DTU Library

New results on multivariate phase-type distributions

Larsen, Nicolai Siim

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Doctor of Philosophy
Applied Mathematics - Statistics and Probability Theory

DTU Compute
Department of Applied Mathematics and Computer Science

New results on multivariate phase-type distributions

Nicolai Siim Larsen

Kongens Lyngby, June 2022

Nicolai Siim Larsen

Richard Petersens Plads
Building 324
Room 210
2800 Kongens Lyngby, Denmark
Phone +45 2684 3077
nisi@dtu.dk

DTU Compute
Department of Applied Mathematics and Computer Science
Section for Statistics and Data Analysis
Technical University of Denmark

Richard Petersens Plads
Building 324
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Preface

This thesis is submitted to the Department of Applied Mathematics and Computer Science (DTU Compute) at the Technical University of Denmark in partial fulfillment of the requirements for the degree Doctor of Philosophy in Statistics and Probability Theory. The research presented in this thesis has been conducted predominantly at the Section of Statistics and Data Analysis of the Department of Applied Mathematics and Computer Science at the Technical University of Denmark and partially at the Department of Actuarial Sciences of the Faculty of Business and Economics at the University of Lausanne (UNIL). The research and the studies conducted during the doctoral programme have been supervised by the following researchers and faculty members:

Principal supervisor
Prof. Bo Friis Nielsen, Department of Applied Mathematics and Computer Science, Technical University of Denmark.

Co-supervisor
Prof. Murat Kulahci, Department of Applied Mathematics and Computer Science, Technical University of Denmark.

Supervisor during external research stay
Prof. Hansjörg Albrecher, Faculty of Business and Economics, University of Lausanne.

The research conducted during the doctoral programme is broadly concerned with probability theory and makes contributions to multivariate distribution theory, in particular with regard to multivariate phase-type distributions. Motivating applications are found across a wide range of engineering and sciences disciplines, although the results are more theoretical than practical in nature.

By signing below, I certify that the dissertation contains no research or other material which have been accepted for the award of any other degree or diploma in any university or tertiary institution, and to the best of my knowledge and belief, contains no research or material previously published or written by another person, except where due reference has been made. Furthermore, my signature indicates that I consent to the terms and conditions set forth by the Technical University of Denmark in connection with the submission of the dissertation.

Nicolai Siim Larsen
20 June 2022
Kongens Lyngby, Denmark

Acknowledgements

I would like to extend my sincere gratitude to my supervisors, my colleagues, and my family.

First, I would like to express my appreciation to my principal supervisor Bo Friis Nielsen for his tireless support and encouragement throughout my doctoral studies. I am grateful for countless fruitful discussions on personal and academic matters as well as his guidance, which has shaped me as a researcher. Finally, I would like to thank Bo for introducing me to probability theory and sharing his comprehensive knowledge of the subject, which has been an important source of inspiration for my research.

I would also like to thank Hansjörg Albrecher for functioning as an extremely friendly and welcoming host during my research stay at the University of Lausanne. The hospitality of the university staff was overwhelming, and I should thank Prof. Albrecher's entire research group for providing a supportive and friendly environment during my stay in Lausanne. Further thanks to Prof. Albrecher for his willingness to share his ideas and perspectives on various matters.

Additionally, I owe thanks to my colleagues and friends at the university for many enjoyable conversations and instructive discussions. Further thanks to everybody at the department who has helped through the doctoral programme one way or another.

I should also thank all the students, whom I have had the pleasure of supervising during the doctoral programme. The good collaborations have brought about some excellent projects, and it is an absolute pleasure to follow the students and watch them succeed in their respective industries.

And last, but definitely not least, a heartfelt thank you to my wife Mia for her never-ending support. She has followed me through good times and bad times, and she never ceased believing in me. She travelled with me to Switzerland and has always tried to accommodate my writing process. Thanks to Mia and the rest of my family for your love, your patience, and your sacrifices, which have helped me complete the doctoral programme.

The doctoral programme has been fully funded by a DTU scholarship. The stipend has covered tuition fees and salary expenses as well as several other expenses. I am very grateful that the university has given me this opportunity to conduct research and attend the doctoral programme.

The scholarship also partially covered some of the expenses related to the external research stay. I would like to give special thanks to the Otto Mønsted Foundation, the Foundation of Reinholdt W. Jorck and Wife, the Vera og Carl Johan Michaelsen Memorial Foundation, and the Chartered Accountant Oluf Christian Olsen and Wife Julie Rasmine Olsen Memorial Foundation for making considerable donations towards the external research stay. Additionally, several other foundations deserve my appreciation for providing financial support during the research stay.

English summary

New results on multivariate phase-type distributions

Phase-type distributions are used for mathematical modelling within numerous scientific fields of both engineering and commercial interests, such as genetics, transport optimisation, and insurance mathematics. Statistical models based on phase-type distributions are often flexible and well-suited for numerical computations and simulations, but the theoretical foundation of the multivariate phase-type distributions is not yet fully developed. Further research in estimation and characterisation of multivariate phase-type distributions is therefore necessary to learn how the useful properties of phase-type distributions can contribute to the advancement of data science and artificial intelligence.

This thesis concerns univariate and multivariate phase-type distributions, and a manuscript for a scientific paper constitutes the primary research contribution of the thesis. The second chapter of the thesis first describes the construction of the univariate phase-type distributions, which are introduced on the basis of Markov processes. Subsequently, several fundamental theorems regarding the univariate phase-type distributions are derived with a particular emphasis on their closure and denseness properties. Finally, the chapter treats reward structures, which are instrumental in the construction of the multivariate distributions, before some concrete examples of univariate phase-type distributions are presented.

Chapter three of the thesis deals with multivariate phase-type distributions. Like in the previous chapter, multiple theorems and properties are derived, many of which follow immediately from a result on concatenations of Markov processes proved in the second chapter. The proofs of the main theorems on the joint survival function and the cross-moments of the multivariate phase-type distributions contain various technical details and arguments, which cannot be not found elsewhere in the literature. The proofs are followed

by a section on phase-type representations of known multivariate exponential and gamma distributions. In said section, phase-type representations are compared to joint density functions and copula representations of different distributions, e.g. Marshall and Olkin's bivariate exponential distribution and the multivariate gamma distribution by Cheriyan and Ramabhadran.

A manuscript for a scientific paper on necessary and sufficient conditions for the existence of the multivariate gamma distribution by Dussauchoy and Berland comprises the last main chapter of the thesis. In the paper, it is shown that the distribution can be decomposed into a convolution of independent multivariate mixture distributions, which are easily recognised as phase-type distributions. This leads to a phase-type representation of the multivariate gamma distribution by Dussauchoy and Berland, which in conjunction with the abovementioned decomposition give rise to some necessary and sufficient conditions for the existence of the distribution. The conditions are expressed as constraints on the shape parameters of the distribution and the weights defining the previously mentioned mixture distributions. Lastly, the paper discusses the divisibility of the distribution and provides a specific example of parameter values leading to an invalid distribution. Based on this example, the paper concludes that the distribution is not always infinitely divisible and conjectures that the distribution generally cannot exist when the shape parameters take non-integer values.

Furthermore, the thesis contains some sections on related univariate and multivariate distributions as well as a short chapter on future research directions relating to the topics discussed throughout the thesis.

Dansk resumé

Nye resultater for multivariate fasetypefordelinger

Fasetypefordelinger bliver anvendt til matematisk modellering inden for flere videnskabelige områder, der både har ingeniørmæssige og forretningsmæssige interesser, såsom genetik, transportoptimering og forsikringsmatematik. Statistiske modeller baseret på fasetypefordelinger er ofte fleksible og velegnet til numeriske beregninger og simulationer, men det teoretiske grundlag for de multivariate fasetypefordelinger er stadig ikke færdigudviklet. Yderligere forskning om estimation og karakterisering af de multivariate fordelinger er derfor nødvendig for at afdække, hvordan fasetypefordelingers brugbare egenskaber kan bidrage til udviklingen inden for datavidenskab og kunstig intelligens.

Denne afhandling omhandler både univariate og multivariate fasetypefordelinger, medens afhandlingens primære forskningsbidrag udgøres af et manuskript til en videnskabelig artikel. Afhandlingens andet kapitel beskriver først konstruktionen af de univariate fasetypefordelinger, som introduceres på baggrund af Markov processer. Dernæst udledes en række hovedsætninger for de univariate fasetypefordelinger med særlig fokus på deres luknings- og tilnærmelsesegenskaber. Kapitlet behandler afslutningsvis konceptet præmiestrukturer, der er grundlæggende for konstruktionen af de multivariate fordelinger, før der præsenteres nogle konkrete eksempler på univariate fasetypefordelinger.

Kapitel tre i afhandlingen vedrører multivariate fasetypefordelinger. Ligesom i det forrige kapitel udledes flere sætninger og egenskaber, hvoraf flere følger umiddelbart fra delresultater om sammenkædninger af Markov processer, som blev bevist i det andet kapitel. Beviserne for hovedsætningerne om den simultane overlevelsesfunktion og krydsmomenterne for de multivariate fasetypefordelinger indeholder flere tekniske detaljer og argumenter, som ikke optræder i andre kilder. Beviserne efterfølges af et underkapitel om fasetype-

repræsentationer af kendte multivariate eksponential- og gammafordelinger. I underkapitlet sammenholdes fasetyperepræsentationerne med simultane tæthedsfunktioner og copulaer for adskillige fordelinger som f.eks. den bivariate eksponentialfordeling af Marshall og Olkin samt den multivariate gammafordeling af Cheriyan og Ramabhadran.

Afhandlingens sidste hovedkapitel udgøres af et manuskript til en videnskabelig artikel om nødvendige og tilstrækkelige betingelser for eksistensen af Dussauchoy og Berlands multivariate gammafordeling. I artiklen påvises det, at fordelingen kan dekomponeres til en foldning af uafhængige multivariate miksturfordelinger, der nemt kan beskrives som fasetypefordelinger. Derved opstilles en fasetyperepræsentation for Dussauchoy og Berlands gammafordeling, der i sammenhæng med ovennævnte dekomponering giver anledning til nogle nødvendige og tilstrækkelige betingelser for eksistensen af fordelingen. Bestingelserne udtrykkes som uligheder for fordelingens formparametre og vægtene, der indgår i de førnævnte miksturfordelinger. Endeligt diskuterer artiklen fordelingens delelighed og der gives et konkret eksempel på parameterværdier, hvor fordelingen ikke kan eksistere. Baseret på eksemplet konkluderes det, at fordelingen ikke altid er uendelig delelig, og eksemplet leder også til en formodning om, at fordelingen generelt ikke kan eksistere, når formparametrene tager ikke-heltallige værdier.

Afhandlingen indeholder desuden nogle underkapitler om beslægtede univariate og multivariate fordelinger samt et kort kapitel om fremtidige forskningsretninger vedrørende nogle af områder, som er blevet berørt i afhandlingen.

Acronyms

PH:	Phase-type
ME:	Matrix-exponential
IPH:	Inhomogeneous phase-type
BPH:	Bilateral phase-type
MPH:	Multivariate phase-type in the sense of Assaf et al. (1984)
MPH*:	Multivariate phase-type in the sense of Kulkarni (1989)
MVPH:	Multivariate phase-type in the sense of Bladt and Nielsen (2010)
MBPH*:	Multivariate bilateral phase-type in the sense of Kulkarni (1989)
MVBPH:	Multivariate bilateral phase-type in the sense of Bladt and Nielsen (2010)
MME*:	Multivariate matrix-exponential in the sense of Kulkarni (1989)
MVME:	Multivariate matrix-exponential in the sense of Bladt and Nielsen (2010)
MBME*:	Multivariate bilateral matrix-exponential in the sense of Kulkarni (1989)
MVBME:	Multivariate bilateral matrix-exponential in the sense of Bladt and Nielsen (2010)
LHS:	Left hand side
RHS:	Right hand side

Notation

$\boldsymbol{M}$	Matrix
$\boldsymbol{\alpha}$	(Greek) Row vector
$\boldsymbol{t}$	(Roman) Column vector
$\mathbf{0}_k$	k-dimensional row vector of zeros
$\mathbf{0}_{k\times k}$	Square matrix of zeros of order k
$\mathbf{1}_k$	k-dimensional row vector of ones
$\boldsymbol{e}_{m:k}$	k-dimensional row vector with a one in entry m and zeros otherwise
$\boldsymbol{I}_k$	Identity matrix of order k
$\boldsymbol{B}_k(\lambda)$	Bidiagonal matrix with $-\lambda$ in the diagonal and λ in the upper bidiagonal
$\boldsymbol{\alpha}^\top, \boldsymbol{M}^\top$	Transpose of the vector $\boldsymbol{\alpha}$ or the matrix $\boldsymbol{M}$
$\boldsymbol{\alpha}^*$	Normalized version of the vector $\boldsymbol{v}$
$\boldsymbol{M}^{-1}$	Inverse of the matrix $\boldsymbol{M}$
$(\boldsymbol{t})_i$	i'th element of the vector $\boldsymbol{t}$
$(\boldsymbol{M})_{ij}$	The element in row i and column j of the matrix $\boldsymbol{M}$
$\boldsymbol{M}_{i.}$	i'th row of the matrix $\boldsymbol{M}$
$\boldsymbol{M}_{.j}$	j'th column of the matrix $\boldsymbol{M}$
$\{X_t; t \in I\}$	Stochastic process with time index I
$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	The natural, integer, rational, real, and complex numbers, respectively
$\mathbb{R}^+$	The positive real numbers
$\mathbb{R}_0^+$	The non-negative real numbers
$\mathbb{1}\{A\}$	Indicator variable of the event A
$\mathrm{Re}(z)$	Real part of the complex number z
$\mathrm{Im}(z)$	Imaginary part of the complex number z

$\mathcal{L}$	Laplace transform
$f^{(n)}$	Derivative of order $n \in \mathbb{N}$ of the function f
$\mathbb{E}[Z]$	Expectation of the random variable Z
$\mathbb{V}[Z]$	Variance of the random variable Z
$\|\|\boldsymbol{v}\|\|_1$	Manhatten norm (1-norm) of the vector $\boldsymbol{v}$
$\boldsymbol{S} \oplus \boldsymbol{T}$	Kronecker sum of the matrices $\boldsymbol{S}$ and $\boldsymbol{T}$
$\boldsymbol{S} \otimes \boldsymbol{T}$	Kronecker product of the matrices $\boldsymbol{S}$ and $\boldsymbol{T}$
$\boldsymbol{S} \odot \boldsymbol{T}$	Hadamard product of the matrices $\boldsymbol{S}$ and $\boldsymbol{T}$
$E \times F$	Cartesian product of the sets E and F
$\mathrm{diag}(a_1, ..., a_n)$	Diagonal matrix of order n with elements $a_1, ..., a_n$ in the diagonal
$\mathrm{diag}(\boldsymbol{\alpha}), \boldsymbol{\Delta}(\boldsymbol{\alpha})$	Diagonal matrix with the elements of $\boldsymbol{\alpha}$ in the diagonal
$\mathrm{diag}(\boldsymbol{S}_1, ..., \boldsymbol{S}_n)$	Block diagonal matrix with matrices $\boldsymbol{S}_1, ..., \boldsymbol{S}_n$ in the diagonal

List of Figures

Contents

CHAPTER 1

Introduction

The fields of data science and machine learning have advanced considerably in recent years. New technologies have enabled researchers to apply large scale computational models and numerical simulations to solve complex problems in natural and engineering sciences. An increased focus on artificial intelligence along with the emergence of comprehensive data sets have caused especially non-parametric models to gain popularity. While the flexible structures of these models allow for highly accurate predictions, inference in such models is often difficult, even when regularization techniques are applied. Both industry and the academic community therefore search for statistical methods with sufficient flexibility to produce precise forecasts and generate reliable inference.

Matrix-analytic methods constitute an interesting avenue of research because they invoke semi-parametric models, which retain some of the structural versatility of non-parametric models while allowing for statistical inference. Modelling dynamical systems using the theory of matrix-analytical methods therefore allows for analytical and numerical approaches. These features and the tractability of the underlying foundation of Markov processes have caused matrix-analytic methods to become fundamental to stochastic modelling. Matrix-analytic methods have been used extensively in modelling and analysis of various stochastic systems of current commercial and engineering interest. The scientific literature on matrix-analytic methods comprises research papers from a variety of fields ranging from actuarial mathematics to telecommunications networks.

Phase-type and matrix-exponential distributions represent two examples of stochastic models within matrix-analytic methods. Phase-type distributions were originally proposed as holding time distributions within the field of telecommunications, but have since found applications across numerous research areas. A brief survey of recently published research papers finds articles in insurance mathematics and risk theory, in biology, and in technology engineering. There is also substantial research on phase-type distributions in various areas of operations research such as queueing theory, maintenance and scheduling, and healthcare modelling. The interested reader can consult the additional reading section of the bibliography, wherein several papers with relevance to the mentioned application areas have been listed. The phase-type and matrix-exponential distributions are also interesting from a purely theoretical standpoint. These distributions admit a fluid flow interpretation, see e.g. section 4.6 of Bladt and Nielsen (2017), and these representations can be used in numerical simulation and integration of stochastic differential equations, cf. Ramaswami (2013). Furthermore, the class of univariate matrix-exponential distributions corresponds to the class of univariate distributions with rational Laplace transforms, which suggests that results on matrix-exponential distributions are fairly general in nature. The classes of univariate phase-type and matrix-exponential distributions have additional closure and denseness properties that are useful when modelling and analyzing stochastic systems.

The attractive properties and the ubiquitous applicability of the univariate distributions have motivated research into constructions of multivariate matrix-exponential and phase-type distributions, and throughout the last couples of decades, several methods of construction have been proposed. The most studied among these distributions is the MPH* distribution propounded in Kulkarni (1989). Kulkarni put forward a multivariate phase-type distribution based on the concept of reward structures, which possesses some closure and denseness properties similar to those of the univariate distributions. Research has further shown that the simple construction lends itself to mathematical analysis while being sufficiently versatile to model numerous known multivariate exponential and gamma distributions. Examples of distributions admitting a MPH* representation include McKay's multivariate gamma distribution, Cheriyan and Ramabhadran's multivariate gamma distribution, Prékopa and Szántai's multivariate gamma distribution, and Dussauchoy and Berland's bivariate gamma distribution as well as Raftery's bivariate exponential distribution and the bivariate exponential distribution due to Marshall and Olkin.

Advances in the theory of MPH* distributions can materialize as new models within certain branches of a scientific field or as improved understanding of statistics in general. Sometimes the advances lead to simplifications of a result, which would otherwise require cumbersome calculations. A specific example could be the result in Gupta and Nadarajah (2006) on the sum of the components in McKay's bivariate gamma distribution, which follows immediately from the MPH* representation of the distribution. The potential for applications and the mathematical generality of the MPH* distributions emphasize the importance of researching and developing the theoretical foundations of the distributions. In particular, there are open questions related to MPH* distributions concerning estimation, characterisation, and general properties, which are of general interest. Such a question is presented in the following section.

1.1 A motivational example

This example is taken from the pharmaceutical industry, where the scientists at a drug development company are interested in evaluating the effects of a new drug during some clinical trials. The scientists have performed an experiment in which a number of subjects have been administered a drug against a serious disease. Following the administration, the scientists have observed the subjects and recorded their progression free survival time (PFS) and overall survival time (OS). Based on medical considerations and a subsequent data analysis of the experiment, the scientists decide to model the association between the PFS and the OS with a bivariate exponential distribution. The data analysis has generated point estimates of the rate parameters in the marginal distributions and the correlation between the two variables, and these numbers should enter into the model. According to the U.S. Food and Drug Administration (2018), the PFS should equal the OS whenever a subject dies before the disease shows signs of progression, and the model should naturally reflect this definition. In summary, the problem is to construct a bivariate exponential distribution (X, Y), where X is the OS and Y is the PFS such that

1. $X \sim \text{Exp}(\lambda_1)$,
2. $Y \sim \text{Exp}(\lambda_2)$,
3. $\text{Corr}(X, Y) = \rho$, and
4. $X \geq Y$.

Bladt and Nielsen (2010) gives a MPH* representation for a bivariate exponential distribution satisfying the first three conditions, but it is not known whether the construction can be extended to include the fourth condition. It is even unknown whether such a bivariate exponential distribution exists for certain combinations of parameter values. The Fréchet-Hoeffding bound shows that the correlation ρ cannot subceed $1 - \pi^2/6$, cf. sec. 2.9 of Kotz et al. (2000), however further nontrivial constraints on the parameters are yet to be found.

Since Sklar's theorem states that any multivariate distribution can be written in terms of its univariate marginal distributions and a copula characterizing the dependence structure in the distribution, the general question on the existence of the above bivariate exponential distribution translates into a question on the existence of certain copulas. The example is therefore of general interest in multivariate distribution theory, and a resolution to this problem could potentially lead to the discovery of a new dependence structure and shed new light on how to model complex associations between random variables.

Addressing the issues in the example might require research into novel methods of creating dependence structures, but prior to conceiving new structures, the possibilities of resolving the issues with currently known constructions should be explored. A natural starting point is thus to examine a simple dependence structure by studying its properties and the extent to which it can model known multivariate exponential and gamma distributions.

The construction underlying the MPH* distributions introduces a particularly interesting dependence structure because of its simplicity and the properties described on the previous pages. Furthermore, the concept of reward structures allows for numerous different extensions, which are yet to be properly studied.

For these reasons, the MPH* distributions constitute a promising research topic, where progress could lead to improvements in practical applications as well as theoretical advances in statistics and probability theory.

1.2 Scope of the thesis

The thesis contains an exposition of univariate and multivariate phase-type distributions along with a manuscript detailing some new findings and results from the doctoral studies. The studies have broadly been concerned with advancing the theoretical framework of the multivariate phase-type distributions, in particular with characterizing the distributions in terms of its properties and possible dependence structures, which is reflected in the two main contributions of the doctoral studies. Based on these results, the doctoral studies have yielded the following two manuscripts:

Paper A
First order moment distributions and concomitants of order statistics for MPH* distributions

The first manuscript of the studies deals with distributions derived from moments and order statistics of MPH* distributions. The first result is the derivation of the first order moment distributions for MPH* distributions based on a repeated time reversal argument. The derivation initially adopts an approach from Bladt and Nielsen (2011) used to calculate moment distributions in the univariate case before invoking an additional time reversal argument to obtain the first order moment distributions in the multivariate setting. The second result of the manuscript is a derivation of the MPH* representation of the concomitants associated with the order statistics of MPH* distributions. Navarro (2020) provides formulas for the probability density functions of concomitants from a sample of bivariate MPH* random variables using a semi-explicit expression for the joint probability density function of a bivariate MPH* distribution given in Breuer (2016). Navarro continues to show that the resulting density functions belongs to a phase-type distribution, but does not establish the corresponding phase-type representation. This lack of a phase-type representation severely complicates further analysis, and without the probabilistic interpretation linked to the representation, the model might be difficult to apply or implement. We remedy this issue by deriving the MPH* representation of the concomitants with probabilistic arguments based on the Markovian framework and the notion of time reversal. The first step of the derivation consists of establishing an absorbing Markov process, which allows for identifying the concomitant based on which state the process is absorbed from. This allows for deriving a representation of the time reversed version of the appropriate process, and a repeated time reversal argument the produces MPH* representation.

The manuscript covers the simplest possible case; the concomitants induced by the maximum and the minimum of two samples of a general multivariate MPH* distribution. However, since these results generalize easily (but with cumbersome and tedious calculations), the theory extends to the general case with an arbitrary number of samples and the concomitant of an arbitrary order statistic. At the time of writing, the manuscript is not sufficiently complete to be included in the thesis, but upon completion the manuscript shall be submitted to the Journal of Applied Probability.

Paper B

Necessary and sufficient conditions for the existence of the multivariate gamma distribution by Dussauchoy and Berland

The second manuscript of the studies constitutes the fourth chapter of the thesis and is concerned with the multivariate gamma distribution by Dussauchoy and Berland. Section 8.2 of Bladt and Nielsen (2017) gives MPH* representations for numerous multivariate exponential and gamma distributions, for instance those mentioned in the penultimate paragraph of the section preceding the motivational example. The exposition by Bladt and Nielsen is comprehensive, but not completely exhaustive compared to multivariate distributions listed in chapters 47 and 48 of Kotz et al. (2000). The most interesting distributions from Kotz et al. are however not necessarily those missing from the book by Bladt and Nielsen, but rather the distributions that Bladt and Nielsen only address partially. The two foremost examples are the multivariate gamma distribution by Dussauchoy and Berland and the multivariate exponential distribution by Krishnamoorthy and Parthasarathy. The distribution by Dussauchoy and Berland was especially interesting because the MPH* representation was only known for the bivariate case, and the distribution by Krishnamoorthy and Parthasarathy is particularly interesting since the three-dimensional distribution cannot be described by a MPH* representation of dimension three, cf. proposition 8.1.11 of Bladt and Nielsen (2017). Consequently, the distributions were prime candidates to potentially settle a conjecture suggesting that the MVPH class is a strict superset of the MPH* class.

In the second manuscript, we examine the dependence structure of the general multivariate gamma distribution by Dussauchoy and Berland, and the analysis in the manuscript produces two main findings. The first result is a derivation of some necessary and sufficient conditions for the existence of the distribution, while the second result is the identification of a MPH* representation of the distribution, which implies that the dependence structure of the distribution can be modelled with a linear reward structure. The established representation indicates that the simple decomposition used to construct the two-dimensional distribution is inadequate to describe the distributions of higher dimensions. Instead the general multivariate distribution must be decomposed into a convolution of distributions including multivariate mixture distributions. The manuscript also includes sections on infinite divisibility and numerical examples of the multivariate distributions. These sections contain a concrete counterexample showing that the distribution is not infinitely divisible along with a few examples illustrating the principal structure behind the multivariate distribution.

The results in the manuscript contribute to general multivariate distribution theory and to the subfield of multivariate phase-type distribution theory. Another interesting discovery that was made during the research process is the striking resemblance between the distribution by Dussauchoy and Berland and the distribution proffered in Carpenter and Diawara (2007). Although Carpenter and Diawara analyze a generalized gamma distribution, the underlying construction is based on a decomposition principle so similar to that invoked by Dussauchoy and Berland in their paper that Carpenter and Diawara should reference or otherwise give credit to Dussauchoy and Berland.

Notice that the manuscript appearing in this thesis is an extended version of the paper that will be submitted to the Journal of Multivariate Analysis.

1.3 Organization of the thesis

The first sections of the thesis have introduced and motivated the research into multivariate phase-type distributions by outlining some relevant applications of the distributions and explaining how the research connects to the broader topic of dependence structures and copulas. Furthermore, the preceding section expounded on the main results and contributions of the doctoral studies by briefly reviewing the different backgrounds and contexts of the manuscripts detailing the results.

The remainder of the thesis is comprised of four chapters and a bibliography. This introductory chapter is followed by two chapters (chapters 2 and 3) which serve as expositions on univariate and multivariate phase-type distributions (respectively), while chapter 4 is Paper B as described above. Chapter 5 is the final chapter and contains the conclusion of the thesis along with a discussion on future research. Before moving on, we give some brief summaries of the three main chapters:

Chapter 2: Univariate phase-type distributions
Chapter 2 covers the fundamental theory of univariate phase-type distributions. The chapter contains the basic definitions related to univariate phase-type distributions and derivations of some properties of the distributions. In addition, the chapter introduces the notion of transforming distribution through reward structures and give a short review of some related distributions such as matrix-exponential distributions.

Chapter 3: MPH* distributions
Chapter 3 reviews some variants of multivariate phase-type distributions and elaborates on the construction of MPH* distributions. The subsequent sections of the chapter are concerned with properties of the MPH* distributions and MPH* representations of multivariate exponential and gamma distributions. Finally, chapter 3 also concludes with a discussion of some related distributions.

Chapter 4: Paper B
Chapter 4 is comprised by paper B, which describes and analyzes the multivariate gamma distribution by Dussauchoy and Berland. The paper recaps the MPH* construction of the bivariate gamma distribution given by Bladt and Nielsen before deriving some necessary conditions for the existence of the general multivariate distribution. Then the MPH* representation of the general multivariate distribution is derived together with some sufficient conditions for the existence of the distribution. The paper concludes with sections on infinite divisibility and numerical examples.

CHAPTER 2

Univariate phase-type distributions

This chapter covers definitions and properties of univariate finite-dimensional continuous time phase-type distributions. The phase-type distribution is first introduced as the probability law governing the time until absorption in certain Markov processes. This interpretation allows for applying the general framework of Markov processes to derive probability functions and properties of the distribution. Emphasis is placed on the closure and denseness properties of the distribution. The closure properties extend to the multivariate setting, whereas the denseness property motivates current research on characterization of multivariate phase-type distributions. The chapter concludes with an introduction to the concept of reward structures and a discussion on related distributions.

2.1 Introduction

The notion of phase-type distributions was introduced by Erlang in 1917 and further formalized by Neuts in a 1975 paper. Erlang considered a process traversing a system with a fixed number of stages, where the process spends an exponentially distributed time in each stage, and derived the distribution of the total time the process spends in the system. This is of course the appropriately named Erlang distribution. Neuts generalized this concept and introduced the phase-type distribution as the distribution of the total sojourn time in the transient states of an absorbing Markov jump process, which is the common definition today.

To formally state the definition, consider a continuous time Markov jump process $\{X_t; t \geq 0\}$ with only transient and absorbing states. We can assume that there is only a single absorbing state without loss of generality, even though for some purposes and applications it is beneficial or necessary to keep the absorbing states separated. The state-space shall thus be denoted by $E = \{1, 2, ..., k, k+1\}$, where the first k states are transient and the final state is absorbing. The process is initiated according to the initial distribution $(\boldsymbol{\pi}, \pi_{k+1})$, where $\boldsymbol{\pi} = (\pi_1, ..., \pi_k)$, that is $\mathbb{P}(X_0 = i) = \pi_i$ for all $i \in E$. Notice that in some text books and papers, the vector $\boldsymbol{\pi}$ alone is specified as the initial distribution since the probability that the process is initiated in the absorbing state can be computed directly from $\boldsymbol{\pi}$. The process then evolves according to the associated generator matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0}_k & 0 \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{T}$ is a sub-generator matrix and $\boldsymbol{t}$ is the exit rate vector containing the transition rates from the transient states to the absorbing state. The exit rate vector is completely determined from the sub-generator through the relation

$$\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{1}_k^\top. \tag{2.2}$$

The initial distribution and the generator matrix completely describe the process and give rise to the definition of a phase-type distribution.

Definition 2.1
Let $\{X_t; t \geq 0\}$ be an absorbing Markov jump process as described above with initial distribution $(\boldsymbol{\pi}, \pi_{k+1})$ and generator matrix $\boldsymbol{\Lambda}$. The time until absorption of the process is defined as

$$\tau = \inf\{t \geq 0 | X_t = k+1\} \tag{2.3}$$

and follows a continuous time phase-type distribution of dimension k with representation $(\boldsymbol{\pi}, \boldsymbol{T})$. This is written as $\tau \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$, where the subscript can be suppressed if the dimension of the distribution is clear from the context.

Figure 2.1 shows two realizations of a phase-type distributed random variable generated according to the above definition. In the example used in the figure, state number 6 (marked in red) is the absorbing state and the realized values of the random variable are the hitting times when the process reaches the absorbing state.

The definition also indicates that phase-type distributed random variables can take the value zero, which happens if the Markov process is initiated in the absorbing state. Phase-type distributions can therefore have an atom in zero, the size of which is given by π_{k+1}. Consequently, a distribution simplifies to a deterministic distribution when $\pi_{k+1} = 1$, while a distribution with $\pi_{k+1} = 0$ is absolutely continuous.

This construction based on the tractable framework of Markov processes makes phase-type distributions a versatile and flexible tool in stochastic modelling with applications across a large number of scientific fields. The structure of Markov jump processes makes efficient numerical simulations of the distributions readily accessible and results from the theory of Markov processes are often useful in derivations concerning the distributions.

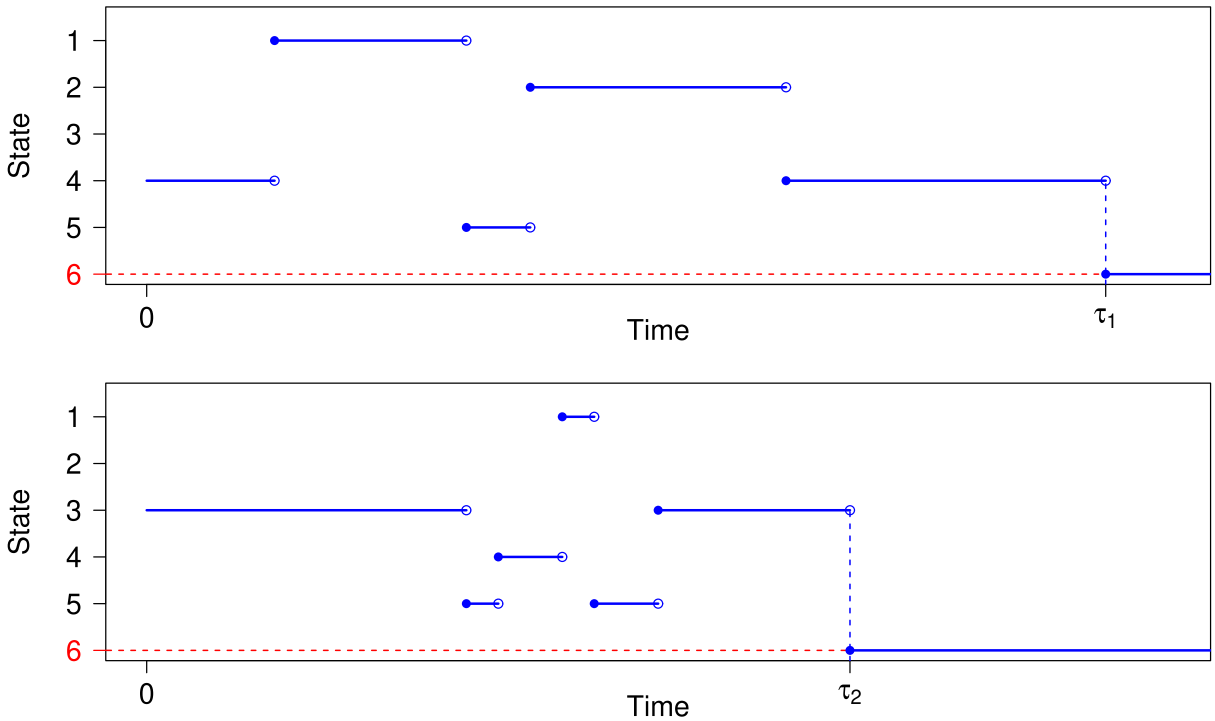

Figure 2.1: *Two realizations of a phase-type distributed random variable and the underlying Markov jump process.*

A result from the theory of Markov processes that will be invoked frequently throughout the thesis is that the transition probability matrix of a Markov jump process can be expressed in terms of the matrix-exponential of its generator matrix. Specifically, the transition probability matrix $\boldsymbol{P}(t)$ with elements $p_{ij}(t)$ defined as

$$p_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i), \quad \forall i, j \in E, \tag{2.4}$$

is given by

$$\boldsymbol{P}(t) = e^{\boldsymbol{\Lambda} t}, \tag{2.5}$$

which follows directly from linear systems theory. The matrix-exponential on the RHS of equation (2.5) is defined through the same power series defining the common exponential function, where the argument is a matrix instead of a number, see e.g. Bladt and Nielsen (2017), corollary 1.3.11.

Another useful property of Markov jump processes is the non-singularity of their generator matrices. Phase-type distribution thus inherit this property from the underlying Markovian structure, which simplify many calculations that are tedious in the more general context of matrix-exponential distributions.

Theorem 2.1

Let $\{X_t; t \geq 0\}$ be an absorbing Markov jump process as described above with initial distribution $(\boldsymbol{\pi}, \pi_{k+1})$ and generator matrix $\boldsymbol{\Lambda}$. The sub-generator matrix $\boldsymbol{T}$ is then non-singular.

Proof. Let the transition probability matrix of the (pseudo) embedded Markov chain derived from the process $\{X_t\}$ be given by

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{q} \\ \boldsymbol{0}_k & 1 \end{pmatrix}, \tag{2.6}$$

where $\boldsymbol{Q}$ is a square matrix of order k containing the transition probabilities among the transient states and $\boldsymbol{q}$ is a k-dimensional column vector containing the transition probabilities from the transient states to the absorbing state. The diagonal elements of $\boldsymbol{Q}$ are zero and the off-diagonal elements are calculated as

$$\boldsymbol{Q}_{ij} = q_{ij} = \frac{T_{ij}}{-T_{ii}}, \quad \forall (i,j) \in {E^*}^2 : i \neq j, \tag{2.7}$$

where $E^* = \{1, ..., k\}$ denotes the state-space of the transient states, cf. e.g. Karlin and Pinsky (2011), chapter 6.6. Since the diagonal elements of the sub-generator matrix $\boldsymbol{T}$ are strictly negative and the rows of $\boldsymbol{T}$ sum to zero, the values generated by the ratios in eq. (2.7) are always in the unit interval and thus probabilities. Consider then a real-valued k-dimensional row vector $\boldsymbol{\nu}$ defined through the equation

$$\boldsymbol{\nu T} = \boldsymbol{0}_k, \tag{2.8}$$

which is shorthand notation for the following system of linear equations

$$\sum_{i=1}^{k} \nu_i T_{ij} = 0, \quad \forall j \in E^*, \tag{2.9}$$

where $E^* = \{1, ..., k\}$ denotes the state-space of the transient states. Subtracting the appropriate term in each equation yields

$$-\nu_j T_{jj} = \sum_{\substack{i=1 \\ i \neq j}}^{k} \nu_i T_{ij}, \quad \forall j \in E^*. \tag{2.10}$$

By applying the result from equation (2.7) in eq. (2.10), the system becomes

$$\nu_j T_{jj} = \sum_{\substack{i=1 \\ i \neq j}}^{k} \nu_i q_{ij} T_{ii} = \sum_{i=1}^{k} \nu_i q_{ij} T_{ii}, \quad \forall j \in E^*, \tag{2.11}$$

where the latter equality holds due to the fact that the diagonal elements of $\boldsymbol{Q}$ are zero. A new row vector $\boldsymbol{\beta}$ is then defined. The vector has elements $\boldsymbol{\beta}_i = \beta_i = \nu_i T_{ii}$ for all $i \in E^*$. Invoking the new vector allows for expressing the system of equations as

$$\beta_j = \sum_{i=1}^{k} \beta_i q_{ij}, \quad \forall j \in E^* \tag{2.12}$$

or equivalently as the matrix equation

$$\boldsymbol{\beta} = \boldsymbol{\beta Q}. \tag{2.13}$$

Straightforward recurrence arguments show that

$$\boldsymbol{\beta} = \boldsymbol{\beta Q}^n, \quad \forall n \in \mathbb{Z}. \tag{2.14}$$

As $\boldsymbol{Q}$ governs the transitions among the transient states, the elements of $\boldsymbol{Q}^n$ will converge to zero as n tends to infinity, i.e.

$$\lim_{n\to\infty} \boldsymbol{Q}^n = \mathbf{0}_{k\times k}. \tag{2.15}$$

Hence,

$$\boldsymbol{\beta} = \boldsymbol{\beta} \lim_{n\to\infty} \boldsymbol{Q}^n = \mathbf{0}_k. \tag{2.16}$$

Since $\boldsymbol{\beta} = 0$ and the diagonal elements of $\boldsymbol{T}$ are strictly negative, the vector $\boldsymbol{\nu}$ must be a vector of zeros. In conclusion, the system in equation (2.8) does not admit any non-trivial solutions, which means that the columns of the sub-generator $\boldsymbol{T}$ are linearly independent. This in turn implies that the sub-generator matrix is non-singular. □

There actually exists a stronger theorem stating that the sub-generator matrix is invertible if, and only if, all the states governed by the sub-generator are transient, cf. e.g. Bladt and Nielsen (2017), theorem 3.1.11.

The non-singularity of the sub-generator matrix ensures the existence of the inverse matrix $\boldsymbol{T}^{-1}$. The negated version of the inverse matrix $\boldsymbol{U} = -\boldsymbol{T}^{-1}$ is dubbed the Green matrix and it describes the expected sojourn time in the transient states prior to absorption[1]. The next theorem verifies this statement and also serves as a prelude to the analysis used to derive the probability functions in the next section.

Theorem 2.2
Let $\{X_t; t \geq 0\}$ be an absorbing Markov jump process as described above with initial distribution $(\boldsymbol{\pi}, \pi_{k+1})$ and generator matrix $\boldsymbol{\Lambda}$, and define τ as the time until absorption of the process like in definition 2.1. The Green matrix $\boldsymbol{U} = -\boldsymbol{T}^{-1}$ then has elements

$$u_{ij} = \mathbb{E}_i\left[\int_0^\tau \mathbb{1}\{X_t = j\}dt \middle| X_0 = i\right], \quad \forall i, j \in E^*, \tag{2.17}$$

i.e. the element u_{ij} describes the expected sojourn time in state j prior to absorption given the process is initiated in state i.

Proof. Define Z_j as the sojourn time in an arbitrary transient state $j \in E^*$ prior to absorption. The expected sojourn time in state j prior to absorption given the process is initiated in state $i \in E^*$ can then be formulated as the conditional expectation

$$\mathbb{E}[Z_j|X_0 = i] = \mathbb{E}\left[\int_0^\tau \mathbb{1}\{X_t = j\}dt \middle| X_0 = i\right] = \mathbb{E}\left[\int_0^\infty \mathbb{1}\{X_t = j\}dt \middle| X_0 = i\right]. \tag{2.18}$$

The last expression in eq. (2.18) can be considered as a double integral. Since the conditional expectation of an indicator variable is a conditional probability, and thus finite, Fubini's theorem (Schilling (2005), corollary 13.9) applies and allows for interchanging the order of integration. The conditions in Fubini's theorem related to measurability are trivially satisfied as the integrand is an indicator variable.

$$\mathbb{E}[Z_j|X_0 = i] = \int_0^\infty \mathbb{E}\left[\mathbb{1}\{X_t = j\}|X_0 = i\right] dt = \int_0^\infty \mathbb{P}\left(X_t = j|X_0 = i\right) dt = \int_0^\infty p_{ij}(t)dt, \tag{2.19}$$

[1] It is convention to include the addendum "*prior to absorption*" although it is redundant since the process cannot visit the transient states after entering the absorbing state.

where the last equality follows from the definition in equation (2.4). The latter integral is evaluated by integrating the transition probability matrix restricted to the transient states, say $\boldsymbol{P}^*(t)$. This restricted version of the probability transition matrix can be expressed as the matrix-exponential of the sub-generator matrix $\boldsymbol{T}$ due to a simple decomposition of the matrix-exponential of the generator $\boldsymbol{\Lambda}$, see e.g. Bladt and Nielsen (2017), lemma 3.1.6. Using this result leads to the integral

$$\int_0^\infty \boldsymbol{P}^*(t)dt = \int_0^\infty e^{\boldsymbol{T}t}dt = \left[\boldsymbol{T}^{-1}e^{\boldsymbol{T}t}\right]_0^\infty, \tag{2.20}$$

which is well-defined since theorem 2.1 guarantees the existence of the inverse matrix $\boldsymbol{T}^{-1}$. As

$$\left[\boldsymbol{T}^{-1}e^{\boldsymbol{T}t}\right]_0^\infty = \left[\boldsymbol{T}^{-1}\boldsymbol{P}^*(t)\right]_0^\infty = -\boldsymbol{T}^{-1}\left(\mathbf{0}_{k\times k} - \boldsymbol{I}_k\right) = -\boldsymbol{T}^{-1} = \boldsymbol{U}, \tag{2.21}$$

the results in equations (2.19) and (2.20) yield that

$$\mathbb{E}[Z_j|X_0 = i] = \int_0^\infty p_{ij}(t)dt = \left(\int_0^\infty \boldsymbol{P}^*(t)dt\right)_{ij} = \boldsymbol{U}_{ij} = u_{ij}, \tag{2.22}$$

which completes the proof. □

The probabilistic interpretation of the Green matrix is useful in the context of phase-type distributions, where the matrix features in several derivations. For instance, the Green matrix appears in formulae for the moments of phase-type distributions, which shall be treated in the next section.

2.2 Probability functions

This section is concerned with the probability function associated with phase-type distributions and some basic results derived from those functions. The starting point of this exposition is the distribution function, which is covered in the following theorem. Then we treat the density function and some kernel transformations of the distribution function before the final part of the section deals with the moments of phase-type distributions.

Theorem 2.3
Let $Z \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$ and define $F : \mathbb{R} \to [0,1]$ as the distribution function of Z. Then F has support on the non-negative real numbers and is given by

$$F(z) = 1 - \boldsymbol{\pi}e^{\boldsymbol{T}z}\mathbf{1}_k^\top, \quad z \geq 0. \tag{2.23}$$

Proof. Definition 2.1 entails that for all $z \in \mathbb{R}_0^+$

$$\mathbb{P}(Z > z) = \mathbb{P}(X_z \in E^*), \tag{2.24}$$

where $\{X_t; t \geq 0\}$ still refers to the underlying process and $E^* = \{1, ..., k\}$ to the state-space of the transient states of the process. Since the process can only sojourn in one state at a time, the RHS of eq. (2.24) can be decomposed into the sum

$$\mathbb{P}(X_z \in E^*) = \sum_{j\in E^*} \mathbb{P}(X_z = j) = \sum_{j=1}^k \mathbb{P}(X_z = j). \tag{2.25}$$

The law of total probability then gives that

$$\sum_{j=1}^{k}\mathbb{P}(X_z=j)=\sum_{j=1}^{k}\sum_{i=1}^{k+1}\mathbb{P}(X_z=j|X_0=i)\mathbb{P}(X_0=i)=\sum_{j=1}^{k}\sum_{i=1}^{k}\mathbb{P}(X_z=j|X_0=i)\mathbb{P}(X_0=i). \tag{2.26}$$

As the process cannot transition from the absorbing state to a transient state, the final term of the inner summation is always zero, which justifies the last step in eq. (2.26). The final expression in equation (2.26) can be restated in terms of the initial distribution and the transition probability matrix as

$$\sum_{j=1}^{k}\sum_{i=1}^{k}\mathbb{P}(X_z=j|X_0=i)\mathbb{P}(X_0=i)=\sum_{j=1}^{k}\sum_{i=1}^{k}p_{ij}(z)\pi_i=\sum_{j=1}^{k}\sum_{i=1}^{k}\boldsymbol{\pi}_i\left(\boldsymbol{P}^*(z)\right)_{ij}, \tag{2.27}$$

where $\boldsymbol{P}^*$ is again the transition probability matrix restricted to the transient states as described earlier in the proof of theorem 2.2. The decomposition from Bladt and Nielsen (2017), lemma 3.1.6, allows for recasting the transition matrix as a matrix-exponential of the sub-generator $\boldsymbol{T}$, which implies that

$$\sum_{j=1}^{k}\sum_{i=1}^{k}\boldsymbol{\pi}_i\left(\boldsymbol{P}^*(z)\right)_{ij}=\sum_{j=1}^{k}\sum_{i=1}^{k}\boldsymbol{\pi}_i\left(e^{\boldsymbol{T}z}\right)_{ij}=\boldsymbol{\pi}e^{\boldsymbol{T}z}\mathbf{1}_k^\top. \tag{2.28}$$

Finally, the distribution function is obtained as

$$F(z)=1-\mathbb{P}(Z>z)=1-\boldsymbol{\pi}e^{\boldsymbol{T}z}\mathbf{1}_k^\top, \quad z\geq 0, \tag{2.29}$$

which coincides with the formula stated in the theorem. □

Theorem 2.3 suggests that the distribution function can have a jump discontinuity at zero. This fact connects back to the short discussion in the paragraph succeeding definition 2.1 about atoms in phase-type distributions. The size of the jump discontinuity is exactly equal to the probability mass allocated to the atom in zero, i.e.

$$\mathbb{P}(Z=0)=\mathbb{P}(Z\leq 0)=F(0)=1-\boldsymbol{\pi}\mathbf{1}_k^\top=\pi_{k+1}, \tag{2.30}$$

where Z and F are as defined in theorem 2.3. The Lebesgue decomposition theorem states that any probability measure can be decomposed into a singular measure, a discrete measure, and a measure absolutely continuous with respect to the Lebesgue measure, cf. e.g. Rudin (1974), section 6.9. Phase-type distributions never have singular components, but have a discrete component whenever $\pi_{k+1}>0$, in which case the atom represents the discrete component of the distribution. If $\pi_{k+1}=1$, the phase-type distribution becomes a degenerate distribution without an absolutely continuous component. This case is obviously not interesting and will be disregarded for the remainder of the thesis.

If a phase-type distribution has an atom, the distribution does not admit a density with respect to the Lebesgue measure. Larsen (2018) adopts a modified Borel-Stieltjes reference measure and calculates the Radon-Nikodym derivative of the image measure induced by the distribution with respect to the modified reference measure using Lebesgue-Stieltjes integration, but for the purposes of this treatise it will suffice to establish the density of the absolutely continuous component with respect to the Lebesgue measure. This is the subject of the next theorem.

Theorem 2.4
Let $Z \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$ and define $f : \mathbb{R} \to \mathbb{R}_0^+$ as the density function of the absolutely continuous component of Z with respect to the Lebesgue measure. Then f has support on the positive real numbers and is given by

$$f(z) = \boldsymbol{\pi} e^{\boldsymbol{T}z} \boldsymbol{t}, \quad z > 0. \tag{2.31}$$

Proof. The Radon-Nikodym theorem [Kallenberg (2002), theorem 2.10] ensures the existence of a density function of the absolutely continuous component of the distribution with respect to the Lebesgue measure. Since the discrete component of the distribution allocates probability mass solely to the point zero, the density is simply found as the derivative of the distribution function, see e.g. page 313 of Pitman (1993):

$$f(z) = F'(z) = -\boldsymbol{\pi} e^{\boldsymbol{T}z} \boldsymbol{T} \mathbf{1}_k^\top = \boldsymbol{\pi} e^{\boldsymbol{T}z} (-\boldsymbol{T} \mathbf{1}_k^\top) = \boldsymbol{\pi} e^{\boldsymbol{T}z} \boldsymbol{t}, \quad z > 0, \tag{2.32}$$

where $\boldsymbol{t}$ is the exit rate vector defined in equation (2.2). □

In Bladt and Nielsen (2017), there is a short remark regarding the range of the density function, where they note that the density function is strictly positive if the phase-type representation is irreducible. This comment essentially reduces to the idea that the density function is strictly positive only if the distribution actually has an absolutely continuous component, i.e. when it is not a degenerate distribution. Since this was assumed in the paragraph below equation (2.30), the remark can safely be disregarded.

The next part of the section examines kernel transforms of the distribution function such as the Laplace transform and the characteristic function. This subject is particularly important for the study of phase-type and matrix-exponential distributions because these classes of distributions can be classified according to the properties of their integral transforms. These characterizations shall be explored in later sections after the transforms have been introduced.

The various integral transforms all uniquely determine the phase-type distribution, but they have different properties and can be useful in different situations. Selecting the appropriate or most suitable transform for a specific problem can be highly dependent on the details of the problem. Different transforms will therefore be encountered throughout the thesis, the first of which will be the Laplace transform.

Theorem 2.5
The Laplace transform $\mathcal{L} : \mathbb{C} \to \mathbb{C}$ of a phase-type distributed random $Z \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$ with the distribution function F defined in theorem 2.3 is given by

$$\mathcal{L}(s) = \pi_{k+1} + \boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-1} \boldsymbol{t}, \tag{2.33}$$

which is well-defined for arguments with $Re(s) > Re(\lambda_{\max})$, where $\lambda_{\max}$ denotes the eigenvalue of $\boldsymbol{T}$ with the largest real part, and in particular for arguments with non-negative real parts.

