MATRIX CALCULATIONS FOR MOMENTS OF MARKOV PROCESSES

ANDREW DAW ,* *University of Southern California*
JAMOL PENDER ,** *Cornell University*

Abstract

Matryoshka dolls, the traditional Russian nesting figurines, are known worldwide for each doll's encapsulation of a sequence of smaller dolls. In this paper, we exploit the structure of a new sequence of nested matrices we call *matryoshkan matrices* in order to compute the moments of the one-dimensional polynomial processes, a large class of Markov processes. We characterize the salient properties of matryoshkan matrices that allow us to compute these moments in closed form at a specific time without computing the entire path of the process. This simplifies the computation of the polynomial process moments significantly. Through our method, we derive explicit expressions for both transient and steady-state moments of this class of Markov processes. We demonstrate the applicability of this method through explicit examples such as shot noise processes, growth–collapse processes, ephemerally self-exciting processes, and affine stochastic differential equations from the finance literature. We also show that we can derive explicit expressions for the self-exciting Hawkes process, for which finding closed-form moment expressions has been an open problem since their introduction in 1971. In general, our techniques can be used for any Markov process for which the infinitesimal generator of an arbitrary polynomial is itself a polynomial of equal or lower order.

Keywords: Markov processes; matryoshkan matrices; moments; polynomial processes

2020 Mathematics Subject Classification: Primary 60J25
Secondary 47D07; 65C40

1. Introduction

In our recent study of the intensity of Markovian Hawkes processes, originally defined in [33], we have been interested in computing all the moments of this process. In surveying the previous literature for this process, there do not seem to be any closed-form transient solutions at the fourth order or higher (see Proposition 5 of [29] for moments one through three), and both steady-state solutions and ordinary differential equations (ODEs) have only been available up to the fourth moment; see [15, 22]. Similarly, [2] give expressions for the fourth transient moment of a self-exciting jump-diffusion model up to squared error in the length of time, and one could simplify these expressions to represent the Markovian Hawkes intensity with the same error. The standard methodology for finding moments is to differentiate the moment

generating function to obtain the moments, but this is intractable for practical reasons; see for example [22]. The problem of finding the moments of the Hawkes process intensity is also the subject of the recent interesting research in [13, 14], works that are concurrent with and independent from this one. In [13], the authors propose a new approach for calculating moments that they construct from elementary probability arguments and also relate to the infinitesimal generator. Like the infinitesimal generator, this new methodology produces differential equations that can be solved algebraically or numerically to yield the process moments, and the authors provide closed-form transient expressions up to the second moment. The paper [14] extends this methodology to cases of gamma decay kernels. In other recent previous works [18, 41], the authors have identified the differential equation for an arbitrary moment of the Hawkes process intensity, although the closed-form solutions for these equations have remained elusive and prompted closer investigation. Upon inspecting the differential equation for a given moment of the Hawkes process intensity, one can notice that this expression depends on the moments of lower order. Thus, to compute a given moment one must solve a system of differential equations with size equal to the order of the moment, meaning one must at least implicitly solve for all the lower-order moments first. This same pattern occurs in [13]. Noticing this nesting pattern leads one to wonder: what other processes have moments that follow this structure?

In this paper, we explore this question by identifying what exactly this nesting structure is. In the sequel, we will define a novel sequence of matrices that captures this pattern. Just as matryoshka dolls—the traditional Russian nesting figurines—stack inside of one another, these matrices are characterized by their encapsulation of their predecessors in the sequence. Hence, we refer to this sequence as **matryoshkan matrices**. As we will show, these matrices can be used to describe the linear system of differential equations that arise in solving for the moments of the Hawkes process, as well as the moments of a large class of other Markov processes. Similar structure has been successfully leveraged previously in the case of discrete state spaces (see, e.g., [44, 43, 5]), but here we consider continuous state spaces as well, such as our original motivating example, the Hawkes process. As we will demonstrate through detailed examples, this includes a wide variety of popular stochastic processes, such as Itô diffusions and shot noise processes. By utilizing this nesting structure we are able to solve for the moments of these processes in closed form. In comparison to traditional methods of solving these systems of differential equations numerically, the advantage of the approach introduced herein is the fact that the moments can be computed at a specific point in time rather than on a path through time.

In fact, the only assumption we make on these processes is that their moments satisfy differential equations that do not depend on any higher-order moments. Notably, this assumption exactly coincides with the definition of *polynomial processes*, hence specifying the exact class of Markov processes we will consider. In [11], this is defined as the collection of Markov processes such that the expected value of any polynomial is 'again a polynomial in the initial value of the process' up to the same order. Hence, the processes we study here are precisely the one-dimensional polynomial processes. The paper [11] demonstrates that the moments of such processes can be computed through a matrix exponential, and thus our contributions to the polynomial process literature are that the sequence of these matrices (and their inverses, exponentials, and decompositions) exhibit a convenient nested structure that enables simple iterative computation. Hence, this paper can be seen as building on a line of work that aims to efficiently and/or recursively compute polynomial process moments, such as [26, 1, 3]. This also provides context and relevance for the various examples on which we demonstrate our

methods. For example, we show the nested matrix structure arising in Itô diffusions, and this connects to a rich history of study on polynomial diffusions; see for example [47, 42, 26, 24, 25] and references therein. Polynomial processes have been used to great success in many contexts in finance, such as [4, 10, 31], in addition to many of the papers mentioned above.

The breadth of applications of polynomial processes suggests that our methodology has the potential to be quite relevant in practice. Of course, these techniques can be used to efficiently calculate the commonly used first four moments, thus obtaining the mean, variance, skewness, and kurtosis. Moreover, though, let us note that the higher moment calculations are also of practical use. For example, these higher moments can be used in Markov-style concentration inequalities, as the higher order should improve the accuracy of the tail bounds. To that end, one can also use the vector of moments to approximate generating functions such as the moment generating function or Laplace–Stieltjes transforms. This could then be used to characterize the stationary distribution of the process, for example, or to provide approximate calculations of quantities such as the cumulative distribution function through transform methods. The calculation of moments is also highly relevant for many applications in mathematical finance. One could also expect this efficiently calculated vector of moments to be of use in estimation through method of moments techniques. Again in this case, access to higher-order moments should improve fit.

The remainder of this paper is organized as follows. In Section 2, we introduce matryoshkan matrix sequences and identify some of their key properties. In Section 3 we use these matrices to find the moments of a large class of general Markov processes—the one-dimensional polynomial processes. We also give specific examples. In Section 4, we conclude. Throughout the course of this study, we make the two following primary contributions:

(i) We define a novel class of matrix sequences, which we call matryoshkan matrix sequences for their nesting structure. We identify key properties of these matrices, such as their inverses and matrix exponentials.

(ii) Through these matryoshkan matrices, we solve for closed-form expressions for the moments of polynomial Markov processes. Furthermore, we demonstrate the general applicability of this technique through application to notable stochastic processes, including Hawkes processes, shot noise processes, Itô diffusions, growth–collapse processes, and linear birth–death–immigration processes. In the case of the Hawkes process and growth–collapse processes this resolves an open problem, as closed-form expressions for these general transient moments were not previously known in the literature.

2. Matryoshkan matrix sequences

For the sake of clarity, let us begin this section by introducing general notation patterns we will use throughout this paper. Because of the heavy use of matrices in this work, we reserve boldface uppercase variables for these objects; for example, we use $\mathbf{I}$ for the identity matrix. Similarly we let boldface lowercase variables denote vectors—for example, using $\mathbf{v}$ for the vector of all ones or $\mathbf{v}_i$ for the unit vector in the ith direction. One can assume that all vectors are column vectors unless otherwise noted. Scalar terms will not be bolded. A special matrix that we will use throughout this work is the diagonal matrix, denoted by $\mathbf{diag}(\mathbf{a})$, which is a square matrix with the values of the vector $\mathbf{a}$ along its diagonal and zeros elsewhere. We will also make use of a generalization of this, denoted by $\mathbf{diag}(\mathbf{a}, k)$, which instead contains the

values of **a** on the kth off-diagonal, with negative k being below the diagonal and positive k being above.

Let us now introduce a sequence of matrices that will be at the heart of this work. We begin as follows: consider a sequence of lower triangular matrices $\{\mathbf{M}_n, n \in \mathbb{Z}^+\}$ such that

$$\mathbf{M}_n = \begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n & m_{n,n} \end{bmatrix}, \tag{1}$$

where $\mathbf{m}_n \in \mathbb{R}^{n-1}$ is a row vector, $m_{n,n} \in \mathbb{R}$, and $\mathbf{M}_1 = m_{1,1}$, an initial value. Taking inspiration from matryoshka dolls, the traditional Russian nesting dolls, we will refer to these objects as *matryoshkan* matrices. Using their nested and triangular structures, we can make four quick observations of note regarding matryoshkan matrices.

Proposition 1. *Each of the following statements is a consequence of the definition of matryoshkan matrices given by Equation* (1):

(i) *If* $\mathbf{X}_n \in \mathbb{R}^{n \times n}$ *and* $\mathbf{Y}_n \in \mathbb{R}^{n \times n}$ *are both matryoshkan matrix sequences, then so are* $\mathbf{X}_n + \mathbf{Y}_n$ *and* $\mathbf{X}_n \mathbf{Y}_n$.

(ii) *If* $m_{i,i} \neq 0$ *for all* $i \in \{1, \ldots, n\}$ *then the matryoshkan matrix* $\mathbf{M}_n \in \mathbb{R}^{n \times n}$ *is nonsingular. Moreover, the inverse of* $\mathbf{M}_n$ *is given by the recursion*

$$\mathbf{M}_n^{-1} = \begin{bmatrix} \mathbf{M}_{n-1}^{-1} & \mathbf{0}_{n-1 \times 1} \\ -\frac{1}{m_{n,n}} \mathbf{m}_n \mathbf{M}_{n-1}^{-1} & \frac{1}{m_{n,n}} \end{bmatrix}. \tag{2}$$

(iii) *If* $m_{i,i} \neq 0$ *for all* $i \in \{1, \ldots, n\}$ *and are all distinct, then the matrix exponential of the matryoshkan matrix* $\mathbf{M}_n \in \mathbb{R}^{n \times n}$ *multiplied by* $t \in \mathbb{R}$ *follows the recursion*

$$e^{\mathbf{M}_n t} = \begin{bmatrix} e^{\mathbf{M}_{n-1} t} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n} \mathbf{I})^{-1} (e^{\mathbf{M}_{n-1} t} - e^{m_{n,n} t} \mathbf{I}) & e^{m_{n,n} t} \end{bmatrix}. \tag{3}$$

(iv) *If* $m_{i,i} \neq 0$ *for all* $i \in \{1, \ldots, n\}$ *and are all distinct, then the matrices* $\mathbf{U}_n \in \mathbb{R}^{n \times n}$ *and* $\mathbf{D}_n \in \mathbb{R}^{n \times n}$ *are such that*

$$\mathbf{M}_n \mathbf{U}_n = \mathbf{U}_n \mathbf{D}_n$$

for the matryoshkan matrix $\mathbf{M}_n \in \mathbb{R}^{n \times n}$, *when defined recursively as*

$$\mathbf{U}_n = \begin{bmatrix} \mathbf{U}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{D}_{n-1} - m_{n,n} \mathbf{I})^{-1} \mathbf{U}_{n-1} & 1 \end{bmatrix}, \ \mathbf{D}_n = \begin{bmatrix} \mathbf{D}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{0}_{1 \times n-1} & m_{n,n} \end{bmatrix}. \tag{4}$$

One can pause to note that in some sense any lower triangular matrix could be considered matryoshkan, or at least could satisfy these properties. However, we note that some of the most significant insights we can gain from the matryoshkan structure are the recursive implications available for sequences of matrices. Moreover, it is the combination of the nested relationship of consecutive matrices and the lower triangular structure that enables us to find these patterns. We will now see how this notion of matryoshkan matrix sequences and the associated properties above can be used to specify element-wise solutions to a sequence of differential equations.

