Poisson approximation of point processes with stochastic intensity, and application to nonlinear Hawkes processes

Giovanni Luca Torrisi

Istituto per le Applicazioni del Calcolo "Mauro Picone", CNR, Via dei Taurini 19, I-00185 Roma. E-mail: torrisi@iac.rm.cnr.it

Abstract. We give a general inequality for the total variation distance between a Poisson distributed random variable and a first order stochastic integral with respect to a point process with stochastic intensity, constructed by embedding in a bivariate Poisson process. We apply this general inequality to first order stochastic integrals with respect to a class of nonlinear Hawkes processes, which is of interest in queueing theory, providing explicit bounds for the Poisson approximation, a quantitative Poisson limit theorem, confidence intervals and asymptotic estimates of the moments.

Résumé. Nous donnons une inégalité générale pour la distance en variation totale entre une variable de Poisson aléatoire et une intégrale stochastique par rapport à un processus ponctuel avec une intensité stochastique, construite par plongement dans un processus de Poisson bivarié. Nous appliquons cette inégalité générale aux intégrales stochastiques par rapport à une classe de processus de Hawkes non linéaires, ce qui a un intérêt en théorie des files d'attente, en fournissant des bornes explicites pour l'approximation Possonienne, ainsi qu'un théorème limite Poissonien quantitatif et des intervalles de confiance et estimatées asymptotiques des moments.

1. Introduction

The aim of this paper is to estimate the error when approximating, by a Poisson random variable, a first order stochastic integral with respect to a point process with stochastic intensity. We accomplish this task for first order stochastic integrals with respect to a point process constructed by embedding in a bivariate Poisson process (see e.g. Lemmas 2 and 3 in [9]; see also Lemma 2.1), combining the Chen–Stein method (see e.g. [3]) with a Clark–Ocone type representation formula for functionals of the Poisson process (see Theorem 1.1 in [23]; see also Lemma 2.2).

One of our achievements is a general inequality (see Theorem 3.1) assessing the total variation distance between a Poisson distributed random variable and a first order stochastic integral with respect to a point process with stochastic intensity, constructed by embedding in a bivariate Poisson process.

The present work is a natural continuation of the findings contained in [37] where the author combined Stein's method (see e.g. [15]) with a Clark–Ocone type representation formula for functionals of the Poisson process in order to compute explicit bounds for the Gaussian approximation of the first chaos (or innovation) of a point process with stochastic intensity (still constructed by embedding in a bivariate Poisson process).

We apply the general inequality for the Poisson approximation to a class of stationary nonlinear Hawkes processes which is of interest in queueing theory, since it encompasses the input process to the Erlang loss system (see Section 5), and in neuroscience, since it may be interpreted as a neuron firing model (see [9,26]). Particularly, we provide an explicit bound for the total variation distance between a Poisson distributed random variable and a first order

stochastic integral with respect to a stationary nonlinear Hawkes process with bounded and Lipschitz dynamics (see Theorem 4.1). For such class of first order stochastic integrals, we provide also a quantitative Poisson limit theorem, confidence intervals and asymptotic estimates of the moments (see Theorem 4.4, Corollary 4.5 and Corollary 4.7, respectively).

Historically, the work [22] was the first on the subject of nonlinear Hawkes processes. In [9] the authors studied the stability properties of these processes such as stationarity and convergence to stationarity, and described nonlinear Hawkes processes as natural extensions of the so-called self-exciting point processes studied in [20,21]. Various generalizations of the results in [9] are provided in [26]. The rate of convergence to equilibrium of nonlinear Hawkes processes is investigated in [11,36], central limit theorems and large deviation principles are given in [7,39–41], mean field limits are provided in [19].

Our general inequality for the Poisson approximation extends the bound of Theorem 3.1 in [32] specialized to first order Poisson stochastic integrals. In [32] the author combines the Chen–Stein method with the integration by parts formula of the Malliavin calculus on the Poisson space (see e.g. [29]) to deduce explicit bounds for the Poisson approximation of functionals of the Poisson process. Here we combine the Chen–Stein method with a Clark–Ocone type representation formula for functionals of the Poisson process in order to get bounds on the total variation distance between a Poisson distributed random variable and a first order stochastic integral with respect to a point process with stochastic intensity (constructed by embedding in a bivariate Poisson process). Our general bound differs in many respects from the bound in Theorem 3.7 of [4], which refers to the Poisson approximation in total variation distance of the number of points on a finite interval of a point process with compensator.

The ideas in this article are inspired by the seminal papers [27] and [31], where the Stein method and the Malliavin calculus are combined in order to study Gaussian approximations of functionals on the Wiener and Poisson spaces, respectively. The works [27,31] have been indeed the seeds of many generalizations and refinements, we cite the papers [24,34,35] and the monograph [28], among others.

The paper is organized as follows. In Section 2 we give some preliminaries on point processes, including the notion of stochastic intensity, the Poisson embedding construction and a Clark–Ocone type representation formula for functionals of the Poisson process. In Section 3 we present a general bound for the Poisson approximation of first order stochastic integrals with respect to point processes with stochastic intensity, constructed by embedding in a bivariate Poisson process. The application of the general inequality to first order stochastic integrals with respect to nonlinear Hawkes processes with bounded and Lipschitz dynamics is provided in Section 4, including explicit bounds, a quantitative Poisson limit theorem, confidence intervals and asymptotic estimates of the moments. Finally, in Section 5, as an illustrative example, we give a quantitative Poisson limit theorem, confidence intervals and asymptotic estimates of the moments for the input process to the $M/D/K/0$ queue or Erlang loss system (see e.g. [2]).

2. Preliminaries on point processes

In this section we give some preliminaries on point processes, and refer the reader to the books [8,16,17] for more insight into this subject.

Let $\{T_n\}_{n\in\mathbb{Z}}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$ with values in $\mathbb{R}$. Given a Borel set $A \in \mathcal{B}(\mathbb{R})$, we define

$$N(A) := \sum_{n\in\mathbb{Z}} \mathbb{1}_A(T_n)$$

and we call $N := \{N(A)\}_{A\in\mathcal{B}(\mathbb{R})}$ the point process with times $\{T_n\}_{n\in\mathbb{Z}}$. We suppose that N has the following properties, which hold almost surely (a.s.):

$T_0 \le 0 < T_1$ and $T_n < T_{n+1}$, $n \in \mathbb{Z}$,

$N(A) < \infty$, for all bounded A.

These conditions guarantee that N is simple, i.e. $N(\{a\}) \le 1$ for any $a \in \mathbb{R}$, and locally finite.

Given a sequence $\{Z_n\}_{n \in \mathbb{Z}}$ of random variables on Ω with values in some measurable space $(E, \mathcal{E})$, we define

$$\overline{N}(C) := \sum_{n \in \mathbb{Z}} \mathbb{1}_C(T_n, Z_n), \quad C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{E}$$

and

$$\int_C \psi(t, z) \overline{N}(\mathrm{d}t \times \mathrm{d}z) := \sum_{n \in \mathbb{Z}} \psi(T_n, Z_n) \mathbb{1}_C(T_n, Z_n)$$

for a measurable function $\psi : \mathbb{R} \times E \to \mathbb{R}$ for which the infinite sum is well defined.

2.1. *Point processes with stochastic intensity*

Let $\mathcal{F} := \{\mathcal{F}_t\}_{t \in \mathbb{R}} \subset \mathcal{A}$ be a filtration such that $\mathcal{F}_t \supseteq \mathcal{F}_t^N$ for any $t \in \mathbb{R}$, where $\mathcal{F}^N := \{\mathcal{F}_t^N\}_{t \in \mathbb{R}}$ is the natural filtration of the point process N, i.e.

$$\mathcal{F}_t^N := \sigma\big\{N(A) : A \in \mathcal{B}(\mathbb{R}), A \subseteq (-\infty, t]\big\}.$$

Let $\{\lambda(t)\}_{t \in \mathbb{R}}$ be a non-negative stochastic process defined on $(\Omega, \mathcal{A}, P)$ which is $\mathcal{F}$-adapted, i.e. $\lambda(t)$ is $\mathcal{F}_t$-measurable for any $t \in \mathbb{R}$, and such that

$$\int_a^b \lambda(t) \, \mathrm{d}t < \infty, \quad \text{a.s., for all } a, b \in \mathbb{R}.$$

We call $\{\lambda(t)\}_{t \in \mathbb{R}}$ $\mathcal{F}$-stochastic intensity of N if, for any $a, b \in \mathbb{R}$,

$$\mathrm{E}\big[N\big((a, b]\big) | \mathcal{F}_a\big] = \mathrm{E}\bigg[\int_a^b \lambda(t) \, \mathrm{d}t | \mathcal{F}_a\bigg], \quad \text{a.s.}$$

Given a filtration $\mathcal{G} := \{\mathcal{G}_t\}_{t \in \mathbb{R}} \subset \mathcal{A}$, we define the σ-field $\mathcal{P}(\mathcal{G})$ on $\mathbb{R} \times \Omega$ by

$$\mathcal{P}(\mathcal{G}) := \sigma\big\{(a, b] \times A : a, b \in \mathbb{R}, A \in \mathcal{G}_a\big\}.$$

We call $\mathcal{P}(\mathcal{G})$ the $\mathcal{G}$-predictable σ-field, and say that a real-valued stochastic process $\{X(t)\}_{t \in \mathbb{R}}$ is $\mathcal{G}$-predictable if the mapping $X : \mathbb{R} \times \Omega \to \mathbb{R}$ is $\mathcal{P}(\mathcal{G})$-measurable. A typical $\mathcal{G}$-predictable process is a $\mathcal{G}$-adapted process with left-continuous trajectories.

Given a point process N on $\mathbb{R}$ with $\mathcal{F}$-stochastic intensity $\{\lambda(t)\}_{t \in \mathbb{R}}$ and a measurable function $u : \mathbb{R} \to \mathbb{R}$, we shall consider the random variables

$$N(u) := \int_{\mathbb{R}} u(t) N(\mathrm{d}t), \qquad \Lambda(u) := \int_{\mathbb{R}} u(t) \lambda(t) \, \mathrm{d}t, \qquad \delta(u) := N(u) - \Lambda(u),$$

any time these quantities are well-defined.

2.2. *Point processes constructed by embedding in a bivariate Poisson process*

Hereafter, $\overline{N}$ denotes a Poisson process on $\mathbb{R} \times [0, \infty)$ defined on a probability space $(\Omega, \mathcal{A}, P)$, with mean measure $\mathrm{d}t \, \mathrm{d}z$. We denote by $\mathcal{F}^{\overline{N}} := \{\mathcal{F}_t^{\overline{N}}\}_{t \in \mathbb{R}}$ the natural filtration of $\overline{N}$, i.e.

$$\mathcal{F}_t^{\overline{N}} := \sigma\big\{\overline{N}(A \times B) : A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}\big([0, \infty)\big), A \subseteq (-\infty, t]\big\}.$$

Point processes with stochastic intensity may be constructed by embedding in a bivariate Poisson process as follows.

Lemma 2.1. *Let* $f, g : \mathbb{R} \times \Omega \to [0, \infty)$ *be two non-negative,* $\mathcal{P}(\mathcal{F}^{\overline{N}})$*-measurable mappings such that*

$$\int_a^b |f(t) - g(t)| \, \mathrm{d}t < \infty, \quad a.s., \text{ for all } a, b \in \mathbb{R}$$

and define the point process on $\mathbb{R}$

$$N(\mathrm{d}t) := \overline{N}\big(\mathrm{d}t \times \big(\min\{f(t), g(t)\}, \max\{f(t), g(t)\}\big]\big), \quad t \in \mathbb{R}.$$

Then N *has* $\mathcal{F}^{\overline{N}}$*-stochastic intensity* $\{|f(t) - g(t)|\}_{t \in \mathbb{R}}$.

This result is an extension of the method proposed in [25] for the simulation of non-homogeneous Poisson processes and was used e.g. in [9] and [26] to study the stability of various classes of point processes, including Hawkes processes.

