

Joint Hitting-Time Densities for Finite State Markov Processes

Tomasz R. Bielecki^{1,*}, Monique Jeanblanc^{2,†}, Ali Devin Sezer^{2,3‡}

¹ Department of Applied Mathematics
Illinois Institute of Technology
Chicago, IL 60616, USA

² Laboratoire de Mathématiques et Modélisation d'Évry
Université d'Évry Val d'Essonne, UMR CNRS 8071
91037 Évry Cedex, France

³ Institute of Applied Mathematics,
Middle East Technical University,
Ankara, Turkey

January 4, 2015

Abstract

For a finite state Markov process and a finite collection $\{\Gamma_k, k \in K\}$ of subsets of its state space, let τ_k be the first time the process visits the set Γ_k . We derive explicit/recursive formulas for the joint density and tail probabilities of the family of stopping times $\{\tau_k, k \in K\}$. In particular, we provide a general solution to the problem that was studied (Assaf et. al., Multivariate phase-type distributions, Operations Research 32 (1984), no. 3, 688-702) in the context of multivariate phase-type distributions. We give a numerical example and indicate the relevance of our results to credit risk modeling.

*The research of T.R. Bielecki was supported by NSF Grant DMS-0604789 and NSF Grant DMS-0908099.

†The research of M. Jeanblanc benefited from the support of the 'Chaire Marchés en mutation', Fédération Bancaire Française.

‡The research of Ali Devin Sezer is supported by Rbuce-up Marie Curie European Project <http://www.rbuceup.eu>

1 Introduction

One of the basic random variables associated with a Markov process X is its first hitting time to a given subset of its state space. In the present work we will confine ourselves to finite state Markov processes. If X has an absorbing state and all of the states can communicate with it, the distribution of the first hitting time to the absorbing state is said to be a *phase-type distribution*. Phase-type distributions, that go back to Erlang [9], are used in modeling of a wide range of phenomena in reliability theory, communications systems, in insurance and finance, to name just a few areas of applications. The literature on these distributions is immense, see, e.g., [2, 3, 1, 12, 15, 17].

To the best of our knowledge, Assaf et al. [4] were the first ones to study multivariate (multidimensional) phase-type distributions. Their setup, for the two dimensional case is as follows: take two proper subsets Γ_1 and Γ_2 of the state space, and assume that with probability 1 the process enters their intersection; let τ_k be the first time the process enters Γ_k . The joint distribution of (τ_1, τ_2) is a two dimensional phase-type distribution. Higher dimensional versions are defined similarly for a finite collection of subsets $\{\Gamma_k, k \in K\}$ of the state space. Denote the number of elements in K by $|K|$. Multidimensional phase-type distributions can put nonzero mass on lower dimensional subsets of $\mathbb{R}_+^{|K|}$, and the density of the distribution when restricted to these subsets is called the “singular part of the distribution.” As far as we know, the only result available in the current literature giving a complete characterization of any multidimensional phase-type density is the case of two dimensions treated in [4]. The same work presents, without proof, a density formula for the nonsingular part of a phase-type distribution of arbitrary finite dimension. The proof of the formula was later given in [10].

The main contribution of the present paper is Theorem 3.2, which gives an explicit formula for the joint density (over appropriate subsets of $\mathbb{R}_+^{|K|}$) of the random vector $\boldsymbol{\tau} := (\tau_k, k \in K)$, covering all possible singular and nonsingular parts. We make no assumptions on whether $\{\Gamma_k, k \in K\}$ are absorbing or not and Theorem 3.2 gives a general formula for the joint density of a collection of first hitting times for any finite state Markov process X . The formula for phase-type densities follows as a special case (see Proposition 4.1 for the one dimensional case and Proposition 4.2 for the general case).

One common method of computing a density is to compute the corresponding tail probability and then to differentiate it to get the density. This is the main method used in the prior literature on our problem (see, for example, [4, 10]). As will be seen in Subsection 3.4, “singular” tail probabilities of $\boldsymbol{\tau}$ (i.e., tail probabilities where some components of $\boldsymbol{\tau}$ are equal) turn out to be much more complicated objects than the corresponding densities and

if one tries to compute these singular tail probabilities with standard methods (for example, using the methodology of taboo probabilities as presented in Syski [17]) one quickly runs into seemingly insurmountable calculations even when $|K|$ is low. Thus, in this paper, we decided to follow the opposite route and compute the density directly using the following idea: for each $\mathbf{t} \in \mathbb{R}_+^{|K|}$, the event $\{\boldsymbol{\tau} = \mathbf{t}\}$ corresponds to the limit of a specific and simple set of trajectories of the Markov process whose (vanishing) probability can be written in terms of the exponentials of submatrices of the intensity matrix.

Subsection 3.1 explains the idea in its simplest form in the derivation of the density of a single τ_k , given as Theorem 3.1 (see also Remark 3.3). The same idea extends to multiple hitting times in Subsection 3.2 and the multidimensional density is given as Theorem 3.2. Subsection 3.4 derives the tail probabilities of $\boldsymbol{\tau}$ using a variant of the density formula; the result is given as Theorem 3.3. The formulas for tail probabilities are more complex compared to densities and are best expressed recursively. We provide a second formula (36) which explicitly states some of the integration that is hidden in the completely recursive (30).

In Section 4, we derive alternative expressions for the density and the tail probability formulas for absorbing $\{\Gamma_k\}$ and indicate the connections between our results and the prior literature. Section 5 gives a numerical example.

From an applied perspective, our primary motivation in deriving the results in the present paper has been to be able to model default times of companies/obligors with first hitting times of a finite state Markov process *where multiple defaults are allowed to happen at the same time*; with the results of our paper this is now possible in great generality (for the case of two obligors one could use the results in [4]). The conclusion explains this application starting from the credit risk model of [10] and the numerical example of Section 5. In addition to credit risk, we expect our results to be useful in reliability theory (see, e.g., [4]), counterparty risk (see, e.g., [7]), and insurance (see, e.g., [5]). From a theoretical perspective, the first motivation of the paper has been the solution of a problem whose two dimensional version was solved in [4], i.e., find simple expressions for the density of $\boldsymbol{\tau}$; surprisingly, prior to the results of the present paper, such expressions were not available in the literature. A second theoretical contribution of the present work is to the line of research originated in [8] and continued in [13] and [11]; the next paragraph explains this side of our contribution.

In [8] the following problem was studied: given a filtration $\mathbb{G} = \{\mathcal{G}_u, u \in \mathbb{R}_+\}$ and a multivariate random time $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$ study the conditional law, say $\mu_u^{\mathbb{G}}$, of $\boldsymbol{\tau}$ given \mathcal{G}_u , in the case that where $P(\tau_i = \tau_j) = 0$ for

$i \neq j$, $i, j = 1, 2, \dots, m$. Thus, a (random) measure was sought so that

$$P(\boldsymbol{\tau} \in B | \mathcal{G}_u) = \int_B \mu_u^{\mathbb{G}}(d\mathbf{t}), \quad (1)$$

for any measurable subset B of \mathbb{R}_+^m . If the measure $\mu_u^{\mathbb{G}}$ is represented as

$$\mu_u^{\mathbb{G}}(d\mathbf{t}) = \varphi_u^{\mathbb{G}}(\mathbf{t})\nu(d\mathbf{t}), \quad (2)$$

where ν is a (possibly random) measure on \mathbb{R}_+^m , then $\varphi_u^{\mathbb{G}}$ is called the density of $P(\cdot | \mathcal{G}_u)$ with respect to ν , and the process $\varphi^{\mathbb{G}}$ is called the conditional density process. This study, was extended in [11], for $n = 2$, to the case where joint default were allowed, that is $P(\tau_1 = \tau_2) > 0$.

Now, let $\mathbb{F} = \{\mathcal{F}_u, u \in \mathbb{R}_+\}$ be the filtration generated by X . The Markov property of X implies that the conditional density of $\boldsymbol{\tau}$ given \mathcal{F}_u directly follows from the density formula (17), as we show in Proposition 3.2. Thus, our results generalize (1) and (2) to the case of arbitrary $m \geq 1$ where the restriction $P(\tau_i = \tau_j) = 0$ for $i \neq j$, $i, j = 1, 2, \dots, m$ is no longer required. It needs to be stressed though that this generalization is only done here in the Markovian case, i.e., when $\mathbb{G} = \mathbb{F}$ and when $\boldsymbol{\tau}$ is defined as first hitting times of the process X . Still, it is an important generalization allowing for studying and modeling probabilities related to simultaneous multivariate trigger events, such as simultaneous defaults in a large pool of obligors.

2 Definitions

Let E be a finite set and X an E -valued continuous time process defined over a measurable space (Ω, \mathcal{F}) equipped with a family of probability measures P_i , $i \in E$, such that $P_i(X_0 = i) = 1$. Under each P_i , X is assumed Markov with intensity matrix $\boldsymbol{\lambda}$. We denote by P the column vector $\{P_i, i \in E\}$. If α is a probability measure on E (written as a row), we will denote by P_α the probability measure $P_\alpha = \alpha P = \sum_{i \in E} \alpha(i)P_i$ on (Ω, \mathcal{F}) . It follows from these definitions that under P_α the initial distribution of X is α , i.e., $P_\alpha(X_0 = i) = \alpha(i)$. The total jump rate of the process when in state i is $-\lambda(i, i) = \sum_{j \neq i} \lambda(i, j)$.

For a finite collection $\{\Gamma_k \subset E, k \in K\}$ of subsets of E , define $\tau_k := \inf\{u \in (0, \infty) : X_u \in \Gamma_k\}$. The index set K can be any finite set, but we will always take it to be a finite subset of the integers. In the next section we derive formulas for the (conditional) joint density and tail probabilities of the stopping times $\{\tau_k, k \in K\}$. To ease notation, unless otherwise noted, we will assume throughout that $E - \bigcup_{k \in K} \Gamma_k$ is not empty and that the initial distribution α puts its full mass on this set, see Remark 3.2 and Subsection 3.3 for comments on how one removes this assumption.

For a set $a \subset E$, a^c will mean its complement with respect to E and $|a|$ will mean the number of elements in it. For two subsets $a, b \subset E$ define $\lambda(a, b)$ as the matrix with elements

$$\begin{cases} \lambda(i, j) & \text{if } i \in a, j \in b, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

For $a \subset E$, we will write $\lambda(a)$ for $\lambda(a, a)$, so that in particular $\lambda = \lambda(E)$.

Throughout we will need to refer to zero matrices and vectors of various dimensions, we will write all as 0; the dimension will always be clear from the context.

2.1 Restriction and extension of vectors and τ as a random function

For any nonempty finite set a , let \mathbb{R}^a be the set of functions from a to \mathbb{R} . \mathbb{R}^a is the same as $\mathbb{R}^{|a|}$, except for the way we index the components of their elements. For two sets $a \subset b$ and $y \in \mathbb{R}^b$ denote y 's restriction to a by $y|_a \in \mathbb{R}^a$:

$$y|_a(i) := y(i) \text{ for } i \in a. \quad (4)$$

The same notation continues to make sense for a of the form $b \times c$, and therefore can be used to write submatrices of a matrix. Thus, for $\mathbf{M} \in \mathbb{R}^{E \times E}$ and nonempty $b, c \subset E$

$$\mathbf{M}|_{b \times c} \quad (5)$$

will mean the submatrix of \mathbf{M} consisting of its components $M(i, j)$ with $(i, j) \in b \times c$. For $b = c$ we will write $\mathbf{M}|_b$.

