Self-exciting jump processes and their asymptotic behaviour

Kristina Rognlien Dahl \(^a\) and Heidar Eyjolfsson \(^b\)

\(^a\)Department of Mathematics, University of Oslo, Oslo, Norway; \(^b\)Department of Engineering, Reykjavik University, Reykjavik, Iceland

**ABSTRACT**

The purpose of this paper is to investigate properties of self-exciting jump processes where the intensity is given by an SDE, which is driven by a finite variation stochastic jump process. The value of the intensity process immediately before a jump may influence the jump size distribution. We focus on properties of this intensity function, and show that for each fixed point in time, \(t \geq 0\), a scaling limit of the intensity process converges in distribution, and the limit equals the strong solution of the square-root diffusion process (Cox–Ingersoll–Ross process) at \(t\). As a particular example, we study the case of a linear intensity process and derive explicit expressions for the expectation and variance in this case.

**ARTICLE HISTORY**

Received 26 May 2021
Accepted 10 January 2022

**KEYWORDS**

Self-exciting stochastic processes; jump processes; stochastic differential equations; asymptotic behaviour; intensity process

**1. Introduction**

Self-exciting processes were first studied by Hawkes [10]. Initially, the main application was in seismology; the modelling of earthquakes and their aftershocks. Over the last decade various versions of self-exciting processes have been used in financial applications (see, e.g. [3,7]), to model group behaviour in social media (see [18]), and for predicting crime and terrorist acts (see [13,15]). The class of self-exciting stochastic process models has the advantage of being very versatile in applications.

In this paper, we investigate the properties of self-exciting jump processes. Our definition of self-exciting processes follows Eyjolfsson and Tjøstheim [9]. The self-exciting process is essentially a counting process, which counts the number of shocks that have occurred at any given time. The intensity process of the self-exciting process, denoted by \(\lambda(t)\), determines the probability of shocks occurring in the infinitesimal interval \((t, t + dt)\) conditioned on the information at time \(t\). We assume that this intensity process has Markovian stochastic differential equation (SDE) dynamics. The self-exciting processes considered in this paper differ from Hawkes processes, see Hawkes [10] and Hawkes and Oakes [11], because the stochastic jump size of the self-exciting process may depend on the current value of the intensity process. As noted in Eyjolfsson and Tjøstheim [9], this kind of self-exciting process is a generalization of the exponential Hawkes model.

The self-exciting model we present in this paper is a finite variation jump process (like, e.g. the compound Poisson process) in the sense that in bounded time intervals it only
produces a finite amount of jumps. However, as we will discuss in Section 3, when model parameters are chosen in a suitable way, and then passed to a limit one obtains a process in the limit which can be thought of as a scaling limit analogue of the self-exciting process. This is similar to how the gamma and inverse Gaussian processes can be obtained as limits of compound Poisson processes. However, we emphasize that the class of self-exciting processes has the ability to produce periods (clusters) of high activity between periods of low activity, which is something that Lévy processes cannot reproduce. Our main result is that for each fixed time-point, the distribution of the aforementioned limit equals the distribution of the strong solution of the square-root diffusion process (also known as the Cox–Ingersoll–Ross (CIR) process [5]) as the fixed time-point.

Jaisson and Rosenbaum [12] derive limit theorems for Hawkes processes which are nearly unstable. These processes are such that the $L^1$-norm of their kernel is close to unity. Jaisson and Rosenbaum [12] show that after a rescaling, the nearly unstable Hawkes counting processes asymptotically behave like integrated CIR models, and the nearly unstable intensity processes asymptotically behave like CIR models. More recently, Erny et al. [8] study so-called mean-field limits for interacting Hawkes processes in a diffusive regime, where they obtain a CIR process limit in distribution, and Abi Jaber et al. [1] prove a convergence result of Hawkes processes to a Volterra process with a CIR-type dynamic.

In the same spirit, we prove limit theorems for our class of self-exciting processes. In Theorem 3.3, we show that for a fixed time, $t \geq 0$, the scaling limit of the intensity process of the self-exciting process at time $t$ behaves like the strong solution of the CIR square-root process at time point $t$ in distribution. We, moreover, prove a similar result for the integrated scaled intensity process, which converges to the integrated CIR square-root process in distribution. We reiterate that our self-exciting processes differ from the Hawkes processes in that our processes are specified by a Markovian SDE with stochastic jumps, whereas the Hawkes intensity process has constant jumps and specified by a kernel function with $L^1$-norm less than one.

The paper is structured as follows: In Section 2, we introduce the framework for self-exciting stochastic processes and illustrate such processes via a numerical example. In Section 3, we show that for each fixed time point as an appropriate scaling limit of the finite variation self-exciting process is taken, the distribution of the limit equals the distribution of the strong solution of the square-root diffusion process at the same time point. In Section 4, we study a particular case where the intensity process of the self-exciting process is assumed to be linear. We derive the expected value and variance of this linear intensity process, as well as the moments of the integrated intensity. Finally, in Section 5, we conclude and sketch ideas on further research.

2. Self-exciting stochastic jump processes

Our definition of self-exciting processes follows Eyjolffson and Tjøstheim [9]. Essentially, the self-exciting process is a counting process, which counts the number of shocks which have occurred at any given time. Let $(\Omega, \mathcal{F})$ denote a measurable space, and let $(T_n)_{n \geq 1}$ be a point process taking values in $\mathbb{R}_+$. The sequence $(T_n)_{n \geq 1}$ represents times of successive events and is assumed non-negative and non-decreasing, i.e. $0 \leq T_1 \leq T_2 \leq \cdots$ holds.
The counting process, \( N(t) \), associated to the point process,

\[
N(t) := \sum_{n \geq 1} 1_{\{T_n \leq t\}},
\]

where \( t \geq 0 \), is the counting process which records all the jumps of the point process. The rate at which the events occur is furthermore dictated by the intensity process, which we define in what follows. We identify a point process with its counting process (1) and let \( \mathcal{F}_t^N := \sigma\{N(s) : 0 \leq s \leq t\} \), where \( t \geq 0 \). That is, \( \{\mathcal{F}_t^N\}_{t \geq 0} \) is the filtration generated by the counting process. Assume that we are given a point process adapted to some filtration \( \{\mathcal{F}_t\}_{t \geq 0} \), with \( \mathcal{F}_t^N \subset \mathcal{F}_t \) for all \( t \geq 0 \). Suppose that \( N(t) \) admits a càdlàg \( \{\mathcal{F}_t\}_{t \geq 0} \)-adapted, and thus predictable, intensity \( \lambda(t) \), such that

\[
E \left[ \int_0^\infty f(s) \, dN(s) \right] = E \left[ \int_0^\infty f(s) \lambda(s) \, ds \right],
\]

holds for all predictable \( f : \Omega \times \mathbb{R}_+ \rightarrow [-\infty, \infty] \). Note that this means the process \( t \mapsto N(t) - \lambda(t) \) is a martingale, and that the intensity process \( \lambda(t) \) determines the probability of shocks occurring in the infinitesimal interval \((t, t + dt)\) conditioned on \( \mathcal{F}_t \). Note in particular that if the intensity is constant, \( \lambda(t) = \lambda_0 > 0 \), holds for all \( t \geq 0 \), then \( N(t) \) is a standard homogeneous Poisson process with intensity \( \lambda_0 \).