Lemma 1. *Let $\mathbf{M}_n \in \mathbb{R}^{n \times n}$, $\mathbf{c}_n \in \mathbb{R}^n$, and $\mathbf{s}_n(t) : \mathbb{R}^+ \to \mathbb{R}^n$ be such that*

$$\mathbf{M}_n = \begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n & m_{n,n} \end{bmatrix}, \quad \mathbf{c}_n = \begin{bmatrix} \mathbf{c}_{n-1} \\ c_n \end{bmatrix}, \quad \text{and} \quad \mathbf{s}_n(t) = \begin{bmatrix} \mathbf{s}_{n-1}(t) \\ s_n(t) \end{bmatrix},$$

where $\mathbf{m}_n \in \mathbb{R}^{n-1}$ is a row vector, $\mathbf{c}_{n-1} \in \mathbb{R}^{n-1}$, $s_n(t) \in \mathbb{R}$, and $\mathbf{M}_1 = m_{1,1}$. Further, suppose that

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{s}_n(t) = \mathbf{M}_n \mathbf{s}_n(t) + \mathbf{c}_n.$$

Then, if $m_{k,k} \neq 0$ for all $k \in \{1, \ldots, n\}$, the vector function $\mathbf{s}_n(t)$ is given by

$$\mathbf{s}_n(t) = e^{\mathbf{M}_n t} \mathbf{s}_n(0) - \mathbf{M}_n^{-1} (\mathbf{I} - e^{\mathbf{M}_n t}) \mathbf{c}_n, \tag{5}$$

and if all $m_{k,k} \neq 0$ for $k \in \{1, \ldots, n\}$ are distinct, the nth scalar function $s_n(t)$ is given by

$$s_n(t) = \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n} \mathbf{I})^{-1} (e^{\mathbf{M}_{n-1} t} - e^{m_{n,n} t} \mathbf{I}) \left(\mathbf{s}_{n-1}(0) + \frac{\mathbf{c}_{n-1}}{m_{n,n}} \right) + e^{m_{n,n} t} s_n(0)$$

$$- \frac{c_n}{m_{n,n}} (1 - e^{m_{n,n} t}) + \frac{\mathbf{m}_n}{m_{n,n}} \mathbf{M}_{n-1}^{-1} (\mathbf{I} - e^{\mathbf{M}_{n-1} t}) \mathbf{c}_{n-1}, \tag{6}$$

where $t \geq 0$.

In Equation (5), we can see that the vector solution to the *n*-dimensional ODE system is written in terms of both the inverse and the exponential of the *n*th matryoshkan matrix. To compute the full vector solution, then, we could simply directly exploit the matryoshkan structure of both the inverse and the exponential from Equations (2) and (3) and calculate each individual matrix function recursively. However, one promising method to compute these at once would be to instead leverage the matryoshkan form of the diagonalized decomposition in Equation (4). That is, both the matrix exponential and the inverse can of course be easily computed from the decomposition matrices, but this way one can also conveniently use the recursion in Equation (4) to simultaneously compute terms needed for both the matrix exponential and the inverse. At each step in the recursion, one needs to compute $\mathbf{U}_n$, $\mathbf{U}_n^{-1}$, and $\mathbf{D}_n$. Because $\mathbf{U}_n$ is itself a matryoshkan matrix, note that we can use Equation (2) to find $\mathbf{U}_n^{-1}$. Starting from $\mathbf{U}_1 = \mathbf{U}_1^{-1} = 1$ and $\mathbf{D}_1 = m_{1,1}$ with pre-allocated space, the update from $n-1$ to n takes $O(1)$ operations to append $m_{n,n}$ to $\mathbf{D}_{n-1}$, $O(n-1)$ operations to compute the divisions needed to form the vector $\mathbf{m}_n (\mathbf{D}_{n-1} - m_{n,n} \mathbf{I})^{-1}$ for $\mathbf{U}_n^{-1}$, and $O((n-1)^2)$ operations to then multiply this vector by $\mathbf{U}_{n-1}$ for $\mathbf{U}_n$. Thus, Step i is $O((i-1)^2)$ for $i \in \{1, \ldots n\}$, yielding that in sum it takes $O(n^3)$ operations to form $\mathbf{U}_n$, $\mathbf{U}_n^{-1}$, and $\mathbf{D}_n$. Naturally, these same ideas apply to the computation of Equation (6), where many matryoshkan forms—the inverses, exponentials, and even products of square matrices—can be recognized and calculated recursively.

With this lemma in hand, we can now move to using these matrix sequences for calculating Markov process moments. To do so, we will use the infinitesimal generator, a key tool for Markov processes, to find the derivatives of the moments through time. By identifying a matryoshkan matrix structure in these differential equations, we are able to apply Lemma 1 to find closed-form expressions for the moments.

3. Calculating moments through matryoshkan matrix sequences

In this section we connect matryoshkan matrix sequences with the moments of Markov processes. In Subsection 3.1, we prove the main result, which is the computation of the moment of

a one-dimensional polynomial Markov process through matryoshkan matrices. To demonstrate the applicability of this result, we now apply it to a series of examples. First, in Subsection 3.2, we obtain the moments of the self-exciting Hawkes process, for which finding moments in closed form has been an open problem. Then, in Subsection 3.3, we study the Markovian shot noise process, a stochastic intensity process that trades self-excitement for external shocks. Next, in Subsection 3.4, we showcase the use of these techniques for diffusive dynamics through application to Itô diffusions. In Subsection 3.5, we consider growth–collapse processes as an application of these techniques to models with state-dependent and randomly sized down jumps. Finally, in Subsection 3.6 we apply this technique to a process with jumps both upwards and downwards, a linear birth–death–immigration process we have studied previously. In each scenario, we describe the process of interest, define the infinitesimal generator, and identify the matrix structure. Through this, we solve for the process moments.

3.1. The moments of general polynomial Markov processes

The connection between matryoshkan matrices and Markov processes is built upon a key tool for Markov processes, the infinitesimal generator. For a detailed introduction to infinitesimal generators and their use in studying Markov processes, see e.g. [23]. For a Markov process X_t on a state space $\mathbb{S}$, the infinitesimal generator on a function $f : \mathbb{S} \to \mathbb{R}$ is defined as

$$\mathcal{L}f(x) = \lim_{\tau \to 0} \frac{\mathbb{E}\big[f(X_\tau) \,|\, X_0 = x\big] - f(x)}{\tau}.$$

In our context and in many others, the power of the infinitesimal generator comes through use of Dynkin's formula, which gives us that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[f(X_t)\big] = \mathbb{E}\big[\mathcal{L}f(X_t)\big].$$

To study the moments of a Markov process, we are interested in functions f that are polynomials. Let us suppose now that $\mathcal{L}x^n$ for any $n \in \mathbb{Z}^+$ is polynomial in the lower powers of x for a given Markov process X_t, meaning that X_t is a one-dimensional polynomial process. Then we can write

$$\mathcal{L}X_t^n = \theta_{0,n} + \sum_{i=1}^n \theta_{i,n} X_t^i,$$

which implies that the differential equation for the nth moment of this process is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[X_t^n\big] = \theta_{0,n} + \sum_{i=1}^n \theta_{i,n}\mathbb{E}\big[X_t^i\big],$$

for some collection of constants $\theta_{0,n}$, $\theta_{1,n}$, $\theta_{n,n}$. Thus, to solve for the nth moment of X_t we must first solve for all the moments of lower order. We can also observe that to solve for the $(n-1)$th moment we must have all the moments below it. In comparing these systems of differential equations, we can see that all of the equations in the system for the $(n-1)$th moment are also in the system for the nth moment. No coefficients are changed in any of these lower moment equations; the only difference between the two systems is the inclusion of the differential equation for the nth moment in its own system. Hence, the nesting matryoshkan structure arises. By expressing each system of linear ODEs in terms of a vector of moments, a matrix of coefficients, and a vector of shift terms, we can use these matrix sequences to

capture how one moment's system encapsulates all the systems below it. This observation then allows us to calculate all the moments of the process in closed form, as we now show in Theorem 1. In steady state, this also implies that the $(n+1)$th moment can be computed as a linear combination of moments 1 through n.

Theorem 1. *Let X_t be a Markov process such that the time derivative of its nth moment can be written as*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t^n] = \theta_{0,n} + \sum_{i=1}^{n} \theta_{i,n}\mathbb{E}[X_t^i], \tag{7}$$

for any $n \in \mathbb{Z}^+$, where $t \geq 0$ and $\theta_{i,n} \in \mathbb{R}$ for all $i \leq n$. Let $\boldsymbol{\Theta}_n \in \mathbb{R}^{n \times n}$ be defined recursively by

$$\boldsymbol{\Theta}_n = \begin{bmatrix} \boldsymbol{\Theta}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \boldsymbol{\theta}_n & \theta_{n,n} \end{bmatrix}, \tag{8}$$

where $\boldsymbol{\theta}_n = [\theta_{1,n}, \ldots, \theta_{n-1,n}]$ and $\boldsymbol{\Theta}_1 = \theta_{1,1}$. Furthermore, let $\boldsymbol{\theta}_{0,n} = [\theta_{0,1}, \ldots, \theta_{0,n}]^{\mathrm{T}}$. Then, if $\theta_{k,k} \neq 0$ for all $k \in \{1, \ldots, n\}$ are distinct, the nth moment of X_t can be expressed as

$$\mathbb{E}[X_t^n] = \boldsymbol{\theta}_n(\boldsymbol{\Theta}_{n-1} - \theta_{n,n}\mathbf{I})^{-1}\left(e^{\boldsymbol{\Theta}_{n-1}t} - e^{\theta_{n,n}t}\mathbf{I}\right)\left(\mathbf{x}_{n-1}(x_0) + \frac{\boldsymbol{\theta}_{0,n-1}}{\theta_{n,n}}\right)$$

$$+ x_0^n e^{\theta_{n,n}t} - \frac{\theta_{0,n}(1 - e^{\theta_{n,n}t})}{\theta_{n,n}} + \frac{\boldsymbol{\theta}_n}{\theta_{n,n}}\boldsymbol{\Theta}_{n-1}^{-1}(\mathbf{I} - e^{\boldsymbol{\Theta}_{n-1}t})\boldsymbol{\theta}_{0,n-1}, \tag{9}$$

where x_0 is the initial value of X_t and where $\mathbf{x}_n(a) \in \mathbb{R}^n$ is such that $(\mathbf{x}_n(a))_i = a^i$. If X_t has a stationary distribution and the nth steady-state moment $\mathbb{E}[X_\infty^n]$ exists, it is given by

$$\mathbb{E}[X_\infty^n] = \frac{1}{\theta_{n,n}}\left(\boldsymbol{\theta}_n\boldsymbol{\Theta}_{n-1}^{-1}\boldsymbol{\theta}_{0,n-1} - \theta_{0,n}\right), \tag{10}$$

and these steady-state moments satisfy the recursive relationship

$$\mathbb{E}[X_\infty^{n+1}] = -\frac{1}{\theta_{n+1,n+1}}\left(\boldsymbol{\theta}_{n+1}\mathbf{s}_n^X(\infty) + \theta_{0,n+1}\right), \tag{11}$$

where $\mathbf{s}_n^X(\infty) \in \mathbb{R}^n$ is the vector of steady-state moments such that $(\mathbf{s}_n^X(\infty))_i = \mathbb{E}[X_\infty^i]$.

