multiple levels; its presentation would double the length of the whole proof – yielding hardly additional insight at that. So we chose to leave it with (A.14). Continuing with (A.11), we obtain that, for  $n \in \mathbb{Z}$ ,

$$\begin{aligned} \tilde{X}_n &\stackrel{\text{(A.11)}}{=} \sum_{i=-\infty}^{n-1} \sum_{j=1}^{\varepsilon_i} \sum_{g=0}^{\infty} G_{n-i}^{(g,i,j)} + \varepsilon_n \\ &\stackrel{\text{(A.12)}}{=} \sum_{k=1}^{\infty} \sum_{(g,i,j,m) \in I^{(n,k)}} \xi_{g,i,j,m}^{(n,k)} + \varepsilon_n = \sum_{k=1}^{\infty} \sum_{l=1}^{|I^{(n,k)}|} \xi_l^{(n,k)} + \varepsilon_n \\ &\stackrel{\text{(A.14)}}{=} \sum_{k=1}^{\infty} \sum_{l=1}^{\tilde{X}_{n-k}} \xi_l^{(n,k)} + \varepsilon_n. \end{aligned}$$

We conclude that  $(\tilde{X}_n)$  indeed solves (2).

For uniqueness, consider two stationary solutions  $(X_n, )$ ,  $(Y_n)$  of (2) – defined on the same probability space and with respect to the same immigration sequence  $(\varepsilon_n)$  and the same offspring sequences  $(\xi_l^{(n,k)})$ ,  $n \in \mathbb{Z}$ ,  $k \in \mathbb{N}$ . It follows from (2) that  $\mathbb{E}X_n = \mathbb{E}Y_n = \alpha/(1-K) < \infty$ . Then

$$\begin{aligned} \mathbb{E}|X_n - Y_n| &\stackrel{\text{(1)}}{=} \mathbb{E} \left| \sum_{k=1}^{\infty} (\alpha_k \circ X_{n-k} - \alpha_k \circ Y_{n-k}) \right| \leq \mathbb{E} \sum_{k=1}^{\infty} |\alpha_k \circ X_{n-k} - \alpha_k \circ Y_{n-k}| \\ &= \sum_{k=1}^{\infty} \mathbb{E} |\alpha_k \circ X_{n-k} - \alpha_k \circ Y_{n-k}|. \end{aligned} \quad (\text{A.15})$$

As the offspring at time  $n$  of  $X_{n-k}$  and  $Y_{n-k}$  is given by same offspring sequence, we have that

$$\begin{aligned} |\alpha_k \circ X_{n-k} - \alpha_k \circ Y_{n-k}| &= \left| \sum_{l=1}^{X_{n-k}} \xi_l^{(n,k)} - \sum_{l=1}^{Y_{n-k}} \xi_l^{(n,k)} \right| = \left| \sum_{l=\min\{X_{n-k}, Y_{n-k}\}}^{\max\{X_{n-k}, Y_{n-k}\}} \xi_l^{(n,k)} \right| \\ &\stackrel{d}{=} \left| \sum_{l=1}^{|X_{n-k}-Y_{n-k}|} \xi_l^{(n,k)} \right|, \quad k \in \mathbb{N}. \end{aligned} \quad (\text{A.16})$$

Plugging (A.16) in (A.15), we obtain

$$\mathbb{E}|X_n - Y_n| \leq \sum_{k=1}^{\infty} \mathbb{E} \sum_{i=1}^{|X_{n-k}-Y_{n-k}|} \xi_i^{(n,k)} = \sum_{k=1}^{\infty} \alpha_k \mathbb{E}|X_{n-k} - Y_{n-k}| \stackrel{\text{stat.}}{=} K \mathbb{E}|X_n - Y_n|, \quad n \in \mathbb{Z}.$$

As  $K < 1$  by assumption and  $\mathbb{E}|X_n - Y_n| < \infty$ , we get that  $\mathbb{E}|X_n - Y_n| = 0$  and therefore  $X_n = Y_n$ ,  $n \in \mathbb{Z}$ , almost surely.  $\square$

## A.2. Proof of Proposition 1

Eq. (3) together with (4) and (5) is exactly the construction of a solution to the defining difference-equations (2) in the proof of Theorem 1; see (A.3) and (A.10). To establish (6), consider the process on the right-hand side:

$$\left( \tilde{F}_n \right)_{n \in \mathbb{Z}} := \left( 1_{\{n=0\}} + \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} F_{n-i}^{(i,j)} \right)_{n \in \mathbb{Z}}.$$

We show that the process  $(\tilde{F}_n)$  is constructed by the same (stochastic) recursion as  $(F_n)$ ; see (4) and (5). Then the equality in distribution follows. For  $n \in \mathbb{Z}$ , we define

$$\tilde{G}_n^{(0)} := 1_{\{n=0\}} \quad \text{and} \quad \tilde{G}_n^{(g)} := \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} G_{n-i}^{(g-1,i,j)}, \quad g \in \mathbb{N}, \quad (\text{A.17})$$

where  $(G_n^{(g,i,j)})$ ,  $g \in \mathbb{N}_0$ , are the generation processes that constitute the family processes  $(F_n^{(i,j)})$ ,  $i \in \mathbb{Z}$ ,  $j \in \mathbb{N}$ , in (6). In particular,  $(G_n^{(g,i,j)})$  are independent copies of the generations  $(G_n^{(g)})$  defined in (5). Then, by construction,  $\tilde{F}_n = \sum_{g \geq 0} \tilde{G}_n^{(g)}$ ,  $n \in \mathbb{Z}$ . This establishes a representation for  $(\tilde{F}_n)$  of the same form as (4) for  $(F_n)$ . Next, we show that the summands  $(\tilde{G}_n^{(g)})$  follow the same recursion (5) as the original generations  $(G_n^{(g)})$ : for  $g = 0$ , we have that  $\tilde{G}_n^{(0)} = 1_{\{n=0\}}$ ,  $n \in \mathbb{Z}$ . So the starting value of the recursion for  $(\tilde{G}_n^{(g)})$  is the same as the starting value of the recursion (5) for  $(G_n^{(g)})$ . For  $g = 1$ , recursion (5) is also analogue:

$$\tilde{G}_n^{(1)} = \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} G_{n-i}^{(0,i,j)} = \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} 1_{\{n-i=0\}} = G_n^{(1)} = \sum_{k=1}^n \alpha_k \circ 1_{\{n-k=0\}} = \sum_{k=1}^n \alpha_k \circ \tilde{G}_{n-k}^{(0)}.$$

And, for any  $g \geq 2$ , we find that

$$\tilde{G}_n^{(g)} = \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} G_{n-i}^{(g-1,i,j)} = \sum_{i=1}^n \sum_{j=1}^{G_i^{(1)}} \sum_{k=1}^{n-i} \alpha_k \circ G_{n-i-k}^{(g-2,i,j)} = \sum_{k=1}^n \sum_{i=1}^{n-k} \sum_{j=1}^{G_i^{(1)}} \alpha_k \circ G_{n-i-k}^{(g-2,i,j)},$$

where in the third equality we use that  $G_{n-i-k}^{(g-2,i,j)} = 0$ ,  $g \geq 2$ ,  $n - i - k \leq 0$ . At this point we avoid the explicit representation of the offspring sequences. We just remind ourselves that all reproductions involved are independent and establish

$$\tilde{G}_n^{(g)} = \sum_{k=1}^n \alpha_k \circ \sum_{i=1}^{n-k} \sum_{j=1}^{G_i^{(1)}} G_{n-i-k}^{(g-2,i,j)} = \sum_{k=1}^n \alpha_k \circ \tilde{G}_{n-k}^{(g-1)}, \quad n \in \mathbb{Z}, \quad g \geq 2.$$

In other words, the processes  $(G_n)$  and  $(\tilde{G}_n)$  and, consequently, the processes  $(F_n)$  and  $(\tilde{F}_n)$  are constructed by the same stochastic recursion. We conclude that  $(F_n) \stackrel{d}{=} (\tilde{F}_n)$ . This establishes (6).  $\square$

