Mogens Bladt, Bo Friis Nielsen

Matrix–exponential methods in Applied Probability

October 25, 2011

Springer

Contents

Preface

This book deals with the theory and application of matrix–analytic methods in the field of Applied Probability. There are only few books available in the field, and our main motivation for writing this book has been to provide an as complete as possible account on the state–of–the–art, which must at the same time give an overview to the researcher in the field as well as an didactic introduction to the new student.

Since the seminal and classic book by Marcel F. Neuts on Matrix–analytic solution in stochastic models, a great amount of new material has emerged, in particular general methods for distribution with a rational Laplace transform such as matrix–exponential distributions and rational arrival processes, statistical methods and even recent developments into the multivariate setting.

We have aimed at writing the book such a way that it essentially should be self–contained for an advanced undergraduate level and up. The book contains the necessary background on stochastic processes from the Poisson processes to Markov processes in general. The emphasis is on the probabilistic method of reasoning the foundation of which is laid out in the introductory chapter on Markov processes.

It is also our hope that the book will appeal to students who are interested in the applications of stochastic processes in general. It is a unique feature of the book that it contains both a thorough theoretical treatment of the matrix–analytic method, that it presents applications to various important areas such as telecommunications, queueing theory, risk theory and finance and at the same time integrate a framework for performing statistical analyses which shall enable the reader to apply the models in practice.

Chapter 1
Rational Laplace transforms of distributions

1.1 Introduction

In this chapter we analyze the structure of distributions which have rational Laplace transforms. We establish the general form of the Laplace transform and show the important characterization that a distribution having a rational Laplace transform is equivalent to its density being a linear combination of the terms in a matrix–exponential. To this end we use a standard trick (though reversed) from theory of differential equation involving the so–called Companion matrix. From this characterization we define the matrix–exponential distributions exactly as those having a rational Laplace transform.

We provide the basic distributional properties relating to matrix–exponential distributions such as density, distribution function and moments. Also considerations concerning minimality of representation is considered using an algebraic approach. The theory of rational functions is by nature closely related to algebra, and we consider further properties involving continued fractions and Hankel matrices. In particular the so–called reduced moments of a matrix–exponential distribution is of great importance concerning the dimensionality problem and also related to the so–called moment problem.

1.2 Basic form of the Laplace transform

Let X be a non–negative random variable with distribution function F. The Laplace transform of X is defined as the function

$$L_X(s) = \mathbb{E}\left(e^{-sX}\right) = \int_0^\infty e^{-sx} dF(x).$$

Since $X \geq 0$ we see that the expression is well defined. The right hand side is the Laplace–Stieltjes transform of F. If X is absolutely continuous and has a density

$f = F'$, then

$$L_X(s) = \int_0^\infty e^{-sx} f(x)dx = \mathscr{L}(f,s),$$

where $\mathscr{L}$ is the Laplace transform of function f. Suppose that X is absolutely continuous and $L_X(s)$ is a rational function, i.e.

$$L_X(s) = \frac{p(s)}{q(s)},$$

where p and q are polynomials

$$\begin{aligned} p(s) &= p_0 s^m + p_1 s^{m-1} + ... + p_m \\ q(s) &= q_0 s^n + q_1 s^{n-1} + ... + q_n. \end{aligned}$$

We shall assume that the coefficients to the leading powers p_0 and q_0 are different from zero. Without loss of generality we may assume that $q_0 = 1$ by dividing through with q_0. By the monotone convergence theorem $L_X(s) \to 0$ as $s \to \infty$ so $m \leq n-1$. This means that the general form of the rational Laplace transform for X may be written as

$$L_X(s) = \frac{p_1 s^{n-1} + ... + p_n}{s^n + q_1 s^{n-1} + ... + q_n}, \tag{1.1}$$

where $p_1, ..., p_{n-1}$ are possibly zero. If L_X is irreducible, i.e. $p(s)$ and $q(s)$ have no common roots, then p_n and q_n must be different from zero. Since $L_X(0) = 1$ we further conclude that $p_n = q_n$.
[1]

1.3 The Companion matrix

Given $L_X(s)$ on the form (1.1), we shall now derive a formula for the density f of X. Multiplying both sides of the expression (1.1) by its denominator we get

$$s^n \mathscr{L}(f,s) + q_1 s^{n-1} \mathscr{L}(f,s) + ... + q_n \mathscr{L}(f,s) = p_1 s^{n-1} + ... + p_n.$$

If we let $f^{(i)}(s)$ denote the i'th derivative of f, then

$$\mathscr{L}(f^{(n)},s) = s^n \mathscr{L}(f,s) - s^{n-1} f(0) - s^{n-2} f'(0) - ... - f^{(n-1)}(0). \tag{1.2}$$

Let $k_1 = f(0)$ and $k_i = q_{i-1} f(0) + q_{i-2} f'(0) + ... + q_1 f^{(i-2)}(0) + f^{(i-1)}(0)$, $i = 2, ..., n$. Then

[1] Insert here: inverse Laplace transform (=density) is infinitely often differentiable, use partial fraction expansion and write the general form of inverse

$$\begin{aligned} &s^n\mathscr{L}(f,s)+q_1s^{n-1}\mathscr{L}(f,s)+...+q_n\mathscr{L}(f,s)\\ &\quad=\mathscr{L}(f^{(n)}+q_1f^{(n-1)}+...+q_nf,s)+k_1s^{n-1}+k_2s^{n-2}+...+k_n. \end{aligned}$$

Hence

$$\mathscr{L}(f^{(n)}+q_1f^{(n-1)}+...+q_nf,s)=(p_1-k_1)s^{n-1}+(p_2-k_2)s^{n-2}+...+p_n-k_n.$$

Consider the Taylor expansion of f about zero, $f(t)=f(0)+f'(0)t+\frac{1}{2}f''(0)t^2+....$ Taking Laplace transform this results in the asymptotic expansion

$$\mathscr{L}(f,s)=\frac{f(0)}{s}+\frac{f'(0)}{s^2}+...,$$

since

$$\mathscr{L}(t^n,s)=\frac{1}{s^{n+1}}\Gamma(n+1). \tag{1.3}$$

This type of asymptotic expansion is rigorously studied in Watson's lemma. Now equating this expression with (1.1) and multiplying through with the denominator we have that

$$p_1s^{n-1}+...+p_n=\left(s^n+q_1s^{n-1}+...+q_n\right)\left(\frac{f(0)}{s}+\frac{f'(0)}{s^2}+...\right).$$

By simple coefficient matching we deduce that

$$\begin{aligned} p_1 &= f(0)=k_1\\ p_2 &= q_1f(0)+f'(0)=k_2\\ p_3 &= q_2f(0)+q_1f'(0)+f''(0)=k_3\\ &... \ ... \ ...\\ p_n &= q_{n-1}f(0)+q_{n-2}f'(0)+...+q_1f^{(n-2)}(0)+f^{(n-1)}(0)=k_n. \end{aligned} \tag{1.4}$$

Hence $p_i-k_i=0$ for $i=1,2,...,n$, which implies that

$$\mathscr{L}(f^{(n)}+q_1f^{(n-1)}+...+q_nf,s)=0.$$

The inverse Laplace transform of the zero–function is again zero, so we conclude that

$$f^{(n)}+q_1f^{(n-1)}+...+q_nf=0. \tag{1.5}$$

This is homogeneous, linear, ordinary differential equation with constant coefficients. Such an equation can be solved in the following way. Define the column vector $\boldsymbol{f}(t)=(f(t),f'(t),...,f^{(n-1)}(t))^t$, where t refers to transpose. Then (1.5) is equivalent to

$$\begin{pmatrix} f'(t) \\ f''(t) \\ f^{(3)}(t) \\ \vdots \\ f^{(n)}(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ 0 & 0 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_n & -q_{n-1} & -q_{n-2} & ... & -q_1 \end{pmatrix} \begin{pmatrix} f(t) \\ f'(t) \\ f''(t) \\ \vdots \\ f^{(n-1)}(t) \end{pmatrix}.$$

The matrix

$$\boldsymbol{C} = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ 0 & 0 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_n & -q_{n-1} & -q_{n-2} & ... & -q_1 \end{pmatrix} \tag{1.6}$$

is called a companion matrix. Then

$$\boldsymbol{f}'(t) = \boldsymbol{C}\boldsymbol{f}(t),$$

the general solution of which is

$$\boldsymbol{f}(t) = e^{\boldsymbol{C}(t-t_0)}\boldsymbol{f}(t_0).$$

Let $t_0 = 0$. Then

$$\boldsymbol{f}(t) = e^{\boldsymbol{C}t}\boldsymbol{f}(0).$$

Let $\boldsymbol{e}_1 = (1,0,0,...,0)^t$. Then the density f can be expressed by

$$f(s) = \boldsymbol{e}_1^t \boldsymbol{f}(s) = \boldsymbol{e}_1^t e^{\boldsymbol{C}s}\boldsymbol{f}(0) \tag{1.7}$$

The initial condition $\boldsymbol{f}(0)$ is found by (1.4),

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ ... \\ p_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ q_1 & 1 & 0 & ... & 0 \\ q_2 & q_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ q_{n-1} & q_{n-2} & q_{n-3} & ... & 1 \end{pmatrix} \begin{pmatrix} f(0) \\ f'(0) \\ f''(0) \\ ... \\ f^{(n-1)}(0) \end{pmatrix}.$$

Let

$$\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ q_1 & 1 & 0 & ... & 0 \\ q_2 & q_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ q_{n-1} & q_{n-2} & q_{n-3} & ... & 1 \end{pmatrix}.$$

Hence $\boldsymbol{f}(0) = \boldsymbol{M}^{-1}\boldsymbol{p}$, where $\boldsymbol{p} = (p_1, p_2, ..., p_n)^t$. Thus

$$\begin{aligned} f(s) &= \boldsymbol{e}_1^t e^{\boldsymbol{C}s}\boldsymbol{M}^{-1}\boldsymbol{p} \\ &= \boldsymbol{e}_1^t \boldsymbol{M}^{-1} e^{(\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}^{-1})s}\boldsymbol{p} \end{aligned}$$

$$= \boldsymbol{e}_1^t e^{\tilde{\boldsymbol{C}}s}\boldsymbol{p},$$

where

$$\tilde{\boldsymbol{C}} = \begin{pmatrix} -q_1 & 1 & 0 & \dots & 0 \\ -q_2 & 0 & 1 & \dots & 0 \\ -q_3 & 0 & 0 & \dots. & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_n & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Transposing and reordering the states in reverse order yields the following representation.

Theorem 1.1. *A distribution with rational Laplace transform (1.1) has a density on the form*

$$f(s) = \boldsymbol{p}_r^t e^{\boldsymbol{C}s}\boldsymbol{e}_n, \tag{1.8}$$

where $\boldsymbol{p}_r = (p_n, p_{n-1}, ..., p_1)^t$, $\boldsymbol{e}_n = (0,0,....,0,1)$ *and* $\boldsymbol{C}$ *is the companion matrix (1.6).*

The determinant of the companion matrix equals q_n. If the rational function is irreducible, then $q_n \neq 0$, and hence the companion matrix is non–singular. If the Laplace transform is irreducible we then have that

$$1 = \int_0^\infty f(x)dx = \boldsymbol{e}_1^t \int_0^\infty e^{\boldsymbol{C}x}dx\boldsymbol{f}(0) = -\boldsymbol{e}_1^t\boldsymbol{C}^{-1}\boldsymbol{f}(0).$$

Theorem 1.2. *The n'th moment of distribution with rationbal Laplace transform (1.1) is given by*

$$\mu_n = n!\boldsymbol{e}_1^t(-\boldsymbol{C})^{-(n+1)}\boldsymbol{f}(0) \tag{1.9}$$

$$= n!\boldsymbol{p}_r^t(-\boldsymbol{C})^{-(n+1)}\boldsymbol{e}_n. \tag{1.10}$$

Proof. Follows by repeated partial integeration of $\int_0^\infty x^n f(x)dx$, where f is given by (1.7) or (1.8). □

1.4 Poles

Theorem 1.3. *Let X be a non–negative random variable which has a distribution with a rational Laplace transform. Let* $\boldsymbol{C}$ *denote the companion matrix in (1.6). Then the real parts of the eigenvalues of* $\boldsymbol{C}$ *are all strictly negative.*

Proof. Write $\boldsymbol{C}$ on Jordan normal form, $\boldsymbol{C} = \boldsymbol{PJP}^{-1}$, where $\boldsymbol{J}$ is a block–diagonal matrix matrix

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{J}_3 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{J}_n \end{pmatrix}$$

for some n, and

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ 0 & 0 & \lambda_i & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix},$$

the dimension of which equals the multiplicity of the eigenvalue λ_i.

By (1.8), we may then write the density of X as

$$\begin{aligned} f(x) &= \boldsymbol{p}_r^t \exp(\boldsymbol{C}x)\boldsymbol{e}_n \\ &= \boldsymbol{p}_r^t \boldsymbol{P} \exp(\boldsymbol{J}x)\boldsymbol{P}^{-1}\boldsymbol{e}_n \\ &= \boldsymbol{p}_r^t \boldsymbol{P} \begin{pmatrix} \exp(\boldsymbol{J}_1 x) & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \exp(\boldsymbol{J}_2 x) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \exp(\boldsymbol{J}_3 x) & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \exp(\boldsymbol{J}_n x) \end{pmatrix} \boldsymbol{P}^{-1}\boldsymbol{e}_n \end{aligned}$$

where

$$\exp(\boldsymbol{J}_i x) = \begin{pmatrix} e^{-\lambda_i x} & x e^{-\lambda_i x} & \frac{x^2}{2} e^{-\lambda_i x} & \dots & \frac{x^{n_i}}{n_i!} e^{-\lambda_i x} \\ 0 & e^{-\lambda_i x} & x e^{-\lambda_i x} & \dots & \frac{x^{n_i-1}}{(n_i-1)!} e^{-\lambda_i x} \\ 0 & 0 & e^{-\lambda_i x} & \dots. & \frac{x^{n_i-2}}{(n_i-2)!} e^{-\lambda_i x} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & e^{-\lambda_i x} \end{pmatrix}. \tag{1.11}$$

Since $f(x) \to 0$ as $x \to \infty$ we conclude that all eigenvalues must have negative. real parts. □

The eigenvalues of $\boldsymbol{C}$ are also the roots of the denominator polynomial and hence the poles of the Laplace transform. Hence all poles must have negative real parts. We also obtain the following important result from the proof of Theorem 1.3.

Corollary 1.1. *A distribution is a matrix–exponential distribution if and only if it has a density f on the form*

$$f(x) = \sum_{j=1}^{q_1} r_j x^j e^{\rho_j x} + \sum_{j=1}^{q_2} s_j x^j \cos(a_j x) e^{\sigma_j x} + \sum_{j=1}^{q_3} t_j x^j \sin(b_j x) e^{\tau_j x}, \tag{1.12}$$

where $q_1, q_2, q_3 \geq 0$ (where $q_i = 0$ indicates that the corresponding sum is zero) and all constants $r_j, s_j, t_j, a_j, b_j, \rho_j, \sigma_j, \tau_j$ are real–valued.

Proof. The result is a direct consequence of (1.11) and the possibly complex eigenvalues. □

Definition 1.1. A dominant pole is a pole which has the largest real part among all the poles and which has the largest multiplicity among the poles which have the largest real parts.

Theorem 1.4. *Let X be random variable with a rational Laplace transform*

$$L_X(s) = \frac{p(s)}{q(s)}.$$

Then L_X has a dominant pole which is real.

Proof. Let $r_j = -\alpha + \mathrm{i}w_j$, $j = 1, 2, ..., 2k$ denote the complex poles which has a maximum real part among all the poles and let r_0 denote a real pole of maximum real part. The complex poles come in conjugate pairs so there is an even number of these. We must show that the multiplicity of the real pole r_0 n_0 is at least that of the maximum of the multiplicities $n_1, ..., n_k$ of the complex poles $r_1, ..., r_{2k}$. Notice that we only define k different multiplicities as the conjugate pairs have the same multiplicity.

Assume the contrary, that there exists an n_i, $i = 1, ..., k$ such that $n_i > n_0$. Without loss of generality we may assume that this n_i is n_1 and that this n_1 is the largest of its kind. We shall prove that the density then assume negative values at certain points.

As stated earlier the roots come in conjugate pairs, so if the multiplicity of one root $-\alpha + \mathrm{i}\beta$ is n, then so is the multiplicity of its conjugate $-\alpha - \mathrm{i}\beta$ and there is a contribution of $(s+\alpha-\mathrm{i}\beta)^n(s+\alpha+\mathrm{i}\beta)^n = ((s+\alpha)^2+\beta^2)^n$ in the denominator polynomial. Hence we may write the Laplace transform on the form

$$L_X(s) = \frac{p(s)}{((s+\alpha)^2+w_1^2)^{n_1} \cdots ((s+\alpha)^2+w_k^2)^{n_k}(s+\alpha)^{n_0} q_1(s)},$$

where $q_1(s)$ is a polynomial with roots which have real parts which are all smaller than $-\alpha$. Then

$$\tilde{L}_X(s) = L_X(s-\alpha) = \frac{p(s-\alpha)}{(s^2+w_1^2)^{n_1} \cdots (s^2+w_k^2)^{n_k} s^{n_0} q_1(s-\alpha)},$$

where $q_1(s-\alpha)$ still is polynomial with roots which have negative real parts.

Using Exercises 1.6 and 1.7 together with partial fraction expansion, we see that the inverse Laplace transform $\tilde{f}$ of $\tilde{L}_X$ may be written as

$$\tilde{f}(s) = \sum_{i=1}^{k} (P_i(s)\sin(w_i s) + Q_i(s)\cos(w_i s)) + P_0(s) + \phi(s),$$

where P_i, Q_i are polynomials the order of which is at most $n_i - 1$, and $\phi(s)$ is function which tends to zero exponentially fast.

