CONVERGENCE OF NONHOMOGENEOUS HAWKES PROCESSES AND FELLER RANDOM MEASURES

TRISTAN PACE AND GORDAN ŽITKOVIĆ

ABSTRACT. We consider a sequence of Hawkes processes whose excitation measures may depend on the generation, and study its scaling limits in the near-unstable limiting regime. The limiting random measures, characterized via a nonlinear convolutional equation, form a family parameterized by a pair consisting of a locally finite measure and a geometrically infinitely divisible probability distribution on the positive real line. These measures can be interpreted as generalizations of the Feller diffusion and fractional Feller (CIR) processes, but also allow for the "driving noise" associated to general Lévy-type operators of order at most 1, including fractional derivatives of any order $\alpha > 0$ (formally corresponding to possibly negative Hurst parameters).

1. INTRODUCTION

Hawkes (point) processes were introduced in [Haw71b, Haw71a] as models for self-exciting stochastic phenomena. Their fundamental property is that new points are generated at the rate that depends on the number and locations of existing points via a function known as the excitation kernel. Initially used as models for seismic events, Hawkes processes have since found numerous applications in various disciplines ranging from epidemiology and criminology, over genetics and neuroscience to economics and finance (see the survey [LLPT24] and its references).

1.1. **Limiting theory of Hawkes processes - an overview of the literature.** The investigation into the limiting theory of Hawkes processes began almost immediately after their introduction. A central limit theorem (as $t \to \infty$) for Hawkes processes whose kernels admit a finite first moment was established already in [HO74] (this paper also introduced the cluster representation we use in the current paper). We start with a brief survey of existing pertinent results split into two classes, based on the scaling regime.

In the first class, the total mass $a = \int_0^\infty \phi(t)\, dt$ of the kernel ϕ is kept constant, while time, space and other parameters are scaled. One of the earliest results here was provided by [BDHM13], where a functional central limit theorem (FCLT) with convergence towards a scaled Brownian motion was established under a finiteness assumption on the $1/2$-th moment of the kernel. Later, [GZ18] introduced a framework where the background (immigrant) intensity is taken to infinity, but only space is scaled to compensate. Under the assumption that the kernel is exponential, it is shown there that the limiting process is no longer Brownian but only Gaussian with a non-Markovian

2020 *Mathematics Subject Classification.* 60F17, 60G55, 60G57, 45D05.

Key words and phrases. Hawkes processes, functional central limit theorem, nearly unstable, convolutional Riccati equation, fractional CIR processes, geometric infinite divisibility, Feller random measures.

During the preparation of this work both authors were supported by the National Science Foundation under Grant DMS-2307729. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF). .

covariance function. More recently, a FCLT for marked Hawkes processes and associated shot noises was established in [HX21]. We also mention a recent preprint [HX24] by the same authors where FCLTs or parallel negative results are established for Hawkes processes in several stability regions defined by the values of the total mass and the first moment of the kernel ϕ.

The second class of results features the "nearly unstable" scaling regime, introduced in [JR15], which is also utilized in the present work. In this regime, both time and space are scaled in a non-Brownian manner, while the total mass of the kernel is sent to 1 - the stability threshold constant. Assuming that the kernel has a finite first moment, these authors establish a functional scaling limit theorem for the integrated intensity process with the Feller (CIR) diffusion, a non-Gaussian process, as the limit. In the follow-up paper [JR16b], the requirement for a finite first moment is relaxed to the finiteness of some moment above $1/2$, and the limiting Feller diffusion is replaced by a fractional Feller (CIR) process, which is neither Markovian nor a semimartingale. The mode of convergence obtained in [JR16b] was considerably strengthened in [HXZ23] under the same assumptions on the kernel. These authors show that the intensity processes themselves converge in the Skorokhod topology, and not only their integrals, as in [JR16b].

1.2. **Our contributions.** The goal of this paper is to add to the existing literature by extending the aforementioned results in several directions. Firstly, we consider nonhomogeneous Hawkes processes, i.e., the generalizations of Hawkes processes where the kernel is allowed to vary from generation to generation (see [FLM15] for a related model). This not only provides additional modeling flexibility, but also unlocks a wider range of possible limiting objects. Additionally, we allow the kernels themselves to serve as scaling parameters in that they may depend on the scaling parameter n. Compared to the existing results, these two extensions can be though of as a transition from scaled sums of iid sequences to sums of (triangular) arrays of independent random variables in classical probability theory. Continuing this analogy, the class of our limiting objects now includes not only the analogues of stable distributions (fractional Feller (CIR) processes), but also the analogues of infinitely-divisible distributions (termed Feller random measures in this paper).

Another direction in which we broaden all existing results is that we do not impose any conditions on the integrability of the kernels; we do not even require them to be functions in $\mathbb{L}^1([0,\infty))$ but permit them to be general finite measures on $[0,\infty)$. This allows us, in particular, to expand the analysis of [JR15, JR16b] down to and below the critical $1/2$-moment threshold imposed in the existing literature. In this regime, the limiting objects are no longer necessarily (integrals of) stochastic processes; they can now be located in the space of nonnegative random measures. Consequently, we cannot talk about convergence in Skorohod's J_1, or any related topology, but need to work with the vague topology this space is naturally endowed with.

A novel difficulty encountered in our approach is that the tools of stochastic analysis and martingale theory, standard in the Hawkes-process literature, seem to lose much of their usefulness. This is largely due to the appearance of genuine random measures as limits, but is also exacerbated by the nonhomogeneity of our model. This dependency makes it challenging to express the conditional intensity process in a convenient form without sacrificing finite dimensionality. Consequently, we are led to the cluster representation of the Hawkes process and the related cascade of relationships among the Laplace functionals associated to a sequence of auxiliary point processes. Here, the process indexed by m represents the progeny of an individual of generation m. The crux of the argument then rests on obtaining tight coupling estimates for pairs of such processes. This results in an convergence theorem which provides scaling constants, gives sufficient conditions under which the scaling limit exists, and characterizes its Laplace functional in terms of the unique solution to a nonlinear convolutional Riccati equation.

With the convergence theorem established, we turn to its conditions in the second part of the paper. There we use the theory of of random summation (see the monograph [GK96]) to give a detailed characterization of the possible limiting random measures and fairly explicit conditions on the arrays of excitation kernels that achieve them. It turns out that the limits are completely described by two measure-valued parameters: a locally finite measure μ and a geometrically infinitely-divisible probability distribution ρ on $[0, \infty)$. We call them Feller random measures because, as shown in [JR15] and [JR16a], their densities are given by the Feller diffusion when ρ is the exponential distribution and fractional Feller process when ρ is the Mittag-Leffler distribution. We go on to show that these random measures admit interesting distributional properties, like infinite divisibility, and allow for simple recursive formulas for the cumulants (and, therefore, moments). We also observe that Feller random measures can be used to produce stochastic representations for a class of Riccati equations where the classical derivative is replaced by a general Lévy-type differential operator of at-most first order.

1.3. **Connections with fractional Brownian motion and rough volatility models.** One of the motivations for this work comes from the role Hawkes processes have in financial modeling. Their self-exciting nature is particularly well-suited for capturing the dependence of market buy and sell orders on past orders (see, e.g., [BMM15] for an overview). A phenomenon well-explained by such modeling is the observed "roughness" (see [CR98] and [GJR18]) of market volatility. Indeed, the fractional Feller (CIR) process that appears in the results of [JR16b] - and corresponds to the squared volatility - can be informally thought of as a continuous stochastic process "driven" by the fractional Brownian motion (fBM). The value of the Hurst parameter $H \in (0, 1)$ of this fBM, used to describe the degree of "roughness" of the volatility process, has been the subject of several empirical studies. Early estimates gave $H \in (1/2, 1)$ ([CR98]) whereas two decades later the consensus shifted towards $H \in (0, 1/2)$ (see [GJR18], [BLP21] and [FTW22]). Many of the latter estimates put H very close to 0, suggesting that $H = 0$ might be the "true" value (see [FFGS22], [BHP21]).

Even though there is no universally accepted way to define the fractional Brownian motion with $H = 0$ either as a stochastic process or as a random measure/field, several authors have proposed models that could play such a role in one sense or another. These include the multifractal random walks (see [BDM01]) and various Gaussian random fields with a logarithmic kernel (see [FKS16], [NR18], and [HN22] for a sample of different approaches). Our framework allows not only to define generalized fractional Feller (CIR) processes corresponding to values of the Hurst parameter H in the interval $(-1/2, 1/2]$, but also corresponding to a much wider range of driving noises beyond the one-dimensional fractional family. Moreover, we only require a single passage to the limit, and do not define a limiting process for $H > 0$ first, and then pass it to a (second) limit $H \to 0$, as is often done in the literature mentioned above. While a full analysis is left for future research, and it is difficult to give a formal definition of the notion of a driving noise for random measures, we do note that the form of the covariance kernel we obtain in subsection 4.3.3 below suggests the log-correlated class (see the survey [DRSV17]).

1.4. **Organization of the paper.** Following this introduction, Section 2 provides the necessary background, established the notation and defines nonhomogeneous Hawkes. Section 3 contains the statement and the proof of the convergence theorem, while Section 4 delves into various properties of the limiting Feller random measures.

1.5. **Notation and conventions.** For a constant $c \in \mathbb{R}$ and a function $f : D \subseteq \mathbb{R}^d \to \mathbb{R}$, we write $f \ll c$ if there exists $\varepsilon > 0$ such that $f \le c - \varepsilon$.

Let D be a subset of a Euclidean space, let $\mathcal{B}(D)$ denote the Borel σ-algebra on D and let Leb denote the Lebesgue measure. Unless otherwise specified, measurability will always be understood with respect to $\mathcal{B}(D)$. The family of all measurable functions on D is denoted by $\mathcal{L}^0(D)$, while $\mathcal{L}^1(D)$ denotes the standard Lebesgue space with respect to the Lebesgue measure on D, except that we do not pass to Leb-a.e.-equivalence classes and $|f|_{\mathcal{L}^1(D)} := \int_D |f(x)|\, dx$ is only a seminorm. We let $\mathcal{S}^\infty(D)$ be the family of all bounded functions in $\mathcal{L}^0(D)$ and let $|f|_{\mathcal{S}^\infty_{\text{loc}}(D)} := \sup_{x \in D} |f(x)|$. The space $\mathcal{L}^1_{\text{loc}}(D)$ (respectively $\mathcal{S}^\infty_{\text{loc}}(D)$) consists of all $f \in \mathcal{L}^0(D)$ such that $f \in \mathcal{L}^1(D \cap B)$ (respectively $\mathcal{S}^\infty_{\text{loc}}(D \cap B)$ for all bounded $B \in \mathcal{B}(D)$. For $\{f_n\}_{n \in \mathbb{N}}, f \in \mathcal{L}^1(D)$, we write $f_n \to f$ in $\mathcal{L}^1_{\text{loc}}(D)$ if $|f_n - f|_{\mathcal{L}^1(D \cap B)} \to 0$ for each bounded $B \in \mathcal{B}(D)$.

$C_{\text{b}}(D)$ and $C_{\text{c}}(D)$ denote the families of bounded and compactly supported continuous functions on D. When $D = [0, \infty)$, $C_0 = C_0([0, \infty))$ denotes the family of continuous function f on $[0, \infty)$ with $f(0) = 0$.

$\mathcal{M}(D)$ and $\mathcal{M}_{\text{s}}(D)$ denote the sets of all positive and signed Borel measures D, respectively, and $\delta_{\{a\}}$ denotes the Dirac measure on $a \in D$. The sets of finite, finite on bounded sets, and probability measures on D are denoted by $\mathcal{M}_{\text{f}}(D)$, $\mathcal{M}_{\text{lf}}(D)$, and $\mathcal{M}_{\text{p}}(D)$, respectively. The total mass of $\mu \in \mathcal{M}(D)$ is denoted by $|\mu|$ and the total variation of $\mu \in \mathcal{M}_{\text{s}}(D)$ by $|\mu|_{\mathcal{M}_{\text{s}}(D)}$.

When $D = [0, \infty)$ we omit it from notation and write, e.g., $\mathcal{S}^\infty$ for $\mathcal{S}^\infty[0, \infty)$, etc.

2. Nonhomogeneous Hawkes processes

In order to introduce the notation and to single out one of several similar (but not entirely equivalent) frameworks found in the literature, we provide a short introduction to Hawkes processes and their nonhomogeneous versions. We believe that most (or all) results here are well-known (at least to specialists) but could not locate precise-enough references, and so we provide self-contained proofs for some of them. The reader is referred to any standard text on random measures and point processes (such as [DVJ03, DVJ08] or [Kal17]) for unexplained details. We also mention the paper [FLM15] which focuses on some aspects of the asymptotic behavior of nonhomogeneous Hawkes processes.

2.1. **Random measures and point processes.** For a subset D of an Euclidean space, we induce the measurable structure on $\mathcal{M}(D)$ by the evaluation maps $\mu \mapsto \mu(A)$, $A \in \mathcal{B}(D)$. The random elements in $\mathcal{M}(D)$ are called *random measures*, while random measures with values in $\mathcal{M}_{\text{lf}}(D)$ are said to be *locally finite*.

For $f \in \mathcal{S}^\infty_{\text{loc}}$ and $\mu \in \mathcal{M}_{\text{lf}}$, the *convolution* $f * \mu \in \mathcal{S}^\infty_{\text{loc}}$ is given by

$$(f * \mu)(t) := \int_{[0,t]} f(t-s)\, \mu(ds) \text{ for } t \ge 0.$$

To enhance legibility, we often use the convention that functions and measures inside a convolution take the value 0 outside their original domain of definition and often simply write $(f * \mu)(t) = \int f(t - \cdot)\, d\mu$. Using this convention, we can define the convolution of two measures $\mu, \nu \in \mathcal{M}_{\text{lf}}$ by

$(\mu * \nu)(A) := \int \mu(A-s)\nu(ds) = \int \nu(A-s)\,\mu(ds)$ and note that it is the unique element of $\mathcal{M}_{\text{lf}}$ such that $f*(\mu*\nu) = (f*\mu)*\nu$ for all $f \in \mathcal{S}^{\infty}_{\text{loc}}$.

The convolutional version of the standard moment-generating functional, defined below, proves to be easier to work with in the context of Hawkes processes than its classical counterpart. The value of the *convolutional moment-generating functional* M_ξ on $f \in \mathcal{S}^{\infty}_{\text{loc}}$ at $t \in [0,\infty)$ is given by

$$M_\xi[f](t) = \mathbb{E}\Big[e^{(f*\xi)(t)}\Big] \in [0,\infty]. \tag{2.1}$$

We note that, unlike in the standard case, the functional M_ξ depends on the additional parameter t. While this dependence does not encode any additional information (it simply shifts the function f), it leads to significantly simpler notation in the sequel.

A locally finite random measure N on $[0,\infty)$ is called a *point process* if $N(A) \in \mathbb{N}_0$ for all bounded $A \in \mathcal{B}([0,\infty))$. Each point process N admits a sequence $\{T_k\}_{k\in\mathbb{N}}$ of $[0,\infty]$-valued random variables called the *points of* N, such that $T_0 \le T_1 \le \dots$ and $T_k \to \infty$, a.s., and

$$N = \textstyle\sum_k \delta_{T_k},$$

where the sum is always taken only over k such that $T_k < \infty$; equivalently, $\delta_{+\infty}$ is identified with the zero measure on $[0,\infty)$. Since $\int f(t)\,N(dt) = \sum_k f(T_k)$, a.s., whenever both sides are well defined, we often use the convenient standard notation

$$\textstyle\sum_{T\in N} f(T) := \int f(t)\,N(dt).$$

We recall that for $\mu \in \mathcal{M}_{\text{lf}}$, the *Poisson process (with the intensity measure μ)* is the unique point process P such that 1) $P(A)$ is a Poisson random variable with expectation (parameter) $\mu(A)$ for each bounded $A \in \mathcal{B}([0,\infty))$, and 2) $P(A_1), \dots, P(A_n)$ are independent random variables whenever $A_1,\dots,A_n \in \mathcal{B}([0,\infty))$ are bounded and disjoint. For such P we have

$$M_P[f] = e^{(\exp(f)-1)*\mu} \text{ for all } f \in \mathcal{S}^{\infty}_{\text{loc}}.$$

We will also need the following expression

$$J_P[f] = e^{(f-1)*\mu} \text{ for } g \in \mathcal{S}^{\infty}_{\text{loc}}, \tag{2.2}$$

for the *convolutional probability-generating functional*

$$J_P[f](t) := \mathbb{E}\Big[\textstyle\prod_{T\in P, T\le t} g(t-T)\Big], t \ge 0,$$

of the Poisson process P with intensity μ.

2.2. **Nonhomogeneous single-progenitor Hawkes processes.** The typical definition of a standard Hawkes process involves two inputs: the background intensity and the excitation kernel. It will be convenient for our later analysis to separate the two and first construct a class of processes without any background intensity, but started, instead, from a single point (progenitor) at time $t=0$. Their distributions are determined by two parameters: a constant $a \in (0,1)$ and a sequence $\pi = \{\pi^m\}_{m\in\mathbb{N}}$ of probability measures on $(0,\infty)$. To relate them to the standard notation, we note that when π^m is absolutely continuous, we can define the excitation kernel ϕ^m (associated to the rate at which the points in generation $m-1$ produce offspring in generation m) by $a\pi^m(dt) = \phi^m(t)\,dt$.