Proof. Using the definition of $\boldsymbol{\Theta}_n$ in Equation (8), Equation (7) gives rise to the system of ODEs given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{s}_n^X(t) = \boldsymbol{\Theta}_n\mathbf{s}_n^X(t) + \boldsymbol{\theta}_{0,n}, \tag{12}$$

where $\mathbf{s}_n^X(t) \in \mathbb{R}^n$ is the vector of transient moments at time $t \geq 0$ such that $(\mathbf{s}_n^X(t))_i = \mathbb{E}[X_t^i]$ for all $1 \leq i \leq n$. We can observe that by definition the matrices $\boldsymbol{\Theta}_n$ form a matryoshkan sequence, and thus by Lemma 1 we achieve the stated transient solution. To prove the steady-state solution, we can first note that if the process has a steady-state distribution then the vector $\mathbf{s}_n^X(\infty) \in \mathbb{R}^n$ defined by $(\mathbf{s}_n^X(\infty))_i = \mathbb{E}[X_\infty^i]$ will satisfy

$$0 = \boldsymbol{\Theta}_n\mathbf{s}_n^X(\infty) + \boldsymbol{\theta}_{0,n}, \tag{13}$$

as this is the equilibrium solution to the differential equation corresponding to each of the moments. This system has a unique solution, since $\boldsymbol{\Theta}_n$ is nonsingular owing to the assumption that the diagonal values are unique and nonzero. Using Proposition 1, we find the nth moment by

$$\mathbb{E}[X_\infty^n] = -\mathbf{v}_n^{\mathrm{T}}\boldsymbol{\Theta}_n^{-1}\boldsymbol{\theta}_{0,n} = \left[\frac{1}{\theta_{n,n}}\boldsymbol{\theta}_n\boldsymbol{\Theta}_{n-1}^{-1} \quad -\frac{1}{\theta_{n,n}}\right]\begin{bmatrix}\boldsymbol{\theta}_{0,n-1}\\\theta_{0,n}\end{bmatrix} = \frac{1}{\theta_{n,n}}\left(\boldsymbol{\theta}_n\boldsymbol{\Theta}_{n-1}^{-1}\boldsymbol{\theta}_{0,n-1} - \theta_{0,n}\right),$$

which completes the proof of Equation (10). To conclude, one can note that each line of the linear system in Equation (13) implies the stated recursion. □

3.2. Application to Hawkes process intensities

For our first example of this method let us turn to our motivating application, the Markovian Hawkes process intensity. Via [33], this process is defined as follows. Let λ_t be a stochastic arrival process intensity such that

$$\lambda_t = \lambda^* + (\lambda_0 - \lambda^*)e^{-\beta t} + \int_0^t \alpha e^{-\beta(t-s)}\mathrm{d}N_s = \lambda^* + (\lambda_0 - \lambda^*)e^{-\beta t} + \sum_{i=1}^{N_t} \alpha e^{-\beta(t-A_i)},$$

where $\{A_i \mid i \in \mathbb{Z}^+\}$ is the sequence of arrival epochs in the point process N_t, with

$$\mathbb{P}(N_{t+s} - N_t = 0 \mid \mathcal{F}_t) = \mathbb{P}(N_{t+s} - N_t = 0 \mid \lambda_t) = e^{-\int_0^s \lambda_{t+u}\mathrm{d}u},$$

where $\mathcal{F}_t$ is the filtration generated by the history of λ_t up to time t. Let us note that the Hawkes process as defined in [33] is not in general a Markov process. If we restrict the excitation kernel to being exponential, then we obtain a Markovian process. In what remains, we will assume that the excitation kernel is exponential. Moreover, we will assume that $\beta > \alpha > 0$, so that the process has a stationary distribution, and we will also let $\lambda^* > 0$ and $\lambda_0 > 0$. Note that the process behaves as follows: at arrivals λ_t increases by α, and in the interims it decays exponentially at rate β towards the baseline level λ^*. Thus, (λ_t, N_t) is referred to as a *self-exciting* point process, as the occurrence of an arrival increases the intensity and thus increases the likelihood that another arrival will occur soon afterwards. Because the intensity λ_t forms a Markov process, we can follow the literature (see, e.g., Equation (7) of [15]) and write its infinitesimal generator for a (sufficiently regular) function $f : \mathbb{R}^+ \to \mathbb{R}$ as follows:

$$\mathcal{L}f(\lambda_t) = \lambda_t(f(\lambda_t + \alpha) - f(\lambda_t)) - \beta(\lambda_t - \lambda^*)f'(\lambda_t),$$

where $f'(x)$ is the first derivative of $f(\cdot)$ evaluated at x. Note that this expression showcases the process dynamics that we have described, as the first term on the right-hand side corresponds to the product of the arrival rate and the change in the process when an arrival occurs, while the second term captures the decay.

To obtain the nth moment we must consider $f(\cdot)$ of the form $f(x) = x^n$. In the simplest case, when $n = 1$, this formula yields an ODE for the mean, which can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_t] = \alpha\mathbb{E}[\lambda_t] - \beta\left(\mathbb{E}[\lambda_t] - \lambda^*\right) = \beta\lambda^* - (\beta - \alpha)\mathbb{E}[\lambda_t].$$

By comparison, for the second moment at $n = 2$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_t^2] = \mathbb{E}\left[\lambda_t\left((\lambda_t + \alpha)^2 - \lambda_t^2\right) - 2\beta\lambda_t(\lambda_t - \lambda^*)\right] = \left(2\beta\lambda^* + \alpha^2\right)\mathbb{E}[\lambda_t] - 2(\beta - \alpha)\mathbb{E}[\lambda_t^2],$$

and we can note that while the ODE for the mean is autonomous, the second moment equation depends on both the mean and the second moment. Thus, to solve for the second moment we must also solve for the mean, leading us to the following system of differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[\lambda_t] \\ \mathbb{E}[\lambda_t^2] \end{bmatrix} = \begin{bmatrix} -(\beta - \alpha) & 0 \\ 2\beta\lambda^* + \alpha^2 & -2(\beta - \alpha) \end{bmatrix} \begin{bmatrix} \mathbb{E}[\lambda_t] \\ \mathbb{E}[\lambda_t^2] \end{bmatrix} + \begin{bmatrix} \beta\lambda^* \\ 0 \end{bmatrix}.$$

Moving on to the third moment, the infinitesimal generator formula yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_t^3] &= \mathbb{E}\left[\lambda_t\big((\lambda_t + \alpha)^3 - \lambda_t^3\big) - 3\beta\lambda_t^2(\lambda_t - \lambda^*)\right] \\ &= \alpha^3 \mathbb{E}[\lambda_t] + 3(\beta\lambda^* + \alpha^2)\mathbb{E}[\lambda_t^2] - 3(\beta - \alpha)\mathbb{E}[\lambda_t^3], \end{aligned}$$

and hence we see that this ODE now depends on all of the first three moments. Thus, to solve for $\mathbb{E}[\lambda_t^3]$ we need to solve the system of ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[\lambda_t] \\ \mathbb{E}[\lambda_t^2] \\ \mathbb{E}[\lambda_t^3] \end{bmatrix} = \begin{bmatrix} -(\beta - \alpha) & 0 & 0 \\ 2\beta\lambda^* + \alpha^2 & -2(\beta - \alpha) & 0 \\ \alpha^3 & 3(\beta\lambda^* + \alpha^2) & -3(\beta - \alpha) \end{bmatrix} \begin{bmatrix} \mathbb{E}[\lambda_t] \\ \mathbb{E}[\lambda_t^2] \\ \mathbb{E}[\lambda_t^3] \end{bmatrix} + \begin{bmatrix} \beta\lambda^* \\ 0 \\ 0 \end{bmatrix},$$

and this now suggests the matryoshkan structure of these process moments: we can note that the system for the second moment is nested within the system for the third moment. That is, the matrix for the three-dimensional system contains the two-dimensional system in its upper left-hand block, just as the vector of the first three moments has the first two moments in its first two coordinates. In general, we can see that the nth moment will satisfy the ODE given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\lambda_t^n] &= \mathbb{E}\left[\lambda_t\big((\lambda_t + \alpha)^n - \lambda_t^n\big) - n\beta\lambda_t^{n-1}(\lambda_t - \lambda^*)\right] \\ &= \sum_{k=1}^{n} \binom{n}{k-1} \alpha^{n-k+1} \mathbb{E}[\lambda_t^k] - n\beta\mathbb{E}[\lambda_t^n] + n\beta\lambda^*\mathbb{E}[\lambda_t^{n-1}], \end{aligned}$$

where we have simplified by use of the binomial theorem. Thus, the system of differential equations needed to solve for the nth moment uses the matrix from the $(n-1)$th system augmented below by the row

$$\begin{bmatrix} \alpha^n & n\alpha^{n-1}, & \binom{n}{2}\alpha^{n-2}, & \ldots, & \binom{n}{n-3}\alpha^3, & \binom{n}{n-2}\alpha^2 + n\beta\lambda^*, & -n(\beta - \alpha) \end{bmatrix}, \quad (14)$$

and buffered on the right by a column of zeros. To collect these coefficients into a coherent structure, let us define the matrix $\mathcal{P}_n(a) \in \mathbb{R}^{n \times n}$ for $a \in \mathbb{R}$ such that

$$(\mathcal{P}_n(a))_{i,j} = \begin{cases} \binom{i}{j-1}a^{i-j+1}, & i \geq j, \\ 0, & i < j. \end{cases} \quad (15)$$

If we momentarily disregard the terms with β in the general augment row in Equation (14), one can observe that the remaining terms in this vector are given by the bottom row of the matrix

$\mathcal{P}_n(\alpha)$. Furthermore, by definition $\{\mathcal{P}_n(a) \mid n \in \mathbb{Z}^+\}$ forms a matryoshkan matrix sequence. We can also note that $\mathcal{P}_n(a)$ can be equivalently defined as

$$\mathcal{P}_n(a) = \sum_{k=1}^{n} a \begin{bmatrix} \mathbf{0}_{n-k \times n-k} & \mathbf{0}_{n-k \times k} \\ \mathbf{0}_{k \times n-k} & \mathbf{L}_k(a) \end{bmatrix},$$

where $\mathbf{L}_k(a) = e^{a\mathbf{diag}(1:k-1,-1)}$ is the kth lower triangular Pascal matrix; i.e. the nonzero terms in $\mathbf{L}_k(1)$ yield the first k rows of Pascal's triangle. Alternatively, $\mathcal{P}_n(a)$ can be found by creating a lower triangular matrix from the strictly lower triangular values in $\mathbf{L}_{n+1}(a)$. For these reasons, we refer to the sequence of $\mathcal{P}_n(a)$ as *matryoshkan Pascal matrices*. For brief overviews and beautiful properties of Pascal matrices, see [8, 9, 49, 21]. As we have seen in the preceding derivation, matryoshkan Pascal matrices arise naturally in using the infinitesimal generator to calculate moments of Markov processes. This follows from the application of the binomial theorem to jump terms. Now, in the case of the Markovian Hawkes process intensity we find closed-form expressions for all transient moments in Corollary 1.

