UNIVERSITY OF COPENHAGEN
DEPARTMENT OF MATHEMATICAL SCIENCES

Master's Thesis

Niels Moctezuma Krarup

Phase-type distributions

Extensions, Implementation and Simulation

Supervisor: Mogens Bladt

Submitted on: June 19th 2019

Abstract

In this thesis, we give an introduction to phase-type distributions which are constructed as the time until absorption of an underlying Markov chain or Markov jump process. In the discrete version we extend the basic theory to cover first occurrence of a general sub-sequence of iid discrete random variables and a basic urn experiment. In the continuous case, the well known result of denseness of the class of phase-type distributions on the non-negative real line, in the sense of weak convergence, is proved. Closure properties, with emphasis on order statistics is investigated. We give a characterization of the probabilities of the n! different orderings of n independent but non-identically distributed phase-type distributions, relying on matrix operations rather than integration. Transformation via linear rewards of the underlying process is discussed, being fundamental in the construction of the class of multivariate phase-type distributions denoted MPH*, which is briefly discussed. Finally, we investigate two methods of estimating functionals of the underlying process conditioned on the time of absorption, an importance sampler and a Markov Chain Monte Carlo method. Quantities of interest are implemented in R.

Notation and Conventions

Besides the table of notation given below, to streamline much of the mathematics we will adopt following notational conventions. Matrices and vectors will be denoted with **bold** with matrices being upper case and vectors being lower case. Entries of a matrix or vectors will be the corresponding non-bold symbol with indices e.g. an $m \times n$ matrix will be denoted

$$\boldsymbol{A} = \{a_{ij}\}_{i=1,\ldots,m,j=1,\ldots,n}$$

while an p-vector will be denoted

$$\boldsymbol{t} = \{t_i\}_{i=1,\ldots,p}$$

Vectors given by Roman letters are column vectors while vectors given by Greek letters are row vectors. i.e. $\boldsymbol{t\pi}$ is a matrix and $\boldsymbol{\pi t}$ is a scalar (the standard dot product).
Reference to a function in the accompanying `R`-package is done using the font special font and parenthesis: `FunName()`.

$1\{A\}, 1_A(x), 1\{x \in A\}$	Indicator function 1 if A is true 0 otherwise.
$:=$	Left hand side defined as right hand side.
$\delta_x(y)$	Kronecker delta, 1 if $y = x$ 0 otherwise.
δ_{ij}	1 if $i = j$ 0 otherwise.
X, Y	Random variables are uppercase nonbold roman letters.
$X_{i:n}$	The ith order statistic, the ith smallest value of $X_1, X_2, \ldots, X_n$.
$\mathbb{N}$	The natural numbers $\{0, 1, \ldots\}$.
$\mathbb{R}, \mathbb{R}^n$	The real line. n-dimensional euclidean
$\boldsymbol{e}$	Vector of 1s of appropriate length.
$\boldsymbol{e}_i$	The ith unit vector of the standard basis of $\mathbb{R}^n$
$\boldsymbol{A}_{i\cdot}$	ith row of $\boldsymbol{A}$.
$\boldsymbol{A}_{\cdot j}$	jth column of $\boldsymbol{A}$.
$\text{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$	Block diagonal matrix with $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ on the diagonal.
$\boldsymbol{\Delta}(\boldsymbol{a})$	Diagonal matrix with the elements of $\boldsymbol{a}$ on the diagonal.
$(\boldsymbol{A}\,\boldsymbol{B}), (\boldsymbol{A}, \boldsymbol{B})$	Concatenation of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
$\mathbb{P}$	Probability (measure).
$\mathbb{E}$	Expectation.
$Var, \mathbb{V}ar, \mathbb{V}$	Variance.
Cov	Covariance.
$a.s.$	Almost surely
$\sim, \stackrel{d}{=}$	Distributed as, equal in distribution.
$X_n \xrightarrow{P} X$	Convergence in probability.
$X_n \xrightarrow{d} X$	Convergence in law, weak convergence.
$\oplus$	Kronecker sum
$\otimes$	Kronecker product
$\bullet$	Schur (Hadamard) entry wise matrix product.
$exp(\lambda)$	The exponential distribution with rate λ i.e. $X \exp(\lambda)$ then $\mathbb{E}(X) = \frac{1}{\lambda}$
$geom(p)$	Geometric distribution with success parameter p. Given as the number of *trials* before first success occurred.
$Er_k(\lambda)$	Generalized Erlang distribution.

Acknowledgement

I would like to thank my supervisor Mogens Bladt for always taking time to discuss my questions and ideas. In some way, every meeting was like an iteration of the EM-algorithm. Mogens was also the one who introduced me to the topic of phase-type distributions which I became quite interested in, as this thesis hopefully demonstrates.

I would also like to thank friends and family for providing a good atmosphere outside the writing periods, in particular Mossa Merhi, Kajwan Rasul and Dina Sofie Grønseth.

Contents

Chapter 1

Introduction

As a student of statistics and probability one is usually presented to probability distributions characterized by their density $f(x)$ from which other quantities of interest can be derived e.g. cumulative distribution function (CDF), moments, Laplace transform or some probabilistic property e.g. the memorylessness of the exponential distribution or the Gaussian as the limit of certain normalized sums of independent identically distributed (iid) random variables as stated by the Central Limit Theorem (CLT) and its variants.

In contrast, for phase-type distributions one start by taking a step back defining the distribution of interest from an underlying stochastic process. To be precise, a phase-type distribution is defined as the time until absorption of a Markov chain in the discrete case or a Markov jump process, also referred to as Continuous Time Markov Chain (CTMC) in the continuous case. Using this approach, which at first glance may seem as a detour, one may derive interesting properties such as density, moments, closure properties etc. based on the underlying process, often in a more streamlined and elegant way than standard analytic methods.

On of the first steps taken in this direction was made by Danish statistician and engineer A.K. Erlang who in the beginning of the 20th century proposed a model for the durations of telephone calls as a decomposition of individual (fictitious) stages each being exponentially distributed, what is known today as the Erlang distribution.

This approach was later generalized by A. Jensen in 1953 who extended the formulation of stages using Markov jump processes. Finally, during the seventies M.F. Neuts and others brought the theory up to the level of today's standard theory which has made phase-type distributions an important and active research area both in application and theoretical studies.

Alluding to the first paragraph, the probabilistic reasoning usually involving sample paths arguments can come across almost as "cheating" the first time encountered but it is truly a powerful tool which after some time of acquiring becomes almost second nature.

This thesis covers basic phase-type theory. It contains many examples, intuition and implementations the reason of which is twofold. First is the simple reason of personal interest in seeing the theory in practice, second is the belief that true understanding of concepts is best conveyed by explanation (examples or illustrations) and computational implementation. It is our hope that this has not resulted in a too long and too text-book-like thesis, but simply demonstrates the mathematical transparency and understanding.

The thesis is organised as follows: In chapter 1, we introduce and walk through basic theory of Markov chains and Markov jump processes, laying the foundation for the construction of the phase-type distributions. Chapter 2 is entirely devoted to the discrete phase-type where we end the first part of the basic theory with some extensions. Chapter 3 goes through the continu-

ous phase-type distribution with emphasis on closure properties in particular order statistics including implementation. Chapter 5 goes though the basic multivariate phase-type extension. Chapter 6 implements two methods proposed in [Bladt and Nielsen, 2017] for conditional simulation of phase-types given the absorption time. The thesis is concluded in chapter 7 followed by an appendix explaining how to install and use the `R`-package `phasetypeUtilsUcph` accompanying this thesis.

Chapter 2

Markov Processes

The theory of phase-type distributions is based on certain underlying Markov processes. Those being Markov jump process in the continuous version and Markov chains in the discrete version. The properties of the underlying process is quintessential in the development and extension of the phase-type[1] as these can often be based on probabilistic arguments. These arguments use the underlying process which makes the calculations more streamlined compared to direct analytical methods. Hence this chapter is dedicated to a review of the basic properties of Markov chains and Markov jump processes with focus on the specific application to phase-type distributions.

For Markov jump processes emphasis will be on infinitesimal probabilities and the Minimal Construction where the process is completely characterised by its intensity matrix $\mathbf{\Lambda}$ and initial distribution vector $\boldsymbol{\pi}$. For our purpose of the phase type we may restrict ourselves to the case of finite state spaces which leaves out phenomena such as explosion and immediate states which will not be discussed although of interest in itself. Further, by nature Markov jump processes are piecewise constant and transitions occur via jumps. We will make the assumption that the process is continuous from the right and has limits from the left, known as a càdlàg process. For an in-depth foundation we refer to [Amussen, 2003] and also [Cinlar, 2013]. However most of the definitions and results presented here are based on [Bladt and Nielsen, 2017].

Formally we will be working on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathscr{F}$ a sigma-algebra of events on Ω and $\mathbb{P}$ a probability measure on the measurable space $(\Omega, \mathscr{F})$. A real random variable X is then a $\mathscr{F} - \mathcal{B}$ measurable function $X : \Omega \to \mathbb{R}$. Where $\mathcal{B}$ is the Borel sigma-algebra. A collection of random variables $\{X_i\}_{i \in I}$ for some index set I is then called a stochastic process, or a sequence of random variables depending on the context. For a random variable X we define $\sigma(X)$ as the smallest sigma-algebra on Ω making X measurable and we say that $\sigma(X)$ is the sigma-algebra generated by X. Similarly, $\sigma(\{X_i\}_{i \in I})$ is the smallest sigma-algebra making X_i measurable for all $i \in I$.

2.1 Markov Chains

A Markov chain, $\{X_n\}_{n \in \mathbb{N}}$ is a stochastic process taking values in a discrete state space E i.e. finite or at most countable. By definition the future of such processes only depend on the past trough the present.

Definition 2.1. *A Markov Chain $\{X_n\}_{n \in \mathbb{N}}$ with state space E is a stochastic process which satisfies*

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i).$$

For all $n \in \mathbb{N}$ and all $i_0, ..., i_{n-1}, i, j \in E$.

[1] We will adopt a slight abuse of language and use "phase-type" both when referring to something being phase-type distributed and for the general class of phase-type distributions.

The above is know as the Markov property. Further, if the transition probabilities do not depend on the time we say that the process is *time-homogeneous*.

Definition 2.2. *A Markov Chain $\{X_n\}_{n\in\mathbb{N}}$ with state space E is called time-homogeneous if it satisfies*

$$\mathbb{P}(X_{n+1} = j|X_n = i) = \mathbb{P}(X_1 = j|X_0 = i).$$

For all $n \in \mathbb{N}$ and all $i, j \in E$.

In this thesis we will only be working with time-homogeneous Markov chains, whose transition probabilities of going from state i to j during one step can simply be expressed by

$$p_{ij} = \mathbb{P}(X_{n+1} = j|X_n = i) = \mathbb{P}(X_1 = j|X_0 = i).$$

For time-homogeneous Markov chains the transition probabilities can be arranged in a $|E| \times |E|$ matrix called the transition matrix given by

$$\boldsymbol{P} = \{p_{ij}\}_{i,j\in E}.$$

To characterize the Markov chain $\{X_n\}_{n\in\mathbb{N}}$ we also need to define the initial distribution vector $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$ with entries given by $\pi_i = \mathbb{P}(X_0 = i)$. By definition $\boldsymbol{\pi}$ is a stochastic vector, which means entries in $[0, 1]$ and $\boldsymbol{\pi e} = 1$.
For a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ the joint distribution of $(X_0, X_1, \ldots, X_n)$ is expressed using the initial vector $\boldsymbol{\pi}$ and the transition matrix $\boldsymbol{P}$ as shown in the next theorem.

Theorem 2.3 (Markov Chain joint distribution)**.** *The stochastic process $\{X_n\}_{n\in\mathbb{N}}$ is a Markov chain if and only if $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \pi_{i_0} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{n-1} i_n}$ for all sets $\{X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$ with positive probability where $n \in \mathbb{N}$ and $i_0, \ldots, i_n \in E$.*

Proof. If $\{X_n\}_{n\in\mathbb{N}}$ is a Markov chain we have

$$\begin{aligned}
\mathbb{P}(X_n = i_n, \ldots, X_0 = i_0) &= \mathbb{P}(X_n = i_n|X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)\mathbb{P}(X_{n-1} = i_{n-1}, \ldots X_0 = i_0) \\
&= \mathbb{P}(X_n = i_n|X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \quad (Markovproperty) \\
&= p_{i_{n-1} i_n}\mathbb{P}(X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \quad (Time - homogeneity) \\
&\vdots \\
&= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdot \ldots \cdot p_{i_{n-1} i_n}.
\end{aligned}$$

The reverse implication follows from the definition of conditional independence, the Markov property and time-homogeneity. □

Consider the n-step transition probabilities given by

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j|X_0 = i) \tag{2.1}$$

which have the nice property of being the ijth entry of the nth power of the transition matrix $\boldsymbol{P}$ as the next theorem shows.

Theorem 2.4 (n-step transition probabilities)**.** *Let $n \in \mathbb{N}$.*
For $p_{ij}^{(n)}$ given by (2.1) then $\left\{p_{ij}^{(n)}\right\}_{i,j\in E} = \boldsymbol{P}^n$ in particular $X_n \sim \boldsymbol{\pi P}^n$ and $\boldsymbol{P}^n$ is the transition matrix for the Markov chain $\{X_{kn}\}_{k\in\mathbb{N}}$.

Proof. By the law of total probability and Theorem 2.3 we have

$$\begin{aligned}
\mathbb{P}(X_n = j, X_0 = i) &= \sum_{i_{n-1}\in E} \cdots \sum_{i_1 \in E} \mathbb{P}\left(X_n = j, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i\right) \\
&= \sum_{i_{n-1}\in E} \cdots \sum_{i_1 \in E} \pi_i p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\
&= \sum_{i_{n-2}\in E} \cdots \sum_{i_1 \in E} \pi_i p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-3} i_{n-2}} \left(\boldsymbol{P}^2\right)_{i_{n-2} j} \\
&\vdots \\
&= \pi_i \left(\boldsymbol{P}^n\right)_{ij}
\end{aligned}$$

and from conditional independence $\mathbb{P}(X_n = j | X_0 = i) = \left(\boldsymbol{P}^n\right)_{ij}$.
The distribution of X_n simply follows from summing over the initial state

$$\mathbb{P}(X_n = j) = \sum_{i\in E} \mathbb{P}(X_n = j, X_0 = i) = \sum_{i\in E} \pi_i \left(\boldsymbol{P}^n\right)_{ij} = \left(\boldsymbol{\pi}\boldsymbol{P}^n\right)_j .$$

Lastly, the stochastic process $\{X_{kn}\}_{k\geq 0}$ which is formed by taking n-sized steps in the $\{X_n\}_{n\geq 0}$ Markov chain, is obviously still a Markov chain on E. Using the time-homogeneity we see that the n-step transition probability has the form

$$\mathbb{P}(X_{kn+n} = j | X_{kn} = i) = \mathbb{P}(X_n = j | X_0 = i).$$

Which we have just seen to be the ijth entry of $\boldsymbol{P}^n$. □

Theorem 2.5 (Chapman-Kolmogorov)**.** *The n-step transition probabilities $p_{ij}^{(n)}$ given by (2.1) satisfy*

$$p_{ij}^{(n+m)} = \sum_{k\in E} p_{ik}^{(n)} p_{kj}^{(m)}$$

Proof. By Theorem 2.4 it follows from matrix multiplication, $\boldsymbol{P}^{n+m} = \boldsymbol{P}^n \boldsymbol{P}^m$. But it can also be shown using the Markov property directly.

$$\begin{aligned}
p_{ij}^{(n+m)} &= \mathbb{P}(X_{n+m} = j | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_{n+m} = j | X_n = k, X_0 = i) \mathbb{P}(X_n = k | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_m = j | X_0 = k) \mathbb{P}(X_n = k | X_0 = i) \\
&= \sum_{k\in E} p_{ik}^{(n)} p_{kj}^{(m)}
\end{aligned}$$

where in the fourth equality we used the Markov- and Time-homogeneous property. □

It is worth noting that the Markov property can be reformulated using generated sigma-algebras. If $\mathscr{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ then the Markov property can we written as

$$\mathbb{P}(X_{n+1} = j | \mathscr{F}_n) = \mathbb{P}(X_{n+1} = j | X_n) \tag{2.2}$$

for all $n \in \mathbb{N}$ and $j \in E$.
Properties of a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ can often be expressed by the initial value X_0 hence we shall adopt the following notation for ease.

$$\begin{aligned}\mathbb{P}_i(\cdot) &= \mathbb{P}(\cdot|X_0 = i)\\ \mathbb{E}_i(\cdot) &= \mathbb{E}(\cdot|X_0 = i)\\ \mathbb{P}_Y(\cdot) &= \mathbb{P}(\cdot|X_0 = Y)\end{aligned}$$

Where Y is a random variable taking values in E.
Using the above notation we can restate the time-homogeneous property as

$$\mathbb{P}(X_{n+1} = j|X_n) = \mathbb{P}_{X_n}(X_1 = j)$$

and we have that $\{X_n\}_{n\in\mathbb{N}}$ is a time-homogeneous Markov chain on E if and only if

$$\mathbb{P}(X_{n+1} = j|\mathscr{F}_n) = \mathbb{P}_{X_n}(X_1 = j)$$

for all $n \geq 1$ and $j \in E$.

The next step in the theory is to ensure that the Markov property also holds at certain well behaved random times, to this end we need the concept of stopping times.

Definition 2.6 (Stopping Time)**.** *A stopping time $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ is a random variable with the property that $\{\tau = n\} \in \mathscr{F}_n$ for all $n \in \mathbb{N}$. For a stopping time τ for the Markov chain $\{X_n\}_{n\in\mathbb{N}}$ note that this implicates that from $X_0, X_1, \ldots, X_n$ we know whether $\tau = n$ or not. The collection of sets $\mathscr{F}_\tau$, which consists of all the events defined by the process up to and including the random time τ is then defined by the relation*

$$A \in \mathscr{F}_\tau \Leftrightarrow A \cap \{\tau = n\} \in \mathscr{F}_n \quad \forall n \in \mathbb{N} \cup \{\infty\}$$

Theorem 2.7 (Sigma-Algebra from stopping time)**.** *The collection of sets $\mathscr{F}_\tau$ defined in Definition 2.6 is a σ-algebra.*

Proof. We need to show three properties of the collection of sets $\mathscr{F}_\tau$. First that it includes Ω.

$$\Omega \cap \{\tau = n\} = \{\tau = n\} \in \mathscr{F}_n \quad \forall n \in \mathbb{N} \cup \{\infty\}$$

Since τ is a stopping time. Next we show that $\mathscr{F}_\tau$ is closed under complementation. Take $A \in \mathscr{F}_\tau$ then

$$A^c \cap \{\tau = n\} = \{\tau = n\} \cap \{A \cap \{\tau = n\}\}^c \in \mathscr{F}_n \quad \forall n \in \mathbb{N} \cup \{\infty\}$$

since $\mathscr{F}_n$ is a σ-algebra for all $n \geq 1$. Lastly let $A_1, A_2, \ldots \in \mathscr{F}_\tau$ then

$$\bigcup_{k\geq 1} A_k \cap \{\tau = n\} = \bigcup_{k\geq 1} \{A_k \cap \{\tau = n\}\} \in \mathscr{F}_n \quad \forall n \in \mathbb{N} \cup \{\infty\}$$

since $\{A_k \cap \{\tau = n\}\} \in \mathscr{F}_n \quad \forall k \geq 1$ and $\mathscr{F}_n$ is a σ-algebra. □

We now prove that the Markov property also holds at stopping times.

Theorem 2.8 (Strong Markov Property)**.** *For a stopping time τ for a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on E then on $\{\tau < \infty\}$ we have*

$$\mathbb{P}(X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k|\mathscr{F}_\tau) = \mathbb{P}_{X_\tau}(X_1 = i_1, \ldots, X_k = i_k)$$

for all $k \in \mathbb{N}$ and $i_1, \ldots, i_k \in E$.

Proof. The measure theoretic definition of conditional- expectation and probability states that $\mathbb{P}(X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k|\mathscr{F}_\tau) = \mathbb{E}1\{X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k\}|\mathscr{F}_\tau$.
Hence by definition of conditional expectation we need to show that the right hand side is $\mathscr{F}_\tau$-mesurable and that for all $D \in \mathscr{F}_\tau$ we have

$$\int_{D\cap\{\tau<\infty\}} 1\{X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k\}d\mathbb{P} = \int_{D\cap\{\tau<\infty\}} \mathbb{P}_{X_\tau}(X_1 = i_1, \ldots, X_k = i_k)d\mathbb{P}$$

The measurability is obvious since $\sigma(X_\tau) \subseteq \sigma(X_0, \ldots, X_\tau)$.
Now splitting up $\{\tau < \infty\} = \bigcup_{n=1}^{\infty}\{\tau = n\}$ we get for $D \in \mathscr{F}_\tau$

$$\begin{aligned}\int_{D\cap\{\tau<\infty\}} 1\{X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k\}d\mathbb{P} &= \sum_{n=1}^{\infty}\int_{D\cap\{\tau=n\}} 1\{X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k\}d\mathbb{P} \\ &= \sum_{n=1}^{\infty}\int_{D\cap\{\tau=n\}} 1\{X_{n+1} = i_1, \ldots, X_{n+k} = i_k\}d\mathbb{P}\end{aligned}$$

Now note that by the standard Markov Property

$$\begin{aligned}\mathbb{P}_{X_n}(X_1 = i_1, \ldots, X_k = i_k) &= \mathbb{P}(X_{n+1} = i_1, \ldots, X_{n+k} = i_k|X_n) \\ &= \mathbb{P}(X_{n+1} = i_1, \ldots, X_{n+k} = i_k|\mathscr{F}_n) \\ &= \mathbb{E}1\{X_{n+1} = i_1, \ldots, X_{n+k} = i_k\}|\mathscr{F}_n\end{aligned}$$

Continuing, since the set $\{D \cap \{\tau = n\}\} \in \mathscr{F}_n$ since $D \in \mathscr{F}_\tau$ we get by definition of conditional expectation

$$\begin{aligned}\int_{D\cap\{\tau<\infty\}} 1\{X_{\tau+1} = i_1, \ldots, X_{\tau+k} = i_k\}d\mathbb{P} &= \sum_{n=1}^{\infty}\int_{D\cap\{\tau=n\}} \mathbb{P}_{X_n}(X_1 = i_1, \ldots, X_k = i_k)d\mathbb{P} \\ &= \sum_{n=1}^{\infty}\int_{D\cap\{\tau=n\}} \mathbb{P}_{X_\tau}(X_1 = i_1, \ldots, X_k = i_k)d\mathbb{P} \\ &= \int_{D\cap\{\tau<\infty\}} \mathbb{P}_{X_\tau}(X_1 = i_1, \ldots, X_k = i_k)d\mathbb{P}\end{aligned}$$

□

Definition 2.9 (First Return Time)**.** *Let $T_i = \inf\{n \geq 1|X_n = i\}$ be the first time of entry in state i (or return in the case that $X_0 = i$). With the convention that $\inf\emptyset = +\infty$.*
Further, let $N_i = \sum_{n=1}^{\infty} 1\{X_n = i\}$ be the total numbers of visits to state i, again excluding the initial state X_0.
We call a state $i \in E$ recurrent if

$$\mathbb{P}_i(T_i < \infty) = 1$$

And if $\mathbb{P}_i(T_i < \infty) < 1$ the state is called transient. i.e. if the Markov chain is initiated in a recurrent state $\in Ei$ eventual return to state i is certain.

Notice that T_i is a stopping time. By the Strong Markov property it seems reasonable to expect the chain to visit a recurrent state infinitely many times, given initiation in that state since eventual return is certain, where after the chain is "resat" and a second return again will be certain. The next theorem makes this intuition precise.

Theorem 2.10 (Characterization of recurrent states)**.** *For a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on E then for a state $i \in E$ the following statements are equivalent.*

$$\begin{aligned}&1)\ \textit{State } i \textit{ is recurrent} \\ &2)\ N_i = \infty \quad \mathbb{P}_i - a.s. \\ &3)\ \mathbb{E}_i(N_i) = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty.\end{aligned}$$

Proof. For 1) $\Rightarrow$ 2) let $T_i^1 = T_i$ and $T_i^k = \inf\{n > T_i^{k-1} \,|X_n = i\}$ be the times of successive visits to state i then T_i^k is a stopping time for the Markov chain $\{X_n\}_{n\in\mathbb{N}}$ for all $k \in \mathbb{N}$ and $i \in E$. Then

$$\begin{aligned}
\mathbb{P}_i(T_i^{k+1} < \infty) &= \mathbb{P}_i(T_i^{k+1} < \infty, T_i^k < \infty) \quad \left(T_i^{k+1} < \infty \Rightarrow T_i^k < \infty\right) \\
&= \mathbb{E}_i(\mathbb{P}_i(T_i^{k+1} < \infty, T_i^k < \infty | \mathscr{F}_{T_i^k})) \quad \text{(Tower property)} \\
&= \mathbb{E}_i\left(1\{T_i^k < \infty\}\mathbb{P}_i\left(T_i^{k+1} < \infty | \mathscr{F}_{T_i^k}\right)\right) \quad \text{(Measurability)} \\
&= \mathbb{E}_i\left(1\{T_i^k < \infty\}\mathbb{P}_{X_{T_i^k}}\left(T_i^1 < \infty\right)\right) \quad \text{(Strong Markov proptery)} \\
&= \mathbb{E}_i\left(1\{T_i^k < \infty\}\mathbb{P}_i\left(T_i^1 < \infty\right)\right) \quad \left(X_{T_i^k} = i\right) \\
&= \mathbb{P}_i\left(T_i^1 < \infty\right)\mathbb{P}_i\left(T_i^k < \infty\right) \\
&\vdots \\
&= \mathbb{P}_i\left(T_i^1 < \infty\right)^{k+1}.
\end{aligned}$$

Now since state i is recurrent we have $\mathbb{P}_i\left(T_i^1 < \infty\right) = 1$ hence $\mathbb{P}_i(T_i^k < \infty) = 1$ i.e. $T_i^k < \infty$ $\mathbb{P}_i$-a.s. for all $k \in \mathbb{N}$. Now since

$$N_i = \sum_{k=1}^{\infty} 1\{T_i^k < \infty\}$$

we have $N_i = \infty$ $\mathbb{P}_i$-a.s.
For 2) $\Rightarrow$ 3) note that

$$\mathbb{E}_i N_i = \mathbb{E}_i \sum_{n=1}^{\infty} 1\{X_j = i\} = \sum_{n=1}^{\infty} \mathbb{P}_i\left(X_n = i\right) = \sum_{n=1}^{\infty} p_{ii}^{(n)}$$

by Beppo Levi. Now if $N_i = \infty$ $\mathbb{P}_i$-a.s. then obviously $\mathbb{E}_i N_i = \infty$.
Lastly, we show that 3) $\Rightarrow$ 1) or equivalently that $\neg 1) \Rightarrow \neg 3)$ that is, assume $i \in E$ is transient i.e. $\mathbb{P}_i\left(T_i < \infty\right) < 1$, and note that

$$\{N_i > n\} = \{N_i \geq n + 1\} = \{T^{n+1} < \infty\}.$$

Then recalling that for non-negative discrete random variables X we have $\mathbb{E}X = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$ we get

$$\mathbb{E}_i\left(N_i\right) = \sum_{n=0}^{\infty} \mathbb{P}_i\left(N_i > n\right) = \sum_{n=0}^{\infty} \mathbb{P}_i\left(T_i^{n+1} < \infty\right) = \sum_{n=0}^{\infty} \mathbb{P}_i\left(T_i < \infty\right)^{n+1} < \infty$$

□

The following corollary then follows immediately from the above theorem.

Corollary 2.11. *For a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on E then for a state i the following statements are equivalent.*

$$\begin{aligned}
&1)\ \textit{State } i \textit{ is transient} \\
&2)\ N_i < \infty \quad \mathbb{P}_i - a.s. \\
&3)\ \mathbb{E}_i(N_i) = \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty.
\end{aligned}$$

Definition 2.12. *Let $i, j \in E$ we say that i leads to j or that j is accessible from i if there exists $n \in \mathbb{N}$ such that $p_{ij}^{(n)} > 0$ and write $i \to j$. We say that i, j communicate if $i \to j$ and $j \to i$ and write $i \leftrightarrow j$ in the affirmative.*

Lemma 2.13. *The relation $\leftrightarrow$ defines an equivalence relation on E, i.e. $\leftrightarrow$ satisfies:*

$$1)\, i \leftrightarrow i \quad (Reflexivity)$$
$$2)\, i \leftrightarrow j \text{ if and only if } j \leftrightarrow i \quad (Symmetry)$$
$$3) \text{ If } i \leftrightarrow j \text{ and } j \leftrightarrow k \text{ then } i \leftrightarrow k \quad (Transitivity)$$

Proof. 1) follows from the fact that $p_{ii}^0 = 1 > 0$.
2) Obvious from the definition.
3) By assumption we have that $i \to j$ and $j \to k$ hence there exist $n_1, n_2 \in \mathbb{N}$ such that $p_{ij}^{(n_1)} > 0$ and $p_{jk}^{(n_2)} > 0$. By Chapman-Kolmogorov we have

$$p_{ik}^{(n_1+n_2)} = \sum_{l \in E} p_{il}^{(n_1)} p_{lk}^{(n_2)} \geq p_{ij}^{(n_1)} p_{jk}^{(n_2)} > 0$$

which shows that $i \to k$ and in a similar fashion we have that $k \to i$ and consequently that $i \leftrightarrow k$. □

Using Lemma 2.13 we can partition the state space E into disjoint equivalence classes. An important result on the nature of these classes is the following.

Theorem 2.14. *If state i is recurrent and $i \leftrightarrow j$ then state j is also recurrent.*

Proof. From the assumptions we have that there exist $n_1, n_2 \in \mathbb{N}$ such that $p_{ij}^{(n_1)} > 0$ and $p_{ji}^{(n_2)} > 0$. Then

$$\mathbb{E}_j(N_j) = \sum_{n=1}^{\infty} p_{jj}^{(n)} \geq \sum_{n=1}^{\infty} p_{ji}^{(n_2)} p_{ii}^{(n)} p_{ij}^{(n_1)} = p_{ij}^{(n_1)} \mathbb{E}_i\left(N_i\right) p_{ji}^{(n_2)} = \infty$$

Where the inequality follows from taking a subset of the paths going from state j and back to j. From Theorem 2.10 we now have that state j is also recurrent. □

Thus, each equivalence class induced by the $\leftrightarrow$-relation is either transient or recurrent. We say that recurrence and transience are *class properties*. It follows that the state space E can be partitioned into disjoint sets $E = T \cup R_1 \cup R_2 \cup \ldots$ Such that T contians all transient states, which need not all communicate and $R_1, R_2, \ldots$ are recurrent communicating classes. The notion of an absorbing state will then be important to us.

Definition 2.15. *If a reccurent class R consists of only one state i, then i is called absorbing and in particular the transition probabilities are $p_{ii} = 1$ and $p_{ij} = 0$ for all $j \neq i$.*

The case of a Markov chain having state space partitioned by $E = T \cup R$ where R is an absorbing state will be the backbone of the (discrete) phase-type.

2.2 Markov Jump Processes

In this section we present the basic definitions and properties of Markov jump processes, which can be seen as the continuous version of the Markov chains discussed in the previous section. Let $\{X_t\}_{t\geq 0}$ be a continuous time stochastic process taking values in a discrete state space E. The definition of the the Markov property is slightly different from the discrete version but the basic nature remains the same; the future of the process depends on the past only through the present.

Definition 2.16 (Markov jump process)**.** *A stochastic process $\{X_t\}_{t\geq 0}$ on a discrete state space E is called a Markov jump process if for every $n \geq 1$ for all $t_n > t_{n-1} > ... > t_1 > 0$ and $i_n, i_{n-1}, ..., i_0 \in E$, we have*

$$\mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}).$$

In this thesis we will only be working with *time-homogeneous* Markov jump processes which satisfy the following definition

Definition 2.17 (Time-homogeneous Markov jump process)**.** *A Markov jump process $\{X_t\}_{t\geq 0}$ on E is called time-homogeneous if*

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \mathbb{P}(X_h = j | X_0 = i)$$

for all $t, h \geq 0$ and $i, j \in E$.

In the time-homogeneous setting we can refer to the above as the h-step transition probability of going from state i to j denoted $p_{ij}(h)$ or p_{ij}^h. The transition probabilities can then be arranged in a matrix

$$\boldsymbol{P}(h) = \{p_{ij}(h)\}_{i,j\in E}, \quad h \geq 0.$$

Note that $\boldsymbol{P}(0) = \boldsymbol{I}$.

Theorem 2.18 (Chapman-Kolmogorov)**.** *The transition matrix $\boldsymbol{P}(h)$ of a Markov jump process $\{X_t\}_{t\geq 0}$ satisfy*

$$\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\boldsymbol{P}(t)$$

for all $s, t \geq 0$.

Proof. As in the discrete case, we condition on the state of the process at an intermediate point in time. For $s, t \geq 0$ and $i, j \in E$ we have

$$\begin{aligned}
p_{ij}(s+t) &= \mathbb{P}(X_{s+t} = j | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_{s+t} = j, X_s = k | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_{s+t} = j | X_s = k, X_0 = i)\mathbb{P}(X_s = k | X_0 = i) \\
&= \sum_{k\in E} \mathbb{P}(X_{s+t} = j | X_s = k)\mathbb{P}(X_s = k | X_0 = i) \quad \textit{(Markov Property)} \\
&= \sum_{k\in E} p_{ik}(s)p_{kj}(t) \quad \textit{(Time-Homogeneity)} \\
&= \{\boldsymbol{P}(s)\boldsymbol{P}(t)\}_{ij}.
\end{aligned}$$

□

Definition 2.19 (Stopping time, continuous version)**.** *Let $\{X_t\}_{t\geq 0}$ be a Markov jump process and let $\mathscr{F}_t = \sigma(X_s : s \leq t)$ i.e. the sigma-algebra generated by the process up to and including time t. A non-negative random variable τ is then called a stopping time for $\{X_t\}_{t\geq 0}$ if $\{\tau \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$.*
The sigma-algebra generated by the process up to τ, $\mathscr{F}_\tau$ is defined by the collection of measurable sets A for which

$$\forall t \geq 0 : A \cap \{\tau \leq t\} \in \mathscr{F}_t$$

Again, we are able to reformulate the definition of a time-homogeneous Markov jump process, $\{X_t\}_{t\geq 0}$, as a process satisfying

$$\mathbb{P}(X_{t+h} = j | \mathscr{F}_t) = \mathbb{P}_{X_t}(X_h = j)$$

For all $t, h \geq 0$ and $j \in E$.

As for Markov chains, a continuous version of the Strong Markov property also holds for Markov jump processes.

Theorem 2.20 (Strong Markov property)**.** *Let $\{X_t\}_{t\geq 0}$ be a Markov jump process on E, further let τ be a stopping time for $\{X_t\}_{t\geq 0}$. Then for all $n \in \mathbb{N}$ and all $0 \leq h_1 \leq h_2 \leq \cdots \leq h_n$ we have on $\{\tau < \infty\}$ that*

$$\mathbb{P}(X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n | \mathscr{F}_\tau) = \mathbb{P}_{X_\tau}(X_{h_1} = i_1, \ldots, X_{h_n} = i_n) \tag{2.3}$$

For $i_1, \ldots, i_n \in E$.

Proof. If τ is deterministic then the result is the standard Markov property. Now assume that τ can take at most a countable number of values, including possibly infinity i.e.

$$\tau \in \{t_1, t_2, \ldots\} \cup \{+\infty\}$$

By definition of conditional probability we have that

$$\mathbb{P}(X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n | \mathscr{F}_\tau) = \mathbb{E}1\left\{X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n\right\} | \mathscr{F}_\tau.$$

By definition of conditional expectation to show (2.3) we need to show that the right hand side is $\mathscr{F}_\tau$-measurable, which is obvious, and that for all $A \in \mathscr{F}_\tau$ we have

$$\int_{A\cap\{\tau<\infty\}} 1\{X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n\} d\mathbb{P} = \int_{A\cap\{\tau<\infty\}} \mathbb{P}_{X_\tau}(X_{h_1} = i_1, \ldots, X_{h_n} = i_n) d\mathbb{P}$$

Since $\{\tau < \infty\} = \bigcup_{i=1}^{\infty} \{\tau = t_i\}$, we have that

$$\begin{aligned}\int_{A\cap\{\tau<\infty\}} 1\{X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n\} d\mathbb{P} &= \sum_{i=1}^{\infty} \int_{A\cap\{\tau=t_i\}} 1\{X_{\tau+h_1} = i_1, \ldots, X_{\tau+h_n} = i_n\} d\mathbb{P} \\ &= \sum_{i=1}^{\infty} \int_{A\cap\{\tau=t_i\}} 1\{X_{t_i+h_1} = i_1, \ldots, X_{t_i+h_n} = i_n\} d\mathbb{P}\end{aligned}$$

By the Markov property

$$\begin{aligned}\mathbb{P}_{X_{t_i}}(X_{h_1} = i_1, \ldots, X_{h_n} = i_n) &= \mathbb{P}(X_{t_i+h_1} = i_1, \ldots, X_{t_i+h_n} = i_n | \mathscr{F}_{t_i}) \\ &= \mathbb{E}1\{X_{t_i+h_1} = i_1, \ldots, X_{t_i+h_n} = i_n\} | \mathscr{F}_t\end{aligned}$$

Then, again by the definition of conditional expectation and the fact that $A \cap \{\tau = t_i\} \in \mathscr{F}_{t_i}$, we have

$$\begin{aligned}\int_{A\cap\{\tau<\infty\}} 1\{X_{\tau+h_1}=i_1,\ldots,X_{\tau+h_n}=i_n\}d\mathbb{P} &= \sum_{i=1}^{\infty}\int_{A\cap\{\tau=t_i\}} \mathbb{P}_{X_{t_i}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)d\mathbb{P}\\ &= \sum_{i=1}^{\infty}\int_{A\cap\{\tau=t_i\}} \mathbb{P}_{X_{\tau}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)d\mathbb{P}\\ &= \int_{A\cap\{\tau<\infty\}} \mathbb{P}_{X_{\tau}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)d\mathbb{P}\end{aligned}$$

Now let τ be any stopping time and define the sequence τ_m by

$$\tau_m = \frac{[2^m\tau+1]}{2^m} > \tau, \quad m \geq 1$$

It is clear that τ_m takes only countably many values and that $\tau_m \downarrow \tau$ as $m \to \infty$. In particular the first part of the proof holds for τ_m i.e.

$$\mathbb{P}(X_{\tau_m+h_1}=i_1,\ldots,X_{\tau_m+h_n}=i_n|\mathscr{F}_{\tau_m}) = \mathbb{P}_{X_{\tau_m}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)$$

on $\{\tau_m<\infty\}$, or equivalently that for all $A\in\mathscr{F}_{\tau_m}$ we have

$$\mathbb{E}\left(1\{X_{\tau_m+h_1}=i_1,\ldots,X_{\tau_m+h_n}=i_n\}1_A\right) = \mathbb{E}\left(\mathbb{P}_{X_{\tau_m}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)1_A\right)$$

Note now that the assertion is equivalent to prove that

$$\mathbb{E}\left(1\{X_{\tau+h_1}=i_1,\ldots,X_{\tau+h_n}=i_n\}1_A\right) = \mathbb{E}\left(\mathbb{P}_{X_{\tau}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)1_A\right)$$

For all $A\in\mathscr{F}_{\tau}$.
So now, pick $A\in\mathscr{F}_{\tau}$ then $A\in\mathscr{F}_{\tau_m}$ as well, since $\tau_m\downarrow\tau$ which then implies $\mathscr{F}_{\tau}\subseteq\mathscr{F}_{\tau_m}$ for all $m\geq 1$. By right-continuity of Markov jump processes $X_{\tau_m}\to X_{\tau}$, then by dominated convergence,

$$\mathbb{E}\left(1\{X_{\tau_m+h_1}=i_1,\ldots,X_{\tau_m+h_n}=i_n\}1_A\right) \to \mathbb{E}\left(1\{X_{\tau+h_1}=i_1,\ldots,X_{\tau+h_n}=i_n\}1_A\right)$$

as $m\to\infty$. On the other hand, since

$$\mathbb{P}(\cdot|X_{\tau_m}=j) = \frac{\mathbb{P}(\cdot\cap\{X_{\tau_m}=j\}}{\mathbb{P}(X_{\tau_m}=j)} \to \frac{\mathbb{P}(\cdot\cap\{X_{\tau}=j\}}{\mathbb{P}(X_{\tau}=j)} = \mathbb{P}(\cdot|X_{\tau}=j)$$

We see that $\mathbb{P}_{X_{\tau_m}}(\cdot)\to\mathbb{P}_{X_{\tau}}(\cdot)$, and again, by dominated convergence

$$\mathbb{E}\left(\mathbb{P}_{X_{\tau_m}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)1_A\right) \to \mathbb{E}\left(\mathbb{P}_{X_{\tau}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)1_A\right)$$

as $m\to\infty$ therefore

$$\mathbb{E}\left(1\{X_{\tau+h_1})=i_1,\ldots,X_{\tau+h_n}=i_n\}1_A\right) = \mathbb{E}\left(\mathbb{P}_{X_{\tau}}(X_{h_1}=i_1,\ldots,X_{h_n}=i_n)1_A\right)$$

For all $A\in\mathscr{F}_{\tau}$. □

By nature transitions occurs by jumps at time points which we will denote $S_0=0<S_1<S_2<...$ i.e $S_n=\inf\{t>S_{n-1}:X_t\neq X_{S_{n-1}}\}$, $n\geq 1$. The time between jumps $T_n=S_{n+1}-S_n$ are called holding or sojourn times. The sequence of visited states $Y_0,Y_1,...$ is then given by $Y_n=X_{S_n}$.
In the case of a last S_n i.e. if $\{X_t\}_{t\geq 0}$ enters an absorbing state, we will define $T_n=T_{n+1}=\ldots=\infty$ and $Y_n=Y_{n+1}=\ldots=k$ if $X_{S_n}=k$. Knowing the sequence of visited states and the holding times in each state, it is clear that we can reconstruct the sample path of $\{X_t\}_{t\geq 0}$ from $\{(Y_n,T_n)\}_{n\in\mathbb{N}}$. We will in the next theorem characterize the joint distribution of (Y_k,T_k) for $k\leq n$.

Theorem 2.21 (Structure of Markov jump process)**.** *There exist real numbers $\lambda_i \geq 0$ and a transition matrix $\boldsymbol{Q} = \{q_{ij}\}$ such that*

$$\mathbb{P}_i\left(Y_n = i_n, T_{n-1} > t_n, \ldots, Y_1 = i_1, T_0 > t_1\right) = \prod_{k=1}^{n} q_{i_{k-1},i_k} e^{-\lambda_{i_{k-1}} t_k} \tag{2.4}$$

where $i_0 = i$.