For $x \in \mathbb{R}^a$, and $a \subset b$, denote by $x|_b \in \mathbb{R}^b$ the following extension of x to b :

$$x|_b(i) = \begin{cases} x(i) & \text{for } i \in a, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

The random vector $\tau = (\tau_k, k \in K)$ can also be thought of as a random function on K , and we will often do so. Thus for $A \subset K$, we may write $\tau|_A$ to denote $(\tau_k, k \in A)$. The advantage of the notation $\tau|_A$ is that we are able to index its components with elements of A rather than with the integers $\{1, 2, 3, \dots, |A|\}$; this proves useful when stating the recursive formulas and proofs below.

2.2 Subpartitions of K

The key aspect of the distribution of τ , already referred to in the introduction, is that it may put nonzero mass on lower dimensional subsets of $\mathbb{R}_+^{|K|}$. This happens, for example, when X can hit $\bigcap_{k \in A} \Gamma_k$ before $\bigcup_{k \in A} \Gamma_k - \bigcap_{k \in A} \Gamma_k$ with positive probability for some $A \subset K$ with $|A| > 1$.

As this example suggests, one can divide $\mathbb{R}_+^{|K|}$ into a number of regions and associate with each an intersection of events of the form “ X hits a before b ” for appropriate subsets of $a, b \subset E$. To write down the various regions and the corresponding events we will use subpartitions of K , which we introduce now.

Recall that K is the set of indices of the stopping times $\{\tau_k\}$ or equivalently the sets $\{\Gamma_k\}$. We call an ordered sequence of disjoint nonempty subsets of K a *subpartition* of K . If the union of all elements of a subpartition is K then we call it a partition. For example, $(\{1, 2\}, \{3\}, \{4\})$ [$(\{1, 2\}, \{4\})$] is a [sub]partition of $\{1, 2, 3, 4\}$. Denote by $|s|$ the number of components in the subpartition and by $s(n)$ its n^{th} component, $n \in \{1, 2, 3, \dots, |s|\}$. In which order the sets appear in the partition matters. For example, $(\{3\}, \{4\}, \{1, 2\})$ is different from the previous partition. In the combinatorics literature this is often called an “ordered partition,” see, e.g., [16]. Only ordered partitions appear in the present work and therefore to be brief we always assume every subpartition to have a definite order and drop the adjective “ordered.” With a slight abuse of notation we will write $s(n_1, n_2)$ to denote the n_2^{nd} element of the n_1^{st} set in the partition.

Two subpartitions s_1 and s_2 are said to be disjoint if $\cup_n s_1(n)$ and $\cup_n s_2(n)$ are disjoint subsets of K . For a given disjoint pair of subpartitions s_1, s_2 let $s_1 \cup s_2$ be their concatenation, for example $(\{1, 2\}, \{3\}) \cup (\{4, 6\}) = (\{1, 2\}, \{3\}, \{4, 6\})$.

For a subpartition s , let Ls be its left shift, i.e., $L(s(1), s(2), \dots, s(|s|)) = (s(2), s(3), \dots, s(|s|))$. Let L^m denote left shift m times. Similarly for $\mathbf{t} \in \mathbb{R}^n$, $n > 1$ let $L\mathbf{t} \in \mathbb{R}^{n-1}$ be its left shift. For $\mathbf{t} \in \mathbb{R}^n$ and $r \in \mathbb{R}$ let $\mathbf{t} - r$ denote $(t_1 - r, t_2 - r, \dots, t_n - r)$.

Given a subpartition s and an index $0 < n \leq |s|$, let $s - s(n)$ be the subpartition which is the same as s but without $s(n)$, e.g., $(\{1, 2\}, \{3\}, \{4, 7\}) - \{3\} = (\{1, 2\}, \{4, 7\})$. Given a nonempty $A \subset K - \bigcup_{n=1}^{|s|} s(n)$ let $s + A$ denote the subpartition that has all the sets in s and A , e.g., $(\{1, 2\}, \{3\}) + \{4, 7\} = (\{1, 2\}, \{3\}, \{4, 7\})$.

Define

$$S(s) := \bigcup_{n=1}^{|s|} \bigcup_{k \in s(n)} \Gamma_k,$$

$S(s)$ is the set of all states of X contained in the partition s . For a *partition* s , define $R_s \subset \mathbb{R}_+^K$ as

$$R_s := \left\{ \mathbf{t} \in \mathbb{R}_+^K : \bigcap_{n=1}^{|s|} \bigcap_{k_1, k_2 \in s(n)} \{t_{k_1} = t_{k_2}\} \right. \\ \left. \cap \{t_{s(1,1)} < t_{s(2,1)} < \dots < t_{s(|s|,1)}\} \right\}.$$

Example 2.1. For $|K| = 6$, and $s = (\{1, 4\}, \{2\}, \{3, 5, 6\})$, we have $|s| = 3$, $s(1) = \{1, 4\}$, $s(2) = \{2\}$, $s(3) = \{3, 5, 6\}$, $s(1, 1) = 1$, $s(2, 1) = 2$, $s(3, 1) = 3$ and

$$R_s = \{\mathbf{t} : t_1 = t_4 < t_2 < t_3 = t_5 = t_6\}.$$

Let \mathcal{S} be the set of all partitions of K . The sets R_s , $s \in \mathcal{S}$, are disjoint and their union is \mathbb{R}_+^K . Our main result, Theorem 3.2 below, shows that for each $s \in \mathcal{S}$, the distribution of $\boldsymbol{\tau}$ restricted to R_s is absolutely continuous with respect to the $|s|$ -dimensional Lebesgue measure on R_s and gives a formula for the corresponding density.

2.3 Restriction of matrices

Let \mathbf{I} be the identity matrix $\mathbf{I} \in \mathbb{R}^{|E| \times |E|}$. For $a \subset E$, we replace its rows whose indices appear in a^c with the 0 vector and call the resulting matrix \mathbf{I}_a , e.g., \mathbf{I}_E is \mathbf{I} itself and \mathbf{I}_\emptyset is the zero matrix. The matrix \mathbf{I}_a has the following action on matrices and vectors:

Lemma 2.1. *Let n be a positive integer. For any $\mathbf{M} \in \mathbb{R}^{|E| \times n}$, $\mathbf{I}_a \mathbf{M}$ is the same as \mathbf{M} except that its rows whose indices are in a^c are replaced by 0 (a zero row vector of dimension n), i.e., if \mathbf{r}_i is the i^{th} row of \mathbf{M} then the i^{th} row of $\mathbf{I}_a \mathbf{M}$ is \mathbf{r}_i if $i \in a$ and 0 otherwise.*

The proof follows from the definitions and is omitted.

Right multiplication by \mathbf{I}_a acts on the columns, i.e., $\mathbf{M} \mathbf{I}_a$ is the same as \mathbf{M} except now that the columns with indices in a^c are set to zero. As an operator on $|E|$ dimensional vectors, \mathbf{I}_a replaces with 0 the coordinates of the vector whose indices are in a^c .

It follows from the definition (3) and Lemma 2.1 that

$$\boldsymbol{\lambda}(a, b) = \mathbf{I}_a \boldsymbol{\lambda} \mathbf{I}_b \tag{7}$$

The operation of setting some of the columns of the identity matrix to zero commutes with set operations, i.e., one has

$$\mathbf{I}_{a \cap b} = \mathbf{I}_a \mathbf{I}_b, \quad \mathbf{I}_{a \cup b} = \mathbf{I}_a + \mathbf{I}_b - \mathbf{I}_a \mathbf{I}_b, \quad \mathbf{I}_{a^c} = \mathbf{I} - \mathbf{I}_a. \tag{8}$$

Using this and Lemma 2.1 one can write any formula involving $\boldsymbol{\lambda}$ in a number of ways. For example, $\boldsymbol{\lambda}(a^c, a)$ can be written as $\mathbf{I}_{a^c} \boldsymbol{\lambda} \mathbf{I}_a = (\mathbf{I} - \mathbf{I}_a) \boldsymbol{\lambda} \mathbf{I}_a = \boldsymbol{\lambda} \mathbf{I}_a - \mathbf{I}_a \boldsymbol{\lambda} \mathbf{I}_a$ or $\boldsymbol{\lambda}(a, b \cap c)$ as $\mathbf{I}_a \boldsymbol{\lambda} \mathbf{I}_{b \cap c} = \mathbf{I}_a \boldsymbol{\lambda} \mathbf{I}_b \mathbf{I}_c = \mathbf{I}_a \boldsymbol{\lambda} \mathbf{I}_c \mathbf{I}_b$.

3 The density of first hitting times

We start by deriving the density of a single hitting time over sets of sample paths that avoid a given subset of the state space until the hitting occurs.

3.1 Density of one hitting time

For any set $d \subset E$ and $u \in \mathbb{R}_+$ define

$$p_{\alpha,d}^u(j) := P_\alpha(X_u = j, X_v \notin d, \forall v \leq u)$$

and $p_d^u(i, j) := P_i(X_u = j, X_v \notin d, \forall v \leq u)$. In addition, let $p^u(i, j) := p_0^u(i, j) = P_i(X_u = j)$. The quantity $\mathbf{p}_{\alpha,d}^u$ is a row vector and \mathbf{p}_d^u and \mathbf{p}^u are $|E| \times |E|$ matrices. Conditioning on the initial state implies $\mathbf{p}_{\alpha,d}^u = \alpha \mathbf{p}_d^u$. It follows from the definition of p^h that

$$\lim_{h \rightarrow 0} p^h(i, j)/h = \lambda(i, j). \quad (9)$$

Lemma 3.1. *Let α be an initial distribution on E with $\alpha|_d = 0$. Then*

$$\mathbf{p}_{\alpha,d}^u = \alpha e^{u\lambda(d^c)}. \quad (10)$$

Proof. We only need to modify slightly the proof of [2, Theorem 3.4, page 48]. The steps are: 1) write down a linear ordinary differential equation (ODE) that the matrix valued function $u \rightarrow \mathbf{p}_d^u|_{d^c}$, $u \in \mathbb{R}_+$, satisfies, 2) the basic theory of ODEs will tell us that the unique solution is $u \rightarrow e^{u\lambda(d^c)}|_{d^c}$.

Let ν_1 be the first jump time of X ; for $X_0 = i \in d^c$, ν_1 is exponentially distributed with rate $-\lambda(i, i) > 0$. Conditioning on ν_1 gives

$$p_d^u(i, j) = P_i(\nu_1 > u)\delta_i(j) + \int_0^u \lambda(i, i)e^{\lambda(i, i)v} \left(\sum_{l \in d - \{i\}} \frac{\lambda(i, l)}{\lambda(i, i)} p_d^{u-v}(l, j) \right) dv \quad (11)$$

for $(i, j) \in d^c \times d^c$. In comparison with the aforementioned proof we have only changed the index set of the last sum to ensure that only paths that keep away from d are included. The unique solution of (11) equals $\mathbf{p}_d^u|_{d^c} = e^{u\lambda|_{d^c}} = e^{u\lambda(d^c)}|_{d^c}$. The equality (10) follows from this and $\alpha|_d = 0$. \square

Remark 3.1. Probabilities that concern sample paths that stay away from a given set are called ‘‘taboo probabilities’’ in [17, Section 1.2]; [17, Equation (F), page 28] is equivalent to (11).

The next result (written in a slightly different form) is well known, see, e.g., [14, 4]. We record it as a corollary here and will use it in subsection 4.3 where we indicate the connections of our results to prior literature. Let $\mathbf{1}$ be the $|E|$ dimensional column vector with all components equal to 1.

