Publ. Math. Debrecen
78/1 (2011), 103–126
DOI: 10.5486/PMD.2011.4637

Identification of almost unstable Hawkes processes

By VILMOS PROKAJ (Budapest) and BALÁZS TORMA (Budapest)

Abstract. Self-exciting point processes, also called Hawkes processes are widely used to model credit events (defaults) on bond markets in financial mathematics. This is a point process whose intensity is defined via a feedback mechanism where the input is the past of the point process itself. The identification (calibration) of Hawkes processes is a hot research area. In this paper we consider Hawkes processes in which the feedback path is defined by a finite dimensional linear system. This feedback system admits a stationary solution, i.e. stable, if the integral of the impulse response function of the feedback path is strictly less than one. In this paper we calculate the limit distribution of the appropriately rescaled state process from which we conclude that the intensity process has a diffusion limit. Simulation results for the standard Hawkes process are also presented.

1. Introduction

The Hawkes process $(N(t))_{t\in\mathbb{R}}$ is a self-exciting point process. Its intensity λ depends on the past of the process through the formula

$$\lambda(t) \quad = \quad m + \int_{(-\infty,t)} g(t-u)dN(u), \tag{1}$$

where $m \geq 0$ and $g : [0,\infty) \to [0,\infty)$.

A necessary condition for (1) to have a stationary solution with $\lambda(t) \in L^1$ is that

$$\int_0^\infty g(u)du < 1. \tag{2}$$

Mathematics Subject Classification: 60J60, 60F17, 60G55.
Key words and phrases: Hawkes process, self-exciting point process, stability, Fisher information.

This condition is sufficient for the existence of a stationary process with the structure given above, see [5].

In this paper we consider a class of Hawkes processes, proposed in [4]. In this model the intensity λ satisfies the linear state equations

$$dx(t) = -Ax(t)dt + bdN(t), \tag{3}$$

$$\lambda(t) = c^T x(t-) + m. \tag{4}$$

with a matrix A, and vectors b, c, such that the system's impulse response is non-negative, i.e.

$$g(u) = c^T e^{-Au} b \geq 0, \quad \text{for all } u \geq 0. \tag{5}$$

The sample path of the state process x is right continuous with left limits. The left limit at a given time t is denoted by $x(t-)$.

It is easy to see that Condition (2) and (5) allow us to consider only stable matrices as $-A$, without loss of generality. The stability condition (2)

$$c^T A^{-1} b < 1. \tag{6}$$

Concerning the expected value of the intensity one can easily see that the intensity process is transient if (6) does not hold.

In this paper, our attention is focused on the identifiability of nearly unstable models. That is, we take a parameter sequence (A_n, b_n, c_n) satisfying (5) and (6) such that

$$(A_n, b_n, c_n) \to (A, b, c) \quad \text{as} \quad n \to \infty,$$

where

$$c^T A^{-1} b = 1, \tag{7}$$

and we analyze the limiting behaviour of the state process. This is a first step in understanding the behaviour of the Fisher information matrix near the boundary of the stability domain.

From the above it is clear that we only have to deal with parameters (A, b, c) from a rather limited set, even when we consider boundary points of the domain of stability. So our working assumption throughout the paper is the following.

Assumption 1.

(i) The real part of any eigenvalue of A is strictly positive,

(ii) $g(u) = c^T e^{-Au} b \geq 0$ for all $u \geq 0$,

(iii)

$$\int_0^\infty g(u)du = c^T A^{-1} b \leq 1.$$

We use the term *feasible parameter* for (A, b, c) satisfying Assumption 1.

The log-likelihood of the observation $(N(t))_{0\le t\le T}$ can be written as

$$L_T(\vartheta) = \int_0^T -\hat{\lambda}(t)dt + \int_0^T \ln \hat{\lambda}(t)dN(t)$$

where $\hat{\lambda}(t) = \hat{\lambda}(t, \vartheta)$ is the solution of (3)–(4) with the observed point process $N(t)$ and parameter vector ϑ, see e.g. in [2]. The Fisher information contained in the observation of $(N(t))_{0\le t\le T}$ is

$$I_T(\vartheta) = \mathbf{E}\left[-\partial_\vartheta^2 L_T(\vartheta)\right] = \mathbf{E}\left[\int_0^T \partial_\vartheta^2 \hat{\lambda}(t)dt - \int_0^T \partial_\vartheta^2 \ln \hat{\lambda}(t)dN(t)\right].$$

It follows from the martingale property of $dN(t) - \lambda(t)dt$, that for a non-negative predictable process C

$$\mathbf{E}\left[\int C(t)dN(t)\right] = \int \mathbf{E}\left[C(t)\lambda(t)\right] dt. \tag{8}$$

For details we refer the reader to the book of Brémaud [1].

By (8) the Fisher information contained in the observation can be written as

$$I_T(\vartheta) = \int_0^T \mathbf{E}\left[\frac{(\partial_\vartheta \hat{\lambda}(t))(\partial_\vartheta \hat{\lambda}(t))^T}{\hat{\lambda}^2(t)}\lambda(t)\right] dt. \tag{9}$$

Here we assume that the initial state $x(0)$ is known, therefore $\hat{\lambda}(t) = \lambda(t)$ when ϑ is the true parameter. From identity (9) and the ergodicity we can see that I_T/T has a limit, which we call the time-normalized Fisher information and denote by $I(\vartheta)$, i.e.

$$I(\vartheta) = \mathbf{E}\left[\frac{\lambda_\vartheta \lambda_\vartheta^T}{\lambda}\right],$$

where the expectation is taken with respect to the stationary distribution and λ_ϑ stands for the derivative $\partial_\vartheta \hat{\lambda}$. Recall that $\lambda_\vartheta(t)$ is the derivative of the calculated intensity. So if we form the extended system $\bar{x} = (x, x_\vartheta)^T$ where x_ϑ is the derivative of the state vector with respect to the parameters, then λ_ϑ can be calculated from $\bar{x}$ with a linear transformation or, in other words, with a read-out matrix C, i.e. $\lambda_\vartheta = C\bar{x}$. The extended state vector fulfills a linear state equation similar to (3)–(4) driven by the same point process N. It is easy to see that the parameter, say $(\bar{A}, \bar{b}, \bar{c})$ appearing in the extended state equation is feasible if the parameter for the original system (A, b, c) were feasible. Moreover, extending the state does not affect the stability, i.e., the extended system admits a stationary solution provided that the original system does so. For the simplest case, the standard Hawkes process, see Proposition 17 below.

Thus, in order to understand the limiting behaviour of the time-normalized Fisher information, first we need to understand the limiting behaviour of the stationary system defined by the equations (3)–(4). In this paper we examine this question, by showing that the appropriately rescaled state process converges in law in the Skorohod space $D_{\mathbb{R}^d}[0,\infty)$.

At this point we have to mention the similarities to branching processes with immigration. Our stability condition, that is in the average, one event "generates" strictly less than one direct descendant event, corresponds to the condition of subcriticality for the branching process, i.e. the expectation of the offspring distribution is strictly less than one. For the critical case, when the average number of offsprings is exactly one, the rescaled branching process has a limiting law similar to the one that we obtain in Theorem 2, see [12]. In the nearly critical case, after rescaling the sequence of branching processes with immigration approaching criticality a diffusion limit is obtained in [11]. This is extended to multitype branching processes with immigration in [6]. In both cases the diffusion limit is the same as in our Theorem 2. Our result for the non-dominating part of the state process, that is Theorem 4 below, is probably simpler then the corresponding fluctuation limit theorem for branching processes with immigration, cf. [7, Theorem 2.4]. For Hawkes processes, the "fluctuation limit" is a time-homogeneous diffusion process, actually it is a random multiple of an Orstein–Uhlenbeck process, while for branching processes it is a time-inhomogeneous process of Orstein–Uhlenbeck type.