Proof. Fix $i \in E$. The first step is to consider tail probabilities of the time of the first jump given initiation in state i.

$$f_i(t) = \mathbb{P}_i(T_0 > t).$$

We then have for $t, s \geq 0$ that

$$\begin{aligned}
f_i(t+s) &= \mathbb{P}_i(T_0 > t+s) \\
&= \mathbb{E}_i(\mathbb{P}_i(T_0 > t+s|\mathscr{F}_t)) \quad \textit{(Tower Property)} \\
&= \mathbb{E}_i(\mathbb{P}_i(T_0 > t+s, T_0 > T|\mathscr{F}_t)) \quad (T_0 > t+s \Rightarrow T_0 > t) \\
&= \mathbb{E}_I(1\{T_0 > t\}\mathbb{P}_i(T_0 > t+s|\mathscr{F}_t)) \quad \textit{(Mesuaribilty)} \\
&= \mathbb{E}_i(1\{T_0 > t\}\mathbb{P}_i(T_0 > s)) \quad \textit{(Markov Property and } X_t = i \textit{ on } \{T_0 > t\}\ \textit{)} \\
&= f_i(t) f_i(s).
\end{aligned}$$

Since f_i is non-decreasing (it is a survival function) and $f_i(0) = 1$ by the characterization of the exponential function f_i must have the form

$$f_i(t) = e^{-\lambda_i t}$$

for some $\lambda_i \geq 0$. Now, consider

$$\begin{aligned}
\mathbb{P}_i(Y_1 = j, T_0 > t) &= \mathbb{E}_i(\mathbb{P}_i(Y_1 = j, T_0 > t|\mathscr{F}_t)) \quad \textit{(Tower property)} \\
&= \mathbb{E}_i(1\{Y_1 = j\}\mathbb{P}_i(T_0 > t|\mathscr{F}_t)) \quad \textit{(Measurability)} \\
&= \mathbb{E}_i(1\{Y_1 = j\}\mathbb{P}_i(T_0 > t|X_t)) \quad \textit{(Markov Property)}
\end{aligned}$$

On $\{T_0 > t\}$ the process is still in the initial state i at time t i.e. $X_t = i$ and since T_0 is exponentially distributed with rate parameter λ_i if $\lambda_i > 0$ or in the case of $\lambda_i = 0$ we have $T_0 = \infty$, then $q_{ij} = \delta_{ij}$. In either case we get, due to the memoryless property of the exponential distribution, that

$$\mathbb{P}_i(Y_1 = j|X_t) = \mathbb{P}_i(Y_1 = j) \text{ on } \{T_0 > t\}$$

Now, setting $q_{ij} = \mathbb{P}_i(Y_1 = j)$ we get

$$\mathbb{P}_i(Y_1 = j, T_0 > t) = \mathbb{E}_i(1\{T_0 > t\}q_{ij}) = q_{ij}e^{-\lambda_i t} \tag{2.5}$$

Finally, we use an induction step, let

$$p_m = \mathbb{P}_i(Y_m = i_m, T_{m-1} > t_m, \ldots, Y_1 = i_1, T_0 > t_1)$$

and assume that

$$p_n = \prod_{k=1}^{n} q_{i_{k-1},i_k} e^{-\lambda_{i_{k-1}} t_k}$$

then

$$\begin{aligned}
p_{n+1} &= \mathbb{P}_i(Y_{n+1} = i_{n+1}, T_n > t_{n+1}, \dots, Y_1 = i_1, T_0 > t_1) \\
&= \mathbb{E}_i\left(\mathbb{P}_i(Y_{n+1} = i_{n+1}, T_n > t_{n+1}, \dots, Y_1 = i_1, T_0 > t_1 | \mathscr{F}_{S_n})\right) \quad \textit{(Tower Prop.)} \\
&= \mathbb{E}_i\left(1\{Y_n = i_n, T_{n-1} > t_n, \dots, Y_1 = i_1, T_0 > t_1\}\mathbb{P}_i(Y_{n+1} = i_{n+1}, T_n > t_{n+1} | \mathscr{F}_{S_n})\right) \quad \textit{(Mesurability)} \\
&= \mathbb{P}_{i_n}(Y_1 = i_{n+1}, T_0 > t_{n+1}) p_n \quad \textit{(Strong Markov Prop. and } X_{S_n} = i_n \textit{ on } \{Y_n = i_n\}) \\
&= q_{i_n, i_{n+1}} e^{-\lambda_{i_n} t_{n+1}} \cdot p_n \quad \textit{(eq. (2.5))} \\
&= \prod_{k=1}^{n+1} q_{i_{k-1}, i_k} e^{-\lambda_{i_{k-1}} t_k}
\end{aligned}$$

since the product expression of p_n is obviously true for $n = 1$ the result follows by induction. □

Corollary 2.22. *The sample path of a Markov jump process,$\{X_t\}_{t\geq 0}$ given by $\{(Y_n, T_n)\}_{n\in\mathbb{N}}$ have the following properties.*
a) The sequence of visited states, $\{Y_n\}_{n\in\mathbb{N}}$ is a Markov chain with transition matrix $\boldsymbol{Q}$ called the embedded chain of $\{X_t\}_{t\geq 0}$.
b) The holding times $T_0, T_1, \dots$ are independent given $\{Y_n\}_{n\in\mathbb{N}}$ in particular there exist intensities $\lambda_i \geq 0$ such that $T_n \sim exp(\lambda_i)$ if $Y_n = i$.

Proof. a) The Markov chain property of the embedded chain follows from setting $t_k = 0$ for all k. Then, by definition of conditional probability, we have

$$\begin{aligned}
\mathbb{P}(Y_n = i_n | Y_{n-1} = i_{n-1}, \dots, Y_0 = i_0) &= \frac{\mathbb{P}(Y_n = i_n, Y_{n-1} = i_{n-1}, \dots, Y_0 = i_0)}{\mathbb{P}(Y_{n-1} = i_{n-1}, \dots, Y_0 = i_0)} \\
&= \frac{\prod_{k=1}^{n} q_{i_{k-1} i_k}}{\prod_{k=1}^{n-1} q_{i_{k-1} i_k}} \\
&= q_{i_{n-1} i_n}
\end{aligned}$$

b) By using (2.4) and conditional probability we have

$$\begin{aligned}
\mathbb{P}_i(T_{n-1} > t_n, \dots, T_0 > t_1 | Y_n = i_n, \dots, Y_1 = i_1) &= \frac{\mathbb{P}_i(Y_n = i_n, T_{n-1} > t_n, \dots, Y_1 = i_1, T_0 > t_1)}{\mathbb{P}_i(Y_n = i_n, \dots, Y_1 = i_1)} \\
&= \frac{\prod_{k=1}^{n} q_{i_{k-1} i_k} e^{-\lambda_{i_{k-1}} t_k}}{\prod_{k=1}^{n} q_{i_{k-1} i_k}} \\
&= \prod_{k=1}^{n} e^{-\lambda_{i_{k-1}} t_k}
\end{aligned}$$

Which both implies conditional independence and the exponential distribution. □

Note, as stated earlier, if state i is absorbing then $\lambda_i = 0$ and in the embedded Markov chain we have transition probabilities $q_{ii} = 1$ and $q_{ij} = 0$ for $j \neq i$.

Definition 2.23 (Intensity Matrix)**.** *The intensity matrix, or infinitesimal generator, of a Markov jump process $\{X_t\}_{t\geq 0}$, is the matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j\in E}$ with entries defined as*

$$\lambda_{ij} = \lambda_i q_{ij},\ j \neq i \quad \textit{and} \quad \lambda_{ii} = -\sum_{j\neq i} \lambda_{ij} = -\lambda_i$$

Some remarks are in order.

Remark 1. 1) By construction the row sums of $\mathbf{\Lambda}$ are always zero and rows corresponding to absorbing states are zero filled.
2) The transition probabilities of the embedded Markov chain $\{Y_n\}_{n\geq 0}$ of a Markov jump process $\{X_t\}_{t\geq 0}$ with intensity matrix $\mathbf{\Lambda}$ can simply be deducted from the intensity matrix:

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = q_{ij} = -\frac{\lambda_{ij}}{\lambda_{ii}} = \frac{\lambda_{ij}}{\lambda_i}$$

if i is not an absorbing state. While $q_{ij} = 1, j = i$ zero otherwise if i is absorbing.
3) Since the holding times in state i are exponentially distributed with rate λ_i, we get the useful characterisation of the transition probabilities of the process jumping out of state i during $[t, t+dt)$ given $X_{t-} = i$ (left limit) is $\lambda_i dt$. By which we mean that the probability for the jump to occur during $[t, t+h)$ is $\lambda_i h + o(h)$ where $o(h)$ is the standard small-o notation of a function such that $\lim_{h\downarrow 0} \frac{o(h)}{h} \to 0$.
Since the probability of a jump from state i to j conditionally on a jump in $[t, t+dt)$ and $X_{t-} = i$ is simply q_{ij} we get that the probability of a jump from state i to j during $[t, t+dt)$ given $X_{t-} = i$ is

$$\lambda_i q_{ij} dt = \lambda_{ij} dt. \tag{2.6}$$

Likewise the probability of the process remaining in state i during $[t, t+dt)$ given $X_{t-} = i$ is

$$1 - \lambda_i dt = 1 + \lambda_{ii} dt. \tag{2.7}$$

We will make extensive use of these simple facts when extending the basic phase-type, since this can often be based on how the process behaves during infinitesimal time intervals.

Also note the calculation rule that $(dt)^\alpha = 0$, $\alpha > 1$ a consequence of this is that two independent Markov jump processes cannot simultaneously jump.

Next follows one of the most powerful theorems of the theory of Markov jump processes because it enables us to describe the distribution of the process at time $t \geq 0$.
Let $\boldsymbol{P}^t = \{p_{ij}^t\}_{i,j\in E}$ denote the transition probabilities of $\{X_t\}_{t\geq 0}$ i.e.

$$p_{ij}^t = \mathbb{P}(X_{s+t} = j | X_s = i) = \mathbb{P}(X_t | X_0 = i), \quad s, t \geq 0$$

We then have the theorem known as Kolmogorovs Differential Equations.

Theorem 2.24 (Kolmogorov's Differential Equations)**.** *The transition matrix $\boldsymbol{P}^t$ of a Markov jump process $\{X_t\}_{t\geq 0}$ with transition matrix $\mathbf{\Lambda}$ satisfy*

$$\frac{d}{dt}\boldsymbol{P}^t = \mathbf{\Lambda}\boldsymbol{P}^t = \boldsymbol{P}^t\mathbf{\Lambda}. \tag{2.8}$$

The first equation is known as the backward differential equation while the second is known as the forward differential equation.

Proof. We start by showing the backward equation. First step is to show that

$$p_{ij}^t = e^{-\lambda_i t}\left(\delta_{ij} + \int_0^t e^{\lambda_i u} \sum_{k\neq i} \lambda_{ik} p_{kj}^u du\right) \tag{2.9}$$

Conditioning on the first jump T_0 there are two options, either the jump happens later than t and $i = j$ or the process jumps to some intermediate state $k \neq i$ after some time $0 \leq u \leq t$ here after the problem is self similar with quantity of interest p_{kj}^{t-u} as a consequence of the Strong Markov property.
If $\lambda_i = 0$ (absorbing state) then $p_{ij}^t = e^{-\lambda_i t}\delta_{ij} = \delta_{ij}$ and the result holds.

Otherwise, for $\lambda_i > 0$ we get

$$\begin{aligned}
p_{ij}^t &= \mathbb{P}(X_t = j | X_0 = i) \\
&= \mathbb{P}_i(T_0 > t)\delta_{ij} + \int_0^t \lambda_i e^{-\lambda_i s} \sum_{k \neq i} q_{ik} p_{kj}^{t-s} ds \\
&= e^{-\lambda_i t}\delta_{ij} + \int_0^t \lambda_i e^{-\lambda_i (t-u)} \sum_{k \neq i} q_{ik} p_{kj}^u \, du \quad (\textit{Change of variable } u = t - s) \\
&= e^{-\lambda_i t} \left(\delta_{ij} + \int_0^t e^{\lambda_i u} \sum_{k \neq i} \lambda_{ik} p_{kj}^u du \right) \quad (q_{ij} = \frac{\lambda_{ij}}{\lambda_i})
\end{aligned}$$

Now, since that for any $T > 0$, the function $u \mapsto e^{\lambda_i u} \sum_{k \neq i} \lambda_{ik} p_{kj}^u$ is bounded on $[0, T]$ implying that $t \mapsto p_{ij}^t$ is continuous (the integral of a bounded integrand is continuous). It then follows that $u \mapsto \sum_{k \neq i} \lambda_{ik} p_{kj}^u$ is also continuous and therefore $t \mapsto p_{ij}^t$ is differentiable (the integral of a continuous function is differentiable). Thus, by the chain rule and the Fundamental theorem of calculus

$$\begin{aligned}
\frac{d}{dt} p_{ij}^t &= -\lambda_i \underbrace{e^{-\lambda_i t} \left(\delta_{ij} + \int_0^t e^{\lambda_i u} \sum_{k \neq i} \lambda_{ik} p_{kj}^u du \right)}_{= p_{ij}^t} + e^{-\lambda_i t} e^{\lambda_i t} \sum_{k \neq i} \lambda_{ik} p_{kj}^t \\
&= -\lambda_i p_{ij}^t + \sum_{k \neq i} \lambda_{ik} p_{kj}^t \\
&= \sum_k \lambda_{ik} p_{kj}^t, \quad (-\lambda_i = \lambda_{ii})
\end{aligned}$$

which in matrix form is equivalent to

$$\frac{d}{dt} \boldsymbol{P}^t = \boldsymbol{\Lambda} \boldsymbol{P}^t$$

which shows the backward differential equations.
For the forwards equations we only prove the statement under the assumption of bounded intensities i.e. we assume that $\sup_{i \in E} \lambda_i = K < \infty$. However the result does hold in general.
The idea of the proof is to use Chapman-Kolmogorov and the backward equation to write

$$\frac{d}{ds} \boldsymbol{P}^{t+s} = \frac{d}{ds} \boldsymbol{P}^t \boldsymbol{P}^s = \boldsymbol{P}^t \frac{d}{ds} \boldsymbol{P}^s = \boldsymbol{P}^t \boldsymbol{\Lambda} \boldsymbol{P}^s$$

And then letting $s \downarrow 0$ such that $\boldsymbol{P}^s \to \boldsymbol{I}$. But special care has to be made to ensure validity of differentiation under the infinite sums (when E is countable infinite).
We will use the following result from analysis: Let $F(x) = \sum_k f_k(x)$, it holds that if f_k' is continuous and the series $\sum_k f_k'$ converges uniformly, then $F'(x) = \sum_k f'(x)$. Which we will apply to $s \mapsto p_{ij}^{s+t} = \sum_k p_{ik}^t p_{kj}^s$

To show continuity of the derivatives note, by the backward equation, that

$$\frac{d}{ds} p_{ik}^t p_{kj}^s = p_{ik}^t \frac{d}{ds} p_{kj}^s = p_{ik}^t \sum_l \lambda_{kl} p_{lj}^s$$

since $|\lambda_{kl} p_{lj}^s| \leq |\lambda_{kl}|$ and $\sum_l |\lambda_{kl}| = |\lambda_{kk}| + \sum_{l \neq k} |\lambda_{kl}| \leq 2K < \infty$ Weierstraus M-test gives that $\sum_l \lambda_{kl} p_{lj}^s$ converges uniformly and since the functions $s \mapsto \lambda_{kl} p_{lj}^s$ are continuous since p_{lj}^s are

differentiable, so is $s \mapsto \frac{d}{ds} p_{kj}^s = \sum_l \lambda_{kl} p_{lj}^s$.
For the uniform convergence of the derivatives, we see that

$$\begin{aligned} |p_{ik}^t \frac{d}{ds} p_{kj}^s| &= |p_{ik}^t \sum_l \lambda_{kl} p_{lj}^s| \\ &\leq p_{ik}^t \sum_l |\lambda_{kl} p_{lj}^s| \\ &\leq p_{ik}^t 2K. \end{aligned}$$

And since $\sum_k p_{ik}^t 2K = 2K \sum_k p_{ik}^t = 2K < \infty$ Weierstrauss' M-test gives uniform convergence of $\sum_k p_{ik}^t \frac{d}{ds} p_{kj}^s$. Therefore

$$\frac{d}{ds} p_{ij}^{t+s} = \sum_k p_{ik}^t \sum_l \lambda_{kl} p_{lj}^s$$

or in matrix notation

$$\frac{d}{ds} \boldsymbol{P}^{t+s} = \boldsymbol{P}^t \boldsymbol{\Lambda} \boldsymbol{P}^s.$$

With the same uniform convergence we can interchange the limits when letting $s \downarrow 0$ from which $\boldsymbol{P}^s \to \boldsymbol{I}$ from which we get

$$\begin{aligned} \lim_{s \downarrow 0} \frac{d}{ds} \boldsymbol{P}^{t+s} &= \lim_{s \downarrow 0} \lim_{h \to 0} \frac{\boldsymbol{P}^{t+s+h} - \boldsymbol{P}^{t+s}}{h} \\ &= \lim_{h \to 0} \lim_{s \downarrow 0} \frac{\boldsymbol{P}^{t+s+h} - \boldsymbol{P}^{t+s}}{h} \quad \textit{(Uniform Convergence)} \\ &= \lim_{h \to 0} \frac{\boldsymbol{P}^{t+h} - \boldsymbol{P}^t}{h} \quad \textit{(Continuity of } t \mapsto \boldsymbol{P}^t\textit{)} \\ &= \frac{d}{dt} \boldsymbol{P}^t \end{aligned}$$

at the same time we have

$$\lim_{s \downarrow 0} \frac{d}{ds} \boldsymbol{P}^{t+s} = \lim_{s \downarrow 0} \boldsymbol{P}^t \boldsymbol{\Lambda} \boldsymbol{P}^s = \boldsymbol{P}^t \boldsymbol{\Lambda} \lim_{s \downarrow 0} \boldsymbol{P}^s = \boldsymbol{P}^t \boldsymbol{\Lambda} \boldsymbol{I} = \boldsymbol{P}^t \boldsymbol{\Lambda}$$

From which we conclude that

$$\frac{d}{dt} \boldsymbol{P}^t = \boldsymbol{P}^t \boldsymbol{\Lambda}$$

□

An immediate corollary, which extends the discussion of Remark 1 follows from the theorem.

Corollary 2.25. *The transition probabilities p_{ij}^h of Markov jump process are differentiable in h and satisfy*

$$p_{ij}^h = \lambda_{ij} h + o(h),\ j \neq i, \quad \textit{and} \quad p_{ii}^h = 1 + \lambda_{ii} h + o(h) = 1 - \lambda_i h + o(h) \tag{2.10}$$

or written more compactly

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \delta_{ij} + \lambda_{ij} h + o(h)$$

For all $i, j \in E$.

Kolmogorov's differential equations are of great importance in the theory of Markov jump processes because they are used to find the form of $\boldsymbol{P}^t$. In the case of finite state space E we get a very simple form, involving the matrix exponential of the intensity matrix, which will be made precise in the following.

Definition 2.26 (Matrix Exponential)**.** *The exponential of an $n \times n$ matrix $\boldsymbol{A}$ is defined by the series*

$$e^{\boldsymbol{A}} = \sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n}{n!},$$

with $\boldsymbol{A}^0 = \boldsymbol{I}$.

Lemma 2.27. *Let $\boldsymbol{A}$ be a real square matrix. The matrix function given by $t \mapsto e^{\boldsymbol{A}t} = \sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n t^n}{n!}$ converges uniformly on $[-h, h]$ for any $h > 0$. Hence it is well defined on $\mathbb{R}$.*

Proof. Let $h > 0$ and $t \in [-h, h]$. Then for any $\epsilon > 0$ we have for a matrix-norm $||\cdot||$ the tail of the series evaluated at t is

$$\begin{aligned}\left|\left|\sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n t^n}{n!} - \sum_{n=0}^{m} \frac{\boldsymbol{A}^n t^n}{n!}\right|\right| &= \left|\left|\sum_{n=m+1}^{\infty} \frac{\boldsymbol{A}^n t^n}{n!}\right|\right| \leq \sum_{n=m+1}^{\infty} \left|\left|\frac{\boldsymbol{A}^n t^n}{n!}\right|\right| \\ &= \sum_{n=m+1}^{\infty} \frac{||\boldsymbol{A}^n|| t^n}{n!} \leq \sum_{n=m+1}^{\infty} \frac{||\boldsymbol{A}||^n h^n}{n!} < \epsilon\end{aligned}$$

for m large enough due to the convergence of the power series of the real exponential function. □

Now, it is not difficult to see by differentiating each term that

$$\frac{d}{dt} e^{\boldsymbol{A}t} = \boldsymbol{A} e^{\boldsymbol{A}t} = e^{\boldsymbol{A}t} \boldsymbol{A}$$

In particular, if $\boldsymbol{A}$ is non-singular, we have the integral given by

$$\int e^{\boldsymbol{A}t} dt = \boldsymbol{A}^{-1} e^{\boldsymbol{A}t}. \tag{2.11}$$

Another obvious but useful fact is that matrix exponentials of a diagonal matrix is the corresponding diagonal matrix of the exponential of each diagonal entry.

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \Rightarrow e^D = \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & \dots & e^{\lambda_n} \end{pmatrix} \tag{2.12}$$

Lemma 2.28. *For all $t, s \in \mathbb{R}$ we have that*

$$e^{\boldsymbol{A}(t+s)} = e^{\boldsymbol{A}t} e^{\boldsymbol{A}s}$$

Proof. By definition we have

$$\begin{aligned} e^{\boldsymbol{A}t} e^{\boldsymbol{A}s} &= \left(\sum_{j=0}^{\infty} \frac{\boldsymbol{A}^j t^j}{j!}\right)\left(\sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k s^k}{k!}\right) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^{j+k} t^j s^k}{j!k!}. \end{aligned}$$

Let $n = j + k$, so $k = n - j$. Then by changing order of the summing of the triangular array and using the binomial theorem we get

$$\begin{aligned} e^{\boldsymbol{A}t} e^{\boldsymbol{A}s} &= \sum_{j=0}^{\infty} \sum_{n \geq j} \frac{\boldsymbol{A}^n t^j s^{n-j}}{j!(n-j)!} = \sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n}{n!} \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} t^j s^{n-j} \\ &= \sum_{n=0}^{\infty} \frac{\boldsymbol{A}^n (t+s)^n}{n!} = e^{\boldsymbol{A}(t+s)}. \end{aligned}$$

□

The above is an example of $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}}$. This is en general not true, but if $\boldsymbol{A}$ and $\boldsymbol{B}$ commute i.e. if $\boldsymbol{AB} = \boldsymbol{BA}$ it holds true. Lets formalize this in a Theorem.

Theorem 2.29. *If $\boldsymbol{AB} = \boldsymbol{BA}$ we have $e^{\boldsymbol{AB}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}}$.*

Proof. By definition by have

$$e^{\boldsymbol{A}}e^{\boldsymbol{B}} = \sum_{i=0}^{\infty} \frac{\boldsymbol{A}^i}{i!} \sum_{j=0}^{\infty} \frac{\boldsymbol{B}^j}{j!} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\boldsymbol{A}^i\boldsymbol{B}^j}{i!j!}$$

The double sum is given by the sum of the array below.

$$\begin{array}{cccc} \frac{\boldsymbol{A}^0\boldsymbol{B}^0}{0!0!} & \frac{\boldsymbol{A}^0\boldsymbol{B}^1}{0!1!} & \frac{\boldsymbol{A}^0\boldsymbol{B}^2}{0!2!} & \cdots \\ \frac{\boldsymbol{A}^1\boldsymbol{B}^0}{1!0!} & \frac{\boldsymbol{A}^1\boldsymbol{B}^1}{1!1!} & \frac{\boldsymbol{A}^1\boldsymbol{B}^2}{1!2!} & \cdots \\ \frac{\boldsymbol{A}^2\boldsymbol{B}^0}{2!0!} & \frac{\boldsymbol{A}^2\boldsymbol{B}^1}{2!1!} & \frac{\boldsymbol{A}^2\boldsymbol{B}^2}{2!2!} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

By summing the diagonals we get

$$e^{\boldsymbol{A}}e^{\boldsymbol{B}} = \sum_{n=0}^{\infty}\sum_{k=0}^{n} \frac{\boldsymbol{A}^k\boldsymbol{B}^{n-k}}{(n-k)!k!}.$$

Now, If $\boldsymbol{A}$ and $\boldsymbol{B}$ commute we can use the binomial theorem to get

$$e^{\boldsymbol{A}}e^{\boldsymbol{B}} = \sum_{n=0}^{\infty} \frac{1}{n!}(\boldsymbol{A}+\boldsymbol{B})^n = e^{(\boldsymbol{A}+\boldsymbol{B})}$$

□

As a consequence we see that $e^{\boldsymbol{A}}e^{-\boldsymbol{A}} = \boldsymbol{I}$ i.e. the inverse of $e^{\boldsymbol{A}}$ is $e^{-\boldsymbol{A}}$.

Finally we have the useful result, that when the state space of a Markov jump process is finite we have a compact way of writing the transition probabilities this will become extremely useful later.

Corollary 2.30. *For a Markov jump process with state space E finite we have*

$$\boldsymbol{P}^t = e^{\boldsymbol{\Lambda}t} = \sum_{n=0}^{\infty} \frac{\boldsymbol{\Lambda}^n t^n}{n!}. \tag{2.13}$$

Proof. This is standard theory of linear systems of differential equations with boundary condition $\boldsymbol{P}^0 = \boldsymbol{I}$. □

Definition 2.31. *For a Markov jump process $\{X_t\}_{t\geq 0}$ a state $i \in E$ is called recurrent (transient) if the state is reccurent (transient) in the embedded chain $\{Y_n\}_{n\in\mathbb{N}}$.*

Later we will need a small result which can be seen as a so-called thinning of the Markov jump process by deleting certain jumps.

Lemma 2.32 (Thinning of Markov jump process.)**.** *Let $\{X_t\}_{t\geq 0}$ be a Markov jump process on a finite state space $E = \{1, 2, ..., p\}$ with intensity matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j\in E}$.*
Let $\boldsymbol{P}^* = \{p^*_{ij}\}_{i,j\in E^*}$ and $\boldsymbol{\alpha}^* = \{\alpha^*_i\}_{i\in E^*}$ be an arbitrary transition matrix and initial distribution on a subset of states $E^* \subseteq E$.
Let $\{\tilde{X}_t\}_{t\geq 0}$ inherit the holding times of $\{X_t\}_{t\geq 0}$ but initiation and transitions are governed by $\boldsymbol{\alpha}^*$ and $\boldsymbol{P}^*$ by which we mean that $\{\tilde{X}_t\}_{t\geq 0}$ is a stochastic process on E^* with $\tilde{X}_0 \sim \boldsymbol{\alpha}^*$. At arrival of a jump while being in state $i \in E^*$, which is exponentially distributed with rate λ_i, the $\tilde{X}_t$ process jumps to state $j \in E^*$ with probability p^*_{ij}. Note that $j = i$ is allowed.
The resulting process $\{\tilde{X}_t\}_{t\geq 0}$, see figure 2.1, is then again a Markov jump process with initial distribution $\boldsymbol{\alpha}^*$, intensity matrix $\boldsymbol{\Lambda}^*$ given by

$$\lambda^*_{ij} = -\lambda_{ii}p_{ij}, \quad j \neq i$$
$$\lambda^*_{ii} = \lambda_{ii}(1 - p_{ii}).$$

for $i, j \in E^*$.

Proof. It is evident that the construction still allows $\{\tilde{X}_t\}_{t\geq 0}$ to enjoy the Markov property, and since the imposed transition matrix $\boldsymbol{P}^*$ is fixed it remains time-homogeneous. Now for the dynamics of the jumps. Recall the transition probabilities given in (2.10) we will use the infinitesimal version of these.
For $i, j \in E^*$ given that $X_t = i$ then in a small time interval $[t, t + dt)$ the constructed Markov jump process can jump to state j by arrival of a jump, which happens with probability $-\lambda_{ii}dt$. Then conditioned on a jump the probability of jumping to state j is governed by $\boldsymbol{P}^*$ i.e. p^*_{ij}. The constructed Markov jump process can remain in state i in two ways: Either the underlying process simply remains with probability $1 + \lambda_{ii}dt$, or it makes a jump to itself. Hence we get

$$\begin{aligned}
\underset{i\to j, j\neq i}{\text{Jump}}: \quad & -\lambda_{ii}dtp_{ij} := \tilde{\lambda}_{ij}dt \\
\underset{i\to i}{\text{Stay}}: \quad & 1 + \lambda_{ii}dt + \lambda_i dtp_{ii} \Leftrightarrow \\
& 1 + \lambda_{ii}(1 - p_{ii})dt := 1 + \tilde{\lambda}_{ii}dt
\end{aligned}$$

Since $\sum_{j\neq i}\tilde{\lambda}_{ij} = \sum_{j\neq i} -\lambda_{ii}p_{ij} = -\lambda_{ii}(1 - p_{ii}) = -\tilde{\lambda}_{ii}$ the rows all sum to zero. Also note that the transition probabilities of the embedded Markov chain of $\{\tilde{X}_t\}_{t\geq 0}$ are given by:

$$q_{ij} = \frac{\tilde{\lambda}_{ij}}{-\tilde{\lambda}_{ii}} = \frac{-\lambda_{ii}p_{ij}}{-\lambda_{ii}(1 - p_{ii})} = \frac{p_{ij}}{(1 - p_{ii})}$$

which is expected since we have essentially conditioned on jumping out of each state. The constructed Markov jump process is then fully described. □

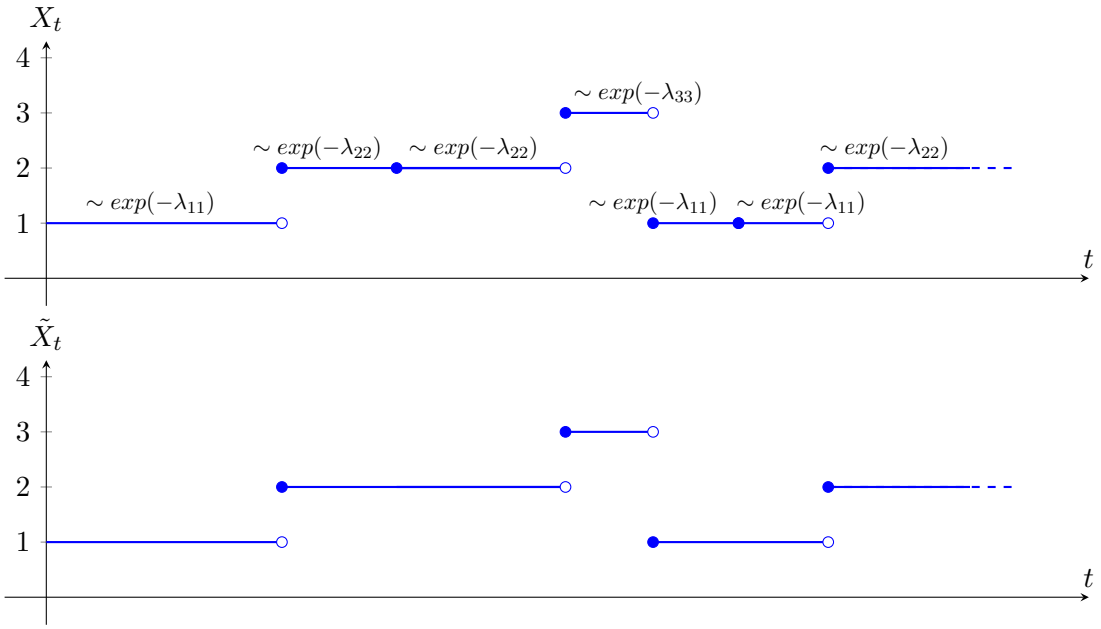

Figure 2.1: The constructed Markov Jump Process,$\{\tilde{X}_t\}_{t\geq 0}$, from Lemma 2.32 with underlying process $\{X_t\}_{t\geq 0}$ on the state space $E = \{1, 2, 3, 4\}$ and $E^* = \{1, 2, 3\}$. Note that the 2nd and 5th jump is deleted.

2.2.1 Multidimensional processes and the Kronecker product.

A natural extension of Markov jump processes is to consider the n-dimensional process $\{Y_t\}_{t\geq 0}$ given by

$$Y_t = (X_t^1, X_t^2, ..., X_t^n)$$

where X_t^i is a Markov jump process on E_i with intensity matrix $\mathbf{\Lambda}_i = \{\lambda_{ij}\}_{i,j\in E_i}$ for $i = 1, ..., n$. The state space of $\{Y_t\}_{t\geq 0}$ will be given by $E = E_1 \times E_2 \times \cdots \times E_n$. The key concept is to consider the multidimensional process as a one-dimensional process on a larger, ordered state space of E. The ordering which is both natural and enabling us to express the transition matrix in a compact way is the *lexicographical* ordering. Which simply means that we change the last entry of E first where after a full cycle of E_n we increase E_{n-1} by one and so forth. Formally we have

Definition 2.33 (Lexicographical ordering)**.** *For $E = E_1 \times E_2 \times \cdots \times E_n$ we say that for $a = (a_1, ..., a_n), b = (b_1, ..., b_n) \in E$ that $a \overset{lex}{<} b$ if there exists a $1 \leq j \leq n$ such that $a_j < b_j$ and $a_1 = b_1, a_2 = b_2, \ldots, a_{j-1} = b_{j-1}$.*

Hence when E is lexicographically ordered we consider the process $\{Y_t\}_{t\geq 0}$ as a Markov jump process on $E' = \{1, 2, ..., d\}$ where $d = |E_1| \cdot |E_2| \cdot ... \cdot |E_n|$ assuming only finite state spaces. For the intensity matrix we will need what is called the Kronecker product which is an matrix operation given by the definition below.

Definition 2.34 (Kronecker- product and sum)**.** *Let $\boldsymbol{A} = \{a_{ij}\}$ be a $m \times n$ matrix and $\boldsymbol{B} = \{b_{ij}\}$ a $r \times s$ matrix. We define the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$ as the $mr \times ns$ matrix given by*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \cdots & a_{1,n}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \cdots & a_{2,n}\boldsymbol{B} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ a_{m,1}\boldsymbol{B} & a_{m,2}\boldsymbol{B} & \cdots & a_{m,n}\boldsymbol{B} \end{pmatrix}$$

When $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of dimension n and r respectively we define the Kronecker sum $\oplus$ by

$$\boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I}_r + \boldsymbol{I}_n \otimes \boldsymbol{B}$$

Where $\boldsymbol{I}_m$ is an $m \times m$ identity matrix.

For simplicity lets look at the case of $n = 2$ that is

$$Y_t = (X_t^1, X_t^2)$$

Where $\{X_t^1\}_{t\geq 0}$ and $\{X_t^2\}_{t\geq 0}$ have state space E_1 and E_2 with intensity matrices $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j\in E_1}$ respectively $\boldsymbol{\Gamma} = \{\gamma_{ij}\}_{i,j\in E_2}$.
Given $Y_t = (i, j)$ then during a small time interval $[t, t+dt]$ simultaneous jumps cannot occur but a jump of the second process while the first remains in state i i.e. $(i, j) \to (i, k)$ is possible with probability $\gamma_{j,k}dt$. Likewise a jump of the first process while the second remains, $(i, j) \to (l, j)$ has probability $\lambda_{i,l}dt$. cf. Remark 1.
It should now be evident that in the 2-dimensional case the intensity matrix of $\{Y_t\}_{t\geq 0}$ is given by

$$\boldsymbol{\Lambda} \oplus \boldsymbol{\Gamma} = \begin{pmatrix} \lambda_{11}\boldsymbol{I} & \lambda_{12}\boldsymbol{I} & \dots & \lambda_{1d_1}\boldsymbol{I} \\ \lambda_{21}\boldsymbol{I} & \lambda_{22}\boldsymbol{I} & \dots & \lambda_{2d_1}\boldsymbol{I} \\ \vdots & \vdots & :::: & \vdots \\ \lambda_{d_1 1}\boldsymbol{I} & \lambda_{d_1 2}\boldsymbol{I} & \dots & \lambda_{d_1 d_1}\boldsymbol{I} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma} & \dots & \boldsymbol{0} \\ \vdots & \vdots & :::: & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{\Gamma} \end{pmatrix} \tag{2.14}$$

The shape of the first matrix in (2.14) assures that jumps of $\{X_t^1\}_{t\geq 0}$ maintain the state of $\{X_t^2\}_{t\geq 0}$, while the diagonal block design of the second matrix restricts jumps of $\{X_t^2\}_{t\geq 0}$ to happen with $\{X_t^1\}_{t\geq 0}$ maintaining its current state.
Using the above it should be clear how the intensity matrix of the general n-dimensional case is constructed.
Regarding the initial distribution $\boldsymbol{\pi}$ of the $\{Y_t\}_{t\geq 0}$ process this can also be described using the Kronecker product on the initial distributions $\boldsymbol{\pi}_i$, $i = 1, 2, ..., n$. I we have lexicographical ordering we clearly have that the initial distribution of the multidimensional process is given by

$$\boldsymbol{\pi} = \boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2 \otimes ... \otimes \boldsymbol{\pi}_n \tag{2.15}$$

and we have proved the following.

Theorem 2.35 (Multidimensional Markov jump process)**.** *Let $\{X_t^i\}_{t\geq 0}$ be independent Markov jump processes with finite state spaces E_i, intensity matrices $\boldsymbol{\Lambda}_i$ and initial distributions $\boldsymbol{\pi}_i$ for $i = 1, 2, ..., n$. Then the joint process*

$$Y_t = (X_t^1, ..., X_t^n)$$

is a Markov jump process, $\{Y_t\}_{t\geq 0}$, on state space $E = E_1 \times E_2 \times ... \times E_n$ and if E is lexicographically ordered, the intensity matrix of $\{Y_t\}_{t\geq 0}$ is given by $\boldsymbol{\Lambda}_1 \oplus \boldsymbol{\Lambda}_2 \oplus ... \oplus \boldsymbol{\Lambda}_n$ and initial distribution $\boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2 \otimes ... \otimes \boldsymbol{\pi}_n$.

Chapter 3

Discrete Phase-Type Distributions

3.1 Discrete Phase-Type

As stated in the previous chapter, the phase-type has a discrete time version, which are distributions on $\mathbb{N} \setminus \{0\}$ constructed using an underlying Markov chain.

Now if we let the Markov chain $\{X_n\}_{n\in\mathbb{N}}$ have state space $E = \{1, 2, ..., p, p+1\}$ such that state $p+1$ is absorbing with the remaining p states being transient, then the transition matrix $\boldsymbol{P}$ has the form

$$\boldsymbol{P} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \tag{3.1}$$

where $\boldsymbol{T}$ is a $p \times p$ sub-transition matrix meaning $\boldsymbol{Te} \leq \boldsymbol{e}$ with strict inequality at at least one entry. $\boldsymbol{t} = \{t_i\}_{i=1,...,p}$ is called the exit probability vector since t_i is the probability of exiting to the absorbing state from state $i \in \{1, \ldots, p\}$. Since $\boldsymbol{Pe} = \boldsymbol{e}$ we have the useful fact $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}$. As in the continuous case we do not allow for immediate absorption hence with $\pi_i = \mathbb{P}(X_0 = i)$ such that $\sum_{i=1}^{p} \pi_i = 1$ the initial distribution of the underlying Markov chain has form $(\boldsymbol{\pi}, 0)$. In this chapter, we will go through the most important properties of the discrete phase-type and summarize and extend the theory with examples in the end of this section.

Definition 3.1 (Discrete Phase-Type)**.** *Let*

$$\tau = \inf\{n \geq 1 \,|\, X_n = p + 1\}$$

be the time until absorption of the Markov chain, $\{X_n\}_{n\in\mathbb{N}}$ as given above, then τ is said to be discrete phase-type distributed, of dimension p, initial distribution $\boldsymbol{\pi}$ and sub-transition matrix $\boldsymbol{T}$. Often we will not make the dimension explicit and simply write

$$\tau \sim DPH(\boldsymbol{\pi}, \mathbf{T}).$$

Markov chain theory allows us to generate formulae for various distributional aspects of the Discrete phase type. The proofs tend to be quite short and instructive hence we will go through proving most of them. First, we need a lemma.

Lemma 3.2. *For $n \geq 1$ and with $\boldsymbol{P}$ defined as in* (3.1) *we have that*

$$\boldsymbol{P}^n = \begin{pmatrix} \boldsymbol{T}^n & \boldsymbol{e} - \boldsymbol{T}^n\boldsymbol{e} \\ \mathbf{0} & 1 \end{pmatrix}.$$

Proof. The above can be shown using a simple induction argument or by straight forward

calculations as below:

$$\begin{aligned}
\boldsymbol{P} &= \begin{pmatrix} \boldsymbol{T} & (\boldsymbol{I}-\boldsymbol{T})\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix} \\
\boldsymbol{P}^2 &= \begin{pmatrix} \boldsymbol{T}^2 & \boldsymbol{T}(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{e} + (\boldsymbol{I}-\boldsymbol{T})\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix} \\
&\vdots \\
\boldsymbol{P}^n &= \begin{pmatrix} \boldsymbol{T}^n & \sum_{k=0}^{n-1} \boldsymbol{T}^k(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{T}^n & \boldsymbol{T}^0\boldsymbol{e} - \boldsymbol{T}^{n-1+1}\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{T}^n & \boldsymbol{e} - \boldsymbol{T}^n\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix}.
\end{aligned}$$

Where we have used that the sum is telescoping and that $\boldsymbol{T}^0 = \boldsymbol{I}$. □

Recall that the n'th power of the transition matrix $\boldsymbol{P}^n = \{p_{ij}^{(n)}\}_{i,j\in E}$ are the n-step transition probabilities

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j | X_0 = i).$$

Observe that $\boldsymbol{P}^n|_{\{1,2,\dots,p\}} = \boldsymbol{T}^n$ and

$$\mathbb{P}_i(X_n = j, \tau > n) = (\boldsymbol{T}^n)_{ij}, \quad i,j \in E$$

since $\{X_n \in \{1,2,\dots,p\}\} \subseteq \{\tau > n\}$.
We make the useful observation that n-step transitions among the transient states $1,2,...,p$ are simply governed by $\boldsymbol{T}^n$ since any path going from transient states i to j cannot go through state $p+1$. That is, prior to absorption the underlying Markov Chain is described by $\boldsymbol{T}$.

Theorem 3.3 (PMF of discrete phase-type)**.** *Let $\tau \sim DPH(\boldsymbol{\pi}, \boldsymbol{T})$ then the probability mass function of τ is given by $f(n) = \mathbb{P}(\tau = n) = \boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t}$ for $n \geq 1$.*

Proof. If $\tau = n$ we must have been in one of the transient states immediately before at time $n-1$. By the Law of Total Probability we get

$$\begin{aligned}
f(n) &= \mathbb{P}(\tau = n) \\
&= \sum_{i=1}^{p} \mathbb{P}(X_n = p+1 | X_{n-1} = i)\mathbb{P}(X_{n-1} = i) \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \mathbb{P}(X_{n-1} = i | X_0 = k)\mathbb{P}(X_0 = k)t_i \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \pi_k(\boldsymbol{P}^{n-1})_{ki}t_i \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \pi_k(\boldsymbol{T}^{n-1})_{ki}t_i \quad \textit{(Lemma 3.2)} \\
&= \boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t}.
\end{aligned}$$

□

Theorem 3.4 (CDF of discrete phase-type)**.** *Let $\tau \sim DPH(\boldsymbol{\pi}, \boldsymbol{T})$ Then $F(n) = \mathbb{P}(\tau \leq n) = 1 - \boldsymbol{\pi}\boldsymbol{T}^n\boldsymbol{e}$*

Proof. The proof relies on the fact that $\{X_n \in \{1,2,...,p\}\} = \{\tau > n\}$. Conditioning on the first state X_0 we have:

$$\begin{aligned}
1 - F(n) &= \mathbb{P}(\tau > n) \\
&= \sum_{i=1}^{p} \mathbb{P}(X_n = i) \\
&= \sum_{k=1}^{p}\sum_{i=1}^{p} \mathbb{P}(X_n = i | X_0 = k)\mathbb{P}(X_0 = k) \\
&= \sum_{k=1}^{p}\sum_{i=1}^{p} \pi_k \boldsymbol{T}^n_{ki} \\
&= \boldsymbol{\pi}\boldsymbol{T}^n\boldsymbol{e}.
\end{aligned}$$

□

Now follows a general result for Markov chains which we will use later.

Theorem 3.5. *For a general Markov chain with $p+1 \times p+1$ transition matrix*

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix}$$

then $(\boldsymbol{I} - \boldsymbol{T})^{-1}$ exists if and only if the states $1, 2, ..., p$ are transient.

Proof. For the "if" part, assume that states $1, 2, ..., p$ are transient. By the previous discussion this implies that we almost surely leave the transient states, $\boldsymbol{T}^n \to \boldsymbol{0}$ as $n \to \infty$. Now consider the equation $(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{x} = \boldsymbol{0}$. Then $\boldsymbol{x} = \boldsymbol{T}\boldsymbol{x} = \boldsymbol{T}^2\boldsymbol{x} = ... = \boldsymbol{T}^n\boldsymbol{x}$ for all $n \geq 1$ now letting $n \to \infty$ shows that the only solution is $\boldsymbol{x} = \boldsymbol{0}$ hence the matrix $(\boldsymbol{I} - \boldsymbol{T})$ is invertible.
For the "only if" part, assume that $(\boldsymbol{I} - \boldsymbol{T})$ is invertible and let a_i be the probability of eventual absorption to the state $p+1$ given that the underlying Markov chain has $X_0 = i$. Then we have

$$a_i = t_i + \sum_{k=1}^{p} t_{ik} a_k$$

since we can split the event into absorption on the first transition with probability t_i or jumping to another transient state and from there reset the problem. The above in matrix notation, for $\boldsymbol{a} = \{a_i\}_{i=1,2,...,p}$ reads

$$\boldsymbol{a} = \boldsymbol{t} + \boldsymbol{T}\boldsymbol{a} \iff (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{a} = \boldsymbol{t}.$$

Using that $\boldsymbol{t}$ is the exit probability vector, we also have

$$(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{a} = \boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}$$

which by the invertibility of $(\boldsymbol{I} - \boldsymbol{T})$ implies that $\boldsymbol{a} = \boldsymbol{e}$, i.e. eventual absorption and thus no return, has probability one for every state and in particular all states $1, 2, ..., p$ are transient. □

Corollary 3.6. *For $DPH_p(\boldsymbol{\pi}, \boldsymbol{T})$ a discrete phase-type distribution, then $(\boldsymbol{I} - \boldsymbol{T})$ is invertible.*

A useful quantity is the so-called *Green Matrix* for the discrete version, which we shall see again in the continuous case.