Corollary 3.1. *For $\tau_d := \inf\{u : X_u \in d\}$, and an initial distribution with $\alpha|_d = 0$*

$$P_\alpha(\tau_d > u) = \alpha e^{u\lambda(d^c)} \mathbf{1}. \quad (12)$$

Proof.

$$P_\alpha(\tau_d > u) = \sum_{j \in d^c} P_\alpha(X_u = j, X_v \notin d, \forall v \leq u) = \alpha e^{u\lambda(d^c)} \mathbf{1},$$

where the last equality is implied by (10). \square

Remark 3.2. One must modify (12) to

$$P_\alpha(\tau_d > u) = \alpha \mathbf{I}_{d^c} e^{u\lambda(d^c)} \mathbf{1}, \quad P_\alpha(\tau_d = 0) = \alpha \mathbf{I}_d \mathbf{1}$$

if one does not assume $\alpha|_d = 0$.

Once $P_\alpha(\tau_d > u)$ is known, one can differentiate it to compute the density of τ_d . This, in essence, has been the strategy employed in previous works. It clearly works well for a single hitting time or for nonsingular parts of the distribution of τ (see [4, 10]). However, as explained in the introduction, the same idea runs into difficulties if one tries to use it to compute the nonsingular parts of the distribution of τ . The next theorem computes the density directly for the case of a single stopping time τ_d . The theorem allows also to specify a subset $b \subset E$ that the process is required to stay away before the hitting time; this generalization turns out to be useful to extend the theorem to multiple hitting times (see the next subsection), setting $b = \emptyset$ gives the actual density of τ_d .

Theorem 3.1. *Let $a, b \subset E$, $a \cap b = \emptyset$ be given. Define $\tau_a := \inf\{u : X_u \in a\}$ and set $d = a \cup b$. Then*

$$\frac{d}{du} [P_\alpha(\tau_a \in (0, u], X_v \notin b, \forall v < \tau_a)] = \alpha e^{u\lambda(d^c)} \lambda(d^c, a) \mathbf{1}, \quad (13)$$

where α is the initial distribution of X with $\alpha|_d = 0$.

add comments: a is the target, d^c the waiting set

In other words, the density of τ_a on the set $\{X_v \notin b, \forall v < \tau_a\}$ (by this we mean that $P_\alpha(\tau_a \in C, X_v \notin b, \forall v < \tau_a) = \int_C \alpha e^{u\lambda(d^c)} \lambda(d^c, a) \mathbf{1} du$) is given by the right side of (13).

The idea behind (13) and its proof is this: for $\tau_a = u$ with X staying out of b until time u , X has to stay in the set d^c until time u and jump exactly at that time into a .

Proof of Theorem 3.1. The definition of the exponential distribution implies that X jumps more than once during the time interval $[u, u + h]$ has probability $O(h^2)$. This, (9) and the Markov property of X (invoked at time u) give

$$\begin{aligned} P_i(X_{\tau_a} = j, \tau_a \in (u, u + h), X_v \notin b, \forall v \leq u) \\ = \left(\sum_{l \in d^c} p_d^u(i, l) \lambda(l, j) \right) h + o(h). \end{aligned} \quad (14)$$

$$P_i(\tau_a \in (u, u+h), X_v \notin d, \forall v \leq u) = \left(\sum_{j \in a} \sum_{l \in d^c} p_d^u(i, l) \lambda(l, j) \right) h + o(h). \quad (15)$$

By the previous lemma $p_d^u(i, j)$ equals exactly the $(i, j)^{th}$ component of $e^{u\lambda(d^c)}$. These imply (13). \square

Remark 3.3. *Setting $b = \emptyset$ in Theorem 3.1 we get the density of τ_a . The formula (13) generalizes the exponential density: if τ' is exponentially distributed with rate $\lambda' \in (0, \infty)$ it has density $e^{-\lambda' t}$.*

The following result will be needed in the proof of Theorem 3.2. We state it here, as it is derived using the same reasoning as the one used for Theorem 3.1.

Proposition 3.1. *Let $a, b \subset E$, $a \cap b = \emptyset$ where the set a is nonempty, be given. Define $\tau_a := \inf\{u : X_u \in a\}$ and $d = a \cup b$. Let α is an initial distribution on E with $\alpha|_d = 0$. Set $\alpha_1 := \alpha e^{\tau_a \lambda(d^c)} \lambda(d^c, a)$ and $\mathcal{V} := \{X_v \notin b, \forall v \leq \tau_a\}$. Then*

$$P_\alpha(X_{\tau_a} = j | (\tau_a, 1_{\mathcal{V}})) = \alpha_1(j) / \alpha_1 \mathbf{1} \text{ on } \mathcal{V},$$

where $1_{\mathcal{V}}$ is the indicator function of the event \mathcal{V} .

Note that \mathcal{V} is the event that X does not visit the set b before time τ_a .

Proof. The proof follows from (14) and the definition of the conditional expectation. \square

3.2 The multidimensional density

One can extend (13) to a representation of the distribution of τ using the subpartition notation of subsection 2.2. For a partition s of K , $n \in \{1, 2, \dots, |s|\}$ and $\mathbf{t} \in R_s \subset \mathbb{R}_+^K$, define

$$\bar{t}_n := t_{s(n,1)}, \quad \bar{t}_0 := 0, \quad W_n := [S(L^{n-1}s)]^c, \quad T_n := \left[\bigcap_{k \in s(n)} \Gamma_k \right] \cap W_{n+1}, \quad (16)$$

where W stands for “waiting” and T for “target.” In particular, $W_1 = [S(L^0 s)]^c = [S(s)]^c = [\bigcup_{k \in K} \Gamma_k]^c = E - \bigcup_{k \in K} \Gamma_k$.

the previous example was completely wrong and had no sense! The key idea of the density formula and its proof is the $|s|$ step version of the one in Theorem 3.1: in order for $\tau = \mathbf{t} \in \mathbb{R}_+^K$, X has to stay in the set W_1 until time \bar{t}_1 and jump exactly at that time into $T_1 \subset W_2$; then stay in the set W_2 until time \bar{t}_2 and jump exactly then into T_2 and so on until all of the pairs (W_n, T_n) , $n \leq |s|$, are exhausted.

Although not explicitly stated, all of the definitions so far depend on the collection $\{\Gamma_k, k \in K\}$. We will express this dependence explicitly in the following theorem by including the index set K as a variable of the density function f . This will be useful in its recursive proof, in the next subsection where we comment on the case when α is an arbitrary initial distribution and in Proposition 3.2 where we give the conditional density of τ given \mathcal{F}_u , $u > 0$. For a sequence $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n$ of square matrices of the same size $\prod_{m=1}^n \mathbf{M}_m$ will mean $\mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_n$.

Theorem 3.2. *For any partition $s \in \mathcal{S}$, the distribution of τ on the set R_s has density*

$$f_s(\alpha, \mathbf{t}, K) = \alpha \left(\prod_{n=1}^{|s|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \right) \mathbf{1}, \quad \mathbf{t} \in R_s \quad (17)$$

with respect to the $|s|$ -dimensional Lebesgue measure on R_s .

Proof. The proof will use induction on $|K|$. For $|K| = 1$, (17) is the same as (13) with $b = \emptyset$. Let $\kappa > 1$ and suppose that (17) holds for all K with $|K| \leq \kappa - 1$; we will now argue that (17) also holds for all K with $|K| = \kappa$. Fix a K with $|K| = \kappa$ and a partition s of K ; we will show that the distribution of τ restricted to R_s has the density (17). Specifically, we will show that for any bounded and measurable function $g : \mathbb{R}^K \rightarrow \mathbb{R}$ the following equality holds

$$\mathbb{E}[1_{R_s}(\tau)g(\tau)] = \int_{R_s} g(\mathbf{t})f_s(\alpha, \mathbf{t}, K)d_s\mathbf{t}, \quad (18)$$

where $d_s\mathbf{t}$ denotes the $|s|$ -dimensional Lebesgue measure on R_s .

Define $\vartheta = \min \tau := \min\{\tau_k, k \in K\} = \tau_{s(1,j)}$ for any $j \in s(1)$, that is, ϑ is the first time X enters the set $\bigcup_{k \in K} \Gamma_k$. In the rest of the proof we will proceed as if $P_\alpha(\vartheta < \infty) = 1$; the treatment of the possibility $P_\alpha(\vartheta = \infty) > 0$ needs no new ideas and the following argument can be extended to handle it by adding several case by case comments.

For a sample path of X that satisfies $\tau \in R_s$ the following must hold: 1) $X_\vartheta \in T_1$ and 2) $X_t \in W_1$ for $t \leq \vartheta$; 1) and 2) also imply $\vartheta = \tau_{s(1,1)}$ for the same sample path. Therefore,

$$\{\tau \in R_s\} \subset \mathcal{W}_1 := \{X_u \in W_1, u < \vartheta\} \cap \{X_\vartheta \in T_1\}. \quad (19)$$

Theorem 3.1 implies that if $\lambda(W_1, T_1)$ is zero, then \mathcal{W}_1 has probability zero. Thus, (19) implies that if $\lambda(W_1, T_1)$ is zero then $P_\alpha(\tau \in R_s) = 0$ and, indeed, $f_s(\alpha, \mathbf{t}, K) = 0$ is the density of τ on R_s . From here on, we will treat the case when $\lambda(W_1, T_1)$ is nonzero.

Next, define the process \widehat{X} by $\widehat{X}_u := X_{u+\vartheta}$, $u \geq 0$, and $\widehat{\tau} = (\widehat{\tau}_k, k \in S(Ls))$ where

$$\widehat{\tau}_k := \inf\{u : \widehat{X}_u \in \Gamma_k\};$$

\widehat{X} is the trajectory of X after time ϑ . The strong Markov property of X implies that \widehat{X} is a Markov process with intensity matrix $\boldsymbol{\lambda}$ and starting from $\widehat{X}_0 = X_\vartheta$. This and (19) imply

$$\widehat{\boldsymbol{\tau}} = \boldsymbol{\tau}|_{L_s} - \vartheta, \quad (20)$$

where $\boldsymbol{\tau}|_{L_s}$ is defined in accordance with (4). Finally, the definition of $\widehat{\boldsymbol{\tau}}$ and that of \mathcal{W}_1 imply

$$\{\boldsymbol{\tau} \in R_s\} = \mathcal{W}_1 \cap \{\widehat{\boldsymbol{\tau}} \in R_{L_s}\}. \quad (21)$$

In words, this display says: for $\boldsymbol{\tau}$ to be partitioned according to s , among all $\{\Gamma_k\}$, X must visit $\bigcap_{k \in s(1)} \Gamma_k$ first and after this visit the rest of the hitting times must be arranged according to the partition L_s .

Denote by $\mathbf{1}'$ the function that maps all elements of K to 1. Define $\widehat{g} : \mathbb{R}_+ \times \mathbb{R}_+^{S(L_s)} \rightarrow \mathbb{R}$ as

$$\widehat{g}(u, \widehat{\boldsymbol{t}}) := g\left(u\mathbf{1}' + \widehat{\boldsymbol{t}}^{S(s)}\right),$$

where we used the function extension notation of (6). Equalities (20) and (21) imply

$$\begin{aligned} \mathbb{E}[1_{R_s}(\boldsymbol{\tau})g(\boldsymbol{\tau})] &= \mathbb{E}[1_{\mathcal{W}_1}1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}})] \\ &= \mathbb{E}[\mathbb{E}[1_{\mathcal{W}_1}1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}})|\mathcal{F}_\vartheta]] \\ &= \mathbb{E}[1_{\mathcal{W}_1}\mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})g(\vartheta, \widehat{\boldsymbol{\tau}})|\mathcal{F}_\vartheta]], \end{aligned} \quad (22)$$

where for the last equality we used the fact that \mathcal{W}_1 is \mathcal{F}_ϑ measurable.