Corollary 1. *Let* λ_t *be the intensity of a Hawkes process with baseline intensity* $\lambda^* > 0$, *intensity jump* $\alpha > 0$, *and decay rate* $\beta > \alpha$. *Then the nth moment of* λ_t *is given by*

$$\mathbb{E}\big[\lambda_t^n\big] = \mathbf{m}_n^\lambda \big(\mathbf{M}_{n-1}^\lambda + n(\beta - \alpha)\mathbf{I}\big)^{-1} \big(e^{\mathbf{M}_{n-1}^\lambda t} - e^{-n(\beta-\alpha)t}\mathbf{I}\big) \left(\mathbf{x}_{n-1}(\lambda_0) - \frac{\beta\lambda^* \mathbf{v}_1}{n(\beta - \alpha)}\right)$$

$$+ \lambda_0^n e^{-n(\beta-\alpha)t} + \mathbb{I}_{\{n=1\}} \frac{\beta\lambda^*}{\beta - \alpha}\big(1 - e^{-(\beta-\alpha)t}\big)$$

$$- \frac{\beta\lambda^*}{n(\beta - \alpha)} \mathbf{m}_n^\lambda \big(\mathbf{M}_{n-1}^\lambda\big)^{-1} \big(\mathbf{I} - e^{\mathbf{M}_{n-1}^\lambda t}\big)\mathbf{v}_1, \tag{16}$$

for all $t \geq 0$ *and* $n \in \mathbb{Z}^+$, *where* $\mathbf{v}_1 \in \mathbb{R}^n$ *is the unit vector in the first coordinate,* $\mathbf{M}_n^\lambda = \beta\lambda^* \mathbf{diag}(2:n,-1) - \beta\mathbf{diag}(1:n) + \mathcal{P}_n(\alpha)$, $\mathbf{m}_n^\lambda = \big[\big(\mathbf{M}_n^\lambda\big)_{n,1}, \ldots, \big(\mathbf{M}_n^\lambda\big)_{n,n-1}\big]$ *is given by*

$$\big(\mathbf{m}_n^\lambda\big)_j = \begin{cases} \binom{n}{j-1}\alpha^{n-j+1} & \text{if } j < n-1, \\ \binom{n}{n-2}\alpha^2 + n\beta\lambda^* & \text{if } j = n-1, \end{cases}$$

and $\mathbf{x}_n(a) \in \mathbb{R}^n$ *is such that* $(\mathbf{x}_n(a))_i = a^i$. *In steady state, the nth moment of* λ_t *is given by*

$$\lim_{t\to\infty} \mathbb{E}\big[\lambda_t^n\big] = -\frac{\beta\lambda}{n(\beta - \alpha)} \mathbf{m}_n^\lambda \big(\mathbf{M}_{n-1}^\lambda\big)^{-1}\mathbf{v}_1, \tag{17}$$

for $n \geq 2$, *with* $\lim_{t\to\infty} \mathbb{E}[\lambda_t] = \frac{\beta\lambda^*}{\beta - \alpha}$. *Moreover, the* $(n+1)$*th steady-state moment of the Hawkes process intensity is given by the recursion*

$$\lim_{t\to\infty} \mathbb{E}\big[\lambda_t^{n+1}\big] = \frac{1}{(n+1)(\beta - \alpha)} \mathbf{m}_{n+1}^\lambda \mathbf{s}_n^\lambda, \tag{18}$$

for all $n \in \mathbb{Z}^+$, *where* $\mathbf{s}_n^\lambda \in \mathbb{R}^n$ *is the vector of steady-state moments defined so that* $(\mathbf{s}_n^\lambda)_i = \lim_{t\to\infty} \mathbb{E}[\lambda_t^i]$ *for* $1 \leq i \leq n$.

By use of the binomial theorem, we can observe that if $\mathbb{E}\left[J_1^n\right] < \infty$ then the nth moment of the shot noise process satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\psi_t^n\right] = \mathbb{E}\left[\lambda\left((\psi_t + J_1)^n - \psi_t^n\right) - n\beta\psi_t^n\right] = \sum_{k=0}^{n-1}\binom{n}{k}\mathbb{E}\left[J_1^{n-k}\right]\mathbb{E}\left[\psi_t^k\right] - n\beta\mathbb{E}\left[\psi_t^n\right],$$

which means that the n-dimensional system is equal to the preceding one augmented below by the row vector

$$\left[n\lambda\mathbb{E}[J_1^{n-1}], \quad \binom{n}{2}\lambda\mathbb{E}[J_1^{n-2}], \quad \binom{n}{3}\lambda\mathbb{E}[J_1^{n-3}], \quad \ldots, \quad \binom{n}{n-2}\lambda\mathbb{E}[J_1^2], \quad n\lambda\mathbb{E}[J_1] \quad -n\beta\right],$$

and to the right by zeros. Bringing this together, this now leads us to Corollary 2.

Corollary 2. *Let ψ_t be the intensity of a shot noise process with epochs given by a Poisson process with rate $\lambda > 0$, jump sizes drawn from the sequence of independent and identically distributed random variables $\{J_i \mid i \in \mathbb{Z}^+\}$, and exponential decay at rate $\beta > 0$. If $\mathbb{E}\left[J_1^n\right] < \infty$, the nth moment of ψ_t is given by*

$$\mathbb{E}\left[\psi_t^n\right] = \mathbf{m}_n^{\psi}\left(\mathbf{M}_{n-1}^{\psi} + n\beta\mathbf{I}\right)^{-1}\left(e^{\mathbf{M}_{n-1}^{\psi}t} - e^{-n\beta t}\mathbf{I}\right)\left(\mathbf{x}_{n-1}(\psi_0) - \frac{\lambda\mathbf{j}_{n-1}}{n\beta}\right) + \psi_0^n e^{-n\beta t}$$

$$+ \frac{\lambda\mathbb{E}\left[J_1^n\right]}{n\beta}\left(1 - e^{-n\beta t}\right) - \frac{\lambda\mathbf{m}_n^{\psi}}{n\beta}\left(\mathbf{M}_{n-1}^{\psi}\right)^{-1}\left(\mathbf{I} - e^{\mathbf{M}_{n-1}^{\psi}t}\right)\mathbf{j}_{n-1}, \qquad (19)$$

for all $t \geq 0$ and $n \in \mathbb{Z}^+$, where $\mathbf{j}_n \in \mathbb{R}^n$ is such that $(\mathbf{j}_n)_i = \mathbb{E}\left[J_1^i\right]$; $\mathbf{M}_n^{\psi} \in \mathbb{R}^{n\times n}$ is recursively defined by

$$\mathbf{M}_n^{\psi} = \begin{bmatrix} \mathbf{M}_{n-1}^{\psi} & \mathbf{0}_{n-1\times 1} \\ \mathbf{m}_n^{\psi} & -n\beta \end{bmatrix},$$

with the row vector $\mathbf{m}_n^{\psi} \in \mathbb{R}^{n-1}$ defined so that $\left(\mathbf{m}_n^{\psi}\right)_i = \binom{n}{i}\lambda\mathbb{E}\left[J_1^{n-i}\right]$, and with $\mathbf{M}_1^{\psi} = -\beta$; and $\mathbf{x}_n(a) \in \mathbb{R}^n$ is such that $(\mathbf{x}_n(a))_i = a^i$. In steady state, the $(n+1)$th moment of the shot noise process is given by

$$\lim_{t\to\infty}\mathbb{E}\left[\psi_t^n\right] = \frac{\lambda}{n\beta}\left(\mathbb{E}\left[J_1^n\right] - \mathbf{m}_n^{\psi}\left(\mathbf{M}_{n-1}^{\psi}\right)^{-1}\mathbf{j}_{n-1}\right) \qquad (20)$$

for $n \geq 2$, where $\lim_{t\to\infty}\mathbb{E}[\psi_t] = \frac{\lambda}{\beta}\mathbb{E}[J_1]$. Moreover, if $\mathbb{E}\left[J_1^{n+1}\right] < \infty$, the $(n+1)$th moment of the shot noise process is given by the recursion

$$\lim_{t\to\infty}\mathbb{E}\left[\psi_t^{n+1}\right] = \frac{1}{(n+1)\beta}\left(\mathbf{m}_{n+1}^{\psi}\mathbf{s}_n^{\psi} + \mathbb{E}\left[J_1^{n+1}\right]\right), \qquad (21)$$

for all $n \in \mathbb{Z}^+$, where $\mathbf{s}_n^{\psi} \in \mathbb{R}^n$ is the vector of steady-state moments defined so that $\left(\mathbf{s}_n^{\psi}\right)_i = \lim_{t\to\infty}\mathbb{E}\left[\psi_t^i\right]$ for $1 \leq i \leq n$.

3.4. Application to Itô diffusions

For our third example, we consider an Itô diffusion; see e.g. [45] for an overview. Let S_t be given by the stochastic differential equation

$$\mathrm{d}S_t = g(S_t)\mathrm{d}t + h(S_t)\mathrm{d}B_t,$$

where B_t is a Brownian motion and $g(\cdot)$ and $h(\cdot)$ are real-valued functions. By Theorem 7.9 of [45], the infinitesimal generator for this process is given by

$$\mathcal{L}f(S_t) = g(S_t)f'(S_t) + \frac{h(S_t)^2}{2}f''(S_t),$$

where $f''(x)$ is the second derivative of $f(\cdot)$ evaluated at x. Because we will be considering functions of the form $f(x) = x^n$ for $n \in \mathbb{Z}^+$, we will now specify the forms of $g(\cdot)$ and $h(\cdot)$ to be $g(x) = \mu + \theta x$ for some $\mu \in \mathbb{R}$ and $\theta \in \mathbb{R}$ and $h(x) = \sigma x^{\gamma/2}$ for some $\sigma \in \mathbb{R}$ and $\gamma \in \{0, 1, 2\}$. One can note that this encompasses a myriad of relevant stochastic processes including many that are popular in the financial models literature, such as Ornstein–Uhlenbeck processes, geometric Brownian motion, and Cox–Ingersoll–Ross processes. In this case, the infinitesimal generator becomes

$$\mathcal{L}f(S_t) = (\mu + \theta S_t)f'(S_t) + \frac{\sigma^2 S_t^\gamma}{2}f''(S_t),$$

meaning that we can express the ODE for the mean as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t] = \mu + \theta\mathbb{E}[S_t],$$

and similarly the second moment will be given by the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t^2] = \mathbb{E}[2(\mu + \theta S_t)S_t + \sigma^2 S_t^\gamma] = 2\mu\mathbb{E}[S_t] + 2\theta\mathbb{E}[S_t^2] + \sigma^2\mathbb{E}[S_t^\gamma].$$

For the sake of example, we now let $\gamma = 1$, as is the case in the Cox–Ingersoll–Ross process. Then the first two transient moments of S_t will be given by the solution to the system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[S_t] \\ \mathbb{E}[S_t^2] \end{bmatrix} = \begin{bmatrix} \theta & 0 \\ 2\mu + \sigma^2 & 2\theta \end{bmatrix}\begin{bmatrix} \mathbb{E}[S_t] \\ \mathbb{E}[S_t^2] \end{bmatrix} + \begin{bmatrix} \mu \\ 0 \end{bmatrix}.$$

By observing that the third moment differential equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t^3] = \mathbb{E}[3(\mu + \theta S_t)S_t^2 + 3\sigma^2 S_t^{\gamma+1}] = 3\mu\mathbb{E}[S_t^2] + 3\theta\mathbb{E}[S_t^3] + 3\sigma^2\mathbb{E}[S_t^{\gamma+1}],$$

we can note that the third moment system for $\gamma = 1$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[S_t] \\ \mathbb{E}[S_t^2] \\ \mathbb{E}[S_t^3] \end{bmatrix} = \begin{bmatrix} \theta & 0 & 0 \\ 2\mu + \sigma^2 & 2\theta & 0 \\ 0 & 3\mu + 3\sigma^2 & 3\theta \end{bmatrix}\begin{bmatrix} \mathbb{E}[S_t] \\ \mathbb{E}[S_t^2] \\ \mathbb{E}[S_t^3] \end{bmatrix} + \begin{bmatrix} \mu \\ 0 \\ 0 \end{bmatrix},$$

and this showcases the matryoshkan nesting structure, as the second moment system is contained within the third. Because the general nth moment for $n \geq 2$ has differential equation given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t^n] = \mathbb{E}\left[n(\mu + \theta S_t)S_t^{n-1} + \frac{n(n-1)\sigma^2}{2}S_t^{n+\gamma-2}\right]$$

$$= n\mu\mathbb{E}[S_t^{n-1}] + n\theta\mathbb{E}[S_t^n] + \frac{n(n-1)\sigma^2}{2}\mathbb{E}[S_t^{n+\gamma-2}],$$

we can see that the $(n-1)$th system can be augmented below by the row vector for $\gamma = 1$,

$$\begin{bmatrix} 0, & 0, & \ldots, & 0, & n\mu + \frac{n(n-1)\sigma^2}{2}, & n\theta \end{bmatrix},$$

and to the right by zeros. Through this observation, we can now give the moments of Itô diffusions in Corollary 3.