Since n_1 is the dominant exponent, we can write

$$\tilde{f}(s) = s^{n_!-1}\left[\sum_{i=1}^{k}\left(p_i(s)\sin(w_i s) + q_i(s)\cos(w_i s)\right) + p_0(s) + \phi(s)\right],$$

where then the functions p_i, q_i are bounded functions, and the limits

$$\bar{p}_i = \lim_{s\to\infty}\frac{p_i(s)}{s^{n_1-1}}$$
$$\bar{q}_i = \lim_{s\to\infty}\frac{q_i(s)}{s^{n_1-1}}$$

exists (they are either constant zero or non–zero). At least one of the $\bar{p}_i$ or $\bar{q}_i$ is different from zero. Now consider the "inside limit" of the sum term,

$$g(s) = \sum_{i=1}^{k}\left[\bar{p}_i\sin(w_i s) + \bar{q}_i\cos(w_i s)\right].$$

This is a periodic function, which changes sign (at an infinite number of points). Since

$$\frac{\tilde{f}(s)}{s^{n_1-1}} = g(s) + o(1)$$

it then follows that the function $\tilde{f}(s)$ also changes sign. Since

$$\tilde{L}_X(s) = \int_0^\infty e^{-(s-\alpha)x} f(x)dx = \int_0^\infty e^{-sx}\left(e^{\alpha x} f(x)\right)dx,$$

then $\tilde{f}(x) = e^{\alpha x} f(x)$ we conclude the that f itself also must change sign. □

1.5 The algebra of rational Laplace transforms

There is a lot of algebra which deals with rational functions. In fact, at the beginning of the 20th century, Oskar Perron in his foreword to a book on algebra stated that modern algebra essentially is the study of rational functions. In this section we derive a number of important results dealing with minimality and representations of distributions with a rational Laplace transform.

Definition 1.2. A distribution on the positive real axis is called matrix–exponential if it has a possible atom at zero, and an absolutely continuous part represented by a density f which can be written on the form

$$f(x) = \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}$$

for some row vector $\boldsymbol{\alpha}$, column vector $\boldsymbol{s}$ and matrix $\boldsymbol{S}$. If X is a random variable with such a distribution we write $X \sim \mathrm{ME}_p(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$, where p is the dimension of the vectors involved. The triple $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is called a representation for the distribution and p its order.

Note 1.1. A possible atom at zero is not explicitly specified in the representation above but will be implicit from $\int_0^\infty \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} dx < 1$, in which case we may define $\alpha_{p+1} = 1 - \int_0^\infty \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} dx$ and interpret this terms as the initial probability of initiating in the absorbing state, $p+1$.

Theorem 1.5. *A random variable X is matrix–exponentially distributed if and only if it has a rational Laplace transform.*

Proof. If X has a rational Laplace transform, then its density may be written on the form (1.8). On the other hand, if $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is a representation of a matrix–exponential distribution, then $\boldsymbol{\alpha} \exp(\boldsymbol{S}x)\boldsymbol{s} \to 0$ as $x \to \infty$, and hence the eigenvalues of $\boldsymbol{S}$ are strictly negative. Hence $\boldsymbol{S}$ is non–singular, and the characteristic polynomial $\det(\lambda \boldsymbol{I} - \boldsymbol{S})$ is zero exactly at the eigenvalues. Hence $(s\boldsymbol{I} - \boldsymbol{S})^{-1}$ exists except for s assuming the values of of the eigenvalues, at which point it is singular. The Laplace transform is given by

$$\begin{aligned} L_X(s) &= \mathbb{E}(e^{-sX}) \\ &= \int_0^\infty e^{-sx} \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} dx \\ &= \int_0^\infty \boldsymbol{\alpha} e^{-(s\boldsymbol{I}-\boldsymbol{S})x} \boldsymbol{s} dx \\ &= \boldsymbol{\alpha} \left[-(s\boldsymbol{I} - \boldsymbol{S})^{-1} e^{-(s\boldsymbol{I}-\boldsymbol{S})x} \right]_0^\infty \boldsymbol{s} \\ &= \boldsymbol{\alpha} (s\boldsymbol{I} - \boldsymbol{S})^{-1} \boldsymbol{s}. \end{aligned}$$

Thus the Laplace transform is obtained by a linear combination of the elements in the inverse matrix of $s\boldsymbol{I} - \boldsymbol{S}$, which is then clearly a rational function since we may obtain the inverse by applying the same sequence of elementary operations to the identity matrix as to which we transform $s\boldsymbol{I} - \boldsymbol{S}$ into the identity matrix. □

Also, the eigenvalues of the matrix $\boldsymbol{S}$ in any representation of a matrix–exponential distribution must be strictly negative. This follows from an argument identical to that of Theorem 1.3. The eigenvalues of $\boldsymbol{S}$ are also the poles of the Laplace transform $\boldsymbol{\alpha} (s\boldsymbol{I} - \boldsymbol{S})^{-1} \boldsymbol{s}$.

Theorem 1.6. *Let $X \sim ME_p(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Then its cumulative distribution function F is given by*

$$F(x) = 1 + \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{S}^{-1} \boldsymbol{s},$$

the moment generating function by

$$M_X(s) = \boldsymbol{\alpha}(-s\boldsymbol{I} - \boldsymbol{S})^{-1} \boldsymbol{s},$$

and its n'th moment by

$$\mu_n = \mathbb{E}(X^n) = n!\boldsymbol{\alpha}(-\boldsymbol{S})^{-(n+1)}\boldsymbol{s}.$$

Proof. Since $\boldsymbol{S}$ is invertible we get that

$$1 = \int_0^\infty \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}dx = \boldsymbol{\alpha}\left[\boldsymbol{S}^{-1}e^{\boldsymbol{S}x}\right]_0^\infty \boldsymbol{s} = -\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s}$$

so that

$$\begin{aligned} F(x) &= \int_0^x \boldsymbol{\alpha} e^{\boldsymbol{S}y}\boldsymbol{s}dy \\ &= \boldsymbol{\alpha}\left[e^{\boldsymbol{S}y}\boldsymbol{S}^{-1}\boldsymbol{s}\right]_0^x \boldsymbol{s} \\ &= \boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{S}^{-1}\boldsymbol{s} - \boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s} \\ &= 1 + \boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{S}^{-1}\boldsymbol{s}. \end{aligned}$$

The result on the moment generating function is trivial, while the n'th moment is easily found by repeated partial integration of $\int_0^\infty x^n \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}dx$. □

Definition 1.3. A representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is called minimial if the vectors and matrix involved have the lowest possible dimension.

If X has a Laplace transform

$$L_X(s) = \frac{p_1 s^{n-1} + ... + p_n}{s^n + q_1 s^{n-1} + ... + q_n},$$

and the latter being irreducible (numerator and denominator do not have any common roots), then the lowest possible dimension is n, and the representation $(\boldsymbol{p}_r, \boldsymbol{C}, \boldsymbol{e}_n)$ is indeed a minimal representation. So one method for checking whether a given representation is minimal is calculating $\boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}$ and reducing to an irreducible rational function. If then the dimension n of the rational function coincide with the dimension p of the representation, then the representation is minimal.

In the following we shall provide some other criteria for minimality through the analysis of certain related vectorspaces. To this end we consider an arbitrary representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ of a matrix–exponential distribution, and define the following vector spaces

$$\begin{array}{ll} R_p = \text{span}\{\boldsymbol{s}, \boldsymbol{S}\boldsymbol{s}, \boldsymbol{S}^2\boldsymbol{s}, ..., \boldsymbol{S}^{p-1}\boldsymbol{s}\} & L_p = \text{span}\{\boldsymbol{s}, \boldsymbol{\alpha}\boldsymbol{S}, \boldsymbol{\alpha}\boldsymbol{S}^2, ..., \boldsymbol{\alpha}\boldsymbol{S}^{p-1}\} \\ R_\infty = \text{span}\{\boldsymbol{s}, \boldsymbol{S}\boldsymbol{s}, \boldsymbol{S}^2\boldsymbol{s}, ...\} & L_\infty = \text{span}\{\boldsymbol{s}, \boldsymbol{\alpha}\boldsymbol{S}, \boldsymbol{\alpha}\boldsymbol{S}^2, ...\} \\ R_e = \text{span}\{e^{\boldsymbol{S}x}\boldsymbol{s} : x \geq 0\} & L_e = \text{span}\{\boldsymbol{\alpha}e^{\boldsymbol{S}x} : x \geq 0\} \end{array}$$

Theorem 1.7. *$R_p = R_\infty = R_e$ and $L_p = L_\infty = L_e$.*

Proof. The characteristic polynomial d of S is given by $d(s) = \det(s\boldsymbol{I} - \boldsymbol{S})$. The Cayley–Hamilton theorem states that $d(\boldsymbol{S}) = 0$. Hence $\boldsymbol{S}^p$ is a linear combination of the powers $\boldsymbol{S}^i$, $0 = 1, 2, ..., p-1$. Hence $R_p = R_\infty$ and $L_p = L_\infty$.

Next we prove that $R_e \subseteq R_\infty$. Let $y \in R_e$. Thus there is a $x \geq 0$ such that $y = e^{\boldsymbol{S}x}$. Now $\boldsymbol{S}^i\boldsymbol{s} \in R_\infty$ for all $i = 0, 1, 2, ...$ Since R_∞ is a vector space, $\sum_{i=1}^m \boldsymbol{S}^i\boldsymbol{s}x^i/i! \in R_\infty$, and since R_∞ is a finite–dimensional vector–space, it is closed and hence the limit $\exp(\boldsymbol{S}x)\boldsymbol{s} = \lim_{m\to\infty}\sum_{i=1}^m \boldsymbol{S}^i\boldsymbol{s}x^i/i! \in R_\infty$.

For the converse $R_\infty \subseteq R_e$, we proceed as follows. Letting $x = 0$, we conclude that $\boldsymbol{s} \in R_e$. We also know that $\exp(\boldsymbol{S}x)\boldsymbol{s} \in R_e$ (by definition), and since R_e is a vector space, $\frac{1}{x}(\exp(\boldsymbol{S}x)\boldsymbol{s} - \boldsymbol{s}) \in R_e$ for all $x \geq 0$. By closedness (we know that R_e is a sub–space of the finite–dimensional vector–space R_∞ and hence itself finite dimensional), $\boldsymbol{S}\boldsymbol{s} = \lim_{x\downarrow 0}\frac{1}{x}(\exp(\boldsymbol{S}x)\boldsymbol{s} - \boldsymbol{s}) \in R_e$. Using the expansion

$$e^{\boldsymbol{S}x} = \boldsymbol{I} + \boldsymbol{S}x + \boldsymbol{S}^2\frac{x^2}{2} + \boldsymbol{S}^3\frac{x^3}{3!} + ...$$

we see that

$$2\frac{e^{\boldsymbol{S}x} - \boldsymbol{S}x - \boldsymbol{I}}{x^2} = \boldsymbol{S}^2 + \boldsymbol{S}^3\frac{x}{3} + ...$$

so

$$\lim_{x\downarrow 0} 2\frac{e^{\boldsymbol{S}x}\boldsymbol{s} - \boldsymbol{S}\boldsymbol{s}x - \boldsymbol{s}}{x^2} = \boldsymbol{S}^2\boldsymbol{s} \in R_e.$$

Continuing this way we conclude that $\boldsymbol{S}^i\boldsymbol{s} \in R_e$ for all i, hence $R_\infty \subseteq R_e$. The case $L_\infty = L_e$ is similar. □

The above vector–spaces plays an important role concerning minimality of representations. Basically the vector–spaces must have full dimension in order for a representation of amatrix–exponential distribution to be minimal. We proceed towards this result in a number of steps.

Definition 1.4. A representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is said to have property (R) if $\dim(R_p) = p$ and property (L) if $\dim(L_p) = p$.

Theorem 1.8. *Let $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ be two representations of the same matrix–exponential distribution. If the representation has the property (R), then $\boldsymbol{\alpha} = \boldsymbol{\beta}$. If the representation does not have property (R), then there exists a $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$.*

Similarly, if $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ and $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{t})$ are representation of the same matrix–exponential distribution. If the representation has property (L), then $\boldsymbol{t} = \boldsymbol{s}$, while if (L) fails there exists a $\boldsymbol{t} \neq \boldsymbol{s}$.

Proof. That $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ and $(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ are representation for the same distribution means equality of measures (distribution functions),

$$\int_A \boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}dx = \int_A \boldsymbol{\beta}e^{\boldsymbol{S}x}\boldsymbol{s}dx$$

for all measurable A, and hence

$$\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s} = \boldsymbol{\beta}e^{\boldsymbol{S}x}\boldsymbol{s} \ \ a.s.$$

By continuity, $\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s} = \boldsymbol{\beta}e^{\boldsymbol{S}x}\boldsymbol{s}$ for all x. Hence $(\boldsymbol{\alpha} - \boldsymbol{\beta})e^{\boldsymbol{S}x}\boldsymbol{s} = 0$ for all x, which means that $\boldsymbol{\alpha} - \boldsymbol{\beta}$ is orthogonal to $R_e = R_p = \mathbb{R}^p$, implying $\boldsymbol{\alpha} = \boldsymbol{\beta}$. If (R) fails, then

$\dim(R_p) = d < p$, and there exists a $\boldsymbol{\gamma} \neq \mathbf{0}$ which is orthogonal to R_p and hence to R_e. Thus $\boldsymbol{\gamma} e^{\boldsymbol{S}x}\boldsymbol{s} = 0$ for all x, and $(\boldsymbol{\gamma} + \boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is a representation for the matrix–exponential distribution different from $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. The case (L) is similar. □

The next result show that if property (R) or (L) fails, then it is possible to reduce the dimension of the representation.

Theorem 1.9. *Let $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ be a representation of a matrix–exponential distribution for which property (R) does not hold. Let $d = dim(R_p) < p$. Let $\boldsymbol{r}_1, ..., \boldsymbol{r}_d$ be a basis for R_p and let $\boldsymbol{A}$ be the $p \times d$ matrix which columns consists of the $\boldsymbol{r}_i$s. Let $\boldsymbol{B}$ denote the change of basis transformation which takes $\boldsymbol{r}_1, ..., \boldsymbol{r}_d$ into $\boldsymbol{e}_1, ..., \boldsymbol{e}_d$, the usual Euclidian basis. Then*

$$(\boldsymbol{\alpha}_d, \boldsymbol{S}_d, \boldsymbol{s}_d) = (\boldsymbol{\alpha}\boldsymbol{A}, \boldsymbol{B}\boldsymbol{S}\boldsymbol{A}, \boldsymbol{B}\boldsymbol{s})$$

is a d–dimensional representation of the same matrix–exponential distribution.

Proof. Let $\boldsymbol{y} \in R_p$. Then $\boldsymbol{y} = \sum_{i=}^{d} \theta_i \boldsymbol{r}_i$ and

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{B}\boldsymbol{y} &= \boldsymbol{A}\boldsymbol{B}\left(\sum_{i=}^{d} \theta_i \boldsymbol{r}_i\right) \\
&= \boldsymbol{A}\sum_{i=1}^{d} \theta_i \boldsymbol{B}\boldsymbol{r}_i \\
&= \boldsymbol{A}\sum_{i=1}^{d} \theta_i \boldsymbol{e}_i \\
&= \boldsymbol{y}.
\end{aligned}$$

Thus $\boldsymbol{A}\boldsymbol{B}\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in R_p$. Then

$$\begin{aligned}
\boldsymbol{\alpha}_d \boldsymbol{S}_d^n \boldsymbol{s}_d &= \boldsymbol{\alpha}\boldsymbol{A}(\boldsymbol{B}\boldsymbol{S}\boldsymbol{A})^n \boldsymbol{B}\boldsymbol{s} \\
&= \boldsymbol{\alpha}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{S})^n \boldsymbol{A}\boldsymbol{B}\boldsymbol{s} \\
&= \boldsymbol{\alpha}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{S})^n \boldsymbol{s} \quad (\boldsymbol{s} \in R_p) \\
&= \boldsymbol{\alpha}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{S})^{n-1} \boldsymbol{A}\boldsymbol{B}\boldsymbol{S}\boldsymbol{s} \\
&= \boldsymbol{\alpha}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{S})^{n-1} \boldsymbol{S}\boldsymbol{s} \quad (\boldsymbol{S}\boldsymbol{s} \in R_p) \\
&= \ldots \\
&= \boldsymbol{\alpha}\boldsymbol{S}^n \boldsymbol{s},
\end{aligned}$$

and hence

$$\boldsymbol{\alpha}_d e^{\boldsymbol{S}_d x}\boldsymbol{s}_d = \sum_{n=0}^{\infty} \frac{x^n}{n!} \boldsymbol{\alpha}_d \boldsymbol{S}_d^n \boldsymbol{s}_d = \sum_{n=0}^{\infty} \frac{x^n}{n!} \boldsymbol{\alpha}\boldsymbol{S}^n \boldsymbol{s} = \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}.$$

□

Theorem 1.10. *Let $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ be a representation of a matrix–exponential distribution for which property (L) does not hold. Let $d = dim(L_p) < p$. Let $\boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_d$ be a basis for L_p and let $\boldsymbol{A}$ denote the $d \times p$ matrix the rows of which consists of the $\boldsymbol{\ell}_i$'s. Let $\boldsymbol{B}$ be the change of basis transformation such that $\boldsymbol{\ell}_i \boldsymbol{B} = \boldsymbol{e}_i^t$ for $i = 1, ..., d$. Then*

$$(\boldsymbol{\alpha}_d, \boldsymbol{S}_d, \boldsymbol{s}_d) = (\boldsymbol{\alpha}\boldsymbol{B}, \boldsymbol{A}\boldsymbol{S}\boldsymbol{B}, \boldsymbol{A}\boldsymbol{s})$$

is a d–dimensional representation of the same matrix–exponential distribution.

Proof. The proof si similar to the above and left as an exercise. □

Later in the book the so–called residual life operator will play an important role both the treatment of Rational Arrival Processes (RAP) as well as for the more intrisic analyses of phase–type distributions. At this point, the residual life distribution is now introduced and will be used in to prove a main result on minimality.

Definition 1.5. Let X be a non–negative random varible with distribution function B. Then we let B_y denote the distribution function of the conditional distribution of $X - y$ given $X > y$.