Proof. The Laplace(-Stieltjes) transform is defined as the Lebesgue-Stieltjes integral

$$\mathcal{L}(s) = \int_0^\infty e^{-sx} dF(x) \tag{2.34}$$

according to Lukacs (1960), section 1.3. Since F is continuous with derivative f for $x > 0$, the Lebesgue-Stieltjes integral evaluates to

$$\int_0^\infty e^{-sx} dF(x) = \left(F(0) - \lim_{x \to 0^-} F(x) \right) + \int_0^\infty e^{-sx} f(x) dx = \pi_{k+1} + \int_0^\infty e^{-sx} f(x) dx. \tag{2.35}$$

The density function from theorem 2.4 is inserted in the above expression, which yields

$$\int_0^\infty e^{-sx} f(x) dx = \int_0^\infty e^{-sx} \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{t} dx = \int_0^\infty \boldsymbol{\pi} e^{-(s\boldsymbol{I}_k - \boldsymbol{T})x} \boldsymbol{t} dx. \tag{2.36}$$

The latter integral is only well-defined when the matrix $(s\boldsymbol{I}_k - \boldsymbol{T})$ is invertible. It suffices to show that all the eigenvalues of the matrix are non-zero since the determinant of the matrix can be found as the product of its eigenvalues. This implies that the matrix $(s\boldsymbol{I}_k - \boldsymbol{T})$ is invertible whenever s is not an eigenvalue of $\boldsymbol{T}$. Under the assumption that s is not an eigenvalue of $\boldsymbol{T}$, it then follows that

$$\int_0^\infty \boldsymbol{\pi} e^{-(s\boldsymbol{I}_k - \boldsymbol{T})x} \boldsymbol{t} dx = \left[-\boldsymbol{\pi} \left(s\boldsymbol{I}_k - \boldsymbol{T} \right)^{-1} e^{-(s\boldsymbol{I}_k - \boldsymbol{T})x} \boldsymbol{t} \right]_0^\infty . \tag{2.37}$$

Citing corollary 3.1.15 of Bladt and Nielsen (2017), the eigenvalues of $\boldsymbol{T}$ all have strictly negative real parts. For all $s \in \mathbb{C}$ such that $\mathrm{Re}(s) > \mathrm{Re}(\lambda_{\max})$, all the eigenvalues of $(s\boldsymbol{I}_k - \boldsymbol{T})$ must therefore have a strictly positive real part. Rewriting the matrix $(s\boldsymbol{I}_k - \boldsymbol{T})$ in terms of its Jordan canonical form, see eq. 4.1 in Bladt and Nielsen (2017), suggests that this condition on the eigenvalues ensures that the matrix-exponential of $-x\left(s\boldsymbol{I}_k - \boldsymbol{T}\right)$ converges to a matrix of zeros when x tends to infinity. In combination, these arguments justify the claim that the Laplace transform is convergent (well-defined) for all arguments with $\mathrm{Re}(s) > \mathrm{Re}(\lambda_{\max})$. Thus, for any argument with $\mathrm{Re}(s) > \mathrm{Re}(\lambda_{\max})$, the RHS of equation (2.37) is computed as

$$\left[-\boldsymbol{\pi} \left(s\boldsymbol{I}_k - \boldsymbol{T} \right)^{-1} e^{-(s\boldsymbol{I}_k - \boldsymbol{T})x} \boldsymbol{t} \right]_0^\infty = \boldsymbol{\pi} (s\boldsymbol{I}_k - \boldsymbol{T})^{-1} \boldsymbol{t}. \tag{2.38}$$

The two previous equations showcase the property that the Laplace transform of a matrix-exponential of a matrix amounts to the resolvent of said matrix. In summary, equations (2.35) to (2.38) produce the Laplace transform

$$\mathcal{L}(s) = \pi_{k+1} + \boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-1} \boldsymbol{t} \tag{2.39}$$

with $\mathrm{Re}(\lambda_{\max})$ as the abscissa of convergence. □

A similar derivation in Larsen (2018) applies probabilistic arguments to evaluate the Laplace transform as an expectation of a function in terms of a phase-type distributed random variable. While the two methods produce comparable derivations, the above approach considering the Laplace transform as a Laplace-Stieltjes transform of the distribution function is more concise.

The moment-generating function associated with a phase-type distribution can also be obtained through theorem 2.5 by changing the sign of the argument and adjusting the domain appropriately. Furthermore, the characteristic function can be considered as the restriction of the Laplace transform to the imaginary axis in the complex plane. These integral transforms provide information about the moments of phase-type distributions, which can be established through simple probabilistic arguments. The next theorem identifies the higher order derivatives of the Laplace transform, which shall be used to calculate the aforementioned moments.

Theorem 2.6
The higher order derivatives of the Laplace transform $\mathcal{L}$ defined in theorem 2.5 are given by

$$\mathcal{L}^{(n)}(s) = (-1)^n n! \boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+1)}\boldsymbol{t}, \quad n \in \mathbb{N}, \tag{2.40}$$

which is well-defined whenever the Laplace transform $\mathcal{L}$ is well-defined, i.e. it is well-defined in the same domain as $\mathcal{L}$.

Proof. The theorem shall be proved using the principle of induction. It is assumed throughout the proof that s is in the region of convergence of the Laplace transform such that the transform is convergent and the matrix $(s\boldsymbol{I}_k - \boldsymbol{T})$ is non-singular. Furthermore, the Laplace transform is analytic in the region of convergence, cf. Lukacs (1960), page 132, which implies that its derivative always exists in the region of convergence.

For the base case $n = 1$, the induction hypothesis is shown through direct calculation of the derivative of the Laplace transform. This computation requires an auxiliary result from Selby (1974) on the derivative of inverse matrices:

$$\frac{d\boldsymbol{M}^{-1}}{dx} = -\boldsymbol{M}^{-1}\frac{d\boldsymbol{M}}{dx}\boldsymbol{M}^{-1}, \tag{2.41}$$

which actually follows immediately from the product rule of differentiation applied to the matrix identity $\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{I}$. The auxiliary result yields the derivative

$$\frac{d\mathcal{L}}{ds}(s) = \frac{d}{ds}\left(\pi_{k+1} + \boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-1}\boldsymbol{t}\right) = \boldsymbol{\pi}\left(s\boldsymbol{I}_k - \boldsymbol{T}\right)^{-1}\boldsymbol{I}_k\left(s\boldsymbol{I}_k - \boldsymbol{T}\right)^{-1}\boldsymbol{t} = \boldsymbol{\pi}\left(s\boldsymbol{I}_k - \boldsymbol{T}\right)^{-2}\boldsymbol{t}, \tag{2.42}$$

which agrees with the induction hypothesis. Assume next that the induction hypothesis is true for some arbitrary $n \in \mathbb{N}$ and consider the derivative

$$\mathcal{L}^{(n+1)}(s) = \frac{d}{ds}\mathcal{L}^{(n)}(s) = \frac{d}{ds}\left((-1)^n n!\boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+1)}\boldsymbol{t}\right) = (-1)^n n!\boldsymbol{\pi}\frac{d}{ds}\left((s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+1)}\right)\boldsymbol{t}. \tag{2.43}$$

Selby (1974) provides the following formula for the derivative of a matrix product

$$\frac{d}{dx}\left(\boldsymbol{M}_1\boldsymbol{M}_2\right) = \left(\frac{d}{dx}\boldsymbol{M_1}\right)\boldsymbol{M}_2 + \left(\frac{d}{dx}\boldsymbol{M_2}\right)\boldsymbol{M}_1. \tag{2.44}$$

A simple recurrence argument using this result repeatedly together with the fact that the derivative of $(s\boldsymbol{I}_k - \boldsymbol{T})$ with respect to s is the identity matrix shows that the chain rule from standard calculus applies to the derivative in equation (2.43). Hence,

$$(-1)^n n!\boldsymbol{\pi}\frac{d}{ds}\left((s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+1)}\right)\boldsymbol{t} = (-1)^n n!\boldsymbol{\pi}\left(-(n+1)(s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+2)}\right)\boldsymbol{t}. \tag{2.45}$$

Collecting terms and simplifying the expression gives the derivative

$$\mathcal{L}^{(n+1)}(s) = (-1)^{n+1}(n+1)!\boldsymbol{\pi}(s\boldsymbol{I}_k - \boldsymbol{T})^{-(n+2)}\boldsymbol{t}, \tag{2.46}$$

which aligns with the induction hypothesis. By the principle of weak induction, the induction hypothesis holds for all $n \in \mathbb{N}$, which in turn completes the proof. □

The result in theorem 2.6 gives rise to a convenient method of calculating the moments. The Laplace transform given in equation (2.34) has the property that

$$\mathcal{L}(s) = \int_0^\infty e^{-sx} dF(x) = \mathbb{E}\left[e^{-sZ}\right], \tag{2.47}$$

where Z is as defined in theorem 2.5. Since the Laplace transform and its derivatives are holomorphic in a neighborhood around the origin, the moments are related to the higher order derivatives through the relationship

$$(-1)^n \mathcal{L}^{(n)}(0) = \mathbb{E}\left[Z^n\right] = \int_0^\infty x^n dF(x), \tag{2.48}$$

cf. pages 412-413 in Feller (1966). The result is derived from the Leibniz integral rule for improper integrals, where the dominated convergence theorem justifies the differentiation under the integral sign, i.e.

$$\mathcal{L}^{(n)}(s) = \frac{d^n}{ds^n}\left(\int_0^\infty e^{-sx} dF(x)\right) = \int_0^\infty \left(\frac{\partial^n}{\partial s^n} e^{-sx}\right) dF(x) = \int_0^\infty (-x)^n e^{-sx} dF(x), \tag{2.49}$$

where the latter expression evaluated in $s = 0$ is recognized as the n'th moment of F multiplied with negative one raised to the power of n. The formulas for the expectation and the variance of a phase-type distribution therefore follow from theorem 2.6 as corollaries.

Corollary 2.1
The moments of a random variable $Z \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$ are given by

$$\mathbb{E}\left[Z^n\right] = n! \boldsymbol{\pi} (-\boldsymbol{T})^{-n} \mathbf{1}_k^\top, \quad n \in \mathbb{N}. \tag{2.50}$$

In particular, the expectation takes the form

$$\mathbb{E}\left[Z\right] = \boldsymbol{\pi} (-\boldsymbol{T})^{-1} \mathbf{1}_k^\top = \boldsymbol{\pi} \boldsymbol{U} \mathbf{1}_k^\top, \tag{2.51}$$

while the variance is found as

$$\mathbb{V}\left[Z\right] = 2\boldsymbol{\pi} (-\boldsymbol{T})^{-2} \mathbf{1}_k^\top - \left(\boldsymbol{\pi} (-\boldsymbol{T})^{-1} \mathbf{1}_k^\top\right)^2 = 2\boldsymbol{\pi} \boldsymbol{U}^2 \mathbf{1}_k^\top - \left(\boldsymbol{\pi} \boldsymbol{U} \mathbf{1}_k^\top\right)^2, \tag{2.52}$$

where $\boldsymbol{U}$ is the previously defined Green matrix.

Proof. The formula in equation (2.50) emerges as a direct consequence of theorem 2.6 and the result in equation (2.48):

$$\mathbb{E}\left[X^n\right] = (-1)^n \mathcal{L}^{(n)}(0) = n! \boldsymbol{\pi} (-\boldsymbol{T})^{-(n+1)} \boldsymbol{t}. \tag{2.53}$$

This result is equivalent with that in the corollary as eq. (2.2) gives that $(-\boldsymbol{T})^{-1}\boldsymbol{t} = \mathbf{1}_k^\top$. The expectation is simply a special instance of the result in (2.50), and the variance formula is derived as the second central moment, see e.g. the computational formula for variance in Pitman (1993). □

Theorem 2.6 and the subsequent analysis show that a phase-type distribution possesses all moments. This property is noteworthy in several regards, which will be illustrated and emphasized later in the thesis. The property further presents a natural segue to the notion of moment distributions.

Moment distributions shall be discussed comprehensively in paper A, but we will nonetheless provide a brief introduction to the subject here. Moment distributions can be defined in terms of either Laplace transforms or density functions. The former representation defines the n'th order moment distribution (with distribution function F_n) of a non-negative random variable $X \in L^n(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F through the Laplace transform

$$\mathcal{L}[F_n](s) = \int_0^\infty e^{-sx} dF_n(x) = \frac{(-1)^n \mathcal{L}^{(n)}[F](s)}{\mathbb{E}[X^n]}. \tag{2.54}$$

The condition placed on the random variable X is merely a technicality ensuring that the n'th moment of X is finite, which in turn also ensures that the Laplace transform can be differentiated n times. It can be shown that the kernel transformation in equation (2.54) is associated with a phase-type distribution, but the proof relies on some theorems that will only be developed later. Additionally, the phase-type representations of the moment distributions can be established through renewal theory, which we shall also expand on in paper A. The two aforementioned results can also be found in Bladt and Nielsen (2017) as theorems 4.4.19 and 5.5.3, respectively. The main conclusion from these results is that phase-type distributions are closed under formation of moment distributions. This feature of phase-type distributions is applicable e.g. in economics, where the first moment distribution is used to calculate Lorenz curves and Gini indices that quantify the income inequality in society. Section 4.4.5 in Bladt and Nielsen (2017) is an excellent reference on this topic as it contains both the necessary derivations and explicit numerical examples. Alternatively, the interested reader may consult the original work in Bladt and Nielsen (2011).

2.3 Properties

Phase-type distributions have several useful properties similar to the one described at the end of the previous section. This section covers some main properties of phase-type distributions including closure and denseness properties. The section is also concerned with some theorems on characterization and order statistics of phase-type distributions, which are closely related to the main contributions of the dissertation. The closure properties discussed in this section are useful when formulating stochastic models for complex systems involving multiple random variables, and while the denseness property does not inherit the same applicability, it does showcase how phase-type distributions can approximate all other non-negative distributions.

The proofs in this section are heavily reliant on probabilistic arguments utilizing the structure induced by the underlying framework of Markov processes. Consequently, the value of the Markovian foundation becomes increasingly evident throughout this section. This is exemplified in the first theorem on convolutions of phase-type distributions.

Theorem 2.7
Let $Z_1 \sim PH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ and $Z_2 \sim PH_n(\boldsymbol{\beta}, \boldsymbol{T})$ be independent. The convolution $Z = Z_1 + Z_2$ then follows a phase-type distribution of dimension $m+n$ with initial distribution $\boldsymbol{\pi} = (\boldsymbol{\alpha}, \alpha_{m+1}\boldsymbol{\beta})$ and sub-generator

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} \\ \boldsymbol{0}_{n\times m} & \boldsymbol{T} \end{pmatrix}, \tag{2.55}$$

where $\boldsymbol{s}$ denotes the exit rate vector associated with underlying Markov process of Z_1.

Proof. We shall first give a heuristic argument that demonstrates how the Markov structure underlying the phase-type distributions can be used constructively to deduce the proposed representation. The found representation will then be verified analytically using the Laplace transform.

The heuristic approach is based on the concept of process concatenation as described in theorem 14.8 of Sharpe (1988). The convolution $Z = Z_1 + Z_2$ can be considered as the total combined time it takes for the two underlying Markov processes of Z_1 and Z_2 to be absorbed, and by concatenating the two underlying processes, we construct a new process whose time until absorption can be modelled by the convolution. The state-space of the concatenated process contains a single absorbing state and two compartments of transient states, where the transient states of the two underlying processes constitute the respective compartments. The resulting state-space thus has $m + n$ transient states along with the absorbing state. The dynamics of the transitions among the transient states of the concatenated process are governed by the sub-generator

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{K} \\ \boldsymbol{0}_{n\times m} & \boldsymbol{T} \end{pmatrix}, \tag{2.56}$$

where $\boldsymbol{K}$ is referred to as the transfer kernel. The idea is then to devise the initial distribution $\boldsymbol{\pi}$ and specify the transfer kernel such that the sojourn times in the respective compartments have the same distributions as Z_1 and Z_2.

The structure of the sub-generator entails that $\boldsymbol{S}$ governs the transitions among the transient states in the first compartment and $\boldsymbol{T}$ governs the dynamics among the transient states of the second compartment. Furthermore, the process can transition from the first compartment to the second compartment according to the transfer kernel, while the process cannot make the converse transition. Consequently, the constructed process can only visit the states in the first compartment if the process is initiated in the first compartment. The initial distribution over the m states in the first compartment must therefore be selected as $\boldsymbol{\alpha}$, i.e. $\pi_i = \alpha_i$ for all $i \in \{1, ..., m\}$. This construction ensures that the process enters the first compartment according to $\boldsymbol{\alpha}$ and evolves within the compartment in accordance with $\boldsymbol{S}$ until it irreversibly transitions to either the second compartment of directly into the absorbing state. The sojourn time in the first compartment is therefore phase-type distributed with initial distribution $\boldsymbol{\alpha}$ and sub-generator $\boldsymbol{S}$ exactly like Z_1. The remaining elements of $\boldsymbol{\pi}$ and the transfer kernel must now be determined such that the constructed process always enters the second compartment according to $\boldsymbol{\beta}$.

The process is not initiated in the first compartment with probability α_{m+1}, in which case the process might be initiated in one of the n states in the second compartment according to $\boldsymbol{\beta}$. The last n elements of the initial distribution $\boldsymbol{\pi}$ are therefore chosen as $\pi_{m+i} = \alpha_{m+1}\beta_i$ for all $i \in \{1, ..., n\}$ to reflect this property. The constructed process might also enter the second compartment after sojourning in the first compartment. The exit rate vector $\boldsymbol{s}$ describes the transition rates out of the states in the first compartment, and since the process should enter the second compartment according to $\boldsymbol{\beta}$ immediately upon exiting the first compartment regardless of which state the process transitions from, the transfer kernel must take the form $\boldsymbol{K} = \boldsymbol{s\beta}$. This is exactly the transfer kernel implied by the construction proposed in the theorem, and in conjunction with the initial distribution $\boldsymbol{\pi}$, it ensures that the process always enters the second compartment according to $\boldsymbol{\beta}$. After entering the second compartment, the process will sojourn within the second compartment under the dynamics specified by $\boldsymbol{T}$ before it is eventually absorbed. In conclusion, the total sojourn time in the second compartment must be phase-type distributed with initial distribution $\boldsymbol{\beta}$ and sub-generator $\boldsymbol{T}$, which is identical to the distribution of the random variable Z_2. The time until absorption of the constructed process must therefore be distributed completely analogously to the convolution $Z = Z_1 + Z_2$, and we conclude that the convolution has a phase-type distribution with the initial distribution $\boldsymbol{\pi}$ and the sub-generator $\boldsymbol{V}$ as proposed in the theorem.

The proposed construction can also be verified through an analytic argument. The Laplace transform $\mathcal{L}$ of the convolution is naturally given as

$$\mathcal{L}(s) = \mathcal{L}_1(s)\mathcal{L}_2(s), \tag{2.57}$$

where $\mathcal{L}_1$ and $\mathcal{L}_2$ are the Laplace transforms of Z_1 and Z_2, respectively. The abscissa of this transform is simply the larger of the two abscissae of the entering transforms. Theorem 2.5 then gives that

$$\mathcal{L}(s) = \left(\alpha_{m+1} + \boldsymbol{\alpha}(s\boldsymbol{I}_m - \boldsymbol{S})^{-1}\boldsymbol{s}\right)\left(\beta_{n+1} + \boldsymbol{\beta}(s\boldsymbol{I}_n - \boldsymbol{T})^{-1}\boldsymbol{t}\right) \tag{2.58}$$

$$= \alpha_{m+1}\beta_{n+1} + \alpha_{m+1}\boldsymbol{\beta}(s\boldsymbol{I}_n - \boldsymbol{T})^{-1}\boldsymbol{t} + \beta_{n+1}\boldsymbol{\alpha}(s\boldsymbol{I}_m - \boldsymbol{S})^{-1}\boldsymbol{s} + \boldsymbol{\alpha}(s\boldsymbol{I}_m - \boldsymbol{S})^{-1}\boldsymbol{s}\boldsymbol{\beta}(s\boldsymbol{I}_n - \boldsymbol{T})^{-1}\boldsymbol{t}. \tag{2.59}$$

This expression can be compared with the Laplace transform $\mathcal{L}^*$ of a random variable following a phase-type distribution with the initial distribution and sub-generator proposed in the theorem:

$$\mathcal{L}^*(s) = \pi_{m+n+1} + \boldsymbol{\pi}\left(s\boldsymbol{I}_{m+n} - \boldsymbol{V}\right)^{-1}\boldsymbol{v}, \tag{2.60}$$

where the probability of instantaneous absorption is computed as $\pi_{m+n+1} = 1 - \boldsymbol{\pi}\boldsymbol{1}_{m+n}^{\top} = \alpha_{m+1}\beta_{n+1}$, and the exit rate vector becomes

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{s}\beta_{n+1} \\ \boldsymbol{t} \end{pmatrix}. \tag{2.61}$$

The transform in equation (2.60) thus takes the form

$$\mathcal{L}^*(s) = \alpha_{m+1}\beta_{n+1} + \begin{pmatrix} \boldsymbol{\alpha} & \alpha_{m+1}\boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} s\boldsymbol{I}_m - \boldsymbol{S} & -\boldsymbol{s}\boldsymbol{\beta} \\ \boldsymbol{0}_{n\times m} & s\boldsymbol{I}_n - \boldsymbol{T} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{s}\beta_{n+1} \\ \boldsymbol{t} \end{pmatrix}. \tag{2.62}$$

Since $\boldsymbol{V}$ is a sub-generator, the inverse matrices in equations (2.60) and (2.62) exist. Lemma A.1.1 in Bladt and Nielsen (2017) further provides a formula for the inverse of a block matrix, which is applied to the block matrix in eq. (2.62)

$$\mathcal{L}^*(s) = \alpha_{m+1}\beta_{n+1} + \begin{pmatrix} \boldsymbol{\alpha} & \alpha_{m+1}\boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} (s\boldsymbol{I}_m - \boldsymbol{S})^{-1} & (s\boldsymbol{I}_m - \boldsymbol{S})^{-1}\boldsymbol{s}\boldsymbol{\beta}(s\boldsymbol{I}_n - \boldsymbol{T})^{-1} \\ \boldsymbol{0}_{n\times m} & (s\boldsymbol{I}_n - \boldsymbol{T})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}\beta_{n+1} \\ \boldsymbol{t} \end{pmatrix}. \tag{2.63}$$

It is a simple exercise to write out the result of the matrix-multiplication and realize that the expressions in equations (2.59) and (2.63) are identical. In order to conclude that the transforms represent identical distributions, the transforms must also have the same abscissa, cf. page 36 of Lukacs (1960). The eigenvalues of the sub-generator $\boldsymbol{V}$ can be identified through the characteristic equation

$$\det\left(\boldsymbol{V} - \lambda\boldsymbol{I}_{m+n}\right) = \det\begin{pmatrix} \boldsymbol{S} - \lambda\boldsymbol{I}_m & \boldsymbol{s}\boldsymbol{\beta} \\ \boldsymbol{0}_{n\times m} & \boldsymbol{T} - \lambda\boldsymbol{I}_n \end{pmatrix} = \det\left(\boldsymbol{S} - \lambda\boldsymbol{I}_m\right)\det\left(\boldsymbol{T} - \lambda\boldsymbol{I}_n\right) = 0, \tag{2.64}$$

which suggests that the eigenvalues of $\boldsymbol{S}$ and $\boldsymbol{T}$ are also eigenvalues of $\boldsymbol{V}$. The latter identity follows from linear systems theory and can be found e.g. in section 9.1.2 of Petersen and Pedersen (2012). The eigenvalue of $\boldsymbol{V}$ with the largest real part is therefore whichever eigenvalue of $\boldsymbol{S}$ and $\boldsymbol{T}$ that has the largest real part, which in turn implies that the abscissae of $\mathcal{L}$ and $\mathcal{L}^*$ are the same. In conclusion, the two Laplace transforms are identical and the convolution must have the proposed phase-type representation. □

The plot in figure 2.2 shows a realization of the concatenated process constructed in the proof of theorem 2.7. As detailed in the proof, the state-space of the constructed process is divided into two compartments, and the sojourn times in the respective compartments represent the two random variables entering the convolution. In the particular realization displayed in figure 2.2, the concatenated process is initiated in the first compartment and evolves within the first compartment until it transitions to second compartment before it is absorbed. Depending on the specific structures of the initial distributions and the sub-generator, the process might be initiated in the second compartment or never visit any of the compartments.

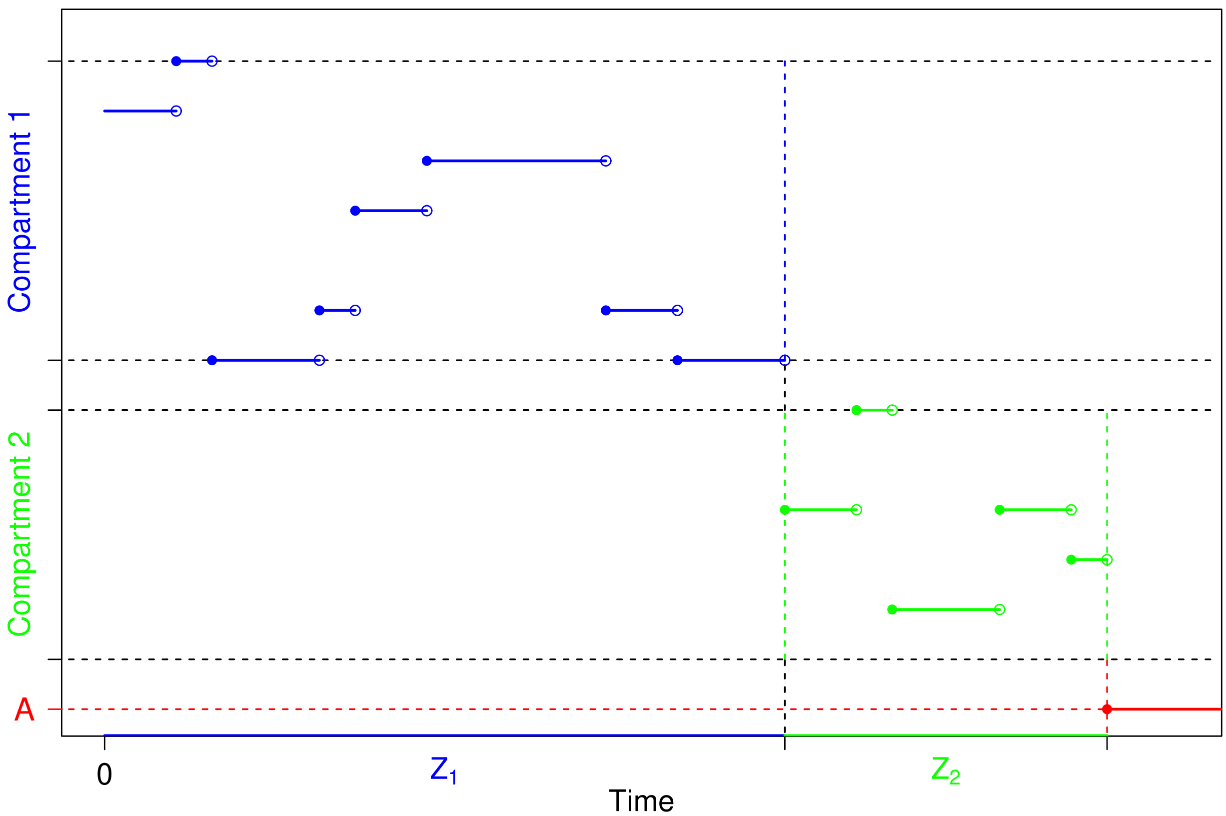

Figure 2.2: *A realization of the concatenated process constructed in the proof of theorem 2.7.*

The scope of theorem 2.7 is restricted to convolutions of independent random variables, however the closure property also holds under certain dependence structures. One particular method of inducing a dependence structure between phase-type distributed random variables is through the transfer kernel which determines how the underlying processes are concatenated. One should however notice that the choice of transfer kernel can effect the distribution of the random variables entering the convolution. A very similar approach is applied in Yera et al. (2021) to model dependent inter-failure times in the context of railway transportation. The paper is concerned with a bivariate model, but the methodology used to model dependence structures carries over to the univariate case. The concept of reward structures, which shall be introduced in a coming section, presents another method of creating dependence structures under which the closure property also applies. This shall be explored further in a later chapter. The next corollary shows that the closure property also extends to finite convolutions.

Corollary 2.2

Any finite convolution of independent phase-type distributed random variables has a phase-type distribution.

Proof. The corollary follows directly from a simple induction argument. Theorem 2.7 covers the base case showing that the convolution of two independent phase-type distributed random variables follows a phase-type distribution. Then assuming that any convolution of n independent phase-type distributed random variables has a phase-type distribution, any $(n+1)$-fold convolution of independent phase-type distributions can be considered a convolution of two independent phase-type distributions, which is again a phase-type distribution. By the principle of induction, the induction hypothesis (the corollary) must hold, and the proof is complete. □

In addition to being closed under finite convolutions, the class of phase-type distributions is also closed under finite mixtures. The mixture properties are similar to the convolution properties and they are derived using the same basic principles, but the mixture properties actually extend further than the convolution properties. An example is that phase-type distributions can be closed under infinite mixtures, i.e. they can arise as compound distributions. However, phase-type distributions are in general not closed under arbitrary infinite mixtures. A concrete counterexample can be obtained by considering a Poisson mixture of Erlang distributions, which results in a compound distribution with a non-rational Laplace transform, cf. problem 3.5.20 in Bladt and Nielsen (2017). For reasons that will be explained in later sections, this is sufficient to argue that the resulting compound distribution cannot be a phase-type distribution.

The next results establish the properties discussed above, the first of which concerns the mixture of two independent phase-type distributions. The presentation of the proofs do however require some preliminaries regarding conditional distributions. The distribution function F of a random variable $Y \sim \mathrm{PH}_k(\boldsymbol{\pi}, \boldsymbol{T})$ is given in equation (2.23) under theorem 2.3 and can be rewritten as

$$F(x) = 1 - ||\boldsymbol{\pi}||_1 + ||\boldsymbol{\pi}||_1 - \boldsymbol{\pi} e^{\boldsymbol{T}x}\mathbf{1}_k^\top = \pi_{k+1} + ||\boldsymbol{\pi}||_1 \left(1 - \frac{\boldsymbol{\pi}}{||\boldsymbol{\pi}||_1} e^{\boldsymbol{T}x}\mathbf{1}_k^\top\right), \quad x \geq 0, \tag{2.65}$$

where $||\cdot||_1$ refers to the Manhatten norm (the 1-norm), which justifies that $1 - ||\boldsymbol{\pi}||_1 = \pi_{k+1}$. By comparing this expression with

$$\begin{aligned} F(x) = \mathbb{P}(Y \leq x) &= \mathbb{P}(Y \leq x | Y = 0)\mathbb{P}(Y = 0) + \mathbb{P}(Y \leq x | Y > 0)\mathbb{P}(Y > 0) & (2.66) \\ &= \pi_{k+1} + \mathbb{P}(Y \leq x | Y > 0)||\boldsymbol{\pi}||_1, \quad x \geq 0, & (2.67) \end{aligned}$$

it follows that the conditional distribution of Y given that Y is positive is a phase-type distribution with initial distribution $\boldsymbol{\pi}^*$ and sub-generator $\boldsymbol{T}$, where $\boldsymbol{\pi}^* = \boldsymbol{\pi}/||\boldsymbol{\pi}||_1$ is the normalized initial distribution.

This result makes sense intuitively as a phase-type distributed random variable is positive if, and only if, the underlying process is initiated in a transient state. This condition can be taken into account when specifying the phase-type representation exactly by normalizing the initial distribution to ensure that the process cannot be absorbed immediately. The close relationship between the distributions (the base and the conditional) can often be utilized in conjunction with the law of total probability or the law of iterated expectation as seen in the below proof.

Theorem 2.8

Let $Z_1 \sim PH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ and $Z_2 \sim PH_n(\boldsymbol{\beta}, \boldsymbol{T})$ be independent and denote their respective distribution functions by F_1 and F_2. The mixture Z with distribution function F defined as

$$F(x) = p_1 F_1(x) + p_2 F_2(x), \tag{2.68}$$

where $p_1 = \mathbb{P}(Z = Z_1)$ and $p_2 = \mathbb{P}(Z = Z_2)$ such that $p_1 + p_2 = 1$, then follows a phase-type distribution of dimension $m + n$ with initial distribution $\boldsymbol{\pi} = (p_1\boldsymbol{\alpha}, p_2\boldsymbol{\beta})$ and sub-generator

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{0}_{m\times n} \\ \boldsymbol{0}_{n\times m} & \boldsymbol{T} \end{pmatrix}. \tag{2.69}$$

Proof. This proof is based on the same principle of process concatenation as appeared earlier. The process concatenation produces a new Markov process with two compartments of transient states consisting of m and n states, respectively. In contrast to the previous proof, we shall not expound upon the method of constructively deducing the required initial distribution and sub-generator, but rather briefly explain how the proposed construction yields the mixture distribution corresponding to the distribution function in equation (2.68).

This distribution function of the mixture states that

$$F(x) = p_1\mathbb{P}(Z_1 \le x) + p_2\mathbb{P}(Z_2 \le x), \tag{2.70}$$

and by applying the law of total probability twice, it takes the form

$$F(x) = p_1\left[\alpha_{m+1} + \mathbb{P}(Z_1 \le x | Z_1 > 0)(1 - \alpha_{m+1})\right] + p_2\left[\beta_{n+1} + \mathbb{P}(Z_2 \le x | Z_2 > 0)(1 - \beta_{n+1})\right], \ x \ge 0, \tag{2.71}$$

which simplifies to

$$F(x) = (p_1\alpha_{m+1} + p_2\beta_{n+1}) + p_1||\boldsymbol{\alpha}||_1\mathbb{P}(Z_1 \le x | Z_1 > 0) + p_2||\boldsymbol{\beta}||_1\mathbb{P}(Z_2 \le x | Z_2 > 0), \quad x \ge 0. \tag{2.72}$$

The appropriate construction should reflect this composition of the mixture. The proposed initial distribution indicates that the constructed process can be initiated in the first (upper) compartment of transient states with probability $p_1||\boldsymbol{\alpha}||_1$ and in the second (lower) compartment with probability $p_2||\boldsymbol{\beta}||_1$, which leaves a probability of immediate absorption amounting to $p_1\alpha_{m+1} + p_2\beta_{n+1}$. Because the proposed sub-generator does not allow for the constructed process to transition between the compartments, the process will only visit the states of the compartment in which the process is initiated. In summary, the proposed phase-type representation suggests that the constructed process can evolve in three distinct ways: The process can be absorbed immediately with probability $p_1\alpha_{m+1} + p_2\beta_{n+1}$, the process can be initialized in the first compartment with probability $p_1||\boldsymbol{\alpha}||_1$ and evolve within this compartment according to $\boldsymbol{S}$ until it is absorbed, or with probability $p_2||\boldsymbol{\beta}||_1$, the process is initiated in the second compartment and evolves there in accordance with $\boldsymbol{T}$ before it is eventually absorbed. Reasoning with the arguments deduced in the latter paragraphs of the previous page, given the constructed process is initiated in the first compartment, the sojourn in the first compartment, and by extension in all the transient states, has a phase-type distribution with initial distribution $\boldsymbol{\alpha}^*$ and sub-generator $\boldsymbol{S}$, which is exactly characterized by the conditional probability in the second term of the RHS of equation (2.72). Due to the symmetric nature of the proposed construction, the conditional probability appearing in the third term of the RHS in equation (2.72) represents the sojourn time in the transient states (the second compartment) given the constructed process is initiated in the second compartment. These arguments justify the conclusion that the constructed process has a time until absorption with the distribution function in equation (2.72), and thus the proposed construction replicates the mechanisms associated with the desired mixture distribution.

Finally, it is worthwhile noting that arguments about the Laplace transform similar to those used in the previous proof can be applied to verify that the proposed construction indeed yields the sought after mixture distribution. □

The generalization of theorem 2.8 to finite mixtures is completely analogous to the extension of theorem 2.7 to corollary 2.2, and therefore we shall give the following corollary without proof.

Corollary 2.3
Any finite mixture of independent phase-type distributed random variables has a phase-type distribution.

Corollary 2.2 states that any finite convolution of independent phase-type distributions has a phase-type distribution, but it does not explicitly provide an adequate phase-type representation. Such a representation can of course be found by repeatedly applying theorem 2.7, but it is often cumbersome to write out the resulting representation explicitly. Conversely, the representation of a finite mixture is readily obtained due to the simple structures of the initial distribution and the sub-generator.

Phase-type representations are however not unique, and there might exist representations of lower dimensions than those produced by the preceding theorems, and example 2.1 presents an illustrative case, where a lower dimensional representation is available.

Example 2.1
Let $Z_1 \sim \text{Erlang}(2,1)$ and $Z_2 \sim \text{Erlang}(3,1)$ be independent. Then the two random variables are phase-type distributed with initial distributions $\boldsymbol{\alpha} = (1,0)$ and $\boldsymbol{\beta} = (1,0,0)$, and sub-generators

$$\boldsymbol{S} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}, \tag{2.73}$$

respectively. Following theorem 2.8, a mixture that selects one of the two random variables with equal probability follows a phase-type distribution with initial distribution $\boldsymbol{\pi} = (0.5, 0, 0.5, 0, 0)$ and sub-generator

$$\boldsymbol{V} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.74}$$

An alternative representation of lower dimensions emerges when utilizing that the two random variables share a common structure in their underlying Markov processes. This representation has the initial distribution $\boldsymbol{\pi}_* = (1, 0, 0)$ and the sub-generator

$$\boldsymbol{V}_* = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0.5 \\ 0 & 0 & -1 \end{pmatrix}, \tag{2.75}$$

which constructs the mixture by letting a process spend an Erlang$(2, 1)$-distributed amount of time in the two first states and with a 50% chance also an exponential(1)-distributed amount of time in the last state. Essentially, the alternative representation simplifies the mixture to a convolution of an Erlang$(2, 1)$ distribution and an exponential(1) distribution with an atom. □

Example 2.1 showcases how considerations about the Markov structure can suggest a representation of lower dimensions. There are other cases, where the reduction in dimensionality derives from direct probabilistic arguments such as in example 2.2.

Example 2.2
Let $Z \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ with initial distribution $\boldsymbol{\pi} = (1, 0)$ and sub-generator

$$\boldsymbol{T} = \begin{pmatrix} -\lambda & \lambda p \\ \lambda p & -\lambda \end{pmatrix}, \tag{2.76}$$

where $\lambda > 0$ and p is a probability. The Laplace transform of Z is then given by theorem 2.5 as

$$\mathcal{L}(s) = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} s+\lambda & -\lambda p \\ -\lambda p & s+\lambda \end{pmatrix}^{-1} \begin{pmatrix} \lambda(1-p) \\ \lambda(1-p) \end{pmatrix} = \frac{\lambda(1-p)}{\lambda(1-p)+s} \tag{2.77}$$

in the domain of the transform. This is of course the Laplace transform associated with an exponential distribution, which shows that $Z \sim \text{Exp}(\lambda(1-p))$. This finding allows for a much simpler phase-type representation with initial distribution $\boldsymbol{\pi}_* = (1)$ and sub-generator $\boldsymbol{T}_* = (-\lambda(1-p))$. □

The examples prompt the introduction of a notion which encompasses that phase-type representations can have different dimensions. To this end, the order of a phase-type distribution is defined as the smallest dimension found among all the possible representations. The notion of order can be helpful since some phase-type distributions can be over-parameterized, cf. page 138 in Bladt and Nielsen (2017).

Example 2.2 further underlines another point, namely that compound distributions arise naturally in the context of phase-type distributions. The random variable considered in example 2.2 is actually constructed as a compound geometric distribution, in particular as an infinite shape mixture of Erlang distributions, which simplifies to an exponential distribution. One interpretation of the random variable in example 2.2 is as the random sum

$$Z = \sum_{i=1}^{N} X_i, \tag{2.78}$$

where $N \sim \text{Geo}(1-p)$ (on the positive integers) and the random variables X_i for $i \in \mathbb{N}$ are independent and identically distributed with exponential(λ) distributions. It turns out that the closure property extends to a fairly broad class of infinite mixtures of phase-type distributions generated as random sums on the form

(2.78), where N follows a discrete phase-type distribution (see section 2.5). The random sum in example 2.2 is actually the archetypical example of such an infinite mixture since the exponential distribution is a continuous phase-type distribution and the geometric distribution is a discrete phase-type distribution. We now consult the next theorem for a general result.

Theorem 2.9
Let $N \sim DPH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ and, independently of N, define the independent and identically distributed random variables $X_i \sim PH_n(\boldsymbol{\beta}, \boldsymbol{T})$, where $||\boldsymbol{\beta}||_1 = 1$, for all $i \in \mathbb{N}$. Then the (possibly infinite) mixture Z generated by the random sum

$$Z = \sum_{i=1}^{N} X_i \tag{2.79}$$

has a phase-type distribution with initial distribution $\boldsymbol{\pi} = \boldsymbol{\alpha} \otimes \boldsymbol{\beta}$ and sub-generator $\boldsymbol{V} = \boldsymbol{I}_m \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t\beta}$, where the resulting distribution has dimension $k = mn$.

Proof. See the proof of theorem 3.1.28 in Bladt and Nielsen (2017). □

We shall not reproduce the proof of theorem 3.1.28 in Bladt and Nielsen (2017) here, but rather present a brief review. The basic idea of the proof is to construct a Markov process that simultaneously captures the dynamics of the processes underlying N and the random variables X_i. The constructed process is a modified version of the process generating N, where each of the m transient states are replaced with a compartment of n transient states, and the total time of every visit to a compartment follows a PH$(\boldsymbol{\beta}, \boldsymbol{T})$ distribution. The state-space of the new process is thus created by amending $E_N^* \times E_X^*$, which is the Cartesian product of the state-spaces of transient states associated with the processes governing N and the random variables X_i, with an absorbing state and ordering the states lexicographically. The proposed initial distribution entails that the process is initiated in one of the compartments (or the absorbing state) according to $\boldsymbol{\alpha}$ and subsequently in a transient state within the chosen compartment according to $\boldsymbol{\beta}$ (unless the process is initiated in the absorbing state). The suggested sub-generator is composed of two matrices, which together induces the desired dynamics:

$$\boldsymbol{V} = \boldsymbol{I}_m \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t\beta} = \begin{pmatrix} \boldsymbol{T} + s_{11}\boldsymbol{t\beta} & s_{12}\boldsymbol{t\beta} & \cdots & s_{1m}\boldsymbol{t\beta} \\ s_{21}\boldsymbol{t\beta} & \boldsymbol{T} + s_{22}\boldsymbol{t\beta} & \cdots & s_{2m}\boldsymbol{t\beta} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1}\boldsymbol{t\beta} & s_{m2}\boldsymbol{t\beta} & \cdots & \boldsymbol{T} + s_{mm}\boldsymbol{t\beta} \end{pmatrix}. \tag{2.80}$$

The first matrix installs the basic dynamics among the transient states within each compartment, i.e. the dynamics among the states within each compartment are governed by $\boldsymbol{T}$. The second matrix governs the transitions among the compartments and adjusts the basic dynamics within the compartments to account for the possibility that the process governing N can transition from a transient state to itself. The structure of the proposed sub-generator thus ensures that the process transitions between the compartments in accordance with the probabilities in $\boldsymbol{S}$. Furthermore, the matrix $\boldsymbol{t\beta}$ ensures that whenever the process exits a compartment, it will enter the new compartment according to $\boldsymbol{\beta}$ (unless it is absorbed). In summary, the constructed process makes $N \sim \text{DPH}_m(\boldsymbol{\alpha}, \boldsymbol{S})$ visits to different compartments until it is absorbed, and the sojourn time for each visit follows a $\text{PH}_n(\boldsymbol{\beta}, \boldsymbol{T})$ distribution. The time until absorption in the constructed process is therefore exactly modelled by the random sum in equation (2.79).

It has now been established that the class of phase-type distributions possesses attractive closure properties in the context of convolutions and mixtures. The next part examines order statistics of samples from phase-type distributions. Like in the analyses of convolutions and mixtures, we shall initially restrict our attention to the simplest case, where a minimum and a maximum is formed from a sample of two phase-type distributions. The main theorem relating to order statistics is then as follows.

Theorem 2.10
Let $X \sim PH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ and $Y \sim PH_n(\boldsymbol{\beta}, \boldsymbol{T})$ be independent random variables. Then

$$\min(X, Y) \sim PH_{mn}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{S} \oplus \boldsymbol{T}), \tag{2.81}$$

and

$$\max(X, Y) \sim PH_{mn+m+n}\left((\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{\alpha}\beta_{n+1}, \alpha_{m+1}\boldsymbol{\beta}), \begin{pmatrix} \boldsymbol{S} \oplus \boldsymbol{T} & \boldsymbol{I}_m \otimes \boldsymbol{t} & \boldsymbol{s} \otimes \boldsymbol{I}_n \\ \boldsymbol{0}_{m \times mn} & \boldsymbol{S} & \boldsymbol{0}_{m \times n} \\ \boldsymbol{0}_{n \times mn} & \boldsymbol{0}_{n \times m} & \boldsymbol{T} \end{pmatrix} \right), \tag{2.82}$$

where $\boldsymbol{s}$ and $\boldsymbol{t}$ are the exit rate vectors associated with the two distributions.

Proof. Denote the underlying Markov processes of X and Y by $\{X_t; t \geq 0\}$ and $\{Y_t; t \geq 0\}$, respectively, and consider the bivariate Markov process $\{\boldsymbol{Z}_t; t \geq 0\}$ defined as $\boldsymbol{Z}_t = (X_t, Y_t)$ with state-space $E_X \times E_Y$, i.e. the Cartesian product of the state-spaces of the two underlying processes, where it can be assumed that the states are ordered lexicographically without loss of generality. The generator of the bivariate process is then the Kronecker sum of the generators of the two underlying processes, and similarly, the sub-generator containing the transition rates among only the states in $E_X^* \times E_Y^*$ is the Kronecker sum $\boldsymbol{S} \oplus \boldsymbol{T}$. Furthermore, the initial distribution of the process $\{\boldsymbol{Z}_t\}$ over the states in $E_X^* \times E_Y^*$ is calculated directly as

$$\mathbb{P}(\boldsymbol{Z}_0 = (i,j)) = \mathbb{P}(X_0 = i, Y_0 = j) = \mathbb{P}(X_0 = i)\mathbb{P}(Y_0 = j) = \alpha_i \beta_j, \quad i \in E_X^*, \quad j \in E_Y^*, \tag{2.83}$$

which implies that the initial distribution is the Kronecker product $\boldsymbol{\alpha} \otimes \boldsymbol{\beta}$.

The minimum of X and Y can be found by starting the two processes $\{X_t\}$ and $\{Y_t\}$ simultaneously and recording the time until one of the processes is absorbed. The time until absorption is exactly the sojourn time of $\{\boldsymbol{Z}_t\}$ among the states in $E_X^* \times E_Y^*$, and consequently, the minimum of X and Y follows a $\text{PH}_{mn}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{S} \oplus \boldsymbol{T})$ distribution. The derivation of the representation associated with the maximum of X and Y takes the construction of the minimum as its starting point. The idea to create a structure that allows for extracting some additional information from the construction of the minimum. Specifically, the new structure should allow for identifying which component of $\{\boldsymbol{Z}_t\}$ is first absorbed while recording the state of the other component. Based on this information, the process underlying the other component can be restarted in the state it was previously in and continue to evolve until it is itself absorbed, which marks the absorption of the bivariate process and thereby the maximum of X and Y. It is the interplay between the Markov property and the memorylessness of the exponential distribution which ensures that restarting the process will not effect its dynamics nor its time until absorption.

To accommodate these specifications, the state-space of the new structure is divided into three compartments of transient states. The first compartment contains mn states and represents the time, where neither $\{X_t\}$ nor $\{Y_t\}$ has been absorbed. The second compartment contains m states and is only accessed when $\{Y_t\}$ is absorbed before $\{X_t\}$, and conversely, the third compartment, which contains n states, is only accessed when $\{X_t\}$ is absorbed before $\{Y_t\}$. The dynamics between the states within the first compartment are governed by $\boldsymbol{S} \oplus \boldsymbol{T}$, which implies that the underlying process exits the first compartment from a state $(i,j) \in E_X^* \times E_Y^*$ with rate $s_i + t_j$. This reflects the property that the process $\{X_t\}$ is absorbed from said state with rate s_i, while the process $\{Y_t\}$ is absorbed with rate t_j. Therefore, the process can transition from a state (i,j) in the first compartment to the i'th state in the second compartment with rate t_j, after which the process would evolve within the second compartment in accordance with $\boldsymbol{S}$ until absorption. Similarly, the process can transition to the third compartment, after which it would evolve according to $\boldsymbol{T}$ until it is eventually absorbed. The proposed sub-generator in eq. (2.82) induces exactly the described dynamics, where the underlying process transitions to a given compartment dependent on whether $\{X_t\}$ or $\{Y_t\}$ is absorbed first, and the structure of the transition rates to the two latter compartments is a direct consequence of the lexicographical ordering of the states in the first compartment. The proof is completed by observing that the proposed initial distribution in eq. (2.82) dictates that the underlying process must start in the second compartment if $Y = 0$, and in the third compartment if $X = 0$. The initial distribution thus appropriately accounts for possible atoms in the distributions of X and Y. In conclusion, the proposed structure in equation (2.82) is a phase-type representation of the random variable $\max(X,Y)$. □

The above derivation of the minimum is straightforward, whereas the justification of the maximum requires some explanation. There is nevertheless an alternative approach, which leads to a slightly different representation. The alternative representation is easier to obtain, but it is much less intuitive than that given in theorem 2.10. Reusing the notation and terminology of the above proof, denote the generators of $\{X_t\}$ and $\{Y_t\}$ by $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, respectively. The generator of $\{\boldsymbol{Z}_t\}$ is then given as

$$\boldsymbol{\Lambda} \oplus \boldsymbol{\Gamma} = \boldsymbol{\Lambda} \otimes \boldsymbol{I}_{n+1} + \boldsymbol{I}_{m+1} \otimes \boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{S} \oplus \boldsymbol{\Gamma} & \boldsymbol{s} \otimes \boldsymbol{I}_{n+1} \\ \boldsymbol{0}_{(n+1)\times(n+1)m} & \boldsymbol{\Gamma} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S} \oplus \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0}_n & 0 \end{pmatrix} & \boldsymbol{s} \otimes \boldsymbol{I}_{n+1} \\ \boldsymbol{0}_{(n+1)\times(n+1)m} & \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0}_n & 0 \end{pmatrix} \end{pmatrix}. \tag{2.84}$$

Assuming the lexicographical ordering of the states, the last state is $(m+1, n+1)$, which is the absorbing state of the bivariate process. Since the maximum of X and Y can be defined as the time until absorption of $\{\boldsymbol{Z}_t\}$, an appropriate sub-generator for a phase-type representation of the maximum can be found by removing the last row and column of the generator in equation (2.84). The resulting sub-generator, say $\boldsymbol{V}$, would take the form

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{S} \oplus \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0}_n & 0 \end{pmatrix} & \boldsymbol{s} \otimes \begin{pmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_n \end{pmatrix} \\ \boldsymbol{0}_{n\times(n+1)m} & \boldsymbol{T} \end{pmatrix}. \tag{2.85}$$

It is not surprising that the sub-generator $\boldsymbol{V}$ resembles the sub-generator given in theorem 2.10, and in fact the connection between the two sub-generators can be more clear by using the bilinearity of the Kronecker product. The Kronecker sum in the upper left block of the sub-generator in equation (2.85) can be recast as

$$\boldsymbol{S}\oplus\begin{pmatrix}\boldsymbol{T} & \boldsymbol{t}\\ \boldsymbol{0}_n & 0\end{pmatrix} = \boldsymbol{S}\otimes\boldsymbol{I}_{n+1}+\boldsymbol{I}_m\otimes\left(\begin{pmatrix}\boldsymbol{T} & \boldsymbol{0}_n^\top\\ \boldsymbol{0}_n & 0\end{pmatrix}+\begin{pmatrix}\boldsymbol{0}_{n\times n} & \boldsymbol{t}\\ \boldsymbol{0}_n & 0\end{pmatrix}\right) = \boldsymbol{S}\oplus\begin{pmatrix}\boldsymbol{T} & \boldsymbol{0}_n^\top\\ \boldsymbol{0}_n & 0\end{pmatrix}+\boldsymbol{I}_m\otimes\begin{pmatrix}\boldsymbol{0}_{n\times n} & \boldsymbol{t}\\ \boldsymbol{0}_n & 0\end{pmatrix}. \quad (2.86)$$

Inserting this into $\boldsymbol{V}$ gives

$$\boldsymbol{V} = \begin{pmatrix}\boldsymbol{S}\oplus\begin{pmatrix}\boldsymbol{T} & \boldsymbol{0}_n^\top\\ \boldsymbol{0}_n & 0\end{pmatrix}+\boldsymbol{I}_m\otimes\begin{pmatrix}\boldsymbol{0}_{n\times n} & \boldsymbol{t}\\ \boldsymbol{0}_n & 0\end{pmatrix} & \boldsymbol{s}\otimes\begin{pmatrix}\boldsymbol{I}_n\\ \boldsymbol{0}_n\end{pmatrix}\\ \boldsymbol{0}_{n\times(n+1)m} & \boldsymbol{T}\end{pmatrix}, \quad (2.87)$$

and by reordering the states in the two upper blocks, the sub-generator is restructured as

$$\begin{pmatrix}\boldsymbol{S}\oplus\begin{pmatrix}\boldsymbol{T} & \boldsymbol{0}_n^\top\\ \boldsymbol{0}_n & 0\end{pmatrix}+\boldsymbol{I}_m\otimes\begin{pmatrix}\boldsymbol{0}_{n\times n} & \boldsymbol{t}\\ \boldsymbol{0}_n & 0\end{pmatrix} & \boldsymbol{s}\otimes\begin{pmatrix}\boldsymbol{I}_n\\ \boldsymbol{0}_n\end{pmatrix}\\ \boldsymbol{0}_{n\times(n+1)m} & \boldsymbol{T}\end{pmatrix} \longrightarrow \begin{pmatrix}\boldsymbol{S}\oplus\boldsymbol{T} & \boldsymbol{I}_m\otimes\boldsymbol{t} & \boldsymbol{s}\otimes\boldsymbol{I}_n\\ \boldsymbol{0}_{m\times mn} & \boldsymbol{S} & \boldsymbol{0}_{m\times n}\\ \boldsymbol{0}_{n\times mn} & \boldsymbol{0}_{n\times m} & \boldsymbol{T}\end{pmatrix}. \quad (2.88)$$

This shows that the sub-generator $\boldsymbol{V}$ derived from $\boldsymbol{\Lambda}\oplus\boldsymbol{\Gamma}$ is the same sub-generator as the one given in theorem 2.10 with a different ordering of the states, which means that the sub-generators in theorem 2.10 and in equation (2.85) do indeed generate the same underlying process. In turn, this reaffirms the validity of the theorem and shows how knowledge about the underlying structure can yield a correct representation without direct calculations.

