Cumulants of Hawkes point processes

Stojan Jovanović*
Bernstein Center Freiburg & Faculty of Biology, University of Freiburg, 79104 Freiburg im Breisgau, Germany and KTH Royal Institute of Technology, 10691 Stockholm, Sweden

John Hertz†
Institute for Neuroscience and Pharmacology and Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark and NORDITA, KTH Royal Institute of Technology and Stockholm University, 10691 Stockholm, Sweden

Stefan Rotter‡
Bernstein Center Freiburg & Faculty of Biology, University of Freiburg, 79104 Freiburg im Breisgau, Germany
(Received 19 September 2014; published 7 April 2015)

We derive explicit, closed-form expressions for the cumulant densities of a multivariate, self-exciting Hawkes point process, generalizing a result of Hawkes in his earlier work on the covariance density and Bartlett spectrum of such processes. To do this, we represent the Hawkes process in terms of a Poisson cluster process and show how the cumulant density formulas can be derived by enumerating all possible "family trees," representing complex interactions between point events. We also consider the problem of computing the integrated cumulants, characterizing the average measure of correlated activity between events of different types, and derive the relevant equations.

DOI: 10.1103/PhysRevE.91.042802 PACS number(s): 05.65.+b, 05.10.Gg, 87.10.Mn, 87.19.ll

I. INTRODUCTION

The Hawkes point process was first introduced in Ref. [1] as a model of chain-reaction-like phenomena, in which the occurrence of an event increases the likelihood of more such events happening in the future. This intrinsic "self-exciting" property has made Hawkes processes appealing to a wide variety of researchers dealing with data exhibiting strong temporal clustering. Although it was originally used to model the dynamics of aftershocks that accompany strong earthquakes [2,3], it has since found application to many other problems, including accretion disk formation [4], gene interactions [5], social dynamics [6], insurance risk [7,8], corporate default clustering [9,10], market impact [11], high-frequency financial data [12], microstructure noise [13], crime [14], generic properties of high-dimensional inverse problems in statistical mechanics [15], and dynamics of neural networks [16,17]. There is also recent theoretical work exploring generic mathematical properties of the process in its own respect [18–20].

As the areas of application of Hawkes models continue to grow, it becomes increasingly important to understand the probabilistic behavior of the process. Unfortunately, despite its ubiquity, the mathematical properties of the Hawkes process are still not fully known. In fact, the same dynamical characteristics that make it such a useful model in practice are the ones that complicate formal analysis. Hawkes processes do not (except in some special cases, see Ref. [21]) possess the Markov property, making it impossible to study them using standard techniques.

Recently, quite a few methods have been devised to circumvent this problem; there are now many well-known results describing Hawkes process stability [22], long-term behavior [7,23], and large deviation properties [24]. Yet, since the early works of Hawkes himself on the covariance density and Bartlett spectrum [25] of self-exciting processes [1,26], few have tried to further elucidate their statistical properties. In his work, Adamopoulos [27], for example, attempts to derive the probability generating functional of the Hawkes process, but manages only to represent it implicitly, as a solution of an intractable functional equation. Errais *et al.* [10], using the elegant theory of affine jump processes, show that the moments of Hawkes processes can be computed by solving a system of nonlinear ODEs. Once again, however, explicit formulas turn out to be unobtainable by analytic means. Last, Saichev and Sornette [28,29], using the alternative Poisson cluster representation of self-exciting processes, show that the moment generating function of the Hawkes process satisfies a transcendental equation that does not admit an explicit solution.

Statistical behavior of, for example, Hawkes process moments and cumulants is of some importance in neuroscience, where the problem of quantifying levels of synchronization of action potentials has become very pertinent. It has been shown that nerve cells can be extremely sensitive to synchronous input from large groups of neurons [30]. More precisely, a neuron's firing rate profile depends, to a large degree, on higher-order correlations among the presynaptic spikes [31]. Of course, which synchronous patterns are favored by the network is also determined by its connection structure. While the contribution of specific structural motifs to the emergence of pairwise correlations (i.e., two-spike patterns) has already been dissected [16], no such result exists in the case of more complex patterns, stemming from correlations of higher order.

In this paper, we derive analytic formulas for the nth-order cumulant densities of a linear, self-exciting Hawkes process with arbitrary interaction kernels, generalizing the result in Ref. [26]. Inspired by the approach of Saichev *et al.*, we do

*stojan.jovanovic@bcf.uni-freiburg.de
†hertz@nbi.dk
‡stefan.rotter@biologie.uni-freiburg.de

this by utilizing the Poisson cluster process representation [32], which simplifies calculations considerably. Furthermore, we show that the cumulant densities admit a natural and intuitive graphical representation in terms of the branching structure of the underlying process and describe an algorithm that facilitates practical computation. Finally, we generalize the result in Ref. [16] by showing that the integrated cumulant densities can be expressed in terms of formal sums of topological motifs of a graph, induced by specifying the physical interactions between different types of point events.

II. PRELIMINARIES

A. Basic definitions

Consider a sequence $T = (T_n)_{n\geqslant 1}$ of positive, random variables, representing times of random occurrences of a certain event. Alternatively, T can be also thought of as as collection of random points on the positive half-line $\mathbb{R}^+$. By superposing all event times in the sequence, we obtain the point process $s = [s(t)]_{t\geqslant 0}$, formally defined by setting

$$s(t) := \sum_{n\geqslant 1} \delta(t - T_n), \tag{1}$$

where $\delta(t - T_n)$ denotes the Dirac δ function, centered at the random point T_n.

It is easy to see that the number of events occurring before time t is given by

$$N(t) := \int_{-\infty}^{t} s(u)du = |\{n : T_n \leqslant t\}|. \tag{2}$$

The conditional probability, given the past activity, of a new event occurring in the interval $(t,t+dt)$ is specified by the conditional rate function $[\lambda(t)]_{t\geqslant 0}$. More specifically, we have, up to first order [33],

$$P\{dN(t) = 1|\mathcal{H}_t\} = \lambda(t)dt, \tag{3}$$

where $\mathcal{H}_t$ represents the history of the point process s up to time t. Additionally, we assume that

$$P\{dN(t) \geqslant 2|\mathcal{H}_t\} = o(dt), \tag{4}$$

i.e., that the probability of two or more events arriving simultaneously is negligibly small. Intuitively, therefore, the conditional rate function represents the probability of a new event occurring in the infinitesimally near future, given the information about all events in the past.

Furthermore, from our previous considerations it also follows that $dN(t)$ is (up to first order) a Bernoulli random variable and, therefore,

$$\langle dN(t)\rangle = P\{dN(t) = 1\} = P\{\text{an event occurs at } t\}. \tag{5}$$

B. The multivariate Hawkes process

As was pointed out in Ref. [32], the Hawkes process can be defined in two equivalent ways: either by specifying its conditional rate function or as a *Poisson cluster process*, generated by a certain branching structure.