It is clear that $B_y(x) = \frac{B(x+y)-B(y)}{1-B(y)}$. If B is absolutely continuous with density b, we then have that

$$b_y(x) = B_y(dx) = \frac{b(y+x)}{1-B(y)}.$$

Theorem 1.11. *If B is matrix–exponentially distributed with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ then the residual life time is again matrix–exponentially distributed with representation $(\boldsymbol{\alpha}_y, \boldsymbol{S}, \boldsymbol{s})$, where*

$$\boldsymbol{\alpha}_y = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}y}}{\boldsymbol{\alpha} e^{\boldsymbol{S}y}(-\boldsymbol{S})^{-1}\boldsymbol{s}} \in L_e.$$

Proof. It is immediate that

$$b_y(x) = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}(y+x)}\boldsymbol{s}}{1-B(y)} = \boldsymbol{\alpha}_y e^{\boldsymbol{S}x}\boldsymbol{s}$$

where

$$\boldsymbol{\alpha}_y = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}y}}{\int_y^\infty \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s}dx} = \frac{\boldsymbol{\alpha} e^{\boldsymbol{S}y}}{\boldsymbol{\alpha} e^{\boldsymbol{S}y}(-\boldsymbol{S})^{-1}\boldsymbol{s}}.$$

That $\boldsymbol{\alpha}_y \in L_e$ follows essentially from the defintion of L_e. □

Now consider the following complex vector space of distributions (measures),

$$V_B = \text{span}\{B_x : x \geq 0\}. \tag{1.13}$$

An element of $y \in V_B$ can then be written as $y = \sum_{i=1}^n c_i B_{x_i}$ for some n, $c_i \in \mathbb{C}$, $i = 1, ..., n$ and $X_1, ..., x_n \geq 0$, which is obviously not a measure itself, but is often considered in the measure theoretic literature as complex signed measures. Let $B_{(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})}$ denote the distribution (function) of a matrix–exponenital distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Then $B_{x_i} = B_{(\boldsymbol{\alpha}_i,\boldsymbol{S},\boldsymbol{s})}$ for some $\boldsymbol{\alpha}_i \in L_e$ so if $B \sim \text{ME}_p(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$, then

$$V_B = \text{span}\{F \sim \text{ME}_p(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s}) : \beta \in L_e\}.$$

Let q denote the dimension of V_B and let $v_1, ..., v_q$ be a basis for V. The basis elements may be taken as distributions, which are then matrix–exponential. If $y \in V_B$ then $y = \sum_{i=1}^{q} c_i v_i$. But v_i is matrix–exponential with some representation $(\boldsymbol{\beta}_i, \boldsymbol{S}, \boldsymbol{s})$. Thus we may write

$$y = \sum_{i=1}^{q} c_i B_{(\boldsymbol{\beta}_i, \boldsymbol{S}_i, \boldsymbol{s}_i)} \tag{1.14}$$

where $\boldsymbol{\beta}_i \in L_e$, $i = 1, ..., q$. From this we conclude that $q \leq \dim(L_e)$ since all the $\boldsymbol{\beta}_i$'s can be written in terms of $\dim(L_e)$ vectors and then so can y. Hence we have proved that for a matrix–exponential distribution B,

$$\dim(V_B) \leq \dim(L_e) \tag{1.15}$$

We now state and prove a characterization of minimality.

Theorem 1.12. *A representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ is minimal if and only if the representation satisfy both properties (R) and (L).*

Proof. To prove that minimality implies both properties (R) and (L), we assume contrarily that at least one of them fails. Then we may apply Theorem 1.9 or Theorem 1.10 and find a representation of a stricly lower order, hence minimality cannot hold.

Now assume that both properties (R) and (L) hold. Let the order of the representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ be p. Then $\dim(V_B) \leq \dim(L_e) \leq p$. If property (R) holds, then by Theorem 1.8, the $\boldsymbol{\beta}_i$'s in (1.14) are unique. Then $\dim(V_B) = \dim(L_e)$ and by property (L) $\dim(L_e) = p$. The vector space V_B does not depend on the representation of the distribution but only on the corresponding form of the distribution function (which is the same for all representations), the equality $\dim(V_B) = \dim(L_e) = p$ holds for all representations for which properties (R) and (L) hold, in particular for minimal representations. If p_0 is the order of a minimal representation we hence conclude that $p = \dim(V_B) = \dim(L_e) = p_0$. □

Corollary 1.2. *Let $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ be a p–dimensional representation of a matrix–exponential distribution with corresponding vector spaces R_p and L_p as previously defined. Then there is a minimal representation of order $p_0 = \min(dim(R_p), dim(L_p)) \leq p$.*

Minimal representation can be obtained by applying Theorem 1.9 if $\dim(R_p) \leq \dim(L_p)$ or Theorem 1.10 if $\dim(R_p) \geq \dim(L_p)$.

Example 1.1. Consider the matrix–exponential distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$, where

$$\boldsymbol{\alpha} = (1\ 0\ 0),\ \boldsymbol{S} = \begin{pmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 0 & 0 & 1 \end{pmatrix},\ \boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then the order of the representation is $p = 3$,

$$R_3 = \text{span}\{\boldsymbol{s}, \boldsymbol{S}\boldsymbol{s}, \boldsymbol{S}^2\boldsymbol{s}\} = \text{span}\{\begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}-2\\-2\\1\end{pmatrix}\}$$
$$L_3 = \text{span}\{\boldsymbol{\alpha}, \boldsymbol{\alpha}\boldsymbol{S}, \boldsymbol{\alpha}\boldsymbol{S}^2\} = \text{span}\{(1\ 0\ 0), (-3\ 2\ 1), (11\ -10\ -2)\}.$$

Forming the matrices $\boldsymbol{M}_1 = (\boldsymbol{s}, \boldsymbol{S}\boldsymbol{s}, \boldsymbol{S}^2\boldsymbol{s})$ and $\boldsymbol{M}_2 = \begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}\boldsymbol{S}\\ \boldsymbol{\alpha}\boldsymbol{S}^2\end{pmatrix}$, we see that $\det(\boldsymbol{M}_1) = 0$ and $\det(\boldsymbol{M}_2) = 6 \neq 0$ while the order of the largest sub–determinant of $\boldsymbol{M}_1$ different from zero is 2. Hence $\dim(R_3) = 2$ and $\dim(L_2) = 3$. Hence the matrix–exponential distribution of minmimal order is 2.

Property (R) is not satisfied, and we may apply Theorem 1.9 in order to generate a 2 dimensional representation. We take $\boldsymbol{r}_1 = \boldsymbol{s}, \boldsymbol{r}_2 = \boldsymbol{S}\boldsymbol{s}$ as a basis for R_3. Then let

$$\boldsymbol{A} = \begin{pmatrix}0 & 1\\ 0 & 1\\ 1 & -1\end{pmatrix}.$$

Then we solve for a matrix 2×3 matrix $\boldsymbol{B}$ which must satisfy

$$\boldsymbol{B}\boldsymbol{r}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \quad \boldsymbol{B}\boldsymbol{r}_2 = \begin{pmatrix}0\\1\end{pmatrix}.$$

This is a fairly simple exercise in solving an equation system with an infinite number of solutions, and we may e.g. choose

$$\boldsymbol{B} = \begin{pmatrix}1\ 0\ 1\\ 1\ 0\ 0\end{pmatrix}.$$

Thus we find a minimal representation given by $(\boldsymbol{\alpha}_2, \boldsymbol{S}_2, \boldsymbol{s}_2)$ where

$$\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}\boldsymbol{S} = (0\ 1), \quad \boldsymbol{S}_2 = \boldsymbol{B}\boldsymbol{S}\boldsymbol{A} = \begin{pmatrix}0 & -1\\ 1 & -2\end{pmatrix}, \quad \boldsymbol{s}_2 = \boldsymbol{B}\boldsymbol{s} = \begin{pmatrix}1\\0\end{pmatrix}.$$

A minimal representation can also be found by a more direct method by calculating the Laplace transform and using the compainon matrix. The Laplace transform is given by

$$\boldsymbol{\alpha}(s\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s} = \frac{1}{(1+s)^2} = \frac{1}{s^2+2s+1}.$$

Applying (1.8), we get that

$$\boldsymbol{p}_r = (1\ 0), \quad \boldsymbol{C} = \begin{pmatrix}0 & 1\\ -1 & -2\end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix}0\\1\end{pmatrix},$$

which is the transposed representation of $(\boldsymbol{\alpha}_2, \boldsymbol{S}_2, \boldsymbol{s}_2)$. □

The following result shall play an important role when dealing with more complex modeling of matrix–exponential distribution using flow methods.

Theorem 1.13. *It is alway possible to choose a (also minimal) representation* $(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ *such that* $\boldsymbol{s}=-\boldsymbol{S}\boldsymbol{e}$. *In this case* $\boldsymbol{\alpha}\boldsymbol{e}=1$. *Furthermore, the vector* $\boldsymbol{\alpha}$ *and matrix* $\boldsymbol{S}$ *may be taken to be real valued (not complex).*

Proof. Start out with a minimal representation (1.8) of dimension n. Let $\boldsymbol{M}$ be any non–singular matrix. Then we have that

$$f(x)=\boldsymbol{p}_r'\boldsymbol{M}\exp\left(\boldsymbol{M}^{-1}\boldsymbol{C}\boldsymbol{M}x\right)\boldsymbol{M}^{-1}\boldsymbol{e}_n.$$

The requirement to the transformation is that $-\boldsymbol{M}^{-1}\boldsymbol{C}\boldsymbol{M}\boldsymbol{e}=\boldsymbol{M}^{-1}\boldsymbol{e}_n$ or equivalently that $\boldsymbol{M}\boldsymbol{e}=-\boldsymbol{C}^{-1}\boldsymbol{e}_n$ (notice that $\boldsymbol{C}$ is invertible since the representation is minimal and hence its determinant which is $q_n\neq 0$). Now

$$\boldsymbol{C}^{-1}=\begin{pmatrix}-\frac{q_{n-1}}{q_n} & -\frac{q_{n-2}}{q_n} & \ldots & -\frac{q_1}{q_n} & -\frac{1}{q_n}\\ 1 & 0 & \ldots & 0 & 0\\ 0 & 1 & \ldots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \ldots & 1 & 0\end{pmatrix},$$

so

$$-\boldsymbol{C}^{-1}\boldsymbol{e}_n=(1/q_n,0,0,\ldots,0)'.$$

Thus any non–singular matrix which first row sums to $1/q_n$ and the rest to 0 is usable for the transformation. We could e.g. choose

$$\boldsymbol{M}=\begin{pmatrix}1/q_n & 0 & 0 & \ldots & 0 & 0\\ 0 & -1 & 1 & \ldots & 0 & 0\\ 0 & 0 & -1 & \ldots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \ldots & -1 & 1\\ -1 & 0 & 0 & \ldots & 0 & 1\end{pmatrix}.$$

If $\boldsymbol{s}=-\boldsymbol{S}\boldsymbol{e}$ then

$$1=\int_0^\infty \boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{s}dx=\left[-\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{e}\right]_0^\infty=\boldsymbol{\alpha}\boldsymbol{e}.$$

The representation is real–valued since it is a transformation of the real–valued canonical representation.

□

Definition 1.6. Let $X\sim \mathrm{ME}_p(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$. If $\boldsymbol{s}=-\boldsymbol{S}\boldsymbol{e}$, then we shall write $X\sim \mathrm{ME}_p(\boldsymbol{\alpha},\boldsymbol{S})$.

1.6 Flow interpretation

In this section we only work with matrix–exponenital distributions $\mathrm{ME}_p(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ which satisfy $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$. We shall establish a deterministic flow system from which we can generate the matrix–exponential distribution and perform further stochastic analysis.

Consider a matrix–exponential distribution $\mathrm{ME}_p(\boldsymbol{\alpha},\boldsymbol{S})$. Define p doubly infinite containers in which the initial amount of liquid in container i is α_i, $i = 1,2,...,p$. We define an additional container $p+1$, which shall serve as a kind of repository for the flow of the other containers, which has an intial amount of $0 \leq \alpha_{p+1} < 1$. We suppose that $\alpha_1 + \alpha_2 + ... + \alpha_{p+1} = 1$.

Liquid flows with rate S_{ij} from container i to container j, $1 \leq i, j \leq p$, $i \neq j$, which means that if $v_i(t)$ is the amount of liquid in container i by time t, then in a small time interval $[t, t+dt)$, the flow of liquid from i to j is $v_i(t)S_{ij}dt$. If S_{ij} is negative then there is a flow from j to i, but the force of the flow is regulated by the amount of liquid in the receptor i and which should not be mistaken for an equivalent from j to i of $v_j(t)(-S_{ij})dt$ since here the liquid is forced by the amount of liquid in container j. Furthermore, we define an "exit" flow from container i to container $p+1$ with rate s_i.

Let $\boldsymbol{v}(t) = (v_1(t),...,v_p(t))$ denote the vector of contents in the first p containers by time t. Then

Theorem 1.14. *For $t \geq 0$,*

$$\boldsymbol{v}(t) = \boldsymbol{\alpha} e^{\boldsymbol{S}t}.$$

Proof. The amount of liquid in container i by time $t+dt$ is obtained from the amount it had at time t, $v_i(t)$, plus all the in–flow from the other containers $j = 1,2...,p$, $j \neq i$ minus all the out–flow to container $j = 1,2,...,p,p+1$, $j \neq i$. Hence

$$v_i(t+dt) = v_i(t) + \sum_{\substack{j=1 \\ j\neq i}}^{p} v_j(t)S_{ji}dt - \sum_{\substack{j=1 \\ j\neq i}}^{p} v_i(t)S_{ij}dt - v_i(t)s_i dt.$$

Now $S_{ii} = -\sum_{\substack{j=1 \\ j\neq i}}^{p} S_{ij} - s_i$ so

$$v_i(t+dt) = v_i(t)(1+S_{ii}dt) + \sum_{\substack{j=1 \\ j\neq i}}^{p} v_j(t)S_{ji}dt$$

from which we conclude that

$$v_i'(t) = \sum_{j=1}^{p} v_j(t)S_{ji}$$

and since $\boldsymbol{v}(0) = (\alpha_1,...,\alpha_p)$, the theorem follows. □

The total amount to liquid in containers $1,2,..,p$ at time t is $\nu_1(t)+...+\nu_p(t) = vect\nu(t)\boldsymbol{e}$. Since the flow is entirely deterministic with initial total amount of $\alpha_1+...+\alpha_p+\alpha_{p+1}=1$, then the total amount must be preserved through all times. Hence the amount of liquid in container $p+1$ by time t is $1-\boldsymbol{\nu}(t)\boldsymbol{e} = 1-\boldsymbol{\alpha}\exp(\boldsymbol{S}t)\boldsymbol{e}$. This amounts to the distribution function of the matrix–exponential distribution and hence the amount of liquid in container $p+1$ is non–decreasing through time, and the initial amount α_{p+1}, if differento from 0, defines an atom at 0 of size α_{p+1}.

Definition 1.7. A flow system defined by initial amounts $\boldsymbol{\alpha}$ and flow rates $\boldsymbol{S}$ is called valid if the content $\nu_{p+1}(t)$ is non–decreasing and tends to 1 as $t\to\infty$.

The following theorem establishes the link between flows and matrix–exponential distributions.

Theorem 1.15. *Let $X\sim ME_p(\boldsymbol{\alpha},\boldsymbol{S})$. Consider the corresponding valid flow generated by $(\boldsymbol{\alpha},\boldsymbol{S})$ and let $T(x)=\inf\{s\geq|\nu_{p+1}(s)=x\}$. If U is uniformly distributed over $[0,1]$*

$$X\sim T(U).$$

Proof. The total in–flow to container $p+1$ during $[t,t+dt)$ is

$$\sum_{j=1}^{p}\nu_j(t)s_j dt.$$

Let $I_t=\sum_{j=1}^{p}\nu_j(t)s_j$. Since $U\sim\text{Uniform}(0,1)$, the probability that U is contained in $[\nu_{p+1}(t),nu_{p+1}(t)+I_t dt]$ is simply $I_t dt$. If we let g denote the density of $T(U)$, then

$$g(t)dt = I_t dt = \sum_{j=1}^{p}\nu_j(t)s_j dt = \sum_{j=1}^{p}\left(\boldsymbol{\alpha}e^{\boldsymbol{S}t}\right)_j s_j dt = \boldsymbol{\alpha}e^{\boldsymbol{S}t}\boldsymbol{s}dt,$$

from which the result follows. □

This result can be used to derive properties concerning matrix–exponential distributions and shall be used throughout the book whenever needed. As a first example of use we derive the form of the distribution function for the matrix–exponential distribution.

Theorem 1.16. *Let $X\sim ME_p(\boldsymbol{\alpha},\boldsymbol{S})$. Then the distribution function F of X is given by*

$$F(x)=1-\boldsymbol{\alpha}e^{\boldsymbol{S}x}\boldsymbol{e}.$$

Proof. The total amount of liquid in container $p+1$ at time t is $1-\boldsymbol{\alpha}\exp(\boldsymbol{S}t)\boldsymbol{e}$. Now

$$T(U)\leq x \iff U\leq 1-\boldsymbol{\alpha}\exp(\boldsymbol{S}t)\boldsymbol{e},$$

and so

$$F(x)=\mathbb{P}(T(U)\leq x)=\mathbb{P}(U\leq 1-\boldsymbol{\alpha}\exp(\boldsymbol{S}t)\boldsymbol{e})=1-\boldsymbol{\alpha}\exp(\boldsymbol{S}t)\boldsymbol{e}.$$

□

1.7 Closure properties of matrix–exponential distributions

Let $X \sim \mathrm{ME}_p(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ and $Y \sim \mathrm{ME}_q(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{t})$ be independent random variables. Then in terms of Laplace transforms, $L_{X+Y}(s) = L_X(s)L_Y(s)$ from which it is clear that $X+Y$ has a rational Laplace transform being a product of two rational functions $L_X(s)$ and $L_Y(s)$. Hence $X+Y$ is again a matrix–exponential distribution. If f_Z denotes the density of a random variable Z, we then have that

$$\begin{aligned} f_{X+Y}(s) &= \int_0^s f_X(u) f_Y(s-u) du \\ &= \int_0^t \boldsymbol{\alpha} e^{\boldsymbol{S}u} \boldsymbol{s} \boldsymbol{\beta} e^{\boldsymbol{T}(s-u)} \boldsymbol{t} du \\ &= \ldots \end{aligned}$$

max,min, moment distributions etc.