Theorem 2.10 concerns only the maximum and the minimum derived from a sample of two phase-type distributions, however the result can be generalized to include any order statistic of a sample of arbitrary size. One might simply apply theorem 2.10 repeatedly to obtain the representation of a certain order statistic, or alternatively, adjust the arguments in the proof to account for additional random variables. The latter approach produces some intuitive results, which are however inconvenient in terms of notation. These results appear in various sources in the literature, e.g. in section 3.1.7 of Bladt and Nielsen (2017), but note that Bladt and Nielsen work with phase-type distributions without atoms.

The closure properties covered so far in this section give stochastic models with phase-type distributions flexibility and tractability. The properties suggest that phase-type distributed random variables can be combined in several different manners to model complex systems and maintain a phase-type structure. The remainder of this section is concerned with a denseness property phase-type distributions, which states that any distribution with support on the non-negative reals can be approximated arbitrarily well with a phase-type distribution. In its bare formulation, the theorem does not hold much practical value, but it indicates that phase-type distributions might be applicable in data fitting or even in black box modelling.

There is a multitude of different proofs of the denseness property in the literature. Many of these proofs apply similar methodologies and differ only in their presentations. Bladt and Nielsen (2017) contains a long and detailed proof that includes several lemmas of general interest, Breuer and Baum (2005) provides a brief and intuitive proof, and Asmussen (2003) gives a condensed and highly technical proof. The proof presented in this thesis is taken verbatim from Larsen (2018), which is an extended version of Asmussen's proof with additional elaborate comments.

Theorem 2.11

The class of continuous phase-type distributions is dense (in the sense of weak convergence) in the space of distributions on $\mathbb{R}_0^+$. In general, for any distribution on $\mathbb{R}_0^+$ with finite p'th moment $\mu_F^{(p)}$, there exists a sequence of phase-type distributions $\{F_k\}$ such that $F_k \xrightarrow{w} F$ and $\mu_{F_k}^{(q)} \to \mu_F^{(q)}$ for all $q \leq p$.

Proof. Let d denote a metric of weak convergence in the set of all distributions on $\mathbb{R}_0^+$ and define for $p \geq 0$ another metric d_p as

$$d_p(F,G) = d(F,G) + \left|\mu_F^{(p)} - \mu_G^{(p)}\right|, \tag{2.89}$$

for F and G in the set of all distributions on $\mathbb{R}_0^+$, here denoted by $\mathcal{P}$. Since convergence in $\mathcal{L}^p$ implies convergence in distribution and convergence in $\mathcal{L}^q$ for $0 < q \leq p$, see Kallenberg (2002), proposition 4.12, it is sufficient to prove that the class of finite mixtures of Erlang distributions is dense with respect to the metric d_p in the set of distributions on $\mathbb{R}_0^+$ with finite p'th moment, say $\mathcal{P}^{(p)} = \{P \in \mathcal{P} : \mu_P^{(p)} < \infty\}$.

Denote next the set of distributions compactly supported on $(0,a]$ for $a > 0$ by $\mathcal{P}^a$. Consider then a distribution $F \in \mathcal{P}$ and let F_a denote F truncated at a, i.e. F_a is defined as

$$F_a(x) = \frac{F(\min(x,a))}{F(a)}, \quad x > 0. \tag{2.90}$$

This is obviously also a distribution, as distributions are monotonically non-decreasing, bounded, and right-continuous. Naturally, $d(F,F_a) \to 0$ for $a \to \infty$. Thus,

$$\bigcup_{a<\infty} \mathcal{P}^a \tag{2.91}$$

is dense in $\mathcal{P}$. It further follows immediately from Christensen (2010), theorem 5.4.2, that the set of distributions with only finite support on $(0,a]$, say $\mathcal{P}_*^a$, is dense in $\mathcal{P}^a$, with respect to the metric d. However, for $F, G \in \mathcal{P}^a$

$$d(F,G) \leq d_p(F,G) \leq d(F,G) + a^p, \tag{2.92}$$

which shows that the two metrics are equivalent, cf. Hansen (2012), definition 4.1.5. Conclusively, $\mathcal{P}_*^a$ is also dense in $\mathcal{P}^a$ with respect to d_p. Thus, the problem has been simplified to concern the space $\mathcal{P}_*^a$ equipped with the metric d.

Consider an arbitrary distribution $G \in \mathcal{P}_*^a$ and denote the atoms and their weights by $t_1, ..., t_k$ and $h_1, ..., h_k$, respectively. Furthermore, for each atom $i \in \{1, ..., k\}$, consider a sequence of random variables $\{Y_{m,i}; m \in \mathbb{N}\}$ with $Y_{m,i} \sim \text{Erlang}(m, m\lambda_i)$, where $\lambda_i = t_i^{-1}$. The expectation and variance are then given by

$$\mathbb{E}[Y_{m,i}] = \frac{1}{\lambda_i} = t_i \;\;\wedge\;\; \mathbb{V}[Y_{m,i}] = \frac{1}{m\lambda^2} = \frac{t_i^2}{m}. \tag{2.93}$$

Note that the variance tends to zero for $m \to \infty$. Chebyshev's inequality implies that the sequence $\{Y_{m,i}\}_{m\in\mathbb{N}}$ converges weakly (in probability) to a degenerate random variable at t_i, cf. for $Y_{m,i}$

$$\mathbb{P}\left(|Y_{m,i} - \mathbb{E}[Y_{m,i}]| \geq c\right) \leq \frac{\mathbb{V}[Y_{m,i}]}{c^2} \Rightarrow \mathbb{P}\left(|Y_{m,i} - t_i| \geq c\right) \leq \frac{t_i^2}{mc^2}, \tag{2.94}$$

which leads to

$$\mathbb{P}\left(|Y_{m,i} - t_i| \geq c\right) \to 0, m \to \infty, \tag{2.95}$$

i.e. convergence in probability and thus in distribution. Consequently, by defining the finite mixture

$$G_m = \sum_{i=1}^{k} h_i Y_{m,i}, \tag{2.96}$$

it follows that $d(G_m, G) \to 0$ for $m \to \infty$. Moreover, the moments converge and conclusively, $d_p(G_m, G) \to 0$. Therefore, by the initial observations, since G_m belongs to the class of finite mixtures of Erlang distributions, the class is dense in $\mathcal{P}^{(p)}$. Consequently, the class of finite mixtures of Erlang distributions is dense in $\mathcal{P}$, and hence so is the class of phase-type distributions.

To finalize the proof, additional arguments should be given to include distributions on $[0, \infty)$ that are different from their restrictions on $(0, \infty)$. In these cases, the distributions must have an atom at zero with a certain weight, say p_0. Such distributions can be considered a mixture of the degenerate distribution at zero with weight p_0 and a phase-type distribution (finite mixture of Erlang distributions) approximating the absolutely continuous part on $(0, \infty)$ with weight $1 - p_0$. That is a mixture of two phase-type distributions, which by theorem 2.8 is a phase-type distribution. Hence, the class of phase-type distributions is dense in the space of distributions on $\mathbb{R}_0^+$. □

Taken in conjunction with closure properties derived earlier in this section, the denseness property constitutes a considerable part of the main theoretical framework of univariate phase-type distributions. Another important subject of research within the field is characterization of phase-type distributions, and in particular, their relation to matrix-exponential distributions (see section 2.5). Matrix-exponential distributions coincide with the distributions with rational transforms, and since the transform in theorem 2.5 is also rational, researchers studied how to classify and distinguish the two classes of distributions based on their transforms.

Complete characterizations have been provided for certain subclasses of phase-type distributions, for example Dehon and Latouche (1982) characterized a broad class of generalized hyperexponential distributions, while O'Cinneide (1993) characterized a certain class of Coxian distributions. The most general characterization is arguably found in O'Cinneide (1990), where elegant geometric arguments are used to derive the following theorem.

Theorem 2.12
A distribution with rational Laplace-Stieltjes transform is of phase type if, and only if, it is either the point mass at zero, or it has a continuous positive density on the positive reals and its Laplace-Stieltjes transform has a unique pole of maximal real part (which is therefore real).

Proof. See O'Cinneide (1990) or theorem 4.7.45 in Bladt and Nielsen (2017). □

Fackrell (2003) presents an algorithm that resolves the characterization problem for matrix-exponential distributions based on the work by Dehon and Latouche referenced earlier. The algorithm works by reducing the problem to a minimization problem and checking whether the minimum satisfies some conditions.

Even though the algorithm by Fackrell solves the characterization problem, there is still ongoing research trying to establish more efficient necessary and sufficient conditions on the transforms for a rational distribution to be a phase-type distribution, and it remains an open problem to formulate an analytical theorem solving the characterization problem. In the multivariate setting, there are however many more open problems in regard to characterization. The multivariate distributions shall be studied in the next chapter, but first the concept of reward structures must be introduced. Before proceeding to the topic of reward structures, we mention a last result on a minimal characterization of phase-type distribution.

Theorem 2.13
Phase-type distributions constitute the smallest class of distributions on the non-negative reals that satisfy the following four conditions:

1. *The class contains the degenerate distribution at zero.*
2. *The class contains all exponential distributions.*
3. *The class is closed under finite convolutions and mixtures.*
4. *The class is closed under geometric compounding.*

Proof. See theorem 3.2.10 of Bladt and Nielsen (2017). □

Theorems 2.12 and 2.13 combine to yield a rather comprehensive characterization of the phase-type distributions. Theorem 2.13 essentially states that the set of distributions which can arise from mixing, convoluting, and geometrically compounding exponential distributions is exactly the class of phase-type distributions, and theorem 2.12 accurately describes these distributions in terms of their probability functions.

2.4 Reward structures

The notion of reward structures in the context of phase-type distributions was introduced by Kulkarni in 1989 to define a multivariate phase-type distribution. Reward structures can be thought of as transformations of the underlying Markov process using various functionals. Yor (2001) contains ten papers on applications of exponential functionals of Brownian motions and related processes in mathematical finance, which underlines the potential impact of studying functionals of absorbing Markov processes. The reward structures (the functionals) will primarily be used to transform phase-type distributions into other phase-type distributions, but they also offer additional modelling possibilities. The most studied transformations are those arising from simple functionals.

The setup involves a random variable $\tau \sim \mathrm{PH}_k(\boldsymbol{\pi}, \boldsymbol{T})$ and its underlying Markov process $\{X_t; t \geq 0\}$ whose state-space of transient states is denoted by E^*. The reward structure $r : E^* \to [0, \infty)$ then generates a new random variable Y through the transformation

$$Y = \int_0^\tau r(X_t)dt. \tag{2.97}$$

If the reward structure maps all arguments to one, the transformation does not change the distribution, cf.

$$Y = \int_0^\tau r(X_t)dt = \int_0^\tau 1dt = \tau. \tag{2.98}$$

Another simple reward structure emerges when each element in the domain is mapped to a constant. This reward structure is especially important because of its interpretation in terms of linear rewards. To clarify this connection, define for all $i \in E^*$ the random variable Z_i as the total time the process $\{X_t\}$ spends in state i. Then $\tau = \sum_{i\in E^*} Z_i$ and the transformation in equation (2.97) becomes

$$Y = \int_0^\tau r(X_t)dt = \sum_{i\in E^*} r(i)Z_i. \tag{2.99}$$

The random variable Y can thus represent the total accumulated reward earned by the process $\{X_t\}$ prior to its absorption, when the process earns reward at a constant rate in every state. In other words, the accumulated reward increases linearly (piece-wise) with a slope dependent on the state of the process. An example is a linear reward structure which maps one element $j \in E^*$ to one and all other elements in $E^* \setminus \{j\}$ to zero. Applying this structure transforms τ to a random variable Y that represents the total sojourn time of $\{X_t\}$ in state j prior to absorption. As already evident, linear reward structures have many applications and allow for modelling a wide variety of setups, and while it is obvious that the transformations associated with non-negative linear reward structures yield new phase-type distributions, the representations of those distributions are not obvious.

The next theorem gives the distributions that arise exactly from transformations with non-negative linear reward structures. To ease notation, from this point onwards, linear reward structures (in the univariate case) will be represented by vectors which contain the reward rates associated with the different states. Specifically, for the state-space $E^* = \{1, ..., k\}$, the linear reward structure is formulated as a vector $\boldsymbol{r}^\top = (r(1), ..., r(k))$ with elements in $\mathbb{R}_0^+$. Furthermore, the theorem invokes some notation which first must be explained. The state-space E^* can be divided into two disjoint sets: $E^+ = \{i \in E^* | r(i) > 0\}$ and $E^0 = \{i \in E^* | r(i) = 0\}$, i.e. a set containing states with positive reward rates and a set containing states with zero reward rates. If the states are reordered such that $E^* = (E^+, E^0)$, the sub-generator must be adjusted accordingly, which allows for a block-partitioning of the sub-generator into two compartments:

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}^{++} & \boldsymbol{T}^{+0} \\ \boldsymbol{T}^{0+} & \boldsymbol{T}^{00} \end{pmatrix}. \tag{2.100}$$

The superscripts indicate which transition rates belong to the different blocks. The block $\boldsymbol{T}^{++}$ holds the transition rates from states in E^+ to states in E^+, whereas the block $\boldsymbol{T}^{0+}$ contains the transition rates from the states in E^0 to the states in E^+. Similarly, the initial distribution must be changed accordingly to reflect the new ordering of the states, which implies that the initial distribution can be partitioned as $\boldsymbol{\pi} = (\boldsymbol{\pi}^+, \boldsymbol{\pi}^0)$. Using this notation, the next theorem reads:

Theorem 2.14

Consider the random variable $\tau \sim PH_k(\boldsymbol{\pi}, \boldsymbol{T})$ and a linear reward structure $\boldsymbol{r} \in (\mathbb{R}_0^+)^k$. Assuming that $\boldsymbol{T}$ takes the form in (2.100), the random variable Y defined as in equation (2.97) follows a phase-type distribution of dimension $d = |E^+|$ with initial distribution

$$\boldsymbol{\alpha} = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0(\boldsymbol{I}_{k-d} - \boldsymbol{Q}^{00})^{-1}\boldsymbol{Q}^{0+} \tag{2.101}$$

and sub-generator $\boldsymbol{T}^* = \{t^*_{ij}\}_{i,j\in\{1,\dots,d\}}$ *whose elements are given by*

$$t^*_{ij} = \begin{cases} \frac{t_{ii}}{r(i)}(1-p_{ii}), & i = j \\ -\frac{t_{ii}}{r(i)}p_{ij}, & i \neq j \end{cases}, \tag{2.102}$$

where $\boldsymbol{Q}^{00}$ *and* $\boldsymbol{Q}^{0+}$ *are blocks of the matrix* $\boldsymbol{Q}$*, which is defined in the proof, and* $\boldsymbol{P} = \{p_{ij}\}_{i,j\in\{1,\dots,d\}}$ *is a sub-transition probability matrix also defined in the proof.*

Proof. We adopt the setup introduced in the section prior to the theorem, i.e. $\{X_t; t \geq 0\}$ is the underlying process of τ with sub-generator $\boldsymbol{T}$ and initial distribution $\boldsymbol{\pi}$, where it is assumed that the state-space is order into two compartments E^+ and E^0 such that $\boldsymbol{T}$ takes the form in equation (2.100). Define then the pseudo embedded Markov chain $\{\widetilde{X}_n; n \in \mathbb{N}_0\}$ governed by the transition probability matrix $\widetilde{\boldsymbol{Q}}$ constructed as

$$\widetilde{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{q} \\ \boldsymbol{0}_k & 1 \end{pmatrix}, \tag{2.103}$$

where $\boldsymbol{Q}$ is a square matrix of order k with zeros in the diagonal and off-diagonal elements

$$q_{ij} = -\frac{t_{ij}}{t_{ii}}, \quad i,j \in E^*, \tag{2.104}$$

and $\boldsymbol{q}$ is a k-dimensional column vector with elements

$$q_i = 1 - \sum_{j=1}^{k} q_{ij}, \quad i \in E^*. \tag{2.105}$$

Since an embedded Markov chain cannot enter the state it transitions from, an absorbing process does not admit an embedded chain. The process $\{\widetilde{X}_n\}$ is dubbed the pseudo embedded Markov chain because its transition probabilities are calculated like a standard embedded chain except for the transitions out of the state $k+1$, which is treated as a standard absorbing state. By using these definitions, an absorbing Markov process can be assigned a (pseudo) embedded Markov chain with the transition probability matrix $\widetilde{\boldsymbol{Q}}$. Due to the assumed ordering of the state-space, the blocks of $\widetilde{\boldsymbol{Q}}$ can be further partitioned into the blocks:

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}^{++} & \boldsymbol{Q}^{+0} \\ \boldsymbol{Q}^{0+} & \boldsymbol{Q}^{00} \end{pmatrix} \quad \text{and} \quad \boldsymbol{q} = \begin{pmatrix} \boldsymbol{q}^{+} \\ \boldsymbol{q}^{0} \end{pmatrix}. \tag{2.106}$$

The process $\{X_t\}$ does not earn reward when it sojourns in the states of E^0 (per definition). To reflect this property, a new process $\{X^*_n; n \in \mathbb{N}_0\}$ is defined on the state-space $F = E^+ \cup \{k+1\}$. For all $n \geq 1$, the new process is defined as

$$X^*_{n-1} = \widetilde{X}_{M_n}, \tag{2.107}$$

where the process $\{M_n; n \in \mathbb{N}\}$ describes the (discrete) times points, where $\{\widetilde{X}_n\}$ visits the states in F. For instance, M_1 denotes the first point in time, where $\{\widetilde{X}_n\}$ visits a state in F, and M_2 denotes the second point in time, where $\{\widetilde{X}_n\}$ visits a state in F. Although it is not technically true, as $\{M_n\}$ is not a Lévy process, it might be helpful to think of $\{X^*_n\}$ as $\{\widetilde{X}_n\}$ subordinated by $\{M_n\}$. The transition probability matrix of the new process must admit the form

$$\boldsymbol{P}^* = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{p} \\ \boldsymbol{0}_d & 1 \end{pmatrix}, \tag{2.108}$$

where d is the cardinality of E^+. The elements of the sub-transition probability matrix $\boldsymbol{P} = \{p_{ij}\}_{i,j \in E^+}$ are found through direct calculations. It follows immediately from the previous definitions that for all $n \geq 1$

$$p_{ij} = \mathbb{P}\left(X_n^* = j | X_{n-1}^* = i\right) = \mathbb{P}\left(\widetilde{X}_{M_{n+1}} = j | \widetilde{X}_{M_n} = i\right), \quad i,j \in E^+. \tag{2.109}$$

The latter probability can be evaluated using a sum of joint probabilities:

$$\mathbb{P}\left(\widetilde{X}_{M_{n+1}} = j | \widetilde{X}_{M_n} = i\right) = \sum_{m=1}^{\infty} \mathbb{P}\left(\widetilde{X}_{M_{n+1}} = j, M_{n+1} = M_n + m | \widetilde{X}_{M_n} = i\right). \tag{2.110}$$

Recasting the expression in terms of joint probabilities allows for a useful interpretation. The conditional joint probabilities appearing in the sum on the RHS of equation (2.110) are certain m-step transition probabilities of the pseudo embedded Markov chain $\{\widetilde{X}_n\}$. In particular, they are the probabilities that the pseudo embedded process will transition from state i to state j in m steps without visiting states in F in between. Thus,

$$p_{ij} = \sum_{m=1}^{\infty} \mathbb{P}\left(\widetilde{X}_{M_n+m} = j, \widetilde{X}_{M_n+(m-1)} \notin F, ..., \widetilde{X}_{M_n+2} \notin F, \widetilde{X}_{M_n+1} \notin F | \widetilde{X}_{M_n} = i\right). \tag{2.111}$$

These transition probabilities can be calculated as

$$\boldsymbol{P} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left[\sum_{m=2}^{\infty} \left(\boldsymbol{Q}^{00}\right)^{m-2}\right]\boldsymbol{Q}^{0+} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left[\sum_{m=0}^{\infty} \left(\boldsymbol{Q}^{00}\right)^{m}\right]\boldsymbol{Q}^{0+}. \tag{2.112}$$

The first term represents the one-step transition probabilities, and the second term represents transitions over more steps. The multi-step transitions require the underlying process to transition from state i to a state in E^0 and remain in E^0 for a number of steps before transitioning to state j, which explains the structure of the second term. Since the matrix $\boldsymbol{Q}^{00}$ is a sub-generator, the arguments from the proof of theorem 2.5 implies that the matrix $\boldsymbol{I}_{k-d} - \boldsymbol{Q}^{00}$ is invertible. The Neumann series in equation (2.112) is therefore convergent, which allows for rewriting the sub-transition probability matrix as

$$\boldsymbol{P} = \boldsymbol{Q}^{++} + \boldsymbol{Q}^{+0}\left(\boldsymbol{I}_{k-d} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+}. \tag{2.113}$$

The initial distribution $(\boldsymbol{\alpha}, \alpha_{d+1})$ associated with $\{X_n^*\}$ is derived in a completely analogous manner:

$$\boldsymbol{\alpha} = \boldsymbol{\pi}^+ + \boldsymbol{\pi}^0\left(\boldsymbol{I}_{k-d} - \boldsymbol{Q}^{00}\right)^{-1}\boldsymbol{Q}^{0+}. \tag{2.114}$$

The process $\{\widetilde{X}_n\}$ can either be initiated in a state in F or it can be initiated in E^0, in which case the process will eventually visit a state in F almost surely. This property ensures that the elements of the initial distribution sum to one. The atom at zero has size α_{d+1}, which is the probability that the process $\{\widetilde{X}_n\}$ is either absorbed directly upon initiation or absorbed prior to visiting a state in E^+. The initial distribution $(\boldsymbol{\alpha}, \alpha_{d+1})$ and the transition matrix $\boldsymbol{P}^*$ completely determines the process $\{X_n^*\}$, which describes the behaviour of the pseudo embedded Markov chain of $\{X_t\}$ on the states in which it receives positive rewards. This process does not constitute a valid embedded Markov chain as the transition matrix allows for intrastate jumps (jumps from a state to itself). The process can however be modified into a pseudo embedded Markov chain by appropriately adjusting the transition matrix.

A continuous time Markov process $\{Y_t; t \geq 0\}$ on F with the desired properties can therefore be devised based on the information from $\boldsymbol{P}^*$, $\boldsymbol{r}$, and the diagonal elements of $\boldsymbol{T}$. The rates with which the continuous process should exit the different states can be found by adjusting the diagonal elements in $\boldsymbol{T}$ to account for the possibility of intrastate jumps. The diagonal elements of $\boldsymbol{T}$ describe the average sojourn time of any visit of $\{X_t\}$ to the different states, and the diagonal elements of $\boldsymbol{P}$ describe the probabilities of intrastate jumps in $\{X_n^*\}$ for the different states. The sojourn times of the visits of the process $\{Y_t\}$ are therefore given as random sums of exponential distributions. Specifically, for any state $i \in E^+$, the probability that $\{X_n^*\}$ makes an intrastate jump from state i is p_{ii}, and the interjump times are independent and identically distributed with exponential$(-t_{ii})$ distributions. By theorem 2.9 or example 2.2, such a random sum follows an exponential distribution with rate $-t_{ii}(1-p_{ii})$. Since the process receives a linear reward at rate $r(i)$, equation (2.99) shows that this is equivalent to an exponential distribution with rate $-t_{ii}(1-p_{ii})/r(i)$ receiving a linear reward at rate one. As demonstrated in equation (2.98), such a linear reward does not change the distribution, and the visits of $\{Y_t\}$ to the different states are therefore exponentially distributed with these rates. Letting $\boldsymbol{T}^* = \{t^*_{ij}\}_{i,j\in E^+}$ denote the sub-generator of $\{Y_t\}$, the diagonal elements are thus given by

$$t^*_{ii} = \frac{t_{ii}}{r(i)}(1-p_{ii}), \quad \forall i \in E^+. \tag{2.115}$$

The remaining elements of $\boldsymbol{T}^*$ are found by multiplying the respective diagonal elements (with a sign change) with the transition probabilities of the (pseudo) embedded Markov chain. These transition probabilities are derived in a standard manner:

$$\mathbb{P}\left(X_1^* = j | X_0^* = i, X_1^* \neq i\right) = \frac{\mathbb{P}\left(X_1^* = j, X_1^* \neq i | X_0^* = i\right)}{\mathbb{P}\left(X_1^* \neq i | X_0^* = i\right)} = \frac{p_{ij}}{1-p_{ii}}, \quad \forall (i,j) \in E^+ \times E^+, \quad i \neq j. \tag{2.116}$$

Notice that the diagonal elements cannot be one, since this would imply that one or more of the states were absorbing. The transition probabilities calculated in equation (2.116) allow for calculating the off-diagonal elements of $\boldsymbol{T}^*$ as

$$t^*_{ij} = -t^*_{ii}\mathbb{P}\left(X_1^* = j | X_0^* = i, X_1^* \neq i\right) = -\frac{t_{ii}}{r(i)}p_{ij}. \tag{2.117}$$

As the initial distribution in the theorem, eq. (2.101), coincides with the initial distribution established in equation (2.114), and the sub-generator in the theorem, eq. (2.102), coincides with the the transition rates given in equations (2.115) and (2.117), equipping the underlying process of τ with the linear reward structure $\boldsymbol{r}$ does indeed yield the proposed phase-type distribution. □

Reward structures can be informative in various ways. For instance, theorem 2.14 gives rise to a number of corollaries. The next corollary concerns the total sojourn time spent in the transient states prior to absorption.

Corollary 2.4
Let $\{X_t; t \geq 0\}$ be an absorbing Markov process underlying a phase-type distribution. Then the total sojourn times of $\{X_t\}$ in each of the transient states follow exponential distributions with possible atoms.

Proof. Let E^* denote the transient states of $\{X_t\}$, and define for all $i \in E^*$ the random variable Z_i as the total sojourn time of the process in state i prior to absorption. Furthermore, let τ denote the time until absorption of $\{X_t\}$ such that τ has a phase-type distribution. Consider then a linear reward structure which maps all elements of E^* to zero except for one element, say $j \in E^*$, which is mapped to one. This reward structure transforms τ into the random variable

$$\int_0^\tau r(X_t)dt = \sum_{i\in E^*} r(i)Z_i = Z_j. \tag{2.118}$$

Since the reward structure only has one state with positive reward rate, theorem 2.14 states that Z_j follows a phase-type distribution of dimension one, which is an exponential distribution with a possible atom. Since $j \in E^*$ was chosen arbitrarily, the result holds for all $j \in E^*$. Hence, the total sojourn time of $\{X_t\}$ in any transient state prior to absorption follows an exponential distribution with a possible atom. □

The concept of reward structures also allows for extending the result on compound distributions presented in theorem 2.9 to include infinite mixtures of phase-type distributions with atoms, i.e. the introduction of reward structures allows for removing the restriction on the initial distribution in the theorem. This is achieved by inserting transient states with zero reward rates that imitate the behaviour of absorbing states.

Corollary 2.5

The result in theorem 2.9 holds for any continuous phase-type distribution. The representation of the resulting phase-type distribution must however be changed accordingly.

Proof. The setup from theorem 2.9 is adopted, but now the random variables $X_i \sim \mathrm{PH}_n(\boldsymbol{\beta}, \boldsymbol{T})$ are allowed to have atoms. To accommodate this change, the representations of the phase-type distributions are altered by introducing transient state with zero reward rate. In particular, the random variables now have the phase-type representation $(\boldsymbol{\beta}_*, \boldsymbol{T}_*)$ with a reward structure $\boldsymbol{r}_*$, where the new initial distribution is given as $\boldsymbol{\beta}_* = (\boldsymbol{\beta}, \beta_{n+1})$, the new sub-generator is given by

$$\boldsymbol{T}_* = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{0}_n^\top \\ \boldsymbol{0}_n & -1 \end{pmatrix}, \tag{2.119}$$

and the reward structure is $\boldsymbol{r}_*^\top = (r_*(1), ..., r_*(n+1)) = (\boldsymbol{1}_n, 0)$ such that the introduced state has zero reward rate. The associated exit rate vector is also changed accordingly, such that $\boldsymbol{t}_*^\top = (\boldsymbol{t}^\top, 1)$. This representation satisfies the condition given in theorem 2.9, which means that the theorem applies. The infinite mixture will thus have a phase-type representation with the initial distribution $\boldsymbol{\pi} = \boldsymbol{\alpha} \otimes \boldsymbol{\beta}_*$ and the sub-generator $\boldsymbol{V} = \boldsymbol{I}_m \otimes \boldsymbol{T}_* + \boldsymbol{S} \otimes \boldsymbol{t}_*\boldsymbol{\beta}_*$, which is equipped with the reward structure $\boldsymbol{r} = \boldsymbol{1}_m^\top \otimes \boldsymbol{r}_*$. Theorem 2.14 then gives the phase-type representation of the transformed distribution induced by the reward structure. The conclusion is that closure property presented in theorem 2.9 also holds for infinite mixtures of phase-type distributions with atoms, and an appropriate representation can be obtained through theorems 2.9 and 2.14. □

The notion of introducing transient states with zero reward rates to obtain additional information also features prominently in paper A, where the reward structure is used to track the order in which components of random vectors are absorbed. In summary, reward structures allow for analyzing and studying phase-type distributions and present a versatile modelling tool. Before we end this section, we give a short example of a reward structure and illustrate graphically how it transforms one phase-type distribution to another.

Example 2.3
Consider a random variable $\tau \sim \mathrm{PH}_5(\boldsymbol{\alpha}, \boldsymbol{T})$ with the underlying process $\{X_t; t \geq 0\}$ and the reward structure

$$\boldsymbol{r}^\top = (0, 0, 1, 2, 3)\,, \tag{2.120}$$

i.e. $r(1) = 0, r(2) = 0, ..., r(5) = 3$. Define then the random variable Y as

$$Y = \int_0^\tau r(X_t)dt. \tag{2.121}$$

Figures 2.3 and 2.4 show how the reward structure transform the distribution of τ into the distribution of Y.

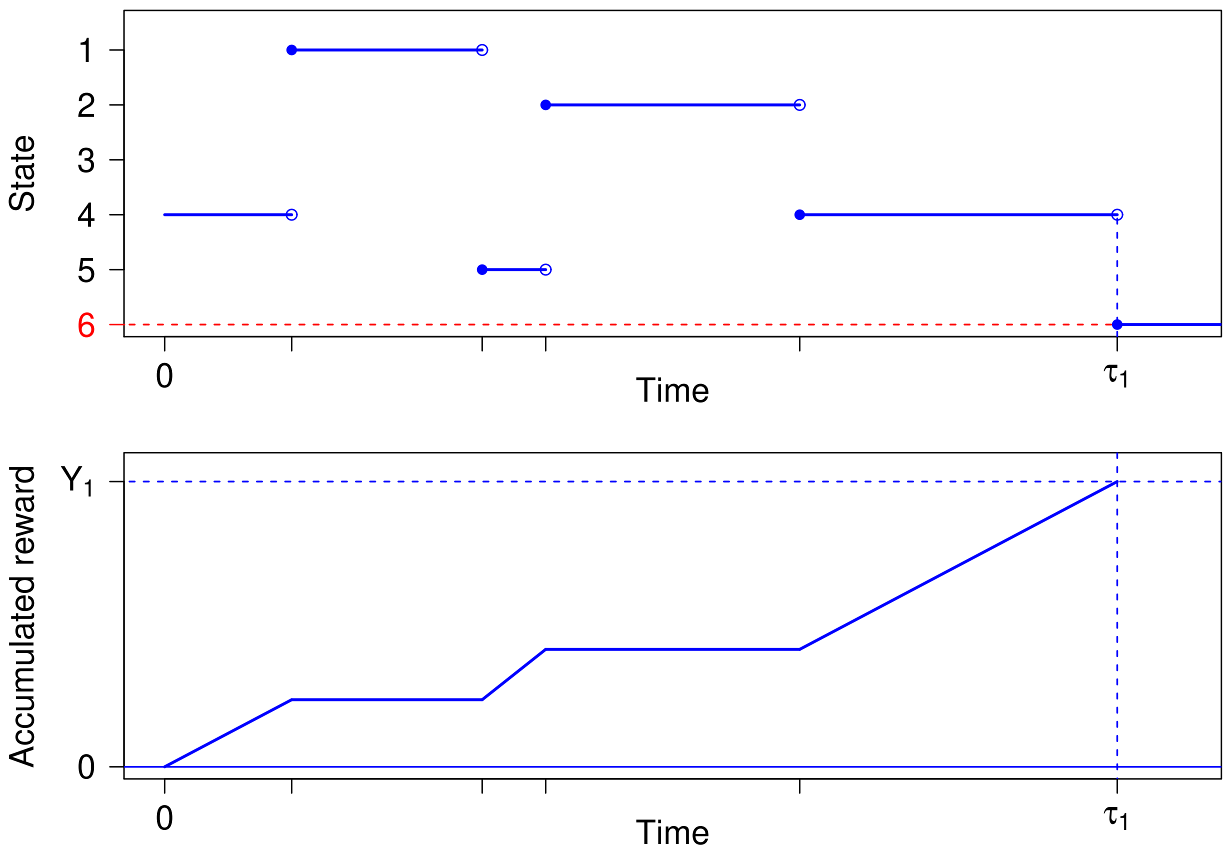

Figure 2.3: *A realization of a random variable generated from a transformation using a reward structure.*

Comparing figures 2.3 and 2.4 underlines a property of the transformations, namely that the relationship between the original random variable τ and the transformed random variable Y is highly complex. Different reward structures can generate vastly different dependence structures between the two random variables τ and Y. □

The study of MPH* distributions is exactly concerned with analyzing and understanding the properties and dependence structures induced by such linear reward structures. The topic of reward structures therefore gives a natural transition from the univariate phase-type distributions to their multivariate relatives.

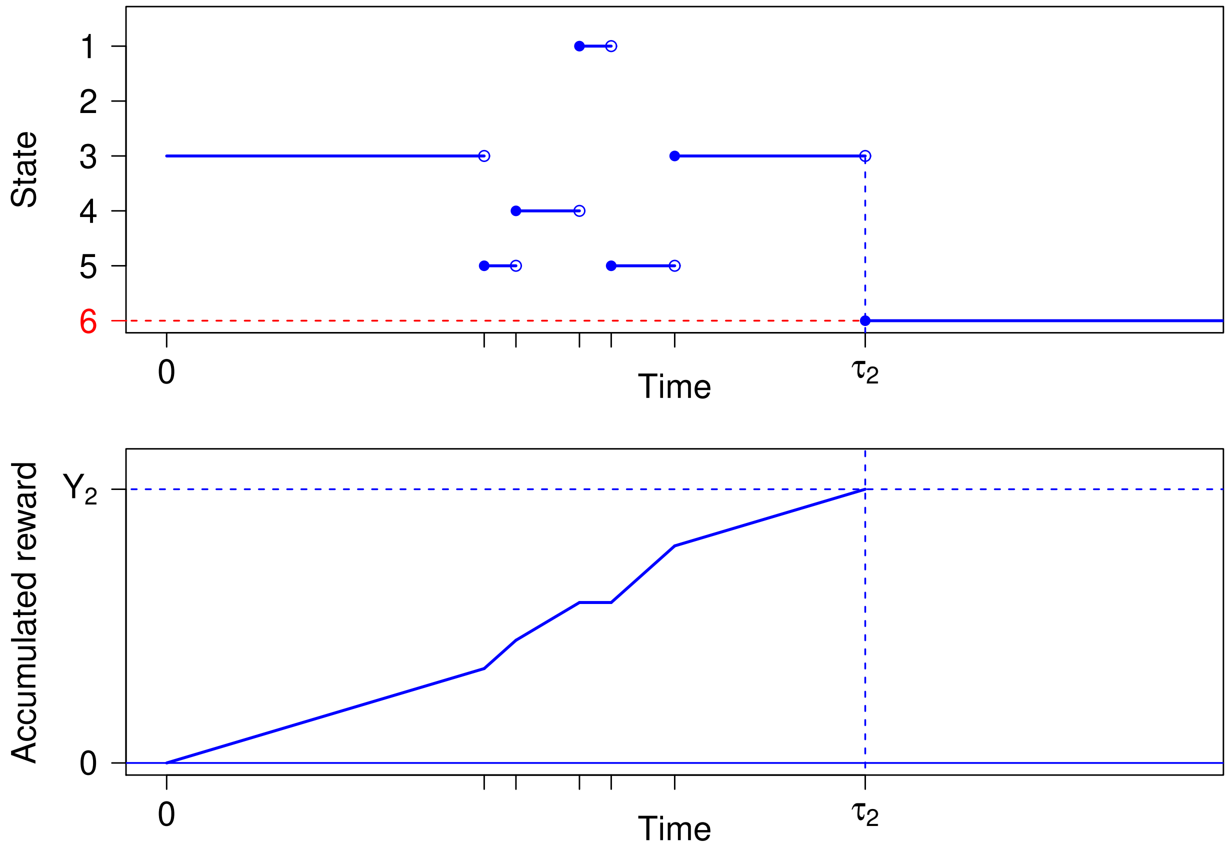

Figure 2.4: *Another realization of a random variable generated from a transformation using a reward structure.*

2.5 Related distributions

The previous sections have covered the most important and fundamental theoretical considerations in regard to the univariate phase-type distributions. The last section of this chapter contains a brief review of some distributions closely related to the phase-type distributions. Some of the examples presented in this section have already been encountered earlier in the thesis, but they are included here for the sake of completeness.

The most basic examples of phase-type distributions of course the exponential distribution and the Erlang distribution. The Erlang distribution is a special instance of the gamma distribution when the shape parameter takes integer (natural) values, and the generalized gamma (Erlang) distribution is also a phase-type distribution when the shape parameter takes integer values. The flowchart associated with a generalized gamma distribution is shown in figure 2.5 below, and the flowchart allows for directly inferring a phase-type structure.

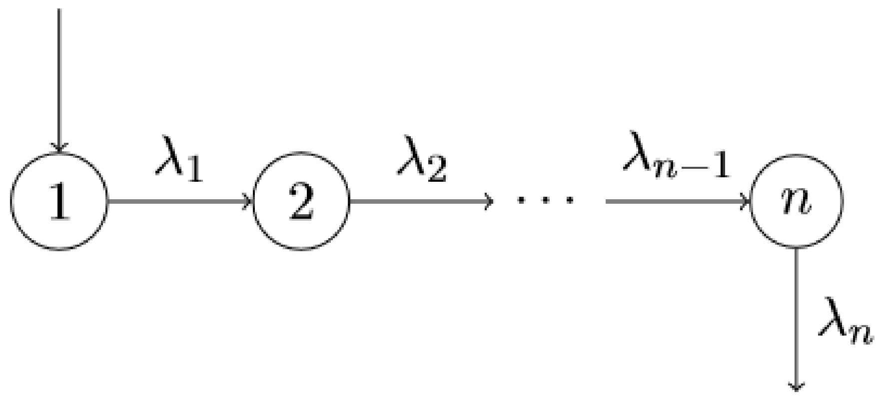

Figure 2.5: *Flowchart of a generalized gamma distribution with integer (natural) shape parameter.*

Example 2.4
A generalized gamma distribution with shape parameter $n \in \mathbb{N}$ and rate parameters $\lambda_1, ..., \lambda_n$ is a phase-type distribution of dimension n with initial distribution $\boldsymbol{\pi} = (1, \mathbf{0}_{n-1})$ and the bidiagonal sub-generator

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & & \\ & -\lambda_2 & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{n-1} & \lambda_{n-1} \\ & & & & -\lambda_n \end{pmatrix}, \tag{2.122}$$

where the empty elements represent zeros. □

The next example is the Coxian distribution, which relaxes the condition of the generalized gamma distribution that the underlying Markov process can only exit to the absorbing state after sequentially traversing the n transient states. The dynamics of the process underlying a Coxian distribution is presented in the flowchart in figure 2.6:

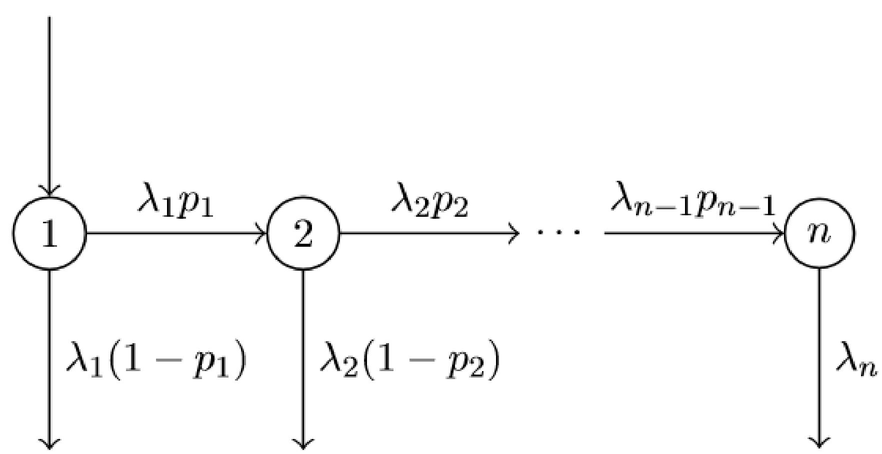

Figure 2.6: *Flowchart of a Coxian distribution.*

Example 2.5

A Coxian distribution with n phases, rate parameters $\lambda_1, ..., \lambda_n$, and probability parameters $p_1, ..., p_n$ is a phase-type distribution with initial distribution $\boldsymbol{\pi} = (1, \mathbf{0}_{n-1})$ and the bidiagonal sub-generator

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_1 & & & \\ & -\lambda_2 & \lambda_2 p_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{n-1} & \lambda_{n-1} p_{n-1} \\ & & & & -\lambda_n \end{pmatrix}, \tag{2.123}$$

where the empty elements represent zeros. □

The generalized Coxian distribution further relaxes the condition of the Coxian distribution that the underlying process can only be initiated in the first state. Consequently, the flowchart and the phase-type representation should be adjusted to reflect these new dynamics. Figure 2.7 shows the updated flowchart and example 2.6 gives the appropriate phase-type representation.

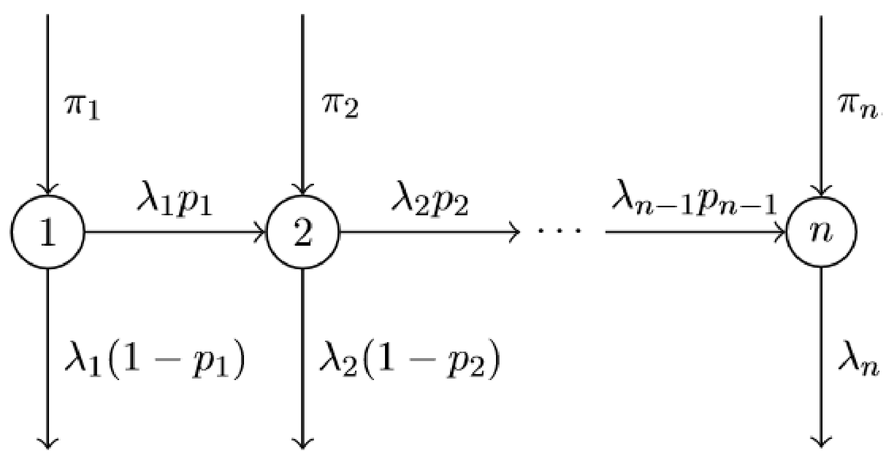

Figure 2.7: *Flowchart of a generalized Coxian distribution.*

Example 2.6

A generalized Coxian distribution with with n phases, rate parameters $\lambda_1, ..., \lambda_n$, probability parameters $p_1, ..., p_n$, and initiation probabilities $\pi_1, ..., \pi_n$ is a phase-type distribution with the obvious initial distribution $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)$ and the same bidiagonal sub-generator as the ordinary Coxian distribution, i.e. with the sub-generator given in equation (2.123). □

When the probability parameters in a generalized Coxian distribution are all zero, i.e. the underlying Markov process cannot transition between the transient states, the distribution reduces to a hyperexponential distribution, which is essentially a mixture of exponential distributions. This case was covered in theorem 2.8 and subsequently in corollary 2.3. However, these results did not explicitly state the phase-type representation in cases with more than two mixture components. The complete representation follows immediately from the aforementioned theorem and is given in the next example.

Example 2.7

A hyperexponential distribution with n components with mixture probabilities $\pi_1, ..., \pi_n$ and rate parameters $\lambda_1, ..., \lambda_n$ is a phase-type distribution with initial distribution $\boldsymbol{\pi} = (\pi_1, ..., \pi_n)$ and the diagonal sub-generator

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_n \end{pmatrix}, \tag{2.124}$$

where the empty elements represent zeros. The associated flowchart reflects this diagonal structure of the sub-generator. The flowchart in figure 2.8 highlights the property that the underlying process cannot transition between the transient states, which exactly characterizes the mixture. □

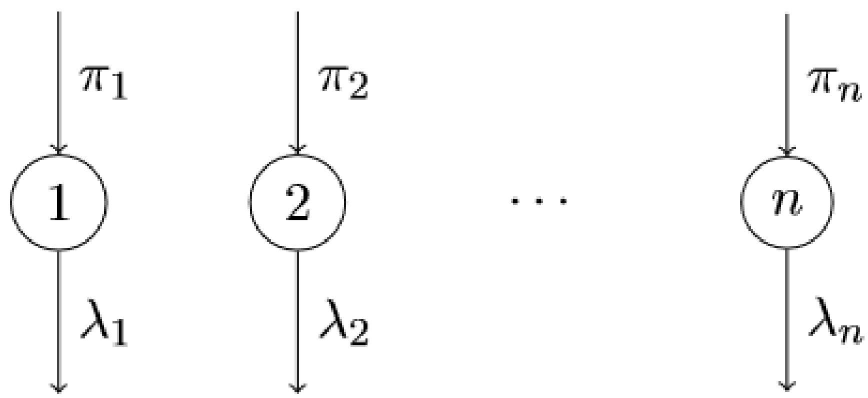

Figure 2.8: *Flowchart of a hyperexponential distribution.*

The examples above showcase the great number of distributions which belong to the class of phase-type distributions. There are nonetheless several variations and extensions of the phase-type distributions, some of which shall be reviewed here.

The discrete phase-type distribution is the discrete time variant of the phase-type distribution, which arises as the time until absorption in discrete time Markov processes (Markov chains). The theoretical foundation of the discrete phase-type distributions is very similar to that of the continuous phase-type distributions, and the classes even share some common results. Since the subject of Markov chains is frequently taught in undergraduate or graduate probability courses, much of the theory related to discrete phase-type distributions can be found in standard textbooks. More comprehensive treatments of the topic can be found in section 1.2.6 of Bladt and Nielsen (2017) or the doctoral dissertation of Navarro (2019).

Many of the extensions of the continuous phase-type distributions have been proposed to remedy the issue that phase-type distributions are light-tailed, see e.g. Bladt and Yslas (2022). Bladt et al. (2015) resolves the issue by constructing a class of infinite-dimensional phase-type distributions called NPH distributions. They define the NPH distributions as the product distributions arising from products of non-negative discrete random variables and phase-type distributed random variables. Specifically, the product of a random variable $Z \sim \text{PH}(\boldsymbol{\alpha}, \boldsymbol{T})$ and a discrete random variable N with distribution $\mathbb{P}(N = j) = q_j$ for all $j \in \mathbb{N}$ has a NPH distribution, which is written as $Y = ZN \sim \text{NPH}(\boldsymbol{q}, \boldsymbol{\alpha}, \boldsymbol{T})$, where $\boldsymbol{q} = \{q_j\}_{j \in \mathbb{N}}$. The random variable Y can then be considered as an infinite mixture of phase-type distributions with mixing probabilities $\boldsymbol{q}$, cf.

$$\mathbb{P}(Y = jZ) = \mathbb{P}(NZ = jZ) = \mathbb{P}(N = j) = q_j, \quad \forall j \in \mathbb{N}, \tag{2.125}$$

and for any natural number j, the random variable jZ is a phase-type distribution with initial distribution $\boldsymbol{\alpha}$ and sub-generator $\boldsymbol{T}/j$. The representation can be derived through the decomposition in equation (2.99) as scaling a phase-type distributed random variable is equivalent to scaling the sojourn times of all the individual visits of the underlying process. Thus, the random variable Y can be considered as the time until absorption of a Markov process on an infinite state-space with initial distribution $\boldsymbol{q} \otimes \boldsymbol{\alpha}$ and the block diagonal sub-generator $\boldsymbol{S} = \text{diag}(\boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{T}_3, ...)$, where $\boldsymbol{T}_j = \boldsymbol{T}/j$ for all $j \in \mathbb{N}$.

An alternative interpretation of a NPH distribution is as a phase-type distribution equipped with a random linear reward structure. This interpretation follows immediately from equation (2.99) as

$$Y = NZ = N \sum_{i \in E^*} Z_i = \sum_{i \in E^*} NZ_i, \tag{2.126}$$

which is exactly a transformation of the phase-type distributed random variable Z using the random reward structure $r(i) = N$ for all $i \in E^*$. Therefore, while the NPH distributions are obtained as product distributions and considered as phase-type distributions with infinite state-spaces, they can also represent an extension of the phase-type distributions through certain random reward structures.

The different variations of phase-type distributions which have been covered so far have all been derived as hitting time distributions based on time-homogeneous Markov processes. If the underlying Markov process is instead time-inhomogeneous, i.e. the transition rates among the transient states are time-dependent, the time until absorption is distributed according to an inhomogeneous phase-type distribution (abbreviated IPH distribution). The inhomogeneous phase-type distributions are more sophisticated mathematically than their homogeneous counterparts as their probability functions are derived using product integrals. The interested reader may consult Goodman and Johansen (1973) for an analysis of time-inhomogeneous Markov processes and Johansen (1986) for an introduction to product integrals.

In Albrecher and Bladt (2019), the authors take another approach to construct a class of heavy-tailed phase-type distributions by transforming the time scales of the states of the underlying Markov process. These transformations lead to time dependent transition rates among the states of the underlying Markov processes, and they introduce a series of distributions, which are all special instances of inhomogeneous phase-type distributions. The authors devote most of the paper to the matrix-Pareto distributions, but they also study matrix-Weibull distributions and several other distributions such as the exponential-PH distributions and the power-PH distributions. Albrecher and Bladt show that the matrix-Pareto distributions have heavy tails and that the distributions are dense in the class of distributions on the positive real line, cf. theorem 3.9 in the aforementioned paper. Furthermore, Albrecher and Bladt showcase the applicability of the distributions by applying the Expectation-Maximization algorithm to fit and model Dutch fire insurance claims with inhomogeneous phase-type distributions. In general, the IPH distributions find several applications within the actuarial sciences, see e.g. Kiersch (2020) for a thorough survey of insurance pricing using time-inhomogeneous Markov processes and IPH distributions.

