GAUSSIAN APPROXIMATION AND MODERATE DEVIATIONS OF POISSON SHOT NOISES WITH APPLICATION TO COMPOUND GENERALIZED HAWKES PROCESSES

MAHMOUD KHABOU,* *Imperial College London*

GIOVANNI LUCA TORRISI,** *Consiglio Nazionale delle Ricerche*

Abstract

In this article, we give explicit bounds on the Wasserstein and Kolmogorov distances between random variables lying in the first chaos of the Poisson space and the standard normal distribution, using the results of Last *et al.* (*Prob. Theory Relat. Fields* **165**, 2016). Relying on the theory developed by Saulis and Statulevicius in *Limit Theorems for Large Deviations* (Kluwer, 1991) and on a fine control of the cumulants of the first chaoses, we also derive moderate deviation principles, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction terms for the same variables. The aforementioned results are then applied to Poisson shot noise processes and, in particular, to the generalized compound Hawkes point processes (a class of stochastic models, introduced in this paper, which generalizes classical Hawkes processes). This extends the recent results of Hillairet *et al.* (*ALEA* **19**, 2022) and Khabou *et al.* (*J. Theoret. Prob.* **37**, 2024) regarding the normal approximation and those of Zhu (*Statist. Prob. Lett.* **83**, 2013) for moderate deviations.

Keywords: Generalized Hawkes processes; moderate deviations; Wasserstein distance; Kolmogorov distance; concentration inequalities

2010 Mathematics Subject Classification: Primary 60G55; 60F05; 60F10
                                          Secondary 60G57; 60H05

1. Introduction

Shot noise models built on one-dimensional Poisson processes are very popular in applied probability. Because of their versatility and mathematical tractability, they find application in many fields, such as insurance, finance, queueing theory, and neuroscience (see e.g. [7, 10, 17, 18, 24, 27, 28, 31, 33, 34, 36, 43, 46]). Shot noise models whose underlying point processes are spatial Poisson processes (hereafter called *spatial* Poisson shot noise models; see Section 2 for a formal definition) are a bit less popular, but they play an important role in wireless communication, where they are exploited as models of the inference in ad hoc networks (see e.g. [1–3, 16, 38, 42]). Furthermore, as explained in detail in the next section, spatial Poisson shot noise models encompass spatial Poisson cluster point processes, which are widely used in many research areas, such as spatial statistics (see e.g. [35]). Since spatial Poisson shot noise

1

models are stochastic integrals with respect to a Poisson random measure, Gaussian approximation bounds for the Wasserstein and Kolmogorov distances between such random variables (properly standardized) and the standard normal law can easily be obtained by applying the general theory developed in the seminal paper [30]. One of the main achievements of the present article are explicit bounds for the Wasserstein and Kolmogorov distances between a properly standardized compound sum, which extends Poisson cluster and Hawkes point processes, and the standard normal law (see Corollaries 1, 3 and 5). These results improve upon and go beyond the findings in [23, 26], exploiting a considerably simpler approach (see also the discussion in Section 7.3.1). Using a well-known link between cumulants and large deviation theory (see [40]), we also provide sufficient conditions which guarantee moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term (see Definition 1 for details) for sequences of random variables which belong to the first chaos on the Poisson space (see Theorem 2). We then transfer such results to sequences of spatial Poisson shot noise models. As a main application, we provide moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for sequences of compound sums, which extend Poisson cluster and Hawkes point processes (see Corollaries 2, 4, and 6). Remarkably, the result on moderate deviations recovers, under an alternative condition on the fertility function, the moderate deviations for the number of points of a classical Hawkes process on the time interval $(0, t]$ proved in [47] (see Section 7.3.2).

The paper is structured as follows. In Section 2 we introduce the Poisson shot noise models considered in the paper, and we show that compound Poisson cluster point processes and generalized compound Hawkes processes are indeed Poisson shot noise models. Furthermore, we recall a simple model of wireless communication, which accounts for interference effects described by a Poisson shot noise. In Section 3 we provide an informal description of our results. In Section 4 we give Gaussian approximation bounds for the Wasserstein and Kolmogorov distances between a random variable belonging to the first chaos of the Poisson space and the standard normal law, and we give moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for sequences of random variables belonging to the first chaos on the Poisson space. Applications of the results of Section 4 to spatial Poisson shot noise models and compound Poisson cluster point processes are provided in Sections 5 and 6, respectively. The general results on Gaussian approximation and moderate deviations are applied to generalized compound Hawkes processes in Sections 7 and 8, respectively.

2. Poisson shot noise random variables

Throughout this article, if x is a point in some set E and $C \subset E$, then $C - x$ denotes the set $\{y - x, y \in C\}$. A Poisson shot noise random variable is a real-valued random variable of the form

$$S(C) := \sum_{n \geq 1} H(C - X_n, Z_n), \quad C \in \mathcal{B}(\mathbb{R}^d). \tag{1}$$

Here $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-field on $\mathbb{R}^d$, $d \geq 1$, $\mathcal{P} \equiv \{(X_n, Z_n)\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d \times \mathbf{Z}$ with mean measure $\lambda(x)\mathrm{d}x\mathbb{Q}(\mathrm{d}z)$, $(\mathbf{Z}, \mathcal{Z})$ is a measurable space, $\lambda : \mathbb{R}^d \to [0, \infty)$ is a locally integrable intensity function, $\mathbb{Q}$ is a probability measure on $\mathbf{Z}$, and $H : \mathcal{B}(\mathbb{R}^d) \times \mathbf{Z} \to \mathbb{R}$ is a mapping such that, for each fixed $C \in \mathcal{B}(\mathbb{R}^d)$, the function

$$(x, z) \in \mathbb{R}^d \times \mathbf{Z} \mapsto H(C - x, z) \in \mathbb{R}$$

is measurable. Poisson shot noise random variables encompass a variety of important stochastic models.

2.1. Compound Poisson cluster point processes

Let $\{X_n\}_{n\geq 1}$ be the points of a Poisson process on $\mathbb{R}^d$, $d \geq 1$, with a locally integrable intensity function $\lambda : \mathbb{R}^d \to [0, \infty)$, and let $\{Z_n(\cdot, \cdot)\}_{n\geq 1}$ be a sequence of independent and identically distributed simple point processes on $\mathbb{R}^d \times \mathbb{R}$, independent of $\{X_n\}_{n\geq 1}$. More concretely, for $(C_1, C_2) \in \mathcal{B}(\mathbb{R}^d) \times \mathcal{B}(\mathbb{R})$, $Z_n(C_1, C_2)$ counts the number of points of the *n*th point process that fall in C_1 and whose marks are in C_2. For each $n \geq 1$, we denote the points of $Z_n(\cdot, \cdot)$ by $\{(Y_{n,k}, M_{n,k})\}_{k\geq 0}$, and we assume that $Y_{n,0} := \mathbf{0}$ (which implies $Z_n(\{\mathbf{0}\}, \mathbb{R}) := 1$) and that the sequence $\{M_{n,k}\}_{k\geq 0}$ is independent of $\{Y_{n,k}\}_{k\geq 0}$. Furthermore, we suppose that the random variables $\{M_{n,k}\}_{n\geq 1, k\geq 0}$ are independent and identically distributed. Throughout the paper we denote by M the generic random variable $M_{n,k}$.

One naturally interprets the first component of each point of $Z_n(\cdot, \cdot)$ as a 'location', and the second component as a 'mark' which describes some characteristic of the location to which it is attached. Hereafter, for $n \geq 1$, we consider the point processes $\theta_{X_n} Z_n(\cdot, \cdot) \equiv \{(X_n + Y_{n,k}, M_{n,k})\}_{k\geq 0}$.

For arbitrarily fixed $n \geq 1$ and $C \in \mathcal{B}(\mathbb{R}^d)$, we define the random variable

$$\upsilon(Z_n)(C) := \sum_{k=0}^{Z_n(C,\mathbb{R})-1} M_{n,k}, \tag{2}$$

which aggregates the marks attached to the locations that fall in C. It turns out that the random variable, say $V(C)$, which aggregates all the marks attached to the points which fall in C of the Poisson cluster point process

$$N \equiv \bigcup_{n\geq 1} \{X_n + Y_{n,k}\}_{k\geq 0}$$

is a Poisson shot noise random variable. Indeed,

$$V(C) := \sum_{n\geq 1} \sum_{k=0}^{\theta_{X_n} Z_n(C,\mathbb{R})-1} M_{n,k} = \sum_{n\geq 1} \upsilon(\theta_{X_n} Z_n)(C) = \sum_{n\geq 1} \upsilon(Z_n)(C - X_n) \tag{3}$$

is a random variable of the form (1) with $H(C - x, z) := \upsilon(z)(C - x), x \in \mathbb{R}^d, z \in \mathbf{Z} := \mathbf{N}_{\mathbb{R}^d \times \mathbb{R}}$. Here $\mathbf{N}_{\mathbb{R}^d \times \mathbb{R}}$ denotes the space of σ-finite counting measures on $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}))$ equipped with the usual σ-field (see Section 4 for details), and

$$v(z)(C) := \sum_{k\geq 0} \mathbf{1}_C(y_k) m_k, \quad \text{for } z \equiv \{(y_k, m_k)\}_{k\geq 0}. \tag{4}$$

Note that if $M_{n,k} := 1$ for every $n \geq 1$ and $k \geq 0$, then the random variable

$$N(C) := V(C) = \sum_{n\geq 1} Z_n(C - X_n, \mathbb{R}) \tag{5}$$

equals the number of points of the Poisson cluster point process N which fall in $C \in \mathcal{B}(\mathbb{R}^d)$.

2.2. Generalized Hawkes processes and generalized compound Hawkes processes

Let $N \equiv \{N(C)\}_{C \in \mathcal{B}(\mathbb{R}^d)}$ be the Poisson cluster point process defined by (5). We shall refer to N as a *generalized Hawkes process* if the random variable $Z := Z_1(\mathbb{R}^d, \mathbb{R})$ is distributed as the total progeny of a sub-critical Galton–Watson process with one ancestor. It is worthwhile to note the following:

(i) *Classical Hawkes processes* on $(0, \infty)$ (respectively, on $\mathbb{R}$) with parameters (λ, g), introduced in the seminal papers [21, 22], are particular examples of generalized Hawkes processes. Indeed, they are Poisson cluster point processes defined as follows. (1) The process of cluster centers $\{X_n\}_{n \geq 1}$ is a Poisson process on $(0, \infty)$ (respectively, on $\mathbb{R}$) with constant intensity equal to $\lambda > 0$. (2) The points of the cluster $\theta_{X_n} Z_n(\cdot, \mathbb{R})$ are partitioned into generations and generated recursively as follows. The ancestor constitutes the generation 0 of the cluster and is located at X_n. Given X_n, the ancestor generates points of the first generation of the cluster according to a non-homogeneous Poisson process on (X_n, ∞) with intensity function $g(\cdot - X_n)$, where $g: \mathbb{R} \to [0, \infty)$ is a measurable function which is null on $(-\infty, 0]$ and such that $h := \int_0^\infty g(x)\mathrm{d}x < 1$. In turn, given the points of the first generation of the cluster, a point of this generation, which is located at X, generates points of the second generation of the cluster according to a non-homogeneous Poisson process on (X, ∞) with intensity function $g(\cdot - X)$; and so on and so forth. Note that $Z_n(\mathbb{R}, \mathbb{R}) = \theta_{X_n} Z_n(\mathbb{R}, \mathbb{R}) = \theta_{X_n} Z_n([X_n, \infty), \mathbb{R}) = Z_n([0, \infty), \mathbb{R})$ is distributed as the total progeny of a sub-critical Galton–Watson process with one ancestor and Poisson offspring law with mean h.

(ii) *Spatial Hawkes processes* on $\mathbb{R}^d$, $d \geq 1$, with parameters (λ, g), introduced in [9] and further studied in [34], are also particular examples of generalized Hawkes processes. Indeed, they are Poisson cluster point processes defined as follows. (1) The process of cluster centers $\{X_n\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d$, $d \geq 1$, with constant intensity equal to $\lambda > 0$. (2) The points of the cluster $\theta_{X_n} Z_n(\cdot, \mathbb{R})$ are partitioned into generations and generated recursively as follows. The ancestor constitutes the generation 0 of the cluster and is located at X_n. Given X_n, the ancestor generates points of the first generation of the cluster according to a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function $g(\cdot - X_n)$, where $g: \mathbb{R}^d \to [0, \infty)$ is a measurable function such that $h := \int_{\mathbb{R}^d} g(x)\mathrm{d}x < 1$. In turn, given the points of the first generation of the cluster, a point of this generation, which is located at X, generates points of the second generation of the cluster according to a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function $g(\cdot - X)$; and so on and so forth. Note that $\theta_{X_n} Z_n(\mathbb{R}^d, \mathbb{R}) = Z_n(\mathbb{R}^d, \mathbb{R})$ is distributed as the total progeny of a sub-critical Galton–Watson process with one ancestor and Poisson offspring law with mean h.

Note that, according to these definitions, classical Hawkes processes on $\mathbb{R}$ are different from spatial Hawkes processes on $\mathbb{R}$.

The collection of random variables $V \equiv \{V(C)\}_{C \in \mathcal{B}(\mathbb{R}^d)}$, where $V(C)$ is defined by (3), will be called a *generalized compound Hawkes process* if the random variable Z is distributed as the total progeny of a sub-critical Galton–Watson process with one ancestor. Note that $V(C)$ aggregates the marks attached to the points of a generalized Hawkes process which fall in C.

2.3. Interference in wireless communication

Consider the following simple model of wireless communication, which accounts for interference effects that arise when several nodes transmit at the same time. Suppose that transmitting nodes (e.g., antennas) are located according to $\{X_n\}_{n \geq 1}$, a Poisson process on the plane with intensity function $\lambda(\,\cdot\,)$—i.e., X_n is the location of node n—and denote by $Z_n \in (0, \infty)$ the signal power of the transmitting node n. Suppose that the sequence $\{Z_n\}_{n \geq 1}$ is independent of the Poisson process, and that the random variables Z_n, $n \geq 1$, are independent and identically distributed. Assume that a receiver is located at the origin $\mathbf{0} \in \mathbb{R}^2$ and that a new transmitter is added at $x \in \mathbb{R}^2$ and has signal power $y \in (0, \infty)$. Suppose that the physical propagation of the signal is described by a measurable positive function $A : \mathbb{R}^2 \to (0, \infty)$, which gives the attenuation or path loss of the signal power. For simplicity, we assume that random fading (due to occluding objects, reflections, multipath interference, etc.) is encoded in the random variables $Z_n \in (0, \infty)$. Thus, $Z_n A(X_n)$ is the power received at the origin from the transmitting node at X_n, and the total interference at the origin, due to simultaneous transmissions, is equal to

$$I(\{\mathbf{0}\}) := \sum_{n \geq 1} Z_n A(X_n).$$

Note that this is a Poisson shot noise random variable of the form (1). Indeed, let $H : \mathcal{B}(\mathbb{R}^2) \times (0, \infty) \to (0, \infty)$ be a mapping which, restricted to $\mathbb{R}^2 \times (0, \infty)$, coincides with $\widetilde{H}(x, z) := z A(- x)$. Then

$$S(\{\mathbf{0}\}) = \sum_{n \geq 1} H(\{\mathbf{0}\} - X_n, Z_n) = \sum_{n \geq 1} \widetilde{H}(- X_n, Z_n) = I(\{\mathbf{0}\}).$$

We refer the reader to [2, 3] for more insight into this model, and limit ourselves to observing that the receiver at the origin can decode the signal of power $y \in (0, \infty)$ from the transmitter at $x \in \mathbb{R}^2$ if and only if the signal-to-interference-plus-noise ratio (SINR) is bigger than a given threshold, i.e.,

$$\mathrm{SINR} := \frac{y A(x)}{I(\{\mathbf{0}\}) + w} \geq \tau,$$

where w is e.g. a thermal noise near the receiver at the origin and τ is the given threshold.

3. Informal description of the results

We start by noting that some results in this paper refer to sequences of Poisson shot noise random variables of the form

$$S_\ell(C_\ell) := \sum_{n \geq 1} H(C_\ell - X_n^{(\ell)}, Z_n^{(\ell)}), \quad \ell \geq 1, \{C_\ell\}_{\ell \geq 1} \subset \mathcal{B}(\mathbb{R}^d), \tag{6}$$

where, for each $\ell \geq 1$, $\mathcal{P}_\ell = \{(X_n^{(\ell)}, Z_n^{(\ell)})\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d \times \mathbf{Z}$ with mean measure $\lambda_\ell(x)\mathrm{d}x\mathbb{Q}_\ell(\mathrm{d}z)$, $\lambda_\ell : \mathbb{R}^d \to [0, \infty)$ is a locally integrable function, and $\mathbb{Q}_\ell(\,\cdot\,)$ is a probability measure on a measurable space $(\mathbf{Z}, \mathcal{Z})$.

The achievements of the paper are the following:

(i) We provide bounds for the Wasserstein and Kolmogorov distances (hereafter denoted by d_W and d_K, respectively; see Section 4.1 for a formal definition of these probability metrics) between a standard Gaussian random variable G and

$$T(C) := \frac{S(C) - \mathbb{E}S(C)}{\sqrt{\mathbb{V}\mathrm{ar}(S(C))}}, \quad C \in \mathcal{B}(\mathbb{R}^d); \tag{7}$$

see Theorem 3. As special cases (see Remark 1), we get bounds for the Wasserstein and Kolmogorov distances between G and

$$W(C) := \frac{V(C) - \mathbb{E}V(C)}{\sqrt{\mathbb{V}\mathrm{ar}(V(C))}}, \quad C \in \mathcal{B}(\mathbb{R}^d), \tag{8}$$

where $V(C)$ is defined by (3), and between G and

$$L(\{\mathbf{0}\}) := \frac{I(\{\mathbf{0}\}) - \mathbb{E}I(\{\mathbf{0}\})}{\sqrt{\mathbb{V}\mathrm{ar}(I(\{\mathbf{0}\}))}}, \quad \mathbf{0} \in \mathbb{R}^2, \tag{9}$$

where $I(\{\mathbf{0}\})$ is defined in Section 2.3.

(ii) We provide moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for the sequence $\{T_\ell(C_\ell)\}_{\ell \geq 1}$, where

$$T_\ell(C_\ell) := \frac{S_\ell(C_\ell) - \mathbb{E}S_\ell(C_\ell)}{\sqrt{\mathbb{V}\mathrm{ar}(S_\ell(C_\ell))}}, \quad C_\ell \in \mathcal{B}(\mathbb{R}^d)$$

see Theorem 4. As particular cases (see Remark 2), we get moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for the sequences

$$W_\ell(C_\ell) := \frac{V_\ell(C_\ell) - \mathbb{E}V_\ell(C_\ell)}{\sqrt{\mathbb{V}\mathrm{ar}(V_\ell(C_\ell))}}, \quad C_\ell \in \mathcal{B}(\mathbb{R}^d), \quad \ell \geq 1, \tag{10}$$

where $V_\ell(C_\ell)$ is defined by (3), with obvious modifications, and

$$L_\ell(\{\mathbf{0}\}) := \frac{I_\ell(\{\mathbf{0}\}) - \mathbb{E}I_\ell(\{\mathbf{0}\})}{\sqrt{\mathbb{V}\mathrm{ar}(I_\ell(\{\mathbf{0}\}))}}, \quad \mathbf{0} \in \mathbb{R}^2, \tag{11}$$

where $I_\ell(\{\mathbf{0}\})$ is defined in Section 2.3, with obvious modifications.