1. The conditional rate representation

Following Refs. [1] and [26], let us consider a d-dimensional point process $\mathbf{s} = [\mathbf{s}(t)]_{t\geqslant 0}$, with rate function $[\boldsymbol{\lambda}(t)]_{t\geqslant 0}$ defined by

$$\boldsymbol{\lambda}(t) := \boldsymbol{\mu} + \int_{-\infty}^{t} \mathbf{G}(t-u)\cdot d\mathbf{N}(u) \tag{6}$$

$$\equiv \boldsymbol{\mu} + \int_{-\infty}^{t} \mathbf{G}(t-u)\cdot \mathbf{s}(u)du, \tag{7}$$

where $\boldsymbol{\mu}$ denotes the d-dimensional base rate vector with positive entries ($\mu^i > 0$) and $\mathbf{G}(t)$ is an $d\times d$ matrix of nonnegative, integrable functions $g^{ij}(t)$, with support on $\mathbb{R}^+$, called the interaction kernel. In principle, therefore, the rate $\boldsymbol{\lambda}(t)$ should always remain positive, but models for which the probability of negative values is sufficiently small may be useful approximations [16].

Rewriting Eq. (6) in terms of the components of the conditional rate function $\lambda^i(t)$, we find that, $\forall i$,

$$\lambda^i(t) = \mu^i + \sum_{j=1}^{d} \int_{-\infty}^{t} g^{ij}(t-u)dN^j(u). \tag{8}$$

From Eqs. (3) and (8), we can now see that

$$\frac{P\{dN^i(t) = 1|\mathcal{H}_t\}}{dt} = \overbrace{\mu^i}^{\text{base rate}} \tag{9}$$

$$+ \underbrace{\sum_{j=1}^{d} \int_{-\infty}^{t} g^{ij}(t-u)dN^j(u)}_{\text{influence of past } \mathcal{H}_t}, \tag{10}$$

i.e., that the probability of an event of type i occurring at time t is simply the sum of a constant base rate and a convolution of the complete history of the process with the interaction kernel $\mathbf{G}(t)$, whose component $g^{ij}(t)$ describes the increase of the likelihood of type i events at t, caused by a type j event, occurring at 0. Note that, in the special case of no interactions $[g^{ij}(t) \equiv 0]$, we recover the definition of a multivariate Poisson process with constant rate $\boldsymbol{\mu}$. In this case, however, the (conditional) rate function is independent both of time and of the history $\mathcal{H}_t$.

2. The cluster process representation

Let us consider a Poisson cluster process C, which evolves in the following way ([34]; see also Fig. 1):

(1) Let I^k be a realization, on the interval $[0,T]$, of a homogeneous Poisson process with rate μ^k. We will call points in I^k *immigrants of type* k.

(2) For every k, each immigrant $x \in I^k$ generates a *cluster* of points C_x^k. All such clusters are mutually independent.

(3) The clusters C_x^k are generated according to the following branching structure:

(a) Each cluster C_x^k consists of generations of offspring of all types of the immigrant x, which itself belongs to generation 0.

(b) Recursively, given the immigrant x and the offspring of generation $1,2,\cdots,n$ of all types, every "child" y of

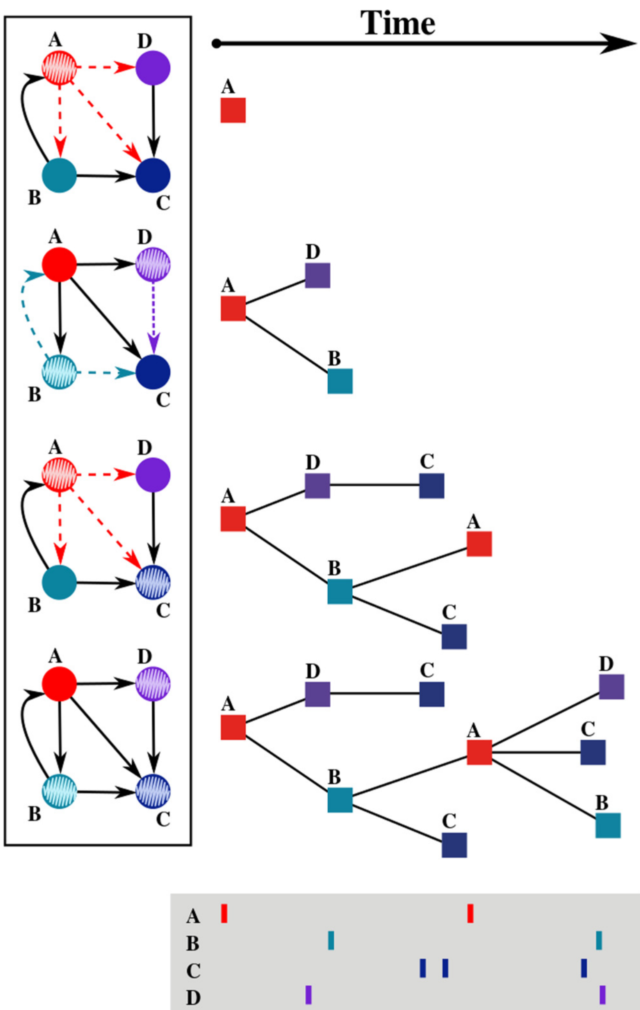

FIG. 1. (Color online) Evolution of a Poisson cluster process on a network with 4 nodes and 6 directed links. Left column: Hatched nodes are active, and dashed links are transmitting a signal to all their respective neighbors. Right column: A cluster, generated by the arrival of an immigrant from node A, evolves sequentially as new offspring are produced. Row 1: A type A immigrant arrives, providing the seed from which a new cluster will emerge. Row 2: The type A immigrant generates 1st generation offspring of types B and D, respectively. Together, these constitute the 1st generation of events. Row 3: The 1st generation, type D, event generates a single offspring event of type C. The event of type B creates two offspring of types A and C, respectively. Row 4: Finally, the 2nd generation, type A, event is the only one to generate 3 offspring events, of types B, C, and D, respectively. Bottom: A time sequence of events generated in this 4-generation-long evolution, here displayed by means of a "raster plot," which indicates the events generated by each node on the time axis.

generation n and type j, produces, $\forall i$, its own offspring of generation $n+1$ and type i by generating a realization of an inhomogeneous Poisson process O^{ij} with rate $\lambda(t) := g^{ij}(t-y)$. In other words, the probability of there being, at time t, a type i offspring event of generation $n+1$, caused by a type j event y of generation n is equal to $g^{ij}(t-y)dt$.

(4) The point process C is equal to the superposition of all points in all generated clusters, i.e.,

$$C = \sum_{k,x} C_x^k. \tag{11}$$

For example, if C were used to model the dynamics of a spiking neuronal network, the immigrants I^k would represent all the spikes of neuron k, caused by constant, external input to the network, and the clusters C_x^k all subsequent spikes, caused by action potential propagation through the network via synaptic connections.