The phase-type distributions are supported on the non-negative reals, however Ahn and Ramaswami introduced a bilateral phase-type distribution in 2005. Ahn and Ramaswami extend the phase-type distribution to the entire real axis by equipping an absorbing Markov process with a linear reward structure that includes negative reward rates. Ahn and Ramaswami further shows that the class of bilateral phase-type distributions (the BPH class) is dense (in the sense of weak convergence) in the class of all univariate distributions, cf. theorem 2.3.2 in their paper. In contrast to case of IPH distributions, the BPH distributions actually inherit several properties from the phase-type distributions including the closure properties under finite convolutions and mixtures. The BPH distributions further have a property that allows for any BPH distribution to be decomposed into a three component mixture consisting of a phase-type distributions on the positive reals, a phase-type distribution on the negative reals, and a possible atom at zero, cf. sections 4 and 5 in Anh and Ramaswami (2005).

In the paragraph on infinite dimensional phase-type distributions, we established that NPH distributions could be obtained as transformations of phase-type distributions using random linear reward structures, and as described in the above paragraph, BPH distributions are defined through linear reward structures with negative reward rates. In summary, transformations using (possibly random) linear reward structures can generate all NPH and BPH distributions. Taking the versatility of the transformations into account, it is not surprising that reward structures can be used to construct a large class of multivariate distributions. These distributions are studied in the next chapter.

The last class of distributions which shall be covered in this section is the class of matrix-exponential (abbrev. ME) distributions. The matrix-exponential distributions are the distributions with rational Laplace-Stieltjes transforms and constitute a superset of the phase-type distributions. Theorem 2.12 states exactly which conditions that matrix-exponential distributions must satisfy in order to be of phase-type. An alternative definition of the matrix-exponential distributions is as all the distributions with distribution functions on the form (2.23), where the entering quantities need not admit a probabilistic interpretation, e.g. the matrix $\boldsymbol{Q}$ needs not be a sub-generator matrix. This lack of probabilistic interpretations often complicates analyses of matrix-exponential distributions, whose properties are usually derived using complex analysis and functional calculus. Nonetheless, matrix-exponential and phase-type distributions share numerous properties. The closure properties given in theorems 2.7 and 2.8 extend to the matrix-exponential distributions, and naturally the matrix-exponential distributions inherit the denseness property directly from the phase-type distributions.

Fackrell (2003) and Bean et al. (2008) address the problem of determining whether a rational expression constitutes a Laplace-Stieltjes transform, and there are several papers on the minimal representation of ME distributions based on their rational transforms, see e.g. Asmussen and Bladt (1996). Chapter 4 of Bladt and Nielsen (2017) serves as an excellent introduction to the topic of matrix-exponential distributions and is one of the most thorough resources currently available. Alternative, Asmussen and O'Cinneide (1998) can be consulted for a short and concise introduction to the topic.

We conclude this chapter by directing the reader to the *Additional reading* section of the bibliography, which contains further references on other related distributions.

CHAPTER 3

MPH* distributions

This chapter is concerned with the MPH* distributions, which are multivariate phase-type distributions defined through linear reward structures. The first section contains a survey of three different types of multivariate phase-type distributions, while the following section serves as a brief introduction to the MPH* distributions. The third sections then covers various properties of the MPH* distributions. The fourth section is concerned with some examples of MPH* distributions obtained through different reward structures and is succeeded by a section on related multivariate distributions.

3.1 Types of multivariate phase-type distributions

The seminal 1975 paper by Neuts inspired research into different methods of constructing multivariate distributions generalizing the univariate phase-type distributions. Most, if not all, multivariate phase-type distributions are derived from or based on one of two constructions.

The first of these constructions is from Assaf et al. (1984). Assaf et al. consider a Markov process and divide the associated state-space into a number of possibly intersecting non-empty sets. They then construct a random vector consisting of the hitting times of the different subsets and define the corresponding distribution as a multivariate phase-type distribution (MPH distribution). They further show that the MPH distributions are dense in the class of multivariate distributions on the closed positive half-spaces. Their paper also includes theorems showing that the class of MPH distributions is closed under finite mixtures and that all multivariate marginal distributions of MPH distributions are themselves MPH distributions. Assaf and the coauthors actually also derive other classes of multivariate phase-type distributions based on the MPH distributions, which have not received much attention in the literature. In the last part of their paper, they demonstrate that the multivariate exponential distribution of Marshall and Olkin (1967), the bivariate extension of the bivariate exponential distribution in Freund (1961), and the bivariate extension of the gamma distribution due to Becker and Roux (1981) are all special instances of MPH distributions.

Despite the useful properties and applicability of the MPH distributions, current research mainly deals with the multivariate distributions introduced in Kulkarni (1989). The concept of reward structures was introduced in section 2.4 of this thesis, and Kulkarni applies such reward structures to define the MPH* class of multivariate phase-type distributions. Specifically, Kulkarni considers an absorbing Markov process and a number of linear reward structures with non-negative reward rates. Each reward structure corresponds to a certain transformation of the underlying Markov process and leads to a certain accumulation of reward. Kulkarni then constructs a random vector whose elements are exactly the total accumulated rewards prior to absorption of the underlying process using the respective reward structures, for example the first vector component is the total accumulated reward according to the first reward structure and the second vector component is the total accumulated reward according to the second reward structure. The precise definition and construction are presented in the next section.

The paper by Kulkarni contains a number of interesting results. The first important result shows that the MPH class by Assaf et al. is a strict subset of the MPH* class. The next result concerns non-negative linear combinations of the vector components of a MPH* distribution, and it leads to a corollary, which asserts that MPH* distributions are closed under finite convolutions. In conjunction, the two results affirmatively settle the conjecture by Assaf et al. that the MPH class is closed under finite convolutions. In addition to these results, Kulkarni also shows that the MPH* class is closed under finite mixtures and is dense in the class of multivariate distributions on the closed positive half-spaces. The paper features some additional results along with a question on whether the minimum formed over the components of a MPH* distribution remains a phase-type distribution.

A possibly even more general variant of a multivariate phase-type distribution was proposed in Bladt and Nielsen (2010). The definition that Bladt and Nielsen propose is based on geometric projections and classifies a multivariate distribution as a MVPH distribution if every conical combination of the components belongs to the class of phase-type distributions. Since the MPH* distributions clearly satisfy these conditions, the class by Bladt and Nielsen includes all distributions in the class of Kulkarni, but it is an open question, whether the MPH* distributions constitute a proper subset of the MVPH distributions. The MVPH distributions are so general in nature that only little is known about them. A few properties are however readily derived, e.g. all univariate marginal distributions are phase-type distributions and all multivariate marginal distributions are themselves MVPH distributions. The question above therefore essentially reduces to a question of whether there exists a dependence structure between phase-type distributions, which cannot be modelled with linear reward structures, such that their convolution is still a phase-type distribution.

The characterization proposed by Bladt and Nielsen is inspired by the characterization of the multivariate Gaussian distribution, and the distributions might be more closely connected than one would expect. Theorem 8.5.16 of Bladt and Nielsen (2017) states that the MVBME and MBME* classes of multivariate bilateral matrix-exponential distributions are different, and the proof is based on a concrete counterexample. They show that a bivariate random vector consisting of two independent Brownian motions observed at an exponentially distributed time is a MVBME distribution, which cannot be expressed as a MBME* distribution. In other words, they show that this Gaussian scale mixture (a Normal-Exponential distribution) is only a multivariate matrix-exponential distribution in the sense of Bladt and Nielsen.

This concludes our brief survey of some of the most common multivariate phase-type constructions. The remainder of the chapter will focus on the distributions by Kulkarni and their properties. As the MVPH distributions introduced by Bladt and Nielsen still require further research and explorative studies to be properly understood, this exposition will only study the MPH* distributions. The multivariate Gaussian distribution will be revisited shortly in the next chapter in relation to the gamma distribution by Dussauchoy and Berland.

3.2 Construction

In this section, we adopt the setup and notation from the former chapter to explain the construction of the MPH* distributions. We consider a random variable $\tau \sim \mathrm{PH}_k(\boldsymbol{\alpha}, \boldsymbol{S})$ with the governing Markov jump process $\{X_t; t \geq 0\}$, which is defined on the state-space $E = \{1, ..., k, k+1\}$. Furthermore, we define the random variable Z_{ij} as the sojourn time of the j'th visit of the underlying process to state $i \in E^*$, and the random variable N_i as the number of visits of the underlying process to state i prior to absorption. The total sojourn time of the governing process in state i prior to absorption will be denoted by Z_i and is calculated as

$$Z_i = \sum_{j=1}^{N_i} Z_{ij} = \int_0^\tau \mathbb{1}\{X_t = i\} dt, \quad \forall i \in E^*. \tag{3.1}$$

The construction also includes a multivariate transformation $r : E^* \times \{1, ..., n\} \to \mathbb{R}_0^+$, where n denotes the dimension of constructed distribution. This transformation defines a reward structure in the multivariate setting, which can be represented through a matrix with n columns and $\dim(E^*) = k$ rows. Therefore, a reward structure is defined as a matrix $\boldsymbol{R} = \{R_{ij}\}$ with elements $R_{ij} = r(i, j)$ for $i \in E^*$ and $j \in \{1, ..., n\}$. This definition also implies that every column of the matrix $\boldsymbol{R}$ defines a univariate reward structure, i.e. for each $j \in \{1, ..., n\}$, we can define a univariate reward structure $r_j : E^* \to \mathbb{R}_0^+$ given by $r_j(i) = r(i, j) = R_{ij}$. By recasting the univariate reward structures as vectors, see the paragraph above theorem 2.14 in section 2.4, the reward matrix $\boldsymbol{R}$ can be expressed as $\boldsymbol{R} = (\boldsymbol{r}_1, ..., \boldsymbol{r}_n)$.

After establishing the preliminary setup, the multivariate phase-type distribution introduced in Kulkarni (1989) can be defined formally.

Definition 3.1
Let $\tau \sim PH_k(\boldsymbol{\alpha}, \boldsymbol{S})$ be governed by the underlying process $\{X_t; t \geq 0\}$ on the state-space E as described above, and let $\boldsymbol{R} = \{R_{ij}\}$ be a matrix of dimension $p \times n$ with non-negative elements. Then $\boldsymbol{R}$ constitutes a reward structure and each column of $\boldsymbol{R}$ can be considered as a function $r_j : E^+ \to \mathbb{R}_0^+$ defined by $r_j(i) = R_{ij}$. The n-dimensional random vector $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$ with elements

$$Y_j = \int_0^\tau r_j(X_t)dt = \sum_{i=1}^{k} R_{ij} Z_i, \quad j \in \{1, ..., n\}, \tag{3.2}$$

then has a MPH$_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ distribution.*

The definition shows that the components of a random vector having a multivariate phase-type distribution in the sense of Kulkarni are obtained by applying different univariate linear reward structures to a common underlying Markov process. For this reason, textbooks and papers often refer to the MPH* distributions as the multivariate phase-type distributions arising from linear reward structures.

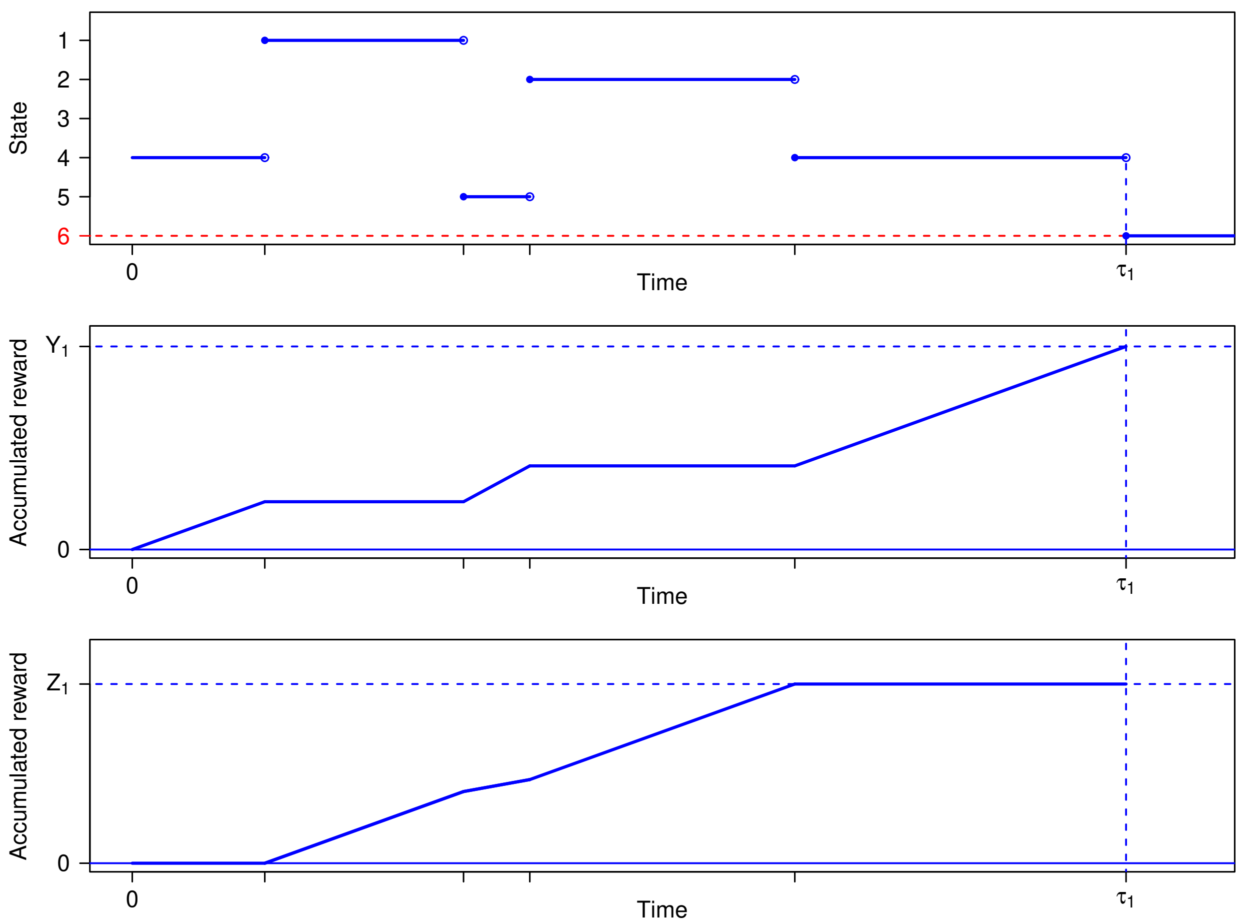

Figure 3.1: *A realization of a random vector (Y, Z) following a MPH* distribution.*

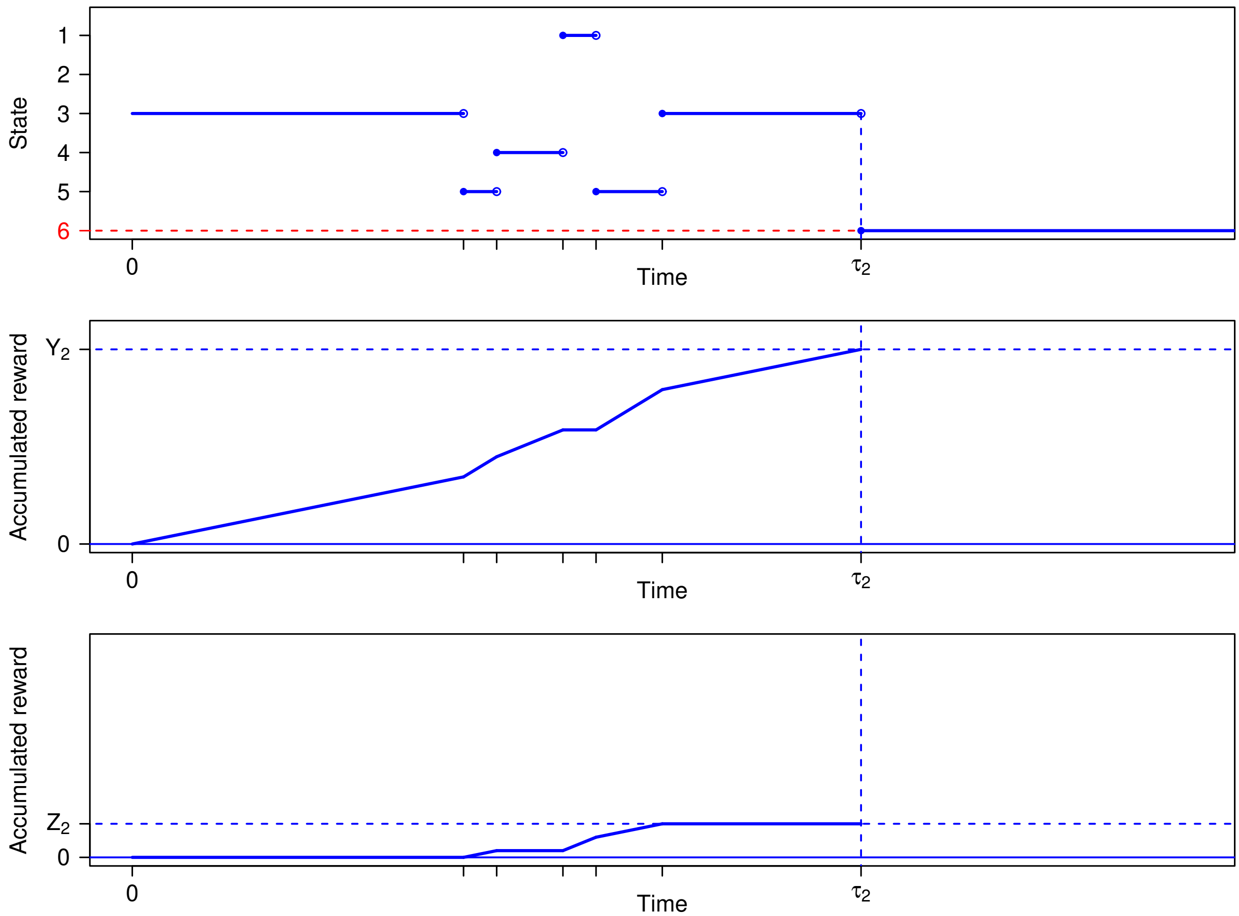

Figure 3.2: *Another realization of a random vector* (Y, Z) *following a MPH* distribution.*

Figures 3.1 and 3.2 depict two realizations of a random vector (Y, Z) having a MPH* distribution. Specifically, the figures show two realizations of the underlying Markov process and how the vector components accumulate reward until the Markov process is absorbed. The reward structure dictates the rates of reward accumulation based on the state of the Markov process, which translates into the varying slopes in the graphics. The slopes indicate that only the Z-component receives a positive reward when the underlying process sojourns in states 1 and 2, while only the Y-component receives positive reward when the process is in either state 3 or 4. Finally, it can deduced that both vector components receive positive reward when the Markov process sojourns in state 5.

The figures suggest how the reward structure in conjunction with the underlying process induces a certain dependence structure between the vector components. The two realizations presented in the figures show how different paths of the underlying process can lead to similar or different values for the vector components. Consequently, the association between the vector components depends on the likelihood of the possible paths of the underlying process combined with the reward structure.

Like in the univariate setting, there are special reward structures that reproduce certain important distributions. An example is when the reward structure is the identity matrix, which results in the joint distribution of the total sojourn times of the underlying process in the transient states prior to absorption. In section 3.4, it will be further explored how multivariate exponential and gamma distributions can be modelled with linear reward structures, which underlines the scope of the MPH* distributions.

3.3 Properties

The construction of the MPH* distributions using reward structures entails that many properties of the univariate distributions carry over to their multivariate counterparts. This includes the closure properties with respect to mixing and convolution as well as the denseness property. There are however also some probability functions which cannot in general be expressed in the multivariate setting and some analysis which is considerably more involved compared to the univariate case. Furthermore, some concepts such as concomitants of order statistics and order statistics of vector components are only defined in the multivariate context. The concomitants are treated in the forthcoming paper A, which shows that the concomitants are themselves distributed according to a phase-type distribution, while O'Cinneide (1990) proves that order statistics of dependent phase-type distributions are not necessarily phase-type distributed, cf. section 4 of the O'Cinneide paper.

Since the joint probability density function and the joint distribution function are generally not available for MPH* distributions, analysis regarding the distributions usually involves transformations of the distributions. The moment-generating function and the Laplace transform are the most commonly applied transforms, but some authors invoke the characteristic functions. The starting point of this exposition on the multivariate phase-type distributions is therefore a derivation of the Laplace transform of a random vector following a MPH* distribution. In the derivation, the notation $\langle\cdot,\cdot\rangle : \mathbb{C}^k \times \mathbb{C}^k \to \mathbb{C}$ shall denote the standard Hermitian inner product, which is equivalent to the Euclidean inner product when the arguments are real-valued vectors.

Theorem 3.1
The Laplace(-Stieltjes) transform $\mathcal{L} : \mathbb{C}^n \to \mathbb{C}$ of an n-dimensional random vector $\boldsymbol{Y} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[e^{-\langle \boldsymbol{Y},\boldsymbol{\theta}\rangle}\right] = \alpha_{k+1} + \boldsymbol{\alpha}\left(\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta}) - \boldsymbol{S}\right)^{-1}\boldsymbol{s} = \alpha_{k+1} + \boldsymbol{\alpha}\left(\boldsymbol{I}_k + \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1}\boldsymbol{1}_k^{\top}, \tag{3.3}$$

and there exists a negative constant θ_0 such that the transform is well-defined for at least all arguments whose elements all have real parts greater than θ_0.

Proof. Define the conditional expectations

$$\mathcal{L}_i(\boldsymbol{\theta}) = \mathbb{E}\left[e^{-\langle \boldsymbol{Y},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right], \quad \forall i \in \{1, ..., k\}, \tag{3.4}$$

and the column vector

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \left(\mathcal{L}_1(\boldsymbol{\theta}), ..., \mathcal{L}_k(\boldsymbol{\theta})\right)^{\top} \tag{3.5}$$

such that

$$\mathcal{L}(\boldsymbol{\theta}) = \alpha_{k+1} + \boldsymbol{\alpha}\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) \tag{3.6}$$

by the law of total expectation. The conditional expectations in equation (3.4) can be calculated through a decomposition. Given the governing process is initiated in a transient state i, the first occupation time in said state Z_{i1} follows an exponential distribution, which allows for rewriting the time until absorption of the underlying process as $\tau = Z_{i1} + \tau_r$, where τ_r is the remainder of the time until absorption after the first visit of the process to state i. The random vector $\boldsymbol{Y}$ can similarly be decomposed into $\boldsymbol{Y} = \boldsymbol{Y}_{i1} + \boldsymbol{Y}_r$, where the terms represent the accumulated reward of the underlying process during the first visit to state i and throughout the remainder of the time until absorption, respectively. Since the two random variables Z_{i1} and τ_r are conditionally independent given the governing process is initiated in state i, the same property applies to the two random vectors $\boldsymbol{Y}_{i1}$ and $\boldsymbol{Y}_r$. Hence,

$$\mathbb{E}\left[e^{-\langle \boldsymbol{Y},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] = \mathbb{E}\left[e^{-\langle \boldsymbol{Y}_{i1},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right]\mathbb{E}\left[e^{-\langle \boldsymbol{Y}_r,\boldsymbol{\theta}\rangle}\middle| X_0 = i\right]. \tag{3.7}$$

The first factor on the RHS of equation (3.7) is derived using the fact that $\boldsymbol{Y}_{i1} = Z_{i1}\boldsymbol{R}_{i.}^{\top}$, where $\boldsymbol{R}_{i.}$ refers to row i of the reward matrix, which leads to

$$\mathbb{E}\left[e^{-\langle \boldsymbol{Y}_{i1},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] = \mathbb{E}\left[e^{-\langle Z_{i1}\boldsymbol{R}_{i.}^{\top},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right]. \tag{3.8}$$

Since the standard Hermitian inner product is bilinear over the field of real numbers, the random variable Z_{i1} may be factored out of the inner product on the RHS of equation (3.8). As Z_{i1} has an exponential distribution with rate $-(\boldsymbol{S})_{ii} = -S_{ii}$, the expression simplifies to the Laplace transform of an exponential distribution:

$$\mathbb{E}\left[e^{-\langle Z_{i1}\boldsymbol{R}_{i.}^{\top},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] = \mathbb{E}\left[e^{-Z_{i1}\langle \boldsymbol{R}_{i.}^{\top},\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] = \frac{1}{1-\langle \boldsymbol{R}_{i.}^{\top},\boldsymbol{\theta}\rangle S_{ii}^{-1}} = \frac{1}{1-\boldsymbol{R}_{i.}\boldsymbol{\theta}S_{ii}^{-1}}, \tag{3.9}$$

which is well-defined for arguments satisfying $\mathrm{Re}(\boldsymbol{R}_{i.}\boldsymbol{\theta}) > S_{ii}$. The second factor on the RHS of equation (3.7) is evaluated using the law of total expectation and the (pseudo) embedded Markov chain $\{\widetilde{X}_m : m \in \mathbb{N}_0\}$ as

$$\mathbb{E}\left[e^{-\langle \boldsymbol{Y}_r,\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] = \mathbb{E}\left[e^{-\langle \boldsymbol{Y}_r,\boldsymbol{\theta}\rangle}\middle| \widetilde{X}_0 = i\right] = \sum_{\substack{j\in E\\ j\neq i}}\mathbb{E}\left[e^{-\langle \boldsymbol{Y}_r,\boldsymbol{\theta}\rangle}\middle| \widetilde{X}_1 = j, \widetilde{X}_0 = i\right]\mathbb{P}\left(\widetilde{X}_1 = j\middle|\widetilde{X}_0 = i\right). \tag{3.10}$$

The transition probabilities of the embedded Markov process are given in sections 2.1 and 2.4, while the conditional expectations can be simplified. Since $\boldsymbol{Y}_r$ constitutes the reward accumulated after leaving the initial state, knowing the state $\widetilde{X}_1$ makes knowledge about the initial state superfluous. Moreover, the conditional distribution of $\boldsymbol{Y}_r$ given $\widetilde{X}_1$ must be identical to the conditional distribution of $\boldsymbol{Y}$ given $\widetilde{X}_0$, which implies that

$$\begin{aligned}\mathbb{E}\left[e^{-\langle \boldsymbol{Y}_r,\boldsymbol{\theta}\rangle}\middle| X_0 = i\right] &= \sum_{\substack{j\in E\\ j\neq i}}\mathbb{E}\left[e^{-\langle \boldsymbol{Y},\boldsymbol{\theta}\rangle}\middle| \widetilde{X}_0 = j\right]\mathbb{P}\left(\widetilde{X}_1 = j\middle|\widetilde{X}_0 = i\right) && (3.11)\\ &= -\frac{s_i}{S_{ii}} - \sum_{\substack{j\in E^*\\ j\neq i}}\mathcal{L}_j(\boldsymbol{\theta})\frac{S_{ij}}{S_{ii}}. && (3.12)\end{aligned}$$

These results allow for recasting equation (3.7) as

$$\mathcal{L}_i(\boldsymbol{\theta}) = \frac{1}{1 - \boldsymbol{R}_{i.}\boldsymbol{\theta} S_{ii}^{-1}} \left(-\frac{s_i}{S_{ii}} - \sum_{\substack{j \in E^* \\ j \neq i}} \mathcal{L}_j(\boldsymbol{\theta}) \frac{S_{ij}}{S_{ii}} \right), \tag{3.13}$$

and it follows directly that

$$\left(1 - \boldsymbol{R}_{i.}\boldsymbol{\theta} S_{ii}^{-1}\right) \mathcal{L}_i(\boldsymbol{\theta}) = -\frac{s_i}{S_{ii}} - \sum_{\substack{j \in E^* \\ j \neq i}} \mathcal{L}_j(\boldsymbol{\theta}) \frac{S_{ij}}{S_{ii}}. \tag{3.14}$$

Simple algebraic manipulations then yield

$$\boldsymbol{R}_{i.}\boldsymbol{\theta}\mathcal{L}_i(\boldsymbol{\theta}) = s_i + \sum_{j \in E^*} \mathcal{L}_j(\boldsymbol{\theta}) S_{ij}. \tag{3.15}$$

The system of equations arising from considering equation (3.15) for all the transient states can be expressed as the matrix equation

$$\boldsymbol{\Delta}(\boldsymbol{R\theta})\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \boldsymbol{s} + \boldsymbol{S}\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) \tag{3.16}$$

or equivalently as

$$\left(\boldsymbol{\Delta}(\boldsymbol{R\theta}) - \boldsymbol{S}\right) \boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \boldsymbol{s}. \tag{3.17}$$

Since the sub-generator matrix is invertible, the Gershgorin circle theorem implies that there exists a neighborhood around the origin, where the matrix $\boldsymbol{\Delta}(\boldsymbol{R\theta}) - \boldsymbol{S}$ is also non-singular. In said neighborhood, the solution to the matrix equation is well-defined and given by

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}) = \left(\boldsymbol{\Delta}(\boldsymbol{R\theta}) - \boldsymbol{S}\right)^{-1} \boldsymbol{s}. \tag{3.18}$$

Substituting this result into equation (3.6) then gives the Laplace(-Stieltjes) transform

$$\mathcal{L}(\boldsymbol{\theta}) = \alpha_{k+1} + \boldsymbol{\alpha} \left(\boldsymbol{\Delta}(\boldsymbol{R\theta}) - \boldsymbol{S}\right)^{-1} \boldsymbol{s}. \tag{3.19}$$

The transform is well-defined whenever the conditions induced in equations (3.9) and (3.18) are satisfied, and due to the nature of the conditions, there must exist a sufficiently small neighborhood around the origin, where all the conditions are always met. This implies the existence of a negative constant θ_0 such that the transform exists for all arguments whose elements all have real parts greater than θ_0. □

The Laplace transform and the related transforms allow for deriving several properties and calculating various quantities associated with the multivariate distributions, for example the cross-moments. The cross-moments can be used to calculate covariances between dependent random variables and derive higher order moment distributions. These features underline the importance of the cross-moments, and the next theorem states the formula for the cross-moments of MPH* distributions.

Theorem 3.2

Consider an n-dimensional random vector $\boldsymbol{Y} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and let $h_1, ..., h_n$ be non-negative integers, whose sum h is at least one, i.e.

$$h = \sum_{j=1}^{n} h_j \geq 1. \tag{3.20}$$

The associated cross-moment of order h is then given by

$$\mathbb{E}\left[\prod_{j=1}^{n} Y_j^{h_j}\right] = \boldsymbol{\alpha} \sum_{\ell=1}^{h!} \left(\prod_{p=1}^{h} \boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\right) \mathbf{1}_k^\top, \tag{3.21}$$

where $\sigma_1, ..., \sigma_{h!}$ are the ordered permutations of h-tuples of derivatives, and $\sigma_\ell(p)$ is the p'th derivative in σ_ℓ.

Proof. Since $h \geq 1$ by assumption, the cross-moments can be found as mixed partial derivatives of the moment-generating function $M : \mathbb{R}^n \to \mathbb{R}_0^+$, which is found as the restriction of the Laplace transform (with a sign change in the argument) to real-valued arguments. The moment-generating function is therefore well-defined for all arguments whose elements are all less than $-\theta_0$ (cf. theorem 3.1). Specifically,

$$\mathbb{E}\left[\prod_{j=1}^{n} Y_j^{h_j}\right] = \frac{\partial^h M}{\partial\theta_1^{h_1} \cdots \partial\theta_n^{h_n}} \left(\mathbf{0}_n^\top\right). \tag{3.22}$$

Using the latter expression for the Laplace transform in equation (3.3), the moment-generating function takes the form

$$M(\boldsymbol{\theta}) = \alpha_{k+1} + \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \mathbf{1}_k^\top. \tag{3.23}$$

Now a proof by induction is invoked to show that the mixed partial derivative in equation (3.22) evaluates to the expression in equation (3.21). The proof is actually concerned with the more general result that

$$\frac{\partial^h M}{\partial\theta_1^{h_1} \cdots \partial\theta_n^{h_n}}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \sum_{\ell=1}^{h!} \left(\prod_{p=1}^{h} \boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1}\right) \mathbf{1}_k^\top. \tag{3.24}$$

The base cases are the cross-moments of order 1, which correspond to the expectations of the different vector components. Consider therefore the partial derivative of the moment-generating function with respect to an arbitrary variable θ_i with $i \in \{1, ..., n\}$

$$\frac{\partial M}{\partial\theta_i}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\left[\alpha_{k+1} + \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \mathbf{1}_k^\top\right] = \boldsymbol{\alpha}\frac{\partial}{\partial\theta_i}\left[\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1}\right] \mathbf{1}_k^\top. \tag{3.25}$$

The partial derivative of the inverse matrix exists whenever the inverse matrix itself exists. Consequently, the derivative can be established using the Neumann series expansion of the matrix $\boldsymbol{U\Delta}(\boldsymbol{R\theta})$, which converges exactly when said inverse matrix is non-singular. This result also appears in Selby (1974), see equation (2.41), and yields that

$$\frac{\partial}{\partial\theta_i}\left[\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1}\right] = -\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \frac{\partial}{\partial\theta_i}\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \tag{3.26}$$

$$= \left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \boldsymbol{U\Delta}(\boldsymbol{R}_{.i})\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1}. \tag{3.27}$$

Inserting this partial derivative into equation (3.25) then produces the expression

$$\frac{\partial M}{\partial\theta_i}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \boldsymbol{U\Delta}(\boldsymbol{R}_{.i})\left(\boldsymbol{I}_k - \boldsymbol{U\Delta}(\boldsymbol{R\theta})\right)^{-1} \mathbf{1}_k^\top, \tag{3.28}$$

which coincides with the formula in equation (3.24) for $h = 1$. The result thus verifies the formula for the base cases. The moments are obtained by evaluating the expression in $\boldsymbol{\theta} = \mathbf{0}_n^\top$, which gives

$$\frac{\partial M}{\partial \theta_i}\left(\mathbf{0}_n^\top\right) = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i})\mathbf{1}_k^\top. \tag{3.29}$$

The expression on the RHS of equation (3.29) does not agree completely with the result of corollary 2.1 in equation (2.51), but the discrepancy can easily be explained. The result from the corollary gives the expected time until absorption of the underlying process, whereas the result in (3.29) gives the expected accumulated reward of the underlying process prior to absorption. This property follows immediately from theorem 2.2 and equation (3.2) as

$$\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[\sum_{j=1}^{k} R_{ji}Z_j\right] = \sum_{m=1}^{k+1}\mathbb{E}\left[\sum_{j=1}^{k} R_{ji}Z_j \,\middle|\, X_0 = m\right]\mathbb{P}(X_0 = m) \tag{3.30}$$

$$= \sum_{m=1}^{k+1}\sum_{j=1}^{k} R_{ji}\mathbb{E}\left[Z_j | X_0 = m\right]\mathbb{P}(X_0 = m), \tag{3.31}$$

which simplifies in accordance with equation (2.22) to

$$\mathbb{E}\left[Y_i\right] = \sum_{m=1}^{k}\sum_{j=1}^{k} R_{ji}u_{mj}\alpha_m = \sum_{j=1}^{k}\left(\sum_{m=1}^{k}\alpha_m u_{mj}\right)R_{ji} = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i})\mathbf{1}_k^\top. \tag{3.32}$$

This completes the first step of the proof by induction. In the next part, the induction step, it is assumed that the formula in equation (3.24) holds for a particular mixed partial derivative of order $h > 1$ such that

$$\frac{\partial^h M}{\partial\theta_1^{h_1}\cdots\partial\theta_n^{h_n}}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1}\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{\cdot\sigma_\ell(p)}\right)\left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1}\right)\mathbf{1}_k^\top. \tag{3.33}$$

Consider then the mixed partial derivative of order $h+1$ arising from taking a partial derivative of the function in equation (3.33) with respect to an arbitrary variable θ_i with $i \in \{1, ..., n\}$, say

$$M_i^{(h+1)}(\boldsymbol{\theta}) = \frac{\partial^{h+1}M}{\partial\theta_1^{h_1}\cdots\partial\theta_i^{h_i+1}\cdots\partial\theta_n^{h_n}}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\left(\frac{\partial^h M}{\partial\theta_1^{h_1}\cdots\partial\theta_n^{h_n}}\right)(\boldsymbol{\theta}). \tag{3.34}$$

It follows directly from the induction hypothesis that

$$M_i^{(h+1)}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\left(\boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1}\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{\cdot\sigma_\ell(p)}\right)\left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1}\right)\mathbf{1}_k^\top\right). \tag{3.35}$$

We introduce the shorthand notation $\boldsymbol{K}(\boldsymbol{\theta}) = \boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})$ to simplify the expressions and note that $\partial\boldsymbol{K}^{-1}(\boldsymbol{\theta})/\partial\theta_i = \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{\cdot i})\boldsymbol{K}^{-1}(\boldsymbol{\theta})$. Applying the product rule of differentiation on the RHS of equation (3.35) then leads to

$$\begin{aligned} M_i^{(h+1)}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\Bigg[&\frac{\boldsymbol{K}^{-1}(\boldsymbol{\theta})}{\partial\theta_i}\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{\cdot\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right) \\ &+ \boldsymbol{K}^{-1}(\boldsymbol{\theta})\frac{\partial}{\partial\theta_i}\left[\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{\cdot\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right)\right]\Bigg]\mathbf{1}_k^\top, \end{aligned} \tag{3.36}$$

which simplifies to

$$M_i^{(h+1)}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\boldsymbol{K}^{-1}(\boldsymbol{\theta})\left[\boldsymbol{U\Delta}(\boldsymbol{R}_{.i})\boldsymbol{K}^{-1}(\boldsymbol{\theta})\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right) + \frac{\partial}{\partial\theta_i}\left[\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right)\right]\right]\mathbf{1}_k^\top. \tag{3.37}$$

The outermost parenthesis in equation (3.37) contains two terms:

$$\boldsymbol{U\Delta}(\boldsymbol{R}_{.i})\boldsymbol{K}^{-1}(\boldsymbol{\theta})\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right) \tag{3.38}$$

and

$$\frac{\partial}{\partial\theta_i}\left[\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right)\right]. \tag{3.39}$$

The first term, which is shown in equation (3.38), has $h!$ terms of products with $h+1$ factors on the form $\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})$. The second term, which is shown in equation (3.39), can be rewritten using the linearity of the differential operator in combination with the product rule;

$$\frac{\partial}{\partial\theta_i}\left[\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right)\right] = \sum_{\ell=1}^{h!}\frac{\partial}{\partial\theta_i}\left(\prod_{p=1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right) \tag{3.40}$$

$$= \sum_{\ell=1}^{h!}\sum_{q=1}^{h}\left(\left[\prod_{p=1}^{q-1}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right]\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(q)}\right)\frac{\partial\boldsymbol{K}^{-1}}{\partial\theta_i}(\boldsymbol{\theta})\left[\prod_{p=q+1}^{h}\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})\right]\right). \tag{3.41}$$

The partial derivatives in equation (3.41) are again given by $\partial\boldsymbol{K}^{-1}(\boldsymbol{\theta})/\partial\theta_i = \boldsymbol{K}^{-1}(\boldsymbol{\theta})\boldsymbol{U\Delta}(\boldsymbol{R}_{.i})\boldsymbol{K}^{-1}(\boldsymbol{\theta})$, which means that the expression in equation (3.41) comprises $h!h$ terms, and each term is a product of $h+1$ factors on the form $\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})$. The outermost parenthesis in equation (3.37) therefore spans $h! + h!h = h!(1+h) = (h+1)!$ terms which are products of $h+1$ factors on the form $\boldsymbol{U\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\boldsymbol{K}^{-1}(\boldsymbol{\theta})$. Consider then the indices (the columns) of the reward matrix appearing in the terms (each a product) of the outermost parenthesis of equation (3.37). The terms in equation (3.38) share a common structure; the first index is i followed by h indices ordered according to the respective permutations σ_ℓ. Listing the indices in order of appearance such that each row corresponds to a term/product gives:

$$h!\left\{\begin{array}{ccccccc} i & \sigma_1(1) & \sigma_1(2) & \sigma_1(3) & \cdots & \sigma_1(h-1) & \sigma_1(h) \\ i & \sigma_2(1) & \sigma_2(2) & \sigma_2(3) & \cdots & \sigma_2(h-1) & \sigma_2(h) \\ & & & & \vdots & & \\ i & \sigma_{h!-1}(1) & \sigma_{h!-1}(2) & \sigma_{h!-1}(3) & \cdots & \sigma_{h!-1}(h-1) & \sigma_{h!-1}(h) \\ i & \sigma_{h!}(1) & \sigma_{h!}(2) & \sigma_{h!}(3) & \cdots & \sigma_{h!}(h-1) & \sigma_{h!}(h) \end{array}\right. . \tag{3.42}$$

The same can be done for the terms in equation (3.41). For each value of ℓ (for each permutation), the summation over q has h terms, whose indices are

$$h\left\{\begin{array}{cccccc} \sigma_\ell(1) & i & \sigma_\ell(2) & \cdots & \sigma_\ell(h-1) & \sigma_\ell(h) \\ \sigma_\ell(1) & \sigma_\ell(2) & i & \cdots & \sigma_\ell(h-1) & \sigma_\ell(h) \\ & & & \ddots & & \\ \sigma_\ell(1) & \sigma_\ell(2) & \sigma_\ell(3) & \cdots & i & \sigma_\ell(h) \\ \sigma_\ell(1) & \sigma_\ell(2) & \sigma_\ell(3) & \cdots & \sigma_\ell(h) & i \end{array}\right. . \tag{3.43}$$

Collecting all the terms represented in equations (3.42) and (3.43), one can observe that the index i has been interjected into every possible position of every permutation. Consequently, the outermost parenthesis in the expression of (3.37) comprises the sum over all $(h+1)!$ permutations of $(h+1)$-tuples of derivatives of the moment-generating function that leads to $M_i^{(h+1)}$, i.e. the $(h+1)!$ different ways to obtain said mixed partial derivative. In conclusion, equation (3.37) simplifies to

$$M_i^{(h+1)}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\boldsymbol{K}^{-1}(\boldsymbol{\theta}) \sum_{\ell=1}^{(h+1)!} \left(\prod_{p=1}^{h+1} \boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right) \boldsymbol{K}^{-1}(\boldsymbol{\theta}) \right) \boldsymbol{1}_k^\top, \tag{3.44}$$

and the back-substitution of the matrix $\boldsymbol{K}(\boldsymbol{\theta})$ then gives

$$M_i^{(h+1)}(\boldsymbol{\theta}) = \boldsymbol{\alpha}\left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1} \sum_{\ell=1}^{(h+1)!} \left(\prod_{p=1}^{h+1} \boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right) \left(\boldsymbol{I}_k - \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{\theta})\right)^{-1} \right) \boldsymbol{1}_k^\top. \tag{3.45}$$

As h and i were chosen arbitrarily in the induction step and the final expression in (3.45) coincides with the induction hypothesis for the order $h+1$, the principle of induction implies that the formula in equation (3.33) holds for all $h \geq 1$ and any underlying collection $\{h_j\}_{1\leq j\leq n}$. The cross-moments are finally established by evaluating the formula (3.33) in $\boldsymbol{\theta} = \boldsymbol{0}_n^\top$:

$$\mathbb{E}\left[\prod_{j=1}^{n} Y_j^{h_j}\right] = \frac{\partial^h M}{\partial\theta_1^{h_1}\cdots\partial\theta_n^{h_n}}\left(\boldsymbol{0}_n^\top\right) = \boldsymbol{\alpha}\sum_{\ell=1}^{h!}\left(\prod_{p=1}^{h}\boldsymbol{U}\boldsymbol{\Delta}\left(\boldsymbol{R}_{.\sigma_\ell(p)}\right)\right)\boldsymbol{1}_k^\top, \tag{3.46}$$

which is indeed the formula given in the theorem. □

The formula in theorem 3.2 naturally does not reproduce the exact results from the previous chapter due to inclusion of the reward structures in the multivariate setting. Instead theorem 3.2 yields the formulas

$$\mathbb{E}[Y_i] = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{1}_k^\top = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{R}_{.i}, \tag{3.47}$$

$$\mathbb{V}[Y_i] = 2\boldsymbol{\alpha}\left(\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\right)^2\boldsymbol{1}_k^\top = 2\boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{R}_{.i}, \tag{3.48}$$

$$\mathbb{E}\left[Y_iY_j\right] = \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{R}_{.j} + \boldsymbol{\alpha}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.j})\boldsymbol{U}\boldsymbol{R}_{.i}. \tag{3.49}$$

A third order cross-moment can be considered to illustrate how the formula in theorem 3.2 works.

Example 3.1

Consider the third order cross-moment $\mathbb{E}\left[Y_i^2Y_j\right]$. In this case is $h_i = 2$ and $h_j = 1$, while $h_p = 0$ for all $p \in \{1, ..., n\}\setminus\{i, j\}$, which implies that $h = 3$. There are therefore $3! = 6$ permutations of the 3-tuple (i, i, j), and these permutations are ordered and assigned a number, e.g. $\sigma_1 = (i, i, j)$, $\sigma_2 = (i, j, i)$, and so on. The permutations represent the different ways the governing moment-generating function can be differentiated twice with regard to θ_i and once with regard to θ_j. Notice that identical permutations are allowed and may be counted as different tuples, e.g. $\sigma_1 = (i, i, j)$ and $\sigma_3 = (i, i, j)$ or $\sigma_2 = (i, j, i)$ and $\sigma_4 = (i, j, i)$. According to the theorem, the cross-moment is given by

$$\mathbb{E}\left[Y_i^2Y_j\right] = \boldsymbol{\alpha}\left(\sum_{\ell=1}^{6}\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.\sigma_\ell(1)})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.\sigma_\ell(2)})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.\sigma_\ell(3)})\right)\boldsymbol{1}_k^\top \tag{3.50}$$

$$= \boldsymbol{\alpha}\left[\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.j}) + ... + \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.j})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\right]\boldsymbol{1}_k^\top \tag{3.51}$$

$$= 2\boldsymbol{\alpha}\left[\left(\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\right)^2\boldsymbol{U}\boldsymbol{R}_{.j} + \left(\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.j}) + \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.j})\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}_{.i})\right)\boldsymbol{U}\boldsymbol{R}_{.i}\right]. \quad \square \tag{3.52}$$

The phase-type representation and the cross-moments describe features of the joint distribution of the random vector components, but as evident from example 3.1 they also provide complete information about the marginal distributions of the components. The main difference between the univariate and the multivariate setting is the reward structures, and since it is customary to specify univariate phase-type distributions without reward structures, the marginal distributions cannot be obtained directly from the MPH* representation.

Definition 3.1 shows that the marginal distribution associated with the random vector component Y_i results from a transform of a univariate phase-type distribution $\text{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$ using the linear reward structure $\boldsymbol{R}_{\cdot i}$. An appropriate application of theorem 2.14 would therefore yield a valid phase-type representation for the marginal distribution. This approach also works in a more general context, and the next theorem is concerned with projections arising from conical combinations of the random vector components. The result alone is of independent interest, but it has a particularly important corollary, which shows that the MPH* distributions belong to the MVPH class proposed by Bladt and Nielsen.

Before stating the theorem, a prerequisite result and some remarks on notation are needed. The theorem considers a row vector $\boldsymbol{w}$ of non-negative real numbers, which can be used to partition the state-space of the governing Markov process. Similarly to what was done in section 2.4, the state-space of transient states is partitioned into $E^+ = \{i \in E^* | (\boldsymbol{R}\boldsymbol{w})_i > 0\}$ and $E^0 = \{i \in E^* | (\boldsymbol{R}\boldsymbol{w})_i = 0\}$, which corresponds to partitioning the transient states according to whether the conical combination generated by $\boldsymbol{w}$ earns a positive reward or not. The arguments preceding theorem 2.14 now apply again and allow for expressing the initial distribution and the sub-generator of the underlying process in the block-partitioned forms. The initial distribution then takes the form $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_+, \boldsymbol{\alpha}_0)$, and the sub-generator takes the form given in equation (2.100), where $\boldsymbol{T}$ is replaced by $\boldsymbol{S}$, and the indices are given in the subscripts. The reward matrix can of course also be partitioned as $\boldsymbol{R} = (\boldsymbol{R}_+, \boldsymbol{R}_0)^\top$, which should be obvious from the construction. Finally, the cardinalities of the two constructed compartments are denoted by $d_+ = |E^+|$ and $d_0 = |E^0|$.

