The Annals of Applied Probability
2003, Vol. 13, No. 2, 628–640

TRANSIENT MARKOV ARRIVAL PROCESSES

BY GUY LATOUCHE, MARIE-ANGE REMICHE[1] AND PETER TAYLOR

Université Libre de Bruxelles, Université Libre de Bruxelles and University of Melbourne

We define the family of transient Markov arrival processes (transient MAPs) which combine features of transient (or terminating) renewal processes and of the well-known MAPs: transient MAPs are point processes on the line, controlled by a finite Markov chain, which almost surely comprise a finite number of points. We analyze their basic properties.

1. Introduction. Markov arrival processes (abbreviated MAPs) are counting processes controlled by a discrete-time or continuous-time Markov chain. In continuous time, a MAP is generated by a two-dimensional Markov chain $\{(N(t), \varphi(t)) : t \in \mathbb{R}^+\}$ on the state space $\{(n, i) : n \in \mathbb{N},\ i \in \{1, \dots, m < \infty\}\}$.

From a given state (n, i), the only possible one-step transitions are to the states $\{(n, j) : 1 \le i \ne j \le m\}$ and to the states $\{(n + 1, j) : 1 \le j \le m\}$. The transition rates from state (n, i) are independent of n. Thus, in continuous time, the rate matrix has the structure

$$Q = \begin{bmatrix} D_0^* & D_1^* & 0 & 0 & \\ 0 & D_0^* & D_1^* & 0 & \ddots \\ 0 & 0 & D_0^* & D_1^* & \ddots \\ 0 & 0 & 0 & D_0^* & \ddots \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{1}$$

with $(D_0^*)_{ij}$, $i \ne j$, and $(D_1^*)_{ij}$, respectively being the instantaneous transition rates from (n, i) to (n, j) and from (n, i) to $(n + 1, j)$. The diagonal entries of D_0^* are strictly negative, with the row sums of Q equal to 0:

$$(D_0^*)_{ii} = -\left\{ \sum_{\substack{1 \le j \le m \\ j \ne i}} (D_0^*)_{ij} + \sum_{1 \le j \le m} (D_1^*)_{ij} \right\}$$

for $1 \le i \le m$.

In discrete time, the transition probability matrix has the structure (1), where D_0^* and D_1^* are nonnegative and $D_0^* + D_1^*$ is stochastic. The theory of discrete-time

Received May 2001; revised June 2002.

[1]Research supported in part by the EC Marie Curie Program Grant HPMF-CT-1999-00024.

AMS 2000 subject classifications. Primary 60J22; secondary 60K37.

Key words and phrases. Markovian arrival process, Markov additive process, transient process, terminating process, quasistationary distribution.

MAPs is similar to that of continuous-time MAPs and we restrict ourselves to the latter in the present paper.

The component $\varphi(t)$ is called the *phase* and $\{\varphi(t) : t \in \mathbb{R}^+\}$ is a Markov process on the state space $\{1, \dots, m\}$ with rate matrix $D^* = D_0^* + D_1^*$. The component $N(t)$ is called the *level*: If $N(0) = 0$, then $N(t)$ is the number of *events* which occur in the interval $(0, t)$ and we refer to the process $\{N(t)\}$ as a MAP. One may think of $\{\varphi(t)\}$ as driving the MAP. There are two kinds of transitions from phase i to phase j, with and without an accompanying event, and the instantaneous rates of such transitions are $(D_1^*)_{ij}$ and $(D_0^*)_{ij}$, respectively. In addition, events may happen even when the phase does not change. This occurs at the constant rate $(D_1^*)_{ii}$ when the phase is i.

Special examples of MAPs are the Poisson process, which can be constructed by taking $m = 1$, the Markov modulated Poisson process, for which D_1^* is a diagonal matrix, and the phase-type (PH) renewal process, for which D_1^* is a matrix of rank 1. Other examples abound in the literature. An extensive treatment may be found in Neuts [(1989), Chapter 10], and also in Lucantoni, Meier-Hellstern and Neuts (1990), Neuts (1979) and Pacheco and Prabhu (1995).

Given that $N(0) = 0$, the process is characterized by the distribution of $\varphi(0)$ in addition to D_0^* and D_1^*. We say that $\{N(t) : t \in \mathbb{R}^+\}$ is MAP($\boldsymbol{\alpha}$, D_0^*, D_1^*), where $\boldsymbol{\alpha}$ is the row vector such that $\alpha_i = P[\varphi(0) = i]$. We use lowercase boldface letters to denote vectors, both row and column.

In the previous literature, it has been assumed that D^* is irreducible and so the phase process $\{\varphi(t)\}$ has a stationary distribution $\boldsymbol{\pi}$ which is the unique solution of the system $\boldsymbol{\pi} D^* = \mathbf{0}$, $\boldsymbol{\pi}\mathbf{1} = 1$. The process MAP($\boldsymbol{\pi}$, D_0^*, D_1^*) is called the *stationary version* of the MAP generated by D_0^* and D_1^*. Its most interesting property is that it has stationary increments, since $\{N(t) - N(t_0) : t \geq t_0\}$ is also MAP($\boldsymbol{\pi}$, D_0^*, D_1^*) for any given t_0.