1.1. Notation. $\ker A$ and $\operatorname{im} A$ denotes the kernel and the range of a linear mapping A, respectively. As usual, I stands for the identity matrix. An idempotent linear mapping P, i.e., $P = P^2$ is called a projection onto $\operatorname{im} P$ along the subspace $\ker P$. We use $|v|$ for the Euclidean norm of the vector v and $|A|$ for the operator norm of the linear mapping A. The Skorohod space is denoted by $D_{\mathbb{R}^d}[0,\infty)$, for details see [3]. Convergence in law of processes is meant in this space.

2. Motivation

In [4] a recursive algorithm was proposed for the identification of standard Hawkes processes. Under general conditions the asymptotic accuracy of the identification is determined by the time-normalized Fisher information.

In the standard case, (3)–(4) reduces to

$$d\lambda(t) = -a(\lambda(t) - m)dt + \sigma dN(t)$$

where a, $\sigma = bc$ and m are positive real parameters. In this model the parameter vector is $\vartheta = (a, \sigma, m)$. For the standard Hawkes process the stability condition simplifies to $\sigma < a$. Notice, that the value of m does not play any role in the stability condition. Moreover, easy calculation shows that the time-normalized Fisher information with respect to m tends to zero if $a - \sigma \to 0$ with m fixed. For this reason we exclude m from the analysis, except for some numerical investigations.

Figure 1 illustrates the effect of approaching criticality. The value of $\sigma = 0.3$ and $m = 0.1$ is kept fixed. On the left hand side of Figure 1, $a = 0.35$, while on the right it is $a = \sigma + 10^{-6}$.

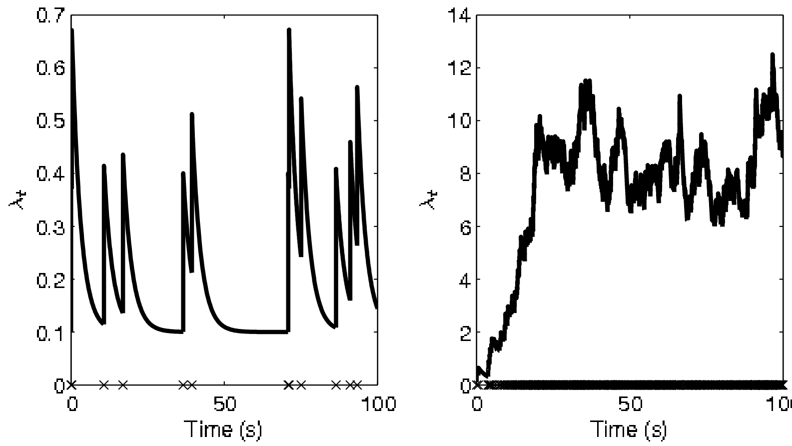

Figure 1. Intensity in the stable (left) and the nearly unstable (right) case.

The density of events and also $\mathbf{E}[\lambda]$ is much larger when the parametrization is close to the boundary of the stability domain, i.e., when $a - \sigma$ is small. Moreover, the nearly unstable intensity process shows diffusion characteristics.

Let us now turn to the time-normalized Fisher information with respect to the parameter a, i.e.,

$$I(a) = \mathbf{E}\left[\frac{\lambda_a^2}{\lambda}\right].$$

To evaluate $I(a)$ the joint law of λ_a and λ is needed. In Figure 2 the scatter plot of λ_a against λ is shown with decreasing decay factor a, where $a = 1$ is the critical value. We can see that the cloud gets narrower as a gets closer to 1.

This indicates an increasing correlation between λ and λ_a. It is easy to calculate the correlation coefficient, which indeed tends to -1 as a goes to 1. Comparing the expected values of λ and λ_a one can see that they have the same order of magnitude. Then, at least at a heuristic level, we can expect that $\lambda_a^2/\lambda \approx \lambda$ and $I(a) \approx \mathbf{E}[\lambda]$ for $a - \sigma \approx 0$. This shows that the time-normalized Fisher information $I(a)$ goes to infinity as a approaches the critical value. This finding is made precise below in Theorem 19.

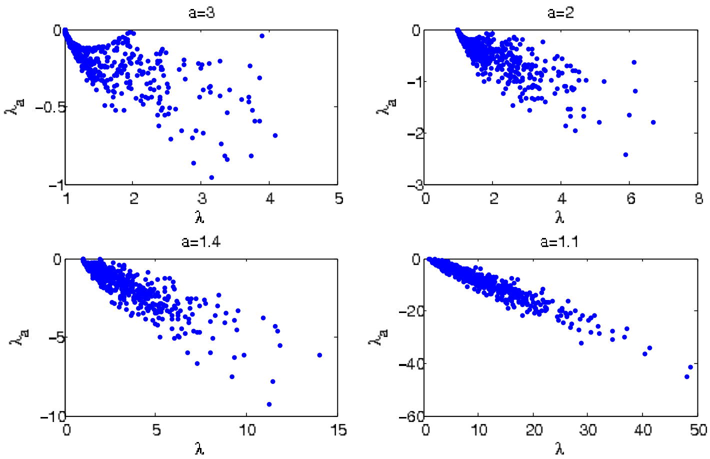

Figure 2. λ_a vs. λ in a standard Hawkes model with $\sigma = m = 1$.

In a similar manner one easily finds that $\lambda_a \approx \lambda_\sigma \approx \lambda \approx (a-\sigma)^{-1}$. Thus, the rescaled Fisher information matrix with respect to parameters a and σ has the form

$$\lim_{a-\sigma \to 0} (a-\sigma) I(a,\sigma) = vv^T$$

where v is a vector with non-zero elements.

Next, we present a simulation experiment, in which $\vartheta = (a, \sigma, m)$ includes m also, since in practice one can not avoid the estimation of m. The time-normalized Fisher information matrix is approximated with a time average

$$\hat{I}(\vartheta) = \frac{1}{T} \int_0^T \frac{\partial_\vartheta \hat{\lambda}(t,\vartheta) \partial_\vartheta \hat{\lambda}(t,\vartheta)^T}{\hat{\lambda}(t,\vartheta)} dt$$

for T large in a long simulation of the standard Hawkes process. We keep the parameters $\sigma = 0.3$ and $m = 0.1$ fixed. Figure 3 shows the diagonal elements of this empirical matrix as a approaches σ from above. The Fisher information with respect to parameters a and σ is a decreasing function of $a - \sigma$, while $\hat{I}(m)$ is increasing. The graph of $\hat{I}(m)$ in accordance with the simple analytical result mentioned above, namely that $I(m)$ tends to zero as $a - \sigma \to 0$ with m fixed.

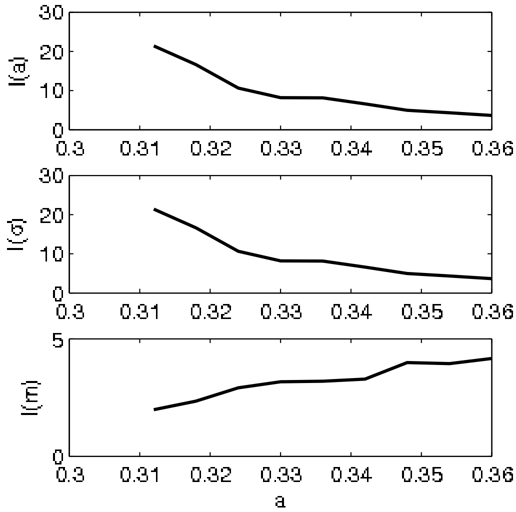

Figure 3. Diagonals of the empirical Fisher information matrix in the standard Hawkes case.

From a practical point of view the inverse of the Fisher information matrix $I^{-1}(\vartheta)$ is even more important then $I(\vartheta)$ itself, since $I^{-1}(\vartheta)$ indicates the accuracy of parameter estimation. For example, in the standard Hawkes model asymptotic normality holds for the maximum likelihood estimator, see [9]. The asymptotic covariance matrix is $I^{-1}(\vartheta)$. Note also that in the standard Hawkes case the overparametrization issue is resolved by introducing $\sigma = bc$.