Theorem 3.3

Let $\boldsymbol{w}$ be an n-dimensional row vector of non-negative real numbers, and let $\boldsymbol{Y} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ be an n-dimensional random vector, where the initial distribution and the sub-generator are ordered as described above. The projection $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ then follows a phase-type distribution with representation $(\boldsymbol{\alpha_w}, \boldsymbol{S_w})$, where

$$\boldsymbol{\alpha_w} = \boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0 \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \tag{3.53}$$

and

$$\boldsymbol{S_w} = diag\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)^{-1} \left(\boldsymbol{S}_{++} + \boldsymbol{S}_{+0}(-\boldsymbol{S}_{00})^{-1}\boldsymbol{S}_{0+}\right). \tag{3.54}$$

Proof. The Laplace-Stieltjes transform of the projection is given by

$$\mathcal{L}_{\boldsymbol{w}}(\theta) = \mathbb{E}\left[e^{-\theta\langle \boldsymbol{Y}, \boldsymbol{w}\rangle}\right] = \mathbb{E}\left[e^{-\langle \boldsymbol{Y}, \boldsymbol{w}\theta\rangle}\right] = \alpha_{k+1} + \boldsymbol{\alpha}\left(\boldsymbol{I}_k + \theta \boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{Rw})\right)^{-1} \boldsymbol{1}_k^\top \tag{3.55}$$

since the inner product is linear in the second argument as implicitly assumed in equation (3.9). The Green matrix in equation (3.55) can be found as the inverse of a block-matrix, see e.g. Lemma A.1.1 on page 711 of Bladt and Nielsen (2017), which results in the Green matrix

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{++} & \boldsymbol{U}_{+0} \\ \boldsymbol{U}_{0+} & \boldsymbol{U}_{00} \end{pmatrix} = (-\boldsymbol{S})^{-1} = \begin{pmatrix} -\boldsymbol{S}_{++} & -\boldsymbol{S}_{+0} \\ -\boldsymbol{S}_{0+} & -\boldsymbol{S}_{00} \end{pmatrix}^{-1} \tag{3.56}$$

with

$$\begin{aligned} \boldsymbol{U}_{++} &= \left(-\boldsymbol{S}_{++} - \boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\right)^{-1}, \\ \boldsymbol{U}_{+0} &= \left(-\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00} - \boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{+0}\right)^{-1}, \\ \boldsymbol{U}_{0+} &= \left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++} - \boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\right)^{-1}, \\ \boldsymbol{U}_{00} &= \left(-\boldsymbol{S}_{00} - \boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{+0}\right)^{-1}. \end{aligned} \tag{3.57}$$

The matrix $\boldsymbol{U}_{0+}$ does not appear directly in the referenced lemma, but arises from a reformulation. According to the lemma, the block is given by

$$\boldsymbol{U}_{0+} = \left(-\boldsymbol{S}_{00} - \boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{+0}\right)^{-1}\boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1} = -\left(\boldsymbol{S}/\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1}, \tag{3.58}$$

where $\boldsymbol{S}/\boldsymbol{S}_{++}$ denotes the Schur complement of $\boldsymbol{S}_{++}$ in $\boldsymbol{S}$. However, by comparing the results on pages 432 and 433 of Gallier (2011), the matrix may be expressed as

$$\boldsymbol{U}_{0+} = -\left(\boldsymbol{S}/\boldsymbol{S}_{++}\right)^{-1}\boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++}\right)^{-1} = -\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\left(\boldsymbol{S}/\boldsymbol{S}_{00}\right)^{-1} \tag{3.59}$$

$$= \left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\left(-\boldsymbol{S}_{++} - \boldsymbol{S}_{+0}\left(-\boldsymbol{S}_{00}\right)^{-1}\boldsymbol{S}_{0+}\right)^{-1}, \tag{3.60}$$

which justifies the result in equation (3.57). Alternatively, the result could have been deduced through symmetry arguments based on the structure and form of $\boldsymbol{U}_{+0}$. The resulting blocks in equation (3.57) are only well-defined if the matrices $\boldsymbol{S}_{++}$ and $\boldsymbol{S}_{00}$ along with their Schur complements are invertible. The matrices $\boldsymbol{S}_{++}$ and $\boldsymbol{S}_{00}$ are indeed invertible since they are themselves sub-generators, and since $\boldsymbol{S}$ is invertible, it follows from the calculations

$$\det\left(\boldsymbol{S}\right) = \det\left(\boldsymbol{S}/\boldsymbol{S}_{++}\right)\det\left(\boldsymbol{S}_{++}\right), \tag{3.61}$$

$$\det\left(\boldsymbol{S}\right) = \det\left(\boldsymbol{S}/\boldsymbol{S}_{00}\right)\det\left(\boldsymbol{S}_{00}\right) \tag{3.62}$$

that their associated Schur complements in $\boldsymbol{S}$ are also non-singular. The transform in equation (3.55) can thus be recast as

$$\mathcal{L}_{\boldsymbol{w}}(\theta) = \alpha_{k+1} + (\boldsymbol{\alpha}_+, \boldsymbol{\alpha}_0)\left(\begin{pmatrix}\boldsymbol{I}_{d_+} & \boldsymbol{0}_{d_+\times d_0} \\ \boldsymbol{0}_{d_0\times d_+} & \boldsymbol{I}_{d_0}\end{pmatrix} + \theta\begin{pmatrix}\boldsymbol{U}_{++} & \boldsymbol{U}_{+0} \\ \boldsymbol{U}_{0+} & \boldsymbol{U}_{00}\end{pmatrix}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{w})\right)^{-1}\boldsymbol{1}_k^\top. \tag{3.63}$$

In addition, the ordering of the states into the two compartments implies that

$$\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{w}) = \operatorname{diag}\left((\boldsymbol{R}\boldsymbol{w})_1, ..., (\boldsymbol{R}\boldsymbol{w})_{d_+}, 0, ..., 0\right), \tag{3.64}$$

where there are d_0 zeros. Therefore,

$$\begin{pmatrix}\boldsymbol{U}_{++} & \boldsymbol{U}_{+0} \\ \boldsymbol{U}_{0+} & \boldsymbol{U}_{00}\end{pmatrix}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{w}) = \begin{pmatrix}\boldsymbol{U}_{++}\operatorname{diag}\left((\boldsymbol{R}\boldsymbol{w})_1, ..., (\boldsymbol{R}\boldsymbol{w})_{d_+}\right) & \boldsymbol{0}_{d_+\times d_0} \\ \boldsymbol{U}_{0+}\operatorname{diag}\left((\boldsymbol{R}\boldsymbol{w})_1, ..., (\boldsymbol{R}\boldsymbol{w})_{d_+}\right) & \boldsymbol{0}_{d_0\times d_0}\end{pmatrix}. \tag{3.65}$$

This simplifies the Laplace-Stieltjes transform to

$$\mathcal{L}_{\boldsymbol{w}}(\theta) = \alpha_{k+1} + (\boldsymbol{\alpha}_+, \boldsymbol{\alpha}_0)\begin{pmatrix}\boldsymbol{I}_{d_+} + \theta\boldsymbol{U}_{++}\operatorname{diag}\left((\boldsymbol{R}\boldsymbol{w})_1, ..., (\boldsymbol{R}\boldsymbol{w})_{d_+}\right) & \boldsymbol{0}_{d_+\times d_0} \\ \theta\boldsymbol{U}_{0+}\operatorname{diag}\left((\boldsymbol{R}\boldsymbol{w})_1, ..., (\boldsymbol{R}\boldsymbol{w})_{d_+}\right) & \boldsymbol{I}_{d_0}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{1}_{d_+}^\top \\ \boldsymbol{1}_{d_0}^\top\end{pmatrix}. \tag{3.66}$$

The inverse matrix in equation (3.66) exists whenever the top left block is non-singular due to the fact that one of the blocks is a matrix of zeros. The matrix $\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)$ is a diagonal matrix whose diagonal elements are positive by definition, which implies that it is invertible, and the inverse of $\boldsymbol{U}_{++}$ is the Schur complement $-\boldsymbol{S}/\boldsymbol{S}_{00}$, whose existence has already been established. Furthermore, both matrices have only eigenvalues with strictly positive real parts (the arguments from theorem 2.14 justify the claim for the matrix $\boldsymbol{U}_{++}$), and consequently, the inverse matrix in equation (3.66) exists for all values of θ with real parts greater than some real negative value $\theta_{\boldsymbol{w}}$. Inverting the matrix and computing the matrix-vector product then yields

$$\begin{aligned}\mathcal{L}_{\boldsymbol{w}}(\theta) &= \alpha_{k+1} + \boldsymbol{\alpha}_+ \left(\boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right)^{-1} \mathbf{1}_{d_+}^\top + \boldsymbol{\alpha}_0 \boldsymbol{I}_{d_0} \mathbf{1}_{d_0}^\top \\ &\quad - \boldsymbol{\alpha}_0 \theta \boldsymbol{U}_{0+}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right) \left(\boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right)^{-1} \mathbf{1}_{d_+}^\top \end{aligned} \tag{3.67}$$

$$\begin{aligned}&= \left(\boldsymbol{\alpha}_+ - \boldsymbol{\alpha}_0 \theta \boldsymbol{U}_{0+}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right) \left(\boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right)^{-1} \mathbf{1}_{d_+}^\top \\ &\quad + \alpha_{k+1} + \boldsymbol{\alpha}_0 \boldsymbol{I}_{d_0} \mathbf{1}_{d_0}^\top .\end{aligned} \tag{3.68}$$

The structure of the blocks in equation (3.57) suggests that $\boldsymbol{U}_{0+} = \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \boldsymbol{U}_{++}$, which prompts the rewriting

$$\theta \boldsymbol{U}_{0+}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right) = \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right). \tag{3.69}$$

By adding and subtracting an identity matrix of appropriate dimensions, the above equation extends to

$$\begin{aligned}&\theta \boldsymbol{U}_{0+}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right) \\ &\quad = \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \left(\boldsymbol{I}_{d_+} - \boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right) &(3.70) \\ &\quad = \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \left(\boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right) - \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+}. &(3.71)\end{aligned}$$

The result in equation (3.71) is inserted back into the transform in equation (3.68) to simplify the expression to

$$\begin{aligned}\mathcal{L}_{\boldsymbol{w}}(\theta) = &\left(\boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0 \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+}\right) \left(\boldsymbol{I}_{d_+} + \theta \boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right)^{-1} \mathbf{1}_{d_+}^\top \\ &+ \alpha_{k+1} + \boldsymbol{\alpha}_0 \boldsymbol{I}_{d_0} \mathbf{1}_{d_0}^\top - \boldsymbol{\alpha}_0 \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \mathbf{1}_{d_+}^\top .\end{aligned} \tag{3.72}$$

Comparing this expression with the formula in theorem 2.5 (when written on the equivalent form as in theorem 3.1), the transform is recognized as the Laplace-Stieltjes transform of a phase-type distributed random variable with initial distribution

$$\boldsymbol{\alpha}_{\boldsymbol{w}} = \boldsymbol{\alpha}_+ + \boldsymbol{\alpha}_0 \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+} \tag{3.73}$$

and sub-generator

$$\boldsymbol{S}_{\boldsymbol{w}} = -\left(\boldsymbol{U}_{++}\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)\right)^{-1} = -\text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)^{-1} \boldsymbol{U}_{++}^{-1}, \tag{3.74}$$

which by equation (3.57) yields the desired result

$$\boldsymbol{S}_{\boldsymbol{w}} = \text{diag}\left((\boldsymbol{Rw})_1, ..., (\boldsymbol{Rw})_{d_+}\right)^{-1} \left(\boldsymbol{S}_{++} + \boldsymbol{S}_{+0} \left(-\boldsymbol{S}_{00}\right)^{-1} \boldsymbol{S}_{0+}\right). \tag{3.75}$$

In conclusion, the projection $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ has the Laplace-Stieltjes transform given in equation (3.72), which is well-defined for all arguments with real parts greater that some negative constant $\theta_{\boldsymbol{w}}$, and the projection therefore follows a phase-type distribution with representation $(\boldsymbol{\alpha}_{\boldsymbol{w}}, \boldsymbol{S}_{\boldsymbol{w}})$. □

The core idea behind the above proof can be obscured by the cumbersome calculations and expressions, but it is actually rather simple. The vector $\boldsymbol{Rw}$ is a reward structure that describes the rate at which the conical combination $\langle \boldsymbol{Y}, \boldsymbol{w} \rangle$ earns reward in the various transient states, and theorem 2.14 can thus be applied to transform the PH($\boldsymbol{\alpha}, \boldsymbol{S}$)-distribution in accordance with the reward structure. The latter approach is comparatively more widespread due to its simplicity, see e.g. section 3.1.3 of Larsen (2018), but the above method is valid also for MME* distributions, cf. problem 8.5.29 in Bladt and Nielsen (2017). Theorem 8.1.8 in the latter reference also demonstrates the reverse implication of theorem 3.1, which follows from a straightforward proof.

The previous theorems showcase how the Laplace-Stieltjes transform describes both the joint and marginal laws of a multivariate phase-type distribution. The Laplace-Stieltjes and related transforms completely determine the multivariate distribution, but there are certain calculations and applications, where it is preferable to express the dependence structure through a joint density function or a copula model. Copulas are closely related to joint distribution functions and joint survival functions, and the derivation of these functions are therefore of great interest.

Kulkarni showed in his 1989 paper that the joint survival function associated with a multivariate phase-type (MPH*) distribution satisfies a system of partial differential equations, but no solution has been established thus far. Junker (2016) is a master thesis concerned with identifying a solution to the system, and Junker devises a method to obtain a semi-explicit solution in terms of some complicated integrals, which are intractable and computationally demanding. Furthermore, Junker presents some more promising results in a few special cases, where certain assumptions are imposed on the reward matrix and the sub-generator. As there are no known results in the general case, the emphasis of this dissertation shall be on presenting the system of partial differential equation that governs the joint survival function.

Theorem 3.4

Consider an n-dimensional random vector $\boldsymbol{Y} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ *and define the conditional joint survival functions*

$$\boldsymbol{G}_i(\boldsymbol{y}) = \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_0 = i\right), \quad \forall i \in E, \tag{3.76}$$

where $\boldsymbol{y} = (y_1, ..., y_n)$ *and* $\{X_t; t \geq 0\}$ *is the governing Markov process with the transient states* E^**. Based on these functions, define the vector function* $\boldsymbol{G}(\boldsymbol{y}) = (G_1(\boldsymbol{y}), ..., G_k(\boldsymbol{y}))^\top$ *with the Jacobian* $\boldsymbol{J}$ *whose elements are given by*

$$(\boldsymbol{J})_{ij} = J_{ij} = \frac{\partial G_i}{\partial y_j}, \quad \forall (i,j) \in E^* \times \{1, ..., n\}. \tag{3.77}$$

Then the conditional joint survival functions satisfy the system

$$\left(\boldsymbol{J} \odot \boldsymbol{R}\right) \boldsymbol{1}_n^\top = \boldsymbol{SG}, \tag{3.78}$$

where $\odot$ *denotes the Hadamard product, whenever the joint survival function is (Fréchet) differentiable.*

Proof. Consider the conditional joint survival function for an arbitrary transient state $i \in E^*$. The law of total probability then gives

$$G_i(\boldsymbol{y}) = \sum_{m=1}^{k+1} \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = m, X_0 = i\right) \mathbb{P}(X_h = m | X_0 = i) \tag{3.79}$$

for some arbitrary $h > 0$. A simple rearrangement of the terms then leads to

$$\begin{aligned} &G_i(\boldsymbol{y}) - \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = i, X_0 = i\right) \mathbb{P}(X_h = i | X_0 = i) \\ &\qquad = \sum_{m=1, m\neq i}^{k+1} \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = m, X_0 = i\right) \mathbb{P}(X_h = m | X_0 = i). \end{aligned} \tag{3.80}$$

The next step is to divide both sides by h and take the limit as h vanishes (approaches zero from the right). The resulting terms on the RHS of equation (3.80) are easily determined as

$$\begin{aligned} &\lim_{h\to 0^+} \left(\frac{\mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = m, X_0 = i\right) \mathbb{P}(X_h = m | X_0 = i)}{h} \right) \\ &\qquad = G_m(\boldsymbol{y}) \lim_{h\to 0^+} \left(\frac{\mathbb{P}(X_h = m | X_0 = i)}{h} \right) \qquad (3.81) \\ &\qquad = \begin{cases} G_m(\boldsymbol{y}) S_{im}, & m \in E^* \setminus \{i\} \\ 0, & m = k+1 \end{cases}, \quad \forall \boldsymbol{y} \in \left(\mathbb{R}_0^+\right)^n. \qquad (3.82) \end{aligned}$$

The first equality follows from interchanging the order of multiplication and taking limits (one of the algebraic limit theorems), which is allowed since both quantities in question converge. The second equality is simply a postulate of the Poisson process (Markov process), cf. section 6.1.1 in Pinksy and Karlin (2011). As all the terms on the RHS of equation (3.80) converge, the expression simplifies to

$$\lim_{h\to 0^+} \left(\sum_{\substack{m=1 \\ m\neq i}}^{k+1} \frac{\mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = m, X_0 = i\right) \mathbb{P}(X_h = m | X_0 = i)}{h} \right) = \sum_{m=1, m\neq i}^{k} G_m(\boldsymbol{y}) S_{im}. \tag{3.83}$$

Evaluating the resulting LHS of equation (3.80) after dividing by h and taking the limit requires a short intermediate step. First, define the random variable N_h as the number of jumps of the underlying Markov process until time h (including the time point h). Using Landau notation, the law of total probability then dictates that

$$\begin{aligned} &\mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | X_h = i, X_0 = i\right) \\ &\qquad = \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | N_h = 0, X_h = i, X_0 = i\right) \mathbb{P}\left(N_h = 0 | X_h = i, X_0 = i\right) + o(h), \end{aligned} \tag{3.84}$$

where $o(h)$ denotes a function such that $\lim_{h\to 0^+} o(h)/h = 0$. The first term on the RHS of equation (3.84) represents the case, where the governing process remains in the initial state throughout the time interval $[0, h]$, while the second term represents the probability that the underlying process has made a multi-step transition from state i to itself throughout the time interval $[0, h]$. By the Poisson postulate referenced above, the latter term is a $o(h)$ function, since a Markov process cannot make multiple transitions in an infinitesimal time interval. The first factor of the first term on the RHS of equation (3.84) can now be recast as

$$\begin{aligned} \mathbb{P}\left(Y_1 > y_1, ..., Y_n > y_n | N_h = 0, X_h = i, X_0 = i\right) &= \mathbb{P}\left(Y_1 > y_1 - R_{i1}h, ..., Y_n > y_n - R_{in}h | X_0 = i\right) \qquad (3.85) \\ &= G_i(\boldsymbol{y} - h\boldsymbol{R}_{i.}) \qquad (3.86) \end{aligned}$$

using the condition that the underlying process remains in the initial state until time h in conjunction with the Markov property. These results imply that the LHS of equation (3.80) changes as

$$\begin{aligned}&\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-\mathbb{P}\left(Y_1>y_1,...,Y_n>y_n|X_h=i,X_0=i\right)\mathbb{P}(X_h=i|X_0=i)}{h}\right)\\&=\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-(G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\mathbb{P}\left(N_h=0|X_h=i,X_0=i\right)+o(h))\,\mathbb{P}(X_h=i|X_0=i)}{h}\right).\end{aligned}\tag{3.87}$$

As

$$\mathbb{P}\left(N_h=0|X_h=i,X_0=i\right)\mathbb{P}(X_h=i|X_0=i)=\mathbb{P}\left(N_h=0,X_h=i|X_0=i\right)=\mathbb{P}\left(N_h=0|X_0=i\right),\tag{3.88}$$

the expression in equation (3.87) simplifies to

$$\begin{aligned}&\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-(G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\mathbb{P}\left(N_h=0|X_h=i,X_0=i\right)+o(h))\,\mathbb{P}(X_h=i|X_0=i)}{h}\right)\\&=\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\mathbb{P}\left(N_h=0|X_0=i\right)}{h}+\frac{o(h)}{h}\right)\qquad(3.89)\\&=\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})}{h}+\frac{G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\mathbb{P}\left(N_h>0|X_0=i\right)}{h}+\frac{o(h)}{h}\right).\qquad(3.90)\end{aligned}$$

When the joint survival function is (Fréchet) differentiable, the associated Jacobian exists, which in turn implies that the total derivatives may be expressed through the partial derivatives as

$$\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})}{h}\right)=\sum_{j=1}^{n}\frac{\partial G_i}{\partial y_j}(\boldsymbol{y})R_{ij}=\boldsymbol{R}_{i.}\nabla G_i(\boldsymbol{y}),\tag{3.91}$$

where ∇ denotes the gradient. The algebraic limit theorems thus apply and yield

$$\begin{aligned}&\lim_{h\to 0^+}\left(\frac{G_i(\boldsymbol{y})-G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})}{h}+\frac{G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\mathbb{P}\left(N_h>0|X_0=i\right)}{h}+\frac{o(h)}{h}\right)\\&=\boldsymbol{R}_{i.}\nabla G_i(\boldsymbol{y})+\lim_{h\to 0^+}\left(G_i(\boldsymbol{y}-h\boldsymbol{R}_{i.})\right)\lim_{h\to 0^+}\left(\frac{\mathbb{P}\left(N_h>0|X_0=i\right)}{h}\right)\qquad(3.92)\\&=\boldsymbol{R}_{i.}\nabla G_i(\boldsymbol{y})-S_{ii}G_i(\boldsymbol{y}).\qquad(3.93)\end{aligned}$$

Dividing by h and taking the limit as h vanishes on both sides of equation (3.80) therefore leads to

$$\boldsymbol{R}_{i.}\nabla G_i(\boldsymbol{y})-S_{ii}G_i(\boldsymbol{y})=\sum_{m=1,m\neq i}^{k}G_m(\boldsymbol{y})S_{im},\tag{3.94}$$

cf. equations (3.83) and (3.93). Moving the second term on the LHS of equation (3.94) to the RHS and collecting the equations for all transient states in E^* produces the following system of partial differential equations governing the joint survival function

$$\boldsymbol{R}_{i.}\nabla G_i(\boldsymbol{y})=\sum_{m=1}^{k}G_m(\boldsymbol{y})S_{im},\quad\forall i\in E^*,\tag{3.95}$$

which is equivalent to the matrix-equation

$$\left(\boldsymbol{J}\odot\boldsymbol{R}\right)\boldsymbol{1}_n^\top=\boldsymbol{S}\boldsymbol{G}.\tag{3.96}$$

The latter equation only holds whenever the result in (3.91) is valid for all $i\in E^*$, which is exactly when the joint survival function is (Fréchet) differentiable. When it is well-defined, the equation in (3.96) coincides with (3.78) in the theorem, which completes the proof. □

Section 4.8 of Junker (2016) discusses conditions for differentiability of the joint survival function (existence of the partial derivatives). The results obtained by Junker have not been peer-reviewed nor verified in any academic journals or conferences, but they suggest that the differentiability can be examined by analysing the possible paths of the underlying process. In particular, the differentiability should be ensured whenever the reward matrix has linearly independent rows.

The system of partial differential equations presented in theorem 3.4 has several boundary conditions, cf. page 443 in Bladt and Nielsen (2017), but as mentioned earlier, there are no known analytical or numerical solutions in the general case. The lack of analytical expressions for the probability functions associated with the joint distribution further explains why most derivations within the theory of phase-type distributions are based on integral transforms.

This section has thus far covered basic concepts related to multivariate distributions. The next parts will be concerned with closure properties similar to those presented in chapter 2. Many of the closure properties from the univariate phase-type distributions carry over to the multivariate setting because the multivariate distributions are derived from a single Markov process just like the univariate distributions. The results on process concatenation applied in chapter 2 are therefore also applicable in the context of multivariate distributions, and consequently, the closure properties generalise easily to phase-type distributions in higher dimensions.

Theorem 3.5

Let $\boldsymbol{X} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $\boldsymbol{Y} \sim MPH^*_m(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{V})$ be independent n-dimensional random vectors. Then the sum (convolution) $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$ has a multivariate phase-type (MPH) distribution with representation $(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{W})$ of dimension $k+m$, where the initial distribution is given by*

$$\boldsymbol{\pi} = (\boldsymbol{\alpha}, \alpha_{k+1}\boldsymbol{\beta}), \tag{3.97}$$

the sub-generator is given by

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s\beta} \\ \boldsymbol{0}_{m\times k} & \boldsymbol{T} \end{pmatrix}, \tag{3.98}$$

and the reward matrix (structure) takes the form

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{V} \end{pmatrix}. \tag{3.99}$$

Proof. Consider an arbitrary component of the random vector $\boldsymbol{Z}$, say Z_j for some $j \in \{1, ..., n\}$. Definition 3.1 states that this component is defined as

$$Z_j = X_j + Y_j = \sum_{h=1}^{k} A_h R_{hj} + \sum_{\ell=1}^{m} B_\ell V_{\ell j}, \tag{3.100}$$

where $\{A_h\}_{h=1}^k$ are the total sojourn times (prior to absorption) of the underlying process governing $\boldsymbol{X}$ in the respective k transient states, and $\{B_\ell\}_{\ell=1}^m$ are the total sojourn times (prior to absorption) of the underlying process governing $\boldsymbol{Y}$ in the respective m transient states. A new Markov process is then constructed by concatenating the two underlying processes using the result from Sharpe (1988) like in theorem 2.7. The constructed process has the initial distribution $\boldsymbol{\pi}$ given in (3.97) and the sub-generator $\boldsymbol{D}$ given in (3.98).

Similarly to the univariate case, the new process has two compartments of transient states, and the sojourn times in transient states belonging to different compartments are naturally independent. These sojourn times of the constructed process in the transient states (prior to absorption) shall be denoted by $\{C_g\}_{g=1}^{k+m}$.

The concatenation thus entails that

$$C_g = \begin{cases} A_g, & 1 \leq g \leq k \\ B_{g-k}, & k < g \leq k+m \end{cases}, \tag{3.101}$$

which allows for rewriting equation (3.100) as

$$Z_j = \sum_{h=1}^{k} A_h R_{hj} + \sum_{\ell=1}^{m} B_\ell V_{\ell j} = \sum_{g=1}^{k} C_g R_{gj} + \sum_{g=k+1}^{k+m} C_g V_{(g-k)j}. \tag{3.102}$$

By introducing the reward structure defined in (3.99), the expression in equation (3.102) simplifies to

$$Z_j = \sum_{g=1}^{m+k} C_g W_{gj}. \tag{3.103}$$

Since j was selected arbitrarily, the statement holds for all values of $j \in \{1, ..., n\}$, which by definition 3.1 implies that the random vector $\boldsymbol{Z}$ has a MPH* distribution with the representation proposed in the theorem. □

As alluded to earlier, theorem 3.5 follows almost immediately from the arguments used to prove theorem 2.7. In particular, the result in (3.101) arising from the principle of process concatenation makes the derivation rather straightforward. The same arguments also apply to the next theorem on mixtures of multivariate phase-type distributions, which shall be presented (almost) without proof.

Theorem 3.6

Let $\boldsymbol{X} \sim MPH^*_k(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $\boldsymbol{Y} \sim MPH^*_m(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{V})$ be independent n-dimensional random vectors. Then a mixture $\boldsymbol{Z}$ with mixture probabilities $\mathbb{P}(\boldsymbol{Z} = \boldsymbol{X}) = p_{\boldsymbol{X}}$ and $\mathbb{P}(\boldsymbol{Z} = \boldsymbol{Y}) = p_{\boldsymbol{Y}}$ has a multivariate phase-type (MPH) distribution with representation $(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{W})$ of dimension $k+m$, where the initial distribution is given by*

$$\boldsymbol{\pi} = (p_{\boldsymbol{X}} \boldsymbol{\alpha}, p_{\boldsymbol{Y}} \boldsymbol{\beta}), \tag{3.104}$$

the sub-generator is given by

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{0}_{k \times m} \\ \boldsymbol{0}_{m \times k} & \boldsymbol{T} \end{pmatrix}, \tag{3.105}$$

and the reward matrix (structure) takes the form

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{V} \end{pmatrix}. \tag{3.106}$$

Proof. The proof generalises theorem 2.8 completely analogously to how theorem 3.5 generalises theorem 2.7. The principle of process concatenation from Sharpe (1988) leads to the same Markov process as the process constructed in theorem 2.8, and then the arguments from theorem 3.5 apply to justify the theorem. □

Theorems 3.5 and 3.6 yield an immediate corollary, which embodies the main closure properties of the multivariate phase-type distributions.

Corollary 3.1
The MPH distributions are closed under finite convolutions and finite mixtures.*

The result in the above corollary gives rise to the question of whether theorem 2.9 also generalises to the multivariate setting. More generally, one might consider how discrete and multivariate phase-type distributions can be combined to construct multivariate distributions. One example could be to consider the joint distribution between a continuous phase-type distributions and the number of jumps (transitions) in the governing Markov process prior to absorption. Another example could be to consider multivariate compound distributions obtained by using a discrete phase-type distribution as the mixing distribution. The latter example is exactly the multivariate counterpart to theorem 2.9, and these infinite mixtures of multivariate distributions are dealt with in the following theorem.

Theorem 3.7
Let $N \sim DPH_m(\boldsymbol{\alpha}, \boldsymbol{S})$ and, independently of N, define the independent and identically distributed random vectors $\boldsymbol{X}_i \sim MPH^*{}_k(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{R})$, where $||\boldsymbol{\beta}||_1 = 1$, for all $i \in \mathbb{N}$. Then the (possibly infinite) mixture $\boldsymbol{Z}$ generated by the random sum

$$\boldsymbol{Z} = \sum_{i=1}^{N} \boldsymbol{X}_i \tag{3.107}$$

has a MPH distribution with initial distribution $\boldsymbol{\pi} = \boldsymbol{\alpha} \otimes \boldsymbol{\beta}$, sub-generator $\boldsymbol{V} = \boldsymbol{I}_m \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t}\boldsymbol{\beta}$, and reward structure $\boldsymbol{W} = \boldsymbol{1}_m^\top \otimes \boldsymbol{R}$, where the resulting representation has dimension $d = mk$.*

Proof. The arguments used in the proof of theorem 2.9 also apply here, which means that the Markov process generating the infinite mixture is constructed like in the univariate case. The initial distribution and the sub-generator in the representation are therefore taken from theorem 2.9. The structure of the sub-generator is specified in equation (2.80), and it is evident from this structure that the reward structure should be a concatenation of m versions of $\boldsymbol{R}$, which is exactly given by $\boldsymbol{W} = \boldsymbol{1}_m^\top \otimes \boldsymbol{R}$. In conclusion, the possibly infinite mixture associated with the random sum in (3.107) has the multivariate phase-type distribution proposed in the theorem. □

The previous three results illustrate an attractive property of the construction propounded by Kulkarni. The construction is derived from a single Markov process and a reward structure, which implies that the Markov processes governing convolutions and mixtures are the same in the univariate and multivariate cases. Proving the closure properties in the multivariate context is therefore primarily a problem of identifying an appropriate reward structure, which is usually a trivial task.

We shall conclude this section with an example on an application of theorem 3.5, which shows how the reward structure combines with the constructed Markov processes to produce a convolution of multivariate phase-type distributions.

Example 3.2

Consider two bivariate random vectors $\boldsymbol{X} = (X_1, X_2) \sim \text{MPH}^*{}_2(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ and $\boldsymbol{Y} = (Y_1, Y_2) \sim \text{MPH}^*{}_2(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{V})$ with initial distributions

$$\boldsymbol{\alpha} = \left(\frac{1}{3}, \frac{1}{3}\right), \quad \boldsymbol{\beta} = \left(\frac{2}{3}, 0\right), \tag{3.108}$$

sub-generators

$$\boldsymbol{S} = \begin{pmatrix} -2 & 2 \\ 1 & -2 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -\frac{1}{10} & \frac{1}{10} \\ 0 & -\frac{1}{5} \end{pmatrix}, \tag{3.109}$$

and reward structures

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 \\ 3 & 5 \end{pmatrix}, \quad \boldsymbol{V} = \begin{pmatrix} 0 & 4 \\ 2 & 0 \end{pmatrix}. \tag{3.110}$$

Figure 3.3 shows two separate realizations of the two random vectors

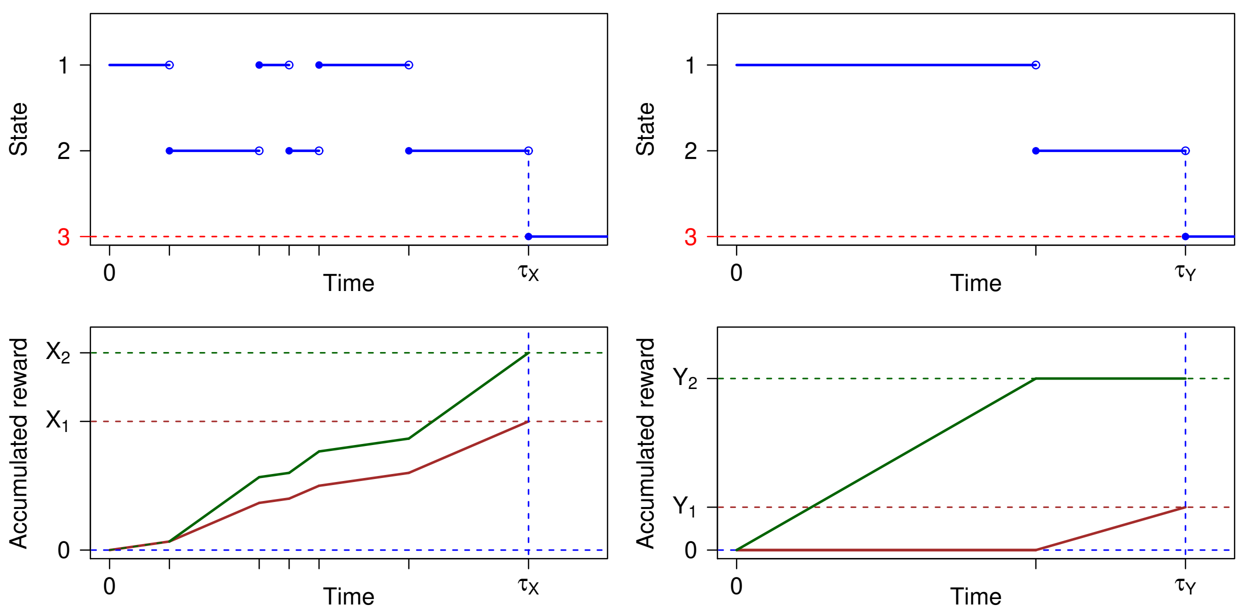

Figure 3.3: *Two separate realizations of the random vectors* $\boldsymbol{X}$ *(Left) and* $\boldsymbol{Y}$ *(Right) from example 3.2.*

Assuming the random vectors are independent, the sum of the two random vectors $\boldsymbol{Z} = (Z_1, Z_2) = \boldsymbol{X}+\boldsymbol{Y}$ then has the following representation according to theorem 3.5: $\boldsymbol{Z} \sim \text{MPH}^*{}_4(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{W})$ with initial distribution

$$\boldsymbol{\pi} = (\boldsymbol{\alpha}, \alpha_3 \boldsymbol{\beta}) = \left(\frac{1}{3}, \frac{1}{3}, \frac{2}{9}, 0\right), \tag{3.111}$$

sub-generator

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} \\ \boldsymbol{0}_{2\times 2} & \boldsymbol{T} \end{pmatrix} = \begin{pmatrix} -2 & 2 & 0 & 0 \\ 1 & -2 & \frac{2}{3} & 0 \\ 0 & 0 & -\frac{1}{10} & \frac{1}{10} \\ 0 & 0 & 0 & -\frac{1}{5} \end{pmatrix}, \tag{3.112}$$

and reward structure

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{V} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & 5 \\ 0 & 4 \\ 2 & 0 \end{pmatrix}. \tag{3.113}$$

A graphical representation of this operation is given below, where the two separate realizations in figure 3.3 is combined to generate a realization of the random vector $\boldsymbol{Z}$. Figure 3.4 thus illustrates how the realization of the concatenated Markov process generates the realization of a random vector which is generated by a sum (the distribution is generated by a convolution).

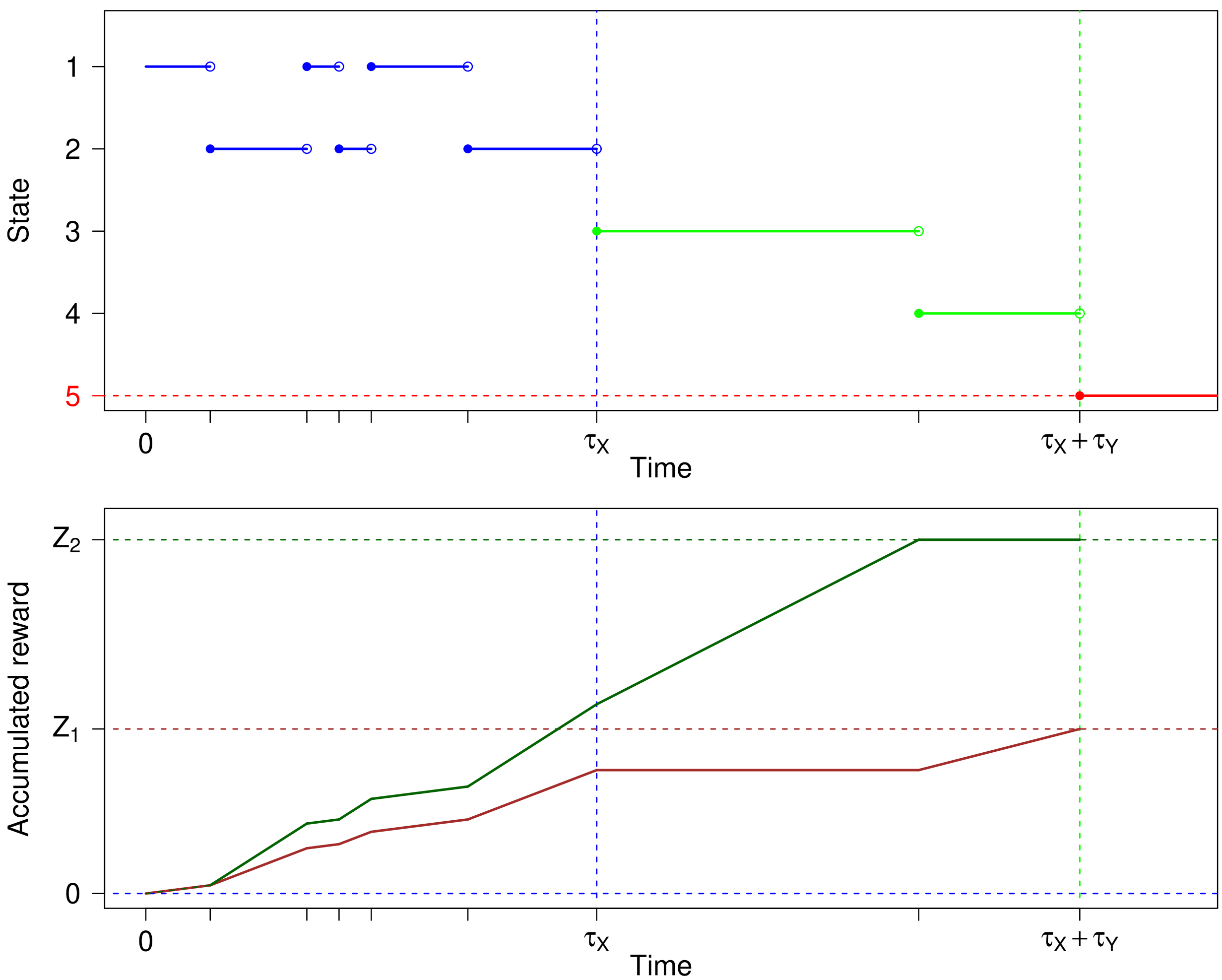

Figure 3.4: *Realization of the concatenated Markov process and the bivariate random vector* $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$ *from example 3.2.*

The uppermost plot in figure 3.4 shows the concatenated Markov process, where the *blue* part of the process is associated with the governing process of $\boldsymbol{X}$ and the *green* part with process governing $\boldsymbol{Y}$. The bottom plot of the figure shows how the two vector components (Z_1: *Brown*, Z_2: *Dark green*) accumulate reward throughout the evolution of process. □

3.4 Representations for multivariate distributions

The previous sections have covered basic definitions and main properties of the multivariate phase-type distributions. The closure properties and the interpretability of the reward structures make MPH* distributions tractable models for multivariate data, and they can further make identifying MPH* representations of known multivariate exponential and gamma distributions easier. In this section, we present MPH* representations of a variety of multivariate exponential and gamma distributions known from the scientific literature. Many of these known distributions are defined through convolution, mixing, or compounding, which already suggests that they might belong to the class of MPH* distributions. The following examples will show how the theoretic foundations established in the former sections can be used to derive representations of a few important multivariate distributions.

Example 3.3

The first example of the section in concerned with McKay's bivariate gamma distribution, whose construction is based on the principle of sharing exponential terms. When the shape parameters in the distribution take natural values, a random vector $\boldsymbol{Y} = (Y_1, Y_2)$ with a bivariate gamma distribution in the sense of McKay can be constructed as:

$$Y_1 = \frac{1}{\lambda_1} \sum_{j=1}^{k_1} Z_j, \tag{3.114}$$

$$Y_2 = \frac{1}{\lambda_2} \sum_{j=1}^{k_1+k_2} Z_j, \tag{3.115}$$

where the random variables $\{Z_j\}_{j=1}^{k_1+k_2}$ are independent and identically distributed with an exponential(1) distribution. The parameters k_1 and k_2 are natural numbers, while the rate parameters λ_1 and λ_2 are positive real numbers. It follows immediately from the construction that

$$Y_2 = \frac{\lambda_1}{\lambda_2} Y_1 + \frac{1}{\lambda_2} \sum_{j=k_1+1}^{k_1+k_2} Z_j, \tag{3.116}$$

which indicates that Y_2 is formed as a sum of a scaling of Y_1 and an independent gamma distribution. The marginal distributions are $Y_1 \sim \gamma(\lambda_1, k_1)$ and $Y_2 \sim \gamma(\lambda_2, k_1 + k_2)$, and the correlation between the two components is given as

$$\rho = \mathrm{Corr}(Y_1, Y_2) = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathbb{V}[Y_1]\mathbb{V}[Y_2]}} = \frac{\mathrm{Cov}\left(Y_1, \frac{\lambda_1}{\lambda_2} Y_1\right)}{\sqrt{\mathbb{V}[Y_1]\mathbb{V}[Y_2]}} = \frac{\lambda_1}{\lambda_2}\sqrt{\frac{\mathbb{V}[Y_1]}{\mathbb{V}[Y_2]}} = \sqrt{\frac{k_1}{k_1 + k_2}}, \tag{3.117}$$

i.e. the square root of the ratio between the shape parameters in the distribution. The structure of the distribution and its associated properties should be compared to those of the bivariate gamma distribution by Dussauchoy and Berland, which has a similar correlation structure and a similar decomposition property. The distribution by Dussauchoy and Berland is studied more in depth in the next chapter, where some of the differences will be further examined.

McKay's distribution can be further classified according to its joint density function. The joint density function follows directly from the joint distribution function

$$F(x, y) = \mathbb{P}(Y_1 \leq x, Y_2 \leq y) = \int_0^x \mathbb{P}\left(\frac{1}{\lambda_2}\sum_{j=k_1+1}^{k_1+k_2} Z_j \leq y - \frac{\lambda_1}{\lambda_2}z\right) f_{Y_1}(z)dz. \tag{3.118}$$

By defining $X = \lambda_2^{-1}\sum_{j=k_1+1}^{k_1+k_2} Z_j \sim \gamma(\lambda_2, k_2)$, equation (3.118) simplifies to

$$F(x, y) = \int_0^x F_X\left(y - \frac{\lambda_1}{\lambda_2}z\right) f_{Y_1}(z)dz, \tag{3.119}$$

which leads to

$$f(x, y) = \frac{\partial^2}{\partial y \partial x}F(x, y) = \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}\int_0^x F_X\left(y - \frac{\lambda_1}{\lambda_2}z\right) f_{Y_1}(z)dz\right) = \frac{\partial}{\partial y}F_X\left(y - \frac{\lambda_1}{\lambda_2}x\right) f_{Y_1}(x), \tag{3.120}$$

which in turn yields the joint density function

$$f(x, y) = f_X\left(y - \frac{\lambda_1}{\lambda_2}x\right) f_{Y_1}(x) = \frac{\left(y - \frac{\lambda_1}{\lambda_2}x\right)^{k_2-1}\lambda_2^{k_2}e^{-\lambda_2\left(y-\frac{\lambda_1}{\lambda_2}x\right)}}{(k_2-1)!}\frac{x^{k_1-1}\lambda_1^{k_1}e^{-\lambda_1 x}}{(k_1-1)!} \tag{3.121}$$

$$= \lambda_1\lambda_2\frac{(\lambda_2 y - \lambda_1 x)^{k_2-1}(\lambda_1 x)^{k_1-1}}{(k_2-1)!(k_1-1)!}e^{-\lambda_2 y}, \quad y > \frac{\lambda_1}{\lambda_2}x > 0. \tag{3.122}$$

The joint distribution function is generally not available for McKay's bivariate gamma distribution, even though it can be expressed through a special transcendental function, cf. page 331 in Balakrishnan and Lai (2009). Since copulas are generally difficult to derive for multivariate gamma distributions, there are only few alternative descriptions of the joint distribution; one of which is through phase-type representations and their associated joint integral transforms. McKay's bivariate distribution thus constitutes a prime example of a distribution, where the framework of multivariate phase-type distributions provide useful information about the dependence structure and the construction of the multivariate distribution that would be difficult to obtain otherwise.

Equations (3.114) through (3.116) give rise to various MPH* representations of the bivariate distributions. In this thesis, two different representations shall be presented because they emphasize flexibility of the phase-type representations. The most natural and commonly applied representation invokes a Markov process whose time until absorption is Erlang distributed and uses the reward structure to generate the appropriate distribution. This representation $(\boldsymbol{\alpha}_1, \boldsymbol{S}_1, \boldsymbol{R}_1)$ has the initial distribution $\boldsymbol{\alpha}_1 = (1, \mathbf{0}_{k_1+k_2-1})$ and the bidiagonal sub-generator $\boldsymbol{S}_1 = \boldsymbol{B}_{k_1+k_2}(1)$, which follows the form:

$$\boldsymbol{B}_k(\lambda) = \begin{matrix} 1 \\ 2 \\ \vdots \\ k-1 \\ k \end{matrix} \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & -\lambda & \lambda \\ & & & & -\lambda \end{pmatrix}, \quad k \in \mathbb{N}, \quad \lambda \in \mathbb{R}^+, \tag{3.123}$$

where empty entries represent zeros. The random variables $\{Z_j\}_{j=1}^{k_1+k_2}$ in equations (3.114) and (3.115) can now be considered as the sojourn times in the $k_1 + k_2$ transient states of the Markov process characterised by $(\boldsymbol{\alpha}_1, \boldsymbol{S}_1)$. The bivariate distribution can now be obtained through the reward structure

$$\boldsymbol{R}_1 = \begin{pmatrix} \frac{1}{\lambda_1}\mathbf{1}_{k_1}^\top & \frac{1}{\lambda_2}\mathbf{1}_{k_1}^\top \\ \mathbf{0}_{k_2}^\top & \frac{1}{\lambda_2}\mathbf{1}_{k_2}^\top \end{pmatrix}. \tag{3.124}$$

This most natural representation reflects the structure in definition 3.1, where the random variables are generated through the governing Markov process and the coefficients in equations (3.114) and (3.115) are considered the reward rates. An alternative representation $(\boldsymbol{\alpha}_2, \boldsymbol{S}_2, \boldsymbol{R}_2)$, which emphasizes the structure and coefficients given in equation (3.116), has the same initial distribution $\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1$, but the sub-generator

$$\boldsymbol{S}_2 = \begin{pmatrix} \boldsymbol{B}_{k_1}(\lambda_1) & \lambda_1 \boldsymbol{e}_{k_1:k_1}^\top \boldsymbol{e}_{1:k_2} \\ \mathbf{0}_{k_2\times k_1} & \boldsymbol{B}_{k_2}(\lambda_2) \end{pmatrix}, \tag{3.125}$$

where $\boldsymbol{e}_{k:n}$ is a n-dimensional row vector with zeros in all entries except entry $1 \leq k \leq n$, which is a one. This sub-generator must be paired with the reward structure

$$\boldsymbol{R}_2 = \begin{pmatrix} \mathbf{1}_{k_1}^\top & \frac{\lambda_1}{\lambda_2}\mathbf{1}_{k_1}^\top \\ \mathbf{0}_{k_2}^\top & \mathbf{1}_{k_2}^\top \end{pmatrix} \tag{3.126}$$

in order to yield a proper representation of the bivariate distribution. The second representation is characterised by having a Markov process, where the sojourn times are divided into two compartments in which the sojourn times have different rates (different exponential distributions), and a reward structure that assigns unit reward rates to the first component for states in the first compartment and unit reward rates to the second component for states in the second compartment.

Regardless of which representation is selected, the Laplace transform associated with the distribution is given by

$$\mathcal{L}(u_1, u_2) = \left(\frac{1}{1 + \frac{u_1}{\lambda_1} + \frac{u_2}{\lambda_2}}\right)^{k_1} \left(\frac{1}{1 + \frac{u_2}{\lambda_2}}\right)^{k_2}, \tag{3.127}$$

which is well-defined for all arguments $(u_1, u_2) \in \mathbb{C}^2$ such that the two denominators have positive real parts. Finally, it should be noted that this example actually covers a more general bivariate distribution than the bivariate gamma distribution by McKay, since the scientific literature usually only refers to McKay's distribution when the two rate parameters are identical, cf. page 451 in Bladt and Nielsen (2017). □

The interaction between the reward structure and the rates of the sojourn times leading to the different representations for the above distribution is a simple consequence of the structure used in the construction of the MPH* distributions. This is evident from equation (3.2) as

$$Y_j = \sum_{i=1}^{k} Z_i R_{ij} = \sum_{i=1}^{k} \left(\frac{Z_i}{\lambda_i}\right)(R_{ij}\lambda_i), \quad j \in \{1, ..., n\}. \tag{3.128}$$

McKay originally derived the distribution in order to improve estimation in schemes including sampling from batches, and the distribution has since found application across many engineering and commercial fields.

The concept of sharing some exponential terms can also be used to construct various other multivariate gamma distributions. Two examples of such multivariate gamma distributions are the distribution by Prékopa and Szántai (1978) and the distribution by Cheriyan and Ramabhadran.

Example 3.4

The distribution in this example was first discovered in Cheriyan (1941) and later further generalised in Ramabhadran (1951), but many of the results pertaining to the distribution were derived by David and Fix in 1961, cf. section 48.2.2 of Kotz et al. (2000). The n-dimensional gamma distribution $\boldsymbol{Y} = (Y_1, ..., Y_n)$ in the sense of Cheriyan and Ramabhadran has components on the form:

$$Y_j = \frac{1}{\lambda_j}\left(\sum_{i=1}^{k_0} Z_{0i} + \sum_{m=1}^{k_j} Z_{jm}\right), \quad j \in \{1, ..., n\}, \tag{3.129}$$

for $\{k_j\}_{j=0}^{n} \in \mathbb{N}^{n+1}$, where all the random variables are independent with exponential(1) distributions. The components thus share k_0 exponential terms and have a varying number of additional exponential terms (k_j), which are then all scaled. The marginal distributions of the components are therefore gamma distributed with $Y_j \sim \gamma(\lambda_j, k_0 + k_j)$. The joint density function associated with the distribution is derived from the joint distribution of $n+1$ independent gamma random variables

$$X_j \sim \gamma(1, k_j), \quad j \in \{0, 1, ..., n\}, \tag{3.130}$$

such that the vector components of $\boldsymbol{Y}$ can be described as

$$Y_j = \frac{1}{\lambda_j}\left(X_0 + X_j\right), \quad j \in \{1, ..., n\}, \tag{3.131}$$

which is equivalent to the matrix equation:

$$\boldsymbol{Y} \odot \boldsymbol{\lambda}^\top = \begin{bmatrix} \boldsymbol{1}_n^\top & \boldsymbol{I}_n \end{bmatrix} \boldsymbol{X}, \tag{3.132}$$

where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ and $\boldsymbol{X} = (X_0, ..., X_n)^\top$. The joint density function of $\boldsymbol{X}$ is naturally given as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=0}^{n}\left[\frac{1}{(k_j-1)!}e^{-x_j}x_j^{k_j-1}\right], \quad \boldsymbol{x} \in (\boldsymbol{R}^+)^{n+1}, \tag{3.133}$$

and the multivariate change of variable theorem thus allows for obtaining the joint density function of the random vector $(X_0, Y_1, ..., Y_n)^\top$ as

$$f_{X_0,Y_1,...,Y_n}(x_0, y_1, ..., y_n) = \left[\frac{1}{(k_0-1)!}e^{-x_0}x_0^{k_0-1}\right]\prod_{j=1}^{n}\left[\lambda_j\frac{1}{(k_j-1)!}e^{-(-x_0+\lambda_j y_j)}\left(-x_0+\lambda_j y_j\right)^{k_j-1}\right] \tag{3.134}$$

$$= \frac{1}{(k_0-1)!}x_0^{k_0-1}e^{(n-1)x_0-\sum_{r=1}^{n}\lambda_r y_r}\prod_{j=1}^{n}\left[\lambda_j\frac{1}{(k_j-1)!}\left(\lambda_j y_j - x_0\right)^{k_j-1}\right]. \tag{3.135}$$

To obtain the joint density function of $\boldsymbol{Y}$, the random variable X_0 has to be marginalised(integrated) out of the expression in equation (3.135). This leaves the marginal joint density function

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_0^{\min(\lambda_1 y_1, ..., \lambda_n y_n)} f_{X_0,Y_1,...,Y_n}(x_0, y_1, ..., y_n)dx_0. \tag{3.136}$$

The integral in equation (3.136) is however rather cumbersome to evaluate and leads to a complicated expression in the general case, cf. Kotz et al. (2000). A few improvements are found in Eagleson (1964) and Mardia (1970), which provide expansions of some probability functions in terms of Laguerre polynomials, while Szantai (1986) contains a semi-explicit formula for the joint density function.

The dependence structure arising from the distribution is therefore usually expressed through the covariance structure between the vector components or a joint integral transform. The covariance between two vector components follows easily from equation (3.131) as

$$\mathrm{Cov}(Y_s, Y_t) = \mathrm{Cov}\left(\frac{1}{\lambda_s}(X_0 + X_s), \frac{1}{\lambda_t}(X_0 + X_t)\right) = \frac{1}{\lambda_s \lambda_t}\mathrm{Cov}(X_0, X_0) = \frac{k_0}{\lambda_s \lambda_t}. \tag{3.137}$$

The joint Laplace transform is also established in a straightforward manner using the independence between the components of the random vector $\boldsymbol{X}$:

$$\mathcal{L}(\boldsymbol{u}) = \mathbb{E}\left[e^{-\langle \boldsymbol{Y}, \boldsymbol{u}\rangle}\right] = \mathbb{E}\left[e^{-\sum_{j=1}^{n}(X_0+X_j)\frac{u_j}{\lambda_j}}\right] = \mathbb{E}\left[e^{-X_0\sum_{j=1}^{n}\frac{u_j}{\lambda_j}}\prod_{r=1}^{n} e^{-X_r\frac{u_r}{\lambda_r}}\right] \tag{3.138}$$

$$= \mathbb{E}\left[e^{-X_0\sum_{j=1}^{n}\frac{u_j}{\lambda_j}}\right]\prod_{r=1}^{n}\mathbb{E}\left[e^{-X_r\frac{u_r}{\lambda_r}}\right] \tag{3.139}$$

$$= \left(1 + \sum_{j=1}^{n}\frac{u_j}{\lambda_j}\right)^{-k_0}\prod_{r=1}^{n}\left(1 + \frac{u_r}{\lambda_r}\right)^{-k_r}, \tag{3.140}$$

which is well-defined for all arguments such that all individual factors in (3.140) have positive real parts.