Throughout this paper we consider point processes N on $\mathbb{R}$ defined by

$$N(\mathrm{d}t) := \overline{N}\big(\mathrm{d}t \times \big(0, \lambda(t)\big]\big), \tag{1}$$

where $\{\lambda(t)\}_{t \in \mathbb{R}}$ is a non-negative stochastic process of the form

$$\lambda(t) := \varphi(t, \overline{N}|_{(-\infty, t)}) \tag{2}$$

such that

$$\int_a^b \lambda(s) \, \mathrm{d}s < \infty, \quad a.s., \text{ for all } a, b \in \mathbb{R}. \tag{3}$$

Here, $\varphi : \mathbb{R} \times \mathcal{N} \to [0, \infty)$ is a measurable functional, $\mathcal{N}$ denotes the space of simple and locally finite counting measures on $\mathbb{R} \times [0, \infty)$ endowed with the vague topology (see e.g. [17]) and, for simplicity, with a little abuse of notation, we denote by $\overline{N}|_{(-\infty, t)}$ the restriction of $\overline{N}$ to $(-\infty, t) \times [0, \infty)$. Let $C \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}([0, \infty))$ be arbitrarily fixed. Since the process $\{\overline{N}|_{(-\infty, t)}(C)\}_{t \in \mathbb{R}}$ is $\mathcal{F}^{\overline{N}}$-adapted and left-continuous, the mapping

$$(t, \omega) \to \overline{N}|_{(-\infty, t)}(C)$$

is $\mathcal{P}(\mathcal{F}^{\overline{N}})$-measurable. Therefore, $\{\lambda(t)\}_{t \in \mathbb{R}}$ is $\mathcal{F}^{\overline{N}}$-predictable (see e.g. Remark 1 in [26]). Consequently, by Lemma 2.1 we deduce that N defined by (1), (2) and (3) has $\mathcal{F}^{\overline{N}}$-stochastic intensity $\{\lambda(t)\}_{t \in \mathbb{R}}$.

As we shall see more in detail later on, Hawkes processes may be constructed by embedding in a bivariate Poisson process, see [9].

2.3. *The finite difference operator on the Poisson space and a Clark–Ocone type representation formula*

Given a measurable functional $\psi : \mathcal{N} \to \mathbb{R}$, we define the finite difference operator D by

$$D_{(t,z)}\psi(\overline{N}) := \psi(\overline{N} + \varepsilon_{(t,z)}) - \psi(\overline{N}),$$

where $\varepsilon_{(t,z)}$ denotes the Dirac measure at $(t, z) \in \mathbb{R} \times [0, \infty)$. We also define the σ-field

$$\mathcal{F}_{t^-}^{\overline{N}} := \sigma\big\{\overline{N}(A \times B) : A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}([0, \infty)), A \subseteq (-\infty, t)\big\}, \quad t \in \mathbb{R}.$$

The following Clark–Ocone type representation formula holds, see Theorem 1.1 in [23] (see also Lemma 1.3 in [38]).

Lemma 2.2. *For any measurable functional* $\psi : \mathcal{N} \to \mathbb{R}$ *such that* $\psi(\overline{N}) \in L^2(\Omega, \mathrm{d}P)$, *we have*

$$\psi(\overline{N}) - \mathrm{E}\big[\psi(\overline{N})\big] = \int_{\mathbb{R} \times [0, \infty)} g(t, z)\big(\overline{N}(\mathrm{d}t \times \mathrm{d}z) - \mathrm{d}t \, \mathrm{d}z\big),$$

where $g(t, z)$ *denotes a* $\mathcal{P}(\mathcal{F}^{\overline{N}}) \otimes \mathcal{B}([0, \infty))$*-measurable version of* $\mathrm{E}[D_{(t,z)}\psi(\overline{N})|\mathcal{F}_{t^-}^{\overline{N}}]$.

3. Poisson approximation of first order stochastic integrals with respect to point processes with stochastic intensity

Given a function $f : \mathbb{N} \to \mathbb{R}$, $\mathbb{N} := \{0, 1, 2, \ldots\}$, we define the operators

$$\Delta f(k) := f(k+1) - f(k)$$

and $\Delta^2 f := \Delta(\Delta f)$. Let $\mathrm{Po}(\theta)$ be a Poisson random variable with mean $\theta > 0$ and $A \subseteq \mathbb{N}$. It turns out that there exists a unique function $f_A : \mathbb{N} \to \mathbb{R}$ such that

$$\mathbb{1}_A(k) - P\big(\mathrm{Po}(\theta) \in A\big) = \theta f_A(k+1) - k f_A(k), \quad k \in \mathbb{N} \tag{4}$$

verifying the boundary condition $f_A(0) = 0$. The above equation is called Chen–Stein's equation (see e.g. [3]). Throughout this section, given $f : \mathbb{N} \to \mathbb{R}$, we set $\|f\|_\infty := \sup_{k \in \mathbb{N}} |f(k)|$. The following bounds for the solution of the Chen–Stein equation hold (see Lemma 1.1.1 and Remark 1.1.2 in [3]):

$$\|f_A\|_\infty \leq \min\left(1, \sqrt{\frac{2}{\theta e}}\right), \qquad \|\Delta f_A\|_\infty \leq \frac{1 - e^{-\theta}}{\theta}. \tag{5}$$

In addition, by the latter inequality in (5) and the relation $\|\Delta^2 f_A\|_\infty \leq 2\|\Delta f_A\|_\infty$ (which is a straightforward consequence of the triangle inequality), we deduce

$$\|\Delta^2 f_A\|_\infty \leq \frac{2(1 - e^{-\theta})}{\theta}. \tag{6}$$

We finally recall that the total variation distance between (the laws of) the random variables X_i, $i = 1, 2$, with values on $\mathbb{N}$ and defined respectively on the probability spaces $(\Omega_i, \mathcal{A}_i, P_i)$, is defined by

$$d_{\mathrm{TV}}(X_1, X_2) := \sup_{A \subseteq \mathbb{N}} \big|P_1(X_1 \in A) - P_2(X_2 \in A)\big|.$$

Of course, the topology induced by d_{TV} on the class of probability measures on $\mathbb{N}$ is strictly stronger than the topology induced by the convergence in distribution.

The following general inequalities hold.

Theorem 3.1. *Assume that $u : \mathbb{R} \to \mathbb{N}$ is a measurable function such that*

$$\int_{\mathbb{R}} u(t)^2 \mathrm{E}\big[\lambda(t)\big] \, \mathrm{d}t < \infty, \tag{7}$$

$$\int_{\mathbb{R} \times \mathbb{R}_+} \left(\int_t^\infty u(s)^2 \mathrm{E}\big[\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s\right) \mathrm{d}t \, \mathrm{d}z < \infty, \tag{8}$$

and

$$\int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\left[\left(\int_t^\infty u(s)\big|D_{(t,z)}\lambda(s)\big| \, \mathrm{d}s\right)^2\right] \mathrm{d}t \, \mathrm{d}z < \infty. \tag{9}$$

Moreover, suppose that

For Lebesgue almost all $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, the (random) function $s \mapsto \big|D_{(t,z)}\lambda(s)\big|$

is locally integrable on (t, ∞), almost surely. $\tag{10}$

Then

$$d_{\mathrm{TV}}\big(N(u), \mathrm{Po}(\theta)\big) \leq \frac{1 - e^{-\theta}}{\theta} \mathfrak{G} + \min\left(1, \sqrt{\frac{2}{\theta e}}\right) \mathrm{E}\big[\big|\Lambda(u) - \theta\big|\big], \tag{11}$$

where

$$\mathfrak{G} := \int_{\mathbb{R}} u(t)\big(u(t)^2 - 1\big) \mathrm{E}\big[\lambda(t)\big] \, \mathrm{d}t + \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\bigg(\int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z$$

$$+ \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\mathrm{E}\bigg[\bigg(\int_t^{\infty} \mathbb{1}_{(0,\lambda(t)]}(z)u(s)\big|D_{(t,z)}\lambda(s)\big| \, \mathrm{d}s\bigg)^2\bigg] \, \mathrm{d}t \, \mathrm{d}z$$

$$+ 2\int_{\mathbb{R} \times \mathbb{R}_+} u(t)\sqrt{\int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s \, \mathrm{E}\bigg[\bigg(\int_t^{\infty} \mathbb{1}_{(0,\lambda(t)]}(z)u(s)\big|D_{(t,z)}\lambda(s)\big| \, \mathrm{d}s\bigg)^2\bigg]} \, \mathrm{d}t \, \mathrm{d}z$$

$$+ 2\int_{\mathbb{R} \times \mathbb{R}_+} u(t)^2 \bigg(\int_t^{\infty} u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z.$$

Corollary 3.2. *Assume that* $u : \mathbb{R} \to \mathbb{N}$ *is a measurable function such that* $u \in L^1(\mathbb{R}, \mathrm{d}x)$, (7), (8) *and* (10) *hold, and*

$$\int_{\mathbb{R} \times \mathbb{R}_+} \bigg(\int_t^{\infty} u(s)\mathrm{E}\big[\big|D_{(t,z)}\lambda(s)\big|^2\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z < \infty. \tag{12}$$

Then

$$d_{\mathrm{TV}}\big(N(u), \mathrm{Po}(\theta)\big) \le \frac{1 - \mathrm{e}^{-\theta}}{\theta} \mathfrak{G}' + \min\bigg(1, \sqrt{\frac{2}{\theta \mathrm{e}}}\bigg) \mathrm{E}\big[\big|\Lambda(u) - \theta\big|\big],$$

where

$$\mathfrak{G}' := \int_{\mathbb{R}} u(t)\big(u(t)^2 - 1\big) \mathrm{E}\big[\lambda(t)\big] \, \mathrm{d}t + \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\bigg(\int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z$$

$$+ \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\bigg(\int_t^{\infty} u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|^2\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z$$

$$+ 2\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}^{1/2}$$

$$\times \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\sqrt{\int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s \int_t^{\infty} u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|^2\big] \, \mathrm{d}s} \, \mathrm{d}t \, \mathrm{d}z$$

$$+ 2\int_{\mathbb{R} \times \mathbb{R}_+} u(t)^2 \bigg(\int_t^{\infty} u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big] \, \mathrm{d}s\bigg) \, \mathrm{d}t \, \mathrm{d}z.$$

Remark 3.3. *If* N *is a Poisson process on* $\mathbb{R}$ *with mean measure* $f(t) \, \mathrm{d}t$ *and* $u : \mathbb{R} \to \mathbb{N}$ *is a function such that*

$$\int_{\mathbb{R}} u(t)f(t) \, \mathrm{d}t > 0 \quad and \quad \int_{\mathbb{R}} u(t)^2 f(t) \, \mathrm{d}t < \infty,$$

then by Theorem 3.1 *(with* $\theta := \|uf\|_{L^1(\mathbb{R},\mathrm{d}x)}$, $\lambda(t) = f(t)$, $t \in \mathbb{R}$, *and therefore* $D_{(t,z)}\lambda(s) = 0$, *for any* $s, t \in \mathbb{R}$ *and* $z \in \mathbb{R}_+$) *we have*

$$d_{\mathrm{TV}}\bigg(N(u), \mathrm{Po}\bigg(\int_{\mathbb{R}} u(t)f(t) \, \mathrm{d}t\bigg)\bigg) \le \frac{1 - \mathrm{e}^{-\int_{\mathbb{R}} u(t)f(t) \, \mathrm{d}t}}{\int_{\mathbb{R}} u(t)f(t) \, \mathrm{d}t} \int_{\mathbb{R}} u(t)\big(u(t)^2 - 1\big)f(t) \, \mathrm{d}t. \tag{13}$$

As expected, the right-hand side of the inequality (13) *is equal to zero if we take* $u(t) := \mathbb{1}_A(t)$, *for a Borel set* $A \subseteq \mathbb{R}$ *such that* $\int_A f(t) \, \mathrm{d}t \in (0, \infty)$.

Remark 3.4. *The bound* (11) *has the classical structure of the error estimates in the Poisson approximation by the total variation distance, see the seminal papers* [4,5] *and* [14]. *Indeed, the bound* (11) *consists of the sum of two terms involving the magic Stein's factors* $\frac{1-e^{-\theta}}{\theta}$ *and* $\min(1, \sqrt{\frac{2}{\theta e}})$.