The property 5A in [6, page 98] implies

$$\mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}})|\mathcal{F}_\vartheta] = h(\vartheta) \quad (23)$$

where

$$h(u) := \mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u, \widehat{\boldsymbol{\tau}})|\mathcal{F}_\vartheta]. \quad (24)$$

The strong Markov property of X and the definition of \widehat{X} imply

$$h(u) = \mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u, \widehat{\boldsymbol{\tau}})|X_\vartheta] = \mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u, \widehat{\boldsymbol{\tau}})|\widehat{X}_0].$$

The random variable \widehat{X}_0 takes values in a finite set and therefore one can compute the conditional expectation $\mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u, \widehat{\boldsymbol{\tau}})|\widehat{X}_0]$ by conditioning on each of these values separately. Since \widehat{X} is a Markov process with initial distribution \widehat{X}_0 with intensity matrix $\boldsymbol{\lambda}$, one can invoke the induction hypothesis for the set $K - s(1)$ to conclude that, on the set $\{\widehat{X}_0 = j\}$,

$$h(u) = \mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u, \widehat{\boldsymbol{\tau}})|\widehat{X}_0 = j] = \int_{R_{L_s}} f_{L_s}(\delta_j, \boldsymbol{t}, K - s(1))g(u, \boldsymbol{t})d_{L_s}\boldsymbol{t} \quad (25)$$

where f_{L_s} is given as in (17) with s changed to L_s and K changed to $K - s(1)$, i.e., for $\mathbf{t} \in R_{L_s}$

$$f_{L_s}(\delta_j, \mathbf{t}, K - s(1)) = \delta_j \left(\prod_{n=2}^{|s|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \boldsymbol{\lambda}(W_n, T_n) \right) \mathbf{1}.$$

This and (25) now give

$$\begin{aligned} \mathbb{E}[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}}) \widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}}) | \mathcal{F}_\vartheta] &= \sum_{j \in s(1)} 1_{\{\widehat{X}_0=j\}} \int_{R_{L_s}} f_{L_s}(\delta_j, \mathbf{t}, K - s(1)) g(\vartheta, \mathbf{t}) d_{L_s} \mathbf{t} \\ &= \int_{R_{L_s}} f_{L_s}(\widehat{\alpha}, \mathbf{t}, K - s(1)) g(\vartheta, \mathbf{t}) d_{L_s} \mathbf{t}. \end{aligned}$$

with $\widehat{\alpha} = \delta_{X_\vartheta}$.

Therefore, the last expectation in (22) involves only three random variables: ϑ , 1_A and $\widehat{X}_0 = X_\vartheta$, where $A = \{X_u \in W_1, u < \vartheta\}$. Theorem 3.1 implies that the density of ϑ on the set A is $\alpha e^{\lambda(W_1)\bar{t}_1} \boldsymbol{\lambda}(W_1, T_1) \mathbf{1}$, and Proposition 3.1 implies that the distribution of \widehat{X}_0 conditioned on ϑ and 1_{W_1} is

$$\frac{\alpha e^{\lambda(W_1)\vartheta} \boldsymbol{\lambda}(W_1, T_1)}{\alpha e^{\lambda(W_1)\vartheta} \boldsymbol{\lambda}(W_1, T_1) \mathbf{1}}.$$

These, the induction hypothesis, (24) and (25) imply that the outer expectation (22) equals (18). This last assertion finishes the proof of the induction step and hence the theorem. \square

In what follows, to ease exposition, we will sometimes refer to f as the ‘‘density’’ of $\boldsymbol{\tau}$ without explicitly mentioning the reference measures d_s , $s \in \mathcal{S}$.

Remark 3.4. If any of the matrices in the product (17) equals the zero matrix, then f will be 0. Therefore, if $\boldsymbol{\lambda}(W_n, T_n) = 0$ for some $n = 1, 2, \dots, |s|$ then $P_\alpha(\boldsymbol{\tau} \in R_s) = 0$. By definition $\boldsymbol{\lambda}(W, T) = 0$ if $T = \emptyset$. Thus, as a special case, we have $P_\alpha(\boldsymbol{\tau} \in R_s) = 0$ if $T_n = \emptyset$ for some $n = 1, 2, 3, \dots, |s|$.

Remark 3.5. The first $\kappa > 0$ jump times of a standard Poisson process with rate $\lambda' \in (0, \infty)$ have the joint density

$$\prod_{n=1}^{\kappa} e^{\lambda'(t_n - t_{n-1})} \lambda',$$

$0 = t_0 < t_1 < t_2 < \dots < t_\kappa$. The density (17) is a generalization of this simple formula.

3.3 When α puts positive mass on $\cup_k \Gamma_k$

If α puts positive mass on $\gamma := \cup_{k \in K} \Gamma_k$, one best describes the distribution of τ proceeding as follows. Define $\bar{\alpha}' := 1 - \sum_{i \in \gamma} \alpha(i)$ and $\alpha' := (\alpha - \sum_{i \in \gamma} \alpha(i)\delta_i)/\bar{\alpha}'$ if $\bar{\alpha}' > 0$; $\bar{\alpha}'$ is a real number and α' , when defined, is a distribution.

First consider the case when $\bar{\alpha}' > 0$. The foregoing definitions imply

$$P_\alpha(\tau \in U) = \bar{\alpha}' P_{\alpha'}(\tau \in U) + \sum_{i \in \gamma} \alpha(i) P_i(\tau \in U) \quad (26)$$

for any measurable $U \subset \mathbb{R}_+^K$. By its definition α' puts no mass on $\gamma = \cup_{k \in K} \Gamma_k$ and therefore Theorem 3.2 is applicable and $f(\alpha', \cdot, K)$ is the density of the distribution $P_{\alpha'}(\tau \in \cdot)$. For the second summand of (26), it is enough to compute each $P_i(\tau \in U)$ separately. Define $K_i := \{k : i \in \Gamma_k\}$, $U_i := \{\mathbf{t} : \mathbf{t} \in U, t_k = 0, k \in K_i\}$, $\bar{U}_i := \{\mathbf{t}|_{K_i^c}, \mathbf{t} \in U_i\}$. Now remember that $i \in \gamma$; thus if $i \in \Gamma_k$ then $\tau_k = 0$ under P_i , and therefore $P_i(\tau \in U) = P_i(\tau \in U_i)$. For $\tau \in U_i$, the stopping times $\tau|_{K_i}$ are all deterministically 0. Thus to compute $P_i(\tau \in U_i)$ it suffices to compute $P_i(\tau|_{K_i^c} \in \bar{U}_i)$. But by definition $i \notin \cup_{k \in K_i^c} \Gamma_k$ and once again Theorem 3.2 is applicable and gives the density of $\tau|_{K_i^c}$ under P_i as $f(\delta_i, \cdot, K_i^c)$.

If $\bar{\alpha}' = 0$ then

$$P_\alpha(\tau \in U) = \sum_{i \in \gamma} \alpha(i) P_i(\tau \in U)$$

and the computation of $P_i(\tau \in U)$ goes as above.

3.4 Tail probabilities of τ

Probabilities of tail events have representations as integrals of densities given in Theorem 3.2 over appropriate subsets of $\mathbb{R}_+^{|K|}$. But to evaluate such integrals directly is nontrivial and inefficient. In the present subsection, we derive a recursive and compact representation of these probabilities that use a version of the density formula and the ideas used in its derivation.

By tail probabilities we will mean probabilities of sets of the form

$$\bigcap_{n=1}^{|s|} \bigcap_{k_1, k_2 \in s(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s(n,1)} > t_n\} \bigcap_{n_1 \neq n_2, n_1, n_2 \leq |s|} \{\tau_{s(n_1,1)} \neq \tau_{s(n_2,1)}\}, \quad (27)$$

where s is a partition of K , and $\mathbf{t} \in \mathbb{R}_+^{|s|}$ is such that $t_n < t_{n+1}$. In (27) all equality and inequality condition are explicitly specified. One can write standard tail events in terms of these e.g., $\{\tau_1 > t_1\} \cap \{\tau_2 > t_2\}$ is the same as the disjoint union

$$(\{\tau_1 > t_1, \tau_2 > t_2\} \cap \{\tau_1 \neq \tau_2\}) \cup \{\tau_1 = \tau_2 > \max(t_1, t_2)\}.$$

Both of these sets are of the form (27). Thus, it is enough to be able to compute probabilities of the events of the form (27).

Remark 3.6. From here on, to keep the notation short, we will assume that, over tail events, unless explicitly stated with an equality condition, all stopping times appearing in them are different from each other (therefore, when writing formulas, we will omit the last intersection in (27)).

A tail event of the form (27) consists of a sequence of constraints of the form

$$\{\tau_{s(n,1)} = \tau_{s(n,2)} = \cdots = \tau_{s(n,|s(n)|)} > t_n\}.$$

There are two types of sub-constraints involved here: that entrance to all Γ_k , $k \in s(n)$, happen at the same time and that this event occurs after time t_n . Keeping track of all of these constraints as they evolve in time requires that we introduce yet another class of events that generalize (27) (see also Remark 4.1). For two **disjoint** subpartitions s_1 and s_2 of K and an element $\mathbf{t} \in \mathbb{R}_+^{|s_1|}$ such that $t_{|s_1|} > t_{|s_1|-1} > \cdots > t_2 > t_1$ (if $|s_1| = 0$ by convention set $t = 0$) define

$$\begin{aligned} \mathcal{T}(s_1, s_2, \mathbf{t}) := & \left(\bigcap_{n=1}^{|s_1|} \bigcap_{k_1, k_2 \in s_1(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s_1(n,1)} > t_n\} \right) \cap \\ & \bigcap_{n=1}^{|s_2|} \bigcap_{k_1, k_2 \in s_2(n)} \{\tau_{k_1} = \tau_{k_2}\}. \end{aligned} \quad (28)$$

In view of our convention (cf. Remark 3.6), setting $s_1 = s$ and $s_2 = \emptyset$ reduces (28) to (27). The indices in s_1 appear both in equality constraints and time constraints while indices in s_2 appear only in equality constraints.

Remark 3.7. The definition (28) implies that if a component of s_2 has only a single element, that component has no influence on $\mathcal{T}(s_1, s_2, \mathbf{t})$. For example, $\mathcal{T}(s_1, (\{1\}, \{2, 3\}), \mathbf{t})$ is the same as $\mathcal{T}(s_1, (\{2, 3\}), \mathbf{t})$.

To express $P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t}))$ we will define a collection of functions \mathbf{p}_i , $i \in \Omega_0$, of s_1 , s_2 and \mathbf{t} . We will denote by \mathbf{p} the column matrix $\{\mathbf{p}_i, i \in \Omega_0\}$. For $s_1 = \emptyset$, and $i \in \Omega_0$ define \mathbf{p}_i as

$$\mathbf{p}_i(\emptyset, s_2, 0) := P_i(\mathcal{T}(\emptyset, s_2, 0)).$$

The definitions of \mathbf{p} and \mathcal{T} and Remark 3.7 imply

$$\mathbf{p}(\emptyset, s_2, 0) = \mathbf{1} \quad (29)$$

if s_2 is empty or it consists of components with single elements. For a given disjoint pair of subpartitions s_1, s_2 define

$$T_n(s_1, s_2) := \bigcap_{k \in s_2(n)} \Gamma_k - S(s_1 \cup s_2 - s_2(n)), \quad T(s_1, s_2) := \bigcup_{n=1}^{|s_2|} T_n(s_1, s_2).$$

If $s_1 \neq \emptyset$, define \mathbf{p} recursively as

$$\begin{aligned} \mathbf{p}(s_1, s_2, \mathbf{t}) := & \tag{30} \\ & \int_0^{t_1} e^{u\lambda(W)} \lambda(W, T(s_1, s_2)) \left(\sum_{n=1}^{|s_2|} \mathbf{I}_{T_n(s_1, s_2)} \mathbf{p}(s_1, s_2 - s_2(n), \mathbf{t} - u) \right) du \\ & + e^{t_1\lambda(W)} \mathbf{p}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1), \end{aligned}$$

where $W = [S(s_1 \cup s_2)]^c$.