### A.3. Proof of Proposition 2

For (9), we first apply representation (3):

$$\begin{aligned} M_{(X_n)}((t_n)) &= \mathbb{E} \exp \left\{ \sum_{n=0}^{\infty} t_n X_n \right\} = \mathbb{E} \exp \left\{ \sum_{n=0}^d t_n X_n \right\} = \mathbb{E} \exp \left\{ \sum_{n=0}^d t_n \sum_{i \in \mathbb{Z}} \sum_{j=1}^{\varepsilon_i} F_{n-i}^{(i,j)} \right\} \\ &= \mathbb{E} \prod_{i \in \mathbb{Z}} \prod_{j=1}^{\varepsilon_i} \exp \left\{ \sum_{n=0}^d t_n F_{n-i}^{(i,j)} \right\}. \end{aligned}$$

In the following, we set  $t_m := 0$  whenever  $m < 0$ . Conditioning on the immigration sequence  $(\varepsilon_i)$  and exploiting its independence from the family processes  $(F_n^{(i,j)})$  yields

$$\begin{aligned} M_{(X_n)}(t_1, \dots, t_d) &= \mathbb{E} \prod_{i \in \mathbb{Z}} \prod_{j=1}^{\varepsilon_i} \mathbb{E} \left[ \exp \left\{ \sum_{n=0}^d t_n F_{n-i}^{(i,j)} \right\} \right] \\ &= \prod_{i \in \mathbb{Z}} \mathbb{E} \left[ \mathbb{E} \left[ \exp \left\{ \sum_{n \in \mathbb{Z}} t_n F_{n-i} \right\} \right]^{\varepsilon_i} \right] \\ &= \prod_{i \in \mathbb{Z}} \mathbb{E} \left[ M_{(F_n)}((t_{i+n})_{n \in \mathbb{N}_0})^{\varepsilon_i} \right] \\ &= \prod_{i \in \mathbb{Z}} \exp \left\{ \alpha_0 \left( M_{(F_n)}((t_{i+n})_{n \in \mathbb{N}_0}) - 1 \right) \right\} \\ &= \exp \left\{ \sum_{i \in \mathbb{Z}} \alpha_0 \left( M_{(F_n)}((t_{i+n})_{n \in \mathbb{N}_0}) - 1 \right) \right\}, \end{aligned} \tag{A.18}$$

where in the last but one step we applied the formula for the probability-generating function of a Poisson random variable. Up to now, (A.18) is only a formal representation of  $M_{(X_n)}$  in terms of  $M_{(F_n)}$ . It is not clear yet, whether and when  $M_{(X_n)}$  is finite. For (10), we apply representation (6) of  $(F_n)$  from Proposition 1:

$$M_{(F_n)}((s_n)) = \mathbb{E} \exp \left\{ \sum_{n=0}^{\infty} s_n F_n \right\} \stackrel{(6)}{=} \mathbb{E} \exp \left\{ \sum_{n=0}^{\infty} s_n \left( 1_{\{n=0\}} + \sum_{k=1}^n \sum_{j=1}^{G_k^{(1)}} F_{n-k}^{(k,j)} \right) \right\}.$$

We note that in the last term, the index  $k$  may run to  $\infty$  instead of  $n$ , because  $F_{n-k}^{(k,j)} = 0$ ,  $j \in \mathbb{N}$ , whenever  $k > n$ . After straightforward calculations, we obtain

$$\begin{aligned} M_{(F_n)}((s_n)_{n \in \mathbb{N}_0}) &= e^{s_0} \mathbb{E} \exp \left\{ \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \sum_{j=1}^{G_k^{(1)}} s_n F_{n-k}^{(k,j)} \right\} \\ &= e^{s_0} \exp \left\{ \sum_{k=1}^{\infty} \alpha_k \left( M_{(F_n)}((s_{n+k})_{n \in \mathbb{N}_0}) - 1 \right) \right\}. \end{aligned} \quad (\text{A.19})$$

Next we derive finiteness of  $M_{(F_n)}$ . Let  $(s_n)$  be a sequence with finite support and  $s := \max\{s_n\}$  so that  $\sum_{n=0}^{\infty} s_n F_n$  is bounded from above by  $sS$ , where  $S := \sum_{n=0}^{\infty} F_n$  denotes the total number of individuals in the generic family  $(F_n)$ . We remind ourselves of the defining Eq. (4) for the family process  $(F_n)$  and find that

$$S = \sum_{n=0}^{\infty} F_n = \sum_{n=0}^{\infty} \sum_{g=0}^{\infty} G_n^{(g)} = \sum_{g=0}^{\infty} \sum_{n=0}^{\infty} G_n^{(g)}. \quad (\text{A.20})$$

We denote the total number of individuals in the  $g$ th generation by  $Y_g := \sum_{n=0}^{\infty} G_n^{(g)}$ ,  $g \in \mathbb{N}_0$ . The sequence  $(Y_g)_{g \in \mathbb{N}_0}$  is the embedded generation process. Applying (5), we find that  $Y_0 = 1$  and, for  $g \geq 2$ ,

$$\begin{aligned} Y_g &= \sum_{n=0}^{\infty} G_n^{(g)} = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \alpha_k \circ G_{n-k}^{(g-1)} = \sum_{k=1}^{\infty} \sum_{n=0}^{\infty} \alpha_k \circ G_{n-k}^{(g-1)} \stackrel{\text{d}}{=} \sum_{k=1}^{\infty} \alpha_k \circ \sum_{n=0}^{\infty} G_{n-k}^{(g-1)} \\ &= \sum_{k=1}^{\infty} \alpha_k \circ Y_{g-1} = \sum_{k=1}^{Y_{g-1}} \xi_k^{(g)}, \end{aligned}$$

where  $\xi_k^{(g)} \stackrel{\text{i.i.d.}}{\sim} \text{Pois}(K)$ ,  $k, g \in \mathbb{N}$ . In other words, the embedded generation process  $(Y_g)$  is a standard Galton–Watson branching process. From (A.20), we see that  $S \stackrel{\text{d}}{=} \sum_{g=0}^{\infty} Y_g$ . In other words,  $S$  is distributed like the cumulative limit of a standard Galton–Watson process. The moment-generating functions of such limits have been considered in the literature: as  $K < 1$ , by Theorem 2.1 in [20], there exists a  $\delta > 0$  such that  $\mathbb{E} \exp\{\delta S\} < \infty$  if and only if there exists a  $\tilde{\delta} > 0$  such that  $\mathbb{E} \exp\{\tilde{\delta} \xi_1^{(1)}\} < \infty$ . The latter is indeed the case because the moment-generating function of a Poisson variable is finite on  $\mathbb{R}$ . Furthermore, by Jensen’s inequality, we see that  $1 \leq \lim_{n \rightarrow \infty} \mathbb{E} \exp\{\delta S/n\} \leq \lim_{n \rightarrow \infty} (\mathbb{E} \exp\{\delta S\})^{1/n} = 1$ . So, for any given  $\epsilon > 0$ , we can

assume the existence of a  $\delta > 0$  such that  $\mathbb{E} \exp(\delta S) < 1 + \epsilon$ . In particular,

$$\begin{aligned} \exists \delta > 0 \text{ such that } M_{(F_n)}((s_n)) &\leq \mathbb{E} \exp(\delta S) < 1 + (1 - K)/(2K) \\ \text{for } (s_n)_{n \in \mathbb{N}_0} &\in c_{00}((-\infty, \delta]), \end{aligned} \quad (\text{A.21})$$