Corollary 3. *Let S_t be an Itô diffusion that satisfies the stochastic differential equation*

$$dS_t = (\mu + \theta S_t)dt + \sigma S_t^{\gamma/2}dB_t, \tag{22}$$

where B_t is a Brownian motion, $\mu, \theta, \sigma \in \mathbb{R}$, and $\gamma \in \{0, 1, 2\}$. Then the nth moment of S_t is given by

$$
\begin{aligned}
\mathbb{E}\left[S_t^n\right] &= \mathbf{m}_n^S \left(\mathbf{M}_{n-1}^S - \chi_n \mathbf{I}\right)^{-1} \left(e^{\mathbf{M}_{n-1}^S t} - e^{\chi_n t}\mathbf{I}\right)\left(\mathbf{x}_{n-1}(S_0) + \frac{\mu \mathbf{v}_1 + \sigma^2 \mathbb{I}_{\{\gamma=0\}}\mathbf{v}_2}{\chi_n}\right) \\
&\quad + S_0^n e^{\chi_n t} - \left(\mu \mathbb{I}_{\{n=1\}} + \sigma^2 \mathbb{I}_{\{\gamma=0, n=2\}}\right)\frac{1 - e^{\chi_n t}}{\chi_n} \\
&\quad + \frac{\mathbf{m}_n^S}{\chi_n}\left(\mathbf{M}_{n-1}^S\right)^{-1}\left(\mathbf{I} - e^{\mathbf{M}_{n-1}^S t}\right)\left(\mu \mathbf{v}_1 + \sigma^2 \mathbb{I}_{\{\gamma=0\}}\mathbf{v}_2\right),
\end{aligned} \tag{23}
$$

for all $t \geq 0$ and $n \in \mathbb{Z}^+$, where $\chi_n = n\theta + \frac{n}{2}(n-1)\sigma^2 \mathbb{I}_{\{\gamma=2\}}$;

$$\mathbf{M}_n^S = \theta \mathbf{diag}(1:n) + \mu \mathbf{diag}(2:n, -1) + \frac{\sigma^2}{2}\mathbf{diag}\left(\mathbf{d}_{n+\gamma-2}^{2-\gamma}, \gamma - 2\right)$$

for $\mathbf{d}_k^j \in \mathbb{R}^k$ such that $\left(\mathbf{d}_k^j\right)_i = (j+i)(j+i-1)$;

$$\mathbf{m}_n^S = \left[\left(\mathbf{M}_n^S\right)_{n,1}, \ \ldots, \ \left(\mathbf{M}_n^S\right)_{n,n-1}\right]$$

is such that

$$
\left(\mathbf{m}_n^S\right)_j = \begin{cases} n\mu + \frac{n(n-1)\sigma^2}{2}\mathbb{I}_{\{\gamma=1\}}, & j = n-1, \\ \frac{n(n-1)\sigma^2}{2}\mathbb{I}_{\{\gamma=0\}}, & j = n-2, \\ 0, & 1 \leq j < n-2; \end{cases}
$$

and $\mathbf{x}_n(a) \in \mathbb{R}^n$ is such that $(\mathbf{x}_n(a))_i = a^i$. If $\theta < 0$ and $\gamma \in \{0, 1\}$, then the nth steady-state moment of S_t is given by

$$\lim_{t \to \infty} \mathbb{E}\left[S_t^n\right] = \frac{\mu}{\chi_n}\mathbf{m}_n^S\left(\mathbf{M}_{n-1}^S\right)^{-1}\mathbf{v}_1, \tag{24}$$

for $n \geq 2$, with $\lim_{t \to \infty} \mathbb{E}[S_t] = -\frac{\mu}{\theta}$. Moreover, the $(n+1)$th steady-state moment of S_t is given by the recursion

$$\lim_{t \to \infty} \mathbb{E}\left[S_t^{n+1}\right] = -\frac{1}{\chi_n}\mathbf{m}_{n+1}^S\mathbf{s}_n^S, \tag{25}$$

for all $n \in \mathbb{Z}^+$, where $\mathbf{s}_n^S \in \mathbb{R}^n$ is the vector of steady-state moments defined so that $\left(\mathbf{s}_n^S\right)_i = \lim_{t \to \infty} \mathbb{E}\left[S_t^i\right]$ for $1 \leq i \leq n$.

As a consequence of these expressions we can also gain insight on the moments of an Itô diffusion in the case of non-integer $\gamma \in [0, 2]$, as is used in volatility models such as the CEV model and the SABR model; see e.g. [34]. This can be achieved through bounding the differential equations, as the nth moment of such a diffusion is again given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t^n] = n\mu\mathbb{E}[S_t^{n-1}] + n\theta\mathbb{E}[S_t^n] + \frac{n(n-1)\sigma^2}{2}\mathbb{E}\left[S_t^{n+\gamma-2}\right],$$

the rightmost term of which can be bounded above and below by

$$\mathbb{E}\left[S_t^{n+\lfloor\gamma\rfloor-2}\right] \le \mathbb{E}\left[S_t^{n+\gamma-2}\right] \le \mathbb{E}\left[S_t^{n+\lceil\gamma\rceil-2}\right],$$

and the differential equations given by substituting these bounded terms form a closed system solvable by Corollary 3. Assuming the true differential equation and the upper and lower bounds all share an initial value, the solution to the bounded equations bounds the solution to the true moment equation; see [32].

3.5. Application to growth–collapse processes

For a fourth example, we consider growth–collapse processes with Poisson-driven shocks. These processes have been studied in a variety of contexts; see e.g. [7, 38, 39, 6]. More recently, these processes and their related extensions have seen renewed interest in connection with the study of the crypto-currency Bitcoin; see for example [28, 40, 37, 27]. While growth–collapse processes can be defined in many different ways, for this example we use a definition in the style of Section 4 from [7]. We let Y_t be the state of the growth–collapse model and let $\{U_i \mid i \in \mathbb{Z}^+\}$ be a sequence of independent Uni(0, 1) random variables that are also independent from the state and history of the growth–collapse process. From Proposition 1 of [7], the infinitesimal generator of Y_t is given by

$$\mathcal{L}f(Y_t) = \lambda f'(Y_t) + \mu \int_0^1 (f(uY_t) - f(Y_t))\,\mathrm{d}u.$$

Thus, Y_t experiences linear growth at rate $\lambda > 0$ throughout time, but it also collapses at epochs given by a Poisson process with rate $\mu > 0$. At the ith collapse epoch the process falls to a fraction of its current level; specifically it jumps down to $U_i Y_t$. Using the infinitesimal generator, we can see that the mean of this growth–collapse process satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Y_t] = \lambda + \mu(\mathbb{E}[U_1 Y_t] - \mathbb{E}[Y_t]) = \lambda - \frac{\mu}{2}\mathbb{E}[Y_t],$$

and its second moment satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[Y_t^2\right] = 2\lambda\mathbb{E}[Y_t] + \mu\left(\mathbb{E}\left[U_1^2 Y_t^2\right] - \mathbb{E}\left[Y_t^2\right]\right) = 2\lambda\mathbb{E}[Y_t] - \frac{2\mu}{3}\mathbb{E}\left[Y_t^2\right].$$

Therefore, we can write the linear system of differential equations for the second moment of this growth–collapse process as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[Y_t] \\ \mathbb{E}\left[Y_t^2\right] \end{bmatrix} = \begin{bmatrix} -\frac{\mu}{2} & 0 \\ 2\lambda & -\frac{2\mu}{3} \end{bmatrix}\begin{bmatrix} \mathbb{E}[Y_t] \\ \mathbb{E}\left[Y_t^2\right] \end{bmatrix} + \begin{bmatrix} \lambda \\ 0 \end{bmatrix}.$$

Moving to the third moment, via the infinitesimal generator we write its differential equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[Y_t^3\right] = 3\lambda\mathbb{E}\left[Y_t^2\right] + \mu\left(\mathbb{E}\left[U_1^3 Y_t^3\right] - \mathbb{E}\left[Y_t^3\right]\right) = 3\lambda\mathbb{E}\left[Y_t^2\right] - \frac{3\mu}{4}\mathbb{E}\left[Y_t^3\right],$$

which shows that the system of differential equations for the third moment is

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[Y_t] \\ \mathbb{E}[Y_t^2] \\ \mathbb{E}[Y_t^3] \end{bmatrix} = \begin{bmatrix} -\frac{\mu}{2} & 0 & 0 \\ 2\lambda & -\frac{2\mu}{3} & 0 \\ 0 & 3\lambda & -\frac{3\mu}{4} \end{bmatrix} \begin{bmatrix} \mathbb{E}[Y_t] \\ \mathbb{E}[Y_t^2] \\ \mathbb{E}[Y_t^3] \end{bmatrix} + \begin{bmatrix} \lambda \\ 0 \\ 0 \end{bmatrix};$$

this obviously encapsulates the system for the first two moments. We can note that the general nth moment will satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Y_t^n] = n\lambda\mathbb{E}[Y_t^{n-1}] + \mu\left(\mathbb{E}[U_1^n Y_t^n] - \mathbb{E}[Y_t^n]\right) = n\lambda\mathbb{E}[Y_t^{n-1}] - \frac{n\mu}{n+1}\mathbb{E}[Y_t^n],$$

and thus the system for the nth moment is given by appending the row vector

$$\begin{bmatrix} 0, & , 0, & \ldots, & 0, & n\lambda, & -\frac{n\mu}{n+1} \end{bmatrix}$$

below the matrix from the $(n-1)$th system augmented by zeros on the right. Following this derivation, we reach the general expressions for the moments in Corollary 4. Furthermore, we can note that because of the relative simplicity of this particular structure, we are able to solve the recursion for the steady-state moments and give these terms explicitly.

