Joint Hitting-Time Densities for Finite State Markov Processes

Tomasz R. Bielecki, Monique Jeanblanc, Devin Sezer

▶ **To cite this version:**

Tomasz R. Bielecki, Monique Jeanblanc, Devin Sezer. Joint Hitting-Time Densities for Finite State Markov Processes. 2015. hal-01250956

HAL Id: hal-01250956
https://hal.science/hal-01250956

Preprint submitted on 8 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

December 18, 2015

Abstract

For a finite state Markov process X 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. In general, X may enter some of the Γ_k at the same time and therefore the vector $\boldsymbol{\tau} := (\tau_k, k \in K)$ may put nonzero mass over lower dimensional regions of $\mathbb{R}_+^{|K|}$; these regions are of the form $R_s = \{t : t_i = t_j,\ \ i,j \in s(1)\} \cap \bigcap_{l=2}^{|s|} \{t : t_m < t_i = t_j,\ \ i,j \in s(l), m \in s(l-1)\}$ where s is any ordered partition of the set K and $s(j)$ denotes the j^{th} subset of K in the partition s. When $|s| < |K|$, the density of the law of $\boldsymbol{\tau}$ over these regions is said to be "singular" because it is with respect to the $|s|$-dimensional Lebesgue measure over the region R_s. We derive explicit/recursive and simple to compute formulas for these singular densities and their corresponding tail probabilities over all R_s as s ranges over ordered partitions of K. We give a numerical example and indicate the relevance of our results to credit risk modeling.

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, which go back to Erlang [11], are used to model

*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`

a wide range of phenomena in, e.g., reliability theory, communications systems, insurance and finance. The literature on these distributions is immense, see, e.g., [2, 3, 1, 14, 20, 22].

To the best of our knowledge, Assaf et al. [5] were the first 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 law 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 the distribution of the random vector $\boldsymbol{\tau} := (\tau_k, k \in K)$ is a $|K|$ dimensional phase-type distribution, where $|K|$ denotes the number of elements in K. In general, the underlying process can hit some of the Γ_k simultaneously and this implies that multidimensional phase-type distributions can put nonzero mass on certain lower dimensional regions of $\mathbb{R}_+^{|K|}$; e.g., for $|K| = 3$ some of these regions are $\{t : t_1 = t_2 < t_3\}$, $\{t : t_1 = t_2 = t_3\}$, $\{t : t_2 < t_1 = t_3\}$, etc. In general, each ordered partition s of K defines an $|s|$-dimensional subset R_s of $\mathbb{R}_+^{|K|}$ (where $|s|$ denotes the number of subsets of K which appears in the partition s; subsection 2.2 gives the precise definitions) over which the law of $\boldsymbol{\tau}$ may put nonzero probability mass. The law of $\boldsymbol{\tau}$, when restricted to one of these lower dimensional regions, turns out to have a density with respect to the $|s|$-dimensional Lebesgue measure of that region; these densities are called "singular" (or "the singular part(s) of the density of $\boldsymbol{\tau}$") because of the lower dimensionality of R_s. *The focus of the present paper is on these singular densities of $\boldsymbol{\tau}$; our goal is to find simple formulas for them and the tail probabilities associated with them.* To the best of our knowledge, the only paper currently available which develops density or tail probability formulas for the singular parts is [5], which focuses on the case of $|K| = 2$ and Γ_1, Γ_2 absorbing. The only currently available density formula for $|K| > 2$ was also derived in [5] and *covers only the absolutely continuous part of the density (i.e., the density over the $|K|$ dimensional region $\{t \in \mathbb{R}_+^{|K|} : t_i \neq t_j$, for $i \neq j\}$) in the case when Γ_k, $k \in K$, are assumed absorbing*; display (39) in subsection 4.3 gives this density formula from [5] (this formula is stated without a proof in [5]; [12] provides a proof for it). Over the last three decades this formula has found use in a range of application areas, e.g., [18, modeling of plant development], [9, insurance] and [12, credit risk].

The main contributions of the present paper are Theorem 3.1, which gives an explicit formula for the singular density of the random vector $\boldsymbol{\tau}$ over each $R_s \subset \mathbb{R}_+^{|K|}$ as s ranges over all partitions of K, covering all possible singular and nonsingular parts and Theorem 3.2 which gives a recursive formula for the tail probabilities of $\boldsymbol{\tau}$ using the density formulas. We make no assumptions on whether $\{\Gamma_k, k \in K\}$ are absorbing and Theorem 3.1 gives a general formula for the joint density of a collection of first hitting times for any finite state Markov process X. The density formula when Γ_k are absorbing follows as a special case (Proposition 4.2).

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 method used in [5, 12]. 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 more complex objects than the corresponding densities and if one tries to compute these singular tail probabilities (for example, using the methodology of taboo probabilities as presented in Syski [22]) one quickly runs into difficult calculations even when $|K|$ is small. For this reason, rather than first trying to compute the tail probabilities of $\boldsymbol{\tau}$, we directly compute the singular densities of $\boldsymbol{\tau}$ over each $R_s \subset \mathbb{R}_+^{|K|}$ using the following idea: for each $\boldsymbol{t} \in R_s \subset \mathbb{R}_+^{|K|}$, the event $\{\boldsymbol{\tau} = \boldsymbol{t}\}$ corresponds to the limit of a specific set of trajectories of the Markov process whose (vanishing) probability can be written in terms of the exponentials of submatrices of the intensity matrix. These sets of

trajectories are of the following form: the process alternates at precise jump times between a sequence of waiting sets W_1, W_2,..., W_k and target sets T_1, T_2,..., T_k where $k = |s|$ is the dimension of R_s. The waiting and the target sets and the jump times are all determined by $\boldsymbol{t}$. Subsection 3.1 gives the one step version of this idea in the computation of the density of a single τ_k, given as Proposition 3.1. The same idea extends to multiple hitting times in subsection 3.2 and the multidimensional density is given as formula (16) in Theorem 3.1. The formula (16) gives the density of all singular parts of the distribution of $\boldsymbol{\tau}$ (over all $2^{|K|}$ hyperplanes of $\mathbb{R}_+^{|K|}$ of dimensions ranging from $|K|$ to 1); the ability to cover all singular parts is the novelty of the result. It accomplishes this by using exponentials of appropriate *sub*matrices of the rate matrix $\boldsymbol{\lambda}$. The sequence of waiting and target sets determine which submatrices of $\boldsymbol{\lambda}$ appear in the density formula. In contrast, the well-known density formula (39) from [5] for the absolutely continuous part involves only exponentials of the full rate matrix.