Our purpose in this paper is to define *transient* MAPs. These are MAPs for which (almost surely) $\lim_{t\to\infty} N(t) < \infty$. In the next section we give a precise definition of a transient MAP. The packet stream model defined in Ramaswami and Latouche (1989), and analyzed in Neuts (1990) and Latouche and Ramaswami (1992), is an example of a transient MAP. We give three other illustrative examples in Section 2. In Section 3 we determine the distribution of the lifetime of a transient MAP and in Section 4 we determine the distribution of the total number of events. In Section 5, we extend the notion of a stationary version to a transient MAP. Some concluding comments are made in Section 6.

Our definition of transient MAPs is motivated by that of PH random variables. A random variable is said to be PH if it has the same distribution as the time to absorption for a Markov process with finitely many transient states and one absorbing state. It is characterized by the pair $(\boldsymbol{\tau}, T)$, where $\boldsymbol{\tau}$ is the initial probability vector and T is the matrix which describes the transitions among the transient states. For further details, see Neuts (1981) and Latouche and Ramaswami (1999).

We shall deal with both continuous and discrete-time PH distributions. In order to avoid any ambiguity, we use the notation $\mathrm{PH_c}(\cdot,\cdot)$ for the former and $\mathrm{PH_d}(\cdot,\cdot)$ for the latter. The distribution function of a $\mathrm{PH_c}(\boldsymbol{\tau}, T)$ random variable is

$$F(x) = 1 - \boldsymbol{\tau} \exp(T)\mathbf{1} \qquad \text{for } x \text{ in } \mathbb{R}_+$$

and that of a $\mathrm{PH_d}(\boldsymbol{\tau}, T)$ random variable is

$$F(n) = 1 - \boldsymbol{\tau} T^n \mathbf{1} \qquad \text{for } n \text{ in } \mathbb{N}.$$

2. Transient MAPs. As mentioned above, it has previously been assumed that D^* is irreducible. Except in the trivial case where $D_1^* = 0$, this implies that, with probability one, there are infinitely many events in the MAP. The same conclusion can be reached if D^* is reducible and the subblocks of D_1^* corresponding to the recurrent classes of D^* are nonzero. However, if D^* has a recurrent class for which the corresponding subblock of D_1^* is zero, then there is a possibility that the corresponding MAP has only finitely many events. In this paper we shall concentrate on the case where D^* has precisely one absorbing state and all other states are transient. The generalization to more complicated cases are obvious.

Assume that there exists an absorbing phase 0 such that, for all j, $(D_1^*)_{0j} = (D_0^*)_{0j} = 0$ and that, in the matrix D^*, there exists a path of positive rate from all other phases to phase 0. Thus, if $\varphi(t_0) = 0$ for some t_0, then $\varphi(t) = 0$ and $N(t) = N(t_0)$ for all $t \geq t_0$ and the process ceases to evolve once phase 0 is reached. We shall say that a *catastrophe* occurs when the process moves into a state of the form $(n, 0)$.

A *transient MAP* is the counting process $\{N(t)\}$ generated by a two-dimensional Markov process $\{(N(t), \varphi(t)) : t \in \mathbb{R}^+\}$ on the state space $\{(n, i) : n \in \mathbb{N}, i \in \{0, \dots, m\}\}$ with transition matrix of the form (1), where

$$D_0^* = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{d}_0 & D_0 \end{bmatrix} \quad \text{and} \quad D_1^* = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{d}_1 & D_1 \end{bmatrix}$$

with $D_1 \geq 0$, $\boldsymbol{d}_0, \boldsymbol{d}_1 \geq \mathbf{0}$, the off-diagonal entries of D_0 are nonnegative and the diagonal entries are strictly negative so that the row sums of Q are equal to zero, that is,

$$D_0\mathbf{1} + D_1\mathbf{1} + \boldsymbol{d}_0 + \boldsymbol{d}_1 = \mathbf{0}. \tag{2}$$

Thus, in addition to the traditional blocks D_0 and D_1, we introduce the rates $\boldsymbol{d}_0$ and $\boldsymbol{d}_1$ at which the catastrophe occurs. Clearly, every state of the form $(n, 0)$ is absorbing.

The transient MAP is characterized by the 4-tuple $(\boldsymbol{\alpha}, D_0, D_1, \boldsymbol{d}_0)$, where $\boldsymbol{\alpha} = (\alpha_i)$ is a vector of size m, with α_i denoting the probability that $\varphi(0) = i$. The probability α_0 that $\varphi(0) = 0$ is implicitly defined by $\alpha_0 = 1 - \boldsymbol{\alpha}\mathbf{1}$ and the vector $\boldsymbol{d}_1$ is defined by (2).

The epochs $\{T_n : n \in \mathbb{N}\}$ at which events of the MAP occur are defined by $T_0 = 0$ and $T_n = \inf\{t \geq 0 : N(t) = n\}$ if such a t exists, otherwise $T_n = \infty$.

Assume that the following conditions are satisfied.

CONDITION 2.1. The matrix $D = D_0 + D_1$ is nonsingular.

This condition implies that states of the form (n, i) with $i \neq 0$ are transient and that, almost surely, the process will enter one of the absorbing states $(n, 0)$ in finite time. Indeed, the transition rate matrix of the process $\{\varphi(t)\}$ is

$$D^* = \begin{bmatrix} 0 & \mathbf{0} \\ \boldsymbol{d}_0 + \boldsymbol{d}_1 & D \end{bmatrix}.$$

The phase 0 is clearly absorbing and the phases 1 to m are all transient if and only if D is nonsingular [Latouche and Ramaswami (1999), Theorem 2.4.3].