Theorem 3.3. *Suppose $\Omega_0 - S(s_1 \cup s_2)$ is not empty and that α is an initial distribution on Ω_0 that puts all of its mass on this set. Then*

$$P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t})) = \alpha\mathbf{p}(s_1, s_2, \mathbf{t}).$$

Proof. The proof is parallel to that of Theorem 3.2 and proceeds by induction; we will only provide the argument that justifies the recursion (30). We shall consider two cases: either X has visited $W^c = S(s_1 \cup s_2)$ before t_1 or not. That is, we split $\mathcal{T}(s_1, s_2, \mathbf{t})$ into disjoint sets

$$\mathcal{T}(s_1, s_2, \mathbf{t}) = \{\vartheta > t_1, \mathcal{T}(s_1, s_2, \mathbf{t})\} \cup \{\vartheta \leq t_1, \mathcal{T}(s_1, s_2, \mathbf{t})\}$$

where ϑ is the first time when X enters in W^c . Since \mathbf{t} satisfies $t_{|s_1|} > t_{|s_1|-1} > \dots > t_2 > t_1$, X must enter at time ϑ into the set s_2 on the set $\vartheta < t_1$, i.e., $\vartheta = \min_{j,k}(\tau_{s_2(j,k)})$.

In a first step, we consider the case where X has not visited $W^c = S(s_1 \cup s_2)$ before t_1 , that is, we are on the set $\Delta_1 := \{\vartheta > t_1\} = \{X_u \notin W^c, \forall u < t_1\}$, so that, in particular $\tau_{s_1(1,1)} > t_1$. Therefore, on Δ_1

$$\begin{aligned} \mathcal{T}(s_1, s_2, \mathbf{t}) &= \left(\bigcap_{n=2}^{|s_1|} \bigcap_{k_1, k_2 \in s_1(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s_1(n,1)} > t_n\} \right) \\ &\quad \bigcap_{k_1, k_2 \in s_1(1)} \{\tau_{k_1} = \tau_{k_2} > t_1\} \bigcap_{n=1}^{|s_2|} \bigcap_{k_1, k_2 \in s_2(n)} \{\tau_{k_1} = \tau_{k_2} > t_1\} \\ &=: \mathcal{T}^{t_1}(s_1, s_2, \mathbf{t}). \end{aligned}$$

where the set

$$\bigcap_{k_1, k_2 \in s_1(1)} \{\tau_{k_1} = \tau_{k_2} > t_1\} \bigcap_{n=1}^{|s_2|} \bigcap_{k_1, k_2 \in s_2(n)} \{\tau_{k_1} = \tau_{k_2} > t_1\}$$

will be, after a shift operation with t_1 acting on the times τ_k , a set with equalities constraints only, of the form

$$\bigcap_{n=1}^{|s|} \bigcap_{k_1, k_2 \in s(n)} \{\tau_{k_1} = \tau_{k_2}\}$$

with $s = s_2 + s_1(1)$.

This, that $\mathcal{T}^{t_1}(s_1, s_2, \mathbf{t})$ is measurable with respect to $\sigma(X_u, u > t_1)$ and the Markov property of X imply

$$\begin{aligned} P_\alpha(X_u \notin W^c, \forall u \leq t_1, X_{t_1} = j, \mathcal{T}(s_1, s_2, \mathbf{t})) \\ &= \mathbb{E}_\alpha \left[1_{\Delta_1} 1_{X_{t_1}=j} \mathbb{E}_\alpha(1_{\mathcal{T}^{t_1}(s_1, s_2, \mathbf{t})} | \mathcal{F}_{t_1}) \right] \\ &= \mathbb{E}_\alpha \left[1_{\Delta_1} 1_{X_{t_1}=j} \mathbb{E}_\alpha(1_{\mathcal{T}^{t_1}(s_1, s_2, \mathbf{t})} | X_{t_1}) \right] \\ &= \mathbb{E}_\alpha \left[1_{\Delta_1} 1_{X_{t_1}=j} P_j(\mathcal{T}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1)) \right]. \end{aligned}$$

It follows that

$$\begin{aligned} P_\alpha(\Delta_1) &= \sum_{j \in W} P_\alpha(X_u \notin W^c, \forall u \leq t_1, X_{t_1} = j, \mathcal{T}(s_1, s_2, \mathbf{t})) \\ &= \sum_{j \in W} P_\alpha(X_u \notin W^c, \forall u \leq t_1, X_{t_1} = j) P_j(\mathcal{T}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1)) \\ &= \sum_{j \in E} \left(\alpha e^{t_1 \lambda(W)} \right)_j \mathbf{p}_j(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1) \\ &= \alpha e^{t_1 \lambda(W)} \mathbf{p}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1), \end{aligned}$$

where we have used the induction hypothesis in the second equality and (10) in the third equality.

In a second step, we work on the set $\vartheta \leq t_1$. Since the sets $T_n(s_1, s_2), n = 1, \dots, |s_2|$ are disjoint, we have

$$P_\alpha(\vartheta \leq t_1, \mathcal{T}(s_1, s_2, \mathbf{t})) = \sum_{n=1}^{|s_2|} P_\alpha(\vartheta \leq t_1, X_\vartheta \in T_n(s_1, s_2), \mathcal{T}(s_1, s_2, \mathbf{t})).$$

Now, we note

$$\begin{aligned} P_\alpha(\vartheta \in du, X_\vartheta \in T_n(s_1, s_2), \mathcal{T}(s_1, s_2, \mathbf{t})) \\ &= \mathbb{E}_\alpha \left[1_{\vartheta \in du} \sum_{j \in T_n(s_1, s_2)} 1_{X_u=j} \mathbb{E}_\alpha(1_{\mathcal{T}(s_1, s_2, \mathbf{t})} | \mathcal{F}_u) \right] \\ &= \mathbb{E}_\alpha \left[1_{\vartheta \in du} \sum_{j \in T_n(s_1, s_2)} 1_{X_u=j} P_j(\mathcal{T}(s_1, s_2 - s_2(n), \mathbf{t} - u)) \right] \\ &= \mathbb{E}_\alpha \left[1_{\vartheta \in du} \sum_{j \in T_n(s_1, s_2)} 1_{X_u=j} P_j(\mathcal{T}(s_1, s_2 - s_2(n), \mathbf{t} - u)) \right] \\ &= \sum_{j \in T_n(s_1, s_2)} (\alpha e^{u \lambda(W)} \lambda(W, T(s_1, s_2)))_j \mathbf{p}_j(s_1, s_2 - s_2(n), \mathbf{t} - u) du \\ &= \alpha e^{u \lambda(W)} \lambda(W, T(s_1, s_2)) \mathbf{I}_{T_n(s_1, s_2)} \mathbf{p}(s_1, s_2 - s_2(n), \mathbf{t} - u) du, \end{aligned}$$

where the equality $\mathbb{E}_\alpha [1_{\vartheta \in du} 1_{X_u=j}] = (\alpha e^{u\lambda(W)} \boldsymbol{\lambda}(W, T(s_1, s_2)))_j$ follows from the proof of Theorem 3.1 and the fourth equality follows from the induction hypothesis. It remains to integrate with respect to du . \square

Theorem 3.3 holds for all finite state Markov processes and does not require that any of the $\{\Gamma_k\}$ be absorbing. The evaluations of \mathbf{p} on the right side of the recursion (30) will have smaller subpartitions in its arguments; then in a finite number of steps these recursions will lead to an evaluation of \mathbf{p} with $s_1 = \emptyset$.

Note that (30) reduces to

$$\mathbf{p}(s_1, \emptyset, \mathbf{t}) = e^{\lambda(S(s_1)^c)t_1} \mathbf{p}(Ls_1, s_1(1), L\mathbf{t} - t_1). \quad (31)$$

if $s_2 = \emptyset$.

Example 3.1. In case where $|s_2| = |s_1| = 1$ we have

$$\begin{aligned} \mathbf{p}(s_1, s_2, \mathbf{t}) &= \int_0^{t_1} e^{u\lambda(W)} \boldsymbol{\lambda}(W, T(s_1, s_2)) \mathbf{I}_{T(s_1, s_2)} \mathbf{p}(s_1, \emptyset, \mathbf{t} - u) du \\ &+ e^{t_1\lambda(W)} \mathbf{p}(\emptyset, s_2 + s_1(1), L\mathbf{t} - t_1). \end{aligned} \quad (32)$$

The quantities $\mathbf{p}(s_1, \emptyset, \mathbf{t})$ and $\mathbf{p}(\emptyset, s_2 + s_1(1), L\mathbf{t} - t_1)$ are known from (31) and (29) for the first and (30) and (29) for the second.

Example 3.2. Let $K = \{1, \dots, 8\}$, $s_1 = (\{1, 4\}, \{2\})$ and $s_2 = (\{3, 6\}, \{7\})$ so that $s_2(1) = \{3, 6\}$. Then, $S(s_1 \cup s_2) = \{\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4 \cup \Gamma_6 \cup \Gamma_7\}$ and $W = \Gamma_5 \cup \Gamma_8$. For $\mathbf{t} = (t_1, t_2)$ with $t_2 > t_1$, one defines

$$\mathcal{T}(s_1, s_2, \mathbf{t}) = \{\tau_1 = \tau_4 > t_1, \tau_2 > t_2, \tau_3 = \tau_6\}$$

By definition, $T_1 = \Gamma_3 \cap \Gamma_6$. Then,

$$\begin{aligned} P(\{\tau_1 = \tau_4 > t_1, \tau_2 > t_2, \tau_3 = \tau_6\}) &= \int_0^{t_1} e^{u\lambda(W)} \boldsymbol{\lambda}(W, T_1) \mathbf{I}_{T_1} \mathbf{p}(s_1, \{7\}, \mathbf{t} - u) du \\ &+ e^{t_1\lambda(W)} \mathbf{p}(\{2\}, s_2 + s_1(1), L\mathbf{t} - t_1), \end{aligned}$$

and $\mathbf{p}(s_1, \{7\}, \mathbf{t} - u) = \mathbf{p}(s_1, \emptyset, \mathbf{t} - u)$. The computation of the quantity $\mathbf{p}(\{2\}, s_2 + s_1(1), L\mathbf{t} - t_1)$ require again recursion, until to obtain an empty set for one of the arguments.

When s_1 has no equality constraints and $s_2 = \emptyset$, one can invoke (31) $|s_1|$ times along with Remark 3.7 and (29) and get

Corollary 3.2. Let α be as in Theorem 3.3. If $|s_1| > 0$ equals the dimension of \mathbf{t} , then

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \mathbf{t})) = \alpha \mathbf{p}(s_1, \emptyset, \mathbf{t}) = \alpha \left(\prod_{n=1}^{|\mathbf{t}|} e^{\lambda(W_n)(t_n - t_{n-1})} \right) \mathbf{1} \quad (33)$$

where $W_n = [S(L^{n-1}(s_1))]^c$.

The formula (33) is a generalization of [4, equation (7)] to general finite state Markov processes.

If $s_1 = \emptyset$, we have no time constraints and $P_\alpha(\mathcal{T}(\emptyset, s_2, 0))$ reduces to the probability that certain equality and inequality constraints hold among the stopping times. This can be written as the solution of a sequence of linear equations whose defining matrices are submatrices of the intensity matrix. The details require further notation and are left to future work (or to the reader) except for the special case of $P_\alpha(\tau_1 = \tau_2)$ which we would like to use in what follows in order to relate our results to earlier works in the literature.