(iii) If the Poisson process $\{X_n\}_{n \geq 1}$ has intensity function of the form $\lambda(x) := \lambda \mathbf{1}_B(x), x \in \mathbb{R}^d$, for some positive constant $\lambda > 0$ and a suitable Borel set $B \subseteq \mathbb{R}^d$, then the bounds on $d_W(W(C), G)$ and $d_K(W(C), G)$ are particularly simple and depend only on λ, the Lebesgue measure of $B \cap C$, and a few moments of M and Z; see Corollary 1. If Z is distributed as the total progeny of a sub-critical Galton–Watson process with one ancestor, then we are able to compute the moments of Z in terms of the moments of the offspring law; see Proposition 1. This allows for explicit bounds when $V = \{V(C)\}_{C \in \mathcal{B}(\mathbb{R}^d)}$ is a generalized compound Hawkes process with Poisson or binomial offspring laws; see Corollaries 3 and 5, respectively.

(iv) If the Poisson process $\{X_n^{(\ell)}\}_{n \geq 1}$ has intensity function of the form $\lambda_\ell(x) = \lambda_\ell \mathbf{1}_{B_\ell}(x)$, $x \in \mathbb{R}^d$, for positive constants $\lambda_\ell > 0$ and suitable Borel sets $B_\ell \in \mathcal{B}(\mathbb{R}^d)$, and $\mathbb{Q}_\ell \equiv \mathbb{Q}$ for each $\ell \geq 1$, then the condition which guarantees a moderate deviation principle, a Bernstein-type concentration inequality, and a normal approximation bound with Cramér correction term for the sequence $\{W_\ell(C_\ell)\}_{\ell \geq 1}$ is quite simple (see the condition (22) of Corollary 2), and it allows for applications to generalized compound Hawkes processes with Poisson and binomial offspring distributions; see Corollaries 4 and 6, respectively.

We conclude this section by emphasizing that the basic idea of this paper is very simple (especially if compared with the techniques exploited in [23, 26] for the Gaussian approximation of classical Hawkes processes on $(0, \infty)$). Since the random variable $S(C)$ can be rewritten as the Poisson integral

$$S(C) = \int_{\mathbb{R}^d \times \mathbf{Z}} H(C - x, z) \mathcal{P}(\mathrm{d}x, \mathrm{d}z), \quad C \in \mathcal{B}(\mathbb{R}^d), \tag{12}$$

we are able to do the following:

(i) We apply the quantitative central limit theorems, in the Wasserstein and Kolmogorov metrics, for functionals of the Poisson measure proved in the seminal paper [30], to provide normal approximation bounds for first chaoses on the Poisson space (see Theorems 1), and then we apply such bounds to the random variable $T(C)$.

(ii) We prove moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction for a sequence of first chaoses on the Poisson space (see e.g. Theorem 2), and then we apply such results to the sequence $\{T_\ell(C_\ell)\}_{\ell \geq 1}$.

4. Gaussian approximation and moderate deviations of the first chaos on the Poisson space

Let $(A, \mathcal{A}, \alpha)$ be a measure space with $\alpha(\,\cdot\,)$ a σ-finite measure, and let $\mathbf{N}_A$ be the set of all σ-finite counting measures on $(A, \mathcal{A})$ equipped with the σ-field generated by the mappings $\nu \mapsto \nu(B)$, $B \in \mathcal{A}$. We say Π is a Poisson measure on $(A, \mathcal{A})$ with mean measure $\alpha(\,\cdot\,)$ if it is a measurable mapping from an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbf{N}_A$ such that (i) for any $B \in \mathcal{A}$, $\Pi(A)$ is Poisson distributed with mean $\alpha(A)$, and (ii) if $B_1, \ldots, B_n \in \mathcal{A}$, $n \in \mathbb{N}$, are pairwise disjoint, then the random variables $\Pi(B_1), \ldots, \Pi(B_n)$ are independent.

Let $\{\Pi_\ell\}_{\ell \in \mathbb{N}}$ be a sequence of Poisson measures on $(A, \mathcal{A})$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that Π_ℓ has a σ-finite mean measure $\alpha_\ell(\,\cdot\,)$, $\ell \in \mathbb{N}$. We denote by $L^m(\alpha_\ell)$ the space of measurable functions $f : A \to \mathbb{R}$ such that $\int_A |f(a)|^m \alpha_\ell(\mathrm{d}a) < \infty$, $m \in \mathbb{N}$, and, for $\{f_\ell\}_{\ell \in \mathbb{N}} \in L^2(\alpha_\ell)$, we consider

$$I^{(\ell)}(f_\ell) := \int_A f_\ell(a)(\Pi_\ell(\mathrm{d}a) - \alpha_\ell(\mathrm{d}a)), \quad \ell \in \mathbb{N},$$

the first chaos on the Poisson space, i.e., the first-order stochastic integral of f_ℓ with respect to the compensated Poisson measure $\Pi_\ell(\mathrm{d}a) - \alpha_\ell(\mathrm{d}a)$. If the law of Π_ℓ does not depend on ℓ, we suppress the dependence on ℓ of the related quantities, and for $f \in L^2(\alpha)$ we simply write

$$I(f) := \int_A f(a)(\Pi(\mathrm{d}a) - \alpha(\mathrm{d}a)).$$

4.1. Bounds on the Wasserstein and Kolmogorov distances between the law of a first chaos on the Poisson space and the standard normal distribution

Let X and Y be two real-valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The Wasserstein and Kolmogorov distances between the law of X and the law of Y, written $d_W(X, Y)$ and $d_K(X, Y)$, respectively, are defined by

$$d_W(X, Y) := \sup_{g \in \mathrm{Lip}(1)} |\mathbb{E}[g(X) - g(Y)]|$$

and

$$d_K(X, Y) := \sup_{x \in \mathbb{R}} |\mathbb{P}(X \le x) - \mathbb{P}(Y \le x)|.$$

Here, $\mathrm{Lip}(1)$ denotes the set of Lipschitz functions $g : \mathbb{R} \to \mathbb{R}$ with Lipschitz constant less than or equal to 1. We recall that throughout this paper, G denotes a random variable distributed according to the standard normal law.

Theorem 1. *Let $f \in L^2(\alpha)$ be such that $\|f\|_{L^2(\alpha)} = 1$. Then*

$$d_W(I(f), G) \le \int_A |f(a)|^3 \alpha(\mathrm{d}a)$$

and

$$d_K(I(f), G) \le \left(1 + \frac{1}{2} \max \left\{ 4, \left[4 \int_A |f(a)|^4 \alpha(\mathrm{d}a) + 2 \right]^{1/4} \right\} \right) \int_A |f(a)|^3 \alpha(\mathrm{d}a)$$

$$+ \left(\int_A |f(a)|^4 \alpha(\mathrm{d}a) \right)^{1/2}.$$

Proof. We refer the reader to [29] for all the notions of stochastic analysis on the Poisson space used in this proof. Let F be a functional of Π, i.e., $F = \mathfrak{f}(\Pi)$, where $\mathfrak{f}$ is a real-valued measurable function defined on $\mathbf{N}_A$. We recall that the difference operator D is defined by

$$D_a F = \mathfrak{f}(\Pi + \delta_a) - \mathfrak{f}(\Pi), \quad a \in A,$$

where δ_a is the Dirac measure at $a \in A$, and that the second difference operator D^2 is defined by

$$D^2_{a_1, a_2} F = D_{a_1}(D_{a_2} F), \quad a_1, a_2 \in A.$$

We also recall that the domain of D, denoted by $\mathrm{dom}(D)$, is the family of square-integrable random variables $F = \mathfrak{f}(\Pi)$ such that

$$\int_A \mathbb{E}|D_a F|^2 \alpha(\mathrm{d}a) < \infty.$$

Setting $F := I(f)$, we have that $\mathbb{E}F = 0$ with $\mathbb{V}\mathrm{ar}(F) = 1$ (as follows by applying the isometry formula for Poisson chaoses) and that $F \in \mathrm{dom}(D)$. Using Theorem 1.1 in [30] we have that

$$d_W(F, G) \le \gamma_1 + \gamma_2 + \gamma_3,$$

where

$$\begin{cases} \gamma_1 & = 2 \left(\int_{A^3} \left(\mathbb{E}[(D_{a_1}F)^2(D_{a_2}F)^2] \right)^{1/2} \left(\mathbb{E}[(D^2_{a_1,a_3}F)^2(D^2_{a_2,a_3}F)^2] \right)^{1/2} \alpha^3(\mathrm{d}(a_1, a_2, a_3)) \right)^{1/2}, \\ \gamma_2 & = \left(\int_{A^3} \mathbb{E}[(D^2_{a_1,a_3}F)^2(D^2_{a_2,a_3}F)^2] \alpha^3(\mathrm{d}(a_1, a_2, a_3)) \right)^{1/2}, \\ \gamma_3 & = \int_A \mathbb{E}|D_a F|^3 \alpha(\mathrm{d}a). \end{cases}$$

Since for $F = I(f)$ we have that $D_a F = f(a)$ and $D_{a_1, a_2} F = 0$, $a_1, a_2 \in A$, the terms γ_1 and γ_2 vanish and $\gamma_3 = \int_A |f(a)|^3 \alpha(\mathrm{d}a)$. Therefore, we obtain the bound on the Wasserstein distance between $I(f)$ and G.

Similarly, Theorem 1.2 in [30] gives

$$d_K(F, G) \leq \left(1 + \frac{1}{2}\left(\mathbb{E}F^4\right)^{1/4}\right)\int_A |f(a)|^3\alpha(\mathrm{d}a) + \left(\int_A f(a)^4\alpha(\mathrm{d}a)\right)^{1/2}.$$

Using Lemma 4.2 in [30] we have

$$\mathbb{E}F^4 \leq \max\left\{256\left(\int_A f(a)^2\alpha(\mathrm{d}a)\right)^2, 4\int_A f(a)^4\alpha(\mathrm{d}a) + 2\right\}$$

$$= \max\left\{4^4, 4\int_A f(a)^4\alpha(\mathrm{d}a) + 2\right\} \quad \text{(because } \|f\|_{L^2(\alpha)} = 1),$$

which yields the upper bound on the Kolmogorov distance. $\qquad\square$

4.2. Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for first chaoses on the Poisson space

We start with a definition.

Definition 1. Let $\{Y_\ell\}_{\ell\in\mathbb{N}}$ be a sequence of real-valued random variables, $\gamma \geq 0$ a non-negative constant, and $\{\Delta_\ell\}_{\ell\in\mathbb{N}}$ a positive numerical sequence. We make the following definitions:

(1) The sequence $\{Y_\ell\}_{\ell\in\mathbb{N}}$ satisfies a moderate deviation principle with parameters γ and $\{\Delta_\ell\}_{\ell\in\mathbb{N}}$ (**MDP**$(\gamma, \{\Delta_\ell\}_{\ell\in\mathbb{N}})$ for short) if, for any sequence of positive numbers $\{a_\ell\}_{\ell\in\mathbb{N}}$ such that $\lim_{\ell\to\infty} a_\ell = +\infty$ and $\lim_{\ell\to\infty}\frac{a_\ell}{\Delta_\ell^{1/(1+2\gamma)}} = 0$, the sequence $\{Y_\ell\}_{\ell\in\mathbb{N}}$ satisfies a large deviation principle with speed a_ℓ^2 and rate function $J(x) := x^2/2$, i.e., for any Borel set $B \subset \mathbb{R}$,

$$-\inf_{x\in\overset{\circ}{B}} J(x) \leq \liminf_{\ell\to\infty} a_\ell^{-2}\log\mathbb{P}\left(Y_\ell/a_\ell \in B\right) \leq \limsup_{\ell\to\infty} a_\ell^{-2}\log\mathbb{P}\left(Y_\ell/a_\ell \in B\right) \leq -\inf_{x\in\overline{B}} J(x),$$

where $\overset{\circ}{B}$ denotes the interior of B and $\overline{B}$ denotes the closure of B.

(2) The sequence $\{Y_\ell\}_{\ell\in\mathbb{N}}$ satisfies a Bernstein-type concentration inequality with parameters γ and $\{\Delta_\ell\}_{\ell\in\mathbb{N}}$ (**BCI**$(\gamma, \{\Delta_\ell\}_{\ell\in\mathbb{N}})$ for short) if, for all $\ell \in \mathbb{N}$ and $x \geq 0$, we have

$$\mathbb{P}(|Y_\ell| \geq x) \leq 2\exp\left(-\frac{1}{4}\min\left\{\frac{x^2}{2^{1+\gamma}}, (x\Delta_\ell)^{1/(1+\gamma)}\right\}\right).$$

(3) The sequence $\{Y_\ell\}_{\ell\in\mathbb{N}}$ satisfies a normal approximation bound with Cramér correction term with parameters γ and $\{\Delta_\ell\}_{\ell\in\mathbb{N}}$ (**NACC**$(\gamma, \{\Delta_\ell\}_{\ell\in\mathbb{N}})$ for short), if there exist positive constants $c_0, c_1, c_2 > 0$ depending only on γ such that for all $\ell \in \mathbb{N}$ and $x \in [0, c_0\Delta_\ell^{1/(1+2\gamma)}]$,

$$\mathbb{P}(Y_\ell \geq x) = e^{L_{\ell,x}^+}[1 - \mathbb{P}(G \leq x)]\left(1 + c_1\theta_{\ell,x}^+\frac{1+x}{\Delta_\ell^{1/(1+2\gamma)}}\right)$$

and

$$\mathbb{P}(Y_\ell \leq -x) = e^{L_{\ell,x}^-}[1 - \mathbb{P}(G \leq -x)]\left(1 + c_1\theta_{\ell,x}^- \frac{1+x}{\Delta_\ell^{1/(1+2\gamma)}}\right),$$

where $\theta_{\ell,x}^\pm \in [-1, 1]$ and $L_{\ell,x}^\pm \in \left(-c_2 \frac{x^3}{\Delta_\ell^{1/(1+2\gamma)}}, c_2 \frac{x^3}{\Delta_\ell^{1/(1+2\gamma)}}\right)$.

As a preliminary result, we provide moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for sequences of first chaoses on the Poisson space.

Theorem 2. *Assume the following: (i) $f_\ell \in L^m(\alpha_\ell)$ for any $m \geq 2$ and any $\ell \in \mathbb{N}$ with $\|f_\ell\|_{L^2(\alpha_\ell)} = 1$ for any $\ell \in \mathbb{N}$; (ii) there exist a constant $\gamma \geq 0$ and a positive numerical sequence $\{\Delta_\ell\}_{\ell \in \mathbb{N}}$ such that*

$$\left|\int_A f_\ell(a)^m \alpha_\ell(\mathrm{d}a)\right| \leq \frac{(m!)^{1+\gamma}}{\Delta_\ell^{m-2}}, \quad \text{for all } m \geq 3 \text{ and } \ell \in \mathbb{N}.$$

Then the sequence $\{I^{(\ell)}(f_\ell)\}_{\ell \in \mathbb{N}}$ satisfies an **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *a* **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *and an* **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$.

Proof. We recall that for real-valued random variables $X_1, \ldots, X_m$, $m \in \mathbb{N}$, the joint cumulant of $X_1, \ldots, X_m$ is defined as

$$\mathrm{cum}(X_1, \ldots, X_m) := (-\mathbf{i})^m \frac{\partial^m}{\partial t_1 \ldots \partial t_m} \log \varphi_{X_1,\ldots,X_m}(t_1, \ldots, t_m)\Big|_{t_1=\ldots=t_m=0},$$

where $\mathbf{i}$ is the imaginary unit and $\varphi_{X_1,\ldots,X_m}$ is the joint characteristic function of $(X_1, \ldots, X_m)$. For a real-valued random variable X and $m \in \mathbb{N}$ we shall write $\mathrm{cum}_m(X) := \mathrm{cum}(X, \ldots, X)$ for the mth cumulant of X.

For an arbitrarily fixed $\ell \in \mathbb{N}$, set $X_\ell := I^{(\ell)}(f_\ell)$. Clearly, $\mathbb{E}X_\ell = 0$ and $\mathbb{E}X_\ell^2 = 1$ (which is a consequence of the isometry formula for Poisson chaoses). Then the claim follows by the theory developed in [40] (see e.g. Proposition 2.1 in [41]; see also [13, 14]) if we prove that

$$\mathbb{E}|X_\ell|^m < \infty \quad \text{for any } \ell \in \mathbb{N} \text{ and } m \geq 3 \tag{13}$$

and

$$\mathrm{cum}_m(X_\ell) = \int_A f_\ell(a)^m \alpha_\ell(\mathrm{d}a) \quad \text{for any } \ell \in \mathbb{N} \text{ and } m \geq 3. \tag{14}$$

To this end we are going to apply Theorem 3.6 of [41]. A partition σ of $\{1, \ldots, m\}$, $m \geq 3$, is a collection $\{B_1, \ldots, B_k\}$ of $1 \leq k \leq m$ pairwise disjoint non-empty sets, called blocks, such that $B_1 \cup \ldots \cup B_k = \{1, \ldots, m\}$. The number k of blocks of a partition σ is denoted by $|\sigma|$. Let $J_j := \{j\}, j \in \{1, \ldots, m\}$. Letting $\Pi(\mathbf{1}_m)$, $\mathbf{1}_m := (1, \ldots, 1) \in \mathbb{R}^m$, denote the set of all partitions σ of $\{1, \ldots, m\}$ whose blocks B are such that $\mathrm{Card}(B \cap J_j) \leq 1$ for every $j \in \{1, \ldots, m\}$, we clearly have that $\Pi(\mathbf{1}_m)$ is the set of all partitions of $\{1, \ldots, m\}$. Letting $\widetilde{\Pi}(\mathbf{1}_m)$ denote the set of all partitions $\sigma \in \Pi(\mathbf{1}_m)$ with $|\sigma| = 1$, we clearly have $\widetilde{\Pi}(\mathbf{1}_m) = \{\{1, \ldots, m\}\}, m \geq 3$. Letting $\widetilde{\Pi}_{\geq 2}(\mathbf{1}_m)$ denote the set of all partitions $\sigma \in \widetilde{\Pi}(\mathbf{1}_m)$ whose blocks have cardinality bigger than or equal to 2, since $m \geq 3$, we clearly have $\widetilde{\Pi}_{\geq 2}(\mathbf{1}_m) = \widetilde{\Pi}(\mathbf{1}_m) = \{\{1, \ldots, m\}\}$. We denote by $\Pi_{\geq 2}(\mathbf{1}_m)$, $m \geq 2$, the family of all partitions $\sigma \in \Pi(\mathbf{1}_m)$, i.e., the family of all partitions σ of $\{1, \ldots, m\}$ whose blocks have cardinality bigger than or equal to 2.