Without loss of generality, we can also impose the following condition:

CONDITION 2.2. For every phase $i \neq 0$ there exists at least one j such that there is a path from (n, i) to $(n + 1, j)$.

If this condition is not satisfied, then there are phases $i \neq 0$ such that no more points of the transient MAP are observed once the process moves into a state of the form (n, i). An equivalent transient MAP can be defined by lumping these phases i together with the absorbing phase 0.

We now give three examples of transient MAPs.

EXAMPLE 2.3. Our first example is a standard MAP $\{M(t)\}$ with representation $(\boldsymbol{\beta}, S_0, S_1)$ observed during a random interval $(0, W)$ which has a continuous $\mathrm{PH_c}(\boldsymbol{\tau}, T)$ distribution. The points of $\{N(t)\}$ are precisely the points of $\{M(t)\}$ that occur before time W.

Assume $t < T$. Then the Markov chain associated with W is in some phase $\zeta(t)$ and the Markov chain associated with $\{M(t)\}$ is in some phase $\xi(t)$. To define the transient MAP $\{N(t)\}$, we need to know both of these pieces of information, so we define the phase associated with $\{N(t)\}$ to be the ordered pair $\varphi(t) = (\xi(t), \zeta(t))$. The initial distribution of $\varphi(t)$ is given by $\boldsymbol{\alpha} = \boldsymbol{\beta} \otimes \boldsymbol{\tau}$, where $\otimes$ denotes the Kronecker product. Moreover, $\varphi(t)$ can change because of a change in $\xi(t)$ or because of a change in $\zeta(t)$. These events occur independently, so the rate matrices of $\{N(t)\}$ are given by $D_0 = S_0 \otimes I + I \otimes T$ and $D_1 = S_1 \otimes I$. The matrix D_0 contains rates of transitions between phases that are not associated with an event of $\{M(t)\}$ and the matrix D_1 contains the rates of transitions that do generate an event of $\{M(t)\}$.

At time point W, $\zeta(t)$ moves to the absorbing state corresponding to W. This coincides with the catastrophe, when $\varphi(t)$ moves to the state 0 and ceases to evolve thereafter. Thus $\boldsymbol{d}_0 = \mathbf{1} \otimes (-T\mathbf{1})$. Almost surely, the catastrophe occurs at a time which is not a point of $\{N(t)\}$, a property which is reflected in the fact that $\boldsymbol{d}_1 = \mathbf{0}$.

EXAMPLE 2.4. A transient MAP can also be defined such that events stop being counted on the basis of a circumstance related to the process itself.

Consider the $M/M/1$ queue and count the number of arrivals until the queue size exceeds m for the first time. The phase $\varphi(t)$ is the state of the queue at time t. An event is recorded at every arrival and the catastrophe occurs when there is an arrival with $\varphi(t) = m$.

In this process, the catastrophe is an event since it corresponds to an increase of $N(t)$.

EXAMPLE 2.5. As an example of a transient MAP in which the catastrophe might or might not be an event, consider a model for the history $\{N(t)\}$ of time points at which claims are made by an individual against a health insurance policy. Assume that claims occur at a rate which is dependent on the person's "underlying state of health," which we denote by $\varphi(t)$. This changes according to a continuous-time Markov chain, and points at which changes occur might or might not be associated with a claim against the policy. Rates of the former type are stored in the matrix D_1 and rates of the latter type are stored in the matrix D_0.

The catastrophe occurs when the person ceases to be insured by the company. This might happen at a time point where there is a claim, for example, a claim associated with a fatal illness, or it might occur at a point when there is no claim, for example, the person transferred to another health fund. The rates of these transitions are stored in the vectors $\boldsymbol{d}_1$ and $\boldsymbol{d}_0$, respectively.

3. Lifetime distributions. For a transient MAP, three random variables of basic interest are the lifetime L of the process, the time V until the catastrophe occurs and the total number K of events. These are defined as

$$L = T_K = \sup\{T_n : T_n < \infty\},$$
$$V = \inf\{t \geq 0 : \varphi(t) = 0\},$$
$$K = \lim_{t\to\infty} N(t).$$

We derive the distribution of the first two of these in this section. The derivation of the distribution of K is deferred until the next section.

If $\boldsymbol{d}_0 = \mathbf{0}$, then $L = V$ since the process can enter the absorbing phase only at a time of increase of $N(t)$. Otherwise, L and V have different distributions and $L \leq V$. We may then think of L as the last time that the process is "externally" observed to be alive.

By (1), $\{\varphi(t)\}$ is a Markov process on $\{0, 1, \dots, m\}$ with transition rate matrix $D^* = D_0^* + D_1^*$ and we readily conclude that V has a phase-type distribution. We state this as a lemma for future reference.

LEMMA 3.1. *For a transient* MAP$(\boldsymbol{\alpha}, D_0, D_1, \boldsymbol{d}_0)$ *such that Condition* 2.1 *holds*, *the time* V *until the catastrophe has a* $\mathrm{PH_c}(\boldsymbol{\alpha}, D)$ *distribution*.