1.8 Characterization in terms of Moments and Hankel matrices

In this section we shall determinate the minimal order of a matrix–exponential distribution in terms of its moments. To this end we shall consider the moment generating function $M_X(s) = \mathbb{E}(e^{sX})$ of a matrix–exponentially distributed random variable X, which of course is also a rational function due to its close relationship with the Laplace transform.

Define moments $\mu_i = \mathbb{E}(X^i)$ and reduced moments $\mu_i' = \mu_i / i!$. Since $M_X(0) = 1$, we may write $M_X(s)$ on the following form, which for reasons which will be obvious in the following is more conveninent for our present purpose than form for the Laplace transform (1.1),

$$M_X(s) = \frac{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + 1}{b_n s^n + b_{n-1} s^{n-1} + \ldots + b_1 s + 1} = 1 + \mu_1' s + \mu_2' s^2 + \ldots. \tag{1.16}$$

Multiplying through with denominator, we get that

$$\begin{aligned} &a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + 1 \\ &\quad = (b_n s^n + b_{n-1} s^{n-1} + \ldots + b_1 s + 1)(1 + \mu_1' s + \mu_2' s^2 + \ldots). \end{aligned}$$

Matching coefficient to the terms s^i, $i = n+1, n+2, \ldots, 2n$ on both sides of the equation results in the following equations

$$0 = \mu_{n+1}' + \mu_n' b_1 + \mu_{n-1}' b_2 + \ldots + \mu_1' b_n$$

$$\begin{aligned}
0 &= \mu'_{n+2} + \mu'_{n+1} b_1 + \mu'_n b_2 + \ldots + \mu'_2 b_n \\
0 &= \mu'_{n+3} + \mu'_{n+2} b_1 + \mu'_{n+1} b_2 + \ldots + \mu'_3 b_n \\
&\ldots \\
0 &= \mu'_{2n} + \mu'_{2n-1} b_1 + \mu'_{2n-2} b_2 + \ldots + \mu'_n b_n.
\end{aligned}$$

Treating the b_i as unknowns and the μ'_i as constants, we solve for b_n, obtaining by Cramérs rule

$$b_n = \frac{\begin{vmatrix} -\mu'_{n+1} & \mu'_2 & \mu'_3 & \ldots & \mu'_n \\ -\mu'_{n+2} & \mu'_3 & \mu'_4 & \ldots & \mu'_{n+1} \\ -\mu'_{n+3} & \mu'_4 & \mu'_5 & \ldots & \mu'_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\mu'_{2n} & \mu'_{n+1} & \mu'_{n+2} & \ldots & \mu'_{2n-1} \end{vmatrix}}{\begin{vmatrix} \mu'_1 & \mu'_2 & \mu'_3 & \ldots & \mu'_n \\ \mu'_2 & \mu'_3 & \mu'_4 & \ldots & \mu'_{n+1} \\ \mu'_3 & \mu'_4 & \mu'_5 & \ldots & \mu'_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu'_n & \mu'_{n+1} & \mu'_{n+2} & \ldots & \mu'_{2n-1} \end{vmatrix}}, \tag{1.17}$$

where $|\cdot|$ denotes the determinant. Define Hankel determinants

$$\boldsymbol{\phi}_n = \begin{vmatrix} \mu'_1 & \mu'_2 & \mu'_3 & \ldots & \mu'_n \\ \mu'_2 & \mu'_3 & \mu'_4 & \ldots & \mu'_{n+1} \\ \mu'_3 & \mu'_4 & \mu'_5 & \ldots & \mu'_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu'_n & \mu'_{n+1} & \mu'_{n+2} & \ldots & \mu'_{2n-1} \end{vmatrix} \tag{1.18}$$

$$\boldsymbol{\psi}_n = \begin{vmatrix} \mu'_2 & \mu'_3 & \mu'_4 & \ldots & \mu'_n \\ \mu'_3 & \mu'_4 & \mu'_5 & \ldots & \mu'_{n+1} \\ \mu'_4 & \mu'_5 & \mu'_6 & \ldots & \mu'_{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu'_n & \mu'_{n+2} & \mu'_{n+3} & \ldots & \mu'_{2n-2} \end{vmatrix}. \tag{1.19}$$

$\boldsymbol{\psi}_n$ is the upper right sub–determinant of one degree less than $\boldsymbol{\phi}_n$. By simple manipulation of the determinant in the numerator of (1.17), we get that

$$b_n = (-1)^n \frac{\boldsymbol{\psi}_{n+1}}{\boldsymbol{\phi}_n}. \tag{1.20}$$

Theorem 1.17. *The minimal order p_0 of a matrix–exponential distribution is given by*

$$p_0 = \sup\{n \in \mathbb{N} \mid \boldsymbol{\psi}_{n+1} \neq 0\}.$$

Proof. If p_0 is the minimal order, then $b_n = 0$ for $n > p_0$ and hence $\boldsymbol{\psi}_{n+1} = 0$ for $n > p_0$. □

Example 1.2. Consider the distribution with density $f(x) = 2e^{-x}(1-\cos(x))$. Then, calculating the first few reduced moments, which amount to $2, \frac{5}{2}, \frac{5}{2}, \frac{9}{4}, 2, \frac{15}{8}, \frac{15}{8}, 2, \frac{65}{32}, \frac{65}{32}, 2, ...$, we get that $\boldsymbol{\phi}_4 = -\frac{1}{16}$, while $0 = \boldsymbol{\phi}_5 = \boldsymbol{\phi}_6 = ...$, so the minimal order is 3, one less than the index of the highest non–zero determinant.

1.9 Continued fractions

The theory of continued fractions provides an elegant framework for representing rational functions and provide further insight into the deeper structures of a theory which also has ramifications towards the moment problem.

By a continued fraction we under an expression on the form

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}},$$

which we shall denote by the more compact Pringsheim notation as

$$b_0 + \frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + ...$$

The n'th order finite continued fraction

$$\frac{A_n}{B_n} = b_0 + \frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + ... + \frac{a_n|}{|b_n} \tag{1.21}$$

is called an approximation to the (possibly infinite order) continued fraction.

A continued fraction, as a function of a variable s, is called a corresponding continued fraction, or C–fraction, if it can be written on the form

$$1 + \frac{a_1 s^{\alpha_1}|}{|1} + \frac{a_2 s^{\alpha_2}|}{|1} + \frac{a_3 s^{\alpha_3}|}{|1} + ... = 1 + \cfrac{a_1 s^{\alpha_1}}{1 + \cfrac{a_2 s^{\alpha_2}}{1 + \cfrac{a_3 s^{\alpha_3}}{1 + ...}}}.$$

A C–fraction is called regular if all $\alpha_i = 1$. A continued fraction is called terminal if the a_i's are zero from a certain point.

Theorem 1.18. *Let $M_X(s) = \frac{p_0(s)}{p_1(s)}$ denote the moment generating function of a matrix–exponential distribution, where p_1 is a polynomial of order d and p_0 a polynomial of order at most $d-1$. Then $M_X(s)$ can be written as a terminal regular C–fraction.*

Proof. In terms of the reduced moments $\mu_i' = \mathbb{E}(X^i)/i!$ we can write

$$M_X(s) = 1 + \mu_1' s + \mu_2' s^2 + \ldots = \frac{p_0(s)}{p_1(s)}.$$

Since $M_X(0) = 1$, we may assume without loss of generality that both p_0 and p_1 have the constant term 1. Let $B_0(s) = \frac{p_0(s)}{p_1(s)}$. Since we aim at developing $M_X(s)$ as a C–fraction, we write

$$B_0(s) = 1 + \frac{a_1 s^{\alpha_1}}{B_1(s)}$$

for some function $B_1(s)$. Then

$$B_1(s) = \frac{a_1 s^{\alpha_1}}{B_0(s) - 1} = \frac{a_1 s^{\alpha_1}}{\frac{p_0(s)}{p_1(s)} - 1} = \frac{p_1(s) a_1 s^{\alpha_1}}{p_0(s) - p_1(s)}.$$

Since both p_0 and p_1 have constant terms 1, it is clear that there exists an α_1 such that $a_1 s^{\alpha_1}$ divides $p_0(s) - p_1(s)$. Now

$$\frac{p_0(s)}{p_1(s)} = 1 + \mu_1' + \mu_2' s^2 + \ldots$$

so

$$p_0(s) - p_1(s) = \mu_1' s p_1(s) + \mu_2 s^2 p_1(s) + \ldots$$

and since $a_1 s^{\alpha_1}$ cannot divide $p_1(s)$ because of its constant term 1, $a_1 s^{\alpha_1}$ must divide $\mu_1 s$, so the largest α_1 we can choose is $\alpha_1 = 1$. Then

$$B_1(s) = \frac{p_1(s)}{p_2(s)}$$

where

$$p_2(s) = \frac{p_0(s) - p_1(s)}{a_1 s}.$$

If we let $a_1 = \mu_1$ then furthermore

$$B_1(s) = \frac{\mu_1' s}{B_0(s) - 1} = \frac{\mu_1' s}{\mu_1' s + \mu_2' s^2 + \ldots} = \frac{1}{1 + \mu_2'/\mu_1' s + \ldots.}$$

will have a power expansion with the constant term 1 and all coefficient to the powers s^i non–zero. Hence the situation is identical to the one for $B_0(s)$. Hence we may proceed with the same proceedure and the n iteration we obtain a rational function

$$B_n(s) = \frac{p_n(s)}{p_{n+1}(s)}.$$

Concerning the orders of the polynomials, we let d_i denote the degree of the polynomial p_i. Since $p_2(s) = (p_0(s) - p_1(s))/(\mu_1' s)$, then $d_2 \leq \max(d_0, d_1) - 1$. Similarly, continuing the proceedure from above, $d_3 \leq \max(d_1, d_2) - 1 \leq \max(d_0, d_1) - 1$ so

$$\max(d_2,d_3) \leq \max(d_0,d_1) - 1.$$

Continuing further this way to polynomial of order 4 and 5,

$$\max(d_4,d_5) = \leq \max(d_2,d_3) - 1 \leq \max(d_0,d_1) - 2.$$

Obviously this can only go on a finite number of times until the degrees of the polynomials are zero (constants). Hence the result has been proved. □

The proof above indicates a method for transforming an expression into a continued fraction. We shall illustrate the procedure more precisely in the following example.

Example 1.3. Consider a random variable X of matrix–exponential distribution with density

$$f(x) = 2e^{-x}(1-\cos(x)).$$

Its moment generating function $M_X(s)$ is given by

$$M_X(s) = \frac{2}{-s^3+3s^2-4s+2}.$$

First we write the moment generating function as a fraction between two polynomials p_0 and p_1 with constant term 1, i.e.

$$M_X(s) = \frac{1}{-\frac{1}{2}s^3+\frac{3}{2}s^2-2s+1} = \frac{p_0(s)}{p_1(s)}.$$

Then consider the difference

$$p_0(s) - p_1(s) = 1 - (-\frac{1}{2}s^3 + \frac{3}{2}s^2 - 2s + 1) = \frac{1}{2}s^3 - \frac{3}{2}s^2 + 2s.$$

In order to produce a polynomial with constant term 1, we divide $p_0 - p_1$ by $2s$. Hence $a_1 = 2$. Let

$$p_2(s) = \frac{p_0(s) - p_1(s)}{2s} = \frac{1}{4}s^2 - \frac{3}{4}s + 1.$$

Then consider

$$p_1(s) - p_2(s) = -\frac{1}{2}s^3 + \frac{5}{4}s^2 - \frac{5}{4}s.$$

Again, in order to produce a polynomial with constant term 1, we divide $p_1 - p_2$ by $-\frac{5}{4}s$. Hence $a_2 = -\frac{5}{4}$. Let

$$p_3(s) = \frac{p_1(s) - p_2(s)}{-\frac{5}{4}s} = \frac{2}{5}s^2 - s + 1.$$

Then consider

$$p_2(s) - p_3(s) = -\frac{3}{20}s^2 + \frac{1}{4}s$$

from which we conclude that $a_3 = \frac{1}{4}$, and

$$p_4(s) = \frac{p_2(s) - p_3(s)}{\frac{1}{4}s} = -\frac{3}{5}s + 1.$$

Hence

$$p_3(s) - p_4(s) = \frac{2}{5}s^2 - \frac{2}{5}s$$

and $a_4 = -\frac{2}{5}$. Thus

$$p_5(s) = \frac{p_3(s) - p_4(s)}{-\frac{2}{5}s} = -s + 1$$

and

$$p_4(s) - p_5(s) = \frac{2}{5}s.$$

Thus $a_5 = \frac{2}{5}$ and $p_6(s) = -s$. Hence $p_7(s) = 1$ and the procedure terminates. Thus we have calculated that

$$M_X(s) = 1 + \frac{2s|}{|1} + \frac{-\frac{5}{4}s|}{|1} + \frac{\frac{1}{4}s|}{|1} + \frac{-\frac{2}{5}s|}{|1} + \frac{\frac{2}{5}s|}{|1} + \frac{-s|}{|1} = 1 + \cfrac{2s}{1 - \cfrac{5/4s}{1 + \cfrac{1/4s}{1 - \cfrac{2/5s}{1 + \cfrac{2/5s}{1-s}}}}}.$$

□

The procedure would work equally well on the Laplace transform or any other rational function which allows for a regular C–fraction expression.

In the following we develop a general formula for the coefficients in the continued fraction in terms of the reduced moments μ_i'. To this end we consider lower approximation to the continued fraction. Consider a (posibly infinite) regular C–fraction

$$1 + \frac{a_1 s|}{|1} + \frac{a_2 s|}{|1} + \frac{a_3 s|}{|1} + \ldots$$

and its n'th order approximants (see (1.21)),

$$\frac{A_n(s)}{B_n(s)} = 1 + \frac{a_1 s|}{|1} + \frac{a_2 s|}{|1} + \frac{a_3 s|}{|1} + \ldots + \frac{a_n s|}{|1}. \tag{1.22}$$

Then it is clear that $A_n(s)/B_n(s)$ are rational functions. Furthermore, the order of A_{2n} and B_{2n} are at most n, while the order of A_{2n-1} and B_{2n-1} are at most n and $n-1$ respectively. This follows trivially from noticing that for each extra term which is added to the continuous fraction, the degree of either the numerator or denominator is increased by 1 alternatingly. Hence we may write

$$\frac{A_{2n}(s)}{B_{2n}(s)} = \frac{1+a_{n,1}s+...+a_{n,n}s^n}{1+b_{n,1}s+...+b_{n,n}s^n}$$
$$\frac{A_{2n-1}(s)}{B_{2n-1}(s)} = \frac{1+c_{n,1}s+...+c_{n,n}s^n}{1+d_{n,1}s+...+d_{n,n-1}s^{n-1}}$$

for some constants $a_{i,j}, b_{i,j}, c_{i,j}$ and $d_{i,j}$.

Now suppose that the regular C–fraction corresponds to the power series

$$1+\mu_1' s+\mu_2' s^2+....$$

Let us consider the power series expansion of $A_{2n}(s)/B_{2n}(s)$ and $A_{2n-1}(s)/B_{2n-1}(s)$. When calculating the coefficients in the continuous fraction we first determinate a_1 by writing

$$B_0(s) = 1+\frac{a_1 s}{B_1(s)} = 1+\mu_1' s+....$$

This gives the same a_1 independently of the order of $B_1(s)$. The second term a_2 is determined posteriorly by noting that

$$B_1(s) = \frac{a_1 s}{B_0(s)-1} = \frac{1}{1+\mu_1'' s+\mu_2'' s^2+...}$$

where $\mu_i'' = \mu_i'/\mu_1'$. Then writing

$$B_1(s) = 1+\frac{a_2 s}{B_s(s)}$$

and expanding $\dfrac{1}{1+\mu_1'' s+\mu_2'' s^2+...}$ into a Taylor series solves for a_2 independently of the order of B_2. Continuing this way we conclude that the a_1 is determined by first non–constant term of the Taylor series, a_2 by the first two non-constant terms etc. Hence we conclude that in the power series expansion of $A_n(s)/B_n(s)$ the coefficients of the first n terms coincide with $\mu_1', \mu_2', ..., \mu_n'$. Hence

$$\frac{A_{2n}(s)}{B_{2n}(s)} = \frac{1+a_{n,1}s+...+a_{n,n}s^n}{1+b_{n,1}s+...+b_{n,n}s^n} = 1+\mu_1' s+...+\mu_{2n}' s^{2n}+k_{2n+1}^1 s^{2n+1}+...$$
$$\frac{A_{2n-1}(s)}{B_{2n-1}(s)} = \frac{1+c_{n,1}s+...+c_{n,n}s^n}{1+d_{n,1}s+...+d_{n,n-1}s^{n-1}} = 1+\mu_1' s+...+\mu_{2n-1}' s^{2n-1}+k_{2n}^2 s^{2n+1}+...$$

for some constants k_n^i.

We need some Lemmas on basic properties of manipulation with continued fraction.