We assume that the intensity process admits Markovian SDE dynamics. Each jump has a particular size, which feeds into (i.e. excites) the intensity, and typically raises the intensity immediately after the shock has occurred, although the intensity will then revert back to some mean level in the absence of further shocks. The size of the shock can furthermore influence how much the likelihood of further shocks is increased (i.e. the level of excitement). Hence, a particularly large shock may, for example, lead to a high likelihood of aftershocks, whereas a small shock may be less likely to excite the intensity and thus cause further aftershocks. Thus, the model class allows the shocks to vary in size, and the size of each shock determines the level of the corresponding intensity process excitation. If the intensity becomes high enough, a cluster of shocks might appear.

Consider the stochastic jump process, \( U(t) \), given by

\[
U(t) = \sum_{k=1}^{N(t)} X_k,
\]

where \( \{N(t)\}_{t \geq 0} \) is the counting process (1), and \( \{X_k\}_{k \in \mathbb{N}} \) is a family of random variables, where \( X_k \) has the probability distribution \( v(\lambda(T_k-), \cdot) \), for a given family \( \{v(\lambda, \cdot)\}_{\lambda > 0} \) of probability distributions, and \( \lambda(t-) := \lim_{s \uparrow t} \lambda(s) \). Thus, we allow the value of the intensity process immediately before the jump to influence the jump size distribution. We introduce the stochastic differential equation (SDE)

\[
d\lambda(t) = \mu(\lambda(t)) \, dt + \beta \, dU(t),
\]

\( \lambda(0) = \lambda_0 \), where \( \beta \in \mathbb{R} \) is a constant and we assume that \( \mu : \mathbb{R}_+ \rightarrow \mathbb{R} \), is Lipschitz continuous.
**Definition 2.1:** An SDE-driven self-exciting jump process is a stochastic jump process (2) with the intensity \( \lambda(t) \), given by the SDE (3), with jump-sizes, \( X_k \), which follow the probability distribution \( \nu(\lambda(T_k-), \cdot) \). Here, \( \{\nu(\lambda, \cdot)\}_{\lambda > 0} \) is a family of probability distributions, and \( \nu(\lambda, \cdot) \) is supported on \([\lambda_0 - \lambda, \infty)\).

From the above Definition 2.1, we see that the jumps feed into the intensity via the jump process \( U(t) \). Furthermore, the value of the intensity process immediately before the jump is a parameter in the jump-size probability distribution. This means that the intensity level prior to a jump can determine the size of the next jump. Note moreover that the self-exciting processes in Definition 2.1 differ from Hawkes processes, see Hawkes [10] and Hawkes and Oakes [11], because of the stochastic jump size modelled via the family \( \{X_k\}_{k \in \mathbb{N}} \) of random variables which may depend on the current value of the intensity process. Actually, this kind of self-exciting process generalizes the exponential Hawkes model (i.e. the exponential Hawkes process is a special case of Definition 2.1), see Eyjolfsson and Tjøstheim [9] for more details.

Given the above definition, we may think about the compound jump process \( U(t) \) along the following lines. Clearly, \( U(0) = 0 \). Since by our definition \( \lambda(t) = \lambda_0 + \int_0^t \mu(\lambda(s)) \, ds \), for all \( t \geq 0 \) such that \( N(t) = 0 \), it follows that that the first jump time, \( T_1 \), is determined by a non-homogeneous Poisson process with the deterministic intensity \( t \mapsto \lambda(t) \), and that the random jump size \( X_1 \) is drawn from the distribution \( \nu(\lambda(T_1-), \cdot) \). Similarly, after \( k \geq 1 \) jumps the \( (k+1) \)st jump time, \( T_{k+1} \), is determined by a non-homogeneous Poisson process with the intensity \( t \mapsto \lambda(T_k + t) \) (which is deterministic between jumps), and the random jump size \( X_{k+1} \) is drawn from the distribution \( \nu(\lambda(T_{k+1}-), \cdot) \). Clearly, one can use this characterization together with for example a thinning algorithm for non-homogeneous Poisson processes to simulate a self-exciting process.

**Example 2.2 (Simulation of a self-exciting process \( U(t) \)):** To illustrate, we simulate two paths of the same self-exciting process and plot the intensity \( \lambda(t) \) as well as the corresponding self-exciting process \( U(t) \). Following Eyjolfsson and Tjøstheim [9], we consider a non-linear intensity process \( \lambda(t) \) given as the solution to the following SDE:

\[
d\lambda(t) = \left( \alpha + \delta \exp(-\gamma \lambda(t)^2) \right) (\lambda_0 - \lambda(t)) \, dt + \beta \, dU(t),
\]

\[\lambda(0) = \lambda_0. \tag{4}\]

As mentioned in Eyjolfsson and Tjøstheim [9], the speed of mean reversion in the SDE (4) varies between \( \alpha + \delta \exp(-\gamma \lambda_0^2) \) for \( \lambda(t) = \lambda_0 \) and decreases towards \( \alpha \) when \( \lambda(t) \) increases. The interpretation of this is that in low activity periods, the effect of a jump fades out faster than in high activity periods.

For the simulation, we choose \( \lambda_0 = 0.05, \alpha = 0.1233, \beta = 0.0399 \) and the jumps are simulated from an inverse Gaussian distribution with parameters 1.9389 (mean) and 5.4943 (shape). The parameter values were chosen based on Eyjolfsson and Tjøstheim [9], but the choice of \( \lambda_0 \) was modified slightly to better display the particular structure of the self-exciting process \( U(t) \). The simulation was performed using a thinning algorithm from Ogata [16].

In Figures 1 and 2, we have plotted two different paths of the self-exciting process \( U(t) \) with the corresponding intensity process \( \lambda(t) \). Periods with a lot of jump activity in the
intensity process correspond to a large increases in the self-exciting process. Periods, where there are no jumps in the intensity process, correspond to plateaus (i.e. no change) in the corresponding self-exciting process.

In Eyjolfsson and Tjøstheim [9], conditions are provided which ensure that the counting process $N(t)$ associated with the intensity SDE (3) does not explode in finite time. This means that the self-exciting process has finitely many jumps on any compact interval, and is thus of finite variation. Let

$$\Lambda(t) := \int_0^t \lambda(s) \, ds.$$ 

Then, a condition for the self-exciting process to be non-explosive is:

**Lemma 2.3:** Assume that $\Lambda(t) < \infty$ holds almost surely for any $t > 0$. Then, the SDE-driven self-exciting process does not explode in finite time.

**Proof:** The Lemma follows from Assumption 3.1, and the comments thereafter, in Eyjolfsson and Tjøstheim [9].
If the integrated intensity process $\Lambda(t)$ in Equation (3) does not satisfy the finiteness assumption of Lemma 2.3, then the counting process $N(t)$ may explode in finite time. Indeed, since the expected number jumps are determined by the value of the (non-negative) intensity process, that is for $0 < s < t$ it holds that $E[N(t) - N(s)] = E[\int_s^t \lambda(s) \, ds]$, it follows that in order for the process to produce an infinite number of jumps in a compact interval, the integrated intensity must tend to infinity.

