stochastic processes and their applications

Point processes with finite-dimensional conditional probabilities

Søren Asmussen[a,*], Mogens Bladt[b]

[a] Department of Mathematical Statistics, Lund University, Box 119, 221 00 Lund, Sweden
[b] IIMAS–UNAM, A.P. 20-726, Mexico, D.F. USA

Received 19 June 1998; received in revised form 29 January 1999; accepted 5 February 1999

Abstract

We study the structure of point processes N with the property that the $\mathbb{P}(\theta_t N \in \cdot \mid \mathscr{F}_t)$ vary in a finite-dimensional space where θ_t is the shift and $\mathscr{F}_t$ the σ-field generated by the counting process up to time t. This class of point processes is strictly larger than Neuts' class of Markovian arrival processes. On the one hand, it allows for more general features like interarrival distributions which are matrix-exponential rather than phase type, on the other the probabilistic interpretation is a priori less clear. Nevertheless, the properties are very similar. In particular, finite-dimensional distributions of interarrival times, moments, Laplace transforms, Palm distributions, etc., are shown to be given by two fundamental matrices $\boldsymbol{C}$, $\boldsymbol{D}$ just as for the Markovian arrival process. We also give a probabilistic interpretation in terms of a piecewise deterministic Markov process on a compact convex subset of $\mathbb{R}^p$, whose jump times are identical to the epochs of N.

MSC: 60J25; 60G55

Keywords: Markovian arrival process; Matrix-exponential distribution; Palm theory; Rational Laplace transform; Semi-group

1. Introduction

Neuts' Markovian arrival process (MAP) (Neuts, 1979) is one of the main specific examples in point process theory. It is defined by a background Markov process $\{J(t)\}$ with $p < \infty$ states and intensity matrix $\boldsymbol{Q}$. On time intervals where $J(t)=i$, the arrivals are Poisson at rate β_i. In addition, there is a probability a_{ij} that an arrival occurs at a jump from i to $j \neq i$.

Examples of the MAP incorporate, e.g., Markov-modulated Poisson processes, renewal processes with phase-type (PHT) (Neuts, 1981; Asmussen, 1987, Chapter III.6) interarrival times and semi-Markov point processes with PHT interarrival times. In fact (Asmussen and Koole, 1963), MAPs are dense in a suitable sense in the space of point processes on $[0, \infty)$. Together with the amenability of models using the MAP to

* Corresponding author.
E-mail addresses: asmus@maths.lth.se (S. Asmussen), bladt@markov.iimas.unam.mx. (M. Bladt)

0304-4149/99/$ - see front matter
PII: S0304-4149(99)00006-X

algorithmic treatment, this has made the MAP a popular tool in queueing theory (a basic paper is Lucantoni (1991); recent surveys and extensive lists of references are given in Neuts (1992, 1995); and Ramaswami (1995)) for recent surveys and extensive lists of references), with a particularly simple analysis being available for the case where the service times are also PHT.

In older literature (e.g. Cohen, 1982; Smith, 1953; Cox, 1955), the larger class of distributions B with a rational Laplace transform is used instead of the PHT class. An equivalent characterization of such a distribution is that it has a matrix-exponential (ME) density b,

$$b(x) = \boldsymbol{\alpha} \mathrm{e}^{\boldsymbol{T}x} \boldsymbol{s} \tag{1.1}$$

(here $\boldsymbol{\alpha}$ is a row vector, $\boldsymbol{T}$ a square matrix and $\boldsymbol{s}$ a column vector). See, e.g., Asmussen and Bladt (1996), Asmussen and O'Cinneide (1998) and Lipsky (1992). The question we ask in this paper is the following: *is there a natural class of point processes extending the MAP in a similar way as ME distributions generalizes PHT distributions*? For example, such point processes should allow for general ME interarrival times, preferably dependent.

Our starting point is the characterization of O'Cinneide (1989) of ME distributions. For a given distribution F on $[0,\infty)$, let F_t denote the distribution $F_t(x) := F(x+t) - F(x)$. In terms of random variables, if X has distribution F, then F_t is the defective distribution of the residual life $X - t$, defined on $\{X > t\}$ only. Let $\mathrm{span}(F)$ denote the linear space of measures consisting of all linear combinations of the F_t. Then:

Proposition 1.1. *A distribution F is ME if and only if* $\mathrm{span}(F)$ *is finite-dimensional.*

For the extension to point processes, let $\mathcal{N}$ be the set of all counting measures on $(0,\infty)$, equipped with the usual vague topology and the corresponding Borel σ-algebra $\mathscr{B}_{\mathcal{N}}$, and $\mathscr{M}(\mathcal{N})$ the set of all finite signed measures on $(\mathcal{N}, \mathscr{B}_{\mathcal{N}})$. A point process is then a random element of $(\mathcal{N}, \mathscr{B}_{\mathcal{N}})$, defined say on $(\Omega, \mathscr{F}; \mathbb{P})$ (see e.g. Franken et al. (1982), Kallenberg (1983) and König and Schmidt (1992) for general background on point processes). Let $0 < T_1 < T_2 < \cdots$ be the event times (measurable functionals $T_i = T_i(N)$ on $\mathcal{N}$), and let $\mathscr{F}_t = \sigma(N(s): s \leqslant t)$ where $\{N(t)\}$ is the (right-continuous) counting process associated with N. Let θ_t be the usual shift operator on $\mathcal{N}$, and write $\mu(t,\cdot)$ for a version of $\mathbb{P}(\theta_t N \in \cdot \mid \mathscr{F}_t)$ and $\mu(t,\omega)$ for $\mu(t,\cdot)$ evaluated at $\omega \in \Omega$. We will freely change between the two equivalent notations $\mathbb{P}(\theta_t N \in \cdot \mid \mathscr{F}_t)$ and $\mu(t,\cdot)$. In the terminology of Knight (1981,1992), $\{\mu(t,\cdot)\}_{t \geqslant 0}$ is the *prediction process*.

Definition 1.1. We call a point process N a rational arrival process (*RAP*) if $\mathbb{P}(N(0,\infty) = \infty) = 1$ and there exists a finite-dimensional subspace V of $\mathscr{M}(\mathcal{N})$ such that for any t, $\mathbb{P}(\theta_t N \in \cdot \mid \mathscr{F}_t)$ has a version $\mu(t,\cdot)$ with $\mu(t,\omega) \in V$ for all $\omega \in \Omega$.

Our aim is to characterize RAPs. For a MAP, the natural choice of V is $\mathrm{span}(v_1, \ldots, v_p)$ where v_i is the distribution of N corresponding to $J(0) = i$. Indeed,

$$\mathbb{P}(\theta_t N \in \cdot \mid \mathscr{F}_t) = \sum_{i=1}^{p} A_i(t) v_i,$$

where $A_i(t) = \mathbb{P}(J(t) = i \mid \mathscr{F}_t)$. According to Proposition 1.1, a renewal process with ME but not PHT interarrival time gives a (trivial) example of a RAP which is not a MAP. We will exhibit more substantial examples in Section 3.

One complete characterization of a RAP is the following. Define $\operatorname{dev}(\boldsymbol{C})$ as the dominant eigenvalue (the one with maximal real part) of a matrix $\boldsymbol{C}$ and let $\boldsymbol{e} = (1\ldots1)'$. For a given point process N, let $f_{N,n}(x_1,\ldots,x_n)$ denote the joint density of $T_1, T_2 - T_1, \ldots, T_n - T_{n-1}$ (the first n interarrival times) at $x_1,\ldots,x_n$.