Definition 3.7 (Green Matrix. Discrete Phase-Type)**.** *For $\tau \sim DPH(\boldsymbol{\pi}, \mathbf{T})$ the associated Green Matrix is defined as*

$$U = \{u_{ij}\}_{i,j=1,...,p} = (\boldsymbol{I} - \boldsymbol{T})^{-1}.$$

The Green matrix has a nice probabilistic interpretation and will become useful in the derivation of moments of discrete phase-types.

Theorem 3.8 (Green Matrix)**.** *The associated Green matrix* $\boldsymbol{U} = \{u_{ij}\}_{i,j=1,2,\ldots,p} = (\boldsymbol{I} - \boldsymbol{T})^{-1}$ *for a discrete phase-type distribution. Then* u_{ij} *is the expected time that the underlying Markov chain spends in state* j *given initiation in state* i*.*

Proof. Let Z_j be the time spend in state j by the underlying Markov chain, which prior to absorption moves according to $\boldsymbol{T}$. Then we have

$$\begin{aligned}
\mathbb{E}_i(Z_j) &= \mathbb{E}_i\left(\sum_{n=0}^{\tau-1} \mathbb{1}\left\{X_n = j\right\}\right) \\
&= \sum_{n=0}^{\tau-1} \mathbb{P}_i\left(X_n = j\right) \quad \text{(Beppo-Levi)} \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i\left(X_n = j, \tau > n\right) \\
&= \sum_{n=0}^{\infty} \left(\boldsymbol{T}^n\right)_{ij} \quad \left(\{X_n = j\} \subseteq \{\tau > n\}\right).
\end{aligned}$$

Since states $\{1, \ldots, p\}$ are transient we have that $\mathbb{E}_i Z_j < \infty$. such that $\sum_{n=1}^{\infty} \boldsymbol{T}^n < \infty$ for all $i, j \in \{1, 2, \ldots, p\}$. Consequently $\sum_{n=1}^{\infty} \boldsymbol{T}^n$ converges. This implies that $\boldsymbol{T}^m \to \boldsymbol{0}$ as $m \to \infty$. Now let

$$\boldsymbol{S}_m = \sum_{k=0}^{m} \boldsymbol{T}^k = \boldsymbol{I} + \boldsymbol{T} + \boldsymbol{T}^2 + \ldots + \boldsymbol{T}^m.$$

Setting $\boldsymbol{S}_\infty = \lim_{m\to\infty} \boldsymbol{S}_m = \sum_{k=0}^{\infty} \boldsymbol{T}^k$ we have

$$\boldsymbol{T}\boldsymbol{S}_m = \boldsymbol{T} + \boldsymbol{T}^2 + \ldots + \boldsymbol{T}^{m+1}$$

implying

$$(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{S}_m = \boldsymbol{I} + \boldsymbol{T}^{m+1}.$$

By letting m tend to infinity we get

$$(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{S}_\infty = \boldsymbol{I}$$

and since $(\boldsymbol{I} - \boldsymbol{T})$ is invertible we conclude that

$$\boldsymbol{U} = \sum_{n=0}^{\infty} \boldsymbol{T}^n = (\boldsymbol{I} - \boldsymbol{T})^{-1}.$$

□

The convergence of the power series above implies the following result.

Corollary 3.9. *For a discrete phase-type distribution* $DPH_p(\boldsymbol{\pi}, \boldsymbol{T})$ *then the eigenvalues of the sub-transition matrix* $\boldsymbol{T}$ *are all contained strictly within the unit circle.*

Proof. If $\boldsymbol{T}$ is diagonalizable, it follows immediately from the decomposition $\boldsymbol{T} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ where $\boldsymbol{D}$ is a diagonal matrix with the eigenvalues on the diagonal. However, things get more complicated if $\boldsymbol{T}$ does not allow itself to diagonalization. Instead, let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{T}$ with corresponding eigenvalue λ hence

$$\boldsymbol{T}\boldsymbol{v} = \lambda\boldsymbol{v}.$$

Iterating we have

$$\boldsymbol{T}^n\boldsymbol{v} = \lambda^n\boldsymbol{v}$$

for all $n \in \mathbb{N}$. By the convergence of $\sum_{n=0}^{\infty}\boldsymbol{T}^n$ we know that

$$\sum_{n=0}^{\infty}\boldsymbol{T}^n\boldsymbol{v} = \sum_{n=0}^{\infty}\lambda^n\boldsymbol{v} < \infty.$$

Since the right hand side converges if and only if $|\lambda| < 1$, we have that the eigenvalues are strictly contained within the unit circle. □

With Theorem 3.8 in mind it is easy to derive the expectation of the discrete phase-type.

Corollary 3.10 (Expectation of DPH)**.** *Let $\tau \sim DPH_p(\boldsymbol{\pi},\boldsymbol{T})$ then the expectation of τ is given by*

$$\mathbb{E}(\tau) = \boldsymbol{\pi}(\boldsymbol{I}-\boldsymbol{T})^{-1}\boldsymbol{e}$$

Proof. Using the properties of the Green matrix, since the expected time spent by the underlying Markov chain in the transient states $1, 2, \ldots, p$ given $X_0 = i$ is given by $\sum_{j=1}^{p} u_{ij}$, we get by conditioning on X_0,

$$\mathbb{E}(\tau) = \sum_{i=1}^{p}\pi_i\sum_{j=1}^{p}u_{ij} = \boldsymbol{\pi}\boldsymbol{U}\boldsymbol{e} = \boldsymbol{\pi}(\boldsymbol{I}-\boldsymbol{T})^{-1}\boldsymbol{e}.$$

□

Theorem 3.11 (Probability Generating Function)**.** *The probability generating function $G(z)$ for $\tau \sim DPH(\boldsymbol{\pi},\boldsymbol{T})$ is given by*

$$G(z) = \mathbb{E}(z^\tau) = z\boldsymbol{\pi}(\boldsymbol{I}-z\boldsymbol{T})^{-1}\boldsymbol{t} = \boldsymbol{\pi}(z^{-1}\boldsymbol{I}-\boldsymbol{T})^{-1}\boldsymbol{t},$$

and exists at least for $|z| \leq 1$.

Proof. We have

$$\begin{aligned}
G(z) &= \mathbb{E}(z^\tau)\\
&= \sum_{n=1}^{\infty} z^n\mathbb{P}(\tau = n)\\
&= \sum_{n=1}^{\infty} z^n\boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t}\\
&= z\sum_{n=0}^{\infty}\boldsymbol{\pi}(z\boldsymbol{T})^n\boldsymbol{t}\\
&= z\boldsymbol{\pi}(\boldsymbol{I}-z\boldsymbol{T})^{-1}\boldsymbol{t}
\end{aligned}$$

since if λ is an eigenvalue of $\boldsymbol{T}$ with eigenvector $\boldsymbol{v}$ then $z\lambda$ is the corresponding eigenvalue of $z\boldsymbol{T}$ with corresponding eigenvector $\boldsymbol{v}$. Since all eigenvalues of $\boldsymbol{T}$ are strictly contained within the unit circle so are the eigenvalues of $z\boldsymbol{T}$ when $|z| \leq 1$. Hence the series $\sum_{n=0}^{\infty}(z\boldsymbol{T})^n$ converges. □

Theorem 3.12. *Let $\tau \sim DPH(\boldsymbol{\pi},\boldsymbol{T})$. Then the kth factorial moment is given by*

$$\mathbb{E}(\tau(\tau-1)\cdots(\tau-k+1)) = k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}(\boldsymbol{I}-\boldsymbol{T})^{-k}\boldsymbol{e}.$$

Proof. Let $|z| \leq 1$ then we know from the previous theorem that the spectral radius of $z\boldsymbol{T}$ is strictly less than 1. Then

$$\begin{aligned}
\frac{d}{dz}(\boldsymbol{I} - z\boldsymbol{T})^{-1} &= \frac{d}{dz}\sum_{n=0}^{\infty}(z\boldsymbol{T})^n \\
&= \sum_{n=1}^{\infty} nz^{n-1}\boldsymbol{T}^n \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{n} z^{n-1}\boldsymbol{T}^n \\
&= \sum_{i=1}^{\infty}\sum_{n=i}^{\infty} z^{n-1}\boldsymbol{T}^n \\
&= \sum_{i=1}^{\infty}\sum_{n=i}^{\infty} z^{n-i}z^{i-1}\boldsymbol{T}^{n-i}\boldsymbol{T}^{i-1}\boldsymbol{T} \\
&= \boldsymbol{T}\sum_{i=1}^{\infty}(z\boldsymbol{T})^{i-1}\sum_{n=i}^{\infty}(z\boldsymbol{T})^{n-i} \\
&= \boldsymbol{T}(\boldsymbol{I} - z\boldsymbol{T})^{-2}.
\end{aligned}$$

Now, we show by induction that

$$\frac{d^k G(z)}{dz^k} = k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}(\boldsymbol{I} - z\boldsymbol{T})^{-k-1}\boldsymbol{t}. \tag{3.2}$$

For $k = 1$, we have by the chain rule for matrix functions and the result above

$$\begin{aligned}
\frac{dG(z)}{dz} &= \boldsymbol{\pi}\left((\boldsymbol{I} - z\boldsymbol{T})^{-1} + z\frac{d}{dz}(\boldsymbol{I} - z\boldsymbol{T})^{-1}\right)\boldsymbol{t} \\
&= \boldsymbol{\pi}\left((\boldsymbol{I} - z\boldsymbol{T})^{-1} + z\boldsymbol{T}(\boldsymbol{I} - z\boldsymbol{T})^{-2}\right)\boldsymbol{t} \\
&= \boldsymbol{\pi}(\boldsymbol{I} - z\boldsymbol{T} + z\boldsymbol{T})(\boldsymbol{I} - z\boldsymbol{T})^{-2}\boldsymbol{t} \\
&= \boldsymbol{\pi}(\boldsymbol{I} - z\boldsymbol{T})^{-2}\boldsymbol{t}
\end{aligned}$$

which coincides with (3.2). Now, assuming the formula holds for k, we get

$$\begin{aligned}
\frac{d^{k+1}G(z)}{dz^{k+1}} &= k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}\frac{d(\boldsymbol{I} - z\boldsymbol{T})^{-k-1}}{dz}\boldsymbol{t} \\
&= k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}\left((k+1)\boldsymbol{T}(\boldsymbol{I} - z\boldsymbol{T})^{-2}(\boldsymbol{I} - z\boldsymbol{T})^{-k}\right)\boldsymbol{t} \\
&= (k+1)!\boldsymbol{\pi}\boldsymbol{T}^{(k+1)-1}(\boldsymbol{I} - z\boldsymbol{T})^{-(k+1)-1}\boldsymbol{t},
\end{aligned}$$

by taking the derivative of the $k+1$ matrix product and using that $\boldsymbol{T}$ and $(\boldsymbol{I} - z\boldsymbol{T})^{-1}$ commute (use the series expression).

Finally, from standard theory of probability generating functions, we have that the kth factorial moment of τ is given by the kth derivative of $G(z)$ evaluated at $z = 1$ i.e.

$$\mathbb{E}\left(\prod_{i=1}^{k}(\tau - i + 1)\right) = \frac{d^k}{dz^k}G(z)\bigg|_{z=1}$$

Using that $\boldsymbol{t} = (\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}$ we get

$$\begin{aligned}
\frac{d^k}{dz^k}G(z)\bigg|_{z=1} &= k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}(\boldsymbol{I} - \boldsymbol{T})^{-k-1}\boldsymbol{t} \\
&= k!\boldsymbol{\pi}\boldsymbol{T}^{k-1}(\boldsymbol{I} - \boldsymbol{T})^{-k}\boldsymbol{e}
\end{aligned}$$

□

Corollary 3.13. *The variance of $\tau \sim DPH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

$$\mathbb{V}(\tau) = 2!\boldsymbol{\pi}\boldsymbol{T}(\boldsymbol{I}-\boldsymbol{T})^{-2}\boldsymbol{e} + \boldsymbol{\pi}(\boldsymbol{I}-\boldsymbol{T})^{-1}\boldsymbol{e} - \left[\boldsymbol{\pi}(\boldsymbol{I}-\boldsymbol{T})^{-1}\boldsymbol{e}\right]^2.$$

In general, if we let $(\tau)_k = \tau(\tau-1)\cdots(\tau-k+1)$ be the falling factorial, it is well known that

$$\mathbb{E}[\tau^k] = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} \mathbb{E}[(\tau)_j]$$

where $\left\{ {k \atop j} \right\}$ is the Stirling numbers of the second kind, i.e. the number of ways to partition a set of k objects into j non-empty subsets.

One of the major drawbacks of the phase-type, both discrete and continuous, is the lack of canonical representations. That is, for two phase-type distributions X, Y with different representations, including different dimensions, we might have $X \overset{d}{=} Y$. A consequence of this is that discrete phase-type might be greatly over-parametrized as the next theorem shows.

Theorem 3.14 (Over parametrizations discrete case)**.** *Let $\tau \sim DPH_m(\boldsymbol{\pi}, \boldsymbol{T})$ such that $\boldsymbol{t} = \boldsymbol{e} - \boldsymbol{T}\boldsymbol{e} = p\boldsymbol{e}$ for some $p \in (0,1)$ then $\tau \sim geom(p)$. That is the m-dimensional discrete phase-type is a geometric distribution, i.e. $P(\tau = k) = (1-p)^{k-1}p$, $k \geq 1$.*

Proof. We will show that the discrete phase-type and the geometric distribution have the same probability generating function.
Recall that for $X \sim geom(p)$ then X has probability generating function $G_X(s) = \mathbb{E}(s^X) = \frac{ps}{1-s(1-p)}$, for $|s| < (1-p)^{-1}$. Also recall the probability generating function of $\tau \sim \mathrm{DPH}(\boldsymbol{\pi}, \boldsymbol{T})$ given by $G_\tau(s) = \mathbb{E}(s^\tau) = s\boldsymbol{\pi}\left(I - s\boldsymbol{T}\right)^{-1}\boldsymbol{t}$ which exists at least for $|s| \leq 1$.
Since $\boldsymbol{T}\boldsymbol{e} = (1-p)\boldsymbol{e}$ we see that,

$$\boldsymbol{\pi}\boldsymbol{T}^n\boldsymbol{e} = \boldsymbol{\pi}\boldsymbol{T}^{n-1}(1-p)\boldsymbol{e} = ... = (1-p)^n. \tag{3.3}$$

Hence,

$$\begin{aligned} G_\tau(s) &= s\boldsymbol{\pi}\left(I - s\boldsymbol{T}\right)^{-1}\boldsymbol{t} \\ &= s\boldsymbol{\pi}\sum_{n=0}^{\infty}(s\boldsymbol{T})^n p\boldsymbol{e} \\ &= sp\sum_{n=0}^{\infty} s^n\boldsymbol{\pi}\boldsymbol{T}^n\boldsymbol{e} \\ &= sp\sum_{n=0}^{\infty}\left[s(1-p)\right]^n \quad (3.3) \\ &= \frac{sp}{1-s(1-p)} = G_X(s). \end{aligned}$$

Where the sum converges if and only if $|s| < (1-p)^{-1}$. □

The result is not surprising since one could also argue from a probabilistic point of view, that at each time step and any state the underlying Markov chain has probability $1-p$ of exiting which is independent of the previous states.

As we shall see in the examples in section 3.2, the class of discrete phase-type distributions includes a wide array of distributions. Although some well known distribution on $\mathbb{N}$ do not

have a discrete phase-type representation e.g. the Poisson distribution with mean $\mu > 0$ as seen by comparing the probability generating function of a discrete phase-type, which is a rational function to the probability generating function of $pois(\mu)$ given by $G(z) = e^{z(\lambda-1)}$. However as the next theorem shows, all discrete distributions with finite support are indeed of discrete phase-type.

Theorem 3.15. *For a discrete random variable X on $\mathbb{N}$ with finite support. Then $X \sim DPH(\boldsymbol{\alpha}, \boldsymbol{T})$ with representation described in the proof below.*

Proof. The idea of the proof is to construct a discrete phase-type τ with underlying Markov chain $\{X_n\}_{n\in\mathbb{N}}$ that moves deterministically and use the initial distribution of the Markov chain to make the distribution of τ coincide with the distribution of X.
Let p_k describe the point masses of X, i.e.

$$\mathbb{P}(X = k) = p_k,\ k \geq 0.$$

Define

$$n = \inf\{n \geq 0 \,|\, \sum_{k=0}^{n} p_k = 1\}.$$

Then $n < \infty$ since the support is assumed finite and X is then concentrated on the set $\{0, 1, \ldots, n\}$.

Now, adhering to the discrete phase-type construction we define a Markov chain with state space $\{1, 2, \ldots, n+1\}$ where state $n+1$ is absorbing and the states $1, 2, \ldots, n$ are transient. We construct the Markov chain to move deterministically upwards by steps of size 1 to the absorbing state $n+1$, i.e.

$$\mathbb{P}(X_1 = j+1 | X_0 = j) = 1, \quad 1 \leq j \leq n$$

and

$$\mathbb{P}(X_1 = n+1 | X_0 = n+1) = 1.$$

Hence the $(n+1) \times (n+1)$ transition matrix has the form

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and $\boldsymbol{P}$ can be partitioned as

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix}$$

where

$$\boldsymbol{T} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \boldsymbol{t} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

For the initial distribution of $\{X_n\}_{n\in\mathbb{N}}$ given by $(\boldsymbol{\alpha}, \alpha_{n+1}) = \{\alpha_k\}_{k\in\{1,2,\ldots,n+1\}}$, i.e. $\boldsymbol{\alpha}$ is the initial distribution on the transient states. Since $\mathbb{P}(X = 0) = p_0$ we let the Markov chain initiate in the absorbing state $n+1$ with probability $\alpha_{n+1} = p_0$. Such that τ takes the value zero with probability p_0.

By the same logic, if the Markov chain is initiated in state j, then it will enter the absorbing state after $n+1-j$ time steps and consequently $\tau = n+1-j$ which should happen with probability p_{n+1-j}. We therefore have the following index equality

$$\alpha_j = \mathbb{P}(\tau = n+1-j) = p_{n+1-j}$$

From which we have $(\alpha_1, \alpha_2, \ldots, \alpha_{n-1}, \alpha_n, \alpha_{n+1}) = (p_n, p_{n-1}, \ldots, p_2, p_1, p_0)$ by construction we have $\tau \sim \mathrm{DPH}_n(\boldsymbol{\alpha}, \boldsymbol{T})$ and

$$\tau \overset{d}{=} X.$$

□

Remark 2. Although the above theorem is for distributions on $\mathbb{N}$ general discrete distributions with finite support on $\mathbb{Z}$ could easily be achieved by a simple shift argument.
For general discrete distributions X on $\mathbb{N}$ which are not of discrete phase-type e.g. the Poisson distribution, by a simply cut-off argument we can construct a sequence of discrete phase-type distributions $\tau_1, \tau_2, \ldots$ such that $\tau_n \xrightarrow{d} X$ for $n \to \infty$.
Lastly, it should be noted that although of interest in itself applications of 3.15 seems cumbersome and insufficient.

3.2 Examples and Extensions

In this part, we will give examples of other stochastic phenomena that can be modelled by a discrete phase-type distribution. The construction almost always comes down to a proper abstraction of the general state space into a smaller one using so-called "macro" states which can greatly reduce the state space that the problem seems to present at first hand. In fact, both in example 3.16 and 3.18 the state space one could be tempted to use is infinite.

Example 3.16. A neat example of how a discrete phase type can be used to tackle even quite complicated problems readily making distributional results available, is in the case of first occurrence of sub-sequences within a general sequence of random variables.
Let $\{X_n\}_{n\geq 1}$ be a Bernoulli process with success probability p i.e. $X_1, X_2, ...$ is an i.i.d. sequence of Bernoulli distributed random variables with $P(X_1 = 1) = 1 - P(X_1 = 0) = p$ for $p \in (0, 1)$. We regard the outcome 1 as a success and 0 as a failure. Our task will be to find the distribution of the first occurrence of k consecutive failures in the sequence which we will denote τ.

To put this into the framework of the discrete phase-type we realise that we can effectively be in $k + 1$ different 'macro' states along the sequence $\{X_n\}_{n\geq 1}$, describing how far along the sub-sequence we are: At any time we have either $0, 1, 2, ..., k - 1$ of the k needed failures. If the length of the sub-sequence reaches k we are done and define this to be the absorbing state.
Now the transition between the transient macro states is given by the fact that if the current length of the sub-sequence is $0 \leq j \leq k - 1$ what can happen in the next step is either that we break our sub-sequence by getting a 1 with probability p, hence having to start over with sub-sequence of length 0, i.e. $j \to 0$ or that we get another 0 with probability $1 - p$ and hence add 1 to the length of the sub-sequence i.e. $j \to j + 1$.
Regarding the first state of X_0 we simply initiate the Markov chain in the first macro state such that we end up with a discrete phase-type of dimension k with initial vector

$$\boldsymbol{\pi} = (1, 0, ..., 0)$$

An important note is that if we instead had $\boldsymbol{\pi}$ being the outcome of the first Bernoulli variable i.e. $\boldsymbol{\pi} = (p, 1 - p, 0, ..., 0)$ then $\tau + 1$ would be our quantity of interest since the definition of the discrete phase type disregards X_0. Therefore, for simplicity we will prefer the first construction.

The sub-transition matrix $\boldsymbol{T}$ is $k \times k$ describing transitions between the macro states given by the current length of the sub-sequence

$$\boldsymbol{T} = \begin{pmatrix} p & 1-p & 0 & 0 & \cdots & 0 \\ p & 0 & 1-p & 0 & \cdots & 0 \\ p & 0 & 0 & 1-p & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots & \ddots & \ddots & \vdots \\ p & 0 & \cdots & \cdots & 0 & 1-p \\ p & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}.$$

By construction, we have that the time of absorption, τ, is the time of observing the sub-sequence at hand. Hence, $\tau \sim \mathrm{DPH}_k(\boldsymbol{\pi}, \mathbf{T})$.

With the whole machinery of discrete phase-types now at our disposal we easily obtain answers to various distributional questions, below we set $n = 200, k = 6, p = 0.5$.

$$\mathbb{P}(\tau \leq n) = 1 - \boldsymbol{\pi}\boldsymbol{T}^n\boldsymbol{e} = 0.8009322$$
$$\mathbb{P}(\tau = m) = \boldsymbol{\pi}\boldsymbol{T}^{m-1}\boldsymbol{t}, \quad m \geq 1$$
$$\mathbb{E}(\tau) = \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{e} = 126$$
$$\mathrm{Var}(\tau) = 2!\boldsymbol{\pi}\boldsymbol{T}(\boldsymbol{I} - \boldsymbol{T})^{-2}\boldsymbol{e} + \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{e} - \left[\boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{e}\right]^2 = 14718$$

The point masses are plotted in Figure 3.1 below for $m = 1, ..., 150$.
Naturally $\mathbb{P}(\tau = m) = 0$, $m \in \{1, 2, ..., 5 = k - 1\}$. Notice the "flat land" at $m \in \{k + 1 = 7, ..., 12 = 2k\}$ these all have probability $p(1 - p)^k$ since conditioned on $\tau = m$ the values of $X_{m-k+1}, ..., X_m$ must be 0 and X_{m-k} must be 1 otherwise the event would have occurred one step before. The values of the remaining part of the sequence, $X_1, ..., X_{m-k-1}$ is thus unrestricted since the only thing disallowed is for another sub-sequence to occur which is not possible since there is less than k remaining variables.

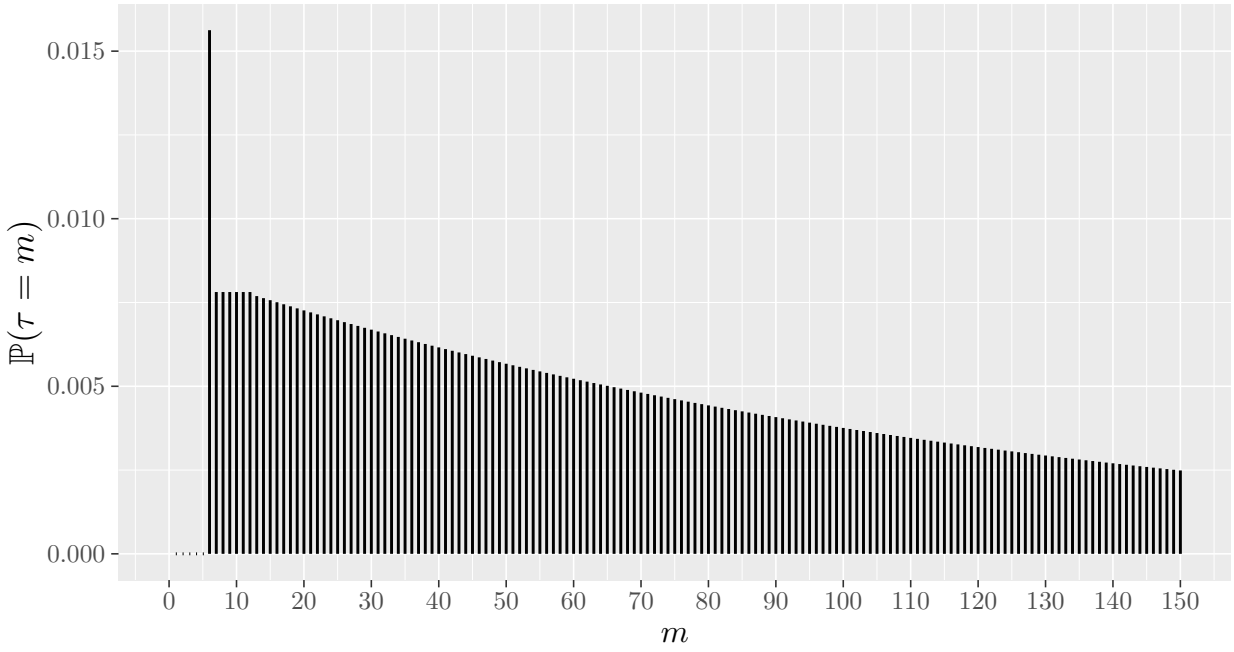

Figure 3.1: Probability mass of τ for $m = 1, \ldots, 150$

It should now be obvious how to construct the sub-transition matrix for other sub-sequences, although care has to be taken when transitioning between the macro-states. e.g. for the first occurrence of the subsequence F, S, S, F denoted τ_{FSSF} the sub-transition matrix is

$$\mathbf{T}_{\text{FSSF}} = \begin{pmatrix} p & 1-p & 0 & 0 \\ 0 & 1-p & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 0 & 0 \end{pmatrix}.$$

and with initial vector $\boldsymbol{\pi}_{\text{FSSF}} = (1, 0, 0, 0)$ then $\tau_{\text{FSSF}} \sim \text{DPH}_4(\boldsymbol{\pi}_{\text{FSSF}}, \boldsymbol{T}_{\text{FSSF}})$.

As a side-note, another neat way of tackling the above problems of the expected time of occurrence of a sub-sequence is by a martingale argument, which goes as follows:
Consider a fair game involving a random number of gamblers. A gambler enters at each time step, betting on the sub-sequence i.e. first he bets \$1 on the first outcome of the sub-sequence and if correct wins $\$\frac{1}{1-p}$ if the outcome needed is F and $\$\frac{1}{p}$ if the outcome needed is S. If not he loses all his bettings.

Each gambler continues to bet all his earnings until either he reaches the sub-sequence and the game is stopped, or he loses. A third possibility is that he becomes a partial winner by betting correctly but not being allowed to finish. When the game is ended there be a winner and a partial winner. See Figure 3.2.

This betting arrangement is then a fair game and if we let N_{FSSF} be the time of observing the sub-sequence, at which point the gamblers will have invested $\$N_{FSSF}$ and the total earnings

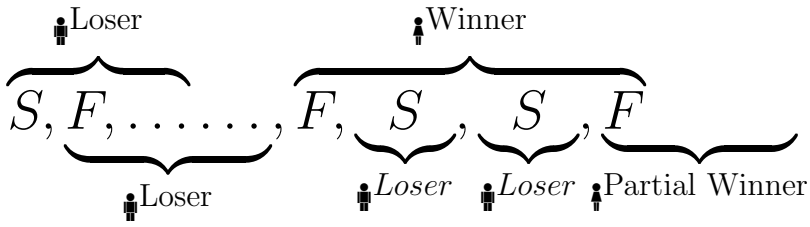

Figure 3.2: The fair game setup discussed in example 3.16

of the two winners will be $\$\frac{1}{1-p}\frac{1}{p}\frac{1}{p}\frac{1}{1-p}+\$\frac{1}{1-p}=\$\frac{1+p^2(1-p)}{p^2(1-p)^2}$. Since the game is fair the net payoff $\frac{1+p^2(1-p)}{p^2(1-p)^2}-N_{FSSF}$ must have expectation zero from which we conclude that

$$\mathbb{E}(N_{FSSF})=\frac{1+p^2(1-p)}{p^2(1-p)^2}.$$

Which coincides with the matrix equations from the phase-type setting, i.e.

$$\mathbb{E}(\tau_{FSSF})=\boldsymbol{\pi}_{\mathrm{FSSF}}(\boldsymbol{I}-\boldsymbol{T}_{\mathrm{FSSF}})^{-1}\boldsymbol{e}=\frac{1}{(p-1)^2}-\frac{1}{p^2(p-1)}-\frac{1}{p(p-1)}-\frac{1}{p-1}=\frac{1+p^2(1-p)}{p^2(1-p)^2}.$$

But note that we do not have any information of other distributional aspects of the time of the first occurrence using the martingale argument.

Having warmed up through Example 3.16 we now notice that we are not restricted to binary outcomes as in the Bernoulli process. The construction only requires the abstraction of setting up an underlying Markov chain on the macro-states, given by how far along the sub-sequence we are. Hence the previous result can be generalized.

Theorem 3.17. *For a sequence $X_1, X_2, \ldots$ of independent identically distributed random variables taking values in a discrete set $A=\{a_1,a_2,\ldots\}$. The first occurrence of some k sub-sequence $(a_{i_1},a_{i_2},\ldots,a_{i_k})$ is of discrete phase-type distribution of dimension k and sub-transition matrix $\boldsymbol{T}$ given by Algorithm 1 and initial vector $\boldsymbol{\pi}=(1,0,\ldots,0)$.*

Proof. As discussed above we let the underlying Markov chain have macro-states corresponding to the progress along the k-subsequence at each time step. Hence the initial distribution is once again given by $\boldsymbol{\pi}=(1,0,\ldots,0)$.

We now describe an algorithmic approach for constructing the $\boldsymbol{T}$ matrix row by row using backwards induction. For the last, kth row of $\boldsymbol{T}$, conditioned on being in the macro state of having observed $k-1$ of the subsequence i.e. $a_{i_1},a_{i_2},\ldots,a_{i_{k-1}}$, there are four types of continuation:

1) The last needed outcome a_{i_k} comes up, with probability p_{i_k} and the underlying Markov chain is absorbed. Since $\boldsymbol{T}$ is the sub-transition matrix, this is not reflected in $\boldsymbol{T}$.

2) Another outcome j, say than the required a_{i_k} comes up, but can be used in conjugation with some of the previous outcomes, the simplest case being if $a_{i_{k-1}}=a_{i_1}$ and $a_{i_2}=j$. We will denote the general case $(a_{i_{k-m}},\ldots,a_{i_{k-1}}=a_{i_1},\ldots,a_{i_m})$ with $a_{i_{m+1}}=j$ a short-cut of length m with needed outcome j. See figure 3.3.
If there exists a short-cut of length m with needed outcome j then the underlying process moves to the macro state of having observed $m+1$ of the needed subsequence with probability p_j.
Note that it is possible to have multiple short-cuts with same needed outcome. In that case only the longest short-cut is used.

3) If no short-cuts of any length exists with needed outcome a_{i_1}, then the underlying process moves to the macro states of having observed 1 with probability p_{i_1}.

4) An outcome comes up which can not be used in any way and hence brings us back to the macro state of having observed zero of the k needed elements of the subsequence. This happens with the complementary of the probabilities mentioned above.

All of the above rules, with 1) being appropriately changed, apply for each macro state from which we get Algorithm 1 below.

Algorithm 1: Sub-transition matrix for DPH-representation.

Input: k-subsequence $(a_{i_1}, a_{i_2}, \ldots, a_{i_k})$, probabilities of each outcome of the subsequence $(p_{i_1}, p_{i_2}, \ldots, p_{i_k})$.

Result: Sub-transition matrix $\boldsymbol{T}$ for DPH representation of first occurrence of k-subsequence.

Initialization;
$\boldsymbol{T} \leftarrow$ zero filled $k \times k$ matrix;
for $n = k$ **to** 2 **do**
  Find all $1 \leq m \leq n-1$ such that $(a_{i_{n-m}}, \ldots, a_{i_{n-1}}) == (a_{i_1}, \ldots, a_{i_m})$ denote M_n ;
  if $m, m' \in M_n$ *with* $a_{i_{m+1}} = a_{i_{m'+1}}$ **then**
    Delete $\min(m, m')$
  forall $m \in M_n$ **do**
    if $m = k-1$ **then**
      skip;
    else
      $T_{n,m+2} \leftarrow p_{i_{m+1}}$;
  if $a_{i_1} \neq a_{i_{m+1}}$ *for all* $m \in M_n$ **then**
    $T_{n,2} \leftarrow p_{i_1}$;
  if $n = k$ **then**
    $T_{n,1} \leftarrow 1 - sum(T_{[n,.]}) - p_{i_{n+1}}$;
  else
    $T_{n,1} \leftarrow 1 - sum(T_{[n,.]})$;
$T_{1,2} \leftarrow p_{i_1}$;
$T_{1,1} \leftarrow 1 - p_{i_1}$;

□

The above algorithm is implemented in the `R`-function `DPHSubSeqGen()` in the `R`-package accompanying this thesis. The syntax is slightly unintuitive, see Appendix A.2 for an example.

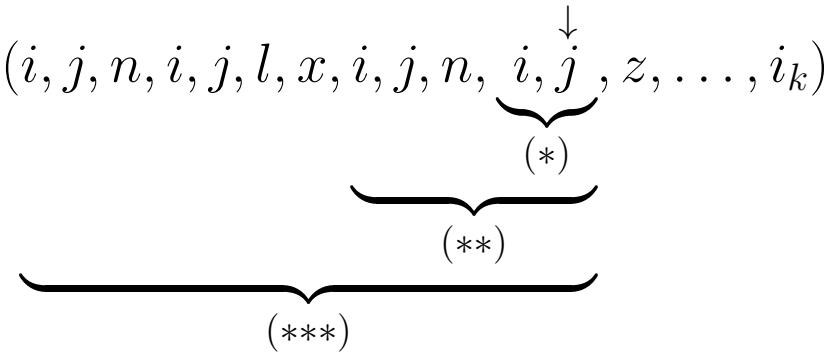

Figure 3.3: Possible movements along the macro states. Consider the sub-sequence above, and consider the case of having reached the outcome in the sub-sequence marked with the down arrow, i.e. the next needed outcome in the sub-sequence is z.
$(*)$ is a short-cut of length 2 with needed outcome n.
$(**)$ is a short-cut of length 4 with needed outcome l.
$(***)$ is the trivial short-cut of the entire sub-sequence up until the current state with needed outcome z.
Also, since none of the short-cuts needs the initial outcome, i, of the sub-sequence, in the case of i being the next outcome the underlying process would move to the macro state of having observed 1 outcome in the k-sub sequence.

Example 3.18. Continuing with the set-up of the first part of Example 3.16. If we, instead of focusing on observing k *successive* failures, simply count the number of failures and let the underlying Markov chain be absorbed at the realisation of the kth failure, we get the $k \times k$ sub-transition matrix

$$\hat{\mathbf{T}} = \begin{pmatrix} p & 1-p & 0 & 0 & \dots & 0 \\ 0 & p & 1-p & 0 & \dots & 0 \\ 0 & 0 & p & 1-p & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & p & 1-p \\ 0 & 0 & \dots & 0 & 0 & p \end{pmatrix}.$$

We see that the exit probability vector is given by $\hat{\mathbf{t}} = \boldsymbol{e} - \hat{\boldsymbol{T}}\boldsymbol{e} = (0 \dots, 0, 1-p)'$.
If we let $\hat{\boldsymbol{\pi}} = (1, 0, \dots, 0)$, we have that $\hat{\tau} \sim DPH_k(\hat{\boldsymbol{\pi}}, \hat{\mathbf{T}})$ has a negative binomial distribution[1] which again is a generalization of the geometric distribution with success parameter p and number of failures k, i.e.

$$\mathbb{P}(n \text{ trials before observing } k \text{ failures}) = \mathbb{P}(\hat{\tau} = n) = \hat{\boldsymbol{\pi}}\hat{\mathbf{T}}^{n-1}\hat{\boldsymbol{t}} = \binom{n-1}{k-1} p^{n-k}(1-p)^k.$$

Knowing the underlying probabilistic interpretation, the last equality does not come as a surprise. Note that $\hat{\boldsymbol{\pi}}\hat{\mathbf{T}}^{n-1}\hat{\boldsymbol{t}}$ is the $(1,k)$th element of $\hat{\mathbf{T}}^{n-1}$ times $(1-p)$. •

[1] The formulation of the Negative Binomial adopted here is the number of trials n before k failures are observed with failure (success) having probability $1-p$ (p).

Example 3.19. The last extension connects experiments of urn type with discrete phase-type distributions. The key observation of the two previous examples has been the construction of macro states, which effectively reduces the state space. However this becomes infeasible an urn set-up because we lose the independence between draws, hence another strategy has to be used, where the state space of the underlying process is effectively expanded to reflect the specific realisation of the possible urn draws, a technique which we will return to in 4.3.1.

Consider an urn with n balls, $1 \leq d \leq n$ of which are black the rest are white. At time $1, 2, \ldots, n$ a ball is picked at random from the urn, the colour observed and set aside, i.e. draws without replacement are made. We are interested in the time of drawing the kth black ball for $1 \leq k \leq d$ which will be concentrated on $\{k, \ldots, n-d+k\}$.

This differs from the two previous examples since the probabilities of drawing white and black balls respectively changes as draws are made. The discrete phase-type framework allows us to keep track of this for us, and the difficulties now lies in describing the appropriate "macro"-state space and the transition probabilities between these. As we have seen the state space is not unique but it seems that the minimal requirement is that we keep track of how many black balls have already been picked and how far into the urn we are.
We divide the state space into k blocks, corresponding to how many black balls, $0, 1, \ldots, k-1$, have been drawn, such that the sub-transition matrix takes the form

$$\check{\boldsymbol{T}} = \begin{pmatrix} \boldsymbol{T}_{(1:n)} & \boldsymbol{T}^0_{(1:n)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_{(2:n)} & \boldsymbol{T}^0_{(2:n)} & \ldots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{T}_{(k-1:n)} & \boldsymbol{T}^0_{(k-1:n)} \\ \boldsymbol{0} & \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{T}_{(k:n)} \end{pmatrix}. \tag{3.4}$$

Where $\boldsymbol{T}_{(j:n)}$ describe draws of the urn when $j-1$ black balls have already been drawn. Since the first draw of a black ball can happen as early as the first draw and latest at draw $n-d+1$ we get a $(n-d+1) \times (n-d+1)$ sub-transition matrix on the form

$$\boldsymbol{T}_{(1:n)} = \begin{pmatrix} 0 & \frac{n-d}{n} & 0 & \cdots & 0 \\ 0 & 0 & \frac{n-d-1}{n-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \frac{1}{d+1} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

Where the last row corresponds to the case of having drawn all the white balls without drawing any black such that the next, and remaining, draws will be black. Hence, leaving the "macro" state of having observed zero black balls.
After the first black ball has been drawn, there are $d-1$ black balls left in the urn. If the black ball was drawn on the first draw there are $n-1$ balls left in the urn, while if it got drawn at the latest time possible, $n-d+1$ there remains $d-1$, all black balls, in the urn such that the next draw, at time $n-d+2$ surely will result in another black ball being drawn and another change in "macro" state. From this it should be clear that $\boldsymbol{T}^0_{(1:n)}$ and $\boldsymbol{T}_{(2:n)}$ will have the form

$$\boldsymbol{T}^0_{(1:n)} = \begin{pmatrix} \frac{d}{n} & 0 & \cdots & 0 \\ 0 & \frac{d}{n-1} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \frac{d}{d} \end{pmatrix}, \quad \boldsymbol{T}_{(2:n)} = \begin{pmatrix} 0 & \frac{n-1-(d-1)}{n-1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{n-2-(d-1)}{n-2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \frac{1}{(d-1)+1} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

And it should now be clear that the general form of $\boldsymbol{T}^0_{(j:n)}$ and $\boldsymbol{T}_{(j:n)}$, $1 \leq j \leq k$ is

$$\boldsymbol{T}^0_{(j:n)} = \begin{pmatrix} \frac{d-(j-1)}{n-(j-1)} & 0 & \cdots & 0 \\ 0 & \frac{d-(j-1)}{n-(j-1)-1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{d-(j-1)}{d-(j-1)} \end{pmatrix}$$

$$\boldsymbol{T}_{(j:n)} = \begin{pmatrix} 0 & \frac{n-(j-1)-(d-j+1)}{n-(j-1)} & 0 & \cdots & 0 \\ 0 & 0 & \frac{n-(j-1)-(d-j+1)-1}{n-(j-1)-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \frac{1}{d-(j-1)+1} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Finally, we simply initiate the Markov chain in the first state, i.e

$$\check{\boldsymbol{\pi}} = (1, 0, \dots, 0)\,.$$

Since all k diagonal blocks of $\check{\boldsymbol{T}}$ have the same dimension $n - d + 1$ the dimension of $\check{\boldsymbol{\pi}}$, and the corresponding discrete phase-type is $k(n - d + 1)$.
From this we see that the time $\check{\tau}$, of our initial question of the kth draw of a black ball is

$$\check{\tau} \sim \mathrm{DPH}\left(\check{\boldsymbol{\pi}}, \check{\boldsymbol{T}}\right).$$

Again, we will summarize our discussion as a theorem.

Theorem 3.20. *In an urn drawing setup with n balls, $1 \leq d \leq n$ of which are black. Then the time of picking the kth, $1 \leq k \leq d$ black ball is a discrete phase-type distribution, with representation given by* (3.4) *and the discussion that follows it.*

Besides of interest in itself of the distribution of the theorem, for reproducibility the function `Urn3()` in the `R`-package accompanying this thesis, takes input n, d, k and returns $\left(\check{\boldsymbol{\pi}}, \check{\boldsymbol{T}}\right)$ as a list. Below we set $n = 100, d = 10$ and plot the density of the corresponding discrete phase type for $k = 1, 3, 5, 10$ along with moments. Notice the symmetry of the two distributions with $k = 1$ and $k = 10$.