In what follows, we are going to consider a sequence of intensity processes, $\{\lambda_k(t)\}$, which tend to infinity as $k \to \infty$, thus producing more jumps, but simultaneously the average jump-size becomes smaller and smaller. Now, in order for this sequence to make sense as $k \to \infty$, we scale it with a decreasing sequence $\{a_k\}$ of positive real numbers, and study the probability distribution of $\lim_{k} a_k \lambda_k(t)$ for each fixed $t \geq 0$.

3. Scaling limits of self-exciting processes

As the parameters which govern the dynamics of the SDE intensity are changed in a way such that $\Lambda(t)$ increases, while the jump-sizes, $X_1, X_2, \ldots$, become smaller and smaller simultaneously, then, as we pass to the limit, we can ensure that in the limit the self-exciting jump process, $U(t)$, has infinite activity, meaning that infinitely many jumps occur on each compact interval.

We study the asymptotic properties of the intensity process as we increase the jump intensity and decrease the jump size distribution simultaneously. To that end, consider a self-exciting process which depends on a parameter, $k \geq 1$, i.e. a process with an intensity,

$$d\lambda_k(t) = \mu_k(\lambda_k(t)) \, dt + \beta_k \, dU_k(t),$$

where $\lambda_k(0) = \lambda_{0k}$, $\mu_k$ is Lipschitz continuous, and the jumps are non-negative and determined by the $\lambda$ dependent probability measure $\nu_k(\lambda, \cdot)$. Suppose furthermore that $\{a_k\}$, $k \geq 1$ is a sequence such that $a_k > 0$ for all $k \geq 1$ and $\lim_{k \to \infty} a_k = 0$. We study the behaviour of the scaled intensity process

$$\hat{\lambda}_k(t) := a_k \lambda_k(t).$$

Now, let $m_{k,n}(t) := E[\lambda_{k}^n(t)]$ and $\hat{m}_{k,n}(t) := E[\hat{\lambda}_k^n(t)]$. According to Eyjolfsson and Tjøstheim [9], the generator of the intensity process $\lambda_k(t)$ is given by

$$\mathcal{A}_k f(\lambda) = \mu_k(\lambda)f'(\lambda) + \int (f(\lambda + \beta_k x) - f(\lambda)) \nu_k(\lambda, dx),$$

whenever $f$ is in the domain of the generator. Furthermore, if $f$ is in the domain of the generator, then the Dynkin formula is verified, i.e.

$$E[f(\lambda_k(t))] = f(\lambda_{0k}) + E\left[\int_0^t \mathcal{A}_k f(\lambda_k(r)) \, dr\right],$$

when it assumed that the initial value is $\lambda_k(0) = \lambda_{0k}$. So if $f(\lambda) = \lambda^n$, where $n \geq 1$,

$$J_{k,n}(\lambda) := \int x^n \nu_k(\lambda, dx).$$

for all \( n, k \geq 1 \) denote the \( n \)th moment of \( \nu_k(\lambda, \cdot) \), then by the employing the binomial theorem, we obtain that

\[
(A_k f)(\lambda) = n(\mu_k(\lambda) + \beta_k J_{k,1}(\lambda)\lambda)^{n-1} + \sum_{j=0}^{n-2} \binom{n}{j} \beta_k^{n-j} J_{k,n-j}(\lambda)\lambda^{j+1}.
\]

So, an application of Dynkin’s formula and Fubini yields

\[
m_{k,n}(t) = \lambda_0 n + \int_0^t \left[ nE[(\mu_k(\lambda(s)) + \beta_k J_{k,1}(\lambda_k(s))\lambda_k(s))\lambda_k^{n-1}(s)] \right] ds
\]

\[
+ \sum_{j=0}^{n-2} \binom{n}{j} \int_0^t E[\beta_k^{n-j} J_{k,n-j}(\lambda_k(s))\lambda_k^{j+1}(s)] ds,
\]

where the sum is dropped when \( n = 1 \).

**Assumption 3.1:** Given the intensity processes (5) and the corresponding scaled processes (6), suppose that the functions \( \lambda \mapsto J_{k,n}(\lambda) \), as defined by Equation (7), are bounded, and that the jump-sizes are non-negative. Moreover, assume that \( \lim_{k \to \infty} a_k \lambda_0 = 0 \), there is a sequence of bounded non-negative functions \( \{g_k\} \) such that \( \lim_{k \to \infty} a_k g_k(\lambda) = c_0 \in \mathbb{R} \), where the convergence is uniform in \( \lambda \), and it holds that

\[
\mu_k(\lambda) + \beta_k J_{k,1}(\lambda)\lambda = g_k(\lambda) - c_1 \lambda
\]

where \( c_1 > 0 \). Suppose furthermore that

\[
\lim_{k \to \infty} \beta_k^{j-1} J_{k,j}(\lambda) = \begin{cases} c_2 & \text{if } j = 2, \\ 0 & \text{if } j > 2, \end{cases}
\]

where the convergence is uniform in \( \lambda \) and \( c_2 > 0 \).

Note, in particular, that according to the above assumption, it follows that if \( f(\lambda) = \lambda \), then

\[
A_k f(\lambda) = \mu_k(\lambda) + \lambda J_{k,1}(\lambda) \leq \|g_k\|_{\infty},
\]

where \( \|f\|_{\infty} = \sup_x |f(x)| \), and thus the generator fulfills the (CD0) non-explosion condition of Meyn and Tweedie [14], which means that the intensity processes are non-explosive.

See Example 3.5, where we demonstrate the preceding Assumption in practice. In the following lemma, we show that the \( n \)th moment of the scaled intensity process, \( \hat{m}_{k,n}(t) = E[\hat{\lambda}_k^n(t)] \), is bounded by an \( n \)th degree polynomial.

**Lemma 3.1:** For each \( n \geq 0 \), the moment \( \hat{m}_{k,n}(t) \) is bounded by an \( n \)th degree polynomial which is independent of \( k \geq 1 \).

**Proof:** By (8) and Assumption 3.1, it holds that

\[
\hat{m}_{k,n}(t) = a_k^n \lambda_0 n + \int_0^t nE[(a_k g_k(\lambda_k(s)) - c_1 \hat{\lambda}_k(s))\hat{\lambda}_k^{n-1}(s)] ds
\]
and induction that \\
Hence, since convergent sequences are bounded, it follows according to Assumption 3.1 \\
where \\
So, we may conclude that \\
Weshallmoreovermakeuseofthefollowingresultintheproofsofourmainresults. \\
Lemma 3.2: Suppose that \(y : \mathbb{R}_+ \rightarrow \mathbb{R}\) verifies the equation \\
y(t) = y_0 - c \int_0^t y(s) \, ds + \int_0^t \phi(s) \, ds, \\
where \(y_0, c \in \mathbb{R}\), and \(\phi \in L^1([0, T])\) for any \(T > 0\). Then it holds that \\
y(t) = e^{-ct} \left(y_0 + \int_0^t e^{cs} \phi(s) \, ds\right). \\
\textbf{Proof:} Note that if \(t, s \geq 0\), then \(|1_{[0,t]} \phi - 1_{[0,s]} \phi| \leq |1_{[0,t]} \phi| + |1_{[0,s]} \phi| \leq 2|\phi|\). Now let \(\{s_k\} \subset \mathbb{R}_+\) be a sequence such that \(\lim_k s_k = t\), then \(\{s_k\}\) is bounded in \(\mathbb{R}_+\). Let \(\tilde{s} := \sup_k s_k\) and apply the dominated convergence theorem to conclude that \\
\[
\lim_{k \to \infty} \left| \int_0^t \phi(r) \, dr - \int_0^{s_k} \phi(r) \, dr \right| \leq \lim_{k \to \infty} \int_0^{\tilde{s}} |(1_{[0,t]}(r) - 1_{[0,s_k]}(r))\phi(r)| \, dr = 0.
\]
So, we may conclude that \(\Phi(t) := \int_0^t \phi(s) \, ds\) is a continuous function. Equation (11) is a Volterra integral equation of the second kind, and since \(\Phi\) is continuous it follows (see e.g. Theorem 5.2.3 in [2]) that there exists a unique solution to the Equation (11). Note that according to Fubini’s theorem \\
c \int_0^t \left( e^{-cs} \left(y_0 + \int_0^s e^{cr} \phi(r) \, dr \right) \right) \, ds = y_0(1 - e^{-ct}) + c \int_0^t \left( \int_r^t e^{-cs} \, ds \right) e^{cr} \phi(r) \, dr
According to Equation (8) and Assumption 3.1, it holds that