Having defined the cluster process C, it is then possible to show (see, e.g., Ref. [32]) that by letting $\forall i$ and $\forall t \geqslant 0$,

$$\lambda^i(t) := \lim_{\delta\to 0}\frac{1}{\delta}P\{\text{event of type } i \text{ in } [t,t+\delta)|\mathcal{H}_t\}, \tag{12}$$

and assuming that the spectral radius $\rho(\mathbf{G})$ (i.e., the largest eigenvalue) of the integrated kernel matrix,

$$\mathbf{G} := \int_{-\infty}^{+\infty}\mathbf{G}(t)dt, \tag{13}$$

is strictly less than 1, then $\lambda^i(t)$ must be equal to the conditional rate function in Eq. (8). Furthermore, the corresponding point process will also be stationary. In other words, we will have

$$\frac{\langle \mathbf{N}(t)\rangle}{t} = \langle \boldsymbol{\lambda}(t)\rangle = \boldsymbol{\lambda} = (\mathbf{I}-\mathbf{G})^{-1}\boldsymbol{\mu}, \tag{14}$$

where $\mathbf{I}$ denotes the $d\times d$ identity matrix. In what follows, we will always assume that we are working with a stationary version of a Hawkes process. More specifically, we will assume that $\rho(\mathbf{G}) < 1$ and, consequently, that the matrix $(\mathbf{I}-\mathbf{G})^{-1}$ can be expanded in terms of powers of the integrated kernel matrix $\mathbf{G}$, i.e.,

$$(\mathbf{I}-\mathbf{G})^{-1} = \sum_{n=0}^{+\infty}\mathbf{G}^n. \tag{15}$$

Note that the matrix $\mathbf{G}$ has a very useful interpretation (which follows from the definition of the Poisson cluster process)—its component g^{ij} represents the average total number of events of type i in the second generation, caused by a first generation, type j event. Thus, the components of the nth matrix power $\mathbf{G}^n$ equal the average total number of type i offspring within n subsequent generations, of a first generation, type j event.

From our previous considerations, it now follows that by requiring that $\rho(\mathbf{G}) < 1$ (or, equivalently, that the series Eq. (15) converges), we, in fact, assume that each event of a given type produces only finitely many events of any other type, after an infinite number of generations.

III. THE HAWKES PROCESS CUMULANT DENSITY

Consider now an arbitrary n-dimensional random vector $\mathbf{X} = (X_1,\cdots,X_n) \equiv X_{\bar{n}}$, where we used the symbol $\bar{n}$ to denote the set $\{1,\cdots,n\}$. The cumulant of order n, denoted by $k(X_{\bar{n}})$, is a general measure of statistical dependence of the components of $\mathbf{X}$. It is defined, combinatorially, as (see

Ref. [35], page 27)

$$k(X_{\vec{n}}) = \sum_{\pi} (|\pi| - 1)!(-1)^{|\pi|-1} \prod_{B\in\pi} \langle X_B \rangle, \qquad (16)$$

where the sum goes over all partitions π of the set $\{1,\cdots,n\}$, $|\cdot|$ denotes the number of blocks of a given partition, and

$$\langle X_B \rangle = \left\langle \prod_{i\in B} X_i \right\rangle. \qquad (17)$$

A dual formula, expressing moments in terms of cumulants, reads

$$\langle X_{\vec{n}} \rangle = \sum_{\pi} \prod_{B\in\pi} k(X_B), \qquad (18)$$

where $k(X_B)$ denotes the cumulant of those components of $\mathbf{X}$, whose indices are in B.

The cumulant $k(X_{\vec{n}})$ is a natural generalization, to higher dimensions, of the covariance $\mathrm{cov}(X_1,X_2)$ of two variables.

Indeed, if we set $n = 2$, $\mathbf{X} = (X_1,X_2)$ and apply Eq. (16), we obtain

$$\begin{aligned} k(X_1,X_2) = {} & (1-1)!(-1)^{1-1}\langle X_1X_2\rangle \\ & + (2-1)!(-1)^{2-1}\langle X_1\rangle\langle X_2\rangle, \end{aligned} \qquad (19)$$

as $\pi_1 = \{\{1,2\}\}$ and $\pi_2 = \{\{1\},\{2\}\}$ are the only partitions of the set $\{1,2\}$. Also, obviously, $|\pi_1| = 1$ and $|\pi_2| = 2$. Thus,

$$k(X_1,X_2) = \langle X_1X_2\rangle - \langle X_1\rangle\langle X_2\rangle = \mathrm{cov}(X_1,X_2). \qquad (20)$$

For a given time vector $\mathbf{t} = (t_1,\cdots,t_n)$ and multi-index $\mathbf{i} = (i_1,\cdots,i_n)$, we now define the $\mathbf{n}$th order cumulant density of the Hawkes process, denoted by $k^{\mathbf{i}}(\mathbf{t})$, by letting

$$k^{\mathbf{i}}(\mathbf{t}) := \frac{k[dN^{i_1}(t_1),\cdots,dN^{i_n}(t_n)]}{d\mathbf{t}}, \qquad (21)$$

where we used $d\mathbf{t}$ to denote the differential $dt_1\cdots dt_n$.

As in the general case, the cumulant density $k^{\mathbf{i}}(\mathbf{t})$ is used to quantify the mutual dependence of random events of types $(i_1,\cdots,i_n)$ at times $(t_1,\cdots,t_n)$.

For example, from Eqs. (20) and (5), we can see that, for $\mathbf{i} = (1,2)$, $\mathbf{t} = (t_1,t_2)$ and $d\mathbf{t} = dt_1dt_2$,

$$\begin{aligned} k^{\mathbf{i}}(\mathbf{t})d\mathbf{t} = {} & P\{\text{type 1 event at } t_1, \text{ type 2 event at } t_2\} \\ & - P\{\text{type 1 event at } t_1\}P\{\text{type 2 event at } t_2\}. \end{aligned} \qquad (22)$$

The formulas for the nth order cumulant density $k^{\mathbf{i}}(\mathbf{t})$, however, get more and more complicated with increasing n, as the number of set partitions involved grows supraexponentially.

To illustrate this point, we set $n = 3$, $\mathbf{i} = (1,2,3)$, and $\mathbf{t} = (t_1,t_2,t_3)$. Then, from Eq. (16), we have

$$\begin{aligned} k^{\mathbf{i}}(\mathbf{t})d\mathbf{t} = {} & \langle dN^1(t_1)dN^2(t_2)dN^3(t_3)\rangle \\ & - \langle dN^1(t_1)dN^2(t_2)\rangle\langle dN^3(t_3)\rangle \\ & - \langle dN^1(t_1)dN^3(t_3)\rangle\langle dN^2(t_2)\rangle \\ & - \langle dN^2(t_2)dN^3(t_3)\rangle\langle dN^1(t_1)\rangle \\ & + 2\langle dN^1(t_1)\rangle\langle dN^2(t_2)\rangle\langle dN^3(t_3)\rangle. \end{aligned} \qquad (23)$$

To alleviate the problem of increasing complexity, we use the cluster process representation to come up with a useful and intuitive expression for the density $k^{\mathbf{i}}(\mathbf{t})$ in terms of the cluster process's branching structure.

First off, note that the only way that events $(t_1,\cdots,t_n)$ [of types $(i_1,\cdots,i_n)$] can be statistically dependent is if they all belong to the same cluster, i.e., if they are all offspring (possibly of different generations) of a single original immigrant.