The inverse of the Fisher information matrix with $a = 0.312$, its eigenvalues z and the condition number κ are

$$\hat{I}^{-1}(a, \sigma, m) = \begin{pmatrix} 0.8737 & 0.8134 & 0.2007 \\ 0.8134 & 0.8059 & 0.1188 \\ 0.2007 & 0.1188 & 0.6605 \end{pmatrix}, \quad z = \begin{pmatrix} 0.0210 \\ 0.6156 \\ 1.7034 \end{pmatrix}, \quad \kappa = 81.11.$$

The parameters a and σ can be estimated by the maximum likelihood method approximately equally accurately, the estimation errors with respect to these two parameters are highly correlated in this nearly unstable case (the correlation coefficient is 0.9694). Moreover, the condition number is moderately high indicating that standard iterative numerical procedures are applicable for maximum likelihood estimation of ϑ in this model.

In Figure 4 the trace of $\hat{I}^{-1}(\vartheta)$ is shown. Theoretical considerations imply that $\mathrm{Tr}\left(I^{-1}(\vartheta)\right)$ should first decrease and then go to infinity as ϑ approaches criticality with m fixed. The curve confirms decreasing but it is incomplete on the left due to the immense computational burden arising very close to criticality.

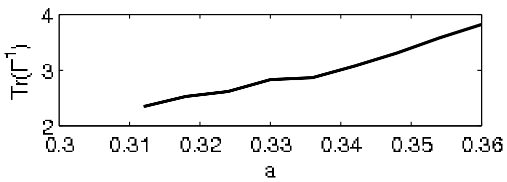

Figure 4. The trace of the empirical asymptotic covariance matrix in the standard Hawkes case.

3. Limit theorems

Theorem 2. *Let* (A_n, b_n, c_n), $n \geq 1$ *and* (A, b, c) *be feasible parameters, with* $c_n^T A_n^{-1} b_n < 1$ *and* $c^T A^{-1} b = 1$. *Assume that* $(A_n, b_n, c_n) \to (A, b, c)$ *and denote* $(x_n(t))_{t\geq 0}$ *the stationary solution of* (3)–(4) *with parameter* (A_n, b_n, c_n).

Let $\bar{x}_n(t) = x_n(\alpha_n t)/\alpha_n$, *where* $\alpha_n = (1 - c_n^T A_n^{-1} b_n)^{-1}$. *Then,*

$$\bar{x}_n \xrightarrow{\mathcal{D}} y^{(1)} A^{-1} b$$

where $(y^{(1)}(t))_{t\geq 0}$ *is the stationary solution of*

$$dy^{(1)}(t) = -a(y^{(1)}(t) - m)dt + a\sqrt{y^{(1)}(t)}dW(t), \tag{10}$$

with $a = (c^T A^{-2} b)^{-1} > 0$ *and* W *is a standard one dimensional Wiener process.*

Note, that the normalizing constant $\alpha_n^{-1} = 1 - c_n^T A_n^{-1} b_n$ is the distance of the integrated impulse response to the critical value 1.

Corollary 3. *Under the assumptions of Theorem 2 the rescaled (stationary) intensity process* $\bar{\lambda}_n(t) = \lambda(\alpha_n t)/\alpha_n$ *converges in distribution to the stationary diffusion process defined by* (10).

Note, that since $a, m > 0$, the equation for $y^{(1)}(t)$ is just the Cox–Ingersoll–Ross (CIR) model for short rate. Its stationary distribution is known to be the Γ-law. With the specific parameters its mean is m and variance is $ma/2$.

Theorem 2 states that when x_n is almost unstable then there is a one-dimensional dominating part of x_n. We will see below that this is due to the spectral properties of the matrix $D_n = -A_n + b_n c_n^T$, namely, that for large n the matrix D_n has a unique small eigenvalue z_n. Moreover, $\ker(D_n - z_n I)$ and $\operatorname{im}(D_n - z_n I)$ are direct summands for n large. This makes it possible to define P_n the projection onto $\ker(D_n - z_n I)$ along the subspace $\operatorname{im}(D_n - z_n I)$. Using P_n we can express the dominating part of x_n as $P_n x_n$. The next statement is about the small, non-dominating part of x_n. With a different scaling, and without speeding up the time it has a limit in distribution as well.

Theorem 4. *Under the assumptions of Theorem 2*

$$\alpha_n^{-1/2}\left((I - P_n)x_n, \sqrt{\lambda_n}\right) \xrightarrow{\mathfrak{D}} \left(y^{(2)}, \sqrt{\lambda}\right),$$

where the process λ has constant sample path, i.e., $\lambda(t) = \lambda(0) \stackrel{d}{=} y^{(1)}(0)$ and $y^{(2)}$ is the stationary solution with values in $\operatorname{im} D$ of the SDE

$$dy^{(2)}(t) = Dy^{(2)}dt + (I - P)b\sqrt{\lambda}dW'(t), \tag{11}$$

where W' is a standard one-dimensional Wiener-process, $D = -A + bc^T$ and P is the projection onto $\ker D$ along the subspace $\operatorname{im} D$.

The process $y^{(2)}/\sqrt{\lambda}$ is a centered Gaussian process, actually it is a multi-dimensional Orstein–Uhlenbeck process. The covariance of the stationary solution is

$$\operatorname{Cov}(y^{(2)}(0)/\sqrt{\lambda}) = \int_0^\infty e^{Ds}(I - P)bb^T(I - P^T)e^{D^T s}ds.$$

4. Proofs

4.1. Some linear algebra. First, we investigate the properties of $D = -A + bc^T$ for a feasible parameter (A, b, c). We denote the spectrum of a matrix A by $\operatorname{sp}(A)$.

Proposition 5. *Under Assumption 1 the real parts of non-zero eigenvalues of $D = -A + bc^T$ are strictly negative. Moreover if $c^T A^{-1} b < 1$ then D is invertible.*

Proof. According to Lemma 20, if $D - zI$ is singular then either $-A - zI$ is singular or $c^T(A + zI)^{-1}b = 1$. Thus, we have

$$\operatorname{sp}(D) \subseteq \operatorname{sp}(-A) \cup \left\{z \in \mathbb{C} : c^T(A + zI)^{-1}b = 1\right\}.$$

$\operatorname{sp}(-A)$ contains eigenvalues with negative real parts by (i) of Assumption 1.

Now let $z = x + iy$ with $x \geq 0$, $y \neq 0$. Then

$$\begin{aligned}\Re c^T(A+zI)^{-1}b &= \int_0^\infty c^T e^{-As} b \Re e^{-zs} ds \\ &= \int_0^\infty c^T e^{-As} b e^{-xs}\cos(ys) ds < c^T A^{-1} b \leq 1.\end{aligned}$$

That is, $c^T(A+zI)^{-1}b \neq 1$, which is also true for $z = 0$ if $c^T A^{-1} b < 1$. Hence the statement follows. □

Proposition 6. *Under Assumption 1 if $c^T A^{-1} b = 1$ then $D = -A + bc^T$ is singular, and the multiplicity of zero in the spectrum is one, i.e.* $\ker D$ *and* $\operatorname{im} D$ *are direct summands, with* $\dim\ker D = 1$.

Proof. $DA^{-1}b = 0$ so D is singular and it can be seen from

$$0 = Dv = -Av + (c^T v)b$$

that any eigenvector with eigenvalue 0 has to be a multiple of $A^{-1}b$. It remains to show that $A^{-1}b \notin \operatorname{im} D$. Note, that by Assumption 1 (i)

$$A^{-n} = \int_0^\infty \frac{s^{n-1}}{(n-1)!} \exp\{-sA\}\, ds, \quad \text{for } n \geq 1.$$

So we get that

$$c^T A^{-1} A^{-1} b = c^T A^{-2} b = \int_0^\infty s c^T e^{-As} b ds > 0.$$

On the other hand

$$c^T A^{-1} Du = -c^T u + c^T A^{-1} b c^T u = 0.$$

Hence $A^{-1}b \notin \operatorname{im} D$ and the proof is complete. □

The projection P onto $\ker D$ along $\operatorname{im} D$ can be given explicitly. Put

$$Pv = \frac{c^T A^{-1} v}{c^T A^{-2} b} A^{-1} b.$$

$P = P^2$ and $\operatorname{im} P = \ker D$ is clear, hence it is enough to show that for any v, we have $v - Pv \in \operatorname{im} D$. Now, take

$$u = (c^T A^{-1} v) A^{-2} b - (c^T A^{-2} b) A^{-1} v.$$

Then, $c^T u = 0$ so

$$Du = -Au = (c^T A^{-2} b) v - (c^T A^{-1} v) A^{-1} b = (c^T A^{-2} b)(v - Pv).$$

Note also, that $Pb = aA^{-1}b$, where $a^{-1} = c^T A^{-2} b > 0$, since $c^T A^{-1} b = 1$.