\[ y_0 - e^{-ct} \left( y_0 + \int_0^t e^{cr} \phi(r) \, dr \right) + \int_0^t \phi(r) \, dr, \]

and by rearranging the terms it follows that (12) verifies the Volterra Equation (11).

We now show that, given our assumption, for each \( t \geq 0 \) the scaling limit of the intensity process (6) as \( k \to \infty \) in distribution is the square-root process which is given by the strong solution of the SDE

\[ dY(t) = (c_0 - c_1 Y(t)) \, dt + \sqrt{c_2 Y(t)} \, dB(t), \tag{13} \]

where \( Y(0) = 0 \) and \( B(t) \) denotes Brownian motion.

**Theorem 3.3:** For any \( t \geq 0 \), it holds that \( \hat{\lambda}_k(t) \to Y(t) \), as \( k \to \infty \) in distribution, where \( Y(t) \) is given by (13).

**Proof:** According to Equation (8) and Assumption 3.1, it holds that

\[
\hat{m}_{k,n}(t) = a_k^n \hat{\lambda}_k^n \cdot 0_k + \int_0^t nE[(a_k g_k(\lambda_k(s)) - c_1 \hat{\lambda}_k(s)) \lambda_k^{n-1}(s)] \, ds \\
+ \sum_{j=0}^{n-2} \binom{n}{j} \int_0^t E[a_k^{n-j-1} \beta^n_{j} \int_{k,n-j}(\lambda_k(s)) \lambda_k^{j+1}(s)] \, ds.
\]

Note that, we may rewrite the above, and get a Volterra integral equation of the second kind:

\[
\hat{m}_{k,n}(t) = a_k^n \hat{\lambda}_k^n \cdot 0_k - c_1 n \int_0^t \hat{m}_{k,n}(s) \, ds + \int_0^t \phi_{k,n}(s) \, ds, \tag{14}
\]

where

\[
\phi_{k,n}(s) = nE[a_k g_k(\lambda_k(s)) \hat{\lambda}_k^{n-1}(s)] + \sum_{j=0}^{n-2} \binom{n}{j} E[a_k^{n-j-1} \beta^n_{j} \int_{k,n-j}(\lambda_k(s)) \lambda_k^{j+1}(s)].
\]

Observe that according to Assumption 3.1, there exists a constant \( C > 0 \) such that

\[
|\phi_{k,n}(s)| \leq C \left( \hat{m}_{k,n-1}(s) + \sum_{j=0}^{n-2} \hat{m}_{k,j+1}(s) \right),
\]

holds for all \( k \), which together with Lemma 3.1 implies that the \( \phi_{k,n} \) is bounded by an integrable function. Hence, we may apply Lemma 3.2 to conclude that

\[
\hat{m}_{k,n}(t) = e^{-nc_1 t} \left( a_k^n \hat{\lambda}_k^n \cdot 0_k + \int_0^t e^{nc_1 s} \phi_{k,n}(s) \, ds \right)
\]

solves Equation (14). Given this representation, we prove by induction over \( n \geq 1 \) that the limits \( \lim_{k \to \infty} \hat{m}_{k,n}(t) \) exist. If \( n = 1 \), then \( \phi_{k,1}(s) = E[a_k g_k(\lambda_k(s))] \), and by Assumption 3.1, \( \phi_{k,1}(s) \to c_0 \), so by the dominated convergence theorem \( \lim_{k \to \infty} \hat{m}_{k,1}(t) \) exists.
Now assume that \( \lim_{k \to \infty} \hat{m}_{k,n-1}(t) \) exists for some \( n > 1 \). Then, according to Assumption 3.1, the limit \( \lim_{k \to \infty} \phi_{k,n}(s) \) exists and is given by

\[
\lim_{k \to \infty} \phi_{k,n}(s) = n c_0 \lim_{k \to \infty} \hat{m}_{k,n-1}(s) + \frac{n(n-1)}{2} c_2 \lim_{k \to \infty} \hat{m}_{k,n-1}(s).
\]

Therefore, we may employ the dominated convergence theorem to conclude that \( \lim_{k \to \infty} \hat{m}_{k,n}(t) \) also exists. So, by induction the limit \( \lim_{k \to \infty} \hat{m}_{k,n}(t) \) exists for all \( n \geq 1 \). If we set \( y_n(t) := \lim_{k \to \infty} \hat{m}_{k,n}(t) \) for all \( n \geq 1 \) and \( y_0 = 1 \), it follows that

\[
y_n(t) = e^{-nc_1 t} \int_0^t e^{nc_1 s} \left( nc_0 + \frac{n(n-1)}{2} c_2 \right) y_{n-1}(s) \, ds
\]

holds for any \( n \geq 1 \). According to this equation \( y_n \) is continuous for any \( n \geq 1 \), so we may apply the fundamental theorem of the calculus to conclude that the \( y_n \) functions verify the following system of ODE’s:

\[
y'_n = -nc_1 y_n + \left( nc_0 + \frac{n(n-1)}{2} c_2 \right) y_{n-1},
\]

where \( y_n(0) = 0 \).

Now, by applying Itô’s lemma with \( f(x) = x^n \) to the stochastic process \( Y(t) \) it follows that the \( n \)th moment of \( Y(t) \) also verifies the above ODE, for every \( n \geq 1 \). Moreover, Dufresne [6] shows that the series

\[
\sum_{n=0}^{\infty} \frac{s^n}{n!} E[Y^n(t)]
\]

converges when \( s \) is small enough. Thus, according to Theorems 30.1 and 30.2 in Billingsley [4], we conclude that for each fixed \( t \geq 0 \), \( \lim_{k \to \infty} \hat{\lambda}_k(t) \) equals \( Y(t) \) in distribution. □

A similar result holds for the integrated intensity process.

**Theorem 3.4:** For any \( t \geq 0 \), it holds that \( \int_0^t \hat{\lambda}_k(s) \, ds \to \int_0^t Y(s) \, ds \), as \( k \to \infty \) in distribution, where \( Y(t) \) is given by (13).