More specifically, we can show that (see Appendix A), for every multi-index $\mathbf{i} = (i_1,\cdots,i_n)$ and every vector $\mathbf{t} = (t_1,\cdots,t_n)$,

$$k^{\mathbf{i}}(\mathbf{t})d\mathbf{t} = P\{E^{\mathbf{i}}_{\mathbf{t}} \cap C^{\mathbf{i}}_{\mathbf{t}}\}, \qquad (24)$$

where

$$E^{\mathbf{i}}_{\mathbf{t}} = \{\forall k, \text{ there is a type } i_k \text{ event at time } t_k\}, \qquad (25)$$

$$C^{\mathbf{i}}_{\mathbf{t}} = \{\exists \text{ cluster } C \text{ such that}, \forall k,\ t_k \in C\}. \qquad (26)$$

This result now provides us with a practical way of computing $k^{\mathbf{i}}(\mathbf{t})d\mathbf{t}$.

For example, in the case when $n = 2$, we have that $k^{ij}(t_1,t_2)dt_1dt_2$ is equal to the probability of there being a type i event at time t_1, a type j event at time t_2, and that both of these events are descendant from a common immigrant. Therefore, in order to compute the 2nd order cumulant density, we need to sum up the probabilities of all possible "family trees," which contain events t_1 and t_2 (see Fig. 2).

In order to formalize this computation, we define

$$R^{ij}_t := \frac{P\{\text{type } j \text{ event at 0 causes type } i \text{ event at } t\}}{dt}. \qquad (27)$$

Then,

$$R^{ij}_t = \left[\sum_{n\geqslant 0} \mathbf{G}^{\star n}(t)\right]_{ij}, \qquad (28)$$

where $[\cdot]_{ij}$ extracts component (i,j) of a given matrix, and $\mathbf{G}^{\star n}(t)$ denotes the $\mathbf{n}$th convolution power of the interaction

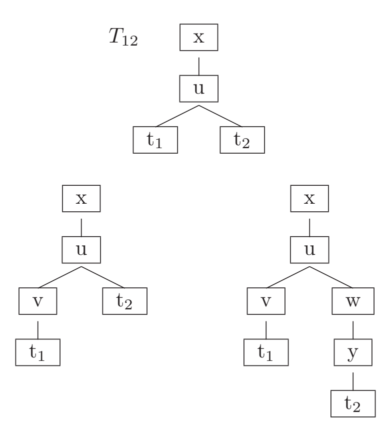

FIG. 2. A schematic representation of all possible family trees, containing events t_1 and t_2. Two concrete examples of family trees, mapping to the same general schema, are given in the second row. The root, x, denotes the original immigrant and u the branching point in the family tree of immigrant x, leading to the appearance of offspring t_1 and t_2. Note that each link connecting two nodes can be, in theory, any number of generations long.

kernel $\mathbf{G}(t)$, defined recursively by

$$\mathbf{G}^{\star 0}(t) = \mathbf{I}\delta(t),$$

$$\mathbf{G}^{\star n}(t) = \int_{-\infty}^{t} \mathbf{G}^{\star(n-1)}(t-s)\cdot\mathbf{G}(s)ds.$$

Indeed, if we define $p_n^{ij}(t)$ to be equal to the probability that an event of type j at 0, after n generations, causes a type i event at t, we have

$$\frac{p_0^{ij}(t)}{dt} = \delta_{ij}\delta(t) = [\mathbf{I}\delta(t)]_{ij}, \tag{29}$$

$$\frac{p_1^{ij}(t)}{dt} = [\mathbf{G}(t)]_{ij}, \tag{30}$$

$$\frac{p_2^{ij}(t)}{dt} = \sum_{k=1}^{d}\int_{-\infty}^{t}[\mathbf{G}(t-s)]_{ik}[\mathbf{G}(s)]_{kj}ds = [\mathbf{G}^{\star 2}(t)]_{ij}, \tag{31}$$

and therefore, by induction,

$$R_t^{ij} = [I\delta(t) + \mathbf{G}(t) + \mathbf{G}^{\star 2}(t) + \mathbf{G}^{\star 3}(t) + \cdots]_{ij} \tag{32}$$

$$= \left[\sum_{n\geqslant 0}\mathbf{G}^{\star n}(t)\right]_{ij}. \tag{33}$$

Furthermore, noting that $P\{\text{type } k \text{ immigrant at } x\}$ is, by construction, equal to $\mu^k dx$, we obtain the probability of an immigrant (arriving at any point in time) generating an event of type m at time u. It equals

$$\sum_{k=1}^{d}\int_{\mathbb{R}}\mu^k R_{u-x}^{mk}dx = \sum_{k=1}^{d}[(\mathbf{I}-\mathbf{G})^{-1}]_{mk}\mu^k = \lambda^m, \tag{34}$$

i.e., it is the mth component of the stationary rate vector $\boldsymbol{\lambda}$ in Eq. (14), where the first equality in the previous equation follows from

$$\int_{\mathbb{R}} R_{u-x}^{mk}dx = \sum_{n\geqslant 0}\int_{\mathbb{R}}[\mathbf{G}^{\star n}(u-x)]_{mk}dx \tag{35}$$

$$= \sum_{n\geqslant 0}[\mathbf{G}^n]_{mk} = [(\mathbf{I}-\mathbf{G})^{-1}]_{mk}. \tag{36}$$

Computing the probability of the family tree T_{12} in Fig. 2 is now straightforward; recalling the definition of R_t^{ij} and taking into account our previous considerations, we get

$$k^{ij}(t_1,t_2) = \frac{P(T_{12})}{dt_1dt_2} = \sum_{m=1}^{d}\lambda^m\int_{\mathbb{R}}R_{t_1-u}^{im}R_{t_2-u}^{jm}du, \tag{37}$$

recovering a classical and well-known result on the covariance density of the Hawkes process (see Ref. [26]).

A big advantage of our approach, however, is that it can be used to compute cumulant densities of orders greater than 2.

For example, in order to compute the 3rd order density $k^{ijk}(t_1,t_2,t_3)$, we start, as in the two-dimensional case, by enumerating all possible family trees with leaves t_1, t_2, and t_3. In this case, however, there are in total four different possibilities (see Fig. 3).

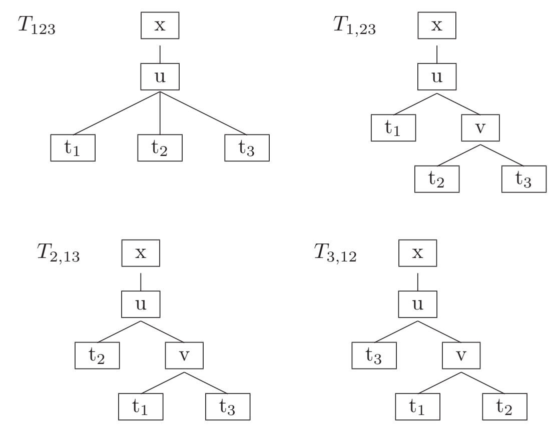

FIG. 3. A schematic representation of all possible family trees, containing events t_1, t_2, and t_3. Once again, the root x denotes the original immigrant, nodes u and v represent the branching points in the family tree of immigrant x, and each link between two nodes can be any number of generations long.