•

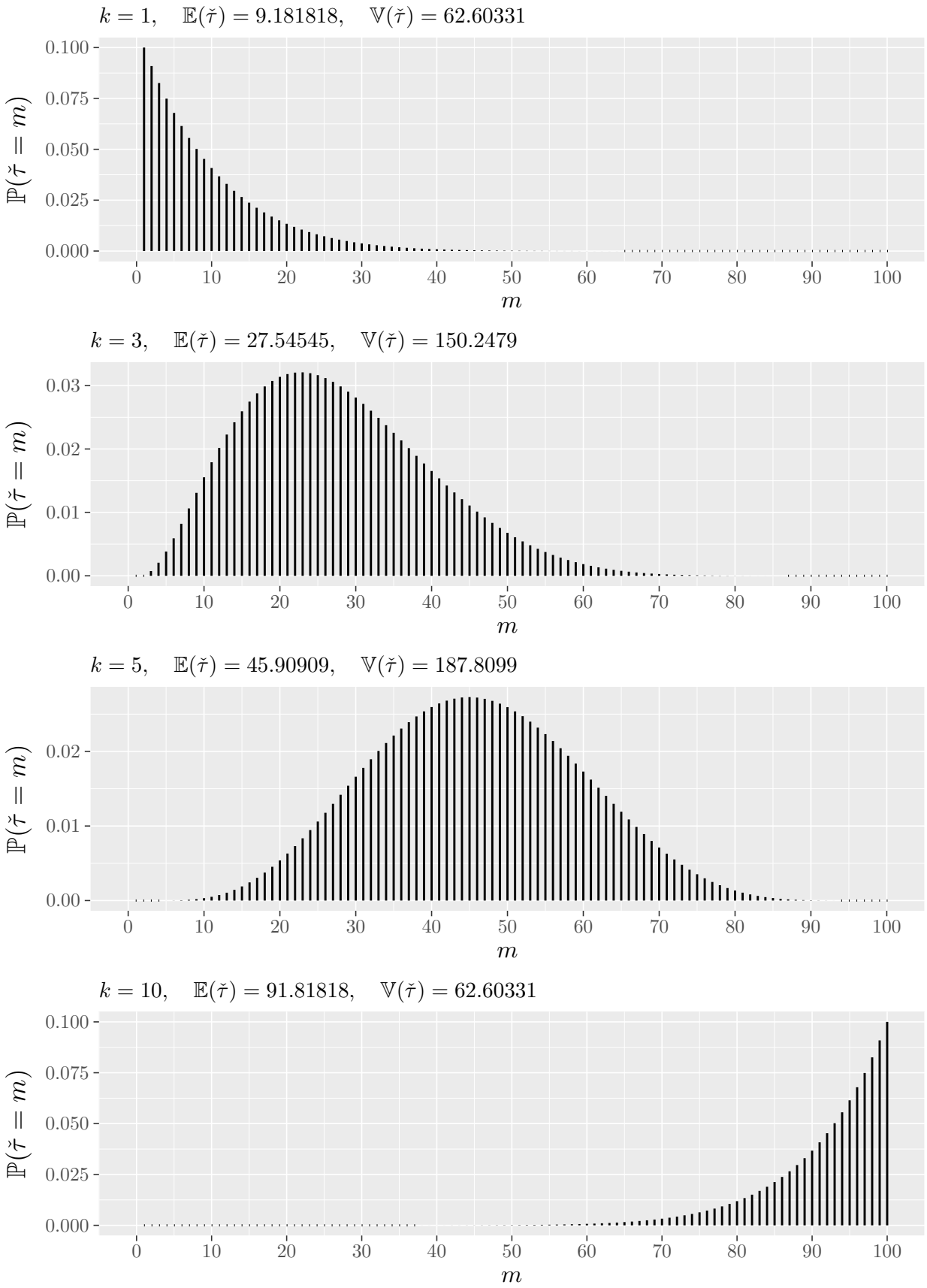

Figure 3.4: Density and moments for $\check{\tau}$ for $k = 1, 3, 5, 10$

Chapter 4

Phase Type Distributions

4.1 Construction and Examples

In this chapter we introduce and develop fundamental properties of Phase Type Distributions. A relative new area of probability being formalised in the 1970s one of the first thorough expositions of phase type distributions can be found in [Neuts, 1981], though in this thesis the exposition and notation is based on [Bladt and Nielsen, 2017].
First, we let $\{X_t\}_{t\geq 0}$ be a Markov jump process with finite state space $E = \{1, 2, ..., p, p+1\}$, ordered such that states $1, ..., p$ are transient while the last state $p+1$ is absorbing, hence the chain will eventually almost surely jump to the absorbing state. Let $(\boldsymbol{\pi}, 0)$ be the process' initial distribution where $\boldsymbol{\pi} = (\mathbb{P}(X_0 = 1), ..., \mathbb{P}(X_0 = p))$ is the initial distribution of the transient states, with $\sum_{i=1}^{p} \pi_i = 1$ we do not allow for immediate absorption, this would create an atom at zero which is most easily dealt with separately. The Markov jump process will then have initial distribution given by the vector $(\boldsymbol{\pi}, 0)$ and intensity matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{4.1}$$

where $\mathbf{T}$ is a $p \times p$ sub-intensity matrix i.e. $\boldsymbol{Te} \leq \mathbf{0}$ with strict inequality in at least one index and $\mathbf{t} = \{t_i\}_{i\in\{1,2,...,p\}}$ is called the exit rate vector since t_i is the rate by which the process jumps to the absorbing state when being in state i. Since the row sums equal 0 we have the simple but useful fact $\mathbf{t} = -\mathbf{Te}$. Notice that all the information regarding initial distribution and intensity matrix is then contained in $\boldsymbol{\pi}$ and $\boldsymbol{T}$.

Definition 4.1 (Phase-Type Distribution)**.** *Let*

$$\tau = \inf\{t > 0 \,|\, X_t = p + 1\}$$

be the time until absorption of the Markov jump process $\{X_t\}_{t\geq 0}$ as given above. Then τ is said to be Phase-Type distributed with dimension p, initial distribution $\boldsymbol{\pi}$ and sub-intensity matrix $\boldsymbol{T}$. Often we will not make the dimension explicit since it can be deducted from $\boldsymbol{\pi}$ or $\boldsymbol{T}$ and simply write

$$\tau \sim PH(\boldsymbol{\pi}, \mathbf{T}).$$

The pair $(\boldsymbol{\pi}, \boldsymbol{T})$ is called the representation of the phase-type distribution.

It is seen directly from the construction that phase-type distributions are distributions on $\mathbb{R}_+$ and with the restriction that we do not start in state $p+1$ the distribution is absolutely continuous. See Figure 4.1 for a graphical representation of the construction of the phase-type.

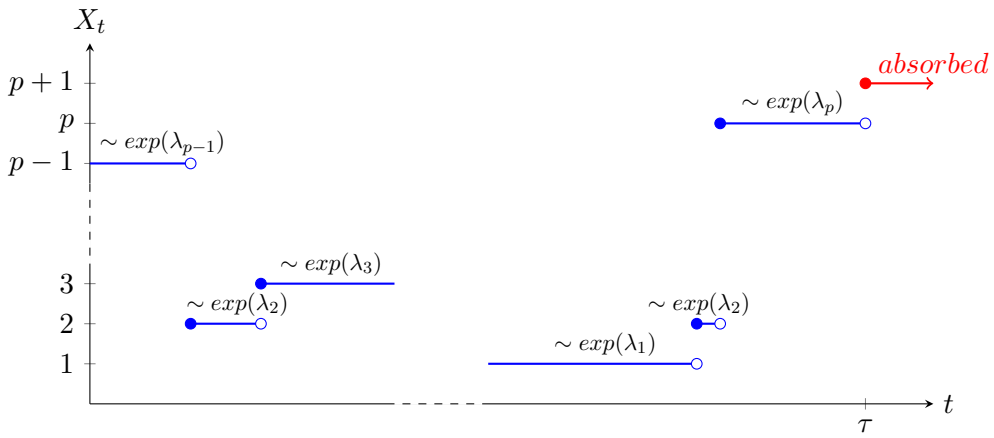

Figure 4.1: Construction of phase-type distribution. Sample paths of a Markov jump process with state space $E = \{1, 2, \dots, p, p+1\}$ with state $p+1$ absorbing. Absorption at time τ.

Lemma 4.2. *For the intensity matrix of the underlying Markov jump process* (4.1) *of a phase-type the matrix exponential is given by*

$$e^{\mathbf{\Lambda} s} = \begin{pmatrix} e^{\boldsymbol{T}s} & \boldsymbol{e} - e^{\boldsymbol{T}s}\boldsymbol{e} \\ \mathbf{0} & 1 \end{pmatrix}.$$

Proof. Since

$$\begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}^2 = \begin{pmatrix} \mathbf{T}^2 & \mathbf{T}\boldsymbol{t} \\ \mathbf{0} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{T}^2 & -\mathbf{T}^2\boldsymbol{e} \\ \mathbf{0} & 0 \end{pmatrix}$$
$$\begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}^3 = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{T}^2 & -\mathbf{T}^2\boldsymbol{e} \\ \mathbf{0} & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{T}^3 & -\mathbf{T}^3\boldsymbol{e} \\ \mathbf{0} & 0 \end{pmatrix}$$
$$\vdots$$

we see that

$$\mathbf{\Lambda}^n = \begin{pmatrix} \boldsymbol{T}^n & -\boldsymbol{T}^n\boldsymbol{e} \\ \mathbf{0} & 0 \end{pmatrix}, \; n \in \mathbb{N}.$$

Where we used that $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$. By definition we have $\mathbf{\Lambda}^0 = \boldsymbol{I}_{p+1}$. Hence

$$\begin{aligned} e^{\mathbf{\Lambda} s} &= \sum_{n=0}^{\infty} \frac{\mathbf{\Lambda}^n s^n}{n!} \\ &= \boldsymbol{I}_{p+1} + \sum_{n=1}^{\infty} \frac{\mathbf{\Lambda}^n s^n}{n!} \\ &= \boldsymbol{I}_{p+1} + \begin{pmatrix} \sum_{n=1}^{\infty} \frac{\boldsymbol{T}^n s^n}{n!} & -\sum_{n=1}^{\infty} \frac{\boldsymbol{T}^n s^n}{n!}\boldsymbol{e} \\ \mathbf{0} & 0 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_p + \sum_{n=1}^{\infty} \frac{\boldsymbol{T}^n s^n}{n!} & -\left(\sum_{n=0}^{\infty} \frac{\boldsymbol{T}^n s^n}{n!} - \boldsymbol{I}_p\right)\boldsymbol{e} \\ \mathbf{0} & 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{\boldsymbol{T}s} & \boldsymbol{e} - e^{\boldsymbol{T}s}\boldsymbol{e} \\ \mathbf{0} & 1 \end{pmatrix}. \end{aligned}$$

□

An important consequence of Lemma 4.2 is that the restriction of $\boldsymbol{P}^s$ to the transient states $\{1, 2, \dots, p\}$ is simple given by $e^{\boldsymbol{T}s}$, formally

$$\mathbb{P}(X_s = j | X_0 = i) = \mathbb{P}(X_s = j, \tau > s | X_0 = i) = \left(e^{\boldsymbol{T}s}\right)_{ij}, \; i, j \in \{1, 2, \dots, p\}, s \geq 0. \tag{4.2}$$

Which follows from $\{X_s = j\} \subseteq \{\tau > s\}$ and Corollary 2.30.

Theorem 4.3 (Density of Phase-Type)**.** *Let $\tau \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ then the density of τ is given by $f(u) = \boldsymbol{\pi} e^{\boldsymbol{T}u}\boldsymbol{t}$, $u \geq 0$.*

Proof. Conditioning on the initial state and the state prior to absorption of the underlying Markov jump process, $\{X_t\}_{t\geq 0}$, we get

$$\begin{aligned} f(u)du &= \mathbb{P}\left(\tau \in (u, u+du]\right) \\ &= \sum_{i,j\in E} \mathbb{P}\left(\tau \in (u, u+du] \,|X_u = j, X_0 = i\right) \mathbb{P}\left(X_u = j | X_0 = i\right) \mathbb{P}\left(X_0 = i\right) \\ &= \sum_{i,j\in E} \mathbb{P}\left(\tau \in (u, u+du] \,|X_u = j\right) \left(e^{\boldsymbol{T}u}\right)_{ij} \pi_i \end{aligned}$$

Now using the infinitesimal interpretation of the exit-rates in Remark 1
we have that $\mathbb{P}\left(\tau \in (u, u+du] | X_u = j\right)$ is given by $t_j du$ hence

$$\begin{aligned} f(u)du &= \sum_{i,j\in E} t_j \left(e^{\boldsymbol{T}u}\right)_{ij} \pi_i du \\ &= \boldsymbol{\pi} e^{\boldsymbol{T}u}\boldsymbol{t} du. \end{aligned}$$

Where we have used that $p_{ij}^u = e_{ij}^{\boldsymbol{T}u}$ for $i, j \in \{1, 2, \ldots, p\}$. □

For the distribution function of τ one may simply use integration of the density however, a more constructive and streamlined proof is obtained by using the underlying Markov jump process in a probabilistic approach.

Theorem 4.4 (CDF of Phase-Type)**.** *Let $\tau \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ Then the cumulative distribution function of τ is given by $F(u) = 1 - \boldsymbol{\pi} e^{\boldsymbol{T}u}\boldsymbol{e}$, $u \geq 0$.*

Proof. Using that $X_u \in \{1, \ldots, p\}$ is equivalent to $\{\tau > u\}$ and conditioning on the initial state we get

$$\begin{aligned} F_\tau(u) &= \mathbb{P}(\tau \leq u) \\ &= 1 - \mathbb{P}(\tau > u) \\ &= 1 - \sum_{j=1}^{p} \mathbb{P}(X_u = j) \\ &= 1 - \sum_{i,j=1}^{p} \mathbb{P}(X_u = j | X_0 = i)\mathbb{P}(X_0 = i) \\ &= 1 - \sum_{i,j=1}^{p} \left(e^{\boldsymbol{T}u}\right)_{ij} \pi_i \\ &= 1 - \boldsymbol{\pi} e^{\boldsymbol{T}u}\boldsymbol{e} \end{aligned}$$

□

Example 4.5 (Generalized Erlang)**.** One example of a phase-type, which will become central in Section 4.4, is the Erlang distribution which is a special case of the generalized Erlang distribution. Let $X_1, X_2, ..., X_p$ be independent exponential distributions with rate $\lambda_1, \lambda_2, \ldots, \lambda_p > 0$. Then $Y = X_1 + X_2 + \ldots + X_p$ has a generalized Erlang distribution written $Y \sim Er_p(\lambda_1, \ldots, \lambda_p)$. We can put this into the phase-type framework by letting the underlying Markov jump process pass through p phases where the sojourn time in phase i is exponential with rate λ_i. We start in phase 1 and then let the process continue to phase 2 and so forth until node p has been left where after we let the process be absorbed. See Figure 4.2.

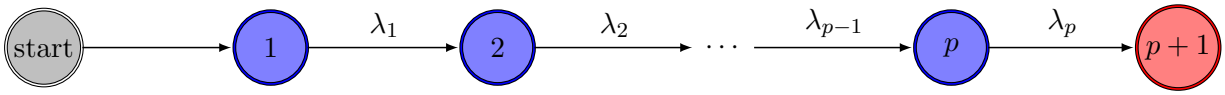

Figure 4.2: Sum of p exponentials visualized as the time to pass through a queue with p nodes. From "start" the costumer goes directly to node 1 then to 2 etc. until reaching the absorbing state $p+1$. In node i the waiting time is exponential with rate λ_i.

Now the states $1, 2, ..., p$ are transient while state $p+1$ is absorbing, hence the sub-intensity matrix of the underlying process is given by the $p \times p$ matrix

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & -\lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & -\lambda_{p-1} & \lambda_{p-1} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_p \end{pmatrix}.$$

As stated earlier we let the process start in state 1 such that the p vector of the initial distribution is

$$\boldsymbol{\pi} = (1, 0, ..., 0)\,.$$

Hence the variable Y described above is seen to be phase type with representation $(\boldsymbol{\pi}, \boldsymbol{T})$ i.e.

$$Y \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T}).$$

An important case is when $\lambda_1 = ... = \lambda_p = \lambda$. In this case we say that Y is p-dimensional Erlang, written $Y \sim Er_p(\lambda)$. Since this is the convolution of p i.i.d exponentials it has mean $\frac{p}{\lambda}$ and variance $\frac{p}{\lambda^2}$, in particular a p-dimensional Erlang with mean m has variance $\frac{m^2}{p}$ which tend to zero as the number of phases increases. This will become quintessential in Section 4.4.

•

A plethora of generalizations can be made from the Erlang distribution by changing certain aspect of how the underlying stages are traversed. We will briefly mention the generalized Erlang with feedback and the generalized Coxian distribution.

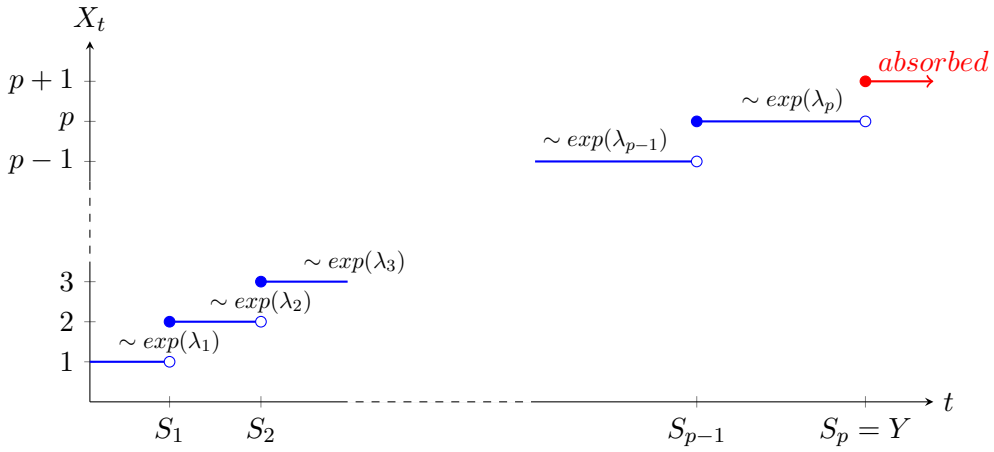

Figure 4.3: Sample path of Markov jump process underlying the generalized Erlang distribution.

Example 4.6 (Generalized Erlang with feedback)**.** A further extension of the generalized Erlang distribution discussed in example 4.5 is made by introducing a feedback parameter $\rho \in [0, 1)$ such that when the final pth stage is reach in the underlying process, return to the initial stage happens with probability ρ. We will denote the above distribution as $Er_p^\rho(\lambda_1, \ldots, \lambda_p)$ or simply $Er_p^\rho(\lambda)$ if all stages have the same intensity. We then have the transition diagram given in Figure 4.4.

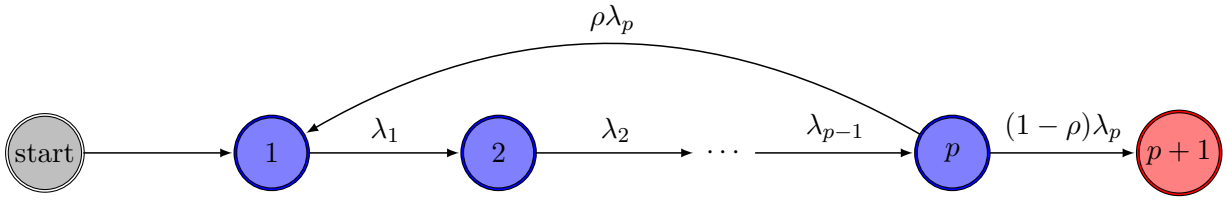

Figure 4.4: Generalized Erlang with feedback, flowdiagram.

The sub-intensity matrix of the underlying Markov jump process is then given by the $p \times p$ matrix

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & -\lambda_{p-1} & \lambda_{p-1} \\ \rho\lambda_p & 0 & 0 & \cdots & 0 & -\lambda_p \end{pmatrix}.$$

Again the initial distribution is $\boldsymbol{\pi} = (1, 0, \ldots, 0)$. Interestingly the simple construction above leads to a wave-phenomena in the density, due to the possibility of returning to the first stage. To illustrate the richness of the densities, in Figure 4.5 we plot the density of $Er_k^\rho(\lambda)$ changing the number of phases k, intensity λ, and feedback parameter ρ, respectively. For implementation, the function `erlang_ph()` returns $\boldsymbol{\pi}, \boldsymbol{T}$ underlying a (generalized) Erlang distribution with or without feedback.

•

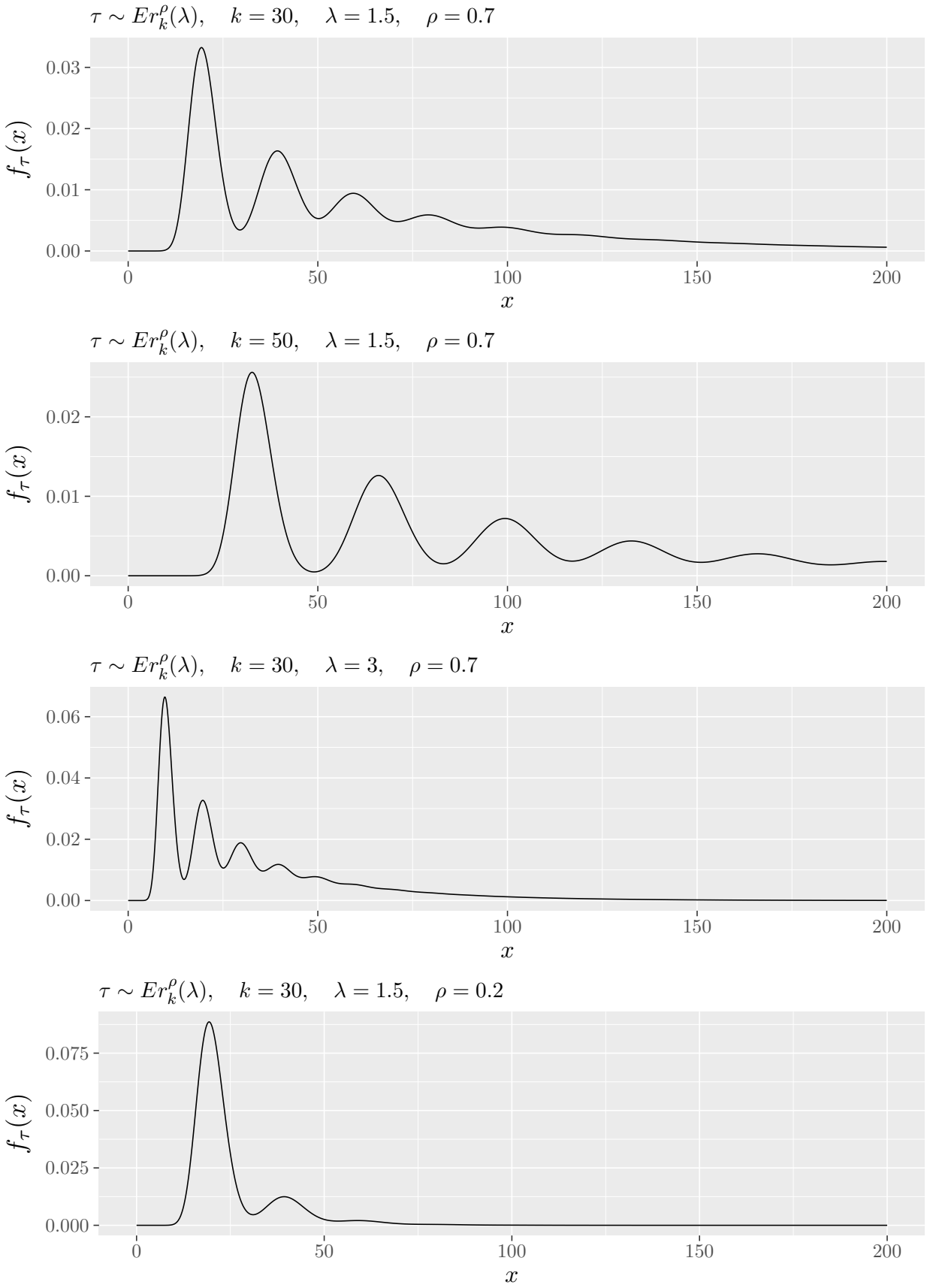

Figure 4.5: Erlang with feedback for different values number of phases, intensity, and feedback parameter.

Example 4.7 (Coxian distribution)**.** In the general setup of the k-stage Erlang distribution, the Coxian structure arises from allowing the process to transit directly to the absorbing state $p+1$ from any of the transient states i with probability g_i. We then have the transition diagram given in Figure 4.6.

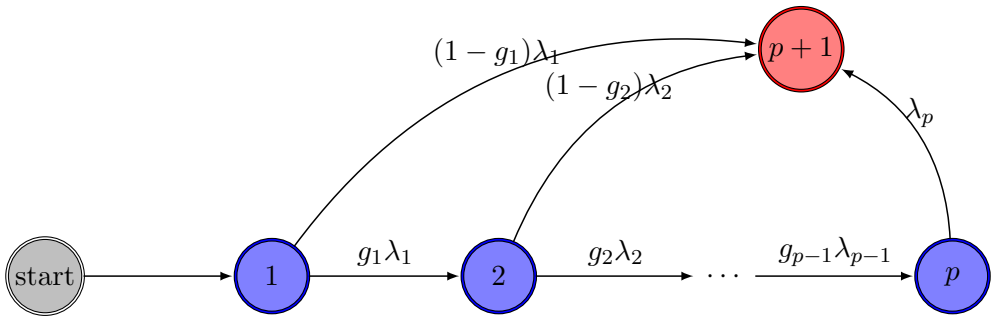

Figure 4.6: Coxian distribution, flowdiagram.

The sub-intensity matrix of the underlying Markov jump process is then given by the $p \times p$ matrix

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & g_1\lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_2 & g_2\lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & -\lambda_{p-1} & g_{p-1}\lambda_{p-1} \\ 0 & 0 & 0 & \cdots & 0 & -\lambda_p \end{pmatrix}$$

and initial distribution $\boldsymbol{\pi} = (1, 0, \ldots, 0)$. •

In Figure 4.3 we see a possible sample path of the Markov jump process underlying the $\mathrm{Er}_p(\lambda_1, \ldots, \lambda_p)$ distribution described above. Just as the order of summation is of no importance, the order that we pass through the states of the underlying Markov jump process will not change the absorption time and consequently the phase-type. That is, a re-indexing of the state space or equivalently symmetric permutations of the rows and columns of the sub-intensity matrix $\boldsymbol{T}$ and corresponding changes applied to $\boldsymbol{\pi}$, will not change the phase-type distribution from which we make the important observation that the representation, $(\boldsymbol{\pi}, \boldsymbol{T})$ of a phase-type is generally not unique, additionally a phase-type distribution can be greatly over-parametrized as we also saw for the discrete case.

Theorem 4.8 (Over parametrizations)**.** *Let $\tau \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$. If $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e} = \lambda \boldsymbol{e}$ for some $\lambda > 0$ then $\tau \sim exp(\lambda)$.*

Proof. Recall that $F(x) = 1 - \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e}$ such that the survival function is $S(x) = 1 - F(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e}$. The density of τ is given by

$$\begin{aligned} f(x) &= \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t} \\ &= \lambda \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e} \\ &= \lambda S(x). \end{aligned}$$

Now the survival function satisfies $-S'(x) = f(x)$ which leads to the differential equation

$$S'(x) = -\lambda S(x), \quad S(0) = 1$$

this has solution $S(x) = e^{-\lambda x}$ which is seen to be the survival function of the exponential distribution with rate λ hence $\tau \sim exp(\lambda)$. □

From the observations of the previous theorem the following definition is made.

Definition 4.9. *For $PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ then p is called the dimension of the representation while the lowest dimension of the phase-type distribution is called the order.*

Theorem 4.8 is an example of a phase-type distribution with arbitrary high dimension $p \geq 1$ while its order is 1.

Theorem 4.10. *Let $\{X_t\}_{t\geq 0}$ be a Markov jump process on state space $E = \{1, 2, \ldots, p, p+1\}$ and intensity matrix*

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 0 \end{pmatrix}$$

then $\boldsymbol{T}$ is invertible if and only if states $1, 2, \ldots, p$ are transient.

Proof. For the "if" part, assume that states $1, 2, \ldots, p$ are transient. An equivalent condition for $\boldsymbol{T}$ being invertible is that the equation $\boldsymbol{vT} = \boldsymbol{0}$ only has the trivial solution $\boldsymbol{v} = \boldsymbol{0}$. Now, $\boldsymbol{vT} = \boldsymbol{0}$ is equivalent to $v_i(-t_{ii}) = \sum_{j\neq i} v_j t_{ji}$ for $i = 1, 2, \ldots, p$.
By definition $-t_{ii} = \lambda_i$ and $t_{ji} = \lambda_j q_{ji}$ where q_{ji} is the transition probability of the embedded chain going from state j to i with $q_{ii} = 0$. Thus, the above equalities can be written

$$v_i\lambda_i = \sum_{j\in E\setminus\{i,p+1\}} v_j\lambda_j q_{ji}$$

Since all states $1, 2, \ldots, p$ are transient they have $\lambda_i > 0$, let $u_i = v_i\lambda_i$ then the equation can be written

$$\boldsymbol{u} = \boldsymbol{uQ}$$

where $\boldsymbol{Q} = \{q_{ij}\}_{i,j\in\{1,2,\ldots,p\}}$ and iterrating we have $\boldsymbol{u} = \boldsymbol{uQ}^n$ for all $n \in \mathbb{N}$. Recall that $\{\boldsymbol{Q}\}_{ij} = \mathbb{P}(Y_n = j | Y_0 = i)$ where Y_n is the embedded process, then since states $1, 2, \ldots, p$ are transient we have $\lim_{n\to\infty} \boldsymbol{Q}^n = \boldsymbol{0}$. Thus

$$\boldsymbol{u} = \boldsymbol{u} \lim_{n\to\infty} \boldsymbol{Q}^n = \boldsymbol{0},$$

and since $\lambda_i > 0$ we must have $v_i = 0$ for all $i = 1, 2, \ldots, p$ which shows that the only solution to $\boldsymbol{vT} = \boldsymbol{0}$ is given by $\boldsymbol{v} = 0$, and therefore $\boldsymbol{T}$ is invertible.

For the "only if" part, assume that $\boldsymbol{T}$ is invertible. Let a_i be the probability of eventual absorption into state $p+1$ given initiation in state $i = 1, 2, \ldots, p$. Then conditioning on the next jump we have the equality

$$a_i = q_{i,p+1} + \sum_{j\neq i} q_{ij} a_j.$$

Since $q_{ij} = \frac{t_{ij}}{-t_{ii}}$ for $i \neq j$ we get

$$-t_{ii}a_i = t_{i,p+1} + \sum_{j\neq i} t_{ij} a_j$$

or written in matrix notation

$$\boldsymbol{t} + \boldsymbol{Ta} = \boldsymbol{0}.$$

Since $\boldsymbol{T}$ is invertible by assumption we can solve for $\boldsymbol{a}$ to get

$$\boldsymbol{a} = -\boldsymbol{T}^{-1}\boldsymbol{T} = -\boldsymbol{T}^{-1}(-\boldsymbol{Te}) = \boldsymbol{e}.$$

Hence $a_i = 1$ for all $i = 1, 2 \ldots, p$, showing that states $1, 2, \ldots, p$ are transient since eventual absorption into state $p+1$, which happens with probability 1, will mean not returning to any of the states $1, 2, \ldots, p$. □

As in the discrete case, for all phase-type representations the sub-intensity matrix $\boldsymbol{T}$ is invertible.

Corollary 4.11. *Let $(\boldsymbol{\pi}, \boldsymbol{T})$ be a phase-type representation, then $\boldsymbol{T}$ is invertible.*

As in the discrete case the continuous phase type also has an assosiated *green matrix* with an interesting property.

Definition 4.12. *For a phase-type distribution PH$(\boldsymbol{\pi}, \boldsymbol{T})$ then $\boldsymbol{U} = (-\boldsymbol{T})^{-1}$ is called the distributions associated Green matrix.*

Theorem 4.13. *For the associated green matrix for a phase-type distribution PH$(\boldsymbol{\pi}, \boldsymbol{T})$ given by $\boldsymbol{U} = \{u_{ij}\}_{i,j=1,2,\ldots,p}$, then u_{ij} is the expected time spent by the underlying Markov jump process in state j conditioned on initiation in state i.*

Proof. Let Z_j denote the time spend in state j by prior to absorption of the underlying Markov jump process.

$$\begin{aligned}
\mathbb{E}_i(Z_j) &= \mathbb{E}_i\left(\int_0^\tau 1\{X_t = j\}dt\right) \\
&= \mathbb{E}_i\left(\int_0^\infty \mathbb{P}_i(X_t = j, \tau > t)\right) dt \\
&= \int_0^\infty (e^{\boldsymbol{T}s})_{ij}\, ds \quad (4.2) \\
&= \left[\left(\boldsymbol{T}^{-1}e^{\boldsymbol{T}s}\right)_{ij}\right]_0^\infty \\
&= \left(-\boldsymbol{T}^{-1}\right)_{ij} \qquad (e^{\boldsymbol{T}s} \to \boldsymbol{0} \text{ for } s \to \infty)
\end{aligned}$$

and we see that $\mathbb{E}_i(Z_j) = u_{ij}$ for all $i, j = 1, 2, \ldots, p$. □

Corollary 4.14. *All eigenvalues for a sub-intensity matrix $\boldsymbol{T}$ of a phase-type distribution have strictly negative real parts.*

Proof. Let $\theta = \max_i(-t_{ii}) > 0$ i.e. the largest exit rate of the underlying states. Define the matrix $\boldsymbol{K}$ by

$$\boldsymbol{T} = \theta(\boldsymbol{K} - \boldsymbol{I}) \iff \boldsymbol{K} = \frac{1}{\theta}\boldsymbol{T} + \boldsymbol{I}$$

Then $\boldsymbol{K}$ is seen to be a sub-transition matrix of a discrete phase-type distribution. To see this we first observe, by virtue of the definition of the intensity matrix of a Markov jump process, that for $i = 1, 2, \ldots, p$ we have

$$\begin{aligned}
&0 > t_{ii} \geq -\theta \quad \iff \\
&0 > \frac{1}{\theta}t_{ii} \geq -1 \quad \iff \\
&1 > \frac{1}{\theta}t_{ii} + 1 \geq 0 \quad \iff \\
&1 > k_{ii} \geq 0
\end{aligned}$$

Hence the diagonal elements of $\boldsymbol{K}$ are all contained in the interval $[0, 1)$ in particular none of the states $1, 2, \ldots, p$ are absorbing, which would contradict $\boldsymbol{K}$ being a sub-transition matrix of a discrete phase-type. Now for $i, j = 1, 2, \ldots, p, i \neq j$

$$\begin{aligned}
&0 \leq t_{ij} \leq -t_{ii} \leq \theta \quad \iff \\
&0 \leq \frac{1}{\theta}t_{ij} \leq 1 \quad \iff \\
&0 \leq k_{ij} \leq 1.
\end{aligned}$$

Lastly, the row sums of $\boldsymbol{K}$ are seen to satisfy

$$\boldsymbol{K}\boldsymbol{e} = \left(\frac{1}{\theta}\boldsymbol{T} + \boldsymbol{I}\right)\boldsymbol{e} = \underbrace{\frac{1}{\theta}\boldsymbol{T}\boldsymbol{e}}_{\leq \boldsymbol{0}} + \boldsymbol{e} \leq \boldsymbol{e}.$$

Since $\boldsymbol{T}\boldsymbol{e} \leq \boldsymbol{0}$ holds with strict inequality in at least one entry, we also have a strict inequality above in at least one entry. Since $\boldsymbol{K}$ is a sub-transition matrix for a discrete phase-type distribution by Corollary 3.9, all eigenvalues of $\boldsymbol{K}$ are contained strictly within the unit circle. Now let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{T}$ with corresponding eigenvalue λ then $\boldsymbol{v}$ is also an eigenvector for $\boldsymbol{K}$ with corresponding eigenvalue μ since

$$\boldsymbol{K}\boldsymbol{v} = \left(\frac{1}{\theta}\boldsymbol{T} + \boldsymbol{I}\right)\boldsymbol{v} = \frac{1}{\theta}\lambda\boldsymbol{v} + \boldsymbol{v} = \underbrace{\left(\frac{1}{\theta}\lambda + 1\right)}_{\mu}\boldsymbol{v}$$

then $\lambda = \theta(\mu - 1)$ whose real part is then strictly less that 0. □

Recall that the Laplace transform of a non-negative random variable X is defined by $\mathcal{L}_X(s) = \mathbb{E}\left(e^{-sX}\right)$ which exists at least for s such that $Re(s) \geq 0$. For $s \in \mathbb{R}_+$ we have the interpretation of the Laplace transform $\mathcal{L}_X(s)$ as the probability that $X < Y$ where $Y \sim exp(s)$. Since

$$\mathbb{P}(X < Y) = \mathbb{E}\left(\mathbb{P}(Y > X|X)\right) = \mathbb{E}(e^{-sX}) = \mathcal{L}_X(s).$$

Theorem 4.15. *The Laplace transform of $\tau \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

$$\mathcal{L}_\tau(s) = \mathbb{E}\left(e^{-s\tau}\right) = \boldsymbol{\pi}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t} = \boldsymbol{\pi}\left(s\boldsymbol{U} + \boldsymbol{I}\right)^{-1}\boldsymbol{e}. \tag{4.3}$$

which is well defined for all s such that $Re(s) > Re(\lambda_{max})$, where λ_{max} is the eigenvalue of $\boldsymbol{T}$ with largest real part. In particular $\mathcal{L}_\tau(s)$ is well defined for $s \in \mathbb{R}_+$

Proof. By definition, if λ is an eigenvalue of $\boldsymbol{T}$ with corresponding eigenvector $\boldsymbol{v}$ then

$$(s\boldsymbol{I} - \boldsymbol{T})\boldsymbol{v} = s\boldsymbol{v} - \lambda\boldsymbol{v} = (s - \lambda)\boldsymbol{v}$$

i.e. $(s - \lambda)$ is an eigenvalue for $(s\boldsymbol{I} - \boldsymbol{T})$.
Let λ_{max} be the eigenvalue of $\boldsymbol{T}$ with largest real part, then we know from corollary 4.14 that $Re(\lambda_{max}) < 0$. Consequently $(s - \lambda_{max})$ is the eigenvalue of $(s\boldsymbol{I} - \boldsymbol{T})$ with smallest real part and if $Re(s) > Re(\lambda_{max})$ then all eigenvalues of $(s\boldsymbol{I} - \boldsymbol{T})$ will have strictly positive real part, in particular every eigenvalue is nonzero which means that $(s\boldsymbol{I} - \boldsymbol{T})$ is invertible. Then for s such that $Re(s) > Re(\lambda_{max})$ we have

$$\begin{aligned}
\mathcal{L}_\tau(s) &= \mathbb{E}(e^{-s\tau}) \\
&= \int_0^\infty e^{-sx}\boldsymbol{\pi}e^{\boldsymbol{T}x}\boldsymbol{t}\,dx \\
&= \int_0^\infty \boldsymbol{\pi}e^{-sx\boldsymbol{I}}e^{\boldsymbol{T}x}\boldsymbol{t}\,dx \quad (a\boldsymbol{T} = a\boldsymbol{I}\boldsymbol{T} \text{ and } (2.12)) \\
&= \boldsymbol{\pi}\int_0^\infty e^{-(s\boldsymbol{I}-\boldsymbol{T})x}dx\,\boldsymbol{t} \quad (\textit{Theorem } 2.29) \\
&= \boldsymbol{\pi}(s\boldsymbol{I} - \boldsymbol{T})^{-1}\boldsymbol{t}. \quad ((\boldsymbol{T} - s\boldsymbol{I}) \textit{ is a sub-intensity Matrix and Eq } (2.11))
\end{aligned}$$

□

Notice that the Laplace transform of the phase-type is a rational function in s, since $(s\boldsymbol{I}-\boldsymbol{T})$ can be inverted by elementary operations.

An important discussion on the theory of phase-type distributions and general non-negative distributions is connected to the analytical form of the Laplace transform. Recall the density of a phase-type distribution.

$$f(x) = \boldsymbol{\alpha} e^{\boldsymbol{T}x}\boldsymbol{t}, \quad x > 0.$$

One may disregard the proabapilitic underlining in the above density and simply regard p-dimensional vectors $\boldsymbol{\alpha}$ and $\boldsymbol{t}$ and $p \times p$ matrices $\boldsymbol{T}$ such that the above defines a density, possibly with an atom at zero of size $\int_0^\infty \boldsymbol{\alpha} e^{\boldsymbol{T}x}\boldsymbol{t}\, dx$. If a random variable X has above density we say that the X has a Matrix-Exponential distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$, where the vector $\boldsymbol{\alpha}$ is called the starting vector, $\boldsymbol{s}$ the closing vector and the matrix $\boldsymbol{S}$ the generator [Buchholz et al., 2014] and [Bladt and Nielsen, 2017].
Consequently, Matrix-Exponential distributions have rational Laplace transform and it can be shown that the reverse implication also holds. This is not the case for phase-types as seen by e.g. the distribution given by the density $f(x) = 2e^{-x}(1 - cos(x))$ which has rational Laplace transform, but since $f(2\pi) = 0$, it is seen that it is not of phase-type. Hence the phase type is a true subset of Matrix-Exponential distributions/Distributions having a rational Laplace transform. Although of interest in itself we will not pursue the study of Matrix-Exponential distributions here.

The non-central moment of any order of a phase-type distribution can be derived by the Laplace transform using the fact that $\mathbb{E}[\tau^n] = (-1)^n \frac{d^n}{ds^n}\mathcal{L}_\tau(s)\big|_{s=0}$ see e.g. [Amussen, 2003].
But to demonstrate the probabilistic argument as opposed to the more analytical method using the underlying process, which will be used extensively later, we will give another proof using a construction based on the underlying Markov jump process. The proof is based on [Bladt and Nielsen, 2017].

Consider $\mathbb{R}_+^n$. From origo an n-dimensional cube is generated with side length equal to $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$, see Figure 4.7 for the case $n = 2$. That is, at time t if X_t has not been absorbed, the cube will have volume t^n and the increment in volume during the infinitesimal time interval $[t, t+dt)$ is $nt^{n-1}dt$. At absorption the total volume of the cube will be τ^n with expected volume $\mathbb{E}(\tau^n)$.

Now let $v_{ij}^{(n)}$ be the expected volume of the n-dimensional cube generated by the underlying Markov jump process while in state j given $X_0 = i$ and let $\boldsymbol{V}^{(n)} = \{v_{ij}^{(n)}\}_{i,j\in\{1,2,\ldots,p\}}$.

We then have

$$\begin{aligned} v_{ij}^{(n)} &= \mathbb{E}_i\left(\int_0^\tau nt^{n-1} 1\{X_t = j\}dt\right) \\ &= \int_0^\infty nt^{n-1}\mathbb{P}_i(X_t = j, \tau > t)dt \\ &= \int_0^\infty nt^{n-1}\left(e^{\boldsymbol{T}t}\right)_{ij} dt \end{aligned}$$

hence

$$\boldsymbol{V}^{(n)} = \int_0^\infty nt^{n-1} e^{\boldsymbol{T}t}\, dt$$

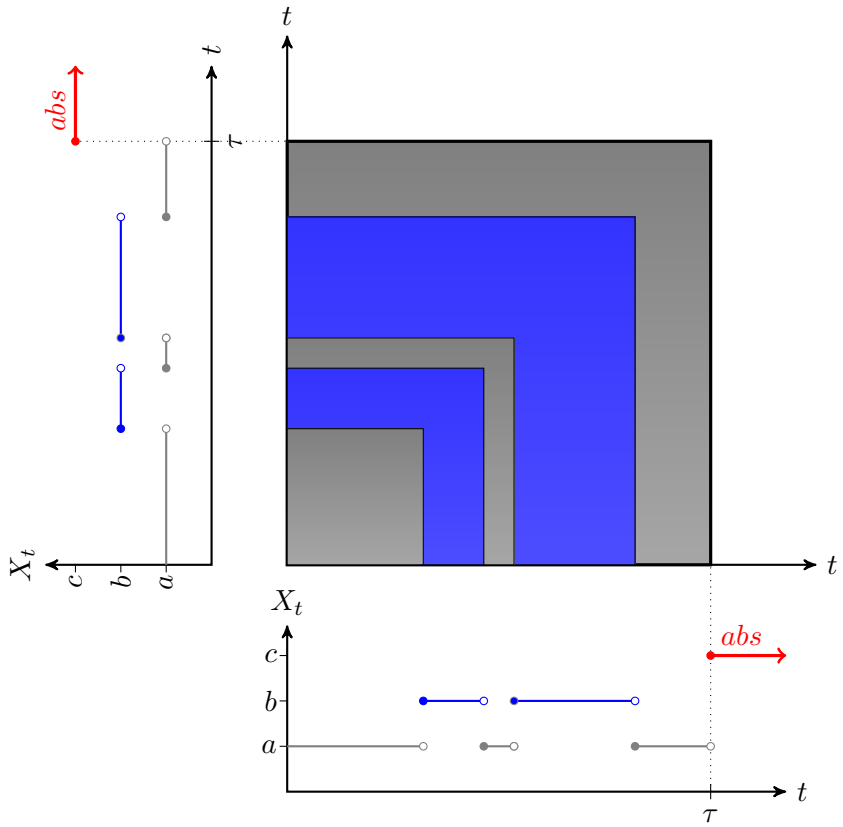

Figure 4.7: A Markov jump process underlying a phase-type distribution with states $\{a, b, c\}$, where the last state is absorbing. A square with side length τ is generated. When X_t is in state a the gray area is generated. The blue area is generated while the underlying Markov jump process is in state b. At absorption τ no more area is generated.

applying integration by parts $n-1$ times we get

$$\begin{aligned}
\boldsymbol{V}^{(n)} &= \underbrace{\left[nt^{n-1}\boldsymbol{T}^{-1}e^{\boldsymbol{T}t}\right]_0^\infty}_{=0} - \int_0^\infty n(n-1)t^{n-2}\boldsymbol{T}^{-1}e^{\boldsymbol{T}t}\,dt \\
&= \ldots = (-1)^{n-1}n!\boldsymbol{T}^{-(n-1)}\int_0^\infty e^{\boldsymbol{T}t}dt \\
&= (-1)^n n!\boldsymbol{T}^{-n} \\
&= n!(-\boldsymbol{T}^{-1})^n \\
&= n!\boldsymbol{U}^n.
\end{aligned}$$

From the above discussion we have proved the following theorem.