Proof of Theorem 3.1. We divide the proof in 4 steps. In the first step we give a Taylor expansion for the discrete Malliavin derivative of $f_A(N(u))$; in the second step we apply the Chen–Stein method; in the third step we conclude the proof; in the fourth step we check the technical condition (20) below.

Step 1: *a Taylor expansion for* $D_{(t,z)} f_A(N(u))$. As shown in the proof of Theorem 3.1 in [32], for any $f : \mathbb{N} \to \mathbb{R}$ and any $k, a \in \mathbb{N}$,

$$\left| f(k) - f(a) - \Delta f(a)(k-a) \right| \leq \frac{\|\Delta^2 f\|_\infty}{2} \left| (k-a)(k-a-1) \right|. \tag{14}$$

For any $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, define

$$R_{(t,z)}(u) := D_{(t,z)} f_A(N(u)) - \Delta f_A(N(u)) D_{(t,z)} N(u). \tag{15}$$

We have

$$\begin{aligned}
D_{(t,z)} f_A(N(u)) &= D_{(t,z)} f_A\left(\int_{\mathbb{R} \times \mathbb{R}_+} u(s) \mathbb{1}_{(0,\lambda(s)]}(v) \overline{N}(\mathrm{d}s \times \mathrm{d}v) \right) \\
&= f_A\left(\int_{\mathbb{R} \times \mathbb{R}_+} u(s) \mathbb{1}_{(0,\varphi(s,(\overline{N}+\varepsilon_{(t,z)})|_{(-\infty,s)})]}(v) (\overline{N}+\varepsilon_{(t,z)})(\mathrm{d}s \times \mathrm{d}v) \right) \\
&\quad - f_A\left(\int_{\mathbb{R} \times \mathbb{R}_+} u(s) \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v) \overline{N}(\mathrm{d}s \times \mathrm{d}v) \right) \\
&= f_A(k) - f_A(a)
\end{aligned}$$

and $D_{(t,z)} N(u) = k - a$, where

$$k := \int_{\mathbb{R} \times \mathbb{R}_+} u(s) \mathbb{1}_{(0,\varphi(s,(\overline{N}+\varepsilon_{(t,z)})|_{(-\infty,s)})]}(v) \overline{N}(\mathrm{d}s \times \mathrm{d}v) + u(t) \mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) \tag{16}$$

and

$$a := N(u) = \int_{\mathbb{R} \times \mathbb{R}_+} u(s) \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v) \overline{N}(\mathrm{d}s \times \mathrm{d}v). \tag{17}$$

Therefore, $R_{(t,z)}(u) = f_A(a) - f_A(k) - \Delta f_A(a)(k-a)$ and so by (14) we deduce

$$\left| R_{(t,z)}(u) \right| \leq \frac{\|\Delta^2 f_A\|_\infty}{2} \left| D_{(t,z)} N(u) \left(D_{(t,z)} N(u) - 1 \right) \right|. \tag{18}$$

Step 2: *application of the Chen–Stein method.* By the Chen–Stein equation (4), for any $A \subseteq \mathbb{N}$, we have

$$\begin{aligned}
P\left(\mathrm{Po}(\theta) \in A \right) - P\left(N(u) \in A \right) &= \mathrm{E}\left[N(u) f_A(N(u)) - \theta f_A(N(u)+1) \right] \\
&= \mathrm{E}\left[\left(N(u) - \Lambda(u) \right) f_A(N(u)) - \Lambda(u) \left(f_A(N(u)+1) - f_A(N(u)) \right) \right] \\
&\quad + \mathrm{E}\left[\left(\Lambda(u) - \theta \right) f_A(N(u)+1) \right] \\
&= \mathrm{E}\left[\delta(u) f_A(N(u)) - \Lambda(u) \Delta f_A(N(u)) \right] \\
&\quad + \mathrm{E}\left[\left(\Lambda(u) - \theta \right) f_A(N(u)+1) \right]. \tag{19}
\end{aligned}$$

Note that, due to (1)–(2), $f_A(N(u))$ is of the form $\psi(\overline{N})$ for some functional ψ and that $f_A(N(u)) \in L^2(\Omega, dP)$ since f_A is bounded. So by Lemma 2.2 we deduce

$$f_A\big(N(u)\big) - \mathrm{E}\big[f_A\big(N(u)\big)\big] = \int_{\mathbb{R}\times[0,\infty)} g(t,z)\big(\overline{N}(dt \times dz) - dt\,dz\big),$$

where $g(t,\omega,z)$ is a $\mathcal{P}(\mathcal{F}^N) \otimes \mathcal{B}([0,\infty))$-measurable version of $\mathrm{E}[D_{(t,z)}f_A(N(u))|\mathcal{F}^N_{t^-}](\omega)$. We shall check later on (see Step 4) that

$$g \text{ is integrable and square integrable with respect to the measure } dt\,dz\,dP(\omega). \tag{20}$$

Note that the mapping $(t,\omega,z) \mapsto u(t)\mathbb{1}_{(0,\lambda(t)]}(z)$ is $\mathcal{P}(\mathcal{F}^N) \otimes \mathcal{B}(\mathbb{R}_+)$-measurable and, due to assumption (7) and the inequality $u(t) \leq u(t)^2$, $t \in \mathbb{R}$ (which holds since the function u takes values on $\mathbb{N}$), it is integrable and square integrable with respect to the measure $dt\,dz\,dP(\omega)$. So by Theorem 3 in [10] (see formulas (19) and (20) therein), the fact that $\{\lambda(t)\}_{t\in\mathbb{R}}$ is adapted with respect to the filtration $\{\mathcal{F}^N_{t^-}\}_{t\in\mathbb{R}}$ and (15), we have

$$
\begin{aligned}
\mathrm{E}\big[\delta(u)f_A\big(N(u)\big)\big] &= \mathrm{E}\big[\delta(u)\big(f_A\big(N(u)\big) - \mathrm{E}\big[f_A\big(N(u)\big)\big]\big)\big] \\
&= \mathrm{E}\bigg[\bigg(\int_{\mathbb{R}\times[0,\infty)} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(\overline{N}(dt \times dz) - dt\,dz\big)\bigg) \\
&\qquad\quad \times \bigg(\int_{\mathbb{R}\times[0,\infty)} g(t,z)\big(\overline{N}(dt \times dz) - dt\,dz\big)\bigg)\bigg] \\
&= \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)g(t,z)\,dt\,dz\bigg] \\
&= \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)D_{(t,z)}f_A\big(N(u)\big)\,dt\,dz\bigg] \\
&= \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(\Delta f_A\big(N(u)\big)D_{(t,z)}N(u) + R_{(t,z)}(u)\big)\,dt\,dz\bigg].
\end{aligned}
$$

Combining this relation with (19), we deduce

$$
\begin{aligned}
P\big(\mathrm{Po}(\theta) \in A\big) - P\big(N(u) \in A\big) &= \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(\Delta f_A\big(N(u)\big)D_{(t,z)}N(u) + R_{(t,z)}(u)\big)\,dt\,dz \\
&\qquad\quad - \Lambda(u)\Delta f_A\big(N(u)\big)\bigg] + \mathrm{E}\big[\big(\Lambda(u) - \theta\big)f_A\big(N(u) + 1\big)\big] \\
&= \mathrm{E}\bigg[\Delta f_A\big(N(u)\big)\bigg(\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)D_{(t,z)}N(u)\,dt\,dz - \Lambda(u)\bigg)\bigg] \\
&\qquad + \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)R_{(t,z)}(u)\,dt\,dz\bigg] + \mathrm{E}\big[\big(\Lambda(u) - \theta\big)f_A\big(N(u) + 1\big)\big].
\end{aligned}
$$

By taking absolute values on both sides, as well as by applying the estimates (5), (6) and (18) we have

$$
\begin{aligned}
d_{\mathrm{TV}}\big(N(u), \mathrm{Po}(\theta)\big) &\leq \frac{1 - e^{-\theta}}{\theta}\bigg(\mathrm{E}\bigg[\bigg|\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(D_{(t,z)}N(u) - 1\big)\,dt\,dz\bigg|\bigg] \\
&\qquad + \mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}N(u)\big(D_{(t,z)}N(u) - 1\big)\big|\,dt\,dz\bigg]\bigg) \\
&\qquad + \min\bigg(1, \sqrt{\frac{2}{\theta e}}\bigg)\mathrm{E}\big[|\Lambda(u) - \theta|\big].
\end{aligned} \tag{21}
$$

Step 3: *conclusion of the proof.* For $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, we have

$$
\begin{aligned}
D_{(t,z)}N(u) &= u(t)\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) \\
&\quad + \int_{\mathbb{R} \times \mathbb{R}_+} u(s)\big(\mathbb{1}_{(0,\varphi(s,(\overline{N}+\varepsilon_{(t,z)})|_{(-\infty,s)})]}(v) - \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v)\big)\overline{N}(\mathrm{d}s \times \mathrm{d}v) \\
&= u(t)\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) \\
&\quad + \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{t<s}u(s)\big(\mathbb{1}_{(0,\varphi(s,(\overline{N}+\varepsilon_{(t,z)})|_{(-\infty,s)})]}(v) - \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v)\big)\overline{N}(\mathrm{d}s \times \mathrm{d}v) \\
&\quad + \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{t\geq s}u(s)\big(\mathbb{1}_{(0,\varphi(s,(\overline{N}+\varepsilon_{(t,z)})|_{(-\infty,s)})]}(v) - \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v)\big)\overline{N}(\mathrm{d}s \times \mathrm{d}v) \\
&= u(t)\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) \\
&\quad + \int_{\mathbb{R} \times \mathbb{R}_+} \mathbb{1}_{t<s}u(s)\big(\mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)}+\varepsilon_{(t,z)})]}(v) - \mathbb{1}_{(0,\varphi(s,\overline{N}|_{(-\infty,s)})]}(v)\big)\overline{N}(\mathrm{d}s \times \mathrm{d}v) \\
&= u(t)\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) + \int_{(t,\infty)} u_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s),
\end{aligned}
\tag{22}
$$

where

$$
u_{(t,z)}(s) := u(s)\,\mathrm{sign}\big(\varphi(s, \overline{N}|_{(-\infty,s)} + \varepsilon_{(t,z)}) - \varphi(s, \overline{N}|_{(-\infty,s)})\big), \quad s > t
$$

and

$$
\begin{aligned}
&N_{(t,z)}(\mathrm{d}s) \\
&\quad := \overline{N}\big(\mathrm{d}s \times \big(\varphi(s, \overline{N}|_{(-\infty,s)} + \varepsilon_{(t,z)}) \wedge \varphi(s, \overline{N}|_{(-\infty,s)}), \varphi(s, \overline{N}|_{(-\infty,s)} + \varepsilon_{(t,z)}) \vee \varphi(s, \overline{N}|_{(-\infty,s)})\big]\big), \quad s > t.
\end{aligned}
$$

Here, for $a, b \in \mathbb{R}$, $a \wedge b$ and $a \vee b$ denote the minimum and the maximum between a and b. For any $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, the processes $\{\varphi(s, \overline{N}|_{(-\infty,s)} + \varepsilon_{(t,z)})\}_{s>t}$ and $\{\varphi(s, \overline{N}|_{(-\infty,s)})\}_{s>t}$ are $\mathcal{P}(\{\mathcal{F}_s^N\}_{s>t})$-measurable. So by Lemma 2.1 (which may be applied due to condition (10)), for Lebesgue almost all $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, the point process $N_{(t,z)}$ has $\{\mathcal{F}_s^N\}_{s>t}$-stochastic intensity $\{|D_{(t,z)}\lambda(s)|\}_{s>t}$. By (22) we have

$$
\begin{aligned}
&u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(D_{(t,z)}N(u) - 1\big) \\
&= u(t)\big(u(t) - 1\big)\mathbb{1}_{(0,\lambda(t)]}(z) \\
&\quad + u(t)\int_{(t,\infty)} \overline{u}_{(t,z)}(s)\,\mathrm{sign}\big(\varphi(s, \overline{N}|_{(-\infty,s)} + \varepsilon_{(t,z)}) - \varphi(s, \overline{N}|_{(-\infty,s)})\big)N_{(t,z)}(\mathrm{d}s),
\end{aligned}
\tag{23}
$$

where

$$
\overline{u}_{(t,z)}(s) := u(s)\mathbb{1}_{(0,\lambda(t)]}(z).
$$

For any $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, the process $\{\overline{u}_{(t,z)}(s)\}_{s>t}$ is $\{\mathcal{F}_s^N\}_{s>t}$-predictable, and therefore by e.g. Theorem 2 in [10], for Lebesgue almost all $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, we have

$$
\mathrm{E}\left[\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s)\right] = \int_t^\infty u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s.
\tag{24}
$$