The proof of Theorem 3.1, although based on elementary ideas, seems to be novel. It starts from the following idea: f is the density of $\boldsymbol{\tau}$ if and only if $\mathbb{E}[g(\boldsymbol{\tau})] = \int f(x)g(x)dx$ for all bounded continuous functions g. Because $\boldsymbol{\tau}$ has an arbitrary finite dimension (since $|K|$ and $|s|$ are arbitrary integers), the proof must be recursive; and to deal with the generality of the result, novel notation needs to be introduced and several conditional expectations need to be treated carefully.

By a tail event we mean events of the form $\cap_{k\in K}\{\tau_k \geq t_k\}$. In general, the singular parts of the distribution of $\boldsymbol{\tau}$ make tail events complex because they may (and in general will) intersect with the lower dimensional regions $R_s \subset \mathbb{R}_+^{|K|}$ on which $\boldsymbol{\tau}$ puts nonzero mass. This complexity precludes the possibility of obtaining simple closed form formulas for the probabilities of these events (and hence for the distribution function of $\boldsymbol{\tau}$). A much more efficient way to represent tail probabilities (in the presence of singularities) is given in Subsection 3.4 as Theorem 3.2; this is a recursive formula involving a single dimensional integral over a one-dimensional version of the density formula.

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 density formulas in [5]. Section 5 gives a numerical example.

A well known generalization of multivariate phase type distributions is given in [17]. Let us point out how the results in [17] can be used to obtain a theoretical characterization of the distribution of $\boldsymbol{\tau}$ and compare this characterization with the formulas given in the present work. [17] generalizes multivariate phasetype distributions (to a class called MPH* in that work) and derives a theoretical representation of this generalization (see [17, Equation(35)]); [17, Equation(35)] is based on an infinite sum characterization of the full occupation measure (the time spent in each state until absorption to an absorbing state) of the process which is given in [17, Equation(34)]. [17] also explains a procedure which represents the distribution of $\boldsymbol{\tau}$ as an MPH* (this involves the expansion of the state space of the original process super-exponentially in the dimension of $\boldsymbol{\tau}$). One can combine these to get an infinite sum representation of the distribution of $\boldsymbol{\tau}$. The use of the complete occupation measure seems necessary for the full MPH* framework but its use to represent the distribution of $\boldsymbol{\tau}$ through the route just indicated would be unnecessarily complex and impractical for the following reasons 1) the resulting formula involves the computation of a high dimensional recursion and an infinite sum over $\mathbb{Z}_+^{|E|\times\mathcal{S}}$, where $\mathcal{S}$ is the set of all subpartitions of K; the terms in the sum involve products of exponentials of arbitrary length and the results of the recursion just mentioned (thus the recursion must be computed over a high dimensional infinite set); hence as is, even when $|E|$ and K are small, the resulting characterization is theoretical and not a practical formula intended to be explicitly computed 2) to get the distribution of $\boldsymbol{\tau}$ one ends up computing a much more complex distribution associated with the process

(the distribution of the full occupation measure of the process). A much simpler route that obviates these computations is the one followed in the present work: instead of trying to compute tail probabilities or the distribution function of $\boldsymbol{\tau}$, we focus on the computation of the density of $\boldsymbol{\tau}$ over each R_s; the exact calculation of these turns out to involve only a finite number $|s| \leq |K|$ of matrix multiplications and exponentiations. For moderate sizes of $|E|$, e.g., $|E| \leq 1500$, these computations can be carried out quickly on a standard modern laptop; we give several numerical examples with $|E| = 27$ in Section 5.

A recent generalization of multivariate phase type distributions is given in [7], in which a distribution on $\mathbb{R}_+^k$ is called "multivariable matrix-exponential" (MVME) if its Laplace transform is the ratio of two multivariate polynomials; as explained in the same work, these can be seen as generalizations of multivariable phase type distributions. The main result of [7] proves that a distribution is MVME if the Laplace transforms of its pullbacks on $\mathbb{R}_+$ under positive linear maps are polynomials. The main computational result of the paper is a formula for the cross-moments of an MVME distribution. While covering a class of distributions connected to the ones treated in the present work, the focus of [7] is on Laplace transforms and moments (rather than densities, tail probabilities or distribution functions) and it uses a completely different set of tools from those used in the present work.

From an applied perspective, our primary motivation in deriving the results in the present paper has been 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*; given the results of our paper this is now possible in great generality (for the case of two obligors one could use the results in [5]). The conclusion explains this application starting from the credit risk model of [12] 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., [5]), counterparty risk (see, e.g., [10]), and insurance (see, e.g., [6]). From a theoretical perspective, the first motivation of the paper has been the solution of a problem whose two dimensional version was solved in [5], 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 [15] and continued in [16] and [13]; the next paragraph explains this side of our contribution.

In [15] the following problem was studied: given a filtration $\mathbb{G} = \{\mathscr{G}_u, u \in \mathbb{R}_+\}$ and a multivariate random time $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$ study the conditional law, say $\mu_u^{\mathbb{G}}$, of $\boldsymbol{\tau}$ given $\mathscr{G}_u$, in the case that where $P(\tau_i = \tau_j) = 0$ for $i \neq j,\ i, j = 1, 2, \ldots, m$. Thus, a (random) measure was sought so that

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

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

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

where ν is a (possibly random) measure on $\mathbb{R}_+^m$, then $\varphi_u^{\mathbb{G}}$ is called the density of $P(\cdot|\mathscr{G}_u)$ with respect to ν, and the process $\varphi_\cdot^{\mathbb{G}}$ is called the conditional density process. This study was extended in [13], for $n = 2$, to the case where joint default were allowed, that is $P(\tau_1 = \tau_2) > 0$.

Now, let $\mathbb{F} = \{\mathscr{F}_u, u \in \mathbb{R}_+\}$ be the filtration generated by X. The Markov property of X implies that the conditional density of $\boldsymbol{\tau}$ given $\mathscr{F}_u$ directly follows from the density formula (16), as we show in Proposition 3.3. 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 $i \neq j,\ i, j = 1, 2, \ldots, m$ 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 allows one to study and model 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 $(\Omega, \mathscr{F})$. If $\alpha \in \mathcal{P}(E)$, where $\mathcal{P}(E)$ is the set of probability measures on E, we denote by P_α the law of X given that the initial distribution of X is α. In the case where $\alpha = \delta_i$, $i \in E$, we simply write P_i, so that $P_i(X_0 = i) = 1$. Consequently, $P_\alpha = \sum_{i \in E} \alpha(i) P_i$.

From now on, we assume that under each P_α, X is a time homogeneous Markov chain with intensity matrix $\boldsymbol{\lambda}$. The jump rate of the process from state i is $-\lambda(i,i) = \sum_{j \neq i} \lambda(i,j)$.