The lifetime and the total number of events also have phase-type distributions as we show in Theorems 3.2 and 4.1. Before proving these, however, we need to establish a technical property.

Let us consider the process embedded at the epochs $\{V\} \cup \{T_k : k \geq 0\}$. This embedded process is a discrete-time Markov chain with possible transitions from (n, i) to $(n, 0)$ or $(n+1, j)$, so that the transition matrix has the form

$$P = \begin{bmatrix} H_0 & H_1 & 0 & 0 & \\ 0 & H_0 & H_1 & 0 & \ddots \\ & 0 & H_0 & H_1 & \ddots \\ & & 0 & H_0 & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \tag{3}$$

with

$$H_0 = \begin{bmatrix} 1 & \mathbf{0} \\ \boldsymbol{f}_0 & 0 \end{bmatrix} \quad \text{and} \quad H_1 = \begin{bmatrix} 1 & \mathbf{0} \\ \boldsymbol{f}_1 & F \end{bmatrix}.$$

By Gaver, Jacobs and Latouche [(1984), Lemma 1], we have that

$$F = (-D_0)^{-1} D_1, \tag{4}$$

$$\boldsymbol{f}_0 = (-D_0)^{-1} \boldsymbol{d}_0 \tag{5}$$

and

$$\boldsymbol{f}_1 = (-D_0)^{-1} \boldsymbol{d}_1. \tag{6}$$

Note that

$$F\mathbf{1} + \boldsymbol{f}_0 + \boldsymbol{f}_1 = \mathbf{1}. \tag{7}$$

It is immediate that Condition 2.2 is satisfied if and only if, for all i, $(\boldsymbol{f}_0)_i < 1$. This follows because $(\boldsymbol{f}_0)_i$ is the probability that, starting from a state of the form (n, i) with $i \neq 0$, the process enters $(n, 0)$ at some future time. There exists a path from (n, i) to the next level, or equivalently Condition 2.2 is satisfied, if and only if this probability is strictly less than 1.

Let $\boldsymbol{f}_+ = \mathbf{1} - \boldsymbol{f}_0$ and $\Delta = \operatorname{diag}(\boldsymbol{f}_+)$. The above observation guarantees that Δ is nonsingular.

THEOREM 3.2. *Consider a transient* MAP$(\boldsymbol{\alpha}, D_0, D_1, \boldsymbol{d}_0)$ *such that Condition* 2.2 *holds. The lifetime L of the process has a* $\mathrm{PH_c}(\boldsymbol{\alpha}\Delta, \Delta^{-1} D \Delta)$ *distribution. If $\boldsymbol{d}_0 = 0$, then L is identical to V and has a* $\mathrm{PH_c}(\boldsymbol{\alpha}, D)$ *distribution.*

PROOF. The second claim is obvious and we concentrate on the first.

The lifetime L is equal to 0 if and only if either $\varphi(0) = 0$ or $1 \leq \varphi(0) \leq m$ and the Markov process is absorbed in $(0, 0)$ before it visits level 1. Thus,

$$P[L = 0] = 1 - \boldsymbol{\alpha}\mathbf{1} + \boldsymbol{\alpha}\boldsymbol{f}_0 = 1 - \boldsymbol{\alpha}\boldsymbol{f}_+. \tag{8}$$

For $x > 0$, the event $[L > x]$ occurs if and only if at time x the process is in one of the transient phases, with distribution given by $\boldsymbol{\alpha}\exp(Dx)$, *and* an event occurs in the future with probability $\boldsymbol{f}_+$. Thus, $P[L > x] = \boldsymbol{\alpha}\exp(Dx)\boldsymbol{f}_+$ and

$$P[L \le x] = 1 - \boldsymbol{\alpha}\exp(Dx)\boldsymbol{f}_+. \tag{9}$$

It is a simple matter to verify that

$$\boldsymbol{\alpha}\exp(Dx)\boldsymbol{f}_+ = \boldsymbol{\alpha}\Delta\exp[\Delta^{-1}D\Delta x]\mathbf{1}$$

so that the proof of the theorem is complete once we show that $\boldsymbol{\alpha}\Delta$ is a (possibly defective) probability density on $\{1, \dots, m\}$ and $\Delta^{-1}D\Delta$ is a generator.

The first of these follows easily because $\mathbf{0} \le \boldsymbol{\alpha}\Delta \le \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}$ is a probability density on $\{1, \dots, m\}$. To get the second, observe that the off-diagonal entries clearly are nonnegative and, since

$$\begin{aligned} D\Delta\mathbf{1} = D\boldsymbol{f}_+ &= -\boldsymbol{d}_0 - \boldsymbol{d}_1 - D(-D_0)^{-1}\boldsymbol{d}_0 = -\boldsymbol{d}_1 - D_1(-D_0)^{-1}\boldsymbol{d}_0 \\ &= -\boldsymbol{d}_1 - D_1\boldsymbol{f}_0, \end{aligned}$$

the row sums are negative or zero. $\square$

4. Number of events. In this section, we determine the distribution of the total number K of events of a transient MAP.

THEOREM 4.1. *Consider a transient* MAP$(\boldsymbol{\alpha}, D_0, D_1, \boldsymbol{d}_0)$ *such that Condition* 2.2 *holds. The number K of events has a* $\mathrm{PH_d}(\boldsymbol{\alpha}\Delta, \Delta^{-1}F\Delta)$ *distribution. If* $\boldsymbol{d}_0 = \mathbf{0}$, *then K is* $\mathrm{PH_d}(\boldsymbol{\alpha}, F)$; *if* $\boldsymbol{d}_1 = \mathbf{0}$, *then K is* $\mathrm{PH_d}(\boldsymbol{\alpha}F, F)$.