Corollary 4. *Let Y_t be a growth–collapse process with growth rate $\lambda > 0$ and uniformly sized collapses occurring according to a Poisson process with rate $\mu > 0$. Then the nth moment of Y_t is given by*

$$\begin{aligned} \mathbb{E}[Y_t^n] = {}&n\lambda\mathbf{v}_{n-1}^{\mathrm{T}}\left(\mathbf{M}_{n-1}^Y + \frac{n\mu}{n+1}\mathbf{I}\right)^{-1}\left(e^{\mathbf{M}_{n-1}^Y t} - e^{-\frac{n\mu t}{n+1}}\mathbf{I}\right)\left(\mathbf{x}_{n-1}(y_0) - \frac{(n+1)\lambda\mathbf{v}_1}{n\mu}\right) \\ &+ y_0^n e^{-\frac{n\mu t}{n+1}} + \frac{(n+1)\lambda\mathbb{I}_{\{n=1\}}}{n\mu}\left(1 - e^{-\frac{n\mu t}{n+1}}\right) \\ &- \frac{(n+1)\lambda^2}{\mu}\mathbf{v}_{n-1}^{\mathrm{T}}\left(\mathbf{M}_{n-1}^Y\right)^{-1}\left(\mathbf{I} - e^{\mathbf{M}_{n-1}^Y t}\right)\mathbf{v}_1, \end{aligned} \tag{26}$$

where y_0 is the initial value of Y_t, $\mathbf{x}_n(a) \in \mathbb{R}^n$ is such that $(\mathbf{x}_n(a))_i = a^i$,

$$\mathbf{M}_n^Y = \lambda\mathbf{diag}(2:n, -1) - \mu\mathbf{diag}\left(\frac{1}{2}:\frac{n}{n+1}\right),$$

and

$$\mathbf{m}_n^Y = \left[\left(\mathbf{M}_n^Y\right)_{n,1}, \ldots, \left(\mathbf{M}_n^Y\right)_{n,n-1}\right]$$

is such that $\mathbf{m}_n^Y = n\lambda\mathbf{v}_{n-1}^{\mathrm{T}}$. Moreover, the nth steady-state moment of Y_t is given by

$$\lim_{t\to\infty}\mathbb{E}[Y_t^n] = (n+1)!\left(\frac{\lambda}{\mu}\right)^n, \tag{27}$$

for $n \in \mathbb{Z}^+$.

3.6. Application to ephemerally self-exciting processes

As a final detailed example of the applicability of matryoshkan matrices, we now consider a stochastic process we have analyzed in [19]. This process is a linear birth–death–immigration process, which has been of interest in classical teletraffic theory; see for example [46], [35], and Section 8.4 of [36]. By comparison, our interests in [19] are in alternate representations of self-exciting processes and alternate constructions of the original self-exciting process, the Hawkes process. Here, the occurrence of an arrival increases the arrival rate by an amount $\alpha > 0$, as in the Hawkes process, and this increase expires after an exponentially distributed duration with some rate $\mu > \alpha$. Accordingly, this process is an *ephemerally self-exciting* process. Given a baseline intensity $\nu^* > 0$, let Q_t be such that new arrivals occur at rate $\nu^* + \alpha Q_t$ and then the overall rate until the next excitement expiration is μQ_t. One can then think of Q_t as the number of entities still causing active excitement at time $t \geq 0$. We will refer to Q_t as the number in system for this ephemerally self-exciting process. The infinitesimal generator for a function $f : \mathbb{N} \to \mathbb{R}$ is thus

$$\mathcal{L}f(Q_t) = \left(\nu^* + \alpha Q_t\right)\left(f(Q_t + 1) - f(Q_t)\right) + \mu Q_t(f(Q_t - 1) - f(Q_t)),$$

which again captures the dynamics of the process, as the first term on the right-hand side is the product of the up-jump rate and the change in function value upon an increase in the process, while the second term is the product of the down-jump rate and the corresponding process decrease. This now yields an ODE for the mean given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Q_t] = \nu^* + \alpha\mathbb{E}[Q_t] - \mu\mathbb{E}[Q_t] = \nu^* - (\mu - \alpha)\mathbb{E}[Q_t],$$

while the second moment will satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Q_t^2] = \mathbb{E}\left[\left(\nu^* + \alpha Q_t\right)\left((Q_t + 1)^2 - Q_t^2\right) + \mu Q_t\left((Q_t - 1)^2 - Q_t^2\right)\right]$$
$$= \left(2\nu^* + \mu + \alpha\right)\mathbb{E}[Q_t] + \nu^* - 2(\mu - \alpha)\mathbb{E}[Q_t^2].$$

Thus, the first two moments are given by the solution to the linear system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[Q_t] \\ \mathbb{E}[Q_t^2] \end{bmatrix} = \begin{bmatrix} -(\mu - \alpha) & 0 \\ 2\nu^* + \mu + \alpha & -2(\mu - \alpha) \end{bmatrix}\begin{bmatrix} \mathbb{E}[Q_t] \\ \mathbb{E}[Q_t^2] \end{bmatrix} + \begin{bmatrix} \nu^* \\ \nu^* \end{bmatrix},$$

and by observing that the third moment differential equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Q_t^3] = \mathbb{E}\left[\left(\nu^* + \alpha Q_t\right)\left((Q_t + 1)^3 - Q_t^3\right) + \mu Q_t((Q_t - 1)^3 - Q_t^3)\right]$$
$$= \left(3\nu^* + 3\alpha + 3\mu\right)\mathbb{E}[Q_t^2] + \left(3\nu^* + \alpha - \mu\right)\mathbb{E}[Q_t] + \nu^* - 3(\mu - \alpha)\mathbb{E}[Q_t^3],$$

we can see that the third moment system does indeed encapsulate the second moment system:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[Q_t] \\ \mathbb{E}[Q_t^2] \\ \mathbb{E}[Q_t^3] \end{bmatrix} = \begin{bmatrix} -(\mu - \alpha) & 0 & 0 \\ 2\nu^* + \mu + \alpha & -2(\mu - \alpha) & 0 \\ 3\nu^* + \alpha - \mu & 3\nu^* + 3\alpha + 3\mu & -3(\mu - \alpha) \end{bmatrix}\begin{bmatrix} \mathbb{E}[Q_t] \\ \mathbb{E}[Q_t^2] \\ \mathbb{E}[Q_t^3] \end{bmatrix} + \begin{bmatrix} \nu^* \\ \nu^* \\ \nu^* \end{bmatrix}.$$

In general, the *n*th moment is given by the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Q_t^n] = \mathbb{E}\big[(\nu^* + \alpha Q_t)\big((Q_t+1)^n - Q_t^n\big) + \mu Q_t\big((Q_t-1)^n - Q_t^n\big)\big]$$

$$= \nu^* + \nu^* \sum_{k=1}^{n-1}\binom{n}{k}\mathbb{E}\big[Q_t^k\big] + \alpha \sum_{k=1}^{n}\binom{n}{k-1}\mathbb{E}\big[Q_t^k\big]$$

$$+ \mu \sum_{k=1}^{n}\binom{n}{k-1}\mathbb{E}\big[Q_t^k\big](-1)^{n-k-1},$$

which means that the *n*th system is given by augmenting the previous system below by

$$\left[n\nu^* + \alpha + \mu(-1)^n, \ \ \binom{n}{2}\nu^* + n\alpha + n\mu(-1)^{n-1}, \ \ \dots, \ \ n\nu^* + \binom{n}{2}\alpha + \binom{n}{2}\mu, \ \ -n(\mu-\alpha)\right],$$

and to the right by zeros. By comparing this row vector to the definition of the matryoshkan Pascal matrices in Equation (15), we arrive at explicit forms for the moments of this process, shown now in Corollary 5.

Corollary 5. *Let Q_t be the number in system for an ephemerally self-exciting process with baseline intensity $\nu^* > 0$, intensity jump $\alpha > 0$, and duration rate $\mu > \alpha$. Then the nth moment of Q_t is given by*

$$\mathbb{E}[Q_t^n] = \mathbf{m}_n^Q\Big(\mathbf{M}_{n-1}^Q + n(\mu-\alpha)\mathbf{I}\Big)^{-1}\Big(e^{\mathbf{M}_{n-1}^Q t} - e^{-n(\mu-\alpha)t}\mathbf{I}\Big)\left(\mathbf{x}_{n-1}(Q_0) - \frac{\nu^*\mathbf{v}}{n(\mu-\alpha)}\right)$$

$$+ Q_0^n e^{-n(\mu-\alpha)t} + \frac{\nu^*}{n(\mu-\alpha)}\Big(1 - e^{-n(\mu-\alpha)t}\Big)$$

$$- \frac{\nu^*\mathbf{m}_n^Q}{n(\mu-\alpha)}\Big(\mathbf{M}_{n-1}^Q\Big)^{-1}\Big(\mathbf{I} - e^{\mathbf{M}_{n-1}^Q t}\Big)\mathbf{v}, \tag{28}$$

for all $t \geq 0$ and $n \in \mathbb{Z}^+$, where $\mathbf{M}_n^Q = \nu^*\mathcal{P}_n(1)\mathbf{diag}(\mathbf{v}, -1) + \alpha\mathcal{P}_n(1) + \mu\mathcal{P}_n(-1)$,

$$\mathbf{m}_n^Q = \left[\Big(\mathbf{M}_n^Q\Big)_{n,1}, \ \dots, \ \Big(\mathbf{M}_n^Q\Big)_{n,n-1}\right]$$

is such that

$$\Big(\mathbf{m}_n^Q\Big)_j = \binom{n}{j}\nu^* + \binom{n}{j-1}\alpha + \binom{n}{j-1}\mu(-1)^{n-j-1},$$

and $\mathbf{x}_n(a) \in \mathbb{R}^n$ is such that $(\mathbf{x}_n(a))_i = a^i$. In steady state, the nth moment of Q_t is given by

$$\lim_{t\to\infty}\mathbb{E}[Q_t^n] = \frac{\nu^*}{\mu-\alpha}\left(1 - \mathbf{m}_n^Q\Big(\mathbf{M}_{n-1}^Q\Big)^{-1}\mathbf{v}\right), \tag{29}$$

for $n \geq 2$, with $\lim_{t\to\infty}\mathbb{E}[Q_t] = \frac{\nu^*}{\mu-\alpha}$. Moreover, the $(n+1)$th steady-state moment of the ephemerally self-exciting process is given by the recursion

$$\lim_{t\to\infty}\mathbb{E}[Q_t^{n+1}] = \frac{1}{(n+1)(\mu-\alpha)}\Big(\mathbf{m}_{n+1}^Q\mathbf{s}_n^Q + \nu^*\Big), \tag{30}$$

for all $n \in \mathbb{Z}^+$, where $\mathbf{s}_n^Q \in \mathbb{R}^n$ is the vector of steady-state moments defined so that $\big(\mathbf{s}_n^Q\big)_i = \lim_{t\to\infty}\mathbb{E}[Q_t^i]$ for $1 \leq i \leq n$.

Let us note that the ephemerally self-exciting process is known to be negative binomially distributed in steady state [19]. Hence, while the vector and recursive computations in Equations (29) and (30), respectively, may be useful for higher moments, for lower moments there may be closed-form expressions available.