Theorem 4.16. *The expected volume of the n-dimensional cube with side length τ generated while the underlying Markov jump process being in state j given initiation in state i is equal to the ijth element of $n!\boldsymbol{U}^n$.*

Corollary 4.17. *The nth moment of $\tau \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

$$\mathbb{E}(\tau^n) = n!\boldsymbol{\pi}(-\boldsymbol{T}^{-1})^n\boldsymbol{e}. \tag{4.4}$$

In particular

$$\mathbb{E}(\tau) = \boldsymbol{\pi}(-\boldsymbol{T}^{-1})\boldsymbol{e} = \boldsymbol{\pi}\boldsymbol{U}\boldsymbol{e} \tag{4.5}$$

$$Var(\tau) = 2\boldsymbol{\pi}(\boldsymbol{T}^{-2})\boldsymbol{e} - \left(\boldsymbol{\pi}(-\boldsymbol{T}^{-1})\boldsymbol{e}\right)^2 \tag{4.6}$$

Proof. From the interpretation of the matrix $n!(-\boldsymbol{T}^{-1})^n$ from Theorem 4.16, conditioning on the initial state the vector $n!\boldsymbol{\pi}(-\boldsymbol{T}^{-1})^n$ is the expected volume generated while the underlying Markov jump process is in states $1, 2, \ldots, p$. Then summing the entries we get the expected total volume i.e. $n!\boldsymbol{\pi}(-\boldsymbol{T}^{-1})^n\boldsymbol{e}$. □

In general it is superfluous to define the phase-type using more than one absorbing state since, as we shall see, it produces the regular phase-type but with formulae being unnecessarily complicated. Hence for the sake of distributional properties it is not useful. However, in the case that we focus on which absorbing state is reached it can be useful for developing the theory, as we will see in Section 4.3.2, hence we go through basic definitions and properties when introducing multiple absorbing states.

4.2 Multiple Absorbing States

The definition of the phase-type easily lends itself to the extension of multiple absorbing states. Formally, the underlying Markov jump process now has state space

$$E = \{1, 2, \ldots, p, p+1, \ldots, p+m\}$$

for some $m \geq 1$. As earlier, the first p states are transient but now states $\{p+1, \ldots, p+m\}$ are absorbing states. We will denote the transient states $S^{trans} = \{1, \ldots, p\}$ while the absorbing states will be called $S^{abs} = \{p+1, \ldots, p+m\}$. The quantity which we will be interested in is, analogously to the standard phase-type, given by the time until absorption in any of the absorbing states of the Markov jump process

$$\tau = \inf\{t > 0 | \, X_t \in S^{abs}\}.$$

Consequently the intensity matrix of the Markov jump process now has block form

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}. \tag{4.7}$$

$\boldsymbol{T}$ retains its usual interpretation from earlier sections, as the intensity matrix of the Markov jump process prior to absorption, but now the lower left zero-matrix has dimension $m \times p$ and the lower right zero-matrix has dimension $m \times m$. The $p \times m$ matrix $\boldsymbol{D}$ is the exit-matrix, $\boldsymbol{D} = \{d_{ij}\}_{i \in S^{trans}, j \in S^{abs}}$ with the usual interpretation that d_{ij} is the rate of leaving for the absorbing state j given the Markov jump process is in state i in particular, given $X_u = i$ then the probability of absorption to state j during $[u, u+du)$ is $d_{ij}du$.
Given that $X_{\tau-} = i$ (left limit) the probability that $X_\tau = j$ is given by $\frac{d_{ij}}{\sum_{k=p+1}^{p+m} d_{ik}}$.
The initial vector is now a $p+m$-vector, and as in the one absorbing state case, we will assume that initiation in absorbing states does not happen since this will create an atom at zero which is easiest dealt with separately. Hence the initial distribution of the underlying Markov jump process has the form

$$(\mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), \ldots, \mathbb{P}(X_0 = p), 0, \ldots, 0) = (\boldsymbol{\pi}, 0, \ldots, 0)$$

with $\pi_i = \mathbb{P}(X_0 = i)$, $i = 1, \ldots, p$.
Since the rows of $\boldsymbol{\Lambda}$ must sum to zero we have the relation $\boldsymbol{De} = -\boldsymbol{Te}$.

The new construction still yields a distribution of phase-type. Quantities of interest like distribution function, density and Laplace transform are very similar to the standard case. We will go through the first two but not state them as theorems.
The distribution function is given by

$$F(u) = \mathbb{P}(\tau \leq u) = \mathbb{P}(X_u \in S^{abs}) = 1 - \mathbb{P}(X_u \in S^{trans}).$$

Conditioning on the initial state we get

$$\begin{aligned}
1 - F(u) &= \mathbb{P}(X_u \in S^{trans}) \\
&= \sum_{i=1}^{p} \mathbb{P}(X_u \in S^{trans} | X_0 = i)\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \mathbb{P}(X_u = k | X_0 = i)\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \pi_i \left(e^{\boldsymbol{T}u}\right)_{ik} \\
&= \boldsymbol{\pi} e^{\boldsymbol{T}u} \boldsymbol{e}.
\end{aligned}$$

Which is exactly the same as the case of one absorbing state. Hence the construction yields the standard phase-type.
The density, on the other hand, is not identical to the formula of Theorem 4.3, since now the exit rate vector is a matrix, but it is on the form expected where the total exit rates for each transient state is used.

$$\begin{aligned}
f(u)du &= \mathbb{P}(\tau \in (u, u + du]) \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \mathbb{P}(\tau \in (u, u + du] | X_u = k, X_0 = i)\mathbb{P}(X_u = k | X_0 = i)\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{p}\sum_{k=1}^{p} \left(\sum_{j=1}^{m} d_{kj}\right) du \left(e^{\boldsymbol{T}u}\right)_{ik} \pi_i \\
&= \boldsymbol{\pi} e^{\boldsymbol{T}u} \boldsymbol{D}\boldsymbol{e}\, du.
\end{aligned}$$

Hence the density is given by $f(u) = \boldsymbol{\pi} e^{\boldsymbol{T}u}\boldsymbol{D}\boldsymbol{e}$ which can be seen as the regular case but using the rowsums of $\boldsymbol{D}$ as exit rate vector.

The nth power of the intensity matrix (4.7) are given by $\boldsymbol{\Lambda}^n = \begin{pmatrix} \boldsymbol{T}^n & \boldsymbol{T}^{n-1}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$. One can convince oneself of this by considering the first few powers

$$\begin{aligned}
\begin{pmatrix} \boldsymbol{T} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} &= \boldsymbol{\Lambda} \\
\begin{pmatrix} \boldsymbol{T} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}^2 &= \begin{pmatrix} \boldsymbol{T}^2 & \boldsymbol{T}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\
\begin{pmatrix} \boldsymbol{T} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}^3 &= \begin{pmatrix} \boldsymbol{T} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{T}^2 & \boldsymbol{T}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{T}^3 & \boldsymbol{T}^2\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}
\end{aligned}$$

following the lines of Lemma 4.2 the matrix exponential of 4.7 is found. Recall that by definition

$\boldsymbol{\Lambda}^0 = \boldsymbol{I}_{p+m}$. We then have

$$\begin{aligned} e^{\boldsymbol{\Lambda}} &= \sum_{n=0}^{\infty} \frac{\boldsymbol{\Lambda}^n}{n!} \\ &= \boldsymbol{I}_{p+m} + \sum_{n=1}^{\infty} \frac{\boldsymbol{\Lambda}^n}{n!} \\ &= \boldsymbol{I}_{p+m} + \sum_{n=1}^{\infty} \begin{pmatrix} \boldsymbol{T}^n & \boldsymbol{T}^{n-1}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} / n! \\ &= \boldsymbol{I}_{p+m} + \begin{pmatrix} \sum_{n=1}^{\infty} \frac{\boldsymbol{T}^n}{n!} & \sum_{n=1}^{\infty} \frac{\boldsymbol{T}^{n-1}}{n!}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\ &= \begin{pmatrix} e^{\boldsymbol{T}} & \left(e^{\boldsymbol{T}} - \boldsymbol{I}\right)\boldsymbol{T}^{-1}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix} \end{aligned}$$

Hence we have the transition matrix for the multiple absorbing states case

$$\boldsymbol{P}^t = \begin{pmatrix} e^{\boldsymbol{T}t} & \left(e^{\boldsymbol{T}t} - \boldsymbol{I}\right)\boldsymbol{T}^{-1}\boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix} \tag{4.8}$$

Where we note that prior to absorption the transitions are given by the restriction of $\boldsymbol{P}^t$ to the transient states i.e. $\boldsymbol{P}^t|_{\{1,2,\ldots,p\}} = e^{\boldsymbol{T}t}$. Likewise that probability of being in absorbing state $p+j$ at time t given initial state i is given by $\left(\left(e^{\boldsymbol{T}t} - \boldsymbol{I}\right)\boldsymbol{T}^{-1}\boldsymbol{D}\right)_{ij}$.

Absorption Probabilities

An important quantity which will become useful to us is section 4.3 is the absorption probabilities for the absorbing states

$$\boldsymbol{\phi} = \{\phi_j\}_{j=1,\ldots,m} = (\mathbb{P}(X_\tau = p+1), \ldots, \mathbb{P}(X_\tau = p+m))$$

i.e. ϕ_j is the probability that the underlying chain is absorbed in the jth absorbing state. From (4.8) we see that for $i \in S^{trans}$ and $j \in S^{abs}$ we have $\boldsymbol{P}^t_{ij} = \left[\left(e^{\boldsymbol{T}t} - \boldsymbol{I}\right)\boldsymbol{T}^{-1}\boldsymbol{D}\right]_{ij}$.

Assuming $X_0 \in S^{trans}$ with probability 1, we can write

$$\begin{aligned} \phi_j &= \sum_{i=1}^{p} \pi_i \lim_{t\to\infty} \boldsymbol{P}^t_{ij} \\ &= \sum_{i=1}^{p} \pi_i \lim_{t\to\infty} \left[\left(e^{\boldsymbol{T}t} - \boldsymbol{I}\right)\boldsymbol{T}^{-1}\boldsymbol{D}\right]_{ij} \\ &= \sum_{i=1}^{p} \pi_i \left[-\boldsymbol{T}^{-1}\boldsymbol{D}\right]_{ij} \end{aligned}$$

since $\lim_{t\to\infty} e^{\boldsymbol{T}t} = \boldsymbol{0}$ since the states $1, 2, \ldots, p$ are transient.
We therefore have

$$\phi_j = \sum_{i=1}^{p} \pi_i \left[-\boldsymbol{T}^{-1}\boldsymbol{D}\right]_{ij}$$

or written in vector form

$$\boldsymbol{\phi} = \boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{D}. \tag{4.9}$$

We have now proved the useful result

Theorem 4.18 (Absorption probabilities for multiple absorbing states.)**.** *In the case of $m \geq 2$ absorbing states of the Markov jump process underlying a phase-type distribution, if the state space is ordered such that the first p states are transient and the remaining m states absorbing, then the m absorption probabilities are given by*

$$\boldsymbol{\phi} = \boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{D}.$$

Note that the result also makes sense for $m = 1$ since in that case $\boldsymbol{D} = \boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ such that $\boldsymbol{\phi} = 1$. Also, since $\boldsymbol{D}\boldsymbol{e} = -\boldsymbol{T}\boldsymbol{e}$ the vector $\boldsymbol{\phi}$ sums to one.

4.3 Closure Properties

A number of operations on phase-type distributions lead again to the phase-type with a new representation. Besides being a nice mathematical and probabilistic property it is also algorithmically useful since what would normally be complicated operations, which in general would require numerical integration comes down to matrix operations.[1] The proofs can often be based on the underlying process or in some cases more analytical approaches e.g. the Laplace transform [Neuts, 1981].
To keep the exposition from being too lengthy we will not prove all of the operations under which the phase-type is closed which include finite convolutions and mixtures. But we will go thoroughly through the stability under order statistics which we will extend and combine with section 4.2 for the main result in section 4.3.2.
As stated earlier finite mixtures of independent phase-types are again of phase-type, but this is in general not true for infinite mixtures. Consider for example the infinite mixture of $Er_n(\lambda)$ with $pois(\mu)$ mixing weights, i.e. the distribution given by the density $f_X(x) = \sum_{n=0}^{\infty} p_n f_{n,\lambda}(x)$ where $f_{n,\lambda}(x)$ is the density of an Erlang distribution with n phases and rate λ and p_n are $Pois(\mu)$ point probabilities.

Theorem 4.19. *The distribution X given by the infinite mixture of $Er_n(\lambda)$ described in the discussion above has Laplace transform $\mathcal{L}_X(s) = e^{-\frac{\mu s}{\lambda+s}}$ in particular it is not a phase-type distribution.*

Proof. Recall that the Laplace transform of an Erlang distribution with n phases and rate λ is given by $\left(\frac{\lambda}{s+\lambda}\right)^n$. Since the Laplace transform of a mixture is simply the mixture of the Laplace transforms we get

$$\begin{aligned}
\mathcal{L}_X(s) &= \sum_{n=0}^{\infty} p_n \mathcal{L}_n(s) \\
&= \sum_{n=0}^{\infty} e^{-\mu} \frac{\mu^n}{n!} \left(\frac{\lambda}{s+\lambda}\right)^n \\
&= e^{-\mu} + e^{-\mu} \sum_{n=1}^{\infty} \left(\frac{\mu\lambda}{s+\lambda}\right)^n /n! \quad (\textit{special case } \mathcal{L}_0(s) = 1) \\
&= e^{-\frac{\mu s}{s+\lambda}}.
\end{aligned}$$

Since the Laplace transform of the mixture is not a rational function the mixture is not of phase-type. □

Notice that the above infinite mixture can be written as the compound sum $Z = \sum_{n=0}^{N} X_n$. where $N \sim pois(\mu)$ and $X_n \sim exp(\lambda)$. Since $exp(\lambda) \stackrel{d}{=} \mathrm{PH}_1(1, \{-\lambda\})$ we see that in general compound sums/mixtures of phase-type distributions are not again of phase-type. However, as the next theorem shows, when the mixing weights are given by a discrete phase-type distribution the infinite mixture is indeed of phase-type.

Theorem 4.20. *Let $N \sim DPH_p(\boldsymbol{\alpha}, \boldsymbol{S})$ and $X_n \sim PH_q(\boldsymbol{\beta}, \boldsymbol{T})$ i.i.d and independent of N for all $n \geq 1$. Then*

$$Z = \sum_{n=1}^{N} X_n \sim PH_{pq}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{I} \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t}\boldsymbol{\beta}).$$

i.e. the compound sum is again of phase-type.

[1] The downside of this of course is that the dimensions of these can be very large

Proof. Let $E_N = \{1, 2, \ldots, p\}$ and $E_X = \{1, 2, \ldots, q\}$ be the transient states of the Markov chain respectively Markov jump process underlying N and X_n, $n \geq 1$. Then Z can be constructed in the following way. First the Markov chain underlying N is initiated on E_N according to $\boldsymbol{\alpha}$ and the Markov jump process underlying X_1, with initial distribution $\boldsymbol{\beta}$ evolves until absorption at time X_1, after which $\{X_n\}_{n\in\mathbb{N}}$ jumps from $i \in E_N$ say, to a new (or the same) transient state $j \in E_N$ with probability s_{ij} after which a new Markov jump process is initiated, again with $\boldsymbol{\beta}$ and run until absorption.
The above is repeated until the underlying Markov chain is absorbed into state $p + 1$ which happens with probability s_i if the current state is $i \in E_N$. See Figure 4.8.

Now let $\{Z_t\}_{t\geq 0}$ be the process on $E = E_N \times E_X$ given by the values of the underlying Markov Chain and Markov jump process at time t. Then $\{Z_t\}_{t\geq 0}$ is clearly a Markov jump process and during the time interval $[t, t + dt)$ it may jump from $(i, j) \to (k, l)$ in two ways.

Since transitions between transient states of the underlying Markov jump process happens while the underlying Markov chain remains in its current state, the Markov jump process may jump from $j \to l$ with $i = k$ which happens with probability $t_{jl}dt$.

The Markov jump process may exit to the absorbing state $q + 1$, with probability $t_j dt$, whereafter the Markov chain moves to $k \in E_N$ and the new Markov jump process is initiated in state $l \in E_X$ which has probability $t_j dt s_{ik} \beta_l$.

If E is ordered lexicographically the initial distribution of Z_t is clearly $\boldsymbol{\alpha} \otimes \boldsymbol{\beta}$ and the $pq \times pq$ sub-intensity matrix of $\{Z_t\}$ is given by

$$\boldsymbol{I} \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t}\boldsymbol{\beta}$$

where the first term only allows jumps of Z_t which maintain the underlying value of the Markov chain. The second term; absorption of the Markov jump process, jump of the Markov chain and initiation of the new Markov jump process.
Notice that Z_t may actually jump to itself, if the underlying Markov chain jumps to the same state, and the Markov jump process is initiated in the same state as the one that it was absorbed from. In said case the point would then be deleted, i.e. resulting in a kind of thinning, see Figure 4.8 where the underlying Markov chain take the value 2 at time $n = 2, 3$ and since the Markov jump process generating X_2 is absorbed while in state 1 and the Markov jump process generating X_3 is initiated in state 1 as well, the corresponding point in Z_t is deleted.

□

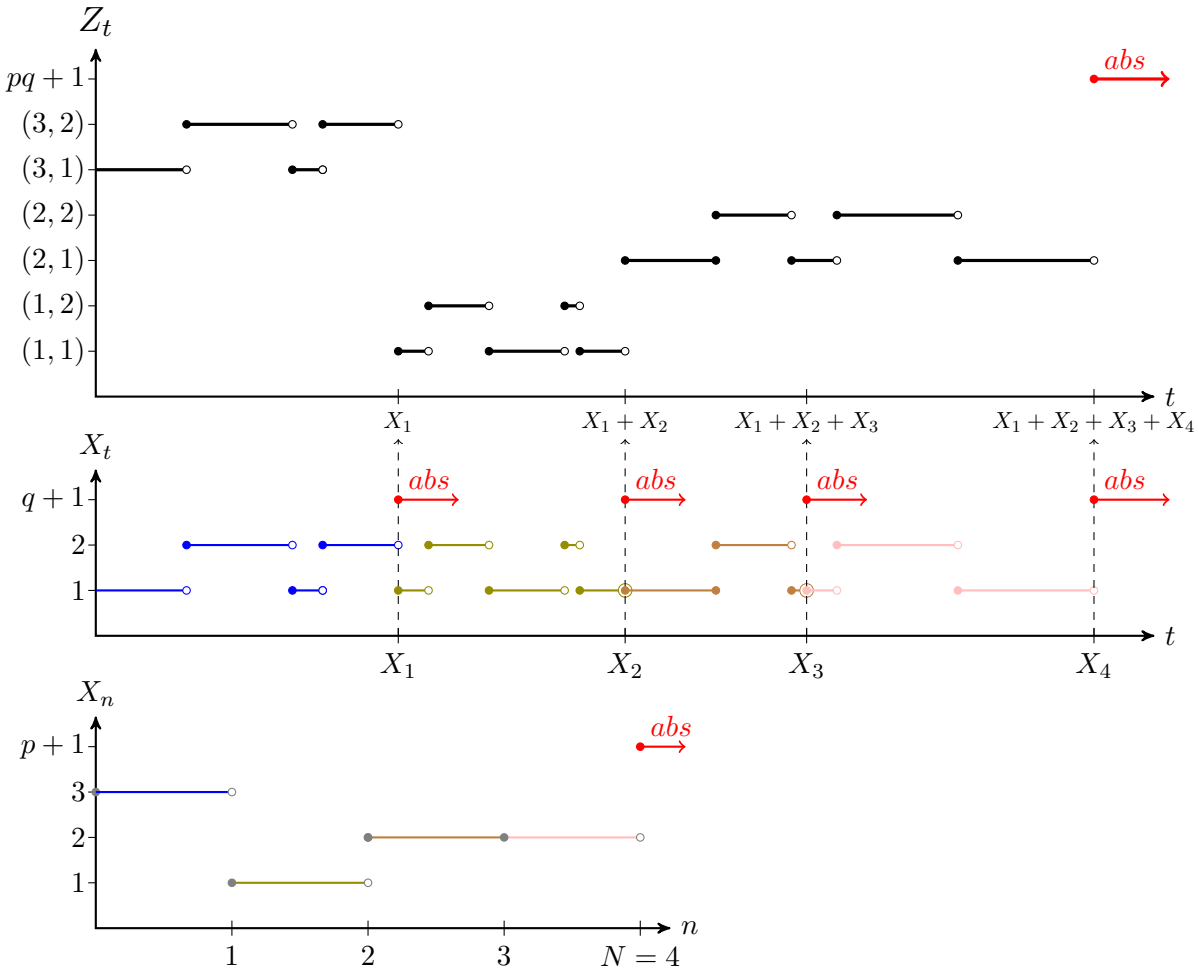

Figure 4.8: Construction of the process Z_t for $p = 3, q = 2$.

4.3.1 Order Statistics

In this section we focus on the closure property for order statistics, that is, for an independent sequence of phase-types $X_1, X_2, ..., X_n$ with $X_i \sim \text{PH}_{p_i}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$ we have the order statistic $X_{k:n}$ defined as the k'th-smallest variable of the sequence. By construction the order statistics have ordering

$$X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}.$$

Assuming that we disallow initiation in the absorbing state p_i+1, $i = 1, ..., n$ in all of the Markov jump processes underlying the phase-types equality between any terms above happens with probability 0. In particular we have that $X_{1:n} = \min\{X_1, ..., X_n\}$ and $X_{n:n} = \max\{X_1, ..., X_n\}$.

This part proves that $X_{k:n}$ for $k = 1, ..., n$ is again phase-type with representation to be deducted in the following. The idea is to simultaneously initiate the n underlying Markov jump processes using the theory of Section 2.2.1. After the first process has been absorbed the remaining $n-1$ processes continue in a reduced state space until k processes have reached their absorbing state the time of which will be $X_{k:n}$.

We will be working with sub-intensity matrices instead of intensity matrices but the results of Section 2.2.1 are still applicable when describing the transitions of the processes prior to absorption. That is, the joint processes $(X_1, X_2, ..., X_n)$, assuming lexicographical ordering on $E = E_1 \times E_2 \times \cdots \times E_n$, prior to absorption of any of the underlying processes moves according to $\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 \oplus \cdots \oplus \boldsymbol{S}_n$. Since the initial distribution of the joint process is the Kronecker product

of the individual initial distributions as seen in (2.15) we have fully described the behaviour of the joint process prior to the first absorption and immediately get that

$$X_{1:n} \sim \mathrm{PH}_{p_{1:n}} \left(\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 \oplus \cdots \oplus \boldsymbol{S}_n \right)$$

where the dimension $p_{1:n}$ is to be specified later.

Proceeding, we now need to continue the remaining $n-1$ processes. Care has to be taken to assure that the remaining processes continue as prior to absorption. Adhering to the lexicographical ordering, we first consider the case of $\{X_t^n\}_{t\geq 0}$ being the first of the absorbed processes. To maintain the states of $(X_t^1, X_t^2, ..., X_t^{n-1})$ at time of absorption, going from the larger state space $E = E_1 \times \cdots \times E_n$ to $E_1 \times \cdots E_{n-1}$, since we do not need to keep track of the absorbed process. the exit-matrix should therefore be on the form $\boldsymbol{I}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \cdots \otimes \boldsymbol{I}_{p_{n-1}} \otimes \boldsymbol{s}_n$ which has dimension $p_1 p_2 \cdots p_n \times p_1 p_2 \cdots p_{n-1}$. The remaining Markov jump processes are now governed by $\boldsymbol{S}_1 \oplus \cdots \oplus \boldsymbol{S}_{n-1}$.

Similarly , if $\{X_t^{n-1}\}_{t\geq 0}$ reaches its absorbing state first, the multidimensional process should transfer from the larger state space E to the smaller marginal state space $E_1 \times \cdots E_{n-2} \times E_n$. Since X^n should retain its current state we now have that the transition matrix should be given by $\boldsymbol{I}_{p_1} \otimes \boldsymbol{I}_{p_2} \otimes \cdots \otimes \boldsymbol{s}_{n-1} \otimes \boldsymbol{I}_{p_n}$ which has dimension $p_1 p_2 \cdots p_n \times p_1 p_2 \cdots p_{n-2} p_n$, now the remaining processes are obviously governed by $\boldsymbol{S}_1 \oplus \cdots \oplus \boldsymbol{S}_{n-2} \oplus \boldsymbol{S}_n$.

We will denote the exit-matrix, $\boldsymbol{S}_{k:n}^0$, governing how the remaining $n-k+1$ processes transfer to the reduced state space of $n-k$ processes, due to the discussion above we have for $k=1$

$$\boldsymbol{S}_{1:n}^0 = \left(\boldsymbol{I}_{p_1} \otimes \cdots \otimes \boldsymbol{I}_{p_{n-1}} \otimes \boldsymbol{s}_n, \boldsymbol{I}_{p_1} \otimes \cdots \otimes \boldsymbol{s}_{n-1} \otimes \boldsymbol{I}_{p_n}, \cdots, \boldsymbol{s}_1 \otimes \boldsymbol{I}_{p_2} \otimes \cdots \otimes \boldsymbol{I}_{p_n} \right)$$

It should now be clear how the procedure is continued until absorption of the kth process.

Regarding the sub-intensity matrices of the reduced state spaces, note that $X_{k:n}$ is the time of the k'th absorption, prior to which $n-k+1$ processes were still active, having indices $j_1, j_2, ..., j_{n-k+1}$. There are $\binom{n}{n-k+1}$ ways to pick the $n-k+1$ processes out of n.
Let $\mathcal{F}(n:k)$ be the set of these $(n-k+1)$-tuples, lexicographically ordered, on the form $(a_1, ..., a_{n-k+1})$ such that $a_i \in \{1, 2, ..., n\}$ and $a_1 < a_2 < \cdots < a_{n-k+1}$ e.g. for $n=5$

$\mathcal{F}(5:1) = \{(1,2,3,4,5)\}$
$\mathcal{F}(5:2) = \{(1,2,3,4), (1,2,3,5), (1,2,4,5), (1,3,4,5), (2,3,4,5)\}$
$\mathcal{F}(5:3) = \{(1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5)(2,3,4), (2,3,5), (2,4,5), (3,4,5)\}$
$\mathcal{F}(5:4) = \{(1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5)\}$
$\mathcal{F}(5:5) = \{1,2,3,4,5\}.$

We are then able to describe the intensity matrix of the $n-k+1$ processes prior to absorption of the k'th process. Define

$$\boldsymbol{S}_{k:n} = \mathrm{diag}\left(\left(\boldsymbol{S}_{j_1} \oplus \boldsymbol{S}_{j_2} \oplus \cdots \oplus \boldsymbol{S}_{j_{n-k+1}} \right)_{(j_1, j_2, ..., j_{n-k+1}) \in \mathcal{F}(n:k)} \right) \tag{4.10}$$

Due to the ordering of $\boldsymbol{S}_{k:n}$ and $\boldsymbol{S}_{k-1:n}$ the transfer from the state space of $n-k+1$ processes to $n-k$, i.e. $\boldsymbol{S}_{k:n}^0$ is uniquely determined.

Let $p_{k:n}$ be the dimension of $X_{k:n}$ we have

$$p_{k:n} = \sum_{j=1}^{k} \left[\sum_{(i_1, ..., i_{n-j+1}) \in \mathcal{F}(n:j)} \prod_{m \in (i_1, ..., i_{n-j+1})} p_m \right] \tag{4.11}$$

Where p_m is the dimension of X_m, $m = 1, ..., n$. Or equivalently

$$p_{k:n} = \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in A_i} \prod_{l=1}^{n} p_l^{x_l} \tag{4.12}$$

where $A_i = \{\boldsymbol{x} \in \{0,1\}^n | \boldsymbol{x}\boldsymbol{e} = n - i + 1\}$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.
In the `R`-package accompanying this thesis the function `dimOrderStat()` takes input $k, (p_1, \ldots, p_n)$, and returns the dimension given by (4.12) or equivalently (4.12).
The dimension increase quite quickly, depending on n the numbers of variables, and $p_1, ..., p_n$ the dimensions of each of the variables. A quick example with $n = 4$, $(p_1, p_2, p_3, p_4) = (6, 9, 2, 5)$ then

k	1	2	3	4
$p_{k:n}$	540	1068	1237	1259

In the case of all dimensions being equal, p say, we have

$$p_{k:n} = \sum_{j=1}^{k} \binom{n}{n-j+1} p^{n-j+1}$$

or equivalently

$$p_{(n-k):n} = (1+p)^n - 1 - \sum_{i=1}^{k} \binom{n}{i} p^i$$

Using the binomial theorem. We have now proved the following

Theorem 4.21 (kth order statistic)**.** *Let $X_1, ..., X_n$ be independent phase-type distributions $X_i \sim PH_{p_i}(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$, $i = 1, 2, ..., n$. Then $X_{k:n}$ is phase-type for $k = 1, \ldots, n$.*

$$X_{k:n} \sim PH_{p_{k:n}}(\boldsymbol{\pi}_{k:n}, \boldsymbol{T}_{k:n}).$$

Where $p_{k:n}$ is given by (4.11) *and*

$$\boldsymbol{\pi}_{k:n} = (\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{0}, ..., \mathbf{0})$$

$$\boldsymbol{T}_{k:n} = \begin{pmatrix} \boldsymbol{S}_{1:n} & \boldsymbol{S}^0_{1:n} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{2:n} & \boldsymbol{S}^0_{2:n} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{S}_{k-1:n} & \boldsymbol{S}^0_{k-1:n} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \boldsymbol{S}_{k:n} \end{pmatrix}.$$

Where $\boldsymbol{S}_{k:n}$ and $\boldsymbol{S}^0_{k:n}$ are defined by (4.10) *and the discussion following it.*

The result of Theorem 4.21 is quite recent and due to [Bladt and Nielsen, 2017], hence computational implementation of the quantities of Theorem 4.21 is not available making practical application less available. In the `R`-package accompanying this thesis the function `knOrderPH()` takes input k, $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n)$ and $(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$ as lists, and returns $(\boldsymbol{T}_{k:n}, \boldsymbol{\pi}_{k:n})$.

For the special case of $n = 2$ we get nice small explicit expressions, and we will summarize this as a corollary.

Corollary 4.22 (max and min of two phase types)**.** *Let* $X_1 \sim PH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ *and* $X_2 \sim PH_n(\boldsymbol{\beta}, \boldsymbol{T})$ *then*

$$X_{1:2} = \min\{X_1, X_2\} \sim PH_{mn}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{S} \oplus \boldsymbol{T})$$

and

$$X_{2:2} = \max\{X_1, X_2\} \sim PH_{mn+m+n}\left((\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \mathbf{0}_n, \mathbf{0}_m), \begin{pmatrix} \boldsymbol{S} \oplus \boldsymbol{T} & \boldsymbol{I}_m \otimes \boldsymbol{t} & \boldsymbol{s} \otimes \boldsymbol{I}_n \\ \mathbf{0}_{m\times mn} & \boldsymbol{S} & \mathbf{0}_{m\times n} \\ \mathbf{0}_{n\times mn} & \mathbf{0}_{n\times m} & \boldsymbol{T} \end{pmatrix}\right)$$

Usually we will not explicitly write the dimensions of the $\mathbf{0}$ and $\boldsymbol{I}$-vectors and matrices. We will return to the max and min of two phase-types in the discussion of Multivariate Phase-Type distributions in Chapter 5.

4.3.2 Order of Absorption

As a motivational example let $X_1 \sim \mathrm{PH}(\boldsymbol{\alpha}_1, \boldsymbol{S}_1), X_2 \sim \mathrm{PH}(\boldsymbol{\alpha}_2, \boldsymbol{S}_2)$ be two independent phase-type distributions. To calculate quantities like

$$\mathbb{P}(X_1 < X_2)$$

One would usually resort to integration of the the joint density

$$\int_0^\infty \int_x^\infty \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 x} \boldsymbol{s}_1 \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_2 x} \boldsymbol{s}_2 dy dx = \int_0^\infty \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 x} \boldsymbol{s}_1 \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_2 x} \boldsymbol{e}\, dx = \int_0^\infty f_1(u)(1 - F_2(u))\, du.$$

For $X_1, \ldots, X_n$ independent phase-types the n-successive integral becomes messy and for all $n!$ different orderings a complete characterization awkward.
In the light of Theorem 4.21 it seems reasonable that we should be not only be able to observe the time of the kth absorption, but also the order of absorptions of the n underlying Markov jump processes. That is, we want to describe the distribution of the $n!$ different orderings that the n Markov jump processes can be absorbed in. This requires some additional notation and definitions which we will go through now, before summarizing the discussion as a theorem.

The setting will be as in Theorem 4.21 and we will utilize the theory of phase-types with multiple absorbing states to find the order of absorption. For simplicity we show the result for $n = 3$ while explaining how the result is generalized for a general n. With X_1, X_2, X_3 independent phase-type distributions with sub-intensity matrices $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3$, initial distributions $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ of dimension p_1, p_2, p_3.
At the outset the 3 underlying Markov jump processes move according to the sub-intensity matrix $\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 \oplus \boldsymbol{S}_3$ and the initial vector is, as in theorem 4.21 given by $\boldsymbol{\pi}_3 = (\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \otimes \boldsymbol{\alpha}_3, \mathbf{0})$. We now carefully construct the overall sub intensity matrix such that we keep track of the order of absorption of the underlying Markov jump processes. In particular we need to note which of the underlying processes gets absorbed first, then move the remaining 2 processes to an *isolated* reduced state space depending on which of the 3 underlying processes that was absorbed.

Adhering to the lexicographical ordering we first handle the case of the 3rd process reaching its absorbing state, then the 2nd and lastly the 1st. Hence the exit rate matrix has the form

$$S^0_{(1:3)} = \begin{pmatrix} \boldsymbol{I} \otimes \boldsymbol{I} \otimes \boldsymbol{s}_3 & \mathbf{0} & \boldsymbol{I} \otimes \boldsymbol{s}_2 \otimes \boldsymbol{I} & \mathbf{0} & \boldsymbol{s}_1 \otimes \boldsymbol{I} \otimes \boldsymbol{I} & \mathbf{0} \end{pmatrix}$$

Where the $\mathbf{0}$-blocks are of dimension such that the underlying processes continue to the immediately following isolated sub-statespace with one underlying process removed. The notation we will adopt in the general case for the exit matrix of underlying processes with index $1, 2, \ldots, n$ will be

$$\boldsymbol{S}^0_{(1:n)}.$$

The renaming processes now have sub-intensity matrix $\boldsymbol{S}_1 \otimes \boldsymbol{S}_2, \boldsymbol{S}_1 \otimes \boldsymbol{S}_3, \boldsymbol{S}_2 \otimes \boldsymbol{S}_3$ depending on which process reached its absorbing state first. It should now be clear that the overall sub-intensity matrix will be given by

$$\boldsymbol{T}_3 = \begin{pmatrix}
\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 \oplus \boldsymbol{S}_3 & \boldsymbol{I} \otimes \boldsymbol{I} \otimes \boldsymbol{s}_3 & \mathbf{0} & \mathbf{0} & \boldsymbol{I} \otimes \boldsymbol{s}_2 \otimes \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \boldsymbol{s}_1 \otimes \boldsymbol{I} \otimes \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I} \otimes \boldsymbol{s}_2 & \boldsymbol{s}_1 \otimes \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \boldsymbol{S}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_1 \oplus \boldsymbol{S}_3 & \boldsymbol{I} \otimes \boldsymbol{s}_3 & \boldsymbol{s}_1 \otimes \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_2 \oplus \boldsymbol{S}_3 & \boldsymbol{I} \otimes \boldsymbol{s}_3 & \boldsymbol{s}_2 \otimes \boldsymbol{I} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_2 & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{S}_3
\end{pmatrix}.$$

The corresponding exit matrix $\boldsymbol{D}_3$ is simply given by the exit rate vectors $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ stacked according to $\boldsymbol{T}_3$ with a new column for every exit rate vector. That is, every column corresponds to a specific path along the sub intensity matrix representing a specific order of absorption.

$$\boldsymbol{D}_3 = \begin{pmatrix}
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\boldsymbol{s}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \boldsymbol{s}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \boldsymbol{s}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{s}_3 & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{s}_2 & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{s}_3
\end{pmatrix}.$$

By construction the columns of $\boldsymbol{D}_3$ corresponds to absorption orders

$$(3,2,1), (3,1,2), (2,3,1), (2,1,3), (1,3,2), (1,2,3)$$

where we by $(3,2,1)$ refer to $\{X_3 < X_2 < X_1\}$. In general we have that the columns of $\boldsymbol{D}$ correspond to the reversely ordered set of all $n!$ combinations of the numbers $1, \ldots, n$ which will be denoted $\mathcal{F}^*(n)$. e.g.

$$\begin{aligned}
\mathcal{F}^*(1) &= \{1\} \\
\mathcal{F}^*(2) &= \{(2,1), (1,2)\} \\
\mathcal{F}^*(3) &= \{(3,2,1), (3,1,2), (2,3,1), (2,1,3), (1,3,2), (1,2,3)\} \\
\mathcal{F}^*(4) &= \{(4,3,2,1), (4,3,1,2), (4,2,3,1), (4,2,1,3), (4,1,3,2), \\
&\quad (4,1,2,3), (3,4,2,1), (3,4,1,2), (3,2,4,1), (3,2,1,4), (3,1,4,2), (3,1,2,4), \\
&\quad (2,4,3,1), (2,4,1,3), (2,3,4,1), (2,3,1,4), (2,1,4,3), (2,1,3,4), (1,4,3,2), \\
&\quad (1,4,2,3), (1,3,4,2), (1,3,2,4), (1,2,4,3), (1,2,3,4)\}
\end{aligned}$$

For the general case, let $\boldsymbol{T}_{(1:n)(-k)}$ correspond to the sub intensity matrix constructed above for sub-intensity matrix indices $1, \ldots, n$ with kth index removed e.g. $\boldsymbol{T}_{(1:4)(-4)} = \boldsymbol{T}_3$ above. We then have the recursive general formula for $\boldsymbol{T}_n$.

$$\boldsymbol{T}_n = \boldsymbol{T}_{(1:n)} = \begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 \oplus \ldots \oplus \boldsymbol{S}_n & & \boldsymbol{S}^0_{(1:n)} & & \\ \mathbf{0} & \boldsymbol{T}_{(1:n)(-n)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{(1:n)(-(n-1))} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots\vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{(1:n)(-2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{(1:n)(-1)} \end{pmatrix} \tag{4.13}$$

With boundary condition for when only two indices $i, j \in \{1, \ldots, n\}, i \neq j$ remain, here without loss of generality $i < j$

$$\boldsymbol{T}_{(i,j)} = \begin{pmatrix} \boldsymbol{S}_i \oplus \boldsymbol{S}_j & \boldsymbol{I} \otimes \boldsymbol{s}_j & \boldsymbol{s}_i \otimes \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{S}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{S}_j \end{pmatrix}.$$

The initial vector is, as in Theorem 4.21 given by

$$\boldsymbol{\pi}_n = (\boldsymbol{\alpha}_1 \otimes \cdots \otimes \boldsymbol{\alpha}_n, \mathbf{0}, ..., \mathbf{0}) \tag{4.14}$$

The dimension of 4.13 and 4.14. is given by

$$\sum_{i=0}^{n-1} \left[i! \sum_{\boldsymbol{x} \in B_{n-i}} \prod_{l=1}^{n} p_l^{x_l} \right] \tag{4.15}$$

where $B_i = \{\boldsymbol{x} \in \{0,1\}^n | \boldsymbol{x}\boldsymbol{e} = i\}$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$
To see this note that each block of kronecker sums corresponds to an entry in B_i.
B_n represents the sub state space of all n processes being alive, the one element of B_n hence only enters once. Entries corresponding to B_{n-1} also only appear once each. But for B_{n-2} each element appears twice, since e.g. the B_{n-2} element with indices k, l removed are generated by either the B_{n-1} element with index k removed, then removing index l when entering the $n-2$ sub-statespace, or the B_{n-1} element with index l removed then removing index k when entering the $n-2$ sub state space and so forth.

In the `R`-package accompanying this thesis the function `dimAbsOrderPH()` takes input $(p_1, \ldots, p_n)$, and returns the dimension given by (4.15).

We have now done the work for applying Theorem 4.18 to get the probabilities of the $n!$ different orders of absorption of the n underlying Markov jump processes, and have now proved the following.

Theorem 4.23 (Absorption order)**.** *For $X_1, \ldots, X_n$ independent phase-type distributions with representations $(\boldsymbol{\alpha}_1, \boldsymbol{S}_1), \ldots, (\boldsymbol{\alpha}_n, \boldsymbol{S}_n)$ the possible orderings of $X_1, \ldots, X_n$ have probabilities given by the corresponding entry in the vector*

$$\boldsymbol{\pi}_n - \boldsymbol{T}_n^{-1} \boldsymbol{D}_n.$$

Where the entries correspond to the orders given by $\mathcal{F}^*(n)$ and $\boldsymbol{\pi}_n, \boldsymbol{T}_n, \boldsymbol{D}_n$ are defined in the discussion above.

For the result of Theorem 4.23 to have any practical relevance computational implementation of the quantities need to be readily available. In the `R`-package accompanying this thesis the function `AbsOrderPH()` takes input $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n)$ and $(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$ and returns $(\boldsymbol{\pi}_n, \boldsymbol{T}_n, \boldsymbol{D}_n)$.

Remark 3. Note that if one disregard the $\boldsymbol{D}$-matrix and simply use $(\boldsymbol{\pi}_n, \boldsymbol{T}_n)$ as representation of a phase-type distribution, this is in particular the time of absorption of the last of the n underlying Markov jump processes hence

$$\mathrm{PH}(\boldsymbol{\pi}_n, \boldsymbol{T}_n) \overset{d}{=} \max\{X_1, \ldots, X_n\} \overset{d}{=} \mathrm{PH}(\boldsymbol{\pi}_{n:n}, \boldsymbol{T}_{n:n}).$$

Where the quantities $(\boldsymbol{\pi}_{n:n}, \boldsymbol{T}_{n:n})$ are from Theorem 4.21. i.e. $(\boldsymbol{\pi}_n, \boldsymbol{T}_n)$ can in particular be seen as a greatly over parametrized representation for the nth order statistic i.e max of n independent phase types.

4.3.3 Incorporating absorptions orders

We will briefly touch upon an application of the absorption orders. Consider

$$\mathbb{P}(X_1 > c \,|\, X_1 > X_2)$$

or more generally the density given by

$$f(u)du = \mathbb{P}(X_1 \in du | X_1 > X_2) = \mathbb{P}(X_1 \in [u, u+du) | X_1 > X_2) \tag{4.16}$$

which in some cases can be more easily be expressed by

$$f(u)du = \frac{\mathbb{P}(X_1 \in du, X_1 > X_2)}{\mathbb{P}(X_1 > X_2)}.$$

For $\mathbb{P}(X_1 \in du, X_1 > X_2)$ we reason probabilistically using that for $X_1 \in du$ then the process underlying X_2 must be absorbed during $[0, u)$ after which the process underlying X_2 continues $u - x$ before absorption, then integrating out the time of absorption of the process underlying X_2 we get

$$\begin{aligned}\mathbb{P}(X_1 \in du, X_1 > X_2) &= \int_{x=0}^{x=u} \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 x} \boldsymbol{I} \otimes \boldsymbol{s}_2 e^{\boldsymbol{S}_1 (u-x)} \boldsymbol{s}_1 dx \\ &= \int_{x=0}^{x=u} \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 (u-x)} \boldsymbol{I} \otimes \boldsymbol{s}_2 e^{\boldsymbol{S}_1 x} \boldsymbol{s}_1 dx\end{aligned}$$

Using Van Loans rule for calculating integrals involving matrix-exponentials, see e.g. appendix A.2.1 in [Bladt and Nielsen, 2017] stating that for $\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{pmatrix}$ then

$$e^{\boldsymbol{\Lambda} u} = \begin{pmatrix} e^{\boldsymbol{A} u} & \int_0^u e^{\boldsymbol{A}(u-x)} \boldsymbol{B} e^{\boldsymbol{C} x} dx \\ \boldsymbol{0} & e^{\boldsymbol{C} u} \end{pmatrix}$$

with $\boldsymbol{A} = \boldsymbol{S}_1 \oplus \boldsymbol{S}_2, \boldsymbol{B} = \boldsymbol{I} \otimes \boldsymbol{s}_2, \boldsymbol{C} = \boldsymbol{S}_1$ we see that we simply need the upper right block of $e^{\boldsymbol{\Lambda} u}$ for the integral, hence

$$\int_{x=0}^{x=u} \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 (u-x)} \boldsymbol{I} \otimes \boldsymbol{s}_2 e^{\boldsymbol{S}_1 x} \boldsymbol{s}_1 dx = \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix} e^{\begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I} \otimes \boldsymbol{s}_2 \\ \boldsymbol{0} & \boldsymbol{S}_1 \end{pmatrix} u} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{s}_1$$

Where $\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}$ is the concatenation of a $(p_1 p_2)$-identity matrix and a $(p_1 p_2) \times p_1$-zero block matrix similairly for $\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{pmatrix}$ with a $(p_1 p_2) \times p_1$ zero-block and a p_1-identity matrix.