Remark 2.1. Until Section 4 we will assume that there are no absorbing states of X, that is $-\lambda(i,i) > 0$ for all $i \in E$. The general case can be treated by straightforward modifications of the arguments presented in the paper, as we shall see in Section 4.

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 $\boldsymbol{\lambda}(a,b)$ as the matrix with elements

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

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

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 $y|_a \in \mathbb{R}^a$:

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

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 $b, c \subset E$

$$\mathbf{M}|_{b \times c} \tag{5}$$

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:

$$x|^b(i) = \begin{cases} x(i) \text{ for } i \in a, \\ 0, \text{ otherwise.} \end{cases} \tag{6}$$

The random vector $\boldsymbol{\tau} = (\tau_k, k \in K)$ 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 $\boldsymbol{\tau}|_A$ to denote $(\tau_k, k \in A)$. The advantage of the notation $\boldsymbol{\tau}|_A$ is that we are able to index its components with elements of A rather than with the integers $\{1, 2, 3, ..., |A|\}$; this proves useful when stating the recursive formulas and proofs below.

2.2 Subpartitions of K

The key aspect of the distribution of $\boldsymbol{\tau}$, already mentioned 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, $n \in \{1, 2, 3, ..., |s|\}$. 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., [21]. Only ordered partitions appear in the present work and therefore to be brief we always assume every subpartition to have a definite order and we drop the adjective "ordered." With a slight abuse of notation we will write $s(n_1, n_2)$ to denote the n_2-th element of the n_1-th 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 shift given as

$$Ls = L(s(1), s(2), ..., s(|s|)) := (s(2), s(3), ..., s(|s|)).$$

Let L^m denote left shift m times. Similarly for $\boldsymbol{t} \in \mathbb{R}^n$, $n > 1$ let $L\boldsymbol{t} \in \mathbb{R}^{n-1}$ be its left shift. For $\boldsymbol{t} \in \mathbb{R}^n$ and $r \in \mathbb{R}$ let $\boldsymbol{t} - r$ denote $(t_1 - r, t_2 - r, ..., t_n - r)$.

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

$$R_s := \bigcap_{n=1}^{|s|} \left\{ \boldsymbol{t} \in \mathbb{R}_+^K : t_{k_1} = t_{k_2}, k_1, k_2 \in s(n) \right\} \cap \left\{ \boldsymbol{t} \in \mathbb{R}_+^K : t_{s(1,1)} < t_{s(2,1)} < \cdots < t_{s(|s|,1)} \right\}. \tag{7}$$

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*

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

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 3.1 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.

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 action of matrix $\mathbf{I}_a$ on matrices is described in the following simple well-known fact:

Lemma 2.1. Let n be a positive integer. For any $\mathbf{M} \in \mathbb{R}^{|E|\times n}$, the left multiplication by $\mathbf{I}_a$, i.e. $\mathbf{I}_a\mathbf{M}$, acts on the rows of $\mathbf{M}$, and $\mathbf{I}_a\mathbf{M}$ 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. Similarly, right multiplication by $\mathbf{I}_a$ acts on the columns of a matrix $\mathbf{M} \in \mathbb{R}^{n\times|E|}$, and $\mathbf{M}\mathbf{I}_a$ is the same as $\mathbf{M}$ except that now the columns with indices in a^c are set to zero.

It follows from (3) and Lemma 2.1 that $\boldsymbol{\lambda}(a,b) = \mathbf{I}_a\boldsymbol{\lambda}\mathbf{I}_b$. The operation of setting some of the columns of the identity matrix to zero commutes with set operations, i.e., one has $\mathbf{I}_{a\cap b} = \mathbf{I}_a\mathbf{I}_b$, $\mathbf{I}_{a\cup b} = \mathbf{I}_a + \mathbf{I}_b - \mathbf{I}_a\mathbf{I}_b$, $\mathbf{I}_{a^c} = \mathbf{I} - \mathbf{I}_a$. 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 $\mathbf{I}_{a^c}\boldsymbol{\lambda}\mathbf{I}_a = (\mathbf{I} - \mathbf{I}_a)\boldsymbol{\lambda}\mathbf{I}_a = \boldsymbol{\lambda}\mathbf{I}_a - \mathbf{I}_a\boldsymbol{\lambda}\mathbf{I}_a$, and $\boldsymbol{\lambda}(a, b\cap c)$ can be written 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 $d \subset E$, $j \in E$, $\alpha \in \mathcal{P}(E)$ and $u \in \mathbb{R}_+$ we define $p^u_{\alpha,d}(j) := P_\alpha(X_u = j, X_v \notin d, \forall v \le u)$, $p^u_d(i,j) := p^u_{\delta_i,d}(j) = P_i(X_u = j, X_v \notin d, \forall v \le u)$, and $p^u(i,j) := P_i(X_u = j)$. The symbol $\mathbf{p}^u_{\alpha,d}$ will denote the row vector with components $p^u_{\alpha,d}(j)$; $\mathbf{p}^u_d$ and $\mathbf{p}^u$ will denote the $|E| \times |E|$ matrices with elements $p^u_d(i,j)$ and $p^u(i,j)$, respectively. Note that $\mathbf{p}^u_{\alpha,d} = \alpha\mathbf{p}^u_d$. It follows from the definition of p^u that

$$\lim_{u\to 0} p^u(i,j)/u = \lambda(i,j). \tag{8}$$

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

$$\mathbf{p}^u_{\alpha,d} = \alpha e^{u\boldsymbol{\lambda}(d^c)}. \tag{9}$$

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 $u \to \mathbf{p}_d^u|_{d^c}$, $u \in \mathbb{R}_+$, satisfies, 2) the basic theory of ODEs will tell us that the unique solution is $u \to e^{u\boldsymbol{\lambda}(d^c)}|_{d^c}$.

Let ν_1 be the first jump time of X; for $X_0 = i \in d^c$, ν_1 is exponentially distributed with rate $-\lambda(i,i) > 0$.[1] Conditioning on ν_1 gives

$$p_d^u(i,j) = P_i(\nu_1 > u)\delta_i(j) + \int_0^u \lambda(i,i)e^{\lambda(i,i)v} \left(\sum_{l \in d^c - \{i\}} \frac{\lambda(i,l)}{\lambda(i,i)} p_d^{u-v}(l,j) \right) dv \tag{10}$$

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 (10) equals $\mathbf{p}_d^u|_{d^c} = e^{u\boldsymbol{\lambda}|_{d^c}} = e^{u\boldsymbol{\lambda}(d^c)}|_{d^c}$. The equality (9) follows from this and $\alpha|_d = 0$. □

Remark 3.1. Probabilities that concern sample paths that stay away from a given set are called "taboo probabilities" in [22, Section 1.2]; [22, Equation (F), page 28] is equivalent to (10).