A phase-type representation of the multivariate gamma distribution by Cheriyan and Ramabhadran now becomes evident from the joint Laplace transform (if it was not already apparent from equation (3.131)). The representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ has dimension $k = \sum_{j=0}^{n} k_j$ with the initial distribution $\boldsymbol{\alpha} = (1, \boldsymbol{0}_{k-1})$, the sub-generator $\boldsymbol{S} = \boldsymbol{B}_k(1)$, and the reward structure

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{1}_{k_0}^{\top}\boldsymbol{\lambda} \\ \boldsymbol{1}_{k_1}^{\top}(\boldsymbol{e}_{1:n}\odot\boldsymbol{\lambda}) \\ \boldsymbol{1}_{k_2}^{\top}(\boldsymbol{e}_{2:n}\odot\boldsymbol{\lambda}) \\ \vdots \\ \boldsymbol{1}_{k_{n-1}}^{\top}(\boldsymbol{e}_{(n-1):n}\odot\boldsymbol{\lambda}) \\ \boldsymbol{1}_{k_n}^{\top}(\boldsymbol{e}_{n:n}\odot\boldsymbol{\lambda}) \end{pmatrix}. \tag{3.141}$$

This representation gives a clear overview of the dependence structure in the distribution and allows for analysing the distribution analytically as well as through efficient simulation schemes.

In the bivariate case, the integral in equation (3.136) becomes sufficiently simple that an explicit formula of the joint density function can be obtained. The explicit expression allows for plotting the bivariate function, and Sen Hu et al. (2019) includes some graphics displaying different dependence structures, which the model by Cheriyan and Ramabhadran can describe. □

The multivariate gamma distribution proposed in Prékopa and Szántai (1978) is based on a construction principle similar to that of Cheriyan and Ramabhadran. Prékopa and Szántai extend the ideas introduced by Cheriyan and Ramabhadran by defining a n-dimensional random vector $\boldsymbol{Y}$ as

$$\boldsymbol{Y} \odot \boldsymbol{\lambda}^\top = \boldsymbol{A}\boldsymbol{W}, \tag{3.142}$$

where $\boldsymbol{W}$ is a random vector consisting of $2^n - 1$ independent gamma random variables, and $\boldsymbol{A}$ is a special $n \times (2^n - 1)$ matrix. The matrix $\boldsymbol{A}$ has $2^n - 1$ columns which each represents a unique element in the set $\{0, 1\}^n \setminus \{0\}^n$, so for $n = 4$ the matrix may take the form:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{3.143}$$

Comparing the expression in (3.142) and (3.143) with that in equation (3.132), the differences between the two distributions become apparent. In example 3.4, each vector component includes a common gamma term that is shared among all the vector components and an individual gamma term, which is unique to said vector component. In contrast, the vector components in Prékopa and Szántai's distribution have additional terms. Specifically, the components in Prékopa and Szántai's distribution share a specific gamma term with each possible subset of the other vector components. In the case $n = 4$, the matrix $\boldsymbol{A}$ given in (3.143) shows that Y_1 has the unique term W_1, while it shares W_5 with Y_2, W_6 with Y_3, and so on until W_{15}, which is shared among all four vector components. The distribution by Prékopa and Szántai is therefore the most general distribution that can be generated using only this concept of sharing terms.

Another method of constructing multivariate distributions is through a principle of decomposition. The decomposition principle in question essentially entails expressing an exponential distribution as a mixture. The foremost example is the decomposition of a random variable Y having an exponential(1) distribution as

$$Y = (1 - p)W + JZ, \tag{3.144}$$

where W and Z are independent and identically distributed with an exponential(1) distribution, and J follows a Bernoulli(p) distribution independent of W and Z. Examples of distributions derived from this decomposition method include the bivariate exponential distribution by Marshall and Olkin as well as the bivariate exponential distribution by Raftery.

Example 3.5

The bivariate exponential distribution studied in this example is taken from two 1967 papers by Marshall and Olkin. Their bivariate distribution and its generalisations have been studied extensively throughout the years, and the distributions remain some of the most applied bivariate models in the field of applied probability.

The model developed by Marshall and Olkin is derived from a two-dimensional Poisson process similarly to how the univariate exponential distribution is derived as the distribution governing inter-arrival times in the univariate Poisson process. Marshall and Olkin consider two components that are subject to shocks which incur failures in the components. These shocks arrive according to two independent Poisson processes and the shocks from the first process always effects the first components, while shocks from the second process always effects the second component. In addition, shocks from the first process effect the second component with a certain probability and shocks from the second process can effect the first component. Marshall and Olkin's bivariate distribution is then that of the joint lifetimes (times until failure) of two such components.

An alternative characterisation presented in e.g. Kotz et al. (2000) uses instead three independent Poisson processes. In this presentation, the first process effects only the first component, the second process effects only the second components, and the third process always effects both components. Denote the arrival rates of these processes by λ_1, λ_2, and λ_{12}, respectively, and the first arrival times in these processes by X_1, X_2, and X_{12}, respectively, such that

$$X_1 \sim \mathrm{Exp}(\lambda_1), \quad X_2 \sim \mathrm{Exp}(\lambda_2), \quad \text{and } X_{12} \sim \mathrm{Exp}(\lambda_{12}) \tag{3.145}$$

are mutually independent. The random vector $\boldsymbol{Y} = (Y_1, Y_2)$ defined as

$$Y_j = \min(X_j, X_{12}), \quad j \in \{1, 2\}, \tag{3.146}$$

then has a bivariate exponential distribution in the sense of Marshall and Olkin. The marginal laws are easily determined to be

$$Y_j \sim \mathrm{Exp}(\lambda_j + \lambda_{12}), \quad j \in \{1, 2\}, \tag{3.147}$$

while the correlation between the two vector components is given by

$$\mathrm{Corr}(Y_1, Y_2) = \mathbb{P}(Y_1 = Y_2) = \frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}} \geq 0. \tag{3.148}$$

This result is usually derived from an integral transform, but it can be established in a straightforward way through the decomposition principle. In this case, the decomposition allows for expressing the vector components as the mixtures

$$(\lambda_j + \lambda_{12}) Y_j = \frac{\lambda_j + \lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}} W + J_j Z_j, \quad j \in \{1, 2\} \tag{3.149}$$

where all the random variables are mutually independent except J_1 and J_2, which have a bivariate Bernoulli distribution with joint probability mass function

$$\begin{aligned}
\mathbb{P}(J_1 = 0, J_2 = 0) &= \frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}} = \mathbb{P}(\min(X_1, X_2, X_{12}) = X_{12}), \\
\mathbb{P}(J_1 = 1, J_2 = 0) &= 1 - \frac{\lambda_1 + \lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}} = \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_{12}} = \mathbb{P}(\min(X_1, X_2, X_{12}) = X_2), \\
\mathbb{P}(J_1 = 0, J_2 = 1) &= 1 - \frac{\lambda_2 + \lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_{12}} = \mathbb{P}(\min(X_1, X_2, X_{12}) = X_1), \\
\mathbb{P}(J_1 = 1, J_2 = 1) &= 0,
\end{aligned} \tag{3.150}$$

and W, Z_1, and Z_2 have exponential(1) distributions as prescribed by the construction in equation (3.144).

The decomposition indicates that Y_j is merely an exponential$(\lambda_1+\lambda_2+\lambda_{12})$ random variable when either $\min(X_1,X_2,X_{12})=X_j$ or $\min(X_1,X_2,X_{12})=X_{12}$, and otherwise the vector component has a convolution of an exponential$(\lambda_1+\lambda_2+\lambda_{12})$ distribution and an exponential$(\lambda_j+\lambda_{12})$ distribution. This description explains the probabilities given in (3.150) and subsequently also the mixture probabilities in the decompositions. The decompositions further lead to a simple MPH* representation of the bivariate distribution as $\boldsymbol{Y}\sim$ MPH*$(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{R})$, where

$$\boldsymbol{\alpha}=(1,0,0),\quad \boldsymbol{S}=\begin{pmatrix}-(\lambda_1+\lambda_2+\lambda_{12}) & \lambda_2 & \lambda_1\\ 0 & -(\lambda_1+\lambda_{12}) & 0\\ 0 & 0 & -(\lambda_2+\lambda_{12})\end{pmatrix},\quad \boldsymbol{R}=\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}. \tag{3.151}$$

As mentioned earlier, the correlation in (3.148) becomes readily available using the decomposition. First, let $p_1=\mathbb{P}(J_1=1,J_2=0)$ and $p_2=\mathbb{P}(J_1=0,J_2=1)$ such that

$$\mathbb{E}[Y_1Y_2]=\mathbb{E}\left[\left((1-p_1)\frac{W}{\lambda_1+\lambda_{12}}+J_1\frac{Z_1}{\lambda_1+\lambda_{12}}\right)\left((1-p_2)\frac{W}{\lambda_2+\lambda_{12}}+J_2\frac{Z_2}{\lambda_2+\lambda_{12}}\right)\right]. \tag{3.152}$$

It then follows from simple manipulations and the mutual independence that

$$\mathbb{E}[Y_1Y_2]=\frac{2(1-p_1)(1-p_2)+(1-p_1)p_2+(1-p_2)p_1}{(\lambda_1+\lambda_{12})(\lambda_2+\lambda_{12})}=\frac{2-p_1-p_2}{(\lambda_1+\lambda_{12})(\lambda_2+\lambda_{12})}, \tag{3.153}$$

which yields the correlation

$$\operatorname{Corr}(Y_1,Y_2)=\frac{\mathbb{E}[Y_1Y_2]-\mathbb{E}[Y_1]\mathbb{E}[Y_2]}{\sqrt{\mathbb{V}[Y_1]\mathbb{V}[Y_2]}}=2-p_1-p_2+1=1-p_1-p_2=\mathbb{P}(Y_1=Y_2), \tag{3.154}$$

since $\mathbb{P}(Y_1=Y_2)=\mathbb{P}(J_1=0,J_2=0)=1-p_1-p_2$. The phase-type representation of the distribution given in (3.151) now provides a natural way to derive the Laplace transform associated with the bivariate exponential distribution. The representation suggests that both components earn reward in the first (uppermost) state after which the governing Markov process can transition into three different states (two transient and the absorbing), which determine if either of the components should receive further reward. This structure dictates the joint Laplace transform

$$\mathcal{L}(u_1,u_2)=\frac{1}{1+\frac{u_1+u_2}{\lambda_1+\lambda_2+\lambda_{12}}}\left((1-p_1-p_2)+p_1\frac{1}{1+\frac{u_1}{\lambda_1+\lambda_{12}}}+p_2\frac{1}{1+\frac{u_2}{\lambda_2+\lambda_{12}}}\right), \tag{3.155}$$

where we shall leave out considerations about the domain of the transform. The phase-type representation and the Laplace transform of the bivariate distribution reveal the underlying structure of the distribution and present a simple approach to simulating the bivariate distributions, but some results are more easily established using the Marshall-Olkin copula or the joint survival function. The latter is found as

$$G(y_1,y_2)=\mathbb{P}(Y_1>y_1,Y_2>y_2)=\mathbb{P}(\min(Z_1,Z_{12})>y_1,\min(Z_2,Z_{12})>y_2) \tag{3.156}$$

$$=\mathbb{P}(Z_1>y_1)\mathbb{P}(Z_2>y_2)\mathbb{P}(Z_{12}>\max(y_1,y_2)) \tag{3.157}$$

$$=e^{-(\lambda_1y_1+\lambda_2y_2+\lambda_{12}\max(y_1,y_2))},\quad y_1,y_2\geq 0, \tag{3.158}$$

and the joint density function thus follows trivially as

$$f(y_1,y_2)=\begin{cases}\lambda_2(\lambda_1+\lambda_{12})e^{-((\lambda_1+\lambda_{12})y_1+\lambda_2y_2)}, & 0\leq y_2<y_1\\ \lambda_1(\lambda_2+\lambda_{12})e^{-((\lambda_2+\lambda_{12})y_2+\lambda_1y_1)}, & 0\leq y_1<y_2\\ \lambda_{12}e^{-(\lambda_1+\lambda_2+\lambda_{12})y_1}, & 0\leq y_1=y_2\end{cases}, \tag{3.159}$$

by taking partial derivatives, cf. Kotz et al. (2000). The Marshall-Olkin copula, also known as the generalised Cuadras-Augé copula, derived from the bivariate distribution can be found in e.g. sec. 3.1.1 of Nelsen (2006), and it is usually given with a different set of parameters than those used above. Standard references normally define $a = \lambda_{12}/(\lambda_1 + \lambda_{12})$ and $b = \lambda_{12}/(\lambda_2 + \lambda_{12})$ and express the copula as

$$C(u, v) = \min\left(u^{1-a}v, uv^{1-b}\right) = \begin{cases} u^{1-a}v, & u^a \geq v^b \\ uv^{1-b}, & u^a \leq v^b \end{cases}, \tag{3.160}$$

which shows that complete dependence occurs when $a = b = 1$ ($\lambda_{12} = 1, \lambda_1 = \lambda_2 = 0$) and independence is achieved when $a = b = \lambda_{12} = 0$. □

The copula derived from the bivariate distribution by Marshall and Olkin has been applied in many models across several scientific fields. Nadarajah et al. (2017) gives as examples pricing of CDO (Collateralized Debt Obligation) contracts and modelling of cross-border bank contagion, while referring to Cherubini et al. (2015) for further examples and application.

Another widely used model based on the decomposition principle is the exponential distribution by Raftery.

Example 3.6

The multivariate exponential distribution by Raftery was introduced in 1984 as one of the first multivariate exponential distribution that could model a full range of correlation structures including both symmetric and asymmetric dependence structures. Furthermore, the distribution is easy to simulate from, and it has several properties analogous to the normal distribution. The bivariate model by Raftery is also a physically inspired shock model, but in contrast to the distribution by Marshall and Olkin, Raftery's distribution is absolutely continuous without a singular component.

This example presents the simplest version of the multivariate construction defined by Raftery, i.e. a bivariate distribution which models the lifetimes of two system components. The stochastic representation of the two lifetimes $\boldsymbol{Y} = (Y_1, Y_2)$ is given by the decomposition

$$\lambda_j Y_j = (1 - p)W_j + JZ, \quad j \in \{1, 2\}, \tag{3.161}$$

where Z, W_1, and W_2 are mutually independent with standard exponential distributions, and independent of these J is a Bernoulli random variable with parameter p. The marginal laws of the distribution are exponential with rates λ_1 and λ_2, respectively, and the correlation between the component lifetimes is derived as

$$\text{Corr}(Y_1, Y_2) = 2p - p^2. \tag{3.162}$$

Notice how the correlation cannot be negative in this specific example due to the simplified structure in equation (3.161). The structure is also sufficiently simple that the correlation calculation become straightforward as

$$\text{Cov}(Y_1, Y_2) = \frac{\text{Cov}((1-p)W_1 + JZ, (1-p)W_2 + JZ)}{\lambda_1\lambda_2} = \frac{\mathbb{V}[JZ]}{\lambda_1\lambda_2} = \frac{\mathbb{E}[(JZ)^2] - \mathbb{E}[JZ]^2}{\lambda_1\lambda_2} = \frac{2p - p^2}{\lambda_1\lambda_2}. \tag{3.163}$$

Furthermore, the simplicity of the decomposition is apparent from the phase-type representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$, where

$$\boldsymbol{\alpha} = (1,0,0), \quad \boldsymbol{S} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & p \\ 0 & 0 & -1 \end{pmatrix}, \quad \text{and } \boldsymbol{R} = \begin{pmatrix} \frac{1-p}{\lambda_1} & 0 \\ 0 & \frac{1-p}{\lambda_2} \\ \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \end{pmatrix}, \tag{3.164}$$

which yields the Laplace transform

$$\mathcal{L}(u_1,u_2) = \left(\frac{1}{1+\frac{1-p}{\lambda_1}u_1}\right)\left(\frac{1}{1+\frac{1-p}{\lambda_2}u_2}\right)\left((1-p)+p\frac{1}{1+\frac{u_1}{\lambda_1}+\frac{u_2}{\lambda_2}}\right). \tag{3.165}$$

Again, the domain of the transform is not of interest in this setting, so it shall remain unspecified. The phase-type representation and the Laplace transform constitute two easy methods of defining the bivariate model, but Raftery actually defines the distribution through the joint probability functions in his paper. In particular, the joint survival function of the bivariate distribution is derived by first applying the law of total probability

$$G(y_1,y_2) = \mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 0)\mathbb{P}(J = 0) + \mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 1)\mathbb{P}(J = 1). \tag{3.166}$$

Then the conditional probabilities are determined

$$\begin{aligned} \mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 0) &= \mathbb{P}\left(W_1 > \frac{\lambda_1 y_1}{1-p}, W_2 > \frac{\lambda_2 y_2}{1-p}\right) \\ \mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 1) &= \mathbb{P}\left(W_1 + \frac{Z}{1-p} > \frac{\lambda_1 y_1}{1-p}, W_2 > \frac{\lambda_2 y_2}{1-p} + \frac{Z}{1-p}\right). \end{aligned} \tag{3.167}$$

Recall that the three entering random variables are mutually independent, which means that the upper joint probability in (3.167) factors into a simple product. The lower joint probability is evaluated by conditioning on the common random variable Z such that

$$\begin{aligned} &\mathbb{P}\left(W_1 + \frac{Z}{1-p} > \frac{\lambda_1 y_1}{1-p}, W_2 > \frac{\lambda_2 y_2}{1-p} + \frac{Z}{1-p}\right) \\ &= \int_0^\infty \mathbb{P}\left(W_1 > \frac{\lambda_1 y_1 - z}{1-p}, W_2 > \frac{\lambda_2 y_2 - z}{1-p}\right) e^{-z} dz \qquad (3.168) \\ &= \int_0^\infty \mathbb{P}\left(W_1 > \frac{\lambda_1 y_1 - z}{1-p}\right) \mathbb{P}\left(W_2 > \frac{\lambda_2 y_2 - z}{1-p}\right) e^{-z} dz. \qquad (3.169) \end{aligned}$$

The interval of integration can then be divided into three disjoint intervals:

$$[0,\infty) = [0, \min(\lambda_1 y_1, \lambda_2 y_2)) \cup [\min(\lambda_1 y_1, \lambda_2 y_2), \max(\lambda_1 y_1, \lambda_2 y_2)) \cup [\max(\lambda_1 y_1, \lambda_2 y_2), \infty), \tag{3.170}$$

which leads to two different cases for the integral in equation (3.169). Whenever $\lambda_1 y_1 \geq \lambda_2 y_2 \geq 0$, the integral simplifies to

$$\begin{aligned} \int_0^\infty \mathbb{P}\left(W_1 > \frac{\lambda_1 y_1 - z}{1-p}\right) \mathbb{P}\left(W_2 > \frac{\lambda_2 y_2 - z}{1-p}\right) e^{-z} dz &= \int_0^{\lambda_2 y_2} e^{-\frac{\lambda_1 y_1 - z}{1-p}} e^{-\frac{\lambda_2 y_2 - z}{1-p}} e^{-z} dz \\ &+ \int_{\lambda_2 y_2}^{\lambda_1 y_1} e^{-\frac{\lambda_1 y_1 - z}{1-p}} e^{-z} dz \\ &+ \int_{\lambda_1 y_1}^{\infty} e^{-z} dz. \end{aligned} \tag{3.171}$$

It is simple to evaluate the three integrals on the RHS of equation (3.171), and collecting the results from the various steps gives the conditional probability that

$$\begin{aligned}\mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 1) = &\frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1 + \lambda_2 y_2}{1-p}} \left[e^{\frac{1+p}{1-p}\lambda_2 y_2} - 1 \right] \\ &+ \frac{1-p}{p} e^{-\frac{\lambda_1 y_1}{1-p}} \left[e^{\lambda_1 y_1 \frac{p}{1-p}} - e^{\lambda_2 y_2 \frac{p}{1-p}} \right] \\ &+ e^{-\lambda_1 y_1}.\end{aligned} \tag{3.172}$$

Although tedious, standard calculations produce a considerable simplification to

$$\mathbb{P}(Y_1 > y_1, Y_2 > y_2 | J = 1) = -\frac{1}{p}\frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1 - p\lambda_2 y_2}{1-p}} + \frac{1}{p} e^{-\lambda_1 y_1} - \frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1 + \lambda_2 y_2}{1-p}}. \tag{3.173}$$

Thus, for the case $\lambda_1 y_1 \geq \lambda_2 y_2 \geq 0$, the joint survival function takes the form

$$\begin{aligned}\mathbb{P}(Y_1 > y_1, Y_2 > y_2) = &\left(-\frac{1}{p}\frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1 - p\lambda_2 y_2}{1-p}} + \frac{1}{p} e^{-\lambda_1 y_1} - \frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1 + \lambda_2 y_2}{1-p}} \right) p \\ &+ \left(e^{-\frac{\lambda_1 y_1 + \lambda_2 y_2}{1-p}} \right)(1-p),\end{aligned} \tag{3.174}$$

which can be simplified into

$$\mathbb{P}(Y_1 > y_1, Y_2 > y_2) = e^{-\lambda_1 y_1} + \frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1}{1-p}} \left(e^{-\frac{\lambda_2 y_2}{1-p}} - e^{\frac{p\lambda_2 y_2}{1-p}} \right). \tag{3.175}$$

Due to symmetry, the result can be summarized as follows: Assuming that $p \neq 0$, the joint survivlal function is given by

$$\mathbb{P}(Y_1 > y_1, Y_2 > y_2) = \begin{cases} e^{-\lambda_1 y_1} + \frac{1-p}{1+p} e^{-\frac{\lambda_1 y_1}{1-p}} \left(e^{-\frac{\lambda_2 y_2}{1-p}} - e^{\frac{p\lambda_2 y_2}{1-p}} \right), & \lambda_1 y_1 \geq \lambda_2 y_2 \geq 0 \\ e^{-\lambda_2 y_2} + \frac{1-p}{1+p} e^{-\frac{\lambda_2 y_2}{1-p}} \left(e^{-\frac{\lambda_1 y_1}{1-p}} - e^{\frac{p\lambda_1 y_1}{1-p}} \right), & \lambda_2 y_2 \geq \lambda_1 y_1 \geq 0 \end{cases}. \tag{3.176}$$

The copula associated with Raftery's bivariate distribution is derived in a similar manner to the above. The copula formulation found in Nadarajah et al. (2017) is more explicit than the expression given in Nelsen (2006), and will therefore be preferred here. In accordance with Nadarajah et al. the survival copula can be expressed as

$$C(u_1, u_2) = \begin{cases} u_1 + \frac{1-p}{1+p} u_1^{\frac{1}{1-p}} \left(u_2^{\frac{1}{1-p}} - u_2^{-\frac{p}{1-p}} \right), & u_1 \leq u_2 \\ u_2 + \frac{1-p}{1+p} u_2^{\frac{1}{1-p}} \left(u_1^{\frac{1}{1-p}} - u_1^{-\frac{p}{1-p}} \right), & u_1 \geq u_2 \end{cases}. \tag{3.177}$$

This reaffirms the notion that complete dependence is achieved when $p = 1$ (although the copula is not valid for $p = 1$), while the two components of the random vector are independent when $p = 0$. □

Examples 3.5 and 3.6 illustrate the usefulness of the decomposition principle and how distributions with complex dependence structures can be conveniently defined through their phase-type representations. The previous examples have included distributions derived from the concepts of shared exponential terms and decomposition, but there several distributions, which cannot be established using those principles.

Typical examples of distributions which cannot be described purely using the above techniques are distributions governed by Farlie–Gumbel–Morgenstern copulas. As shown in section 8.3.2 of Bladt and Nielsen (2017), a bivariate random vector whose marginal distributions have rational transforms will have a rational joint transform, which implies that the joint distribution belongs to the class of multivariate matrix-exponential distributions (and often they will also belong to the subclass of multivariate phase-type distributions).

The Farlie–Gumbel–Morgenstern copulas are based on mixtures of order statistics, and the concept of mixing actually gives rise to another method of constructing multivariate phase-type distributions. In chapter 2, it was shown how an exponential distribution could arise from an infinite (geometric) mixture of Erlang distributions, and this principle construction extends to other distributions.

Example 3.7

One example of a distribution constructed using mixing (compounding) is the gamma distribution due to Gaver (1970). The bivariate distribution is a special case of the Kibble-Moran bivariate gamma distribution, cf. page 437 in Kotz et al. (2000), and is derived from a mixing methodology originally developed by Feller (1966), see page 414. The idea is to consider a random variable X with a (negative binomial) $\mathrm{NB}(r,p)$ distribution whose probability generating function is

$$G_*(z) = \sum_{k=0}^{\infty} \mathbb{P}(X=k)z^k = \left(\frac{p}{1-(1-p)z}\right)^r, \quad |z| < \frac{1}{1-p}. \tag{3.178}$$

Define then the random variable $Y \sim \gamma(1, X+r)$ and consider the characteristic function

$$\phi(u) = \mathbb{E}\left[e^{iuY}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{iuY}|X\right]\right] = \mathbb{E}\left[(1-iu)^{-(X+r)}\right] = (1-iu)^{-r}\sum_{k=0}^{\infty}\mathbb{P}(X=k)\,(1-iu)^{-k}. \tag{3.179}$$

The last summation can be recognised as the probability generating function, i.e.

$$\phi(u) = (1-iu)^{-r}\,G_*\left((1-iu)^{-1}\right) = (1-iu)^{-r}\left(\frac{p}{1-(1-p)(1-iu)^{-1}}\right)^r = \left(\frac{p}{p-iu}\right)^r, \tag{3.180}$$

which is the characteristic function of a $\gamma(p,r)$ distribution. The compounding thus leads to a marginal gamma distribution for Y, and this concept extends to higher dimensions. Gaver therefore considers a random vector $\boldsymbol{Y} = (Y_1, ..., Y_n)$ with components,

$$Y_j = X_j + Z_j, \quad j \in \{1,..,n\}, \tag{3.181}$$

where $Z_j \sim \gamma(1,r)$ and $X_j \sim \gamma(1,N)$ are independent. The random variable $N \sim \mathrm{NB}(r,p)$ is independent of the other random variables and is common for all the X_j random variables, which makes it the sole source of covariation in this setting. Another consequence of the construction in equation (3.181) is that all the n variates are identically distributed; it is even symmetrical in all the n variates. The covariance between two different vector components therefore reduces to the variance of the random variable N. To show this, suppose that $j \neq m$, such that

$$\mathrm{Cov}(Y_j, Y_m) = \mathrm{Cov}(X_j + Z_j, X_m + Z_m) = \mathrm{Cov}(X_j, X_m) = \mathbb{E}[X_j X_m] - \mathbb{E}[X_j]\mathbb{E}[X_m]. \tag{3.182}$$

Due to the conditional independence of X_j and X_m given N and the linearity of the expectation operator, the last expression in (3.182) evaluates to

$$\mathbb{E}[X_j X_m] - \mathbb{E}[X_j]\mathbb{E}[X_m] = \mathbb{E}\left[\mathbb{E}[X_j X_m|N]\right] - \mathbb{E}[N]^2 = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \mathbb{V}[N] = \frac{r(1-p)}{p^2}. \tag{3.183}$$

The joint characteristic function associated with the construction in equation (3.181) is obtained using the same arguments as in the univariate case, and the characteristic function of the random vector is thus given by

$$\phi_{\boldsymbol{Y}}(u_1, ..., u_n) = \left(\frac{p}{\prod_{j=1}^{n}(1 - iu_j) - (1-p)} \right)^r . \tag{3.184}$$

Finally, as the negative binomial distribution belongs to the class of discrete phase-type distributions and the gamma distribution is a special instance of a continuous phase-type distribution, the MPH* representation of Gaver's multivariate gamma distribution might be established using theorems 3.7 and 3.5. □

As mentioned in the previous example, Gaver's distribution is an instance of the Kibble-Moran [Kibble (1941), Moran (1967)] distribution in the bivariate case. The general Kibble-Moran bivariate distribution is however also a MPH* distribution arising from mixing, and a representation of the Kibble distribution is found in e.g. example 3.3 of Meisch (2014).

Yet another example of a distribution arising from compounding is the Moran-Downton distribution proposed in Moran (1967) and Downton (1970). Their bivariate exponential distribution is the two-dimensional generalisation of the construction in example 2.2 and bears quite some similarity to the Gaver's distribution. A random vector $\boldsymbol{Y} = (Y_1, Y_2)$ following a Moran-Downton distribution has components on the form $Y_1 \sim \gamma(\lambda_1/(1-p), N)$ and $Y_2 \sim \gamma(\lambda_2/(1-p), N)$, where $N \sim \text{Geo}(1-p)$, cf. example 8.2.8 in Bladt and Nielsen (2017). Kotz et al. (2000) provides a comprehensive analysis and discussion of the distribution in section 2.7 of the book. The section further details how the model can be interpreted as a shock model and provides several derivations of probability functions and the associated transforms as well as numerous properties of some conditional distributions. Moreover, an appropriate phase-type representation can be found in the above-mentioned reference by Bladt and Nielsen. Since the distribution is described so well in other sources, we shall not cover the distribution more in depth here.

The examples and discussions presented in this section underline the flexibility of the multivariate phase-type distributions. As evident from the examples, the phase-type distributions include numerous distributions that have found applications in many scientific and engineering disciplines, and the phase-type distributions arise from a number of different constructions, e.g. compounding and decompositions. The distributions listed in this section do in no way constitute an exhaustive exposition of known multivariate distribution with phase-type representations, but there are still many distributions where potential phase-type representations have not been established. Consequently, there is active research on identifying representations of known distributions. In the next chapter, we present a paper on analysis and phase-type representation of the multivariate gamma distribution by Dussauchoy and Berland.

3.5 Related multivariate distributions

In this short section, we present a discussion of multivariate distributions related to the multivariate phase-type distributions. Many of the generalisations of the univariate phase-type distributions described in section 2.5 also apply to the multivariate phase-type distributions. In particular, there exist the more general MME* and MVME classes of multivariate matrix-exponential distributions along with their bilateral extensions, the so-called MBME* and MVBME. The differences between the classes of matrix-exponential distributions are completely analogous to the differences between the different types of multivariate phase-type distributions discussed in the first section of the chapter. There are of course also bilateral variants of the multivariate phase-type distributions, and the resulting classes are termed the MBPH* and MVBPH classes.

The inhomogeneous phase-type distributions have also been generalised to a multivariate setting. For example, Albrecher et al. (2020) presents a definition of multivariate IPH distributions along with several examples of distributions included in the resulting class. Furthermore, the authors provide an EM algorithm for parameter estimation in certain bivariate distributions and apply the method to fit various models to different data sets.

These distinctions are not of purely academic and theoretical interest as some very useful distributions belong to the classes of multivariate matrix-exponential distributions without being a multivariate phase-type distribution. One particularly interesting example is the real noncentral Wishart distribution as given in example 8.5.6 of Bladt and Nielsen (2017), which is reproduced below. The Wishart distribution is an important distribution in multivariate statistics, because it is used in the estimation of covariance matrices and is the conjugate prior of the inverse covariance matrix of a multivariate normal random vector.

Example 3.8

Consider k mutually independent random vectors of dimension m, $\boldsymbol{Y}_j = (Y_{j1}, ..., Y_{jm})^\top$ for $j \in \{1, ..., k\}$, such that the random vectors are normally distributed with possibly different mean vectors and a common covariance matrix, i.e. $\boldsymbol{Y}_j \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ for all appropriate values of j. These random vectors constitute the columns of a random matrix $\boldsymbol{Y}$ whose j'th column is $\boldsymbol{Y}_j$, and this allows for defining a new random matrix as $\boldsymbol{W} = \boldsymbol{Y}\boldsymbol{Y}^\top$. Similarly, a matrix $\boldsymbol{\mu}$ can be constructed, where the j'th column of $\boldsymbol{\mu}$ corresponds to $\boldsymbol{\mu}_j$. If $\boldsymbol{\mu}\boldsymbol{\mu}^\top \neq \boldsymbol{0}_{m\times m}$, the random matrix $\boldsymbol{W}$ has a noncentral Wishart distribution $\mathcal{W}_k(\boldsymbol{\mu}\boldsymbol{\mu}^\top, \boldsymbol{\Sigma})$, cf. James (1955). The moment-generating function associated with $\boldsymbol{W}$ is then given as

$$M(\boldsymbol{\theta}) = \det\left(\boldsymbol{I}_m - 2\boldsymbol{\theta}\boldsymbol{\Sigma}\right)^{-\frac{k}{2}}, \tag{3.185}$$

which Bladt and Nielsen show is a rational transform for even values of k, and they thereby conclude that the distribution is of the MVBME type. The moment generating function also reveals an interesting link to the multivariate exponential distribution by Krishnamoorthy and Parthasarathy, whose moment generating function emerges when $k = 2$ and the off-diagonal elements of $\boldsymbol{\theta}$ are zero. The distribution by Krishnamoorthy and Parthasarathy is of particular interest to us as this is a distribution with a rational transform that does not admit an MME* representation of the same dimension as the degree of the distribution. The distribution is therefore a natural starting point in the search for a distribution belonging to the MVME class without an MME* representation. □

This discussion concludes chapter 3, in which we have presented a review and an exposition of the most important properties and examples of multivariate phase-type distributions. The next chapter deals with the multivariate gamma distribution constructed by Dussauchoy and Berland. The research into this distribution was originally motivated by the MME*-MVME conjecture referenced in example 3.8, but our work instead led to a MPH* representation of the distribution in combination with a series of other results. The research utilises several of the theorems from this chapter and shows how complex dependence structures the MPH* distributions accommodate.

CHAPTER 4

Paper

Necessary and sufficient conditions for the existence of the multivariate gamma distribution by Dussauchoy and Berland

Nicolai Siim Larsen and Bo Friis Nielsen

ABSTRACT

This paper presents some new results concerning the multivariate gamma distribution by Dussauchoy and Berland. The two first results are a decomposition of the joint characteristic function and a necessary condition for the existence of the distribution expressed as constraints on the shape parameters. These results imply that the distribution can be represented as a convolution of gamma mixtures, which belongs to the class of multivariate matrix-exponential distributions when the shape parameters take integer values. Further analysis of the decomposition leads to an additional necessary condition for the existence of the distribution, which improves the former result to conclude that the distribution can only exist as a multivariate phase-type distribution when the shape parameters take integer values. The phase-type representation of the distribution is then presented along with simulation results and a brief discussion about the distribution for non-integer valued shape parameters

1. Introduction

The multivariate normal distribution is presumingly the most studied and applied model within the field of multivariate analysis and statistics. The distribution arises naturally in many areas of applications and it has numerous properties that are useful in statistical and stochastic modeling. There are however applications, where data suggests that the multivariate normal distribution is not an appropriate model. These cases have motivated research into a variety of models such as multivariate gamma distributions, which have been applied extensively in actuarial mathematics and engineering sciences.

Several methodologies for constructing multivariate gamma distributions have been proposed throughout time leading to constructions with different properties and dependence structures. The property of infinite divisibility has been studied for a wide range of these constructions in order to establish necessary and sufficient conditions for the existence of the distributions. Vere-Jones (1967) and Griffiths (1969a, 1969b) resolve the matter for a broad class of bivariate gamma distributions, whereas characterization of infinitely divisible multivariate gamma distribution remains an active research area. Recent research conducted on the subject includes Griffiths (1970, 1984), Bernardoff (2006, 2016, 2018) and Perez-Abreu et al. (2014). Barndorff-Nielsen et al. (2006) presents research on infinite divisibility of general multivariate distributions and F. W. Steutel has produced numerous results on infinite divisibility, several of which can be found in Steutel et al. (2004).

Dussauchoy and Berland (1972) introduces a bivariate gamma distribution with a decomposition property also found in the bivariate normal distribution based on results due to Griffiths (1969a). The distribution has received only limited attention in scientific literature, but it appears in various compendia such as Kotz et al. (2000) and Balakrishnan and Lai (2009). The paper has also been included curisorily in Anderson (1979) and Saboor et al. (2010), while Bernardoff (2016, 2018) has developed theory on Laplace copulas associated with general bivariate gamma distributions including the distribution by Dussauchoy and Berland.

The distribution was first proposed as a statistical model of electrical micro-discharges studied under industrial vacuum, see Dussauchoy and Berland (1973), but has since been applied in financial risk management and reliability theory. Nadarajah (2005) applies the bivariate distribution as a stress-strenght model and gives several examples of applications within different engineering sciences. Combes et al. (2008) develops a simulation and planning program assisting hospitals with managing operation theatres based partially on the bivariate distribution.

Dussauchoy and Berland (1975) extends the distribution to a multivariate setting based on the earlier mentioned decomposition concept. The general multivariate distribution is considerably more complex than the bivariate model since the distribution does not in general admit a density function and the abovementioned results on infinite divisibility do not apply to this construction. Consequently, there are no available sufficient or necessary conditions for the existence of the distribution. However, the distribution is derived recursively and the construction allows for efficient numerical simulation using the conditional distribution approach as described in Johnson (1987). One application of the multivariate distribution is found in Jiang (2018), who derives formulas for loss aggregations and capital allocations in portfolios with dependent investments, where the investment losses are modelled by gamma distributions, e.g. that of Dussauchoy and Berland.

The contributions of this paper are constituted by two new findings; necessary conditions for the existence of the multivariate distribution, which are sufficient conditions when the shape parameters take integer values, and a phase-type representation for said distribution, in the sense of Kulkarni (1989). We hope that this improved understanding of the multivariate distribution can inspire further studies and applications of the distribution.

The remaining contents of this paper are structured as follows: We shall proceed with a short introduction to the multivariate gamma distribution defined by Dussauchoy and Berland. In section 3, we briefly review some preliminary results before presenting the decomposition of the joint characteristic function. Section 4 is concerned with necessary and sufficient conditions for the existence of the distribution when the shape parameters take integer values, while sections 5 and 6 deal with divisibility and numerical simulation of the distribution, respectively. The last section contains a discussion on the results and our concluding remarks.

2. The multivariate gamma distribution by Dussauchoy and Berland

Dussauchoy and Berland (1972) constructs a bivariate gamma distribution, which in some sense generalizes the bivariate gamma distribution by McKay (1934). In the simplest case, as a bivariate exponential distribution, the distribution is a special case of the bivariate exponential distribution by Marshall and Olkin (1967a, 1967b). Dussauchoy and Berland introduce their bivariate gamma distribution through the joint characteristic function $\phi : \mathbb{R} \to \mathbb{C}$ defined as

$$\phi(u_1, u_2) = \left(\frac{1 - i\frac{\beta_{12} u_2}{\alpha_1}}{1 - i\frac{u_1 + \beta_{12} u_2}{\alpha_1}} \right)^{\ell_1} \left(\frac{1}{1 - i\frac{u_2}{\alpha_2}} \right)^{\ell_2}, \quad u_1, u_2 \in \mathbb{R}, \tag{4.1}$$

where the rate parameters (α_1, α_2) are non-negative real numbers subject to $\alpha_1 \geq \beta_{12}\alpha_2 \geq 0$, and the shape parameters (ℓ_1, ℓ_2) are non-negative real numbers subject to the condition $\ell_2 \geq \ell_1$. In their paper, Dussauchoy and Berland prove that the bivariate distribution is infinitely divisible and derive several properties of the bivariate distribution. For a random vector $\boldsymbol{Z} = (Z_1, Z_2)$ having the characteristic function in (4.1), they establish that the distribution has marginal gamma laws, specifically $Z_i \sim \gamma(\alpha_i, \ell_i)$, and they calculate the correlation between the elements of the vector as

$$\rho_{12} = \text{Corr}(Z_1, Z_2) = \beta_{12} \frac{\alpha_2}{\alpha_1} \sqrt{\frac{\ell_1}{\ell_2}}. \tag{4.2}$$

They also present some results on conditional distributions related to the bivariate distribution along with additional analysis of the distribution. For instance, they derive the joint probability density function of the bivariate distribution, which in general has support on the area defined by $\{(x, y) \in \mathbb{R}^2 : y \geq \beta_{12} x \geq 0\}$.

The bivariate distribution also has one particular decomposition property which is fundamental for the construction of the multivariate distribution examined in the coming section, namely that the random variable $X_{12} := Z_2 - \beta_{12} Z_1$ is independent of Z_1. Notice that the random variable X_{12} is non-negative, which follows immediately from the support of the distribution. Consequently, we can decompose the second component of the distribution into a sum of two non-negative independent random variables as $Z_2 = \beta_{12} Z_1 + X_{12}$. We can thus write the bivariate random vector on the form

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = Z_1 \begin{bmatrix} 1 \\ \beta_{12} \end{bmatrix} + X_{12} \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{4.3}$$

By rearranging the factors in (4.1) as

$$\begin{aligned} \phi(u_1, u_2) &= \left(\frac{1}{1 - i\frac{u_1 + \beta_{12} u_2}{\alpha_1}} \right)^{\ell_1} \left(\frac{1}{1 - i\frac{u_2}{\alpha_2}} \right)^{\ell_2 - \ell_1} \left(\frac{1 - i\frac{\beta_{12} u_2}{\alpha_1}}{1 - i\frac{u_2}{\alpha_2}} \right)^{\ell_1} \\ &= \left(\frac{1}{1 - i\frac{u_1 + \beta_{12} u_2}{\alpha_1}} \right)^{\ell_1} \left(\frac{1}{1 - i\frac{u_2}{\alpha_2}} \right)^{\ell_2 - \ell_1} \left(\beta_{12} \frac{\alpha_2}{\alpha_1} + \left(1 - \beta_{12} \frac{\alpha_2}{\alpha_1} \right) \frac{1}{1 - i\frac{u_2}{\alpha_2}} \right)^{\ell_1}, \end{aligned} \tag{4.4}$$

we identify the first factor of (4.4) as the characteristic function associated with the first term of (4.3) and the product of the two latter factors in (4.4) as the characteristic function associated with the second term of (4.3). When the shape parameters of the bivariate distribution take integer values, we can represent X_{12} as a sum in terms of independent random variables, and we can then expand the vector representation in (4.3) to

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \sum_{k=1}^{\ell_1} \frac{U_k}{\alpha_1} \begin{bmatrix} 1 \\ \beta_{12} \end{bmatrix} + \sum_{k=1}^{\ell_2 - \ell_1} \frac{V_k}{\alpha_2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \sum_{k=1}^{\ell_1} I_k \frac{W_k}{\alpha_2} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{4.5}$$

where all the random variables are independent with $U_k, V_k, W_k \sim \exp(1)$ and $I_k \sim \text{Ber}\left(1 - \beta_{12}\frac{\alpha_2}{\alpha_1}\right)$ for all values of k, and we note that the condition $\alpha_1 \geq \beta_{12}\alpha_2 \geq 0$ ensures that the parameter of the Bernoulli distribution is restricted to the interval $[0, 1]$. Since the random vector consists of sums and mixtures of linearly scaled exponential distributions, the random vector belongs to the class of multivariate phase-type distributions (in the sense of Kulkarni), and Bladt and Nielsen (2017) gives a phase-type representation of the bivariate random vector using the structure given in (4.5).

The expression in (4.5) suggests that the random variable X_{12} can be considered as a mixture of gamma (Erlang) distributions with binomial probabilities. Indeed, we have that

$$X_{12} = \sum_{k=1}^{\ell_2 - \ell_1} \frac{V_k}{\alpha_2} + \sum_{k=1}^{\ell_1} I_k \frac{W_k}{\alpha_2} \stackrel{d}{=} \sum_{k=1}^{\ell_2 - \ell_1} \frac{V_k}{\alpha_2} + \sum_{k=1}^{N} \frac{W_k}{\alpha_2} \stackrel{d}{=} \sum_{k=1}^{\ell_2 - \ell_1 + N} \frac{V_k}{\alpha_2}, \tag{4.6}$$

where $N \sim \text{Bin}\left(\ell_1, 1 - \beta_{12}\frac{\alpha_2}{\alpha_1}\right)$, i.e. a binomial distribution with ℓ_1 trials and success probability parameter $1 - \beta_{12}\frac{\alpha_2}{\alpha_1}$. This form allows us to derive the distribution function and subsequently the probability density function of X_{12}, and simple calculations show that the probability density function is simply a weighted sum of gamma densities. The representations of the distribution with random variables in (4.5) and (4.6) are not valid when the shape parameters take non-integer values, but they provide an alternative method to show that the bivariate disitribution is infinitely divisible. Dussauchoy and Berland apply the canonical representation theorem by Kolmogorov, see e.g. theorem 5.5.3. in Lukacs (1960), to show the infinite divisibility. The property also follows directly from a result by Goldie (1967) stating that mixtures of exponential distributions are infinitely divisible. Since infinite divisibility is preserved under convolution of independent distributions, the random vector in (4.5) is indeed infinitely divisible.

Dussauchoy and Berland (1975) extends the bivariate construction to a multivariate setting using the decomposition properties presented on the previous pages. The underlying concept is inspired by a decomposition property of characteristic functions, which is especially applicable to multivariate normal distributions. The property applies to a random column vector $\boldsymbol{W}$ composed (concatenated) of two random vectors $(\boldsymbol{U};\boldsymbol{V})$ and allows for factoring the characteristic function of $\boldsymbol{W}$ if there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{V}-\boldsymbol{BU}$ is independent of $\boldsymbol{U}$, in which case the characteristic function of $\boldsymbol{W}$ can be constructed from the characteristic functions of $\boldsymbol{U}$ and $\boldsymbol{V}$. In particular, the characteristic function of $\boldsymbol{W}$ is given by

$$\varphi_{\boldsymbol{W}}(\boldsymbol{u},\boldsymbol{v})=\varphi_{\boldsymbol{V}}(\boldsymbol{v})\frac{\varphi_{\boldsymbol{U}}(\boldsymbol{u}+\boldsymbol{B}^\top\boldsymbol{v})}{\varphi_{\boldsymbol{U}}(\boldsymbol{B}^\top\boldsymbol{v})}. \tag{4.7}$$

A result due to Cramer (1963) gives the matrix $\boldsymbol{B}$ in terms of the covariance matrix of $\boldsymbol{W}$ when the random vector is distributed according to a multivariate normal distribution under mild regularity conditions. This leads to theorem 2 of Dussauchoy and Berland (1975), which states that for an n-dimensional normal random vector $\boldsymbol{W}$ with covariance matrix $\boldsymbol{A}=\{a_{ij}\}_{i,j\in\{1,\ldots,n\}}$ having only strictly positive entries in the diagonal, the characteristic function of $\boldsymbol{W}$ takes the form

$$\varphi_{\boldsymbol{W}}(\boldsymbol{w})=\prod_{j=1}^{n}\frac{\varphi_j\left(w_j+\frac{\sum_{k=j+1}^{n}a_{jk}w_k}{a_{jj}}\right)}{\varphi_j\left(\frac{\sum_{k=j+1}^{n}a_{jk}w_k}{a_{jj}}\right)},\quad \boldsymbol{w}=(w_1,...,w_n)\in\mathbb{R}^n, \tag{4.8}$$

where φ_j refers to the characteristic function of the j'th component. Dussauchoy and Berland provide a proof by successively applying the result in (4.7) and thereby constructing the distribution recursively.

Based on this notion of constructing a distribution recursively, they construct an n-dimensional multivariate gamma distribution and obtain a joint characteristic function $\varphi:\mathbb{R}^n\to\mathbb{C}$ on the form

$$\varphi(\boldsymbol{u})=\prod_{j=1}^{n}\frac{\varphi_j\left(u_j+\sum_{k=j+1}^{n}\beta_{jk}u_k\right)}{\varphi_j\left(\sum_{k=j+1}^{n}\beta_{jk}u_k\right)}=\prod_{j=1}^{n}\frac{\left(1-i\frac{\sum_{k=j+1}^{n}\beta_{jk}u_k}{\alpha_j}\right)^{\ell_j}}{\left(1-i\frac{u_j+\sum_{k=j+1}^{n}\beta_{jk}u_k}{\alpha_j}\right)^{\ell_j}},\quad \boldsymbol{u}=\{u_m\}_{m=1}^{n}, \tag{4.9}$$

where the rate parameters $\{\alpha_m\}_{m=1}^n$ are non-negative real numbers subject to the inequality constraints $\alpha_j\geq\beta_{jk}\alpha_k\geq 0,\ \forall(j,k):1\leq j<k\leq n$, and the shape parameters $\{\ell_m\}_{m=1}^n$ are non-negative real numbers subject to the conditions $\ell_k\geq\ell_j,\forall(j,k):1\leq j<k\leq n$. Furthermore, the functions φ_j still refer to the characteristic functions of the marginal distributions. The construction begins by obtaining a bivariate gamma distribution (Z_{n-1},Z_n) and proceeds by adding variables successively until the complete random vector $\boldsymbol{Z}=(Z_1,...,Z_n)$ is obtained. The property in (4.7) is used repeatedly to establish the characteristic function of the resulting distribution each time an additional variable is included. This approach is exemplified in the following example.

Example 1 - Construction of the three-dimensional distribution

Consider the bivariate gamma distribution $\boldsymbol{V}=(Z_2,Z_3)^\top$ with the characteristic function $\varphi_{\boldsymbol{V}}$ and the univariate gamma distribution $\boldsymbol{U}=Z_1$ with the characteristic function $\varphi_{\boldsymbol{U}}$. Assuming that the three-dimensional distribution $\boldsymbol{W}=(Z_1,Z_2,Z_3)^\top$ exists and that there exists a matrix $\boldsymbol{B}=(\beta_{12},\beta_{13})^\top$ such that $\boldsymbol{V}-\boldsymbol{BU}$ is independent of $\boldsymbol{U}$, the characteristic function of $\boldsymbol{W}$ is given by (4.7):

$$\varphi_{\boldsymbol{W}}(w_1,w_2,w_3)=\varphi_{\boldsymbol{V}}(w_2,w_3)\frac{\varphi_{\boldsymbol{U}}\left(w_1+\beta_{12}w_2+\beta_{13}w_3\right)}{\varphi_{\boldsymbol{U}}(\beta_{12}w_2+\beta_{13}w_3)}, \tag{4.10}$$

which coincides with the expression in (4.9) for a three dimensional distribution. Dussauchoy and Berland write that the second member, which we interpret as the second factor, of (4.10) is effectively a characteristic function, which would be incorrect. If instead the function in (4.10) is recast as

$$\varphi_{\boldsymbol{W}}(w_1, w_2, w_3) = \varphi_{\boldsymbol{U}}\left(w_1 + \beta_{12}w_2 + \beta_{13}w_3\right) \frac{\varphi_{\boldsymbol{V}}(w_2, w_3)}{\varphi_{\boldsymbol{U}}(\beta_{12}w_2 + \beta_{13}w_3)}, \tag{4.11}$$

the structure reflects the orthogonal decomposition principle applied in the construction. Noting that the second factor of (4.11) corresponds to the characteristic function of $\boldsymbol{V} - \boldsymbol{B}\boldsymbol{U}$, this version emphasizes that

$$\boldsymbol{W} = \begin{bmatrix} 1 \\ \beta_{12} \\ \beta_{13} \end{bmatrix} \boldsymbol{U} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} (\boldsymbol{V} - \boldsymbol{B}\boldsymbol{U})\,. \tag{4.12}$$

Defining the random variables $X_{12} = Z_2 - \beta_{12}Z_1$ and $X_{13} = Z_3 - \beta_{13}Z_1$ allows for writing (4.12) in terms of random variables:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \beta_{12} \\ \beta_{13} \end{bmatrix} Z_1 + \begin{bmatrix} 0 \\ X_{12} \\ X_{13} \end{bmatrix}, \tag{4.13}$$

where the first factor of (4.11) can be associated with the first term of (4.12)/(4.13) and the second factor of (4.11) with the second term of (4.12)/(4.13). Finally, it is straightforward to verify the assumption that $\boldsymbol{V} - \boldsymbol{B}\boldsymbol{U}$ is independent of $\boldsymbol{U}$ in this construction. □

Dussauchoy and Berland do not address the issue of existence. The parameter constraints given by Dussauchoy and Berland are necessary and sufficient to ensure the existence of the univariate and bivariate marginal distributions, but they are not sufficient to guarantee the existence of the joint distribution. The following example illustrates a problem occuring in some constructions.