More precisely, the *nonhomogeneous single-progenitor Hawkes process* $\tilde{H}$ *with parameters* a *and* $\pi = \{\pi^m\}_{m\in\mathbb{N}}$ is defined by

$$\tilde{H} := \cup_{m\in\mathbb{N}_0} \tilde{H}^m, \tag{2.3}$$

where the sequence $\{\tilde{H}^m\}_{m\in\mathbb{N}_0}$ of "generations" is built from a double sequence $\tilde{P}^m(k)$, $m \in \mathbb{N}$, $k \in \mathbb{N}_0$ of independent Poisson processes, where $\tilde{P}^m(k)$ has intensity $a\pi^m$, for each $k \in \mathbb{N}_0$. The zero-th generation $\tilde{H}^0$ is simply the Dirac mass δ_0 at 0, i.e., a deterministic point process with a single point at 0, representing the lone progenitor. Once the first m generations $\tilde{H}^0, \dots, \tilde{H}^{m-1}$, $m \in \mathbb{N}$, have been constructed, we set

$$\tilde{H}^m := \bigcup_{k\in\mathbb{N}_0} \bigcup_{S\in\tilde{P}^{m-1}(k)} \Big(T^{m-1}(k) + S\Big) \tag{2.4}$$

where $\{T^{m-1}(k)\}_{k\in\mathbb{N}_0}$ denotes the point sequence of $\tilde{H}^{m-1}$. In keeping with the convention introduced above, the first union is taken over k such that $T^{m-1}(k) < \infty$.

In the sequel, we often identify a point process with its (random) point set. Moreover, we abuse the notation and write, for example, $\tilde{P}^{m-1}(T)$ for the Poisson process $\tilde{P}^{m-1}(k)$ whose index k is such that $T = T^{m-1}(k)$. This way, (2.4) takes the more legible form

$$\tilde{H}^m = \bigcup_{T\in\tilde{H}^{m-1}} \Big(T + \tilde{P}^{m-1}(T)\Big).$$

The parameters a and $\{\pi^m\}_{m\in\mathbb{N}}$ of a single-progenitor Hawkes process can be used to construct a double sequence of *partial single-progenitor Hawkes processes* $\tilde{H}^{[m,m+k)}$, $m \in \mathbb{N}_0$, $k \in \mathbb{N}$ which will be needed in sequel. The process $\tilde{H}^{[m,m+k)}$ starts with a single individual in generation $m \in \mathbb{N}_0$, and acrues individuals over the next $k-1$ generations. This is distributionally equivalent to collecting the first k generations of a single-progenitor process with parameters a and $(\pi^{m+1}, \pi^{m+2}, \dots)$. These, individual, generations are denoted by $\tilde{H}^{m,(m+j)}$, $j = 0, \dots, k-1$ so that $\tilde{H}^{[m,m+k)} = \bigcup_{j=0}^{k-1} \tilde{H}^{m,(m+j)}$.

For $k \geq 1$, conditioning on the first generation $\tilde{H}^{m,(m+1)}$ of $\tilde{H}^m$ gives the following fundamental recursive distributional equality

$$\tilde{H}^{[m,m+k)} \overset{(d)}{=} \{0\} \cup \bigcup_{T\in\tilde{P}^{m+1}} \Big(T + \tilde{H}^{[m+1,m+k)}(T)\Big) \text{ for } k \geq 1, \tag{2.5}$$

where $\tilde{P}^{m+1}$ is a Poisson process with intensity $a\pi^{(m+1)}$, and $(\tilde{H}^{[m+1,m+k)}(T))_{T\in\tilde{P}^{m+1}}$ are independent partial single-progenitor Hawkes processes. We accumulate over all k in (2.5) to obtain

$$\tilde{H}^m \overset{(d)}{=} \{0\} \cup \bigcup_{T\in\tilde{P}^{m+1}} \Big(T + \tilde{H}^{m+1}(T)\Big). \tag{2.6}$$

2.3. **The moment-generating functional and moments.**

Proposition 2.1. *Given* $f \in \mathcal{S}^\infty_{\text{loc}}$ *and* $m \in \mathbb{N}_0$*, we have* $M_{\tilde{H}^{[m,m)}}[f] = 1$ *and*

$$M_{\tilde{H}^{[m,m+k)}}[f] = \exp\big(f + a(M_{\tilde{H}^{[m+1,m+k)}}[f] - 1) * \pi^{m+1}\big) \text{ for } k \geq 1. \tag{2.7}$$

Proof. The equality $M_{\tilde{H}^{[m,m)}}[f] = 1$ follows trivially from the definition. Assuming that $k \geq 1$ and conditioning on generation $m+1$ in (2.5), we get

$$\begin{aligned} M_{\tilde{H}^{[m,m+k)}}[f](t) &= \mathbb{E}\Big[e^{f(t)+\sum_{T\in\tilde{P}^{m+1}}\sum_{S\in\tilde{H}^{[m+1,m+k)}(T)} f(t-T-S)}\Big] \\ &= e^{f(t)}\mathbb{E}\Big[\prod_{T\in\tilde{P}^{m+1}}\mathbb{E}\Big[e^{\sum_{S\in\tilde{H}^{[m+1,m+k)}(T)} f(t-T-S)}\,\Big|\,\sigma(\tilde{P}^{m+1})\Big]\Big] \\ &= e^{f(t)}\mathbb{E}\big[\prod_{T\in\tilde{P}^{m+1}} M_{\tilde{H}^{[m+1,m+k)}}[f](t-T)\big] \\ &= e^{f(t)} J_{\tilde{P}^{m+1}}[M_{\tilde{H}^{[m+1,m+k)}}](t) \\ &= \exp\Big(f(t) + a\int \big(M_{\tilde{H}^{[m+1,m+k)}}[f](t-s) - 1\big)\,\pi^{m+1}(ds)\Big), \end{aligned}$$

where the last equality follows from (2.2). □

Let $W_0 : [-e^{-1},\infty) \to \mathbb{R}$ denote the principal branch of Lambert's W-function (see, e.g. [DLM, Section 4.13]).

Proposition 2.2. *For $\beta \in \mathbb{R}$, let $l(\beta) := \mathbb{E}[\exp(\beta|\tilde{H}|)] \in (0,\infty]$ be the moment generating function of the total number $|\tilde{H}|$ of points in $\tilde{H}$. Then*

$$l(\beta) = \begin{cases} +\infty, & \beta > a - 1 - \log(a), \\ -\frac{1}{a}W_0(-\exp(\beta - a + \log(a))), & \beta \leq a - 1 - \log(a). \end{cases} \tag{2.8}$$

Proof. Since $|\tilde{H}|$ depends only on the parameter a, and not on the sequence $\{\pi^m\}_{m\in\mathbb{N}}$, we have $|\tilde{H}^m| \overset{(d)}{=} |\tilde{H}|$ for all m. Virtually the same argument as in the proof of Proposition 2.1 above can be used to conclude that the functions $l_k(\beta) = \mathbb{E}\Big[\exp\Big(\beta|\tilde{H}^{[0,k)}|\Big)\Big]$, $k \in \mathbb{N}_0$ satisfy

$$l_k(\beta) = \exp(\beta + a(l_{k-1}(\beta) - 1)) \text{ for } k \in \mathbb{N}.$$

Assuming, first, that $\beta \geq 0$, the fact that $|\tilde{H}|$ is the nondecreasing limit of $|\tilde{H}^{[0,k)}|$, as $k \to \infty$ implies that $l_k(\beta) \nearrow l(\beta)$. Since $l_0(\beta) = 1$, it follows that $l(\beta)$ is the smallest fixed point above 1, if one exists, of the function $F(x) = \exp(\beta + a(x-1))$; otherwise, $l(\beta) = +\infty$. That latter case happens, in particular, when $\beta > a - 1 - \log(a)$, as can be easily seen by inspection. For $\beta \leq a - 1 - \log(a)$, the equation $F(x) = x$ transforms into

$$(-ax)\exp(-ax) = -a\exp(\beta - a), \tag{2.9}$$

with solutions given by

$$x_0 = -\frac{1}{a}W_0(-\exp(-a+\beta+\log(a))) \text{ and } x_{-1} = -\frac{1}{a}W_{-1}(-\exp(-a+\beta+\log(a))),$$

where W_0 and W_{-1} are the two branches of the Lambert's W-function on $[-e^{-1},0)$. The principal solution W_0 is increasing and W_{-1} is decreasing on $(-e^{-1},0)$, while $W_0(-e^{-1}) = W_{-1}(-e^{-1}) = -1$; this is easily seen directly, but we also refer the reader to [DLM, Section 4.13] for a more comprehensive treatment of the W-function. It follows that the smallest solution of $F(x) = x$ above $x = 1$ is given by x_0 defined in (2.9) above, which completes the proof of (2.8).

The case $\beta < 0$ is almost identical, with the distinction that the sequence $\{l_k(\beta)\}_{k\in\mathbb{N}}$ is now nonincreasing and bounded from above by 1, so we are looking for the largest fixed point of F

below 1. Since $F(1) < 1$ in this case, $l_0(\beta) = 1$ is located between the two solutions of $F(x) = x$ which leads us to choose the principal branch W_0 again. $\square$

Corollary 2.3. *Let $f \in \mathcal{S}^{\infty}_{\mathrm{loc}}$ be such that $f \le a - 1 - \log(a)$. Then*

$$M_{\tilde{H}^{[m,m+k)}}[f] \le \frac{1}{a} \text{ and } M_{\tilde{H}^m}[f] \le \frac{1}{a} \text{ for all } m \in \mathbb{N}_0, k \in \mathbb{N}. \tag{2.10}$$

Moreover, we have

$$M_{\tilde{H}^m}[f] = \exp\big(f + a\big(M_{\tilde{H}^{m+1}}[f] - 1\big) * \pi^{m+1}\big) \text{ for all } m \in \mathbb{N}_0. \tag{2.11}$$

Lastly, we use the recursive nature of the Hawkes process to derive an expression for its moments. For $f \in \mathcal{S}^{\infty}_{\mathrm{loc}}$, we let

$$\tilde{e}^{[m,m+k)}[f] = \mathbb{E}\Big[f * \tilde{H}^{[m,m+k)}\Big] \text{ and } \tilde{e}^{m}[f] = \mathbb{E}\Big[f * \tilde{H}^{m}\Big], \text{ for } m \in \mathbb{N}_0, k \in \mathbb{N}.$$

Just like in the proof of Proposition 2.1, the relation (2.5) implies that

$$\tilde{e}^{[m,m)}[f] = f \text{ and } \tilde{e}^{[m,m+k)}[f] = f + a\,\tilde{e}^{[m+1,m+k)} * \pi^{m+1} \text{ for } k \ge 1. \tag{2.12}$$

Therefore,

$$\tilde{e}^{[m,m+k)}[f] = f * \left(\textstyle\sum_{j=0}^{k} a^j \pi^{(m,m+j]}\right)$$

where

$$\pi^{(m,m+j]} := \begin{cases} \delta_0, & j = 0, \\ \pi^{m+1} * \cdots * \pi^{m+j}, & j > 0. \end{cases} \tag{2.13}$$

Since $a < 1$ and $|\pi^{(m,m+j]}| = 1$ for all $m, j \in \mathbb{N}_0$, we have convergence in total variation in

$$\rho^m := \sum_{k=0}^{\infty} (1-a) a^k \pi^{(m,m+k]} \in \mathcal{M}_{\mathrm{p}}, \tag{2.14}$$

and the following identity holds

$$\tilde{e}^{m}[f] = \frac{1}{1-a} f * \rho^m \text{ for } f \in \mathcal{S}^{\infty}_{\mathrm{loc}}, m \in \mathbb{N}_0. \tag{2.15}$$

2.4. **The genealogical tree of a single-progenitor Hawkes process.** The *genealogical tree* of a single-progenitor Hawkes process is a random rooted directed tree associated to its construction via generations $\tilde{H}^0, \tilde{H}^1, \dots$ as in subsection (2.2) above. The vertices of the tree are the points of the process, the root is the initial point at 0, and edges connect each point, except the root, to its "parent" in the previous generation. We note that the structure of this tree does not depend on the position of the individual points in each generation, only their number and the parent-child relationship. This means, in particular, that the distribution of the genealogical tree depends only on the value of the parameter a, but not on the choice of the of the sequence $\{\pi^m\}_{m\in\mathbb{N}}$ of probability measures. It will be important in the sequel to observe that the number of points in each generation of the genealogy is a Bienaymé-Galton-Watson process with the Poisson offspring distribution with parameter a. Moreover, one can construct the single-progenitor Hawkes process starting from the genealogical tree as follows:

(1) Construct a random directed tree corresponding to a Bienaymé-Galton-Watson process with Poisson offspring distribution with parameter a. This will be the genealogical tree of the Hawkes process.
(2) For each directed edge in the tree, connecting generations m and $m-1$, sample an independent random variable - which we call the *length* of the edge - with distribution π^m.
(3) Construct the single-progenitor Hawkes process $\tilde{H}$ by starting with the point at 0 and for each non-root vertex v of the tree add a point to $\tilde{H}$ at the position obtained by adding together the lengths of the edges forming the unique path from the root to v.

2.5. **Hawkes processes.** The Hawkes process is defined as a superposition of independent single-progenitor Hawkes processes, started as different points of an underlying Poisson process. More precisely, in addition to the parameters a and $\pi = \{\pi^m\}_{m\in\mathbb{N}}$ of a single-progenitor process, let a *background intensity* measure $\mu \in \mathcal{M}_{\text{lf}}$ be given. The *Hawkes process with parameters* μ*,* a *and* π is defined by

$$H := \bigcup_{T\in P} \Big(T + \tilde{H}(T)\Big), \tag{2.16}$$

where P is a Poisson process with intensity μ and $\{\tilde{H}(j)\}_{j\in\mathbb{N}}$ is a sequence of independent single-progenitor Hawkes processes with parameters $(a, \{\pi^m\}_{m\in\mathbb{N}})$, independent of P. Thanks to local finiteness of the Poisson process and the finiteness of single-progenitor Hawkes processes (guaranteed by the assumption that $a < 1$), H is a locally finite random measure, too. We have the following continuation of Proposition 2.1

Proposition 2.4. *Given* $f \in \mathcal{S}^\infty_{\text{loc}}$ *with* $f \le a - 1 - \log(a)$*, we have* $M_H[f] \in \mathcal{S}^\infty$ *and*

$$M_H[f] = \exp\big((M_{\tilde{H}}[f] - 1) * \mu\big). \tag{2.17}$$

Proof. The defining relation (2.16) implies that

$$\begin{aligned} M_H[f](t) &= \mathbb{E}[\exp(f * H(t))] = \mathbb{E}\Big[\exp\Big(\textstyle\sum_{T\in P}\sum_{S\in\tilde{H}_T} f(t-(T+S))\Big)\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\Big[\textstyle\prod_{T\in P}\exp\Big(\sum_{S\in\tilde{H}_T} f((t-T)-S)\Big)\ \Big|\ \sigma(P)\Big]\Big] \\ &= \mathbb{E}\big[\textstyle\prod_{T\in P} M_{\tilde{H}}[f](t-T)\big] = J_P[M_{\tilde{H}}[f]](t) = \exp\Big(\big((M_{\tilde{H}}[f]-1)*\mu\big)(t)\Big), \end{aligned}$$

where the last equality follows from (2.2) and we use our standard convention that the functions f and $M_{\tilde{H}}[f] - 1$ take the value 0 for $t < 0$. □

An argument similar to the one leading to (2.17) above implies that we have the following expression for the first moment $e[f] = \mathbb{E}[f * H]$ of the Hawkes process:

$$e[f] = \tilde{e}[f] * \mu = \frac{1}{1-a} f * (\rho * \mu). \tag{2.18}$$

where $\rho = \rho^0$ and ρ^0 is given by (2.14).

3. A convergence theorem for nonhomogeneous Hawkes processes

3.1. **A convergence theorem.** We start with a sequence $\{(a_n, \{\pi_n^m\}_{m\in\mathbb{N}}, \mu_n)\}_{n\in\mathbb{N}}$ of parameter triplets of nonhomogeneous Hawkes processes. In particular, for each $n \in \mathbb{N}$, $a_n \in (0,1)$, μ_n is a locally bounded measure on $[0,\infty)$, and $\{\pi_n^m\}_{m\in\mathbb{N}}$ is a sequence of probability measures on $(0,\infty)$. For each n, we denote a Hawkes process with parameters $(a_n, \{\pi_n^m\}_{m\in\mathbb{N}}, \mu_n)$ by H_n, and the associated single-progenitor process by $\tilde{H}_n$. We also use the partial versions of these processes together with the notation introduced in and after subsection 2.2, but additionally indexed by n.