Theorem 1.1. *Let N be a RAP. Then there exist matrices $\boldsymbol{C}, \boldsymbol{D}$, a row vector $\boldsymbol{\alpha}$ and a column vector $\boldsymbol{s}$, such that* $\operatorname{dev}(\boldsymbol{C}) < 0$, $\operatorname{dev}(\boldsymbol{C}+\boldsymbol{D}) = 0$, $(\boldsymbol{C}+\boldsymbol{D})\boldsymbol{e} = 0$, *and*

$$f_{N,n}(x_1,\ldots,x_n) = \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{C}x_1}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{D}\ldots\mathrm{e}^{\boldsymbol{C}x_n}\boldsymbol{s}. \tag{1.2}$$

Here $\boldsymbol{s}$ can be taken as $\boldsymbol{D}$e. In particular, the nth interarrival time $T_n - T_{n-1}$ is ME with density $\boldsymbol{\alpha}(-\boldsymbol{C}^{-1}\boldsymbol{D})^{n-1}\mathrm{e}^{\boldsymbol{C}x}\boldsymbol{s}$. Conversely, if a point process N has the property (1.2), *then it is a RAP.*

For a MAP, (1.2) is standard and the matrix $\boldsymbol{Q} = \boldsymbol{C} + \boldsymbol{D}$ is the intensity matrix of a continuous-time Markov chain with finitely many states. The matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ correspond to a decomposition of $\boldsymbol{Q}$ where $\boldsymbol{D}$ gives the “intensities of state change with arrivals”, and $\boldsymbol{C}$ those of “state changes without arrivals”. That is,

$$d_{ij} = \begin{cases} \beta_i, & i = j \\ q_{ij}a_{ij} & i \neq j \end{cases}, \qquad c_{ij} = \begin{cases} -\sum_{k\neq i} c_{ik} - \sum_{k=1}^{p} d_{ik}, & i = j, \\ q_{ij}(1 - a_{ij}), & i \neq j. \end{cases}$$

The proof of Theorem 1.1 and other general results on the structure of a RAP is given in Section 2. One main result is that the RAP N can be seen as generated by a piecewise deterministic Markov process $\{\boldsymbol{A}(t)\}_{t\geqslant 0}$ on a compact convex subset of $\mathbb{R}^p$, such that the epochs of N are identical to the jump times of $\{\boldsymbol{A}(t)\}$. Note that it is not surprising that a finite-state space does not suffice – it is well known that the MAP is the most general point process such that the counting process is an additive process on a finite Markov chain. Section 3 contains some examples illustrating the general theory. In Section 4, we show that in a manner similar to (1.2), analytical formulas for the moments $\mathbb{E}N(t)$ of the counting process, its Laplace transform $\mathbb{E}\mathrm{e}^{-\theta N(t)}$, etc., carry over from the MAP to the RAP. We also consider Palm theory.

From the point of view of performance evaluation, say for queues, the usefulness of modeling via MAPs or PHT distributions stems from the fact that algorithmic solutions are available which require basically only finite matrix algebra. For finite matrix algebra to be applicable, one can conversely argue heuristically that some finite-dimensionality property must be available. This provides one possible motivation for Definition 1.1. In fact, we believe that queues like RAP/ME/1 can be solved by largely the same algorithms as MAP/PHT/1, and this topic is currently under investigation. This may be useful even for MAPs, since a RAP representation may have much smaller dimension than a MAP representation, as discussed in Asmussen and Bladt (1996) for the special case of renewal processes. However, we do not want to insist too much on the performance evaluation aspects of the present paper. Rather, it is the mathematical problem which is in the center.

2. Main results and proofs

In this section we will need the following lemma.

Lemma 2.1. *Let $v_1,\dots,v_p$ be linearly independent probability measures on an arbitrary space. Then there exists a constant $\bar{a}$ such that $|a_i|\leqslant\bar{a}$ whenever $\sum_1^p a_i v_i$ is a probability measure.*

Proof. Assume there exists a sequence $\{\boldsymbol{a}^{(n)}\}$ such that $\mu^{(n)}=\boldsymbol{a}^{(n)}\boldsymbol{v}$ is a probability measure and the largest component of $\boldsymbol{a}^{(n)}$ is unbounded, i.e. $||\boldsymbol{a}^{(n)}||\to\infty$ where $||\boldsymbol{a}||=\max_{1,\dots,p}|a_i|$. Then $\mu^{(n)}/||\boldsymbol{a}^{(n)}||\to 0$. Choosing a subsequence $\{n_k\}$ such that $\boldsymbol{a}^{(n_k)}/||\boldsymbol{a}^{(n_k)}||\to\boldsymbol{a}$ for some $\boldsymbol{a}$ with $||\boldsymbol{a}||=1$, we have $\boldsymbol{a}\boldsymbol{v}=0$, contradicting the linear independence. □

Consider a fixed RAP N with distribution $\mathbb{P}$. Assume in the following that V in Definition 1.1 is chosen with minimal dimension (it is easy to see that this V is unique). Choose some basis $v_1,\dots,v_p$ for V. Without loss of generality, we can take the v_i as probability measures (say $\mu_i=\mu(t_i,\omega_i)$ for some t_i and ω_i).

Write

$$\mu(t,\cdot)=\mathbb{P}(\theta_t N\in\cdot\,|\,\mathscr{F}_t)=\sum_{i=1}^{p}A_i(t)v_i=\boldsymbol{A}(t)\boldsymbol{v}, \tag{2.1}$$

where $\boldsymbol{A}(t)=(A_1(t)\ \dots\ A_p(t))$, $\boldsymbol{v}=(v_1\ \dots\ v_p)'$.

The process $\{\boldsymbol{A}(t)\}_{t\geqslant 0}$ will play a fundamental role in the following; its state space is contained in the hyperplane (affine space) $\{\boldsymbol{a}\in\mathbb{R}^p\colon \boldsymbol{a}\boldsymbol{e}=1\}$. More precisely, we may take the state space as

$$\mathscr{A}=\left\{\boldsymbol{a}\in\mathbb{R}^p\colon \boldsymbol{a}\boldsymbol{e}=1,\ \sum_1^p a_i v_i(F)\geqslant 0 \textit{ for all } F\in\mathscr{B}_{\mathscr{N}}\right\};$$

by Lemma 2.1, $\mathscr{A}$ is compact and convex. In particular, $\boldsymbol{A}(t)\in\mathscr{A}$ implies that the $A_i(t)$ are integrable. For $\boldsymbol{a}\in\mathscr{A}$, we define $\mathbb{P}_{\boldsymbol{a}}=\sum_1^p a_i v_i$.

Let $v_i\circ\theta_s$ denote the probability measure $v_i(\theta_s N\in\cdot)$.

Proposition 2.1. (a) *There exists a $p\times p$ matrix $\boldsymbol{Q}$ such that $\boldsymbol{v}\circ\theta_s=\mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{v}$.*

(b) $\mathrm{dev}(\boldsymbol{Q})=0$, $\boldsymbol{Q}\boldsymbol{e}=\boldsymbol{0}$.

(c) $\mathbb{E}[\boldsymbol{A}(t+s)|\mathscr{F}_t]=\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{Q}s}$. *Equivalently, $\{\boldsymbol{A}(t)\mathrm{e}^{-\boldsymbol{Q}t}\}_{t\geqslant 0}$ is a vector-valued martingale.*

Proof. (a) By (2.1),

$$\mathbb{P}(\theta_{t+s}N\in\cdot\,|\mathscr{F}_t)=\sum_{i=1}^{p}A_i(t)v_i\circ\theta_s.$$

On the other hand, by the chain rule for conditional probabilities

$$\begin{aligned}\mathbb{P}(\theta_{t+s}N \in \cdot\,|\mathscr{F}_t) &= \mathbb{E}[\mathbb{P}(\theta_{t+s}N \in \cdot\,|\mathscr{F}_{t+s})|\mathscr{F}_t]\\ &= \sum_{i=1}^{p} \mathbb{E}[A_i(t+s)|\mathscr{F}_t]v_i.\end{aligned}$$

Hence we have the following important identity:

$$\boldsymbol{A}(t,\omega)(\boldsymbol{v}\circ\theta_s) = \boldsymbol{B}(t,s,\omega)\boldsymbol{v},$$

where $\boldsymbol{B}(t,s,\cdot)$ is a version of $\mathbb{E}(\boldsymbol{A}(t+s,\omega)|\mathscr{F}_t)$; for a fixed s, this is valid for all t and for all $\omega \notin S(t)$ where $S(t) \in \mathscr{F}$ is a null set. Now we can choose $t_1,\dots,t_p$ and $\omega_1 \notin S(t_1),\dots,\omega_p \notin S(t_p)$ such that $\boldsymbol{A}(t_1,\omega_1),\dots,\boldsymbol{A}(t_p,\omega_p)$ are linearly independent. Indeed, if span$(\boldsymbol{A}(t,\omega)\colon t>0, \omega \notin S(t))$ is a proper subspace L of $\mathbb{R}^p$, we obtain a new version $\tilde{\mu}(t,\cdot)$ of $\mathbb{P}(\theta_t N \in \cdot\,|\mathscr{F}_t)$ by changing $\boldsymbol{A}(t,\omega)$ to an element of L for any $\omega \in S(t)$. Then $\tilde{\mu}(t,\omega) \in \{\sum_1^p a_i v_i\colon (a_1,\dots,a_p) \in L\}$ which is a proper subspace of V, contradicting the minimality of V.