Example 2 - Insufficient conditions

Coonsider a supposed three-dimensional distribution with shape parameters $(\ell_1, \ell_2, \ell_3) = (2, 3, 4)$, rate parameters $(\alpha_1, \alpha_2, \alpha_3) = (6, 2, 1)$, and $(\beta_{12}, \beta_{13}, \beta_{23}) = (2, 1, 1)$. Then the alleged characteristic function, say φ^*, evaluated along the path $\gamma(u) = ([\beta_{12}\beta_{23} - \beta_{13}]\,u, -\beta_{23}u, u) = (u, -u, u)$ is

$$\varphi^*\left(\gamma(u)\right) = \frac{\left(1 + i\frac{u}{6}\right)^2 \left(1 - i\frac{u}{2}\right)^3}{\left(1 - iu\right)^4}. \tag{4.14}$$

Since the leading coefficient of the numerator polynomial is greater than that of the denominator polynomial, the modulus of the characteristic function diverges toward infinity when u tends to infinity. A characteristic function is however uniformly bounded by one and the fuction φ^* is therefore not a characteristic function, which implies that the conditions set fourth by Dussauchoy and Berland are insufficient to ensure the existence of the joint distribution. □

Dussauchoy and Berland derive numerous properties of their multivariate distribution. As mentioned earlier, the univariate marginal distributions are gamma distributied similarly to the bivariate case, i.e. the k'th element of a random vector $\boldsymbol{Z} = (Z_1, ..., Z_n)$ having the characteristic function in (4.9) follows a gamma distribution with rate parameter α_k and shape parameter ℓ_k. Any pair of components of the random vector $\boldsymbol{Z}$ has a bivariate gamma distribution as described in the previous paragraphs. Consequently, higher order components can be decomposed into sums of independent random variables including lower order components. In particular, the multivariate distribution allows for decompositions of the type

$$Z_k = \beta_{jk} Z_j + X_{jk}, \quad \forall (j,k) : 1 \leq j < k \leq n, \tag{4.15}$$

where $X_{jk} := Z_k - \beta_{jk} Z_j$ is independent of Z_j. There is currently no closed form expression for the joint density function of the multivariate distribution, but Dussauchoy and Berland device a method to obtain the joint density function recursively in their 1975 paper. Therefore, analysis of the distribution is primarily concerned with the characteristic function and the decompositions. Furthermore, the rate parameters of the distribution can be assumed to be one without loss of generality. Letting $W_k = \alpha_k Z_k \sim \gamma(1, \ell_k)$ for $k \in \{1, ..., n\}$, the distribution of $(W_1, ..., W_n)$ has gamma distributed marginals that admit the decompositions

$$W_k = q_{jk} W_j + Y_{jk}, \quad \forall (j,k) : 1 \leq j < k \leq n, \tag{4.16}$$

where $q_{jk} = \alpha_k \beta_{jk} / \alpha_j$ and $Y_{jk} = \alpha_k X_{jk}$. Since the random variable X_{ij} is a difference of gamma distributed random variables, cf. (4.15), and the scaled version Y_{jk} too is a difference of gamma distributed random variables, the distribution of $(W_1, ..., W_n)$ is also a multivariate gamma distribution in the sense of Dussauchoy and Berland. Notice also that the parameter constraints of the distribution simplify to $1 \geq q_{jk} \geq 0$.

3. Decomposition and representation

In this section, we present our first result showing that the characteristic function admits a decomposition, which under simple conditions represents a mixture of gamma distributions. We proceed to introduce some definitions before presenting the decomposition.

Consider the random vector $\boldsymbol{Z} = (Z_1, ..., Z_n)$ with the characteristic function

$$\varphi(\boldsymbol{u}) = \prod_{j=1}^{n} \frac{\left(1 - i\left(\sum_{k=j+1}^{n} q_{jk} u_k\right)\right)^{\ell_j}}{\left(1 - i\left(u_j + \sum_{k=j+1}^{n} q_{jk} u_k\right)\right)^{\ell_j}}, \quad \boldsymbol{u} = \{u_m\}_{m=1}^{n}, \tag{4.17}$$

i.e. it is assumed that all the rate parameters take the value one. To simplify notitation, the following auxiliary function are defined.

Definition 1 - Auxiliary functions
For $j \in \{1, ..., n\}$, define $A_j^*, A_j : \mathbb{R}^n \to \mathbb{C}$ as

$$A_j^*(\boldsymbol{u}) = \sum_{k=j+1}^{n} q_{jk} u_k, \tag{4.18}$$

$$A_j(\boldsymbol{u}) = A_j^*(\boldsymbol{u}) + u_j, \tag{4.19}$$

where $\boldsymbol{u} = \{u_k\}_{k=1}^{n}$.

Additionally, let the function $g : \mathbb{R} \to \mathbb{C}$ be defined as

$$g(u;\ell) = (1 - iu)^{-\ell}, \tag{4.20}$$

which means that for non-negative values of ℓ, the function is a characteristic function of a gamma distributed random variable.

Based on these functions, define for $j \in \{1, ..., n\}$ the functions $\psi_j, \psi_j^{(s)} : \mathbb{R}^n \to \mathbb{C}$ as

$$\psi_j(\boldsymbol{u}) = \prod_{m=j+1}^{n} g\left(A_m(\boldsymbol{u}); 1\right), \tag{4.21}$$

$$\psi_j^{(s)}(\boldsymbol{u}) = \prod_{\substack{m=j+1 \\ m \neq s}}^{n} g\left(A_m(\boldsymbol{u}); 1\right), \quad s \in \{j+1, ..., n\}. \tag{4.22}$$

The two latter functions are featured in the decomposition result. □

Following this definition, the first theorem of the paper can be formulated.

Theorem 1 - Decomposition

The characteristic function of the multivariate gamma distribution by Dussauchoy and Berland can be written as

$$\varphi(\boldsymbol{u}) = \prod_{j=1}^{n} \left[g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r \right) \left(w_0^{(j)} \psi_j(\boldsymbol{u}) + \sum_{s=j+1}^{n} w_{s-j}^{(j)} \psi_j^{(s)}(\boldsymbol{u}) \right)^{\ell_j} \right], \tag{4.23}$$

where the weights $\{w_k^{(j)}\}$ are found as solutions to linear systems of equations.

Proof. Using the auxiliary functions, the characteristic function in (4.17) can be expressed as

$$\varphi(\boldsymbol{u}) = \prod_{j=1}^{n} \frac{g\left(A_j(\boldsymbol{u}); \ell_j\right)}{g\left(A_j^*(\boldsymbol{u}); \ell_j\right)}. \tag{4.24}$$

Basic arithmetic yields

$$\varphi(\boldsymbol{u}) = \prod_{j=1}^{n} \left[\frac{g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r\right)}{g\left(A_j^*(\boldsymbol{u}); \ell_j\right)} \prod_{m=1}^{j-1} g\left(A_j(\boldsymbol{u}); \ell_m\right) \right] \tag{4.25}$$

$$= \left[\prod_{j=1}^{n} \frac{g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r\right)}{g\left(A_j^*(\boldsymbol{u}); \ell_j\right)} \right] \left[\prod_{j=1}^{n} \prod_{m=1}^{j-1} g\left(A_j(\boldsymbol{u}); \ell_m\right) \right] \tag{4.26}$$

$$= \left[\prod_{j=1}^{n} \frac{g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r\right)}{g\left(A_j^*(\boldsymbol{u}); \ell_j\right)} \right] \left[\prod_{j=1}^{n} \prod_{m=j+1}^{n} g\left(A_m(\boldsymbol{u}); \ell_j\right) \right] \tag{4.27}$$

$$= \prod_{j=1}^{n} \left[g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r \right) \frac{\prod_{m=j+1}^{n} g\left(A_m(\boldsymbol{u}); \ell_j\right)}{g\left(A_j^*(\boldsymbol{u}); \ell_j\right)} \right] \tag{4.28}$$

$$= \prod_{j=1}^{n} \left[g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r \right) \left(\frac{\psi_j(\boldsymbol{u})}{g\left(A_j^*(\boldsymbol{u}); 1\right)} \right)^{\ell_j} \right]. \tag{4.29}$$

The main idea behind the decomposition result is to rewrite the fraction in each factor of (4.29). Specifically, the denominator can be expressed as a linear combination of the functions in the numerator, i.e. for each $j \in \{1, ..., n\}$

$$\frac{1}{g\left(A_j^*(\boldsymbol{u}); 1\right)} = w_0^{(j)} + \sum_{m=j+1}^{n} \frac{w_{m-j}^{(j)}}{g\left(A_m(\boldsymbol{u}); 1\right)}. \tag{4.30}$$

The weights in the linear combination are found as the solution to a linear system of equations:

$$1 = w_0^{(j)} + \sum_{m=j+1}^{n} w_{m-j}^{(j)}, \tag{4.31}$$

$$q_{jk} u_k = \left(w_{k-j}^{(j)} + \sum_{m=j+1}^{k-1} w_{m-j}^{(j)} q_{mk} \right) u_k, \quad \forall k \in \{j+1, ..., n\}. \tag{4.32}$$

The system can be solved either resurively or directly using standard algebra. The system can also be represented in terms of a matrix equation

$$\boldsymbol{q}_j = \boldsymbol{A}_j \boldsymbol{w}_j, \tag{4.33}$$

which is shorthand notation for

$$\begin{bmatrix} q_{j,j+1} \\ q_{j,j+2} \\ q_{j,j+3} \\ \vdots \\ q_{j,n-1} \\ q_{jn} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q_{j+1,j+2} & 1 & 0 & \cdots & 0 & 0 & 0 \\ q_{j+1,j+3} & q_{j+2,j+3} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ q_{j+1,n-1} & q_{j+2,n-1} & q_{j+3,n-1} & \cdots & 1 & 0 & 0 \\ q_{j+1,n} & q_{j+2,n} & q_{j+3,n} & \cdots & q_{n-1,n} & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w_1^{(j)} \\ w_2^{(j)} \\ w_3^{(j)} \\ \vdots \\ w_{n-j+1}^{(j)} \\ w_{n-j}^{(j)} \\ w_0^{(j)} \end{bmatrix}. \tag{4.34}$$

The matrix $\boldsymbol{A}_j$ is a triangular matrix with ones in the diagonal, which implies that the determinant of the matrix is one. Consequently, it follows directly from Cramer's rule that the weights are given in terms of determinants of the matrix with certain columns replaced with the LHS column vector. Another feature of the matrix is that the matrices $\boldsymbol{A}_m$ for $j < m < n$ can be found as principal submatrices of $\boldsymbol{A}_j$. Similarly, the matrix $\boldsymbol{A}_j$ would be a principal submatrix of all matrices $\boldsymbol{A}_m$ for $1 \leq m < j$. As the determinant of $\boldsymbol{A}_j$ is non-zero, it follows from Cramer's rule that

$$\boldsymbol{w}_j = \boldsymbol{A}_j^{-1} \boldsymbol{q}_j = \frac{\text{adj}(\boldsymbol{A}_j)}{\det(\boldsymbol{A}_j)} \boldsymbol{q}_j = \text{adj}(\boldsymbol{A}_j) \boldsymbol{q}_j. \tag{4.35}$$

Applying these weights in (4.29) gives

$$\varphi(\boldsymbol{u}) = \prod_{j=1}^{n} \left[g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r \right) \left(\psi_j(\boldsymbol{u}) \left[w_0^{(j)} + \sum_{m=j+1}^{n} \frac{w_{m-j}^{(j)}}{g\left(A_m(\boldsymbol{u}); 1\right)} \right] \right)^{\ell_j} \right] \tag{4.36}$$

$$= \prod_{j=1}^{n} \left[g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r \right) \left(w_0^{(j)} \psi_j(\boldsymbol{u}) + \sum_{s=j+1}^{n} w_{s-j}^{(j)} \psi_j^{(s)}(\boldsymbol{u}) \right)^{\ell_j} \right], \tag{4.37}$$

which completes the proof of theorem 1. □

It follows naturally from definition 1 that the functions ψ_j and $\psi_j^{(s)}$ are characteristic functions. Any convex combination of the functions is therefore also a characteristic function. In particular, it is the characteristic function of a mixture distribution. If said mixture distribution is infinitely divisible, then raising the characteristic function of the mixture to any positive power will result in another characteristic function. Furthermore, definition 1 also states that the function g is a characteristic function whenever the parameter is non-negative.

Under the assumption that the shape parameters take integer values satisfying the condition $\ell_j \geq \sum_{r=1}^{j-1} \ell_r$ for all $j \in \{1, ..., n\}$, the characteristic function is associated with a multivariate matrix-exponential distribution in the sense of Kulkarni, i.e. an MME* distribution. Hence, under the given assumptions, a multivariate gamma distribution in the sense of Dussauchoy and Berland belongs to the class of MME* distributions. Additionally, if the weights are non-negative, any multivariate gamma distribution is a multivariate phase-type distribution (MPH*).

Neuts (1975) introduces a phase-type distribution as the probability law governing the time until absorption in a finite state absorping Markov jump process. As described in Bladt and Nielsen (2017), a phase-type distribution can also be obtained by considering the accumulated reward of an absorping Markov jump process which receives rewards linearly while traversing the transient states. Kulkarni (1989) invokes this concept to introduce MPH* distributions. By assigning different reward rates for the different components in the transient states, the accumulation of rewards varries across the components. Accordingly, MPH* distributions have three parameters; an initial distribution of the Markov jump process, a sub-generator matrix governing the transitions among the transient states of the process, and a reward structure (matrix) containing the reward rates. Formally, for a random vector $\boldsymbol{Z}$ having an MPH* distribution, one writes $\boldsymbol{Z} \sim \text{MPH*}(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$, where the parameters are the initial distribution, the sub-generator matrix, and the reward matrix, respectively. The multivariate phase-type distributions are closed under convolution and mixing, which follows immediately from the analogous closure properties of the univariate phase-type distributions. Bladt and Nielsen (2017), section 3.1.6, provides the formulae for convolutions and mixtures of univariate phase-type distributions, which generalize easily to the multivariate setting. The exact same properties hold for the univariate and multivariate matrix-exponential distributions (MME*), see e.g. Bladt and Nielsen (2017), section 4.4.1. The multivariate matrix-exponential distributions (MME*) generalize the MPH* distributions. The MME* distributions can be characterized as the distributions with density functions (equivalently, distributions functions) on the same form as the MPH* distributions, but where the parameters do not need to have probabilistic interpretations. Consequently, MME* distributions are parametrized exactly like MPH* distributions. Moverover, it is sometimes helpful to indicate the dimension of the representation (the number of transient states in the underlying Markov process) as a subscript after MPH*/MME* when declaring the distribution of random vectors.

Since there are formulae available for obtaining the distributions of convolutions and mixtures, we shall simply provide the MPH* representation of the separate distributions in the mixtures and convolutions in (4.23). The two following lemmas provide the necessary results for obtaining the representations of convolutions and mixtures of MPH* distributions.

Lemma 1 - Convolutions of MPH* distributions
Let $\boldsymbol{X} \sim \text{MPH*}_{m_1}(\boldsymbol{a}_1, \boldsymbol{S}_1, \boldsymbol{R}_1)$ and $\boldsymbol{Y} \sim \text{MPH*}_{m_2}(\boldsymbol{a}_2, \boldsymbol{S}_2, \boldsymbol{R}_2)$ be independent random vectors. Then the convolution $\boldsymbol{X} + \boldsymbol{Y}$ follows an $\text{MPH*}_{m_1+m_2}(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ distribution, where

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, (1 - ||\boldsymbol{\alpha}_1||_1)\boldsymbol{\alpha}_2), \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{s}_1\boldsymbol{\alpha}_2 \\ \boldsymbol{0}_{m_2 \times m_1} & \boldsymbol{S}_2 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \end{bmatrix}, \tag{4.38}$$

where $\boldsymbol{0}_{m_2}$ is a row vector of dimension m_2, $\boldsymbol{0}_{m_2 \times m_1}$ is an $(m_2 \times m_1)$ matrix containing zeros, and $\boldsymbol{s}_1$ is the exit rate vector of $\boldsymbol{X}$ defined as $\boldsymbol{s}_1 = -\boldsymbol{S}_1\boldsymbol{1}_{m_1}^{\top}$. The symbol $\boldsymbol{1}_{m_1}^{\top}$ represents the transposition of an m_1-dimensional row vector of ones. Applying this argument successively implies that any finite convolution of independent MPH* distributions is a MPH* distribution.

Proof. See section 5 of Kulkarni (1989) or sections 3.1.6 and 8.1 of Bladt and Nielsen (2017). □

Lemma 2 - Mixtures of MPH* distributions
Let $\boldsymbol{X} \sim \text{MPH*}_{m_1}(\boldsymbol{a}_1, \boldsymbol{S}_1, \boldsymbol{R}_1)$ and $\boldsymbol{Y} \sim \text{MPH*}_{m_2}(\boldsymbol{a}_2, \boldsymbol{S}_2, \boldsymbol{R}_2)$ be independent random vectors. Then the mixture of $\boldsymbol{X}$ and $\boldsymbol{Y}$ with weights/probabilities $(w_{\boldsymbol{X}}, w_{\boldsymbol{Y}})$ follows an $\text{MPH*}(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ distribution, where

$$\boldsymbol{\alpha} = (w_{\boldsymbol{X}}\boldsymbol{\alpha}_1, w_{\boldsymbol{Y}}\boldsymbol{\alpha}_2), \quad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0}_{m_1 \times m_2} \\ \boldsymbol{0}_{m_2 \times m_1} & \boldsymbol{S}_2 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 \\ \boldsymbol{R}_2 \end{bmatrix}, \tag{4.39}$$

where the same notation from lemma 1 applies. Repeating this argument arbitrarily implies that any finite mixture of MPH* distributions is a MPH* distribution.

Proof. See section 5 of Kulkarni (1989) or sections 3.1.6 and 8.1 of Bladt and Nielsen (2017). Kulkarni does not provide a proof, but simply states the result. □

MPH* distributions can contain atoms, which explains the form of the initial distribution in lemma 1. For the purposes of this paper however, usually the cases without atoms are considered. The lemmas provide a method to obtain a MPH* representation for the multivariate gamma distribution under certain conditions, which are discussed in the next sections.

Theorem 2 - MPH* representation of the multivariate gamma distribution
A multivariate gamma distribution by Dussauchoy and Berland associated with the characteristic function in (4.23), where the shape parameters take integer values satisfying the condition $\ell_j \geq \sum_{r=1}^{j-1} \ell_r$ for all $j \in \{1, ..., n\}$ and the weights are non-negative, can be obtained as mixtures and convolutions of independent multivariate gamma distributions. A random vector $\boldsymbol{X}$ following a multivariate gamma distribution associated with the characteristic function $g(\sum_{j=1}^{n} r_j u_j, \ell)$ with integer shape parameter and non-negative values of the rates $\{r_j\}_{j=1}^{n}$ has a $\text{MPH*}_{\ell}(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ distribution with:

$$\boldsymbol{\alpha} = (1, \boldsymbol{0}_{\ell-1}), \quad \boldsymbol{R} = \begin{bmatrix} r_1 & r_2 & \cdots & r_{n-1} & r_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_1 & r_2 & \cdots & r_{n-1} & r_n \end{bmatrix}, \tag{4.40}$$

and the subgenerator $\boldsymbol{S}$ is an upper bidiagonal matrix with negative ones in the diagonal and positive ones in the diagonal above, i.e. the subgenerator of an absorbing Markov jump process with ℓ successive phases, each with an exponential(1) sojourn time.

Proof. The characteristic function of a MPH*$_\ell(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{R})$ distribution (without an atom) is given as

$$h(\boldsymbol{u}) = \boldsymbol{\alpha}\left(-\mathrm{diag}(i\boldsymbol{R}\boldsymbol{u}) - \boldsymbol{S}\right)^{-1}\boldsymbol{s}, \tag{4.41}$$

where $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{1}_\ell^\top$ and $\mathrm{diag}(i\boldsymbol{R}\boldsymbol{u})$ is a diagonal matrix with elements from the vector $i\boldsymbol{R}\boldsymbol{u}$. Inserting the suggested parameters of the MPH* distribution then yields exactly the function $g(\sum_{j=1}^n r_j u_j; \ell)$. The functions ψ_j and $\psi_j^{(s)}$ given in definition 1 are obtained as convolutions of independent random vectors with characteristic functions on the above form. The characteristic function in (4.23) is associated with an n-fold convolution of independent random vectors whose characteristic functions are completely similar:

$$g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1}\ell_r\right)\left(w_0^{(j)}\psi_j(\boldsymbol{u}) + \sum_{s=j+1}^{n} w_{s-j}^{(j)}\psi_j^{(s)}(\boldsymbol{u})\right)^{\ell_j}, \quad j \in \{1, ..., n\}. \tag{4.42}$$

The assumption that $\ell_j \geq \sum_{r=1}^{j-1}\ell_r$ for all $j \in \{1, ..., n\}$ ensures that the first factor of (4.42) is a valid characteristic function on the abovementioned form $g(\sum_{j=1}^n r_j u_j; \ell)$. The second factor of (4.42) is independent from the first factor and can be identified as an ℓ_j-fold convolution of independent and identically distributed random vectors. These random vectors are distributed according to a mixture distribution due to the assumption that all weights are non-negative and the fact that the weights sum to one, i.e. the weights constitute a convex linear combination. This leads to the implication that the characteristic function in (4.23) can be obtained purely through convolutions and mixtures of MPH* distributions with representations as in (4.40). Lemmas 1 and 2 show that the class of MPH* distributions is closed under finite convolutions and mixtures and how to find the resulting MPH* representation. Therefore, the lemmas imply that the characteristic function in (4.23) is associated with a MPH* distribution and give an appropriate representation. It would be too cumbersome to write out an exact representation, which shall therefore not feature in this proof. □

4. Necessary and sufficient conditions for the existence of the distribution

In theorem 2, three assumptions are necessary to identify the multivariate gamma distribution by Dussauchoy and Berland as a MPH* distribution. The most fundamental assumption is that the shape parameters of the gamma distribution take integer values, which naturally are assumed to be positive. This assumption will be discussed later, but gamma distributions with non-integer shape parameters rarely belong to the MPH* class, which is the reason this case shall not receive further attention. The second assumption involve constraints on the shape parameters of the gamma distribution, while the third assumption ensures that the weights as defined in (4.31) and (4.32) generate certain convex linear combinations. The purpose of this section is to prove the validity of the two latter assumptions.

Prior to approaching the assumptions, some preliminary results must be introduced. Firstly, a lemma on the weights is needed.

The weights are founds in equations (4.31) and (4.32). The weights in (4.31) can be found through direct computation when the other weights are given. Therefore, if those weights are temporarily disregarded, i.e. we remove the bottom line of (4.33)/(4.34) and consider the reduced systems

$$\boldsymbol{q}_j^* = \boldsymbol{A}_j^* \boldsymbol{w}_j^*, \tag{4.43}$$

for $j \in \{1, ..., n\}$, the systems can be solve simultaneously by solving the matrix equation

$$\boldsymbol{Q} = \boldsymbol{A}_1^* \boldsymbol{W}, \tag{4.44}$$

where $\boldsymbol{Q}$ and $\boldsymbol{W}$ are lower triangular matrices with columns $(\boldsymbol{q}_1^*, ..., \boldsymbol{q}_{n-1}^*)$ and $(\boldsymbol{w}_1^*, ..., \boldsymbol{w}_{n-1}^*)$ in the lower halves, respectively. The weights might then be obtained alternatively by solving the system

$$\boldsymbol{Q} = \boldsymbol{W}\boldsymbol{B}, \tag{4.45}$$

in which case the matrix $\boldsymbol{B}$ would be given as $\boldsymbol{B} = \boldsymbol{Q}^{-1}\boldsymbol{A}_1^*\boldsymbol{Q}$. However, the matrix $\boldsymbol{Q}$ is not required to be invertible, and consequently the matrix $\boldsymbol{B}$ must be identified through another approach, which is explored in the below lemma.

Lemma 3 - Recurrence relations for the weights
The weights defined in (4.32) satisfy the equations

$$w_m^{(j)} = q_{j,j+m} - \sum_{h=1}^{m-1} q_{j,j+h} w_{m-h}^{(j+h)}, \quad \forall (j,m) : 1 \leq m \leq n-j. \tag{4.46}$$

Proof. We shall invoke a proof by strong induction using the induction hypothesis given by (4.46). The hypothesis is first verified for the base cases, which are $m = 1$ and all values of $j \in \{1, ..., n-1\}$. For the base cases, the hypothesis yields

$$w_1^{(j)} = q_{j,j+1}, \tag{4.47}$$

which coincides with the results produced by the definition in (4.32). The hypothesis is now assumed to hold for all cases for $j > t$ and for all $m \leq s < n-t$ for $j = t$. It is now the objective to show the hypothesis for $m = s+1$ when $j = t$. Writing out the systems generated by (4.44) gives the recurrence relations

$$q_{j,j+m} = w_m^{(j)} + \sum_{h=1}^{m} w_h^{(j)} q_{j+h,j+m}. \tag{4.48}$$

For $m = s+1$ when $j = t$, simple rearrangements lead to

$$w_{s+1}^{(t)} = q_{t,t+(s+1)} - \sum_{h=1}^{(s+1)-1} w_h^{(t)} q_{t+h,t+(s+1)}, \tag{4.49}$$

and applying the induction hypothesis to the weights on the RHS yields

$$w_{s+1}^{(t)} = q_{t,t+(s+1)} - \sum_{h=1}^{s} \left[q_{t,t+h} - \sum_{k=1}^{h-1} w_{h-k}^{(t+k)} q_{t,t+k} \right] q_{t+h,t+(s+1)} \tag{4.50}$$

$$= q_{t,t+(s+1)} - \sum_{h=1}^{s} \left[q_{t,t+h} q_{t+h,t+(s+1)} \right] + \sum_{h=1}^{s} \sum_{k=1}^{h-1} \left[w_{h-k}^{(t+k)} q_{t,t+k} q_{t+h,t+(s+1)} \right]. \tag{4.51}$$

The derivation proceeds by interchanging the order of summation of the last term and changing the index. Thus,

$$w_{s+1}^{(t)} = q_{t,t+(s+1)} - \sum_{h=1}^{s} \left[q_{t,t+h} q_{t+h,t+(s+1)} \right] + \sum_{k=1}^{s-1} \sum_{r=1}^{s-k} \left[w_r^{(t+k)} q_{t,t+k} q_{t+(k+r),t+(s+1)} \right]. \tag{4.52}$$

The expression simplifies further when the last term of the first summation is taken out:

$$w_{s+1}^{(t)} = q_{t,t+(s+1)} - \sum_{h=1}^{s-1} q_{t,t+h} \left[q_{t+h,t+(s+1)} - \sum_{r=1}^{s-h} w_r^{(t+h)} q_{(t+h)+r,t+(s+1)} \right] - q_{t,t+s} q_{t+s,t+(s+1)}. \tag{4.53}$$

The quantity in the brackets is recognised as the weight $w_{(s+1)-h}^{(t+h)}$ in accordance with (4.48). And furthermore, $q_{t+s,t+(s+1)} = w_1^{(t+s)} = w_{(s+1)-s}^{(t+s)}$, which implies that

$$w_{s+1}^{(t)} = q_{t,t+(s+1)} - \sum_{h=1}^{s-1} q_{t,t+h} w_{(s+1)-h}^{(t+h)} - q_{t,t+s} w_{(s+1)-s}^{(t+s)} \tag{4.54}$$

$$= q_{t,t+(s+1)} - \sum_{h=1}^{s} q_{t,t+h} w_{(s+1)-h}^{(t+h)}. \tag{4.55}$$

Hence, the hypothesis is also true for $j = t$ and $m = s + 1$. By the principle of induction, we can conclude that the hypothesis holds for all (j, m) such that $1 \leq m \leq n - j$. □

The reccurence relations provided in lemma 3 are important in the proof of the next theorem concerning some necessary conditions for the existence of the distribution.

Theorem 3 - Necessary conditions on the shape parameters
A multivariate gamma distribution of the Dussauchoy-Berland type associated with the characteristic function in (4.17) exists if, and only if, the shape parameters satisfy the inequalities

$$\ell_j \geq \sum_{\substack{r=1 \\ r \notin \mathcal{N}_j}}^{j-1} \ell_r \tag{4.56}$$

for all $j \in \{1, ..., n\}$, where $\mathcal{N}_j := \{r \in \{1, ..., j-1\} : w_{j-r}^{(r)} = 0\}$.

Proof. Let $\boldsymbol{Z}$ have the characteristic function in (4.17) with weights defined by (4.31) and (4.32). Consider then the univariate projection Y given as a linear combination of the vector components of $\boldsymbol{Z}$:

$$Y = \sum_{k=1}^{n} y_k Z_k, \tag{4.57}$$

where the coefficients are defined by

$$y_n = 1, \tag{4.58}$$

$$y_k = -w_{n-k}^{(k)}, \quad k \in \{1, ..., n-1\}. \tag{4.59}$$

The existence of the distribution implies the existence of the projection, whose characteristic function is found as

$$\varphi_Y(u) = \mathbb{E}\left[e^{iuY}\right] = \mathbb{E}\left[e^{iu\sum_{k=1}^{n} y_k Z_k}\right] = \varphi(u\boldsymbol{y}), \tag{4.60}$$

where $\boldsymbol{y} = \{y_k\}_{k=1}^{n}$. The characteristic function of the projection thus evaluates to

$$\varphi_Y(u) = \varphi(u\boldsymbol{y}) = \prod_{j=1}^{n} \frac{\left(1 - iu\left(\sum_{k=j+1}^{n} q_{jk} y_k\right)\right)^{\ell_j}}{\left(1 - iu\left(y_j + \sum_{k=j+1}^{n} q_{jk} y_k\right)\right)^{\ell_j}}. \tag{4.61}$$

For $j \in \{1, ..., n-1\}$, the factors of the product can be simplified using lemma 3. According to lemma 3, for $j \in \{1, ..., n-1\}$, the coefficients satisfy the recursions

$$y_j = -w_{n-j}^{(j)} = -q_{jn} + \sum_{h=1}^{n-j-1} w_{n-j-h}^{(j+h)} q_{j,j+h} = -q_{jn} y_n - \sum_{h=1}^{n-j-1} y_{j+h} q_{j,j+h} \tag{4.62}$$

$$= -\sum_{k=j+1}^{n} y_k q_{jk}. \tag{4.63}$$

This result allows for simplifying the characteristic function to

$$\varphi_Y(u) = \prod_{j=1}^{n} \frac{\left(1 - iu\left(\sum_{k=j+1}^{n} q_{jk} y_k\right)\right)^{\ell_j}}{\left(1 - iu\left(y_j + \sum_{k=j+1}^{n} q_{jk} y_k\right)\right)^{\ell_j}} = \frac{\prod_{j=1}^{n-1} \left(1 + iuy_j\right)^{\ell_j}}{\left(1 - iu\right)^{\ell_n}}. \tag{4.64}$$

The numerator only simplifies if, and only if, some of the shape parameters or the weights are zero. Since the distribution would simply reduce to a distribution of lower dimension if some of the shape parameters were zero, this case can be disregarded without loss of generality. Thus, the numerator simplifies only if $\mathcal{N}_n \neq \emptyset$. For any $j \in \mathcal{N}_n$, it holds that $w_{n-j}^{(j)} = -y_j = 0$, which implies that the corresponding factor in the numerator of (4.64) is simply one. Thus,

$$\prod_{j=1}^{n-1} \left(1 + iuy_j\right)^{\ell_j} = \prod_{\substack{j=1 \\ j \notin \mathcal{N}_n}}^{n-1} \left(1 + iuy_j\right)^{\ell_j}. \tag{4.65}$$

The limiting behaviour of the characteristic function in (4.64) can thus be described as

$$\varphi_Y(u) = \frac{\mathcal{O}\left(u^L\right)}{\mathcal{O}\left(u^{\ell_n}\right)}, \tag{4.66}$$

which will diverge towards to infinity unless the condition

$$L = \sum_{\substack{j=1 \\ j \notin \mathcal{N}_n}}^{n-1} \ell_j \leq \ell_n \tag{4.67}$$

is satisfied. Since characteristic functions are uniformly bounded by one, the condition in (4.67) is necessary for the existence of the distribution. It is however not a sufficient condition for the existence of the distribution. This argument justifies the claim in (4.56) for $j = n$. The proofs of the claims for the remaining cases are obtained by applying the same analysis to the lower dimensional marginal distributions $(Z_1, ..., Z_{n-1}), (Z_1, ..., Z_{n-2}), ..., (Z_1, Z_2), (Z_1)$, which must also exist for $\boldsymbol{Z}$ to exist. These cases give rise to the remaining inequalities in the theorem, which in turn completes the proof. □

Theorem 3 adresses the second assumption of theorem 2, whose purpose is to ensure that the first factor of (4.42) is a characteristic function, but the conditions in this assumption can actually be relaxed to the conditions in theorem 3, i.e. the constraint

$$\ell_j \geq \sum_{r=1}^{j-1} \ell_r \tag{4.68}$$

can be relaxed to the condition stated in (4.56). Notice firstly that the conditions are the same if $\mathcal{N}_j = \emptyset$. Secondly, if $\mathcal{N}_j$ is non-empty, a mixture in the distribution simplifies due to a common factor between the mixture components. In particular, for any $k \in \mathcal{N}_j$, the weight $w_{j-k}^{(k)} = 0$, which implies that the mixture

$$w_0^{(k)} \psi_k(\boldsymbol{u}) + \sum_{s=k+1}^{n} w_{s-k}^{(k)} \psi_k^{(s)}(\boldsymbol{u}) \tag{4.69}$$

has zero weight on the component $\psi_k^{(j)}$. Consequently, all the components of the mixture share the common factor $g(A_j(\boldsymbol{u}); 1)$. As seen from (4.42), this mixture is raised to the power ℓ_k, which means that the factor $g(A_j(\boldsymbol{u}); \ell_k)$ can be factored out of the mixture and instead be multiplied with other factors in the characteristic function (4.23), e.g.

$$g\left(A_j(\boldsymbol{u}); \ell_k\right) g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{r=1}^{j-1} \ell_r\right) = g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{\substack{r=1 \\ r \neq k}}^{j-1} \ell_r\right), \tag{4.70}$$

and repeating the argument for all $k \in \mathcal{N}_j$ thus leaves the factor

$$g\left(A_j(\boldsymbol{u}); \ell_j - \sum_{\substack{r=1 \\ r \notin \mathcal{N}_j}}^{j-1} \ell_r\right), \tag{4.71}$$

which is a characteristic function exactly when the condition in theorem 3, (4.56), is satisfied.

The next assumption invoked in theorem 2 is the assumption that the weights are non-negative. This property is ensured by the next theorem.

Theorem 4 - Necessary conditions on the weights

A multivariate gamma distribution in the sense of Dussauchoy and Berland associated with the characteristic function in (4.23) exists if, and only if, all the weights defined in (4.31) and (4.32) are non-negative. This implies that all the weights are probabilities.

Proof. The theorem shall be shown using strong induction. The induction hypothesis is that the weights $(w_0^{(j)}, w_1^{(j)}, ..., w_{n-j}^{(j)})$ are non-negative for any $j \in \{1, ..., n-1\}$. The base case is $j = n-1$, for which it follows directly by definition that $w_1^{(n-1)} = q_{n-1,n}$ and $w_0^{(n-1)} = 1 - q_{n-1,n}$ are both probabilities. Thus, the induction hypothesis is satisfied for the base case. In the induction step, it is assumed that the hypothesis holds for all $j \geq k$ with $n-1 \geq k > 1$. This assumption implies that all the weights $(w_1^{(k-1)}, ..., w_{n-k}^{(k-1)})$ are probabilities, since the structures of the weights are completely similar to e.g. the weights $(w_1^{(k)}, ..., w_{n-k}^{(k)})$, which are probabilities.

It just remains to show that $w_0^{(k-1)}$ and $w_{n-(k-1)}^{(k-1)}$ are probabilities. Consider therefore the two univariate projections of the random vector $\boldsymbol{Z}$:

$$Y_1 = Z_n - \sum_{m=k-1}^{n-1} w_{n-m}^{(m)} Z_m, \tag{4.72}$$

$$Y_2 = X_{k-1,n} + \sum_{m=k}^{n-1} w_0^{(m)} X_{k-1,m}. \tag{4.73}$$

The first projection is completely analogous to the projection in (4.57) and has the characteristic function

$$\varphi_{Y_1}(u) = \frac{\prod_{r=k-1}^{n-1} \left(1 - i w_{n-r}^{(r)} u\right)^{\ell_r}}{(1-iu)^{\ell_n}}, \tag{4.74}$$

cf. the derivation in the proof of theorem 3. Since this function coincides with a regular analytic function in some neighborhood around the origin of the complex plane, the function is an analytic characteristic function according to Lukacs (1960), p. 130. Chapter 7 of Lukacs is concerned with analytic characteristic functions and contains several useful theorems. An application of Theorem 7.1.1 of Lukacs shows that φ_{Y_1} can be extended to a regular analytic function in a horizontal strip of the complex plane (the strip of regularity) whose boundaries are defined by the singularities of φ_{Y_1}. Hence, φ_{Y_1} can be extended to a regular analytic function inside the strip $\{z \in \mathbb{C} : -i < \mathrm{Im}(z)\}$. Corollary to Theorem 7.1.2 (p. 134) of Lukacs states that an analytic characteristic function cannot have any zeros on the imaginary axis inside its strip of regularity. Therefore, the zeros of (4.74) must have imaginary parts no greater than -1. In particular, the zero u_0 calculated from $1 - i w_{n-(k-1)}^{(k-1)} u_0 = 0$ must satisfy that $\mathrm{Im}(u_0) \leq -1$, which implies that $w_{n-(k-1)}^{(k-1)} \in (0,1]$. If the weight is zero, then u_0 does not exist and the expression in (4.74) simplifies. In conclusion, the admissible values of the weight $w_{n-(k-1)}^{(k-1)}$ are all the values in the unit interval $[0,1]$, which means that the weight is a probability.

A similar approach using the projection Y_2 gives the result for the weight $w_0^{(k-1)}$. Rewriting the projection to clarify the linear combination yields

$$Y_2 = Z_n + \sum_{m=k}^{n-1} w_0^{(m)} Z_m - \left(q_{k-1,n} + \sum_{t=k}^{n-1} q_{k-1,t} w_0^{(t)}\right) Z_{k-1}. \tag{4.75}$$

The coefficients of the linear combination can be found through the representation $Y_2 = \langle \boldsymbol{y}_2, \boldsymbol{Z} \rangle$ in terms of an Euclidean inner product. Evaluating the auxiliary functions entering in (4.23) to establish the characteristic function of the projection gives

$$A_{k-1}(\boldsymbol{y}_2 u) = -\left(q_{k-1,n} + \sum_{t=k}^{n-1} q_{k-1,t} w_0^{(t)}\right) u + \sum_{m=k}^{n-1} q_{k-1,m} w_0^{(m)} u + q_{k-1,n} u = 0, \tag{4.76}$$

and

$$A_s(\boldsymbol{y}_2 u) = w_0^{(s)} u + \sum_{m=s+1}^{n-1} q_{sm} w_0^{(m)} u + q_{sn} u \tag{4.77}$$

$$= \left[\left(1 - \sum_{h=s+1}^{n} w_{h-s}^{(s)}\right) + \sum_{m=s+1}^{n-1} q_{sm} \left(1 - \sum_{g=m+1}^{n} w_{g-m}^{(m)}\right) + q_{sn}\right] u \tag{4.78}$$

for $s \in \{k, ..., n\}$. Rearranging the terms in the last expression gives the following:

$$A_s(\boldsymbol{y}_2 u) = \left[1 + \sum_{h=s+1}^{n} \left(q_{sh} - w_{h-s}^{(s)}\right) - \sum_{m=s+1}^{n} \sum_{g=m+1}^{n} q_{sm} w_{g-m}^{(m)}\right] u. \tag{4.79}$$

Interchanging the order of summation and adding an empty term yields

$$A_s(\boldsymbol{y}_2 u) = \left[1 + \sum_{h=s+1}^{n} \left(q_{sh} - w_{h-s}^{(s)}\right) - \sum_{g=s+1}^{n} \sum_{m=s+1}^{g-1} q_{sm} w_{g-m}^{(m)}\right] u. \tag{4.80}$$

A simple change of variable is applied before simplifying the expression

$$\begin{aligned} A_s(\boldsymbol{y}_2 u) &= \left[1 + \sum_{h=s+1}^{n} \left(q_{sh} - w_{h-s}^{(s)}\right) - \sum_{g=s+1}^{n} \sum_{d=1}^{g-s-1} q_{s,s+d} w_{g-(s+d)}^{(s+d)}\right] u & (4.81)\\ &= \left[1 + \sum_{h=s+1}^{n} \left(q_{sh} - w_{h-s}^{(s)} - \sum_{d=1}^{(h-s)-1} q_{s,s+d} w_{h-(s+d)}^{(s+d)}\right)\right] u & (4.82)\\ &= u, & (4.83)\end{aligned}$$

according to lemma 3 (4.46). These intermediate results allow for calculating the characteristic function of the projection Y_2 using the form in (4.24)

$$\begin{aligned} \varphi_{Y_2}(u) = \varphi(\boldsymbol{y}_2 u) = \prod_{s=1}^{n} \frac{g\left(A_s(\boldsymbol{y}_2 u); \ell_s\right)}{g\left(A_s^*(\boldsymbol{y}_2 u); \ell_s\right)} &= \prod_{s=k-1}^{n} \frac{g\left(A_s(\boldsymbol{y}_2 u); \ell_s\right)}{g\left(A_s^*(\boldsymbol{y}_2 u); \ell_s\right)} \\ &= \frac{1}{g\left(A_{k-1}^*(\boldsymbol{y}_2 u); \ell_{k-1}\right)} \prod_{s=k}^{n} \frac{g\left(A_s(\boldsymbol{y}_2 u); \ell_s\right)}{g\left(A_s^*(\boldsymbol{y}_2 u); \ell_s\right)} \\ &= \frac{1}{g\left(A_{k-1}^*(\boldsymbol{y}_2 u); \ell_{k-1}\right)} \prod_{s=k}^{n} \frac{g\left(u; \ell_s\right)}{g\left(\left(1 - w_0^{(s)}\right) u; \ell_s\right)}. \end{aligned} \tag{4.84}$$

The first factor of the above expression can be restated as in (4.30) such that the characteristic function becomes

$$\begin{aligned} \varphi_{Y_2}(u) &= \left(w_0^{(k-1)} + \sum_{c=k}^{n} \frac{w_{c-(k-1)}^{(k-1)}}{g\left(A_c(\boldsymbol{y}_2 u); 1\right)}\right)^{\ell_{k-1}} \prod_{s=k}^{n} \frac{g\left(u; \ell_s\right)}{g\left(\left(1 - w_0^{(s)}\right) u; \ell_s\right)} & (4.85)\\ &= \left(w_0^{(k-1)} + \sum_{c=k}^{n} \frac{w_{c-(k-1)}^{(k-1)}}{g\left(u; 1\right)}\right)^{\ell_{k-1}} \prod_{s=k}^{n} \frac{g\left(u; \ell_s\right)}{g\left(\left(1 - w_0^{(s)}\right) u; \ell_s\right)} & (4.86)\\ &= \left(w_0^{(k-1)} + \frac{1 - w_0^{(k-1)}}{g\left(u; 1\right)}\right)^{\ell_{k-1}} \prod_{s=k}^{n} \frac{g\left(u; \ell_s\right)}{g\left(\left(1 - w_0^{(s)}\right) u; \ell_s\right)} & (4.87)\\ &= \frac{1}{g\left(\left(1 - w_0^{(k-1)}\right) u; \ell_{k-1}\right)} \prod_{s=k}^{n} \frac{g\left(u; \ell_s\right)}{g\left(\left(1 - w_0^{(s)}\right) u; \ell_s\right)} & (4.88)\\ &= \left(1 - iu\left(1 - w_0^{(k-1)}\right)\right)^{\ell_{k-1}} \prod_{s=k}^{n} \left(\frac{1 - iu\left(1 - w_0^{(s)}\right)}{1 - iu}\right)^{\ell_s}. & (4.89)\end{aligned}$$

Applying the same corollary from Lukacs as previously, the function in (4.89) cannot have zeros with imaginary parts greater than one. Consequently, the function is a characteristic function if, and only if,

$$1 - w_0^{(s)} \in [0,1], \quad \forall s \in \{k-1, k, ..., n\}, \tag{4.90}$$

and in particular, the weight $w_0^{(k-1)}$ must be a probability. Hence, the result shows that all the weights $w_0^{(k-1)}, w_1^{(k-1)}, ..., w_{n-(k-1)}^{(k-1)}$ are all probabilities. The principle of induction now justifies the conclusion that all the weights in (4.23) are probabilities. □

Theorems 3 and 4 are necessary conditions, but they are also sufficient conditions for the existence of the multivariate distribution if the shape parameters take integer values. Under this assumption, the theorems ensure that the components resulting from the decomposition in theorem 1 are all MPH* distributed. Theorem 2 then shows that the multivariate distribution is indeed a MPH* distribution and provides the associated representation. The next section concerns infinite divisibility and discusses the existence of the distribution when the shape parameters take non-integer values.

5. Discussion on infinite divisibility

The theorems established in the previous sections provide necessary and sufficient conditions for the distribution to exist whenever the shape parameters take integer values. The necessary conditions are valid even when the shape parameters take non-integer values, but the same does not apply to the sufficiency of the conditions.

In the preceding sections, the arguments and proofs on the weights have relied on properties of characteristic functions of various projections of the distribution. This approach cannot necessarily be invoked in the discussion on infinite divisibility, since infinite divisibility of the projections does not imply infinite divisibility of the joint distribution, see e.g. Dwass et al. (1957), where the Wishart distribution is mentioned as an explicit example. The converse statement is however true, which means that if a linear combination of the vector components is not infinitely divisible then neither is the joint distribution of the random vector. It is thus sufficient to identify one linear combination of the components that is not infinitely divisible in order to show that the distribution is not infinitely divisible.

In the bivariate case, the distribution is infinitely divisible. Dussauchoy and Berland proved this using the Kolmogorov Canonical Representation theorem, but left out considerable parts of the proof in their paper. The proof comes down to showing that the function

$$\eta(t) = \frac{1 - itq}{1 - it}, \quad q \in [0,1], \tag{4.91}$$

is infinitely divisible. To confirm this, define the function $K : \mathbb{R} \to \mathbb{R}$ as

$$K(x) = \begin{cases} (1-q^2) + q^2 e^{-\frac{x}{q}} \left(1 + \frac{x}{q}\right) - e^{-x}(1+x), & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.92}$$

with the corresponding derivative

$$\frac{dK}{dx}(x) = \begin{cases} x\left(e^{-x} - e^{-\frac{x}{q}}\right), & x \geq 0 \\ 0, & x < 0 \end{cases}, \tag{4.93}$$

as for $x \geq 0$:

$$\frac{dK}{dx}(x) = q^2\left(\frac{1}{q}e^{-\frac{x}{q}} - \frac{1}{q}e^{-\frac{x}{q}}\left(1 + \frac{x}{q}\right)\right) - e^{-x} + e^{-x}(1+x) = x\left(e^{-x} - e^{-\frac{x}{q}}\right). \tag{4.94}$$

The function K is bounded, and since the derivative is non-negative, the function must be non-decreasing. The function therefore satisfy the two conditions stated in the Kolmogorov Canonical Representation theorem, and we may evaluate the integral

$$\int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} dK(x) = \int_{0}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} x\left(e^{-x} - e^{-\frac{x}{q}}\right) dx \tag{4.95}$$

$$= \int_{0}^{\infty} \frac{e^{itx} - 1}{x}\left(e^{-x} - e^{-\frac{x}{q}}\right) dx - \int_{0}^{\infty} it\left(e^{-x} - e^{-\frac{x}{q}}\right) dx. \tag{4.96}$$

The latter integral in equation (4.96) is handled easily as

$$\int_{0}^{\infty} it\left(e^{-x} - e^{-\frac{x}{q}}\right) dx = it\left[-e^{-x} + qe^{-\frac{x}{q}}\right]_{0}^{\infty} = it(1-q). \tag{4.97}$$

The former integral in equation (4.96) requires more technique, but can be obtained by rewriting the integral as

$$\int_{0}^{\infty} \frac{e^{itx} - 1}{x}\left(e^{-x} - e^{-\frac{x}{q}}\right) dx = \int_{0}^{\infty} \frac{e^{-(1-it)x} - e^{-x}}{x} dx - \int_{0}^{\infty} \frac{e^{-(1-itq)\frac{x}{q}} - e^{-\frac{x}{q}}}{x} dx \tag{4.98}$$

and using a table base result from section 3.411 of Gradshteyn et al. (2007), which gives

$$\int_{0}^{\infty} \frac{e^{-(1-it)x} - e^{-x}}{x} dx - \int_{0}^{\infty} \frac{e^{-(1-itq)\frac{x}{q}} - e^{-\frac{x}{q}}}{x} dx = -\ln(1-it) + \ln(1-itq). \tag{4.99}$$

Rearranging the results yield the Kolmogorov canonical representation form:

$$\ln(\eta(t)) = -it(1-q) + \int_{-\infty}^{\infty} \frac{e^{itx} - 1 - itx}{x^2} dK(x), \tag{4.100}$$

which implies that η is infinitely divisible.

Our analysis of the divisibility of the general multivariate distribution has been inconclusive, however there are some noteworthy findings from the analysis. In the bivariate case, the property follows almost immediately from the fact that mixtures of univariate exponential distributions are infinitely divisible, cf. Steutel (2004) or Goldie (1967), in conjunction with the decomposition from theorem 1. The decomposition suggests that the distribution is obtained from convolutions of gamma mixtures, which are themselves convolutions of a gamma distribution and mixtures of univariate exponential distributions. Since infinite divisibility is preserved under convolutions, the bivariate distribution is infinitely divisible (as verified above).

In higher dimensions, the decomposition suggests that the distribution contains mixtures of multivariate exponential distributions, which are generally not infinitely divisible, cf. Steutel (2004). Indeed, using a simplified notation, the general multivariate distribution will contain a mixture with the Laplace transform:

$$\mathcal{L}(u_1, u_2) = \left(\frac{p}{1+u_1} + \frac{q}{1+u_2} + \frac{1-p-q}{(1+u_1)(1+u_2)}\right)^k, \quad k \geq 0, \quad p, q, 1-p-q \in [0,1], \tag{4.101}$$

which is well-defined at least whenever the two arguments are positive. Based on the work by Steutel, it is reasonable to expect that such mixtures are not infinitely divisible, but even if this is the case, it is insufficient to disprove the infinite divisibility of the distribution as factors of an infinitely divisible characteristic function need not be infinitely divisible themselves.

The mixture in equation (4.101) is also relevant with regard to the existence of the distribution for non-integer shape parameters. To realize this, define the function $\mathcal{H} : \mathbb{C}^2 \to \mathbb{C}$ as

$$\mathcal{H}(u_1, u_2) = \frac{p}{1+u_1} + \frac{q}{1+u_2} + \frac{1-p-q}{(1+u_1)(1+u_2)}, \quad u_1, u_2 > -1, \tag{4.102}$$

which is the joint Laplace transform of a bivariate exponential mixture. Consequently, the transform in (4.101) is defined as $\mathcal{L}(u_1, u_2) = \mathcal{H}(u_1, u_2)^k$. The mixed partial derivative of this Laplace transform is established through simple, but somewhat tedious, calculations and is given by

$$\frac{\partial^2 \mathcal{L}}{\partial u_1 \partial u_2}(u_1, u_2) = \frac{k\mathcal{H}(u_1, u_2)^{k-2}}{(1+u_1)^2(1+u_2)^2}\left[(k-1)qp + k\mathcal{H}(u_1, u_2)\right]. \tag{4.103}$$

As $\mathcal{H}$ is positive for real $u_1, u_2 > 0$ and tends to zero when the arguments tend to infinity, the mixed partial derivative in (4.103) is negative for $k < 1$ when the arguments tend to infinity along the real axis. According to theorem 5 of Ditkin (1962), the function $\mathcal{L}$ cannot be a joint Laplace transform for $k < 1$.

We may therefore conclude that several of the factors (those associated with mixtures) in the characteristic function of the multivariate gamma distribution are not themselves characteristic functions when the shape parameters take non-integer values, which could be a strong indication that the multivariate distribution cannot exist when certain shape parameters take non-integer values. Furthermore, it also strongly suggest that when the distribution exists in higher dimensions, it is not infinitely divisible.

A concrete example of this behaviour is observed for the three-dimensional distribution with shape parameters

$$\ell_1 = \frac{1}{3}, \quad \ell_2 = \frac{13}{30}, \quad \ell_3 = \frac{26}{30}, \tag{4.104}$$

and q-parameters

$$q_{12} = \frac{1}{3}, \quad q_{23} = \frac{1}{2}, \quad q_{13} = \frac{1}{2}, \tag{4.105}$$

which implies the weights

$$w_1^{(1)} = q_{12} = \frac{1}{3}, \quad w_2^{(1)} = q_{13} - q_{12}q_{23} = \frac{1}{3}, \quad w_1^{(2)} = q_{23} = \frac{1}{2}. \tag{4.106}$$

The above counterexample illustrates that the necessary conditions given in the previous section can be satisfied without ensuring the existence of the distribution. Furthermore, the example further proves that there are combinations of parameters, specifically with non-integer shape parameters, which do not lead to a valid characteristic function. Finally, the example proves that some instances of existing distributions are not infinitely divisible, since this otherwise would imply the existence of a distribution with the above specified parameters.