The next result (written in a slightly different form) is well known, see, e.g., [19, 5]; we record it as a corollary here.

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

$$P_\alpha(\tau_d > u) = \alpha e^{u\boldsymbol{\lambda}(d^c)}\mathbf{1}. \tag{11}$$

Proof. $P_\alpha(\tau_d > u) = \sum_{j \in d^c} P_\alpha(X_u = j, X_v \notin d, \forall v \le u) = \alpha e^{u\boldsymbol{\lambda}(d^c)}\mathbf{1}$, where the last equality is implied by (9). □

Remark 3.2. One must modify (11) to $P_\alpha(\tau_d > u) = \alpha \mathbf{I}_{d^c} e^{u\boldsymbol{\lambda}(d^c)}\mathbf{1}$, $\quad P_\alpha(\tau_d = 0) = \alpha \mathbf{I}_d \mathbf{1}$ 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 seems to be the main method of derivation in most of the prior literature on phase-type distributions. This works well for a single hitting time or for nonsingular parts of the distribution of $\boldsymbol{\tau}$ (see [5, 12]). However, as explained in the introduction, the same idea runs into difficulties if one tries to use it to compute the *singular* parts of the distribution of $\boldsymbol{\tau}$. 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 in extending the theorem to multiple hitting times (see the next subsection).

Proposition 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*

$$\frac{d}{du}\left[P_\alpha(\tau_a \in (0,u], X_v \notin b, \forall v < \tau_a)\right] = \alpha e^{u\boldsymbol{\lambda}(d^c)}\boldsymbol{\lambda}(d^c, a)\mathbf{1}, \tag{12}$$

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

In other words, the density of τ_a on the set $\{X_v \notin b, \forall v < \tau_a\}$ is given by the right side of (12).

[1] See Remark 2.1.

Proof of Proposition 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, (8) and the Markov property of X (invoked at time u) give

$$P_i(X_{\tau_a} = j, \tau_a \in (u, u+h),\ X_v \notin b, \forall v \le u) = \left(\sum_{l \in d^c} p_d^u(i,l)\ \lambda(l,j) \right) h + o(h). \tag{13}$$

$$P_i(\tau_a \in (u, u+h), X_v \notin d, \forall v \le u) = \left(\sum_{j \in a} \sum_{l \in d^c} p_d^u(i,l)\ \lambda(l,j) \right) h + o(h). \tag{14}$$

By the previous lemma $p_d^u(i,l)$ equals exactly the $(i,l)^{th}$ component of $e^{u\boldsymbol{\lambda}(d^c)}$. These imply (12). □

Setting $b = \emptyset$ in Proposition 3.1 we get the density of τ_a. The following result will be needed in the proof of Theorem 3.1.

Proposition 3.2. *Let $a, b \subset E$, $a \cap b = \emptyset$. 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 $\alpha_1 := \alpha e^{\tau_a \boldsymbol{\lambda}(d^c)} \boldsymbol{\lambda}(d^c, a)$ and $\mathcal{V} := \{X_v \notin b, \forall v \le \tau_a\}$. 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 (13) and the definition of the conditional expectation. □

3.2 The multidimensional density

One can extend (12) to a representation of the distribution of $\boldsymbol{\tau}$ using the subpartition notation of subsection 2.2 as follows. For a partition s of K, $n \in \{1, 2, ..., |s|\}$ and $\boldsymbol{t} \in R_s \subset \mathbb{R}_+^K$, define

$$\bar{t}_n := t_{s(n,1)}, \quad \bar{t}_0 := 0, \quad W_n := [S(L^{n-1}s)]^c, T_n := \left[\bigcap_{k \in s(n)} \Gamma_k \right] \cap W_{n+1}, \tag{15}$$

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 key idea of the density formula and its proof is the $|s|$ step version of the one in Proposition 3.1: in order for $\boldsymbol{\tau} = \boldsymbol{t} \in \mathbb{R}_+^K$, 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), $n \le |s|$, 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 3.3. For a sequence $\mathbf{M}_1, \mathbf{M}_2, ..., \mathbf{M}_n$ 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 3.1. *For any partition $s \in \mathcal{S}$, the distribution of $\boldsymbol{\tau}$ on the set R_s has density*

$$f_s(\alpha, \boldsymbol{t}, K) = \alpha \left(\prod_{n=1}^{|s|} e^{\boldsymbol{\lambda}(W_n)(\bar{t}_n - \bar{t}_{n-1})} \boldsymbol{\lambda}(W_n, T_n) \right) \mathbf{1}, \quad \boldsymbol{t} \in R_s \tag{16}$$

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

Proof. The proof will use induction on $|K|$. For $|K| = 1$, (16) is the same as (12) with $b = \emptyset$. Let $\kappa > 1$ and suppose that (16) holds for all K with $|K| \leq \kappa - 1$; we will now argue that (16) 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 $\boldsymbol{\tau}$ restricted to R_s has the density (16). Specifically, we will show that for any bounded and measurable function $g : \mathbb{R}^K \to \mathbb{R}$ the following equality holds

$$\mathbb{E}[1_{R_s}(\boldsymbol{\tau}) g(\boldsymbol{\tau})] = \int_{R_s} g(\boldsymbol{t}) f_s(\alpha, \boldsymbol{t}, K) d_s \boldsymbol{t}, \tag{17}$$

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

Define $\vartheta := \min\{\tau_k, k \in K\}$, i.e., 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. On the set $\{\boldsymbol{\tau} \in R_s\}$ the following conditions hold:

$$1)\ X_\vartheta \in T_1, \qquad 2)\ X_u \in W_1 \text{ for } u \leq \vartheta.$$

These imply $\vartheta = \tau_{s(1,1)}$ holds on the same set. Therefore,

$$\{\boldsymbol{\tau} \in R_s\} \subset \mathcal{W}_1 := \{X_u \in W_1, u < \vartheta\} \cap \{X_\vartheta \in T_1\}. \tag{18}$$

Proposition 3.1 implies that if $\boldsymbol{\lambda}(W_1, T_1)$ is zero, then $\mathcal{W}_1$ has probability zero. Thus, (18) implies that if $\boldsymbol{\lambda}(W_1, T_1)$ is zero then $P_\alpha(\boldsymbol{\tau} \in R_s) = 0$ and, indeed, $f_s(\alpha, \boldsymbol{t}, K) = 0$ is the density of $\boldsymbol{\tau}$ on R_s. From here on, we will treat the case when $\boldsymbol{\lambda}(W_1, T_1)$ is nonzero.