where, as before,  $K = \sum_{k=1}^{\infty} \alpha_k < 1$ . We now have established finiteness of  $M_{(F_n)}$  in a neighborhood of zero. It remains to establish finiteness of  $M_{(X_n)}$ . Our goal is to bound the series representation (A.18) of  $M_{(X_n)}$ . To that aim, we need to refine the bound (A.21) for  $M_{(F_n)}((s_n))$ . To that aim, with the constant  $\delta > 0$  from (A.21), for  $i \in \mathbb{Z}$ , we introduce the sequences  $(\delta_n^{(i)})_{n \in \mathbb{Z}}$  defined by

$$\delta_n^{(i)} := \begin{cases} \delta, & n = i, i-1, \dots, i-d, \\ 0, & \text{else,} \end{cases} \quad (\text{A.22})$$

where  $d = \text{supp}((t_n)) + 1$ , as before, with  $(t_n)$  the considered argument sequence. Note that, for  $i < 0$ , we have that  $\delta_n^{(i)} = 0$ ,  $n \in \mathbb{N}_0$ . Consequently, by definition of  $M_{(F_n)}$ , for  $i < 0$ , we have that  $M_{(F_n)}((\delta_n^{(i)})_{n \in \mathbb{N}_0}) = 1$ . Furthermore, observe that  $(\delta_{n+k}^{(i)})_n = (\delta_n^{(i-k)})_n$ ,  $i, k \in \mathbb{Z}$ . For  $(t_n) \in c_{00}((-\infty, \delta])$ , we have by component-wise monotonicity of  $M_{(X_n)}$  that

$$\begin{aligned} M_{(X_n)}((t_n)) &\leq M_{(X_n)}((\delta_{n+i}^{(d)})_{n \in \mathbb{N}_0}) \stackrel{(\text{A.18})}{=} \exp \left\{ \sum_{i \in \mathbb{Z}} \alpha_0 (M_{(F_n)}((\delta_{n+i}^{(d)})_{n \in \mathbb{N}_0}) - 1) \right\} \\ &= \exp \left\{ \sum_{i=-\infty}^d \alpha_0 (M_{(F_n)}((\delta_n^{(d-i)})_{n \in \mathbb{N}_0}) - 1) \right\} \\ &= \exp \left\{ \alpha_0 \sum_{i=0}^{\infty} m_i \right\}, \end{aligned} \quad (\text{A.23})$$

where we set  $m_i := M_{(F_n)}((\delta_n^{(i)})_{n \in \mathbb{N}_0}) - 1$ ,  $i \in \mathbb{Z}$ . Note that, by (A.21), we have that

$$0 \leq m_i \leq (1 - K)/(2K), \quad i \in \mathbb{Z}. \quad (\text{A.24})$$

and, in particular,  $m_i = 0$  for  $i < 0$ . In the following, we only consider  $m_i$  with  $i > d$ . In this case, we get from (A.22) that  $e^{\delta_0^{(i)}} = e^0 = 1$  and we obtain the recursion

$$\begin{aligned} m_i &= M_{(F_n)}((\delta_n^{(i)})_{n \in \mathbb{N}_0}) - 1 \stackrel{(\text{A.19})}{=} e^{\delta_0^{(i)}} \exp \left\{ \sum_{k=1}^{\infty} \alpha_k (M_{(F_n)}((\delta_{n+k}^{(i)})_{n \in \mathbb{N}_0}) - 1) \right\} - 1 \\ &\stackrel{(\text{A.22})}{=} \exp \left\{ \sum_{k=1}^i \alpha_k m_{i-k} \right\} - 1, \quad i > d. \end{aligned} \quad (\text{A.25})$$

For the sum in the exponential, we find that

$$\sum_{k=1}^i \alpha_k m_{i-k} \stackrel{(\text{A.24})}{\leq} \sum_{k=1}^{\infty} \alpha_k (1 - K)/(2K) = K(1 - K)/(2K) < 1, \quad i \in \mathbb{Z}.$$

Therefore, we may apply the exponential inequality  $\exp(x) \leq (1-x)^{-1}$ ,  $x < 1$ , in (A.25):

$$\begin{aligned} m_i &\leq \frac{1}{1 - \sum_{k=1}^i \alpha_k m_{i-k}} - 1 \stackrel{(A.24)}{=} \frac{\sum_{k=1}^i \alpha_k m_{i-k}}{1 - \sum_{k=1}^i \alpha_k m_{i-k}} \leq \frac{\sum_{k=1}^i \alpha_k m_{i-k}}{1 - \sum_{k=1}^i \alpha_k (1-K)/(2K)} \\ &\leq \frac{2 \sum_{k=1}^{\infty} \alpha_k m_{i-k}}{1+K}, \quad i > d. \end{aligned} \quad (A.26)$$

Summing both sides of (A.26) over  $i > d$ , we obtain

$$\begin{aligned} \sum_{i=d+1}^{\infty} m_i &\leq \frac{2}{1+K} \sum_{i=d+1}^{\infty} \sum_{k=1}^{\infty} \alpha_k m_{i-k} = \frac{2}{1+K} \sum_{k=1}^{\infty} \alpha_k \sum_{i=d+1}^{\infty} m_{i-k} \\ &\stackrel{(A.24)}{\leq} \frac{2}{1+K} \sum_{k=1}^{\infty} \alpha_k \sum_{l=0}^{\infty} m_l = \frac{2K}{1+K} \left( \sum_{l=0}^d m_l + \sum_{l=d+1}^{\infty} m_l \right). \end{aligned} \quad (A.27)$$

Keeping in mind that  $1 - 2K/(1+K) > 0$ , we solve (A.27) for  $\sum_{i=d+1}^{\infty} m_i$  and find that

$$\sum_{i=d+1}^{\infty} m_i \leq \left(1 - \frac{2K}{1+K}\right)^{-1} \frac{2K}{1+K} \sum_{l=0}^d m_l \stackrel{(A.24)}{\leq} \frac{2K}{1-K} d \frac{1-K}{2K} = (1+d). \quad (A.28)$$

For the summation of  $(m_i)_{i \in \mathbb{N}_0}$  over  $i \in \mathbb{N}_0$ , we finally obtain

$$\sum_{i=0}^{\infty} m_i \stackrel{(A.28)}{\leq} \sum_{i=0}^d m_i + d + 1 \stackrel{(A.24)}{\leq} (d+1) \left( \frac{1-K}{2K} + 1 \right) = (d+1) \frac{1+K}{2K}. \quad (A.29)$$

We conclude that, for all  $(t_n) \in c_{00}((-\infty, \delta])$  with  $\text{supp}((t_n)) = d$ , we have that

$$M_{(X_n)}(t_1, \dots, t_d) \stackrel{(A.23)}{\leq} \exp \left\{ \alpha_0 \sum_{i=0}^{\infty} m_i \right\} \stackrel{(A.29)}{\leq} \exp \left\{ \alpha_0 d \frac{1+K}{2K} \right\} < \infty. \quad (A.30)$$

Uniqueness of  $M_{(F_n)}$  follows by induction over the (finite) support of the argument sequence. In that sense, the implicit equation (10) specifies  $M_{(F_n)}$  recursively.  $\square$

#### A.4. Proof of Proposition 3

The sequence values  $u_n$ ,  $n \in \mathbb{Z}$ , are well-defined because the partial sums of  $\sum_{k=1}^{\infty} \alpha_k X_{n-k}$  are nondecreasing and their expectations have a finite limit. Stationarity of  $(u_n)$  follows from stationarity of  $(X_n)$ . Denote  $\mathcal{F}_n := \sigma\{X_k : k \leq n\}$ . From  $\mathbb{E}[\alpha_k \circ X_{n-k} | \mathcal{F}_n] = \alpha_k X_{n-k}$ ,  $k \in \mathbb{N}$ , we get that

$$\mathbb{E}[u_n | \mathcal{F}_{n-1}] = \mathbb{E}[X_n | \mathcal{F}_{n-1}] - \alpha_0 - \sum_{k=1}^{\infty} \alpha_k X_{n-k} = 0, \quad (A.31)$$

and, consequently,  $\mathbb{E}u_n = 0$ ,  $n \in \mathbb{Z}$ . For the autocovariances of the errors, note that, for  $n' < n$  (and then, by symmetry, for  $n' \neq n$ ),

$$\mathbb{E}[u_n u_{n'}] = \mathbb{E}\left[\mathbb{E}[u_n u_{n'} | \mathcal{F}_{n-1}]\right] = \mathbb{E}\left[u_{n'} \underbrace{\mathbb{E}[u_n | \mathcal{F}_{n-1}]}_{\stackrel{(A.31)}{=} 0}\right] = 0.$$