**Proof:** First, we show that the moments of the integrated intensity, \( \hat{\lambda}_k(t) = \int_0^t \hat{\lambda}_k(s) \, ds \), are uniformly bounded with respect to the parameter \( k \geq 1 \). Note that

\[
\hat{\lambda}_k^n(t) = \left( \int_0^t \hat{\lambda}_k(s) \, ds \right)^n \leq t^n \left( \sup_{s \in [0,t]} \hat{\lambda}_k(s) \right)^n.
\]

Since the jumps are non-negative, it follows that the intensity, \( s \mapsto \hat{\lambda}_k(s) \) (and thus \( s \mapsto \hat{\lambda}_k^n(s) \)) is a.s. upper semi-continuous. A property of upper semi-continuous functions is that it attains its supremum on a compact set. A consequence of this is that for any fixed \( \omega \in \Omega \) and \( \epsilon > 0 \), the set

\[
A_\epsilon = \left\{ s \in [0,t] : \hat{\lambda}_k^n(s) \geq \sup_{r \in [0,t]} \hat{\lambda}_k^n(r) - \epsilon \right\}
\]

is not empty and closed. Hence, since \( \epsilon > 0 \) is arbitrary, it must hold that \( \sup_{s \in [0,t]} E[\hat{\lambda}_k^n(s)] \geq E[\sup_{s \in [0,t]} \hat{\lambda}_k^n(s)] \), and conversely it clearly holds that \( \sup_{s \in [0,t]} E[\hat{\lambda}_k^n(s)] \leq E[\sup_{s \in [0,t]} \hat{\lambda}_k^n(s)] \).
\( \hat{\lambda}_k^n(s) \). Similar arguments can moreover be employed to show that \(( \sup_{s \in [0,t]} \hat{\lambda}_k^n(s) )^n = \sup_{s \in [0,t]} \hat{\lambda}_k^n(s) \). It follows that

\[
E[\hat{\lambda}_k^n(t)] \leq t^n E \left[ \left( \sup_{s \in [0,t]} \hat{\lambda}_k(s) \right)^n \right] = t^n \sup_{s \in [0,t]} \hat{m}_{k,n}(s),
\]

so, according to Lemma 3.1, it holds that

\[
\text{so, according to Lemma 3.1, it holds that}
\]

\[
E[\hat{\lambda}_k^n(t)] \leq C(t) < \infty \text{ where } C(t) > 0 \text{ is independent of } k. \text{ Now, let } M_{k,m,n}(t) := E[\hat{\lambda}_k^m(t)\hat{\lambda}_k^n(t)], \text{ for } m, n \geq 0. \text{ Then, according to the Cauchy–Schwarz inequality it holds that}
\]

\[
M_{k,m,n}(t) \leq (E[\hat{\lambda}_k^{2m}(t)]\hat{m}_{k,2n}(t))^{1/2} < C(t),
\]

where the constant \( C(t) > 0 \) is independent of \( k \).

Suppose that \( m, n \geq 0 \). According to Itô’s formula (see Theorem II.33 in Protter [17]) applied to the product \( t \mapsto \hat{\lambda}_k^m(t)\hat{\lambda}_k^n(t) \) it holds that

\[
\hat{\lambda}_k^m(t)\hat{\lambda}_k^n(t) = \int_0^t \hat{\lambda}_k^n(s) \, d(\hat{\lambda}_k^m(s)) + \int_0^t \hat{\lambda}_k^m(s) \, d(\hat{\lambda}_k^n(s)) + \sum_{0 \leq s \leq t} \{ \hat{\lambda}_k^m(s)\hat{\lambda}_k^n(s) - \hat{\lambda}_k^m(s-)\hat{\lambda}_k^n(s-) - \hat{\lambda}_k^m(s-)\Delta\hat{\lambda}_k^n(s) - \hat{\lambda}_k^n(s-)\Delta\hat{\lambda}_k^m(s) \}
\]

\[
= \int_0^t m\hat{\lambda}_k^{m-1}(s)\hat{\lambda}_k^{n+1}(s) \, ds + \int_0^t \hat{\lambda}_k^m(s)\hat{\lambda}_k^n(s-j) a_k \mu_k(\lambda_k(s)) \, ds
\]

\[
+ \sum_{0 \leq s \leq t} \hat{\lambda}_k^m(s)(\hat{\lambda}_k^n(s) - \hat{\lambda}_k^n(s-))
\]

Note moreover that, since \( t \mapsto N_k(t) - \Lambda_k(t) \) is a martingale, it holds that

\[
E \left[ \sum_{0 \leq s \leq t} \hat{\lambda}_k^m(s)(\hat{\lambda}_k^n(s) - \hat{\lambda}_k^n(s-)) \right]
\]

\[
= E \left[ \int_0^t \hat{\lambda}_k^m(s) \left( \int (\hat{\lambda}_k(s-) + a_k \beta_k x)^n v_k(\lambda_k(s-), \lambda_k(s)) \, dx \right) dN_k(s) \right]
\]

\[
= E \left[ \int_0^t \hat{\lambda}_k^m(s) \sum_{j=0}^{n-1} \binom{n}{j} \hat{\lambda}_k^j(s) a_k^{n-j} \beta_k^{n-j} J_{k,n-j}(\lambda_k(s)), \lambda_k(s)) \, ds \right]
\]

\[
= \sum_{j=0}^{n-1} \binom{n}{j} \int_0^t E[ a_k^{n-j} \beta_k^{n-j} J_{k,n-j}(\lambda_k(s)) \hat{\lambda}_k^m(s)\hat{\lambda}_k^n(s-j) ] \, ds,
\]

where we have applied the binomial theorem and Fubini’s theorem. By Assumption 3.1, it holds that

\[
n \int_0^t E[ (a_k \mu_k(\lambda_k(s)) + \beta_k J_{k,1}(\lambda_k(s)) \hat{\lambda}_k(s)) \hat{\lambda}_k^m(s)\hat{\lambda}_k^n(s-j) ] \, ds
\]
Assumption 3.1: There exists a constant \( c_1 \) such that the limit exists, from which it follows by the dominated convergence theorem that the limit \( \lim_{k \to \infty} M_{k,m,0}(t) \) exists. Now, if the limit \( \lim_{k \to \infty} M_{k,m,n-1}(t) \) exists for a fixed \( n \geq 1 \), then the limit \( \lim_{k \to \infty} \psi_{k,m,n}(s) \) also exists and is given by

\[
\lim_{k \to \infty} \psi_{k,m,n}(s) = m \lim_{k \to \infty} M_{k,m,n-1}(s) + \left( nc_0 + \frac{n(n-1)}{2} c_2 \right) \lim_{k \to \infty} M_{k,m,n-1}(s).
\]

Therefore, we may apply the dominated convergence theorem to conclude that the limit \( \lim_{k \to \infty} M_{k,m,n}(s) \) also exists. We shown that by induction that the limit \( \lim_{k \to \infty} M_{k,m,n}(s) \)
exists for all non-negative $m, n \geq 0$. If we set $y_{m,n}(t) := \lim_{k \to \infty} M_{k,m,n}(t)$, $y_{0,0} = 1$ and $y_{-1,n} = 0$ for all $n \geq 1$, it follows that

$$y_{m,n}(t) = e^{-nc_{1}t} \int_{0}^{t} e^{nc_{1}s} \left( my_{m-1,n+1}(s) + \left( nc_{0} + \frac{n(n-1)}{2} c_{2} \right) \lim_{k \to \infty} y_{m,n-1}(s) \right) ds$$

for all $(m, n) \geq (0, 1)$.