We can now proceed in much the same way as before, summing up the probabilities of all possible trees in order to derive the desired formula. We define $\mathbf{t} = (t_1,t_2,t_3)$, $d\mathbf{t} = dt_1dt_2dt_3$, and

$$\Psi_t^{ij} = R_t^{ij} - \delta_{ij}\delta(t) = \left[\sum_{n\geqslant 1}\mathbf{G}^{\star n}(t)\right]_{ij}, \tag{38}$$

finally obtaining

$$\begin{aligned} k^{ijk}(\mathbf{t}) &= \frac{P(T_{1,23})}{d\mathbf{t}} + \frac{P(T_{2,13})}{d\mathbf{t}} + \frac{P(T_{3,12})}{d\mathbf{t}} + \frac{P(T_{123})}{d\mathbf{t}} \\ &= \sum_{m,n=1}^{d}\lambda^n\int_{\mathbb{R}}R_{t_1-u}^{in}\left(\int_{\mathbb{R}}R_{t_2-v}^{jm}R_{t_3-v}^{km}\Psi_{v-u}^{mn}dv\right)du \\ &\quad + \sum_{m,n=1}^{d}\lambda^n\int_{\mathbb{R}}R_{t_2-u}^{jn}\left(\int_{\mathbb{R}}R_{t_1-v}^{im}R_{t_3-v}^{km}\Psi_{v-u}^{mn}dv\right)du \\ &\quad + \sum_{m,n=1}^{d}\lambda^n\int_{\mathbb{R}}R_{t_3-u}^{kn}\left(\int_{\mathbb{R}}R_{t_1-v}^{im}R_{t_2-v}^{jm}\Psi_{v-u}^{mn}dv\right)du \\ &\quad + \sum_{m=1}^{d}\lambda^m\int_{\mathbb{R}}R_{t_1-u}^{im}R_{t_2-u}^{jm}R_{t_3-u}^{km}du. \end{aligned} \tag{39}$$

It is important to point out that Eq. (39) can be derived in a different, albeit a more tedious way using martingale theory arguments, generalizing the derivation of Bacry *et al.* in Ref. [36] for the second-order cumulant density.

The newly introduced function Ψ_t^{ij} corresponds to the probability of a type j event at 0 generating a type i event at t, after at least one generation.

The appearance of such a term in the above equations is a consequence of the fact that, for instance, contracting the link between nodes u and v in tree $T_{1,23}$ to a point turns it into T_{123}, which is already accounted for. Thus, in order to avoid counting certain configurations twice, we must introduce a "stiff" link between the two internal nodes u and v in trees $T_{1,23}$, $T_{2,13}$, and $T_{3,12}$.

TABLE I. Number of terms in $k^{\mathbf{i}}(\mathbf{t})$ for a given n (from Ref. [37]).

n	Terms in $k^{\mathbf{i}}(\mathbf{t})$
2	1
3	4
4	26
5	236
6	2,752
7	39,208
8	660,302
9	12,818,912
10	282,137,824

By generalizing the above considerations, it is possible to construct a general procedure for computing the nth order cumulant density $k^{i_1\cdots i_n}(t_1,\cdots,t_n)$.

(1) For a given $n \geqslant 2$, generate all possible rooted trees T with n leaves.

(2) Label the leaves of T with ordered pairs (i_k,t_k), in arbitrary order. Label the internal nodes (including the root) of T arbitrarily.

(3) For every tree T, construct an integral term I_T, according the the following pseudo-algorithm:

(a) Set $I_T \leftarrow 1$;

(b) For every edge in T, connecting a node v of type j_v to a leaf t_k of type i_k:

$$I_T \leftarrow I_T \cdot R^{i_k j_v}_{t_k - v} dv;$$

(c) For every edge in T, connecting an internal node u of type j_u to another internal node v of type j_v:

$$I_T \leftarrow I_T \cdot \Psi^{j_v j_u}_{v-u} du;$$

(d) Let x be the root of T. Set

$$I_T \leftarrow I_T \cdot \lambda^{j_x};$$

(e) Integrate I_T with respect to the variable du, for every internal node u.

(f) Sum over all j_u for every internal node u.

(g) Sum over all j_x

(4) Add up all integral terms I_T for every rooted tree T, generated in the first step, to obtain the nth order cumulant density.

The principal difficulty of the above procedure lies its first step, i.e., in the enumeration of all topologically distinct rooted trees with n labeled leaves. While there are known algorithms that can tackle this problem (see, e.g., the classic text by Felsenstein [37]), the number of terms grows very quickly with increasing n (see Table I) and thus computing $k^{\mathbf{i}}(\mathbf{t})$ quickly becomes impractical.

IV. INTEGRATED CUMULANTS AS SUMS OF TOPOLOGICAL MOTIFS

Let $k^{\mathbf{i}}(\mathbf{t})$ be, for a given time vector $\mathbf{t} = (t_1,\cdots,t_n)$ and multi-index $\mathbf{i} = (i_1,\cdots,i_n)$, the nth-order cumulant density of a d-dimensional Hawkes process. We define the integrated cumulant of order $\mathbf{n}$, denoted simply by $k^{\mathbf{i}}$, by setting

$$k^{\mathbf{i}} := \int_{\mathbb{R}^n_+} k^{\mathbf{i}}(\mathbf{t}) d\mathbf{t}. \tag{40}$$

Note that $k^{\mathbf{i}}$ can be seen as the n-dimensional Laplace transform, "at zero," of $k^{\mathbf{i}}(\mathbf{t})$. Indeed, if we denote

$$\mathcal{L}_{\boldsymbol{\omega}}[k^{\mathbf{i}}(\mathbf{t})] = \int_{\mathbb{R}^n_+} e^{-\boldsymbol{\omega}\cdot\mathbf{t}} k^{\mathbf{i}}(\mathbf{t}) d\mathbf{t}, \tag{41}$$

where $\boldsymbol{\omega} = (\omega_1,\cdots,\omega_n) \in \mathbb{C}^n$ and $\boldsymbol{\omega}\cdot\mathbf{t} = \sum_i \omega_i t_i$, we have, clearly,

$$k^{\mathbf{i}} = \mathcal{L}_{\mathbf{0}}[k^{\mathbf{i}}(\mathbf{t})]. \tag{42}$$

Thus, if we define

$$\mathbf{R}_t := \begin{pmatrix} R^{11}_t & \cdots & R^{1d}_t \\ \vdots & \ddots & \vdots \\ R^{d1}_t & \cdots & R^{dd}_t \end{pmatrix}, \qquad \boldsymbol{\Psi}_t = \mathbf{R}_t - \mathbf{I}\delta(t), \tag{43}$$

we can, by Laplace transforming the covariance density $k^{ij}(t_1,t_2)$, prove (see Appendix B) that,

$$k^{ij} = \sum_{m=1}^{d} \lambda^m [\mathbf{R}]_{im} [\mathbf{R}]_{jm}, \tag{44}$$

where we set

$$\mathbf{R} = (\mathbf{I} - \mathbf{G})^{-1} = \mathcal{L}_0(\mathbf{R}_t). \tag{45}$$