Corollary 7. *Assume that* $\vartheta_n = (A_n, b_n, c_n)$, $n \geq 1$ *and* $\vartheta = (A, b, c)$ *are feasible parameters, with* $c_n^T A_n^{-1} b_n < 1$ *and* $c^T A^{-1} b = 1$, *and* $\vartheta_n \to \vartheta$.

Then, there is a positive constant γ, *such that for* n *large enough*

$$\mathrm{sp}(D_n) \cap \{z \in \mathbb{C} : \Re z < -\gamma\} = \{z_n\}\,.$$

Moreover, $\ker(D_n - z_n I)$ *and* $\mathrm{im}(D_n - z_n I)$ *are direct summands with* $\dim\ker(D_n - z_n I) = 1$, *and we have* $\lim_{n\to\infty} z_n = 0$.

Proof. Since $D_n \to D$ and the spectrum is continuous in the Hausdorff distance, we have $\mathrm{sp}(D_n) \to \mathrm{sp}(D)$. By Proposition 6 the matrix D is singular and by Proposition 5 there is a $\gamma > 0$ such that each nonzero eigenvalue lies in the half plane $\{z \in \mathbb{C} : \Re z < -2\gamma\}$. With this choice we can define the projection onto the kernel of D along $\mathrm{im}\, D$ with the Cauchy integral formula

$$P = \frac{1}{2\pi i} \oint_{|z|=\gamma} (zI - D)^{-1} dz.$$

At this point we used that $\ker D$ and $\mathrm{im}\, D$ are direct summands.

Since the inversion is continuous and $D_n \to D$ we have that

$$P_n = \frac{1}{2\pi i} \oint_{|z|=\gamma} (zI - D_n)^{-1} dz \to P.$$

Observe that P_n is a projection, that is $P_n^2 = P_n$.

For a projection P we have $\mathrm{Tr}\,(P) = \dim\mathrm{im}\, P$. This can be easily seen by taking first a complete orthonormed system in the range of P and extending it to a basis of the whole space. Calculating the trace in this basis gives immediately that $\dim\mathrm{im}\, P = \mathrm{Tr}\,(P)$.

By the continuity of the trace, $\mathrm{Tr}\,(P_n) = \dim\mathrm{im}\, P_n \to 1$, so if n is large enough then $\dim\mathrm{im}\, P_n = 1$. Since P_n is the projection onto the root space $\sum_{k\geq 1,\, |z|<\gamma} \ker(D_n - zI)^k$ we have that $\mathrm{im}\, P_n = \ker(D_n - z_n I)$ and $P_n D_n = z_n P_n$ with some $z_n \in \mathbb{C}$. Then, $\sum_{k\geq 1} \ker(D_n - z_n I)^k = \ker(D_n - z_n I)$ and $\mathrm{im}(D_n - z_n I) \cap \ker(D_n - z_n I) = \{0\}$. This easily gives that they are direct summands.

To see that $\lim_{n\to\infty} z_n = 0$, note that $\mathrm{sp}(D_n) \to \mathrm{sp}(D) \ni 0$, and z_n is the only element of $\mathrm{sp}(D_n)$ in the ball $|z| < \gamma$. The rest of the statement follows from the fact that $\mathrm{sp}(D) \setminus \{0\} \subset \{z \in \mathbb{C} : \Re z < -2\gamma\}$. □

Proposition 8. *With the notations of Corollary 7 we have*

$$\lim_{n\to\infty} \frac{z_n}{1 - c_n^T A_n^{-1} b_n} = -\frac{1}{c^T A^{-2} b}.$$

Proof. Let $v_n \in \ker(D_n - z_n I) = \operatorname{im} P_n$ so that $c_n^T v_n = 1$. Such a normalization is possible, since $c^T A^{-1} b = 1$, $A^{-1}b \in \operatorname{im} P$ and $P_n \to P$. It also follows that $v_n \to A^{-1}b$. The matrix D_n is invertible by Proposition 5 and

$$c_n^T D_n^{-1} v_n = \frac{1}{z_n} c_n^T v_n = \frac{1}{z_n}.$$

Using Lemma 20 we can write D_n^{-1} as

$$D_n^{-1} = -A_n^{-1} - \frac{A_n^{-1} b_n c_n^T A_n^{-1}}{1 - c_n^T A_n^{-1} b_n}.$$

Thus, with $\alpha_n = (1 - c_n^T A_n^{-1} b_n)^{-1}$ we have that $\alpha_n \to \infty$ and

$$\frac{1}{\alpha_n z_n} = \frac{1}{\alpha_n} c_n^T D_n^{-1} v_n = -\frac{1}{\alpha_n} c_n^T A_n^{-1} v_n - (c_n^T A_n^{-1} b_n) c_n^T A_n^{-1} v_n.$$

Here the first term goes to zero, since $c_n^T A_n^{-1} v_n \to c^T A^{-2} b$. In the second term $c_n^T A_n^{-1} b_n \to c^T A^{-1} b = 1$. So taking $n \to \infty$ gives the statement. □

Corollary 9. *With the notations of Corollary 7*

$$\limsup_{n\to\infty} \sup_{t\geq 0} e^{\frac{at}{2\alpha_n}} \left| e^{D_n t} \right| < \infty,$$

where $a = (c^T A^{-2} b)^{-1}$ *and* $\alpha_n = (1 - c_n^T A_n^{-1} b_n)^{-1}$.

Proof. Since $P_n \to P$ there are positive constants L_1, L_2 such that the sequence of norms $|x|_n = |P_n x| + |(I - P_n)x|$ satisfy

$$L_1 |x| \leq |x|_n \leq L_2 |x| .$$

On the other hand

$$\left| e^{D_n t} P_n x \right| = e^{\Re z_n t} |P_n x| ,$$
$$\left| e^{D_n t} (I - P_n) x \right| \leqq L e^{-\gamma t/2} |(I - P_n)x| ,$$

where γ is from Corollary 7 and L is the constant from Proposition 21. As $z_n \to 0$, for large n the first term will be dominant and we obtain the statement, since $\lim_{n\to\infty} \Re z_n \alpha_n = -a$ by Proposition 8. □

4.2. Proof of Theorem 2. We use the following convention: overbar indicates that the process is obtained with rescaling both the space and time coordinates with $1/\alpha_n$, e.g. $\bar{\lambda}_n = \lambda_n(\alpha_n t)/\alpha_n$, upper indexes (1) and (2) indicate that we have applied the projection P_n and $(I - P_n)$, respectively.

We can write the system in the following form:

$$dx_n(t) = D_n x_n(t)dt + b_n\left(dM_n(t) + mdt\right), \tag{12}$$

$$\lambda_n(t) = c_n^T x_n(t-) + m, \tag{13}$$

where M_n is a martingale. Actually $dM_n(t) = dN_n(t) - \lambda_n(t)dt$. This implies that the quadratic variation process $d[M_n]_t = dN_n(t) = dM_n(t) + \lambda_n(t)dt$. The conditional quadratic variation is $d\left\langle M_n\right\rangle_t = \lambda_n(t)dt$. For details of the calculus used, see [10]. Then the rescaled system has the form:

$$d\bar{x}_n(t) = \alpha_n D_n \bar{x}_n(t)dt + b_n\left(d\bar{M}_n(t) + mdt\right), \tag{14}$$

$$\bar{\lambda}_n(t) = c_n^T \bar{x}_n(t-) + \frac{m}{\alpha_n}, \tag{15}$$

where $\bar{M}_n$ is a martingale with $d\left\langle \bar{M}_n\right\rangle_t = \bar{\lambda}_n(t)dt$.