We now get

$$\boldsymbol{v}\circ\theta_s = \begin{pmatrix}\boldsymbol{A}(t_1,\omega_1)\\ \boldsymbol{A}(t_2,\omega_2)\\ \cdots\\ \boldsymbol{A}(t_p,\omega_p)\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{B}(t_1,s,\omega_1)\\ \boldsymbol{B}(t_2,s,\omega_2)\\ \cdots\\ \boldsymbol{B}(t_p,s,\omega_p)\end{pmatrix}\boldsymbol{v}$$

which means that $\boldsymbol{v}\circ\theta_s = \tilde{\boldsymbol{Q}}(s)\boldsymbol{v}$ for some matrix $\tilde{\boldsymbol{Q}}(s)$ which is unique by linear independence of the v_i. Furthermore, $\tilde{\boldsymbol{Q}}(s_1+s_2) = \tilde{\boldsymbol{Q}}(s_1)\tilde{\boldsymbol{Q}}(s_2)$ by the semi-group property of θ_s, and since $\boldsymbol{v}\circ\theta_s$ is right-continuous in s, we conclude that $\tilde{\boldsymbol{Q}}(s)= \mathrm{e}^{\boldsymbol{Q}s}$ for some $\boldsymbol{Q}$, and (a) is proved. Then (c) follows immediately since

$$\sum_{i=1}^{p}\mathbb{E}[A_i(t+s)|\mathscr{F}_t]v_i = \sum_{i=1}^{p}A_i(t)v_i\circ\theta_s = \sum_{i,j=1}^{p}A_i(t)(\mathrm{e}^{\boldsymbol{Q}s})_{ji}v_j$$

so that $\mathbb{E}[\boldsymbol{A}(t+s)|\mathscr{F}_t]\boldsymbol{v} = \boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{v}$ which by linear independence of the v_i implies $\mathbb{E}[\boldsymbol{A}(t+s)|\mathscr{F}_t] = \boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{Q}s}$.

In (b), $\boldsymbol{Qe}=\boldsymbol{0}$ follows from

$$\mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{e} = \mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{v}(\mathscr{N}) = \boldsymbol{v}\circ\theta_s(\mathscr{N}) = \boldsymbol{e}.$$

Furthermore, if $\Re\,(\mathrm{dev}(\boldsymbol{Q}))>0$, then some element of $\mathrm{e}^{\boldsymbol{Q}s}$ is unbounded in s which by Lemma 2.1 contradicts $\mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{v} = \boldsymbol{v}\circ\theta_s$. □

Corollary 2.1. $\{A(t)\}_{t\geqslant 0}$ *is time-homogeneous Markov on* $\mathscr{A}$ *with paths which are left-continuous with right limits.*

Proof. The Markov property follows from Knight (1981, 1992) and the path property by general martingale results. □

The Markov property will follow also directly from the analysis of $\{\boldsymbol{A}(t)\}$ to be given below where we shall show more precisely that jumps occur exactly at the event times of N. First we need to introduce a decomposition $\boldsymbol{Q} = \boldsymbol{C} + \boldsymbol{D}$. For $\mu \in \mathscr{M}(\mathscr{N})$, define

$$R_t\mu(A) = \mu\{N \in \mathscr{N}\colon \theta_t N \in A,\ \ T_1 > t\}.$$

That is: if N is a point process (a random element of $\mathscr{N}$) with distribution $\mathbb{P}$, then $R_t\mathbb{P}$ is the defective distribution of $\theta_t N$, defined on $\{T_1 > t\}$ only. Thus if $\mu(\cdot) = \mathbb{P}(N \in \cdot)$ is the distribution of N then

$$R_t\mu(\cdot) = R_t\mathbb{P}(N \in \cdot) = \mathbb{P}(\theta_t N \in \cdot,\ \ T_1 > t).$$

Proposition 2.2. (a) $\{R_t\}_{t\geqslant 0}$ *is a semi-group*: $R_{t+s} = R_t \circ R_s$.

(b) $\mu \in V \Rightarrow R_t\mu \in V$.

(c) *There exists a matrix* $\boldsymbol{C}$ *such that* $R_t(\boldsymbol{a}\boldsymbol{v}) = \boldsymbol{a}\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{v}$ *for all* $\boldsymbol{a} \in \mathbb{R}^p$. *Furthermore* $\operatorname{dev}(\boldsymbol{C}) < 0$.

Proof. First we prove (b). To this end notice that $\{T_1 > t\} = \{T_1 \leqslant t\}^{\mathrm{c}}$ is $\mathscr{F}_t$-measurable, and hence

$$I\{T_1 > t\}\mathbb{P}(\theta_t N \in \cdot \,|\mathscr{F}_t) = \mathbb{P}(\theta_t N \in \cdot, T_1 > t \,|\, \mathscr{F}_t).$$

If μ is the distribution of N then

$$\begin{aligned} R_t\mu(\cdot) &= R_t\mathbb{P}(N \in \cdot) = \mathbb{P}(\theta_t N \in \cdot,\ \ T_1 > t) \\ &= \mathbb{E}[\mathbb{P}(\theta_t N \in \cdot,\ \ T_1 > t|\mathscr{F}_t)] = \mathbb{E}[I\{T_1 > t\}\mathbb{P}(\theta_t N \in \cdot\,|\mathscr{F}_t)] \\ &= \mathbb{E}\left[\sum_{i=1}^{p} I\{T_1 > t\}A_i(t)v_i(\cdot)\right] \\ &= \sum_{i=1}^{p} \mathbb{E}[I\{T_1 > t\}A_i(t)]v_i(\cdot) \in V. \end{aligned}$$

The semi-group property (a) follows from the following consideration. Let $G = \{N \in \mathscr{N}\colon \theta_t N \in A, T_1 > t\}$. Then by definition of R_t we get that

$$\begin{aligned} R_s R_t \mu(A) &= R_s\mu(G) \\ &= \mu\{N \in \mathscr{N}\colon \theta_s N \in G, T_1 > s\}. \end{aligned}$$

But if $\theta_s N \in G$ this implies that $\theta_s\theta_t N \in A$ and $T_1(\theta_s N) > t$, where $T_1(\cdot)$ is the operator that for a given point process input returns the first arrival time. The latter property together with $T_1 > s$ further implies that $T_1(N) > s + t$. The opposite consideration also holds, so we conclude that $R_s R_t \mu = R_{s+t}\mu$ and consequently also equal to $R_{t+s}\mu = R_t R_s \mu$. This proves the semi-group property. The existence of $\boldsymbol{C}$ in (b) is a standard consequence of (a) as in O'Cinneide (1989). That $\operatorname{dev}(\boldsymbol{C}) < 0$ follows since otherwise $R_t\mu$ would not converge to 0 as follows from $\mathbb{P}(T_1 < \infty) = 1$. □

Lemma 2.2.

$$\mathbb{P}(\theta_{t+h} N \in \cdot,\ \ N(t, t+h) = 0 \,|\, \mathscr{F}_t) = \boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}.$$

Proof. Simply notice that

$$\mathbb{P}(\theta_{t+h}N \in \cdot,\ \ N(t,t+h)=0\,|\,\mathscr{F}_t)=R_h\mathbb{P}(\theta_t N \in \cdot\,|\,\mathscr{F}_t). \qquad \Box$$

Now let $\boldsymbol{D}=\boldsymbol{Q}-\boldsymbol{C}$. The following result shows that we can think of $\boldsymbol{A}(t-)\boldsymbol{De}$ as the predictable intensity of the counting process N.

Proposition 2.3. $\mathbb{P}(\theta_{t+h}N \in \cdot, N(t,t+h]>0\,|\,\mathscr{F}_t)=\boldsymbol{A}(t)\boldsymbol{Dv}h+\mathrm{o}(h)$. *In particular*, $\mathbb{P}(N(t,t+h]>0\,|\,\mathscr{F}_t)=\boldsymbol{A}(t)\boldsymbol{De}h+\mathrm{o}(h)$.