Expanding $\mathbf{R}$ in powers of $\mathbf{G}$, we get

$$k^{ij} = \sum_{m=1}^{d} \sum_{k=0}^{+\infty} \sum_{l=0}^{+\infty} \lambda^m [\mathbf{G}^k]_{im} [\mathbf{G}^l]_{jm}. \tag{46}$$

Interpreting now the matrix power $\mathbf{G}^l$ in the sense of graph theory, i.e., as a matrix whose component (i,j) corresponds to the sum of lengths of all paths from node j to node i in exactly l steps, we see that the integrated covariance density k^{ij} can be equivalently represented as

$$k^{ij} = \sum_{T \in \mathcal{T}^m_{ij}} w(T), \tag{47}$$

where the sum goes over the set $\mathcal{T}^m_{ij}$ of all rooted trees T with root m, containing nodes i,j. Here, $w(T)$ denotes the weight of tree T, defined as the product of weights of all edges, contained in T, times the weight of the root m, defined as being equal to λ^m.

The graph H with adjacency matrix $\mathbf{G}$ can be thought of as follows. Each node $i \in \{1,\cdots,n\}$ in H corresponds to a type of event in the underlying Hawkes process, and the existence of an edge e_{ij} from j to i indicates the possibility of generating type i events from those of type j. Starting in node j, traversing the corresponding edge to reach node i is equivalent to generating g^{ij} type i offspring of a type j immigrant. Therefore, each path through graph H represents a specific "bloodline" of a type m immigrant, while a tree $T \in \mathcal{T}^m_{ij}$ accounts for the possibility of the bloodline splitting somewhere along the way, concluding in, after a certain number of generations, in offspring of both types i and j. The previous formula tells us that the sum of weights of all such trees is equal to the integrated covariance k^{ij}.

Now, reasoning in much the same way as before we have, for k^{ijk},

$$
\begin{aligned}
k^{ijk} = & \sum_{m=1}^{d} \lambda_m [\mathbf{R}]_{im} [\mathbf{R}]_{jm} [\mathbf{R}]_{km} \\
& + \sum_{m,n=1}^{d} \lambda_n [\mathbf{R}]_{im} [\mathbf{R}]_{jm} [\mathbf{\Psi}]_{mn} [\mathbf{R}]_{kn} \\
& + \sum_{m,n=1}^{d} \lambda_n [\mathbf{R}]_{jm} [\mathbf{R}]_{km} [\mathbf{\Psi}]_{mn} [\mathbf{R}]_{in} \\
& + \sum_{m,n=1}^{d} \lambda_n [\mathbf{R}]_{im} [\mathbf{R}]_{km} [\mathbf{\Psi}]_{mn} [\mathbf{R}]_{jn},
\end{aligned} \tag{48}
$$

where $\mathbf{\Psi} = (\mathbf{I} - \mathbf{G})^{-1} - \mathbf{I} = \mathcal{L}_0(\mathbf{\Psi}_t)$. Once again, expanding $\mathbf{R}$ and $\mathbf{\Psi}$ in powers of $\mathbf{G}$ yields

$$k^{ijk} = \sum_{T \in \mathcal{T}_{ijk}^m} w(T), \tag{49}$$

where $\mathcal{T}_{ijk}^m$ is the set of all rooted trees with root m, containing nodes i,j,k, and $w(\cdot)$ is the already defined weight function.

It is now easy to see that the general result is of the form

$$k^{\mathbf{i}} = \sum_{T \in \mathcal{T}_{\mathbf{i}}^m} w(T), \tag{50}$$

where $\mathbf{i} = (i_1, \cdots, i_n)$ and $\mathcal{T}_{\mathbf{i}}^m = \mathcal{T}_{i_1 \cdots i_n}^m$ is the set of all rooted trees with root m, containing nodes $i_1, \cdots, i_n$.

V. DISCUSSION

In this paper we described the method for computing a class of statistics of linear Hawkes self-exciting point processes with arbitrary interaction kernels. By using the Poisson cluster process representation, we were able to obtain a general procedure for deriving formulas for nth order cumulant densities. Furthermore, we have shown there is a one-to-one correspondence between the integral terms, appearing in said densities, and all topologically distinct rooted trees with n labeled leaves.

We also considered the problem of computing time-integrated cumulants and showed this can be done by simplifying the expressions for the corresponding cumulant densities. Moreover, and not surprisingly, we demonstrated that integrated cumulants likewise admit a representation in terms of a formal sum of topological motifs, generalizing previous work on the topological expansion of the integrated covariance [16].

The problem of quantifying higher-order correlations is of some importance in theoretical neuroscience. Indeed, it has long been suggested [38,39] that understanding the cooperative dynamics of populations of neurons would provide fundamental insight into the nature of neuronal computation. However, while direct experimental evidence for coordinated activity on the spike train level mostly relies on the correlations between pairs of nerve cells [40–44], it is becoming increasingly clear that such pairwise correlations cannot completely resolve the cooperative dynamics of neuronal populations [31,45–47] and that higher-order cumulants need to be taken into account.

One possible shortcoming of our work is the (supraexponentially) increasing complexity of the closed-form expressions for the densities $k^{\mathbf{i}}(\mathbf{t})$ for higher values of n. This "explosion," however, is mostly due to combinatorial factors that arise in many problems involving cumulants. As their definition naturally involves objects such as set partitions, it seems to us that these sorts of issues would be quite difficult to avoid.

Another limitation of the present model is that it only allows for excitatory interactions—an arrival of an event at a given time can only *increase* the likelihood of future event, never decrease it. We hope, in the future, to be able to extend our analysis to include models in which there also exists a possibility of mutual inhibition between points of different types.

Further generalizations of our results might involve computing cumulants (and other important statistics) of a nonlinear Hawkes processes (see, e.g., Ref. [22] for the definition), whose conditional rate function involves a nonlinear transformation of Eq. (6), thus allowing for, for example, multiplicative interaction between point events [48]. However, in this case, the resulting process no longer admits an immigrant-offspring representation, meaning an alternative approach would be necessary.

ACKNOWLEDGMENTS

Supported by the Erasmus Mundus Joint Doctoral programme EuroSPIN and the German Federal Ministry of Education and Research (BFNT-Freiburg*Tübingen, Grant No. 01GQ0830). Stojan Jovanović acknowledges the hospitality of NORDITA. The authors also thank the referees for many helpful comments and suggestions, Marcel Sauerbier for useful discussion, and Gunnar Grah for making Fig. 1.

APPENDIX A: PROOF OF EQUATION (24)

Let $\mathbf{t}$ be an arbitrary time vector $\mathbf{t} = (t_1, \cdots, t_n)$ and $\mathbf{i}$ an arbitrary multi-index $\mathbf{i} = (i_1, \cdots, i_n)$.