PROOF. To determine the distribution of K, we use the discrete-time Markov chain with transition matrix (3).

Since $K = 0$ if and only if $L = 0$, we have by (8) that

$$P[K = 0] = 1 - \boldsymbol{\alpha}\boldsymbol{f}_+. \tag{10}$$

For $k \ge 1$, we have $K = k$ in one of two cases: either the Markov chain (3) reaches the level $k - 1$ and is absorbed in $(k, 0)$ immediately upon leaving that level or it enters one of the states (k, i) with $i \ne 0$ from which it moves to $(k, 0)$ without visiting the level $k + 1$. This decomposition gives us

$$P[K = k] = \boldsymbol{\alpha}F^{k-1}\boldsymbol{f}_1 + \boldsymbol{\alpha}F^k\boldsymbol{f}_0,$$

which is easily seen to be equivalent to

$$P[K = k] = \boldsymbol{\alpha}F^{k-1}(I - F)\boldsymbol{f}_+. \tag{11}$$

If $\boldsymbol{d}_0 = \mathbf{0}$, then $\boldsymbol{f}_+ = \mathbf{1}$ and K is obviously $\mathrm{PH_d}(\boldsymbol{\alpha}, F)$. If $\boldsymbol{d}_1 = \mathbf{0}$, then $\boldsymbol{f}_+ = F\mathbf{1}$ and K is $\mathrm{PH_d}(\boldsymbol{\alpha}F, F)$, which proves the last statement in the theorem.

When we do not have either $\boldsymbol{d}_0 = \mathbf{0}$ or $\boldsymbol{d}_1 = \mathbf{0}$, it takes more effort to recognize in (10) and (11) the $\mathrm{PH_d}(\boldsymbol{\alpha}\Delta, \Delta^{-1}F\Delta)$ distribution. We construct this representation by adapting a procedure used, in a slightly different context, by Maier and O'Cinneide [(1992), Theorem 4.1].

First, the argument that we used in the proof of Theorem 3.2 shows that $\boldsymbol{\alpha}\Delta$ is a probability density on $\{1, \dots, m\}$. Then, simple algebraic manipulations yield

$$\begin{aligned}\boldsymbol{\alpha}\Delta[\Delta^{-1}F\Delta]^{k-1}[I - \Delta^{-1}F\Delta]\mathbf{1} &= \boldsymbol{\alpha}F^{k-1}\Delta[I - \Delta^{-1}F\Delta]\mathbf{1}\\ &= \boldsymbol{\alpha}F^{k-1}[I - F]\Delta\mathbf{1}\\ &= \boldsymbol{\alpha}F^{k-1}[I - F]\boldsymbol{f}_+ = P[K = k]\end{aligned}$$

for all $k \ge 1$. Finally, $\Delta^{-1}F\Delta$ is a substochastic matrix since it is nonnegative and

$$\begin{aligned}\Delta^{-1}F\Delta\mathbf{1} &= \Delta^{-1}\Delta\mathbf{1} - \Delta^{-1}(I - F)\Delta\mathbf{1}\\ &= \mathbf{1} - \Delta^{-1}(I - F)\boldsymbol{f}_+\\ &= \mathbf{1} - \Delta^{-1}(\mathbf{1} - F\boldsymbol{f}_+ + F\boldsymbol{f}_0)\\ &= \mathbf{1} - \Delta^{-1}(\boldsymbol{f}_1 + F\boldsymbol{f}_0)\\ &\le \mathbf{1},\end{aligned}$$

which completes the proof. □

The representation (10) and (11) has an easy interpretation since the ith component of the vector $(I - F)(\boldsymbol{f}_+) = \boldsymbol{f}_1 + F\boldsymbol{f}_0$ is the probability that, starting from phase i, exactly one more event will be recorded before the catastrophe.

To analyze the number of events in a finite interval, one may follow the same approach as for traditional nontransient MAPs [Latouche and Ramaswami (1999), Chapter 3; Narayana and Neuts (1992); Neuts (1989), Chapter 10]. Defining $N(x)$ to be the number of events in the interval $(0, x]$, and $P^*(z, x)$ to be the matrix of generating functions

$$P^*_{ij}(z, x) = \sum_{n\ge 0} \Pr[N(x) = n \text{ and } \varphi(x) = j | \varphi(0) = i]z^n,$$

we have $P^*(z, x) = \exp[(D^*_0 + zD^*_1)x]$, which, after simple manipulations, may be written as

$$P^*(z, x) = \begin{bmatrix} 1 & \mathbf{0}\\ \{\exp[D(z)x] - I\}D(z)^{-1}\boldsymbol{d}(z) & \exp[D(z)x]\end{bmatrix}$$

with $\boldsymbol{d}(z) = \boldsymbol{d}_0 + z\boldsymbol{d}_1$ and $D(z) = D_0 + zD_1$. We premultiply the matrix $P^*(z, x)$ by the vector $(1 - \boldsymbol{\alpha}\mathbf{1}, \boldsymbol{\alpha})$, postmultiply it by $\mathbf{1}$ and find that the generating function $\phi(z, x)$ of $N(x)$ is given by

$$\phi(z, x) = 1 + \boldsymbol{\alpha}\{\exp[D(z)x] - I\}\{\mathbf{1} + D(z)^{-1}\boldsymbol{d}(z)\}. \tag{12}$$

From this expression, one may determine various moments by differentiation and one may use the algorithmic procedures developed in Narayana and Neuts (1992) and Remiche (1999, 2000a).