The density (4.16) is therefore given by

$$f(u) = \frac{\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix} e^{\begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I} \otimes \boldsymbol{s}_2 \\ \boldsymbol{0} & \boldsymbol{S}_1 \end{pmatrix} u} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{s}_1}{\mathbb{P}(X_1 > X_2)} \tag{4.17}$$

The probability $\mathbb{P}(X_1 > X_2)$ can then either be calculated using Theorem 4.23 as the first entry of the vector i.e. $\boldsymbol{\pi}_2 - \boldsymbol{T}_2^{-1}\boldsymbol{D}_2\boldsymbol{e}_1$ or simply by integrating out the absorption time of the process underlying X_2 when both processes are active i.e.

$$\int_0^\infty \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 u} \boldsymbol{I} \otimes \boldsymbol{s}_2 du \boldsymbol{e} = \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 (-\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1} \boldsymbol{I} \otimes \boldsymbol{s}_2 \boldsymbol{e}.$$

Lastly, we note that we now have a simple way of expressing the initial conditional probability given by

$$\mathbb{P}(X_1 > c \,|\, X_1 > X_2)$$

which due to the phase-like structure of the conditional process can simply be viewed as the regular case of survival probabilities but now under the construction leading to (4.17) hence

$$\mathbb{P}(X_1 > c \,|\, X_1 > X_2) = \frac{\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2 \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix} e^{\begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I} \otimes \boldsymbol{s}_2 \\ \boldsymbol{0} & \boldsymbol{S}_1 \end{pmatrix} c} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{e}}{\mathbb{P}(X_1 > X_2)}. \tag{4.18}$$

To get some intuition and practical experience with Theorem 4.21 and 4.23 lets look at a thorough example.

Example 4.24. For $n = 4$ independent phase-type distributions given by

$$X_1 \sim \mathrm{PH}_3\left(\left(\frac{1}{10}, \frac{4}{10}, \frac{5}{10}\right), \begin{pmatrix} -10 & 2 & 3 \\ 1 & -12 & 10 \\ 6 & 1 & -8 \end{pmatrix}\right)$$

$$X_2 \sim \mathrm{PH}_4\left(\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right), \begin{pmatrix} -4 & 1 & 1 & 1 \\ 1 & -5 & 1 & 1 \\ 1 & 1 & -6 & 1 \\ 1 & 1 & 1 & -7 \end{pmatrix}\right)$$

$$X_3 \sim \mathrm{PH}_2\left(\left(\frac{1}{10}, \frac{9}{10}\right), \begin{pmatrix} -4 & 3 \\ 1 & -5 \end{pmatrix}\right)$$

$$X_4 \sim \mathrm{PH}_5\left(\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right), \begin{pmatrix} -7 & 1 & 1 & 1 & 2 \\ 1 & -7 & 1 & 1 & 2 \\ 1 & 1 & -10 & 1 & 1 \\ 1 & 1 & 1 & -8 & 3 \\ 3 & 3 & 2 & 1 & -10 \end{pmatrix}\right)$$

using Theorem 4.21 we get the phase-type representation of the order statistics $X_{1:4}, X_{2:4}, X_{3:4}, X_{4:4}$. Below we plot the densities.

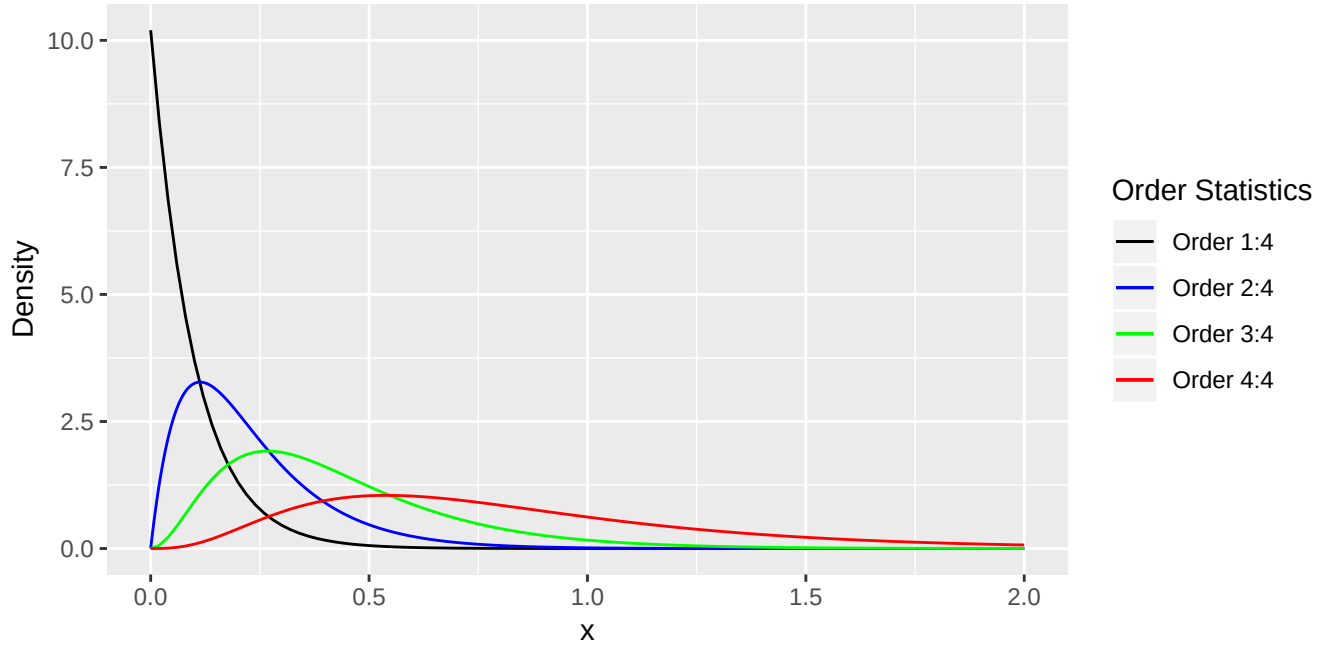

Figure 4.9: Densities for $1st, 2nd, 3rd$ and $4th$ order statistic of X_1, X_2, X_3, X_4 independent phase-type distributions.

The moments are calculated using the theory from the beginning of this section e.g.

$$\mathbb{E}X_{1:4} = \boldsymbol{\pi}_{1:4}(-\boldsymbol{T}_{1:4})^{-1}\boldsymbol{e} = 0.0967$$

$$\mathbb{V}X_{1:4} = 2!\boldsymbol{\pi}_{1:4}\left(-\boldsymbol{T}_{1:4}\right)^{-2}\boldsymbol{e} - \left(\boldsymbol{\pi}_{1:4} - \boldsymbol{T}_{1:4}^{-1}\boldsymbol{e}\right)^2 = 0.00934$$

summarized below in Table 4.1.

Order Statistic	Expectation	Variance
$X_{1:4}$	0.097	0.00934
$X_{2:4}$	0.228	0.0268
$X_{3:4}$	0.428	0.0681
$X_{4:4}$	0.837	0.241

Table 4.1: Moments of order statistics

The most complex of the matrices is obviously $\boldsymbol{T}_{4:4}$, the maximum of the 4 phase-type distributions. As discussed in the proof of Theorem 4.18 the sub-intensity matrix describes the 4 underlying Markov jump processes of X_1, X_2, X_3, X_4 until one is absorbed where-after the state space is reduced to only contain 3 processes and so forth.
From (4.11) or (4.12) the sub-intensity matrix $\boldsymbol{T}_{4:4}$ has dimension 359×359 and consists of 4 block matrices where each block corresponds to the number of underlying processes still being active. This is visualized below where the block matrix where all 4 processes are active is coloured dark green the block matrix describing 3 remaining processes dark orange , the block matrix describing 2 remaining processes dark blue, and the final block matrix for a single process red. The exit matrices have a corresponding lighter color.

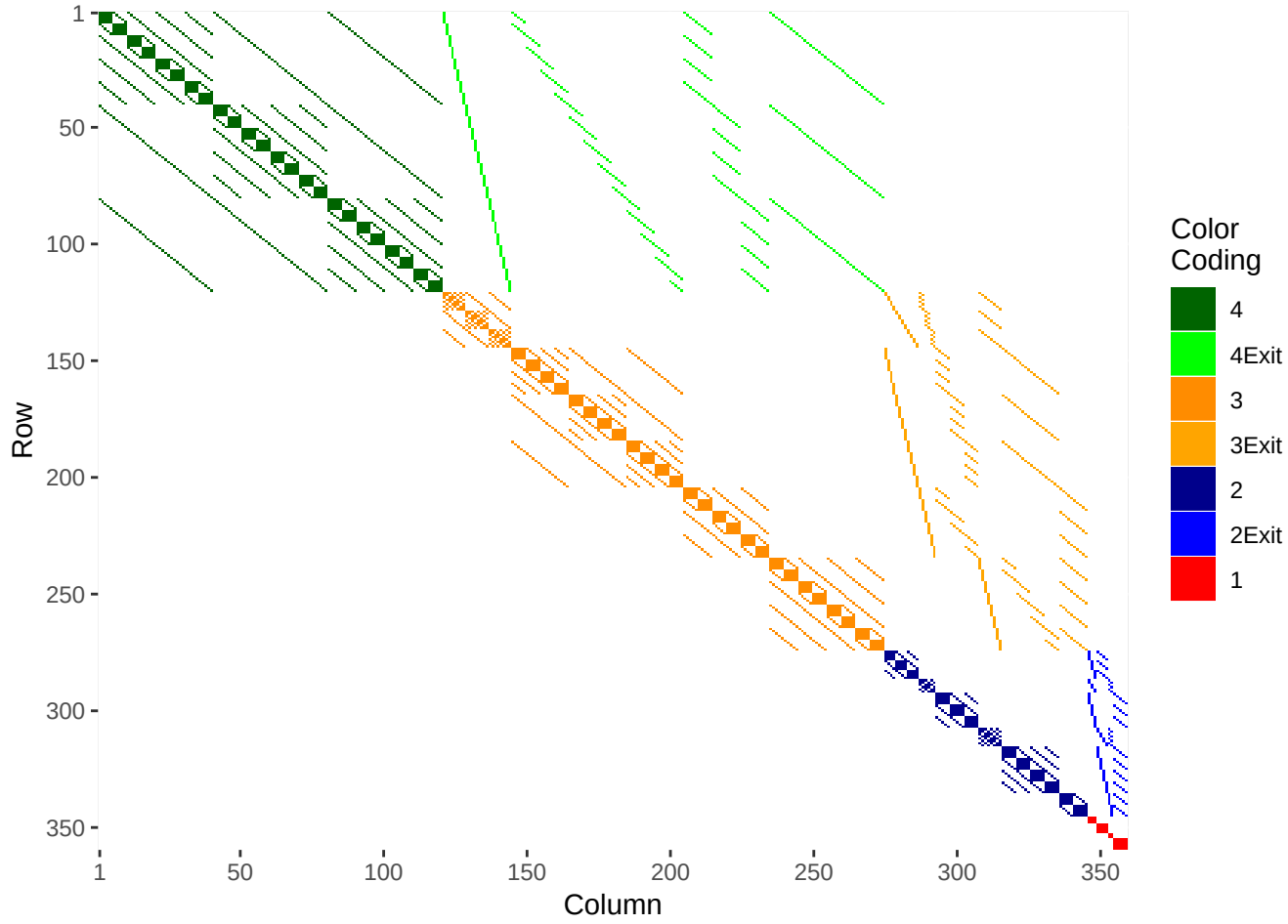

Figure 4.10: $\boldsymbol{T}_{4:4}$ sub intensity matrix visualized. $n = 4$ dark green, $n = 3$ dark orange, $n = 2$ dark blue, $n = 1$ red.

Next we look at the total of $4! = 24$ different orderings of X_1, X_2, X_3, X_4 which we shall

express using Theorem 4.23. Using the `R`-function `AbsOrderPH` we calculate $\boldsymbol{\pi}_4, \boldsymbol{T}_4, \boldsymbol{D}_4$ of Theorem 4.23. The sub-intensity matrix $\boldsymbol{T}_4$ which is now, using (4.15), of dimension 500×500 is visualized by colouring the entries according to whether these are different from zero. Again we colour matrix blocks according the which sub state space they correspond to, as in Figure 4.10.

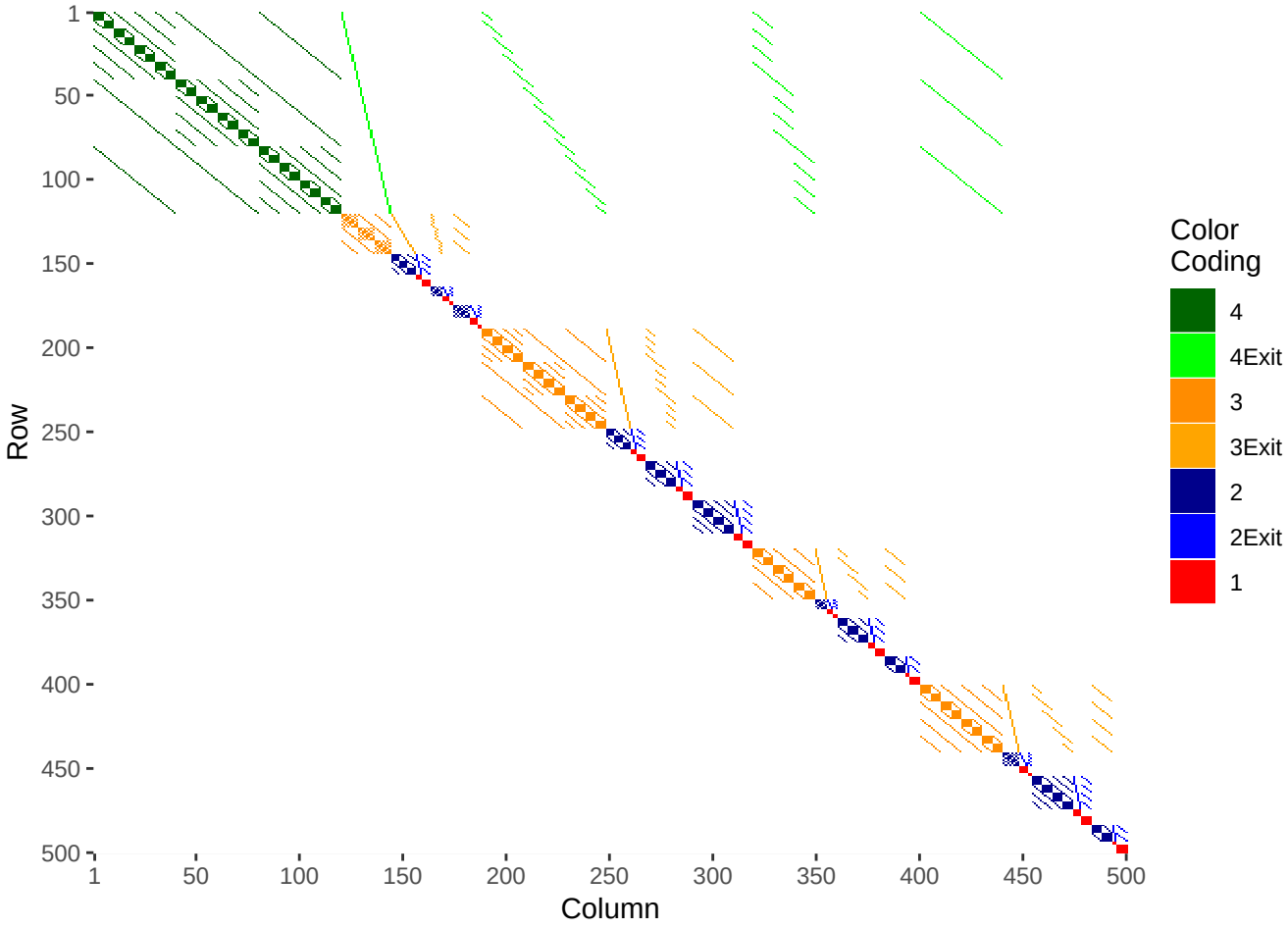

Figure 4.11: $\boldsymbol{T}_4$ sub-intensity matrix for absorption order visualized. The number of active underlying processes are color coded as: $n = 4$ dark green, $n = 3$ dark orange, $n = 2$ dark blue, $n = 1$ red.

The probabilities of each ordering of X_1, X_2, X_3, X_4 is given by

$$\boldsymbol{\pi}_4(-\boldsymbol{T}_4)^{-1}\boldsymbol{D}_4$$

with probability masses shown in Figure 4.12.

Notice that orders where X_3 is early in the order are more likely. This makes sense at least from a heuristic argument since the expected values of the phase types are $\mathbb{E}X_1 = 0.427$, $\mathbb{E}X_2 = 0.434, \mathbb{E}X_3 = 0.312, \mathbb{E}X_4 = 0.416$ hence, since the expected value of X_3 is relatively lower than the rest, orders in which X_3 is appearing as the first variable are more probable.

Figure 4.12: Probabilities of all possible orderings of $n = 4$ phase types. See Appendix A.2.4 for code for reproduction.

Lastly the conditional density given by (4.16) i.e. $f(u)du = \mathbb{P}(X_1 \in du | X_1 > X_2)$ is plotted below in Figure 4.13 along with the unconditional density of X_1.

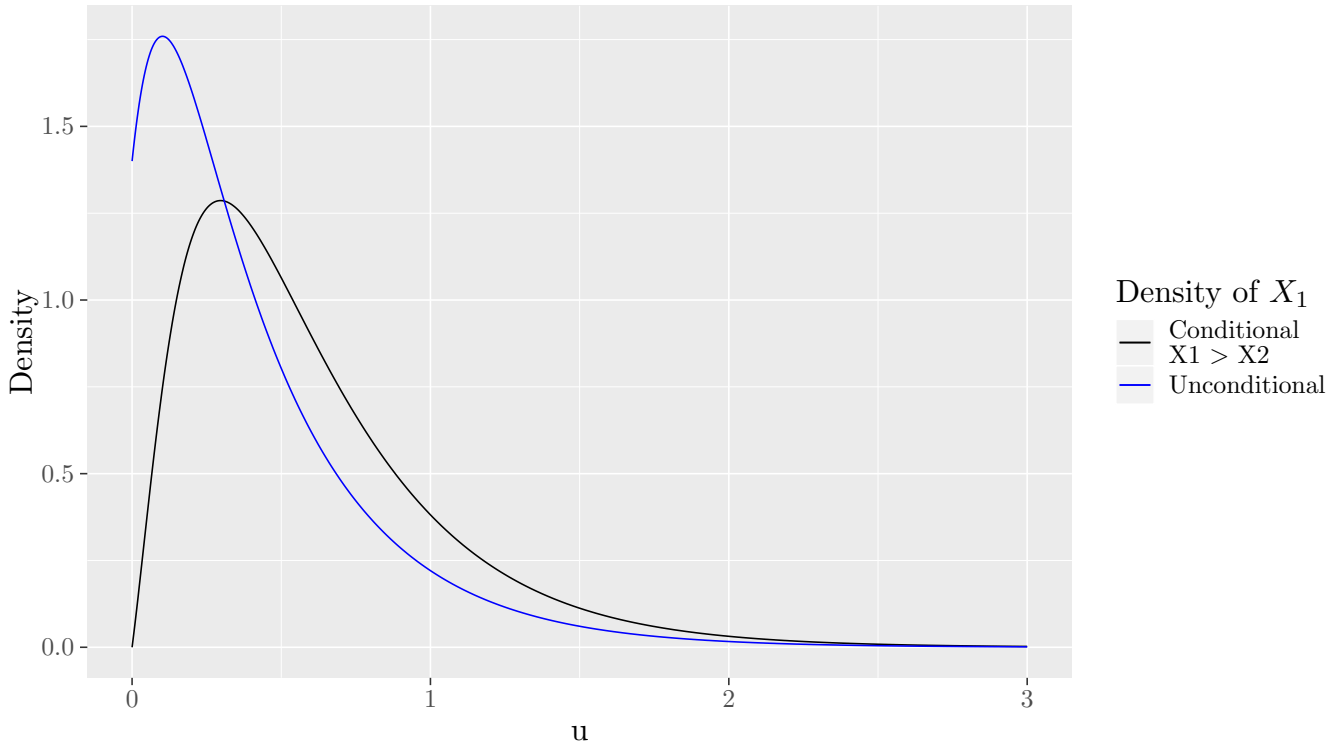

Figure 4.13: Density of X_1 unconditional and conditioned on $X_1 > X_2$.

4.4 The Denseness Property

As we saw in the previous section the class of phase-type distributions possesses many attractive properties when performing various operations. This together with the fact that, as we shall see next, the phase-type is an extremely versatile and flexible class of probability distributions is a mayor motivation both for applications and mathematical research.
This section proves the fact that the class of phase-type distributions is dense in the class of distributions on $[0, \infty)$ in the sense of weak convergence that is, for every non-negative random variable X there exists a sequence $X_1, X_2, \ldots$ of phase-type distributed random variables such that $X_n \xrightarrow{d} X$ meaning that $\lim_{n\to\infty} F_n(x) = F(x)$ for all continuity points of F, or equivalent that for all bounded, continuous mappings $f : \mathbb{R}_+ \to \mathbb{R}$ then $\lim_{n\to\infty} Ef(X_n) = Ef(X)$.

From a practical perspective this is extremely useful since this means that any distribution on $\mathbb{R}_+$ can, at least in principal, be approximated arbitrarily well by a phase-type distribution while retaining the properties of the phase-type from which we are able to analyse a variety of stochastic systems in a unified way and in an algorithmically tractable manner.
As we have seen earlier Erlang distributions are in the class of phase-type distributions and consequently finite mixtures of Erlang are again of phase-type. We now prove that the class of finite mixtures of Erlang distributions are dense in the class of non-negative distributions implying denseness of the phase type too. The proof is based on [Bladt and Nielsen, 2017] and [Johnson and Taaffe, 1988] the latter was made available by one of the authors.

Lemma 4.25. *Let $\{X_{n,\theta}\}_{n\in\mathbb{N}}$ be a sequence of non-negative random variables with distribution $\{F_{n,\theta}\}_{n\in\mathbb{N}}$. If $\mathbb{E}X_{n,\theta} = \theta > 0$ for all $n \in \mathbb{N}$ and*

$$Var\,(X_{n,\theta}) = \sigma^2_{n,\theta} \to 0 \quad as \quad n \to \infty$$

then for every bounded function u continuous in θ, we have that

$$\mathbb{E}\,(u(X_{n,\theta})) \to u(\theta).$$

If u is bounded continuous this is equivalent to $X_{n,\theta} \xrightarrow{d} \theta$ as $n \to \infty$.

Proof. Since u is assumed continuous in θ, given $\varepsilon > 0$ there exists some $\delta > 0$ such that $|u(x) - u(\theta)| < \varepsilon/2$ whenever $|x - \theta| < \delta$. Furthermore, u is bounded by a constant $M/2$, say then

$$\begin{aligned}
|\mathbb{E}\,(u(X_{n,\theta})) - u(\theta)| &\leq \int_0^\infty |u(x) - u(\theta)| dF_{n,\theta} \\
&= \int_{\{x:|x-\theta|<\delta\}} |u(x) - u(\theta)| dF_{n,\theta} + \int_{\{x:|x-\theta|\geq\delta\}} |u(x) - u(\theta)| dF_{n,\theta} \\
&\leq \frac{\varepsilon}{2} + \int_{\{x:|x-\theta|\geq\delta\}} |u(x) - u(\theta)| dF_{n,\theta} \\
&\leq \frac{\varepsilon}{2} + M\mathbb{P}(|X_{n,\theta} - \theta| \geq \delta) \quad \textit{(Boundedness of } u\textit{)} \\
&\leq \frac{\varepsilon}{2} + M\frac{\sigma^2_{n,\theta}}{\delta^2} \quad \textit{(Chebyshev's inequality)}
\end{aligned}$$

By assumption $\sigma^2_{n,\theta} \to 0$ in particular there exists an n_0 such that for $n \geq n_0$ then $\sigma^2_{n,\theta} \leq \frac{\varepsilon}{2M}\delta^2$. Hence for $n \geq n_0$ we have

$$|\mathbb{E}\,(u(X_{n,\theta})) - u(\theta)| \leq \varepsilon$$

and by definition $\mathbb{E}\,(u(X_{n,\theta})) \to u(\theta)$ as $n \to \infty$. □

Lemma 4.26. *Let $\theta > 0$ and u a bounded function continuous in θ. Then*

$$\lim_{n\to\infty} e^{-n\theta} \sum_{k=0}^{\infty} u\left(\frac{k}{n}\right) \frac{(n\theta)^k}{k!} = u(\theta)$$

Proof. Let $Y_n \sim pois(n\theta)$ and $X_n = \frac{1}{n}Y_n$. Then $\mathbb{E}X_n = \theta$, $\mathrm{Var}(X_n) = \frac{\theta}{n}$ and since $X_n \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots\}$ with $\mathbb{P}(X_n = \frac{k}{n}) = e^{-n\theta}\frac{(n\theta)^k}{k!}$ for $k \geq 0$ we have that

$$\mathbb{E}\left(u(X_n)\right) = \sum_{k=0}^{\infty} u\left(\frac{k}{n}\right) e^{-n\theta}\frac{(n\theta)^k}{k!}$$

and the result follows from Lemma 4.25. □

Now, let $Er_k(t;\lambda)$ be the distribution function of an $Er_k(\lambda)$-distribution. Since this is equal to a k-convolution of exponential distributions with rate λ $Er_k(t;\lambda)$ can be viewed as the probability of the kth arrival in a poisson process with rate λ occurring before time t. It is well known that the distribution of arrivals in the time interval $[0,t)$ in a Poisson process with rate λ is poisson distributed with mean λt then with $p_i(t) = e^{-\lambda t}\frac{(\lambda t)^i}{i!}$, we get

$$Er_k(t;\lambda) = 1 - \sum_{i=0}^{k-1} p_i(t).$$

To avoid cumbersome notation in the following introduce for an arbitrary distribution function F, $\lambda > 0$

$$d_F(\lambda, k) = F\left(\frac{k}{\lambda}\right) - F\left(\frac{k-1}{\lambda}\right)$$

notice that for $X \sim F$ then $d_F(\lambda,k) = \mathbb{P}\left(X \in \left(\frac{k-1}{\lambda}, \frac{k}{\lambda}\right]\right)$ and if F has support on $(0,\infty)$ then $\sum_{k=1}^{\infty} d_F(\lambda,k) = 1$ and in particular $d_F(\lambda,k)$, $k = 1,2,\ldots$ can be seen as infinite mixing weights.

Lemma 4.27. *Let F be a distribution with support $(0,\infty)$ and $\lambda > 0$ then*

$$\sum_{k=1}^{\infty} d_F(\lambda,k) Er_k(t;\lambda) = e^{-\lambda t} \sum_{k=0}^{\infty} F\left(\frac{k}{\lambda}\right)\frac{(\lambda t)^k}{k!}.$$

Proof.

$$\begin{aligned}
\sum_{k=1}^{\infty} d_F(\lambda,k) Er_k(t;\lambda) &= \sum_{k=1}^{\infty} d_F(\lambda,k)\left(1 - \sum_{j=0}^{k-1} p_j\right) \\
&= \underbrace{-F\left(\frac{0}{\lambda}\right)}_{=0} + \underbrace{\lim_{k\to\infty} F\left(\frac{k}{\lambda}\right)}_{=1} - \sum_{k=1}^{\infty} d_F(\lambda,k)\sum_{j=0}^{k-1} p_j \\
&= 1 - \sum_{k=1}^{\infty} d_F(\lambda,k)\sum_{j=0}^{k-1} p_j
\end{aligned}$$

Since F has support on $(0,\infty)$ and the first sum is telescoping. Now changing the order of

summation

$$\begin{aligned}\sum_{k=1}^{\infty} d_F(\lambda,k)\sum_{j=0}^{k-1} p_j &= \sum_{j=0}^{\infty} p_j \sum_{k=j+1}^{\infty} d_F(\lambda,k)\\ &= \sum_{j=0}^{\infty} p_j \left(1 - F\left(\frac{j}{\lambda}\right)\right)\\ &= 1 - \sum_{j=0}^{\infty} p_j F\left(\frac{j}{\lambda}\right)\\ &= 1 - e^{-\lambda t}\sum_{j=0}^{\infty} F\left(\frac{j}{\lambda}\right)\frac{(\lambda t)^j}{j!}.\end{aligned}$$

From which we get the result. □

Theorem 4.28. *Let F be any distribution concentrated on $(0,\infty)$. Define for $n \in \mathbb{N}$,*

$$F_n(x) = \sum_{k=1}^{\infty} d_F(n,k) Er_k(x;n).$$

Then F_n is a distribution function, moreover

$$F_n(t) = F(t) \quad as \quad n \to \infty$$

for every continuity point t of F.

Proof. As discussed earlier $F_n(x)$ is an infinite mixture of Erlang distribution functions and is therefore itself a distribution function.
Now, using Lemma 4.27 on F_n with $\lambda = n$ and then Lemma 4.26 with $u = F$ since F is clearly bounded by e.g. 1 we get that for all continuity points t of F that

$$F_n(t) \to F(t) \quad as \quad n \to \infty$$

□

The above construction is given by infinite mixtures of Erlang distributions with the same intensity parameter n. Thus we have proved that the class of infinite mixtures of Erlang distributions are dense in the class of distributions on $[0,\infty)$ which was well known prior to the denseness property of phase-type distributions. See e.g. [Schassberger, 1973]. Though the infinite-mixture denseness result intuitively motivates the denseness of phase-type distributions, the result does not follow trivially since phase-type distributions have finite dimension.

Notice that for a distribution F with support on $[0,\infty)$ but not on $(0,\infty)$ then this must be due to an atom at zero of size p, say. Clearly this point is then not a continuity point of F hence the result still holds for distributions of $[0,\infty)$. Alternatively we could modify the above argument and consider a mixture between an atom at zero with mixing weight p and an infinite Erlang mixture with weight $1-p$.

The final step in the proof is now to consider finite mixtures of Erlang distributions, by a limiting argument of the finite mixtures we are are able to regain the result for infinite mixtures.

Theorem 4.29. *Let F be any any distribution on $(0,\infty)$. Define*

$$F_{m,n}(t) = F\left(\frac{m}{n}\right)^{-1} \sum_{k=1}^{m} d_F(n,k) Er_k(t;n).$$

Then $F_{m,n} \to F_n$ uniformly on $(0,\infty)$. With F_n defined as in Theorem 4.28.

Proof. Let $t > 0$ then

$$\begin{aligned}
|F_n(t) - F_{m,n}(t)| &= \left| \sum_{k=1}^{\infty} d_F(n,k) Er_k(t;n) - F\left(\frac{m}{n}\right)^{-1} \sum_{k=1}^{m} d_F(n,k) Er_k(t;n) \right| \\
&= \left| \left[1 - F\left(\frac{m}{n}\right)^{-1} \right] \sum_{k=1}^{m} d_F(n,k) Er_k(t;n) + \sum_{k=m+1}^{\infty} d_F(n,k) Er_k(t;n) \right| \\
&\leq \left| \left[1 - F\left(\frac{m}{n}\right)^{-1} \right] \sum_{k=1}^{m} d_F(n,k) Er_k(t;n) \right| + \left| \sum_{k=m+1}^{\infty} d_F(n,k) Er_k(t;n) \right| \\
&< \left| \left[1 - F\left(\frac{m}{n}\right)^{-1} \right] \sum_{k=1}^{m} d_F(n,k) \right| + \left| \sum_{k=m+1}^{\infty} d_F(n,k) \right| \quad (0 \leq Er_k(t;n) < 1) \\
&= \left| \left[1 - F\left(\frac{m}{n}\right)^{-1} \right] F\left(\frac{m}{n}\right) \right| + 1 - F\left(\frac{m}{n}\right) \\
&\to 0
\end{aligned}$$

as $m \to \infty$ since F is a distribution function. Since the convergence does not depend on $t \in (0, \infty)$ the convergence is uniform. □

Since $F_{m,n}$ converges uniformly to F_n which in turn converges to F for every continuity point we get the following corollary and theorem.

Corollary 4.30. *Let F be an arbitrary distribution on $(0, \infty)$ and t a point of continuity of F. Then*

$$\lim_{n\to\infty} \lim_{m\to\infty} F_{m,n}(t) = F(t).$$

In particular if $X \sim F$ then there exists a sequence $X_1, X_2, \ldots$ of finite Erlang mixtures such that $X_n \xrightarrow{d} X$.

Since Erlang distribution is of phase-type and finite phase-type mixtures is again of phase-type we immediately get the theorem.

Theorem 4.31. *The class of phase-type distributions is dense in the class of distributions on $[0, \infty)$ the the sense of weak convergence.*

The above theorem motives the use of phase-type distributions as a form of semi-parametric method when working with distributions of $\mathbb{R}_+$. Being able to efficiently fit phase-types from data has therefore been an active research area see e.g. [Asmussen et al., 1996] which develop and implement an EM-algorithm for phase-type fitting.

4.5 Transformation via Rewards

Given $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ a powerful and useful extension of the basic phase-type is constructed by letting the underlying Markov jump process, $\{X_t\}_{t\geq 0}$, be rewarded at rate given by a function $r : \{1, 2, ..., p\} \to \mathbb{R}_+$ depending on the state of the underlying process. That is, at time $t \geq 0$ reward is earned at rate $r(X_t)$. We collect the reward rates in a vector $\boldsymbol{r} = (r(1), ..., r(p)) \in \mathbb{R}_+^p$. Note that we restrict ourselves to non-negative rewards. Interestingly, using negative reward is fundamental in creating the so-called Bilateral phase-type which is an extension of the phase-type onto $(-\infty, \infty)$, see e.g. [Ramaswami, 2005], also Fluid Flow models is an example of negative rewards being used although we will not pursue any of these topics here.
At absorption of the underlying process we look at the total collected reward.

Definition 4.32 (Rewards)**.** *With the setup above we define the random variable*

$$Y = \int_0^\tau r(X_t)dt$$

As the total reward earned by X_t until absorption under reward function r.

We will spend the next few pages proving that Y is in fact also of phase-type.
We start by dividing the state space according to whether the reward is positive or zero

$$E_+ = \{i \in \{1, 2, ..., p\} : r(i) > 0\}, \quad E_0 = \{i \in \{1, 2, ..., p\} : r(i) = 0\}.$$

Define now the transition matrix, $\boldsymbol{Q}$ for the embedded markov chain,$\{Y_n\}_{n\in\mathbb{N}}$ on the transient states $\{1, 2, ..., p\}$.

$$q_{ij} = -\frac{t_{ij}}{t_{ii}},\ i \neq j. \quad q_{ii} = 0.$$

Since state $p + 1$ is absorbing the embedded chain of X_t has transition matrix

$$\tilde{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{q} \\ \boldsymbol{0} & 1 \end{pmatrix}.$$

Where $\boldsymbol{q} = \boldsymbol{e} - \boldsymbol{Q}\boldsymbol{e}$ is the exit probabilities. Now, we order the states by E_+ or E_0 and write

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}^{++} & \boldsymbol{Q}^{+0} \\ \boldsymbol{Q}^{0+} & \boldsymbol{Q}^{00} \end{pmatrix} \tag{4.19}$$

likewise for the exit probability vector $\boldsymbol{q} = (\boldsymbol{q}^+, \boldsymbol{q}^0)'$.
Since only time spent in E_+ contributes to the value of Y we look at the subset of states $E^* = E_+ \cup \{p + 1\}$. Define the sequence $\{M_n\}_{n\in\mathbb{N}}$ by

$$M_0 = \inf\{k \geq 0 : Y_k \in E^*\}$$
$$M_{n+1} = \inf\{k > M_n : Y_k \in E^*\}$$

then the Markov chain $\{X_n^*\}_{n\in\mathbb{N}}$ on E^* can be defined by $X_n^* = Y_{M_n}$. Clearly the corresponding $|E^*| \times |E^*|$ transition matrix has form

$$\boldsymbol{P}^* = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{p} \\ \boldsymbol{0} & 1 \end{pmatrix}.$$

To calculate the entries of $\boldsymbol{P} = \{p_{ij}\}_{i,j\in S^+}$ we note that jumps between $i, j \in E_+$ can happen either by a direct jump of the embedded chain governed by $\boldsymbol{Q}^{++}$ or the embedded chain jumps out of E_+ to E_0 and remain there for $n \geq 0$ steps before returning hence we get:

$$p_{ij} = q_{ij}^{++} + \sum_{k\in E_0} q_{ik}^{+0} \sum_{n=0}^{\infty} \sum_{l\in E_0} \left(q^{00}\right)_{kl}^{(n)} q_{lj}^{0+}.$$

Written in matrix notation:

$$\boldsymbol{P} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0} \sum_{n=0}^{\infty} (\boldsymbol{Q}^{00})^n \boldsymbol{Q}^{0+} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0} \left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} \boldsymbol{Q}^{0+}$$

where we have used Lemma **??**. Now the exit probability vector $\boldsymbol{p} = \{p_i\}_{i \in E_+}$ has a slightly different interpretation since it is not the probability that the embedded chain will jump from state $i \in E_+$ to the absorbing state, rather that state $i \in E_+$ is the last visited state in E_+. Due to the embedded chain being discrete phase-type we can write

$$\boldsymbol{p} = \boldsymbol{e} - \boldsymbol{P}\boldsymbol{e}.$$

For the initial distribution of $\{X_n^*\}_{n \in \mathbb{N}}$ which we will denote $\boldsymbol{\alpha}^* = \{\alpha_i^*\}_{i \in E_+}$ we split the initial distribution vector $\boldsymbol{\pi}$ in the same fashion as before such that $\boldsymbol{\pi} = (\boldsymbol{\pi}^+, \boldsymbol{\pi}^0)$.
Then $\boldsymbol{\alpha}^*$ can be deducted by an argument equal to before: Either the underlying Markov jump process is initiated in state $i \in E_+$, $X_0 = i$ or $X_0 \in E_0$ eventually reaching $i \in E_+$ as the first of the E_+ states.

$$\alpha_i^* = \pi_i^+ + \sum_{k \in E_0} \pi_k^0 \sum_{n=0}^{\infty} \sum_{l \in E_0} (q^{00})_{kl}^{(n)} q_{li}^{0+}$$

which, using Lemma **??**, in matrix notation reads

$$\boldsymbol{\alpha}^* = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0 \left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} \boldsymbol{Q}^{0+}.$$

Notice that even though $\boldsymbol{\pi}$ is a stochastic vector $\boldsymbol{\alpha}^*$ might not sum to 1. This is due to the fact that as long as there is a positive probability of starting in a zero reward state, the chain might get absorbed prior to having earned any reward whatsoever. The size of this zero atom is $\alpha_{d+1}^* = 1 - \boldsymbol{\alpha}^* \boldsymbol{e}$ where $d = |E_+|$.
Now, having determined the transitions of $\{X_n^*\}_{n \in \mathbb{N}}$ we construct a Markov jump process on E^* whose holding times corresponds to the reward earned. When $X_t = i$ the holding time is exponential distributed with rate $-t_{ii}$, reward is earned at rate $r(i)$ hence the total earned reward will again follow an exponential distribution with rate $-t_{ii}/r(i)$. Hence we have now described the reward earning process' holding times on E^* and since jumps are governed by $\boldsymbol{P}$ we can use the thinning Lemma 2.32 to find the $|E^*| \times |E^*|$ intensity matrix $\boldsymbol{\Lambda}^*$, of the reward earning Markov jump process, which due to the absorbing state $p+1$ has the form

$$\boldsymbol{\Lambda}^* = \begin{pmatrix} \boldsymbol{T}^* & \boldsymbol{t}^* \\ \boldsymbol{0} & 0 \end{pmatrix}$$

where $\boldsymbol{T}^* = \{t_{ij}^*\}_{i,j \in E_+}$, $\boldsymbol{t}^* = \{t_i^*\}_{i \in E_+}$ are given by

$$t_{ij}^* = -\frac{t_{ii}}{r(i)} p_{ij} \quad j \neq i, \tag{4.20}$$

$$t_{ii}^* = \frac{t_{ii}}{r(i)} (1 - p_{ii}). \tag{4.21}$$

$$t_i^* = -\frac{t_{ii}}{r(i)} p_i. \tag{4.22}$$

Lets summarize the above.

Theorem 4.33 (Rewards)**.** *The random variable Y of Definition 4.32 is a possibly defect phase-type, the defect being an atom at zero of size $\alpha^*_{d+1} = 1 - \boldsymbol{\alpha}^* \boldsymbol{e}$. The phase-type distribution has representation $(\boldsymbol{\alpha}^*, \boldsymbol{T}^*)$ with dimension $d = |E_+|$ and*

$$\boldsymbol{\alpha}^* = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0 \left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} \boldsymbol{Q}^{0+} \tag{4.23}$$

and $\boldsymbol{T}^*$ and $\boldsymbol{t}^*$ are given by equation (5.11), (5.12), (4.22)*. Where*

$$\boldsymbol{P} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0} \left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} \boldsymbol{Q}^{0+} \quad \textit{and} \quad \boldsymbol{p} = \boldsymbol{e} - \boldsymbol{P}\boldsymbol{e}. \tag{4.24}$$

Remark 4. 1) In chapter 5 theorem 5.6 we will see an alternative way of describing the reward collection.

2) In the case of only positive rewards $r(i) > 0$ for all $i \in E$ the problem simplifies considerably and it is only a question of scaling the holding times in each state by the corresponding entry in $\boldsymbol{r}$ to obtain the new $\boldsymbol{T}^*$ sub-intensity matrix. That is if $\boldsymbol{r} > 0$ then

$$\boldsymbol{T}^* = \begin{pmatrix} \frac{t_{11}}{r(1)} & \frac{t_{12}}{r(1)} & \cdots & \frac{t_{1p}}{r(1)} \\ \frac{t_{21}}{r(2)} & \frac{t_{22}}{r(2)} & \cdots & \frac{t_{2p}}{r(2)} \\ \vdots & \vdots & \vdots\vdots & \vdots \\ \frac{t_{p1}}{r(p)} & \frac{t_{p2}}{r(p)} & \cdots & \frac{t_{pp}}{r(p)} \end{pmatrix} = \boldsymbol{\Delta}(\boldsymbol{r})^{-1} \boldsymbol{T}$$

Where $\boldsymbol{\Delta}(\boldsymbol{r}) = \operatorname{diag}(r(1), ..., r(p))$ i.e. a diagonal matrix with the i'th diagonal element equal to the i'th element of the vector $\boldsymbol{r}$. From this we get that scaling of a phase-type by a scalar $c > 0$ is again of phase-type with sub transition matrix scaled by $\frac{1}{c}$ i.e. for $\tau \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ then

$$c\tau \sim \mathrm{PH}\left(\boldsymbol{\pi}, \frac{1}{c}\boldsymbol{T}\right)$$

in concordance with standard transformation theory of random variables with density.

3) It is seen that if $r(\cdot) = 1$ then $Y = \tau$ and if $r(k) = \delta_{i-k}$ then Y is the total time spent in state i prior to absorption. The latter we will formalise as a corollary.

Corollary 4.34. *The total time spent in state i prior to absorption of the Markov jump process underlying $PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ is a mixture of an atom at zero and an exponential distribution.*
The exponential has rate $-t_{ii}(1 - \boldsymbol{P})$ where $\boldsymbol{P}$ is given by (4.24) *and the size of the atom is $\alpha^*_2 = 1 - \boldsymbol{\alpha}^*$ where $\boldsymbol{\alpha}^*$ is given by* (4.23)

Proof. As stated above if one set $r(i) = 1$, and 0 otherwise, the total reward earned is exactly the time spent in state i prior to absorption.
Using theorem 4.33 we have $E_+ = \{i\}$ and $E_0 = \{1, 2, ..., p\} \setminus \{i\}$. Hence $d = |E_+| = 1$ as we have seen earlier, by construction a phase-type with only one phase is exponential i.e.

$$\mathrm{PH}_1(1, \{-\lambda\}) \overset{d}{=} exp(\lambda).$$

Also, with $d = 1$ the matrices and vectors of theorem 4.33 become real numbers, regardless we will denote them by boldface to maintain consistency. The embedded chain has transition matrix

$$\tilde{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{q} \\ \boldsymbol{0} & 1 \end{pmatrix}$$

with

$$\boldsymbol{Q} = \left(\begin{array}{c|c} \boldsymbol{Q}^{++} & \boldsymbol{Q}^{+0} \\ \hline \boldsymbol{Q}^{0+} & \boldsymbol{Q}^{00} \end{array}\right)$$

then $\boldsymbol{Q}^{++}$ is a real number in particular it is 0 since the embedded chain does not jump to itself. $\boldsymbol{Q}^{+0}$ is a $1 \times (p-1)$ row vector describing the transition probabilities of jumping from state i to any of the zero-reward earning states. $\boldsymbol{Q}^{0+}$ is a $(p-1) \times 1$ column vector describing transitions from the zero-reward states to state i and $\boldsymbol{Q}^{00}$ is a $(p-1) \times (p-1)$ matrix describing jumps between zero-reward states.