Consequently,

$$
\mathrm{E}\left[\left|\int_{\mathbb{R} \times \mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big(D_{(t,z)}N(u) - 1\big)\,\mathrm{d}t\,\mathrm{d}z\right|\right] \leq \mathfrak{G}_1,
\tag{25}
$$

where

$$\mathfrak{G}_1 := \int_{\mathbb{R}} u(t)\big(u(t) - 1\big)\mathrm{E}\big[\lambda(t)\big]\,\mathrm{d}t + \int_{\mathbb{R}\times\mathbb{R}_+} u(t)\bigg(\int_t^\infty u(s)\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\bigg)\mathrm{d}t\,\mathrm{d}z.$$

By (22) and (23) we have, respectively,

$$\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}N(u)\big| \le u(t) + \int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s)$$

and

$$u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}N(u) - 1\big| \le u(t)\big(u(t) - 1\big)\mathbb{1}_{(0,\lambda(t)]}(z) + u(t)\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s).$$

Therefore

$$u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}N(u)\big(D_{(t,z)}N(u) - 1\big)\big|$$

$$\le u(t)^2\big(u(t) - 1\big)\mathbb{1}_{(0,\lambda(t)]}(z) + u(t)\bigg(\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s)\bigg)^2 + u(t)\big(2u(t) - 1\big)\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s).$$

Setting $\delta_{(t,z)}(\mathrm{d}s) := N_{(t,z)}(\mathrm{d}s) - |D_{(t,z)}\lambda(s)|\,\mathrm{d}s$, we have

$$\bigg(\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s)\bigg)^2 = \bigg(\int_{(t,\infty)} \overline{u}_{(t,z)}(s)\delta_{(t,z)}(\mathrm{d}s) + \int_t^\infty \overline{u}_{(t,z)}(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s\bigg)^2$$

$$= \bigg(\int_{(t,\infty)} \overline{u}_{(t,z)}(s)\delta_{(t,z)}(\mathrm{d}s)\bigg)^2 + \bigg(\int_t^\infty \overline{u}_{(t,z)}(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s\bigg)^2$$

$$+ 2\int_{(t,\infty)} \overline{u}_{(t,z)}(s)\delta_{(t,z)}(\mathrm{d}s)\int_t^\infty \overline{u}_{(t,z)}(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s.$$

Taking the expectation, by formula (20) in [10] (which may be applied due to (8)) and the Cauchy–Schwarz inequality, for Lebesgue almost all $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, we have

$$\mathrm{E}\bigg[\bigg(\int_{(t,\infty)} \overline{u}_{(t,z)}(s)N_{(t,z)}(\mathrm{d}s)\bigg)^2\bigg]$$

$$\le \int_t^\infty u(s)^2\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s + \mathrm{E}\bigg[\bigg(\int_t^\infty \mathbb{1}_{(0,\lambda(t)]}(z)u(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s\bigg)^2\bigg]$$

$$+ 2\sqrt{\int_t^\infty u(s)^2\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s}\sqrt{\mathrm{E}\bigg[\bigg(\int_t^\infty \mathbb{1}_{(0,\lambda(t)]}(z)u(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s\bigg)^2\bigg]}.$$

Therefore, employing (24) we deduce

$$\mathrm{E}\bigg[\int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}N(u)\big(D_{(t,z)}N(u) - 1\big)\big|\,\mathrm{d}t\,\mathrm{d}z\bigg] \le \mathfrak{G}_2, \tag{26}$$

where

$$\mathfrak{G}_2 := \int_{\mathbb{R}} u(t)^2\big(u(t) - 1\big)\mathrm{E}\big[\lambda(t)\big]\,\mathrm{d}t + \int_{\mathbb{R}\times\mathbb{R}_+} u(t)\bigg(\int_t^\infty u(s)^2\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\bigg)\mathrm{d}t\,\mathrm{d}z$$

$$+ \int_{\mathbb{R}\times\mathbb{R}_+} u(t)\mathrm{E}\bigg[\bigg(\int_t^\infty \mathbb{1}_{(0,\lambda(t)]}(z)u(s)\big|D_{(t,z)}\lambda(s)\big|\,\mathrm{d}s\bigg)^2\bigg]\mathrm{d}t\,\mathrm{d}z$$

$$+ 2 \int_{\mathbb{R} \times \mathbb{R}_+} u(t) \sqrt{\int_t^\infty u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|\big] \mathrm{d}s \, \mathrm{E}\bigg[\bigg(\int_t^\infty \mathbb{1}_{(0,\lambda(t)]}(z) u(s)\big| D_{(t,z)}\lambda(s)\big| \mathrm{d}s\bigg)^2\bigg]} \, \mathrm{d}t \, \mathrm{d}z$$

$$+ \int_{\mathbb{R} \times \mathbb{R}_+} u(t)\big(2u(t)-1\big)\bigg(\int_t^\infty u(s) \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|\big] \mathrm{d}s\bigg) \mathrm{d}t \, \mathrm{d}z.$$

The claim follows by (21), (25), (26) and noticing that $\mathfrak{G} = \mathfrak{G}_1 + \mathfrak{G}_2$.

Step 4: *proof of (20)*. Let k and a be the non-negative and integer-valued random variables defined by (16) and (17). For all $(t, z) \in \mathbb{R} \times \mathbb{R}_+$, if $k \geq a$ we have

$$D_{(t,z)} f_A\big(N(u)\big) = f_A(k) - f_A(a) = \sum_{i=a+1}^k \big(f_A(i) - f_A(i-1)\big) = \sum_{i=a+1}^k \Delta f_A(i-1)$$

and if $k < a$ we have

$$D_{(t,z)} f_A\big(N(u)\big) = f_A(k) - f_A(a) = -\big(f_A(a) - f_A(k)\big) = -\sum_{i=k+1}^a \Delta f_A(i-1).$$

Therefore, using the second inequality in (5) we deduce $|D_{(t,z)} f_A(N(u))| \leq |k - a| = |D_{(t,z)} N(u)|$. So (20) follows if we show

$$\int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\big[\big| D_{(t,z)} N(u)\big|^2\big] \mathrm{d}t \, \mathrm{d}z < \infty \tag{27}$$

(indeed $|D_{(t,z)} N(u)|$ takes values on $\mathbb{N}$ and so $|D_{(t,z)} N(u)| \leq |D_{(t,z)} N(u)|^2$). We have

$$\int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\big[\big| D_{(t,z)} N(u)\big|^2\big] \mathrm{d}t \, \mathrm{d}z \leq 2 \int_\mathbb{R} u(t)^2 \mathrm{E}\big[\lambda(t)\big] \mathrm{d}t + 2 \int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\bigg[\bigg(\int_{(t,\infty)} u_{(t,z)}(s) N_{(t,z)}(\mathrm{d}s)\bigg)^2\bigg] \mathrm{d}t \, \mathrm{d}z \tag{28}$$

$$\leq 2 \int_\mathbb{R} u(t)^2 \mathrm{E}\big[\lambda(t)\big] \mathrm{d}t + 4 \int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\bigg[\bigg(\int_{(t,\infty)} u(s) \delta_{(t,z)}(\mathrm{d}s)\bigg)^2\bigg] \mathrm{d}t \, \mathrm{d}z$$

$$+ 4 \int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\bigg[\bigg(\int_t^\infty u(s)\big| D_{(t,z)}\lambda(s)\big| \mathrm{d}s\bigg)^2\bigg] \mathrm{d}t \, \mathrm{d}z \tag{29}$$

$$= 2 \int_\mathbb{R} u(t)^2 \mathrm{E}\big[\lambda(t)\big] \mathrm{d}t + 4 \int_{\mathbb{R} \times \mathbb{R}_+} \bigg(\int_t^\infty u(s)^2 \mathrm{E}\big[\big| D_{(t,z)}\lambda(s)\big|\big] \mathrm{d}s\bigg) \mathrm{d}t \, \mathrm{d}z$$

$$+ 4 \int_{\mathbb{R} \times \mathbb{R}_+} \mathrm{E}\bigg[\bigg(\int_t^\infty u(s)\big| D_{(t,z)}\lambda(s)\big| \mathrm{d}s\bigg)^2\bigg] \mathrm{d}t \, \mathrm{d}z, \tag{30}$$

and this latter term is finite due to assumptions (7), (8) and (9). Here, (28) follows by (22) and the elementary inequality $(b + c)^2 \leq 2b^2 + 2c^2$, $b, c \in \mathbb{R}$; (29) is a consequence of the relation

$$\bigg(\int_{(t,\infty)} u_{(t,z)}(s) N_{(t,z)}(\mathrm{d}s)\bigg)^2 \leq \bigg(\int_{(t,\infty)} u(s) N_{(t,z)}(\mathrm{d}s)\bigg)^2$$

$$= \bigg(\int_{(t,\infty)} u(s) \delta_{(t,z)}(\mathrm{d}s) + \int_t^\infty u(s)\big| D_{(t,z)}\lambda(s)\big| \mathrm{d}s\bigg)^2$$

$$\leq 2\bigg(\int_{(t,\infty)} u(s) \delta_{(t,z)}(\mathrm{d}s)\bigg)^2 + 2\bigg(\int_t^\infty u(s)\big| D_{(t,z)}\lambda(s)\big| \mathrm{d}s\bigg)^2;$$

finally, (30) follows applying formula (20) in [10]. □

Proof of Corollary 3.2. By Jensen's inequality, for any non-negative stochastic process $\{Z(s)\}_{s \geq 0}$, we have

$$
\begin{aligned}
\mathrm{E}\left[\left(\int_t^\infty u(s)Z(s)\,\mathrm{d}s\right)^2\right] &= \|u\|_{L^1(\mathbb{R},\mathrm{d}x)}^2 \mathrm{E}\left[\left(\int_0^\infty \mathbb{1}_{(t,\infty)}(s)Z(s)\big(u(s)/\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\big)\,\mathrm{d}s\right)^2\right] \\
&\leq \|u\|_{L^1(\mathbb{R},\mathrm{d}x)}^2 \mathrm{E}\left[\int_0^\infty \mathbb{1}_{(t,\infty)}(s)Z(s)^2\big(u(s)/\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\big)\,\mathrm{d}s\right] \\
&= \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \int_t^\infty u(s)\mathrm{E}\big[Z(s)^2\big]\,\mathrm{d}s.
\end{aligned}
\tag{31}
$$

Setting $Z(s) := |D_{(t,z)}\lambda(s)|$ in (31) we have that condition (9) is guaranteed by the integrability of u and condition (12). Furthermore, setting $Z(s) := \mathbb{1}_{(0,\lambda(t)]}(z)|D_{(t,z)}\lambda(s)|$ in (31) we easily have $\mathfrak{G} \leq \mathfrak{G}'$, and so the claim follows by Theorem 3.1. $\qquad\square$

4. Poisson approximation of stationary nonlinear Hawkes processes with bounded and Lipschitz dynamics

A nonlinear Hawkes process with parameters (ϕ, h) is a point process N on $\mathbb{R}$ with $\mathcal{F}^N$-stochastic intensity of the form

$$
t \mapsto \phi\left(\int_{(-\infty,t)} h(t-s)N(\mathrm{d}s)\right), \quad t \in \mathbb{R},
\tag{32}
$$

where $\phi : \mathbb{R} \to [0, \infty)$ and $h : (0, \infty) \to \mathbb{R}$ are measurable functions. A particular case is the self-exciting process (or linear Hawkes process) with parameters (ν, h), for which $\phi(x) := \nu + x$, for some constant $\nu > 0$, and h is non-negative.