The stationarity of the system implies that there is no change in the expected value of $x_n(t)$ and $\lambda_n(t)$, i.e.,

$$0 = D_n \mathbf{E}\left[x_n(t)\right] + b_n m,$$
$$0 = -A_n \mathbf{E}\left[x_n(t)\right] + b_n \mathbf{E}\left[\lambda_n(t)\right].$$

This implies $b_n \mathbf{E}\left[\lambda_n(t)\right] = -A_n D_n^{-1} b_n m$. Easy calculation gives that $D_n^{-1} b_n = -\alpha_n A_n^{-1} b_n$, hence $\mathbf{E}\left[\lambda_n(t)\right] = m\alpha_n$, and $\mathbf{E}\left[\bar{\lambda}_n(t)\right] = m$ for all $t \geq 0$ and each n.

We decompose $\bar{x}_n$ into two components $\bar{x}_n^{(1)} = P_n \bar{x}_n$ and $\bar{x}_n^{(2)} = (I - P_n)\bar{x}_n$ and show that $\bar{x}_n^{(2)} \xrightarrow{p} 0$ while $\bar{x}_n^{(1)}$ has the limit indicated in Theorem 2. Before doing so, we prove a useful moment estimation for the stationary solution of (3)–(4). It will be proved by induction on the exponent. The induction step is formulated in the next Proposition.

Proposition 10. *Assume that x is a stationary solution of*

$$dx(t) = Dx(t)dt + b\left(dM(t) + mdt\right).$$

where M is a martingale, with $d[M]_t = dM(t) + \lambda(t)dt$ and $d\left\langle M\right\rangle_t = \lambda(t)dt$. Let $p = 2^k$ with some positive integer k and assume that there are $C, L > 1$,

$0<\gamma\le 1$ such that $\mathbf{E}\left[\lambda^{p/2}(t)\right]\le C$ for all and $t\ge 0$, and $\left|e^{Dt}\right|\le Le^{-\gamma t}$. Then, there is a constant c'_p depending on p only such that

$$\mathbf{E}\left[\left|x(t)\right|^p\right]\le c'_p C\left(\frac{L^2\left|b\right|^2}{\gamma}\right)^{p/2}+\frac{1}{2}\left(\frac{2mL\left|b\right|}{\gamma}\right)^p.$$

Proof. The solution x can be given with the following formula

$$x(t)=e^{Dt}x(0)+\int_0^t e^{D(t-s)}b(dM(s)+mds).$$

Then

$$\begin{aligned}\mathbf{E}&\left[\left|x(t)-e^{Dt}x(0)\right|^p\right]\\&\le 2^{p-1}\mathbf{E}\left[\left|\int_0^t e^{D(t-s)}bdM(s)\right|^p\right]+2^{p-1}m^p\left(\int_0^t\left|e^{D(t-s)}b\right|ds\right)^p.\end{aligned}$$

The second term can be estimated obviously with $(2mL\left|b\right|/\gamma)^p/2$. For the first term we use the Burkholder-Davis-Gundy inequality, see [10, Theorem 48 of Chapter IV],

$$\begin{aligned}\mathbf{E}\left[\left|\int_0^t e^{D(t-s)}bdM(s)\right|^p\right]&\le c_p\mathbf{E}\left[\left(\int_0^t\left|e^{D(t-s)}b\right|^2 d[M](s)\right)^{p/2}\right]\\&\le 2^{p/2-1}c_p\Bigg\{\mathbf{E}\left[\left(\int_0^t\left|e^{D(t-s)}b\right|^2 dM(s)\right)^{p/2}\right]\\&\quad+\mathbf{E}\left[\left(\int_0^t\left|e^{D(t-s)}b\right|^2\lambda(s)ds\right)^{p/2}\right]\Bigg\}.\end{aligned}$$

For the first term in the curly bracket we can write a similar estimation, and iteration of this procedure yields

$$\begin{aligned}\mathbf{E}&\left[\left|\int_0^t e^{D(t-s)}bdM(s)\right|^p\right]\\&\le\sum_{i=1}^k\mathbf{E}\left[\left(\int_0^t\left|e^{D(t-s)}b\right|^{2^i}\lambda(s)ds\right)^{2^{k-i}}\right]\prod_{j=0}^{i-1}c_{2^{k-j}}2^{2^{k-j-1}}.\end{aligned}$$

Applying Hölder inequality we can estimate each term and we get a constant c'_p depending only on p such that

$$2^{p-1}\mathbf{E}\left[\left|\int_0^t e^{D(t-s)}bdM(s)\right|^p\right]\le c'_p C\left(\frac{L^2\left|b\right|^2}{\gamma}\right)^{p/2}.$$

Now we can use the fact that the process is stationary and $e^{Dt}x(0) \xrightarrow{p} 0$ as $t \to \infty$. Thus

$$\left|x(t) - e^{Dt}x(0)\right|^p \xrightarrow{\mathcal{D}} |x(0)|^p\,, \quad \text{as } t \to \infty$$

and the result follows. □

Corollary 11. *For $p \geq 1$ we have*

$$\limsup_{n\to\infty} \frac{1}{\alpha_n^p} \sup_{t\geq 0} \mathbf{E}\left[|x_n(t)|^p\right] < \infty. \tag{16}$$

Proof. Let n_0 so large that for $n > n_0$ we have $\left|e^{tD_n}\right| \leq Le^{-\gamma t/\alpha_n}$ for all $t \geq 0$, where $\gamma > 0$ and $L \geq 1$ not depending on n. This choice is possible by Corollary 9.

We prove (16) for $p = 2^k$, with induction on k. For $k = 1$, i.e., $p = 2$ (16) follows from Proposition 10. Assume that (16) holds for $p/2 = 2^{k-1}$. Then $\mathbf{E}\left[\lambda_n^{p/2}(t)\right] < C\alpha_n^{p/2}$ with some $C \geq 1$ and we apply Proposition 10 to obtain (16) for p. □

Corollary 12. *For $p \geq 1$ we have*

$$\limsup_{n\to\infty} \frac{1}{\alpha_n^p} \sup_{t\geq 0} \mathbf{E}\left[\lambda_n^p(t)\right] < \infty.$$ □

Corollary 13. *For $p \geq 1$ we have*

$$\limsup_{n\to\infty} \frac{1}{\alpha_n^{p/2}} \sup_{t\geq 0} \mathbf{E}\left[\left|x_n^{(2)}(t)\right|^p\right] < \infty.$$

Proof. $x_n^{(2)}$ is the stationary solution of

$$d\bar{x}_n^{(2)}(t) = D_n'\bar{x}_n^{(2)}(t)dt + b_n^{(2)}\left(dM_n(t) + mdt\right),$$

where $b_n^{(2)} = (I - P_n)b_n$. The linear mapping $D_n' = D_n|_{\ker P_n}$ is considered as the homomorphism of $\ker P_n$. By Corollary 7 and Proposition 21, there is a $L, \gamma > 0$ such that for sufficiently large n we have $\left|e^{D_n' t}\right| \leq Le^{-\gamma t}$. Now the statement follows from the application of Proposition 10 and Corollary 12. □

Next, we prove that the growth rate of $\sup_{t\leq T}\left|x^{(2)}(t)\right|^2/\alpha_n$ is not greater then $T^{1/2}$.

Proposition 14. *There is $K > 0$ such that for sufficiently large n and for an arbitrary stopping time τ we have*

$$\mathbf{E}\left[\left|x_n^{(2)}(\tau \wedge T)\right|^4\right] \leq K^2 \alpha_n^2 (1+T), \tag{17}$$

$$\mathbf{E}\left[\sup_{t\leq T}\left|x_n^{(2)}(t)\right|^2\right] \leq 2K\alpha_n (1+T)^{1/2}. \tag{18}$$

Proof. (17) implies (18) with the usual argument. Indeed, let $\tau_u = \inf\{t > 0 : |x^{(2)}(t)|^2 > u\}$. Since $x^{(2)}$ is right continuous we have

$$\begin{aligned}\mathbf{E}\left[\sup_{t\leq T}\left|x_n^{(2)}(t)\right|^2\right] &= \int_0^\infty \mathbf{P}\left[\sup_{t\leq T}\left|x_n^{(2)}(t)\right|^2 > u\right] du \\ &\leq v + \int_v^\infty \frac{\mathbf{E}\left[\left|x_n^{(2)}(T\wedge\tau_u)\right|^4\right]}{u^2} du \leq v + \frac{K^2\alpha_n^2(1+T)}{v}.\end{aligned}$$

Taking infimum in $v > 0$ we get the upper bound in (18).