Due to the fact that the total number of events is finite, the first moment of $N(x)$ can also be obtained by a direct argument. Let $K(x)$ denote the total number of events in (x,∞). Whereas the probability distribution of the transient phases at time x is

$$\boldsymbol{\alpha}(x)=\boldsymbol{\alpha}\exp(Dx), \tag{13}$$

we conclude from Theorem 4.1 that $K(x)$ has a PH distribution with representation $(\boldsymbol{\alpha}(x)\Delta,\Delta^{-1}F\Delta)$. Thus, $K(x)$ has the same structure as K. The only difference is in the initial distribution.

Since $N(x)=K-K(x)$ and since the first moment of a $\mathrm{PH_d}(\boldsymbol{\tau},T)$ random variable is $\boldsymbol{\tau}(I-T)^{-1}\mathbf{1}$ [Latouche and Ramaswami (1999), Section 2.5], we have by Theorem 4.1 that

$$\begin{aligned} E[N(x)] &= \bigl(\boldsymbol{\alpha}-\boldsymbol{\alpha}(x)\bigr)\Delta(I-\Delta^{-1}F\Delta)^{-1}\mathbf{1} \\ &= \boldsymbol{\alpha}[I-\exp(Dx)](I-F)^{-1}\boldsymbol{f}_+. \end{aligned} \tag{14}$$

The second moment can be determined as

$$\begin{aligned} E[N^2(x)] &= E[K^2]+E[K^2(x)]-2E[KK(x)] \\ &= E[K^2]-E[K^2(x)]-2E[N(x)K(x)] \\ &= \boldsymbol{\alpha}[I-\exp(Dx)](I+F)(I-F)^{-2}\boldsymbol{f}_+-2E[N(x)K(x)] \end{aligned}$$

since the second moment of a random variable with $\mathrm{PH_d}(\boldsymbol{\tau},T)$ distribution is $\boldsymbol{\tau}(I+T)(I-T)^{-2}\mathbf{1}$. By conditioning on the phase at time x, we find that

$$E[N(x)K(x)]=\sum_{1\le i,j\le m}\alpha_i E[N(x)\mathbb{I}\{\varphi(x)=j\}|\varphi(0)=i]E[K(x)|\varphi(x)=j],$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. As $E[N(x)\mathbb{I}\{\varphi(x)=j\}|\varphi(0)=i]$ is given by the (i,j)th entry of $\partial/\partial z\exp[D(z)x]|_{z=1}$ and $E[K(x)|\varphi(x)=j]=((I-F)^{-1}\boldsymbol{f}_+)_j$, we finally obtain that

$$\begin{aligned} E[N^2(x)] &= \boldsymbol{\alpha}[I-\exp(Dx)](I+F)(I-F)^{-2}\boldsymbol{f}_+ \\ &\quad -2\boldsymbol{\alpha}\bigl[\partial/\partial z\exp[D(z)x]\bigr]_{z=1}(I-F)^{-1}\boldsymbol{f}_+. \end{aligned} \tag{15}$$

Thus, the second moment of $N(x)$ may be obtained with just one differentiation of $\exp[D(z)x]$ instead of the two we would need if we used (12). As we shall see in the next section, further simplifications occur for certain choices of $\boldsymbol{\alpha}$; otherwise, the matrix $\partial/\partial z\exp[D(z)x]|_{z=1}$ must be evaluated by one of the procedures described in Narayana and Neuts (1992) and Remiche (1999, 2000a).

More generally, the generating function and moments of the number of events in the finite interval $(t,t+x]$ are given by the expressions (12), (14) and (15), where the initial vector $\boldsymbol{\alpha}$ is replaced by $\boldsymbol{\alpha}\exp(Dt)$.

5. Quasistationary versions. As stated earlier, Condition 2.1 implies that phases 1 to m are transient, so that the state probability vector $\boldsymbol{\alpha}(x)$ tends to $\mathbf{0}$ as x tends to infinity. The normalized vector $(\boldsymbol{\alpha}(x)\mathbf{1})^{-1}\boldsymbol{\alpha}(x)$, however, tends to a nonzero limit, which is a limiting-conditional or quasistationary distribution of $\{\varphi(x)\}$. We refer the reader to Anderson [(1991), Section 5.2] for a general discussion of quasistationary distributions. Here, we note only that quasistationary distributions are nonnegative solutions of systems of the form

$$\boldsymbol{\delta} D = -\eta \boldsymbol{\delta}, \qquad \boldsymbol{\delta}\mathbf{1} = 1, \tag{16}$$

where $-\eta$ is a real, strictly negative eigenvalue of D. There is always at least one such pair $(\eta, \boldsymbol{\delta})$ and if the matrix D is irreducible, then the pair is unique. In this case, $-\eta$ is the eigenvalue of D of maximal real part and $\boldsymbol{\delta}$ is the corresponding eigenvector, appropriately normalized. We have

$$\boldsymbol{\delta} = \lim_{x\to\infty} \big(\boldsymbol{\alpha} \exp(Dx)\mathbf{1}\big)^{-1} \boldsymbol{\alpha} \exp(Dx), \tag{17}$$

independently of $\boldsymbol{\alpha}$.