Next, define the process $\widehat{X}$ by $\widehat{X}_u := X_{u+\vartheta}, u \geq 0$, and $\widehat{\boldsymbol{\tau}} = (\widehat{\tau}_k, k \in S(Ls))$ where $\widehat{\tau}_k := \inf\{u : \widehat{X}_u \in \Gamma_k\}$; $\hat{X}$ 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 (18) imply

$$\widehat{\boldsymbol{\tau}} = \boldsymbol{\tau}|_{Ls} - \vartheta, \tag{19}$$

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

$$\{\boldsymbol{\tau} \in R_s\} = \mathcal{W}_1 \cap \{\widehat{\boldsymbol{\tau}} \in R_{Ls}\}. \tag{20}$$

In words, this equality 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 Ls.

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

$$\widehat{g}(u, \widehat{\boldsymbol{t}}) := g\left(u\mathbb{1} + \widehat{\boldsymbol{t}}|^{S(s)}\right),$$

where we used the vector extension notation of (6). Equalities (19) and (20) imply

$$\begin{aligned} \mathbb{E}[1_{R_s}(\boldsymbol{\tau}) g(\boldsymbol{\tau})] &= \mathbb{E}[1_{\mathcal{W}_1} 1_{R_{Ls}}(\widehat{\boldsymbol{\tau}}) \widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}})] \\ &= \mathbb{E}[\mathbb{E}[1_{\mathcal{W}_1} 1_{R_{Ls}}(\widehat{\boldsymbol{\tau}}) \widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}}) | \mathscr{F}_\vartheta]] \\ &= \mathbb{E}[1_{\mathcal{W}_1} \mathbb{E}[1_{R_{Ls}}(\widehat{\boldsymbol{\tau}}) \widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}}) | \mathscr{F}_\vartheta]], \end{aligned} \tag{21}$$

where for the last equality we used the fact that the set $\mathcal{W}_1$ is $\mathscr{F}_\vartheta$ measurable. The property 5A in [8, page 98] implies

$$\mathbb{E}[1_{R_{Ls}}(\widehat{\boldsymbol{\tau}})\widehat{g}(\vartheta,\widehat{\boldsymbol{\tau}})|\mathscr{F}_\vartheta] = h(\vartheta)$$

where

$$h(u) := \mathbb{E}[1_{R_{Ls}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u,\widehat{\boldsymbol{\tau}})|\mathscr{F}_u]. \tag{22}$$

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

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

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

$$h(u) = \mathbb{E}[1_{R_{Ls}}(\widehat{\boldsymbol{\tau}})\widehat{g}(u,\widehat{\boldsymbol{\tau}})|\widehat{X}_0 = j] = \int\limits_{R_{Ls}} f_{Ls}(\delta_j, \boldsymbol{t}, K - s(1))g(u,\boldsymbol{t})d_{Ls}\boldsymbol{t} \tag{23}$$

where f_{Ls} is given as in (16) with s changed to Ls and K changed to $K - s(1)$. Once we substitute (23) in (21) we get an expectation involving only three random variables: ϑ, 1_A and $\widehat{X}_0 = X_\vartheta$, where $A = \{X_u \in W_1, u < \vartheta\}$. Proposition 3.1 implies that the density of ϑ on the set A is $\alpha e^{\boldsymbol{\lambda}(W_1)t_1}\boldsymbol{\lambda}(W_1,T_1)\mathbf{1}$, and Proposition 3.2 implies that the law of $\widehat{X}_0$ conditioned on ϑ and $1_{\mathcal{W}_1}$ is

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

These, the induction hypothesis, (22) and (23) imply that the outer expectation (21) equals (17). This last assertion finishes the proof of the induction step and hence the theorem. □

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 3.3. The first $\kappa > 0$ jump times of a standard Poisson process with rate $\lambda \in (0,\infty)$ have the joint density $\prod_{n=1}^{\kappa} e^{\lambda(t_n - t_{n-1})}\lambda$, $0 = t_0 < t_1 < t_2 < \cdots < t_\kappa$. Similarly, the first $\kappa > 0$ jump times of a Markov arrival process with intensity matrix $C + D$ (where C [D] is the matrix of transition intensities with [without] arrivals) have joint density $\boldsymbol{\alpha}\left(\prod_{n=1}^{\kappa} e^{D(t_n - t_{n-1})}C\right)\mathbf{1}$, see [4] or [2, page 304]. The density (16) can also be interpreted as a generalization of these formulas.

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

If α puts positive mass on $\gamma := \bigcup_{k\in K}\Gamma_k$, one best describes the law of $\boldsymbol{\tau}$ 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 probability measure.

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

$$P_\alpha(\boldsymbol{\tau} \in U) = \bar{\alpha}' P_{\alpha'}(\boldsymbol{\tau} \in U) + \sum_{i\in\gamma}\alpha(i)P_i(\boldsymbol{\tau} \in U) \tag{24}$$

for any measurable set $U \subset \mathbb{R}_+^K$. By its definition α' puts no mass on $\gamma = \cup_{k\in K}\Gamma_k$ and therefore Theorem 3.1 is applicable and $f(\alpha',\cdot,K)$ is the density of the probability measure $P_{\alpha'}(\boldsymbol{\tau} \in \cdot)$. For the second summand of (24), it is enough to compute each $P_i(\boldsymbol{\tau} \in U)$

separately. Define $K_i := \{k : i \in \Gamma_k\}$, $U_i := \{\boldsymbol{t} : \boldsymbol{t} \in U, t_k = 0, k \in K_i\}$, $\bar{U}_i := \{\boldsymbol{t}|_{K_i^c}, \boldsymbol{t} \in U_i\}$. Now remember that $i \in \gamma$; thus if $i \in \Gamma_k$ then $\tau_k = 0$ under P_i, and therefore $P_i(\boldsymbol{\tau} \in U) = P_i(\boldsymbol{\tau} \in U_i)$. For $\boldsymbol{\tau} \in U_i$, the stopping times $\boldsymbol{\tau}|_{K_i}$ are all deterministically 0. Thus to compute $P_i(\boldsymbol{\tau} \in U_i)$ it suffices to compute $P_i(\boldsymbol{\tau}|_{K_i^c} \in \bar{U}_i)$. But by definition $i \notin \cup_{k \in K_i^c} \Gamma_k$ and once again Theorem 3.1 is applicable and gives the density of $\boldsymbol{\tau}|_{K_i^c}$ under P_i as $f(\delta_i, \cdot, K_i^c)$.