According to this equation $y_{m,n}$ is continuous for all $(m, n) \geq (0, 0)$, so an application of the fundamental theorem of the calculus thus yields that

$$y'_{m,n} = -nc_{1}y_{m,n} + \left( nc_{0} + \frac{n(n-1)}{2} c_{2} \right) y_{m,n-1} + my_{m-1,n+1},$$

with $y_{m,n}(0) = 0$. By applying Itô’s formula to $t \mapsto (\int_{0}^{t} Y(s) ds)^{m} Y^{n}(t)$ one can moreover show that the moments $t \mapsto E[(\int_{0}^{t} Y(s) ds)^{m} Y^{n}(t)]$ verify the same system of ODE’s. Hence, since the Laplace transform of $Z(t) = \int_{0}^{t} Y(s) ds$ is known in closed form (see [5]) and is finite in a radius around zero as noted by Dufrense [6]. Hence (according to section 30 in Billingsley [4]), the moments of $Z(t)$ determine its distribution and $\hat{\Lambda}_{k}(t) \to Z(t)$ as $k \to \infty$ in distribution. \hfill \blacksquare

**Example 3.5 (Gamma density):** Suppose that the jump-size distribution is independent of the intensity level, and that $dF(x) = f(x) dx$, where $f(x)$ is the PDF of a gamma distribution:

$$f(x) = \frac{u^{\nu}}{\Gamma(\nu)} x^{\nu-1} e^{-ux}, \quad (18)$$

where $u > 0$ and $\nu > 0$ are constants. By letting $\nu \downarrow 0$, the jumps become smaller and smaller, and if the parameters of the intensity process are adjusted simultaneously so that the intensity becomes higher and higher, the self-exciting process tends to the square root process in distribution at each fixed time point.

To that end, suppose that $\{v_{k}\}_{k \geq 1}$ is a sequence such that $v_{k} > 0$ for all $k \geq 1$, and $v_{k} \to 0$ as $k \to \infty$, and let $a_{k} := 1/\Gamma(v_{k})$, for $k \geq 1$. Furthermore, suppose that for each $k \geq 1$, the stochastic intensity, $\lambda_{k}(t)$, defined in (5), has

$$\lambda_{0k} = c_{0} \frac{\Gamma(v_{k})(1 + v_{k})}{v_{k}}, \quad \beta_{k} = \sqrt{\frac{\Gamma(v_{k})}{v_{k}(1 + v_{k})}},$$

and that the drift rate is linear,

$$\mu_{k}(\lambda) = (\beta_{k} E[X_{k}] + c_{1})(\lambda_{0k} - \lambda),$$

where $c_{0}, c_{1} > 0$ are constants (and $X_{k}$ is gamma distributed with density (18) and $\nu = v_{k}$). Then, since the moments of the gamma distribution are given by

$$E[X_{k}^{j}] = \frac{v_{k}(v_{k} + 1) \cdots (v_{k} + j - 1)}{u},$$

it follows that

$$\mu_{k}(\lambda) + \beta_{k} E[X_{k}] \lambda = (\beta_{k} E[X_{k}] + c_{1})\lambda_{0k} - c_{1}\lambda,$$
and clearly it holds that $a_k(\beta k E[X_k] + c_1)\lambda_0k \to c_0/u$ as $k \to \infty$. Finally, note that

$$\lim_{k \to \infty} d_k^{-1} a_k^j E[X_k^j] = \begin{cases} u^{-2} & \text{if } j = 2, \\ 0 & \text{if } j > 2. \end{cases}$$

So, according to Theorem 3.3, we may conclude that the scaled intensity process tends to a process $Y(t)$ in distribution, where the dynamics of $Y(t)$ are given by the SDE

$$dY(t) = \left( c_0 - \frac{c_1}{u} Y(t) \right) \, dt + \frac{1}{u} \sqrt{Y(t)} \, dB(t),$$

and $Y(0) = 0$.

4. **A particular case: linear intensity process**

We now consider the special case where the intensity process, $\lambda(t)$, is linear. That is,

$$d\lambda(t) = \alpha(\lambda_0 - \lambda(t)) \, dt + \beta \, dU(t). \quad (19)$$

In this section, we will study the expected value and variance of the intensity process in this special case. In this section, we also assume that the jump-size distribution associated to each event is independent of the current value of the intensity process. We shall refer to the constant $\alpha > 0$ as the speed of mean reversion, since it determines the speed of which the intensity is pushed back towards $\lambda_0$. Similarly, we refer to the constant $\beta > 0$ as the jump discontinuity scale, since it scales each jump discontinuity.

4.1. **The expected value of the intensity in the linear case**

Let $m(t)$ denote the expected value of the intensity process, viewed as a function of time, i.e.

$$m(t) := E[\lambda(t)]. \quad (20)$$

Define $\rho := \beta E[X] - \alpha$ and assume that $E[\lambda(0)] = \eta$. Eyjolfsson and Tjøstheim [9] (see their Equation (13)), use Dynkin’s formula and Fubini’s theorem to derive that

$$m'(t) = \alpha \lambda_0 + \rho m(t),
\quad m(0) := \eta. \quad (21)$$

If $\rho \neq 0$, then Equation (21) is an ordinary differential equation (ODE) with solution

$$m(t) = m_0 + m_1 e^{\rho t}, \quad (22)$$

where $m_0 = -\alpha \lambda_0 \rho^{-1}$ and $m_1 = \alpha \lambda_0 \rho^{-1} + \eta$. See Eyjolfsson and Tjøstheim [9] for the details of this derivation, or the following Section 4.2, for how to apply Dynkin’s formula to obtain the ODE.

Based on the ordinary differential Equation (21), we separate the long term behaviour of the expected value, $m(t)$, into three cases:

- **If $\rho > 0$:** In this case, $E[\lambda(t)]$ grows exponentially with time.
If \( \rho = 0 \): In this case, from (21) we get \( m(t) = \eta + \alpha \lambda_0 t \), so \( E[\lambda(t)] \) grows linearly with time.

If \( \rho < 0 \): In this case, \( E[\lambda(t)] \) is bounded and \( E[\lambda(t)] \to m_0 \) as \( t \to \infty \).

Recall that \( \rho := \beta E[X] - \alpha \). Hence, \( \rho > 0 \) means that \( E[X] > \frac{\alpha}{\beta} \), where \( \alpha \) is the speed of mean reversion and \( \beta \) is the jump discontinuity scale for the intensity SDE. Hence, the interpretation is that if the jump sizes are independent and identically distributed, and the expected jump size is larger than the fraction of the speed of mean reversion over the jump discontinuity scale, then the expected intensity rate will grow exponentially with time. Similarly, \( \rho = 0 \) means that \( E[X] = \frac{\alpha}{\beta} \). Hence, if the expected jump size is in perfect balance with the fraction of the speed of mean reversion over the jump discontinuity scale, then the expected intensity will grow linearly with time. Finally, \( \rho < 0 \) means that \( E[X] < \frac{\alpha}{\beta} \).

So, if the expected jump size is smaller than the fraction of the speed of mean reversion over the jump discontinuity scale, then the expected intensity is bounded (and we know that it converges too).