Proof.

$$\begin{aligned}
&\mathbb{P}(\theta_{t+h}N \in \cdot,\ \ N(t,t+h]>0\,|\,\mathscr{F}_t)\\
&\quad=\mathbb{P}(\theta_{t+h}N \in \cdot\,|\,\mathscr{F}_t)-\mathbb{P}(\theta_{t+h}N \in \cdot,\ \ N[t,t+h]=0\,|\,\mathscr{F}_t)\\
&\quad=\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{Q}h}\boldsymbol{v}-\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}\\
&\quad=\boldsymbol{A}(t)(\boldsymbol{I}+h\boldsymbol{Q}+\mathrm{o}(h)-\boldsymbol{I}-h\boldsymbol{C}-\mathrm{o}(h))\boldsymbol{v}. \qquad \Box
\end{aligned}$$

Proposition 2.4.

$$\boldsymbol{A}(t)=\boldsymbol{A}(0)+\int_0^t\{\boldsymbol{A}(s)\boldsymbol{C}-\boldsymbol{A}(s)\boldsymbol{Ce}\cdot\boldsymbol{A}(s)\}\,\mathrm{d}s+\sum_{i:T_i\leqslant t}\left\{\frac{\boldsymbol{A}(T_i-)\boldsymbol{D}}{\boldsymbol{A}(T_i-)\boldsymbol{De}}-\boldsymbol{A}(T_i-)\right\}.$$

Proof. We first show that

$$\boldsymbol{A}(t+h)=\frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{e}} \quad \text{on } \{N(t,t+h]=0\}. \tag{2.2}$$

This means that

$$I(N(t,t+h]=0)\boldsymbol{A}(t+h)=I(N(t,t+h]=0)\frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{e}},$$

i.e. (multiply by $\boldsymbol{v}$) that the r.v.'s

$$Z_1=\mathbb{P}(\theta_{t+h}N \in \cdot,\ \ N(t,t+h]=0|\mathscr{F}_{t+h}), \quad Z_2=I(N(t,t+h]=0)\frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{e}}$$

are equal. Since Z_1,Z_2 are $\mathscr{F}_{t+h}$-measurable, it suffices that $\mathbb{E}[Z_1Z_3]=\mathbb{E}[Z_2Z_3]$ for all bounded $\mathscr{F}_{t+h}$-measurable Z_3, which in turn holds if $\mathbb{E}[Z_1Z_4Z_5]=\mathbb{E}[Z_2Z_4Z_5]$ for all bounded $\mathscr{F}_t$-measurable Z_4 and all bounded Z_5 which are measurable w.r.t. $\sigma(N(t,t+v]\colon 0<v\leqslant h)$. Since such a Z_5 is constant a.s., say $Z_5=b$, on $\{N(t,t+h]=0\}$ and $Z_1=Z_2=0$ on $\{N(t,t+h]>0\}$, it suffices that $\mathbb{E}[Z_1Z_4]=\mathbb{E}[Z_2Z_4]$. But by Lemma 2.2,

$$\begin{aligned}
\mathbb{E}[Z_1Z_4]&=\mathbb{E}[Z_4\mathbb{E}[Z_1|\mathscr{F}_t]]=\mathbb{E}[Z_4\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}],\\
\mathbb{E}[Z_2Z_4]&=\mathbb{E}\left[Z_4\frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{e}}\mathbb{P}[N(t,t+h]=0\,|\,\mathscr{F}_t]\right]\\
&=\mathbb{E}[Z_4\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{v}].
\end{aligned}$$

It follows from (2.2) that the right derivative of $\boldsymbol{A}(t)$ exists for all t and is given by

$$\boldsymbol{A}'(t) = \boldsymbol{A}(t)\boldsymbol{C} - \boldsymbol{A}(t)\boldsymbol{C}\boldsymbol{e} \cdot \boldsymbol{A}(t). \tag{2.3}$$

Furthermore, (2.3) is also a left derivative unless t is one of the epochs T_i for N.

Combining with Proposition 2.3, a similar argument now gives

$$\mathbb{P}(\theta_{t+h}N \in \cdot \mid \mathscr{F}_{t+h}) = \frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}\tau - t}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}(t+h-\tau)}\boldsymbol{v}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}\tau - t}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}(t+h-\tau)}\boldsymbol{e}}$$

on $\{N(t,t+h]=1\}$ where $\tau \in (t,t+h]$ is the epoch. Here the r.h.s. is $\boldsymbol{A}(t)\boldsymbol{D}\boldsymbol{v}/\boldsymbol{A}(t)\boldsymbol{D}\boldsymbol{e}+\mathrm{O}(h)$ where the $\mathrm{O}(h)$ term is uniform in t. Taking t,h rational (to avoid trouble with null sets) and considering a sequence $\{t_n,h_n\}$ such that $(t_n,t_n+h_n]\downarrow\{T_i\}$ shows that $\boldsymbol{A}(T_i) = \boldsymbol{A}(T_i-)\boldsymbol{D}\boldsymbol{v}/\boldsymbol{A}(T_i-)\boldsymbol{D}\boldsymbol{e}$. Combining with (2.3), this completes the proof. □

The above results show that we can think of $\{\boldsymbol{A}(t)\}$ as a piecewise deterministic Markov process (Davis, 1993). In between jumps, $\{\boldsymbol{A}(t)\}$ moves according to the deterministic differential equation

$$\dot{\boldsymbol{a}}(t) = \boldsymbol{a}(t)\boldsymbol{C} - \boldsymbol{a}(t)\boldsymbol{C}\boldsymbol{e} \cdot \boldsymbol{a}(t). \tag{2.4}$$

Jumps occur at intensity $\boldsymbol{a}\boldsymbol{D}\boldsymbol{e}$ in state $\boldsymbol{a}$, and the new state after the jump is $\boldsymbol{a}\boldsymbol{D}/\boldsymbol{a}\boldsymbol{D}\boldsymbol{e}$ (deterministic).

This description also immediately yields the following explicit representation of $\boldsymbol{A}(t)$:

Corollary 2.2.

$$\boldsymbol{A}(t) = \frac{\boldsymbol{\alpha}\left(\prod_{i=1}^{N(t)} \mathrm{e}^{\boldsymbol{C}(T_i - T_{i-1})}\boldsymbol{D}\right)\mathrm{e}^{\boldsymbol{C}(t - T_{N(t)})}}{\boldsymbol{\alpha}\left(\prod_{i=1}^{N(t)} \mathrm{e}^{\boldsymbol{C}(T_i - T_{i-1})}\boldsymbol{D}\right)\mathrm{e}^{\boldsymbol{C}(t - T_{N(t)})}\boldsymbol{e}}.$$

Proof of Theorem 1.1. Let first N be a RAP and define $\boldsymbol{\alpha} = \boldsymbol{A}(0)$. Using the strong Markov property of $\{\boldsymbol{A}(t)\}$ (see Davis, 1993), we get

$$\begin{aligned}
\mathbb{P}(T_1 > x_1, T_2 - T_1 > x_2) &= \mathbb{E}_{\boldsymbol{\alpha}}[\mathbb{P}_{\boldsymbol{a}(T_1)}(T_1 > x_2); T_1 > x_1] \\
&= \mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{A}(T_1)\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{e}; T_1 > x_1] \\
&= \mathbb{E}_{\boldsymbol{\alpha}}[\mathbb{E}_{\boldsymbol{A}(x_1)}[\boldsymbol{A}(T_1)\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{e}]; T_1 > x_1] \\
&= \mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{A}(x_1)\int_0^\infty \mathrm{e}^{\boldsymbol{C}t}\boldsymbol{D}\,\mathrm{d}t\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{e}; T_1 > x\right] \\
&= \mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{A}(x_1)(-\boldsymbol{C})^{-1}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{e}; T_1 > x] \\
&= \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{C}x_1}(-\boldsymbol{C})^{-1}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}x_2}\boldsymbol{e}.
\end{aligned}$$

Differentiating, (1.2) follows for $n=2$ with $\boldsymbol{s} = -\boldsymbol{C}\boldsymbol{e} = \boldsymbol{D}\boldsymbol{e}$. The case of $n > 2$ is similar.

Assume conversely that (1.2) holds. Then it is clear that $f_{\theta_t N|\mathscr{F}_t,n}$ again has the form (1.2), but with $\boldsymbol{\alpha}$ replaced by a vector $\boldsymbol{\alpha}^{(t)}$ proportional to

$$\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{C}T_1}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}(T_2-T_1)}\ldots\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}(T_k-T_{k-1})}\boldsymbol{D}\mathrm{e}^{\boldsymbol{C}(t-T_k)}$$

on $\{N(t)=k\}$. Since the $\boldsymbol{\alpha}^{(t)}$ vary in a finite-dimensional space, N is thus a RAP. □

3. Example

Simple examples of RAPs which are not necessarily MAPs are ME renewal processes and semi-Markov point processes with ME interarrival times.

The analysis of Section 2 suggests the following constructive way to exhibit RAPs:

(1) Consider a finite set $f_1,\dots,f_q$ of ME densities.