Finally, we have that

$$\begin{aligned} \text{Var}(u_n) &= \mathbb{E}\left[\underbrace{\text{Var}(u_n | \mathcal{F}_{n-1})}_{\stackrel{(11)}{=} \text{Var}(X_n | \mathcal{F}_{n-1})}\right] + \underbrace{\text{Var}\left(\mathbb{E}[u_n | \mathcal{F}_{n-1}]\right)}_{\stackrel{(A.31)}{=} 0} = \mathbb{E}\left[\alpha_0 + \sum_{k=1}^{\infty} \alpha_k X_{n-k}\right] \\ &= \alpha_0 + K \frac{\alpha_0}{1-K} = \frac{\alpha_0}{1-K}. \quad \square \end{aligned} \quad (\text{A.32})$$

### A.5. Proof of Proposition 4

Let  $B$  be the backward shift operator defined by  $B^k x_n := x_{n-k}$ ,  $k \in \mathbb{Z}$ , for any sequence  $(x_n)_{n \in \mathbb{Z}}$ . Consider the power series  $\phi(z) := 1 - \sum_{k=1}^{\infty} \alpha_k z^k$ . With these notations, we may rewrite (13) as  $\phi(B)(X_n - \mu_X) = u_n$ ,  $n \in \mathbb{Z}$ , where  $(u_n)$  is the white-noise sequence from Proposition 3. One can show that the power series  $\phi(z)$  is (absolutely) convergent for all  $|z| \leq 1$ . So  $\phi$  is analytic on the open unit disc. Furthermore, we have that

$$|\phi(z)| = \left| 1 - \sum_{k=1}^{\infty} \alpha_k z^k \right| \geq 1 - \left| \sum_{k=1}^{\infty} \alpha_k z^k \right| \geq 1 - \sum_{k=1}^{\infty} \alpha_k |z|^k \geq 1 - K > 0, \quad |z| \leq 1.$$

As  $\phi(z) \neq 0$  for  $|z| \leq 1$ , we may define the function  $\psi(z) := 1/\phi(z)$ ,  $|z| \leq 1$ , which is also analytic on the open unit disc and which, consequently, has a power-series representation  $\psi(z) = \sum_{k=0}^{\infty} \beta_k z^k$ ,  $|z| < 1$ . As  $1 = \psi(z)\phi(z)$  by definition, it follows that, for  $|z| < 1$ ,

$$1 = \sum_{k=0}^{\infty} \beta_k z^k \left( 1 - \sum_{l=1}^{\infty} \alpha_l z^l \right) = \sum_{k=0}^{\infty} \left( \beta_k - \sum_{j=1}^k \alpha_j \beta_{k-j} \right) z^k. \quad (\text{A.33})$$

Comparing coefficients in (A.33), one obtains the recursion

$$\beta_0 = 1 \quad \text{and} \quad \beta_k = \sum_{j=1}^k \alpha_j \beta_{k-j}, \quad k \in \mathbb{N}.$$

We note that  $\beta_k \geq 0$  because  $\alpha_k \geq 0$ . Formally, we can write

$$X_n - \mu_X = \psi(B)u_n, \quad n \in \mathbb{Z}. \quad (\text{A.34})$$

For the well-definedness of the right-hand side of this equation, it suffices to show that  $\sum_{k=0}^{\infty} |\beta_k| < \infty$ ; see Proposition 3.1.2 in [6]. To that aim, we apply Wiener's Lemma; see Lemma IIc. in [26]. Let  $\tilde{\phi}(\theta) = 1 - \sum_{k=1}^{\infty} \alpha_k e^{ik\pi\theta}$ ,  $\theta \in (-\pi, \pi]$ . The lemma states that if  $\sum_{k=1}^{\infty} |\alpha_k| < 1$ , then the Fourier series of the function  $1/\tilde{\phi}(\theta)$  is absolutely convergent. By the same calculation as in (A.33), we find that the Fourier-coefficients of  $1/\tilde{\phi}(\theta)$  are exactly the

$\beta_k$ ,  $k \in \mathbb{N}_0$ , from our  $\psi$  function. From this we get

$$\sum_{k=0}^{\infty} |\beta_k| \stackrel{\text{Lemma}}{=} \frac{1}{\tilde{\phi}(0)} = \frac{1}{\phi(1)} = \frac{1}{1-K} < \infty.$$

We conclude that (A.34) is a meaningful family of equations. In other words,  $(X_n - \mu_X)$  can be represented as a moving-average process with respect to the white-noise sequence  $(u_n)$ .  $\square$

#### A.6. Proof of Proposition 5

With the notation from the moving-average representation of the INAR( $\infty$ ) sequence in Proposition 4, we find that, for  $j, n \in \mathbb{Z}$ ,

$$R(j) = \text{Cov}(X_n - \mu_X, X_{n+j} - \mu_X) = \text{Cov}\left(\sum_{k=0}^{\infty} \beta_k u_{n-k}, \sum_{k=0}^{\infty} \beta_k u_{n+j-k}\right).$$

From Proposition 3, we know that  $\text{Cov}(u_n, u_{n+j}) = 1_{\{j=0\}}\alpha_0/(1-K)$ ,  $j \in \mathbb{Z}$ . Furthermore, from Proposition 4, we have that the coefficients  $\beta_k$  are absolutely summable. So (15) follows from Proposition 3.1.2. in [6]. For the sum of the autocovariance sequence, we observe

$$\begin{aligned} \sum_{k=0}^{\infty} R(k) &= \frac{\alpha_0}{1-K} \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \beta_i \beta_{i+k} = \frac{\alpha_0}{1-K} \sum_{i=0}^{\infty} \beta_i \sum_{k=0}^{\infty} \beta_{i+k} \\ &\leq \frac{\alpha_0}{1-K} \sum_{i=0}^{\infty} \beta_i \sum_{k=-i}^{\infty} \beta_{i+k} = \frac{\alpha_0}{(1-K)^3}. \end{aligned}$$

The last equality re-uses the result  $\sum_{i=0}^{\infty} \beta_i = 1/(1-K)$  from Proposition 4.  $\square$

#### A.7. Proof of Proposition 9

Plugging in definitions, we obtain

$$\Psi_{N^{(\Delta)}}[f] = \mathbb{E} \exp \left\{ - \int_{\mathbb{R}} f(t) N^{(\Delta)}(dt) \right\} = \mathbb{E} \exp \left\{ - \sum_{n \in \mathbb{Z}} X_n^{(\Delta)} f(n\Delta) \right\}.$$

By stationarity of  $(X_n^{(\Delta)})$ , we may assume without loss of generality that  $f(t) = 0$ ,  $t < 0$ . We apply formulas (9) and (10) for the joint moment-generating function of the INAR( $\infty$ ) process  $(X_n^{(\Delta)})$  and the corresponding generic family process  $(F_n^{(\Delta)})$ :

$$\begin{aligned} \Psi_{N^{(\Delta)}}[f] &= M_{(X_n^{(\Delta)})}((-f(0), -f(\Delta), -f(2\Delta), \dots)) \\ &\stackrel{(9)}{=} \exp \left\{ \Delta \eta \sum_{i \in \mathbb{Z}} \left( M_{(F_n^{(\Delta)})}((-f((n+i)\Delta))_{n \in \mathbb{N}_0}) - 1 \right) \right\}. \end{aligned}$$

We set  $\Psi_{F^{(\Delta)}}^{(\Delta)}((s_n)) := M_{(F^{(\Delta)})}((-s_n))$ ,  $(s_n) \in c_{00}([0, \infty))$ . This establishes the lemma.  $\square$