In the seminal paper [9], the authors proved that if ϕ is Lipschitz continuous with $\phi(0) > 0$ and Lipschitz constant α such that $\alpha\mu < 1$, where $\mu := \|h\|_{L^1(\mathbb{R}_+,\mathrm{d}x)}$, then there exists a unique stationary distribution of N with dynamics (32) and finite intensity $\lambda := \mathrm{E}[N((0,1])]$.

The stationary solution is constructed by embedding in a bivariate Poisson process, i.e. it turns out

$$
N(\mathrm{d}t) = \overline{N}\big(\mathrm{d}t \times \big(0, \lambda(t)\big]\big), \qquad \lambda(t) = \phi\left(\int_{(-\infty,t)} h(t-s)N(\mathrm{d}s)\right), \quad t \in \mathbb{R}
$$

$\lambda \in (0, \infty)$ and $\lambda(t) = \varphi(t, \overline{N}|_{(-\infty,t)})$, $t \in \mathbb{R}$, for some measurable functional $\varphi : \mathbb{R} \times \mathcal{N} \to [0, \infty)$.

4.1. *Poisson bounds*

The following theorem, whose proof is given below, provides a bound for the total variation distance between a first order stochastic integral, with respect to a stationary and nonlinear Hawkes process with bounded and Lipschitz dynamics, and a Poisson random variable having the same mean of the stochastic integral.

Theorem 4.1. *Assume that $\phi : \mathbb{R} \to [0, \infty)$, $\phi(0) > 0$, is bounded and Lipschitz continuous with Lipschitz constant α and that $h : [0, \infty) \to \mathbb{R}$ is measurable and such that $\alpha\mu < 1$. Let N be the stationary nonlinear Hawkes process with parameters (ϕ, h) and finite intensity $\lambda \in (0, \infty)$. If $u : \mathbb{R} \to \mathbb{N}$, $\|u\|_{L^1(\mathbb{R},\mathrm{d}x)} > 0$ and $u \in L^2(\mathbb{R}, \mathrm{d}x)$, then*

$$
d_{\mathrm{TV}}\big(N(u), \mathrm{Po}\big(\lambda\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\big)\big) \leq \frac{1 - \mathrm{e}^{-\lambda\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}}{\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}\mathfrak{D}_1 + \min\left(1, \sqrt{\frac{2}{\lambda\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\mathrm{e}}}\right)\mathfrak{D}_2,
\tag{33}
$$

where

$$
\begin{aligned}
\mathfrak{D}_1 := \;&\|u\|_{L^3(\mathbb{R},\mathrm{d}x)}^3 - \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \\
&+ 3\min\left\{\|\phi\|_\infty\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\|u\|_{L^2(\mathbb{R},\mathrm{d}x)}^2, \frac{\alpha\mu\|u\|_{L^2(\mathbb{R},\mathrm{d}x)}\|u\|_{L^4(\mathbb{R},\mathrm{d}x)}^2}{1 - \alpha\mu}\right\}
\end{aligned}
$$

$$+ \min\left\{ \|\phi\|_\infty^2 \|u\|_{L^1(\mathbb{R},\mathrm{d}x)}^3, \frac{\alpha\mu\|\phi\|_\infty \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \|u\|_{L^2(\mathbb{R},\mathrm{d}x)}^2}{1 - \alpha\mu} \right\}$$

$$+ 2\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}^{1/2} \min\left\{ \|\phi\|_\infty^2 \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \|u\|_{L^2(\mathbb{R},\mathrm{d}x)}^2, \frac{\alpha\mu\|\phi\|_\infty}{1 - \alpha\mu} \|u\|_{L^2(\mathbb{R},\mathrm{d}x)} \|u\|_{L^4(\mathbb{R},\mathrm{d}x)}^2 \right\}$$

and

$$\mathfrak{D}_2 := \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \min\left\{ \|\phi\|_\infty, \lambda\alpha\mu + \big|\phi(0) - \lambda\big| \right\}.$$

Remark 4.2. *Relation* (33) *depends explicitly on the function* u, *the constants* α, μ, $\|\phi\|_\infty$ *and* $\phi(0)$, *which are known given the parameters* (ϕ, h), *and the intensity* λ, *which is in general unknown. However,* λ *may be estimated by Monte-Carlo simulation as follows. By the results in* [9], *under the assumptions of Theorem* 4.1 *there exists a nonlinear Hawkes process on* $\mathbb{R}_+$ *with parameters* (ϕ, h), *i.e. a point process* N' *on* $\mathbb{R}_+$ *with stochastic intensity*

$$t \mapsto \phi\left(\int_{(0,t)} h(t - s) N'(\mathrm{d}s) \right), \quad t \in \mathbb{R}_+.$$

We note that, due to the boundedness of the stochastic intensity, N' *may be simulated over finite intervals by using the algorithms in* [25] *and* [30], *see also Lemma* 2 *in* [9]. *Moreover, by the stability results in* [9] *we have that, for any* $t > 0$, $N'((s, s + t]) \to N((0, t])$ *in law, as* $s \to \infty$ *(as usual, we denote by* N *the stationary nonlinear Hawkes process with parameters* (ϕ, h)*). Consequently,*

$$\lim_{n \to \infty} \lim_{s \to \infty} \mathrm{E}\big[\min\{N'((s, s+1]), n\}\big] = \lim_{n \to \infty} \mathrm{E}\big[\min\{N((0,1]), n\}\big] = \lambda.$$

We also note that, under the assumptions of Theorem 4.1, *we have*

$$\lambda \in \big[\phi(0)/(1 + \alpha\mu), \min\{\phi(0)/(1 - \alpha\mu), \|\phi\|_\infty\}\big].$$

Indeed, we clearly have $\lambda \leq \|\phi\|_\infty$. *Moreover, by the Lipschitz property of* ϕ

$$\big|\lambda(t) - \phi(0)\big| \leq \alpha \int_{(-\infty,t)} \big|h(t-s)\big| N(\mathrm{d}s),$$

and therefore

$$\phi(0) - \alpha \int_{(-\infty,t)} \big|h(t-s)\big| N(\mathrm{d}s) \leq \lambda(t) \leq \phi(0) + \alpha \int_{(-\infty,t)} \big|h(t-s)\big| N(\mathrm{d}s).$$

Taking the mean, from the first inequality we deduce $\lambda \geq \phi(0)/(1 + \alpha\mu)$ *and from the second inequality we get* $\lambda \leq \phi(0)/(1 - \alpha\mu)$.

Letting $\widehat{\lambda} \in [\phi(0)/(1 + \alpha\mu), \min\{\phi(0)/(1 - \alpha\mu), \|\phi\|_\infty\}]$ denote an estimate of λ (see Remark 4.2 above), we have the following explicit Poisson bound, whose proof is provided below.

Theorem 4.3. *Under assumptions and notation of Theorem* 4.1, *we have*

$$d_{\mathrm{TV}}\big(N(u), \mathrm{Po}\big(\widehat{\lambda}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\big)\big) \leq \frac{1 - \mathrm{e}^{-\widehat{\lambda}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}}{\widehat{\lambda}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}} \min\big\{\phi(0)/(1 - \alpha\mu), \|\phi\|_\infty\big\} \mathfrak{D}_1$$

$$+ \min\left(1, \sqrt{\frac{2}{\widehat{\lambda}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\mathrm{e}}}\right) \mathfrak{D}_3, \tag{34}$$

where $\mathfrak{D}_1$ *is defined as in the statement of Theorem* 4.1 *and*

$$\mathfrak{D}_3 := \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \min\big\{\|\phi\|_\infty, \min\{\phi(0)/(1 - \alpha\mu), \|\phi\|_\infty\}\alpha\mu + \big|\phi(0) - \widehat{\lambda}\big|\big\}.$$

Proof of Theorem 4.1. We apply Corollary 3.2. For the sake of readability, we divide the proof in 3 steps.

Step 1: *preliminary inequalities*. In this step we derive 4 inequalities (see relations (35), (38), (39) and (41)) which will be used later on. By the embedding construction, for $s, t \in \mathbb{R}$, we have

$$
\lambda(s) = \varphi(s, \overline{N}|_{(-\infty,s)}) = \phi\left(\int_{(-\infty,s)\times\mathbb{R}_+} h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right)
$$

$$
= \phi\left(\int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u\leq t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right.
$$

$$
\left. + \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u>t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right).
$$

Consequently, for $s, t \in \mathbb{R}$, $z \in \mathbb{R}_+$,

$$
\left|D_{(t,z)}\lambda(s)\right| \leq \|\phi\|_\infty \tag{35}
$$

and by the Lipschitz continuity of ϕ

$$
\left|D_{(t,z)}\lambda(s)\right| \leq \alpha\left|\int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u\leq t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)(\overline{N}+\varepsilon_{(t,z)})(\mathrm{d}u \times \mathrm{d}v)\right.
$$

$$
+ \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u>t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)}+\varepsilon_{(t,z)})]}(v)(\overline{N}+\varepsilon_{(t,z)})(\mathrm{d}u \times \mathrm{d}v)
$$

$$
- \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u\leq t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)
$$

$$
\left. - \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u>t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right|
$$

$$
= \alpha\left|h(s-t)\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z)\right.
$$

$$
+ \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u>t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)}+\varepsilon_{(t,z)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)
$$

$$
\left. - \int_{(-\infty,s)\times\mathbb{R}_+} \mathbb{1}_{u>t}h(s-u)\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right|
$$

$$
\leq \alpha\left(\left|h(s-t)\right|\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z)\right.
$$

$$
\left. + \int_{(t,s)\times\mathbb{R}_+} \left|h(s-u)\right|\left|\mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)}+\varepsilon_{(t,z)})]}(v) - \mathbb{1}_{(0,\varphi(u,\overline{N}|_{(-\infty,u)})]}(v)\right|\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right)
$$

$$
\leq \alpha\left(\left|h(s-t)\right|\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) + \int_{(t,s)\times\mathbb{R}_+} \left|h(s-u)\right|\right.
$$

$$
\left. \times \mathbb{1}_{(\varphi(u,\overline{N}|_{(-\infty,u)})\wedge\varphi(u,\overline{N}|_{(-\infty,u)}+\varepsilon_{(t,z)}),\varphi(u,\overline{N}|_{(-\infty,u)})\vee\varphi(u,\overline{N}|_{(-\infty,u)}+\varepsilon_{(t,z)})]}(v)\overline{N}(\mathrm{d}u \times \mathrm{d}v)\right)
$$

$$
= \alpha\left(\left|h(s-t)\right|\mathbb{1}_{(0,\varphi(t,\overline{N}|_{(-\infty,t)})]}(z) + \int_{(t,s)} \left|h(s-u)\right|N_{(t,z)}(\mathrm{d}u)\right), \tag{36}
$$

where for $a, b \in \mathbb{R}$ we denote by $a \wedge b$ and $a \vee b$ the minimum and the maximum of a and b, respectively, and $N_{(t,z)}$ is the point process on (t, ∞) defined by

$$
N_{(t,z)}(\mathrm{d}u) := \overline{N}\left(\mathrm{d}u \times \left(\varphi(u, \overline{N}|_{(-\infty,u)}) \wedge \varphi(u, \overline{N}|_{(-\infty,u)} + \varepsilon_{(t,z)}), \varphi(u, \overline{N}|_{(-\infty,u)}) \vee \varphi(u, \overline{N}|_{(-\infty,u)} + \varepsilon_{(t,z)})\right]\right).
$$

The processes $\{\varphi(u, \overline{N}|_{(-\infty,u)} + \varepsilon_{(t,z)})\}_{u>t}$ and $\{\varphi(u, \overline{N}|_{(-\infty,u)})\}_{u>t}$ are $\{\mathcal{F}_u^N\}_{u>t}$-predictable and bounded, therefore by Lemma 2.1 the point process $N_{(t,z)}$ has $\{\mathcal{F}_u^N\}_{u>t}$-stochastic intensity $\{|D_{(t,z)}\lambda(u)|\}_{u>t}$. Taking the mean in (36), we deduce

$$\mathrm{E}\big[\big|D_{(t,z)}\lambda(s)\big|\big] \leq \alpha\left(\big|h(s-t)\big|P\big(\lambda(t) \geq z\big) + \int_t^s \big|h(s-u)\big|\mathrm{E}\big[\big|D_{(t,z)}\lambda(u)\big|\big]\,\mathrm{d}u\right).$$