(2) Extend $f_1,\dots,f_q$ to a set $f_1,\dots,f_p$ of ME densities such that any excess life density of a f_i $(i=1,\dots,p)$ is a linear combination of the $f_1,\dots,f_p$, and $p\geqslant q$ is minimal.

(3) For each f_i, write $f_i(x+y)=\sum_1^p \tilde{c}_{ij}(y)f_j(x)$ (note that then $\tilde{\boldsymbol{C}}(y)=(\tilde{c}_{ij}(y))=\mathrm{e}^{\boldsymbol{C}y}$ for some real matrix $\boldsymbol{C}$ with $\mathrm{dev}(\boldsymbol{C})<0$, and any $f_i(x)$ is ME with density $\boldsymbol{e}_i'\mathrm{e}^{\boldsymbol{C}x}\boldsymbol{e}$).

(4) Compute $\boldsymbol{C}=(c_{ij})$ by $c_{ij}=\tilde{c}_{ij}'(0)$.

(5) Determine the region

$$\mathscr{A}=\left\{\boldsymbol{a}\in\mathbb{R}^p\colon \boldsymbol{ae}=1,\sum_{i=1}^p a_i f_i(x)\geqslant 0 \text{ for all } x\geqslant 0\right\}$$

(by Lemma 2.1, $\mathscr{A}$ is a convex compact subset of $\mathbb{R}^p$).

(6) Let $d_{i.}=-\sum_1^p c_{ij}$. As a candidate for $\boldsymbol{D}$, consider

$$\boldsymbol{D}=\boldsymbol{D}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_p)=\begin{pmatrix} d_{1.}\boldsymbol{a}_1\\ \vdots\\ d_{p.}\boldsymbol{a}_p\end{pmatrix}, \tag{3.1}$$

where $\boldsymbol{a}_1,\dots,\boldsymbol{a}_p\in\mathscr{A}'$, where $\in\mathscr{A}'$ is defined under the next item.

(7) Choose $\mathscr{A}'\subseteq\mathscr{A}$ such that

(a) $\mathscr{A}'$ is invariant under orbit movements, that is for any point $\boldsymbol{a}\in\mathscr{A}'$ the orbit starting from this point remains within $\mathscr{A}'$. Technically the condition amounts to

$$\forall \boldsymbol{a}\in\mathscr{A}'\ \forall s\geqslant 0\colon\ \boldsymbol{a}(s)=\frac{\boldsymbol{a}\mathrm{e}^{\boldsymbol{C}s}}{\boldsymbol{a}\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{e}}\in\mathscr{A}'.$$

(b) For the selected $\boldsymbol{a}_1,\dots,\boldsymbol{a}_p$,

$$\forall \boldsymbol{a}\in\mathscr{A}'\colon\ \frac{\boldsymbol{aD}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_p)}{\boldsymbol{aD}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_p)\boldsymbol{e}}\in\mathscr{A}'.$$

(8) Choose $\boldsymbol{\alpha}=\boldsymbol{A}(0)\in\mathscr{A}'$.

Example 3.1. In step 1, start with $\{f_3(x)\}$ and extend as in 2 by taking

$$f_1(x)=\mathrm{e}^{-x},\quad f_2(x)=\tfrac{2}{3}(1+\sin x)\mathrm{e}^{-x},\quad f_3(x)=\tfrac{2}{3}(1+\cos x)\mathrm{e}^{-x}.$$

Then

$$\tilde{\boldsymbol{C}}(y)=(\tilde{c}_{ij}(y))=\begin{pmatrix}\mathrm{e}^{-y} & 0 & 0\\ \tfrac{2}{3}(1-\cos y-\sin y)\mathrm{e}^{-y} & \cos y\mathrm{e}^{-y} & \sin y\mathrm{e}^{-y}\\ \tfrac{2}{3}(1-\cos y+\sin y)\mathrm{e}^{-y} & -\sin y\mathrm{e}^{-y} & \cos y\mathrm{e}^{-y}\end{pmatrix},$$

$$\boldsymbol{C}=\begin{pmatrix}-1 & 0 & 0\\ -\tfrac{2}{3} & -1 & 1\\ \tfrac{2}{3} & -1 & -1\end{pmatrix},$$

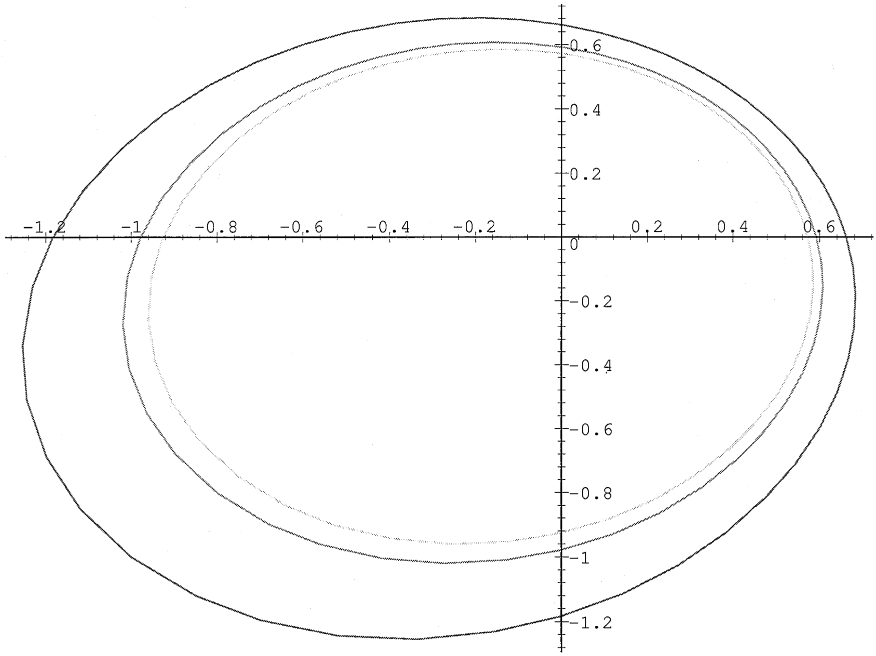

Fig. 1. Three orbits through respectively $(-1,-1)$, $(-0.8,-0.8)$ and $(-0.75,-0.75)$.

$$\mathscr{A}=\left\{(a_1,a_2,a_3)\colon\ a_1+a_2+a_3=1,\ a_1\geqslant -3-\sqrt{a_2^2+a_3^2}\right\}. \tag{3.2}$$

Since $a_1=1-a_2-a_3$ is uniquely determined by a_2,a_3, we can represent $\mathscr{A}$ by

$$\mathscr{A}_0=\left\{(a_2,a_3)\colon\ 1-a_2-a_3\geqslant -3-\sqrt{a_2^2+a_3^2}\right\}, \tag{3.3}$$

which is the region enclosed by an ellipse obtained by translating the ellipse with centre at (0.0) and radia $\frac{5}{7}\sqrt{2}$ and $\sqrt{42}/7$ along the abscissa, $-2/7\sqrt{2}$ units to the left (i.e. in the negative direction), and followed by a rotation of $+45^\circ(=\pi/4)$. One may visualize the Markov process $\{\boldsymbol{A}(t)\}$ by the (a_1,a_2) component moving on these orbits (which in this case are ellipses), with a change of orbit taking place with intensity $\boldsymbol{aDe}$ in state $\boldsymbol{a}$. The orbits through $(-1,-1),(-0.8,-0.8)$ and $(-0.75,-0.75)$ are shown in Fig. 1.

Take any point (a_1,a_2,a_3) such that $(a_2,a_3)\in\mathscr{A}_0$. Then the orbit through $(1-a_2-a_3,a_2,a_3)$ moves in the second and third coordinate according to

$$\boldsymbol{a}(s)=\left[\frac{a_1\cos(s)-a_2\sin(s)}{\frac{5}{3}-\frac{2}{3}\cos(s)},\ \frac{a_1\sin(s)+a_2\cos(s)}{\frac{5}{3}-\frac{2}{3}\cos(s)}\right].$$

The row sums of $\boldsymbol{C}$ are $-1,-2/3,-4/3$. Let us first consider the renewal process with interarrival density $\boldsymbol{b}\mathrm{e}^{\boldsymbol{C}x}\boldsymbol{e}$ where $\boldsymbol{b}=(3,-1,-1)$. Here $\boldsymbol{a}_1=\boldsymbol{a}_2=\boldsymbol{a}_3=\boldsymbol{b}$. The corresponding

D-matrix is

$$\boldsymbol{D}=\begin{pmatrix}\boldsymbol{b}\\ 2/3\boldsymbol{b}\\ 4/3\boldsymbol{b}\end{pmatrix}=\begin{pmatrix}3 & -1 & -1\\ 2 & -2/3 & -2/3\\ 4 & -4/3 & -4/3\end{pmatrix}.$$

When there are jumps, then no matter where in the orbit the jump takes place, the process will return to the original point $(3,-1,-1)$.