Our first goal is to give sufficient conditions on the sequences $\{a_n\}_{n\in\mathbb{N}}$, $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ and $\{\mu_n\}_{n\in\mathbb{N}}$ that will ensure that, when properly scaled, the processes H_n converge, and to characterize the distribution of the limit.

We remind the reader that the *vague topology*, with convergence denoted by $\xrightarrow{v}$, is the coarsest topology on $\mathcal{M}_{\rm lf}$ such that the map $\mu \mapsto \int f\, d\mu$ is continuous for each $f \in C_{\rm c}$. The *weak topology* on the space $\mathcal{M}_{\rm f}$, with convergence denoted by $\xrightarrow{w}$, is defined similarly, but with $C_{\rm c}$ replaced by $C_{\rm b}$.

Since $[0,\infty)$ is complete and separable, the vague topology on $\mathcal{M}_{\rm lf}$ is Polish (completely metrizable and separable). For a sequence of locally finite random measures $\{\xi_n\}_{n\in\mathbb{N}}$, we say that ξ_n *converges to* ξ *in distribution*, and write $\xi_n \xrightarrow{d} \xi$, if ξ_n converges to ξ weakly when interpreted as a sequence of random elements in $\mathcal{M}_{\rm lf}$ metrized by the vague topology. We refer the reader to [Kal17, Section 4.1, p. 111] for a textbook treatment and proofs of various properties of the vague and weak topologies used throughout the paper.

Using the notation introduced in (2.14) above, we set

$$\rho_n := \sum_{k=0}^{\infty} (1-a_n)(a_n)^k \pi_n^{(0,k]} \in \mathcal{M}_{\rm p}. \tag{3.1}$$

It is convenient to express ρ_n as the expected value

$$\rho_n = \mathbb{E}\Big[\pi_n^{(0,G_n]}\Big],$$

where G_n is an $\mathbb{N}_0$-valued geometrically distributed random variable with parameter (probability of success) $1-a_n$.

For $\nu_1, \nu_2 \in \mathcal{M}_{\rm lf}$ and $T \geq 0$, we set

$$W^1_{[0,T]}(\nu_1,\nu_2) = \int_0^T |F_{\nu_1}(t) - F_{\nu_2}(t)|\, dt,$$

where $F_{\nu_i} = \nu_i([0,\cdot])$, $i = 1,2$, are distribution functions of ν_1 and ν_2. When ν_1 and ν_2 are probability measures with supports in $[0,T]$, $W^1_{[0,T]}(\nu_1,\nu_2)$ coincides with the 1-Wasserstein distance between ν_1 and ν_2 (see, e.g., [San15, Proposition 2.17, p. 66]).

The array $\{\pi_n^m\}$ is said to be *null* if, for each $\varepsilon > 0$,

$$\lim_n \sup_m \pi_n^m([\varepsilon,\infty)) = 0.$$

We recall that C_0 denotes the family of all continuous functions f on $[0,\infty)$ with $f(0) = 0$.

Theorem 3.1. *Suppose that*

(1) $(1-a_n)\mu_n \xrightarrow{v} \mu \in \mathcal{M}_{\mathrm{lf}}$.
(2) The array $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ *is null and* $\rho_n \xrightarrow{w} \rho$ *for some* $\rho \in \mathcal{M}_p$ *with* $\rho(\{0\}) = 0$.
(3) Either one of the following two conditions holds:
  (a) there exists a constant $d \in \mathbb{N}$ *such that*

$$\pi_n^{m+d} = \pi_n^m \text{ for all } m \in \mathbb{N}, n \in \mathbb{N}_0, \text{ or} \tag{3.2}$$

  (b) for each $T \geq 0$

$$\lim_n (1-a_n)^{-2} \sup_m W^1_{[0,T]}(\pi_n^m, \pi_n^1) = 0.$$

Then, there exists a locally finite random measure ξ *on* $[0,\infty)$ *such that*

$$(1-a_n)^2 H_n \xrightarrow{d} \xi. \tag{3.3}$$

It is characterized by

$$M_\xi[f] = \exp(h[f] * \mu) \text{ for all } f \in C_0 \text{ with } f \leq 1/2, \tag{3.4}$$

where $h[f]$ *the unique solution in* $\mathcal{S}^\infty_{\mathrm{loc}}$ *to the convolutional Riccati equation*

$$h = (f + \tfrac{1}{2}h^2) * \rho. \tag{3.5}$$

The proof is divided into several lemmas. Before we state them, we introduce the necessary notation and terminology. A quantity is said to be a *universal constant* if it depends only on the primitives $\{a_n\}_{n\in\mathbb{N}}$, $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ and $\{\mu_n\}_{n\in\mathbb{N}}$. We will always denote a generic universal constant with the letter C, even though it might change from use to use.

We retain all the notation from section 2 (subsections 2.2 and 2.5 in particular), but add a subscript n to signal the association with the parameters a_n, $\{\pi_n^m\}_{m\in\mathbb{N}}$ and μ_n. With that in mind, and using the shortcut

$$\varepsilon_n := 1 - a_n,$$

we define the random measures ξ_n^m and $\tilde{\xi}_n^m$ by

$$\xi_n^m := \varepsilon_n^2 H_n^m \quad \text{and} \quad \tilde{\xi}_n^m := \varepsilon_n^2 \tilde{H}_n^m. \tag{3.6}$$

For $f \in \mathcal{S}^\infty_{\mathrm{loc}}$ and $t \geq 0$, we set

$$h_n^m[f](t) := \frac{1}{\varepsilon_n}\Big(M_{\tilde{\xi}_n^m}[f](t) - 1\Big) = \frac{1}{\varepsilon_n}\Big(M_{\tilde{H}_n^m}[\varepsilon_n^2 f](t) - 1\Big) \in (-\infty, \infty]. \tag{3.7}$$

When the function f is clear from the context, we often omit it from notation and write, e.g., h_n^m for $h_n^m[f]$.

An $\mathbb{N}_0$-valued geometrically distributed random variable with parameter (probability of success) $1 - a_n$ will be denoted by G_n throughout the proof.

Given $T > 0$, a function $f \in \mathcal{S}^\infty[0,T]$ is said to be *of bounded variation* if there exists a signed measure $Df \in \mathcal{M}_{\mathrm{s}}[0,T]$ such that $f(t) = Df([0,t])$ for all $t \in [0,T]$. The Hahn decomposition of Df is denoted by $Df = D^+f - D^-f$ and the total variation measure $D^+f + D^-f$, associated to Df, by $|D|f$. The map $|f|_{\mathrm{BV}[0,T]} := |Df|_{\mathcal{M}_{\mathrm{s}}[0,T]} = ||D|f|$ is a Banach norm on $\mathrm{BV}[0,T]$ and the set of all $f \in \mathcal{S}^\infty_{\mathrm{loc}}$ such that $f|_{[0,T]} \in \mathrm{BV}[0,T]$ for all $T \geq 0$ is denoted by $\mathrm{BV}_{\mathrm{loc}}$.

We start with a lemma about the asymptotic behavior of the total number $|\tilde{H}_n^m|$ of points of $\tilde{H}_n^m$. We note that the distribution of $|\tilde{H}_n^m|$ does not depend on m, so we denote it in the sequel plainly by $|\tilde{H}_n|$ to simplify the notation and stress the uniformity of the obtained bounds in m.

Lemma 3.2. *For $\delta \in [0, 1)$ we have*

$$\lim_n \frac{1}{\varepsilon_n}\Big(\mathbb{E}\Big[\exp\Big(\tfrac{1}{2}(1-\delta)\varepsilon_n^2|\tilde{H}_n|\Big)\Big] - 1\Big) = 1 - \sqrt{\delta}, \tag{3.8}$$

and, for $\delta \in (0, 1)$,

$$\lim_n \varepsilon_n^{2k-1}\mathbb{E}\Big[|\tilde{H}_n|^k \exp\Big(\varepsilon_n^2 \tfrac{1-\delta}{2}|\tilde{H}_n|\Big)\Big] = \frac{(k-1)!}{2^{k-1}}\binom{2(k-1)}{k-1}\delta^{1/2-k} \text{ for } k \in \mathbb{N}. \tag{3.9}$$

Proof. Let $w(x) = -W_0(-e^{-1-x^2})$ for $x \geq 0$, where W_0 is the principal branch of Lambert's W-function; see [DLM, Section 4.13] for the standard properties used in this proof. Proposition 2.2 above states that for $\beta < \beta_n^{max} := -\log(1-\varepsilon_n) - \varepsilon_n$ we have

$$\mathbb{E}\Big[\exp\Big(\beta|\tilde{H}_n|\Big)\Big] = \frac{1}{1-\varepsilon_n} w(b_n(\beta)) \text{ where } b_n(\beta) = (\beta_n^{max} - \beta)^{1/2}. \tag{3.10}$$

We have

$$b_n\big(\tfrac{1}{2}\varepsilon_n^2\big) = \Big(-\log(1-\varepsilon_n) - \varepsilon_n - \tfrac{1}{2}\varepsilon_n^2\Big)^{1/2} = O(\varepsilon_n^{3/2})$$

Since w is continuously differentiable on $[0, \infty)$ and $w(0) = 1$, this implies that

$$\frac{1}{1-\varepsilon_n} w\big(b_n\big(\tfrac{1}{2}\varepsilon_n^2\big)\big) = 1 + \varepsilon_n + O(\varepsilon_n^{3/2}),$$

which, in turn, yields (3.8) for $\delta = 0$. Similarly, when $\delta \in (0, 1)$ we have

$$b_n\big(\tfrac{1}{2}(1-\delta)\varepsilon_n^2\big) = \sqrt{\tfrac{\delta}{2}}\varepsilon_n + O(\varepsilon_n^2),$$

and (3.8) when $\delta \in (0, 1)$ follows from

$$\frac{1}{1-\varepsilon_n} w\big(b_n\big(\tfrac{1}{2}(1-\delta)\varepsilon_n^2\big)\big) = 1 + (1-\sqrt{\delta})\varepsilon_n + o(\varepsilon_n).$$

Suppose that $\delta \in (0, 1)$ for the rest of the proof. Standard properties of moment-generating functions allow us to differentiate $k \in \mathbb{N}$ times inside the expectation sign at each $\beta < \beta_n^{max}$ in (3.10) above to obtain

$$\mathbb{E}\Big[|\tilde{H}_n^m|^k \exp\Big(\beta|\tilde{H}_n^m|\Big)\Big] = \frac{1}{1-\varepsilon_n}(w \circ b_n)^{(k)}(\beta) \text{ for } \beta < \beta_n^{max}, \tag{3.11}$$

where $(\cdot)^{(k)}$ denotes the k-th derivative in β. The formula of Faá di Bruno states that $(w \circ b_n)^{(k)}(\beta)$ admits a representation of the form

$$\sum \frac{k!}{m_1! \dots m_k!} w^{(m_1+\dots+m_k)}\Big(b_n(\beta)\Big) \prod_{j=1}^k \left(\frac{b_n^{(j)}(\beta)}{j!}\right)^{m_j}, \tag{3.12}$$

where the sum is taken over all $m_1, \dots, m_k \in \mathbb{N}_0$ such that $m_1 + 2m_2 + \dots + km_k = k$.

We have $b_n^{(j)}(\beta) = (-1)^j j! \binom{1/2}{j} (\beta_n^{max} - \beta)^{1/2-j}$ so that

$$\prod_{j=1}^{k} \left(\frac{b_n^{(j)}(\beta)}{j!} \right)^{m_j} = K \left(\beta_n^{max} - \beta\right)^{\frac{m_1+\dots+m_k}{2} - k} \text{ where } K = \prod_{j=1}^{k} \left((-1)^j \binom{1/2}{j} \right)^{m_j}.$$

The lowest power of $(\beta_n^{max} - \beta)$ appearing in (3.12) is $1/2 - k$ and is attained precisely at $m_1 = \dots = m_{k-1} = 0$, $m_k = 1$. Furthermore, all functions $w^{(m_1+\dots+m_k)}(x)$ converge towards a finite limit as $x \searrow 0$ which implies that

$$(w \circ b_n)^{(k)}(\beta) = k! w'(b_n(\beta))(-1)^k \binom{1/2}{k} (\beta_n^{max} - \beta)^{1/2-k} + o\Big((\beta_n^{max} - \beta)^{1/2-k}\Big).$$

We have

$$b_n\big(\tfrac{1}{2}(1-\delta)\varepsilon_n^2\big) = \sqrt{\tfrac{\delta}{2}}\, \varepsilon_n + o(\varepsilon_n), \tag{3.13}$$

and using (3.13), together with the fact that $\lim_{x \searrow 0} w'(x) = -\sqrt{2}$, we get

$$\lim_n \frac{(w \circ b_n)^{(k)}(\beta_n)}{(\delta \varepsilon_n^2/2)^{1/2-k}} = \sqrt{2}(-1)^{k-1} k! \binom{1/2}{k},$$

which, in turn, implies (3.9). □

The inequalities in Lemma 3.3 below are well known (the first one is the simplest special case of Young's inequality), but we give short proofs for completeness.

Lemma 3.3. *Suppose that* $T \geq 0$ *and* $\nu, \nu' \in \mathcal{M}_{\mathrm{f}}([0,T])$.

(1) If $h \in \mathcal{L}^1[0,T]$*, then* $h * \nu \in \mathcal{L}^1[0,T]$ *and*

$$|h * \nu|_{\mathcal{L}^1[0,T]} \leq |h|_{\mathcal{L}^1[0,T]} |\nu|. \tag{3.14}$$

(2) If $h \in \mathrm{BV}[0,T]$*, then* $h * \nu \in \mathrm{BV}[0,T]$ *with* $D(h * \nu) = Dh * \nu$ *and*

$$|h * \nu|_{\mathrm{BV}[0,T]} \leq |h|_{\mathrm{BV}[0,T]} |\nu| \tag{3.15}$$

as well as

$$\big|h * (\nu - \nu')\big|_{\mathcal{L}^1[0,T]} \leq |h|_{\mathrm{BV}[0,T]}\, W^1_{[0,T]}(\nu, \nu') \tag{3.16}$$

Proof. (1) By Fubini's theorem

$$\begin{aligned} \int_0^T |h * \nu(t)|\, dt &\leq \int_0^T \int |h(t-u)|\, \nu(du)\, dt \\ &= \int \int_u^T |h(t-u)|\, dt\, \nu(du) \leq |h|_{\mathcal{L}^1[0,T]} \nu[0,T]. \end{aligned}$$

(2) We have $(h * \nu)(t) = (Dh * \nu)[0,t]$, which implies that $h * \nu \in \mathrm{BV}[0,T]$ with $D(h * \nu) = Dh * \nu$. Therefore,

$$\begin{aligned} |h * \nu|_{\mathrm{BV}[0,T]} &= |D(h * \nu)|_{\mathcal{M}_{\mathrm{s}}[0,T]} = |Dh * \nu|_{\mathcal{M}_{\mathrm{s}}[0,T]} \\ &= \big|Dh^+ * \nu - Dh^- * \nu\big|_{\mathcal{M}_{\mathrm{s}}[0,T]} \leq (Dh^+ * \nu)[0,T] + (Dh^- * \nu)[0,T] \\ &\leq Dh^+[0,T]\nu[0,T] + Dh^-[0,T]\nu[0,T] = |Dh|_{\mathcal{M}_{\mathrm{s}}[0,T]} |\nu| = |h|_{\mathrm{BV}[0,T]} |\nu|. \end{aligned}$$

To show (3.16), we observe that

$$(Dh * \nu)[0,t] = (\nu * Dh)[0,t] = (F_\nu * Dh)(t) \text{ for } t \geq 0,$$

where $F_\nu(t) = \nu([0,t])$ denotes the distribution function of ν. Hence,

$$\big|h * \nu(t) - h * \nu'(t)\big| \leq \int |F_\nu(t-u) - F_{\nu'}(t-u)|\,|Dh|(du) \text{ for } t \geq 0,$$

and so,

$$\begin{aligned}\int_0^T \big|h * \nu(t) - h * \nu'(t)\big|\,dt &\leq \int_0^T \int |F_\nu(t-u) - F_{\nu'}(t-u)|\,|Dh|(du)\,dt \\ &\leq \int \int_0^T |F_\nu(t) - F_{\nu'}(t)|\,dt\,|Dh|(du) = |h|_{\mathrm{BV}[0,T]} W^1_{[0,T]}(\nu,\nu').\end{aligned}$$

□

Lemma 3.4. *For each $\delta > 0$, there exists a universal constant C such that for all $m \in \mathbb{N}_0, n \in \mathbb{N}, T \geq 0$ and all $f \in \mathcal{S}^\infty_{\mathrm{loc}}$ with $f \leq \frac{1}{2}(1-\delta)$ we have*

(1) $h^m_n \in \mathcal{S}^\infty_{\mathrm{loc}}$ and

$$\inf_{u\in[0,T]} f(u) \leq h^m_n(t) \leq C \sup_{u\in[0,T]} f(u) \text{ for } t \in [0,T]. \tag{3.17}$$

(2) If, additionally, $f \in \mathrm{BV}_{\mathrm{loc}}$, then $h^m_n \in \mathrm{BV}_{\mathrm{loc}}$ and

$$|h^m_n|_{\mathrm{BV}[0,T]} \leq C|f|_{\mathrm{BV}[0,T]}. \tag{3.18}$$

Proof. We pick $\delta \in (0,1)$ and $f \in \mathcal{S}^\infty_{\mathrm{loc}}$ with $f \leq \frac{1}{2}(1-\delta)$ as in the statement and set

$$Z^m_n = \exp\Big(\tfrac{1-\delta}{2}|\tilde{\xi}^m_n|\Big) \tag{3.19}$$

with $\tilde{\xi}^m_n$ as defined in (3.6), noting that the distribution of Z^m_n does not depend on m.