Following the lines of Theorem 4.33 we then get

$$\boldsymbol{\alpha}^* = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0 \left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1} \boldsymbol{Q}^{0+} \in [0,1].$$

Describing the probability of either starting in the reward-earning state i or starting in one of the zero-reward earning states and eventually reach state i before absorption. Hence $\alpha^*_{d+1} = 1 - \boldsymbol{\alpha}^*$ is the size of the atom at zero. Now for the 1×1 sub-transition matrix $\boldsymbol{T}^*$ of the possibly defect phase-type.

$$t^*_{11} = \frac{t_{ii}}{1}(1 - p_{11})$$

where p_{11} is the entry of the 1×1 matrix $\boldsymbol{P}$ of (4.24). □

Example 4.35. Let

$$\tau \sim \mathrm{PH}_3 \left(\left(\frac{2}{10}, \frac{3}{10}, \frac{5}{10} \right), \begin{pmatrix} -4 & 2 & 1 \\ 3 & -6 & 1 \\ 3 & 1 & -7 \end{pmatrix} \right).$$

Lets look at the distribution of the time spent in state 2 prior to absorption which we will denote $R_2(\tau)$, then

$$\boldsymbol{Q} = \left(\begin{array}{c|c} \boldsymbol{Q}^{++} & \boldsymbol{Q}^{+0} \\ \hline \boldsymbol{Q}^{0+} & \boldsymbol{Q}^{00} \end{array}\right) = \begin{array}{c} \\ 2 \\ 1 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 1 & 3 \end{array} \\ \begin{pmatrix} 0 & 3/6 & 1/6 \\ 2/4 & 0 & 1/4 \\ 1/7 & 3/7 & 0 \end{pmatrix} \end{array}$$

$$\left(\boldsymbol{\pi}^+, \boldsymbol{\pi}^0\right) = \left(\begin{array}{c|cc} \frac{3}{10} & \frac{2}{10} & \frac{5}{10} \end{array} \right)$$

and we get

$$\boldsymbol{\alpha} = \frac{3}{10} + \left(\frac{2}{10}, \frac{5}{10}\right) \left(\boldsymbol{I} - \begin{pmatrix} 0 & \frac{1}{4} \\ \frac{3}{7} & 0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \frac{2}{4} \\ \frac{1}{7} \end{pmatrix} = \frac{31}{50}$$

$$\boldsymbol{P} = 0 + \left(\frac{3}{6}, \frac{1}{6}\right) \left(\boldsymbol{I} - \begin{pmatrix} 0 & \frac{1}{4} \\ \frac{3}{7} & 0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \frac{2}{4} \\ \frac{1}{7} \end{pmatrix} = \frac{11}{30}.$$

Hence the rate of the exponential is $\lambda = 6(1 - \frac{11}{30}) = 19/5$ and the size of the atom at zero is $1 - \boldsymbol{\alpha} = 19/50$. From which we get $R_2(\tau) \sim \mathrm{PH}\left(\frac{31}{50}, \{-\frac{19}{5}\}\right)$ the mixture is given by the CDF

$$P(R_2(\tau) \leq t) = \begin{cases} 0, \, t < 0 \\ 1 - \boldsymbol{\alpha}, \, t = 0 \\ 1 - \boldsymbol{\alpha} + \boldsymbol{\alpha}(1 - e^{-\lambda t}), \, t > 0 \end{cases}$$

The mean of the mixture is $(1 - \boldsymbol{\alpha}) \cdot 0 + \boldsymbol{\alpha} \frac{1}{\lambda} = 31/190$ which is in accordance with the $\boldsymbol{\pi}$- weighted 2nd column of the Green matrix

$$\boldsymbol{\pi} \left(-\boldsymbol{T}\right)^{-1} \boldsymbol{e}_2 = 31/190.$$

In Chapter 6 we will return to the problem of estimating the time spend in a state $i \in E$ but conditioned on $\tau = x$. •

The reward transformation of the phase-type is the cornerstone of the multivariate phase-type as we will explore in the next chapter.

Chapter 5

Multivariate Phase-Type Distributions

A natural extension of the phase-type is the construction of multivariate phase-type distributions i.e. vectors $(Y_1, ..., Y_n)$ with the property of the marginals being phase-type, $Y_i \sim PH$. One of the first constructions of the above was given by [Assaf et al., 1984], which defined the class MPH by partitioning the state space E into partitions (Γ_i, Γ_i^c) for $i = 1, 2, ..., n$ with $\Gamma_i \cup \Gamma^c{}_i = E$ and the further assumption that Γ_i is stochastically closed i.e. once we enter Γ_i we never leave it, from which we define the n first hitting times $Y_i = \inf\{t > 0 : X_t \in \Gamma_i\}$. Without loss of generality we let $\cap_{i=1}^n \Gamma_i = \{p+1\}$. It is now evident that Y_i is of phase-type and that for $\tau = \inf\{t > 0 : X_t = p+1\}$ we have $\tau = \max_i Y_i$.
Besides heavy notation, a major drawbacks of the MPH construction is the fact that once we enter Γ_i we cannot re-enter Γ_i^c which puts non-trivial restrictions on the intensity matrix of the underlying Markov jump process.

Later [Kulkarni, 1989] developed the class MPH* which builds on the reward extension of the univariate phase-type discussed in section 4.5. Here it was shown that MPH $\subset$ MPH* while still retaining nice probabilistic interpretations and closed form formulae for many quantities of interest. Later, [Bladt and Nielsen, 2010] contributed to the theory of the MPH* class and further extended the class of Matrix-exponential distributions to multivariate matrix-exponential distributions denoted MVME as well as extending the MPH* class to the class denoted MVPH given by all random vectors $\boldsymbol{Y}$ such that $\langle \boldsymbol{Y}, \boldsymbol{a} \rangle$ has a univariate phase-type distribution for $\boldsymbol{a} \neq \boldsymbol{0}$. It was argued that the MPH* distributions belong to MVPH. However, it is still an open question whether the two classes are identical.
Recall the set-up of section 4.5. For $\tau \sim \text{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$ constructed from the underlying Markov jump process $\{X_t\}_{t \geq 0}$ we gather $n \geq 1$ reward functions $r_j(\cdot)$ in a $p \times n$ reward matrix $\boldsymbol{R}$ by $R_{ij} = r_j(i)$ and have the following definition.

Definition 5.1 (MPH*)**.** *Let*

$$Y_j = \int_0^\tau r_j(X_t)dt, \quad j = 1, 2, ..., n.$$

Then the random vector $(Y_1, ..., Y_n)$ is said to be multivariate phase-type. With representation $\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R}$ written $Y \sim MPH_p^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$. Usually we will not include p when the dimension is not of importance or implicitly known from the other quantities.

If $n = 1$ then $\boldsymbol{R}$ is a p-column vector and the above is equal to transformation by rewards under the reward vector $\boldsymbol{r} = \boldsymbol{R}$. Note that if we define Z_{ij} the duration of the jth visit to state i and let $N_i \geq 0$ be the total number of visits to state i.
Then with $Z_i = \sum_{j=1}^{N_i} Z_{ij} = \int_0^\tau 1\{X_u = i\}du$ the total time spent in state i an equivalent

definition of Y_j is

$$Y_j = \sum_{i=1}^{p} R_{ij} Z_i, \quad j = 1, ..., n$$

Note that if $n = p$ and $R_{ij} = \delta_{i-j}$ then Y_j describes the time spent in state $j = 1, 2, ..., p$ of the underlying Markov jump process, which we from Corollary 4.34 know is exponentially distributed possibly with an atom at zero but not necessarily independent, further $\sum_{j=1}^{p} Y_j = \tau$ where $\tau \sim \mathrm{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$.

Lemma 5.2. *Let $\boldsymbol{Y} \sim MPH^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ then for $\boldsymbol{A} = \{a_{ij}\}$ an $n \times k$ matrix of non-negative real numbers we have that $\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{A} \sim MPH^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R}\boldsymbol{A})$.*

Proof. By definition we have

$$X_i = \sum_{j=1}^{n} Y_j a_{ji} = \sum_{j=1}^{n} \left(\sum_{l=1}^{p} R_{lj} Z_l \right) a_{ji} = \sum_{l=1}^{p} \left(\sum_{j=1}^{n} R_{lj} a_{ji} \right) Z_l = \sum_{l=1}^{p} (\boldsymbol{R}\boldsymbol{A})_{li} Z_l$$

from which we see that $\boldsymbol{X} \sim \mathrm{MPH}^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R}\boldsymbol{A})$. □

Let $\langle \cdot, \cdot \rangle$ denote the usual inner product i.e. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}'\boldsymbol{y}$.

Theorem 5.3. *Let $Y \sim MPH^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ then there exists some $\theta_0 > 0$ such that the moment-generating function of $\boldsymbol{Y}$ exists and is given by*

$$H(\boldsymbol{\theta}) = \mathbb{E}\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} \right) = \boldsymbol{\alpha} \left(-\Delta(\boldsymbol{R}\boldsymbol{\theta}) - \boldsymbol{S} \right)^{-1} \boldsymbol{s} = \boldsymbol{\alpha} \left(\boldsymbol{I} - \boldsymbol{U}\, \Delta(\boldsymbol{R}\boldsymbol{\theta}) \right)^{-1} \boldsymbol{e}. \tag{5.1}$$

For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ whenever $\theta_i < \theta_0$ and as usual $\boldsymbol{U} = -\boldsymbol{S}^{-1}$ and $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$.

Proof. We start by noticing that, conditioning on the initial state of the underlying Markov jump process $\{X_t\}_{t \geq 0}$ we can write

$$\mathbb{E}\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} \right) = \sum_{i=1}^{p} \mathbb{E}\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} | X_0 = i \right) \mathbb{P}(X_0 = i).$$

From which we define

$$H_i(\boldsymbol{\theta}) = \mathbb{E}\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} | X_0 = i \right) = \mathbb{E}_i\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} \right).$$

Now, given $X_0 = i$ then Z_{i1}, the time spent in state i during the 1st visit, will be a.s. positive since it is exponentially distributed with rate $-s_{ii}$, in particular, the time of the first jump away from state i is a stopping time for $\{X_t\}_{t \geq 0}$. We can then decompose $\boldsymbol{Y} = \boldsymbol{Y}_{i1} + \boldsymbol{Y}_r$ where $\boldsymbol{Y}_{i1}$ is the accumulated reward from the sojourn time in state i under the different rewards given by the ith row of $\boldsymbol{R}$ which we denote $\boldsymbol{R}_{i\cdot}$ hence $\boldsymbol{Y}_{i1} = Z_{i1}\boldsymbol{R}_{i\cdot}$ and $\boldsymbol{Y}_r$ is then the remaining accumulated reward after the first holding time. Then

$$\mathbb{E}_i\left(e^{\langle \boldsymbol{Y}, \boldsymbol{\theta} \rangle} \right) = \mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_{i1} + \boldsymbol{Y}_r, \boldsymbol{\theta} \rangle} \right) = \mathbb{E}_i\left(\mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_{i1} + \boldsymbol{Y}_r, \boldsymbol{\theta} \rangle} \right) | \mathscr{F}_{Z_{i1}} \right) = \mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_{i1}, \boldsymbol{\theta} \rangle} \right) \mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_r, \boldsymbol{\theta} \rangle} \right). \tag{5.2}$$

Since Z_{i1} is exponential with rate $-s_{ii}$ we have

$$\mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_{i1}, \boldsymbol{\theta} \rangle} \right) = \mathbb{E}_i\left(e^{Z_{i1} \boldsymbol{R}_{i1} \boldsymbol{\theta}} \right) = \frac{-s_{ii}}{-s_{ii} - \boldsymbol{R}_{i\cdot}\boldsymbol{\theta}} = \frac{1}{1 + \boldsymbol{R}_{i\cdot}\boldsymbol{\theta} s_{ii}^{-1}}, \quad \text{for } \boldsymbol{R}_{i\cdot}\boldsymbol{\theta} < -s_{ii}. \tag{5.3}$$

Now, for $\boldsymbol{Y}_r$ conditioning on the next state after the initial time in state i we have with probability $\frac{s_i}{-s_{ii}}$ that $\mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_r, \boldsymbol{\theta} \rangle} \right) = 1$ since $\boldsymbol{Y}_r = \boldsymbol{0}$. On the other hand the underlying Markov jump process may jump to state $j \neq i$ with probability $\frac{s_{ij}}{-s_{ii}}$ from which we get

$$\mathbb{E}_i\left(e^{\langle \boldsymbol{Y}_r, \boldsymbol{\theta}\rangle}\right) = \frac{s_i}{-s_{ii}} + \sum_{j\neq i} \frac{s_{ij}}{-s_{ii}} H_j(\boldsymbol{\theta})$$

from which we can write, using (5.2)

$$\begin{aligned} H_i(\boldsymbol{\theta}) &= \frac{1}{1 + \boldsymbol{R}_{i\cdot}\boldsymbol{\theta} s_{ii}^{-1}} \left(\frac{s_i}{-s_{ii}} + \sum_{j\neq i} \frac{s_{ij}}{-s_{ii}} H_j(\boldsymbol{\theta}) \right) \iff \\ H_i(\boldsymbol{\theta}) + H_i(\boldsymbol{\theta}) \boldsymbol{R}_{i\cdot}\boldsymbol{\theta} s_{ii}^{-1} &= \frac{s_i}{-s_{ii}} + \sum_{j\neq i} \frac{s_{ij}}{-s_{ii}} H_j(\boldsymbol{\theta})) \iff \\ -\boldsymbol{R}_{i\cdot}\boldsymbol{\theta} H_i(\boldsymbol{\theta}) &= s_i + \sum_{j=1}^{p} s_{ij} H_j(\boldsymbol{\theta}). \end{aligned}$$

For all $i = 1, \ldots, p$. Now, introducing $\boldsymbol{H}(\boldsymbol{\theta}) = (H_1(\boldsymbol{\theta}), \ldots, H_p(\boldsymbol{\theta}))'$ the last equality can be written on matrix form as

$$-\Delta(\boldsymbol{R}\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{s} + \boldsymbol{S}\boldsymbol{H}(\boldsymbol{\theta})$$

i.e.

$$-\left(\Delta(\boldsymbol{R}\boldsymbol{\theta}) + \boldsymbol{S}\right)\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{s}.$$

Since $\boldsymbol{S}$ is a sub-intensity matrix of a phase-type distribution we know from theorem 4.10 that it is invertible, from which we have that there exists a neighbourhood about $\boldsymbol{0}$ such that if $\boldsymbol{\theta}$ is contained in the neighbourhood the matrix $(\Delta(\boldsymbol{R}\boldsymbol{\theta}) + \boldsymbol{S})$ is invertible as well and (5.3) is fulfilled. In that case we can write

$$\boldsymbol{H}(\boldsymbol{\theta}) = -\left(\Delta(\boldsymbol{R}\boldsymbol{\theta}) + \boldsymbol{S}\right)^{-1}\boldsymbol{s}$$

and since

$$H(\boldsymbol{\theta}) = \boldsymbol{\alpha}\boldsymbol{H}(\boldsymbol{\theta})$$

the theorem follows. The second equality of the theorem simply follows from $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ and basic matrix operations. □

Note that the proof of Lemma 5.2 can also be done using the moment-generating function, since

$$\mathbb{E}\left(e^{\langle \boldsymbol{Y}\boldsymbol{A}, \boldsymbol{\theta}\rangle}\right) = \mathbb{E}\left(e^{\langle \boldsymbol{Y}, \boldsymbol{A}\boldsymbol{\theta}\rangle}\right) = \boldsymbol{\alpha}\left(-\Delta(\boldsymbol{R}\boldsymbol{A}\boldsymbol{\theta}) - \boldsymbol{S}\right)^{-1}\boldsymbol{s}$$

is the moment-generating function for the MPH* with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R}\boldsymbol{A})$.

The main work horse in our investigation of the class of MPH* distributions will be the moment-generating function $H(\boldsymbol{\theta})$ as in the next theorem, where the cross moments of an MPH* distribution is characterized.

Theorem 5.4. *For $\boldsymbol{Y} \sim MPH_p^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $h_j \in \mathbb{N}$ for $j = 1, \ldots, p$ with $h = \sum_{j=1}^{n} h_j$, then the cross moments are given by*

$$\mathbb{E}\left(\prod_{j=1}^{n} Y_j^{h_j}\right) = \boldsymbol{\alpha} \sum_{l=1}^{h!} \left(\prod_{i=1}^{h} \boldsymbol{U}\, \Delta(\boldsymbol{R}_{\cdot\sigma_l(i)}) \right) \boldsymbol{e}. \tag{5.4}$$

Where $\boldsymbol{R}_{\cdot k}$ is the kth column of $\boldsymbol{R}$ and $\sigma_l(i)$ is the value at the ith index of values among $1, \ldots, n$ of the lth permutation of the ordered set of $h!$ h-tuples $(i_1, i_2, \ldots, i_h)$ where each entry corresponds to a value $j \in \{1, \ldots, n\}$ and entries corresponding to j appears h_j times for $j = 1, \ldots, n$.

Proof. We wish to evaluate the partial derivative of $H(\boldsymbol{\theta})$ with respect to θ_i i.e.

$$\frac{\partial H(\boldsymbol{\theta})}{\partial \theta_i} = \frac{\partial \boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta}))^{-1}\boldsymbol{e}}{\partial \theta_i}.$$

To this end we start by looking at

$$\frac{\partial \left((\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta}))^{-1}\right)}{\partial \theta_i}.$$

Since $\boldsymbol{U}\Delta(\boldsymbol{R\theta})$ is invertible we can write

$$\frac{\partial \left((\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta}))^{-1}\right)}{\partial \theta_i} = \frac{\partial \left(\sum_{l=0}^{\infty}(\boldsymbol{U}\Delta(\boldsymbol{R\theta})^l\right)}{\partial \theta_i} = \sum_{l=1}^{\infty}\frac{\partial \left(\boldsymbol{U}\Delta(\boldsymbol{R\theta})^l\right)}{\partial \theta_i}.$$

Recall for two matrices $\boldsymbol{U} = \boldsymbol{U}(x), \boldsymbol{V} = \boldsymbol{V}(x)$ which depend on the scalar x then, if the product is well defined

$$\frac{\partial \boldsymbol{U}(x)\boldsymbol{V}(x)}{\partial x} = \frac{\partial \boldsymbol{U}(x)}{\partial x}\boldsymbol{V}(x) + \boldsymbol{U}(x)\frac{\partial \boldsymbol{V}(x)}{\partial x}. \tag{5.5}$$

Using this repetitively we have that

$$\frac{\partial \boldsymbol{U}(x)^n}{\partial x} = \sum_{i=0}^{n-1}\boldsymbol{U}(x)^i\frac{\partial \boldsymbol{U}(x)}{\partial x}\boldsymbol{U}(x)^{n-i-1}.$$

From this we get

$$\begin{aligned}
\frac{\partial \left((\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta}))^{-1}\right)}{\partial \theta_i} &= \sum_{l=1}^{\infty}\sum_{k=0}^{l-1}\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^k\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^{l-k-1} \\
&= \sum_{k=0}^{\infty}\sum_{l=i+1}^{\infty}\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^k\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^{l-k-1} \\
&= \sum_{k=0}^{\infty}\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^k\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)\sum_{l=0}^{\infty}\left(\boldsymbol{U}\,\Delta(\boldsymbol{R\theta})\right)^l \\
&= \left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}
\end{aligned}$$

and the result follows from the properties of the moment-generating function:

$$\mathbb{E}\left(\prod_{j=1}^{n}Y_j^{h_j}\right) = \left.\frac{\partial^h H(\boldsymbol{\theta})}{\partial\theta_1^{h_1}\cdots\partial\theta_n^{h_n}}\right|_{\boldsymbol{\theta}=\boldsymbol{0}}$$

and induction. To see this notice that using the first result for the partial derivative and (5.5) we have that

$$\begin{aligned}
\frac{\partial^2 \left((\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta}))^{-1}\right)}{\partial\theta_j\partial\theta_i} &= \frac{\partial}{\partial\theta_j}\left[\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\right] \\
&= \left[\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot j})\right)\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\right]\boldsymbol{U}\Delta\left(\boldsymbol{R}_{\cdot i}\right)\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1} \\
&+ \left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\left[\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot j})\right)\left(\boldsymbol{I} - \boldsymbol{U}\Delta(\boldsymbol{R\theta})\right)^{-1}\right]
\end{aligned}$$

which evaluated at $\boldsymbol{\theta} = \boldsymbol{0}$ yields

$$\Delta(\boldsymbol{R}_{\cdot j})\Delta(\boldsymbol{R}_{\cdot i}) + \Delta(\boldsymbol{R}_{\cdot i})\Delta(\boldsymbol{R}_{\cdot j})$$

□

To be clear on the $\sigma_l(i)$ notation: Say $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ is MPH* distributed and we wish to calculate $\mathbb{E}(Y_1 Y_2^2)$. This corresponds to $h_1 = 1, h_2 = 2, h_3 = 0$ and hence we look at 3-tuples (i_1, i_2, i_3) where i_1 corresponds to 1 and i_2, i_3 corresponds to 2 (the specific order here is irrelevant). The 3! permutations are given by

$$(i_1, i_2, i_3), (i_1, i_3, i_2), (i_2, i_1, i_3), (i_2, i_3, i_1), (i_3, i_1, i_2), (i_3, i_2, i_1)$$

corresponding to

$$(1, 2, 2), (1, 2, 2), (2, 1, 2), (2, 2, 1), (2, 1, 2), (2, 2, 1).$$

Since the 4th permutation $(2, 2, 1)$ has the value 2 at the 1st position we have $\sigma_4(1) = 2$.

Corollary 5.5. *For $\boldsymbol{Y} \sim$ MPH$^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ then*

$$\begin{aligned}
\mathbb{E}(Y_i) &= \boldsymbol{\alpha}\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\boldsymbol{e} = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{R}_{\cdot i} \\
\mathbb{E}(Y_i^2) &= 2\boldsymbol{\alpha}\left(\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right)^2\boldsymbol{e} = 2\boldsymbol{\alpha}\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\boldsymbol{U}\boldsymbol{R}_{\cdot i} \\
\mathbb{E}(Y_i Y_j) &= \boldsymbol{\alpha}\left[\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot j}) + \boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot j})\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\right]\boldsymbol{e} = \boldsymbol{\alpha}\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot i})\boldsymbol{U}\boldsymbol{R}_{\cdot j} + \boldsymbol{\alpha}\boldsymbol{U}\Delta(\boldsymbol{R}_{\cdot j})\boldsymbol{U}\boldsymbol{R}_{\cdot i}
\end{aligned}$$

Next, we look at linear combinations of the marginals i.e. for a n-column vector $\boldsymbol{w} \geq \boldsymbol{0}$ we look at the quantity $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ for $\boldsymbol{Y} \sim$ MPH$^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$. To this end as usual let $E = \{1, \ldots, p\}$ be the state space of the Markov jump process underlying $\boldsymbol{Y}$. We partition E into E_+ and E_0 by the relation $i \in E_+$ if $(\boldsymbol{R}\boldsymbol{w})_i > 0$ and $i \in E_0$ if $(\boldsymbol{R}\boldsymbol{w})_i = 0$. We may then reorder the state space E and also the initial vector and sub-intensity matrix as

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_+, \boldsymbol{\alpha}_0) \tag{5.6}$$

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{++} & \boldsymbol{S}_{+0} \\ \boldsymbol{S}_{0+} & \boldsymbol{S}_{00} \end{pmatrix} \tag{5.7}$$

Where $\boldsymbol{\alpha}_+ = \{\alpha_i\}_{i \in E_+}$, $\boldsymbol{\alpha}_0 = \{\alpha_i\}_{i \in E_0}$ and
$\boldsymbol{S}_{++} = \{s_{ij}\}_{i,j \in E_+}$, $\boldsymbol{S}_{+0} = \{s_{ij}\}_{i \in E_+, j \in E_0}$, $\boldsymbol{S}_{0+} = \{s_{ij}\}_{i \in E_0, j \in E_+}$, $\boldsymbol{S}_{00} = \{s_{ij}\}_{i,j \in E_0}$.

We now prove that $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ is a, possibly defect, phase-type distribution with representation and atom size to be deducted in the following. As earlier let $Z_i = \int_0^\tau 1\{X_u = i\} du$ be the total time spent in state i by the underlying Markov jump process. We start by noting that

$$\langle \boldsymbol{Y}, \boldsymbol{w} \rangle = \sum_{j=1}^{n} Y_j w_j = \sum_{j=1}^{n} \sum_{i=1}^{p} R_{ij} Z_i w_j = \sum_{i=1}^{p} Z_i (\boldsymbol{R}\boldsymbol{w})_i$$

Since $Z_i = \sum_{k=1}^{N_i} Z_{ik}$ where N_i is the total number of visits to state i by the underlying Markov jump process and $Z_{ik} \sim exp(-s_{ii})$. Introducing $\tilde{Z}_{ik} := Z_i (\boldsymbol{R}\boldsymbol{w})_i \sim exp(-s_{ii}/(\boldsymbol{R}\boldsymbol{w})_i)$ if $i \in E_+$ and degenerate at zero if $i \in E_0$ we have that

$$\langle \boldsymbol{Y}, \boldsymbol{w} \rangle = \sum_{i=1}^{p} \sum_{k=1}^{N_i} \tilde{Z}_{ik} = \sum_{i=1}^{p} \tilde{Z}_i.$$

From this we see that $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ is the time until absorption of a Markov jump process $\{\tilde{X}_t\}_{t \geq 0}$ on E_+ since only time spent in E_+ is counted. Similar to the arguments regarding transformation by rewards of the standard phase-type, we now compute the transitions of $\{\tilde{X}_t\}_{t \geq 0}$.

Recall that $\{X_t\}_{t \geq 0}$ is the underlying Markov jump process on E with sub-intensity matrix $\boldsymbol{S}$. A transition from $i \in E_+$ to $j \in E_+$ can occur either by a direct transition of $\{X_t\}_{t \geq 0}$ governed by $\boldsymbol{S}_{++}$ where each row needs to be scaled by the corresponding entry in $(\boldsymbol{R}\boldsymbol{w})_+$. Second, the underlying Markov jump process may jump to some state $j \in E_0$, stay in E_0 until a new state in E_+ is entered, which evidently is phase-type distributed with sub-intensity matrix $\boldsymbol{S}_{00}$ hence we that the sub-intensity matrix, $\boldsymbol{S}_w$ of $\{\tilde{X}_t\}_{t \geq 0}$ is given by

$$\begin{aligned}\boldsymbol{S}_w &= \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{++} + \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{+0}\int_0^\infty e^{\boldsymbol{S}_{00}u}du\boldsymbol{S}_{0+} \\ &= \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{++} + \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}. \end{aligned} \tag{5.8}$$

Regarding the initial vector $\boldsymbol{\alpha}_w$, we note that $\{\tilde{X}_t\}_{t\geq 0}$ can initiate in $i \in E_+$ either simply by the underlying process $X_0 = i$ which happens with probability according to $\boldsymbol{\alpha}_+$, or secondly, if $X_0 = j$ for some $j \in E_0$ and remains in E_0 before entering E_i hence, written compactly in matrix form we have

$$\boldsymbol{\alpha}_w = \boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0\int_0^\infty e^{\boldsymbol{S}_{00}u}du\boldsymbol{S}_{0+} = \boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0\left(-\boldsymbol{S}_{00}^{-1}\right)\boldsymbol{S}_{0+}. \tag{5.9}$$

Notice that $\boldsymbol{\alpha}_w\boldsymbol{e}$ may be strictly less that 1 corresponding to the case of starting in E_0 and never entering E_+ before absorption. Hence $(\alpha_w)_{|E_+|+1} = 1 - \boldsymbol{\alpha}_w\boldsymbol{e}$ will be the size of the atom at zero of the phase-type distribution of $\langle \boldsymbol{Y}, \boldsymbol{w}\rangle$. We have now proved the following

Theorem 5.6. *Let $\boldsymbol{Y} \sim MPH^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $\boldsymbol{w} \geq 0$. Assume that $\boldsymbol{\alpha}$ and $\boldsymbol{S}$ of the underlying Markov jump process is partitioned according to* (5.6) *and* (5.7) *induced by $\boldsymbol{Rw}$. Then the distribution of $\langle \boldsymbol{Y}, \boldsymbol{w}\rangle$ is give by a possible defect phase-type distribution with an atom at zero of size $1 - \boldsymbol{\alpha}_w\boldsymbol{e}$ and representation $(\boldsymbol{\alpha}_w, \boldsymbol{S}_w)$ where*

$$\begin{aligned}\boldsymbol{\alpha}_w &= \boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0\left(-\boldsymbol{S}_{00}^{-1}\right)\boldsymbol{S}_{0+}. \\ \boldsymbol{S}_w &= \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{++} + \Delta((\boldsymbol{Rw})_+)^{-1}\boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}.\end{aligned}$$

Remark 5. It should be obvious that if we let $\boldsymbol{w} = \boldsymbol{e}_i$ then $\langle \boldsymbol{Y}, \boldsymbol{w}\rangle = Y_i$ which as we have just seen is a possibly defect phase-type distribution with representation given in Theorem 5.6 but we can also view it as a special case of transformation by rewards under reward vector $\boldsymbol{r} = \boldsymbol{Rw} = \boldsymbol{R}_{\cdot i}$ that is, the ith column of $\boldsymbol{R}$. Recall that the result of transformation by rewards also followed the same idea of partitioning the state space according to zero or positive reward, but in in the theorem of reward transformations, 4.33, the proof was build upon the embedded Markov chain, which led to the expression

$$\boldsymbol{\alpha}^* = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+} \tag{5.10}$$

for the initial distribution of the resulting phase-type distribution. Here $\boldsymbol{Q}^{00}$ was the transition matrix of the embedded Markov chain restricted to zero-reward states.
To see that this is indeed equal to $\boldsymbol{\alpha}_w$ of (5.9) simply note that $\boldsymbol{Q}^{00} = \left(-\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)^{-1}\boldsymbol{S}^{00} + \boldsymbol{I}$ where $\bullet$ is the entry-wise matrix product also known as the Schur (or Hadamard) product. Likewise $\boldsymbol{Q}^{0+} = (-\boldsymbol{S}^{00} \bullet \boldsymbol{I})^{-1}\boldsymbol{S}^{0+}$. Then

$$\begin{aligned}\boldsymbol{\pi}^+ + \boldsymbol{\pi}^0\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+} &= \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0\left(\boldsymbol{I} + \left(\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)^{-1}\boldsymbol{S}^{00} - \boldsymbol{I}\right)^{-1}\left(\left(-\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)^{-1}\boldsymbol{S}^{0+}\right) \\ &= \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0(-\boldsymbol{S}^{00})^{-1}\left(\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)\left(\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)^{-1}\boldsymbol{S}^{0+} \\ &= \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0(-\boldsymbol{S}^{00})^{-1}\boldsymbol{S}^{0+}.\end{aligned}$$

Which, save the notational differences, is equal to the expression $\boldsymbol{\alpha}_w$ from the MPH* set-up. Secondly, to see that the representation for the sub-intensity matrix, $\boldsymbol{T}^*$ in the transformation by rewards section, recall that with

$$\boldsymbol{P} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+}$$

then the sub-intensity matrix $\boldsymbol{T}^*$ of the resulting phase-type from transformation under reward vector $\boldsymbol{r}$ was given by

$$t^*_{ij} = -\frac{t_{ii}}{r(i)}p_{ij} \quad j \neq i, \tag{5.11}$$

$$t^*_{ii} = \frac{t_{ii}}{r(i)}(1-p_{ii}) = -\frac{t_{ii}}{r(i)}p_{ii} + \frac{t_{ii}}{r(i)} \tag{5.12}$$

for $i, j \in E_+$. Written in matrix notation

$$\boldsymbol{T}^* = (-\boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1})\boldsymbol{P} + \boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1}$$

where $\boldsymbol{r}_+ = \{r_i\}_{i\in E_+}$. Then inserting the expression for $\boldsymbol{P}$, and using that diagonal matrices commute we get

$$\begin{aligned}
\boldsymbol{T}^* &= \left(-\boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1}\right)\left[\boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left(\boldsymbol{I} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+}\right] + \boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1} \\
&= \left(-\boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1}\right)\left[(-\boldsymbol{S}^{++} \bullet \boldsymbol{I} + \boldsymbol{I}) + (-\boldsymbol{S}^{++} \bullet \boldsymbol{I})\boldsymbol{S}^{+0}\left(\boldsymbol{I} - \left(-\boldsymbol{S}^{00} \bullet \boldsymbol{I}\right)^{-1}\boldsymbol{S}^{00} - \boldsymbol{I}\right)^{-1}(-\boldsymbol{S}^{00} \bullet \boldsymbol{I})^{-1}\right] \\
&\quad + \boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1} \\
&= \left(-\boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1}\right)\left[(-\boldsymbol{S}^{++} \bullet \boldsymbol{I})^{-1}\boldsymbol{S}^{++} + \boldsymbol{I} + (-\boldsymbol{S}^{++} \bullet \boldsymbol{I})^{-1}\boldsymbol{S}^{+0}\left(-\boldsymbol{S}^{00}\right)^{-1}\boldsymbol{S}^{0+}\right] + \boldsymbol{S}^{++} \bullet \boldsymbol{I}\Delta(\boldsymbol{r}_+)^{-1} \\
&= \Delta(\boldsymbol{r}_+)^{-1}\boldsymbol{S}^{++} + \Delta(\boldsymbol{r}_+)^{-1}\boldsymbol{S}^{+0}\left(\boldsymbol{S}^{00}\right)^{-1}\boldsymbol{S}^{0+}.
\end{aligned}$$

Which again, save for the different notation, is equal to the expression $\boldsymbol{S}_w$ from the MPH* set-up (5.8), induced by $\boldsymbol{Rw} = \boldsymbol{r}$.

Similar to that fact that a random vector $\boldsymbol{X}$ being multidimensional normal i.e. $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is equivalent to all linear combinations $Y = \langle \boldsymbol{a}, \boldsymbol{X}\rangle$ being univariate normal (with a zero variance normal distribution corresponding to a degenerate distribution at the mean) the same applies for the class MPH* as stated in the theorem below.

Theorem 5.7. *Let $\boldsymbol{Y}$ be a non-negative random vector. If $\langle \boldsymbol{Y}, \boldsymbol{w}\rangle \sim PH\left(\boldsymbol{\alpha}, \Delta(\boldsymbol{Rw})^{-1}\boldsymbol{S}\right)$ for all vectors $\boldsymbol{w} > \boldsymbol{0}$ and $\boldsymbol{R}$ a non-negative matrix with $\boldsymbol{Rw} > 0$ then*

$$\boldsymbol{Y} \sim MPH^*\left(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R}\right)$$

Proof. Substituting $\boldsymbol{w} = \boldsymbol{\theta}$ then the Laplace transform of $\langle \boldsymbol{Y}, \boldsymbol{\theta}\rangle$ with $s = 1$ given in Theorem 4.15 equals the multidimensional moment-generating function of $\boldsymbol{Y}$ given above whenever $\boldsymbol{w} > \boldsymbol{0}$. The result now follows from the continuity theorem of moment-generating functions. □

Example 5.8 (Joint distribution of order statistics)**.** Recall the notation and set-up of section 4.3, here we saw that max and min of two independent phase-types, $X_1 \sim \mathrm{PH}_{p_1}(\boldsymbol{\alpha}_1, \boldsymbol{S}_1)$ and $X_2 \sim \mathrm{PH}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{S}_2)$ with underlying Markov jump processes $\{X_t^1\}_{t\geq 0}$ and $\{X_t^2\}_{t\geq 0}$ on state space $E_1 = \{1, \ldots, p_1, p_1+1\}$, $E_2 = \{1, \ldots, p_2, p_2+1\}$ were again of phase-type with

$$X_{1:2} = \min\{X_1, X_2\} \sim \mathrm{PH}_{p_1 p_2}\left((\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2), (\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)\right)$$

and

$$X_{2:2} = \max\{X_1, X_2\} \sim \mathrm{PH}_{p_1 p_2 + p_1 + p_2}\left((\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2, \boldsymbol{0}, \boldsymbol{0}), \begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I}_{p_1} \otimes \boldsymbol{s}_2 & \boldsymbol{s}_1 \otimes \boldsymbol{I}_{p_2} \\ \boldsymbol{0} & \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_2 \end{pmatrix}\right)$$

Using a reward matrix with two columns, corresponding to variables Y_1 and Y_2. If Y_1 earns reward at rate 1 when the underlying process is in the states corresponding to both the processes being active and zero otherwise, while Y_2 earns reward in all states, corresponding to both processes active, only $\{X_t^1\}_{t\geq 0}$ active, and only $\{X_t^2\}_{t\geq 0}$ active, i.e.

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{e}_{p_1 p_2} & \boldsymbol{e}_{p_1 p_2} \\ \boldsymbol{0}_{p_1} & \boldsymbol{e}_{p_1} \\ \boldsymbol{0}_{p_2} & \boldsymbol{e}_{p_2} \end{pmatrix}.$$

Then let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2, \boldsymbol{0}_{p_1}, \boldsymbol{0}_{p_2})$ and

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_1 \oplus \boldsymbol{S}_2 & \boldsymbol{I}_{p_1} \otimes \boldsymbol{s}_2 & \boldsymbol{s}_1 \otimes \boldsymbol{I}_{p_2} \\ \boldsymbol{0} & \boldsymbol{S}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_2 \end{pmatrix}.$$

Then $(Y_1, Y_2) = (\min\{X_1, X_2\}, \max\{X_1, X_2\}) \sim \mathrm{MPH}^*(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ i.e. the joint distribution of the order statistics is MPH^* . To calculate moments of (Y_1, Y_2) we need the associated green matrix, $-\boldsymbol{S}^{-1}$. By using the block matrix inversion formula

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \\ -\left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \left(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{pmatrix}$$

(with $\boldsymbol{C} = \boldsymbol{0}$) iteratively, we get

$$-\boldsymbol{S}^{-1} = \begin{pmatrix} -(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1} & -(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{I} \otimes \boldsymbol{s}_2(-\boldsymbol{S}_1)^{-1} & -(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{s}_1 \otimes \boldsymbol{I}(-\boldsymbol{S}_2)^{-1} \\ \boldsymbol{0} & -\boldsymbol{S}_1^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -\boldsymbol{S}_2^{-1} \end{pmatrix}$$

A remark on the structure of $-\boldsymbol{S}^{-1}$ is in order. Recalling the probabilistic interpretation of the green matrix, the form of the diagonal and zero blocks is obvious.
The block $-(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{I} \otimes \boldsymbol{s}_2(-\boldsymbol{S}_1)^{-1}$ describes expected time spent by $\{X_t^1\}_{t\geq 0}$ in $\{1, \ldots, p_1\}$ *after* the process underlying X_2 has been absorbed, given initiation in $\{1, \ldots, p_1\} \times \{1, \ldots, p_2\}$ lexicographically ordered. Conveniently, $-\boldsymbol{S}_1^{-1}$ gives the expected time spent in $(1, \ldots, p_1)$ but we need to account for the fact that the chain is initiated with both processes active. The probability that $\{X_t^2\}_{t\geq 0}$ is absorbed during $[u, u+du)$ with $X_u^1 = j$ given initiation in $i \in \{1, \ldots, p_1\} \times \{1, \ldots, p_2\}$ lexicographically ordered, is

$$\left(e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 u}\boldsymbol{I} \otimes \boldsymbol{s}_2\right)_{ij} du.$$

Integrating out over time of absorption yields

$$\int_0^\infty e^{\boldsymbol{S}_1 \oplus \boldsymbol{S}_2 u}\boldsymbol{I} \otimes \boldsymbol{s}_2 \, du = (-\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{I} \otimes \boldsymbol{s}_2.$$

Another way of formulating the above is by letting

$$\tau_1 = \inf\{u > 0 \,|\, X_u^1 = p_1 + 1\}$$
$$\tau_2 = \inf\{u > 0 \,|\, X_u^2 = p_2 + 1\}$$

then

$$\left(\mathbb{P}\left((X_{\tau_2}^1, X_{\tau_2}^2) = (1, p_2+1)\right), \ldots, \mathbb{P}\left((X_{\tau_2}^1, X_{\tau_2}^2) = (p_1, p_2+1)\right)\right) = \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2(-\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{I} \otimes \boldsymbol{s}_2$$

Likewise for τ_1 i.e. $-(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{s}_1 \otimes \boldsymbol{I}(-\boldsymbol{S}_2)^{-1}$.

Using the block structure of $-\boldsymbol{S}^{-1}$ and $\boldsymbol{R}$ we can then simplify the general expressions for the moments into terms involving the blocks.

$$\begin{aligned}
\mathbb{E}Y_1 &= \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{R}_{\cdot 1} = \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2(-\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\boldsymbol{e} \\
\mathbb{E}Y_2 &= \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{R}_{\cdot 2} = \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2(-\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-1}\left[\boldsymbol{e} + \boldsymbol{I} \otimes \boldsymbol{s}_2(-\boldsymbol{S}_1)^{-1}\boldsymbol{e} + \boldsymbol{s}_1 \otimes \boldsymbol{I}(-\boldsymbol{S}_2)^{-1}\boldsymbol{e}\right] \\
\mathbb{E}Y_1Y_2 &= \boldsymbol{\alpha}\boldsymbol{U} \,\Delta\, (\boldsymbol{R}_{\cdot 1})\boldsymbol{U}\boldsymbol{R}_{\cdot 2} + \boldsymbol{\alpha}\boldsymbol{U} \,\Delta\, (\boldsymbol{R}_{\cdot 2})\boldsymbol{U}\boldsymbol{R}_{\cdot 1} \\
&= \boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha}_2(\boldsymbol{S}_1 \oplus \boldsymbol{S}_2)^{-2}\left[2\boldsymbol{e} + \boldsymbol{I} \otimes \boldsymbol{s}_2(-\boldsymbol{S}_1)^{-1}\boldsymbol{e} + \boldsymbol{s}_1 \otimes \boldsymbol{I}(-\boldsymbol{S}_2)^{-1}\boldsymbol{e}\right]
\end{aligned}$$

Below we use `simMPH()` to simulate $\boldsymbol{Y} = (Y_1, Y_2) = (\min(X_1, X_2), \max(X_1, X_2))$ where the representation of X_1, X_2 is given in Example 4.24.

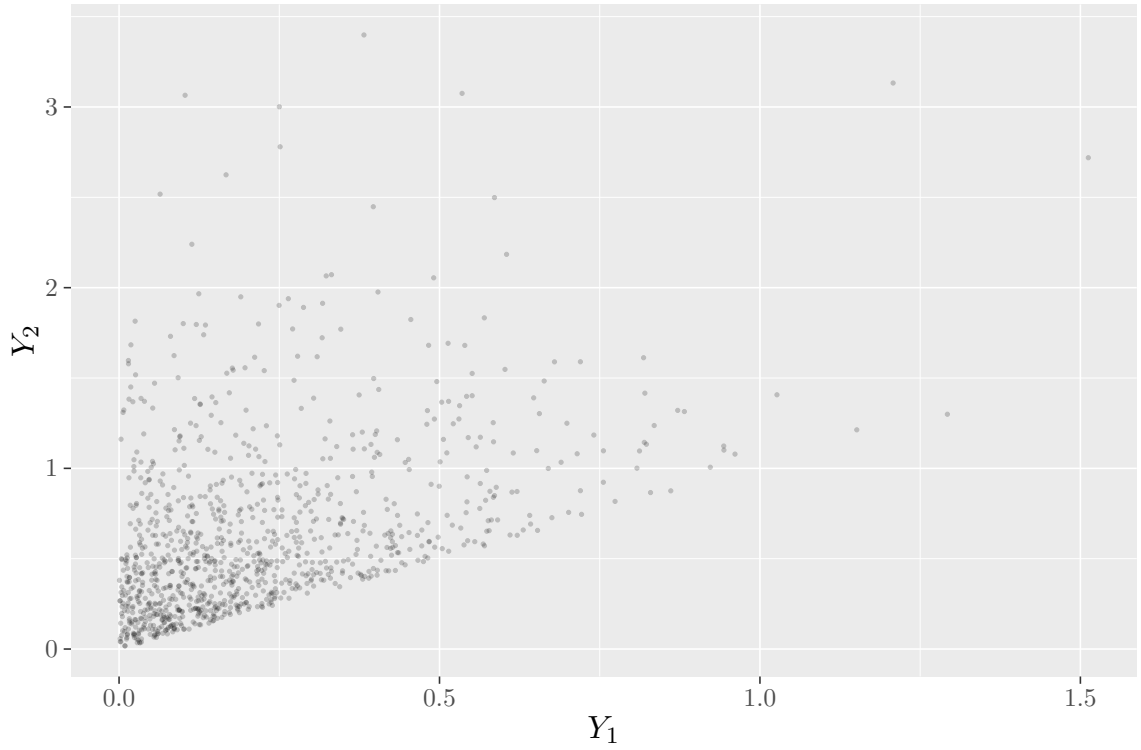

Figure 5.1: $N = 1e3$ simulation of (Y_1, Y_2).