It is a non-trivial matter to construct non-trivial RAPs but a safe way is to start with a renewal process and perturb the chosen renewal point $\boldsymbol{b}$ in a close neighborhood.

Take as our candidate for a non-renewal RAP the following D-matrix:

$$\boldsymbol{D}=\begin{pmatrix}\frac{14}{5} & -\frac{9}{10} & -\frac{9}{10}\\ \frac{26}{15} & -\frac{8}{15} & -\frac{8}{15}\\ \frac{58}{15} & -\frac{19}{15} & -\frac{19}{15}\end{pmatrix},$$

which is obtained by replacing $\boldsymbol{b}$ with the points

$$(2.8,-0.9,-0.9),$$
$$(2.6,-0.8,-0.8),$$
$$(2.9,-0.95,-0.95).$$

When a jump occurs from some point $(1-x-y,x,y)$, it goes to

$$\left(\frac{\frac{14}{5}-\frac{16}{15}x+\frac{16}{15}y}{1-\frac{1}{3}x+\frac{1}{3}y},\frac{-\frac{9}{10}+\frac{11}{30}x-\frac{11}{30}y}{1-\frac{1}{3}x+\frac{1}{3}y},\frac{-\frac{9}{10}+\frac{11}{30}x-\frac{11}{30}y}{1-\frac{1}{3}x+\frac{1}{3}y}\right).$$

Thus the second and third coordinates always coincide, and all possible jumps from an orbit will hence form a straight line with slope $+1$ in the plane of the second and third coordinates. The initial point for any orbit will hence be of the form $(1-2a,a,a)$. If $\boldsymbol{a}$ is the initial point after a jump, then the next jump, after time s, say, will go to

$$\frac{\boldsymbol{a}\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{D}}{\boldsymbol{a}\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{D}\boldsymbol{e}}=(1-2y(s),y(s),y(s)),$$

where

$$y(s)=-\frac{27-18a+18a\cos(s)+22a\sin(s)}{30-20a+20a\cos(s)+20a\sin(s)}.$$

In Fig. 2 all possible jump points are plotted for all possible orbits and we see that they are contained in the area surrounded by the original orbit through $(-1,-1)$ (second and third coordinates), which we henceforth define as $\mathscr{A}'$ and which coincides with $\mathscr{A}$. Indeed, from the orbit through $(-1,-1)$ ($a=-1$ in Fig. 2) the second and third coordinates of the possible jump points, which are equal, are between about -0.74 and -0.95. Now taking any point $(1-2a,a,a)$ as starting value for a new orbit with a in the interval $[-0.95,-0.74]$, we see that new possible jump points will all be contained in this same interval again. Hence property (b) in 7 holds.

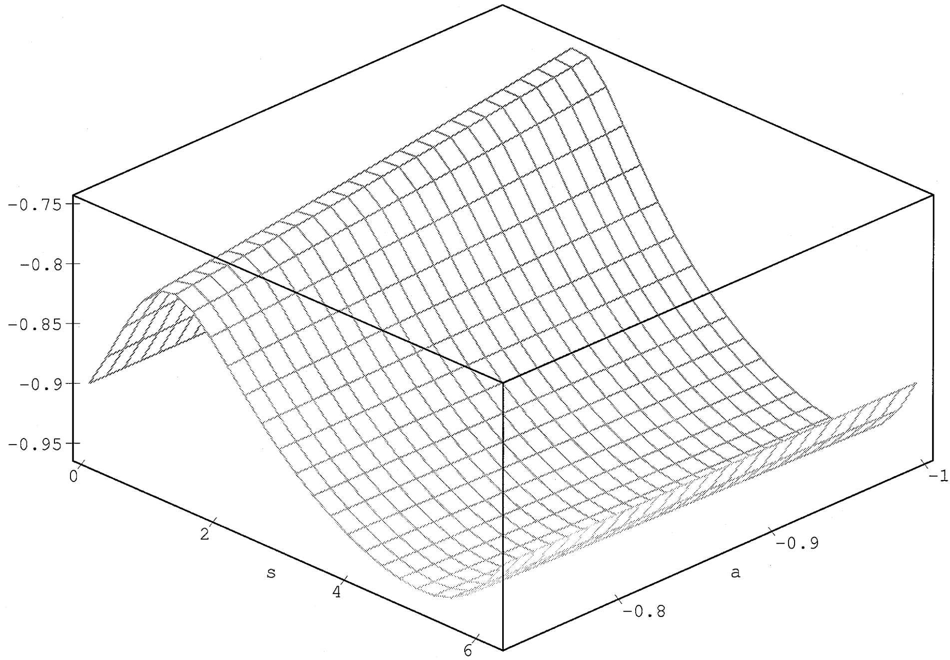

Fig. 2. Point reached by jumps when a process that started in (a,a) jumps at time s.

4. Further properties of RAPs

We consider a RAP N with $\boldsymbol{\alpha}, \boldsymbol{C}, \boldsymbol{D}$ as constructed in Section 2 and will show some further analytic similarities with the MAP which do not directly follow from the results presented so far.

Proposition 4.1. *$\mathbb{E}\mathrm{e}^{\theta N(t)}$ exists for all $\theta > 0$ and is given by $\mathbb{E}\mathrm{e}^{\theta N(t)} = \boldsymbol{\alpha}\mathrm{e}^{t(\boldsymbol{C}+\mathrm{e}^{\theta}\boldsymbol{D})}\boldsymbol{e}$.*

Proof. Clearly, $\{N(t)\}$ is stochastically bounded by a Poisson process with intensity $\max_{\boldsymbol{a}\in\mathscr{A}} \boldsymbol{aDe}$ so $\mathbb{E}\mathrm{e}^{\theta N(t)}$ exists for all $\theta > 0$. Now up to $o(h)$ terms,

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{A}(t+h)\mathrm{e}^{\theta N(t+h)}|\mathscr{F}_t] \\
&= \mathrm{e}^{\theta N(t)}\left\{\frac{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}}{\boldsymbol{A}(t)\mathrm{e}^{\boldsymbol{C}h}\boldsymbol{e}}(1-\boldsymbol{A}(t)\boldsymbol{De}\,h) + \frac{\boldsymbol{A}(t)\boldsymbol{D}\mathrm{e}^{\theta}}{\boldsymbol{A}(t)\boldsymbol{De}}\boldsymbol{A}(t)\boldsymbol{De}\,h\right\} \\
&= \mathrm{e}^{\theta N(t)}\{\boldsymbol{A}(t)(\boldsymbol{I}+\boldsymbol{C}h)(1-(\boldsymbol{A}(t)\boldsymbol{Ce}+\boldsymbol{A}(t)\boldsymbol{De})\,h) + \boldsymbol{A}(t)\boldsymbol{D}\mathrm{e}^{\theta}h\}.
\end{aligned}$$

Using $\boldsymbol{Ce}+\boldsymbol{De}=\boldsymbol{Qe}=\boldsymbol{0}$ and letting $\boldsymbol{b}(t)=\mathbb{E}[\boldsymbol{A}(t)\mathrm{e}^{\theta N(t)}]$, we get $\boldsymbol{b}'(t)=\boldsymbol{b}(t)(\boldsymbol{C}+\boldsymbol{D}\mathrm{e}^{\theta})$ so that $\boldsymbol{b}(t)=\boldsymbol{A}(0)\mathrm{e}^{t(\boldsymbol{C}+\mathrm{e}^{\theta}\boldsymbol{D})}$. Now just note that $\mathbb{E}\mathrm{e}^{\theta N(t)}=\boldsymbol{b}(t)\boldsymbol{e}$ and $\boldsymbol{\alpha}=\boldsymbol{A}(0)$. □

In the rest of this section, we shall for simplicity work in part subject to

Condition 4.1. *Any right eigenvector of $\boldsymbol{Q}$ corresponding to an eigenvalue λ with $\Re\,\lambda = 0$ is proportional to $\boldsymbol{e}$.*

In the MAP setting, this simply means that the background Markov process $\{J(t)\}$ is ergodic. Condition 4.1 and Proposition 2.1 implies that there exists a *unique* left eigenvector $\boldsymbol{\alpha}^*$ corresponding to $\lambda = 0$, which we normalize by $\boldsymbol{\alpha}^* \boldsymbol{e} = 1$.