Lemma 1.1. *(Euler–Wallis)*
Let

$$\frac{A_n}{B_n} = b_0+\frac{a_1|}{|b_1}+...+\frac{a_n|}{|b_n}$$

denote the n'th order approximation to the (possibly infinity) continued fraction

$$b_0 + \frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + \ldots.$$

Then with $A_{-1} = 1, A_0 = b_0, B_{-1} = 0$ *and* $B_0 = 1$ *the following recursion scheme holds true*

$$\begin{aligned} A_n &= b_n A_{n-1} + a_n A_{n-2} \\ B_n &= b_n B_{n-1} + a_n B_{n-2}. \end{aligned}$$

Proof. The Lemma is true for $n = 1$ and $n = 2$. Assume that the formulas are true for up to n. If we let $b'_n = b_n + a_{n+1}/b_{n+1}$, then

$$\begin{aligned} \frac{A_{n+1}}{B_{n+1}} &= b_0 + \frac{a_1|}{|b_1} + \ldots + \frac{a_{n+1}|}{|b_{n+1}} \\ &= b_0 + \frac{a_1|}{|b_1} + \ldots + \frac{a_n|}{|b_n + a_{n+1}/b_{n+1}} \\ &= b_0 + \frac{a_1|}{|b_1} + \ldots + \frac{a_n|}{|b'_n} \\ &= \frac{A'_n}{B'_n}, \end{aligned}$$

where $a'_i = a_i$ for $i = 1, \ldots, n$ and $b'_i = b_i$ for $i = 1, \ldots, n-1$. This is simply condensing the last two terms of the continued fraction into one by redefining the constants adequately. Then by the induction hypothesis, we can write

$$\begin{aligned} \frac{A'_n}{B'_n} &= \frac{b'_n A'_{n-1} + a'_n A'_{n-2}}{b'_n B'_{n-1} + a'_n B'_{n-2}} \\ &= \frac{(b_n + a_{n+1}/b_{n+1}) A_{n-1} + a_n A_{n-2}}{(b_n + a_{n+1}/b_{n+1}) B_{n-1} + a_n B_{n-2}} \\ &= \frac{A_n + a_n/b_n A_{n-1}}{B_n + a_n/b_n B_{n-1}} \\ &= \frac{b_n A_n + a_n A_{n-1}}{b_n B_n + a_n B_{n-1}}, \end{aligned}$$

which proves the lemma. □

Consider the expressions on the form

$$\begin{aligned} P_n &= b_0 b_1 \cdots b_n \left(1 + \frac{a_1}{b_0 b_1}\right)\left(1 + \frac{a_2}{b_1 b_2}\right) \cdots \left(1 + \frac{a_n}{b_{n-1} b_n}\right) \\ Q_n &= b_1 \cdots b_n \left(1 + \frac{a_2}{b_1 b_2}\right)\left(1 + \frac{a_3}{b_2 b_3}\right) \cdots \left(1 + \frac{a_n}{b_{n-1} b_n}\right). \end{aligned}$$

Multiplying out P_n and Q_n, it is clear that they both contain "integer" terms as well as "fractional" terms. For example, for P_n the term

$$b_0 b_1 \cdots b_n \frac{a_1}{b_0 b_1} = a_1 b_2 \cdots b_n$$

is an "integer" term and

$$b_0 b_1 \cdots b_n \frac{a_1}{b_0 b_1} \frac{a_2}{b_1 b_2} = \frac{a_2}{b_1} b_3 \cdots b_n$$

is a fractional "term". Hence we can split P_n and Q_n into an "integer" part, which we shall denote $[[P_n]]$ and a fractional part which shall be denoted $\{\{P_n\}\}$, i.e.

$$P_n = [[P_n]] + \{\{P_n\}\}.$$

Lemma 1.2. *(Euler)*

$$A_n = [[P_n]], \;\; B_n = [[Q_n]].$$

Proof. $P_0 = b_0 = A_0$ and $P_1 = b_0 b_1 (1 + a_1/b_0 b_1) = b_0 b_1 + a_1 = A_1$. So for $n = 0$ and $n = 1$ the assertion holds since both expressions are of "integer" type. Since the initial conditions hold, then it is sufficient to prove that the "integer" parts $[[P_n]]$ satisfy the Euler–Wallis recursion in Lemma 1.1, i.e. we need to show that

$$[[P_n]] = b_n [[P_{n-1}]] + a_n [[P_{n-2}]].$$

Now

$$\begin{aligned} P_n &= b_0 b_1 \cdots b_n \left(1 + \frac{a_1}{b_0 b_1}\right) \left(1 + \frac{a_2}{b_1 b_2}\right) \cdots \left(1 + \frac{a_n}{b_{n-1} b_n}\right) \\ &= b_n \left(1 + \frac{a_n}{b_{n-1} b_n}\right) P_{n-1} \\ &= b_n P_{n-1} + \frac{a_n}{b_{n-1}} P_{n-1} \\ &= b_n P_{n-1} + \frac{a_n}{b_{n-1}} \left(b_{n-1} P_{n-2} + \frac{a_{n-1}}{b_{n-2}} P_{n-2}\right) \\ &= b_n P_{n-1} + a_n P_{n-2} + \frac{a_n a_{n-1}}{b_{n-1} b_{n-2}} P_{n-2}. \end{aligned}$$

P_{n-1} does not depend on b_n so $[[b_n P_{n-1}]] = b_n [[P_{n-1}]]$ and P_{n-2} does not depend on a_n so $[[a_n P_{n-2}]] = a_n [[P_{n-2}]]$. On the other hand, P_{n-1} does not depend on b_{n-1} so the last term is "fractional". Hence

$$[[P_n]] = b_n [[P_{n-1}]] + a_n [[P_{n-2}]].$$

The proof for Q_n is similar. □

Lemma 1.3. *(Euler-Minding)*

$$A_n = b_0 b_1 \cdots b_n \left(1 + \sum_{i}^{0,n-1} \frac{a_{i+1}}{b_i b_{i+1}} + \sum_{i<k}^{0,n-2} \frac{a_{i+1}}{b_i b_{i+1}} \frac{a_{k+2}}{b_{k+1} b_{k+2}} + \right.$$
$$\left. + \sum_{i<k<\ell}^{0,n-3} \frac{a_{i+1}}{b_i b_{i+1}} \frac{a_{k+2}}{b_{k+1} b_{k+2}} \frac{a_{\ell+3}}{b_{\ell+2} b_{\ell+3}} + \ldots \right)$$
$$B_n = b_1 \cdots b_n \left(1 + \sum_{i}^{1,n-1} \frac{a_{i+1}}{b_i b_{i+1}} + \sum_{i<k}^{1,n-2} \frac{a_{i+1}}{b_i b_{i+1}} \frac{a_{k+2}}{b_{k+1} b_{k+2}} + \right.$$
$$\left. + \sum_{i<k<\ell}^{1,n-3} \frac{a_{i+1}}{b_i b_{i+1}} \frac{a_{k+2}}{b_{k+1} b_{k+2}} \frac{a_{\ell+3}}{b_{\ell+2} b_{\ell+3}} + \ldots \right),$$

where the sum $\sum_{i_1<\ldots<i_m}^{a,b}$ *sums over all combinations of* $i_1, \ldots, i_m$ *taking values* $a, a+1, a+2, \ldots, b$ *and without repitition* ($i_1 < i_2 < \ldots < i_m$).

Proof. Direct consequence of Lemma 1.2 . □

Applying the Euler–Minding formulas to the case of regular C–fractions we get that

$$A_n(s) = 1 + s \sum_{i}^{0,n-1} a_{i+1} + s^2 \sum_{i<k}^{0,n-2} a_{i+1} a_{k+2} + s^3 \sum_{i<k<\ell}^{0,n-3} a_{i+1} a_{k+2} a_{\ell+3} + \ldots$$
$$B_n(s) = 1 + s \sum_{i}^{1,n-1} a_{i+1} + s^2 \sum_{i<k}^{1,n-2} a_{i+1} a_{k+2} + s^3 \sum_{i<k<\ell}^{1,n-3} a_{i+1} a_{k+2} a_{\ell+3} + \ldots$$

Then with

$$\frac{A_{2n}(s)}{B_{2n}(s)} = \frac{1 + a_{n,1} s + \ldots + a_{n,n} s^n}{1 + b_{n,1} s + \ldots + b_{n,n} s^n}$$
$$\frac{A_{2n-1}(s)}{B_{2n-1}(s)} = \frac{1 + c_{n,1} s + \ldots + c_{n,n} s^n}{1 + d_{n,1} s + \ldots + d_{n,n-1} s^{n-1}}$$

we see that

$$b_{n,n} = a_2 a_4 \cdots a_{2n}, \;\; c_{n,n} = a_1 a_3 \cdots a_{2n-1}.$$

On the other hand, writing

$$\frac{A_{2n}(s)}{B_{2n}(s)} = \frac{1 + a_{n,1} s + \ldots + a_{n,n} s^n}{1 + b_{n,1} s + \ldots + b_{n,n} s^n} = 1 + \mu'_1 s + \ldots + \mu'_{2n} s^{2n} + k^1_{2n+1} s^{2n+1} + \ldots$$
$$\frac{A_{2n-1}(s)}{B_{2n-1}(s)} = \frac{1 + c_{n,1} s + \ldots + c_{n,n} s^n}{1 + d_{n,1} s + \ldots + d_{n,n-1} s^{n-1}} = 1 + \mu'_1 s + \ldots + \mu'_{2n-1} s^{2n-1} + k^2_{2n} s^{2n+1} + \ldots$$

we may solve for $b_{n,n}$ and $c_{n,n}$ by multiplying the denominator through in each equation and matching coefficients to respectively the terms $s^{n+1}, s^{n+2}, \ldots$ and to $s^n, s^{n+1}, \ldots$. The first term $b_{n,n}$ has already been solved for in (1.20) giving

$$\boldsymbol{\phi}_n b_{n,n} = (-1)^n \boldsymbol{\psi}_{n+1}$$

and hence

$$a_2 a_4 \cdots a_{2n} \boldsymbol{\phi}_n = (-1)^n \boldsymbol{\psi}_{n+1}, \;\; n \geq 1. \tag{1.23}$$

Similarly we obtain that

$$a_1 a_3 \cdots a_{2n-1} \boldsymbol{\psi}_n = (-1)^{n-1} \boldsymbol{\phi}_n, \;\; n \geq 1 \tag{1.24}$$

where we have set $\boldsymbol{\psi}_1 = 1$.

From (1.23) and (1.24) we may extract the following information. We know that $\boldsymbol{\psi}_1 = 1 \neq 0$ and $\boldsymbol{\psi}_2 = \mu_2' \neq 0$. If $\boldsymbol{\psi}_n \neq 0$, then from (1.24), all $a_1, ..., a_{2n-1} \neq 0$ and $\boldsymbol{\phi}_n \neq 0$. If $\boldsymbol{\psi}_n = 0$ for the first time, then $\boldsymbol{\phi}_n = 0$ but then by formula (1.23) also $\boldsymbol{\psi}_{n+1} = 0$ and hence all subsequent $\boldsymbol{\phi}_m = \boldsymbol{\psi}_m = 0$ for $m \geq n$.

We now state one our main theorem of this section.

Theorem 1.19. *The power series $1 + \mu_1' + \mu_2' + ...$ has a regular corresponding continued fraction (C–fraction) which is infinite if and only if all determinants*

$$\boldsymbol{\phi}_n = \begin{vmatrix} \mu_1' & \mu_2' & ... & \mu_n' \\ \mu_2' & \mu_3' & ... & \mu_{n+1}' \\ \mu_3' & \mu_4' & ... & \mu_{n+2}' \\ \vdots & \vdots & \vdots & \vdots \\ \mu_n' & \mu_{n+1}' & ... & \mu_{2n-1}' \end{vmatrix} \neq 0 \;\; \boldsymbol{\psi}_n = \begin{vmatrix} \mu_2' & \mu_3' & ... & \mu_n' \\ \mu_3' & \mu_4' & ... & \mu_{n+1}' \\ \mu_4' & \mu_5' & ... & \mu_{n+2}' \\ \vdots & \vdots & \vdots & \vdots \\ \mu_n' & \mu_{n+1}' & ... & \mu_{2n-2}' \end{vmatrix} \neq 0$$

for all $n = 1, 2, ...$ (recall that $\boldsymbol{\psi}_1 = 1$ by definition), in which case

$$a_1 = \boldsymbol{\phi}_1, \;\; a_{2n} = -\frac{\boldsymbol{\psi}_{n+1} \boldsymbol{\phi}_{n-1}}{\boldsymbol{\psi}_n \boldsymbol{\phi}_n}, \;\; a_{2n-1} = -\frac{\boldsymbol{\phi}_n \boldsymbol{\psi}_{n-1}}{\boldsymbol{\phi}_{n-1} \boldsymbol{\psi}_n}. \tag{1.25}$$

If $\boldsymbol{\psi}_n =$ for some n, then so are all the subsequent determinants, the C–fraction is consequently finite and the formulas (1.25) still holds true.

Proof. In $a_2 a_4 \cdots a_{2n} \boldsymbol{\phi}_n = (-1)^n \boldsymbol{\psi}_{n+1}$ we substitute $a_2 a_4 \cdots a_{2n-2}$ by $(-1)^{n-1} \boldsymbol{\psi}_n / \boldsymbol{\phi}_n$ which yields the formula for a_{2n}. That trick is similar for a_{2n-1}. The the determinants remain zero from a certain point up was noted prior to this theorem. That the formulas hold true also in the finite case is evident. □

1.10 The residual life operator

Let $(X, \mathscr{F}_X)$ be a measurable space, i.e. $\mathscr{F}_X$ is a σ–algebra defined on the space X.

Definition 1.8. A signed measure μ on $(X, \mathscr{F}_X)$ is a function from $\mathscr{F}$ to $[-\infty, \infty]$ such that

1. $\mu(\emptyset) = 0$.
2. μ can at most assume one of the values $+\infty$ or $-\infty$.

3. If $A_1, A_2, \ldots$ are disjoint $\in \mathscr{F}$, then

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

A non–negative signed measure is simply a measure in the usual sense.

Definition 1.9. Let μ be a signed measure on the Borel σ–algebra of $[0,\infty)$. The residual life operator R_t acting on μ is defined by

$$R_t\mu(A) = \mu(A+t) + \mu([0,s))\delta_0(A),$$

where $A+t := \{s+t | s \in A\}$.

Example 1.4. Let $\mu(E) = \mathbb{P}(X \in E)$ for some non–negative random variable X. Then $R_t\mu(E) = \mu(E+t) = \mathbb{P}(X-t \in E) + \mathbb{P}(X < t)\delta_0(E)$, so $R_t\mu$ is the measure obtained by defective distribution of $X-t$, defined on $\{X \geq t\}$, only and adding the defective part $\mathbb{P}(X < t)$ to an atom at zero.

Indeed, if we let $E = [0,x]$, then $\mu(E) = F(x)$, where F denotes the distribution function of X, and $R_tF(x) = R_t\mu([0,x]) = \mu([t,t+x]) + \mu([0,t))\delta_0([0,x]) = F(t+x) - F(t) + F(t) = F(t+x)$. Thus

$$||R_tF|| := \lim_{x\to\infty} R_tF(x) = \lim_{x\to\infty} F(x+t) = 1$$

so R_tF is again a non–defective distribution function.

If F is absolutely continuous with density f, then R_tF consist of an atom at zero of size $\mathbb{P}(X < t)$ and an absolute continuous part which has a defective density given by $R_tf(x) = f(x+t)$. The total mass of the absolute continuous part is

$$\int_0^{\infty} f(x+t)dx = \int_t^{\infty} f(x)dx = \mathbb{P}(X \geq t).$$

Note 1.2. In [1] the residual life operator is defined by

$$R_t\mu(E) = \mu(E+t). \tag{1.26}$$

Here $R_tF(x) = F(x+t) - F(t)$, which means that $R_tF(x)$ does not converge to 1 as $x \to \infty$ but in turn to $1 - F(t)$. The missing mass $F(t)$ is here left unassigned. The difference is not important and the following theory could have been carried out with this definition as well, however, we shall stick to the definition where probability measures are invariant under an R_t action and not left as defective probability measures.

Theorem 1.20. *R_t is a semi–group operator on the space of signed measures on $[0,\infty)$, i.e. $R_tR_s = R_sR_t = R_{t+s}$.*

Proof. Let E be a Borel subset of $[0,\infty)$. Then for $s,t > 0$ an if $0 \in E$, then

$$\begin{aligned} R_t R_s \mu(E) &= R_t \left(\mu(E+s) + \mu([0,s))\right) \\ &= R_t \mu(E+s) + R_t \mu([0,s)) \\ &= \mu(E+s+t) + \mu([0,t))\delta_0(E+t) + \mu([t,t+s)) + \mu([0,t))\delta_0([0,s)) \\ &= \mu(E+s+t) + \mu([0,t+s)) \end{aligned}$$

since $0 \notin E+t$ and $0 \in [0,s)$. If $0 \notin E$, then same calculation gives us that $R_t R_s \mu(E) = \mu(E+s+t)$. Hence we conclude that

$$R_t R_s \mu(E) = \mu(E+t+s) + \mu([0,s+t))\delta_0(E) = R_{t+s}\mu(E) = R_s R_t \mu(E).$$

□

We are going to consider the vector space spanned by $R_t\mu$ for $t \geq 0$. Hence differences must be contained in space as well and therefore we cannot simply define the vector space for measures only since some differences will be negative. Hence the extension to signed measures is necessary. The Hahn and Jordan decompositions provide details about the structure of signed measures in terms of "ordinary" measures.

Theorem 1.21. *(Hahn decomposition) If μ is a signed measure, then there is a decomposition of X into disjoint sets X_+ and X_- such that $X = X_+ \cup X_-$, $\mu(A) \geq 0$ for all $A \in \mathscr{F}_X \cap X_+$ and $\mu(A) \leq 0$ for all $A \in \mathscr{F}_A \cap X_-$.*

A signed measure is supported by set E if $\mu(A) = \mu(A \cap E)$ for all A. This implies that $\mu(E^c) = 0$, and that $\mu(B) = 0$ for all measurable subsets of E^c. We say that two signed measures μ and ν are mutually singular if they are supported by two disjoint sets and we write $\mu \perp \nu$.

Theorem 1.22. *(Jordan decomposition) A signed measure can be decomposed as $\mu = \mu_+ - \mu_-$, where $\mu_+(A) = \mu(A \cap X_+)$ and $\mu_-(A) = \mu(A \cap X_-)$, and where X_+ and X_- are obtained as for the Hahn decomposition.*

A direct consequence of this theorem is that $|\mu| = \mu_+ + \mu_-$. In order to provide a norm on the space of signed measures we define the total variation norm by

$$\|\mu\| = |\mu|(X) = \mu_+(X) + \mu_-(X) \tag{1.27}$$

The following vector space $V_R(\mu)$ generalizes the vector space V_B of (1.13).