Since $f * \tilde{\xi}^m_n \leq \frac{1}{2}(1-\delta)|\tilde{\xi}^m_n|$, (3.9) with $k=1$ implies that

$$\begin{aligned} h^m_n(t) &\leq \varepsilon_n^{-1}\mathbb{E}\Big[Z^m_n (f * \tilde{\xi}^m_n)(t)\Big] \leq \Big(\sup\nolimits_{u\in[0,T]} f(u)\Big)\varepsilon_n \mathbb{E}\Big[Z^m_n|\tilde{H}^m_n|\Big] \\ &\leq C \sup\nolimits_{u\in[0,T]} f(u) \text{ for } t \in [0,T]. \end{aligned} \tag{3.20}$$

To get a lower bound, we use Jensen's inequality and (2.15):

$$\begin{aligned} h^m_n(t) &\geq \varepsilon_n^{-1}\Big(\exp\Big(\mathbb{E}\Big[(f * \tilde{\xi}^m_n)(t)\Big]\Big) - 1\Big) \geq \varepsilon_n^{-1}\mathbb{E}\Big[(f * \tilde{\xi}^m_n)(t)\Big] \\ &= (f * \rho^m_n)(t) \geq \inf_{u\in[0,T]} f(u). \end{aligned}$$

To establish (3.18), we pick $0 \leq r \leq s \leq T$ and observe that

$$\begin{aligned} |h^m_n(s) - h^m_n(r)| &\leq \varepsilon_n^{-1}\mathbb{E}\Big[\Big|\exp\Big(f * \tilde{\xi}^m_n(s)\Big) - \exp\Big(f * \tilde{\xi}^m_n(r)\Big)\Big|\Big] \\ &\leq \varepsilon_n^{-1}\mathbb{E}\Big[Z^m_n\Big|f * \tilde{\xi}^m_n(s) - f * \tilde{\xi}^m_n(r)\Big|\Big] \\ &\leq \varepsilon_n^{-1}\Big(F^m_n(r,s) + \hat{F}^m_n(r,s)\Big), \end{aligned} \tag{3.21}$$

where

$$F_n^m(r,s) = \mathbb{E}\Big[Z_n^m \int_0^r |f(s-u) - f(r-u)|\, \tilde{\xi}_n^m(du)\Big], \text{ and}$$
$$\hat{F}_n^m(r,s) = \mathbb{E}\Big[Z_n^m \int_r^s |f(s-u)|\, \tilde{\xi}_n^m(du)\Big].$$

Since $|f(b)-f(a)| \leq \int 1_{(a,b]}\, d|D|f$, for all $a<b$ in $[0,T]$, we have

$$F_n^m(r,s) \leq \mathbb{E}\Big[Z_n^m \iint 1_{\{v\in(r-u,s-u],u\leq r\}}\, |D|f(dv)\, \tilde{\xi}_n^m(du)\Big].$$

Hence, for $\kappa \in (0,T)$,

$$\begin{aligned}\frac{1}{\kappa}\int_0^{T-\kappa} F_n^m(r, r+\kappa)\, dr &= \\ &= \mathbb{E}\Big[Z_n^m \iint \frac{1}{\kappa}\int 1_{\{r\in[v+u-\kappa, v+u)\cap[T-\kappa,u]\}}\, dr\, |D|f(dv)\tilde{\xi}_n^m(du)\Big] \\ &\leq \mathbb{E}\Big[Z_n^m \iint |D|f(dv)\tilde{\xi}_n^m(du)\Big] \leq |f|_{\mathrm{BV}[0,T]}\mathbb{E}\Big[Z_n^m|\tilde{\xi}_n^m|\Big] \\ &\leq \varepsilon_n C|f|_{\mathrm{BV}[0,T]},\end{aligned}$$

where the last inequality follows from (3.9). Similarly,

$$\begin{aligned}\frac{1}{\kappa}\int_0^{T-\kappa} \hat{F}_n^m(r,r+\kappa)\, dr &\leq \frac{1}{\kappa}\int_0^T |f|_{\mathcal{S}^\infty[0,T]}\mathbb{E}\Big[Z_n^m \tilde{\xi}_n^m([u,u+\kappa])\Big]\, du \\ &\leq |f|_{\mathcal{S}^\infty[0,T]}\mathbb{E}\Big[Z_n^m \int \frac{1}{\kappa}\int 1_{\{s\in[u,u+\kappa]\}}\, du\, \tilde{\xi}_n^m(ds)\Big] \\ &\leq \varepsilon_n C|f|_{\mathcal{S}^\infty[0,T]} \leq \varepsilon_n C|f|_{\mathrm{BV}[0,T]}.\end{aligned}$$

Therefore,

$$\frac{1}{\kappa}\int_0^{T-\kappa} |h_n^m(u+\kappa) - h_n^m(u)|\, du \leq C|f|_{BV[0,T]} \tag{3.22}$$

uniformly in $m\in\mathbb{N}_0$, $n\in\mathbb{N}$ and $\kappa\in(0,T]$. Thanks to [Leo17, Corollary 2.51, p. 53], this implies that for all $m\in\mathbb{N}_0$, $n\in\mathbb{N}$ there exists a signed measure ν_n^m on $(0,T]$ with $|\nu_n^m|_{\mathcal{M}_s[0,T]} \leq C|f|_{\mathrm{BV}[0,T]}$ such that

$$h_n^m(t) - h_n^m(0) = \nu_n^m((0,t]) \text{ a.e., for all } t\in[0,T].$$

Thanks to part (2) of Lemma 3.3, for $f\in\mathrm{BV}_{\mathrm{loc}}$ we have $f*\xi_n^m \in \mathrm{BV}_{\mathrm{loc}}$, a.s, and, in particular, $(f*\xi)(s)\to(f*\xi)(t)$, a.s., when $s\searrow t$. Since, as above,

$$\sup_t \Big|\exp\Big(f*\tilde{\xi}_n^m)(t)\Big) - 1\Big| \leq CZ_n^m|\tilde{\xi}_n^m| \in \mathbb{L}^1,$$

we can use the dominated convergence theorem to conclude that the function h_n^m is right continuous. This implies that $h_n^m(t) - h_n^m(0) = \nu_n^m((0,t])$ everywhere, and, consequently, that $h_n^m \in \mathrm{BV}_{\mathrm{loc}}$. Lastly, by (3.17), we have

$$|h_n^m|_{\mathrm{BV}[0,T]} \leq |h_n^m - h_n^m(0)|_{\mathrm{BV}[0,T]} + |h_n^m(0)| \leq C|f|_{\mathrm{BV}[0,T]}. \qquad \square$$

Lemma 3.5. *For each $\delta \in (0,1)$, there exists a universal constant C such that for all $f \in \mathrm{BV}_{\mathrm{loc}}$ with $f \leq \frac{1}{2}(1-\delta)$ we have*

$$\left|h_n^m - h_n^{m+1}\right|_{\mathcal{L}^1[0,T]} \leq C\Big(\varepsilon_n + W^1_{[0,T]}(\pi_n^m, \delta_0)\Big), \tag{3.23}$$

as well as

$$h_n^m = \mathbb{E}\Big[f * \pi_n^{(m,m+G_n]} + \tfrac{1}{2}(h_n^{m+G_n})^2 * \pi_n^{(m,m+G_n]}\Big] + r_n^m, \tag{3.24}$$

for all $m \in \mathbb{N}_0$, $n \in \mathbb{N}$ and $T \geq 0$, where the remainders r_n^m satisfy

$$\lim_n r_n^m = 0 \text{ in } \mathrm{BV}[0,T] \text{ uniformly in } m.$$

Proof. Thanks to the relation (2.11) of Corollary 2.3, we have

$$\varepsilon_n^2 f + \varepsilon_n a_n h_n^{m+1} * \pi_n^{m+1} = \log(1 + \varepsilon_n h_n^m). \tag{3.25}$$

Assuming that n is so large that $\varepsilon_n h_n^m \geq -1/2$, we can use the mean-value theorem to conclude that there exists a sequence $\{\theta_n^m\}_{n\in\mathbb{N}}$ in $(0,2)$ such that $\theta_n^m \to 1$ uniformly in m, and

$$\log(1+\varepsilon_n h_n^m) = \varepsilon_n h_n^m - \tfrac{1}{2}\varepsilon_n^2 \theta_n^m (h_n^m)^2.$$

Therefore, (3.25) can be rewritten as

$$h_n^m - a_n h_n^{m+1} * \pi_n^{m+1} = \varepsilon_n f + \tfrac{1}{2}\varepsilon_n \theta_n^m (h_n^m)^2. \tag{3.26}$$

It follows that

$$\begin{aligned} \left|h_n^m - h_n^{m+1}\right|_{\mathcal{L}^1[0,T]} \leq \varepsilon_n |f|_{\mathcal{L}^1[0,T]} &+ \varepsilon_n \left|h_n^{m+1} * \pi_n^{m+1}\right|_{\mathcal{L}^1[0,T]} + \\ &+ \varepsilon_n \left|(h_n^m)^2\right|_{\mathcal{L}^1[0,T]} + \left|h_n^{m+1} - h_n^{m+1} * \pi_n^{m+1}\right|_{\mathcal{L}^1[0,T]}. \end{aligned} \tag{3.27}$$

The three terms on the right-hand side of (3.27) above are bounded by universal constants by Lemma 3.4 and the inequality (3.14) of Lemma 3.3. To bound the fourth term in (3.27) by the second term on the right-hand side of (3.23), we use Lemma 3.3 again, but now with the inequality (3.16).

Moving on to the proof of (3.24), given $k \geq 0$ we replace m by $m+k$ in (3.26), convolve it with $\pi^{(m,m+k]}$ and multiply by $(a_n)^k$ to obtain:

$$\begin{aligned} (a_n)^k h_n^{m+k} * \pi^{(m,m+k]} - (a_n)^{k+1} h_n^{m+k+1} * \pi^{(m,m+k+1]} &= \\ = (1-a_n)(a_n)^k f * \pi^{(m,m+k]} &+ \tfrac{1}{2}(1-a_n)(a_n)^k \theta_n^{m+k} (h_n^{m+k})^2 * \pi^{(m,m+k]}. \end{aligned}$$

Summing over $k \geq 0$ gives

$$\begin{aligned} h_n^m &= \sum_{k\geq 0} (1-a_n)(a_n)^k f * \pi_n^{(m,m+k]} \\ &\quad + \tfrac{1}{2} \sum_{k\geq 0} (1-a_n)(a_n)^k \theta_n^{m+k} (h_n^{m+k})^2 * \pi_n^{(m,m+k]} \\ &= \mathbb{E}\Big[f * \pi_n^{(m,m+G_n]} + \tfrac{1}{2}\theta_n^{m+G_n} (h_n^{m+k})^2 * \pi_n^{(m,m+G_n]}\Big]. \end{aligned}$$

The choice

$$r_n^m = \tfrac{1}{2}\mathbb{E}\Big[(\theta_n^{m+G_n} - 1)(h_n^{m+G_n})^2 * \pi_n^{(m,m+G_n]}\Big]$$

yields (3.24). To show that $r^m_n \to 0$, we start with

$$\sup_m |r^m_n|_{\mathrm{BV}[0,T]} \le \tfrac{1}{2}\Big(\sup_k \Big|\theta^k_n - 1\Big|\Big)\Big(\sup_{k\ge 0}\Big|(h^{m+k}_n)^2 * \pi^{(m,m+k]}_n\Big|_{\mathrm{BV}[0,T]}\Big).$$

The first supremum converges to 0 as $n \to \infty$, so it remains to show that the second one is uniformly bounded in m and n. This follows from the inequality (3.14) together with the observation that $(h^m_n)^2 \in \mathrm{BV}_{\mathrm{loc}}$ and

$$\big|(h^m_n)^2\big|_{\mathrm{BV}[0,T]} \le 2|h^m_n|_{\mathcal{S}^\infty[0,T]}||h^m_n|_{\mathrm{BV}[0,T]} \le 2|h^m_n|^2_{\mathrm{BV}[0,T]} \le C|f|^2_{\mathrm{BV}[0,T]}. \tag{3.28}$$

□

Lemma 3.6. *For each $f \in \mathrm{BV}_{\mathrm{loc}}$ with $f \ll 1/2$, there exist sequences $\{r_n\}_{n\in\mathbb{N}}$ and $\{R_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^1_{\mathrm{loc}}$, as well as a universal constant C such that*

$$h^0_n = \mathbb{E}\Big[f * \pi^{(0,G_n]}_n\Big] + \tfrac{1}{2}\mathbb{E}\Big[(h^0_n)^2 * \pi^{(0,G_n]}_n\Big] + r_n + R_n,$$

and

$$|r_n|_{\mathcal{L}^1[0,T]} \to 0 \text{ and } |R_n|_{\mathcal{L}^1[0,T]} \le C\mathbb{E}\Big[\big|h^{G_n}_n - h^0_n\big|_{\mathcal{L}^1[0,T]}\Big] \text{ for all } T \ge 0.$$

Proof. The equation (3.24) of Lemma 3.5 can be rewritten for $m = 0$ as

$$h^0_n = \mathbb{E}\Big[f * \pi^{(0,G_n]}_n\Big] + \tfrac{1}{2}\mathbb{E}\Big[(h^0_n)^2 * \pi^{(0,G_n]}_n\Big] + r_n + R_n$$

where $r_n = r^0_n$ of Lemma 3.5 and

$$R_n = \tfrac{1}{2}\mathbb{E}\Big[\big((h^{G_n}_n)^2 - (h^0_n)^2\big) * \pi^{(0,G_n]}_n\Big].$$

By Lemma 3.5, we have $\lim_n r^0_n = 0$ in $\mathrm{BV}[0,T]$, and, therefore, also in $\mathcal{L}^1[0,T]$. Thanks to inequality (3.14), the fact that $|h^m_n|_{\mathcal{S}^\infty[0,T]}$ is uniformly bounded over m, n, Lemma 3.3 and equation (3.28), we have

$$\begin{aligned}|R_n|_{\mathcal{L}^1[0,T]} &\le \tfrac{1}{2}\mathbb{E}\Big[\Big|\big((h^{G_n}_n)^2 - (h^0_n)^2\big) * \pi^{(0,G_n]}_n\Big|_{\mathcal{L}^1[0,T]}\Big] \\ &\le \tfrac{1}{2}\mathbb{E}\Big[\big|(h^{G_n}_n)^2 - (h^0_n)^2\big|_{\mathcal{L}^1[0,T]}\pi^{(0,G_n]}_n[0,T]\Big] \le C\mathbb{E}\Big[\big|h^{G_n}_n - h^0_n\big|_{\mathcal{L}^1[0,t]}\Big].\end{aligned}$$

□

Lemma 3.7. *Suppose that the condition (3) of Theorem 3.1 holds, the array $\{\pi^m_n\}_{m,n\in\mathbb{N}}$ is null, and that $f \in \mathrm{BV}_{\mathrm{loc}}$ with $f \ll 1/2$ is given. Then*

$$\lim_n \sup_m \big|h^m_n - h^0_n\big|_{\mathcal{L}^1[0,T]} = 0 \text{ for each } T \ge 0. \tag{3.29}$$

Proof. We assume, first, that *(3a)* holds with period $d \in \mathbb{N}$. The inequality (3.23) of Lemma 3.5 implies that for each $T \ge 0$ we have

$$\begin{aligned}\limsup_n \sup_{m,k}\big|h^{m+k}_n - h^m_n\big|_{\mathcal{L}^1[0,T]} &= \limsup_n \max_{m_1,m_2\le d}|h^{m_1}_n - h^{m_2}_n|_{\mathcal{L}^1[0,T]} \\ &\le 2C(d-1)\limsup_n\Big(\varepsilon_n + \max_{m\le d} W^1_{[0,T]}(\pi^m_n, \delta_0)\Big).\end{aligned}$$

Since $\{\pi^m_n\}_{m,n\in\mathbb{N}}$ is a null array, for each $\varepsilon > 0$ there exists a sequence $\kappa_n(\varepsilon) \searrow 0$ such that

$$\pi^m_n([0,t]) \ge 1 - \kappa_n(\varepsilon) \text{ for } t \ge \varepsilon \text{ and } m \in \mathbb{N}.$$

Therefore, for each $\varepsilon > 0$ we have

$$\limsup_n \sup_m W^1_{[0,T]}(\pi^m_n, \delta_0) = \limsup_n \sup_m \int_0^T (1 - \pi^m_n([0,t]))\, dt$$
$$\leq \varepsilon + (T-\varepsilon)^+ \limsup_n \kappa_n(\varepsilon) = \varepsilon,$$

and (3.29) follows.