For a function $g : A \to \mathbb{R}$, set

$$(\otimes_{j=1}^{m} g)(x_1, \ldots, x_m) := g(x_1) \ldots g(x_m).$$

For $\sigma \in \Pi(\mathbf{1}_m)$, define the function $(\otimes_{j=1}^{m} g)_\sigma : A^{|\sigma|} \to \mathbb{R}$ by replacing in $(\otimes_{j=1}^{m} g)(x_1, \ldots, x_m)$ all the variables whose indexes belong to the same block of σ by a new common variable. Note that for $\sigma \in \Pi(\mathbf{1}_m)$, $(\otimes_{j=1}^{m} g)_\sigma : A^{|\sigma|} \to \mathbb{R}$ can be represented as

$$(\otimes_{j=1}^{m} g)_\sigma(a_1, \ldots, a_{|\sigma|}) = \prod_{i=1}^{|\sigma|} g(a_i)^{|B_i|}, \quad a_1, \ldots, a_{|\sigma|} \in A,$$

where $B_1, \ldots, B_{|\sigma|}$ are the blocks of σ and $|B_i|$, $i = 1, \ldots, |\sigma|$, is the cardinality of the block B_i. In particular, for $\sigma \in \widetilde{\Pi}_{\geq 2}(\mathbf{1}_m)$, $(\otimes_{j=1}^{m} g)_\sigma(a) := g(a)^m$, $a \in A$, and, for $\sigma \in \Pi_{\geq 2}(\mathbf{1}_m)$,

$$(\otimes_{j=1}^{m} g)_\sigma(a_1, \ldots, a_{|\sigma|}) := \prod_{i=1}^{|\sigma|} g(a_i)^{|B_i|}, \quad a_1, \ldots, a_{|\sigma|} \in A,$$

where $B_1, \ldots, B_{|\sigma|}$ are the blocks of σ and $|B_i| \geq 2$ for any $i = 1, \ldots, |\sigma|$. Therefore the hypothesis (i) implies the assumptions (3.4) and (3.5) of Theorem 3.6 in [41], and so (13) and (14) hold. $\square$

5. Application to Poisson shot noise random variables

5.1. Gaussian approximation

In this section, we apply Theorem 1 to the standardized random variables $T(C)$, $C \in \mathcal{B}(\mathbb{R}^d)$, defined by (7). Note that, for $C \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbb{E}S(C) = \int_{\mathbb{R}^d} \lambda(x) \mathbb{E}H(C - x, Z_1) \mathrm{d}x,$$

and that, by the isometry formula for Poisson chaoses, if $\int_{\mathbb{R}^d} \lambda(x) \mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x < \infty$, then

$$\mathbb{V}\mathrm{ar}(S(C)) = \int_{\mathbb{R}^d} \lambda(x) \mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x.$$

Therefore the random variable $T(C)$ is well defined and finite for any $C \in \mathcal{B}(\mathbb{R}^d)$ such that

$$0 < \int_{\mathbb{R}^d} \lambda(x) \mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x < \infty. \tag{15}$$

The following theorem holds.

Theorem 3. *Let $C \in \mathcal{B}(\mathbb{R}^d)$ be such that (15) holds.*
Then

$$d_W(T(C), G) \leq \frac{\int_{\mathbb{R}^d} \lambda(x) \mathbb{E}|H(C - x, Z_1)|^3 \mathrm{d}x}{\left(\int_{\mathbb{R}^d} \lambda(x) \mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x\right)^{3/2}}$$

and

$d_K(T(C), G)$

$$\leq \left[1 + \frac{1}{2} \max\left\{ 4, \left[4 \frac{\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^4 \mathrm{d}x}{\left(\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x\right)^2} + 2 \right]^{1/4} \right\} \right] \frac{\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}|H(C - x, Z_1)|^3 \mathrm{d}x}{\left(\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x\right)^{3/2}}$$

$$+ \left(\frac{\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^4 \mathrm{d}x}{\left(\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x\right)^2} \right)^{1/2}.$$

Proof. By (1) we have

$$T(C) = \frac{1}{\sqrt{\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x}} \int_{\mathbb{R}^d \times \mathbf{M}} H(C - x, z)(\mathcal{P}(\mathrm{d}x, \mathrm{d}z) - \lambda(x)\mathrm{d}x\mathbb{Q}(\mathrm{d}z)),$$

$C \in \mathcal{B}(\mathbb{R}^d)$. Therefore $T(C)$ belongs to the first chaos of $\mathcal{P}$ with kernel

$$t(x, z) := \frac{H(C - x, z)}{\sqrt{\int_{\mathbb{R}^d} \lambda(x)\mathbb{E}H(C - x, Z_1)^2 \mathrm{d}x}}.$$

The claim follows from applying Theorem 1. □

Remark 1. As particular cases of Theorem 3 we have the following:

(i) If $H(C - x, z) := \upsilon(z)(C - x)$, with $C \in \mathcal{B}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, $z \in \mathbf{Z} := \mathbf{N}_{\mathbb{R}^d \times \mathbb{R}^p}$, and $\upsilon(z)(C)$ is defined by (4), then $S(C) = V(C)$ where the random variable $V(C)$ is defined by (3). So Theorem 3 provides Gaussian approximation bounds for the random variable $W(C)$ defined by (8). An interesting special case is investigated in Section 6.1.

(ii) If $H : \mathcal{B}(\mathbb{R}^2) \times (0, \infty) \to (0, \infty)$ is a mapping such that its restriction to $\mathbb{R}^2 \times (0, \infty)$ coincides with $\widetilde{H}(x, z) := zA(-x)$, $x \in \mathbb{R}^2$ and $z \in \mathbf{Z} := (0, \infty)$, then $S(\{\mathbf{0}\}) = I(\{\mathbf{0}\})$ where the random variable $I(\{\mathbf{0}\})$ is defined in Section 2.3. So Theorem 3 provides Gaussian approximation bounds for the random variable $L(\{\mathbf{0}\})$ defined by (9). An interesting special case is investigated in Section 8.1.

5.2. Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term

In this section, we apply Theorem 2 to the sequence of standardized random variables $\{T_\ell(C_\ell)\}_{\ell \geq 1}$, $\{C_\ell\}_{\ell \geq 1} \subset \mathcal{B}(\mathbb{R}^d)$, defined by (6).

The following theorem holds.

Theorem 4. *Let* $\{C_\ell\}_{\ell \in \mathbb{N}} \subset \mathcal{B}(\mathbb{R}^d)$ *be a sequence of Borel sets such that*

$$0 < \int_{\mathbb{R}^d} \lambda_\ell(x)\mathbb{E}H(C_\ell - x, Z_1^{(\ell)})^2 \mathrm{d}x < \infty, \quad \ell \geq 1, \tag{16}$$

and assume that there exist a non-negative constant $\gamma \geq 0$ *and a positive numerical sequence* $\{\Delta_\ell\}_{\ell \in \mathbb{N}}$ *such that*

$$\frac{\int_{\mathbb{R}^d} \lambda_\ell(x)\mathbb{E}|H(C_\ell - x, Z_1^{(\ell)})|^m \mathrm{d}x}{\left(\int_{\mathbb{R}^d} \lambda_\ell(x)\mathbb{E}H(C_\ell - x, Z_1^{(\ell)})^2 \mathrm{d}x\right)^{\frac{m}{2}}} \leq \frac{(m!)^{1+\gamma}}{\Delta_\ell^{m-2}}, \quad \text{for all } m \geq 3 \text{ and } \ell \in \mathbb{N}. \tag{17}$$

Then the sequence $\{T_\ell(C_\ell)\}_{\ell \geq 1}$ *satisfies an* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$*, a* **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$*, and an* **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$*.*

Proof. Similarly to the proof of Theorem 3, by (6) we have

$$T_\ell(C_\ell) = \frac{1}{\sqrt{\int_{\mathbb{R}^d} \lambda_\ell(x)\mathbb{E}H(C_\ell - x, Z_1^{(\ell)})^2 \mathrm{d}x}} \int_{\mathbb{R}^d \times \mathbf{Z}} H(C_\ell - x, z)(\mathcal{P}_\ell(\mathrm{d}x, \mathrm{d}z) - \lambda_\ell(x)\mathrm{d}x\mathbb{Q}_\ell(\mathrm{d}z)),$$

$C_\ell \in \mathcal{B}(\mathbb{R}^d)$. Therefore $T_\ell(C_\ell)$ belongs to the first chaos of $\mathcal{P}_\ell$ with kernel

$$t_\ell(x, z) := \frac{H(C_\ell - x, z)}{\sqrt{\int_{\mathbb{R}^d} \lambda_\ell(x)\mathbb{E}H(C_\ell - x, Z_1^{(\ell)})^2 \mathrm{d}x}}.$$

The claim follows by Theorem 2. $\qquad\square$

Remark 2. As particular cases of Theorem 4 we have the following:

(i) If $H(C_\ell - x, z) := \upsilon(z)(C_\ell - x)$, with $C_\ell \in \mathcal{B}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, $z \in \mathbf{Z} := \mathbf{N}_{\mathbb{R}^d \times \mathbb{R}}$, and $\upsilon(z)(\,\cdot\,)$ is defined by (4), then $S_\ell(C_\ell) = V_\ell(C_\ell)$ where the random variable $V_\ell(C_\ell)$ is defined by (3) with λ_ℓ, $\mathbb{Q}_\ell$, and C_ℓ in place of λ, $\mathbb{Q}$, and C, respectively. So Theorem 4 provides moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for the sequence $\{W_\ell(C_\ell)\}_{\ell \geq 1}$, where $W_\ell(C_\ell)$ is defined by (8), with obvious modifications. An interesting special case is investigated in Section 6.2.

(ii) If $H : \mathcal{B}(\mathbb{R}^2) \times (0, \infty) \to (0, \infty)$ is a mapping such that its restriction to $\mathbb{R}^2 \times (0, \infty)$ coincides with $\widetilde{H}(x, z) := zA(-x)$, $x \in \mathbb{R}^2$ and $z \in \mathbf{Z} := (0, \infty)$, then $S_\ell(\{\mathbf{0}\}) = I_\ell(\{\mathbf{0}\})$ where the random variable $I_\ell(\{\mathbf{0}\})$ is defined as in Section 2.3, with λ_ℓ and $\mathbb{Q}_\ell$ in place of λ and $\mathbb{Q}$, respectively. So Theorem 4 provides moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term for the sequence $\{L_\ell(\{\mathbf{0}\})\}_{\ell \geq 1}$, where $L_\ell(\{\mathbf{0}\})$ is defined by (9), with obvious modifications. An interesting special case is investigated in Section 8.2.

6. Application to a class of compound Poisson cluster point processes

6.1. Gaussian approximation

In this section we apply Theorem 3 to the class of standardized compound Poisson cluster point processes $\{W(C)\}_{C \in \mathcal{B}(\mathbb{R}^d)}$ with $\{X_n\}_{n \geq 1}$ having a piecewise constant intensity function. In such a case we have more explicit upper bounds on the Wasserstein and Kolmogorov distances, which pave the way to explicit bounds for some classes of generalized compound Hawkes processes (see Corollaries 3 and 5). We recall that Z denotes the random total number of points of the progeny process, i.e. $Z = Z_1(\mathbb{R}^d, \mathbb{R})$, and that M is a generic random variable with the same distribution as the independent and identically distributed marks.

Hereafter, we denote by $\mathrm{Leb}(\,\cdot\,)$ the Lebesgue measure on $\mathbb{R}^d$.

Corollary 1. *Let* $(B, C) \in \mathcal{B}(\mathbb{R}^d)^2$ *be such that* $0 < \mathrm{Leb}(B \cap C) < +\infty$*. If* $\lambda(x) = \lambda \mathbf{1}_B(x)$ *for any* $x \in \mathbb{R}^d$ *and some positive constant* $\lambda > 0$*,* $\mathbb{E}Z^2 < \infty$*, and* $\mathbb{E}M^2 \in (0, \infty)$*, then*

$$d_W(W(C), G) \leq \frac{\mathbb{E}|M|^3 \mathbb{E}Z^3}{(\mathbb{E}M^2)^{3/2}\sqrt{\lambda\mathrm{Leb}(B \cap C)}} \tag{18}$$

and

$$d_K(W(C), G) \le \left[1 + \frac{1}{2} \max \left\{ 4, \left[4 \frac{\mathbb{E}M^4 \mathbb{E}Z^4}{\lambda \mathrm{Leb}(B \cap C)(\mathbb{E}M^2)^2} + 2 \right]^{1/4} \right\} \right] \frac{\mathbb{E}|M|^3 \mathbb{E}Z^3}{(\mathbb{E}M^2)^{3/2} \sqrt{\lambda \mathrm{Leb}(B \cap C)}}$$

$$+ \left(\frac{\mathbb{E}M^4 \mathbb{E}Z^4}{\lambda \mathrm{Leb}(B \cap C)(\mathbb{E}M^2)^2} \right)^{1/2}. \tag{19}$$

We point out that many articles in the literature (e.g. [4, 23, 26]) consider Hawkes processes with an empty history, that is, with no points in $(-\infty, 0]$, which corresponds to the piecewise constant intensity function $\lambda(x) = \mathbf{1}_{[0,+\infty)}(x)$ in Corollary 1.

The proof of Corollary 1 exploits the following lemma, which is proved in Appendix A.

Lemma 1. *For any $(B, C) \in \mathcal{B}(\mathbb{R}^d)^2$ and $m \in \mathbb{N}$, we have*

$$\int_B \mathbb{E}|v(Z_1)(C - x)|^m \mathrm{d}x \le \mathrm{Leb}(B \cap C)\mathbb{E}Z^m \mathbb{E}|M|^m. \tag{20}$$

Proof of Corollary 1. In order to apply Theorem 3 we need to verify (15) with $H(C - x, z) := v(z)(C - x)$, $C \in \mathcal{B}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, $z \in \mathbf{Z} := \mathbf{N}_{\mathbb{R}^d \times \mathbb{R}}$, and $\lambda(\cdot) \equiv \lambda \mathbf{1}_B(\cdot)$. For the lower bound we note that

$$\int_B \mathbb{E}v(Z_1)(C - x)^2 \mathrm{d}x = \int_B \mathbb{E} \left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} M_{1,k} \right)^2 \mathrm{d}x \ge \int_{B \cap C} \mathbb{E} \left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} M_{1,k} \right)^2 \mathrm{d}x.$$

Expanding the square of the sum and using the independence we have that

$$\mathbb{E} \left(\sum_{k=1}^{Z_1(C-x,\mathbb{R})} M_{1,k} \right)^2 = \mathbb{E}[Z_1(C - x, \mathbb{R})]\mathbb{E}M^2 + \mathbb{E}[Z_1(C - x, \mathbb{R})(Z_1(C - x, \mathbb{R}) - 1)](\mathbb{E}M)^2.$$

For $x \in B \cap C$, we have $x \in C$, and so the set $C - x$ contains the origin. Since $Z_1(\{\mathbf{0}\}, \mathbb{R}) = 1$, we then have

$$\mathbb{E} \left(\sum_{k=1}^{Z_1(C-x,\mathbb{R})} M_{1,k} \right)^2 \ge \mathbb{E}M^2.$$

Therefore,

$$\int_B \mathbb{E}v(Z_1)(C - x)^2 \mathrm{d}x \ge \int_{B \cap C} \mathbb{E}v(Z_1)(C - x)^2 \mathrm{d}x \ge \mathrm{Leb}(B \cap C)\mathbb{E}M^2 > 0. \tag{21}$$

As far as the upper bound is concerned, we note that by the bound on the Wasserstein distance in Theorem 3 and the inequalities (21) and (20), it immediately follows that

$$d_W(F(C), G) \le \frac{\lambda \int_B \mathbb{E}|v(Z_1)(C - x)|^3 \mathrm{d}x}{\left(\int_B \lambda \mathbb{E}v(Z_1)(C - x)^2 \mathrm{d}x \right)^{3/2}} \le \frac{\lambda \mathrm{Leb}(B \cap C)\mathbb{E}Z^3 \mathbb{E}|M|^3}{\left(\lambda \mathbb{E}M^2 \mathrm{Leb}(B \cap C) \right)^{3/2}}$$

$$= \frac{\mathbb{E}|M|^3 \mathbb{E}Z^3}{(\mathbb{E}M^2)^{3/2} \sqrt{\lambda \mathrm{Leb}(B \cap C)}}.$$

Similarly, the upper bound on the Kolmogorov distance follows from the upper bound on the Kolmogorov distance in Theorem 3, and again the inequalities (21) and (20). $\quad\square$

6.2. Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term

In this section we apply Theorem 4 to the sequence $\{W_\ell(C_\ell)\}_{\ell \geq 1}$, when the Poisson processes $\{X_n^{(\ell)}\}$, $\ell \geq 1$, have a piecewise deterministic intensity function and $\mathbb{Q}_\ell \equiv \mathbb{Q}$ for every $\ell \geq 1$. In such a case the assumption (17) is greatly simplified. Moreover, the next corollary paves the way to the application of the theorem to some classes of generalized compound Hawkes processes (see Corollaries 4 and 6).

Corollary 2. *Let* $\{(B_\ell, C_\ell)\}_{\ell \in \mathbb{N}} \subset \mathcal{B}(\mathbb{R}^d)^2$ *be such that* $0 < \mathrm{Leb}(B_\ell \cap C_\ell) < +\infty$, $\ell \in \mathbb{N}$. *Assume that* $\lambda_\ell(x) = \lambda_\ell \mathbf{1}_{B_\ell}(x)$, $x \in \mathbb{R}^d$, *for positive constants* $\lambda_\ell > 0$, $\ell \in \mathbb{N}$, $\mathbb{Q}_\ell \equiv \mathbb{Q}$, $\mathbb{E}M^2 > 0$, *and*

$$\frac{\mathbb{E}|M|^m \mathbb{E}Z^m}{(\mathbb{E}M^2)^{m/2} \sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{(m!)^{1+\gamma}}{\Delta_\ell^{m-2}}, \quad \text{for all } m \geq 3 \text{ and } \ell \in \mathbb{N}, \qquad (22)$$

for some $\gamma \geq 0$ *and a positive numerical sequence* $\{\Delta_\ell\}_{\ell \in \mathbb{N}}$. *Then the sequence* $\{W_\ell(C_\ell)\}_{\ell \geq 1}$ *satisfies an* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *a* **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *and an* **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$.