Define $\nu_0 := 0$, and for $n > 0$ $\nu_n := \inf\{u > \nu_{n-1}, X_u \neq X_{u-}\}$. The sequence $\{\nu_n\}$ is the jump times of the process X . Define a discrete time Markov chain \bar{X} with state space E as $\bar{X}_n := X_{\nu_n}$; it is called the embedded Markov chain of the process X . It follows from (9) that the one step transition matrix of \bar{X} is

$$\bar{\lambda} := \begin{cases} -\lambda(i, j)/\lambda(i, i), & \text{for } i \neq j, \\ 0, & \text{otherwise.} \end{cases}$$

Define $\mathbf{D} \in \mathbb{R}^{E \times E}$ as the diagonal matrix

$$D(i, j) = \begin{cases} -1/\lambda(i, i), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Left multiplying a matrix by \mathbf{D} divides its i^{th} row by $-\lambda(i, i)$. Therefore, $\bar{\lambda} = I + \mathbf{D}\lambda$.

Define $\bar{\tau}_k := \inf\{n : \bar{X}_n \in \Gamma_k\}$. The event $\{\tau_1 = \tau_2\}$ means that X hits the set Γ_1 and Γ_2 at the same time; because this event makes no reference to how time is measured, it can also be expressed in terms of \bar{X} as $\{\bar{\tau}_1 = \bar{\tau}_2\}$.

Define the column vector $\mathbf{q} \in \mathbb{R}^E$, $q(i) := P_i(\bar{\tau}_1 = \bar{\tau}_2)$. Conditioning on the initial position of X implies $P_\alpha(\bar{\tau}_1 = \bar{\tau}_2) = \alpha\mathbf{q}$. From here on we derive a formula for \mathbf{q} . Parallel to the arguments so far, we know that this event happens if and only if \bar{X} hits $\Gamma_1 \cap \Gamma_2$ before $B := (\Gamma_1 - \Gamma_2) \cup (\Gamma_2 - \Gamma_1)$. Set $w := (\Gamma_1 \cup \Gamma_2)^c$. The vector \mathbf{q} satisfies the boundary conditions

$$\mathbf{q}|_{\Gamma_1 \cap \Gamma_2} = 1 \text{ and } \mathbf{q}|_B = 0 \tag{34}$$

and is to be determined only for the states in w . If a state $i \in w$ cannot communicate with $\Gamma_1 \cap \Gamma_2$, $q(i)$ is trivially 0; let w' denote the set of states in w that can communicate with $\Gamma_1 \cap \Gamma_2$. The Markov property of \bar{X} implies that for $i \in w'$

$$q(i) = \sum_{j \in w'} \bar{\lambda}(i, j)q(j) + \sum_{j \in (\Gamma_1 \cap \Gamma_2)} \bar{\lambda}(i, j);$$

or in matrix notation (see (5)):

$$\begin{aligned}\mathbf{q}|_{w'} &= (\bar{\lambda}|_{w'}) \mathbf{q}|_{w'} + (\bar{\lambda}|_{w' \times (\Gamma_1 \cap \Gamma_2)}) \mathbf{1}|_{\Gamma_1 \cap \Gamma_2} \\ (I - \bar{\lambda})|_{w'} \mathbf{q}|_{w'} &= (\bar{\lambda}|_{w' \times (\Gamma_1 \cap \Gamma_2)}) \mathbf{1}|_{\Gamma_1 \cap \Gamma_2} \\ (-\mathbf{D}\lambda)|_{w'} \mathbf{q}|_{w'} &= (\bar{\lambda}|_{w' \times (\Gamma_1 \cap \Gamma_2)}) \mathbf{1}|_{\Gamma_1 \cap \Gamma_2}.\end{aligned}$$

For $i \neq j$, $\bar{\lambda}(i, j) = -\lambda(i, j)/\lambda(i, i) = (\mathbf{D}\lambda)(i, j)$ and in particular the same holds for $(i, j) \in w' \times (\Gamma_1 \cap \Gamma_2)$ and therefore

$$(-\mathbf{D}\lambda)|_{w'} \mathbf{q}|_{w'} = (-\mathbf{D}\lambda|_{w' \times (\Gamma_1 \cap \Gamma_2)}) \mathbf{1}|_{\Gamma_1 \cap \Gamma_2}.$$

There is no harm in taking the diagonal \mathbf{D} out of the projection operation on both sides of the last display:

$$\lambda|_{w'} \mathbf{q}|_{w'} = \lambda|_{w' \times (\Gamma_1 \cap \Gamma_2)} \mathbf{1}|_{\Gamma_1 \cap \Gamma_2}.$$

That all states in w' can communicate with $\Gamma_1 \cap \Gamma_2$ implies that the matrix on the left is invertible and therefore

$$\mathbf{q}|_{w'} = (\lambda|_{w'})^{-1} \lambda|_{w' \times (\Gamma_1 \cap \Gamma_2)} \mathbf{1}|_{\Gamma_1 \cap \Gamma_2}. \quad (35)$$

3.5 A second representation of tail probabilities

For a nonnegative integer n , denote by $\mathcal{P}(n)$ the set of all *sub*permutations of $\{1, 2, 3, \dots, n\}$, e.g., $\mathcal{P}(2) = \{\emptyset, (1), (2), (1, 2), (2, 1)\}$. The tail probability formula (30) conditions on the first time τ' that one of the equality constraints is attained in the time interval $[0, t_1]$ and writes what happens after that as a recursion. What can happen between τ' and t_1 ? A number of other equalities can be attained and rather than pushing these into the recursion, one can treat them inside the integral using a density similar to (17):

$$\begin{aligned}\mathbf{p}(s_1, s_2, \mathbf{t}) &= \sum_{\pi \in \mathcal{P}(|s_2|)} \left(\int_{A_\pi} \left(\prod_{n=1}^{|\pi|} e^{(v_n - v_{n-1})\lambda(W_n)} \mathbf{J}_n \right) e^{(t_1 - v_{|\pi|})\lambda(W)} dv \right) \\ &\quad \cdot \mathbf{p}(Ls_1, s_2 - s_2(\pi) + s_1(1), L\mathbf{t} - t_1),\end{aligned} \quad (36)$$

where $v_0 = 0$ and

$$\begin{aligned}W_n &:= [S(s_1 \cup s_2 - \cup_{n_1=1}^n s_2(\pi(n_1)))]^c, \quad T_n := \left[\bigcap_{k \in s_2(\pi(n))} \Gamma_k \right] \cap W_{n+1}, \\ s_2(\pi) &:= \cup_{m=1}^{|\pi|} s_2(\pi(m)), \quad W := [S(s_1 \cup s_2 - s_2(\pi))]^c, \\ A_\pi &:= \left\{ v \in \mathbb{R}^{|\pi|} : 0 < v_1 < v_2 < \dots < v_{|\pi|} \leq t_1 \right\}, \\ \mathbf{J}_n &:= \lambda(W_n, T_n),\end{aligned}$$

dv is the $|\pi|$ dimensional Lebesgue measure on $\mathbb{R}^{|\pi|}$ for $|\pi| > 0$; $A_\pi := \{0\}$ and dv is the trivial measure on $\{0\}$ for $|\pi| = 0$. The proof involves no additional ideas and is omitted.

3.6 Conditional formulas

The proof of Theorem 3.2 shows how one can use the density formula (17) to write down the regular conditional distribution of τ given $\mathcal{F}_{\tau'}$. One can do the same for \mathcal{F}_{u_0} , where $u_0 \in \mathbb{R}_+$ is a given deterministic time. To that end, introduce the set valued process

$$V_u := \{k \in K, \tau_k < u\}.$$

K is finite, then so is its power set 2^K , thus V_u takes values in a finite set. The set V_u is the collection of Γ_k that X has visited up to time u . For ease of notation we will denote the complement of V_u by \bar{V}_u . The times $\tau|_{V_{u_0}}$ are known by time u_0 and hence they are constant given \mathcal{F}_{u_0} . Thus, we only need to write down the regular conditional density of $\tau|_{\bar{V}_{u_0}}$, i.e., the hitting times to the Γ_k that have not been visited by time u_0 . From here on the idea is the same as in the proof of Theorem 3.2. Define $\hat{X}_u := X_{u+u_0}$ and for $k \in \bar{V}_{u_0}$

$$\hat{\tau}_k := \inf\{u : \hat{X}_u \in \Gamma_k\}.$$

The definitions of \hat{X} and $\hat{\tau}$ imply

$$\hat{\tau} = \tau|_{\bar{V}_{u_0}} - u_0. \quad (37)$$

$\hat{X}_0 = X_{u_0}$ is a constant given \mathcal{F}_{u_0} . Thus the process \hat{X} has exactly the same distribution as X with initial point X_{u_0} and Theorem 3.2 applies and gives the density of $\hat{\tau}$, which is, by (37), the regular conditional distribution of $\tau|_{\bar{V}_{u_0}} - u_0$. Therefore, for any bounded measurable $g : \mathbb{R}^{\bar{V}_{u_0}} \rightarrow \mathbb{R}$ and a partition s' of \bar{V}_{u_0}

$$\mathbb{E} \left[g \left(\tau|_{\bar{V}_{u_0}} \right) 1_{R_{s'}} \left(\tau|_{\bar{V}_{u_0}} \right) | \mathcal{F}_{u_0} \right] = \int_{R_{s'}} g(u_0 + u) f(\delta_{X_{u_0}}, u, \bar{V}_{u_0}) d_{s'} u.$$

We record this as

Proposition 3.2. *The regular conditional density of $\tau|_{\bar{V}_{u_0}} - t_0$ given \mathcal{F}_{u_0} is $f(\delta_{X_{u_0}}, t, \bar{V}_{u_0})$.*

4 Absorbing $\{\Gamma_k\}$ and connections to earlier results

In this section, we show how our formula can be simplified in the case where the Γ_k are absorbing. Then, we prove that the results obtained in the literature in [4, 10] are a particular case of our results. Let us emphasize that in the literature, only a formula for the non singular part of the density is obtained in a multidimensional setting, under the assumption that all the

Γ_k are absorbing, and that the density of singular parts is derived for the two dimensional case only. As we shall see, one of the difficulty to compare the two results is that we are working with $|E| \times |E|$ matrices while the authors of the quoted papers, using the absorbing property have chosen to reduce the dimension. They also introduce some notation to simplify the matrices they are working with to take into account this absorbing property, and we have to recall these notation.

4.1 Density formula for absorbing $\{\Gamma_k\}$

We recall that a nonempty subset $a \subset E$ is said to be *absorbing* if $\lambda(i, j) = 0$ for all $i \in a$ and $j \in a^c$, i.e., if $\boldsymbol{\lambda}(a, a^c) = 0$.

In a first step, we derive an alternative expression for the density formula (17) under the assumption that all $\{\Gamma_k, k \in K\}$ are absorbing.