The assumption that D is irreducible is, however, somewhat restrictive. For example, it is not satisfied in Example 2.3 if W has an Erlang distribution.

The transient MAP$(\boldsymbol{\delta}, D_0, D_1, \boldsymbol{d}_0)$ is said to be *quasistationary* if $\boldsymbol{\delta}$ is a nonnegative solution of (16) for some η. A quasistationary transient MAP has the property that the conditional phase distribution at time x, given that the process has not been absorbed yet, is $\boldsymbol{\delta}$. This follows since

$$\boldsymbol{\delta} \exp(Dx) = e^{-\eta x}\boldsymbol{\delta} \tag{18}$$

for all $x \geq 0$.

The following properties are immediate.

LEMMA 5.1. *If the transient* MAP$(\boldsymbol{\delta}, D_0, D_1, \boldsymbol{d}_0)$ *is quasistationary, then*:

- *The time V until the catastrophe has an exponential distribution with parameter η*;
- *The distribution function of the lifetime L is*

$$P[L \leq x] = 1 - (1 - \boldsymbol{\delta}\boldsymbol{f}_0)e^{-\eta x}.$$

- *The expected number of events in* $(0, x)$ *is*

$$E[N(x)] = (1 - e^{-\eta x})\boldsymbol{\delta}(I - F)^{-1}\boldsymbol{f}_+.$$

PROOF. These statements are direct consequences of (18), Lemma 3.1 and equations (9) and (14), respectively. The lifetime distribution is thus the mixture of an atom at 0 (with mass $\boldsymbol{\delta}\boldsymbol{f}_0$) and an exponential distribution with parameter η. $\square$

We can also derive the following expression for the second moment of $N(x)$.

LEMMA 5.2. *If the eigenvalue* $-\eta$ *of* D *given by* (16) *has algebraic multiplicity* 1, *then*

$$
\begin{aligned}
E[N^2(x)] &= (1 - e^{-\eta x})\boldsymbol{\delta}(I + F)(I - F)^{-2}\boldsymbol{f}_+ \\
&\quad + 2\eta^{-1}e^{-\eta x}\boldsymbol{\delta} D_1\{I - \exp[(I - M)\eta x]\}R(I - F)^{-1}\boldsymbol{f}_+ \qquad (19)\\
&\quad - 2xE[K(x)](\boldsymbol{\delta} D_1\boldsymbol{v}),
\end{aligned}
$$

where $\boldsymbol{v}$ *is the right eigenvector of* D *for to the eigenvalue* $-\eta$, *normalized by* $\boldsymbol{\delta}\boldsymbol{v} = 1$, $M = -\eta^{-1}D$ *and* $R = (I - M + \boldsymbol{v}\boldsymbol{\delta})^{-1}$.

PROOF. The first term in (19) is easily seen to be equal to the first term in (15), so we concentrate on the evaluation of the second term. Since

$$\partial/\partial z \exp[D(z)x]|_{z=1} = \sum_{n\geq 1} x^n/n! \sum_{0\leq\nu\leq n-1} D^\nu D_1 D^{n-1-\nu},$$

we obtain by (16) that

$$
\begin{aligned}
\boldsymbol{\delta}\partial/\partial z \exp[D(z)x]|_{z=1} &= \boldsymbol{\delta} D_1 \sum_{n\geq 1} x^n/n! \sum_{0\leq\nu\leq n-1} (-\eta)^\nu D^{n-1-\nu} \\
&= \boldsymbol{\delta} D_1 \sum_{n\geq 1} x^n/n!(-\eta)^{n-1} \sum_{0\leq\nu\leq n-1} M^{n-1-\nu}
\end{aligned}
$$

with $M = -\eta^{-1}D$ having an eigenvalue equal to 1, and the corresponding eigenvectors being $\boldsymbol{\delta}$ and $\boldsymbol{v}$.

By our assumption, the eigenvalue 1 has multiplicity 1, so that the matrix $I - M + \boldsymbol{v}\boldsymbol{\delta}$ is nonsingular. One directly verifies that

$$\sum_{0\leq\nu\leq n-1} M^{n-1-\nu} = (I - M^n)R + n\boldsymbol{v}\boldsymbol{\delta}$$

by postmultiplying the two sides of this equation by $I - M + \boldsymbol{v}\boldsymbol{\delta}$.