From Eq. (5), for every vector $\mathbf{t} = (t_1, \cdots, t_n)$ and multi-index $\mathbf{i} = (i_1, \cdots, i_n)$, we have that

$$\langle dN^{i_1}(t_1) \cdots dN^{i_n}(t_n) \rangle = P\{E_{\mathbf{t}}^{\mathbf{i}}\}, \tag{A1}$$

$$P\{E_{\mathbf{t}}^{\mathbf{i}}\} = P\{\forall k, \text{there is a type } i_k \text{ event at } t_k\}. \tag{A2}$$

Furthermore, it is clear that

$$P\{E_{\mathbf{t}}^{\mathbf{i}}\} = P\{E_{\mathbf{t}}^{\mathbf{i}} \cap C_{\mathbf{t}}^{\mathbf{i}}\} + P\{E_{\mathbf{t}}^{\mathbf{i}} \cap \bar{C}_{\mathbf{t}}^{\mathbf{i}}\}, \tag{A3}$$

where $\bar{C}_{\mathbf{t}}^{\mathbf{i}}$ denotes the complement of the set

$$C_{\mathbf{t}}^{\mathbf{i}} = \{\exists \text{ cluster } C \text{ such that}, \forall k,\, t_k \in C\}. \tag{A4}$$

Indeed, events $\mathbf{t}$ of type $\mathbf{i}$ either are or aren't all in some cluster C. We now proceed by induction in n. For $n = 2$, we have

$$P\{E^{ij}_{t_1t_2}\} = P\{E^{ij}_{t_1t_2} \cap C^{ij}_{t_1t_2}\} + P\{E^{ij}_{t_1t_2} \cap \bar{C}^{ij}_{t_1t_2}\}. \tag{A5}$$

But, as the only way that two events are not in the same cluster is if they each belong to a different one; say, if $t_1 \in C$ and $t_2 \in D$,

$$P\{E^{ij}_{t_1t_2} \cap \bar{C}^{ij}_{t_1t_2}\} = \sum_{C,D} P\{E^i_{t_1} \cap E^j_{t_2} \cap C^i_{t_1} \cap D^j_{t_2}\} \tag{A6}$$

$$= \sum_{C,D} P\{E^i_{t_1} \cap C^i_{t_1}\} P\{E^j_{t_2} \cap D^j_{t_2}\} \tag{A7}$$

$$= P\{E^i_{t_1}\} P\{E^j_{t_2}\} = \langle dN^i(t_1)\rangle\langle dN^j(t_2)\rangle, \tag{A8}$$

because of independence of different clusters C and D. Thus,

$$P\{E^{ij}_{t_1t_2} \cap C^{ij}_{t_1t_2}\} = P\{E^{ij}_{t_1t_2}\} - \langle dN^i(t_1)\rangle\langle dN^j(t_2)\rangle \tag{A9}$$

$$= k^{ij}(t_1,t_2)dt_1dt_2, \tag{A10}$$

proving that Eq. (24) is true for $n = 2$.

Next, we assume that Eq. (24) is true for $n - 1$ and prove that it then must also be true for n.

Consider the complementary set $\bar{C}^{\mathbf{i}}_{\mathbf{t}}$. If events $\mathbf{t}$ are not all in the same cluster, how could they be distributed? One possibility is that they are divided up between two different clusters, like in the previous case. In fact, they could potentially be distributed in c different clusters, where $2 \leqslant c \leqslant n$. Therefore,

$$P\{E^{\mathbf{i}}_{\mathbf{t}} \cap \bar{C}^{\mathbf{i}}_{\mathbf{t}}\} = \sum_{c=2}^{n} \prod_{r=1}^{c} P\{E^{\mathbf{i}_r}_{\mathbf{t}_r} \cap B(r)^{\mathbf{i}_r}_{\mathbf{t}_r}\}, \tag{A11}$$

where the first sum goes over all possible numbers of different clusters that events $\mathbf{t}$ could be partitioned in, while $\mathbf{t}_r$ denotes the subset of $\mathbf{t}$ that belong to the rth cluster $B(r)$ (and $\mathbf{i}_r$ denotes their types).

Now, note that the previous equation is, in fact, a sum over all partitions π of the set $\{1,\cdots,n\}$ with at least two blocks (i.e., $|\pi| > 1$). Let us now fix one such partition $\pi = \{B(1),\cdots,B(|\pi|)\}$. As $|\pi| > 1$, we must have, $\forall r$, $1 \leqslant r \leqslant |\pi|$, that $|B(r)| < n$. But then, by the inductive assumption,

$$P\{E^{\mathbf{i}_r}_{\mathbf{t}_r} \cap B(r)^{\mathbf{i}_r}_{\mathbf{t}_r}\} = k^{\mathbf{i}_r}(\mathbf{t}_r), \tag{A12}$$

and, therefore,

$$P\{E^{\mathbf{i}}_{\mathbf{t}} \cap \bar{C}^{\mathbf{i}}_{\mathbf{t}}\} = \sum_{\pi:|\pi|>1} \prod_{B\in\pi} k^{\mathbf{i}_r}(\mathbf{t}_r)d\mathbf{t}_r. \tag{A13}$$

Finally, from Eqs. (A3), (A13), and (18), we get

$$k^{\mathbf{i}}_{\mathbf{t}}d\mathbf{t} + \sum_{\pi:|\pi|>1} \prod_{B\in\pi} k^{\mathbf{i}_r}(\mathbf{t}_r)d\mathbf{t}_r = P\{E^{\mathbf{i}}_{\mathbf{t}} \cap C^{\mathbf{i}}_{\mathbf{t}}\} + P\{E^{\mathbf{i}}_{\mathbf{t}} \cap \bar{C}^{\mathbf{i}}_{\mathbf{t}}\}, \tag{A14}$$

which completes the proof.

APPENDIX B: FORMULAS FOR INTEGRATED CUMULANTS

Let $\mathcal{T}^m_{\mathbf{i}}$ be the set of all rooted trees with root m and leaves $\mathbf{i} = (i_1,\cdots,i_n)$. Next, let $T \in \mathcal{T}^m_{\mathbf{i}}$ and let I_T be the corresponding integral term. In order to compute the Laplace transform $\mathcal{L}_{\boldsymbol{\omega}}(I_T)$, we first consider the leaves of T.

Each leaf i_k contributes a term $R^{i_k j_v}_{t_k - v}$, for some internal node v. For simplicity, let us assume that leaves $i_1,\cdots,i_{s(v)}$ all descend from a single internal node, which we denote v. Then, applying to I_T the Laplace transform with respect to variables $t_1,\cdots,t_{s(v)}$, we obtain

$$\prod_{l=1}^{s(v)} \mathcal{L}_{\omega_l}\left(R^{i_l j_v}_{t_l - v}\right)e^{-v\sigma_v}, \tag{B1}$$

where we denote $\sigma_v = \sum_{l=1}^{s(v)} \omega_l$.

Of course, in general the leaves $i_1,\cdots,i_n$ are divided into several groups, according to which internal node they descend from. In that case, applying to each such group the Laplace transform in the already described way, yields several terms of Eq. (B1).