Extending the definition of h for nonpositive times by setting $h(t) = 0$, $t \leq 0$, from the above inequality we have

$$q_{(t,z)}(s) \leq p_{(t,z)}(s) + r * q_{(t,z)}(s), \quad s, t \in \mathbb{R}, z \in \mathbb{R}_+,$$

where for ease of notation we set $q_{(t,z)}(s) := \mathrm{E}[|D_{(t,z)}\lambda(s)|]\mathbb{1}_{s>t}$, $p_{(t,z)}(s) := \alpha|h(s-t)|P(\lambda(t) \geq z)$, $r(s) := \alpha|h(s)|$ and $*$ denotes the convolution product between functions. Iterating this inequality, we deduce, for $n \geq 1$,

$$q_{(t,z)}(s) \leq \sum_{i=0}^{n-1} p_{(t,z)} * r^{*i}(s) + q_{(t,z)} * r^{*n}(s), \quad s, t \in \mathbb{R}, z \in \mathbb{R}_+,$$

where r^{*0} is by definition the Dirac delta function. By (35) and the stability condition $\alpha\mu < 1$ it follows

$$q_{(t,z)} * r^{*n}(s) = \int_{\mathbb{R}} r^{*n}(s-u)q_{(t,z)}(u)\,\mathrm{d}u \leq \|\phi\|_\infty \int_{\mathbb{R}} r^{*n}(s-u)\,\mathrm{d}u$$
$$\leq \|\phi\|_\infty(\alpha\mu)^n \to 0, \quad n \to \infty,$$

where the latter inequality follows by a standard property of convolutions, see e.g. Theorem IV.15 in [12]. Therefore,

$$q_{(t,z)}(s) \leq \sum_{i\geq 0} p_{(t,z)} * r^{*i}(s) = P\big(\lambda(t) \geq z\big) \sum_{i\geq 0} \alpha^{i+1} \int_{\mathbb{R}} \big|h(s-u-t)\big||h|^{*i}(u)\,\mathrm{d}u$$
$$= P\big(\lambda(t) \geq z\big) \sum_{i\geq 1} \alpha^i |h|^{*i}(s-t), \quad s, t \in \mathbb{R}, z \in \mathbb{R}_+. \tag{37}$$

Consequently, for any measurable f, we have

$$\int_{\mathbb{R}\times\mathbb{R}_+}\left(\int_t^{+\infty} |f(s)|\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\right)\mathrm{d}t\,\mathrm{d}z \leq \int_{\mathbb{R}\times\mathbb{R}_+}\left(\int_t^{+\infty} |f(s)|\mathrm{E}\big[\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\right)\mathrm{d}t\,\mathrm{d}z$$
$$\leq \|f\|_{L^1(\mathbb{R},\mathrm{d}x)}\frac{\lambda\alpha\mu}{1-\alpha\mu} \tag{38}$$

and, for any f and g integrable, we have

$$\int_{\mathbb{R}\times\mathbb{R}_+} |f(t)|\left(\int_t^{+\infty} |g(s)|\mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\right)\mathrm{d}t\,\mathrm{d}z$$
$$\leq \int_{\mathbb{R}\times\mathbb{R}_+} |f(t)|\left(\int_t^{+\infty} |g(s)|\mathrm{E}\big[\big|D_{(t,z)}\lambda(s)\big|\big]\,\mathrm{d}s\right)\mathrm{d}t\,\mathrm{d}z$$
$$\leq \|f\|_{L^2(\mathbb{R},\mathrm{d}x)}\|g\|_{L^2(\mathbb{R},\mathrm{d}x)}\frac{\lambda\alpha\mu}{1-\alpha\mu}. \tag{39}$$

Here, we only check this latter inequality, the previous one may be checked similarly. Assume that f and g integrable and square integrable (note that if f or/and g are not square integrable, then the inequality is trivially true). Defining

$\widetilde{f}(x) := f(-x)$, by (37), the stationarity of the nonlinear Hawkes process, the finiteness of its intensity λ, the Cauchy–Schwarz inequality and the properties of the convolution product (see again Theorem IV.15 in [12]), we have

$$
\int_{\mathbb{R}\times\mathbb{R}_+} |f(t)| \left(\int_t^{+\infty} |g(s)| \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|\big] \,\mathrm{d}s \right) \mathrm{d}t \,\mathrm{d}z
$$

$$
\leq \int_{\mathbb{R}\times\mathbb{R}_+} |f(t)| \left(\int_t^{+\infty} |g(s)| \mathrm{E}\big[\big| D_{(t,z)}\lambda(s)\big|\big] \,\mathrm{d}s \right) \mathrm{d}t \,\mathrm{d}z
$$

$$
\leq \lambda \sum_{i\geq 1} \alpha^i \int_{\mathbb{R}} |f(t)| \big((|\widetilde{h}|^{*i} * |g|)(t) \big) \,\mathrm{d}t
$$

$$
\leq \lambda \|f\|_{L^2(\mathbb{R},\mathrm{d}x)} \sum_{i\geq 1} \alpha^i \big\| |\widetilde{h}|^{*i} * |g| \big\|_{L^2(\mathbb{R},\mathrm{d}x)}
$$

$$
\leq \lambda \|f\|_{L^2(\mathbb{R},\mathrm{d}x)} \|g\|_{L^2(\mathbb{R},\mathrm{d}x)} \sum_{i\geq 1} \alpha^i \big\| |\widetilde{h}|^{*i} \big\|_{L^1(\mathbb{R},\mathrm{d}x)}
$$

$$
\leq \lambda \|f\|_{L^2(\mathbb{R},\mathrm{d}x)} \|g\|_{L^2(\mathbb{R},\mathrm{d}x)} \sum_{i\geq 1} \alpha^i \mu^i = \|f\|_{L^2(\mathbb{R},\mathrm{d}x)} \|g\|_{L^2(\mathbb{R},\mathrm{d}x)} \frac{\lambda\alpha\mu}{1-\alpha\mu}.
$$

Finally, we also have

$$
\int_{\mathbb{R}\times\mathbb{R}_+} u(t) \sqrt{\int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|\big] \,\mathrm{d}s \int_t^{\infty} u(s) \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|^2\big] \,\mathrm{d}s} \,\mathrm{d}t \,\mathrm{d}z
$$

$$
\leq \|\phi\|_\infty \int_{\mathbb{R}\times\mathbb{R}_+} u(t) \int_t^{\infty} u(s)^2 \mathrm{E}\big[\mathbb{1}_{(0,\lambda(t)]}(z)\big| D_{(t,z)}\lambda(s)\big|\big] \,\mathrm{d}s \,\mathrm{d}t \,\mathrm{d}z \tag{40}
$$

$$
\leq \min\left\{ \lambda\|\phi\|_\infty^2 \|u\|_{L^1(\mathbb{R},\mathrm{d}x)} \|u\|_{L^2(\mathbb{R},\mathrm{d}x)}^2, \frac{\lambda\alpha\mu\|\phi\|_\infty}{1-\alpha\mu} \|u\|_{L^2(\mathbb{R},\mathrm{d}x)} \|u\|_{L^4(\mathbb{R},\mathrm{d}x)}^2 \right\}, \tag{41}
$$

where (40) follows by (35) and $u \leq u^2$ (recall that u is $\mathbb{N}$-valued), and (41) is a consequence of (39) and again (35).

Step 2: *checking conditions* (7), (8) *and* (12). Condition (7) is an immediate consequence of the square integrability of u, the stationarity of the nonlinear Hawkes process and the finiteness of its intensity λ. As far as conditions (8) and (12), note that, due to (35) and the inequality $u \leq u^2$, it suffices to check (8). For this, by (38) with $f = u^2$, we have

$$
\int_{\mathbb{R}\times\mathbb{R}_+} \left(\int_t^{\infty} u(s)^2 \mathrm{E}\big[\big| D_{(t,z)}\lambda(s)\big|\big] \,\mathrm{d}s \right) \mathrm{d}t \,\mathrm{d}z \leq \|u\|_{L^2(\mathbb{R},\mathrm{d}x)}^2 \frac{\lambda\alpha\mu}{1-\alpha\mu} < \infty.
$$

So the assumptions of Corollary 3.2 are satisfied.

Step 3: *conclusion of the proof.* By the Lipschitz property of ϕ, for any $t \in \mathbb{R}$, we have

$$
\big|\lambda(t) - \lambda\big| \leq \min\big\{ \|\phi\|_\infty, \big|\lambda(t) - \phi(0) + \phi(0) - \lambda\big| \big\}
$$

$$
\leq \min\big\{ \|\phi\|_\infty, \big|\lambda(t) - \phi(0)\big| + \big|\phi(0) - \lambda\big| \big\}
$$

$$
\leq \min\left\{ \|\phi\|_\infty, \alpha \int_{(-\infty,t)} \big|h(t-s)\big| N(\mathrm{d}s) + \big|\phi(0) - \lambda\big| \right\},
$$

and so by the stationarity of N

$$
\mathrm{E}\big[\big|\lambda(t) - \lambda\big|\big] \leq \min\big\{ \|\phi\|_\infty, \lambda\alpha\mu + \big|\phi(0) - \lambda\big| \big\}.
$$

Therefore

$$
\min\left(1, \sqrt{\frac{2}{\lambda \mathrm{e}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}}\right) \mathrm{E}\big[\big|\Lambda(u) - \lambda\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}\big|\big] \leq \min\left(1, \sqrt{\frac{2}{\lambda \mathrm{e}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}}\right) \int_{\mathbb{R}} u(t)\mathrm{E}\big[\big|\lambda(t) - \lambda\big|\big]\,\mathrm{d}t
$$

$$
\leq \min\left(1, \sqrt{\frac{2}{\lambda \mathrm{e}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}}}\right)\mathfrak{D}_2. \tag{42}
$$

By the stationarity of N, (35), (39) and (41) we have

$$
\mathfrak{G}'/\lambda \leq \mathfrak{D}_1. \tag{43}
$$

The claim easily follows by the bound in Corollary 3.2, with $\theta := \lambda\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}$, and the inequalities (42) and (43). $\square$

Proof of Theorem 4.3. We only sketch the proof since it is quite similar to the proof of Theorem 4.1. Conditions (7), (8) and (12) and the inequality (43) may be checked exactly as in the proof of Theorem 4.1. The claim follows by first applying the bound in Corollary 3.2 with $\theta := \widehat{\lambda}\|u\|_{L^1(\mathbb{R},\mathrm{d}x)}$, and then using the inequalities (43), $\lambda \leq \min\{\phi(0)/(1-\alpha\mu), \|\phi\|_\infty\}$ and

$$
\mathrm{E}\big[\big|\lambda(t) - \widehat{\lambda}\big|\big] \leq \min\big\{\|\phi\|_\infty, \lambda\alpha\mu + \big|\phi(0) - \widehat{\lambda}\big|\big\}
$$
$$
\leq \min\big\{\|\phi\|_\infty, \min\big\{\phi(0)/(1-\alpha\mu), \|\phi\|_\infty\big\}\alpha\mu + \big|\phi(0) - \widehat{\lambda}\big|\big\}. \qquad \square
$$

4.2. *A quantitative Poisson limit theorem, construction of confidence intervals and asymptotic estimates of the moments*

The following quantitative Poisson limit theorem holds.