Proof. By (21), (20), the choice of the Borel sets B_ℓ and C_ℓ, $\ell \in \mathbb{N}$, the assumption (22), and the fact that $\mathbb{E}M^2 > 0$, we have

$$0 < \mathrm{Leb}(B_\ell \cap C_\ell)\mathbb{E}M^2 \leq \int_{B_\ell} \mathbb{E}\upsilon(Z_1)(C_\ell - x)^2 \mathrm{d}x$$

and

$$\int_{B_\ell} \mathbb{E}|\upsilon(Z_1)(C_\ell - x)|^m \mathrm{d}x \leq \mathrm{Leb}(B_\ell \cap C_\ell)\mathbb{E}|M|^m \mathbb{E}Z^m < \infty, \quad \text{for every } m \geq 3.$$

Therefore the condition (16) holds, and using again the assumption (22), we have

$$\frac{\lambda_\ell \int_{B_\ell} \mathbb{E}|\upsilon(Z_1)(C_\ell - x)|^m \mathrm{d}x}{\left(\lambda_\ell \int_{B_\ell} \mathbb{E}(\upsilon(Z_1)(C_\ell - x))^2 \mathrm{d}x\right)^{\frac{m}{2}}} \leq \frac{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)\mathbb{E}|M|^m \mathbb{E}Z^m}{(\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)\mathbb{E}M^2)^{\frac{m}{2}}}$$

$$= \frac{\mathbb{E}|M|^m \mathbb{E}Z^m}{(\mathbb{E}M^2)^{m/2} \sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}}$$

$$\leq \frac{(m!)^{1+\gamma}}{\Delta_\ell^{m-2}}, \quad \text{for all } m \geq 3 \text{ and } \ell \in \mathbb{N}.$$

The claim follows by Theorem 4. $\qquad\square$

7. Application to generalized compound Hawkes processes

We start with a proposition which expresses the moments of the total progeny of a Galton–Watson process with one ancestor in terms of moments of the offspring distribution.

Proposition 1. *Suppose that* Z *is distributed as the total progeny of a Galton–Watson process with one ancestor, and let* P *be a random variable distributed according to the offspring law of the Galton–Watson process. Assume that the Galton–Watson process is sub-critical, i.e.*

$$\mathbb{E}P \in (0, 1). \qquad (23)$$

Then, for any $n \in \mathbb{N}$ such that

$$\mathbb{E}[P^n] < +\infty, \tag{24}$$

we have that $\mathbb{E}Z^n < \infty$ and

$$\mathbb{E}[Z^n] = 1 + \sum_{k=1}^{n} k! \binom{n}{k} \sum_{i=1}^{k} \frac{\mathbb{E}[(P)_i]}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\mathbb{E}[Z^{m_1}]}{m_1!} \cdots \frac{\mathbb{E}[Z^{m_i}]}{m_i!}, \tag{25}$$

where $\mathbb{E}(P)_1 = \mathbb{E}P$,

$$\mathbb{E}(P)_i = \mathbb{E}P(P-1) \cdot \ldots \cdot P(P-(i-1)), \quad 2 \le i \le n,$$

and the third sum in (25) is taken over all the $m_1, \ldots, m_i \in \mathbb{N}$ such that $m_1 + \ldots + m_i = k$.

In particular,

$$\mathbb{E}Z^2 = \frac{\mathbb{V}\mathrm{ar}(P) + 1 - \mathbb{E}P}{(1 - \mathbb{E}P)^3},$$

$$\mathbb{E}Z^3 = \frac{1}{1 - \mathbb{E}P} \left(1 + 3\frac{\mathbb{E}P}{1 - \mathbb{E}P} + 3\frac{\mathbb{E}(P)_2}{(1 - \mathbb{E}P)^2} + \frac{\mathbb{E}(P)_3 + 3\mathbb{V}\mathrm{ar}(P)}{(1 - \mathbb{E}P)^3} + 3\frac{\mathbb{V}\mathrm{ar}(P)^2}{(1 - \mathbb{E}P)^4} \right)$$

and

$$\mathbb{E}Z^4 = \frac{1}{1 - \mathbb{E}P} \left[1 + 4\frac{\mathbb{E}P}{1 - \mathbb{E}P} + 6\frac{\mathbb{E}(P)_2}{(1 - \mathbb{E}P)^2} + 4\frac{\mathbb{E}(P)_3}{(1 - \mathbb{E}P)^3} + \frac{\mathbb{E}(P)_4}{(1 - \mathbb{E}P)^4} \right.$$

$$\left. + 3\mathbb{E}Z^2 \left(2\mathbb{E}P + 4\frac{\mathbb{E}(P)_2}{1 - \mathbb{E}P} + \mathbb{E}(P)_2 \mathbb{E}Z^2 + 2\frac{\mathbb{E}(P)_3}{(1 - \mathbb{E}P)^2} \right) + 4\mathbb{E}Z^3 \frac{\mathbb{V}\mathrm{ar}(P)}{1 - \mathbb{E}P} \right].$$

As an immediate consequence of this proposition and Corollary 1, we have that if the point processes $Z_i(\cdot, \mathbb{R})$ are such that Z is distributed as the total progeny of a Galton–Watson process with one ancestor and offspring law satisfying (23), then the following two statements hold: (i) if the offspring law satisfies (24) with $n = 3$ and $\mathbb{E}M^2 \in (0, \infty)$, then the relation (18) holds and the upper bound on $d_W(W(C), G)$ is explicit and depends only on λ, the Lebesgue measure of $B \cap C$, the first three moments of the offspring law and the second and third moments of $|M|$; (ii) if the offspring law satisfies (24) with $n = 4$ and $\mathbb{E}M^2 \in (0, \infty)$, then the relation (19) holds and the upper bound on $d_K(W(C), G)$ is explicit and depends only on λ, the Lebesgue measure of $B \cap C$, the first four moments of the offspring law, and the second, third, and fourth moments of $|M|$.

The cases of the Poisson offspring law (which includes compound Hawkes processes) and the binomial offspring law are treated in detail in Sections 7.1.1 and 7.4.1, respectively.

The proof of Theorem 1 exploits the following lemma. Hereafter, for a sufficiently smooth function f, we denote by $f^{(n)}$ its derivative of order $n \in \mathbb{N}$.

Lemma 2. *(Faà di Bruno formula.) For any sufficiently smooth functions g and h,*

$$(g \circ h)^{(j)}(x) = j! \sum_{i=1}^{j} \frac{g^{(i)}(h(x))}{i!} \sum_{m_1+m_2+\ldots+m_i=j} \frac{h^{(m_1)}(x)}{m_1!} \cdots \frac{h^{(m_i)}(x)}{m_i!}, \quad j \in \mathbb{N},$$

where the second sum is taken over all the $m_1, \ldots, m_i \in \mathbb{N}$ such that $m_1 + \ldots + m_i = j$.

Proof of Proposition 1. We divide the proof into three steps. In the first step we provide a functional equation for $\mathbb{E}e^{\theta Z}$, $\theta \in (-\infty, 0)$. In the second step we prove the finiteness of the moments of Z and the formula (25). In the third step we compute the second, third, and fourth moments of Z.

Step 1: A functional equation for $\mathbb{E}e^{\theta Z}$, $\theta \in (-\infty, 0)$.

We note that Z can be represented as $Z = \sum_{n \geq 0} K_n$, where $K_0 = 1$ and K_n is the number of offspring in the nth generation of the related Galton–Watson process. Let $\{Z_j\}_{j \geq 1}$ be independent copies of Z. For any $\theta \in (-\infty, 0)$, by standard computations we have

$$\mathbb{E}[e^{\theta Z}] = e^{\theta} \sum_{k \geq 0} \mathbb{E}[e^{\theta \sum_{j=1}^{k} Z_j} \mid K_1 = k] p_k$$

$$= e^{\theta} \sum_{k \geq 0} \mathbb{E}[e^{\theta Z}]^k p_k = e^{\theta} G_P(\mathbb{E}[e^{\theta Z}]) < \infty, \tag{26}$$

where $\{p_k\}_{k \geq 0}$ is the law of P and G_P is the probability generating function of P.

Step 2: Proof of $\mathbb{E}Z^n < \infty$, $n \in \mathbb{N}$, and of (25).

As far as the moments of Z are concerned, we start by showing that they coincide with the left derivative of the moment generating function at zero. For any $\theta < 0$, the theorem of differentiation under the expected value yields, for any non-negative integer n,

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[e^{\theta Z}] = \mathbb{E}[Z^n e^{\theta Z}] < \infty.$$

The family $(Z^n e^{\theta Z})_{\theta < 0}$ is nonnegative and increasing in θ; hence, using the Beppo Levi theorem,

$$\lim_{\theta \uparrow 0} \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[e^{\theta Z}] = \lim_{\theta \uparrow 0} \mathbb{E}\left[Z^n e^{\theta Z}\right]$$

$$= \mathbb{E}\left[\lim_{\theta \uparrow 0} Z^n e^{\theta Z}\right]$$

$$= \mathbb{E}\left[Z^n\right], \tag{27}$$

where the equality holds whether the quantities are finite or infinite. Next, we combine the Faà di Bruno formula with the following elementary relation:

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}(f(x)g(x)) = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x), \quad n \in \mathbb{N} \cup \{0\}, \tag{28}$$

for sufficiently smooth functions f and g. For any $\theta \in (-\infty, 0)$, by (26) and (28), for any non-negative integer n we have

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[e^{\theta Z}] = e^{\theta} \sum_{k=0}^{n} \binom{n}{k} \frac{\mathrm{d}^k}{\mathrm{d}\theta^k} G_P(\mathbb{E}[e^{\theta Z}]).$$

By the Faà di Bruno formula we have

$$\frac{\mathrm{d}^k}{\mathrm{d}\theta^k} G_P(\mathbb{E}[e^{\theta Z}]) = k! \sum_{i=1}^{k} \frac{G_P^{(i)}(\mathbb{E}[e^{\theta Z}])}{i!} \sum_{m_1 + m_2 + \ldots + m_i = k} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[e^{\theta Z}]}{m_1!} \ldots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[e^{\theta Z}]}{m_i!}, \quad k \in \mathbb{N},$$

where the sum is taken over all the $m_1, \ldots, m_i \in \mathbb{N}$ such that $m_1 + \ldots + m_i = k$. Then, for any $\theta \in (-\infty, 0)$ and $n \in \mathbb{N} \cup \{0\}$,

$$\frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[\mathrm{e}^{\theta Z}]$$

$$= \mathrm{e}^\theta G_P(\mathbb{E}\mathrm{e}^{\theta Z}) + \mathrm{e}^\theta \sum_{k=1}^{n} \binom{n}{k} \frac{\mathrm{d}^k}{\mathrm{d}\theta^k} G_P(\mathbb{E}[\mathrm{e}^{\theta Z}])$$

$$= \mathrm{e}^\theta G_P(\mathbb{E}\mathrm{e}^{\theta Z}) + \mathrm{e}^\theta \sum_{k=1}^{n} \binom{n}{k} k! \sum_{i=1}^{k} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$= \mathrm{e}^\theta G_P(\mathbb{E}\mathrm{e}^{\theta Z}) + \mathrm{e}^\theta \sum_{k=1}^{n-1} \binom{n}{k} k! \sum_{i=1}^{k} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$+ n! \mathrm{e}^\theta \sum_{i=1}^{n} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=n} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$= \mathrm{e}^\theta G_P(\mathbb{E}\mathrm{e}^{\theta Z}) + \mathrm{e}^\theta \sum_{k=1}^{n-1} \binom{n}{k} k! \sum_{i=1}^{k} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$\tag{29}$$

$$+ n! \mathrm{e}^\theta \sum_{i=2}^{n} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=n} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$+ \mathrm{e}^\theta G_P'(\mathbb{E}[\mathrm{e}^{\theta Z}]) \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[\mathrm{e}^{\theta Z}].$$

Therefore,

$$(1 - \mathrm{e}^\theta G_P'(\mathbb{E}[\mathrm{e}^{\theta Z}])) \frac{\mathrm{d}^n}{\mathrm{d}\theta^n} \mathbb{E}[\mathrm{e}^{\theta Z}]$$

$$= \mathrm{e}^\theta G_P(\mathbb{E}\mathrm{e}^{\theta Z}) + \mathrm{e}^\theta \sum_{k=1}^{n-1} \binom{n}{k} k! \sum_{i=1}^{k} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}$$

$$+ n! \mathrm{e}^\theta \sum_{i=2}^{n} \frac{G_P^{(i)}(\mathbb{E}[\mathrm{e}^{\theta Z}])}{i!} \sum_{m_1+m_2+\ldots+m_i=n} \frac{\frac{\mathrm{d}^{m_1}}{\mathrm{d}\theta^{m_1}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_1!} \cdots \frac{\frac{\mathrm{d}^{m_i}}{\mathrm{d}\theta^{m_i}} \mathbb{E}[\mathrm{e}^{\theta Z}]}{m_i!}.$$

Letting $\theta \uparrow 0$ in this relation we have

$$(1 - G_P'(1)) \mathbb{E}[Z^n] = 1 + \sum_{k=1}^{n-1} \binom{n}{k} k! \sum_{i=1}^{k} \frac{G_P^{(i)}(1)}{i!} \sum_{m_1+m_2+\ldots+m_i=k} \frac{\mathbb{E}[Z^{m_1}]}{m_1!} \cdots \frac{\mathbb{E}[Z^{m_i}]}{m_i!}$$

$$+ n! \sum_{i=2}^{n} \frac{G_P^{(i)}(1)}{i!} \sum_{m_1+m_2+\ldots+m_i=n} \frac{\mathbb{E}[Z^{m_1}]}{m_1!} \cdots \frac{\mathbb{E}[Z^{m_i}]}{m_i!}.$$

Since

$$G_P^{'}(1) = \mathbb{E}(P)_1 = \mathbb{E}P \quad \text{and} \quad G_P^{(i)}(1) = \mathbb{E}[P(P-1)\dots(P-(i-1))], \quad 2 \le i \le n, \quad (30)$$

we have

$$\mathbb{E}[Z^n] = \frac{1}{1-\mathbb{E}P}\left(1 + \sum_{k=1}^{n-1}\binom{n}{k}k!\sum_{i=1}^{k}\frac{\mathbb{E}(P)_i}{i!}\sum_{m_1+m_2+\dots+m_i=k}\frac{\mathbb{E}[Z^{m_1}]}{m_1!}\cdots\frac{\mathbb{E}[Z^{m_i}]}{m_i!}\right.$$

$$\left. + n!\sum_{i=2}^{n}\frac{\mathbb{E}(P)_i}{i!}\sum_{m_1+m_2+\dots+m_i=n}\frac{\mathbb{E}[Z^{m_1}]}{m_1!}\cdots\frac{\mathbb{E}[Z^{m_i}]}{m_i!}\right), \quad n \in \mathbb{N}. \quad (31)$$

Reasoning by induction on $n \in \mathbb{N}$, by the relation (31) we immediately have that $\mathbb{E}Z^n < \infty$ (note that, for $n=1$, we have $\mathbb{E}Z = 1/(1-\mathbb{E}P) < \infty$, and that, for $n \ge 2$, all the moments of Z involved in the right-hand side of (31) are of order less than or equal to $n-1$). The formula (25) follows by letting $\theta \uparrow 0$ in (29) and using the equalities in (30).

Step 3: Computing $\mathbb{E}Z^2$, $\mathbb{E}Z^3$, *and* $\mathbb{E}Z^4$. The claimed expressions for the second, third, and fourth moments of Z easily follow by (25) (or (31)). For instance, for the second moment, the formula gives

$$\mathbb{E}Z^2 = 1 + 2\mathbb{E}P\mathbb{E}Z + \mathbb{E}P\mathbb{E}Z^2 + \mathbb{E}P(P-1)(\mathbb{E}Z)^2,$$

from which the claimed expression for $\mathbb{E}Z^2$ immediately follows (recalling that $\mathbb{E}Z = (1-\mathbb{E}P)^{-1}$). We omit the computations for the third and fourth moments of Z. $\square$

7.1. Generalized compound Hawkes processes with Poisson offspring distribution

7.1.1. *Gaussian approximation.* In this paragraph we suppose that Z is distributed as the total progeny of a Galton–Watson process with one ancestor and offspring distribution the Poisson law with mean $h \in (0, 1)$, and that $\{X_n\}_{n \ge 1}$ is a Poisson process on $\mathbb{R}^d$ with intensity function $\lambda(x) = \lambda\mathbf{1}_B(x)$, $x \in \mathbb{R}^d$, for some $\lambda > 0$ and some Borel set $B \subseteq \mathbb{R}^d$. We denote by V_{Poisson} the corresponding generalized compound Hawkes process and by W_{Poisson} the functional (8) with V_{Poisson} in place of V.

Corollary 3. *Under the foregoing assumptions and notation, if the Borel sets B and C are such that $0 < \text{Leb}(B \cap C) < +\infty$ and $\mathbb{E}M^2 \in (0, \infty)$, then the bounds (18) and (19) hold with W_{Poisson} in place of W,*

$$\mathbb{E}Z^3 = \frac{1+2h}{(1-h)^5}, \quad (32)$$

and

$$\mathbb{E}Z^4 = \frac{1}{1-h}\left[1 + \frac{4h}{1-h} + \frac{6h^2}{(1-h)^2} + \frac{4h^3+6h}{(1-h)^3} + \frac{h^4+12h^2}{(1-h)^4} + \frac{6h^3}{(1-h)^5} + \frac{11h^2+4h}{(1-h)^6}\right]. \quad (33)$$

Proof. Note that the conditions (23) and (24) are satisfied (the latter holds for any $n \in \mathbb{N}$). Therefore, the expressions for the third and fourth moments of the total progeny are given by Proposition 1. The claim follows by Corollary 1. $\square$

7.1.2. *Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term.* In this paragraph we suppose that, for each $\ell \in \mathbb{N}$, $Z = Z_1^\ell(\mathbb{R}^d, \mathbb{R})$ is distributed as the total progeny of a Galton–Watson process with one ancestor and offspring distribution the Poisson law with mean $h \in (0, 1)$, and that $\{X_n^{(\ell)}\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d$ with intensity function $\lambda_\ell(x) = \lambda_\ell \mathbf{1}_{B_\ell}(x)$, $x \in \mathbb{R}^d$, for positive constants $\lambda_\ell > 0$ and Borel sets $B_\ell \subseteq \mathbb{R}^d$, $\ell \in \mathbb{N}$. We denote by $V_{\text{Poisson}}^{(\ell)}$ the corresponding generalized compound Hawkes process and by $W_{\text{Poisson}}^{(\ell)}$ the functional (10) with $V_{\text{Poisson}}^{(\ell)}$ in place of V_ℓ.