Next, we establish the statement under the condition *(3b)*. Let $m \in \mathbb{N}$ be fixed throughout the proof. Using the terminology of subsection 2.4 above, we note that the genealogical trees corresponding to $\tilde{H}^m_n$ and $\tilde{H}^0_n$ have the same distribution. Indeed, this distribution depends only on the parameter a_n, and not on the probabilities $\{\pi^m_n\}_{m\in\mathbb{N}}$. This observation allows us to couple $\tilde{H}^m_n$ and $\tilde{H}^0_n$ on the same probability space by following the alternative construction of a Hawkes process in items (1)-(3) of subsection 2.4 as follows. We start by constructing a single Bienaymé-Galton-Watson tree $\mathcal{T}$ as in of item (1) which will be common to both $\tilde{H}^m_n$ and $\tilde{H}^0_m$. For each vertex v of the tree $\mathcal{T}$, let $(0, v]$ denote the set of edges on the (unique) directed path from v to the root, and let $|v|$ denote the generation number of v, i.e., the cardinality of $(0, v]$. For a directed edge e from child v to parent w, we set $|e| = |w|$.

It is straightforward to see that, given two probability measures ν_1 and ν_2 on $[0,\infty)$, we have

$$W^1_{[0,T]}(\nu_1, \nu_2) = \inf \mathbb{E}\Big[|L_1 \wedge T - L_2 \wedge T|\Big] \tag{3.30}$$

where the infimum is taken over all random vectors (L_1, L_2) with marginals ν_1 and ν_2. The same argument as in the classical ($T = +\infty$) case (see [San15, Theorem 1.4., p. 5, Proposition 2.17, p. 66]) can be used to show that the infimum is attained at some pair (L_1, L_2). Conditionally on the tree $\mathcal{T}$, for each edge e we construct the "lengths" $L(e)$ and $L'(e)$ so that their joint distribution attains the infimum in (3.30) for $\nu_1 = \pi^{m+|e|}_n$ and $\nu_2 = \pi^{|e|}_n$. Conditionally on the tree $\mathcal{T}$, the pairs $(L(e), L'(e))$ are then chosen independently of each other over all edges e.

Since the total number of points depends only on the genealogical tree, we have $|\tilde{H}^m_n| = |\tilde{H}^0_n|$ in this coupling, and, consequently $Z^m_n = Z^0_n$, where Z^m_n, $m \in \mathbb{N}_0$, are defined in (3.19). We denote these common values by $|\tilde{H}_n|$ and Z_n, respectively and use $\mathbb{E}[\cdot]$ to denote the expectation operator on the common probability space for $\tilde{H}^m_n$ and $\tilde{H}^0_n$. As at the beginning of the proof of Lemma 3.4 above, we have

$$\left|h^m_n - h^0_n\right|_{\mathcal{L}^1[0,T]} = \varepsilon_n^{-1} \int_0^T \mathbb{E}\Big[\Big|e^{(f*\tilde{\xi}^m_n)(t)} - e^{(f*\tilde{\xi}^0_n)(t)}\Big|\Big]\, dt$$
$$\leq \varepsilon_n^{-1} \mathbb{E}\Big[Z_n \int_0^T \Big|f * (\tilde{\xi}^m_n - \tilde{\xi}^0_n)(t)\Big|\, dt\Big]$$
$$= \varepsilon_n \mathbb{E}\Big[Z_n \Big|f * \tilde{H}^m_n - f * \tilde{H}^0_n\Big|_{\mathcal{L}^1[0,T]}\Big].$$

By inequality (3.16) of Lemma 3.3 above, we have

$$\left|h^m_n - h^0_n\right|_{\mathcal{L}^1[0,T]} \leq \varepsilon_n |f|_{\mathrm{BV}[0,T]} \mathbb{E}\Big[Z_n W^1_{[0,T]}(\tilde{H}^m_n, \tilde{H}^0_n)\Big]. \tag{3.31}$$

Seen ω-by-ω, the measures H^m_n and H^0_n are, respectively, sums of Dirac masses at points

$$T(v) = \sum_{e\in(0,v]} L(e) \text{ and } T'(v) = \sum_{e\in(0,v]} L'(e),\ v \in \mathcal{T}.$$

It follows easily from the definition that

$$W^1_{[0,T]}(\tilde{H}^m_n, \tilde{H}^0_n) \le \sum_v \big|T(v)\wedge T - T'(v)\wedge T\big| \le \sum_v \big|T(v) - T'(v)\big|$$

so that

$$\mathbb{E}\Big[W^1_{[0,T]}(\tilde{H}^m_n, \tilde{H}^0_n)\ \Big|\ \mathcal{T}\Big] \le \sum_{v\in\mathcal{T}} \mathbb{E}\big[\big|T(v) - T'(v)\big|\ \big|\ \mathcal{T}\big]$$

For $v \in \mathcal{T}$, we observe that we can write the difference $T(v)-T'(v)$ as a sum of $|v|$ random variables $L(e) - L'(e)$, $e \in (0, v]$ so that, by condition *(3b)*, we have

$$\begin{aligned}\mathbb{E}\big[\big|T(v) - T'(v)\big|\ \big|\ \mathcal{T}\big] &\le \sum_{e\in(0,v]} \mathbb{E}\big[\big|L(e) - L(e')\big|\big] \le |H_n| \sup_{m,k} W^1_{[0,T]}(\pi^{m+k}_n, \pi^m_n)\\ &\le b_n\varepsilon^2_n|H_n|, \text{ where } b_n \to 0 \text{ as } n \to \infty.\end{aligned}$$

Consequently, by (3.9) with $k = 2$, we have

$$\varepsilon_n\mathbb{E}\Big[Z_nW^1_{[0,T]}(\tilde{H}^m_n, \tilde{H}^0_n)\Big] \le C\varepsilon^3_n b_n\mathbb{E}\Big[Z_n|\tilde{H}_n|^2\Big] \le Cb_n \to 0,$$

which, in view of (3.31), completes the proof. □

Lemma 3.8. *Suppose that conditions* (2) *and* (3) *of Theorem 3.1 hold, and that* $f \in \mathrm{BV}_{\mathrm{loc}}$ *with* $f \ll 1/2$ *is given. If* $\{Dh^0_n\}_{n\in\mathbb{N}}$ *converges vaguely towards* $Dh \in \mathcal{M}_{\mathrm{s}}$*, possibly only through a subsequence, then* $h = Dh^0[0,\cdot] \in \mathrm{BV}_{\mathrm{loc}}$ *is the unique solution in* $\mathcal{S}^\infty_{\mathrm{loc}}$ *of the equation*

$$h = (f + \tfrac{1}{2}h^2) * \rho. \tag{3.32}$$

Proof. We assume for notational reasons and without loss of generality that the whole sequence $\{Dh^0_n\}_{n\in\mathbb{N}}$ converges vaguely towards Dh. The portmanteau theorem then ensures that for all $t \ge 0$, except at most countably many, we have

$$h^0_n(t) = Dh^0_n[0,t] \to Dh[0,t] = h(t).$$

Thanks to the uniform bound (3.17) of Lemma 3.4, this establishes the convergence $h^0_n \to h$ in $\mathcal{L}^1_{\mathrm{loc}}$. Lemma 3.7 implies that for each $T \ge 0$

$$\mathbb{E}\big[\big|h^{G_n} - h^0_n\big|\big]_{\mathcal{L}^1[0,T]} \to 0,$$

and so, thanks to Lemma 3.6, we have

$$h^0_n - f * \rho_n - \tfrac{1}{2}(h^0_n)^2 * \rho_n \to 0 \text{ in } \mathcal{L}^1[0,T] \text{ for each } T \ge 0.$$

Condition *(2)* of Theorem 3.1 implies that $W^1_{[0,T]}(\rho_n, \rho) \to 0$ for all $T \ge 0$. Therefore, by the estimate (3.16) of Lemma 3.3 we have

$$|f * (\rho_n - \rho)|_{\mathcal{L}^1[0,T]} \le |f|_{\mathrm{BV}[0,T]}W^1_{[0,T]}(\rho_n, \rho) \to 0 \text{ for each } T \ge 0.$$

Similarly, inequalities (3.14) and (3.16) of Lemma 3.3, together with Lemma 3.4 and the estimate (3.28), yield

$$\begin{aligned}\left|(h_n^0)^2 * \rho_n - h^2 * \rho\right|_{\mathcal{L}^1[0,t]} &\leq \left|\left((h_n^0)^2 - h^2\right) * \rho\right|_{\mathcal{L}^1[0,T]} + \left|(h_n^0)^2 * (\rho - \rho_n)\right|_{\mathcal{L}^1[0,T]} \\ &\leq \left|(h_n^0)^2 - h^2\right|_{\mathcal{L}^1[0,T]} + \left|(h_n^0)^2\right|_{\mathrm{BV}[0,T]} W^1_{[0,T]}(\rho, \rho_n) \\ &\leq C|f|_{\mathcal{L}^1[0,T]}\left|h_n^0 - h\right|_{\mathcal{L}^1[0,T]} + C|f|_{\mathcal{S}^\infty[0,T]}|f|_{\mathrm{BV}[0,T]} W^1_{[0,T]}(\rho, \rho_n) \to 0,\end{aligned}$$

for each $T \geq 0$. It follows that h satisfies (3.32), Leb-a.e. Since h is in $\mathrm{BV}_{\mathrm{loc}}$, and, therefore, right continuous and the same is true for $(f + \frac{1}{2}h^2) * \rho$, we conclude that (3.32) holds everywhere. Uniqueness is established in Proposition A.4 in the Appendix. ☐

Conclusion of the proof of Theorem 3.1. Condition *(1)* of the theorem, expression (2.18) and the portmanteau theorem imply that

$$\begin{aligned}\limsup_n \mathbb{E}[\xi_n[0,T]] = \limsup_n \varepsilon_n^2 \mathbb{E}\big[H_n^0[0,T]\big] &= \limsup_n \varepsilon_n (\rho_n^0 * \mu_n)[0,T] \\ &\leq \limsup_n \rho_n^0[0,T]\ \limsup_n (\varepsilon_n \mu_n[0,T]) \\ &\leq \rho[0,T]\,\mu[0,T] < \infty.\end{aligned} \tag{3.33}$$

Thanks to the standard tightness criterion (see, e.g., [Kal17, Theorem 4.10, p. 118]) for weak convergence of random measures, the bound (3.33) implies that the sequence $\{\xi_n\}_{n\in\mathbb{N}} := \{\xi_n^0\}_{n\in\mathbb{N}}$ is tight. We pick an arbitrary convergent subsequence of $\{\xi_n\}_{n\in\mathbb{N}}$, and, taking the usual liberty of not relabeling the indices, we denote this sequence by $\{\xi_n\}_{n\in\mathbb{N}}$ and its limit by ξ. To show the convergence of the original sequence it will be enough to show that the limit ξ does not depend on the choice of the convergent subsequence of $\{\xi_n\}_{n\in\mathbb{N}}$.

We pick $f \in \mathrm{BV}_{\mathrm{loc}}$ with $f \ll 1/2$ and associate to it the array $\{h_n^m\}_{m,n} = \{h_n^m[f]\}_{m,n}$, as in (3.7) above. For each $T \geq 0$, Lemma 3.4 implies that the restrictions to $[0,T]$ of the elements of the sequence $\{h_n^0\}_{n\in\mathbb{N}}$ are bounded in $\mathrm{BV}[0,T]$. By Prokhorov's theorem, we can pass to a subsequence, if necessary, to conclude that the restrictions $Dh_n^0|_{[0,T]}$ converge weakly on $[0,T]$ towards a signed measure Dh with $|Dh|_{\mathcal{M}_{\mathrm{s}}[0,T]} \leq C|f|_{\mathrm{BV}[0,T]}$. Lemma 3.8 then implies that the limit $h := Dh[0,\cdot]$ uniquely satisfies (3.32). In particular, h does not depend on the specific choices of subsequences made above so that no passage to a subsequence is necessary when Prokhorov's theorem is used.

The vague convergence $\varepsilon_n\mu_n \to \mu$ implies that $W^1_{[0,T]}(\varepsilon_n\mu_n, \mu) \to 0$ for each $T > 0$ and we can use inequalities (3.14) and (3.16) to conclude that

$$\begin{aligned}|\varepsilon_n h_n * \mu_n - h * \mu|_{\mathcal{L}^1[0,T]} &\leq |(h_n - h) * \varepsilon_n\mu_n|_{\mathcal{L}^1[0,T]} + |h * (\varepsilon_n\mu_n - \mu)|_{\mathcal{L}^1[0,T]} \\ &\leq C|h_n - h|_{\mathcal{L}^1[0,T]} + |h|_{\mathrm{BV}[0,T]} W^1_{[0,T]}(\varepsilon_n\mu_n, \mu) \to 0.\end{aligned}$$

Assume, next, that $f \in C_0 \cap \mathrm{BV}_{\mathrm{loc}}$ and $f \leq 0$. The weak convergence $\xi_n \to \xi$ implies that

$$M_\xi[f] = \lim_n M_{\xi_n}[f] = \lim_n \exp(\varepsilon_n h_n * \mu_n) = \exp(h * \mu), \text{a.e.,} \tag{3.34}$$

where a passage to a subsequence, if necessary, is made to guarantee a.e.-convergence of $\varepsilon_n h_n * \mu_n$ to $h * \mu$. The convolution $f * \xi$ is in $\mathrm{BV}_{\mathrm{loc}}$ and, therefore, right continuous a.s. Since $f \leq 0$, this right continuity is inherited by $M_\xi[f]$ by the bounded convergence theorem. Consequently, (3.34) can be strengthened to its pointwise version

$$M_\xi[f] = \exp(h[f] * \mu) \text{ for all } f \in C_0 \cap \mathrm{BV}_{\mathrm{loc}} \text{ with } f \le 0. \tag{3.35}$$

Hence, given any two subsequential limits ξ and ξ' of $\{\xi_n\}_{n\in\mathbb{N}}$, the functionals M_ξ and $M_{\xi'}$ agree on the set of all $f \in C \cap \mathrm{BV}_{\mathrm{loc}}$ with $f \le 0$. By density, the same is true for all $f \in C$ with $f \le 0$ which is, in turn, enough to conclude that ξ and ξ' have the same law as random elements in $\mathcal{M}$ (see [Kal17, Theorem 2.2, p. 52]). As mentioned at the beginning, this implies that the full sequence $\{\xi_n\}_{n\in\mathbb{N}}$ converges in law towards the random measure ξ.

Next, we show that (3.35) holds when $f \ll 1/2$, and not only for $f \le 0$. The vague convergence $\xi_n \to \xi$ implies that for any $f \in C_0 \cap \mathrm{BV}_{\mathrm{loc}}$, we have

$$(f * \xi_n)(t) \xrightarrow{w} (f * \xi)(t) \text{ for all } t \ge 0.$$

The expression (2.17) for the moment-generating function of H_n implies that for $p = (1-\delta)^{-1} > 1$ we have

$$\begin{aligned}
\log \mathbb{E}\Big[\Big(e^{(f*\xi_n)(t)}\Big)^p\Big] &= \log M_{H_n}[\varepsilon_n^2 p f](t) = \Big((M_{\tilde{H}_n}[\varepsilon_n^2 p f] - 1) * \mu_n\Big)(t) \\
&\le \left(\mathbb{E}\left[e^{\frac{1}{2}\varepsilon_n^2 \tilde{H}_n[0,\cdot]} - 1\right] * \mu_n\right)(t) \\
&\le \varepsilon_n^{-1}\mathbb{E}\left[e^{\frac{1}{2}\varepsilon_n^2 |\tilde{H}_n|} - 1\right] \varepsilon_n \mu_n[0,t].
\end{aligned}$$

Both terms after the last inequality are bounded in n; the first one by (3.8), and the second one as in (3.33) above. It follows that $\{\exp((f * \xi_n)(t))\}_{n\in\mathbb{N}}$ is uniformly integrable for each $t \ge 0$ so that $M_{\xi_n}[f] \to M_\xi[f]$ pointwise for all $f \in C_0 \cap \mathrm{BV}_{\mathrm{loc}}$ with $f \ll 1/2$. Moreover, it the monotone convergence theorem implies that

$$\mathbb{E}\big[\exp\big(\tfrac{1}{2}\xi[0,T]\big)\big] < \infty \text{ for all } T \ge 0, \tag{3.36}$$

This allows us to conclude, as above, that $M_\xi[f]$ is a right-continuous function for all $f \in \mathrm{BV}_{\mathrm{loc}}$ with $f \le 1/2$ and reuse the argument following (3.35).