The statement is now proved after simple, albeit tedious, algebraic manipulations. Note that $E[K(x)] = e^{-\eta x}\boldsymbol{\delta}(I - F)^{-1}\boldsymbol{f}_+$. $\square$

It is worth noting that there exist initial distributions for the phase such that the distribution of the total number of events has a modified geometric distribution: Take $\boldsymbol{\xi}$ to be a quasistationary distribution for the transition matrix F defined in (4); that is, $\boldsymbol{\xi}$ is a quasistationary distribution for the phase process embedded at epochs of events. This distribution is such that $\boldsymbol{\xi}F = p\boldsymbol{\xi}$ for some $p < 1$. One immediately concludes from (11) that for the transient MAP($\boldsymbol{\xi}$, D_0, D_1, $\boldsymbol{d}_0$), the distribution of K is given by

$$P[K = k] = (1 - \boldsymbol{\xi}\boldsymbol{f}_0)(1 - p)p^{k-1}$$

for all $k \geq 1$.

6. Conclusion. In this paper, we have generalized the traditional Markovian arrival process, which almost surely has infinitely many events, to the transient case in which the number of events is almost surely finite. In Section 3 we gave expressions for the distribution of the time until such a process reaches its absorbing state and for the time until the last observed event, while in Section 4 we concentrated on the number of events in given intervals. Specifically, we gave expressions for the distribution of the total number of events and for the generating function and moments of the number of events occurring in bounded intervals.

There are many possible applications of our results. One in particular occurs when we use a transient MAP to generate a PH-type point process in the plane according to the procedure outlined in Latouche and Ramaswami (1997). To derive measures of such a planar process, such as the distribution of the total number of points in a region of the plane, it is necessary to discard or keep the events of a transient MAP with a probability which depends on their location; see Remiche (2000b) for details. Thus we are motivated to study a marked version of a transient MAP in which a nonnegative random variable Z_n is associated with each event T_n. We shall defer this study to a later paper.

REFERENCES

ANDERSON, W. J. (1991). *Continuous-Time Markov Chains*: *An Applications-Oriented Approach*. Springer, New York.

GAVER, D. P., JACOBS, P. A. and LATOUCHE, G. (1984). Finite birth-and-death models in randomly changing environments. *Adv. in Appl. Probab.* **16** 715–731.

LATOUCHE, G. and RAMASWAMI, V. (1992). A unified stochastic model for the arrival of packets from periodic sources. *Performance Evaluation* **14** 103–121.

LATOUCHE, G. and RAMASWAMI, V. (1997). Spatial point processes of phase type. In *Teletraffic Contributions for the Information Age. Proceedings of the 15th International Teletraffic Congress* (V. Ramaswami and P. Wirth, eds.) 381–390. North-Holland, Amsterdam.

LATOUCHE, G. and RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM, Philadelphia.

LUCANTONI, D. M., MEIER-HELLSTERN, K. S. and NEUTS, M. F. (1990). A single-server queue with server vacations and a class of non-renewal arrival processes. *Adv. in Appl. Probab.* **22** 676–705.

MAIER, R. S. and O'CINNEIDE, C. A. (1992). A closure characterisation of phase-type distributions. *J. Appl. Probab.* **29** 92–103.

NARAYANA, S. and NEUTS, M. F. (1992). The first two moment matrices of the counts for the Markovian arrival process. *Comm. Statist. 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. An Algorithmic Approach*. Johns Hopkins Univ. Press.

NEUTS, M. F. (1989). *Structured Stochastic Matrices of M/G/1 Type and Their Applications*. Dekker, New York.

NEUTS, M. F. (1990). On the packet stream generated by a random flow of messages of random durations. *Comm. Statist. Stochastic Models* **6** 445–470.

PACHECO, A. and PRABHU, N. U. (1995). Markov-additive processes of arrivals. In *Advances in Queueing: Theory, Methods, and Open Problems* (J. H. Dshalalow, ed.) 167–194. CRC Press, Boca Raton, FL.

RAMASWAMI, V. and LATOUCHE, G. (1989). Modeling packet arrivals from asynchronous input lines. In *Teletraffic Science for New Cost-Effective Systems, Networks and Services. Proceedings of the 12th International Teletraffic Congress* (M. Bonatti, ed.) 721–727. North-Holland, Amsterdam.

REMICHE, M.-A. (1999). Efficiency of an IPhP^3 illustrated through a model in cellular networks. In *Numerical Solution of Markov Chains* (*NSMC'99*): *3rd International Workshop* (B. Plateau, W. J. Stewart and M. Silva, eds.) 296–311. Prensas Univ., Zaragoza, Spain.

REMICHE, M.-A. (2000a). On the exact distribution of the isotropic planar point processes of phase-type. *J. Comput. Appl. Math.* **116** 77–91.

REMICHE, M.-A. (2000b). Towards an extension of Campbell's theorem. Technical report, Aachen-RWTH.

G. LATOUCHE
DÉPARTEMENT D'INFORMATIQUE
UNIVERSITÉ LIBRE DE BRUXELLES
CP 212, BOULEVARD DU TRIOMPHE
1050 BRUXELLES
BELGIUM
E-MAIL: guy.latouche@ulb.ac.be

M.-A. REMICHE
FACULTÉ DES SCIENCES APPLIQUÉES
UNIVERSITÉ LIBRE DE BRUXELLES
CP 165/15, AVENUE F.D. ROOSEVELT
1050 BRUXELLES
BELGIUM
E-MAIL: mremiche@ulb.ac.be

P. TAYLOR
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
VICTORIA 3010
AUSTRALIA
E-MAIL: p.taylor@ms.unimelb.edu.au