Corollary 4. *Let the foregoing assumptions and notation prevail, and let the Borel sets B_ℓ and C_ℓ, $\ell \in \mathbb{N}$, be such that $0 < \text{Leb}(B_\ell \cap C_\ell) < +\infty$, $\ell \in \mathbb{N}$, $\mathbb{E}M^2 > 0$, and*

$$\frac{\mathbb{E}|M|^m}{(\mathbb{E}M^2)^{m/2}} \leq (m!)^\gamma, \quad \text{for any } m \geq 3 \text{ and some } \gamma \geq 0. \tag{34}$$

Then the following hold:

(i) *If $h - 1 - \log h \geq 1$, then the sequence $\{W_{\text{Poisson}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies $\mathbf{MDP}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, $\mathbf{BCI}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, $\mathbf{NACC}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, where*

$$\Delta_\ell := h\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

(ii) *If $h - 1 - \log h < 1$, then the sequence $\{W_{\text{Poisson}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies $\mathbf{MDP}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, $\mathbf{BCI}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, $\mathbf{NACC}(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, where*

$$\Delta_\ell := h(h - 1 - \log h)^3\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

Example 1. Note that the condition (34) is trivially satisfied with $\gamma = 0$ if M is a constant different from zero. Similarly, if M is a uniform variable on $[0, D]$ with $D > 0$, the moments of M satisfy

$$\frac{\mathbb{E}|M|^m}{(\mathbb{E}M^2)^{m/2}} = \frac{3^{\frac{m}{2}}}{m+1} < m!.$$

Assumption (34) holds with $\gamma = 1$ even if M is exponentially distributed with mean μ^{-1}, for some $\mu > 0$. Indeed, in such a case we have

$$\frac{\mathbb{E}|M|^m}{(\mathbb{E}M^2)^{m/2}} = \frac{m!}{2^{m/2}} < m!.$$

Another example is M Gaussian distributed with mean zero and variance σ^2. Indeed, in such a case we have

$$\frac{\mathbb{E}|M|^m}{(\mathbb{E}M^2)^{m/2}} \leq (m - 1)!!.$$

By distinguishing the two cases $m = 2p + 1$ and $m = 2p + 2$ for $p = 2, 3, \cdots$, we conclude that

$$\frac{\mathbb{E}|M|^m}{(\mathbb{E}M^2)^{m/2}} \leq \sqrt{m!},$$

which implies that the condition (34) is satisfied with $\gamma = \frac{1}{2}$.

The proof of the corollary exploits the following lemmas, which are proved in Appendix A.

Lemma 3. *For any $v > 0$ and any integer $m \geq 2$ we have*

$$\sum_{k=1}^{+\infty} e^{-vk} k^{m-1} = v^{-m}(m-1)! + R_m,$$

with

$$|R_m| \leq \frac{1}{\pi(m-1)} + \frac{2(m-1)!}{\pi^m}.$$

Lemma 4. *The function*

$$f(x) := x(x-1-\log x)^2, \quad x \in (0, 1),$$

is such that $f(x) \in (0, 1)$.

Proof of Corollary 4. It is well-known that the total progeny Z of a sub-critical Galton–Watson process with one ancestor and Poisson offspring law with mean $h \in (0, 1)$ follows the Borel distribution (cf. [37] and the references therein), i.e.,

$$\mathbb{P}(Z = k) = \frac{e^{-hk}(hk)^{k-1}}{k!}, \quad k = 1, 2, \ldots.$$

Therefore, by Stirling's inequality, for $m \geq 3$ we have

$$\mathbb{E}Z^m = \sum_{k=1}^{+\infty} \frac{e^{-hk}(hk)^{k-1}}{k!} k^m \leq \sum_{k=1}^{+\infty} e^{-hk}(hk)^{k-1} k^m \left(\frac{e}{k}\right)^k \frac{1}{\sqrt{2\pi k}}$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{+\infty} e^{(1-h)k} \frac{k^m}{k\sqrt{k}} h^{k-1}$$

$$= \frac{1}{h\sqrt{2\pi}} \sum_{k=1}^{+\infty} e^{-(h-1-\log h)k} k^{m-1.5}$$

$$\leq \frac{1}{h\sqrt{2\pi}} \sum_{k=1}^{+\infty} e^{-(h-1-\log h)k} k^{m-1}.$$

Using Lemma 3 with $v := h - 1 - \log h > 0$, we have

$$\mathbb{E}Z^m \leq \frac{(m-1)!}{h\sqrt{2\pi}} \left(v^{-m} + \frac{2}{\pi^m} + \frac{1}{\pi(m-1)(m-1)!}\right), \quad m \geq 3. \tag{35}$$

We now give separately the proofs of Parts (i) and (ii); in both cases we shall apply Corollary 2.

Proof of Part (i).
If $v \geq 1$, then, for any $m \geq 3$,

$$v^{-m} + \frac{2}{\pi^m} + \frac{1}{\pi(m-1)(m-1)!} \leq 3 \leq m. \tag{36}$$

Combining this inequality with (35), for $m \geq 3$ and $\ell \in \mathbb{N}$, we have

$$
\begin{aligned}
\frac{\mathbb{E} Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} &\leq \frac{1}{h\sqrt{2\pi}} \frac{m!}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \\
&= \frac{h^{m-3}}{\sqrt{2\pi}} \frac{m!}{(h\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} \\
&\leq \frac{m!}{(h\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}}
\end{aligned}
$$

since $h < 1$ and $\sqrt{2\pi} > 1$. Combining this latter inequality with the assumption (34) we have that the condition (22) is satisfied with $\Delta_\ell = h\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}$, and the claim follows by Corollary 2.

Proof of Part (ii).
Now we suppose $\nu < 1$. Since

$$
\frac{2}{\pi^m} + \frac{1}{\pi(m-1)(m-1)!} \leq \frac{2}{\pi}\left(\frac{1}{\pi^2} + \frac{1}{8}\right) < 0.16, \quad \text{for any } m \geq 3, \tag{37}
$$

by (35) we have

$$
\begin{aligned}
\frac{\mathbb{E} Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} &\leq \frac{1}{h\sqrt{2\pi}} \frac{(m-1)!}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \left(\frac{1}{\nu^m} + 0.16\right) \\
&= \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}\nu)^{m-2}} \frac{\nu^m}{h\nu^2\sqrt{2\pi}}\left(\frac{1}{\nu^m} + 0.16\right), \quad m \geq 3, \ell \in \mathbb{N}.
\end{aligned}
$$

Since $\nu < 1$, we have

$$
\frac{\nu^m}{h\nu^2\sqrt{2\pi}}\left(\frac{1}{\nu^m} + 0.16\right) \leq \frac{1.16}{h\nu^2\sqrt{2\pi}} < \frac{1}{h\nu^2} := u_h.
$$

Therefore

$$
\begin{aligned}
\frac{\mathbb{E} Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} &\leq \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}\nu)^{m-2}} u_h \\
&= \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}\nu u_h^{-1})^{m-2}} u_h^{3-m}, \quad m \geq 3, \ell \in \mathbb{N}.
\end{aligned}
$$

By Lemma 4 we have $u_h^{-1} < 1$. Therefore, for any $m \geq 3$ and $\ell \in \mathbb{N}$, we have

$$
\frac{\mathbb{E} Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{m!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}\nu u_h^{-1})^{m-2}}.
$$

Combining this latter inequality with the assumption (34) we have that the condition (22) is satisfied with $\Delta_\ell := h\nu^3\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}$, and the claim follows by Corollary 2. $\qquad \square$

7.2. On the Gaussian approximation bound in the Kolmogorov distance and the normal approximation with Cramér correction term

The aim of this section is to illustrate, by means of a simple example, the differences and the analogies between Gaussian approximation bounds in the Kolmogorov distance and normal approximations with Cramér correction term.

Let $W_{\text{Poisson}}^{(\ell)}$, $\ell \in \mathbb{N}$, be defined as at the beginning of Section 7.1.2, with $\{(B_\ell, C_\ell)\}_{\ell \in \mathbb{N}} \subset \mathcal{B}(\mathbb{R}^d)^2$ a sequence of Borel sets such that $0 < \text{Leb}(B_\ell \cap C_\ell) < \infty$, $\ell \in \mathbb{N}$, and $\lambda_\ell \text{Leb}(B_\ell \cap C_\ell) \to +\infty$ as $\ell \to +\infty$. Assume $M \equiv 1$ (so that (34) holds with $\gamma = 0$), and let Δ_ℓ be defined as in Part (i) or (ii) of Corollary 4. We know that $\{W_{\text{Poisson}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies **NACC**$(0, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$.

For a fixed $0 < r < 1/3$, let ℓ^* be sufficiently large so that

$$0 < \Delta_\ell^r < c_0 \Delta_\ell, \quad \text{for all } \ell \geq \ell^*,$$

where c_0 is the positive constant which appears in the definition of **NACC**$(0, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$. Setting $x_\ell := \Delta_\ell^r$, for all $\ell \geq \ell^*$, we have

$$\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x_\ell) - \mathbb{P}(G \geq x_\ell) = e^{L_{\ell,x_\ell}^+}\mathbb{P}(G \geq x_\ell)\left(1 + c_1\theta_{\ell,x_\ell}^+ \frac{1+x_\ell}{\Delta_\ell}\right) - \mathbb{P}(G \geq x_\ell)$$

$$= (e^{L_{\ell,x_\ell}^+} - 1)\mathbb{P}(G \geq x_\ell) + c_1\theta_{\ell,x_\ell}^+ e^{L_{\ell,x_\ell}^+}\mathbb{P}(G \geq x_\ell)\frac{1+x_\ell}{\Delta_\ell}.$$

Since $|L_{\ell,x_\ell}^+| \leq c_2\Delta_\ell^{3r-1}$ and $|\theta_{\ell,x_\ell}^+| \leq 1$, we have

$$|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x_\ell) - \mathbb{P}(G \geq x_\ell)| \leq \left(c_2 \left|\frac{e^{L_{\ell,x_\ell}^+} - 1}{L_{\ell,x_\ell}^+}\right| \Delta_\ell^{3r-1} + c_1 O(1)\left|\frac{1+x_\ell}{\Delta_\ell}\right|\right)\mathbb{P}(G \geq x_\ell)$$

$$= \left(c_2 O(1)\Delta_\ell^{3r-1} + c_1 O(1)\left|\frac{1+\Delta_\ell^r}{\Delta_\ell}\right|\right)\mathbb{P}(G \geq x_\ell)$$

$$= O(\Delta_\ell^{3r-1})\mathbb{P}(G \geq x_\ell).$$

Bounding the Gaussian tail from above, we obtain

$$|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x_\ell) - \mathbb{P}(G \geq x_\ell)| \leq O(\Delta_\ell^{3r-1})\frac{e^{-x_\ell^2/2}}{x_\ell\sqrt{2\pi}}$$

$$= O(\Delta_\ell^{2r-1}e^{-\Delta_\ell^{2r}/2})$$

$$= O\left(\frac{\exp\left(-\frac{u_h^{2r}(\lambda_\ell \text{Leb}(B_\ell \cap C_\ell))^r}{2}\right)}{(\lambda_\ell \text{Leb}(B_\ell \cap C_\ell))^{\frac{1}{2}-r}}\right), \quad \text{as } \ell \to +\infty,$$

$$\tag{38}$$

where either $u_h := h$ or $u_h := h(h - 1 - \log h)^3$. On the other hand, the bound (19) and the relations (32) and (33) yield

$$|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x_\ell) - \mathbb{P}(G \geq x_\ell)| = O\left(\frac{1}{(\lambda_\ell \text{Leb}(B_\ell \cap C_\ell))^{1/2}}\right), \quad \text{as } \ell \to +\infty. \tag{39}$$

Clearly the rate (38) is much faster than that of (39). Now let $x \in (0, \infty)$ be arbitrarily fixed. For ℓ large enough we have $x \in [0, c_0 \Delta_\ell]$, and so by Corollary 4, for all ℓ large enough, we have

$$|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x) - \mathbb{P}(G \geq x)| \leq |e^{L_{\ell,x}^+} - 1|\mathbb{P}(G \geq x) + c_1 e^{L_{\ell,x}^+}\mathbb{P}(G \geq x)\theta_{\ell,x}^+ \frac{1+x}{\Delta_\ell}$$

$$\leq c_2 \left|\frac{e^{L_{\ell,x}^+} - 1}{L_{\ell,x}^+}\right| \frac{x^3}{\Delta_\ell}\mathbb{P}(G \geq x) + c_1 e^{L_{\ell,x}^+}\mathbb{P}(G \geq x)\theta_{\ell,x}^+ \frac{1+x}{\Delta_\ell}$$

$$= O\left(\frac{1}{\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}}\right).$$

Clearly, the same rate is provided by the bound (19) and the relations (32) and (33). We emphasize that (i) the inequality (19) and the relations (32) and (33) do indeed yield an explicit bound on the quantity $|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x) - \mathbb{P}(G \geq x)|$, for any $x \in \mathbb{R}$ and any $\ell \in \mathbb{N}$; (ii) an explicit bound on the quantity $|\mathbb{P}(W_{\text{Poisson}}^{(\ell)}(C_\ell) \geq x) - \mathbb{P}(G \geq x)|$, for any $x \in \mathbb{R}$ and any $\ell \in \mathbb{N}$, is not amenable via the normal approximation with Cramér correction term (for various obvious reasons).

7.3. Comparison with some related literature

7.3.1. *Gaussian approximation.* Let N be a classical Hawkes process on $(0, \infty)$ with parameters (λ, g). Then Corollary 3 with $B = (0, \infty)$ and $C = (0, t]$, $t > 0$, gives *explicit bounds* for the Gaussian approximation of

$$\frac{N((0, t]) - \mathbb{E}N((0, t])}{\sqrt{\mathbb{V}\text{ar}(N((0, t]))}},$$

both in the Wasserstein and in the Kolmogorov distance. Note that for the fertility function g we assume only the standard stability condition $h := \int_0^\infty g(t) \, dt \in (0, 1)$.

It is worthwhile to compare these bounds with the ones in [23, 26]. Theorem 3.13 in [23] gives a bound of the kind

$$d_W\left(\frac{N((0, t]) - \mathbb{E}N((0, t])}{\sqrt{t}}, G\right) \leq c/\sqrt{t}, \quad t > 0,$$

for some constant $c > 0$ which is not explicitly computed and for specific choices of g (exponential and Erlang). This result has been extended in [26] to fertility functions $g : [0, \infty) \to [0, \infty)$ such that $h \in (0, 1)$ and $\int_0^\infty tg(t)dt < \infty$. The techniques used in [23, 26] are based on the Poisson embedding construction of Hawkes processes and the Malliavin calculus on the Poisson space. These ideas were previously used in [44, 45] for the purpose of Gaussian and Poisson approximation of some classes of nonlinear Hawkes processes. Note that, in contrast with classical Hawkes processes, nonlinear Hawkes processes (introduced in [8]) do not have a Poisson cluster representation. For Poisson cluster processes (such as classical Hawkes processes), the number of points on some measurable set can be represented as an integral with respect to a suitable Poisson random measure. As a consequence, results on the Gaussian approximation of the number of points on a measurable set can be obtained by applying the general results in [30].

7.3.2. *Moderate deviations.* In this section we compare Corollary 4 with a couple of related results in the literature. Firstly, we prove that Corollary 4, when specialized to a classical Hawkes process on $(0, \infty)$, implies the same moderate deviation principle provided in [47] (see Theorem 1 therein), with an alternative assumption on the fertility function of the process. We refer the reader to [19] for sample-path moderate deviation principles, on the space of càdlàg functions on $[0, 1]$ equipped with the Skorokhod topology, for Poisson cluster point processes on the line; see [19, Theorem 2]. Secondly, we compare Corollary 4 (again specialized to a classical Hawkes process on $(0, \infty)$) with Theorem 8 in [20]. Hereafter, N denotes a classical Hawkes process on $(0, \infty)$ with parameters (λ, g), and we assume that g satisfies the usual stability condition $h := \int_0^\infty g(t)\, \mathrm{d}t \in (0, 1)$.

Corollary 4 with $\lambda_\ell = \lambda > 0$, $B_\ell = (0, \infty)$, and $C_\ell = (0, \ell]$, $\ell \in \mathbb{N}$, and $M \equiv 1$ yields that, for any sequence of positive numbers $\{a_\ell\}_{\ell \in \mathbb{N}}$ such that $\lim_{\ell \to \infty} a_\ell = +\infty$ and $\lim_{\ell \to \infty} \frac{a_\ell}{\sqrt{\ell}} = 0$, for any Borel set $B \subseteq \mathbb{R}$,

$$- \inf_{x \in \overset{\circ}{B}} \frac{x^2}{2} \le \liminf_{\ell \to \infty} a_\ell^{-2} \log \mathbb{P}\left(\frac{N((0, \ell]) - \mathbb{E}N((0, \ell])}{a_\ell \sqrt{\mathbb{V}\mathrm{ar}(N((0, \ell]))}} \in B \right)$$

$$\le \limsup_{\ell \to \infty} a_\ell^{-2} \log \mathbb{P}\left(\frac{N((0, \ell]) - \mathbb{E}N((0, \ell])}{a_\ell \sqrt{\mathbb{V}\mathrm{ar}(N((0, \ell]))}} \in B \right) \le - \inf_{x \in \overline{B}} \frac{x^2}{2}.$$

Reasoning by contradiction, one has that, for any function $a(\,\cdot\,)$ such that $\lim_{t \to \infty} a(t) = +\infty$ and $a(t) = o(\sqrt{t})$, as $t \to +\infty$, and any Borel set $B \subseteq \mathbb{R}$,

$$- \inf_{x \in \overset{\circ}{B}} \frac{x^2}{2} \le \liminf_{t \to \infty} a(t)^{-2} \log \mathbb{P}\left(\frac{N((0, t]) - \mathbb{E}N((0, t])}{a(t) \sqrt{\mathbb{V}\mathrm{ar}(N((0, t]))}} \in B \right)$$

$$\le \limsup_{\ell \to \infty} a(t)^{-2} \log \mathbb{P}\left(\frac{N((0, t]) - \mathbb{E}N((0, t])}{a(t) \sqrt{\mathbb{V}\mathrm{ar}(N((0, t]))}} \in B \right) \le - \inf_{x \in \overline{B}} \frac{x^2}{2}.$$

It is easily realized (cf. [4] for example) that, under the stability condition,

$$\mathbb{E}N((0, t])/t \to \frac{\lambda}{1 - h} \quad \text{and} \quad \mathbb{V}\mathrm{ar}(N((0, t]))/t \to \frac{\lambda}{(1 - h)^3}, \quad \text{as } t \to +\infty.$$

So, letting $b(\,\cdot\,)$ denote a function such that $\sqrt{t} = o(b(t))$ and $b(t) = o(t)$, as $t \to \infty$, and setting $a(t) := b(t)/\sqrt{\mathbb{V}\mathrm{ar}(N((0, t]))}$, we have that, for any Borel set $B \subseteq \mathbb{R}$,

$$- \inf_{x \in \overset{\circ}{B}} \frac{x^2 (1 - h)^3}{2\lambda} \le \liminf_{t \to \infty} \frac{t}{b(t)^2} \log \mathbb{P}\left(\frac{N((0, t]) - \mathbb{E}N((0, t])}{b(t)} \in B \right)$$

$$\le \limsup_{\ell \to \infty} \frac{t}{b(t)^2} \log \mathbb{P}\left(\frac{N((0, t]) - \mathbb{E}N((0, t])}{b(t)} \in B \right) \le - \inf_{x \in \overline{B}} \frac{x^2 (1 - h)^3}{2\lambda}.$$

$$(40)$$

By Lemma 5 in [4], we have that, if in addition to the stability condition $h \in (0, 1)$ we assume

$$\int_0^\infty \sqrt{t} g(t)\, \mathrm{d}t < \infty, \qquad (41)$$

then

$$\frac{\mathbb{E}N((0, t]) - \lambda t/(1 - h)}{b(t)} \to 0, \quad \text{as } t \to \infty.$$

So, for an arbitrarily fixed $\delta > 0$, there exists t_δ such that for any $t > t_\delta$ it holds that

$$\left| \frac{N((0, t]) - \mathbb{E}N((0, t])}{b(t)} - \frac{N((0, t]) - \frac{\lambda t}{1-h}}{b(t)} \right| = \left| \frac{\mathbb{E}N((0, t]) - \lambda t/(1 - h)}{b(t)} \right| < \delta.$$

Therefore, for an arbitrarily fixed $\delta > 0$, there exists t_δ such that for any $t > t_\delta$ we have

$$\log \mathbb{P}\left(\left| \frac{N((0, t]) - \mathbb{E}N((0, t])}{b(t)} - \frac{N((0, t]) - \frac{\lambda t}{1-h}}{b(t)} \right| > \delta \right) = -\infty.$$

Hence the processes $\{\frac{N((0,t]) - \mathbb{E}N((0,t])}{b(t)}\}_{t>0}$ and $\{\frac{N((0,t]) - \lambda t/(1-h)}{b(t)}\}_{t>0}$ are exponentially equivalent (see [12, Definition 4.2.10, p. 130]). Therefore, by [12, Theorem 4.2.13, p. 130], the relation (40) holds with $\mathbb{E}N((0, t])$ replaced by $\lambda t/(1 - h)$. Thus we recover the moderate deviation principle proved in [47] under an alternative condition on g (the latter paper assumes the stability condition and $\sup_{t>0} t^{3/2} g(t) < \infty$, which is clearly different from (41)).