The last step is to show that $M_\xi[f] = \exp(h[f] * \mu)$ for $f \in C_0$ with $f \le 1/2$ without the additional requirement that $f \in \mathrm{BV}_{\mathrm{loc}}$. We pick such f, and choose a sequence $\{f_n\}_{n\in\mathbb{N}}$ in C^1 with $f_n(0) = 0$ such that $f \ll 1/2$ on $[0,T]$ and $|f - f_n|_{\mathcal{S}^\infty[0,T]} \to 0$, for each $T \ge 0$; note also that $f_n \in C \cap \mathrm{BV}_{\mathrm{loc}}$. The dominated convergence theorem implies, via (3.36), that $M_\xi[f_n] \to M_\xi[f]$ pointwise on $[0,\infty)$. On the other hand, Proposition A.4 in the Appendix guarantees that

$$\int_0^T |h[f_n] - h[f]|_{\mathcal{S}^\infty[0,t]}\, dt \to 0,$$

for each $T \ge 0$. The monotonicity of $t \mapsto |h[f_n] - h[f]|_{\mathcal{S}^\infty[0,t]}$ implies that $h[f_n] \to h[f]$ in $\mathcal{S}^\infty[0,T]$ for each $T \ge 0$. Since μ is locally bounded and $\{h[f_n]\}_{n\in\mathbb{N}}$ admits a uniform $\mathcal{S}^\infty[0,T]$ bound, the bounded convergence theorem implies that $h[f_n] * \mu \to h[f] * \mu$ pointwise, and, consequently, that $M_\xi[f] = \exp(h[f] * \mu)$.

4. Feller random measures and their properties

4.1. **Attainable limiting distributions ρ.**

We start by characterizing the family of all probability measures ρ that can arise as limits in Theorem 3.1.

Definition 4.1. A random variable X is said to be *geometrically infinitely divisible (GID)* if for each $p \in (0,1)$ there exists an iid sequence $\{X_m(p)\}_{m\in\mathbb{N}}$ of random variables such that

$$X \overset{(d)}{=} \sum_{m=1}^{G(p)} X_m(p),$$

where $G(p)$ is an $\mathbb{N}_0$-valued geometrically distributed random variable with parameter (probability of success) p, independent of the sequence $\{X_m(p)\}_{m\in\mathbb{N}}$.

This notion has been introduced in [KMM84] as a part of an answer to the following question of Zolotarev: characterize the family $\mathcal{Y}$ of distributions of random variables Y such that, for any $p \in (0,1)$, there exists a random variable $X(p)$ such that

$$Y \overset{(d)}{=} X(p) + B(p)Y$$

where Y, $X(p)$ and $B(p)$ are independent and $\mathbb{P}[B(p)=1] = 1 - \mathbb{P}[B(p)=0] = p$. In the same paper, the authors show that $\mathcal{Y}$ coincides with the set of all GID distributions. Furthermore, they show that a probability measure ρ on $[0,\infty)$ is GID if and only if its Laplace transform $\hat{\rho}$ has the form

$$\hat{\rho}(\lambda) = \frac{1}{1 - \log(\hat{\rho}'(\lambda))}, \tag{4.1}$$

where $\hat{\rho}'$ is the Laplace transform of some infinitely divisible distribution ρ' on $[0,\infty)$. Thanks to the Lévy-Khinchine representation, this is further equivalent to $\hat{\rho}$ admitting the following form

$$\hat{\rho}(\lambda) = \frac{1}{1 + L\lambda + \int_0^\infty (1 - e^{-\lambda t})\,\nu(dt)}\,. \tag{4.2}$$

for some constant $L \geq 0$ and some measure ν on $(0,\infty)$ with $\int_0^\infty \min(1,t)\nu(dt)$.

Proposition 4.2. *Let ρ be a GID distribution on $[0,\infty)$ with $\rho(\{0\}) = 0$ and let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,1)$ with $a_n \to 1$. Then there exists a sequence $\{\pi_n\}_{n\in\mathbb{N}}$ in $\mathcal{M}_{\mathrm{p}}$ such that the array $\{\pi_n^m\}_{m,n\in\mathbb{N}}$, given by $\pi_n^m = \pi_n$ satisfies the conditions of Theorem 3.1 and we have $\rho_n \xrightarrow{w} \rho$.*

Proof. Let $\rho \in \mathcal{M}_{\mathrm{p}}$ be a GID distribution, and let ρ' be an infinitely divisible probability measure on $[0,\infty)$ such that (4.1) holds. By infinite divisibility, for each $n \in \mathbb{N}$ we can find a sequence $\{X_n^m\}_{m\in\mathbb{N}}$ of iid nonnegative random variables such that

$$\sum_{k=1}^{m_n} X_n^k \sim \rho' \text{ where } m_n = \lfloor (1-a_n)^{-1} \rfloor. \tag{4.3}$$

We define $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ to be the row-wise constant array given by $\pi_n^m = \pi_n$, where π_n is the law of X_n^1.

First, we show that $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ satisfies the conditions of Theorem 3.1. Since π_n^m does not depend on m, the condition *(3a)* of Theorem 3.1 is trivially satisfied, so we are left to argue that $\pi_n^m(\{0\}) = 0$

for all m and n, and that $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ is null. Since $\rho(\{0\}) = 0$, we have $\lim_{\lambda\to\infty} \hat{\rho}(\lambda) = 0$ and so, by (4.1),

$$\lim_{\lambda\to\infty} \hat{\rho}'(\lambda) = 0.$$

Since $\hat{\pi}_n(\lambda) = (\hat{\rho}'(\lambda))^{1/m_n}$, we have $\lim_{\lambda\to\infty} \hat{\pi}_n(\lambda) = 0$ which implies that $\pi_n^m(\{0\}) = \pi_n(\{0\}) = 0$. Moreover, since $m_n \to \infty$ and $\hat{\rho}'(\lambda) > 0$ for $\lambda > 0$, we have

$$\lim_n \hat{\pi}_n(\lambda) = \lim_n (\hat{\rho}'(\lambda))^{1/m_n} = 1 \text{ for } \lambda > 0.$$

Therefore, $\pi_n \xrightarrow{w} \delta_{\{0\}}$ which implies that $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ is null.

Turning to the convergence $\rho_n \xrightarrow{w} \rho$, we notice that it is equivalent to the convergence

$$\sum_{m=1}^{G_n} X_n^m \to \rho \text{ in distribution,} \tag{4.4}$$

where $\{G_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathbb{N}_0$-geometric random variables with parameters $\{1-a_n\}_{n\in\mathbb{N}}$ and independent of $\{X_n^m\}_{m\in\mathbb{N}}$. We start from the following expressions

$$\mathbb{E}\left[\exp\left(-\lambda\sum_{m=1}^{m_n} X_n^m\right)\right] = (\hat{\pi}_n(\lambda))^{m_n} \text{ and } \mathbb{E}\left[\exp\left(-\lambda\sum_{m=1}^{G_n} X_n^m\right)\right] = \mathbb{E}\Big[(\hat{\pi}_n(\lambda))^{G_n}\Big],$$

and a straightforward-to-check fact that $G_n/m_n \to E$ in distribution, where E is an exponentially distributed random variable with parameter 1. Assuming, without loss of generality, that that $\{G_n\}_{n\in\mathbb{N}}$ and E are all coupled on a probability space where $G_n/m_n \to E$, a.s., we obtain

$$(\hat{\pi}_n(\lambda))^{G_n} = \big(\hat{\rho}'(\lambda)\big)^{G_n/m_n} \to \big(\hat{\rho}'(\lambda)\big)^E, \text{ a.s.} \tag{4.5}$$

The dominated convergence theorem allows us to pass the limit in (4.5) outside the expectation. This implies that

$$\mathbb{E}\left[\exp\left(-\lambda\sum_{m=1}^{G_n} X_n^m\right)\right] \to \mathbb{E}\Big[\big(\hat{\rho}'(\lambda)\big)^E\Big] = \hat{\rho}(\lambda),$$

which, in turn, implies (4.4). □

While it does add to modeling flexibility, allowing for general $\{\pi_n^m\}_{m,n\in\mathbb{N}}$, as the following proposition shows, does not enlarge the class of attainable distributions ρ.

Proposition 4.3. *Suppose that the measure ρ can arise as a limit $\rho = \lim_n \rho_n$ associated to an array $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ which satisfies the conditions of Theorem 3.1. Then ρ is GID.*

Proof. See [KK93, Theorem 5.1, p.116]. □

Remark 4.4.

(1) Propositions 4.2 and 4.3 above should be viewed in the context of the general "theory of random summation" (see, e.g., the monograph [GK96]) which establishes an almost complete analogy with the classical theory of triangular arrays of independent random variables.

(2) The proof of Proposition 4.2 above allows us to give sufficient conditions on the array $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ so that a particular GID distribution ρ is attained as a limit. Indeed, it suffices to choose them so that (4.3) holds for the infinitely divisible "counterpart" ρ' of ρ, given via its Laplace transform

$$\hat{\rho}'(\lambda) = \exp(1 - 1/\hat{\rho}(\lambda)), \lambda \geq 0.$$

Sufficient conditions for that, in turn, are classical and have been very well understood since the early days of probability (see, e.g., [GK54] or [Kal21, Chapter 7] for a more accessible modern treatment).

Proposition 4.2 states that any GID distribution can arise in the row-constant case $\pi_n^m = \pi_n$, for any scaling sequence $\{a_n\}_{n\in\mathbb{N}}$. When the dependence on n is restricted further, namely, so that each π_n is a rescaled version of the same probability distribution π (as is the case, e.g., in [JR15, JR16b]), the limiting distribution must belong, up to scaling, to a specific one-parameter family, and the sequence $\{a_n\}_{n\in\mathbb{N}}$ is essentially determined by it.

Definition 4.5. A probability measure ρ on $[0,\infty)$ is called the *Mittag-Leffler* distribution with parameter $\alpha \in (0,1]$ if its Laplace transform $\hat{\rho}$ takes the form

$$\hat{\rho}(\lambda) = \frac{1}{1+\lambda^\alpha}. \tag{4.6}$$

The Mittag-Leffler distribution admits an explicit density

$$p^\alpha(t) = t^{\alpha-1}E_{\alpha,\alpha}(-t^\alpha), t \geq 0, \tag{4.7}$$

where, for $\alpha, \beta > 0$, the *Mittag-Leffler function* $E_{\alpha,\beta}$ is given by

$$E_{\alpha,\beta}(x) = \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(\alpha n + \beta)} \; x \geq 0,$$

and Γ denotes the Gamma-function. In the special case $\alpha = 1$ this distribution is exponential, with rate 1, while for $\alpha = 1/2$ its density takes an especially simple form. Indeed, a straightforward, if a bit tedious, derivation directly from the definition yields

$$p^{1/2}(t) = \sqrt{\frac{2}{\pi}}\left(\sqrt{2t} - m\left(\frac{1}{\sqrt{2t}}\right)\right), \tag{4.8}$$

where $m(x) = \frac{1-\Phi(x)}{\varphi(x)}$ is the ratio (known as the Mill's ratio) of the survival function $1-\Phi$ and the density φ of the standard normal distribution.

Proposition 4.6. *Let ψ be a probability measure on $[0,\infty)$ with $\psi(\{0\}) = 0$ and let the array $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ be given by*

$$\pi_n^m(B) = \pi_n(B) = \psi(nB) \text{ for all } B \in \mathcal{B}[0,\infty). \tag{4.9}$$

Then $\{\pi_n^m\}_{m,n\in\mathbb{N}}$ satisfies the condition (2) of Theorem 3.1 if and only if one of the following two conditions are met:

(1) $\lim_{t\to\infty} \frac{\psi[t,\infty)}{\frac{1}{t}\int_0^t \psi[s,\infty)\,ds} = 0$, or

(2) there exists $\alpha \in (0,1)$ such that $\lim_{t\to\infty} \frac{\psi[t,\infty)}{\psi[ct,\infty)} = c^\alpha$ for all $c > 0$.

In either case, if $a_n = 1 - \kappa_n n^{-\alpha}$ for some $\kappa_n \to \kappa \in (0, \infty)$, ρ_n converges towards a (possibly scaled) Mittag-Leffler distribution with parameter α (where $\alpha = 1$ in case (1)).

Proof. This is, essentially, given in [GK96, Theorem 2.5.1, p. 35] for case *(1)* and [GK96, Theorem 2.5.2, p. 37] for case *(2)*. While not explicitly mentioned in the statements of these theorems, the behavior of the scaling sequence can be read off their proofs. □

Remark 4.7.

(1) In the case *(1)* the limiting distribution ρ is exponential and the condition is satisfied, in particular, if the probability measure ψ admits a finite first moment $\int_0^\infty t\, \psi(dt)$ as in [JR15].
(2) The case *(2)* covers all ψ such that $\psi[t, \infty)$ is a regularly varying function with a nontrivial tail, i.e.,

$$\psi[t, \infty) \sim l(t) t^{-\alpha} \text{ as } t \to \infty$$

for some $\alpha \in (0, 1)$ and some slowly varying (e.g., constant) function l.
(3) The choice of n as the scaling factor in (4.9) is simply a convenient normalization and can be easily generalized.

4.2. **Feller random measures.** If we combine the results of Theorem 3.1 and Proposition 4.2, we can conclude that for each $\mu \in \mathcal{M}_{\text{lf}}$ and each GID probability measure ρ with $\rho(\{0\}) = 0$, there exists a locally finite random measure ξ whose law is characterized by

$$M_\xi[f] = \exp(h[f] * \mu),\ f \in C_0 \text{ with } f \le 1/2, \tag{4.10}$$

where $h = h[f]$ is the unique solution in $\mathcal{S}^\infty_{\text{loc}}$ to the following convolutional Riccati equation

$$h = (f + \tfrac{1}{2} h^2) * \rho \text{ on } [0, \infty). \tag{4.11}$$

We call ξ the *Feller random measure* with parameters (μ, ρ), and denote this by $\xi \sim F(\mu, \rho)$. When it exists, a nonnegative measurable process $\{Y_t\}_{t \ge 0}$ such that

$$\xi[A] = \int_A Y_t\, dt \text{ for all } A \in \mathcal{B}([0, \infty)),\ \text{a.s.},$$

is called the *density* of ξ.

Remark 4.8. It has been shown in [JR15] and [JR16b] that when ρ is the Mittag-Leffler distribution with index $\alpha > 1/2$ and μ is the Lebesgue measure on $[0, \infty)$, the Feller random measure ξ admits a density Y which has the distribution of a solution to a Volterra-type stochastic differential equation of the form

$$Y_t = Y_0 + c_1 \int_0^t (t - s)^{\alpha - 1} (\theta - Y_s)\, ds + c_2 \int_0^t (t - s)^{\alpha - 1} \sqrt{Y_t}\, dB_t, \tag{4.12}$$

where c_1, c_2 and θ are constants and B is a Brownian motion. The form of (4.12) explains why Y is called the fractional CIR (or Feller) process in the literature, and also why we adopted the name Feller random measure for the general case. In addition to [JR15, JR16b], we refer the reader to [EER19] for further information on the fractional CIR process and to [JLP19] for a treatment of more general stochastic differential equations of the Volterra type.

4.3. **Distributional properties.**

4.3.1. *Cumulants.* We recall that a real sequence $\{\kappa_n[Y]\}_{n\in\mathbb{N}}$ is called the *sequence of cumulants* of (the distribution of) the random variable Y if

$$\mathbb{E}[\exp(\varepsilon Y)] = \exp\left(\sum_{n\geq 1} \frac{\varepsilon^n}{n!}\kappa_n[Y]\right)$$

for ε in some neighborhood of 0. For a pair (X, Y), we also define the *partial cumulants* $\kappa_n[X, Y]$ by

$$\mathbb{E}[\exp(X + \varepsilon Y)] = \exp\left(\sum_{n\geq 0} \frac{\varepsilon^n}{n!}\kappa_n[X, Y]\right)$$

provided the series converges for ε in some neighborhood of 0. As is well known, knowledge of cumulants of a distribution is tantamount to the knowledge of its moments. Indeed, the two sequences are related to one another via an explicit formulas based on Faà di Bruno's formula and involving Bell polynomials (see [Smi95]).

As our next result shows, cumulants of random variables of the form $f * \xi$, where $\xi \sim F(\rho, \mu)$, satisfy a simple recursive relationship and admit explicit representation. Partial cumulants admit a representation in terms of a solution of a system of convolutional equations.