Definition 1.10.

$$V_R(\mu) = \text{span}\{R_t\mu | t \geq 0\} = \left\{ \sum_{j=1}^{n} c_j R_{t_j}\mu : n \in \mathbb{N}, t_1, ..., t_n \geq 0 \right\}.$$

If μ is a distribution with rational Laplace transform and has density (of the absolute continuous part) f, then the density corresponding to $R_t\mu$ is $x \to f(x+t)$. Hence the density corresponding to $R_t\mu$ is also on the form (1.12). If μ is a signed measure, then by the Jordan decomposition (Theorem 1.22) we may write $\mu = c_1\mu_1 - c_2\mu_2$, where μ_1 and μ_2 are distributions (for some pathological cases we may choose $c_1 =$

∞ or $c_2 = \infty$, but not both) and hence it follows that any element $y = \sum_{j=1}^{n} c_j R_{t_j}\mu \in V_R(\mu)$ can be written as a linear combination of the terms given in Theorem 1.12.

Expressing signed measures $R_t\mu$ in terms of linear combination of densities which correspond to their Jordan decomposition , we have proved that

Theorem 1.23. *If μ is a signed measure which has a rational Laplace transform $\hat{\mu}$, i.e.*

$$\hat{\mu}(s) = \int_0^\infty e^{-sx}\mu(dx)$$

is a rational function in s. Then

$$V_R(\mu) \subseteq span\left\{x^j e^{\rho_j x}, x^j \cos(a_j)e^{\sigma_j x}, x^j \sin(b_j)e^{\tau_j x}, j = 1,...,m\right\}$$

for some finite m.[2]

Let $\mathscr{S}$ denote the space of signed measures. Then R_t is an operator from $\mathscr{S}$ into $\mathscr{S}$. It trivially satisfies the semi–group property $R_tR_s = R_sR_t = R_{t+s}$. The norm of the operator R_t is defined by

$$\|R_t\| = \sup\left\{\frac{\|R_t\mu\|}{\|\mu\|} : \mu \in \mathscr{S}\right\}. \tag{1.28}$$

The operator R_t is linear since $R_t(\alpha\mu + \beta\nu) = \alpha\mu(t+\cdot) + \beta\nu(t+\cdot) = \alpha R_t\mu + \beta R_t\mu$. To prove that it is bounded we may assume without loss of generality that μ is given by a distribution function F. Then $\|F\| = 1$ and $\|R_t\| = \sup\{\|R_tF\|\| : F$ distribution function on $\mathbb{R}_+\} = F(t+\infty) = 1 < \infty$. [3]. Hence R_t is a bounded linear semi–group operator on the space of signed measures $\mathscr{S}$.

Now consider a signed measure for which $V_R(\mu)$ is a finite–dimensional vector–space, and consider the semi–group R_t acting on $V_R(\mu)$. Let d denote the dimension of $V_R(\mu)$ and let $\nu_1,...,\nu_d$ be a basis for $V_R(\mu)$. Then $R_t\nu_i \in V_R(\mu)$ since if $\nu_i = \sum_{j=1}^{n} c_j R_{t_j}\mu$ for some numbers $n \in \mathbb{N}$, $t_1,...,t_n \geq 0$ and $c_1,...,c_n \in \mathbb{R}$, then $R_t\nu_i = \sum_{j=1}^{n} c_i R_{t+t_j}\mu \in V_R(\mu)$. Let $\nu \in V_R(\mu)$. Then $\nu = a_1\nu_1 + ... + a_d\nu_d$ for some constants $a_1,...,a_d \in \mathbb{R}$ and hence $R_t\nu = a_1R_t\nu_1 + ... + a_dR_t\nu_d \in V_K(\mu)$. Thus we have also that $R_t\nu - \nu \in V_K(\mu)$ and $(R_t\nu - \nu)/t \in V_K(\mu)$. Since $V_K(\mu)$ is a finite–dimensional vector space it is closed, and hence

$$\lim_{t\downarrow 0} \frac{R_t\nu - \nu}{t} \in V_K(\mu) \tag{1.29}$$

for any $\nu \in V_K(\mu)$ and where the convergence is in the $\|\cdot\|$–norm on the signed measures. Thus with I denoting the identity operator,

$$\lim_{t\downarrow 0} \frac{R_t - I}{t}$$

[2] Do we need to distinguish between measures and densities and formulate all spaces as spanned by densities? Consider this important issue which is not attended in OCinneide

[3] Not correct; find norm in measure space

is a linear operator on $V_K(\mu)$ (R_t being linear).

We may think of ν as a vector expressed in terms of the basis $\nu_1, ..., \nu_d$, $\nu = a_1\nu_1 + ... + a_d\nu_d$ so $\nu = (a_1, ..., a_d)'$. Now since the limit (1.29) exists and is in $V_R(\mu)$ it then follows that there exists a $d \times d$ matrix $\boldsymbol{G}$ such that

$$\lim_{t \downarrow 0} \frac{R_t - I}{t} \nu = \boldsymbol{G}\nu. \tag{1.30}$$

This proves that R_t is (right–)differentiable at 0 with (right) derivative $\boldsymbol{G}$.

In order to proceed formally we shall need the following technical lemma from functional analysis.

Lemma 1.4. *Let U be a finite–dimensional vector space with norm $\|\cdot\|$. Let $\{\Psi_s : s \geq 0\}$ be a set of linear functionals from U into $\mathbb{R}$. Let $\{\mu_n\}_{n \in \mathbb{N}} \subset U$ and $\mu \in U$. Suppose that $\{\Psi_s : s \geq 0\}$ separates the points of U, i.e. if for any $\nu \in U$ it holds that $\forall s \geq 0 : \Psi_s(\nu) = 0$ implies that $\nu = 0$. Then the following are equivalent.*

1. *μ_n converges in $\|\cdot\|$–norm to μ as $n \to \infty$.*
2. *$\Psi_s(\mu_n) \to \Psi_s(\mu)$ as $n \to \infty$ for all $s \geq 0$.*

Proof. Assume that $\|\mu_n - \mu\| \to 0$ as $n \to \infty$ and that Ψ is a linear functional on U. Since U is of finite dimension, d say, then there exists basis $\nu_1, ..., \nu_d$ which spans U, and the linear functional Ψ can then be expressed as $\Psi\nu = \boldsymbol{B}\nu$ for some matrix $\boldsymbol{B}$. Thus Ψ is obviously continuous so 2. follows.

Now assume that 2. holds. Again we assume that $\dim(U) = d$. If the linear functionals in the set $\{\Psi_s : s \geq 0\}$ separates the points of U, we may find only d such linear functionals which separates the points of U. The reason is again that we may write the functionals Ψ_s as $\Psi_s(\nu) = \boldsymbol{B}_s\nu$ for some matrices $\boldsymbol{B}_s$. Since the dimension of U is d, then we can at most find d linearly independent such functional. Let $\psi_1, ..., \psi_d$ denote such linear functionals. Now define a linear mapping $C(\nu) = (\psi_1(\nu), ..., \psi_d(\nu))$. This functional C also separates the points of U. Obviously the functional is a mapping from U onto U, and if $\nu_1 \neq \nu_2$, then $C(\nu_1) \neq C(\nu_2)$ since assuming the contrary we get that $0 = C(\nu_1 - \nu_2) = (\psi_1(\nu_1 - \nu_2), ..., \psi_d(\nu_1 - \nu_2))$ so all $\psi_i(\nu_1 - \nu_2) = 0$ and hence $\nu_1 - \nu_2 = 0$ which is a contradiction. Hence C is a bijection, and C provides a coordinate system of U, the ψ_i acting as a basis for U. Hence 2. states that we have point wise convergence in this finite coordinate system, but this obviously implies convergence in the norm $\|\cdot\|$. □

Theorem 1.24. *The residual life operator $\{R_t\}_{t \geq 0}$ is an uniformly continuous operator on $V_R(\mu)$, i.e. for any $\nu \in V_R(\mu)$,*

$$R_t\nu \to \nu \quad as \quad t \downarrow 0,$$

where the convergence is in $\|\cdot\|$–norm.

Proof. This is a two–step proof. First we prove that we have convergence of the Laplace–Stieltjes transform of $R_t\nu$ to the Laplace–Stieltjes transform of ν, and next

we prove that this convergence of Laplace–Stieltjes transforms implies convergence in $\|\cdot\|$–norm.

The Laplace–Stieltjes transform of $R_t\nu$, $\widehat{R_t\nu}(s)$, is given by

$$\begin{aligned}\widehat{R_t\nu}(s) &= \int_0^\infty e^{-sx} R_t\nu(dx)\\ &= \left[-e^{-sx}\nu([x+t,\infty))\right]_0^\infty - \int_0^\infty se^{-sx}\nu([x+t,\infty))dx\\ &= \nu([t,\infty)) - s\int_0^\infty e^{-sx}\mu([x+t,\infty))dx.\end{aligned}$$

As $t\downarrow 0$, $\nu([x+t,\infty)) \to \nu([x,\infty))$ so by dominated convergence (ν being bounded by the total variation bound) we conclude that

$$\lim_{t\downarrow 0}\widehat{R_t\nu}(s) = \nu([0,\infty)) - s\int_0^\infty e^{-sx}\nu([x,\infty))dx = \hat{\nu}(s).$$

Next step in the proof is to establish the $\|\cdot\|$–norm convergence of $R_t\nu$ to ν. To this end we define a class of functionals Ψ_s on $V_R(\mu)$ by

$$\Psi_s(\nu) = \hat{\nu}(s),$$

where $\hat{\nu}(s)$ is the Laplace–Stieltjes transform of ν. Then $\{\Psi_s : s \geq 0\}$ separates the points of $V_R(\mu)$. Indeed, by the continuity theorem for Laplace(–Stieltjes) transforms, the transform characterizes the distributions uniquely and we conclude that if $\Psi_s(\nu) = \hat{\nu}(s) = 0$ for all $s \geq 0$ it is because $\nu = 0$. But then by Lemma 1.4, since $V_R(\mu)$ is of finite dimension, it follows that since $\Psi_s(R_t\nu) \to \Psi_s(\nu)$ as $t\downarrow 0$ for all s, then $R_t\nu \to \nu$ as $t\downarrow 0$ in $\|\cdot\|$–norm as well. □

Definition 1.11. Let $A : X \to Y$ denote a linear operator, where X and Y are Banach spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ respectively. The range of A is $\mathrm{Im}(A) = \{y \in Y | \exists x \in X : Ax = y\}$. Then the inverse operator to A is defined as the linear operator from $A^{-1} : \mathrm{Im}(A) \to X$ such that $A^{-1}(y) = x$ where x is the solution to $Ax = y$.

Invertibility of A is then equivalent to A being one–to–one.

Lemma 1.5. *If $\|Ax\|_Y \geq \|x\|_Y$ for some $c > 0$, then A is invertible.*

Proof. Simply notice that $\|Ax\|_Y = 0$ implies that $\|x\|_X = 0$ and hence $x = 0$, and since A is linear this is the same as the operator being one–to–one. □

Lemma 1.6. *If $A : X \to X$ is a linear operator on the Banach space X with norm $\|\cdot\|$ and let I denote the identity operator, i.e. $Ix = x$. If $\|A - I\| < 1$ then A is invertible with inverse*

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

Proof. Let $1 - c = \|A - I\| < 1$ (implying that $c > 0$). Then

$$\|x\| = \|x - Ax + Ax\| \leq \|(I-A)x\| + \|Ax\| \leq (1-c)\|x\| + \|Ax\|$$

implying that $\|Ax\| \geq c\|x\|$. Hence by Lemma 1.5 A is invertible.

To prove the explicit form of A^{-1} we proceed as follows. Recall that the operator norm is defined by (1.28), so for two operators A and B we have that

$$\|ABx\| \leq \|A(Bx)\| \leq \|A\|\|Bx\| \leq \|A\|\|B\|\|x\|$$

so

$$\|AB\| \leq \|A\|\|B\|.$$

Hence $\|(I-A)^n\| \leq \|I-A\|^n$. Let

$$B_n = \sum_{i=0}^{n} (I-A)^i.$$

Then for $m < n$ we have that

$$\begin{aligned}
\|B_n - B_m\| &= \| \sum_{i=m+1}^{n} (I-A)^i \\
&\leq \sum_{i=m+1}^{n} \|(I-A)^i\| \\
&\leq \sum_{i=m+1}^{n} \|I-A\|^n \\
&\leq \sum_{i=m+1}^{n} (1-c)^i
\end{aligned}$$

and since $0 \leq 1-c < 1$ the last sum converges to 0 as $m, n \to \infty$. Hence $\{B_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X, and since X is a Banach space and thus complete, there exists a limit B such that $B_n \to B$ as $n \to \infty$ in $\|\cdot\|$–norm. Furthemore, $B = \sum_{n=0}^{\infty}(I-A)^n$.

To see that B coincides with the proposed form of A^{-1} we let $y \in X$ and $x = By$. then we must show that $Ax = y$. But

$$\begin{aligned}
(I-A)x &= (I-A)By \\
&= (I-A)\sum_{n=0}^{\infty}(I-A)^n y \\
&= \sum_{n=1}^{\infty}(I-A)^n y \\
&= \sum_{n=0}^{\infty}(I-A)^n y - y \\
&= By - y = x - y,
\end{aligned}$$

from which is follows that $Ax = y$ as was to be proved. □

Theorem 1.25. *Assume that μ is a signed measure such that $V_R(\mu)$ is finite dimensional. Then the residual life operator on $V_R(\mu)$ may be written as*

$$R_t = e^{\boldsymbol{G}t}$$

for some finite matrix $\boldsymbol{G}$. The matrix $\boldsymbol{G}$ is called the generator of the semi–group $\{R_t\}_{t\geq 0}$.

Proof. Consider the functional $V(t) = \int_0^t R_s ds$, meaning

$$V(t)\nu(A) = \int_0^t \mu(s+A)ds.$$

Define the identity functional $I = R_0$. Then

$$\begin{aligned} \|\frac{1}{t}V(t) - I\| &= \|\frac{1}{t}\int_0^t (R_s - I)ds\| \\ &\leq \frac{1}{t}\int_0^t \|R_s - I\|ds, \end{aligned}$$

which by the mean–value theorem converges to 0 as $t \downarrow 0$ by the continuity of R_t (here we invoke that R_t acts on the finite dimensional space $V_R(\mu)$). Then for t sufficiently small, $t \leq t_0$ say, define operator $W(t) = \frac{1}{t}V(t)$. Then

$$\|\frac{1}{t}V(t) - I\| < 1.$$

and hence the inverse operator to $W(t) := \frac{1}{t}V^(t)$ exists and is given by

$$W^{-1}(t) = \sum_{n=0}^{\infty} (I - W(t))^n.$$

Hence also $V(t)$ is invertible for $t \leq t_0$. Let t be such a number. Then

$$\begin{aligned} R_s &= V(t)^{-1}V(t)R_s \\ &= V^{-1}(t)\int_0^t R_u du\, R_s \\ &= V^{-1}(t)\int_0^t R_{u+s}du \\ &= V^{-1}(t)\int_s^{s+t} R_u du \\ &= V^{-1}(t)\left(V(s+t) - V(s)\right). \end{aligned}$$

Hence since R_s is continuous, then $V(t) = \int_0^t R_u du$ is differentiable and hence R_s is differentiable with derivative

$$\frac{d}{ds}R_s = \lim_{h\to 0}\frac{R_{s+h} - R_s}{h}$$

$$= \lim_{h \downarrow 0} \frac{R_h - I}{h} R_s$$
$$= R_0' R_s$$

from which it follows that $R_s = e^{\boldsymbol{G}s}$ where $\boldsymbol{G} = R_0'$. The matrix $\boldsymbol{G}$ has been discussed earlier in (1.30). □

We can now prove one of the main results of this section.

Theorem 1.26. *A signed measure ν on $[0,\infty)$ has a rational Laplace–Stieltjes transform if and only if $V_R(\mu) = span\{R_t\mu : t \geq 0\}$ is of finite dimension.*

Proof. If μ has a rational Laplace transform, then $V_R(\mu)$ is of finite dimension as it follows directly from Theorem 1.23.

We now suppose that $V_R(\mu)$ is of finite dimension. Then the residual life operator acting on $V_R(\mu)$ is on the form $R_t = \exp(\boldsymbol{G}t)$ for some finite matrix $\boldsymbol{G}$. Let s_0 denote the eigenvalue of $\boldsymbol{G}$ the real part of which is the largest among all the eigenvalues of $\boldsymbol{G}$. Now consider the Laplace–Stieltjes transform $\hat{\mu}$ of μ. Then

$$\begin{aligned}
\hat{\mu}(s) &= \int_0^\infty e^{-sx}\mu(dx) \\
&= \left[-e^{-sx}\mu([x,\infty))\right]_0^\infty - \int_0^\infty s e^{-sx}\mu([x,\infty))dx \\
&= \mu([0,\infty)) - s\int_0^\infty s^{-sx}\mu([x,\infty))dx \\
&= \mu([0,\infty)) - s\int_0^\infty s^{-sx}R_x\mu([0,\infty))dx.
\end{aligned}$$

Now define the linear functional on the signed measures, $\phi : \mathscr{S} \to \mathbb{R}$, by $\phi(\nu) = \nu((0,\infty))$. Then for $s > s_0$,

$$\begin{aligned}
\hat{\mu}(s) &= \mu([0,\infty)) - s\int_0^\infty s^{-sx}R_x\mu([0,\infty))dx \\
&= \mu([0,\infty)) - s\int_0^\infty s^{-sx}\phi(R_x\mu)dx \\
&= \mu([0,\infty)) - s\phi\left(\int_0^\infty e^{-sx}R_x\mu dx\right) \\
&= \mu([0,\infty)) - s\phi\left(\int_0^\infty e^{-sx}e^{\boldsymbol{G}x}\mu dx\right) \\
&= \mu([0,\infty)) - s\phi\left(\int_0^\infty e^{(-s\boldsymbol{I}+\boldsymbol{G}))x}\mu dx\right) \\
&= \mu([0,\infty)) - s\phi\left((s\boldsymbol{I}-\boldsymbol{G})^{-1}\mu\right)
\end{aligned}$$

the latter obviously being a rational function in s. A few comments on this derivation. If λ is an eigenvalue for $\boldsymbol{G}$, then $\lambda - s < 0$ (since $s > s_0$) is an eigenvalue for $-s\boldsymbol{I} + \boldsymbol{G}$, so the following integral exists and is given by

$$\int_0^\infty e^{(-s\boldsymbol{I}+\boldsymbol{G})x}dx = \left[-(s\boldsymbol{I}-\boldsymbol{G})^{-1}e^{-(s\boldsymbol{I}-\boldsymbol{G})x}\right]_0^\infty = (s\boldsymbol{I}-\boldsymbol{G})^{-1}.$$

Now if the dimension of $V_R(\mu)$ is d (same dimension as $\boldsymbol{G}$!) and $\mu_1,...\mu_d$ is a basis for $V_R(\mu)$, then

$$\mu = c_1\mu_1 + ... + c_d\mu_d$$

and

$$(s\boldsymbol{I}-\boldsymbol{G})^{-1}\mu = r_1(s)\mu_1 + ...r_d(s)\mu_d,$$

where $r_i(s)$ are rational functions in s, so

$$\phi\left((s\boldsymbol{I}-\boldsymbol{G})^{-1}\mu\right) = r_1(s)\phi(\mu_1) + ... + r_d(s)\phi(\mu_d),$$

which is hence a rational function in s. Hence $\hat{\mu}(s)$ is a rational function in s as well.