Proposition 4.1. *If a is absorbing and $\alpha|_a = 0$, denoting*

$$p_{\alpha,a}^u(j) := P_\alpha(X_u = j, X_v \notin a, \forall v \leq u)$$

we have

$$p_{\alpha,a}^u = \alpha e^{\boldsymbol{\lambda}(a^c)u} = \alpha e^{\boldsymbol{\lambda}u} \mathbf{I}_{a^c} \quad (38)$$

Proof. We already know from Lemma 3.1 that the first equality holds. Therefore, it only remains to show $p_{\alpha,a}^u = \alpha e^{\boldsymbol{\lambda}u} \mathbf{I}_{a^c}$. It is well known (see e.g., [2, Theorem 3.4, page 48]) that the distribution of X at time u is $\alpha e^{\boldsymbol{\lambda}u}$, i.e., $P_\alpha(X_u = j) = \alpha e^{\boldsymbol{\lambda}u}(j)$ for all $j \in E$. The fact that a is absorbing implies that if $X_{u_0} \in a$ then $X_u \in a$ for all $u \geq u_0$. Therefore for $j \in a^c$

$$P_\alpha(X_u = j) = P_\alpha(X_u = j, X_v \notin a, v \leq u),$$

i.e.,

$$(\alpha p_{\alpha,a}^u)|_{a^c} = (\alpha e^{\boldsymbol{\lambda}u} \mathbf{I}_{a^c})|_{a^c}. \quad (39)$$

The definition of $p_{\alpha,a}^u$ and $\alpha|_a = 0$ imply $(\alpha p_{\alpha,a}^u)|_a = 0$; The definition of \mathbf{I}_{a^c} implies $(\alpha e^{\boldsymbol{\lambda}u} \mathbf{I}_{a^c})|_a = 0$. This and (39) imply (38). \square

Proposition 4.1 says the following: if a is absorbing then $\alpha e^{\boldsymbol{\lambda}(a^c)u}$ is the same as $\alpha e^{\boldsymbol{\lambda}u} \mathbf{I}_{a^c}$ and both describe the probability of each state in a^c at time t over all paths that avoid a in the time interval $[0, t]$. The first expression ensures that all paths under consideration avoid the set a by setting the jump rates into a to 0. The second expression does this by striking out those paths that end up in one of the states in a (the \mathbf{I}_{a^c} term does this); this is enough because a is absorbing: once a path gets into a it will stay there and one can look at the path's position at time t to figure out whether its weight should contribute to $p_{\alpha,a}^u$. In the general case this is not possible, hence the $\boldsymbol{\lambda}(a^c)$ in the exponent.

The previous proposition implies that one can replace the $\boldsymbol{\lambda}(W_n)$ that appears in the density formula (17) with $\boldsymbol{\lambda}$:

Proposition 4.2. *Assume that all Γ_k are absorbing, and let α be such that $\alpha|_{W_1^c} = 0$. Then, for any $s \in \mathcal{S}$ and $\mathbf{t} \in R_s$ we have*

$$f_s(\alpha, \mathbf{t}, K) = \alpha \left(\prod_{n=1}^{|\mathbf{s}|} e^{\lambda(\bar{t}_n - \bar{t}_{n-1})} \boldsymbol{\lambda}(W_n, T_n) \right) \mathbf{1}, \quad (40)$$

where f_s is the density given in Theorem 3.2.

Proof. First, assume that $|\mathbf{s}| = 1$. Then we have

$$\begin{aligned} f_s(\alpha, \mathbf{t}, K) &= \alpha e^{\lambda(W_1)(\bar{t}_1 - \bar{t}_0)} \boldsymbol{\lambda}(W_1, T_1) \mathbf{1} \\ &= \alpha e^{\lambda(\bar{t}_1 - \bar{t}_0)} \mathbf{I}_{W_1} \boldsymbol{\lambda}(W_1, T_1) \mathbf{1} = \alpha e^{\lambda(\bar{t}_1 - \bar{t}_0)} \mathbf{I}_{W_1} \mathbf{I}_{W_1} \boldsymbol{\lambda}_{T_1} \mathbf{1} \quad (41) \\ &= \alpha e^{\lambda(\bar{t}_1 - \bar{t}_0)} \mathbf{I}_{W_1} \boldsymbol{\lambda}_{T_1} \mathbf{1} = \alpha e^{\lambda(\bar{t}_1 - \bar{t}_0)} \boldsymbol{\lambda}(W_1, T_1) \mathbf{1}, \end{aligned}$$

where the second equality follows from (38) since $\alpha|_{W_1^c} = 0$, and the third and the last equalities come from (7). So, the result is true for s such that $|\mathbf{s}| = 1$.

We now give a proof in the case $|\mathbf{s}| = 2$. In this case, we have

$$\begin{aligned} f_s(\alpha, \mathbf{t}, K) &= \alpha e^{\lambda(W_1)(\bar{t}_1 - \bar{t}_0)} \boldsymbol{\lambda}(W_1, T_1) e^{\lambda(W_2)(\bar{t}_2 - \bar{t}_1)} \boldsymbol{\lambda}(W_2, T_2) \mathbf{1} \\ &= \alpha_1 e^{\lambda(W_2)(\bar{t}_2 - \bar{t}_1)} \boldsymbol{\lambda}(W_2, T_2) \mathbf{1}, \end{aligned}$$

where $\alpha_1 := \alpha e^{\lambda(W_1)(\bar{t}_1 - \bar{t}_0)} \boldsymbol{\lambda}(W_1, T_1)$. Now, in view of the definition of T_1 , and using (7), we see that α_1 satisfies $\alpha_1|_{W_2^c} = 0$. Consequently, invoking (38), we conclude that

$$\alpha_1 e^{\lambda(W_2)(\bar{t}_2 - \bar{t}_1)} \boldsymbol{\lambda}(W_2, T_2) = \alpha_1 e^{\lambda(\bar{t}_2 - \bar{t}_1)} \boldsymbol{\lambda}(W_2, T_2),$$

which, together with (41), demonstrates the result in case $|\mathbf{s}| = 2$. The general result follows by induction on the value of $|\mathbf{s}|$. \square

Let us briefly point out another possible modification of the density formula for absorbing $\{\Gamma_k\}$. Define

$$\widehat{T}_0 = E - S(s), \quad \widehat{T}'_n := \bigcap_{\cup_{m \leq n} s(m)} \Gamma_k, \quad \widehat{T}_n = \widehat{T}'_n - S(L^n(s)), \quad \widehat{W}_n = \widehat{T}_{n-1},$$

where $s \in \mathcal{S}$ and $n \in \{1, 2, 3, \dots, |\mathbf{s}|\}$. If $\{\Gamma_k\}$ are absorbing one can replace the target and waiting sets T_n and W_n of (16) with \widehat{T}_n and \widehat{W}_n defined above. One can prove that the density formula continues to hold after this modification with an argument parallel to the proof of Proposition 4.2 using in addition that the intersection of absorbing sets is again absorbing.

4.2 Tail probabilities for absorbing $\{\Gamma_k\}$

When $\{\Gamma_k, k \in K\}$ are absorbing, then, in view of (38), one can write the tail probability that appears in Theorem 3.3, as

$$P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t})) = \alpha \int_0^{t_1} e^{\lambda u} \boldsymbol{\lambda}(W, T(s_1, s_2)) \left(\sum_{n=1}^{|s_2|} \mathbf{I}_{T_n(s_1, s_2)} \mathbf{p}(s_1, s_2 - s_2(n), \mathbf{t} - u) \right) du + \alpha e^{\lambda t_1} \mathbf{I}_W \mathbf{p}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1)$$

and, in particular,

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \mathbf{t})) = \alpha e^{\lambda t_1} \mathbf{I}_{S(s_1)^c} \mathbf{p}(s_1 - s_1(1), s_1(1), L\mathbf{t} - t_1). \quad (42)$$

4.3 Connections with earlier results

This subsection gives several examples of how to express density/distribution formulas from the prior phase-type distributions literature as special cases of the ones derived in the present work.

4.3.1 Examples related to [4]

We begin with relating to our results three formulas given in [4]. The first two concern a single hitting time and the last one a pair of hitting times. In [4], the authors assume that E has an absorbing element called Δ , they define $T := \inf\{u : X_u = \Delta\}$ and they denote by A what in our paper is given as $\boldsymbol{\lambda}|_{\{\Delta\}^c}$. Moreover, [4] also uses the letter α to denote the initial distribution of X , but on the set $\widehat{E} := E - \{\Delta\}$, rather than on the set E as it is done here; in particular, in [4] it is implicitly assumed that $P(X_0 = \Delta) = 0$. We will use the symbol $\widehat{\alpha}$ to denote the ‘ α of [4].’ The relation between α and $\widehat{\alpha}$ is $\alpha|_{\{\Delta\}^c} = \widehat{\alpha}$.

The first line of [4, display (2), page 690] says $P_\alpha(T > u) = \widehat{\alpha} e^{Au} \mathbf{e}$ where \mathbf{e} is the $|E| - 1$ dimensional vector with all components equal to 1. The corresponding formula in the present work is (12) where one takes $d = \{\Delta\}$. From

$$\boldsymbol{\lambda}(d^c)|_{d^c} = A \quad (43)$$

and from the fact that

$$\text{the row of } \boldsymbol{\lambda}(d^c) \text{ corresponding to } \Delta \text{ is } 0 \quad (44)$$

we obtain the equality of these formulas.

The second line of the same display gives $-\widehat{\alpha} e^{uA} A \mathbf{e}$ as the density of T . The corresponding formula here is (13) with $b = \emptyset$, and $a = \{\Delta\}$ for which it reduces to $e^{u\boldsymbol{\lambda}(a^c)} \boldsymbol{\lambda}(a^c, a) \mathbf{1}$. This time, (43), (44) and the fact that the row

sums of $\boldsymbol{\lambda}$ are zero (so that, $\boldsymbol{\lambda}(a^c, \Delta)|_{a^c} = \boldsymbol{\lambda}|_{a^c} \mathbf{e} = A\mathbf{e}$), imply the equality of the formulas.

The matrix $\boldsymbol{\lambda}(a^c, a)$ is the column of $\boldsymbol{\lambda}$ corresponding to Δ ; one way to write it is as the negative of the sums of the rest of the columns, this is what the last equality says.

The case of a pair of hitting times appears in [4]: using the notation of that paper, we are given two sets $\Gamma_1, \Gamma_2 \subset E$ with $\Gamma_1 \cap \Gamma_2 = \{\Delta\}$, T_k is the first hitting time to Γ_k . The formula [4, Equation (5), page 692] says

$$P_\alpha(T_1 = T_2 > u) = \widehat{\alpha} e^{Au} A^{-1} (Ag_1g_2 - [A, g_1] - [A, g_2])\mathbf{e}, \quad (45)$$

where $g_k = \mathbf{I}_{\Gamma_k}|_{\{\Delta\}^c}$ and for two matrices B and C , $[B, C] := BC - CB$. The absorbing property of Γ_1 and Γ_2 implies that the matrix inside the parenthesis in (45) equals $g'A$, where $g' = \mathbf{I}_{(\Gamma_1 \cup \Gamma_2)^c}|_{\widehat{E}}$ i.e., the same matrix as A except that the rows whose indices appear in $\Gamma_1 \cup \Gamma_2$ are replaced with 0. Thus $(Ag_1g_2 - [A, g_1] - [A, g_2])\mathbf{e}$ is another way to take the Δ column of $\boldsymbol{\lambda}$ and replace its components whose indices appear in $\Gamma_1 \cup \Gamma_2$ with 0. Denote this vector by C_Δ . Then the right side of (45) is

$$\alpha|_{\widehat{E}} \left(e^{\lambda u}|_{\widehat{E}} \right) A^{-1} C_\Delta. \quad (46)$$

The same probability is expressed by a special case of (42); for the present case one sets $K = \{1, 2\}$, $s_1 = (\{1, 2\})$; for these values, (31) and conditioning on the initial state gives

$$P_\alpha(\tau_1 = \tau_2 > u) = \alpha e^{\lambda u} \mathbf{I}_w \mathbf{p}(\emptyset, (\{1, 2\}), 0), \quad (47)$$

where $w = (\Gamma_1 \cup \Gamma_2)^c$.

Remark 4.1. Note that the event $\{\tau_1 = \tau_2 > u\}$ is of the form (27), but the event $\mathcal{T}(\emptyset, (\{1, 2\}), 0)$ is of the form (28).