### 4.2. The variance of the intensity in the linear case

Let \( \nu(t) \) denote the second order moment of the intensity process, viewed as a function of time, so

\[
\nu(t) := E[\lambda^2(t)].
\]  

To determine the second moment, we use the same idea as in Section 4.1, and as in Eyjolfsson and Tjøstheim [9]: We use Dynkin’s formula and Fubini’s theorem to derive an ordinary differential equation for \( \nu(t) \). Assume that \( E[\lambda(0)] = \eta \), by Dynkin’s formula with \( f(x) = x^2 \),

\[
E[\lambda^2(t)] = \eta^2 + E\left[ \int_0^t A\lambda(r) \, dr \right].
\]  

where \( A(\cdot) \) is the infinitesimal (or extended) generator of \( \lambda(t) \). Note that according to (8) it holds that

\[
(Af)(\lambda) = 2\lambda\alpha(\lambda_0 - \lambda) + \lambda \int ((\lambda + \beta x)^2 - \lambda^2) \nu(\lambda, dx)
\]

\[
= \lambda(2\alpha\lambda_0 + \beta^2 E[X^2]) + \lambda^2(2\beta E[X] - 2\alpha)
\]

\[
= A\lambda + 2\rho\lambda^2,
\]

where we define \( A := 2\alpha\lambda_0 + \beta^2 E[X^2] \) and \( \rho = \beta E[X] - \alpha \) as before. By inserting this into the application of Dynkin’s formula above in Equation (24), and using Fubini’s theorem to change the order of integration and expectation, we find

\[
\nu(t) = \eta^2 + \int_0^t E[A\lambda(r) + 2\rho\lambda(r)^2] \, dr
\]

\[
= \eta^2 + \int_0^t (Am(r) + 2\rho \nu(r)) \, dr.
\]
Hence,
\[ v'(t) = Am(t) + 2\rho v(t) \]
\[ v(0) = \eta^2. \]  
(25)

Recall that we have an explicit expression for \( m(t) \) from Equation (22). Hence, Equation (25) is an ordinary differential equation in \( v(t) \) which can be solved by standard techniques by inserting the expression for \( m(t) \) from (22). If \( \rho \neq 0 \), then the second moment differential Equation (25) has the solution
\[ v(t) = v_0 + v_1 e^{\rho t} + v_2 e^{2\rho t}, \]  
(26)
where
\[ v_0 = A\alpha \lambda_0 (2\rho^2)^{-1}, \]
\[ v_1 = -A\rho^{-1}(\alpha \lambda_0 \rho^{-1} + \eta), \]  
and \( v_2 = A(2\rho^2)^{-1}(\alpha \lambda_0 + 2\rho \eta) + \eta^2 \) are constants. We now consider the same three cases as in Section 4.1

- If \( \rho > 0 \): By observing the solution (26), we see that the second moment is exponentially increasing with time.
- If \( \rho = 0 \): In this case, \( v'(t) = A(\eta + \alpha \lambda_0 t) \), so \( v(t) = A(\eta t + \alpha \lambda_0 t^2) \), i.e. \( v(t) \) has quadratic growth.
- If \( \rho < 0 \): In this case, \( v(t) \) is bounded and \( v(t) \to v_0 \) as \( t \to \infty \).

The interpretations of these items are similar to those of Section 4.1: Since \( \rho := \beta E[X] - \alpha \), \( \rho > 0 \) means that \( E[X] > \frac{\alpha}{\beta} \), where \( \alpha \) is the speed of mean reversion and \( \beta \) is the jump discontinuity scale for the intensity SDE. Hence, the interpretation is that if the jump sizes are independent and identically distributed, and the expected jump size is larger than the fraction of the speed of mean reversion over the jump discontinuity scale, then the second moment of the intensity process grows exponentially with time. Similarly, \( \rho = 0 \) means that \( E[X] = \frac{\alpha}{\beta} \). Hence, even when the expected jump size is in perfect balance with the fraction of the speed of mean reversion over the jump discontinuity scale, the second moment of the intensity process grows quadratically with time. Finally, \( \rho < 0 \) means that \( E[X] < \frac{\alpha}{\beta} \). So, if the expected jump size is smaller than the fraction of the speed of mean reversion over the jump discontinuity scale, the intensity process is stable, in the sense that the second moment converges to a constant as \( t \to \infty \).

### 4.3. The second moment of the integrated intensity in the linear case

In this subsection, we employ (26) to determine the second moment of the integrated intensity, \( \Lambda(t) \). Note that by an application of Fubini it holds that
\[ E[\Lambda^2(t)] = \int_0^t \int_0^t E[\lambda(r)\lambda(s)] \, dr \, ds. \]
Suppose that \( s \geq r \). Then, it holds that
\[ E[\lambda(r)\lambda(s)] = v(r) + E[\lambda(r)(\lambda(s) - \lambda(r))], \]
where \( v(t) = E[\lambda^2(t)] \). From what we know about the first moment of a linear intensity, given an initial value,
\[ E[\lambda(r)(\lambda(s) - \lambda(r))] = E[\lambda(r)E[(\lambda(s) - \lambda(r)|\lambda(r))]] \]
Therefore, using that 

\[ \lambda(r) \left( \frac{\alpha \lambda_0}{\rho} + \lambda(r) \right) e^{\rho(s-r)} - \frac{\alpha \lambda_0}{\rho} - \lambda(r) \]

may conclude that

\[ \left( \frac{\alpha \lambda_0}{\rho} m(r) + v(r) \right) e^{\rho(s-r)} - \left( \frac{\alpha \lambda_0}{\rho} m(r) + v(r) \right). \]

To simplify notation, suppose that \( \phi(r) := \frac{\alpha \lambda_0}{\rho} m(r) + v(r) \), for any \( r > 0 \), then it follows that

\[
E[\Lambda^2(t)] = \int_0^t \left( \int_0^s (v(r) + \phi(r)(e^{\rho(s-r)} - 1)) \, dr + \int_s^t (v(s) + \phi(s)(e^{\rho(r-s)} - 1)) \, dr \right) \, ds
\]

\[
= 2\int_0^t \int_s^t (v(s) + \phi(s)(e^{\rho(r-s)} - 1)) \, ds \, dr
\]

\[
= 2\int_0^t (t - s)(v(s) - \phi(s)) + \frac{e^{\rho(t-s)} - 1}{\rho} \phi(s) \, ds.
\]

By writing \( m(s) = m_0 + m_1 e^{\rho s} \), like we do in (22), it holds that

\[
\int_0^t (t - s)(v(s) - \phi(s)) \, ds = m_0 \int_0^t (t - s)(m_0 + m_1 e^{\rho s}) \, ds
\]

\[
= m_0 \left( \frac{m_0}{2} t^2 + m_1 \frac{e^{\rho t} - 1 - \rho t}{\rho^2} \right).
\]

Similarly, there exist constants \( c_0, c_1 \) and \( c_2 \) such that \( \phi(s) = c_0 + c_1 e^{\rho s} + c_2 e^{2\rho s} \), so

\[
\int_0^t \frac{e^{\rho(t-s)} - 1}{\rho} \phi(s) \, ds
\]

\[
= \int_0^t \frac{e^{\rho(t-s)} - 1}{\rho} (c_0 + c_1 e^{\rho s} + c_2 e^{2\rho s}) \, ds
\]

\[
= \frac{1}{\rho} \int_0^t (c_0 e^{\rho t} e^{-\rho s} + (c_1 e^{\rho t} - c_0) + (c_2 e^{\rho t} - c_1) e^{\rho s} - c_2 e^{2\rho s}) \, ds
\]

\[
= \frac{1}{\rho} \left( c_0 \frac{e^{\rho t} - 1}{\rho} + (c_1 e^{\rho t} - c_0)t + (c_2 e^{\rho t} - c_1) \frac{e^{\rho t} - 1}{\rho} - c_2 \frac{e^{2\rho t} - 1}{2\rho} \right).
\]