Lemma 4.1. *Assume that Condition* 4.1 *holds. Then*
(a) *the algebraic multiplicity of the eigenvalue* 0 *is* 1;
(b) $\mathrm{e}^{\boldsymbol{Q}t} = \boldsymbol{e}\boldsymbol{\alpha}^* + \mathrm{O}(\mathrm{e}^{-\varepsilon t})$ *for some* $\varepsilon > 0$;
(c) $\boldsymbol{\alpha}^* \in \mathscr{A}$.

Proof. (a) Assume that the algebraic multiplicity of 0 is at least 2. Then there exists $\boldsymbol{f}$ such that $\boldsymbol{Q}\boldsymbol{f} = \boldsymbol{e}$, and hence $\mathrm{e}^{\boldsymbol{Q}t}\boldsymbol{f} = \boldsymbol{f} + t\boldsymbol{e}$. Then

$$\mathbb{E}\boldsymbol{A}(t)\boldsymbol{f} = \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}t}\boldsymbol{f} = \boldsymbol{\alpha}\boldsymbol{f} + t$$

which is impossible since $\boldsymbol{A}(t)\boldsymbol{f}$ is bounded.

(b) is a standard consequence of (a), Condition 4.1 and Proposition 2.1. For (c), note that $\int_0^t \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s/t = \int_0^t \mathbb{E}\boldsymbol{A}(s)\,\mathrm{d}s/t \in \mathscr{A}$ since $\mathscr{A}$ is convex and compact. Hence we can find a subsequence $\{t_k\}$ with a limit in $\mathscr{A}$. But by (b), such a limit must be $\boldsymbol{\alpha}^*$. □

Proposition 4.2. *Assume that Condition* 4.1 *holds. Then*

$$\mathbb{E}N(t) = t \cdot \boldsymbol{\alpha}^*\boldsymbol{D}\boldsymbol{e} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)(\boldsymbol{I} - \mathrm{e}^{\boldsymbol{Q}t})(\boldsymbol{e}\boldsymbol{\alpha}^* - \boldsymbol{Q})^{-1} \tag{4.1}$$

$$= t \cdot \boldsymbol{\alpha}^*\boldsymbol{D}\boldsymbol{e} + (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)(\boldsymbol{e}\boldsymbol{\alpha}^* - \boldsymbol{Q})^{-1} + \mathrm{O}(\mathrm{e}^{-\varepsilon t}). \tag{4.2}$$

Proof. Obviously,

$$\mathbb{E}N(t) = \mathbb{E}\int_0^t \boldsymbol{A}(s)\mathrm{e}^{\boldsymbol{Q}s}\boldsymbol{D}\boldsymbol{e}\,\mathrm{d}s = \boldsymbol{\alpha}\int_0^t \mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s\,\boldsymbol{D}\boldsymbol{e}.$$

(4.1) follows then by inserting the expression

$$\int_0^t \mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s = \boldsymbol{e}\boldsymbol{\alpha}^* t + (\boldsymbol{I} - \mathrm{e}^{\boldsymbol{Q}t})(\boldsymbol{e}\boldsymbol{\alpha}^* - \boldsymbol{Q})^{-1}$$

(see Eq. (20) of Narayana and Neuts (1992); Lemma 4.1 ensures that the conditions of Narayana and Neuts (1992) are satisfied) and noting that $\boldsymbol{\alpha} = \boldsymbol{\alpha}^* + \boldsymbol{\beta}$ where $\boldsymbol{\beta}\boldsymbol{e} = 0$. Using Lemma 4.1(b) then gives (4.2). □

Next consider Palm theory, i.e. the question of existence of versions N^*, N_0 of N which are time- (resp. event) stationary, and of the relation between N^* and the Palm version N_0 (see e.g. Franken et al. (1982), Sigman (1994) or Baccelli and Bremaud (1994) for the general background). For any $\boldsymbol{\beta} \in \mathscr{A}$, we denote by $N_{\boldsymbol{\beta}}$ the RAP obtained by starting $\{\boldsymbol{A}(t)\}$ with $\boldsymbol{A}(0) = \boldsymbol{\beta}$, and by $\mathbb{P}_{\boldsymbol{\beta}}$ the corresponding probability measure. We first give a general existence result, and next impose Condition 4.1 to ensure uniqueness and some limiting properties.

Proposition 4.3. *For any* $\boldsymbol{\alpha}^* \in \mathscr{A}$, $N_{\boldsymbol{\alpha}^*}$ *is time-stationary if and only if* $\boldsymbol{\alpha}^*\boldsymbol{Q} = \boldsymbol{0}$. *Then the Palm version is* $N_{\boldsymbol{\alpha}_0}$ *where* $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}^*\boldsymbol{D}/\boldsymbol{\alpha}^*\boldsymbol{D}\boldsymbol{e}$. *Furthermore, such a* $\boldsymbol{\alpha}^*$ *always exists.*

Proof. Since $\mathbb{P}_{\boldsymbol{\alpha}^*}(\theta_t N\in\cdot)=\mathbb{P}_{\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{Q}t}}$ and $\mathbb{P}_{\boldsymbol{\alpha}}\neq\mathbb{P}_{\boldsymbol{\beta}}$ for $\boldsymbol{\alpha}\neq\boldsymbol{\beta}$, $N_{\boldsymbol{\alpha}^*}$ is time-stationary if and only if $\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{Q}t}=\boldsymbol{\alpha}^*$ for all t, i.e. if and only if $\boldsymbol{\alpha}^*\boldsymbol{Q}=\boldsymbol{0}$.

By general point process theory, the Palm version is then determined by

$$\begin{aligned}\mathbb{P}_0(N\in\cdot)&=\frac{1}{\mathbb{E}_{\boldsymbol{\alpha}^*}N(h)}\mathbb{E}_{\boldsymbol{\alpha}^*}\sum_{i:T_i\leqslant h}I(\theta_{T_i}N\in\cdot)\\&=\frac{1}{\mathbb{E}_{\boldsymbol{\alpha}^*}N(h)}\{\mathbb{P}_{\boldsymbol{\alpha}^*}(\theta_{T_1}N\in\cdot;T_1\leqslant h)+\mathrm{o}(h)\}\\&=\frac{1}{\boldsymbol{\alpha}^*\boldsymbol{De}\,h}\int_0^h\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{De}\,\mathbb{P}_{\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{D}/\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{C}s}\boldsymbol{De}}\,\mathrm{d}s+\mathrm{o}(1)\\&=\mathbb{P}_{\boldsymbol{\alpha}^*\boldsymbol{D}/\boldsymbol{\alpha}^*\boldsymbol{De}}+\mathrm{o}(1)=\mathbb{P}_{\boldsymbol{\alpha}_0}+\mathrm{o}(1).\end{aligned}$$

Thus $\mathbb{P}_0=\mathbb{P}_{\boldsymbol{\alpha}_0}$.

For existence of $\boldsymbol{\alpha}^*$, note first that as above we can find $\boldsymbol{\alpha}^*\in\mathscr{A}$ as limit of $\int_0^{t_k}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s/t_k$. Then

$$\begin{aligned}\boldsymbol{\alpha}^*\mathrm{e}^{\boldsymbol{Q}h}&=\lim_{k\to\infty}\frac{1}{t_k}\int_0^{t_k}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}(s+h)}\,\mathrm{d}s=\lim_{k\to\infty}\frac{1}{t_k}\int_h^{t_k+h}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s\\&=\lim_{k\to\infty}\frac{1}{t_k}\int_0^{t_k}\boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}s}\,\mathrm{d}s=\boldsymbol{\alpha}^*\end{aligned}$$

for all h which implies $\boldsymbol{\alpha}^*\boldsymbol{Q}=\boldsymbol{0}$. □

Proposition 4.4. *For any $\boldsymbol{\alpha}_0\in\mathscr{A}$, $N_{\boldsymbol{\alpha}_0}$ is event-stationary if and only if $\boldsymbol{\alpha}_0(-\boldsymbol{C}^{-1}\boldsymbol{D})=\boldsymbol{\alpha}_0$. Then the corresponding time-stationary version is $N^*=N_{\boldsymbol{\alpha}^*}$ where $\boldsymbol{\alpha}^*=\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}/\boldsymbol{\alpha}\boldsymbol{C}^{-1}\boldsymbol{e}$. Furthermore, such a $\boldsymbol{\alpha}_0$ always exists.*