6. Examples and simulations

In this section, we present some examples of the multivariate gamma distribution and show some basic visualizations of the distributions.

Example 3 - Invalid distribution due to negative weight
Consider the function $f : \mathbb{R}^4 \to \mathbb{C}$ defined by

$$f(\boldsymbol{u}) = \left(\frac{1 - i\frac{5u_2+u_3+8u_4}{10}}{1 - i\frac{u_1+5u_2+u_3+8u_4}{10}} \right) \left(\frac{1 - i\frac{0.4u_3+u_4}{2}}{1 - i\frac{u_2+0.4u_3+u_4}{2}} \right)^2 \left(\frac{1 - i\frac{2u_4}{4}}{1 - i\frac{u_3+2u_4}{4}} \right)^4 \left(\frac{1}{1 - iu_4} \right)^8. \tag{4.107}$$

First, the q-parameters are calculated, i.e. the parameters in the associated distribution with rate one gamma marginal distributions. These are calculated as $q_{ij} = \beta_{ij}\alpha_j/\alpha_i$ and must be probabilities:

$$\begin{aligned} q_{12} &= 5 \cdot \frac{2}{10} = 1, \qquad & q_{13} &= 1 \cdot \frac{4}{10} = 0.4, \qquad & q_{14} &= 8 \cdot \frac{1}{10} = 0.8, \\ q_{23} &= 0.4 \cdot \frac{4}{2} = 0.8, & q_{24} &= 1 \cdot \frac{1}{2} = 0.5, & q_{34} &= 2 \cdot \frac{1}{4} = 0.5. \end{aligned}$$

The above computations verify that the q-parameters are indeed probabilities. Next, the weights used in the decomposition (Theorem 1) must be calculated:

$$\begin{aligned} w_1^{(1)} &= q_{12} = 1, \qquad & w_2^{(1)} &= q_{13} - w_1^{(1)} q_{23} = -0.4, \qquad & w_3^{(1)} &= q_{14} - w_1^{(1)} q_{24} - w_2^{(1)} q_{34} = 0.5, \\ w_1^{(2)} &= q_{23} = 0.8, & w_2^{(2)} &= q_{24} - w_1^{(2)} q_{34} = 0.1, \\ w_1^{(3)} &= q_{34} = 0.5. \end{aligned}$$

These calculations show that one of the weights is negative, which implies that the function f is not a characteristic function per the necessary condition in Theorem 4. □

In the next example, the function under consideration cannot be a characteristic function due to the shape parameters.

Example 4 - Invalid distribution due to low shape parameter
Consider the function $f : \mathbb{R}^4 \to \mathbb{C}$ defined by

$$f(\boldsymbol{u}) = \left(\frac{1 - i\frac{u_2+0.5u_3+8u_4}{10}}{1 - i\frac{u_1+u_2+0.5u_3+8u_4}{10}} \right) \left(\frac{1 - i\frac{0.5u_3+4u_4}{5}}{1 - i\frac{u_2+0.5u_3+4u_4}{5}} \right)^2 \left(\frac{1 - i\frac{1.5u_4}{2}}{1 - i\frac{u_3+1.5u_4}{2}} \right)^2 \left(\frac{1}{1 - iu_4} \right)^4. \tag{4.108}$$

Based on the expression, the q-parameters are:

$$\begin{aligned} q_{12} &= 1 \cdot \frac{5}{10} = 0.5, \qquad & q_{13} &= 0.5 \cdot \frac{2}{10} = 0.1, \qquad & q_{14} &= 8 \cdot \frac{1}{10} = 0.8, \\ q_{23} &= 0.5 \cdot \frac{2}{5} = 0.2, & q_{24} &= 4 \cdot \frac{1}{5} = 0.8, & q_{34} &= 1.5 \cdot \frac{1}{2} = 0.75, \end{aligned}$$

which produces the weights:

$$\begin{aligned} w_1^{(1)} &= q_{12} = 0.5, \qquad & w_2^{(1)} &= q_{13} - w_1^{(1)} q_{23} = 0, \qquad & w_3^{(1)} &= q_{14} - w_1^{(1)} q_{24} - w_2^{(1)} q_{34} = 0.4, \\ w_1^{(2)} &= q_{23} = 0.2, & w_2^{(2)} &= q_{24} - w_1^{(2)} q_{34} = 0.65, \\ w_1^{(3)} &= q_{34} = 0.75, \end{aligned}$$

and

$$w_0^{(1)} = 1 - w_1^{(1)} - w_2^{(1)} - w_3^{(1)} = 0.1, \qquad w_0^{(2)} = 1 - w_1^{(2)} - w_2^{(2)} = 0.15, \qquad w_0^{(3)} = 1 - w_1^{(3)} = 0.25. \quad (4.109)$$

All the weights are non-negative (and probabilities). The next element of the distribution that needs to be examined is the shape parameters of the distribution. The shape parameters must satisfy the inequalities given in Theorem 3. In order to evaluate these constraints, the folllowing sets are computed:

$$\mathcal{N}_1 = \emptyset, \qquad \mathcal{N}_2 = \emptyset, \qquad \mathcal{N}_3 = \{1\}, \qquad \mathcal{N}_4 = \emptyset. \tag{4.110}$$

The conditions in Theorem 3 are then checked for the shape parameters $\ell_1 = 1$, $\ell_2 = 2$, $\ell_3 = 2$, and $\ell_4 = 4$. For this purpose, the function $S : \{1, ..., n\} \to \mathbb{N}$ is defined as

$$S(j) = \sum_{\substack{r=1 \\ r \notin \mathcal{N}_j}}^{j-1} \ell_r, \tag{4.111}$$

which allows for the constraints in Theorem 3 to be expressed as $S(j) \leq \ell_j$ for appropriate values of j. Using this notation, the following calculations show that the function f does not satisfy the necessary conditions of Theorem 3, and hence is not a characteristic function

$$S(1) = 0 \leq \ell_1, \qquad S(2) = 2 \leq \ell_2, \qquad S(3) = 2 \leq \ell_3, \qquad S(4) = 5 > \ell_4. \tag{4.112}$$

Specifically, it is the last inequality, which violates the necessary condition from Theorem 3. □

The final example presents a valid distribution and shows how to obtain the MPH* representation of the distribution. Furthermore, the example contains a numerical simulation of the distribution, and some visuals of the results are displayed.

Example 5 - Representation and simulation

Consider the same function as in example 4 with the adjustment that $\ell_4 = 7$, i.e. the function is now

$$f(\boldsymbol{u}) = \left(\frac{1 - i\frac{u_2 + 0.5u_3 + 8u_4}{10}}{1 - i\frac{u_1 + u_2 + 0.5u_3 + 8u_4}{10}} \right) \left(\frac{1 - i\frac{0.5u_3 + 4u_4}{5}}{1 - i\frac{u_2 + 0.5u_3 + 4u_4}{5}} \right)^2 \left(\frac{1 - i\frac{1.5u_4}{2}}{1 - i\frac{u_3 + 1.5u_4}{2}} \right)^2 \left(\frac{1}{1 - iu_4} \right)^7. \tag{4.113}$$

The conditions in Theorems 3 and 4 in combination are sufficient conditions for the existence of the distribution. Therefore, let $\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4)$ be a multivariate gamma distribution in the sense of Dussauchoy and Berland with the associated characteristic function given in (4.113). We choose to consider the scaled version of the distribution $(10Z_1, 5Z_2, 2Z_3, Z_4) = (W_1, W_2, W_3, W_4)$, which has the characteristic function $f^* : \mathbb{R}^4 \to \mathbb{C}$ defined by

$$\begin{aligned} f^*(\boldsymbol{u}) = &\left(\frac{1 - i\left(0.5u_2 + 0.1u_3 + 0.8u_4\right)}{1 - i\left(u_1 + 0.5u_2 + 0.1u_3 + 0.8u_4\right)} \right) \left(\frac{1 - i\left(0.2u_3 + 0.8u_4\right)}{1 - i\left(u_2 + 0.2u_3 + 0.8u_4\right)} \right)^2 \\ &\times \left(\frac{1 - i0.75u_4}{1 - i\left(u_3 + 0.75u_4\right)} \right)^2 \left(\frac{1}{1 - iu_4} \right)^7, \end{aligned} \tag{4.114}$$

where the values are the q-parameters found in Example 4. This characteristic function can be decomposed using the result in Theorem 1. After appropriate simplifications, the decomposition becomes

$$\begin{aligned} f^*(\boldsymbol{u}) &= g\left(A_1(\boldsymbol{u});1\right)\left(w_0^{(1)}\psi_1(\boldsymbol{u}) + w_1^{(1)}\psi_1^{(2)}(\boldsymbol{u}) + w_3^{(1)}\psi_1^{(4)}(\boldsymbol{u})\right) \\ &\times g\left(A_2(\boldsymbol{u});1\right)\left(w_0^{(2)}\psi_2(\boldsymbol{u}) + w_1^{(2)}\psi_2^{(3)}(\boldsymbol{u}) + w_2^{(2)}\psi_2^{(4)}(\boldsymbol{u})\right)^2 \\ &\times \left(w_0^{(3)}\psi_3(\boldsymbol{u}) + w_1^{(3)}\psi_3^{(4)}(\boldsymbol{u})\right)^2 g\left(A_4(\boldsymbol{u});2\right) \end{aligned} \tag{4.115}$$

The weights are found in Example 4 and the auxiliary functions can be found in Definition 1. The MPH* representations of the various elements in the characteristic function are easily established and Theorem 2 describes how to obtain the MPH* representation of the complete distribution using Lemmas 1 and 2.

First, the MPH* representations of the different mixtures are determined. The first mixture (the mixture with most components) has representation $(\boldsymbol{\pi}_1^m, \boldsymbol{\Lambda}_1^m, \boldsymbol{R}_1^m)$ with:

$$\boldsymbol{\pi}_1^m = \left(w_0^{(1)}, 0, 0, w_1^{(1)}, 0, w_2^{(1)}, 0, w_3^{(1)}, 0\right), \tag{4.116}$$

$$\boldsymbol{\Lambda}_1^m = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{R}_1^m = \begin{bmatrix} 0 & 1 & q_{23} & q_{24} \\ 0 & 0 & 1 & q_{34} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & q_{34} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & q_{23} & q_{24} \\ 0 & 0 & 1 & q_{34} \end{bmatrix}. \tag{4.117}$$

The representations of the other mixtures can be obtained in a completely similar manner, and their representations shall be denoted by $(\boldsymbol{\pi}_2^m, \boldsymbol{\Lambda}_2^m, \boldsymbol{R}_2^m)$ and $(\boldsymbol{\pi}_3^m, \boldsymbol{\Lambda}_3^m, \boldsymbol{R}_3^m)$. The last mixture is special in the sense that it contains an atom. Therefore, the above construction is not the canonical (minimal order) representation. The canonical representation is however easily obtainable, which readers familiar with phase-type distributions will know. The representation of the remaining component(s) of the distribution is denoted by $(\boldsymbol{\pi}_R, \boldsymbol{\Lambda}_R, \boldsymbol{R}_R)$ and is given as

$$\boldsymbol{\pi}_R = (1, \boldsymbol{0}_4), \tag{4.118}$$

$$\boldsymbol{\Lambda}_R = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{R}_R = \begin{bmatrix} 1 & q_{12} & q_{13} & q_{14} \\ 0 & 1 & q_{23} & q_{24} \\ 0 & 0 & 1 & q_{34} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.119}$$

The first row of the matrices corresponds to the characteristic function $g(A_1(\boldsymbol{u});1)$, while the following row of the matrices is derived from the function $g(A_2(\boldsymbol{u});1)$. The construction of the matrices are completed by following the same pattern. Using these components, the representation of the full multivariate distribution

$(\boldsymbol{\pi}, \boldsymbol{\Lambda}, \boldsymbol{R})$ can be found through an application of Lemma 1. The initial distribution can be constructed as

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_R, \mathbf{0}_9, \mathbf{0}_4, \mathbf{0}_4, \mathbf{0}_2, \mathbf{0}_2). \tag{4.120}$$

Before describing the generator and reward matrices, it is advantageous to define the exit rate vectors

$$\boldsymbol{s}_1^m = -\boldsymbol{\Lambda}_1^m \mathbf{1}_9^\top, \qquad \boldsymbol{s}_2^m = -\boldsymbol{\Lambda}_2^m \mathbf{1}_4^\top, \qquad \boldsymbol{s}_3^m = -\boldsymbol{\Lambda}_3^m \mathbf{1}_2^\top, \qquad \boldsymbol{s}_R = -\boldsymbol{\Lambda}_R \mathbf{1}_5^\top. \tag{4.121}$$

The generator and reward matrices are then given by:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_R & \boldsymbol{s}_R\boldsymbol{\pi}_1^m & \mathbf{0}_{5\times4} & \mathbf{0}_{5\times4} & \mathbf{0}_{5\times2} & \mathbf{0}_{5\times2} \\ \mathbf{0}_{9\times5} & \boldsymbol{\Lambda}_1^m & \boldsymbol{s}_1^m\boldsymbol{\pi}_2^m & \mathbf{0}_{9\times4} & \mathbf{0}_{9\times2} & \mathbf{0}_{9\times2} \\ \mathbf{0}_{4\times5} & \mathbf{0}_{4\times9} & \boldsymbol{\Lambda}_2^m & \boldsymbol{s}_2^m\boldsymbol{\pi}_2^m & \mathbf{0}_{4\times2} & \mathbf{0}_{4\times2} \\ \mathbf{0}_{4\times5} & \mathbf{0}_{4\times9} & \mathbf{0}_{4\times4} & \boldsymbol{\Lambda}_2^m & \boldsymbol{s}_2^m\boldsymbol{\pi}_3^m & \mathbf{0}_{4\times2} \\ \mathbf{0}_{2\times5} & \mathbf{0}_{2\times9} & \mathbf{0}_{2\times4} & \mathbf{0}_{2\times4} & \boldsymbol{\Lambda}_3^m & \boldsymbol{s}_3^m\boldsymbol{\pi}_3^m \\ \mathbf{0}_{2\times5} & \mathbf{0}_{2\times9} & \mathbf{0}_{2\times4} & \mathbf{0}_{2\times4} & \mathbf{0}_{2\times2} & \boldsymbol{\Lambda}_3^m \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_R \\ \boldsymbol{R}_1^m \\ \boldsymbol{R}_2^m \\ \boldsymbol{R}_2^m \\ \boldsymbol{R}_3^m \\ \boldsymbol{R}_3^m \end{bmatrix}. \tag{4.122}$$

In conclusion, the random vector $\boldsymbol{W} = (W_1, W_2, W_3, W_4)$ follows an MPH*$(\boldsymbol{\pi}, \boldsymbol{\Lambda}, \boldsymbol{R})$. Random samples from such a distribution is generated through a simulation of the underlying Markov process. The underlying process is simulated until absorption and the aggregated sojourn times are recorded. The random variates of $\boldsymbol{W}$ are then obtained by scaling the aggregated sojourn times according to the reward rates in the reward matrix. Finally, the samples of $\boldsymbol{Z}$ are obtained by multiplying the components of $\boldsymbol{W}$ with the appropriate rate parameter. This procedure is efficient and easy to implement.

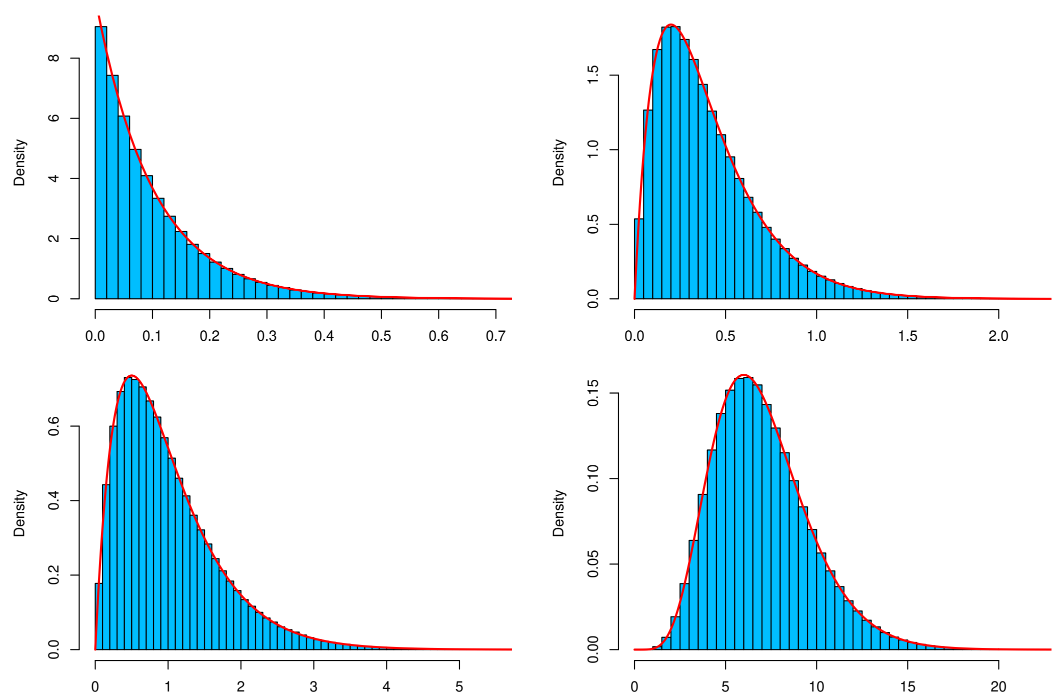

Figure 4.1: *Normalized histograms and theoretical densities of the marginal distributions. The panes are as follow: TL: Z_1, TR: Z_2, BL: Z_3, BR: Z_4.*

We generated 1,000,000 samples of the random vector $\boldsymbol{Z}$ associated with the characteristic function in (4.114). Histograms of the univariate marginal distributions are displayed together with their theoretical density functions in Figure 4.1 (The scatter plots are based on the first 1,000 samples).

The scatter plots of the bivariate marginal distributions presented in Figure 4.2 further showcase the similarity of the bivariate distributions. The scatter plots with bivariate distributions including the fourth component illustrate the effect of the large shape parameter of Z_4. As suggested in eq. (4.5), the difference in shape parameters of the components in a bivariate marginal distribution indicates the number of exponential random variables that is guaranteed to be included in $X_{ij} = Z_j - \beta_{ij} X_i$. Hence, a large difference in shape parameters will tend to increase the vertical distance from the red line ($z_j = \beta_{ij} z_i$ - boundary of the support) to the points. Consequently, the points in the scatter plots of the vectors (Z_1, Z_2), (Z_1, Z_3), and (Z_2, Z_3) are substantially closer to their respective red lines than the points in the scatter plots of the random vectors (Z_1, Z_4), (Z_2, Z_4), and (Z_3, Z_4).

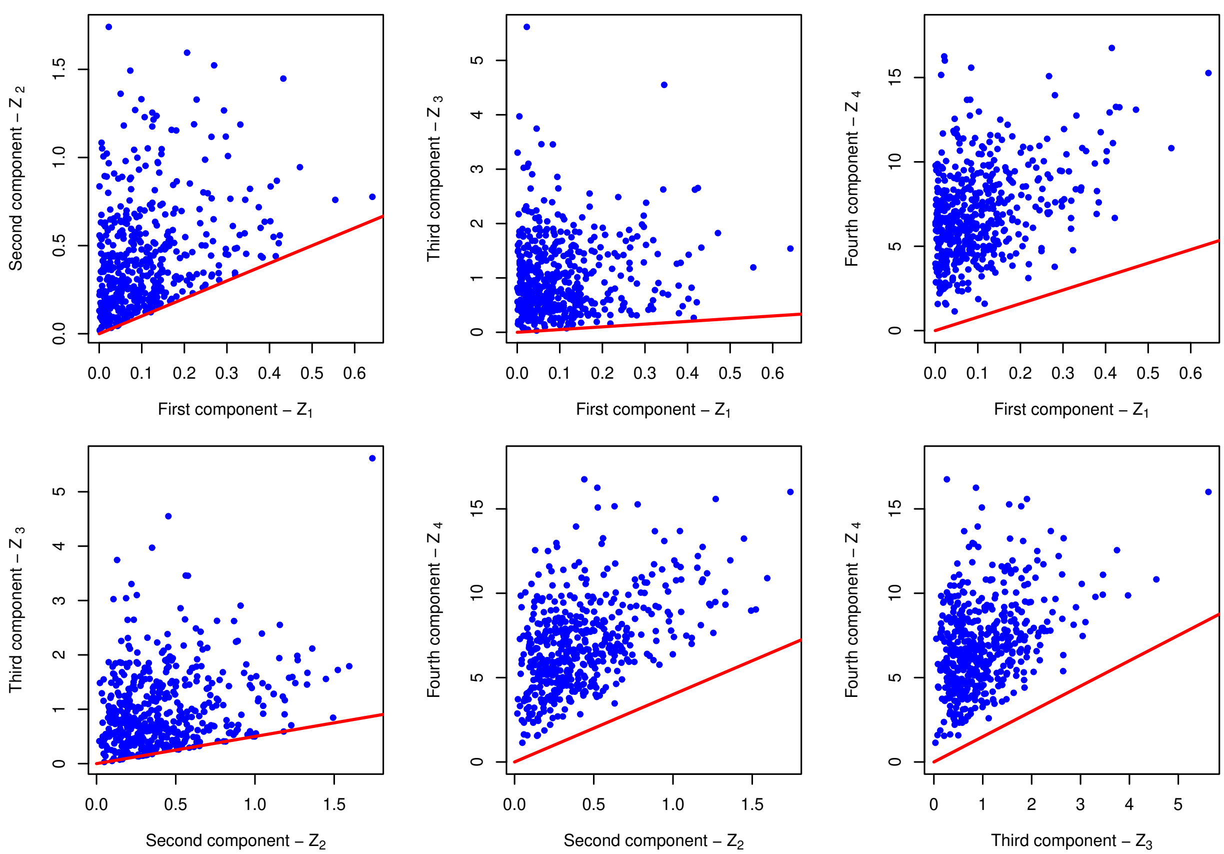

Figure 4.2: *Scatter plots of all pairs of components from the distribution. The red lines indicate the boundaries of the supports of the bivariate distributions, which are defined by* $z_j \geq \beta_{ij} z_i$ *for* $j > i$.

A numerical check also verifies that approximately $q_{23}^2 = 4\%$ of the points in the bivariate distribution of (Z_2, Z_3) are located on the line $z_3 = \beta_{23} z_2$. This is due to the fact that X_{23} has an atom at zero with the aforementioned probability.

The bivariate scatter plots also conform to the theoretical joint densities given in Berland and Dussauchoy (1972). In conclusion, simulations based on the decomposition and representation derived in the paper recreate the theoretical univariate and bivariate marginal distributions (in terms of correlation and joint density) as prescribed in the paper by Dussauchoy and Berland.

7. Conclusion

In this paper, we have derived a decomposition of the multivariate gamma distribution proposed by Dussauchoy and Berland. The decomposition suggests that the distribution arises from convolutions of gamma mixtures, and this was used to derive an appropriate multivariate phase-type representation for distributions with integer shape parameters. We also presented some necessary conditions for the existence of the distribution, and by invoking the MPH* representations, we showed that these conditions are sufficient to ensure the existence of the distribution when the shape parameters take integer values. The phase-type representation also inspired a discussion of the divisibility of the distribution, and we gave a concrete counterexample showing that the distribution cannot exist for all combinations of non-integer shape parameters. In addition, the counterexample proved that some instances of the distribution are not infinitely divisible. Finally, we showed examples of some applications of the necessary conditions along with some numerical results obtained through simulations based on the phase-type representation and verified that the simulations agree with the theoretical results.

8. References

Anderson, D. A. (1979), Some problems of inference in two-dimensional distributions, PhD thesis, Australian National University.

Balakrishnan, N., Lai, C. D. (2009), Continuous Bivariate Distributions, 2nd edition, Springer.

Barndorff-Nielsen, O. E., Maejima M. Sato, K. (2006), Some classes of multivariate infinitely divisible distributions admitting stochastic integral representations, Bernoulli, Volume 12, Issue 1, 1-33.

Bernardoff, P. (2006), Which Multivariate Gamma Distributions Are Infinitely Divisible?, Bernoulli, Volume 12, Issue 1, 169-189.

Bernardoff, P. (2016), Laplace copulas of multifactor gamma distributions are new generalized Farlie-Gumbel-Morgenstern copulas, arXiv:1611.07242.

Bernardoff, P. (2018), Marshall–Olkin Laplace transform copulas of multivariate gamma distributions, Communications in Statistics - Theory and Methods, Volume 47, Issue 3, 655-670.

Bladt, M, Nielsen, B. F. (2017), Matrix-Exponential Distributions in Applied Probability, 1st edition, Springer.

Combes, C., Dussauchoy, A., Meskens, N. (2008), Un modèle de loi bidimensionnelle : application aux différentes durées liées à l'activité chirurgicale, Journal de la Société française de statistique & Revue de statistique appliquée, Volume 149, Issue 3, 3-22.

Cramér, H. (1963), Mathematical Methods of Statistics, 10th edition, Princeton University Press.

Ditkin, V. A., Prudnikov, A. P. (1962), Operational Calculus in Two Variables and Its Applications, Pergaman Press. Translated from Russian to English by D. M. G. Wishart.

Dussauchoy, A., Berland, R. (1972), Lois gamma á deux dimensions, Comptes Rendus de l'Académie des Sciences de Paris. Série A, Volume 274, 1946–1949.

Dussauchoy, A., Berland, R. (1973), Aspects statistiques des régimes de microdécharges electriques entre électrodes métalliques placées dans un vide industriel, Vacuum, Volume 23, Issue 11, 415-421.

Dussauchoy, A., Berland, R. (1975), A Multivariate Gamma Type Distribution Whose Marginal Laws are Gamma, and which has a Property Similar to a Characteristic Property of the Normal Case. In A Modern Course on Statistical Distributions in Scientific Work, Editors: Patil, G. P., Kotz, S., Ord, J. K., Springer, Volume 17, 319–328.

Dwass, M. (1968), A theorem about infinitely divisible characteristic functions, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, Volume 9, 287-289.

Goldie, C. (1967), A class of infinitely divisible random variables, Mathematical Proceedings of the Cambridge Philosophical Society, Volume 63, Issue 4, 1141 - 1143.

Gradshteyn, I. S., Ryzhik, I. M. (2007), Table of Integrals, Series, and Products, 7th edition, Elsevier.

Griffiths, R. C. (1969a), The Canonical Correlation Coefficients of Bivariate Gamma Distributions, The Annals of Mathematical Statistics, Volume 40, Issue 4, 1401-1408.

Griffiths, R. C. (1969b), A Class of Infinitely Divisible Bivariate Gamma Distributions, Sankhyā: The Indian Journal of Statistics, Series A, Volume 31, Issue 4, 473-476.

Griffiths, R. C. (1970), Infinitely divisible multivariate gamma distributions, Sankhyā: The Indian Journal of Statistics, Series A, Volume 32, Issue 4, 393-404.

Griffiths, R. C. (1984), Characterization of infinitely divisible multivariate gamma distributions, Volume 15, Issue 1, 13-20.

Jiang, X., Nadarajah, S., Chu, J. (2018), Aggregation and Capital Allocation Formulas for Bivariate Distributions, Probability in the Engineering and Informational Sciences, Volume 32, Issue 4, 556 - 566.

Johnson, M. E. (1987), Multivariate statistical simulation, 1st edition, John Wiley and Sons,

McKay, A. T. (1934), Sampling from Batches, Supplement to the Journal of the Royal Statistical Society, Volume 1, Issue 2, 207-216.

Kotz, S., Balakrishnan, N., Johnson, N.L. (2000), Continuous Multivariate Distributions, Volume 1: Models and Applications, 2nd edition, John Wiley and Sons.

Kulkarni, V. G. (1989), A New Class of Multivariate Phase Type Distributions, Operations Research, Volume 37, Issue 1, 151-158.

Lukacs, E. (1960), Characteristic Functions, 1st edition, C. Griffin and Co.

Marshall, A. W., Olkin, I. (1967a), A Multivariate Exponential Distribution, Journal of the American Statistical Association, Volume 62, Issue 317, 30-44.

Marshall, A. W., Olkin, I. (1967b), A Generalized Bivariate Exponential Distribution, Journal of Applied Probability, Volume 4, Issue 2, 291–302.

Nadarajah, S. (2005), Reliability For Some Bivariate Exponential Distributions, Mathematical Problems in Engineering, Volume 2005, Article ID 924843, 151-163. *Updated version: Nadarajah, S., Kotz, S. (2006), Reliability For Some Bivariate Exponential Distributions, Mathematical Problems in Engineering, Volume 2006, Article ID 041652, 1-14.*

Neuts, M. F. (1975), Probability distributions of phase type, Liber Amicorum Prof. Emeritus H. Florin, 173–206.

Perez-Abreu, V., Stelzer, R. (2014), Infinitely divisible multivariate and matrix Gamma distributions, Journal of Multivariate Analysis, Volume 130, 155-175.

Saboor, A., Provost, S. B., Ahmad, M. (2010), Univariate and Bivariate Gamma-Type Distributions, LAP Lambert Academic Publishing.

Steutel, F. W., Harn, K. v. (2004), Infinite Divisibility of Probability Distributions on the Real Line, 1st edition, CRC Press.

Vere-Jones, D. (1967), The infinite divisibility of a bivariate gamma distribution, Sankhyā: The Indian Journal of Statistics, Series A, Volume 29, Issue 4, 421-422.

CHAPTER 5

Research directions and conclusion

Research in the field of matrix-analytic methods has many exciting prospects, and the field has seen several new developments in recent years. These developments include both novel theoretical results for and new applications of multivariate phase-type and matrix-exponential distributions. In this final chapter, we shall outline some directions of future research with relevance to the topics discussed throughout the thesis.

Characterisations of the MVPH and MPH* classes
Our research on the multivariate gamma distribution by Dussauchoy and Berland presented in chapter 4 has shown that the distribution is not an example of a MVPH distribution without a MPH* representation, and it therefore remains to be answered whether the two classes are different. One possible approach to the problem mentioned in chapter 3 is to study whether the multivariate gamma distribution by Krishnamoorthy and Parthasarathy admits MPH* or MME* representations. Bladt and Nielsen have already established that such a representation in some cases must have a higher dimension than the degree of the distribution, but concrete representations of the distribution are yet to be identified. A proof that the distribution does not admit a MPH*/MME* representation would imply that either the MVPH-MPH* conjecture or the MVME-MME* conjecture would be settled.

An alternative to searching for counterexamples to the conjectures is a theoretical study of the MVPH distributions with the aim of giving an analytical characterisation of the distribution and deriving some properties of the distribution. The MVPH distributions have phase-type distributed marginals and include all distributions with the property that any conical combination of the components follows a univariate phase-type distribution. A theoretical examination of the distributions might therefore naturally begin by analysing the immediate consequences of the latter condition on the conical combinations. Another natural starting point of the study could be the decompositions of the marginal distributions into convolutions of dependent exponential distributions. One could hope that such an investigation into the MVPH distributions would yield a characterisation theorem akin to the theorem which characterises those distributions with rational transforms that belong to the class of phase-type distributions in the univariate setting.

A third research direction focuses on the relationship between stochastic processes and the phase-type/matrix-exponential distributions. Bladt and Nielsen used a bivariate Brownian motion observed at an exponentially distributed stopping time to construct a bivariate distribution of the MVBME type, which did not belong to the MBME* class. Similar arguments could potentially be applied to other classes of stochastic processes and yield similar results for the MVME and MVPH classes.

The connection between stochastic processes and the matrix-exponential distributions has also been explored in the preprint Peralta (2021), where the author claims to have identified a link between Markov jump processes and the matrix-exponential distributions similar to the link between the processes and the phase-type distributions. The connection between stochastic processes and these distributions are also highly relevant with respect to the property of infinite divisibility. The concept of infinite divisibility is closely related to Lévy processes as any infinitely divisible distribution arises as a marginal distribution in a Lévy process, and it is therefore interesting to establish a link between phase-type or matrix-exponential distributions and Lévy processes.

Connection to stochastic processes and stochastic differential equations

In continuation of the above discussion, future research could explore the connection between phase-type or matrix-exponential distributions and diffusion processes governed by stochastic differential equations.

Bibby et al. (2005) shows how to construct diffusion-type models with prespecified marginal distributions and autocorrelation structure, and shows how to model time series data typically found in financial contexts using the introduced methodology. If the methodology presented in the paper can be extended to include phase-type distributed marginals, the approximation schemes described in chapter 2 might be used in conjunction with the methods devised by Bibby et al. to construct diffusion-type processes with a wide range of possible marginal distributions approximated from data.

The research by Bibby et al. considers stationary processes, but we are of course also interested in modelling processes, which have not reached a stationary state. For such cases, it might be interesting to approximate different sets of marginal distributions with multivariate phase-type distributions and identify the dependence structure from the phase-type representation.

The research should therefore first try to construct a diffusion process given certain multivariate marginal distributions follow a multivariate phase-type distribution or conversely, try to approximate some multivariate phase-type distributions from certain known diffusion models. In doing so, the research might reveal some mappings between the parameters of the diffusion processes and the parameters in the representations of the multivariate marginals. To give a hypothetical example, one could imagine that a diffusion process which eventually converges to a certain state might have phase-type distributed marginals, where the temporal indices of the marginal distributions are closely related to the absorption rates or initial distributions of the Markov processes generating the marginals.

In the long term, the research could produce a data-driven method (based on phase-type distributions) that approximate diffusion process parameters based on empirical marginal distributions. This would allow researchers to analyse time series data and stochastic processes numerically through computer simulations and the theoretical framework of phase-type distributions. Such new methodologies of course also require new developments in estimation and data fitting techniques associated with phase-type distributions, which luckily is already an active area of research.

Alternative reward structures

The previous paragraphs were concerned with new directions for research on MPH* distributions, but one could instead examine distributions arising from considering different reward structures than those leading to the MPH* distributions. Essentially, the reward structure used in MPH* distribution can be given as the mapping

$$r(X_t) = r_{X_t}, \tag{5.1}$$

which simply maps each transient state of the governing Markov process to some constant. However, there is no reason not to consider alternative functions of the governing Markov process such as

$$r(X_t) = r_{X_t} e^{-t}, \tag{5.2}$$

which would imply that the rewards diminish over time. Other reward structures of interest could be dependent on individual sojourn times rather than the temporal index of the process. Using the notation from section 2.4, the random variable with such a reward structure would be constructed as

$$Y = \sum_{i \in E^*} \sum_{j=1}^{N_i} r(i, Z_{ij}), \tag{5.3}$$

where N_i is the number of visits to state i prior to absorption, and Z_{ij} is the sojourn time of the j'th visit to state i. Combining this with the idea from (5.2) could yield the reward structure $r(i, s) = r_i e^{-s}$, and the random variable above would be defined as

$$Y = \sum_{i \in E^*} \sum_{j=1}^{N_i} r_i e^{-Z_{ij}}. \tag{5.4}$$

Alternatively, the reward function could take the form

$$r(i, s) = \int_0^\infty \mathbb{1}\{s \geq t\} r_i e^{-t} dt = \int_0^s r_i e^{-t} dt. \tag{5.5}$$

leading to a random variable

$$Y = \sum_{i \in E^*} \sum_{j=1}^{N_i} \int_0^{Z_{ij}} r_i e^{-t} dt. \tag{5.6}$$

The random variable defined in equation (5.6) differs fundamentally from a phase-type distribution, e.g. in that it can be bounded from above by using an appropriate underlying Markov process. This construction also allows for modelling other distributions such as uniform distributions, which are usually not associated with phase-type distributions. Similarly, the class of distributions that would arise from considering random variables on the form of (5.6) with a sign change in the exponent would be substantially different from the ordinary class of phase-type distributions.

In some cases, introducing such time dependent reward functionals is equivalent with considering certain inhomogeneous phase-type distributions, and it would be interesting to explore the correspondence between these distributions and more complex reward structures. This line of research could further draw inspiration from Yor (2001), where exponential functionals of Brownian motions are considered both from a theoretical standpoint and in the context of pricing financial products. The book by Yor suggests that considering these complex reward structures can indeed yield some useful results.

Existence of Dussauchoy and Berland's distribution for non-integer shape parameters

In the paper constituting chapter 4 of the thesis, we briefly discussed the divisibility of the distribution. As described in the paper, we managed to prove that the distribution is not always infinitely divisible by providing a concrete counterexample. However, it remains to be shown whether the distribution can even exist when the shape parameters take non-integer values. We currently conjecture that the distribution generally cannot exist in these cases based on the notion that mixtures of multivariate gamma distributions are rarely infinitely divisible. This research might also investigate the divisibility of general univariate and multivariate phase-type distributions and try to establish some conditions on the distributions in order to ensure infinite divisibility. Extending the results on the existence of the distribution to include non-integer shape parameters would yield a complete description of the distribution, although the interpretation of the decomposition becomes considerably more complicated when the shape parameters are not integers.

Concluding remarks

The dissertation presents some new results on multivariate phase-type distributions along with several detailed derivations of known results not found elsewhere in the literature. The derivations of the dissertation include technical remarks and additional arguments, which make the proofs more comprehensive. These technicalities of course constitute only minor contributions and generate only few new insights, but they do provide clarifications and specifications in numerous places. The main contributions of the dissertation are naturally those found in the paper of chapter 5 and summarized in section 7 of the paper. The results of the paper establish the multivariate gamma distribution by Dussauchoy and Berland within the framework of multivariate phase-type distributions and provide additional insights into the divisibility of the distribution. As outlined above, the research of the doctoral studies may be continued in a variety of different ways, but the results of the dissertation are perhaps most useful in the search of phase-type representations in order to characterise the different classes of multivariate phase-type and matrix-exponential distributions.

Bibliography

Ahn, S., Ramaswami, V. (2005), Bilateral Phase Type Distributions, Stochastic Models, Volume 21, Issue 2-3, 239-259.

Albrecher, H., Bladt, M. (2019), Inhomogeneous phase-type distributions and heavy tails, Journal of Applied Probability, Volume 56, Issue 4, 1044-1064.

Albrecher, H., Bladt, M., Yslas, J. (2020), Fitting inhomogeneous phase-type distributions to data: the univariate and the multivariate case, Scandinavian Journal of Statistics, Volume 49, Issue 1, 44-77.

Asmussen, S., Bladt, M. (1996), Renewal theory and queueing algorithms for matrix-exponential distributions. In Matrix-Analytic Methods in Stochastic Models, Editors: Chakravarthy, S. R., Alfa, A. S., Marcel Dekker, Volume 183, 313–341.

Asmussen, S., O'Cinneide, C.A. (1998), Matrix-exponential distributions. In Encyclopedia of Statistical Science Update, Editors: Kotz, S., Read, C. B., Banks, D. L., John Wiley and Sons, Volume 2, 435–440.

Asmussen, S. (2003), Applied Probability and Queues, 2nd edition, Springer.

Assaf, D., Langberg, N. A., Savits, T. H., Shaked, M. (1984), Multivariate Phase-Type Distributions, Operations Research, Volume 32, Issue 3, 688-702.

Balakrishnan, N., Lai, C. D. (2009), Continuous Bivariate Distributions, 2nd edition, Springer.

Bean, N. G., Fackrell, M., Taylor, P. (2008), Characterization of Matrix-Exponential Distributions, Stochastic Models, Volume 24, Issue 3, 339-363.

Becker, P. J., Roux, J. J. J. (1981), A Bivariate Extension of the Gamma Distribution, South African Statistical Journal, Volume 15, 1–12.

Bibby, B. M., Skovgaard, I. M., Sørensen, M. (2005), Diffusion-Type Models with Given Marginal Distribution and Autocorrelation Function, Bernoulli, Volume 11, Issue 2, 191-220.

Bladt, M., Nielsen, B. F. (2010), Multivariate Matrix-Exponential Distributions, Stochastic Models, Volume 26, Issue 1, 1-26.

Bladt, M., Nielsen, B. F. (2010), On the Construction of Bivariate Exponential Distributions with an Arbitrary Correlation Coefficient, Stochastic Models, Volume 26, Issue 2, 295-308.

Bladt, M., Nielsen, B. F. (2011), Moment Distributions of Phase Type, Stochastic Models, Volume 27, Issue 4, 651-663.

Bladt, M., Nielsen, B. F., Samorodnitsky, G. (2015), Calculation of ruin probabilities for a dense class of heavy tailed distributions, Scandinavian Actuarial Journal, Issue 7, 573–591.

Bladt, M, Nielsen, B. F. (2017), Matrix-Exponential Distributions in Applied Probability, 1st edition, Springer.

Bladt, M., Yslas, J. (2022), Heavy-tailed phase-type distributions: a unified approach, Extremes 25, 529–565.

Breuer, L., Baum, D. (2005), An Introduction to Queueing Theory, 1st edition, Springer.

Breuer, L. (2016), A Semi-Explicit Density Function for Kulkarni's Bivariate Phase-Type Distribution, Stochastic Models, Volume 32, Issue 4, 632–642.

Carpenter, M., Diawara, N. (2007), A Multivariate Gamma Distribution and its Characterizations, American Journal of Mathematical and Management Sciences, Volume 27, Issue 3-4, 499-507.

Cheriyan, K. C. (1941), A Bi-Variate Correlated Gamma-Type Distribution Function, Journal of the Indian Mathematical Society, Volume 5, 133-144.

Cherubini, U., Durante, F., Mulinacci, S. (2015), Marshall Olkin Distributions - Advances in Theory and Applications, 1st edition, Springer.

Christensen, O. (2010), Functions, Spaces, and Expansions (Mathematical Tools in Physics and Engineering), 1st edition, Birkhäuser.

David, F. N., Fix, E. (1961), Rank Correlation and Regression in a Nonnormal Surface, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, 177-197.

Dehon, M., Latouche, G. (1982), A Geometric Interpretation of the Relations between the Exponential and Generalized Erlang Distributions, Advances in Applied Probability, Volume 14, Issue 4, 885-897.

Downton, F. (1970), Bivariate Exponential Distributions in Reliability Theory, Journal of the Royal Statistical Society. Series B (Methodological), Volume 32, Issue 3, 408-417.

Eagleson, G. K. (1964), Polynomial Expansions of Bivariate Distributions, The Annals of Mathematical Statistics, Volume 35, Issue 3, 1208-1215.

Fackrell, M. W. (2003), Characterization of matrix-exponential distributions, PhD thesis, University of Adelaide.

FDA - U.S. Food and Drug Administration (2018), Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics (Guidance for Industry), *https://www.fda.gov/media/71195/download*: Retrieved May 2022.

Feller, W. (1966), An Introduction to Probability Theory and Applications, 2nd edition, John Wiley and Sons.

Freund, J. E. (1961), A Bivariate Extension of the Exponential Distribution, Journal of the American Statistical Association, Volume 56, Issue 296, 971-977.

Gallier, J. (2011), Geometric Methods and Applications (For Computer Science and Engineering), 2nd edition, Springer.

Gaver, D. P. (1970), Multivariate Gamma Distributions Generated by Mixture, Sankhyā: The Indian Journal of Statistics. Series A, Volume 32, Issue 1, 123-126.

Goodman, G. S., Johansen, S. (1973), Kolmogorov's differential equations for non-stationary, countable state Markov processes with uniformly continuous transition probabilities, Mathematical Proceedings of the Cambridge Philosophical Society, Volume 73, Issue 1, 119-138.

Gupta, A. K., Nadarajah, S. (2006), Sums, products and ratios for Mckay's bivariate gamma distribution, Mathematical and Computer Modelling, Volume 43, Issue 1–2, 185-193.

Hansen, V. L. (2012), Entrance to advanced mathematics (The metric foundation of modern analysis), 4th edition, Rosendahls.

Hu, S., Murphy, T. B., O'Hagan, A. (2019), Bivariate Gamma Mixture of Experts Models for Joint Insurance Claims Modeling, arXiv:1904.04699.

James, A. T. (1955), The Non-Central Wishart Distribution, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, Volume 229, Issue 1178, 364-366.

Johansen, S. (1986), Product Integrals and Markov Processes, CWI Newsletter, Issue 12, 3-13.

Junker, R. G. (2016), Computing the survival function for an important sub-class of multivariate matrix-exponential distributions, Master thesis, Technical University of Denmark.

Kallenberg, O. (2002), Foundations of Modern Probability, 2nd edition, Springer.

Kibble, W. F. (1941), A Two-Variate Gamma Type Distribution, Sankhyā: The Indian Journal of Statistics, Volume 5, Issue 2 (Proceedings of the Indian Statistical Conference 1940), 137-150.

Kiersch, J. L. (2020), Applications of time inhomogeneous Markov processes in insurance mathematics, Master thesis, Technical University of Denmark.

Kotz, S., Balakrishnan, N., Johnson, N.L. (2000), Continuous Multivariate Distributions, Volume 1: Models and Applications, 2nd edition, John Wiley and Sons.

Kulkarni, V. G. (1989), A New Class of Multivariate Phase Type Distributions, Operations Research, Volume 37, Issue 1, 151-158.

Larsen, N. S. (2018), Reward structures for multivariate phase-type distributions, Master thesis, Technical University of Denmark.

Lukacs, E. (1960), Characteristic Functions, 1st edition, C. Griffin and Co.

Mardia, K. V. (1970), Families of Bivariate Distributions, Griffin.

Marshall, A. W., Olkin, I. (1967a), A Multivariate Exponential Distribution, Journal of the American Statistical Association, Volume 62, Issue 317, 30-44.

Marshall, A. W., Olkin, I. (1967b), A Generalized Bivariate Exponential Distribution, Journal of Applied Probability, Volume 4, Issue 2, 291–302.

McKay, A. T. (1934), Sampling from Batches, Supplement to the Journal of the Royal Statistical Society, Volume 1, Issue 2, 207-216.

Meisch, D. (2014), Multivariate phase type distributions - Applications and parameter estimation, PhD thesis, Technical University of Denmark.

Moran, P. A. P. (1967), Testing for Correlation Between Non-Negative Variates, Biometrika, Volume 54, Issue 3/4, 385-394.

Nadarajah, S., Afuecheta, E., Chan, S. (2017), A Compendium of Copulas, Statistica, Volume 77, Issue 4, 279–328.

Navarro, A. C. (2019), Order Statistics and Multivariate Discrete Phase-type Distributions, PhD thesis, Technical University of Denmark.

Navarro, A. C. (2020), Concomitants of order statistics from bivariate phase-type distributions with continuous density functions, Stochastic Models, Volume 36, Issue 4, 574-601.

Nelsen, R. B. (2006), An Introduction to Copulas, 2nd edition, Springer.

Neuts, M. F. (1975), Probability distributions of phase type, Liber Amicorum Prof. Emeritus H. Florin, 173–206.

O'Cinneide, C. A. (1990), Characterization of phase-type distributions, Communications in Statistics, Stochastic Models, Volume 6, Issue 1, 1-57.

O'Cinneide, C. A. (1990), On the Limitations of Multivariate Phase-Type Families, Operations Research, Volume 38, Issue 3, 519–526.

O'Cinneide, C. A. (1993), Triangular order of triangular phase-type distributions, Communications in Statistics, Stochastic Models, Volume 9, Issue 4, 507-529.

Peralta, O. (2021), A Markov jump process associated with the matrix-exponential distribution, arXiv:2103.02722.

Petersen, K. B., Pedersen, M. S. (2012), The Matrix Cookbook, Technical University of Denmark.

Pinksy, M., Karlin, S. (2011), An Introduction to Stochastic Modeling, 4th edition, Elsevier.

Pitman, J. (1993), Probability, 1st edition, Springer.

Prékopa, A., Szántai, T. (1978), A new multivariate gamma distribution and its fitting to empirical streamflow data, Water Resources Research, Volume 14, Issue 1, 19-24.

Raftery, A. E. (1984), A continuous multivariate exponential distribution, Communications in Statistics - Theory and Methods, Volume 13, Issue 8, 947-965.

Raftery, A. E. (1985), Some properties of a new continuous bivariate exponential distribution. Statistics and Decisions, Supplement Issue 2, 53-58.

Ramabhadran, V. K. (1951), A Multivariate Gamma-Type Distribution, Sankhyā: The Indian Journal of Statistics, Volume 11, Issue 1, 45-46.

Ramaswami, V. (2013), A Fluid Introduction to Brownian Motion and Stochastic Integration. In: Matrix-Analytic Methods in Stochastic Models, Editors: Latouche, G., Ramaswami, V., Sethuraman, J., Sigman, K., Squillante, M., Yao, D., Volume 27, 209–225.

Rudin, W. (1974), Real and Complex Analysis, 2nd edition, McGraw-Hill Book Company.

Schilling, R. (2005), Measures, Integrals and Martingales, 1st edition, Cambridge University Press.

Selby, S. M. (1974), Standard Mathematical Tables, 22nd edition, CRC Press.

Sharpe, M. (1988), General Theory of Markov Processes, 1st edition, Academic Press.

Szantai, T. (1986), Evaluation of a special multivariate gamma distribution function, Mathematical Programming Studies, Volume 27, 1-16.

Yera, Y. G., Lillo, R. E., Nielsen, B. F., Ramirez-Cobo, P., Ruggeri, F. (2021), A bivariate two-state Markov modulated Poisson process for failure modeling, Reliability Engineering & System Safety, Volume 208, Article ID 107318.

Yor, M. (2001), Exponential Functionals of Brownian Motion and Related Processes, 1st edition, Springer.

Additional reading

Actuarial mathematics and risk theory

There are several research groups that conduct research on phase-type and matrix-exponential distributions within actuarial sciences and risk theory. Some of the foremost examples include the research groups led by Hansjörg Albrecher at the University of Lausanne and Mogens Bladt at the University of Copenhagen. Søren Asmussen's group at Aarhus University has also made substantial contributions to the fields including considerable theoretical advances.

Asmussen, S., Laub, P. J., Yang, H. (2019), Phase-Type Models in Life Insurance: Fitting and Valuation of Equity-Linked Benefits, Risks, Volume 7, Issue 1.

Badescu, A., Cheung, E., Landriault, D. (2009), Dependent Risk Models with Bivariate Phase-Type Distributions, Journal of Applied Probability, Volume 46, Issue 1, 113-131.

Bladt, M., Nielsen, B. F., Peralta, O. (2019), Parisian types of ruin probabilities for a class of dependent risk-reserve processes, Scandinavian Actuarial Journal, Volume 2019, Issue 1, 32-61.

Biology

Hobolth, A., Siri-Jégousse, A., Bladt, M. (2019), Phase-type distributions in population genetics, Theoretical Population Biology, Volume 127, 16-32.

Hobolth, A., Bladt, M., Andersen, L.N. (2021), Multivariate phase-type theory for the site frequency spectrum, Journal of Mathematical Biology, Volume 83, Article no. 63.

Technology engineering

Albrecher, H., Finger, D., Goffard, P. (2022), Blockchain mining in pools: Analyzing the trade-off between profitability and ruin, Insurance: Mathematics and Economics, Volume 105, 313-335.

Queueing theory

Fackrell, M., Taylor, P., Wang, J. (2021), Strategic customer behavior in an M/M/1 feedback queue, Queueing Systems, Volume 97, 223–259.

He, Q., Zhang, H., Ye, Q. (2018), An M/PH/K queue with constant impatient time, Mathematical Methods of Operations Research, Volume 87, 139–168.

Wu, H., He, Q. (2021), Double-sided queues with marked Markovian arrival processes and abandonment, Stochastic Models, Volume 37, Issue 1, 23-58.

Healthcare technology

Almehdawe, E., Jewkes, B., He, Q. (2016), Analysis and optimization of an ambulance offload delay and allocation problem, Omega, Volume 65, 148-158.

Fackrell, M. (2009), Modelling healthcare systems with phase-type distributions, Health Care Management Science, Volume 12, Article no. 11.

Transportation

Gardner, C. B., Nielsen, S. D., Eltved, M., Rasmussen, T. K., Nielsen, O. A., Nielsen, B. F. (2021), Calculating conditional passenger travel time distributions in mixed schedule- and frequency-based public transport networks using Markov chains, Transportation Research Part B: Methodological, Volume 152, 1-17,

Related distributions

Bladt, M., Esparza, L. J. R., Nielsen, B. F. (2013), Bilateral Matrix-Exponential Distributions. In: Matrix-Analytic Methods in Stochastic Models, Editors: Latouche, G., Ramaswami, V., Sethuraman, J., Sigman, K., Squillante, M., Yao, D., Volume 27, 41–56.

He, Q., Zhang, H. (2007), On Matrix Exponential Distributions, Advances in Applied Probability, Volume 39, Issue 1, 271–292.

Shanthikumar, J. G. (1985), Bilateral phase type distributions, Naval Research Logistics Quarterly, Volume 32, Issue 1, 119–136.