#### A.8. Proof of Lemma 2

Claims (22) and (23) directly follow from the definition of  $N^{(\Delta)}$  in (21) together with Theorem 1 and Lemma 1. For (24), we find that the number of grid points in the interval



$[a, b]$  is less or equal  $\lceil (b - a)/\Delta \rceil + 1$ . To get rid of the ceiling function, we observe that  $\lceil (b - a)/\Delta \rceil + 1 < (b - a)/\Delta + 2$ . Combining this with the facts that  $K^{(\Delta)} \leq \tilde{K}$  and  $\Delta < \delta$  together with (22) and (23) yields inequality (24).  $\square$

#### A.9. Proof of Lemma 3

Let  $A \in \mathcal{B}_b$  be a bounded Borel set. Note that

$$\begin{aligned} \text{Var}\left(N^{(\Delta)}(A)\right) &= \text{Var}\left(\sum_{n: n\Delta \in A} X_n^{(\Delta)}\right) = \sum_{m: m\Delta \in A} \sum_{n: n\Delta \in A} \text{Cov}\left(X_n^{(\Delta)}, X_m^{(\Delta)}\right) \\ &= \sum_{m: m\Delta \in A} \sum_{n: n\Delta \in A} R^{(\Delta)}(n - m), \end{aligned}$$

where  $R^{(\Delta)}$  is the autocovariance function of the INAR( $\infty$ ) sequence from Proposition 5. From this proposition, we know that  $R^{(\Delta)}(k) \geq 0$ ,  $k \in \mathbb{Z}$  and  $\sum_{k=0}^{\infty} R^{(\Delta)}(k) \leq \eta\Delta(1 - K^{(\Delta)})^{-3}$ . Applying these results yields

$$\begin{aligned} \text{Var}\left(N^{(\Delta)}(A)\right) &\leq \sum_{m: m\Delta \in A} \sum_{n \in \mathbb{Z}} R^{(\Delta)}(n - m) \leq \sum_{m: m\Delta \in A} \frac{2\eta\Delta}{(1 - K^{(\Delta)})^3} \\ &\leq \left(2 + \frac{\sup A - \inf A}{\Delta}\right) \frac{2\eta\Delta}{(1 - \tilde{K})^3} \\ &\leq (2\delta + \sup A - \inf A) \frac{2\eta}{(1 - \tilde{K})^3}, \end{aligned}$$

where  $\tilde{K} < 1$  does not depend on the choice of  $\Delta \in (0, \delta)$ ; see Lemma 1.  $\square$

#### A.10. Proof of Lemma 4

The claim follows with Proposition 11.1.VI. from [7] if for all compact intervals  $[a, b] \subset \mathbb{R}$  and for all  $\epsilon > 0$  there exists an  $M < \infty$  such that

$$\sup_{\Delta \in (0, \delta)} \mathbb{P}\left[N^{(\Delta)}([a, b]) > M\right] < \epsilon.$$

The uniform boundedness of these probabilities is a consequence of Lemma 2 and Markov inequality: for any  $\epsilon > 0$  and  $a < b$ , let  $M_\epsilon := (b - a + 2\delta)\eta/(1 - \tilde{K})$ , where  $\delta > 0$  and  $\tilde{K} < 1$  as in Lemma 1. Then we have that

$$\mathbb{P}\left[N^{(\Delta)}([a, b]) > M_\epsilon\right] \leq \frac{\mathbb{E}N^{(\Delta)}([a, b])}{M_\epsilon} < (b - a + 2\delta) \frac{\eta}{M_\epsilon(1 - \tilde{K})} = \epsilon, \quad \Delta \in (0, \delta). \quad \square$$

#### A.11. Proof of Theorem 2

As a consequence of Lemma 5, the family of probability measures  $(\mathbb{P}^{(\Delta)})_{\Delta \in (0, \delta)}$  that corresponds to the family of point processes  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$  is relatively compact for weak convergence by Prokhorov's theorem; see [7, Theorem A.2.4.I]. So every sequence in  $(\mathbb{P}^{(\Delta)})_{\Delta \in (0, \delta)}$ , respectively,  $(N^{(\Delta)})_{\Delta \in (0, \delta)}$ , contains a weakly convergent subsequence. In particular, for any

zero sequence in  $(0, \delta)$ , we can find a subsequence  $(\Delta_n)$  such that  $(N^{(\Delta_n)})$  converges weakly to some point process  $N^*$ . If the distribution of  $N^*$  does not depend on the initial choice of the subsequence, it follows that the original sequence converges weakly to  $N^*$ ; see Theorem 2.3. in [2]. Reconsider the implicit defining-equation (19) from Definition 5. By Proposition 6, we know that this equation determines the distribution of the solving process. So, for the proof of Theorem 2, it suffices to show that any subsequential limit candidate  $N^*$  solves (19). Furthermore, one can show that it suffices to prove (19) for  $A^* \in \mathfrak{B}_a^{N^*}$ , where  $\mathfrak{B}_a^{N^*}$  is a semiring of sets that generates the  $\sigma$ -algebra  $\mathcal{H}_a^{N^*}$ ; see (17). A semiring is a class of sets  $\mathcal{A}$  such that for any pair  $A, B \in \mathcal{A}$  one has (i)  $A \cap B \in \mathcal{A}$  and (ii)  $(A \cup B) \setminus (A \cap B) = \bigcup_{i=1}^n A_i$  for some  $n \in \mathbb{N}$ ,  $(A_i) \subset \mathcal{A}$  and  $A_i \cap A_j = \emptyset$ ,  $i, j = 1, \dots, n$ . We consider, for any  $a \in \mathbb{R}$  and any point process  $N$ ,

$$\mathfrak{B}_a^N := \left\{ \left\{ \omega \in \Omega : N((s_1, t_1])(\omega) \in D_1, \dots, N((s_k, t_k])(\omega) \in D_k \right\} : \right. \\ \left. -\infty < s_i < t_i \leq a, D_i \subset \mathbb{N}_0, k \in \mathbb{N} \right\}. \quad (\text{A.35})$$

One can check that the set system  $\mathfrak{B}_a^N$  is indeed a semiring. Summarizing the above, for the proof of Theorem 2, it suffices to establish

$$\mathbb{E} \left[ 1_{A^*} N^*((a, b]) \right] = \mathbb{E} \left[ 1_{A^*} \int_a^b \lambda(s|N^*) ds \right], \quad a < b, A^* \in \mathcal{H}_a^{N^*}. \quad (\text{A.36})$$

First, let us establish a discrete version of (A.36) for the approximating sequence: set  $N_n := N^{(\Delta_n)}$  for all  $\Delta_n$  in the chosen subsequence. For  $a < b$  and  $A_n \in \mathfrak{B}_a^{N_n}$ , we find that

$$\begin{aligned} \mathbb{E} \left[ 1_{A_n} N_n((a, b]) \right] &= \mathbb{E} \left[ 1_{A_n} \sum_{k: k\Delta_n \in (a, b]} X_k^{(\Delta_n)} \right] \\ &= \mathbb{E} \left[ 1_{A_n} \sum_{k: k\Delta_n \in (a, b]} \left( \varepsilon_k^{(\Delta_n)} + \sum_{l=1}^{\infty} (\Delta_n h(l\Delta_n)) \circ X_{k-l}^{(\Delta_n)} \right) \right] \\ &= \mathbb{E} \left[ 1_{A_n} \sum_{k: k\Delta_n \in (a, b]} \left( \Delta_n \eta + \sum_{l=1}^{\infty} \Delta_n h(l\Delta_n) X_{k-l}^{(\Delta_n)} \right) \right], \quad n \in \mathbb{N}. \end{aligned}$$

The last step follows by the observation that the immigrations  $\varepsilon_k^{(\Delta_n)}$  as well as the reproductions that contribute to  $X_k^{(\Delta_n)}$ ,  $k\Delta_n > a$ , are independent of  $X_{k-1}^{(\Delta_n)}, X_{k-2}^{(\Delta_n)}, \dots$ .