Moving one level up in tree T, we are now in a situation in which several internal nodes, each with its own group of dependent leaves, all descend from a common node u, residing one level above them. We denote these internal nodes by $v_1,\cdots,v_{r(u)}$. Each such internal node v_l contributes to I_T a term $\boldsymbol{\Psi}^{j_{v_l} j_u}_{v_l - u}$. Transforming the exponential term in Eq. (B1),

$$e^{-v_l\sigma_{v_l}} = e^{-(v_l - u)\sigma_{v_l}} e^{-u\sigma_{v_l}}, \tag{B2}$$

and multiplying with $\boldsymbol{\Psi}^{j_{v_l} j_u}_{v_l - u}$, we get

$$\prod_{l=1}^{r(u)} \mathcal{L}_{\sigma_{v_l}}\left(R^{j_{v_l} j_u}_{v_l - u}\right)e^{-u\Sigma_u}, \tag{B3}$$

where $\Sigma_u = \sum_{l=1}^{r(u)} \sigma_{v_l}$.

By induction, we can now see that this procedure must end after a finite number of steps (equal to the “depth” of tree T), at which point we are left with a product of various terms of Eqs. (B1) and (B3), integrated with respect to the position x of the root m (as this is the last node we reach by “climbing up” T). The exponential terms in this product can be combined to form

$$\int_{\mathbb{R}} e^{-x\sum_{i=1}^n \omega_i} dx = \delta(\omega_1 + \cdots + \omega_n), \tag{B4}$$

the integral representation of a Dirac δ function.

By setting $\boldsymbol{\omega} = \mathbf{0}$, we now see that the formulas for $k^{\mathbf{i}}$ can be obtained from formulas for the cumulant densities by simply “erasing” all the integral signs and replacing all the functional terms with their integrated counterparts.

[1] A. G. Hawkes, Biometrika **58**, 83 (1971).
[2] Y. Ogata, J. Am. Stat. Assoc. **83**, 9 (1988).
[3] D. Vere-Jones, J. Roy. Stat. Soc. Ser. B (Methodological) **32**, 1 (1970).
[4] T. Pechácek, V. Karas, and B. Czerny, Astron. Astrophys. **487**, 815 (2008).
[5] P. Reynaud-Bouret and S. Schbath, Ann. Stat. **38**, 2781 (2010).

[6] L. Mitchell and M. E. Cates, J. Phys. A: Math. Theor. **43**, 045101 (2010).
[7] D. Karabash and L. Zhu, arXiv:1211.4039v2.
[8] L. Zhu, Insur. Math. Econ. **53**, 544 (2013), arXiv:1304.1940v2.
[9] S. Azizpour, K. Giesecke, and G. Schwenkler, Social Science Research Network (2014), http://dx.doi.org/10.2139/ssrn.1127792.
[10] E. Errais, K. Giesecke, and L. R. Goldberg, SIAM J. Finan. Math. **1**, 642 (2010).
[11] E. Bacry and J. Muzy, arXiv:1401.0903v1.
[12] L. Bauwens and N. Hautsch, *Modelling Financial High-Frequency Data Using Point Processes* (Springer, Berlin, 2009).
[13] E. Bacry, S. Delattre, H. Marc, and J.-F. Muzy, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, New York, 2011), pp. 5740–5743.
[14] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, J. Am. Stat. Assoc. **106**, 100 (2011).
[15] I. Mastromatteo and M. Marsili, J. Stat. Mech.: Theory Exp. (2011) P10012.
[16] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter, PLoS Comput. Biol. **7**, e1002059 (2011).
[17] T. Onaga and S. Shinomoto, Phys. Rev. E **89**, 042817 (2014).
[18] A. Saichev and D. Sornette, Phys. Rev. E **89**, 012104 (2014).
[19] A. Saichev and D. Sornette, Phys. Rev. E **87**, 022815 (2013).
[20] S. J. Hardiman and J.-P. Bouchaud, Phys. Rev. E **90**, 062807 (2014).
[21] A. Dassios and H. Zhao, Adv. Appl. Probab. **43**, 814 (2011).
[22] P. Brémaud and L. Massoulié, Ann. Prob. **24**, 1563 (1996).
[23] L. Zhu, J. Appl. Prob. **50**, 760 (2013).
[24] L. Zhu, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 50 (Institut Henri Poincaré, Paris, 2014), pp. 845–871.
[25] M. S. Bartlett, J. Roy. Stat. Soc. Ser. B (Methodological) **25**, 264 (1963).
[26] A. G. Hawkes, J. Roy. Stat. Soc. Ser. B (Methodological) **33**, 438 (1971).
[27] L. Adamopoulos, J. Appl. Prob. **12**, 78 (1975).
[28] A. Saichev, T. Maillart, and D. Sornette, Eur. Phys. J. B **86**, 124 (2013).
[29] A. Saichev and D. Sornette, Eur. Phys. J. B-Cond. Matter Complex Syst. **83**, 271 (2011).
[30] C. Rossant, S. Leijon, A. K. Magnusson, and R. Brette, J. Neurosci. **31**, 17193 (2011).
[31] A. Kuhn, A. Aertsen, and S. Rotter, Neural Comput. **15**, 67 (2003).
[32] A. Hawkes and D. Oakes, J. Appl. Prob. **11**, 493 (1974).
[33] D. R. Cox and V. Isham, *Point Processes*, Vol. 12 (CRC Press, Boca Raton, 1980).
[34] J. Rasmussen, Method. Comput. Appl. Prob. **15**, 623 (2013).
[35] E. Lukacs, *Characteristic Functions* (Hafner Publishing Company, New York, 1970).
[36] E. Bacry, K. Dayri, and J.-F. Muzy, Eur. Phys. J. B-Cond. Matter Complex Syst. **85**, 1 (2012).
[37] J. Felsenstein, *Inferring Phylogenies* (Sinauer Associates, Sunderland, MA, 2004).
[38] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Wiley, New York, 1949).
[39] G. Gerstein, P. Bedenbaugh, and A. M. Aertsen, IEEE Trans. Biomed. Eng. **36**, 4 (1989).
[40] C. M. Gray and W. Singer, Proc. Natl. Acad. Sci. USA **86**, 1698 (1989).
[41] E. Vaadia, I. Haalman, M. Abeles, H. Bergman, Y. Prut, H. Slovin, and A. Aertsen, Nature **373**, 515 (1995).
[42] A. Riehle, S. Grün, M. Diesmann, and A. Aertsen, Science **278**, 1950 (1997).
[43] W. Bair, E. Zohary, and W. T. Newsome, J. Neurosci. **21**, 1676 (2001).
[44] A. Kohn and M. A. Smith, J. Neurosci. **25**, 3661 (2005).
[45] L. Martignon, H. Von Hassein, S. Grün, A. Aertsen, and G. Palm, Biol. Cybern. **73**, 69 (1995).
[46] S. M. Bohté, H. Spekreijse, and P. R. Roelfsema, Neural Comput. **12**, 153 (2000).
[47] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, Nature **466**, 617 (2010).
[48] S. Cardanobile and S. Rotter, J. Comput. Neurosci. **28**, 267 (2009).