3.7. Additional applications by combination and permutation

While the preceding examples are the the only detailed examples we include in this paper, we can note that these matryoshkan matrix methods can be applied to many other settings. In fact, one can observe that these example derivations can be applied directly to processes that feature a combination of their structures, such as the dynamic contagion process introduced in [17]. The dynamic contagion process is a point process that is both self-excited and externally excited, meaning that its intensity experiences jumps driven both by its own activity and by the activity of an exogenous Poisson process. In this way, the process combines the behavior of the Hawkes and shot noise processes. Hence, its infinitesimal generator can be written using a combination of expressions used in Subsections 3.2 and 3.3, implying that all moments of the process can be calculated through this methodology. Similarly, these methods can also be readily applied to processes that combine dynamics from Hawkes processes and from Itô diffusions, such as affine point processes. These processes, studied in e.g. [22, 48, 30], feature both self-excitement and diffusion behavior and thus have an infinitesimal generator that can be expressed using terms from the generators for Hawkes and Itô processes. One could also study the combination of externally driven jumps and diffusive behavior, such as in affine jump diffusions; see e.g. [20]. Of course, one can also consider permutations of the model features seen in our examples, such as trading fixed-size jumps for random ones to form marked Hawkes processes or changing to randomly sized batches of arrivals in the ephemerally self-exciting process. In general, the key requirement from the assumptions in Theorem 1 is the closure of the system of moment differential equations specified in Equation (7). This is equivalent to having the infinitesimal generator of any polynomial being a polynomial of order no more than the original, which again connects this work to the literature on polynomial processes. In fact, the notion of closure of polynomial processes under combinations of infinitesimal generators has been studied by [12]. In summary, infinitesimal generators of the form

$$\mathcal{L}f(X_t) = \underbrace{(\alpha_0 + \alpha_1 X_t) \int_0^\infty (f(X_t + a) - f(X_t)) \mathrm{d}F_A(a)}_{\text{Up-jumps}} + \underbrace{(\alpha_2 + \alpha_3 X_t) f'(X_t)}_{\text{Drift, decay, or growth}}$$

$$+ \underbrace{(\alpha_4 + \alpha_5 X_t) \int_0^\infty (f(X_t - b) - f(X_t)) \mathrm{d}F_B(b)}_{\text{Down-jumps}} + \underbrace{(\alpha_6 + \alpha_7 X_t + \alpha_8 X_t^2) f''(X_t)}_{\text{Diffusion}}$$

$$+ \underbrace{\alpha_9 \int_0^\infty (f(cX_t) - f(X_t)) \mathrm{d}F_C(c)}_{\text{Expansion or collapse}}$$

can be handled by this methodology, where $\alpha_j \in \mathbb{R}$ for all j and where the sequences $\{A_i\}$, $\{B_i\}$, and $\{C_i\}$ are of mutually independent random variables with respective cumulative distribution functions $F_A(\cdot)$, $F_B(\cdot)$, and $F_C(\cdot)$. Finally we note that this example generator need not be exclusive, as it is possible that other dynamics may also meet the closure requirements in Equation (7).

4. Conclusion

In this work, we have defined a novel sequence of matrices called matryoshkan matrices, which stack like their Russian nesting doll namesakes. In doing so, we have found an efficient manner of calculating the moments of a large class of Markov processes that satisfy a closure condition for the time derivatives of their transient moments, namely the one-dimensional polynomial processes. Furthermore, this has yielded closed-form expressions for the transient and steady-state moments of these process. Notably, this includes the intensity of the Hawkes process, for which finding an expression for the *n*th moment was previously an open problem. Other examples we have discussed include Itô diffusions from the mathematical finance literature and shot noise processes from the physics literature, which showcases the breadth of this methodology.

We can note that there are many applications of this methodology that we have not explored in this paper; these present opportunities for future work. For example, the vector form of the moments arising from this matrix-based method naturally lends itself to use in the method of moments. Thus, matryoshkan matrices have the potential to greatly simplify estimation for the myriad of Markov processes to which they apply. Additionally, this vector of solutions may also be of use in providing computationally tractable approximations of moment generating functions. That is, by a Taylor expansion one can approximate a moment generating function by a weighted sum of its moments. Because this paper's matryoshkan matrix methods enable efficient calculation of higher-order moments, this enables higher-order approximations of the moment generating function.

As another important direction of future work, we are also interested in extending these techniques to multivariate Markov processes. This is of practical relevance in many of the settings we have described, such as point processes driven by the Hawkes or shot noise process intensities. The challenge in this case arises in the fact that a moment's differential equation now depends on the lower product moments rather than just the lower moments, so the nesting structure is not as neatly organized. Nevertheless, addressing this generalization is an extension worth pursuing, as this would render these techniques even more widely applicable.

Appendix A. Proof of Proposition 1

Proof. For the sake of clarity and ease of reference, we will enumerate the proofs of the four statements.

(i) Suppose $\mathbf{X}_n$ and $\mathbf{Y}_n$ are both matryoshkan matrices. Then, by Equation (1), we have that

$$\mathbf{X}_n + \mathbf{Y}_n = \begin{bmatrix} \mathbf{X}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{x}_n & x_{n,n} \end{bmatrix} + \begin{bmatrix} \mathbf{Y}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{y}_n & y_{n,n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{n-1} + \mathbf{Y}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{x}_n + \mathbf{y}_n & x_{n,n} + y_{n,n} \end{bmatrix},$$

and

$$\mathbf{X}_n \mathbf{Y}_n = \begin{bmatrix} \mathbf{X}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{x}_n & x_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{y}_n & y_{n,n} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{n-1}\mathbf{Y}_{n-1} & \mathbf{0}_{n-1\times 1} \\ \mathbf{x}_n\mathbf{Y}_{n-1} + x_{n,n}\mathbf{y}_n & x_{n,n}y_{n,n} \end{bmatrix}.$$

We can now again invoke Equation (1) to observe that these forms satisfy this definition and thus are also matryoshkan matrices.

(ii) Let $\mathbf{M}_n \in \mathbb{R}^{n \times n}$ be a matryoshkan matrix with all nonzero diagonal elements $m_{i,i}$ for $i \in \{1, \ldots, n\}$. By definition $\mathbf{M}_n$ is lower triangular and hence its eigenvalues are on its diagonal. Since all the eigenvalues are nonzero by assumption, $\mathbf{M}_n$ is invertible. Moreover, it is known that the inverse of a lower triangular matrix is lower triangular as well. Thus, we will now solve for a lower triangular matrix $\mathbf{W}_n \in \mathbb{R}^{n \times n}$ such that $\mathbf{I}_n = \mathbf{M}_n \mathbf{W}_n$, where $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ is the identity. This can be written as

$$\begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{0}_{1 \times n-1} & 1 \end{bmatrix} = \mathbf{I}_n = \mathbf{M}_n \mathbf{W}_n = \begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n & m_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{b} & c \end{bmatrix},$$

where $\mathbf{A} \in \mathbb{R}^{n-1 \times n-1}$, $b \in \mathbb{R}^{1 \times n-1}$, and $c \in \mathbb{R}$. Because $m_{i,i} \neq 0$ for all $i \in \{1, \ldots, n-1\}$, we also know that $\mathbf{M}_{n-1}$ is nonsingular. Thus, we can see that $\mathbf{A} = \mathbf{M}_{n-1}^{-1}$ from $\mathbf{M}_{n-1} \mathbf{A} = \mathbf{I}_{n-1}$. Likewise, $c m_{n,n} = 1$ implies $c = \frac{1}{m_{n,n}}$. Then, we have that

$$\mathbf{0}_{1 \times n-1} = \mathbf{m}_n \mathbf{A} + m_{n,n} \mathbf{b} = \mathbf{m}_n \mathbf{M}_{n-1}^{-1} + m_{n,n} \mathbf{b},$$

and so $\mathbf{b} = -\frac{1}{m_{n,n}} \mathbf{m}_n \mathbf{M}_{n-1}^{-1}$. This completes the solution for $\mathbf{W}_n$, and hence provides the inverse of $\mathbf{M}_n$.

(iii) To begin, we will prove that

$$\mathbf{M}_n^k = \begin{bmatrix} \mathbf{M}_{n-1}^k & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n \sum_{j=0}^{k-1} \mathbf{M}_{n-1}^j m_{n,n}^{k-1-j} & m_{n,n}^k \end{bmatrix}$$

for $k \in \mathbb{Z}^+$. We proceed by induction. The base case, $k = 1$, holds by definition. Therefore we suppose that the hypothesis holds at k. Then at $k+1$ we can observe that

$$\mathbf{M}_n^{k+1} = \mathbf{M}_n \mathbf{M}_n^k$$

$$= \begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n & m_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{M}_{n-1}^k & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n \sum_{j=0}^{k-1} \mathbf{M}_{n-1}^j m_{n,n}^{k-1-j} & m_{n,n}^k \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{M}_{n-1}^{k+1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n \mathbf{M}_{n-1}^k + \mathbf{m}_n \sum_{j=0}^{k-1} \mathbf{M}_{n-1}^j m_{n,n}^{k-j} & m_{n,n}^{k+1} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{M}_{n-1}^{k+1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n \sum_{j=0}^{k} \mathbf{M}_{n-1}^j m_{n,n}^{k-j} & m_{n,n}^{k+1} \end{bmatrix},$$

which completes the induction. We now observe further that for matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times n}$ such that $\mathbf{AB} = \mathbf{BA}$ and $\mathbf{A} - \mathbf{B}$ is nonsingular,

$$\sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{k-1-j} = (\mathbf{A} - \mathbf{B})^{-1} \left(\mathbf{A}^k - \mathbf{B}^k \right).$$

This relationship can verified by multiplying the left-hand side by $\mathbf{A} - \mathbf{B}$:

$$(\mathbf{A} - \mathbf{B}) \sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{k-1-j} = \sum_{j=0}^{k-1} \mathbf{A}^{j+1} \mathbf{B}^{k-1-j} - \sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{B}^{k-j} = \mathbf{A}^k - \mathbf{B}^k.$$

This allows us to observe that

$$\mathbf{M}_n^k = \begin{bmatrix} \mathbf{M}_{n-1}^k & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1} (\mathbf{M}_{n-1}^k - m_{n,n}^k \mathbf{I}) & m_{n,n}^k \end{bmatrix},$$

and thus

$$e^{\mathbf{M}_n t} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{M}_n^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{bmatrix} \mathbf{M}_{n-1}^k & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1} (\mathbf{M}_{n-1}^k - m_{n,n}^k \mathbf{I}) & m_{n,n}^k \end{bmatrix}$$

$$= \begin{bmatrix} e^{\mathbf{M}_{n-1} t} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1} (e^{\mathbf{M}_{n-1} t} - e^{m_{n,n} t} \mathbf{I}) & e^{m_{n,n} t} \end{bmatrix},$$

which completes the proof. Note that because $\mathbf{M}_{n-1}$ is triangular and because we have assumed $m_{1,1}, \ldots, m_{n,n}$ are distinct, we know that $\mathbf{M}_{n-1} - m_{n,n}\mathbf{I}$ is invertible.

(iv) From the statement, we seek a matrix $\mathbf{A} \in \mathbb{R}^{n-1 \times n-1}$, a row vector $\mathbf{b} \in \mathbb{R}^{1 \times n-1}$, and a scalar $c \in \mathbb{R}$ such that

$$\begin{bmatrix} \mathbf{M}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n & m_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{b} & c \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{b} & c \end{bmatrix} \begin{bmatrix} \mathbf{D}_{n-1} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{0}_{1 \times n-1} & m_{n,n} \end{bmatrix},$$

where $\mathbf{D}_{n-1} \in \mathbb{R}^{n-1 \times n-1}$ is a diagonal matrix with values $m_{1,1}, \ldots, m_{n-1,n-1}$. From the triangular structure of $\mathbf{M}_n$, we know that $\mathbf{D}_n$ contains all the eigenvalues of $\mathbf{M}_n$. We will now solve the resulting sub-systems. From $\mathbf{M}_{n-1}\mathbf{A} = \mathbf{A}\mathbf{D}_{n-1}$, we take $\mathbf{A} = \mathbf{U}_{n-1}$. Substituting this forward, we see that

$$\mathbf{m}_n \mathbf{U}_{n-1} + m_{n,n}\mathbf{b} = \mathbf{m}_n \mathbf{A} + m_{n,n}\mathbf{b} = \mathbf{b}\mathbf{D}_{n-1}$$

and so $b = \mathbf{m}_n \mathbf{U}_{n-1}(\mathbf{D}_{n-1} - m_{n,n}\mathbf{I})^{-1}$, where as in Step (iii) we are justified in inverting $\mathbf{D}_{n-1} - m_{n,n}\mathbf{I}$ because of the fact that $m_{1,1}, \ldots, m_{n,n}$ are distinct. Finally, we take $c = 1$, as any value will satisfy $cm_{n,n} = cm_{n,n}$. □