Rewriting the inner sum of the last term as an integral with respect to the random measure  $N_n$ , we obtain, for  $a < b$ ,

$$\begin{aligned} &\mathbb{E} \left[ 1_{A_n} N_n((a, b]) \right] \\ &= \mathbb{E} \left[ 1_{A_n} \sum_{k: k\Delta_n \in (a, b]} \Delta_n \left( \eta + \int_{-\infty}^{k\Delta_n} h(k\Delta_n - s) N_n(ds) \right) \right], \quad A_n \in \mathfrak{B}_a^{N_n}, n \in \mathbb{N}. \end{aligned} \quad (\text{A.37})$$

Note that here and throughout the proof the upper integration bounds in the Hawkes intensities do not require special attention due to the assumption  $h(0) = 0$  for reproduction intensities

$h$  in Definition 4, respectively, Lemma 1. Now we show that (A.37) converges to (A.36) corresponding to the Hawkes process. For both sides of Eq. (A.37), this is achieved in three steps:

- First, we establish that the random variable in the expectation can be written as  $\Phi(N_n)$ , where  $\Phi : (M_p, \mathcal{M}_p) \rightarrow (\mathbb{R}, \mathcal{B})$  denotes some measurable mapping with set of discontinuities  $D_\Phi \subset M_p$ .
- Next, we show that  $\mathbb{P}[N^* \in D_\Phi] = 0$ .
- Finally, we prove that the random variables in question are uniformly integrable.

By Proposition 8, the first two points together imply that  $\Phi(N_n) \xrightarrow{w} \Phi(N^*)$ . The additional uniform-integrability property yields that the corresponding expectations also converge; see Theorem 5.4 in [2].

*Left-hand side of (A.37):*

Consider the map

$$\Phi : (M_p, \mathcal{M}_p) \rightarrow (\mathbb{R}, \mathcal{B}), \quad m \mapsto 1_{\{m((s_1, t_1]) \in D_1, \dots, m((s_k, t_k]) \in D_k\}} m((a, b]); \quad (\text{A.38})$$

see the definition of  $\mathfrak{B}_a^N$  in (A.35) for the notation. We claim that  $\Phi$  is vaguely continuous on  $M_p \setminus \{m : m(D_\Phi) > 0\}$ , where  $D_\Phi := \left(\{a, b\} \cup \bigcup_{i=1}^k \{s_i, t_i\}\right)$ . Indeed: the map  $m \mapsto m((a, b])$  is vaguely continuous on  $M_p \setminus \{m : m(\{a, b\}) > 0\}$  and, for  $i = 1, \dots, k$ , the maps  $m \mapsto m((s_i, t_i])$  are vaguely continuous on  $M_p \setminus \{m : m(\{s_i, t_i\}) > 0\}$ . The map  $\mathbb{N}_0^k \ni (l_1, \dots, l_k) \mapsto 1_{\{l_1 \in D_1, \dots, l_k \in D_k\}}$  is trivially continuous, so that  $m \mapsto 1_{\{m((s_1, t_1]) \in D_1, \dots, m((s_k, t_k]) \in D_k\}}$  is continuous on  $M_p \setminus \bigcup_{i=1}^k \{s_i, t_i\}$ . From Proposition 8, we have that  $\Phi(N_n) \xrightarrow{w} \Phi(N^*)$  if  $\mathbb{P}[N^*(D_\Phi) > 0] = 0$ . Because  $D_\Phi$  is finite, it suffices to show that  $\mathbb{P}[N^*(\{t\}) > 0] = 0$  for any  $t \in \mathbb{R}$ .

$$\begin{aligned} \mathbb{P}[N^*(\{t\}) > 0] &= \mathbb{E} 1_{N^*(\{t\}) > 0} \leq \mathbb{E} N^*(\{t\}) \stackrel{\text{Lemma 4}}{=} \lim_{n \rightarrow \infty} \mathbb{E} N_n(\{t\}) \\ &\stackrel{(23)}{<} \Delta \frac{\eta}{(1 - \tilde{K})}, \quad t \in \mathbb{R}, \Delta \in (0, \delta). \end{aligned}$$

So  $\mathbb{P}[N^*(\{t\}) > 0] = 0$  and, consequently,  $\mathbb{P}[N^*(D_\Phi) > 0]$  must also be zero. This establishes  $\Phi(N_n) \xrightarrow{w} \Phi(N^*)$ , respectively,  $1_{A_n} N_n(a, b) \xrightarrow{w} 1_{A^*} N^*(a, b)$ . From Lemma 4, we know that  $(N_n(a, b))$  is uniformly integrable, so  $(1_{A_n} N_n(a, b))$  is also uniformly integrable. Combining weak convergence and uniform integrability yields convergence of expectations

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ 1_{A_n} N_n((a, b]) \right] = \mathbb{E} \left[ 1_{A^*} N^*((a, b]) \right].$$

We have established the convergence of the left-hand side of (A.37) to the left-hand side of (A.36).

*Right-hand side of (A.37):*

Note that the right-hand side of (A.37) converges to the right-hand side of (A.36) if

$$\begin{aligned} &\sum_{k \in (a, b]} \Delta_n \mathbb{E} \left[ 1_{A_n} \int_{-\infty}^{k \Delta_n} h(k \Delta_n - s) N_n(ds) \right] \\ &\xrightarrow{n \rightarrow \infty} \int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^t h(t - s) N^*(ds) \right] dt. \end{aligned} \quad (\text{A.39})$$

As a first step for establishing (A.39), note that, for all choices of  $M$  with  $-M < a$ , and, for  $t \in [a, b]$ ,

$$1_{A_n} \int_{-M}^t h(t-s) N_n(ds) \xrightarrow{w} 1_{A^*} \int_{-M}^t h(t-s) N^*(ds), \quad n \rightarrow \infty. \quad (\text{A.40})$$

This is due to a continuous-mapping argument similar to the one we have used for the left-hand side of (A.37). We establish that the variances of the random variables  $\int_{-M}^t h(t-s) N_n(ds)$ ,  $n \in \mathbb{N}_0$ , are uniformly bounded:

$$\begin{aligned} \text{Var} \int_{-M}^t h(t-s) N_n(ds) &\leq \text{Var} \sum_{l=1}^{\lceil M/\Delta_n \rceil} h(l\Delta_n) X_{-l}^{(\Delta_n)} \\ &= \sum_{l=1}^{\lceil M/\Delta_n \rceil} \sum_{m=1}^{\lceil M/\Delta_n \rceil} h(l\Delta_n) h(m\Delta_n) \text{Cov} \left( X_{-l}^{(\Delta_n)}, X_{-m}^{(\Delta_n)} \right). \end{aligned}$$

At this point, we write  $R^{(\Delta_n)}$  for the autocovariance function of the INAR( $\infty$ ) process  $(X_l^{(\Delta_n)})$ . Applying Proposition 5 yields

$$\begin{aligned} \text{Var} \int_{-M}^t h(t-s) N_n(ds) &\leq \sum_{l=1}^{\lceil M/\Delta_n \rceil} h(l\Delta_n) \sup h \sum_{m=1}^{\lceil M/\Delta_n \rceil} R^{(\Delta_n)}(|l-m|) \\ &\leq \sum_{l=1}^{\lceil M/\Delta_n \rceil} h(l\Delta_n) \sup h \sum_{i=-\infty}^{\infty} R^{(\Delta_n)}(i) \\ &\stackrel{(16)}{\leq} \sum_{l=1}^{\lceil M/\Delta_n \rceil} h(l\Delta_n) \sup h \frac{2\eta\Delta_n}{(1-K^{(\Delta_n)})^3} \\ &\leq \left( \frac{M}{\Delta_n} + 1 \right) (\sup h)^2 \frac{2\eta\Delta_n}{(1-K^{(\Delta_n)})^3} \\ &\stackrel{(20)}{\leq} (M + \Delta_n) (\sup h)^2 \frac{2\eta}{(1-\tilde{K})^3} \\ &\stackrel{\Delta_n \leq \delta}{\leq} (M + \delta) (\sup h)^2 \frac{2\eta}{(1-\tilde{K})^3} \leq c_M < \infty, \end{aligned} \quad (\text{A.41})$$

where  $c_M$  is a constant independent of  $n$ , respectively,  $\Delta_n$ . We may conclude that the random variables  $1_{A_n} \int_{-M}^t h(t-s) N_n(ds)$ ,  $n \in \mathbb{N}$ , are uniformly integrable. Weak convergence together with uniform integrability yields convergence of expectations. We have established that, for  $M$  with  $-M < a$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ 1_{A_n} \int_{-M}^t h(t-s) N_n(ds) \right] = \mathbb{E} \left[ 1_{A^*} \int_{-M}^t h(t-s) N^*(ds) \right]. \quad (\text{A.42})$$