Next we prove (17). To simplify the notation we drop the index n and the superscript (2), we also assume that n is sufficiently large, thus we can apply the previous results.

Let $V(t) = |x(t)|^2$. The dynamics of V is

$$V(t) = V(0) + \int_0^t 2x^T(s-)dx(s) + \sum_{0\leq s\leq t} \Delta V(s) - 2x^T(s-)\Delta x(s),$$

The jumps of V and x come from the point process $N(t)$, so we have that

$$\Delta V(s) - 2x^T(s-)\Delta x(s) = b^T b dN(s) = |b|^2 \left(dM(s) + \lambda(s)ds\right).$$

Thus,

$$V(t) = V(0) + \int_0^t 2x^T(s)Dx(s) + |b|^2 \lambda(s)ds + \int_0^t 2x^T(s-)b + |b|^2 dM(s).$$

Denote the coefficients of ds and $dM(s)$ by

$$\begin{aligned}\beta(s) &= 2x^T(s)Dx(s) + |b|^2 \lambda(s), \\ \sigma(s) &= x^T(s-)b + |b|^2.\end{aligned}$$

Then $|\beta(s)| \leq C_1 V(s)$ and $\sigma^2(s) \leq C_2(1 + V(s))$ with some positive constants C_1, C_2.

For $V^2(t)$ we obtain

$$V^2(t) = V^2(0) + \int_0^t V(s-)dV(s) + \sum_{0 \leq s \leq t} (\Delta V(s))^2. \tag{19}$$

Here $\Delta V(s) = (2x^T(s-)b + |b|^2)dN(s) = \sigma(s)dN(s)$, hence

$$(\Delta V(s))^2 = \sigma^2(s)(dM(s) + \lambda(s)ds).$$

Hence we obtain that

$$V^2(t) = V^2(0) + \int_0^t V(s)\beta(s) + \sigma^2(s)\lambda(s)ds + \int_0^t V(s-)\sigma(s) + \sigma^2(s)dM(s).$$

The last term is a local martingale. Let τ be a stopping time reducing the last term to a martingale and $T > 0$. Then

$$\mathbf{E}\left[V^2(T \wedge \tau)\right] = \mathbf{E}\left[V^2(0)\right] + \int_0^T \mathbf{E}\left[\chi_{(s \leq \tau)}\left(V(s)\beta(s) + \sigma^2(s)\lambda(s)\right)\right] ds.$$

Now using the moment estimates from Corollary 11 and 12 for $\lambda(s)$ and $V(s) = |x(s)|^2$ we obtain that

$$\mathbf{E}\left[V^2(T \wedge \tau)\right] \leq (1 + T)\alpha^2 C,$$

where C depends only on the values $\sup_t \mathbf{E}\left[V(t)\right]/\alpha$, $\sup_t \mathbf{E}\left[V^2(t)\right]/\alpha^2$ and $\sup_t \mathbf{E}\left[\lambda^2(t)\right]/\alpha^2$, which are uniformly bounded for large n. For general stopping time we can use the usual localization argument and the Fatou lemma to obtain the statement. □

Corollary 15. *For each $T > 0$*

$$\sup_{t \leq T} \left|\bar{x}_n^{(2)}(t)\right| \xrightarrow{p} 0$$

Proof. For n large enough we have

$$\mathbf{E}\left[\sup_{t \leq T} \left|\bar{x}_n^{(2)}(t)\right|^2\right] = \mathbf{E}\left[\sup_{t \leq \alpha_n T} \frac{1}{\alpha_n^2} \left|x_n^{(2)}(t)\right|^2\right] \leq \frac{2K(1 + T\alpha_n)^{1/2}}{\alpha_n} \to 0,$$

where K is a finite constant from Proposition 14. □

Proposition 16. *Write the one-dimensional process* $\bar{x}_n^{(1)}(t)$ *as* $y_n(t)v_n$ *where* $v_n \in \operatorname{im} P_n$ *is such that* $c_n^T v_n = 1$, *i.e.* $y_n(t) = c_n^T \bar{x}_n^{(1)}(t)$. *Then,* $y_n \xrightarrow{\mathcal{D}} y^{(1)}$, *where* $y^{(1)}$ *is the stationary solution of* (10).

Proof. We use a result from the book of Ethier and Kurtz, see [3, Theorem 4.1 of Chapter 7 pp. 354], which we recall for reader's convenience in the appendix, see Theorem 22.

We check the condition of this theorem. Observe that by (14)

$$dy_n(t) = \alpha_n z_n y_n(t)dt + \beta_n m dt + \beta_n dM_n(t),$$

where $z_n P_n = D_n P_n$, M_n is a martingale with

$$d\left\langle M_n\right\rangle_t = \bar{\lambda}_n(t)dt = \left(y_n(t) + \frac{c^T x_n^{(2)}(\alpha_n t) + m}{\alpha_n}\right)dt$$

and $\beta_n = c_n^T P_n b_n$. From Proposition 8 we know that $\alpha_n z_n \to -a$ as $n \to \infty$. Also $\beta_n \to a$, since $\beta_n = c_n^T P_n b_n \to c^T P b = c^T a A^{-1} b = a$.

The jump size of y_n and M_n is uniformly bounded by β_n/α_n and $1/\alpha_n$, respectively, so condition (24) and (25) is obvious.

The drift of $y(t)$ is given by $-a(y-m)$. To check (26) we need that

$$\sup_{t\le T\wedge\tau_n^r} \int_0^t \alpha_n z_n y_n(s) + \beta_n m + a(y_n(s) - m)ds \xrightarrow{p} 0$$

for all $T, r > 0$, where $\tau_n^r = \inf\{t > 0 : |y_n(t)| \vee |y_n(t-)| > r\}$. Since $\alpha_n z_n \to -a$ and $\beta_n \to a$ this is obvious.

We need a similar relation for the diffusion term as well, see assumption (27) below. The quadratic variation of y is $a^2 y(t)dt$ so we need that

$$\sup_{t\le T\wedge\tau_n^r} \int_0^t (c_n^T P b_n)^2 \left(y_n(s) + \frac{c^T x_n^{(2)}(t\alpha_n) + m}{\alpha_n}\right) - a^2 y_n(s)ds \xrightarrow{p} 0$$

for all $T, r > 0$. This is also obvious, since $c_n^T P b_n \to a$, and by Corollary 13 $\mathbf{E}\left[\left|x_n^{(2)}(t)\right|^2\right]/\alpha_n$ is uniformly bounded in n and t.

Now, take a subsequence, say $(y_{n'})$, such that $(y_{n'}(0))$ is convergent in distribution. For this subsequence we can apply Theorem 22 to see that $y_{n'} \xrightarrow{\mathcal{D}} y$. The process y_n is stationary for each n, then so is the limit, hence y is the stationary solution of (10).

It follows now easily, that $y_n \xrightarrow{\mathcal{D}} y$, since by the tightness of $\{y_n(0) : n \ge 1\}$ each subsequence has a sub-subsequence which is convergent in distribution, and as we see, each such convergent subsequence has the same limit, namely the stationary solution of (10). This proves the statement. □

4.3. Proof of Theorem 4. It remains to prove the convergence of

$$\left(\frac{1}{\sqrt{\alpha_n}}(I-P_n)x_n, \frac{1}{\alpha_n}\lambda_n\right).$$

The proof goes along the same line as for $\bar{x}$ and, as before, we denote $(I-P_n)x_n$ by $x_n^{(2)}$.