Theorem 8 in [20] states that, under the assumption

$$\int_0^\infty t g(t) \mathrm{d}t < +\infty$$

(which is clearly stronger than (41)), for any $y(t) = o(t^{1/2 - 1/m})$ as $t \to +\infty$, any integer $m \geq 3$, and any positive function $b(\cdot)$, it holds that

$$\mathbb{P}\left(\frac{N((0, t]) - \frac{\lambda t}{1-h}}{b(t)} \geq \frac{\sqrt{t} y(t)}{b(t)} \frac{\sqrt{\lambda}}{(1 - h)^{3/2}} \right) = \frac{(1 + o(1))}{y(t)\sqrt{2\pi}} \mathrm{e}^{-\sum_{i=2}^{m-1} c_i \frac{y(t)^i}{t^{(i-2)/2}}}, \quad \text{as } t \to +\infty,$$

where $\{c_i\}_{i=1,\cdots,m-2}$ are real coefficients that can be computed explicitly; for instance, one has $c_2 = \frac{1}{2}$. In particular, if $b(t) = o(t^{2/3})$, as $t \to +\infty$, then by choosing

$$y(t) = K\frac{(1 - h)^{3/2}}{\sqrt{\lambda}} \frac{b(t)}{\sqrt{t}} = o(t^{1/2 - 1/3}), \quad \text{as } t \to +\infty, \text{ for some } K > 0,$$

we have

$$\mathbb{P}\left(\frac{N((0, t]) - \frac{\lambda t}{1-h}}{b(t)} \geq K \right) = \frac{\sqrt{\lambda}(1 + o(1))}{(1 - h)^{3/2} K} \frac{\sqrt{t}}{b(t)} \mathrm{e}^{-\frac{K^2}{2} \frac{(1-h)^3}{\lambda} \frac{b(t)^2}{t}}, \tag{42}$$

which is a more precise form of the relation (40) for the Borel set $B = [K, +\infty)$. Note that, unlike the formula (40), which is valid for any Borel set B, the formula (42) gives asymptotic estimates only for half-lines.

7.3.3. *Bernstein-type concentration inequalities.* In this section we present some consequences of Corollary 4 concerning stationary compound Hawkes processes on the line, i.e., $B_T := \mathbb{R}$, observed on the time interval $C_T := (0, T]$; here T replaces ℓ to emphasize the dependence on time.

If we interpret the mark M as the claim that an insurer must pay to an insurance policy holder, then the variable $V((0, T])$ (defined by (3)) represents the total loss incurred by the insurer in the time interval $(0, T]$.

Assume that the claims arrivals are modeled by the points of a Hawkes process of baseline intensity $\lambda > 0$ and Poisson offspring distribution with mean $h \in (0, 1)$ satisfying $h - 1 - \log h \geq 1$. Assume moreover that the mark M follows the exponential distribution of parameter $\mu^{-1} \in (0, +\infty)$. Then Corollary 4 yields

$$\mathbb{P}\left(\left|\frac{V((0, T]) - \mathbb{E}V((0, T])}{\sqrt{\mathbb{V}\mathrm{ar}V((0, T])}}\right| \geq x\right) \leq 2 \exp\left(-\frac{1}{4}\min\left\{\frac{x^2}{2^{1+\gamma}}, (x\Delta_T)^{\frac{1}{1+\gamma}}\right\}\right), \quad x \geq 0,$$

where $\Delta_T = h\sqrt{\lambda T}$ and $\gamma = 1$ by virtue of Example 1. By stationarity, this inequality can be rewritten as

$$\mathbb{P}\left(V((0, T]) \in \left[\frac{\lambda\mu}{1-h}T - x\sqrt{\frac{2\mu^2\lambda}{(1-h)^3}T}, \frac{\lambda\mu}{1-h}T + x\sqrt{\frac{2\mu^2\lambda}{(1-h)^3}T}\right]\right)$$

$$\geq 1 - 2\exp\left(-\frac{1}{4}\min\left\{\frac{x^2}{4}, (xh\sqrt{\lambda T})^{1/2}\right\}\right), \quad x \geq 0,$$

which yields a non-asymptotic lower bound on the probability that the total loss is within x times its standard deviation.

Another quantity of interest for insurers is the probability that the total loss greatly exceeds its expected value. The Bernstein-type concentration inequality, being valid for any $x \geq 0$, yields an upper bound on this probability. Indeed, by choosing

$$x = x_T = (k-1)\sqrt{\frac{\lambda(1-h)T}{2}}, \quad \text{for some } k > 1,$$

we have

$$\mathbb{P}\left(V((0, T]) \geq k\frac{\lambda\mu}{1-h}T\right) = \mathbb{P}\left(\frac{V((0, T]) - \mathbb{E}V((0, T])}{\sqrt{\mathbb{V}\mathrm{ar}V((0, T])}} \geq x_T\right)$$

$$\leq \mathbb{P}\left(\left|\frac{V((0, T]) - \mathbb{E}V((0, T])}{\sqrt{\mathbb{V}\mathrm{ar}V((0, T])}}\right| \geq x_T\right)$$

$$\leq 2\exp\left(-\frac{1}{4}\min\left\{\frac{(k-1)^2\lambda(1-h)T}{8}, \left((k-1)\lambda h\sqrt{\frac{1-h}{2}}T\right)^{1/2}\right\}\right).$$
$$(43)$$

If the time horizon T satisfies $T \geq \frac{2^{11/2}h}{(k-1)^3\lambda(1-h)^{3/2}}$, then the inequality (43) simplifies to

$$\mathbb{P}\left(V((0, T]) \geq k\frac{\lambda\mu}{1-h}T\right) \leq 2\exp\left(-\frac{1}{4}\left((k-1)\lambda h\sqrt{\frac{1-h}{2}}T\right)^{1/2}\right). \quad (44)$$

A similar (non-asymptotic) inequality appears in Proposition 2.1 of [39], albeit working only for stationary Hawkes processes on the line whose fertility functions have compact support, and involving quantities that are not explicitly known. We also point out that, by specializing the inequality (44) to the simple Hawkes process, that is, with constant marks ($\gamma = 0$ by virtue of Example 1), we can find a decay rate for the tail probability similar to the one given in [39], but with explicit constants.

7.4. Generalized compound Hawkes processes with binomial offspring distribution

7.4.1. *Gaussian approximation.* In this paragraph we suppose that Z is distributed as the total progeny of a Galton–Watson process with one ancestor and offspring distribution the binomial law with parameters (h, p), with $h \in \mathbb{N}$ and $p \in (0, 1)$ such that $hp \in (0, 1)$. We assume that $\{X_n\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d$ with intensity function $\lambda(x) = \lambda \mathbf{1}_B(x)$, $x \in \mathbb{R}^d$, for some positive constant $\lambda > 0$ and some Borel set $B \subseteq \mathbb{R}^d$. We denote by V_{binomial} the corresponding generalized compound Hawkes process and by W_{binomial} the functional (8) with V_{binomial} in place of V.

Corollary 5. *Under the foregoing assumptions and notation, if the Borel sets B and C are such that $0 < \text{Leb}(B \cap C) < +\infty$ and $\mathbb{E}M^2 \in (0, \infty)$, then the bounds (18) and (19) hold with W_{binomial} in place of W,*

$$\mathbb{E}Z^3 = \frac{1}{1 - hp}\left(1 + \frac{3hp}{1 - hp} + 3\frac{p^2(h)_2}{(1 - hp)^2} + \frac{p^3(h)_3 + 3hp(1 - p)}{(1 - hp)^3} + 3\frac{h^2p^2(1 - p)^2}{(1 - hp)^4}\right),$$

and

$$\mathbb{E}Z^4 = \frac{1}{1 - hp}\left[1 + \frac{4hp}{1 - hp} + \frac{6p^2(h)_2}{(1 - hp)^2} + 4\frac{p^3(h)_3}{(1 - hp)^3} + \frac{p^4(h)_4}{(1 - hp)^4}\right.$$
$$\left. + \frac{3(1 - hp^2)}{(1 - hp)^3}\left(2hp + \frac{4p^2(h)_2}{1 - hp} + p^2(h)_2\frac{1 - hp^2}{(1 - hp)^3} + \frac{2p^3(h)_3}{(1 - hp)^2}\right) + 4\mathbb{E}Z^3\frac{hp(1 - p)}{1 - hp}\right].$$

Here $(h)_n := h(h - 1) \ldots (h - (n - 1))\mathbf{1}_{\{h \geq n\}}$.

Proof. Similar to the proof of Corollary 3. $\square$

7.4.2. *Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term.* In this paragraph we suppose that, for each $\ell \in \mathbb{N}$, $Z = Z_1^\ell(\mathbb{R}^d, \mathbb{R})$ is distributed as the total progeny of a Galton–Watson process with one ancestor and offspring distribution the binomial law with parameters (h, p), with $h \in \mathbb{N}$ and $p \in (0, 1)$ such that $hp \in (0, 1)$. We assume that $\{X_n^{(\ell)}\}_{n \geq 1}$ is a Poisson process on $\mathbb{R}^d$ with intensity function $\lambda_\ell(x) = \lambda_\ell \mathbf{1}_{B_\ell}(x)$, $x \in \mathbb{R}^d$, for positive constants $\lambda_\ell > 0$ and Borel sets $B_\ell \subseteq \mathbb{R}^d$, $\ell \in \mathbb{N}$. We denote by $V_{\text{binomial}}^{(\ell)}$ the corresponding generalized compound Hawkes process and by $W_{\text{binomial}}^{(\ell)}$ the functional (10) with $V_{\text{binomial}}^{(\ell)}$ in place of V_ℓ.

Corollary 6. *Let the foregoing assumptions and notation prevail, and let the Borel sets B_ℓ and C_ℓ, $\ell \in \mathbb{N}$, be such that $0 < \text{Leb}(B_\ell \cap C_\ell) < +\infty$, $\ell \in \mathbb{N}$, $\mathbb{E}M^2 > 0$. Assume (34). Then the following hold:*

(i.1) *If $h = 1$ and $p \leq e^{-1}$, then the sequence $\{W_{\text{binomial}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *where*

$$\Delta_\ell := \frac{p}{1.05(1 - p)}\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

(i.2) *If $h = 1$ and $p > e^{-1}$, then the sequence $\{W_{\text{binomial}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *where*

$$\Delta_\ell := \frac{p(\log p)^4}{1.05(1-p)}\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

(ii.1) *If $h \geq 2$ and $ph(h(1-p)/(h-1))^{h-1} \leq e^{-1}$, then the sequence $\{W_{\text{binomial}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *where*

$$\Delta_\ell := \left(1 + \sqrt{1 + \frac{1}{h-1}}\,\frac{e^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}}\right)^{-1}\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

(ii.2) *If $h \geq 2$ and $ph(h(1-p)/(h-1))^{h-1} > e^{-1}$, then the sequence $\{W_{\text{binomial}}^{(\ell)}(C_\ell)\}_{\ell \geq 1}$ satisfies* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *where*

$$\Delta_\ell := -\frac{\left(1 + \sqrt{1 + \frac{1}{h-1}}\,\frac{e^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}}\right)^{-1}\left[\log\left(ph\left(\frac{h(1-p)}{h-1}\right)^{h-1}\right)\right]^3}{1.16}\sqrt{\lambda_\ell \text{Leb}(B_\ell \cap C_\ell)}.$$

Proof. It is well known that the total progeny Z of a sub-critical Galton–Watson process with one ancestor and binomial offspring law with parameters (h, p) follows the Consul distribution, i.e.,

$$\mathbb{P}(Z = k) = \frac{1}{k}\binom{kh}{k-1}p^{k-1}(1-p)^{k(h-1)+1}, \quad k = 1, 2, \ldots. \tag{45}$$

By Stirling's upper and lower bounds on the factorial, for $k \geq 1$ and $h \geq 2$, we have

$$\binom{kh}{k-1}$$

$$= \frac{(kh)!}{(k-1)!(k(h-1)+1)!}$$

$$\leq \frac{1}{(k-1)!} \times \frac{\sqrt{2\pi kh}\left(\frac{kh}{e}\right)^{kh}e^{\frac{1}{12kh}}}{\sqrt{2\pi(k(h-1)+1)}\left(\frac{k(h-1)+1}{e}\right)^{k(h-1)+1}e^{\frac{1}{12[k(h-1)+1]+1}}}$$

$$= \frac{e^{-k+1}}{(k-1)!}\sqrt{\frac{kh}{k(h-1)+1}}\frac{(kh)^{kh}}{(k(h-1)+1)^{k(h-1)+1}}e^{\frac{1}{12kh}-\frac{1}{12[k(h-1)+1]+1}}$$

$$= \frac{e^{-k+1}}{(k-1)!}\sqrt{\frac{kh}{k(h-1)+1}}\frac{(k(h-1))^{kh}}{(k(h-1)+1)^{k(h-1)+1}}\left(\frac{h}{h-1}\right)^{kh}e^{\frac{1}{12kh}-\frac{1}{12[k(h-1)+1]+1}}$$

$$\leq \frac{e^{-k+1}}{(k-1)!}\sqrt{1 + \frac{1}{h-1}}\frac{(k(h-1))^{kh}}{(k(h-1)+1)^{k(h-1)+1}}\left(\frac{h}{h-1}\right)^{kh}e^{\frac{1}{12kh}-\frac{1}{12[k(h-1)+1]+1}} \tag{46}$$

$$= \frac{e^{-k+1}}{(k-1)!}\sqrt{1 + \frac{1}{h-1}}\left(\frac{k(h-1)}{k(h-1)+1}\right)^{k(h-1)+1}[k(h-1)]^{k-1}\left(\frac{h}{h-1}\right)^{kh}e^{\frac{1}{12kh}-\frac{1}{12[k(h-1)+1]+1}}$$

$$\leq \frac{e^{\frac{1}{24\cdot25}}}{h-1}\frac{e^{-k+1}k^{k-1}}{(k-1)!}\sqrt{1 + \frac{1}{h-1}}e^{-1}\left(\frac{h^h}{(h-1)^{h-1}}\right)^k,$$

where the inequality (46) follows if we notice that $kh/[k(h-1)+1] \leq h/(h-1)$, and the last inequality follows if we notice that

$$\mathrm{e}^{\frac{1}{12kh} - \frac{1}{12[k(h-1)+1]+1}} \leq \mathrm{e}^{\frac{1}{24\cdot25}}, \quad \text{for each } k \geq 1 \text{ and } h \geq 2,$$

and that $\left(1 \pm \frac{1}{n}\right)^n \leq \mathrm{e}^{\pm1}$, $n \geq 1$. By this latter inequality (with the sign $+$) and the Stirling lower bound on the factorial, we have

$$\frac{\mathrm{e}^{-k+1}}{(k-1)!} k^{k-1} \leq \frac{\mathrm{e}}{\sqrt{2\pi(k-1)}}, \quad k \geq 2.$$

Therefore, for $k \geq 2$ and $h \geq 2$, we have

$$\binom{kh}{k-1} \leq \frac{\mathrm{e}^{\frac{1}{24\cdot25}}}{h-1} \frac{1}{\sqrt{2\pi(k-1)}} \sqrt{1 + \frac{1}{h-1}} \left(\frac{h^h}{(h-1)^{h-1}}\right)^k =: C_{h,k}. \tag{47}$$

We now distinguish between two cases: $h = 1$ and $h \geq 2$.

Case h=1.

Since $\binom{kh}{k-1} = k$, by (45), for any $m \in \mathbb{N}$ we have

$$\mathbb{E}Z^m = \frac{1-p}{p} \sum_{k\geq1} k^m p^k = \frac{1-p}{p} \sum_{k\geq1} k^m \mathrm{e}^{-\nu_1 k}, \quad \text{where } \nu_1 := -\log p > 0.$$

Using Lemma 3 we have

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{(1-p)m!}{p\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \left(\frac{1}{\nu_1^{m+1}} + \frac{1}{\pi m m!} + \frac{2}{\pi^{m+1}}\right), \quad m, \ell \in \mathbb{N}. \tag{48}$$

Proof of Part (i.1).

If $p \leq \mathrm{e}^{-1}$, then $\nu_1 \geq 1$; therefore, setting $u_p := 1.05\frac{1-p}{p}$, by (48), for any $m \geq 3$ and $\ell \in \mathbb{N}$, we have

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{m!}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} 1.05\frac{1-p}{p}$$

$$= \frac{m!}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} u_p$$

$$= \frac{m!}{((u_p^{-1})\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} (u_p^{-1})^{m-3}$$

$$\leq \frac{m!}{((u_p^{-1})\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}}, \tag{49}$$

where the latter inequality follows if we notice that $u_p^{-1} \leq 1$ (indeed $\frac{1}{p} - 1 \geq \log\left(\frac{1}{p}\right) \geq 1$), and so $u_p \geq 1$. Combining (49) with (34), we have that the condition (22) is satisfied with $\Delta_\ell := \frac{p}{1.05(1-p)}\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}$, and the claim follows by Corollary 2.