Proposition 4.9. *Let ξ be a Feller random measure with parameters ρ and μ, and let $f \in C_0$.*

(1) The cumulants $\kappa_n[f * \xi]$ of $f * \xi$ are given by

$$\kappa_n[f * \xi] = n!\, K_n * \mu \text{ for } n \geq 1,$$

where the functions $K_n \in \mathcal{S}^\infty_{\mathrm{loc}}$ are defined recursively by

$$K_1 = f * \rho,\ K_n = \tfrac{1}{2}\Big(\textstyle\sum_{i=1}^{n-1} K_i K_{n-i}\Big) * \rho \text{ for } n \geq 2. \tag{4.13}$$

(2) For $f_0 \in C_0$ with $f \ll 1/2$, the partial cumulants $\kappa_n[f_0 * \xi, f * \xi]$ are given by

$$\kappa_n[f_0 * \xi, f * \xi] = n!\, K'_n * \mu \text{ for } n \geq 1,$$

where $\{K'_n\}_{n\in\mathbb{N}_0}$ is the unique solution in $(\mathcal{S}^\infty_{\mathrm{loc}})^{\mathbb{N}_0}$ of the system

$$\begin{aligned} K'_0 &= h[f_0],\\ K'_1 &= f + \big(K'_0 K'_1\big) * \rho \\ K'_n &= \tfrac{1}{2}\big(\textstyle\sum_{i=0}^{n} K'_i K'_{n-i}\big) * \rho \text{ for } n > 2. \end{aligned} \tag{4.14}$$

Proof. To obtain *(2)*, we simply combine the representation (4.10) with Proposition A.8 with $F = h[f_0]$ and $G = f$. The assertion in *(1)* is a special case of *(2)* with $F = h[0] = 0$ and $G = f$. In that case, the system (4.14) simplifies and turns into the recursive definition given in (4.13). □

The first three cumulants/moments are given below:

$$\begin{aligned} \kappa_1 &= \mathbb{E}[f * \xi] = (f * \rho) * \mu \\ \kappa_2 &= \mathrm{Var}[f * \xi] = \Big((f * \rho)^2 * \rho\Big) * \mu \\ \kappa_3 &= \mathbb{E}\Big[(f * \xi - \mathbb{E}[f * \xi])^3\Big] = 3\bigg(\Big(\big((f * \rho)((f * \rho)^2 * \rho)\big)\Big) * \rho\bigg) * \mu \end{aligned} \tag{4.15}$$

4.3.2. *Infinite divisibility.* The following "branching" property of $F(\mu,\rho)$ follows directly from the characterization (4.10), (4.11).

Proposition 4.10. *Suppose that $\xi_1 \sim F(\mu_1,\rho)$ and $\xi_2 \sim F(\mu_2,\rho)$ where μ_1,μ_2 are locally finite and ρ is a probability measure with $\rho(\{0\})=0$. If ξ_1 and ξ_2 are independent then*

$$\xi_1+\xi_2 \sim F(\mu_1+\mu_2,\rho).$$

Corollary 4.11. *Given $\xi \in F(\mu,\rho)$, the random variable $(f*\xi)(t)$ is infinitely divisible for all $f \in C_0$ and $t \geq 0$.*

We say that a stochastic process $\{Y_t\}_{t\geq 0}$ is infinitely divisible if the random vector $(Y_{t_1},\dots,Y_{t_n})$ is infinitely divisible for any $n\in\mathbb{N}$, $0 \leq t_1 < t_2 < \cdots < t_n < \infty$.

Corollary 4.12. *Suppose that $F(\frac{1}{N}\mu,\rho)$ admits a right-continuous density $Y^{(N)}$ for each $N \in \mathbb{N}$. Then $Y^{(1)}$ is infinitely divisible.*

4.3.3. *The Covariance Structure.* The polarization identity and the expression for κ_2 in (4.15) yield

$$\mathrm{Cov}[f*\xi, g*\xi] = \Big(\big((f*\rho)(g*\rho)\big)*\rho\Big)*\mu, \tag{4.16}$$

for $f,g\in C_0$. We can rewrite (4.16) as

$$\mathrm{Cov}[f*\xi, g*\xi](t) = \iint f(t-r)g(t-s)\gamma(dr,ds), \tag{4.17}$$

where

$$\gamma(dr,ds) = \int \rho(dr-u)\rho(ds-u)(\rho*\mu)(du)$$

i.e., $\gamma(B) = \int\Big(\iint 1_{B+(u,u)}(s,r)\rho(ds)\rho(dr)\Big)(\rho*\mu)(du)$ for $B\in\mathcal{B}([0,\infty)\times[0,\infty))$. In the special case when ρ admits a density p with respect to Lebesgue measure, the measure γ is absolutely continuous and

$$\gamma(dr,ds) = \Sigma(r,s)\,dr\,ds \text{ where } \Sigma(r,s) = \int p(r-u)p(s-u)(p*\mu)(u)\,du. \tag{4.18}$$

A further specialization yields tight asymptotics around the "diagonal" $r=s$. For two functions $f: D_f\subseteq\mathbb{R}^d\to\mathbb{R}$ and $g: D_g\subseteq\mathbb{R}^d\to\mathbb{R}$ we write $f\approx g$ if for each bounded $B\in\mathcal{B}(\mathbb{R}^d)$ there exists a strictly positive constant C such that $f\leq Cg$ and $g\leq Cf$ on $D_f\cap D_g\cap B$.

Proposition 4.13. *Suppose that ρ is a Mittag-Leffler distribution with parameter $\alpha\in(0,1]$, and that μ is the Lebesgue measure on $[0,\infty)$. Then*

$$\Sigma(r,s) \approx_\alpha (r,s) \tag{4.19}$$

where $\Gamma_\alpha : \big\{(s,r)\in(0,\infty)^2 \,:\, s\neq r\big\}\to(0,\infty)$ is a symmetric function defined for $r<s$ by

$$B_\alpha(r,s) = r^{2\alpha}s^{\alpha-1}\begin{cases}\left(1-\frac{r}{s}\right)^{2\alpha-1}, & \alpha<\frac{1}{2},\\ 1-\log(1-\frac{r}{s}), & \alpha=\frac{1}{2},\\ 1, & \alpha>\frac{1}{2}.\end{cases} \tag{4.20}$$

Proof. Being entire, the Mittag-Leffler function $E_{\alpha,\alpha}$ satisfies $E_{\alpha,\alpha} \approx 1$, and so, thanks to (4.7), we have $p(t) \approx t^{\alpha-1}$. Moreover, since μ is the Lebesgue measure, we have $(p * \mu)(t) \approx t^{\alpha}$ so that

$$p(r-u)p(s-u)(p*\mu)(u) \approx (r-u)^{\alpha-1}(s-u)^{\alpha-1}u^{\alpha}.$$

Therefore, by (4.18), we have

$$\Sigma(r,s) \approx \int_0^r (r-u)^{\alpha-1}(s-u)^{\alpha-1}u^{\alpha}\,du = r^{2\alpha}s^{\alpha-1}\int_0^1 (1-w)^{\alpha-1}(1-\tfrac{r}{s}w)^{\alpha-1}w^{\alpha}dw$$

According to [DLM, eq. (15.6.1)], we have

$$\int_0^1 (1-w)^{\alpha-1}(1-\tfrac{r}{s}w)^{\alpha-1}w^{\alpha}dw = \Gamma(\alpha+1)\Gamma(\alpha)\,{}_2F_1\Big(1-\alpha,\alpha+1;2\alpha+1;\frac{r}{s}\Big),$$

where ${}_2F_1$ denotes the hypergeometric function. Since ${}_2F_1$ is entire, we have ${}_2F_1 \approx 1$ which, coupled with different asymptotic regimes described in [DLM, §15.4.2], implies that

$$\begin{aligned} &{}_2F_1(1-\alpha,\alpha+1;2\alpha+1;x) \approx (1-x)^{2\alpha-1} && \text{for } \alpha<\tfrac12,\\ &{}_2F_1(1-\alpha,\alpha+1;2\alpha+1;x) \approx 1-\log(1-x) && \text{for } \alpha=\tfrac12, \text{ and}\\ &{}_2F_1(1-\alpha,\alpha+1;2\alpha+1;x) \approx 1 && \text{for } \alpha>\tfrac12, \end{aligned}$$

which, in turn, establishes (4.19). □

Corollary 4.14. *Suppose that ρ is a Mittag-Leffler distribution with parameter $\alpha \in (0,1]$, and that μ is the Lebesgue measure on $[0,\infty)$. Then $\xi \sim F(\mu,\rho)$ admits a square-integrable density if and only if $\alpha > 1/2$.*

Appendix A. The convolutional Riccati equation

We collect in this appendix several properties of the solutions of the convolutional Riccati equation

$$K = F + \tfrac12 K^2 * \rho \tag{A.1}$$

used throughout the paper. We fix $T \geq 0$ and focus on functions defined $[0,T]$. Extensions to locally defined spaces, such as $\mathcal{S}^{\infty}_{\text{loc}}$ are straightforward. We assume that $\rho \in \mathcal{M}_{\text{p}}$ satisfies $\rho(\{0\}) = 0$, but do not put any other restrictions on it. A constant depending only on a quantities $q_1, q_2, \dots$ is denoted by $C(q_1, q_2, \dots)$ and can change from occurrence to occurrence.

A.1. **Comparison, bounds and stability.**

Lemma A.1. *Suppose that $K \in \mathcal{S}^{\infty}[0,T]$ satisfies*

$$K \geq (QK) * \rho \text{ on } [0,T]. \tag{A.2}$$

for some $Q \in \mathcal{S}^{\infty}[0,T]$ with $Q \geq 0$. Then $K \geq 0$.

Proof. We set $F = (QK)*\rho - K \geq 0$, so that $K = F + (QK)*\rho$ on $[0,T]$. Given $t > 0$, we define the operator $\mathcal{A}_t$ by $\mathcal{A}_t L = F + (QL)*\rho$ for $L \in \mathcal{S}^{\infty}[0,t]$ so that,

$$\begin{aligned} |\mathcal{A}_t L_2 - \mathcal{A}_t L_1|_{\mathcal{S}^{\infty}[0,t]} &\leq\in \varepsilon^t Q(t-s)|L_2(t-s) - L_1(t-s)|\rho(ds)\\ &\leq \Big(|Q|_{\mathcal{S}^{\infty}}\rho[0,t]\Big)\,|L_2 - L_1|_{\mathcal{S}^{\infty}[0,t]} \text{ for } L_1, L_2 \in \mathcal{S}^{\infty}[0,t]. \end{aligned}$$

Since $\rho[0,t] \to \rho(\{0\}) = 0$ as $t \searrow 0$, there exists $\varepsilon > 0$ (which depends only on an upper bound on $|Q|_{\mathcal{S}^\infty[0,T]}$ and ρ, but not on F) such that $\mathcal{A}_\varepsilon$ is a contraction on $\mathcal{S}^\infty_{[0,\varepsilon]}$. For such ε, we have $K = \lim_n \mathcal{A}^n_\varepsilon(F)$ in $\mathcal{S}^\infty_{[0,\varepsilon]}$, and, since $\mathcal{A}_\varepsilon$ is a positive operator and $F \geq 0$, it follows that $K \geq 0$ on $[0,\varepsilon]$.

To extend the conclusion of the previous paragraph from $[0,\varepsilon]$ to the entire $[0,T]$, we assume, without loss of generality, that $T = N\varepsilon$ for some $N \in \mathbb{N}$. If $N = 1$, we are done. Otherwise, for $i \in \{0,\dots,N-1\}$ we set $K^i(t) = K(i\varepsilon + t)$, $Q^i(t) = Q(i\varepsilon + t)$ and $F^i(t) = F(i\varepsilon + t) + \int_0^{i\varepsilon} Q(s)K(s)\rho(i\varepsilon + t - ds)$ for $t \in [0,\varepsilon]$ and observe that

$$\begin{aligned}
K^i(t) &= F(i\varepsilon + t) + \left(\int_0^{i\varepsilon} + \int_{i\varepsilon}^{i\varepsilon+t}\right) Q(s)K(s)\,\rho(i\varepsilon + t - ds)\\
&= F^i(t) + \int_{i\varepsilon}^{i\varepsilon+t} Q(s)K(s)\rho(i\varepsilon + t - ds)\\
&= F^i(t) + \int_0^t Q^i(u)K^i(u)\rho(t-du) = F^i(t) + \big((Q^iK^i)*\rho\big)(t).
\end{aligned}$$

Assuming, as the induction hypothesis, that $K \geq 0$ on $[0,i\varepsilon]$, we have $F^i(t) \geq 0$ on $[0,\varepsilon]$. Moreover, $\left|Q^i\right|_{\mathcal{S}^\infty[0,\varepsilon]} \leq |Q|_{\mathcal{S}^\infty[0,T]}$, so we can use the result of the previous paragraph to conclude that $K^i \geq 0$ on $[0,\varepsilon]$, i.e., that $K \geq 0$ on $[0,(i+1)\varepsilon]$. Therefore, $K \geq 0$ on $[0,T]$. □

Proposition A.2. *Suppose that that $F_1, F_2, K_1, K_2 \in \mathcal{S}^\infty[0,T]$ are such that*

$$K_1 \leq F_1 + \tfrac{1}{2}K_1^2 * \rho \text{ and } K_2 \geq F_2 + \tfrac{1}{2}K_2^2 * \rho.$$

If $F_1 \leq F_2$ and $K_1 + K_2 \geq 0$ then $K_1 \leq K_2$.

Proof. We observe that the function $K = K_2 - K_1$ satisfies the inequality

$$K \geq F_2 - F_1 + \tfrac{1}{2}(K_2^2 - K_1^2) \geq (QK) * \rho \text{ where } Q = \tfrac{1}{2}(K_1 + K_2).$$

Since $Q \geq 0$, by the assumption, Lemma A.1 above can be applied to conclude that $K \geq 0$, i.e., $K_2 \geq K_1$. □

Lemma A.3. *Suppose that $F \in \mathcal{S}^\infty[0,T]$ satisfies $|F| \leq 1/2$ and that K solves (A.1). Then*

$$-|F|_{\mathcal{S}^\infty[0,t]} \leq K(t) \leq 1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,t]}} \text{ for all } t \in [0,T]. \tag{A.3}$$

Proof. To get the lower bound, we simply observe that $K \geq F$. For the upper bound, we define

$$K_2(t) = 1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,t]}}.$$

Since K_2 is nonnegative and nondecreasing, we have $K_2^2 * \rho \leq K_2^2$, and, so,

$$K_2(t) - \tfrac{1}{2}(K_2^2 * \rho)(t) \geq K_2(t) - \tfrac{1}{2}K_2^2(t) = |F|_{\mathcal{S}^\infty[0,t]} \geq F(t)$$

We have

$$(K + K_2)(t) \geq -|F|_{\mathcal{S}^\infty[0,t]} + 1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,t]}} = \tfrac{1}{2}\Big(1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,t]}}\Big)^2 \geq 0,$$

which allows us to use Proposition A.2 above with $K_1 = K$ and $F_1 = F_2 = F$ to conclude that $K \leq K_2$. □

Proposition A.4. *Suppose that $M \geq 0$ and $K_i, F_i \in \mathcal{S}^\infty[0,T]$ are such that $|K_i|_{\mathcal{S}^\infty[0,T]} \leq M$ for $i = 1,2$ and*

$$K_i = F_i + \tfrac{1}{2} K_i^2 * \rho \text{ for } i = 1, 2, \tag{A.4}$$

then

$$\int_0^T |K_2 - K_1|_{\mathcal{S}^\infty[0,t]}\, dt \leq C(\rho, M, T) \int_0^T |F_1 - F_2|_{\mathcal{S}^\infty[0,t]}\, dt. \tag{A.5}$$

In particular, (A.1) *has at most one solution in $\mathcal{S}^\infty[0,T]$ for $F \in \mathcal{S}^\infty[0,T]$.*

Proof. For $t \in [0,T]$ we define $m(t) = |K_2 - K_1|_{\mathcal{S}^\infty[0,t]}$ and $m_F(t) = |F_2 - F_1|_{\mathcal{S}^\infty[0,t]}$. For $s \leq t \leq T$ we have

$$\begin{aligned} |K_2(s) - K_1(s)| &\leq |F_2(s) - F_1(s)| + \tfrac{1}{2} \int_0^s \left| K_2^2(s-u) - K_1^2(s-u) \right| \rho(du) \\ &\leq m_F(s) + M \int_0^s m(s-u)\, \rho(du), \end{aligned}$$

so that

$$m \leq m_F + M(m * \rho) \text{ on } [0,T].$$

We multiply both sides by $\exp(-\lambda \cdot)$ and integrate on $[0,T]$ to obtain

$$\begin{aligned} \int_0^T e^{-\lambda t} m(t)\, dt &\leq \int_0^T e^{-\lambda t} m_F(t)\, dt + M \int_0^T \int_0^t e^{-t-u} m(t-u)\, e^{-\lambda u} \rho(du)\, dt \\ &= \int_0^T e^{-\lambda t} m_F(t)\, dt + M \int_0^T \int_0^{T-u} e^{-\lambda s} m(s)\, ds\, e^{-\lambda u} \rho(du) \\ &\leq \int_0^T e^{-\lambda t} m_F(t)\, dt + M \left(\int_0^T e^{-\lambda t} m(t)\, dt \right) \left(\int_0^T e^{-\lambda u} \rho(du) \right) \\ &\leq \int_0^T e^{-\lambda t} m_F(t)\, dt + \left(M \int_0^T e^{-\lambda u} \rho(du) \right) \int_0^T e^{-\lambda t} m(t)\, dt \end{aligned}$$

Since $\rho(\{0\}) = 0$, there exists $\lambda_0 = \lambda_0(\rho, M) \geq 0$ such that

$$M \int_0^T e^{-\lambda_0 u} \rho(du) \leq 1/2.$$

With such λ_0, we have

$$e^{-\lambda_0 T} \int_0^T m(t)\, dt \leq \int_0^T e^{-\lambda_0 t} m(t)\, dt \leq 2 \int_0^T e^{-\lambda_0 t} m_F(t)\, dt \leq 2 \int_0^T m_F(t)\, dt,$$

which implies (A.5) with the constant $2\exp(\lambda_0 T)$.