This calculation above holds for $s > s_0$, which was necessary in order to make the integral converge (which can be seen by diagonalizing the matrix exponential into a Jordan form of the eigenvalues). On the other hand it is clear, that the resulting formula involving the inverse of $s\boldsymbol{I}-\boldsymbol{G}$ is valid for all $s \geq 0$ apart for some possible singular values.

We now claim that $\hat{\mu}(s)$ can be defined by this formula for all$s$ such that $\mathrm{Re}(s) \geq 0$ by the use of analytic continuation, where as usual $\mathrm{Re}(s)$ denotes the real part of s. Indeed, Laplace(–Stieltjes) transforms are analytic functions in the positive half–plane except at singularities, and rational functions are analytic and can be extended to the whole complex plane except for the their singularities. Hence by uniqueness of analytic continuation, the Laplace transform is a rational function for arguments with real parts larger or equal to zero. □

The matrix $\boldsymbol{G}$ in Theorem 1.25 is the restriction of the infinitesimal generator Γ of the semi–group $\{R_t\}_{t\geq 0}$ to $V_R(\mu)$. The infinitesimal generator Γ is defined as the operator

$$\Gamma = \lim_{h\downarrow 0}\frac{R_h - R_0}{h}. \tag{1.31}$$

Exercises

1.1. Let X be a non–negative random variable with a rational Laplace transform. Prove that if $p_0 \neq 0$ then the distribution has an atom (a probability mass) at zero. Prove that X cannot have atoms at points other than zero.

1.2. Prove that the Laplace transform of the n'th derivative of a function f is given by

$$\mathscr{L}(f^{(n)}(t),s) = s^n\mathscr{L}(f,s) - s^{n-1}f(0) - s^{n-2}f'(0) - ... - f^{(n-1)}(0).$$

1.3. Prove that the Laplace transform of the function $f(t) = t^n$ is given by is given by $\Gamma(n+1)/s^{n+1}$.

1.4. Let $g(z) = z^n + q_1 z^{n-1} + ... + q_n$. Prove that the roots of g coincide with the eigenvalues of the Companion matrix

$$\boldsymbol{C} = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ 0 & 0 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_n & -q_{n-1} & -q_{n-2} & \cdots & -q_1 \end{pmatrix}.$$

Furthermore, find the corresponding eigenvectors.

1.5. Consider an irreducible rational Laplace transform $L_X(s)$ and the corresponding companion matrix $\boldsymbol{C}$. Prove that

$$\int_0^\infty x^n e^{\boldsymbol{C}x} dx = n!(-\boldsymbol{C})^{-(n+1)}, \;\; n = 0,1,2,....$$

Conclude that the formulas (1.9) and (1.10) are valid.

1.6. Prove that $f(x) = x^{n-1}e^{-ax}$ has Laplace transform

$$\hat{f}(s) = \frac{(n-1)!}{(s+a)^n},$$

and that the formula is valid whenever $\mathrm{Re}(s) > \mathrm{Re}(a)$.

1.7. Prove that the Laplace transform of $f_1(x) = x^{n-1}\cos(ax)$ is given by

$$\hat{f}_1(s) = (n-1)!\mathrm{Re}\left(\frac{(s+\mathrm{i}a)^n}{(s^2+a^2)^n}\right)$$

while the Laplace transform of $f_2(x) = x^{n-1}\sin(ax)$ is

$$\hat{f}_2(s) = (n-1)!\mathrm{Im}\left(\frac{(s+\mathrm{i}a)^n}{(s^2+a^2)^n}\right).$$

Hint: Use Exercise 1.6, replacing a with $-\mathrm{i}a$.

1.8. Prove that the function

$$g(s) = \sum_{i=1}^{k} [\bar{p}_i \sin(w_i s) + \bar{q}_i \cos(w_i s)]$$

is periodic. Show that its mean $M(g)$,

$$M(g) = \lim_{T\to\infty} \frac{1}{T} \int_a^{a+T} g(s)ds$$

is well defined, i.e. does not depend on a, and that $M(g) = 0$. Conclude that g must both take positive and negative values.

1.9. Consider matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Show that the matrix–exponential $\exp(A+B) = \exp(A)\exp(B)$ if A and B commute. Conclude that

$$e^{-\lambda x}e^{\boldsymbol{A}x} = \exp\left(-(\lambda \boldsymbol{I} - \boldsymbol{S})x\right).$$

1.10. Consider the Jordan block

$$\boldsymbol{J} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ 0 & 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}.$$

Calculate $\exp(\boldsymbol{J}x)$.

1.11. Consider the density function

$$f(x) = ce^{-ax}(1-\cos(bx)).$$

Find the constant c. Calculate the Laplace transform of the distribution, show that the dominant eigenvalue is real and equals $-a$. Find the companion matrix $\boldsymbol{C}$ and write f on the form (1.8).

1.12. Prove that if $\boldsymbol{S}$ is an invertible matrix, then

$$\int e^{\boldsymbol{S}x}dx = \boldsymbol{S}^{-1}e^{\boldsymbol{S}x}.$$

1.13. Let $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ and $(\boldsymbol{\beta}, \boldsymbol{T}, \boldsymbol{t})$ be two representations for the same matrix–exponential distribution. Prove that

$$\boldsymbol{\alpha}\boldsymbol{S}^n\boldsymbol{s} = \boldsymbol{\beta}\boldsymbol{T}^n\boldsymbol{t} \quad \forall n \in \mathbb{N}.$$

1.14. Consider the matrix–exponential distribution with representation

$$\boldsymbol{\alpha} = (1\ 0\ 0),\ \boldsymbol{S} = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -2 & 2 \\ 1 & 0 & -2 \end{pmatrix},\ \boldsymbol{s} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Calculate its Laplace transform and find the regular C–fraction expansion.

1.15. Prove the formula (1.24).

1.16. Consider the density function

$$f(x) = ce^{-ax}(1-\cos(bx)).$$

Calculate its (reduced) moments and determinate its minimum dimension according to the Hankel matrix criterium (Theorem 1.17).

1.17. (Hahn decomposition)

1.18. Prove that the residual life operator defined by (**??**) satisfies the semi–group property, $R_t R_s \mu = R_s R_t \mu = R_{t+s}\mu$.

1.19. Let P_1 and P_2 be two R–invariant subsets of the set of signed measures. Prove that $P_1 \cap P_2$ and $P_1 \cup P_2$ are both R–invariant.

Chapter 2
Markov processes and their properties

2.1 Introduction

The reason for including a chapter on Markov chains and processes is partly because large parts of the book depend on a precise knowledge of certain parts of this theory and to our opinon easier to obtain a streamlined treatment providing the necessary background material. The other reason for writing this chapter is that much of the theory is rather specialized and not easily available in standard textbooks. This is for example the case for the section where we calculate the joint distribution of the number of jumps and holding times in a Markov jump process with finitely many states. The section on the Karlin–MacGregor theory on coincidence probabilities is another example.

The use of probabilistic reasoning is used throughout in the rest of the book whenever possible and this chapter is written in this spirit. The reader may benefit from an intuitively appealing presentation of the material, whether new to the subject or brushing up key results, where sample paths arguments are preferred to more analytic methods involving e.g. transition kernels and semi–groups.

2.2 The Poisson process

The Poisson process plays an important role as a basic building block for the further development of the theory, in particular concerning phase–type distributions and random environments such as the Markov modulated Poisson process.

Consider the times of arrival of certain events, $0 < S_1 < S_2 < ...$ and let $N(t)$ be the number of arrivals observed in $(0,t]$, that is

$$N(t) = \sup\{n \in \mathbb{N} | S_n \leq t\} = \sum_{n=0}^{\infty} 1\{S_n \leq t\},$$

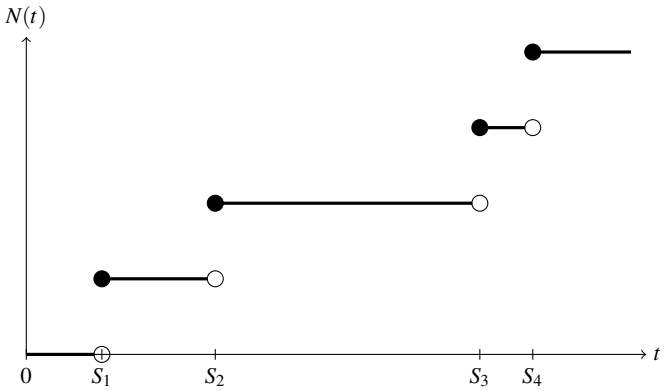

Fig. 2.1 A sample path $t \to N(t)$ from a counting process. It is easy to see that $N(t)$ counts the number of arrivals prior to time t.

where we for convenience let $S_0 = 0$. Then $N(t)$ are random variables and the collection $\{N(t)\}_{t\geq 0}$ a stochastic process. This type of process is called a counting process since it increases step–wise with unit increments (see Figure 2.1). We shall use the notation

$$N(A) = \sum_{n=0}^{\infty} 1\{S_n \in A\}$$

for any (measurable) set A.

Definition 2.1. $\{N(t)\}_{t\geq 0}$ has independent increments if for all $n \in \mathbb{N}$ and all $0 < s_1 < s_2 < ... < s_n$ the random variables $N(s_1), N(s_2) - N(s_1), ..., N(s_n) - N(s_{n-1})$ are independent.

Definition 2.2. $\{N(t)\}_{t\geq 0}$ has stationary increments if for all $n \in \mathbb{N}$, all $0 < s_1 < s_2 < ... < s_n$ and $h \geq 0$

$$(N(s_1+h), N(s_2+h) - N(s_1+h), ..., N(s_n+h) - N(s_{n-1}+h))$$

has a distribution which does not depend on h.

We shall use the common notation that $o(h)$ denotes any function for which $\lim_{h\to 0} \frac{o(h)}{h} = 0$.

Definition 2.3. $\boldsymbol{N} = \{N(t)\}_{t\geq 0}$ is a Poisson process with intensity $\lambda > 0$ if

(i) $\boldsymbol{N}$ has independent and stationary increments.
(ii) $\mathbb{P}(N(h) = 0) = 1 - \lambda h + o(h)$.
(iii) $\mathbb{P}(N(h) = 1) = \lambda h + o(h)$.

As an immediate consequence $\mathbb{P}(N(h) \geq 2) = o(h)$. In the following we shall frequently make use of the infintesimal notation like $\mathbb{P}(N(t, t+dt) = 1) = \lambda dt$ meaning $\mathbb{P}(N(t, t+h) = 1) = \lambda h + o(h)$ as $h \to 0$. Hence in infinitesimal notation we

treat terms of the order $o(h)$ as being zero. This is usually not dangerous since in most arguments we shall conclude the derivation by dividing by h and taking the limit $h \to 0$.

Define the times between arrivals $T_1 = S_1, T_2 = S_2 - S_1, ..., T_n = S_n - S_{n-1}, ...$

Theorem 2.1. *The following statements are equivalent*

(i) $\{N(t)\}_{t\geq 0}$ *is a Poisson process with intensity* λ.
(ii)$\{N(t)\}_{t\geq 0}$ *has independent increments and* $N(t) \sim Po(\lambda t)$ *for all* t.
(iii) $T_1, T_2, ...$ *are i.i.d. (identically and independently distributed)* $\sim \exp(\lambda)$.

Proof. $(i) \Longrightarrow (ii)$: Define $p_n(t) = \mathbb{P}(N(t) = n)$. Then

$$\begin{aligned} p_n(t+dt) &= \mathbb{P}(N(t+dt) = n) \\ &= \mathbb{E}\left(\mathbb{P}(N(t+dt) = n | N(t))\right) \\ &= p_{n-1}(t)\lambda dt + p_n(t)(1-\lambda dt) \end{aligned}$$

implying

$$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t).$$

Let $G(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n$ for complex z. Then $G(z,t) = \mathbb{E}(z^{N(t)})$ and

$$\begin{aligned} \frac{\partial}{\partial t} G(z,t) &= \sum_{n=0}^{\infty} p_n'(t) z^n \\ &= \sum_{n=0}^{\infty} \left(-\lambda p_n(t) + \lambda p_{n-1}(t)\right) z^n \\ &= -\lambda G(z,t) + \lambda z G(z,t) \\ &= (\lambda z - \lambda) G(z,t). \end{aligned}$$

The inital condition $G(z,0) = \mathbb{E}(z^N(0)) = 1$ implies the following solution

$$G(z,t) = \exp((\lambda z - \lambda)t).$$

This is exactly the generating function $\mathbb{E}(z^N)$ of a random variable having a Poisson distribution with parameter λt. Since generating functions characterize discrete disctributions we conclude that $N(t)$ must have son distribution with parameter λt.

$(ii) \Longrightarrow (iii)$: First we prove that $\{N(t)\}_{t\geq 0}$ has stationary increments. Since $N(t+h) \sim \text{Po}(\lambda(t+h))$ we have that

$$\begin{aligned} e^{(\lambda z - \lambda)(t+h)} &= \mathbb{E}\left(z^{N(t+h)}\right) \\ &= \mathbb{E}\left(z^{N(t+h)-N(h)} z^{N(h)}\right) \\ &= \mathbb{E}\left(z^{N(t+h)-N(h)}\right) \mathbb{E}\left(z^{N(h)}\right) \end{aligned}$$

$$= \mathbb{E}\left(z^{N(t+h)-N(h)}\right) e^{(\lambda z-\lambda)h}$$

giving

$$\mathbb{E}\left(z^{N(t+h)-N(h)}\right) = e^{(\lambda z-\lambda)t}$$

from which it follows that $N(t+h)-N(h) \sim \text{Po}(\lambda t)$. Hence the distribution of $N(t+h)-N(h)$ does not depend on h. By the independent increment property it hence follows that the process stationary increments.

Next we calculate the joint density f of the times between arrivals. Let $t_0 = 0 \leq s_1 < t_1 \leq s_2 < t_2 \leq s_3 < t_3 \leq ... \leq s_n < t_n$. Then

$$\begin{aligned}
&\mathbb{P}(s_k < S_k < t_k, k = 1, ..., n) \\
&= \mathbb{P}(N(t_{k-1}, s_k) = 0, N(s_k, t_k) = 1, k = 1, ..., n-1, \\
&\quad N(t_{n-1}, s_n) = 0, N(s_n, t_n) \geq 1).
\end{aligned}$$

This follows since the event $S_n \in [s_n, t_n]$ does not exclude that $S_m \in [s_n, t_n]$ for $m \geq n+1$. This phenomenon does, however, not occur in the preceeding $n-1$ interarrivals since by construction the arrivals are positioned in disjoint intervals.

Using $N(a,b) = N(b) - N(a)$ and the independent and stationary increments we get that

$$\begin{aligned}
&\mathbb{P}(s_k < S_k \leq t_k, k = 1, ..., n) \\
&= (1 - e^{-\lambda(t_n - s_n)}) e^{-\lambda(s_n - t_{n-1})} \prod_{k=1}^{n-1} \left(\lambda(t_k - s_k) e^{-\lambda(t_k - s_k)} e^{-\lambda(s_k - t_{k-1})}\right) \\
&= (e^{-\lambda s_n} - e^{-\lambda t_n}) \lambda^{n-1} \prod_{k=1}^{n-1} (t_k - s_k)
\end{aligned}$$

Now,

$$\int_{s_n}^{t_n} e^{-\lambda x} dx = \frac{1}{\lambda}(e^{-\lambda s_n} - e^{-\lambda t_n})$$

gives

$$\begin{aligned}
\mathbb{P}(s_k < S_k \leq t_k, k = 1, ..., n) &= \lambda^n \prod_{k=1}^{n-1} (t_k - s_k) \int_{s_n}^{t_n} e^{-\lambda y_n} dy_n \\
&= \int_{s_1}^{t_1} ... \int_{s_{n-1}}^{t_{n-1}} \int_{s_n}^{t_n} \lambda^n e^{-\lambda y_n} dy_n dy_{n-1} ... dy_1.
\end{aligned}$$

Hence the joint density of $(S_1, ..., S_n)$ is given by

$$f_{(S_1,...,S_n)}(y_1, ..., y_n) = \begin{cases} \lambda^n \exp(-\lambda y_n) \text{ si } 0 \leq y_1 < y_2 < ... < y_n \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

In order to calcualte the density of $(T_1, T_2, ..., T_n)$ we make use of a standard transformation argument. If $g : (S_1, S_2, ..., S_n) \to (S_1, S_2 - S_1, ..., S_n - S_{n-1})$ then g is a linear transformation given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & ... & 0 & 0 \\ -1 & 1 & 0 & 0 & ... & 0 & 0 \\ 0 & -1 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & -1 & 1 & ... & 0 & 0 \\ . & . & . & . & ... & . & . \\ 0 & 0 & 0 & 0 & ... & -1 & 1 \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ . \\ S_n \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 - S_1 \\ S_3 - S_2 \\ S_4 - S_3 \\ . \\ S_n - S_{n-1} \end{pmatrix}.$$

The Jacobian of g is coefficient matrix above ($\boldsymbol{T}$, say) and has inverse

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

Its determinant $\det(\boldsymbol{T}^{-1}) = 1$ and the transformation theorem states that

$$f_{(T_1,...,T_n)}(x_1, ..., x_n) = \frac{f_{(S_1,....,S_n)}(g^{-1}(x_1, ..., x_n))}{J_g(g^{-1}(x_1, ..., x_n))}.$$

The denominator is 1 and

$$f_{(T_1,....,T_n)}(x_1, ..., x_n) = f_{(S_1,....,S_n)}(x_1, x_1 + x_2, ..., x_1 + ... + x_n)$$

for all $x_1, x_2, ..., x_n \geq 0$. Hence

$$f_{(T_1,...,T_n)}(x_1, ..., x_n) = \lambda^n \exp(-\lambda(x_1 + ... + x_n)) = \prod_{k=1}^{n} \lambda e^{-\lambda x_k}$$

which is the product of densities of exponential distributions. Hence $T_1, T_2, ...$ are i.i.d. $\sim \exp(\lambda)$.