Proof of Part (i.2).
If $p > \mathrm{e}^{-1}$, then $\nu_1 < 1$; therefore, setting $u_p := \frac{1-p}{p\nu_1^3}$, by (48), for any $m \geq 3$ and $\ell \in \mathbb{N}$,

$$
\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{(1-p)m!}{p(\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} \nu_1^{m-2} \left(\frac{1}{\nu_1^{m+1}} + \frac{1}{\pi mm!} + \frac{2}{\pi^{m+1}} \right)
$$

$$
= \frac{1-p}{p\nu_1^3} \frac{m!}{(\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} \left(1 + \frac{\nu_1^{m+1}}{\pi mm!} + 2\left(\frac{\nu_1}{\pi}\right)^{m+1} \right)
$$

$$
\leq \frac{1-p}{p\nu_1^3} \frac{m!}{(\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} 1.05
$$

$$
= u_p \frac{m!}{(\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}} 1.05
$$

$$
= [(1.05u_p)^{-1}]^{m-3} \frac{m!}{((1.05u_p)^{-1}\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}}.
$$

Since the function $(\mathrm{e}^{-1}, 1) \ni p \mapsto u_p = \frac{1-p}{-p(\log p)^3}$ is increasing and $\lim_{p \to \mathrm{e}^{-1}} u_p = \mathrm{e} - 1 > 1$, we have that $u_p > 1$ and so $1.05u_p > 1$, $p \in (\mathrm{e}^{-1}, 1)$. Therefore, for all $m \geq 3$ and $\ell \in \mathbb{N}$, we have

$$
\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{m!}{((1.05u_p)^{-1}\nu_1\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)})^{m-2}}.
$$

Combining this latter inequality with (34), we have that the condition (22) is satisfied with $\Delta_\ell := \frac{p(\log p)^4}{1.05(1-p)}\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}$, and the claim follows by Corollary 2.

Case $h \geq 2$.
If $h \geq 2$, then by (47) we have

$$
\binom{kh}{k-1} \leq 11_{\{k=1\}} + C_{h,k}11_{\{k\geq 2\}}.
$$

Combining this relation with (45) we have

$$
\mathbb{E}Z^m \leq (1-p)^h + \sqrt{1 + \frac{1}{h-1}} \frac{\mathrm{e}^{\frac{1}{24\cdot25}}}{(h-1)\sqrt{2\pi}} \sum_{k\geq 2} k^{m-1} \frac{1}{\sqrt{k-1}} \left(\frac{h^h}{(h-1)^{h-1}}\right)^k p^{k-1}(1-p)^{k(h-1)+1}
$$

$$
= (1-p)^h + \sqrt{1 + \frac{1}{h-1}} \frac{\mathrm{e}^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}} \sum_{k\geq 2} k^{m-1} \frac{1}{\sqrt{k-1}} \left(\frac{h^h}{(h-1)^{h-1}}\right)^k p^k(1-p)^{k(h-1)}
$$

$$
= (1-p)^h + \sqrt{1 + \frac{1}{h-1}} \frac{\mathrm{e}^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}} \sum_{k\geq 2} k^{m-1} \frac{1}{\sqrt{k-1}} \left(ph\left(\frac{h(1-p)}{h-1}\right)^{h-1}\right)^k
$$

$$
\leq (1-p)^h + \sqrt{1 + \frac{1}{h-1}} \frac{\mathrm{e}^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}} \sum_{k\geq 2} k^{m-1}\mathrm{e}^{-\nu_2 k}, \quad m \geq 3,
$$

where

$$v_2 := -\log\left(ph\left(\frac{h(1-p)}{h-1}\right)^{h-1}\right).$$

Now we are going to verify that $v_2 > 0$, i.e.,

$$ph\left(\frac{h(1-p)}{h-1}\right)^{h-1} < 1. \tag{50}$$

Setting $x := ph \in (0, 1)$, we have

$$x\left(\frac{h-x}{h-1}\right)^{h-1} = x\left(\frac{h-1+1-x}{h-1}\right)^{h-1}$$

$$= x\left(1+\frac{1-x}{h-1}\right)^{h-1}$$

$$= xe^{(h-1)\log\left(1+\frac{1-x}{h-1}\right)}$$

$$\leq xe^{1-x}.$$

The relation (50) follows if we notice that the mapping $x \in (0, 1) \mapsto xe^{1-x}$ is an increasing bijection from (0,1) to itself. Therefore, by Lemma 3, for any $m \geq 3$, we have

$$\mathbb{E}Z^m \leq (1-p)^h + (m-1)!\sqrt{1+\frac{1}{h-1}}\frac{e^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}}\left(v_2^{-m} + \frac{1}{\pi(m-1)(m-1)!} + \frac{2}{\pi^m}\right)$$

$$\leq (m-1)!\left((1-p)^h + \sqrt{1+\frac{1}{h-1}}\frac{e^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}}\right)\left(v_2^{-m} + \frac{1}{\pi(m-1)(m-1)!} + \frac{2}{\pi^m}\right)$$

$$\leq (m-1)!u_{p,h}\left(v_2^{-m} + \frac{1}{\pi(m-1)(m-1)!} + \frac{2}{\pi^m}\right), \tag{51}$$

where

$$u_{p,h} := 1 + \sqrt{1+\frac{1}{h-1}}\frac{e^{\frac{1}{24\cdot25}}(1-p)}{p(h-1)\sqrt{2\pi}}.$$

Proof of Part (ii.1).

If $ph(h(1-p)/(h-1))^{h-1} \leq e^{-1}$, i.e., $v_2 \geq 1$, then combining (36) (with v_2 in place of v) with (51), we have

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell\text{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{((u_{p,h})^{-1})^{m-3}m!}{((u_{p,h})^{-1}\sqrt{\lambda_\ell\text{Leb}(B_\ell \cap C_\ell)})^{m-2}}$$

$$\leq \frac{m!}{((u_{p,h})^{-1}\sqrt{\lambda_\ell\text{Leb}(B_\ell \cap C_\ell)})^{m-2}}, \quad m \geq 3, \tag{52}$$

where we used that $(u_{p,h})^{-1} < 1$. Combining (52) with (34), we have that the condition (22) is satisfied with $\Delta_\ell := (u_{p,h})^{-1}\sqrt{\lambda_\ell\text{Leb}(B_\ell \cap C_\ell)}$, and the claim follows by Corollary 2.

Proof of Part (ii.2).

If $ph(h(1-p)/(h-1))^{h-1} > e^{-1}$, i.e., $v_2 < 1$, then combining (37) (with v_2 in place of v) with (51), we have

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq u_{p,h} \frac{(m-1)!}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \left(v_2^{-m} + 0.16\right)$$

$$= \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}v_2)^{m-2}} \frac{v_2^m}{(u_{p,h})^{-1}v_2^2} \left(v_2^{-m} + 0.16\right), \quad m \geq 3.$$

Since $v_2 < 1$, we have

$$\frac{v_2^m}{(u_{p,h})^{-1}v_2^2}\left(v_2^{-m} + 0.16\right) \leq \frac{1.16}{(u_{p,h})^{-1}v_2^2} := \widetilde{u}_{p,h}.$$

Therefore

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}v_2)^{m-2}}\widetilde{u}_{p,h}$$

$$= \frac{(m-1)!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}v_2(\widetilde{u}_{p,h})^{-1})^{m-2}}((\widetilde{u}_{p,h})^{-1})^{m-3}, \quad m \geq 3.$$

Using Lemma 4, we have that $(\widetilde{u}_{p,h})^{-1} < (u_{p,h})^{-1}v_2^2 < (u_{p,h})^{-1} < 1$. Therefore

$$\frac{\mathbb{E}Z^m}{\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}^{m-2}} \leq \frac{m!}{(\sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}v_2(\widetilde{u}_{p,h})^{-1})^{m-2}}, \quad m \geq 3.$$

Combining this latter inequality with (34), we have that the condition (22) is satisfied with $\Delta_\ell := \sqrt{\lambda_\ell \mathrm{Leb}(B_\ell \cap C_\ell)}v_2(\widetilde{u}_{p,h})^{-1}$, and the claim follows by Corollary 2. $\square$

8. Application to a class of interferences in a wireless communication model

8.1. Gaussian approximation

In this section we apply Theorem 3 to the interference $I(\{\mathbf{0}\})$ (see e.g. Remark 1) when the Poisson process of node locations has a piecewise constant intensity function of the form $\lambda(x) := \lambda \mathbf{1}_B(x)$, for some $\lambda > 0$ and $B \in \mathcal{B}(\mathbb{R}^2)$. In such a case we have quite explicit upper bounds on the Wasserstein and Kolmogorov distances. The following corollary (whose proof is straightforward, and therefore omitted) allows for explicit bounds for some classes of signal power distributions and attenuation functions.

Corollary 7. *Let $\lambda(x) := \lambda \mathbf{1}_B(x)$, $x \in \mathbb{R}^2$, for some $\lambda > 0$ and $B \in \mathcal{B}(\mathbb{R}^2)$ such that*

$$0 < \mathbb{E}Z_1^2 \int_B A^2(x)\mathrm{d}x < \infty.$$

Then

$$d_W(L\{\mathbf{0}\}), G) \leq \frac{1}{\sqrt{\lambda}} \frac{\mathbb{E}Z_1^3}{(\mathbb{E}Z_1^2)^{3/2}} \frac{\int_B A(x)^3\mathrm{d}x}{\left(\int_B A(x)^2\mathrm{d}x\right)^{3/2}}$$

and

$$d_K(L(\{\mathbf{0}\}), G)$$

$$\leq \frac{1}{\sqrt{\lambda}} \left[1 + \frac{1}{2} \max\left\{ 4, \left[4\frac{1}{\lambda} \frac{\mathbb{E}Z_1^4}{(\mathbb{E}Z_1^2)^2} \frac{\int_B A(x)^4 \mathrm{d}x}{\left(\int_B A(x)^2 \mathrm{d}x\right)^2} + 2 \right]^{1/4} \right\} \right] \frac{\mathbb{E}Z_1^3}{(\mathbb{E}Z_1^2)^{3/2}} \frac{\int_B A(x)^3 \mathrm{d}x}{\left(\int_B A(x)^2 \mathrm{d}x\right)^{3/2}}$$

$$+ \frac{1}{\sqrt{\lambda}} \frac{\sqrt{\mathbb{E}Z_1^4}}{\mathbb{E}Z_1^2} \frac{\sqrt{\int_B A(x)^4 \mathrm{d}x}}{\int_B A(x)^2 \mathrm{d}x}.$$

Example 2. If the path loss function is the Hertzian attenuation function, i.e., $A(x) := \max\{R, \|x\|\}^{-\alpha}$, $x \in \mathbb{R}^2$, for some $R > 0$ and $\alpha > 1$, $B := \mathbb{R}^2$, and $\mathbb{E}Z_1^2 \in (0, \infty)$, then Corollary 7 applies with

$$\int_B A(x)^m \mathrm{d}x = 2\pi \left(\frac{1}{2} - \frac{1}{2 - \alpha m} \right) R^{2-\alpha m}, \quad \text{for } m = 2, 3, 4.$$

8.2. Moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction term

In this section we apply Theorem 4 to the sequence $\{I_\ell(\{\mathbf{0}\})\}_{\ell \geq 1}$ (defined in Remark 2) when the Poisson processes of node locations have piecewise deterministic intensity functions of the form $\lambda_\ell(x) := \lambda_\ell \mathbf{1}_{B_\ell}(x)$, $x \in \mathbb{R}^2$, for some sequences $\{\lambda_\ell\}_{\ell \geq 1} \subset (0, \infty)$ and $\{B_\ell\}_{\ell \geq 1} \subset \mathcal{B}(\mathbb{R}^2)$. In such a case the assumption (17) is greatly simplified. The following corollary (whose proof is straightforward, and therefore omitted) holds.

Corollary 8. *Let* $\{B_\ell\}_{\ell \in \mathbb{N}} \subset \mathcal{B}(\mathbb{R}^d)$ *and* $\{\mathbb{Q}_\ell\}_{\ell \geq 1}$ *be such that*

$$0 < \mathbb{E}(Z_1^{(\ell)})^2 \int_{B_\ell} A(x)^2 \mathrm{d}x < \infty, \quad \ell \geq 1, \tag{53}$$

and assume that there exist a non-negative constant $\gamma \geq 0$ *and a positive numerical sequence* $\{\Delta_\ell\}_{\ell \in \mathbb{N}}$ *such that*

$$\frac{1}{\lambda_\ell^{\frac{m}{2}-1}} \frac{\mathbb{E}(Z_1^{(\ell)})^m}{(\mathbb{E}(Z_1^{(\ell)})^2)^{m/2}} \frac{\int_{B_\ell} A(x)^m \mathrm{d}x}{\left(\int_{B_\ell} A(x)^2 \mathrm{d}x\right)^{m/2}} \leq \frac{(m!)^{1+\gamma}}{\Delta_\ell^{m-2}}, \quad \text{for all } m \geq 3 \text{ and } \ell \in \mathbb{N}. \tag{54}$$

Then the sequence $\{I_\ell(\{\mathbf{0}\})\}_{\ell \geq 1}$ *satisfies an* **MDP**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *a* **BCI**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, *and an* **NACC**$(\gamma, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$.

Example 3. Under the notation of Corollary 8, let us set $B_\ell \equiv \mathbb{R}^2$ for any $\ell \geq 1$, suppose that $\mathbb{Q}_\ell$ is the exponential law with mean μ^{-1}, for some $\mu > 0$, and assume that the attenuation of the signal is Hertzian, i.e., $A(x) := \max\{R, \|x\|\}^{-\alpha}$, $x \in \mathbb{R}^2$, for some constants $R > 0$ and $\alpha > 1$. Then

$$\frac{\mathbb{E}(Z_1^{(\ell)})^m}{(\mathbb{E}(Z_1^{(\ell)})^2)^{m/2}} = \frac{m!}{2^{m/2}} \quad \text{and} \quad \int_{\mathbb{R}^2} A(x)^m \mathrm{d}x = \frac{\pi \alpha m}{\alpha m - 2} R^{2-\alpha m}, \quad \text{for any } \ell \in \mathbb{N} \text{ and } m \geq 2.$$

So the assumption (53) of Corollary 8 is satisfied, and the left-hand side of the relation (54) reads, for any $\ell \in \mathbb{N}$ and $m \geq 3$,

$$\frac{m!}{\left(R\sqrt{2\lambda_\ell \frac{\pi \alpha}{\alpha-1}}\right)^{m-2}} \times \frac{m(2\alpha - 2)}{4(\alpha m - 2)}.$$

Since $m \geq 3$ we have $\frac{\alpha m - m}{\alpha m - 2} < 1$. So Corollary 8 yields that the sequence $\{I_\ell(\{\mathbf{0}\})\}_{\ell \geq 1}$ satisfies an **MDP**$(0, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, a **BCI**$(0, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$, and an **NACC**$(0, \{\Delta_\ell\}_{\ell \in \mathbb{N}})$ with

$$\Delta_\ell := R\sqrt{2\lambda_\ell \frac{\pi\alpha}{\alpha - 1}}, \quad \ell \geq 1.$$

9. Conclusion

Exploiting the theory developed in [30], we have provided explicit bounds on the Wasserstein and Kolmogorov distances between random variables lying in the first chaos of the Poisson space and the standard normal distribution. Relying on the findings in [40] and on a fine control of the cumulants of the first chaos on the Poisson space, we have also provided moderate deviations, Bernstein-type concentration inequalities, and normal approximation bounds with Cramér correction terms for the same random variables. We have applied these results to Poisson shot noise random variables, and in particular to generalized compound Hawkes point processes. As far as Hawkes processes are concerned, the results proven in this paper generalize many of the asymptotic theorems found in the literature [20, 23, 26, 47] to the spatial case, eventually with a varying baseline intensity and with less constraining assumptions on the excitation kernels.

We point out that some Hawkes processes have a Galton–Watson representation but cannot easily be expressed as a Poisson integral of the type (12). The main example is that of a multivariate Hawkes process exhibiting both self-excitation and cross-excitation between many interacting nodes. Indeed, such a process does have a branching structure [15], but a priori it does not fall within the context of this paper. To the best of our knowledge, we only have bounds on the Wasserstein metric between multivariate Hawkes processes with exponential kernels and their multivariate Gaussian limit, which is of order $O\left(1/\sqrt{t}\right)$ [25].

Another interesting development of the results proven in this paper would be their extension to the whole path of the process, rather than the process evaluated at one instant. More specifically, we would like to find upper bounds on the distance between the centered and normalized *path* of the Poisson shot noise process, and its limiting Gaussian *process* in the space of càdlàg functions equipped with the Skorokhod metric, for example using the results provided in [5]. These approximation results are obviously more delicate to obtain, and to the best of our knowledge they have been studied only in a few works, such as [6].