For the proof of (A.39), we consider a truncated part of  $h$  and its remaining tail separately. For the truncated part, we use (A.42); for the tail part, we use the integrability condition  $\int h \, dt < 1$ :

for any  $M > -a$ , we have

$$\begin{aligned}
 & \left| \sum_{k \in (a, b]} \Delta_n \mathbb{E} \left[ 1_{A_n} \int_{-\infty}^{k\Delta_n} h(k\Delta_n - s) N_n(ds) \right] - \int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^t h(t - s) N^*(ds) \right] dt \right| \\
 & \leq \left| \sum_{k \in (a, b]} \Delta_n \mathbb{E} \left[ 1_{A_n} \int_{-M}^{k\Delta_n} h(k\Delta_n - s) N_n(ds) \right] \right. \\
 & \quad \left. - \int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-M}^t h(t - s) N^*(ds) \right] dt \right| \\
 & \quad + \sum_{k \in (a, b]} \Delta_n \mathbb{E} \left[ 1_{A_n} \int_{-\infty}^{-M} h(k\Delta_n - s) N_n(ds) \right] \\
 & \quad + \int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^{-M} h(t - s) N^*(ds) \right] dt. \tag{A.43}
 \end{aligned}$$

Let  $\varepsilon > 0$ . We show that we can find  $M_\varepsilon$  and  $N_\varepsilon \in \mathbb{N}$  such that each of the three summands in (A.43) is bounded by  $\varepsilon/3$  for  $n \geq N_\varepsilon$ . First, consider the integrand of the last summand in (A.43). By arguing similarly to the  $m((a, b])$  part in (A.38), we find that  $\mathbb{E}N^*(dt)/dt \leq \eta/(1 - \tilde{K})$ . So we can choose  $M_\varepsilon^{(1)} > 0$  so large that, for  $t \in [a, b]$ ,

$$\begin{aligned}
 \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^{-M_\varepsilon^{(1)}} h(t - s) N^*(ds) \right] & \leq \mathbb{E} \int_{-\infty}^{-M_\varepsilon^{(1)}} h(t - s) N^*(ds) \\
 & \leq \frac{\eta}{1 - \tilde{K}} \int_{M_\varepsilon^{(1)} + a}^{\infty} h(s) ds < \frac{\varepsilon}{3(b - a)}.
 \end{aligned}$$

The summands of the second term in (A.43) can be bounded by  $\varepsilon/(3[b - a])$  in a similar and even more direct way – uniformly over  $n$  and possibly with respect to another  $M_\varepsilon^{(2)} > 0$ . We set  $M_\varepsilon := \max\{M_\varepsilon^{(1)}, M_\varepsilon^{(2)}\}$ . So taking the integral over the interval  $[a, b]$  of the last summand, respectively, the Riemann sum of the second summand in (A.43), yields

$$\int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^{-M_\varepsilon} h_M(t - s) N^*(ds) \right] dt < \frac{\varepsilon}{3}, \tag{A.44}$$

respectively,

$$\sum_{k: k\Delta_n \in (a, b]} \Delta_n \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^{-M_\varepsilon} h_M(k\Delta_n - s) N^*(ds) \right] < \frac{\varepsilon}{3}. \tag{A.45}$$

For the first term in (A.43), denote

$$E_n(t) := \mathbb{E} \left[ 1_{A_n} \int_{-M_\varepsilon}^t h(t - s) N_n(ds) \right] \quad \text{and} \quad E^*(t) := \mathbb{E} \left[ 1_{A^*} \int_{-M_\varepsilon}^t h(t - s) N^*(ds) \right].$$

From (A.40), we already know that, for any choice of  $M_\varepsilon$ ,  $\lim_{n \rightarrow \infty} |E_n(t) - E^*(t)| = 0$ ,  $t \in (a, b]$ . However, the convergence of the Riemann-like sums  $\sum_{k: k\Delta_n \in (a, b]} E_n(k\Delta_n) \Delta_n$  to the integral  $\int_a^b E^*(s) ds$  is nontrivial as the functions  $E_n$  are themselves part of the sequence. We

write

$$\left| \sum_{k: k\Delta_n \in (a,b]} \Delta_n E_n(k\Delta_n) - \int_a^b E^*(t) dt \right| \leq \left| \sum_{k: k\Delta_n \in (a,b]} \Delta_n E_n(k\Delta_n) - \int_a^b E_n(t) dt \right| + \left| \int_a^b E_n(t) dt - \int_a^b E^*(t) dt \right|. \quad (\text{A.46})$$

The second absolute difference in (A.46) converges to zero because of dominated convergence of  $(E_n)$ . Indeed, for  $t \in [a, b]$ ,

$$\begin{aligned} \mathbb{E} \left[ 1_{A_n} \int_{-M_\varepsilon}^t h(t-s) N_n(ds) \right] &\leq \mathbb{E} \int_{-M_\varepsilon}^b h(t-s) N_n(ds) \leq \sup h \sum_{k: k\Delta_n \in (-M_\varepsilon, b]} \mathbb{E} X_k^{(\Delta_n)} \\ &\leq \sup h \frac{[M_\varepsilon + b]\eta}{1 - \tilde{K}}. \end{aligned} \quad (\text{A.47})$$

Note that  $\sup h < \infty$  follows from the piecewise-continuity assumption. In view of the upper bound (A.47), we apply the dominated convergence theorem and choose  $N_\varepsilon^{(1)} \in \mathbb{N}$  so large that, for the second absolute difference in (A.46), we have

$$\left| \int_a^b E_n(t) dt - \int_a^b E^*(t) dt \right| < \frac{\varepsilon}{6}, \quad n \geq N_\varepsilon^{(1)}. \quad (\text{A.48})$$

For the first absolute difference in (A.46), we assume that, without loss of generality, the piecewise continuous function  $h$  is uniformly continuous on  $(0, \infty)$ . Otherwise, we note that any piecewise continuous function on  $\mathbb{R}$  that is vanishing at infinity is uniformly continuous on each of its continuous pieces and do the following calculation once for every uniformly continuous piece of  $h$ . Uniform continuity gives us a constant  $\delta_h > 0$ , so small, that, for any  $t_0 > 0$ ,

$$|t - t_0| < \delta_h \wedge t > 0 \Rightarrow |h(t) - h(t_0)| < \frac{\varepsilon(1 - \tilde{K})}{12\eta(b - a + \delta)(M_\varepsilon + b + \delta)}. \quad (\text{A.49})$$

Now, choose  $N_\varepsilon^{(2)}$  so large that

$$\Delta_n < \min \left\{ \delta_h, \frac{\varepsilon(1 - \tilde{K})}{12\eta(b - a + \delta) \sup h} \right\} \quad \text{for } n \geq N_\varepsilon^{(2)}. \quad (\text{A.50})$$