Following the lines of the above, the general construction for the joint distribution of $(X_{1:n}, \ldots X_{n:n})$ involves the sub-intensity matrix of the highest order statistic $\boldsymbol{T}_{n:n}$ with corresponding reward matrix with first column rewarding states corresponding to all n underlying Markov jump processes being active, the second column rewards states where n or $n-1$ of the underlying processes are active and so forth untill the last nth column being equal to $\boldsymbol{e}$, rewarding all states. •

Chapter 6

Simulation

6.1 General Simulation

General non-optimized simulation of phase-type distribution either discrete, continuous or from the MPH* class is straight forward to implement using the underlying processes which rely on drawing exponential distributions for the holding times and probability point sampling for the next state after a jump. Three functions are available in the accompanying `R`-package:

- `simMC()` Simulation of general finite state space Markov chain or Markov jump process given initial distribution $\boldsymbol{\pi}$ and transition matrix $\boldsymbol{P}$ or intensity matrix $\boldsymbol{\Lambda}$.
- `simPH()` Simulation of discrete or continuous phase-type distribution given $(\boldsymbol{\pi}, \boldsymbol{T})$.
- `simMPH()` Simulation of a MPH* distribution given $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$. Note that if $\boldsymbol{R}$ is a column vector this corresponds to a transformation via rewards of a phase-type distribution.

Besides, for general utility

- `RejMJP()` Simulation of sample path $\{j_t\}_{0 \leq t \leq x}$ of Markov jump process underlying phase-type $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ that surpasses time point x. Takes input $(\boldsymbol{\pi}, \boldsymbol{T}, x)$.

The usage of the functions are very similar, an example of `SimPH()` can be found in Appendix A.2.2. Besides the absorption time τ the simulation functions also return the sequence of visited states and jump times, which will be useful in the coming.

6.2 Conditional Simulation

The problem we will discus in this section concerns the process underlying a phase-type distribution conditioned on the time of absorption. If $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ with underlying Markov jump process $J = \{J_t\}_{t \geq 0}$, then conditioned on the absorption time $\tau = x$ we want to simulate realizations of $\{J_t\}_{0 \leq t \leq x}$ which for instance is useful for estimating functionals of the process e.g. total time spent in state i or number of jumps from i to j. Unlike regular simulation this is non-trivial since the event $\{\tau = x\}$ has probability zero.

Altough many approaches exists for conditional simulation see e.g. [Hobolth and Stone, 2009] these approaches are mainly for Markov jump process conditioned at fixed time points e.g. $J_a = i, J_b = j$ and irreducible and positive recurrent (the expected return time for a recurrent state has finite mean) Markov jump processes. Which is not compatible with the phase-type framework.

As an alternative we will implement two computational methods suggested in [Bladt and Nielsen, 2017]. An importance sampler and a Markov Chain Monte Carlo (MCMC) sampler. In the end of the chapter we perform various simulations to asses the correctness and performance of the two methods.

6.2.1 Importance sampler

The basic idea of importance sampling is to change the sampling distribution (change of measure) to one that allows simulation of the desired outcome more easily, then compensating using importance weights. See e.g. [Givens and Hoeting, 2012] for a good source on importance sampling , MCMC and other computer intensive statistical methods. Let

$$\mathbb{P}_s(\cdot) = \mathbb{P}_{\boldsymbol{\pi}}(\cdot|\tau = s)$$
$$\mathbb{P}_s^*(\cdot) = \mathbb{P}_{\boldsymbol{\pi}}(\cdot|\tau \geq s)$$

and $\Psi(\{J_u\}_{0\leq u\leq s}))$ be some functional of the underlying Markov jump process. Then

$$\mathbb{E}\left(\Psi(\{J_t\}_{0\leq t\leq s})\,|\tau = s\right) = \int \Psi(\{J_t\}_{0\leq t\leq s})d\mathbb{P}_s = \int \frac{d\mathbb{P}_s}{d\mathbb{P}_s^*}\Psi(\{J_t\}_{0\leq t\leq s})d\mathbb{P}_s^*.$$

From which we need to calculate the importance weights $\frac{d\mathbb{P}_s}{d\mathbb{P}_s^*}$ which should be understood as densities given by the Radon–Nikodym derivative. Clearly $\mathbb{P}_s << \mathbb{P}_s^*$ i.e. $\mathbb{P}_s$ is absolutely continuous with respect to $\mathbb{P}_s^*$ from which the existence of the Radon-Nikodyn derivative follows. We start by finding the distribution of the Markov jump process just prior to absorption given by $\mathbb{P}(J_{s-} = i|\tau = s)$. Since the unconditional probability of being in state i at time $s \geq 0$ is given by

$$q_i(s) = \mathbb{P}(J_s = i) = \boldsymbol{\pi}e^{\boldsymbol{T}s}\boldsymbol{e}_i$$

and the density f_τ of τ can be split up by the state prior to absorption

$$f_\tau(s) = \sum_{i=1}^{p} q_i(s)t_i ds.$$

Then using the memoryless property of the process we have

$$q_i(s)t_i ds = \mathbb{P}(J_{s-} = i, \tau \in [s, s+ds)) = \mathbb{P}(J_{s-} = i|\tau = s)f_\tau(s)ds$$

from which we obtain

$$\mathbb{P}_s(J_{s-} = i) = \mathbb{P}(J_{s-} = i\,|\,\tau = s) = \frac{q_i(s)t_i}{f_\tau(s)} = \frac{\boldsymbol{\pi}e^{\boldsymbol{T}s}\boldsymbol{e}_i t_i}{\boldsymbol{\pi}e^{\boldsymbol{T}s}\boldsymbol{t}}. \tag{6.1}$$

Likewise for $\mathbb{P}_s^*$ here we get that

$$\begin{aligned}
\mathbb{P}_s^*(J_{s-} = i) &= \mathbb{P}(J_{s-} = i\,|\,\tau \geq s) \\
&= \frac{\mathbb{P}(J_{s-} = i, \tau \geq s)}{\mathbb{P}(\tau \geq s)} \\
&= \frac{\mathbb{P}(J_{s-} = i)}{\mathbb{P}(\tau \geq s)} \qquad (\{J_{s-} = i\} \subseteq \{\tau \geq s\}) \\
&= \frac{q_i(s)}{\sum_{j=1}^{p} q_j(s)}.
\end{aligned} \tag{6.2}$$

The advantage of $\mathbb{P}_s^*$ is that to simulate from this distribution one can simply keep simulating realizations of Markov jump processes until a trajectory has absorption time after s, rejecting the ones with absorption prior to s. Since the Markov jump process is Strong Markov and τ is a stopping time with respect to the process we get

$$\mathbb{P}(\tau \in ds\,|\,\{J_t\}_{0\leq t<s-}, J_{s-}) = \mathbb{P}(\tau \in ds\,|\,J_{s-})$$

i.e. $\{\tau \in ds\}$ is conditionally independent of $\{J_t\}_{0\leq t<s-} \in B$ given J_{s-}, for some set B. Similarly, we also have that $\{\tau \geq s\}$ is conditionally independent of $\{J_t\}_{0\leq t<s-} \in B$ given J_{s-}, from which we get

$$\begin{aligned}\mathbb{P}_s(\{J_t\}_{0\leq t<s-} \in \cdot \,|\, J_{s-}) &= \mathbb{P}(\{J_t\}_{0\leq t<s-} \in \cdot \,|\, J_{s-}, \tau = s) \\ &= \mathbb{P}(\{J_t\}_{0\leq t<s-} \in \cdot \,|\, J_{s-}) \\ &= \mathbb{P}(\{J_t\}_{0\leq t<s-} \in \cdot \,|\, J_{s-}, \tau \geq s) \\ &= \mathbb{P}^*_s(\{J_t\}_{0\leq t<s-} \in \cdot \,|\, J_{s-}). \end{aligned} \tag{6.3}$$

Combining the previous calculations, we get an expression for the Radon-Nikodyn derivative

$$\begin{aligned}\frac{d\mathbb{P}_s}{d\mathbb{P}^*_s}(\{j_t\}_{0\leq t<s-}) &= \frac{\mathbb{P}_s(\{J_t\}_{0\leq t<s-} = \{j_t\}_{0\leq t<s-})}{\mathbb{P}^*_s(\{J_t\}_{0\leq t<s-} = \{j_t\}_{0\leq t<s-})} \\ &= \frac{\mathbb{P}_s(\{J_t\}_{0\leq t<s-} = \{j_t\}_{0\leq t<s-} \,|\, J_{s-} = j_{s-})\mathbb{P}_s(J_{s-} = j_{s-})}{\mathbb{P}^*_s(\{J_t\}_{0\leq t<s-} = \{j_t\}_{0\leq t<s-} \,|\, J_{s-} = j_{s-})\mathbb{P}^*_s(J_{s-} = j_{s-})} \\ &= \frac{\mathbb{P}_s(J_{s-} = j_{s-})}{\mathbb{P}^*_s(J_{s-} = j_{s-})} \quad \text{(by eq. (6.3))}\end{aligned}$$

then using (6.1) and (6.2) we get

$$\frac{d\mathbb{P}_s}{d\mathbb{P}^*_s}(\{j_t\}_{0\leq t<s-}) = \frac{\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{e}\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{e}_{j_{s-}} t_{j_{s-}}}{\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{e}_{j_{s-}}\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{t}} = \frac{\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{e}}{\boldsymbol{\pi} e^{\boldsymbol{T}s}\boldsymbol{t}} t_{j_{s-}}. \tag{6.4}$$

Notice the simple expression which only depends on the exit rate of the Markov jump process at time $s-$ and the matrix exponentials $e^{\boldsymbol{T}s}$ which remain the same for all realizations of $\{J_t\}_{0\leq t<s-}$ since they only depend on the end time s. We now have the following importance sampler for estimating expectations of functionals of the Markov jump process underlying a phase-type distribution conditioned on the absorption time.

Algorithm 2: Importance Sampler.

Input: Initial distribution vector $\boldsymbol{\pi}$ and sub-intensity matrix $\boldsymbol{T}$ of Markov jump process underlying $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$. Functional $O()$ of the underlying process. Fixed $x > 0$.

Result: Approximation of $\mathbb{E}(O(\{J_t\}_{0\leq t\leq x}) \,|\, \tau = x)$

1. Simulate N realizations of $\{J_t\}_{t\geq 0}$ that surpasses x. The resulting sample is called $\{j_s(n)\}_{0\leq s\leq x}$, $n = 1, \ldots, N$.

2. For each $\{j_s(n)\}_{0\leq s\leq x}$ calculate the functional $O(\{j_s(n)\}_{0\leq s<x})$ of the simulated process up to time x denoted $O(n)$.

3. Calculate

$$\mu_{IS} = \frac{1}{N}\sum_{n=1}^{N} \frac{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e}}{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t}} t_{j_{x-}(n)} O(n) \tag{6.5}$$

Algorithm 2 is implemented in the `R`-package accompanying this thesis in the function called `ImpSampler()`. The function takes arguments: $\boldsymbol{\pi}, \boldsymbol{T}, x, N, O()$ and additional arguments which are passed on to $O()$. It returns the sample mean μ_{IS} along with the sample variance σ^2_{IS} of the summands of (6.5).

Note that, as discussed earlier, in (6.5) one may take $\frac{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e}}{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t}}$ out of the sum, which saves a lot of computational power away from evaluations of matrix exponentials. The main computational burdens of Algorithm 2 is 1) a possible large number of needed simulations of the underlying Markov jump process to obtain N samples, this depends on x and/or the dynamics of the

underlying Markov jump process. 2) Evaluations of $O(n)$ depending on the complexity of the functional. At the end of this chapter we will investigate the performance of the importance sampler.

6.2.2 MCMC

The basic idea of MCMC is to construct a general state space Markov chain with limiting distribution equal to some target distribution. One of the simplest and most popular construction methods is the Metropolis-Hastings algorithm.

Theorem 6.1 (Metropolis-Hastings)**.** *A general state space Markov chain $\{Z_n\}_{n\geq 0}$ on S with stationary measure given by $\Pi(A) = \int_A \pi(x)\eta(dx)$ can be constructed by the following algorithm. Let $q(\cdot|Z_n)$ be a proposal distribution with the property that $q(x|Z_n) > 0$ if $\pi(x) > 0$. Let $Z_0 \in S$ be any starting value in the state space and set $n = 0$.*

Algorithm 3: Metropolis-Hastings

1. Draw proposal $Y \sim q(\cdot|Z_n)$.

2. Set

$$Z_{n+1} = \begin{cases} Y, & \text{with probability } \alpha(Z_n, Y) := \min\left\{1, \frac{\pi(Y)q(Z_n|Y)}{\pi(Z_n)q(Y|Z_n)}\right\}, \\ Z_n, & \text{otherwise.} \end{cases}$$

3. Set $n = n + 1$ and GOTO **1.**

One of the main source of variants of the Metropolis-Hastings algorithm arises from using different proposal distributions. If $q(y|x) = q(y)$ i.e. the candidate Y is drawn independently of the current state of the chain, then the resulting sampler is called an *independence sampler*. In said case the acceptance ratio $\alpha(Y, Z_n)$ reduces to

$$\alpha(Y, Z_n) = \min\left\{1, \frac{\pi(Y)q(Z_n)}{\pi(Z_n)q(Y)}\right\} = \min\left\{1, \frac{w'}{w_n}\right\}$$

Where $w' = \frac{\pi(Y)}{q(Y)}$ and $w_n = \frac{\pi(Z_n)}{q(Z_n)}$ that is, the ratio between the importance weights between the proposal and target distributions $\mathbb{P}_x$ and $\mathbb{P}^*_x$, given by

$$\frac{d\mathbb{P}_x}{d\mathbb{P}^*_x}(\{j_t\}_{0\leq t<x-}) = \frac{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{e}}{\boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t}} t_{j_{x-}}.$$

We will implement an independence sampler by using the proposal distribution

$$q(\{j'_t\}_{0\leq t\leq x}|\{j_t\}_{0\leq t\leq x}) = \mathbb{P}^*_x\left(\{J_t\}_{0\leq t\leq x} = \{j'_t\}_{0\leq t\leq x}\right)$$

i.e. we propose a new Markov jump process $\{j'_t\}_{0\leq t\leq x}$ that surpasses time x independently of the previous point $\{j_t\}_{0\leq t\leq x}$. Using the independence sampler the acceptance ratio reduces to

$$\alpha(\{j'_t\}_{0\leq t\leq x}, \{j_t\}_{0\leq t\leq x}) = \min\left\{1, \frac{\frac{d\mathbb{P}_x}{d\mathbb{P}^*_x}(\{j'\}_{0\leq t<x-})}{\frac{d\mathbb{P}_x}{d\mathbb{P}^*_x}(\{j_t\}_{0\leq t<x-})}\right\} = \min\left\{1, \frac{t_{j'_{x-}}}{t_{j_{x-}}}\right\}$$

i.e. the acceptance ratio is simply given by the ratio between the exit rates of the Markov jump process at time x. This has some immediate consequences

Remark 6. 1) New proposals $\{j'_t\}_{0\leq t\leq x}$ with $t_{j'_{x-}} = 0$ will never be accepted as the acceptance ratio becomes 0. This is desired since $\tau = x$ can only occur if the exit rate at time x is positive. However, it has the drawback that the Metropolis Hastings chain might be stuck for long periods if processes surpassing x often have $t_{j_x} = 0$. Known as bad *mixing* properties.
2) The first draw $Z_0 = \{j_t\}_{0\leq t\leq x}$ can in fact have $t_{j_{x-}} = 0$ since the only requirement is that the process surpasses x. In this case we will adopt the convention $\frac{0}{0} = 0$ and $\frac{c}{0} = \infty, c > 0$. It might be desired that the starting state has $t_{j_{x-}} \neq 0$. To achieve this, in the implemented sampler, the argument `ForcePosExitRateStart` should be set to TRUE (default is FALSE).

In practice the the Metropolis-Hastings chain is run for some N steps resulting in the Markov chain $\{Z_n\}_{0\leq n\leq N}$ with stationary distribution $\mathbb{P}_x$. We can then approximate expectations of a functional by averaging realizations from the stationary distribution of the Metropolis-Hastings chain by

$$\mathbb{E}\left(O(\{J_t\}_{0\leq t\leq x}) \,|\, \tau = x\right) = \frac{1}{N+1}\sum_{n=0}^{N} O(n) \tag{6.6}$$

Where $O(n)$ is the value of the functional applied to Z_n.

The Metropolis-Hastings Markov chain may need a so called *burn in* before realizations of the chain are sufficiently close to the target distribution. This consists of discarding a certain number, B of the first realizations of Z_n. In this case the estimator is given by

$$\mathbb{E}\left(O(\{J_t\}_{0\leq t\leq x}) \,|\, \tau = x\right) = \frac{1}{N-B-1}\sum_{n=B+1}^{N} O(n) \tag{6.7}$$

The MCMC approach developed in this section is implemented in the `R`-package accompanying this thesis as the function `MCMC_PH()` which takes arguments: $\boldsymbol{\pi}, \boldsymbol{T}, x, N,$ `BurnIn`$, O()$ and additional arguments which are passed on to the functional $O()$. It returns the sample mean μ_{IS} along with the sample variance σ^2_{IS} of the summands of (6.6) or (6.7) if burn in is used.

6.3 Performance

To evaluate the correctness and efficiency of the importance sampler and the MCMC, we will go through some numerical tests.

6.3.1 Time spent in a transient state.

Recall example 4.35, here we saw that for

$$\tau \sim \mathrm{PH}_3\left(\left(\frac{2}{10}, \frac{3}{10}, \frac{5}{10}\right), \begin{pmatrix} -4 & 2 & 1 \\ 3 & -6 & 1 \\ 3 & 1 & -7 \end{pmatrix}\right) \tag{6.8}$$

then the unconditional time spent in state 2 prior to absorption which we will denote $O_2^{time}(\{J_t\}_{0\leq t\leq \tau}) = O_2^{time}$ was a defect phase-type given by

$$O_2^{time} \sim \mathrm{PH}\left(\frac{31}{50}, \left\{-\frac{19}{5}\right\}\right)$$

in particular $E\left(O_2^{time}\right) = \frac{31}{191}$. Following the lines of this, we now use the importance sampler and MCMC to estimate the expected time spent in state 2 of the Markov jump process underlying τ given $\tau = x$.

Since analytical result to ensure the correctness of the results are unobtainable, for reference and benchmark we simulate realizations of $\{J_t\}_{t\geq 0}$ such that $\tau \in [x-\varepsilon, x+\varepsilon]$ by simple rejection sampling. Choosing ε such that influence on the properties under investigation remain approximately intact while still allowing for feasible simulation is quite difficult, we will adopt an ad hoc solution given by choosing ε as the solution to

$$\varepsilon \mapsto F(x+\varepsilon) - F(x-\varepsilon) = 0.01$$

Where F is the CDF of τ.

The simulation study run below is conducted as follows: In the grid $x = (0.25, 0.5, \ldots, 2.75, 3.0)$ (note that $x = 3$ is approximately 0.99-percentile of τ) the following has been run:

- The benchmark sampler, $\tau \in [x-\varepsilon, x+\varepsilon]$, is run with $N = 1e5$ simulations, the sample mean being a benchmark estimate of $E\left(O_2^{time} \,|\, \tau = x\right)$.
- The Importance sampler is run `NBoot` $= 99$ times with `N_ImpS` $= 1e2, 1e3, 1e4$ each time yielding `Nboot` calculations of (6.5) which are plotted as a box plot.
- The MCMC algorithm is run `NBoot` $= 99$ times with `N_MCMC` $= 1e2, 1e3, 1e4$ and no burn in, each time yielding `NBoot` calculations of (6.6) which are plotted as a box plot.

The simulation study is shown below in Figure 6.1. Both the importance sampler and MCMC are quite close to the benchmark with similar variation that decrease substantially when the number of simulations is increased, although the results for only `N` $= 1e2$ simulations are still relatively close to the benchmark. The samplers seem to have a slight positive bias especially conditioned on large absorption times, although the benchmark should not be viewed as an exact solution.

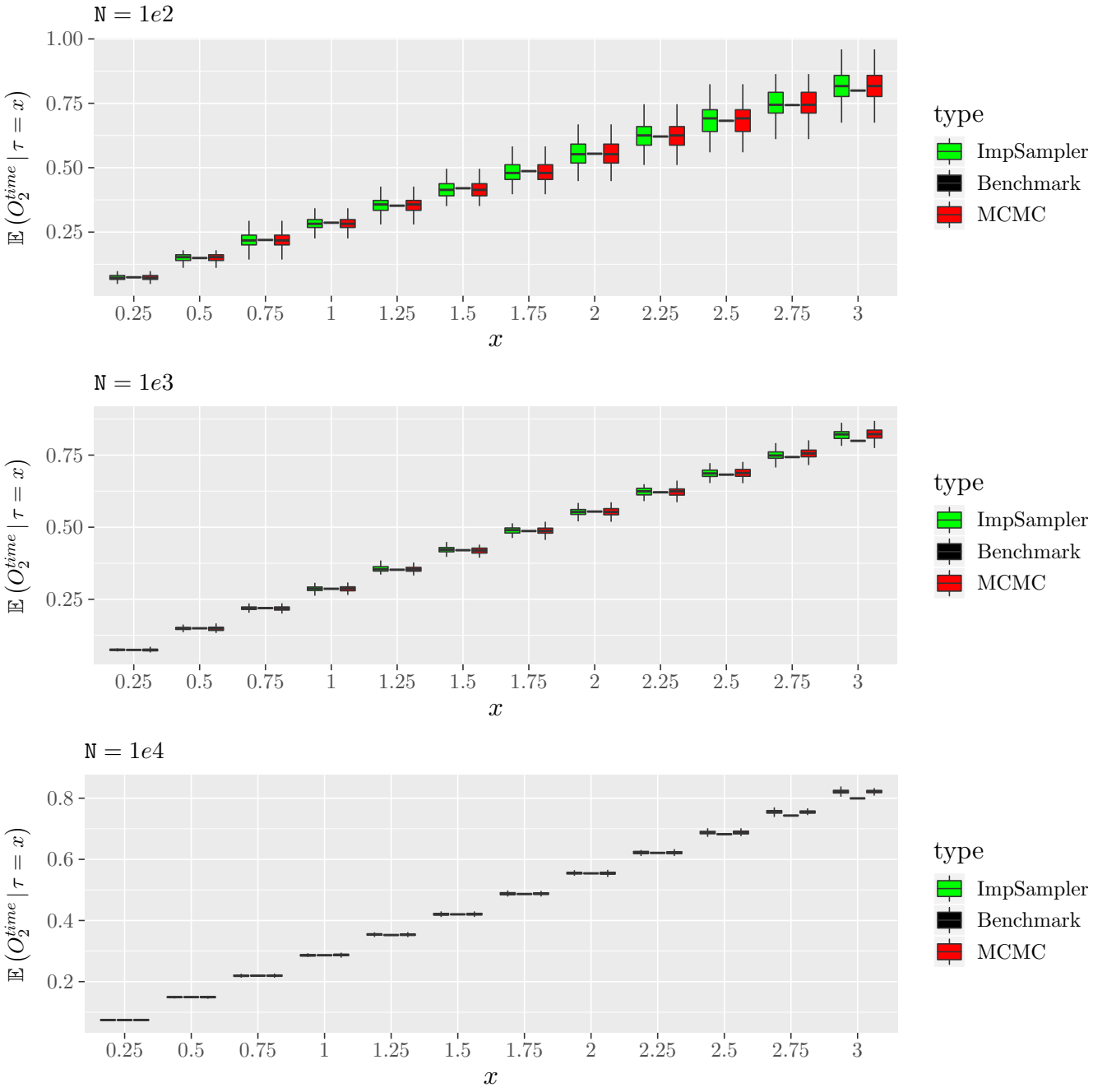

Figure 6.1: Simulations of the importance sampler and MCMC approach. The benchmark is the flat line between the green ImpSampler box plot and the red MCMC box plot.

6.3.2 Number of Jumps between states.

To rely less on the simulated benchmark for assessment of correctness we will exploit deterministic functionals. An example is the number of jumps $i \to j$ of the underlying process $\{J_t\}_{0 \le t \le x}$ which we will denote O_{ij}^{Njmp}. For a Markov jump process underlying a p-stage generalized Erlang distribution, see example 4.5 we obviously have that

$$O_{ij}^{Njmp} = \begin{cases} 1, & \text{if } j = i+1, \quad i \in E, j \in \{1, \dots, p+1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, the total number of jumps, denoted O^{Ntotal} will be equal to p at absorption.

Let $\tau \sim Er_5((1,1,2,2,3))$ i.e. a generalized Erlang distribution with 5 phases where the exponential holding times in the first two stages have rate 1, stage 3 and 4 have rate 2 and the last stage has rate 3. Below we estimate the number of jumps from state $4 \to 5$

$$\mathbb{E}\left(O_{45}^{Njmp} \,|\, \tau = x\right) \tag{6.9}$$

which equals 1. The estimates of (6.9) are plotted on the grid $x = 0.5, 1.0, \ldots, 8.5$, note that 8.5 is approximately the 0.99 quantile, for $N = 1e3, 1e4, 1e5$. We do not deploy the MCMC since, after a chain $\{j_t\}_{0 \le t \le x}$ has been generated that reaches state 5 with corresponding $O_{45}^{Njmp} = 1$ with $t_{j_x} = 3$ all remaining steps in the chain in the will have $O_{45}^{Njmp} = 1$. Hence the importance sampler is more interesting since the correctness depends on the ability to correctly compensate for the few processes reaching state 5 at time x. The results are shown below in Figrue 6.2.

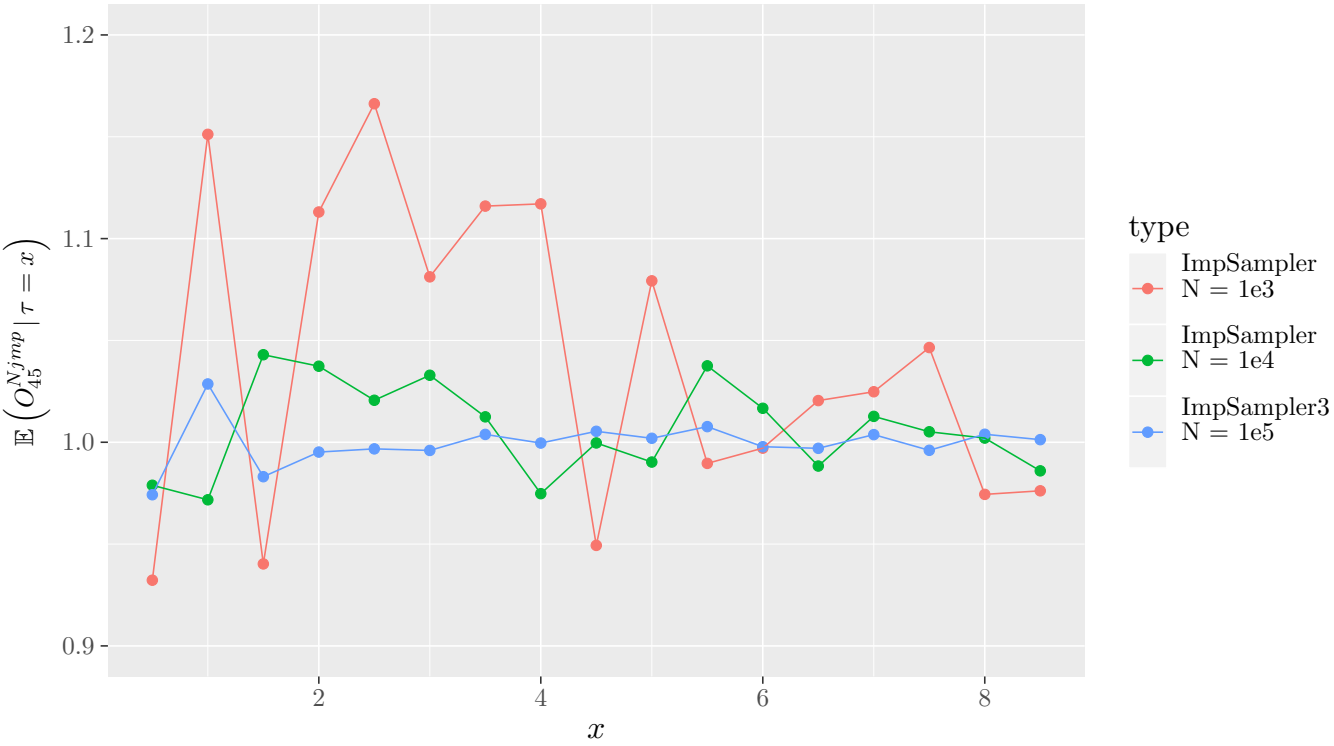

Figure 6.2: Simulation study of the importance sampler. Estimation of number of jumps $4 \to 5$. For $N = 1e3, 1e4, 1e5$ number of simulations.

We see that the importance sampler correctly estimates the number of jumps from state 4 to 5 conditioned of $\tau = x$. However, for x small the underlying process will rarely be in state 5 at time x and therefore yields a zero, rendering the estimates more volatile for small x, although increasing the number of simulations corrects this.

6.4 Conditional distribution function

Lastly, of more practical utility, we will use the samplers for estimating the conditional distribution function of the time spent in state $i \in E$ given $\tau = x$ specifically

$$F_i(y|\tau = x) = \mathbb{P}\left(\int_0^\tau 1\{J_t = i\}dt \leq y \,|\, \tau = x\right), \quad 0 \leq y \leq x. \tag{6.10}$$

Recall the unconditional case, here the time spend in any of the transient states was seen to be a mixture of an atom at zero and an exponential distribution, as seen in Example 4.35, where the CDF is given in the end of the example.
In the conditional case, we must have $F_i(x|\tau = x) = 1$. On the other hand, $F_i(0\,|\tau = x) \geq 0$. The strict inequality is due to the case of the underlying Markov jump process initiating in state $j \neq i$ and never entering state i prior to absorption at time $\tau = x$. Although for large x the probability of having spent no time in state i becomes very small. The equality is due to the fact that some phase-type distributions have underlying processes which will always spent some positive amount of time in each state, e.g. Erlang.
Below we estimate $F_2(y|\tau = x)$ for the phase-type distribution τ of (6.8), using both the importance sampler and the MCMC ($N = 1e4$) for $x = 0.2, 0.8, 2.8$ (note that 2.8 is the 0.99-quantile) using the functional

$$O_i^y(\{j_t\}_{0\leq t\leq x}) = 1\left\{\int_0^x 1\{j_t = i\}\right\} dt \leq y\} = 1\{O_i^{time} \leq y\}.$$

In the `R`-package, the above functional is called `OcondCDF()`.
Notice that the same sample can be used when evaluating $O_i^y(\{j_t\}_{0\leq t\leq x})$ for different y-values.

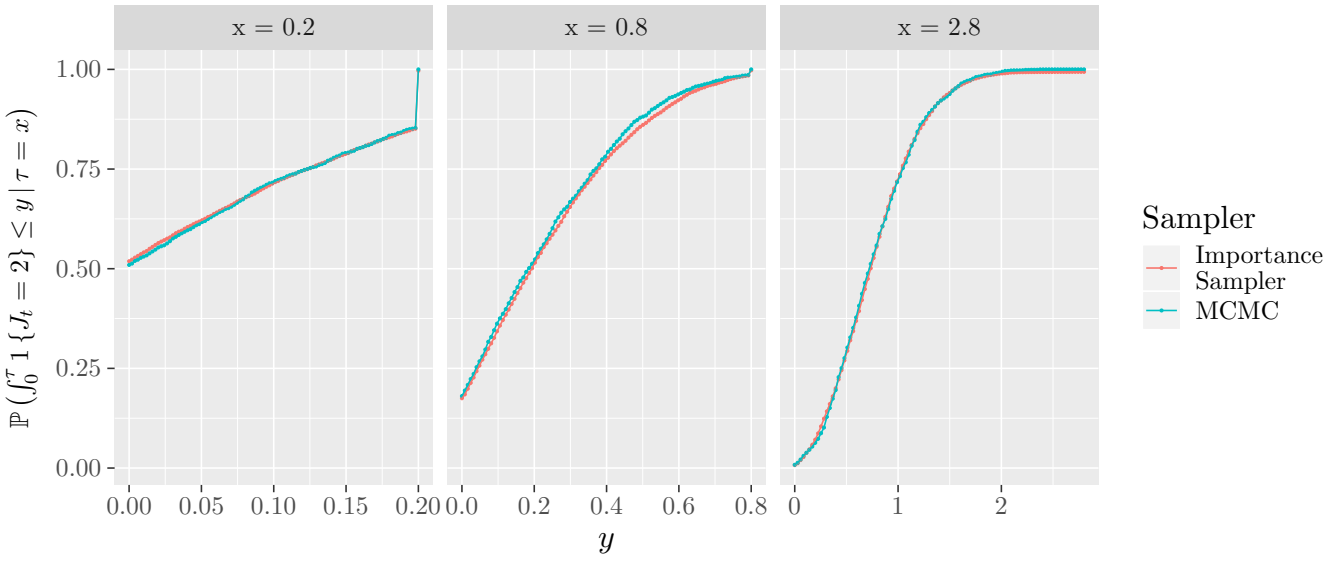

Figure 6.3: Simulation study of the importance sampler and MCMC for estimating the conditional distribution function.

The two samplers are very similar. A discontinuity at the conditioned time point is seen, which is especially pronounced for small x. This is due to the proportions of Markov jump processes having *only* been in state 2.
This concludes our exposition.

Chapter 7

Conclusion

This thesis gives an introduction to phase-type distributions both discrete, continuous and the multivariate MPH* class with thorough proofs of theorems and (hopefully) illustrative examples. We made implementations of some phase-type related quantities which hopefully can help make the theory even more accessible for application. The `R`-functions have been gathered in a package available on Github, see Appendix A.1.1 for how-to-install.

Topics of further research which we consider to be of interest, but not permitted by time, include, but is not limited to

- Relaxation of the assumption of the underlying process being time-homogeneous.
- Canonical representations and identifiability of phase-type distributions in particularly in connection with estimation and fitting of phase-type distributions from data.
- Properties of the MPH* class, e.g. analytical or numerical solutions to certain partial differential equations related to multivariate phase-type distributions.
- Extension and improvement of the Importance sampler and MCMC techniques for conditional simulation.

With regard to novel material to the best of our knowledge the specific urn-type construction of a discrete phase-type given in Theorem 3.20 along implementation, has not previously been presented in the format as given here. Likewise for the general result of discrete phase-type structure of first occurrence of sub-sequences of iid sequences of discrete random variables given in Theorem 3.17 and the accompanying implementation of algorithm 1.

Theorem 4.23 gives a precise characterization of probabilities of the $n!$ possible arrangements of n independent non-identically distributed phase-types using matrix operations instead of integration, for application the quantities of the theorem were implemented in `R`.
We also provided an implementation of the phase-type representation of the kth order statistic given in theorem 4.21 due to [Bladt and Nielsen, 2017] which has previously not been available.

Appendix A

Appendix

A.1 R-package phasetypeUtilsUcph

A.1.1 Installation

To install the R-package make sure that you have the `devtools`-package installed. Then run the following line in `R`

```
devtools::install_github("NielsKrarup/phasetypeUtilsUcph")
```

Alternatively, make sure you have the `remotes`-package installed and run the line

```
remotes::install_github("NielsKrarup/phasetypeUtilsUcph")
```

After the package has been installed it can be loaded by

```
library(phasetypeUtilsUcph)
```

The following objects should now be available

```
> ls("package:phasetypeUtilsUcph")
 [1] "AbsOrderPH"    "dimAbsOrderPH" "dimOrderStat"  "DPHSubSeqGen"  "erlang_ph"
 [6] "ImpSampler"    "knOrderPH"     "listAlpha"     "listS"         "MCMC_PH"
[11] "OcondCDF"      "Ojumpsij"      "OtimeSpent"    "OtotalNrJumps" "RejMJP"
[16] "simMC"         "simMPH"        "simPH"         "sumKronecker"  "Urn3"
```

The functions are described in the thesis when introduced. Except for the following:

- `listAlpha` - A list of 6 initial distribution vectors of phase-type representations for testing. The first 4 correspond to the ones used in Example 4.35. The 6th coresond to the phase-type used in Example 4.35. The corresponding sub-intensity matrices are found in `listS`.
- `listS` - A list of 6 sub-intensity matrices of phase-type representations for testing. The first 4 correspond to the ones used in Example 4.35. The 6th coresond to the phase-type used in Example 4.35. The corresponding initial distribution vectors are found in `listAlpha`.
- `sumKronecker` - The Kronecker sum between two square matrices.

In the next section we go through some examples of how to use the less self-explanatory functions.

A.2 Examples

A.2.1 DPHSubSeqGen()

The syntax is as follows. The subsequence should be stated as a sequence starting with 1 followed by an increment for every new unique value, with repetitions for values already introduced. The probability vector supplied is then the probability of each unique element in the order of appearance. e.g. say we want to investigate the first time we see the 6-sequence $10, 9, 8, 8, 9, 10$ in a sequence $X_1, X_2, \ldots$ of iid $pois(7)$ random variables. The code and output is:

```
> DPHSubSeqGen(pvec = dpois(c(10,9,8), lambda = 7), k_sub_seq = c(1,2,3,3,2,1))

          [,1]       [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.9290167 0.07098327 0.0000000 0.0000000 0.0000000 0.0000000
[2,] 0.8276121 0.07098327 0.1014047 0.0000000 0.0000000 0.0000000
[3,] 0.7986393 0.07098327 0.0000000 0.1303774 0.0000000 0.0000000
[4,] 0.7986393 0.07098327 0.0000000 0.0000000 0.1303774 0.0000000
[5,] 0.8276121 0.07098327 0.0000000 0.0000000 0.0000000 0.1014047
[6,] 0.9290167 0.00000000 0.0000000 0.0000000 0.0000000 0.0000000
```

A.2.2 SimPH()

Returns an object with visited states, times of jumps, whether the underlying process reached absorption (always TRUE, it is present since the function is a wrapper of `simMC()`), the sub-intensity matrix, and the time of absorption. The sample path can then easily be plotted as seen below.

```
> set.seed(123456)
> (obj <- simPH(SubIntMat = listS[[1]], phi0 = listAlpha[[1]]))
$states
[1] 2 3 1 3 4

$t
[1] 0.00000000 0.04226085 0.17468260 0.37198782 0.69997050

$abs
[1] TRUE

$SubIntMat
     [,1] [,2] [,3] [,4]
[1,]  -10    2    3    5
[2,]    1  -12   10    1
[3,]    6    1   -8    1
[4,]    0    0    0    0

$tau
[1] 0.6999705

> plot(obj$t, obj$states, type = 's')
```

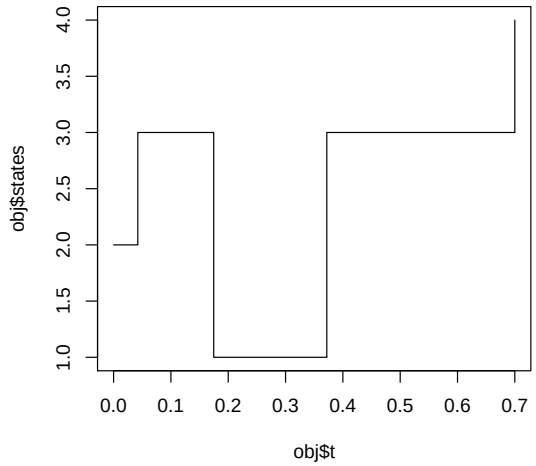

Figure A.1: Sample path of Markov jump process underlying a phase-type distribution.

A.2.3 `ImpSampler()`

As in Section 6.3.1, to estimate the time spent in state 2 conditioned on $\tau = 2.1$, of the phase-type (6.8). Using $N = 1e3$. The representation can be found as the 6th element in `listAlpha` and `listS`. Then the code and output is

```
> set.seed(123)
> ImpSampler(pi = listAlpha[[6]], SubIntMat = listS[[6]], N = 1e3,
TimePoint = 2.1, OFun = OtimeSpent, State = 2)

$N
[1] 1000

$sample_mean
[1] 0.6058142

$sample_var
[1] 0.2443598

$sample_sd
[1] 0.4943276
```

A.2.4 AbsOrderPH()

To reproduce Figure 4.12.

```
> aa <- AbsOrderPH(listS = listS[1:4], listAlpha = listAlpha[1:4])
> pvec <- aa$pi%*%solve(-aa$SubMat)%*%aa$D
> sum(pvec)
[1] 1
> plot(rev(as.vector(pvec)))
```

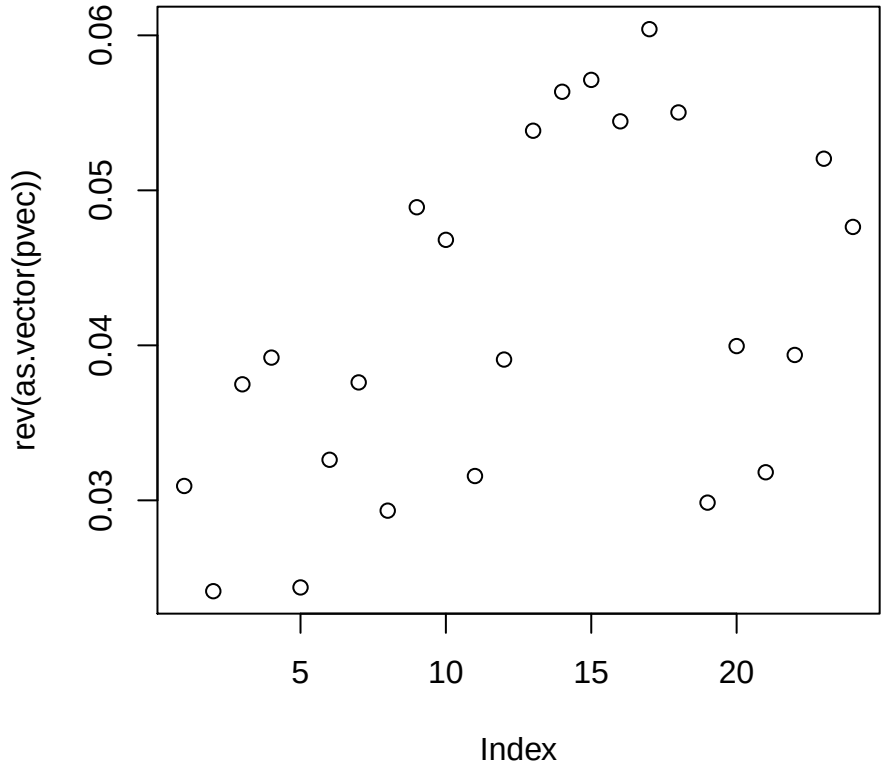

Figure A.2: Probabilities of 4! different absorption orders.

References

[Amussen, 2003] Amussen, S. (2003). *Applied Probability and Queues*. Springer, 2nd edition.

[Asmussen et al., 1996] Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441.

[Assaf et al., 1984] Assaf, D., Langberg, N. A., Savits, T. H., and Shaked, M. (1984). Multivariate phase-type distributions. *Operations Research*, 32(3):688–702.

[Bladt and Nielsen, 2010] Bladt, M. and Nielsen, B. F. (2010). Multivariate matrix-exponential distributions. *Stochastic Models*, 26(1):1–26.

[Bladt and Nielsen, 2017] Bladt, M. and Nielsen, B. F. (2017). *Matrix-Exponential Distributions in Applies Probability*. Springer US, 1st edition.

[Buchholz et al., 2014] Buchholz, P., Kriege, J., and Felko, I. (2014). *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications*. Springer Publishing Company, Incorporated.

[Cinlar, 2013] Cinlar, E. (2013). *Introduction to Stochastic Processes*. Dover Books on Mathematics Series. Dover Publications, Incorporated.

[Givens and Hoeting, 2012] Givens, G. and Hoeting, J. (2012). *Computational Statistics*. Wiley Series in Computational Statistics. Wiley.

[Hobolth and Stone, 2009] Hobolth, A. and Stone, E. A. (2009). Simulation from endpoint-conditioned, continuous-time markov chains on a finite state space, with applications to molecular evolution. *Ann. Appl. Stat.*, 3(3):1204–1231.

[Johnson and Taaffe, 1988] Johnson, M. and Taaffe, M. (1988). *The Denseness of Phase Distributions*. Research memorandum. School of Industrial Engineering, Purdue University.

[Kulkarni, 1989] Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operations Research*, 37(1):151–158.

[Neuts, 1981] Neuts, M. F. (1981). *Matrix-Geometric solutions in stochastic models*. John Hopkins University Press, 1st edition.

[Ramaswami, 2005] Ramaswami, S. A. . V. (2005). Bilateral phase type distributions. *Stochastic Models*, 21(1):239–259.

[Schassberger, 1973] Schassberger, R. (1973). *Warteschlangen*. Springer-Verlag.