To prove uniqueness note that for $F_1 = F_2 = F$ (A.5) yields

$$\int_0^T |K_2 - K_1|_{\mathcal{S}^\infty[0,t]}\, dt = 0,$$

which, in turn, implies that $|K_2 - K_1|_{\mathcal{S}^\infty[0,t]} = 0$, a.e., on $[0,T]$. It follows, by monotonicity, that $K_1 = K_2$ on $[0,T)$. The missing equality $K_1(T) = K_2(T)$ is a consequence of the assumption $\rho(\{0\}) = 0$ since

$$K_2(T) - K_1(T) = \tfrac{1}{2} \int_{(0,T]} \big(K_2^2(T-u) - K_1^2(T-u)\big)\, \rho(du) = 0. \qquad \square$$

A.2. **Existence and series representation.** We start from an infinite triangular system of convolutional equations:

$$\begin{aligned} K_0 &= B, \\ K_1 &= F + (K_0 K_1) * \rho \\ K_n &= \left(\frac{1}{2}\sum_{i=0}^{n} K_i K_{n-i}\right) * \rho \text{ for } n > 2. \end{aligned} \tag{A.6}$$

in the unknown functions $\{K_n\}_{n\in\mathbb{N}_0}$, where $B, F \in \mathcal{S}^\infty[0,T]$.

In order to establish well-posedness of (A.6), we list in Lemma A.5 below a few well-known basic facts about linear convolution equations. We omit the standard argument based on Banach's fixed-point theorem.

Lemma A.5. *Suppose that $F, B \in \mathcal{S}^\infty[0,T]$ are such that $|B|_{\mathcal{S}^\infty[0,T]} < 1$. Then the equation*

$$K = F + (BK) * \rho \tag{A.7}$$

admits a unique solution K in $\mathcal{S}^\infty[0,T]$. Moreover K satisfies

$$|K|_{\mathcal{S}^\infty[0,T]} \leq \frac{|F|_{\mathcal{S}^\infty[0,T]}}{1 - |B|_{\mathcal{S}^\infty[0,T]}}. \tag{A.8}$$

Proposition A.6. *Suppose that $|B|_{\mathcal{S}^\infty[0,T]} < 1$. Then the system (A.6) has a unique solution in $(\mathcal{S}^\infty[0,T])^{\mathbb{N}_0}$, denoted by $\{K_n[B,F]\}_{n\in\mathbb{N}_0}$. Moreover,*

$$|K_n[B,F]|_{\mathcal{S}^\infty[0,T]} \leq C n^{-3/2} \left(\frac{2|F|_{\mathcal{S}^\infty[0,T]}}{(1 - |B|_{\mathcal{S}^\infty[0,T]})^2}\right)^n \text{ for all } n \in \mathbb{N}. \tag{A.9}$$

Proof. We observe that for $n \geq 1$, the n-th equation in the system (A.6) can be written in the form

$$K_n = F_n + (BK_n) * \rho, \text{ where } F_n = \begin{cases} F, & n = 1, \\ \frac{1}{2}\sum_{i=1}^{n-1} (K_i K_{n-i}) * \rho, & n \geq 2. \end{cases} \tag{A.10}$$

We also observe that F_n does not involve K_n or any K_m with $m > n$. This allows us to argue inductively, using Lemma A.5 in each step, that the system (A.6) has a unique solution $\{K_n[B,F]\}_{n\in\mathbb{N}_0}$ in $(\mathcal{S}^\infty[0,T])^{\mathbb{N}_0}$ and that

$$|K_n[B,F]|_{\mathcal{S}^\infty[0,T]} \leq M|F_n|_{\mathcal{S}^\infty[0,T]} \text{ where } M = (1 - |B|_{\mathcal{S}^\infty[0,T]})^{-1}.$$

This implies that $|K_1[B,F]|_{\mathcal{S}^\infty[0,T]} \leq M|F|_{\mathcal{S}^\infty[0,T]}$ and that

$$|K_n[B,F]|_{\mathcal{S}^\infty[0,T]} \leq \tfrac{1}{2} M \sum_{i=1}^{n-1} |K_i[B,F]|_{\mathcal{S}^\infty[0,T]} |K_{n-i}[B,F]|_{\mathcal{S}^\infty[0,T]} \text{ for } n \geq 2. \tag{A.11}$$

If $F = 0$ then $K_n[B, F] = 0$ for all $n \geq 1$. Otherwise, we set

$$c_n = \frac{2^{n-1} |K_n[B, F]|_{\mathcal{S}^\infty[0,T]}}{M^{2n-1} |F|^n_{\mathcal{S}^\infty[0,T]}}, \text{ for } n \geq 1,$$

so that, by (A.11),

$$c_1 \leq 1 \text{ and } c_n \leq \sum_{i=1}^{n-1} c_i c_{n-i} \text{ for } n \geq 2.$$

We recall that the sequence $\{C_n\}_{n \in \mathbb{N}_0}$ of Catalan numbers satisfies (see [Rom15, eq. (1.2), p. 3]) the recurrence relation

$$C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}, \; C_0 = 1,$$

Hence, by induction, $c_n \leq C_{n-1}$, for all $n \geq 1$ and so, the standard asymptotics (see e.g. [Rom15, Theorem 3.1, p. 15]) for Catalan numbers implies that

$$c_n \sim \frac{4^{n-1}}{n^{3/2}\sqrt{\pi}} \text{ as } n \to \infty,$$

which, in turn, implies (A.9). □

Lemma A.7. *Let $B, F \in \mathcal{S}^\infty[0, T]$ be such that $|B|_{\mathcal{S}^\infty[0,T]} < 1$ and $|F|_{\mathcal{S}^\infty[0,T]} \leq \frac{1}{2}(1 - |B|_{\mathcal{S}^\infty[0,T]})^2$, and let $\{K_n[B, F]\}_{n \in \mathbb{N}_0} \in (\mathcal{S}^\infty[0, T])^{\mathbb{N}_0}$ be the unique solution to the system (A.6). Then the series $\sum_{n \geq 0} K_n[B, F]$ converges absolutely in $\mathcal{S}^\infty[0, T]$ and its sum*

$$K[B, F] := \textstyle\sum_{n \geq 0} K_n[B, F]$$

satisfies the equation

$$K[B, F] = F + \tfrac{1}{2}(K[B, F])^2 * \rho + B - \tfrac{1}{2} B^2 * \rho \tag{A.12}$$

Proof. The assumption on the size of F implies, via Proposition A.6, that, using the shortcuts $K = K[B, G]$ and $K_n = K_n[B, F]$, we have

$$|K_n|_{\mathcal{S}^\infty[0,T]} \leq C n^{-3/2} \text{ for } n \in \mathbb{N}.$$

This, in turn, implies that the series $\sum_{n \geq 0} K_n$ converges absolutely in $\mathcal{S}^\infty[0, T]$. Moreover,

$$\begin{aligned} \tfrac{1}{2} K^2 * \rho &= \tfrac{1}{2} \Big(\textstyle\sum_{n \geq 0} K_n \Big)^2 * \rho = \tfrac{1}{2} K_0^2 * \rho + (K_0 K_1) * \rho + \textstyle\sum_{n \geq 2} \big(\tfrac{1}{2} \sum_{i=0}^n K_i K_{n-1} \big) * \rho \\ &= \tfrac{1}{2} B^2 * \rho + (K_1 - F) + \textstyle\sum_{n \geq 2} K_n = \tfrac{1}{2} B^2 * \rho - B + K - F. \end{aligned}$$

□

Proposition A.8. *If $|F|_{\mathcal{S}^\infty[0,T]} \leq 1/2$ the function*

$$K[F] = \textstyle\sum_{n \geq 1} K_n[0, F] \tag{A.13}$$

defines the unique solution of (A.2) in $\mathcal{S}^\infty[0, T]$. Moreover if $G \in \mathcal{S}^\infty[0, T]$ and $\varepsilon \in \mathbb{R}$ are such that $|F|_{\mathcal{S}^\infty[0,T]} + |\varepsilon| |G|_{\mathcal{S}^\infty[0,T]} \leq 1/2$ we have

$$K[F + \varepsilon G] = K[F] + \textstyle\sum_{n \geq 1} \varepsilon^n K_n[K[F], G].$$

with absolute convergence in $\mathcal{S}^\infty[0, T]$.

Proof. In the special case $B = 0$, the conditions of Lemma A.7 are satisfied as soon as $|F|_{\mathcal{S}^\infty[0,T]} \leq 1/2$. Therefore (A.13) defines a solution to (A.1). Uniqueness is the content of Proposition A.4 above.

Let F, G and ε be as in the second part of the statement. Thanks to Lemma A.3 above, we have

$$|K[F]|_{\mathcal{S}^\infty[0,T]} \leq \max\Big(|F|_{\mathcal{S}^\infty[0,T]}, 1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,T]}}\Big) = 1 - \sqrt{1 - 2|F|_{\mathcal{S}^\infty[0,T]}}.$$

so that

$$\tfrac{1}{2}(1 - |K[F]|_{\mathcal{S}^\infty[0,T]})^2 \geq 1/2 - |F|_{\mathcal{S}^\infty[0,T]} \geq |\varepsilon G|_{\mathcal{S}^\infty[0,T]},$$

which is exactly what is needed for Lemma A.7 to apply. Therefore,

$$h[F + \varepsilon G] = K[F] + \textstyle\sum_{n\geq 1} K_n[K[F], \varepsilon G].$$

It remains to observe that functions $\varepsilon^{-n} K_n[K[F], \varepsilon G]$ solve the system (A.6) for any $\varepsilon \neq 0$, so that, by uniqueness, we have

$$K_n[K[F], \varepsilon G] = \varepsilon^n K_n[K[F], G] \text{ for } \varepsilon \in \mathbb{R}. \qquad \square$$

References

[BDHM13] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy, *Some limit theorems for Hawkes processes and application to financial statistics*, Stoch. Process. Appl. **123** (2013), no. 7, 2475–2499.

[BDM01] E. Bacry, J. Delour, and J. F. Muzy, *Multifractal random walk*, Phys. Rev. E **64** (2001), 026103.

[BHP21] Christian Bayer, Fabian A. Harang, and Paolo Pigato, *Log-modulated rough stochastic volatility models*, SIAM Journal on Financial Mathematics **12** (2021), no. 3, 1257–1284.

[BLP21] Mikkel Bennedsen, Asger Lunde, and Mikko S Pakkanen, *Decoupling the Short- and Long-Term Behavior of Stochastic Volatility*, Journal of Financial Econometrics (2021).

[BMM15] E. Bacry, I. Mastromatteo, and J. F. Muzy, *Hawkes processes in finance*, Market Microstructure and Liquidity **1** (2015), no. 1, 1550005.

[CR98] Fabienne Comte and Eric Renault, *Long memory in continuous-time stochastic volatility models*, Math. Finance **8** (1998), no. 4, 291–323.

[DLM] *Nist digital library of mathematical functions*, https://dlmf.nist.gov/, Release 1.2.1 of 2024-06-15, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[DRSV17] Bertrand Duplantier, Rémi Rhodes, Scott Sheffield, and Vincent Vargas, *Log-correlated gaussian fields: An overview*, pp. 191–216, Springer International Publishing, 2017.

[DVJ03] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. I*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2003, Elementary theory and methods.

[DVJ08] ———, *An introduction to the theory of point processes. Vol. II*, second ed., Probability and its Applications (New York), Springer, New York, 2008, General theory and structure.

[EER19] Omar El Euch and Mathieu Rosenbaum, *The characteristic function of rough heston models*, Mathematical Finance **29** (2019), no. 1, 3–38.

[FFGS22] Martin Forde, Masaaki Fukasawa, Stefan Gerhold, and Benjamin Smith, *The riemann–liouville field and its gmc as h→0, and skew flattening for the rough bergomi model*, Statistics & Probability Letters **181** (2022), 109265.

[FKS16] Y. V. Fyodorov, B. A. Khoruzhenko, and N. J. Simm, *Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian Unitary Ensemble*, The Annals of Probability **44** (2016), no. 4, 2980 – 3031.

[FLM15] Raúl Fierro, Víctor Leiva, and Jesper Moller, *The Hawkes process with different exciting functions and its asymptotic behavior*, Journal of Applied Probability **52** (2015), no. 1, 37 – 54.

[FTW22] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal, *Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics*, Mathematical Finance **32** (2022), no. 4, 1086–1132.

[GJR18] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum, *Volatility is rough*, Quant. Finance **18** (2018), no. 6, 933–949.

[GK54] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley Publishing Co., Inc., Cambridge, Mass., 1954, Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob.

[GK96] Boris V. Gnedenko and Victor Yu. Korolev, *Random summation*, CRC Press, Boca Raton, FL, 1996, Limit theorems and applications.

[GZ18] X. Gao and L. Zhu, *Limit theorems for Markovian Hawkes processes with a large initial intensity*, Stoch. Process. Appl. **128** (2018), no. 11, 3807–3839.

[Haw71a] Alan G. Hawkes, *Point Spectra of Some Mutually Exciting Point Processes*, Journal of the Royal Statistical Society. Series B (Methodological) **33** (1971), no. 3, 438–443.

[Haw71b] ———, *Spectra of Some Self-Exciting and Mutually Exciting Point Processes*, Biometrika **58** (1971), no. 1, 83–90.

[HN22] Paul Hager and Eyal Neuman, *The multiplicative chaos of $H = 0$ fractional Brownian fields*, The Annals of Applied Probability **32** (2022), no. 3, 2139 – 2179.

[HO74] A. G. Hawkes and D. Oakes, *A cluster process representation of a self-exciting process*, J. Appl. Probab. **11** (1974), no. 3, 493–503.

[HX21] U. Horst and W. Xu, *Functional limit theorems for marked hawkes point measures*, Stoch. Process. Appl. **134** (2021), 94–131.

[HX24] Ulrich Horst and Wei Xu, *Functional limit theorems for hawkes processes*, 2024.

[HXZ23] U. Horst, W. Xu, and R. Zhang, *Convergence of heavy-yailed Hawkes processes and the microstructure of rough volatility*, arXiv preprint arXiv:2312.08784 (2023).

[JLP19] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido, *Affine volterra processes*, The Annals of Applied Probability **29** (2019), no. 5, 3155 – 3200.

[JR15] Thibault Jaisson and Mathieu Rosenbaum, *Limit theorems for nearly unstable Hawkes processes*, Ann. Appl. Probab. **25** (2015), no. 2, 600–631.

[JR16a] ———, *The different asymptotic regimes of nearly unstable autoregressive processes*, The fascination of probability, statistics and their applications, Springer, Cham, 2016, pp. 283–301.

[JR16b] ———, *Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes*, Ann. Appl. Probab. **26** (2016), no. 5, 2860–2882.

[Kal17] Olav Kallenberg, *Random measures*, Akademie-Verlag, Berlin; Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 2017.

[Kal21] ———, *Foundations of modern probability*, Springer-Verlag, 2021.

[KK93] V.Yu. Korolev and V. M. Kruglov, *Limit theorems for random sums of independent random variables*, Stability Problems for Stochastic Models (Berlin, Heidelberg) (Vladimir V. Kalashnikov and Vladimir M. Zolatarev, eds.), Springer Berlin Heidelberg, 1993, pp. 100–120.

[KMM84] L. B. Klebanov, G. M. Maniya, and I. A. Melamed, *A problem of V. M. Zolotarev and analogues of infinitely divisible and stable distributions in a scheme for summation of a random number of random variables*, Teor. Veroyatnost. i Primenen. **29** (1984), no. 4, 757–760.

[Leo17] Giovanni Leoni, *A first course in sobolev spaces*, second ed., American Mathematical Soc., 2017.

[LLPT24] Patrick J. Laub, Young Lee, Philip K. Pollett, and Thomas Taimre, *Hawkes models and their applications*, Annual Review of Statistics and Its Application (2024).

[NR18] Eyal Neuman and Mathieu Rosenbaum, *Fractional Brownian motion with zero Hurst parameter: a rough volatility viewpoint*, Electronic Communications in Probability **23** (2018), no. none, 1 – 12.

[Rom15] Steven Roman, *An introduction to catalan numbers*, Birkhäuser, 2015.

[San15] Filippo Santambrogio, *Optimal transport for applied mathematicians - calculus of variations, pdes, and modeling*, Birkhäuser, 2015.

[Smi95] Peter J. Smith, *A recursive formulation of the old problem of obtaining moments from cumulants and vice versa*, The American Statistician **49** (1995), no. 2, 217–218.

Tristan Pace, Department of Mathematics, The University of Texas at Austin
Email address: tpace4288@utexas.edu

Gordan Žitković, Department of Mathematics, The University of Texas at Austin
Email address: gordanz@math.utexas.edu