Theorem 4.4. *For $n \geq 1$, assume that $\phi_n : \mathbb{R} \to [0,\infty)$, $\phi_n(0) > 0$, is bounded and Lipschitz continuous with Lipschitz constant α_n and that $h_n : [0,\infty) \to \mathbb{R}$ is measurable and such that $\alpha_n\mu_n < 1$, where $\mu_n := \|h_n\|_{L^1(\mathbb{R}_+,\mathrm{d}x)}$. Let N_n be the stationary nonlinear Hawkes process with parameters (ϕ_n, h_n) and finite intensity $\lambda_n \in (0,\infty)$. If $u_n : \mathbb{R} \to \mathbb{N}$, $\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} > 0$ and $u_n \in L^2(\mathbb{R}, \mathrm{d}x)$, then*

$$
d_{\mathrm{TV}}\big(N_n(u_n), \mathrm{Po}(\gamma)\big) \leq \mathfrak{U}_1^{(n)} \leq \mathfrak{U}_2^{(n)}, \quad n \geq 1. \tag{44}
$$

Here $\gamma > 0$ is a positive constant,

$$
\mathfrak{U}_1^{(n)} := \frac{1 - \mathrm{e}^{-\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}}}{\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}} \min\big\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\big\}\mathfrak{D}_1^{(n)}
$$

$$
+ \min\left(1, \sqrt{\frac{2}{\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}\mathrm{e}}}\right)\mathfrak{D}_3^{(n)} + \big|\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} - \gamma\big|, \tag{45}
$$

$$
\widehat{\lambda}_n \in \big[\phi_n(0)/(1+\alpha_n\mu_n), \min\big\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\big\}\big] \tag{46}
$$

is an estimate of λ_n, $\mathfrak{D}_i^{(n)}$, $i = 1, 3$, is defined as $\mathfrak{D}_i$ in the statements of Theorems 4.1 and 4.3, with u_n, ϕ_n, α_n, μ_n and $\widehat{\lambda}_n$ in place of u, ϕ, α, μ and $\widehat{\lambda}$, respectively, and

$$
\mathfrak{U}_2^{(n)} := \frac{(1+\alpha_n\mu_n)\min\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\}(1 - \mathrm{e}^{-\min\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\}\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}})}{\phi_n(0)\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}}\mathfrak{D}_1^{(n)}
$$

$$
+ \|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} \min\left(1, \sqrt{\frac{2(1+\alpha_n\mu_n)}{\phi_n(0)\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)}\mathrm{e}}}\right)
$$

$$\times \min\big\{\|\phi_n\|_\infty, \min\big\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\big\}\alpha_n\mu_n$$
$$+ \max\big\{\big(\phi_n(0)\alpha_n\mu_n\big)/(1+\alpha_n\mu_n), \big|\phi_n(0)-\min\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\}\big|\big\}\big\}$$
$$+ \max\big\{\big|\big(\phi_n(0)/(1+\alpha_n\mu_n)\big)\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} - \gamma\big|,$$
$$\big|\min\big\{\phi_n(0)/(1-\alpha_n\mu_n), \|\phi_n\|_\infty\big\}\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} - \gamma\big|\big\}.$$

If moreover

$$\lim_{n\to\infty}\alpha_n\mu_n\max\big\{1, \|\phi_n\|_\infty\|u_n\|^2_{L^2(\mathbb{R},\mathrm{d}x)}, \|u_n\|_{L^2(\mathbb{R},\mathrm{d}x)}\|u_n\|^2_{L^4(\mathbb{R},\mathrm{d}x)}\max\{1, \|\phi_n\|_\infty\}\big\} = 0, \tag{47}$$

$$\lim_{n\to\infty}\phi_n(0) = \gamma_1 > 0, \tag{48}$$

$$\lim_{n\to\infty}\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} = \lim_{n\to\infty}\|u_n\|^3_{L^3(\mathbb{R},\mathrm{d}x)} = \gamma_2 > 0 \tag{49}$$

and

$$\gamma = \gamma_1\gamma_2, \tag{50}$$

then $\mathfrak{U}_2^{(n)} \to 0$. *In particular,* $N_n(u_n) \to \mathrm{Po}(\gamma_1\gamma_2)$ *in distribution, as* $n \to \infty$.

Proof. By e.g. Corollary 3.1 in [1] we have

$$d_{\mathrm{TV}}\big(\mathrm{Po}(b), \mathrm{Po}(b')\big) \le |b - b'|, \quad b, b' > 0. \tag{51}$$

Combining this inequality with the triangular inequality we deduce

$$d_{\mathrm{TV}}\big(N_n(u_n), \mathrm{Po}(\gamma)\big) \le d_{\mathrm{TV}}\big(N_n(u_n), \mathrm{Po}(\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)})\big) + \big|\widehat{\lambda}_n\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} - \gamma\big|.$$

The first inequality in (44) then follows by Theorem 4.3. The second inequality in (44) follows by bounding the term $\mathfrak{U}_1^{(n)}$ using (46). Finally, the convergence to zero of $\mathfrak{U}_2^{(n)}$ is a simple consequence of the assumptions (47), (48), (49) and (50). □

Theorem 4.4 may be used to construct confidence intervals for some classes of $\mathbb{N}$-valued first order stochastic integrals of nonlinear Hawkes processes.

Corollary 4.5. *For $n \ge 1$, let N_n be the stationary nonlinear Hawkes process considered in Theorem 4.4 with parameters (ϕ_n, h_n) such that $\phi_n(0) > 0$, $\alpha_n\mu_n < 1$ and let $u_n : \mathbb{R} \to \mathbb{N}$ be such that $\|u_n\|_{L^1(\mathbb{R},\mathrm{d}x)} > 0$ and $u_n \in L^2(\mathbb{R}, \mathrm{d}x)$. Moreover, let $\beta \in (0, 1)$ be arbitrarily fixed and let $k_\beta^{(1)}, k_\beta^{(2)} \in \mathbb{N}$ be such that $k_\beta^{(1)} \le k_\beta^{(2)}$ and*

$$\mathrm{e}^{-\gamma}\sum_{k=k_\beta^{(1)}}^{k_\beta^{(2)}}\frac{\gamma^k}{k!} > 1 - \frac{\beta}{2}.$$

Then: 1

(i) *If $n \in \mathcal{I}_\beta := \{k : \mathfrak{U}_2^{(k)} < \beta/4\}$, then*

$$P\big(k_\beta^{(1)} \le N_n(u_n) \le k_\beta^{(2)}\big) > 1 - \beta,$$

   i.e. $[k_\beta^{(1)}, k_\beta^{(2)}]$ is a confidence interval of $N_n(u_n)$ at level $1 - \beta$.
(ii) *If in addition (47), (48), (49) and (50) hold, then there exists $n_\beta \ge 1$ such that for any $n \ge n_\beta$ we have $n \in \mathcal{I}_\beta$ (and so, in particular, $\mathcal{I}_\beta \ne \varnothing$).*

Remark 4.6. *Let $\beta \in (0, 1)$ and $n \geq 1$ be arbitrarily fixed. Since the term $\mathfrak{U}_2^{(n)}$ is explicitly known in terms of the data, i.e. the parameters (ϕ_n, h_n) of the nonlinear Hawkes process and the function u_n, one can always determine if n belongs to $\mathcal{I}_\beta$ or not.*

The exact computation of the moments of stochastic integrals with respect to point processes is in general a difficult task. To the best of our knowledge, closed (but complicated) formulas exist for the moments of stochastic integrals with respect to point processes with Papangelou conditional intensity, see [33] for the Poisson case and [18] for the general case. Theorem 4.4 may be used also to provide asymptotic estimates of the moments (and, in particular, of the variance) of some classes of $\mathbb{N}$-valued first order stochastic integrals of nonlinear Hawkes processes.

Recall that, for $r \in \mathbb{N}$ and $\lambda > 0$, $\mathrm{E}[\mathrm{Po}(\lambda)^r] = T_r(\lambda)$ where $T_r(\lambda)$ is the Touchard polynomial of order r, defined by $T_0(\lambda) := 1$ and the recurrence relation

$$T_{r+1}(\lambda) = \lambda \sum_{k=0}^{r} \binom{r}{k} T_k(\lambda), \quad r \in \mathbb{N}.$$

The Touchard polynomials are also called exponential polynomials, see e.g. [13]. The following corollary holds.

Corollary 4.7. *Let the notation and the assumptions of Corollary 4.5(ii) prevail. Then, for any $r \in \mathbb{N}$,*

$$\mathrm{E}[N_n(u_n)^r] \to T_r(\gamma_1 \gamma_2), \quad as \ n \to \infty.$$

Proof of Corollary 4.5. *Proof of* (i). For $n \in \mathcal{I}_\beta$, by Theorem 4.4 we have

$$d_{\mathrm{TV}}(N_n(u_n), \mathrm{Po}(\gamma)) \leq \mathfrak{U}_2^{(n)} < \beta/4$$

and so

$$\left| P(N_n(u_n) \in A) - P(\mathrm{Po}(\gamma) \in A) \right| < \beta/4, \quad \text{for any } A \subseteq \mathbb{N}.$$

Therefore

$$P(k_\beta^{(1)} \leq N_n(u_n) \leq k_\beta^{(2)}) = P(N_n(u_n) \leq k_\beta^{(2)}) - P(N_n(u_n) < k_\beta^{(1)})$$

$$= \left(P(N_n(u_n) \leq k_\beta^{(2)}) - P(\mathrm{Po}(\gamma) \leq k_\beta^{(2)}) \right) + \left(P(\mathrm{Po}(\gamma) < k_\beta^{(1)}) - P(N_n(u_n) < k_\beta^{(1)}) \right)$$

$$+ \left(P(\mathrm{Po}(\gamma) \leq k_\beta^{(2)}) - P(\mathrm{Po}(\gamma) < k_\beta^{(1)}) \right)$$

$$> -\beta/4 - \beta/4 + 1 - \beta/2 = 1 - \beta.$$

Proof of (ii). The claim follows by Theorem 4.4 which guarantees that $\mathfrak{U}_2^{(n)} \to \infty$, as $n \to \infty$. $\qquad\square$

Proof of Corollary 4.7. By Theorem 4.4 we have that $N_n(u_n)$ converges to $\mathrm{Po}(\gamma_1 \gamma_2)$ in distribution. Moreover, the random variables $N_n(u_n)$, $n \geq 1$, and $\mathrm{Po}(\gamma_1 \gamma_2)$ are non-negative and

$$\mathrm{E}[N_n(u_n)] = \lambda_n \int_{\mathbb{R}} u_n(t) \, \mathrm{d}t \to \gamma_1 \gamma_2 = \mathrm{E}[\mathrm{Po}(\gamma_1 \gamma_2)], \quad \text{as } n \to \infty.$$

So, by e.g. (the second part of) Theorem 5.4, p. 32 in [6], we have that the random variables $N_n(u_n)$, $n \geq 1$, are uniformly integrable and the proof is completed (see again Theorem 5.4, p. 32, and the comments on p. 33 of [6]). $\square$

Example 4.8. *For $n \geq 1$, assume that $\phi_n : \mathbb{R} \to \mathbb{R}_+$, $\phi_n(0) > 0$, is bounded and Lipschitz continuous with Lipschitz constant α_n and that $h_n : [0, \infty) \to \mathbb{R}$ is measurable and such that $\alpha_n \mu_n < 1$, where $\mu_n := \|h_n\|_{L^1(\mathbb{R}_+, \mathrm{d}x)}$. Let N_n be the stationary nonlinear Hawkes process with parameters (ϕ_n, h_n) and finite intensity $\lambda_n \in (0, \infty)$. For a bounded Borel set $B \subset \mathbb{R}$ with positive Lebesgue measure, we define $u_n(t) := \mathbb{1}_B(t)$, $n \geq 1$, $t \in \mathbb{R}$. Then by Theorem 4.4 we have*

$d_{\mathrm{TV}}(N_n(\mathbb{1}_B), \mathrm{Po}(\gamma)) \leq \mathfrak{U}_2^{(n)}$ *where $\gamma > 0$ is a positive constant and $\mathfrak{U}_2^{(n)}$ is defined as in the statement of Theorem* 4.4 *with*

$$\|u_n\|_{L^1(\mathbb{R}, \mathrm{d}x)} = \ell(B), \qquad \|u_n\|_{L^2(\mathbb{R}, \mathrm{d}x)} = \ell(B)^{1/2}, \qquad \|u_n\|_{L^3(\mathbb{R}, \mathrm{d}x)} = \ell(B)^{1/3}, \qquad \|u_n\|_{L^4(\mathbb{R}, \mathrm{d}x)} = \ell(B)^{1/4},$$

where $\ell(B)$ denotes the Lebesgue measure of B.