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

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

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

3.4 Tail probabilities of τ

Probabilities of tail events have representations as integrals of densities given in Theorem 3.1 over appropriate subsets of $\mathbb{R}_+^{|K|}$. But to try to evaluate such integrals directly would not be a good idea. 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 mean probabilities of sets of the form $\{\tau_2 = \tau_4 > t_1, \tau_3 > t_2, \tau_1 = \tau_5 > t_2, \tau_3 \neq \tau_2, \tau_1 \neq \tau_2, \tau_1 \neq \tau_3\}$, or more generally

$$\bigcap_{n=1}^{|s|} \bigcap_{k_1,k_2 \in s(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s(n,1)} > t_n\} \bigcap_{n_1 \neq n_2, n_1, n_2 \leq |s|} \{\tau_{s(n_1,1)} \neq \tau_{s(n_2,1)}\}, \tag{25}$$

where s is a partition of K, and $\boldsymbol{t} \in \mathbb{R}_+^{|s|}$ is such that $t_n < t_{n+1}$. In (25) 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 (25). Thus, it is enough to be able to compute probabilities of events of the form (25).

Remark 3.4. From here on, to keep 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 (25)).

A tail event of the form (25) 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 entrances 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 (25). For two **disjoint** subpartitions s_1 and s_2 of K and an element $\boldsymbol{t} \in \mathbb{R}_+^{|s_1|}$ such that $t_{|s_1|} > t_{|s_1|-1} > \cdots > t_2 > t_1$ (if $|s_1| = 0$ by convention set $t = 0$) define

$$\mathcal{T}(s_1, s_2, \boldsymbol{t}) := \left(\bigcap_{n=1}^{|s_1|} \bigcap_{k_1,k_2 \in s_1(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s_1(n,1)} > t_n\} \right) \cap \bigcap_{m=1}^{|s_2|} \bigcap_{\ell_1,\ell_2 \in s_2(m)} \{\tau_{\ell_1} = \tau_{\ell_2}\}. \tag{26}$$

In view of Remark 3.4, setting $s_1 = s$ and $s_2 = \emptyset$ reduces (26) to (25). The indices in s_1 appear both in equality constraints and time constraints while indices in s_2 appear only in equality constraints.

Remark 3.5. The definition (26) implies that if a component of s_2 has only a single element, that component has no influence on $\mathcal{T}(s_1, s_2, \boldsymbol{t})$. For example, $\mathcal{T}(s_1, (\{1\}, \{2, 3\}), \boldsymbol{t})$ is the same as $\mathcal{T}(s_1, (\{2, 3\}), \boldsymbol{t})$.

To express $P_\alpha(\mathcal{T}(s_1, s_2, t))$ we will define a collection of functions p_i, $i \in E$, of s_1, s_2 and $\boldsymbol{t}$. We will denote by $\mathbf{p}$ the column vector with components $p_i, i \in E$.

For $s_1 = \emptyset$, and $i \in E$ define $p_i(\emptyset, s_2, 0)$ as $p_i(\emptyset, s_2, 0) := P_i(\mathcal{T}(\emptyset, s_2, 0))$. If s_2 is empty or if it consists of components with single elements, then the definitions of $\mathbf{p}$ and $\mathcal{T}$ and Remark 3.5 imply

$$\mathbf{p}(\emptyset, s_2, 0) = \mathbf{1}. \tag{27}$$

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

$$\begin{aligned} &\mathbf{p}(s_1, s_2, \boldsymbol{t}) := \\ &\int_0^{t_1} e^{u\boldsymbol{\lambda}(W)} \boldsymbol{\lambda}(W, T(s_1, s_2)) \left(\sum_{n=1}^{|s_2|} \mathbf{I}_{T_n(s_1, s_2)} \ \ \mathbf{p}(s_1, s_2 - s_2(n), \boldsymbol{t} - u) \right) du \\ &+ e^{t_1 \boldsymbol{\lambda}(W)} \mathbf{p}\left(Ls_1, s_2 + s_1(1), L\boldsymbol{t} - t_1\right), \end{aligned} \tag{28}$$

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

Theorem 3.2. *Suppose $E - S(s_1 \cup s_2)$ is not empty and that α is an initial distribution on E that puts all of its mass on this set. Then*

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

We omit the proof which is parallel to that of Theorem 3.1 and proceeds by induction. Theorem 3.2 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 (28) 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 (28) reduces to

$$\mathbf{p}(s_1, \emptyset, \boldsymbol{t}) = e^{\boldsymbol{\lambda}(S(s_1)^c)t_1} \mathbf{p}(Ls_1, s_1(1), L\boldsymbol{t} - t_1), \tag{29}$$

if $s_2 = \emptyset$.

When s_1 has no equality constraints and $s_2 = \emptyset$, one can invoke (29) $|s_1|$ times along with Remark 3.5 and (27) and get

Corollary 3.2. *Let α be as in Theorem 3.2. If $|s_1| > 0$ equals the dimension of $\boldsymbol{t}$, (in particular, there are no equality constraints) then*

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \boldsymbol{t})) = \alpha \mathbf{p}(s_1, \emptyset, \boldsymbol{t}) = \alpha \left(\prod_{n=1}^{|s_1|} e^{\boldsymbol{\lambda}(W_n)(t_n - t_{n-1})} \right) \mathbf{1} \tag{30}$$

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

The formula (30) is a generalization of [5, equation (7)] to general finite state Markov processes.

3.5 Conditional formulas

Here we shall compute the conditional density of $\boldsymbol{\tau}|_{V_{u_0}^c} - u_0$ given $\mathscr{F}_{u_0}$, where u_0 is a deterministic time. To this 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 V_u^c. The times $\boldsymbol{\tau}|_{V_{u_0}}$ are known by time u_0 and hence they are constant given $\mathscr{F}_{u_0}$. Thus, we only need to write down the regular conditional density of $\boldsymbol{\tau}|_{V_{u_0}^c}$, 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 3.1. Define $\widehat{X}_u := X_{u+u_0}$ and for $k \in V_{u_0}^c$

$$\widehat{\tau}_k := \inf\{u \in (0,\infty) : \widehat{X}_u \in \Gamma_k\}.$$

The definitions of $\widehat{X}$ and $\widehat{\boldsymbol{\tau}}$ imply

$$\widehat{\boldsymbol{\tau}} = \boldsymbol{\tau}|_{V_{u_0}^c} - u_0. \tag{31}$$

$\widehat{X}_0 = X_{u_0}$ is a constant given $\mathscr{F}_{u_0}$. Thus the process $\widehat{X}$ has exactly the same distribution as X with initial point X_{u_0} and Theorem 3.1 applies and gives the density of $\widehat{\boldsymbol{\tau}}$, which is, by (31), the regular conditional distribution of $\boldsymbol{\tau}|_{V_{u_0}^c} - u_0$. Therefore, for any bounded measurable $g : \mathbb{R}^{V_{u_0}^c} \to \mathbb{R}$ and a partition s' of $V_{u_0}^c$

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

We record this as

Proposition 3.3. *The regular conditional density of $\boldsymbol{\tau}|_{V_{u_0}^c} - u_0$ given $\mathscr{F}_{u_0}$ is $f(\delta_{X_{u_0}}, \boldsymbol{t}, V_{u_0}^c)$.*

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

The next subsection shows how the formulas in the previous sections can be simplified when Γ_k are absorbing; this assumption is made in [5, 12]. The subsection following it shows how formulas from [5] can be interpreted as special cases of the formulas derived in the present work.