Recall that we have denoted the last probability as q and derived for it the formulas (34) and (35). The article [4] assumes that all states can communicate with Δ , which implies that w' of (35) equals w . This and $\Gamma_1 \cap \Gamma_2 = \{\Delta\}$ imply $\boldsymbol{\lambda}|_{w' \times \Gamma_1 \cap \Gamma_2} \mathbf{1}$ in (35) equals $\boldsymbol{\lambda}|_{(\Gamma_1 \cup \Gamma_2)^c \times \Delta}$ i.e., the Δ column of $\boldsymbol{\lambda}$ projected to its indices in $(\Gamma_1 \cup \Gamma_2)^c$, i.e., $C_\Delta|_w$. The only difference between C_Δ and $C_\Delta|_w$ is that the former has zeros in its extra dimensions. This and the absorbing property of Γ_k imply

$$(\boldsymbol{\lambda}|_w)^{-1} C_\Delta|_w = (A^{-1} C_\Delta)|_w.$$

Note that we are commuting the projection operation and the inverse operation; this is where the absorbing property is needed. The last display, (34) and (35) give $I_w \mathbf{p}(\emptyset, (\{1, 2\}), 0)|_{\widehat{E}} = q|_{\widehat{E}} = A^{-1} C_\Delta$. This and $\alpha(\{\Delta\}) = 0$ imply that one can rewrite the right side of (47) as

$$\alpha e^{\lambda u} (A^{-1} C_\Delta)|^E.$$

Once again $\alpha(\Delta) = 0$ implies that the last expression equals (46).

4.3.2 Example related to [10]

The density formula [10, (3.1.11)] will provide our last example. The process X of [10] is a Markov jump process (with absorbing boundary) taking values in $\mathbb{Z}_2^m := \{0, 1\}^m$ (the m -fold Cartesian product), with jumps valued in $\{-e_k, k = 1, 2, 3, \dots, m\}$, where e_k is the unit vector with k^{th} coordinate equal to 1 ([10] uses different but equivalent notation, in particular the name of the process is Y and its state space is represented by subsets of $\{1, 2, 3, \dots, m\}$; the notation of this paragraph is chosen to ease the presentation). In [10] the absorbing sets are denoted as Δ_i , see the display after [10, (2.3)], and they correspond to $\Gamma_k = \{z \in \mathbb{Z}_2^m : z_k = 0\}$ in our present set-up. The jump rate for the increment $-e_k$ is assumed to be $\langle X, b_k \rangle + a_k$ for fixed $b_k \in \mathbb{R}^m$ and $a_k \in \mathbb{R}$ (given in [10, (2.1)]). A key property of this setup is this: take any collection $\{\Gamma_{k_1}, \Gamma_{k_2}, \dots, \Gamma_{k_n}\}$ with $n > 1$; because the only increments of X are the $\{-e_k\}$, the process cannot enter the sets in the collection at the same time. Thus, in this formulation, X must hit the $\{\Gamma_k\}$ at separate times and the distribution of τ has no singular part, i.e., $P(\tau \in R_s) = 0$ for $|s| < m$, and one needs only the density of τ with respect to the full Lebesgue measure in \mathbb{R}^m (the ‘‘absolutely continuous part’’). As noted in [10], this is already available in [4] (see the display following (7) on page 694) and is given in [10, display (3.1.1)] as follows:

$$f(\mathbf{t}) = (-1)^m \alpha \left(\prod_{n=1}^{m-1} e^{\lambda(\bar{t}_n - \bar{t}_{n-1})} (\lambda G_{k_n} - G_{k_n} \lambda) \right) e^{\lambda(\bar{t}_m - \bar{t}_{m-1})} \lambda G_{k_m} \mathbf{1}, \quad (48)$$

for $\mathbf{t} \in R_s$ with $|s| = m$; here $G_k = \mathbf{I}_{\Gamma_k^c}$ and k_n is the index for which $t_{k_n} = \bar{t}_n$ ([10] uses the letter Q for the rate matrix λ). We briefly indicate why (40) is equivalent to the last formula with the assumptions of this paragraph, i.e., when the dynamics of X precludes it to enter more than one of the $\{\Gamma_k\}$ at the same time and in particular when $|s|$ equals the dimension of τ (denoted by m in the current paragraph). Lemma 2.1 and the absorbing property of Γ_k imply

$$\begin{aligned} \lambda G_k - G_k \lambda &= \lambda(\Gamma_k, \Gamma_k^c) - \lambda(\Gamma_k^c, \Gamma_k) \\ &= -\lambda(\Gamma_k^c, \Gamma_k). \end{aligned}$$

On the other hand again Lemma 2.1 and the absorbing property of Γ_k imply $\lambda G_k = -\lambda(\Gamma_k^c, \Gamma_k^c)$. The row sums of λ equal 0. The last two facts imply $\lambda G_k \mathbf{1} = -\lambda(\Gamma_k^c, \Gamma_k) \mathbf{1}$. These imply that one can write (48) as

$$f(\mathbf{t}) = \alpha \left(\prod_{n=1}^m e^{\lambda(\bar{t}_n - \bar{t}_{n-1})} \lambda(\Gamma_{k_n}^c, \Gamma_{k_n}) \right) \mathbf{1}.$$

As we noted above, for $k \neq k'$ the dynamics of X imply that it cannot enter Γ_k and $\Gamma_{k'}$ at the same time. Furthermore, by definition $t_n \neq t_{n'}$ for

Figure 1: The level curves of the density f for $\tau_2 = \tau_3 < \tau_1$. On the right: the values of f over the line segment connecting $(0, 0)$ to $(0.5, 1)$

$n \neq n'$. Finally, the initial distribution α is assumed to be such that it puts zero mass on $\cup_{k \in K}^m \Gamma_k$. These imply that one can replace $\lambda(W_n, T_n)$ of (40) with $\lambda(\Gamma_{k_n}^c, \Gamma_{k_n})$ (a full argument requires an induction similar to the proof of Proposition 4.2), and therefore under the current assumptions the last display and (40) are equal.

5 Numerical Example

The state space of our numerical example is $E = \mathbb{Z}_3^3$. For $z \in \mathbb{Z}_3^3$ and $k \in K = \{1, 2, 3\}$ let z_k denote the k^{th} component of z . For the collection $\{\Gamma_k\}$ take

$$\Gamma_k = \{z : z_k = 0\}.$$

τ_k , as before, is the first time the process X hits the set Γ_k . The initial distribution α will be the uniform distribution over the set

$$E - \bigcup_{k \in K} \Gamma_k = \left\{ z : \min_{k \in K} z_k > 0 \right\}.$$

We will compute the density of $\tau = (\tau_1, \tau_2, \tau_3)$ over the sets $R_{s_1}, R_{s_2} \subset \mathbb{R}_+^3$ defined by the partitions $s_1 = (\{2, 3\}, \{1\})$ and $s_2 = (\{1, 2, 3\})$; the first corresponds to the event $\{\tau \in R_{s_1}\} = \{\tau_2 < \tau_1 = \tau_3\}$ and the second to $\{\tau \in R_{s_2}\} = \{\tau_1 = \tau_2 = \tau_3\}$.

The dynamics of X on \mathbb{Z}_3^3 for our numerical example will be that of a constrained random walk with the following increments:

$$\pm e_k, \pm(e_1 + e_2), \pm(e_1 + e_2 + e_3), k \in K, \quad (49)$$

where $e_1 := (1, 0, 0)$, $e_2 := (0, 1, 0)$ and $e_3 := (0, 0, 1)$; the $\{\Gamma_k\}$ are assumed to be absorbing, i.e., if $X_{u_0} \in \Gamma_k$ any increment involving $\pm e_k$ can no longer be an increment of X for $u > u_0$. The sets $B_k := \{z : z_k = 2\}$ are “reflecting” in the sense that if $X_t \in B_k$ for some t , increments involving $+e_k$ cannot be the first increment of X in the time interval $[t, \infty)$. We assume the following jump rates for the increments listed in (49):

$$2, 1, 2, 1, 3, 1, 0.5, 0.5, 0.2, 0.2.$$

These rates and the aforementioned dynamics give a 27×27 λ matrix. The level sets $f(\alpha, \cdot, K)|_{R_{s_1}}$ are depicted in Figure 1 and the graph of $f(\alpha, \cdot, K)|_{R_{s_2}}$ is depicted in Figure 2.

For the parameter values of this numerical example, $P_\alpha(\cap_{k \neq k'} \tau_k \neq \tau_{k'}) = 0.899$ and thus the singular parts account for around 10% of the distribution of τ .

Figure 2: The density f for $\tau_1 = \tau_2 = \tau_3$

6 Conclusion

Our primary motivation in deriving the formulas in the present paper has been their potential applications to credit risk modeling. Let us comment on this potentiality starting from the credit risk model of [10]. With the results in the present work one can extend the modeling approach of [10] in two directions. Remember that the underlying process in [10] can only move by increments of $\{-e_k\}$ i.e., the model assumes that the obligors can default only one at a time. However, for highly correlated obligors it may make sense to allow simultaneous defaults, i.e., allow increments of the form $-\sum_n e_{k_n}$. Once multiple defaults are allowed the default times will have nonzero singular parts and the formulas in the present work can be used to compute them, as is done in the numerical example of Section 5. Secondly, the default sets $\{\Gamma_k\}$ no longer have to be assumed to be absorbing. Thus, with our formulas, one can treat models that allow recovery from default.

As $|E|$ increases (17) and other formulas derived in the present paper can take too long a time to compute (the same holds for earlier density formulas in the prior literature). Thus it is of interest to derive asymptotic approximations for these densities.

References

- [1] David Aldous and Larry Shepp, *The least variable phase type distribution is Erlang*, *Stochastic Models* **3** (1987), no. 3, 467–473.
- [2] Søren Asmussen, *Applied probability and queues*, Springer, 2003.
- [3] Søren Asmussen and Hansjörg Albrecher, *Ruin probabilities*, vol. 14, World Scientific, 2010.
- [4] David Assaf, Naftali A. Langberg, Thomas H. Savits, and Moshe Shaked, *Multivariate phase-type distributions*, *Operations Research* **32** (1984), no. 3, 688–702.
- [5] Mathieu Bargs, Stéphane Loisel, and Xavier Venel, *On finite-time ruin probabilities with dependence between reinsurance cycles and the claim arrival process*, *Scandinavian Actuarial Journal* **3** (2013), 163–185.
- [6] Alexandr Borovkov, *Probability theory*, Springer, 2009.
- [7] Stéphane Crépey, Tomasz R. Bielecki, and Damiano Brigo, *Counterparty risk and funding: A tale of two puzzles*, Chapman & Hall/CRC Financial Mathematics Series, CRC Press, Boca Raton, FL, 2014.

- [8] Nicole El Karoui, Monique Jeanblanc, and Ying Jiao, *Density approach in modelling successive defaults*, preprint (2014).
- [9] Agner Krarup Erlang, *Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges*, *Elektroteknikereren* **13** (1917), 5–13.
- [10] Alexander Herbertsson, *Modelling default contagion using multivariate phase-type distributions*, *Review of Derivatives Research* **14** (2011), no. 1, 1–36.
- [11] Ying Jiao and Shanqiu Li, *Generalized density approach in progressive enlargement of filtrations*, preprint (2014).
- [12] Mary A. Johnson and Michael R. Taaffe, *Matching moments to phase distributions: Mixtures of Erlang distributions of common order*, *Stochastic Models* **5** (1989), no. 4, 711–743.
- [13] Younes Kchia and Martin Larsson, *Credit contagion and risk management with multiple non-ordered defaults*, arxiv.org/abs/1104.5272 (2013).
- [14] Marcel F. Neuts, *Probability distributions of phase type*, *Liber Amicorum Prof. Emeritus H. Florin* **173** (1975), 206.
- [15] ———, *Matrix-geometric solutions in stochastic models: an algorithmic approach*, Courier Dover Publications, 1981.
- [16] Richard P. Stanley, *Enumerative combinatorics*, vol. 49, Cambridge university press, 2011.
- [17] Ryszard Syski, *Passage times for Markov chains*, vol. 1, Ios Press, 1992.