Appendix B. Proof of Lemma 1

Proof. The vector solution in Equation (5) is known and is thus displayed for reference. Expanding this expression in bracket-notation form, by use of Proposition 1 this is

$$\begin{bmatrix} \mathbf{s}_{n-1}(t) \\ s_n(t) \end{bmatrix} = \begin{bmatrix} e^{\mathbf{M}_{n-1} t} & \mathbf{0}_{n-1 \times 1} \\ \mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1} (e^{\mathbf{M}_{n-1} t} - e^{m_{n,n} t} \mathbf{I}) & e^{m_{n,n} t} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{n-1}(0) \\ s_n(0) \end{bmatrix}$$

$$- \begin{bmatrix} \mathbf{M}_{n-1}^{-1} & \mathbf{0}_{n-1 \times 1} \\ -\frac{1}{m_{n,n}} \mathbf{m}_n \mathbf{M}_{n-1}^{-1} & \frac{1}{m_{n,n}} \end{bmatrix} \begin{bmatrix} \mathbf{I} - e^{\mathbf{M}_{n-1} t} & \mathbf{0}_{n-1 \times 1} \\ -\mathbf{m}_n (\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1} (e^{\mathbf{M}_{n-1} t} - e^{m_{n,n} t} \mathbf{I}) & 1 - e^{m_{n,n} t} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{n-1} \\ c_n \end{bmatrix}.$$

Thus, we can find $s_n(t)$ by multiplying each left side of the equality by a unit row vector in the direction of the nth coordinate, which we denote by $\mathbf{v}_n^{\mathrm{T}}$. This yields

$$
s_n(t) = \mathbf{v}_n^{\mathrm{T}} \begin{bmatrix} \mathbf{s}_{n-1}(t) \\ s_n(t) \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I}) & e^{m_{n,n}t} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{n-1}(0) \\ s_n(0) \end{bmatrix}
$$

$$
- \begin{bmatrix} -\frac{1}{m_{n,n}}\mathbf{m}_n\mathbf{M}_{n-1}^{-1} & \frac{1}{m_{n,n}} \end{bmatrix} \begin{bmatrix} \mathbf{I} - e^{\mathbf{M}_{n-1}t} & \mathbf{0}_{n-1\times 1} \\ -\mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I}) & 1 - e^{m_{n,n}t} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{n-1} \\ c_n \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I}) & e^{m_{n,n}t} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{n-1}(0) \\ s_n(0) \end{bmatrix}
$$

$$
- \begin{bmatrix} -\frac{1}{m_{n,n}}\mathbf{m}_n\mathbf{M}_{n-1}^{-1} & \frac{1}{m_{n,n}} \end{bmatrix} \begin{bmatrix} (\mathbf{I} - e^{\mathbf{M}_{n-1}t})\,\mathbf{c}_{n-1} \\ -\mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I})\,\mathbf{c}_{n-1} + c_n(1 - e^{m_{n,n}t}) \end{bmatrix}.
$$

Then, by taking these inner products, we obtain

$$
s_n(t) = \mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}\left(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I}\right)\mathbf{s}_{n-1}(0) + s_n(0)e^{m_{n,n}t}
$$

$$
+ \mathbf{m}_n\mathbf{M}_{n-1}^{-1}\left(\mathbf{I} - e^{\mathbf{M}_{n-1}t}\right)\frac{\mathbf{c}_{n-1}}{m_{n,n}}
$$

$$
+ \mathbf{m}_n(\mathbf{M}_{n-1} - m_{n,n}\mathbf{I})^{-1}\left(e^{\mathbf{M}_{n-1}t} - e^{m_{n,n}t}\mathbf{I}\right)\frac{\mathbf{c}_{n-1}}{m_{n,n}} - \frac{c_n}{m_{n,n}}\left(1 - e^{m_{n,n}t}\right),
$$

and this simplifies to the stated solution. □

Acknowledgements

Funding information

Competing interests

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

References

[1] ACKERER, D., FILIPOVIĆ, D. AND PULIDO, S. (2018). The Jacobi stochastic volatility model. *Finance Stoch.* **22**, 667–700.

[2] AT-SAHALIA, Y., CACHO-DIAZ, J. AND LAEVEN, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *J. Financial Econom.* **117**, 585–606.

[3] BENTH, F. E. AND LAVAGNINI, S. (2021). Correlators of polynomial processes. *SIAM J. Financial Math.* **12**, 1374–1415.

[4] BIAGINI, F. AND ZHANG, Y. (2016). Polynomial diffusion models for life insurance liabilities. *Insurance Math. Econom.* **71**, 114–129.

[5] BLADT, M., ASMUSSEN, S. AND STEFFENSEN, M. (2019). Matrix calculations for inhomogeneous Markov reward processes, with applications to life insurance and point processes. Preprint. Available at https://arxiv.org/abs/1905.04605.

[6] BOXMA, O., KELLA, O. AND PERRY, D. (2011). On some tractable growth–collapse processes with renewal collapse epochs. *J. Appl. Prob.* **48**, 217–234.

[7] BOXMA, O., PERRY, D., STADJE, W. AND ZACKS, S. (2006). A Markovian growth–collapse model. *Adv. Appl. Prob.* **38**, 221–243.

[8] BRAWER, R. AND PIROVINO, M. (1992). The linear algebra of the Pascal matrix. *Linear Algebra Appl.* **174**, 13–23.

[9] CALL, G. S. AND VELLEMAN, D. J. (1993). Pascal's matrices. *Amer. Math. Monthly* **100**, 372–376.

[10] CUCHIERO, C. (2019). Polynomial processes in stochastic portfolio theory. *Stoch. Process. Appl.* **129**, 1829–1872.

[11] CUCHIERO, C., KELLER-RESSEL, M. AND TEICHMANN, J. (2012). Polynomial processes and their applications to mathematical finance. *Finance Stoch.* **16**, 711–740.

[12] CUCHIERO, C., LARSSON, M. AND SVALUTO-FERRO, S. (2018). Polynomial jump-diffusions on the unit simplex. *Ann. Appl. Prob.* **28**, 2451–2500.

[13] CUI, L., HAWKES, A. AND YI, H. (2020). An elementary derivation of moments of Hawkes processes. *Adv. Appl. Prob.* **52**, 102–137.

[14] CUI, L., WU, B. AND YIN, J. (2021). Moments for Hawkes processes with gamma decay kernel functions. To appear in *Methodology Comput. Appl. Prob.*

[15] DA FONSECA, J. AND ZAATOUR, R. (2014). Hawkes process: fast calibration, application to trade clustering, and diffusive limit. *J. Futures Markets* **34**, 548–579.

[16] DALEY, D. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. **I**, *Elementary Theory and Methods*. Springer, New York.

[17] DASSIOS, A. AND ZHAO, H. (2011). A dynamic contagion process. *Adv. Appl. Prob.* **43**, 814–846.

[18] DAW, A. AND PENDER, J. (2018). Queues driven by Hawkes processes. *Stoch. Systems* **8**, 192–229.

[19] DAW, A. AND PENDER, J. (2022). An ephemerally self-exciting point process. *Adv. Appl. Prob.* **54**, 340–403.

[20] DUFFIE, D., PAN, J. AND SINGLETON, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **68**, 1343–1376.

[21] EDELMAN, A. AND STRANG, G. (2004). Pascal matrices. *Amer. Math. Monthly* **111**, 189–197.

[22] ERRAIS, E., GIESECKE, K. AND GOLDBERG, L. R. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1**, 642–665.

[23] ETHIER, S. N. AND KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence*. John Wiley, Hoboken, NJ.

[24] FILIPOVIĆ, D. AND LARSSON, M. (2016). Polynomial diffusions and applications in finance. *Finance Stoch.* **20**, 931–972.

[25] FILIPOVIĆ, D. AND LARSSON, M. (2020). Polynomial jump-diffusion models. *Stoch. Systems* **10**, 71–97.

[26] FORMAN, J. L. AND SØRENSEN, M. (2008). The Pearson diffusions: a class of statistically tractable diffusion processes. *Scand. J. Statist.* **35**, 438–465.

[27] FRALIX, B. (2020). On classes of Bitcoin-inspired infinite-server queueing systems. *Queueing Systems* **95**, 29–52.

[28] FROLKOVA, M. AND MANDJES, M. (2019). A Bitcoin-inspired infinite-server model with a random fluid limit. *Stoch. Models* **35**, 1–32.

[29] GAO, X. AND ZHU, L. (2018). Limit theorems for Markovian Hawkes processes with a large initial intensity. *Stoch. Process. Appl.* **128**, 3807–3839.

[30] GAO, X. AND ZHU, L. (2019). Affine point processes: refinements to large-time asymptotics. Preprint. Available at https://arxiv.org/abs/1903.06371.

[31] GUASONI, P. AND WONG, K. C. (2020). Asset prices in segmented and integrated markets. *Finance Stoch.* **24**, 939–980.

[32] HALE, J. K. AND LUNEL, S. M. V. (2013). *Introduction to Functional Differential Equations*. Springer, New York.

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

[34] HENRY-LABORDÈRE, P. (2008). *Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing*. Chapman and Hall/CRC, Boca Raton, FL.

[35] IVERSEN, V. B. (1985). A generalization of the classical teletraffic theory. In *Proc. Eleventh International Teletraffic Congress*, Elsevier, Amsterdam, pp. 58–64.

[36] IVERSEN, V. B. (ed.) (2005). *Teletraffic Engineering Handbook*. International Telecommunication Union, Geneva.

[37] JAVIER, K. AND FRALIX, B. (2020). A further study of some Markovian Bitcoin models from *Göbel et al. Stoch. Models* **36**, 223–250.

[38] KELLA, O. (2009). On growth–collapse processes with stationary structure and their shot-noise counterparts. *J. Appl. Prob.* **46**, 363–371.

[39] KELLA, O. AND LÖPKER, A. (2010). A Markov-modulated growth collapse model. *Prob. Eng. Inf. Sci.* **24**, 99–107.

[40] KOOPS, D. (2018). Predicting the confirmation time of Bitcoin transactions. Preprint. Available at https://arxiv.org/abs/1809.10596.

[41] KOOPS, D. T., SAXENA, M., BOXMA, O. J. AND MANDJES, M. (2018). Infinite-server queues with Hawkes input. *J. Appl. Prob.* **55**, 920–943.

[42] MAZET, O. (1997). Classification des semi-groupes de diffusion sur IR associés à une famille de polynômes orthogonaux. In *Séminaire de Probabilités XXXI*, Springer, Berlin, Heidelberg, pp. 40–53.

[43] NASR, W. W., CHARANEK, A. AND MADDAH, B. (2018). MAP fitting by count and inter-arrival moment matching. *Stoch. Models* **34**, 292–321.

[44] NIELSEN, B. F., NILSSON, L. F., THYGESEN, U. H. AND BEYER, J. E. (2007). Higher order moments and conditional asymptotics of the batch Markovian arrival process. *Stoch. Models* **23**, 1–26.

[45] ØKSENDAL, B. (2013). *Stochastic Differential Equations: an Introduction with Applications*. Springer, Berlin, Heidelberg.

[46] WALLSTRÖM, B. (1964). A distribution model for telephone traffic with varying call intensity, including overflow traffic. *Ericsson Technics* **20**, 183–202.

[47] WONG, E. (1964). The construction of a class of stationary Markoff processes. *Stoch. Process. Math. Phys. Eng.* **17**, 264–276.

[48] ZHANG, X., BLANCHET, J., GIESECKE, K. AND GLYNN, P. W. (2015). Affine point processes: approximation and efficient simulation. *Math. Operat. Res.* **40**, 797–819.

[49] ZHANG, Z. (1997). The linear algebra of the generalized Pascal matrix. *Linear Algebra Appl.* **250**, 51–60.