Appendix A. Proofs of Lemmas 1, 3, and 4

A.1. Proof of Lemma 1

The claim is clearly true if $\max\{\text{Leb}(B \cap C), \mathbb{E}Z^m, \mathbb{E}M^m\} = +\infty$. Therefore we assume $\max\{\text{Leb}(B \cap C), \mathbb{E}Z^m, \mathbb{E}M^m\} < +\infty$. We start with the obvious inequality

$$\mathbb{E}|v(Z_1)(C - x)|^m \leq \mathbb{E}\left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|\right)^m. \tag{55}$$

Using Hölder's inequality we have

$$\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}| \leq \left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} 1\right)^{\frac{m-1}{m}} \left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|^m\right)^{\frac{1}{m}}.$$

Raising this to the mth power, we obtain

$$\left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|\right)^m \le (Z_1(C-x,\mathbb{R}))^{m-1} \sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|^m.$$

Using the independence between $Z_1(C-x,\mathbb{R})$ and $(|M_{k,1}|)_{k\in\mathbb{N}}$ and Wald's identity, we have

$$\mathbb{E}\left(\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|\right)^m \le \mathbb{E}\left[(Z_1(C-x,\mathbb{R}))^{m-1} \sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|^m\right]$$

$$= \mathbb{E}\left[(Z_1(C-x,\mathbb{R}))^{m-1}\, \mathbb{E}\left[\sum_{k=0}^{Z_1(C-x,\mathbb{R})-1} |M_{k,1}|^m \middle| Z_1(C-x,\mathbb{R})\right]\right]$$

$$= \mathbb{E}\left[(Z_1(C-x,\mathbb{R}))^m |M|^m\right],$$

and finally the inequality (55) yields

$$\mathbb{E}|v(Z_1)(C-x)|^m \le \mathbb{E}|M|^m \mathbb{E} Z_1(C-x,\mathbb{R})^m.$$

Recalling that we denote by $\{Y_{1,k}\}_{k\ge 0}$, $Y_{1,0} := \mathbf{0}$ the first components of the points of $Z_1(\cdot,\cdot)$, we have

$$Z_1(C-x,\mathbb{R})^m = \operatorname{card}\left\{k \in \mathbb{N} \cup \{0\} : Y_{1,k} \in C-x\right\}^m$$

$$= \left(\sum_{k=0}^{Z-1} \mathbf{1}_{Y_{1,k}\in C-x}\right)^m, \quad m \ge 1. \tag{56}$$

The mth power of the sum of indicators can be expanded by using the multinomial theorem, which yields

$$\left(\sum_{k=0}^{Z-1} \mathbf{1}_{Y_{1,k}\in C-x}\right)^m = \sum_{\substack{k_0,\ldots,k_{Z-1}\ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \prod_{i=0}^{Z-1} \mathbf{1}_{Y_{1,i}\in C-x}^{k_i}$$

$$= \sum_{\substack{k_0,\ldots,k_{Z-1}\ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \prod_{i=0}^{Z-1} \mathbf{1}_{Y_{1,i}\in C-x}$$

$$\le \sum_{\substack{k_0,\ldots,k_{Z-1}\ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \mathbf{1}_{Y_{1,0}\in C-x}, \quad m \ge 1. \tag{57}$$

Here

$$\binom{m}{k_0,\ldots,k_{Z-1}} = \frac{m!}{k_0!\ldots k_{Z-1}!}$$

denotes the multinomial coefficient. By (57) we have

$$\int_B \left(\sum_{k=0}^{Z-1} \mathbf{1}_{Y_{1,k} \in C-x} \right)^m \mathrm{d}x \le \sum_{\substack{k_0,\ldots,k_{Z-1} \ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \int_B \mathbf{1}_{Y_{1,0} \in C-x} \mathrm{d}x$$

$$= \sum_{\substack{k_0,\ldots,k_{Z-1} \ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \int_B \mathbf{1}_{\mathbf{0} \in C-x} \mathrm{d}x$$

$$= \sum_{\substack{k_0,\ldots,k_{Z-1} \ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \mathrm{Leb}(B \cap C)$$

$$= \mathrm{Leb}(B \cap C) \sum_{\substack{k_0,\ldots,k_{Z-1} \ge 0 \\ k_0+\cdots+k_{Z-1}=m}} \binom{m}{k_0,\ldots,k_{Z-1}} \prod_{i=0}^{Z-1} 1^{k_i}$$

$$\le \mathrm{Leb}(B \cap C) Z^m, \quad m \ge 1, \tag{58}$$

where the latter inequality follows from another application of the multinomial theorem. The claim easily follows by (55), (56), and (58).

A.2. Proof of Lemma 3

Set $D := \{z \in \mathbb{C} : \mathrm{Re}\, z > 0\}$ and define $f(z) := z^{m-1} e^{-vz}$, $z \in D$, $m \ge 2$, $v > 0$.
Clearly, f is analytic on D; we shall check later on that the following statements hold:

$$\text{For any compact set } K \subset (0, \infty), \lim_{y \to +\infty} \sup_{x \in K} |f(x \pm \mathbf{i}y)| e^{-2\pi y} = 0. \tag{59}$$

$$\text{For any } x > 0, \int_0^{+\infty} |f(x + \mathbf{i}y) - f(x - \mathbf{i}y)| e^{-2\pi y} \mathrm{d}y < \infty \quad \text{and}$$

$$\lim_{x \to +\infty} \int_0^{+\infty} |f(x + \mathbf{i}y) - f(x - \mathbf{i}y)| e^{-2\pi y} \mathrm{d}y = 0. \tag{60}$$

Therefore, by the Abel–Plana formula (see e.g. [11]) we have (note that $f(0) = 0$)

$$\sum_{k \ge 1} f(k) = \int_0^\infty f(t) \mathrm{d}t + \mathbf{i} \int_0^\infty \frac{f(\mathbf{i}t) - f(-\mathbf{i}t)}{e^{2\pi t} - 1} \mathrm{d}t$$

$$= \int_0^{+\infty} e^{-vt} t^{m-1} \mathrm{d}t + \mathbf{i} \int_0^{+\infty} \frac{e^{-\mathbf{i}vt}(\mathbf{i}t)^{m-1} - e^{\mathbf{i}vt}(-\mathbf{i}t)^{m-1}}{e^{2\pi t} - 1} \mathrm{d}t$$

$$= v^{-m}(m-1)! + R_m, \tag{61}$$

where we used that

$$\frac{v^m}{\Gamma(m)} \int_0^\infty t^{m-1} e^{-vt} dt = 1 \tag{62}$$

and that the Euler gamma function $\Gamma(\cdot)$ computed at the integer m is equal to $(m-1)!$. We proceed by bounding $|R_m|$ from above. We distinguish two cases: $m = 2p$ and $m = 2p + 1$, $p \in \mathbb{N}$. If $m = 2p$ we have

$$\begin{aligned}
R_{2p} &= \mathbf{i} \int_0^{+\infty} \frac{e^{-\mathbf{i}vt}(\mathbf{i}t)^{2p-1} - e^{\mathbf{i}vt}(-\mathbf{i}t)^{2p-1}}{e^{2\pi t} - 1} dt \\
&= \mathbf{i} \int_0^{+\infty} \frac{e^{-\mathbf{i}vt}(\mathbf{i}t)^{2p-1} + e^{\mathbf{i}vt}(\mathbf{i}t)^{2p-1}}{e^{2\pi t} - 1} dt \\
&= \mathbf{i}^{2p} \int_0^{+\infty} \frac{t^{2p-1}(e^{-\mathbf{i}vt} + e^{\mathbf{i}vt})}{e^{2\pi t} - 1} dt \\
&= 2(-1)^p \int_0^{+\infty} \frac{t^{2p-1} \cos(vt)}{e^{2\pi t} - 1} dt.
\end{aligned}$$

Thus

$$\begin{aligned}
|R_m| &\leq 2 \int_0^{+\infty} \frac{t^{m-1}}{e^{2\pi t} - 1} dt \\
&= 2 \int_0^1 \frac{t^{m-1}}{e^{2\pi t} - 1} dt + 2 \int_1^{+\infty} \frac{t^{m-1}}{e^{2\pi t} - 1} dt \\
&\leq \frac{1}{\pi} \int_0^1 \frac{t^{m-1}}{t} dt + 2 \int_1^{+\infty} e^{-\pi t} t^{m-1} dt \\
&\leq \frac{1}{\pi(m-1)} + 2\pi^{-m}(m-1)!, \quad m = 2p, \tag{63}
\end{aligned}$$

where the latter inequality follows from (62) with π in place of v. Similarly, we have

$$\begin{aligned}
|R_m| &= \left| \mathbf{i} \int_0^{+\infty} \frac{e^{-\mathbf{i}vt}(\mathbf{i}t)^{m-1} - e^{\mathbf{i}vt}(-\mathbf{i}t)^{m-1}}{e^{2\pi t} - 1} dt \right| \\
&= 2 \left| \int_0^{+\infty} \frac{t^{m-1} \sin(vt)}{e^{2\pi t} - 1} dt \right| \\
&\leq 2 \int_0^{+\infty} \frac{t^{m-1}}{e^{2\pi t} - 1} dt \\
&\leq \frac{1}{\pi(m-1)} + 2\pi^{-m}(m-1)!, \quad m = 2p + 1. \tag{64}
\end{aligned}$$

The claim follows by the relations (61), (63) and (64).

It remains to prove (59) and (60). We start by proving (59). Let $K \subset (0, \infty)$ be an arbitrary compact set. We have

$$\sup_{x \in K} |f(x \pm \mathbf{i}y)| e^{-2\pi y} = e^{-2\pi y} \sup_{x \in K} (x^2 + y^2)^{\frac{m-1}{2}} e^{-\nu x}$$

$$\leq ((\sup K)^2 + y^2)^{\frac{m-1}{2}} e^{-2\pi y} \to 0, \quad \text{as } y \to +\infty.$$

Finally we prove (60). For any $x, y > 0$, we have

$$|f(x + \mathbf{i}y) - f(x - \mathbf{i}y)| e^{-2\pi y} = |(x + \mathbf{i}y)^{m-1} e^{-\nu \mathbf{i}y} - (x - \mathbf{i}y)^{m-1} e^{\nu \mathbf{i}y}| e^{-2\pi y} e^{-\nu x}$$

$$= \left| \sum_{k=0}^{m-1} \binom{m-1}{k} x^k \left((\mathbf{i}y)^{m-1-k} e^{-\nu \mathbf{i}y} - (-\mathbf{i}y)^{m-1-k} e^{\nu \mathbf{i}y} \right) \right| e^{-2\pi y} e^{-\nu x}$$

$$\leq 2 \sum_{k=0}^{m-1} \binom{m-1}{k} x^k e^{-\nu x} y^{m-1-k} e^{-2\pi y}.$$

Therefore, for any $x > 0$, we have

$$\int_0^{+\infty} |f(x + \mathbf{i}y) - f(x - \mathbf{i}y)| e^{-2\pi y} \mathrm{d}y \leq 2 \sum_{k=0}^{m-1} \binom{m-1}{k} x^k e^{-\nu x} \int_0^{+\infty} y^{m-k-1} e^{-2\pi y} \mathrm{d}y$$

$$= 2 \sum_{k=0}^{m-1} \binom{m-1}{k} x^k e^{-\nu x} \frac{(m-k-1)!}{(2\pi)^{m-k}}, \quad (65)$$

where the latter equality follows from the relation (62) with $m - k$ in place of m and 2π in place of ν. Clearly, the right-hand side of the relation (65) is finite and tends to zero as $x \to +\infty$. The proof is completed.

A.3. Proof of Lemma 4

A simple computation shows

$$f'(x) = (x - 1 - \log x)(3x - 3 - \log x), \quad x \in (0, 1).$$

Since $x - 1 - \log x > 0$ for every $x \in (0, 1)$, the sign of f' coincides with the sign of $g(x) := 3x - 3 - \log x$, $x \in (0, 1)$. Studying the derivative of g, we see that g is increasing on $(1/3, 1)$ and decreasing on $(0, 1/3)$, with a minimum at $x = 1/3$. Since $\lim_{x \to 0^+} g(x) = +\infty$, $g(1/3) < 0$, and $\lim_{x \to 1^-} g(x) = 0$, we then have that there exists a unique $x^* \in (0, 1/3)$ with $g(x^*) = 0$, g is positive on $(0, x^*)$, and it is negative on $(x^*, 1)$. Therefore f has a maximum at x^*, and consequently, for any $x \in (0, 1)$, we have

$$f(x) \leq f(x^*) = x^*(x^* - 1 - \log x^*)^2 = 4x^*(1 - x^*)^2 \leq 4x^*(1 - x^*) \leq 8/9 < 1.$$

Here we used that $g(x^*) = 0$, that the mapping $(0, 1) \in x \mapsto x(1 - x)$ is increasing on $(0, 1/2)$, and that $x^* < 1/3$.

Acknowledgements

Funding information

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BACCELLI, F. AND BŁASZCZYSZYN, B. (2001). On a coverage process ranging from the Boolean model to the Poisson–Voronoi tessellation with applications to wireless communications. *Adv. Appl. Prob.* **33**, 293–323.

[2] BACCELLI, F. AND BŁASZCZYSZYN, B. (2009). *Stochastic Geometry and Wireless Networks*, Vol. I, *Theory*. Now Foundations and Trends, Norwell, MA.

[3] BACCELLI, F. AND BŁASZCZYSZYN, B. (2009). *Stochastic Geometry and Wireless Networks*, Vol. II, *Applications*. Now Foundations and Trends, Norwell, MA.

[4] BACRY, E., DELATTRE, S., HOFFMANN, M. AND MUZY, J. F. (2013). Some limit theorems for Hawkes processes and applications to financial statistics. *Stoch. Process. Appl.* **123**, 2475–2499.

[5] BARBOUR, A. D., ROSS, N. AND ZHENG, G. (2023). Stein's method, Gaussian processes and Palm measures, with applications to queueing. *Ann. Appl. Prob.* **33**, 3835–3871.

[6] BESANÇON, E., COUTIN, L., DECREUSFEOND, L. AND MOYAL, P. (2024). Diffusive limits of Lipschitz functionals of Poisson measures. *Ann. Appl. Prob.* **34**, 555–584.

[7] BRÉMAUD, P. (2000). An insensitivity property of Lundberg's estimate for delayed claims. *J. Appl. Prob.* **37**, 914–917.

[8] BRÉMAUD, P. AND MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24**, 1563–1588.

[9] BRÉMAUD, P., MASSOULIÉ, L. AND RIDOLFI, A. (2005). Power spectra of random spike fields and related processes. *Adv. Appl. Prob.* **37**, 1116–1146.

[10] BRIGHAM, M. AND DESTEXHE, A. (2015). Nonstationary filtered shot-noise processes and applications to neuronal membranes. *Phys. Rev. E* 91, article no. 062102.

[11] BUTZER, P. L., FERREIRA, P. J. S. G., SCHMEISSER, G. AND STENS, R. L. (2011). The summation formulae of Euler–Maclaurin, Abel–Plana, Poisson, and their interconnections with the approximate sampling formula of signal analysis. *Results Math.* **59**, 359–400.

[12] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*. Springer, New York.

[13] DÖRING, H. AND EICHELSBACHER, P. (2013). Moderate deviations via cumulants. *J. Theoret. Prob.* **26**, 360–385.

[14] DÖRING, H., JANSEN, S. AND SCHUBERT, K. (2022). The method of cumulants for normal approximation. *Prob. Surveys* **19**, 185–270.

[15] EMBRECHTS, P., LINIGER, T. AND LIN, L. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Prob.* **48A**, 367–378.

[16] GANESH, A. AND TORRISI, G. L. (2008). Large deviations of the interference in a wireless communication model. *IEEE Trans. Inf. Theory* **54**, 3505–3517.

[17] GANESH, A., MACCI, C. AND TORRISI, G. L. (2005). Sample path large deviations for Poisson shot noise processes, and applications. *Electron. J. Prob.* **10**, 1026–1043.

[18] GANESH, A., MACCI, C. AND TORRISI, G. L. (2007). A class of risk processes with reserve-dependent premium rate: sample path large deviations and importance sampling. *Queueing Systems* **55**, 83–94.

[19] GAO, F. AND WANG, Y. (2020). Functional central limit theorems and moderate deviations for Poisson cluster processes *Adv. Appl. Prob.* **52**, 916–941.

[20] GAO, F. AND ZHU, L. (2021). Precise deviations for Hawkes processes. *Bernoulli* **27**, 221–248.

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

[22] HAWKES, A. G. AND OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 493–503.

[23] HILLAIRET, C., HUANG, L., KHABOU, M. AND RÉVEILLAC, A. (2022). The Malliavin–Stein method for Hawkes functionals. *ALEA* **19**, 1293–1328.

[24] KONSTANTOPOULOS, T. AND LIN, S. J. (1998). Macroscopic models for long-range dependent network traffic. *Queueing Systems* **28**, 214–243.

[25] KHABOU, M. (2021). *Malliavin–Stein method for the multivariate compound Hawkes process*. Preprint. Available at https://arxiv.org/abs/2109.07749.

[26] KHABOU, M., PRIVAULT, N. AND RÉVEILLAC, A. (2024). Normal approximation of compound Hawkes functionals. *J. Theoret. Prob.* **37**, 549–581.

[27] KLÜPPELBERG, C. AND MIKOSCH, T. (1995). Explosive Poisson shot noise processes with applications to risk reserves. *Bernoulli* **1**, 125–147.

[28] KLÜPPELBERG, C. AND MIKOSCH, T. (1995). Delay in claim settlement and ruin probability approximations. *Scand. Actuarial J.* **1995**, 154–168.

[29] LAST, G. AND PENROSE, M. (2017). *Lectures on the Poisson Process*. Cambridge University Press.

[30] LAST, G., PECCATI, G. AND SCHULTE, M. (2016). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Prob. Theory Relat. Fields* **165**, 667–723.

[31] LEONARDI, E. AND TORRISI, G. L. (2017). Modeling LRU caches with shot noise request processes. *SIAM J. Appl. Math.* **77**, 361–383.

[32] MACCI, C. AND TORRISI, G. L. (2004). Asymptotic results for perturbed risk processes with delayed claims. *Insurance Math. Econom.* **34**, 307–320.

[33] MACCI, C., STABILE, G. AND TORRISI, G. L. (2005). Lundberg parameters for non standard risk processes. *Scand. Actuarial J.* **6**, 417–432.

[34] MØLLER, J. AND TORRISI, G. L. (2007). The pair correlation function of spatial Hawkes processes. *Statist. Prob. Lett.* **77**, 995–1003.

[35] MØLLER, J. AND WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, FL.

[36] PRIVAULT, N. (2020). Nonstationary shot noise modeling of neuron membrane potentials by closed-form moments and Gram–Charlier expansions. *Biol. Cybernetics* **114**, 499–518.

[37] PRIVAULT, N. (2021). Recursive computation of the Hawkes cumulants. *Statist. Prob. Lett.* **177**, article no. 109161.

[38] PRIVAULT, N. AND TORRISI, G. L. (2011). Density estimation of functionals of spatial point processes with application to wireless networks. *SIAM J. Math. Anal.* **43**, 1311–1344.

[39] REYNAUD-BOURET, P. AND ROY, E. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belgian Math. Soc. Simon Stevin* **13**, 883–896.

[40] SAULIS, L. AND STATULEVICIUS, V. A. (1991). *Limit Theorems for Large Deviations*. Kluwer, Amsterdam.

[41] SCHULTE, M. AND THÄLE, C. (2023). *Moderate deviations on Poisson chaos*. Preprint. Available at https://arxiv.org/abs/2304.00876v2.

[42] TORRISI, G. L. AND LEONARDI, E. (2013). Simulating the tail of the interference in a Poisson network model. *IEEE Trans. Inf. Theory* **59**, 1773–1787.

[43] TORRISI, G. L. (2004). Simulating the ruin probability of risk processes with delay in claim settlement. *Stoch. Process. Appl.* **112**, 225–244.

[44] TORRISI, G. L. (2016). Gaussian approximation of nonlinear Hawkes processes. *Ann. Appl. Prob.* **26**, 2106–2140.

[45] TORRISI, G. L. (2017). Poisson approximation of point processes with stochastic intensity, and application to nonlinear Hawkes processes. *Ann. Inst. H. Poincaré Prob. Statist.* **53**, 679–700.

[46] TORRISI, G. L. AND LEONARDI, E. (2022). Asymptotic analysis of Poisson shot noise processes, and applications. *Stoch. Process. Appl.* **144**, 229–270.

[47] ZHU, L. (2013). Moderate deviations for Hawkes processes. *Statist. Prob. Lett.* **83**, 885–890.