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

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 $\boldsymbol{\lambda}(a, a^c) = 0$. Let us derive an alternative expression for the density formula (16) under the assumption that all $\{\Gamma_k, k \in K\}$ are absorbing. For this the following proposition is useful:

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

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

we have

$$p_{\alpha,a}^u = \alpha e^{\boldsymbol{\lambda}(a^c)u} = \alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_{a^c} \tag{32}$$

Proof. We already know from Lemma 3.1 that the first equality holds. Therefore, it only remains to show $p^u_{\alpha,a} = \alpha e^{\boldsymbol{\lambda} u}\mathbf{I}_{a^c}$. 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.,

$$(\alpha p^u_{\alpha,a})|_{a^c} = (\alpha e^{\boldsymbol{\lambda} u}\mathbf{I}_{a^c})|_{a^c}. \tag{33}$$

The definition of $p^u_{\alpha,a}$ and $\alpha|_a = 0$ imply $(\alpha p^u_{\alpha,a})|_a = 0$; The definition of $\mathbf{I}_{a^c}$ implies $(\alpha e^{\boldsymbol{\lambda} u}\mathbf{I}_{a^c})|_a = 0$. This and (33) imply (32). □

The previous proposition implies that, when all Γ_k are absorbing, one can replace the quantity $\boldsymbol{\lambda}(W_n)$ that appears in the density formula (16) 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 $\boldsymbol{t} \in R_s$ we have*

$$f_s(\alpha, \boldsymbol{t}, K) = \alpha \left(\prod_{n=1}^{|s|} e^{\boldsymbol{\lambda}(\bar{t}_n - \bar{t}_{n-1})} \boldsymbol{\lambda}(W_n, T_n) \right) \mathbf{1}\,, \tag{34}$$

where f_s is the density given in Theorem 3.1.

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

When $\{\Gamma_k, k \in K\}$ are absorbing, then, in view of (32), one can write the tail probability that appears in Theorem 3.2, as

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

and, in particular,

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \boldsymbol{t})) = \alpha e^{\boldsymbol{\lambda} t_1} \mathbf{I}_{S(s_1)^c} \mathbf{p}(s_1 - s_1(1), s_1(1), L\boldsymbol{t} - t_1). \tag{35}$$

4.3 Connections with earlier results

This subsection relates the phasetype density/ tail probability formulas from [5] to the formulas derived in the present work. In [5], the authors assume that E has a single absorbing state called Δ and they denote by A what in our paper is denoted by $\boldsymbol{\lambda}|_{\{\Delta\}^c}$. Moreover, [5] uses the letter α to denote the initial distribution of X, but on the set $\widehat{E} := E - \{\Delta\}$, rather then on the set E as it is done here; in particular, [5] implicitly assumes $P(X_0 = \Delta) = 0$. We will use the symbol $\widehat{\alpha}$ to denote the 'α of [5].' The relation between α and $\widehat{\alpha}$ is $\alpha|_{\{\Delta\}^c} = \widehat{\alpha}$.

As far as the singular densities / tail probabilities of $\boldsymbol{\tau}$ [5] treats only the case of $|K| = 2$. 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 [5, Equation (5), page 692] says

$$P_\alpha(T_1 = T_2 > u) = \widehat{\alpha} e^{Au} A^{-1}(A g_1 g_2 - [A, g_1] - [A, g_2])\mathbf{e}, \tag{36}$$

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 (36) 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 (36) is

$$\alpha|_{\widehat{E}} \left(e^{\boldsymbol{\lambda} u}|_{\widehat{E}} \right) A^{-1} C_\Delta. \tag{37}$$

In the present work, the same probability is expressed by a special case of (35); for the present case one sets $K = \{1, 2\}$, $s_1 = (\{1, 2\})$; for these values, (29) and conditioning on the initial state gives

$$P_\alpha(\tau_1 = \tau_2 > u) = \alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_w \mathbf{p}(\emptyset, (\{1, 2\}), 0), \tag{38}$$

where $w = (\Gamma_1 \cup \Gamma_2)^c$. One sees that this is equivalent to (37) as follows. On the right side of the last display $s_1 = \emptyset$ and we have no time constraints (the inequality constraints related to $\boldsymbol{t}$) and thus $\mathbf{p}(\emptyset, (\{1, 2\}), 0)$ is the probability of the event $\{\tau_1 = \tau_2\}$; the expression following the matrix exponential in (37) represents this probability. Finally, the absorbing property of the underlying chain and $X_0 \neq \Delta$ imply that we can ignore the restriction to $\hat{E}$ in (37).

The second density formula from [5] is for the absolutely continuous part of the distribution of $\boldsymbol{\tau}$; [12] makes use of this formula in the following context. The process X of [12] 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 in $\{-e_k, k = 1, 2, 3, ..., m\}$, where e_k is the unit vector with k^{th} coordinate equal to 1. In [12] the absorbing sets are denoted as Δ_i, see the display after [12, (2.3)], and they correspond to $\Gamma_k = \{z \in \mathbb{Z}_2^m : z_k = 0\}$ in our present setup. The key property of the setup in [12] is this: take any collection $\{\Gamma_{k_1}, \Gamma_{k_2}, ..., \Gamma_{k_n}\}$ 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 $\boldsymbol{\tau}$ has no singular part, i.e., $P(\boldsymbol{\tau} \in R_s) = 0$ for $|s| < m$, and one needs only the density of $\boldsymbol{\tau}$ with respect to the full Lebesgue measure in $\mathbb{R}^m$ (the "absolutely continuous part"); thus, for the purposes of [12] the density of the absolutely continuous part of the distribution of $\boldsymbol{\tau}$ is sufficient and a formula for this is already available in [5] and is given in [12, display (3.1.1)] as follows:

$$f(\boldsymbol{t}) = (-1)^m \alpha \left(\prod_{n=1}^{m-1} e^{\boldsymbol{\lambda}(\bar{t}_n - \bar{t}_{n-1})} (\boldsymbol{\lambda} G_{k_n} - G_{k_n} \boldsymbol{\lambda}) \right) e^{\boldsymbol{\lambda}(\bar{t}_m - \bar{t}_{m-1})} \boldsymbol{\lambda} G_{k_m} \mathbf{1}, \tag{39}$$

for $\boldsymbol{t} \in R_s$ with $|s| = m$; here $G_k = I_{\Gamma_k^c}$ and k_n is the index for which $t_{k_n} = \bar{t}_n$ ([12] uses the letter Q for the rate matrix $\boldsymbol{\lambda}$). We have derived the full density formula (34) in the absorbing case in Section 4 describing the density of $\boldsymbol{\tau}$ over its all possible parts (singular and nonsingular). Arguments similar to those given for th two dimensional formula can be used to show that (34) reduces to (39) when all components of $\boldsymbol{t}$ are distinct.