Here again,  $\delta$  and  $\tilde{K}$  are the constants from Lemma 1. Let  $a \leq s < t \leq b$  with  $t - s < \Delta_n (< \delta_h)$ , then

$$\begin{aligned} &|E_n(t) - E_n(s)| \\ &= \left| \sum_{k\Delta_n \in (-M_\varepsilon, t)} \mathbb{E} \left[ 1_{A_n} h(t - k\Delta_n) X_k^{(\Delta_n)} \right] - \sum_{k\Delta_n \in (-M_\varepsilon, s]} \mathbb{E} \left[ 1_{A_n} h(s - k\Delta_n) X_k^{(\Delta_n)} \right] \right| \\ &\leq \left( \sum_{k\Delta_n \in (-M_\varepsilon, s]} |h(t - k\Delta_n) - h(s - k\Delta_n)| + \sum_{k\Delta_n \in (s, t)} h(t - k\Delta_n) \right) \mathbb{E} X_0^{(\Delta_n)} \\ &\stackrel{(\text{A.49})}{\leq} \left( \frac{M_\varepsilon + s + \Delta_n}{\Delta_n} \frac{\varepsilon(1 - \tilde{K})}{12\eta(b - a + \delta)(M_\varepsilon + b + \delta)} + \sup h \right) \frac{\Delta_n \eta}{1 - K^{(\Delta)}} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{\varepsilon}{12(b-a+\delta)} + \Delta_n \frac{\eta \sup h}{1-\tilde{K}} \\
&\stackrel{(A.50)}{\leq} \frac{\varepsilon}{12(b-a+\delta)} + \frac{\varepsilon}{12(b-a+\delta)} = \frac{\varepsilon}{6(b-a+\delta)}.
\end{aligned} \tag{A.51}$$

Summarizing the above calculation, we have established the existence of an  $N_\varepsilon^{(2)} \in \mathbb{N}$  such that, for all  $n \geq N_\varepsilon^{(2)}$ , we have  $|E_n(t) - E_n(s)| \leq \varepsilon/(6(b-a+\delta))$  whenever  $|t-s| < \Delta_n$  and  $s, t \in [a, b]$ . For the first absolute difference in (A.46), we therefore get

$$\begin{aligned}
&\left| \sum_{k: k\Delta_n \in (a,b]} \Delta_n E_n(k\Delta_n) - \int_a^b E_n(t) dt \right| \\
&\leq \sum_{k: k\Delta_n \in (a,b]} \int_{k\Delta_n}^{(k+1)\Delta_n} |E_n(k\Delta_n) - E_n(t)| dt \\
&\leq \sum_{k: k\Delta_n \in (a,b]} \int_{k\Delta_n}^{(k+1)\Delta_n} \frac{\varepsilon}{6(b-a+\delta)} dt \\
&\stackrel{(A.51)}{\leq} \frac{(b-a+\Delta_n)}{\Delta_n} \Delta_n \frac{\varepsilon}{6(b-a+\delta)} \\
&\leq \frac{\varepsilon}{6}, \quad n \geq N_\varepsilon^{(2)}.
\end{aligned} \tag{A.52}$$

Set  $N_\varepsilon := \max\{N_\varepsilon^{(1)}, N_\varepsilon^{(2)}\}$ . Combining (A.48) and (A.52), we get that

$$\left| \sum_{k: k\Delta \in (a,b]} \Delta_n E_n(t) - \int_a^b E^*(t) dt \right| < \frac{\varepsilon}{3}, \quad n \geq N_\varepsilon. \tag{A.53}$$

Combining (A.44), (A.45) and (A.53), shows that (A.43) is smaller than the given  $\varepsilon$ , for  $n \geq N_\varepsilon$  and  $M := M_\varepsilon$ , i.e.,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \sum_{k: k\Delta_n \in (a,b]} \Delta_n \mathbb{E} \left[ 1_{A_n} \int_{-\infty}^{\Delta_n} h(k\Delta_n - s) N_n(ds) \right] \\
&= \int_a^b \mathbb{E} \left[ 1_{A^*} \int_{-\infty}^t h(t-s) N^*(ds) \right] dt.
\end{aligned}$$

We have established that the right-hand side of (A.37) also converges to the right-hand side of (A.36). With the result from Proposition 6 on the uniqueness property of (A.36), we find that every subsequential limit  $N^*$  has the same distribution as the Hawkes process  $N$ . We may then conclude that, for  $\Delta \rightarrow 0$ , the approximating sequence of point processes  $(N^{(\Delta)})$  converges weakly to the Hawkes process  $N$ .  $\square$

## References

- [1] M. Al-Osh, A. Alzaid, First order integer valued autoregressive process, *J. Time Series Anal.* 8 (1987) 261–275.
- [2] P. Billingsley, *Convergence of Probability Measures*, John Wiley and Sons, 1968.
- [3] G. Box, G. Jenkins, *Time Series Analysis—Forecasting and Control*, Holden Day, San Francisco, 1970.

- [4] P. Brémaud, L. Massoulié, Stability of nonlinear Hawkes processes, *Ann. Probab.* 24 (3) (1996) 1563–1588.
- [5] P. Brémaud, L. Massoulié, Hawkes branching processes without ancestors, *J. Appl. Probab.* 38 (2001) 122–135.
- [6] P. Brockwell, R. Davis, *Time Series: Theory and Applications*, second ed., Springer, 1991.
- [7] D. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Vol. I and II, second ed., Springer, 2003.
- [8] I.M. da Silva, Contributions to the analysis of discrete-valued time series (Ph.D. Thesis), Departamento de Matematica Aplicada Faculdade de Ciencias da Universidade do Porto, 2005.
- [9] K. Fokianos, B. Kedem, *Regression Models for Time Series Analysis*, Wiley, 2012.
- [10] K. Fokianos, D. Tjøstheim, Nonlinear Poisson autoregression, *Ann. Inst. Statist. Math.* 64 (2012) 1205–1225.
- [11] A. Hawkes, Point spectra of some mutually-exciting point processes, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 33 (1971) 438–443.
- [12] A. Hawkes, Spectra of some self-exciting and mutually-exciting point processes, *Biometrika* 58 (1971) 83–90.
- [13] A. Hawkes, A cluster representation of a self-exciting point process, *J. Appl. Probab.* 11 (1974) 493–503.
- [14] O. Kallenberg, *Random Measures*, third ed., Akademie-Verlag, 1983.
- [15] M. Kirchner, An estimation procedure for the Hawkes process, *Quant. Finance* (2016) Submitted for publication.
- [16] A. Latour, The multivariate GINAR( $p$ ) process, *Adv. Appl. Probab.* 29 (1997) 228–248.
- [17] J.-G. Du, Y. Li, The integer-valued autoregressive (INAR( $p$ )) model, *J. Time Series Anal.* 12 (2) (1991) 129–142.
- [18] T. Liniger, *Multivariate Hawkes processes* (Ph.D. Thesis), ETH Zurich, 2009.
- [19] E. McKenzie, Some simple models for discrete variate time series, *Water Resour. Bull.* 21 (4) (1985) 645–650.
- [20] M. Nakayama, P. Shahabuddin, K. Sigman, On finite exponential moments for branching processes and busy periods for queues, *J. Appl. Probab.* 41 (2004) 273–280.
- [21] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.* 83 (401) (1988) 9–27.
- [22] S. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Springer, 1987.
- [23] E. Seneta, Functional equations and the Galton–Watson process, *Adv. Appl. Probab.* 1 (1969) 1–42.
- [24] F. Steutel, K. van Harn, Discrete analogues of self-decomposability and stability, *Ann. Probab.* 7 (5) (1979) 893–899.
- [25] P. Whittle, Hypothesis testing in time series analysis (Ph.D. Thesis), University of Uppsala, 1951.
- [26] N. Wiener, Tauberian theorems, *Ann. of Math.* 33 (1) (1932) 1–100.
- [27] H. Zhao, A dynamic contagion process for modelling contagion risk in finance and insurance (Ph.D. Thesis), The London School of Economics and Political Science, 2012.