Assume moreover (48) *and*

$$\lim_{n \to \infty} \alpha_n \mu_n \max\{1, \|\phi_n\|_\infty\} = 0, \tag{52}$$

and take $\gamma := \gamma_1 \ell(B)$. Then: (i) *by Theorem* 4.4 *we have that $\mathfrak{U}_2^{(n)} \to 0$, and so $N_n(\mathbb{1}_B) \to \mathrm{Po}(\gamma_1 \ell(B))$ in distribution, as $n \to \infty$;* (ii) *by Corollary* 4.5 *it follows that, for any arbitrarily fixed $\beta \in (0, 1)$, the set $[k_\beta^{(1)}, k_\beta^{(2)}]$ with $k_\beta^{(1)}, k_\beta^{(2)} \in \mathbb{N}$ chosen in such a way that $k_\beta^{(1)} \leq k_\beta^{(2)}$ and*

$$\mathrm{e}^{-\gamma_1 \ell(B)} \sum_{k=k_\beta^{(1)}}^{k_\beta^{(2)}} \frac{(\gamma_1 \ell(B))^k}{k!} > 1 - \beta/2,$$

is a confidence interval at level $1 - \beta$ of $N_n(\mathbb{1}_B)$, for all n such that $\mathfrak{U}_2^{(n)} < \beta/4$; (iii) *by Corollary* 4.7 *we have that, for any $r \geq 1$, $\mathrm{E}[N_n(\mathbb{1}_B)^r] \to T_r(\gamma_1 \ell(B))$, as $n \to \infty$.*

5. An illustrative example: Poisson bounds for the input process to the $M/D/K/0$ queue

The $M/D/K/0$ queue or Erlang loss system is described as follows (see e.g. [2]). The system is formed by K servers and customer n, arrived at time S_n, is accepted in the system if and only if she/he finds a not busy server, i.e. the number of busy servers is less than or equal to $K - 1$. Customer k brings a service request $z \in \mathbb{R}_+$, which means that the server will need z units of time for processing the request of customer k. A server attends one customer at time and do not remain idle as long as there is at least one customer in the system. Once the service is started it can not be interrupted before completion and, at the end of the service, the server chooses his next customer at the head of the line. Assuming $\{S_k\}_{k \in \mathbb{Z}}$ to be a stationary Poisson process on the line with intensity $\alpha > 0$, then the input process $N = \{T_k\}_{k \in \mathbb{Z}}$ to the Erlang loss system described above is a nonlinear Hawkes process with parameters (ϕ, h) given by

$$\phi(x) := \alpha \mathbb{1}_{[0,K)}(x), \qquad h(t) := \mathbb{1}_{(0,z]}(t).$$

Since we are indeed interested to the values of ϕ at the points

$$x = \int_{(-\infty, t)} h(t - s) N(\mathrm{d}s) = \sum_{k \in \mathbb{Z}} \mathbb{1}_{(0,z]}(t - T_k) \mathbb{1}_{(-\infty, t)}(T_k),$$

which are integer-valued random variables, we can replace the indicator function $\mathbb{1}_{[0,K)}(x)$ with a Lipschitz continuous function f with Lipschitz constant 1 which is equal to zero for $x \geq K$ and equal to 1 for $x \in [0, K-1]$ (we can choose for instance $f(x) := \min\{\max\{K - x, 0\}, 1\}$).

For $n \geq 1$, let $N_n = \{T_k^{(n)}\}_{k \in \mathbb{Z}}$ be the input process to the Erlang loss system with exogenous arrivals according to a Poisson process with intensity $\alpha_n > 0$, $K_n \in \mathbb{N} \setminus \{0\}$ servers and service request $z_n > 0$, and set $u_n(t) := \mathbb{1}_{(0,\tau_n]}(t)$, $\tau_n > 0$. If the service request z_n is such that $z_n < \alpha_n^{-1}$, then by Theorem 4.4 we have

$$d_{\mathrm{TV}}(N_n(\mathbb{1}_{(0,\tau_n]}), \mathrm{Po}(\gamma)) \leq \mathfrak{U}_2^{(n)},$$

where $\gamma > 0$ is a positive constant and

$$
\mathfrak{U}_2^{(n)} := \left(1 - \mathrm{e}^{-\alpha_n \tau_n}\right)(1 + \alpha_n z_n)\left(3\alpha_n + \alpha_n^2 \tau_n + 2\alpha_n^2 \tau_n^{1/2}\right) \min\left\{\tau_n, \frac{z_n}{1 - \alpha_n z_n}\right\}
$$

$$
+ \alpha_n \tau_n \min\left(1, \sqrt{\frac{2(1 + \alpha_n z_n)}{\alpha_n \tau_n \mathrm{e}}}\right) \min\left\{1, \alpha_n z_n\left(1 + (1 + \alpha_n z_n)^{-1}\right)\right\}
$$

$$
+ \max\left\{\left|\left(\alpha_n/(1 + \alpha_n z_n)\right)\tau_n - \gamma\right|, |\alpha_n \tau_n - \gamma|\right\}.
$$

Assume moreover that $\alpha_n \to \gamma_1 > 0$, $\tau_n \to \gamma_2 > 0$, $z_n \to 0$, as $n \to \infty$, and take $\gamma := \gamma_1 \gamma_2$. Then: (i) by Theorem 4.4 we have that $\mathfrak{U}_2^{(n)} \to 0$, and so $N_n(\mathbb{1}_{(0,\tau_n]}) \to \mathrm{Po}(\gamma_1 \gamma_2)$ in distribution, as $n \to \infty$; (ii) by Corollary 4.5 it follows that, for any arbitrarily fixed $\beta \in (0, 1)$, the set $[k_\beta^{(1)}, k_\beta^{(2)}]$ with $k_\beta^{(1)}, k_\beta^{(2)} \in \mathbb{N}$ chosen in such a way that $k_\beta^{(1)} \le k_\beta^{(2)}$ and

$$
\mathrm{e}^{-\gamma_1 \gamma_2} \sum_{k=k_\beta^{(1)}}^{k_\beta^{(2)}} \frac{(\gamma_1 \gamma_2)^k}{k!} > 1 - \beta/2,
$$

is a confidence interval at level $1 - \beta$ of $N_n(\mathbb{1}_{(0,\tau_n]})$, for all n such that $\mathfrak{U}_2^{(n)} < \beta/4$; (iii) by Corollary 4.7 we have that, for any $r \ge 1$, $\mathrm{E}[N_n(\mathbb{1}_{(0,\tau_n]})^r] \to T_r(\gamma_1 \gamma_2)$, as $n \to \infty$.

Acknowledgements

References

[1] J. A. Adell and A. Lekuona. Sharp estimates in signed Poisson approximation of Poisson mixtures. *Bernoulli* **11** (2005) 47–65. MR2121455

[2] S. Asmussen. *Applied Probability and Queues*. Springer, New York, 2003. MR1978607

[3] A. D. Barbour, L. Holst and S. Jason. *Poisson Approximation*. Oxford Univ. Press, Oxford, 1992. MR1163825

[4] A. D. Barbour and T. C. Brown. The Stein–Chen method, point processes and compensators. *Ann. Probab.* **20** (1992) 1504–1527. MR1175275

[5] A. D. Barbour and T. C. Brown. Stein's method and point process approximation. *Stochastic Process. Appl.* **43** (1992) 9–31. MR1190904

[6] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968. MR0233396

[7] C. Bordenave and G. L. Torrisi. Large deviations of Poisson cluster processes. *Stoch. Models* **23** (2007) 593–625. MR2362700

[8] P. Brémaud. *Point Processes and Queues*. Springer, New York, 1981. MR0636252

[9] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** (1996) 1963–1988. MR1411506

[10] P. Brémaud and L. Massoulié. Power spectra of general shot noises and Hawkes point processes with a random excitation. *Adv. in Appl. Probab.* **34** (2002) 205–222. MR1895338

[11] P. Brémaud, G. Nappo and G. L. Torrisi. Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Probab.* **39** (2002) 123–136. MR1895148

[12] H. Brezis. *Analyse Fonctionnelle*. Dunod, Paris, 2005.

[13] C. A. Charalambides. *Enumerative Combinatorics*. Chapman & Hall, Boca Raton, 2002. MR1937238

[14] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Appl. Probab.* **3** (1975) 534–545. MR0428387

[15] L. H. Y. Chen, L. Goldstein and Q.-M. Shao. *Normal Approximation by Stein's Method*. Springer, Heidelberg, 2011. MR2732624

[16] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. I*. Springer, New York, 2003. MR1950431

[17] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. II*. Springer, New York, 2008. MR2371524

[18] L. Decreusefond and I. Flint. Moment formulae for general point processes. *J. Funct. Anal.* **267** (2015) 452–476. MR3210036

[19] S. Delattre, N. Fournier and M. Hoffmann. Hawkes processes on large networks. *Ann. Appl. Probab.* To appear, 2016.

[20] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** (1971) 83–90. MR0278410

[21] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting point process. *J. Appl. Probab.* **11** (1974) 493–503. MR0378093

[22] J. Kesrtan. Teilprozesse Poissonscher prozesse In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes* 377–403. Publ. House Czech. Acad. Sci, Prague, 1964. MR0166826

[23] G. Last and M. D. Penrose. Martingale representation for Poisson processes with applications to minimal variance hedging. *Stochastic Process. Appl.* **121** (2011) 1588–1606. MR2802467

[24] G. Last, G. Peccati and M. Schulte. Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields*. To appear, 2016. DOI: 10.1007/s00440-015-0643-7.

[25] P. A. W. Lewis and G. S. Shedler. Simulation of non-homogeneous Poisson processes with log linear rate function. *Biometrika* **63** (1976) 501–506. MR0464448

[26] L. Massoulié. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Process. Appl.* **75** (1998) 1–30. MR1629010

[27] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** (2009) 75–118. MR2520122

[28] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge Univ. Press, Cambridge, 2012. MR2962301

[29] D. Nualart and J. Vives. Anticipative calculus for the Poisson process based on the Fock space. In *Séminaire de Probabilités XXIV, 1988/1989* 154–165. *Lecture Notes in Math.* **1426**. Springer, Berlin, 1990. MR1071538

[30] Y. Ogata. On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* **27** (2001) 23–31.

[31] G. Peccati, J. L. Solé, M. S. Taqqu and F. Utzet. Stein's method and normal approximation of Poisson functionals. *Ann. Probab.* **38** (2010) 443–478. MR2642882

[32] G. Peccati The Chen–Stein method for Poisson functionals. Preprint, 2011. Available at https://sites.google.com/site/giovannipeccati/Home/Publications-by-G-Peccati.

[33] N. Privault. Invariance of Poisson measures under random transformations. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (2012) 947–972. MR3052400

[34] N. Privault and G. L. Torrisi. Probability approximation by Clark–Ocone covariance representation. *Electron. J. Probab.* **18** (2013) 1–25. MR3126574

[35] M. Reitzner and M. Schulte. Central limit theorems for U-statistics of Poisson point processes. *Ann. Probab.* **41** (2013) 3879–3909. MR3161465

[36] G. L. Torrisi. A class of interacting marked point processes: Rate of convergence to equilibrium. *J. Appl. Probab.* **39** (2002) 137–160. MR1895149

[37] G. L. Torrisi. Gaussian approximation of nonlinear Hawkes processes. *Ann. Appl. Probab.* To appear, 2016.

[38] L. Wu. A new modified logarithmic Sobolev inequality for Poisson point processes and several applications. *Probab. Theory Related Fields* **118** (2000) 427–438. MR1800540

[39] L. Zhu. Central limit theorem for nonlinear Hawkes processes. *J. Appl. Probab.* **50** (2013) 760–771. MR3102513

[40] L. Zhu. Process-level large deviations for nonlinear Hawkes processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (3) (2014) 845–871. MR3224291

[41] L. Zhu. Large deviations for Markovian nonlinear Hawkes processes. *Ann. Appl. Probab.* **25** (2015) 548–581. MR3313748