5 Numerical Example

The state space of our numerical example is $E = \mathbb{Z}_3^3$, where $\mathbb{Z}_3 = \{0, 1, 2\}$; the state space has 27 elements. 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

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

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

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

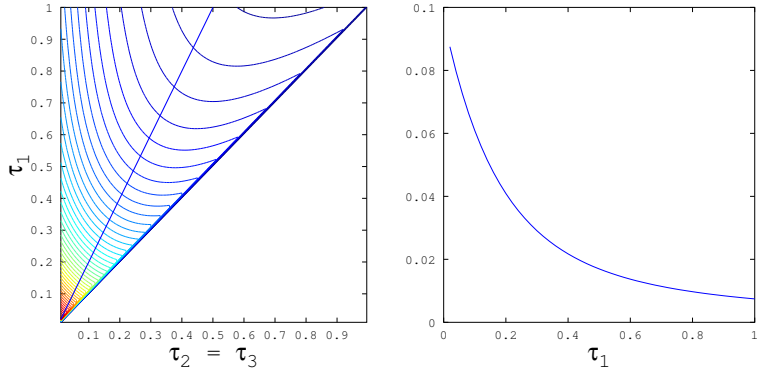

Figure 1: The level curves of the density f for $\tau_2 = \tau_3 < \tau_1$. On the right: the values of f over the line segment connecting $(0,0)$ to $(0.5,1)$

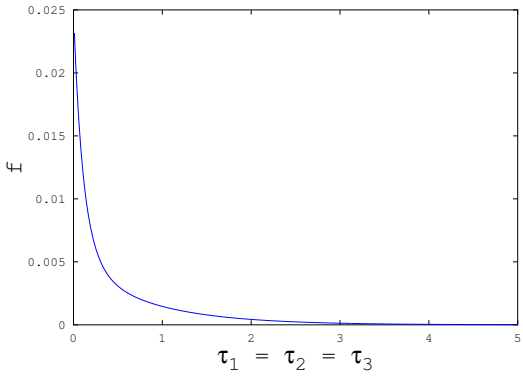

Figure 2: The density f for $\tau_1 = \tau_2 = \tau_3$

We will compute the density of $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3)$ 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 $\{\boldsymbol{\tau} \in R_{s_1}\} = \{\tau_2 < \tau_1 = \tau_3\}$ and the second to $\{\boldsymbol{\tau} \in R_{s_2}\} = \{\tau_1 = \tau_2 = \tau_3\}$.

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:

$$\pm e_k, \pm(e_1 + e_2), \pm(e_1 + e_2 + e_3), k \in K, \tag{40}$$

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 (40):

$$2\ ,1\ ,2\ ,1\ ,3\ ,1\ ,0.5\ ,0.5\ ,0.2\ ,0.2;$$

e.g., if $X_0 = (1,1,1)$ and σ_1 denotes the first jump time of X, σ_1 is exponentially distributed with rate s where s is the sum of the rates in the above display and $P(X_{\sigma_1} = X_0 + e_1 + e_2 + e_3) = 0.2/s$. These rates and the aforementioned dynamics give a 27×27 $\boldsymbol{\lambda}$ 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 $\boldsymbol{\tau}$.

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 [12]. With the results in the present work one can extend the modeling approach of [12] in two directions. Remember that the underlying process in [12] 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 (16) 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 would be 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] Søren Asmussen and Mogens Bladt, *Point processes with finite-dimensional conditional probabilities*, Stochastic Processes and their Applications **82** (1999), no. 1, 127–142.

[5] David Assaf, Naftali A. Langberg, Thomas H. Savits, and Moshe Shaked, *Multivariate phase-type distributions*, Operations Research **32** (1984), no. 3, 688–702.

[6] 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.

[7] Mogens Bladt and Bo Friis Nielsen, *Multivariate matrix-exponential distributions*, Stochastic models **26** (2010), no. 1, 1–26.

[8] Alexandr A. Borovkov, *Probability theory*, Springer, 2009.

[9] Jun Cai and Haijun Li, *Conditional tail expectations for multivariate phase-type distributions*, Journal of Applied Probability (2005), 810–825.

[10] 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.

[11] Agner Krarup Erlang, *Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges*, Elektrotkeknikeren **13** (1917), 5–13.

[12] Alexander Herbertsson, *Modelling default contagion using multivariate phase-type distributions*, Review of Derivatives Research **14** (2011), no. 1, 1–36.

[13] Ying Jiao and Shanqiu Li, *Generalized density approach in progressive enlargement of filtrations*, Electronic Journal of Probability **20** (2015), no. 85, 1–21.

[14] 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.

[15] Nicole El Karoui, Monique Jeanblanc, and Ying Jiao, *Density approach in modeling successive defaults*, SIAM Journal on Financial Mathematics **6** (2015), no. 1, 1–21.

[16] Younes Kchia and Martin Larsson, *Credit contagion and risk management with multiple non-ordered defaults*, arxiv.org/abs/1104.5272 (2013).

[17] Vidyadhar G Kulkarni, *A new class of multivariate phase type distributions*, Operations research **37** (1989), no. 1, 151–158.

[18] Cédric Loi and Paul Henry Cournede, *Generating functions of stochastic l-systems and application to models of plant development*, DMTCS proceedings (2008), no. 1.

[19] Marcel F. Neuts, *Probability distributions of phase type*, Liber Amicorum Prof. Emeritus H. Florin **173** (1975), 206.

[20] ———, *Matrix-geometric solutions in stochastic models: an algorithmic approach*, Courier Dover Publications, 1981.

[21] Richard P. Stanley, *Enumerative combinatorics*, vol. 49, Cambridge university press, 2011.

[22] Ryszard Syski, *Passage times for Markov chains*, Ios Press, 1992.