First we show that $\lambda_n(t)/\alpha_n$ has a constant limit, by showing that $(\lambda_n(t)-\lambda_n(0))/\alpha_n \to 0$ uniformly on compact intervals in probability. Fix $T>0$ and write

$$\begin{aligned}\frac{1}{\alpha_n}\sup_{t\le T}|\lambda_n(t)-\lambda_n(0)| \\ \le \frac{1}{\alpha_n}\int_0^T \left|c_n^T D_n x_n(s)\right| ds + \frac{1}{\alpha_n}\left|c^T b\right|\sup_{t\le T}|M(t)| + \frac{\left|c^T bmT\right|}{\alpha_n}.\end{aligned} \tag{20}$$

It is enough to show that all the three terms on the right hand side go to zero in probability. It is clear for the last term as $\alpha_n \to \infty$. For the second term we can use the Doob inequality and the fact that $\mathbf{E}\left[\lambda_n(t)/\alpha_n\right]$ is bounded uniformly for large n. Finally, the first term can be split into a sum

$$\begin{aligned}\left|c_n^T D_n x_n(s)\right| &\le \left|c_n^T D_n P_n x_n(s)\right| + \left|c_n^T D_n (I-P_n) x_n(s)\right| \\ &\le |z_n|\left|c_n^T\right||x_n(s)| + \left|c_n^T D_n\right|\left|x_n^{(2)}(s)\right|.\end{aligned}$$

Taking expectation and using Proposition 8, Corollary 11 and 13 we obtain

$$\mathbf{E}\left[\left|c_n^T D_n x_n(s)\right|\right] \le C(1+\alpha_n^{1/2})$$

with some finite C. This proves that the first term on the right hand side of (20) goes to zero in probability as well.

Next we check the conditions of Theorem 22 for $\alpha_n^{-1/2}(I-P_n)x_n$. The asymptotic continuity follows from the fact that the jump size goes to zero uniformly in t. The processes A_n, B_n are given by the identities

$$\begin{aligned}A_n(t) &= \int_0^t (I-P_n)bb^T(I-P_n)^T\lambda_n(s)ds, \\ B_n(t) &= \int_0^t D_n x^{(2)}(s)ds.\end{aligned}$$

For the limit diffusion we have that $\beta(y) = Dy$ and $\sigma(y) = (I-P)bb^T(I-P)^T\lambda$. We have to check Condition (26) and (27), i.e.

$$\sup_{t\le T\wedge\tau_n^r}\left|\int_0^t (D_n - D)\frac{x_n^{(2)}(s)}{\alpha_n^{1/2}}ds\right| \xrightarrow{p} 0$$

and

$$\sup_{t\le T\wedge\tau_n^r}\left|\int_0^t \left((I-P_n)b_nb_n^T(I-P_n)^T - (I-P)bb^T(I-P)^T\right)\frac{\lambda_n(s)}{\alpha_n}ds\right| \xrightarrow{p} 0.$$

Both follow from the facts that $\mathbf{E}\left[\alpha_n^{-1/2}|x_n^{(2)}(t)|\right]$ and $\mathbf{E}\left[\lambda_n(t)/\alpha_n\right]$ are uniformly bounded in n and t and

$$(D_n - D) \to 0,$$

$$(I-P_n)b_nb_n^T(I-P_n)^T - (I-P)bb^T(I-P)^T \to 0.$$

The proof of Theorem 4 is completed in the same way as it was done for Theorem 2 using the fact that our processes are stationary and $\alpha_n^{-1/2}x_n^{(2)}(0)$ is tight.

5. Additional results for standard Hawkes process

In this section we focus on the simplest case, when x is scalar-valued. We call the intensity process generated by the one-dimensional model the standard Hawkes process. In this case (3)–(4) reduces to

$$d\lambda(t) = -a(\lambda(t) - m)dt + \sigma dN(t) \tag{21}$$

where a, m, σ are positive real parameters. In this model the parameter vector is $\vartheta^T = (a, \sigma, m)$. In the standard Hawkes model the stability condition simplifies to $\sigma < a$. As indicated in the introduction, the extended system has a linear state space dynamics. The next Proposition shows this for this simplest case. Recall that $\lambda(0)$ is assumed to be known.

Proposition 17. *Let $x(t) = (\lambda(t) - m, \lambda_a(t), \lambda_\sigma(t))^T$, where λ satisfies (21). Then,*

$$dx(t) = -Ax(t)dt + bdN_t,$$

where

$$A = \begin{pmatrix} a & 0 & 0 \\ 1 & a & 0 \\ 0 & 0 & a \end{pmatrix}, \quad b = \begin{pmatrix} \sigma \\ 0 \\ 1 \end{pmatrix}. \qquad \square$$

For the standard Hawkes process with parameter $a > 0$ we can give rather precise estimation for the Fisher information. This is based on the identities given in the next statement.

Proposition 18. *Consider the stationary point process given in* (21) *with* $m = \sigma > 0$. *Then, for any* $k, l \in \mathbb{Z}$, $k \geq 0$ *we have* $\lambda_a^k(\lambda - m)^l \in L^1(\Omega)$ *and*

$$\begin{aligned}\mathbf{E}\left[\lambda_a^k \lambda^{l+1}\right] &= (a(k+l)+m)\mathbf{E}\left[\lambda_a^k(\lambda-m)^l\right] \\ &\quad + k\mathbf{E}\left[\lambda_a^{k-1}(\lambda-m)^{l+1}\right] + \mathbf{E}\left[\lambda_a^k(\lambda-m)^{l+1}\right]. \qquad (22)\end{aligned}$$

Proof. For $k, l \in \mathbb{Z}$, $k \geq 0$ and $\varepsilon > 0$, the integrability of $\lambda_a^k(\lambda - m + \varepsilon)^l$ follows from Proposition 10.

Write the dynamics of $\lambda_a^k(t)(\lambda(t) - m + \varepsilon)^l = x_2^k(t)(x_1 + \varepsilon)^l(t)$ using Proposition 17 and the change of variable formula:

$$\begin{aligned}d(x_1+\varepsilon)^l(t) &= -al(x_1+\varepsilon)^{l-1}(t)x_1(t)dt \\ &\quad + \left((x_1(t-)+\sigma+\varepsilon)^l - (x_1(t-)+\varepsilon)^l\right)dN(t), \\ dx_2^k(t) &= -kx_2^{k-1}(t)\left(x_1(t)+ax_2(t)\right)dt \\ d(x_1(t)+\varepsilon)^l x_2^k(t) &= (x_1+\varepsilon)^l(t)dx_2^k(t) + x_2^k(t)d(x_1+\varepsilon)^l(t)\end{aligned}$$

Since the process $(x_1(t) + \varepsilon)^l x_2^k(t)$ is stationary and in L^1 for all t we have that the mean change is zero. Writing this out, but omitting the actual time t, we obtain that

$$\begin{aligned}&-k\mathbf{E}\left[x_2^{k-1}x_1(x_1+\varepsilon)^l\right] - ka\mathbf{E}\left[x_2^k(x_1+\varepsilon)^l\right] - la\mathbf{E}\left[x_1(x_1+\varepsilon)^{l-1}x_2^k\right] \\ &\qquad + \mathbf{E}\left[x_2^k(x_1+\sigma+\varepsilon)^l\lambda\right] - \mathbf{E}\left[x_2^k(x_1+\varepsilon)^l\lambda\right] = 0.\end{aligned}$$

Rearranging and letting $\varepsilon \to 0+$ gives the relation (22) by $\sigma = m$. For $l \geq 0$ the dominated convergence theorem, for $l < 0$ the Beppo–Levi theorem can be used to see that we can take the limit inside the expectation.

For a given l, the integrability of $\lambda_a^k(\lambda - m)^l$ for all $k \geq 0$ follows from Proposition 10 if $l \geq 0$, while for $l < 0$ from (22) by induction on $-l$. $\square$

Theorem 19. *In the model* (21) *with parameter* $a > 1$ *and fixed* $m = \sigma = 1$, *we have*

$$\frac{2}{(a-1)a(a+1)} - 1 < \mathbf{E}\left[\frac{\lambda_a^2}{\lambda}\right] < \frac{2}{(a-1)a(a+1)}.$$

Proof. By Proposition 18 with $k=2$, $l=-1$ we have

$$\mathbf{E}\left[\frac{\lambda_a^2}{\lambda-1}\right]=-\frac{2}{a+1}\mathbf{E}\left[\lambda_a\right],\quad \mathbf{E}\left[\lambda_a\right]=-\frac{1}{a(a-1)}.$$

This gives the upper bound as $\lambda_a^2/\lambda<\lambda_a^2/(\lambda-1)$.