Therefore, using that \( c_0 = v_0 - m_0^2 \), \( c_1 = v_1 - m_0 m_1 \) and \( c_2 = v_2 \) where \( m_0, m_1, v_0, v_1, v_2 \) are the constants in the first and second moment functions, (22) and (26), respectively, we may conclude that

\[
E[\Lambda^2(t)] = k_0 + k_1 t + k_2 t^2 + (C_0 + C_1 t) e^{\rho t} + C_2 e^{2\rho t},
\]

where the constants \( k_0, k_1, k_2, C_0, C_1, C_2 \) are given by

\[
k_0 = \frac{2m_0(m_0 - 2m_1) - 2v_0 + 2v_1 + v_2}{\rho^2},
\]

\[
k_1 = \frac{2m_0(m_0 - m_1) - 2v_0}{\rho}.
\]
\[ k_2 = m_0, \]
\[ C_0 = \frac{2(m_0(2m_1 - m_0) + v_0 - v_1 - v_2)}{\rho^2}, \]
\[ C_1 = \frac{2(v_1 - m_0m_1)}{\rho}, \]
\[ C_2 = \frac{v_2}{\rho^2}, \]

and \( m_0, m_1, v_0, v_1, v_2 \) are the constants in the first and second moment functions, (22) and (26), respectively. It follows that if \( \rho < 0 \) is close to zero, then the effects of a jump fade out slower, than if the \( \rho < 0 \) is further away from zero.

### 4.4. Convergence to deterministic intensity

In this subsection, we will study what happens to the intensity process \( \lambda(t) \) if \( \beta > 0 \) and \( \alpha > 0 \) both converge towards zero while \( \rho \) is kept constant. Note that according to the definition of \( \rho \)

\[
d\lambda(t) = -\alpha\lambda(t) \, dt + \alpha\lambda_0 \, dt + \beta \, dU(t) \\
\quad = \rho\lambda(t) \, dt + \alpha\lambda_0 \, dt + \beta \, d(U(t) - E[X]t) \to \rho\lambda(t) \, dt
\]

as \( \alpha, \beta \to 0 \) while \( \rho \) is kept constant. This means that the stochastic differential equation which determined the intensity process converges towards an ordinary (deterministic) differential equation as \( \alpha, \beta \to 0 \) while \( \rho \) is kept constant. This ODE is \( \lambda' = \rho\lambda \), which, if \( \mathbb{E}[\lambda(0)] = \eta \), means that \( \lambda(t) = \eta e^{\rho t} \). From this, it follows that the self-exciting process \( U(t) \) converges to a non-homogeneous Poisson process with intensity process \( \lambda(t) = \eta e^{\rho t} \).

Hence, if \( \rho < 0 \), then the intensity converges to zero, and no more jumps occur. If \( \rho = 0 \), then the intensity converges to a constant \( \lambda(t) = \eta \), and if \( \rho > 0 \), then the intensity tends to infinity as \( t \to \infty \).

### 5. Conclusions and future work

To conclude, the purpose of this paper has been to investigate the properties of self-exciting processes with intensity processes given by an SDE. The following are the main contributions of the paper. We have:

- Proved that the scaling limit of the intensity process equals the square root process in distribution. We have also proved a similar result for the integrated intensity process.
- Derived explicit expressions for the expectation and variance of the intensity in the case of an intensity given by a linear SDE.
- Discussed conditions under which the first and second moments of the intensity process tend to a constant as time tends to infinity.
- Proved that the linear intensity process converges to a deterministic intensity as \( \alpha \) and \( \beta \) (and \( E[X] \)) go to zero, as long as \( \rho = \beta \mathbb{E}[X] - \alpha \) remains constant.
There is still much to be done in investigating SDE driven self-exciting processes. Work in progress is finding the moments of $U(t)$ and deriving scaling limit results. Furthermore, applications, for instance, looking into stochastic optimal control problems where the state process is given by an SDE driven self-exciting process would be interesting. This is left for future research.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the Research Council of Norway under the SCROLLER project, project number 299897.

**ORCID**

Kristina Rognlien Dahl  http://orcid.org/0000-0001-9564-907X

Heidar Eyjolfsson  http://orcid.org/0000-0002-5751-3574

**References**

[1] E. Abi Jaber, C. Cuchiero, M. Larsson, and S. Pulido, *A weak solution theory for stochastic Volterra equations of convolution type*, Ann. Appl. Probab. 31 (2021), pp. 2924–2952.

[2] K. Atkinson and W. Han, *Theoretical Numerical Analysis: A Functional Analysis Framework*, 3rd ed., Texts in Applied Mathematics, Vol. 39, Springer, New York, 2009.

[3] E. Bacry, I. Mastromatteo, and J.F. Muzy, *Hawkes processes in finance*, Market Microstruct. Liq. 1 (2015), p. 1550005.

[4] P. Billingsley, *Probability and Measure*, Wiley, New York, 1995.

[5] J.C. Cox, J.E. Ingersoll, and S.A. Ross, *A theory of the term structure of interest rates*, Econometrica 53 (1985), pp. 385–407.

[6] D. Dufresne, *The integrated square-root process*, Working Paper No. 90, Centre for Actuarial Studies, University of Melbourne, 2001.

[7] P. Embrecht, T. Liniger, and L. Lin, *Multivariate Hawkes processes: An application to financial data*, J. Appl. Probab. 48 (2011), pp. 367–378.

[8] X. Erny, E. Löcherbach, and D. Loukianova, *Mean field limits for interacting Hawkes processes in a diffusive regime*, Bernoulli 28 (2022) 125–149.

[9] H. Eyjolfsson and D. Tjøstheim, *Self-exciting jump processes with applications to energy markets*, Ann. Inst. Stat. Math. 70 (2018), pp. 373–393.

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

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

[12] T. Jaisson and M. Rosenbaum, *Limit theorems for nearly unstable Hawkes processes*, Ann. Appl. Probab. 25 (2015), pp. 600–631.

[13] E. Lewis and G. Mohler, *A nonparametric EM algorithm for multiscale Hawkes processes*, J. Nonparametr. Stat. 1 (2011), pp. 1–20.

[14] S.P. Meyn and R.L. Tweedie, *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Adv. Appl. Probab. 25(3) (1993), pp. 518–548.

[15] G. Mohler, *Modeling and estimation of multi-source clustering in crime and security data*, Ann. Appl. Stat. 7 (2013), pp. 1525–1539.

[16] Y. Ogata, *On Lewis’ simulation method for point processes*, IEEE Trans. Inf. Theory 27 (1981), pp. 23–31.
[17] P.E. Protter, Stochastic Integration and Differential Equations, Stochastic Modelling and Applied Probability, Vol. 21, Springer-Verlag, Berlin, 2005.

[18] M.A. Rizoiu, L. Xie, S. Sanner, M. Cebrian, H. Yu, and P. Van Hentenryck, Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity, Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 735–744.