Proof. The first claim follows since $\boldsymbol{A}(t)=\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{D}/\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{De}$ given $T_1=t$ and the density of T_1 is $\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{De}$ so that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\alpha}_0}\boldsymbol{A}(T_1)&=\int_0^\infty\frac{\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{D}}{\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{De}}\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{De}\,\mathrm{d}t\\&=\int_0^\infty\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\boldsymbol{D}\,\mathrm{d}t=-\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}\boldsymbol{D}.\end{aligned}$$

The general point process theory then gives that $N^*=N_{\boldsymbol{\alpha}^*}$ where

$$\begin{aligned}\boldsymbol{\alpha}^*&=\frac{1}{\mathbb{E}_{\boldsymbol{\alpha}_0}T_1}\mathbb{E}_{\boldsymbol{\alpha}_0}\int_0^{T_1}\boldsymbol{A}(t)\,\mathrm{d}t=\frac{1}{-\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}\boldsymbol{e}}\int_0^\infty\mathbb{E}_{\boldsymbol{\alpha}_0}[\boldsymbol{A}(t);T_1>t]\,\mathrm{d}t\\&=\frac{1}{-\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}\boldsymbol{e}}\int_0^\infty\boldsymbol{\alpha}_0\mathrm{e}^{\boldsymbol{C}t}\,\mathrm{d}t=\frac{\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}}{\boldsymbol{\alpha}_0\boldsymbol{C}^{-1}\boldsymbol{e}}.\qquad\square\end{aligned}$$

Proposition 4.5. *Assume that Condition* 4.1 *holds. Then*

(a) *the solution of $\boldsymbol{\alpha}^*\in\mathscr{A}$, $\boldsymbol{\alpha}^*\boldsymbol{Q}=\boldsymbol{0}$ is unique*;

(b) *for $\boldsymbol{\beta}\in\mathscr{A}$, $N_{\boldsymbol{\beta}}$ is time-stationary if and only if $\boldsymbol{\beta}=\boldsymbol{\alpha}^*$*;

(c) $\theta_t N\xrightarrow{\mathscr{D}}N_{\boldsymbol{\alpha}^*},\ t\to\infty.$

Proof. (a) and (b) are easy. For (c), note that $\theta_t N \stackrel{\mathscr{D}}{=} N_{\boldsymbol{\alpha}^{(t)}}$ where $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}\mathrm{e}^{\boldsymbol{Q}t}$. But by Lemma 4.1, $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}\boldsymbol{e}\boldsymbol{\alpha}^* + \mathrm{o}(1) = \boldsymbol{\alpha}^* + \mathrm{o}(1)$. This implies $f_{N_{\boldsymbol{\alpha}^{(t)}},n}(x_1,\ldots,x_n) \to f_{N_{\boldsymbol{\alpha}^*},n}(x_1,\ldots,x_n)$ which by general point process theory is sufficient for $N_{\boldsymbol{\alpha}^{(t)}} \stackrel{\mathscr{D}}{\to} N_{\boldsymbol{\alpha}^*}$. □

It is reasonable to ask whether also $\theta_{T_k} N \stackrel{\mathscr{D}}{\to} N_{\boldsymbol{\alpha}_0}$, $k \to \infty$. However, even for a MAP this is not the case: here $-\boldsymbol{C}^{-1}\boldsymbol{D}$ is a transition matrix which may be periodic even when $\boldsymbol{Q}$ is ergodic. But:

Proposition 4.6. *Assume in addition to Condition* 4.1 *that the only eigenvalue* λ *of* $-\boldsymbol{C}^{-1}\boldsymbol{D}$ *with* $|\lambda| = 1$ *is* $\lambda = 1$. *Then* $\theta_{T_k} N \stackrel{\mathscr{D}}{\to} N_{\boldsymbol{\alpha}_0}$.

Proof. If $\boldsymbol{h}$ is a right eigenvector of $-\boldsymbol{C}^{-1}\boldsymbol{D}$ corresponding to an eigenvalue λ with $|\Re\, \lambda| > 1$, then for $\boldsymbol{\alpha} \in \mathscr{A}$

$$\mathbb{E}_{\boldsymbol{\alpha}} \boldsymbol{A}(T_n)\boldsymbol{h} = \boldsymbol{\alpha}(-\boldsymbol{C}^{-1}\boldsymbol{D})^n \boldsymbol{h} = \lambda^n \boldsymbol{\alpha}\boldsymbol{h}$$

is by compactness only possible if $\boldsymbol{\alpha}\boldsymbol{h} = 0$.

Thus the restriction of $-\boldsymbol{C}^{-1}\boldsymbol{D}$ to $\mathscr{A}$ has 1 as a simple eigenvector with corresponding left eigenvector $\boldsymbol{\alpha}_0$ and right eigenvector $\boldsymbol{e}$, and any other eigenvalue λ must have $|\Re\, \lambda| < 1$. Thus

$$\mathbb{E}_{\boldsymbol{\alpha}} \boldsymbol{A}(T_n) = \boldsymbol{\alpha}\boldsymbol{e}\boldsymbol{\alpha}_0 + \mathrm{o}(1) = \boldsymbol{\alpha}_0 + \mathrm{o}(1)$$

which implies the assertion. □

References

Asmussen, S., 1987. Applied Probability and Queues. Wiley, Chichester.

Asmussen, S., Bladt, M., 1996. Renewal theory and queueing algorithms for matrix-exponential distributions. In: Alfa, A.S., Chakravarty, S. (Eds.), Matrix-Analytic Methods in Stochastic Models. Marcel Dekker, New York, pp. 313–341.

Asmussen, S., Koole, G., 1963. Marked point processes as limits of Markovian arrival streams. J. Appl. Probab. 30, 365–372.

Asmussen, S., O'Cinneide, C.A., 1998. Matrix-exponential distributions [Distributions with a rational Laplace transform]. In: Kotz, S., Read, C.B., Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Supplementary Volume. Wiley, New York.

Baccelli, F., Bremaud, P., 1994. Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurrences. Springer, New York.

Cohen, J.W., 1982. The Single Server Queue, second ed. North-Holland, Amsterdam.

Cox, D.R., 1955. Use of complex probabilities in the theory of stochastic processes. Proc. Cambridge Philos. Soc. 51, 313–319.

Davis, M.H.A.D., 1993. Markov Processes and Optimization. Chapman & Hall, London.

Franken, P., König, D., Arndt, U., Schmidt, V., 1982. Queues and Point Processes. Wiley, Chichester.

Kallenberg, O., 1983. Random Measures. Akademie-Verlag, Berlin, and Academic Press, London.

Knight, F.B., 1981. Essays on the Prediction Process, IMS Lecture Notes series. The Institute of Mathematical Statistics, Hayward, CA, vol. 1.

Knight, F.B., 1992. Foundations of the Prediction Process. Clarendon Press, Oxford.

König, D., Schmidt, V., 1992. Random Point Processes. Teubner, Stuttgart (in German).

Lipsky, L., 1992. Queueing Theory: A Linear Algebraic Approach. Macmillan, New York.

Lucantoni, D., 1991. New results on the single server queue with a batch Markovian arrival process. Stochastic Models 7, 1–41.

Narayana, S., Neuts, M.F., 1992. The first two moment matrices of the counts for the Markovian arrival process. Stochastic Models 8, 459–477.

Neuts, M.F., 1979. A versatile Markovian point process. J. Appl. Probab. 16, 764–779.

Neuts, M.F., 1981. Matrix-Geometric Solutions in Stochastic Models. Johns Hopkins University Press, Baltimore.

Neuts, M.F., 1992. Models based on the Markovian arrival process. IEICE Trans. Commun. E-75, 1255–1265.

Neuts, M.F., 1995. Matrix-analytic methods in stochastic models. In: Dshalalow, J. (Ed.), Advances in Queueing. CRC Press, Boca Raton, pp. 265–293.

O'Cinneide, C.A., 1989. Characterization of phase-type distributions. Stochastic Models 6, 1–57.

Ramaswami, V., 1995. Matrix-analytic methods: a tutorial overview with some extensions and new results. In: Chakravarty, S., Alfa, A.S. (Eds.), Matrix-Analytic Methods in Stochastic Models. Marcel Dekker, New York.

Sigman, K., 1994. Stationary Marked Point Processes: An Intuitive Approach. Chapman & Hall, New York.

Smith, W.L., 1953. Distribution of queueing times. Proc. Cambridge Philos. Soc. 49, 449–461.