For the lower bound we apply Proposition 18 with $k=2$, $l=-2$ and with $k=1$, $l=-1$:

$$\begin{aligned}\mathbf{E}\left[\frac{\lambda_a^2}{\lambda}\right]&=\mathbf{E}\left[\frac{\lambda_a^2}{(\lambda-1)^2}\right]+2\mathbf{E}\left[\frac{\lambda_a}{\lambda-1}\right]+\mathbf{E}\left[\frac{\lambda_a^2}{\lambda-1}\right],\\ \mathbf{E}\left[\lambda_a\right]&=\mathbf{E}\left[\frac{\lambda_a}{\lambda-1}\right]+1+\mathbf{E}\left[\lambda_a\right].\end{aligned}$$

Thus $\mathbf{E}\left[\lambda_a/(\lambda-1)\right]=-1$ and

$$\mathbf{E}\left[\frac{\lambda_a^2}{\lambda}\right]=\mathbf{E}\left[\frac{\lambda_a^2}{\lambda-1}\right]-1+\mathbf{D^2}\left[\frac{\lambda_a}{\lambda-1}\right]\geq\mathbf{E}\left[\frac{\lambda_a^2}{\lambda-1}\right]-1. \tag{23}$$

This proves the statement. □

6. Appendix

Lemma 20 (Sherman–Morrison). *Let A be a matrix of dimension $d\times d$, b and c column vectors of dimension d. Then, the matrix*

$$\left(A+bc^T\right)$$

is invertible if A is invertible and $c^TA^{-1}b\neq-1$, in which case we have

$$\left(A+bc^T\right)^{-1}=A^{-1}-\frac{A^{-1}bc^TA^{-1}}{1+c^TA^{-1}b}.$$

Proposition 21 (Proposition A.2.1 of [8]). *For a $q\times q$ matrix D denote $\gamma(D)=\max\left\{\Re z\,:\,z\in\operatorname{sp}(D)\right\}$. Then, for $t\geq0$*

$$\left|e^{Dt}\right|\leq e^{\gamma(D)t}\left(1+2\left|D\right|\sum_{j=1}^{q-1}\frac{\left(2t\left|D\right|\right)^j}{j!}\right).$$

In particular, if $\mathcal{D}$ is a norm bounded collection of $q\times q$ matrices such that for all $D\in\mathcal{D}$ we have $\gamma(D)\leq\gamma<0$ then there is a constant L such that

$$\left|e^{Dt}\right|\leq Le^{\gamma t/2},\quad \text{for all } D\in\mathcal{D}.$$

Next we state a useful result for the weak convergence of processes in a form that is suitable for our purposes.

Theorem 22 (Theorem 4.1 of Chapter 7 in [3])**.** *Assume that β, σ are continuous functions and the equation $e(\beta, \sigma)$*

$$dy(t) = \beta(y(t))dt + \sigma(y(t))dW(t)$$

has a solution, which is unique in law, i.e. the corresponding martingale problem has a unique solution.

For $n \geq 1$ let y_n, B_n be processes with sample path in $D_{\mathbb{R}^d}[0, \infty)$ and $A_n = (A_n^{i,j})$ be a symmetric $d \times d$ matrix valued process, such that A_n has positive semidefinite increments and continuous sample path. Set

$$\mathcal{F}_t^n = \sigma\left\{y_n(s), B_n(s), A_n(s) \,:\, s \leq t\right\}$$

and let $M_n = y_n - B_n$ and $\tau_n^r = \inf\left\{t > 0 \,:\, |y_n(t)| \vee |y_n(t-)| \geq r\right\}$.

Suppose that

(i) *M_n, $M_n M_n^T - A_n$ are $\mathcal{F}^n$-local martingales,*

(ii) *for each $r, T > 0$*

$$\lim_{n\to\infty} \mathbf{E}\left[\sup_{t\leq T\wedge\tau_n^r} |y_n(t) - y_n(t-)|^2\right] = 0, \tag{24}$$

(iii) *for each $r, T > 0$*

$$\lim_{n\to\infty} \mathbf{E}\left[\sup_{t\leq T\wedge\tau_n^r} |B_n(t) - B_n(t-)|^2\right] = 0, \tag{25}$$

(iv) *for each $r, T > 0$*

$$\sup_{t\leq T\wedge\tau_n^r} \left|B_n(t) - \int_0^t \beta(X_n(s))ds\right| \xrightarrow{p} 0, \tag{26}$$

(v) *for each $r, T > 0$*

$$\sup_{t\leq T\wedge\tau_n^r} \left|A_n(t) - \int_0^t \sigma(X_n(s))ds\right| \xrightarrow{p} 0, \tag{27}$$

(vi) *$y_n(0) \xrightarrow{\mathcal{D}} y(0)$.*

Then $y_n \xrightarrow{\mathcal{D}} y$ in $D_{\mathbb{R}^d}[0, \infty)$, where y is the solution of $e(\beta, \sigma)$ with initial value $y(0)$.

Acknowledgments. The authors wish to express their gratitude to László Gerencsér for suggesting the problem and for many stimulating conversations and to thank one of the anonymous referees for drawing the attention of the authors to the papers [6], [7], [11], [12].

References

[1] Pierre Brémaud, Point Processes and Queues, *Springer-Verlag*, *New York*, 1981.

[2] D. J. Daley and D. Vere-Jones, An Introduction to the Theory of Point Processes, Vol. I, Probability and its Applications (New York), *Springer-Verlag*, *New York*, 2003.

[3] Stewart N. Ethier and Thomas G. Kurtz, Markov processes, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, *John Wiley* & *Sons Inc.*, *New York*, 1986.

[4] L. Gerencsér, C. Matias, Zs. Vágó, B. Torma and B. Weiss, Self-exciting point processes with applications in finance and medicine, 18th International Symposium on Mathematical Theory of Networks and Systems, 2008.

[5] Alan G. Hawkes and David Oakes, A cluster process representation of a self-exciting process, *J. Appl. Probability* **11** (1974), 493–503.

[6] Márton Ispány and Gyula Pap, Weak convergence of step processes and an application for critical multitype branching processes with immigration, arXiv:math/0701803v1.

[7] Márton Ispány, Gyula Pap and M. C. A. van Zuijlen, Fluctuation limit of branching processes with immigration and estimation of the means, *Adv. in Appl. Probab.* **37** (2005), 523–538.

[8] Yuri Kabanov and Sergei Pergamenshchikov, Two-Scale Stochastic Systems, Vol. 49, Applications of Mathematics (New York), *Springer-Verlag*, *Berlin*, 2003.

[9] Yosihiko Ogata, The asymptotic behaviour of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math.* **30** (1978), 243–261.

[10] Philip E. Protter, Stochastic Integration and Differential Equations, Vol. 21, Applications of Mathematics (New York), *Springer-Verlag*, *Berlin*, 2004.

[11] T. N. Sriram, Invalidity of bootstrap for critical branching process with immigration, *Ann. Statist.* **22** (1994), 1013–1023.

[12] C. Z. Wei and J. Winnicki, Some asymptotic results for the branching process with immigration, *Stochastic Processes and their Applications* **31** (1989), 261–282.

VILMOS PROKAJ
EÖTVÖS UNIVERSITY BUDAPEST
DEPARTMENT OF PROBABILITY AND STATISTICS
COMPUTER AND AUTOMATION INSTITUTE OF THE HUNGARIAN
ACADEMY OF SCIENCES

E-mail: prokaj@cs.elte.hu

BALÁZS TORMA
COMPUTER AND AUTOMATION INSTITUTE OF THE HUNGARIAN
ACADEMY OF SCIENCES

E-mail: torma@sztaki.hu

(Received June 18, 2009; revised February 27, 2010)