$(iii) \implies (i)$: If $T_1 \sim \exp(\lambda)$ then $\mathbb{P}(T > h) = \exp(-\lambda h) = 1 - \lambda h + o(h)$ by Taylor expansion, but this is the same as $\mathbb{P}(N(h) = 0)$. Similarly, $\mathbb{P}(T_1 \leq h) = 1 - \exp(-\lambda h) = \lambda h + o(h)$. The memoryless property of the exponential distributions implies that the increments are both independent and stationary. □

2.3 Markov chains

Let $\boldsymbol{X} = \{X_n\}_{n\in\mathbb{N}}$ be a stochastic process taking values in some discrete set E. We shall refer to E as the state–space of the process.

Definition 2.4. $\{X_n\}_{n\in\mathbb{N}}$ is called a *Markov chain* if

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

for all $n \in \mathbb{N}, i_0, ..., i_{n-1}, i, j \in E$. Furthermore, $\{X_n\}_{n\geq 0}$ is said to be *time homogeneous* if the probabilities $\mathbb{P}(X_{n+1} = j|X_n = i)$ do not depend on n. In this case we write

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

and refer to p_{ij} as the *transition probability* of going from state i to state j. The transition matrix $\boldsymbol{P}$ of a Markov chain $\{X_n\}_{n\geq 0}$ is defined by

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

The distribution of a Markov chain is characterized by its initial distribution and its transition probabilities as shows the next theorem.

Theorem 2.2. *$\{X_n\}_{n\geq 0}$ is a time–homogeneous Markov chain if and only if*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = \mathbb{P}(X_0 = i_0)p_{i_1 i_2} p_{i_2 i_3} ... p_{i_{n-1} i_n}.$$

for all $n \in \mathbb{N}$.

Proof.

$$\begin{aligned}
&\mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_n = i_n) \\
&\quad = \mathbb{P}(X_n = i_n|X_{n-1} = i_{n-1}, ..., X_0 = i_0)\mathbb{P}(X_{n-1} = i_{n-1}, ..., X_0 = i_0) \\
&\quad = \mathbb{P}(X_n = i_n|X_{n-1} = i_{n-1})\mathbb{P}(X_{n-1} = i_{n-1}, ..., X_0 = i_0) \\
&\quad = p_{i_{n-1} i_n}\mathbb{P}(X_{n-1} = i_{n-1}, ..., X_0 = i_0) \\
&\quad = ... \\
&\quad = p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} ... p_{i_0 i_1}\mathbb{P}(X_0 = i_0).
\end{aligned}$$

□

Let $\mathscr{F}_n = \sigma(X_0.X_1, ..., X_n)$ be the σ–algebra generated by the Markov chain and let $\mathscr{F}_\infty = \sigma(X_0, X_1, ...)$. Then

Theorem 2.3. *$\{X_n\}_{n\geq 0}$ is a Markov chain if and only if*

$$\mathbb{E}(f(X_n, X_{n+1}, ...)|\mathscr{F}_n) = \mathbb{E}(f(X_0, X_1, ...)|X_0 = X_n) \tag{2.1}$$

for every bounded and measurable function $f : E \times E \times ... \to \mathbb{R}$.

Proof. If (2.1) holds, we simply take $f(X_0, X_1, ...) = \prod_j 1_{\{X_j = i_j\}}$. Now suppose that$\{X_n\}_{n\geq 0}$ is a Markov chain. Then (2.1) follows from a standard argument in measure theory. Since f is bounded and measurable we can write $f = f_+ - f_-$ where f_+ and f_- are bounded, positive and measurable. Also f_+ and f_- can be obtained as an increasing limit of simple functions (linear combinations of indicator functions). Since (2.1) by assumption holds for indicator functions, by linearity of the expectation it also holds for simple functions. By the monotone convergence theorem (2.1) then holds for f_+ and f_- and hence for f. □

Checking the Markov property for particular cases may be a tedious task, but fortunately many chains are constructed in the following way.

Theorem 2.4. *Processes on the form*

$$X_{n+1} = f(X_n, Z_{n+1})$$

where X_0 is independent of $\{Z_n\}_{n\in\mathbb{N}}$ and where $Z_1, Z_2, ...$ are independent and identically distributed (i.i.d.) defines a time–homogeneous Markov chain.

Proof. Left to the reader. □

Theorem 2.5. *The n'th power of the transition matrix, $\boldsymbol{P}^n = \left(p_{ij}^{(n)}\right)_{i,j\in E}$, is a transition matrix for the Markov chain $\{X_{kn}\}_{k\geq 0}$.*

Proof. We must prove that

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

First we notice that

$$\mathbb{P}(X_n = j, X_0 = i) = \sum_{k=1}^{n-1} \sum_{j_k\in E} \mathbb{P}(X_0 = i, X_1 = j_1, ..., X_{n-1} = j_{n-1}, X_n = j).$$

Then from Theorem 2.2 we have that

$$\mathbb{P}(X_n = j, X_0 = i) = \sum_{k=1}^{n-1} \sum_{j_k\in E} \mathbb{P}(X_0 = i) p_{j_1 j_2} p_{j_2 j_3} ... p_{j_{n-1} j}.$$

Hence

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{k=1}^{n-1} \sum_{j_k\in E} p_{i j_1} p_{j_1 j_2} ... p_{j_{n-2} j_{n-1}} p_{j_{n-1} j}.$$

The right hand side of this expression is exactly the ij'th element of $\boldsymbol{P}^n$. □

The next result is the celebrated Chapman–kolmogorov equation.

Theorem 2.6. *The n–step transition probabilities $p_{ij}^{(n)}$ satisfy*

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

Proof. Easy consequence of the Markov property. □

The next important step in the development of Markov chains is to ensure that the Markov property also holds at certain random times, e.g. first hitting times. To this end we shall introduce the concept of stopping times.

Definition 2.5. A stopping time τ for the Markov chain $\{X_n\}_{n\geq 0}$ is a random variable taking values in $\mathbb{N}\cup\{+\infty\}$ with the property that $\{\tau = n\} \in \mathscr{F}_n$ for all n.

The σ–algebra $\mathscr{F}_\tau$ is defined by the relation

$$A \in \mathscr{F}_\tau \iff A\cap\{\tau = n\} \in \mathscr{F}_n \ \ \forall n \in \mathbb{N}\cap\{+\infty\}.$$

We now prove what is referred to as the *strong Markov property*.

Theorem 2.7. *Let τ be a stopping time and let $h : E\times E\times ... \to \mathbb{R}$ be a measurable function. Then on $\{\tau < \infty\}$,*

$$\mathbb{E}\left(h(X_\tau, X_{\tau+1}, ...)|\mathscr{F}_\tau\right) = \mathbb{E}\left(h(X_0, X_1, ...)|X_0 = X_\tau\right).$$

Proof. We use the (measure theoretic) definition of conditional expectation. In order to prove the identity we must prove that the right hand side formula satisfies the definition of the conditional expectation provided by the left hand side, i.e. that $\mathbb{E}\left(h(X_0, X_1, ...)|X_0 = X_\tau\right)$ is $\mathscr{F}_\tau$–measurable and that

$$\int_A h(X_\tau, X_{\tau+1}, ...)dP = \int_A \mathbb{E}\left(h(X_0, X_1, ...)|X_0 = X_\tau\right)dP \ .$$

$\forall A \in \mathscr{F}_\tau \cap \{\tau < \infty\}$.

The measurability is obvious. Furthermore, as $A \in \mathscr{F}_\tau$, we have that $A\cap\{\tau = n\} \in \mathscr{F}_n$ for all n, and since we are on $\{\tau < \infty\}$, then

$$A = \bigcup_{n=0}^{\infty} \left(A\cap\{\tau = n\}\right).$$

But,

$$\begin{aligned}
\int_{A\cap\{\tau=n\}} h(X_\tau, X_{\tau+1}, ...)dP &= \int_{A\cap\{\tau=n\}} h(X_n, X_{n+1}, ...)dP \\
&= \int_{A\cap\{\tau=n\}} \mathbb{E}\left(h(X_n, X_{n+1}, ...)|\mathscr{F}_n\right)dP \\
&= \int_{A\cap\{\tau=n\}} \mathbb{E}\left(h(X_0, X_1, ...)|X_0 = X_n\right)dP \\
&= \int_{A\cap\{\tau=n\}} \mathbb{E}\left(h(X_0, X_1, ...)|X_0 = X_\tau\right).
\end{aligned}$$

Summing over n then yields the result. □

Definition 2.6.

$$T_i = \inf\{n \geq 1 | X_n = i\}$$
$$N_i = \sum_{j=1}^{\infty} 1_{\{X_j = i\}}$$

T_i is the first entrance time to state i and N_i is the number of visits to state i. In both cases the intial state at time $n = 0$ is not included.

We shall use the following notation

$$\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot | X_0 = i), \;\; \mathbb{E}_i(\cdot) = \mathbb{E}(\cdot | X_0 = i).$$

Definition 2.7. A state i of a Markov chain $\{X_n\}_{n\geq 0}$ is called recurrent if $\mathbb{P}_i(T_i < \infty) = 1$. A state which is not recurrent is called transient.

Theorem 2.8. *Let i denote any state. Then following statements are equivalent:*

1. i is recurrent.
2. $N_i = \infty$ a.s..
3. $\mathbb{E}(N_i) = \sum_{m=1}^{\infty} p_{ii}^{(m)} = \infty$.

Proof. Let

$$T_i^n = \inf\{n > T_i^{n-1} | X_n = i\}, \; T_i^1 = T_i$$

be the times of successive visits to state i. Then

$$\begin{aligned}
\mathbb{P}_i(T_i^{k+1} < \infty) &= \mathbb{P}_i(T_i^{k+1} < \infty, T_i^k < \infty) \\
&= \mathbb{E}_i\left(\mathbb{P}_i(T_i^{k+1} < \infty, T_i^k < \infty | \mathscr{F}_{T_i^k})\right) \\
&= \mathbb{E}_i\left(I\{T_i^k < \infty\}\mathbb{P}_i(T_i^{k+1} < \infty | \mathscr{F}_{T_i^k})\right) \;\; \text{(measurable)} \\
&= \mathbb{E}_i\left(I\{T_i^k < \infty\}\mathbb{P}_{X_{T_i^k}}(T_i^1 < \infty)\right) \;\; \text{(strong Markov)} \\
&= \mathbb{E}_i\left(I\{T_i^k < \infty\}\mathbb{P}_i(T_i^1 < \infty)\right) \;\; (X_{T_i^k} = i) \\
&= \mathbb{P}_i(T_i^1 < \infty)\mathbb{P}_i\left(T_i^k < \infty\right) \\
&= \ldots \\
&= \mathbb{P}_i(T_i^1 < \infty)^{k+1}
\end{aligned}$$

If i is recurrent then $\mathbb{P}_i(T_i^{k+1} < \infty) = 1$ for all k, i.e. $T_i^k < \infty$ a.s. for all k. As

$$N_i = \sup\{k \in \mathbb{N} | T_i^k < \infty\}$$

we than have that $N_i = \infty$ a.s. It is clear that if $N_i = \infty$ a.s.. then $\mathbb{E}_i N_i = \infty$. We need, however, to verify the expression

$$\mathbb{E}_i N_i = \sum_{m=1}^{\infty} p_{ii}^{(m)}.$$

This follows directly from (Beppo–Levi)

$$\begin{aligned}\mathbb{E}_i(N_i) &= \mathbb{E}_i\left(\sum_{n=1}^{\infty} I\{X_n = i\}\right)\\ &= \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = i)\\ &= \sum_{n=1}^{\infty} p_{ii}^{(n)}.\end{aligned}$$

Finally we prove that 3. implies 1., or equivalently, its negation. Suppose that i is transient. Then $\mathbb{P}_i(T_i < \infty) < 1$ and

$$\begin{aligned}\mathbb{E}_i(N_i) &= \sum_{n=0}^{\infty} \mathbb{P}_i(N_i > n)\\ &= \sum_{n=1}^{\infty} \mathbb{P}_i(T_i^n < \infty)\\ &= \sum_{n=1}^{\infty} \mathbb{P}_i(T_i < \infty)^n\\ &< \infty.\end{aligned}$$

□

Corollary 2.1. *The following statements are equivalent:*

1. *i is transient*
2. *$N_i < \infty$ a.s.*
3. *$\mathbb{E}(N_i) = \sum_{m=1}^{\infty} p_{ii}^{(m)} < \infty$.*

Definition 2.8. A state i leads to a state j if there exists a $m \in \mathbb{N}$ such that $p_{ij}^{(m)} > 0$ and we write $i \to j$. Two states i and j communicate if $i \to j$ and $j \to i$ in which case we write $i \sim j$.

The relation $\sim$ defines an equivalence relation on the state–space E, i.e. $\sim$ satisfies that $i \sim i$, $i \sim j \Leftrightarrow j \sim i$ and $i \sim j$ and $j \sim k$ implies $i \sim k$. The equivalence relation parts the state–space E into disjoint equivalence classes. We now investigate the nature of these classes.

Theorem 2.9. *If i is recurrent and $i \sim j$ then j is again recurrent.*

Proof. Let n_1, n_2 be such that $p_{ij}^{(n_1)} > 0$ and $p_{ji}^{(n_2)} > 0$. Then

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

where the inequality follows from picking out one particular term from the Chapman–Kolmogorov equation. Recurrence then follows from Theorem 2.8. □

Hence we conclude that if an equivalence class contains an recurrent state i, then all its states are recurrent. Suppose that i is transient and $i \sim j$. Then j must again be transient because if j were recurrent, then $i \sim j$ would imply i to be recurrent as well. Hence transience is a class property as well. Let T denote the set of transient states. They need not all communicate. Then we may split the state–space into disjoint equivalence classes $R_1, R_2, \ldots$ of recurrent states and T of transient states

$$E = T \cup R_1 \cup R_2 \cup \ldots.$$

Definition 2.9. A Markov chain is called irreducible if all its states communicate.

For an irreducible Markov chain we conclude that either all its states are recurrent or transient, and we call the chain either recurrent or transient respectively.

In the following we consider equilibrium properties of Markov chains.

Definition 2.10. A (row) vector $\boldsymbol{\nu} = (\nu_i)_{i \in E}$ is called a stationary measure of the Markov chain $\{X_n\}_{n \geq 0}$ if $\boldsymbol{\nu}$ is finite, non–negative, non–zero and

$$\boldsymbol{\nu} \boldsymbol{P} = \boldsymbol{\nu}.$$

The last condition implies that if the distribution of X_n is $\boldsymbol{\nu}$ then so is the distribution of X_{n+1}.

Theorem 2.10. *If a state i is recurrent then we can define a stationary measure in $\boldsymbol{\nu}$ in the following way*

$$\nu_j = \mathbb{E}_i \left(\sum_{n=0}^{T_i - 1} I\{X_n = j\} \right).$$

ν_j is the expected number of visits to state j between two consecutive visits to state i.

Proof.

$$\begin{aligned}
\nu_j &= \mathbb{E}_i \left(\sum_{n=0}^{T_i - 1} I\{X_n = j\} \right) \\
&= \mathbb{E}_i \left(\sum_{n=1}^{T_i} I\{X_n = j\} \right) \quad (X_0 = X_{T_i} = i) \\
&= \mathbb{E}_i \left(\sum_{n=1}^{\infty} I\{X_n = j, T_i \geq n\} \right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}_i \left(I\{X_n = j, T_i \geq n\} \right) \\
&= \sum_{n=1}^{\infty} \mathbb{P} \left(X_n = j, T_i > n - 1 \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} \mathbb{E}_i \left(\mathbb{P} \left(X_n = j, T_i > n-1 | \mathscr{F}_{n-1} \right) \right) \\
&= \sum_{n=1}^{\infty} \mathbb{E}_i \left(I\{T_i > n-1\} \mathbb{P} \left(X_n = j | \mathscr{F}_{n-1} \right) \right) \quad \text{(measurable)} \\
&= \sum_{n=1}^{\infty} \mathbb{E}_i \left(I\{T_i > n-1\} p_{X_{n-1} j} \right) \quad \text{(Markov)}
\end{aligned}
$$

Now

$$
\begin{aligned}
\mathbb{E}_i \left(I\{T_i > n-1\} p_{X_{n-1} j} \right) &= \mathbb{E}_i \left(\sum_{k \in E} I\{X_{n-1} = k\} I\{T_i > n-1\} p_{X_{n-1} j} \right) \\
&= \sum_{k \in E} p_{kj} \mathbb{P}_i (X_{n-1} = k, T_i > n-1).
\end{aligned}
$$

Inserting the above expression

$$
\begin{aligned}
\nu_j &= \sum_{n=1}^{\infty} \sum_{k \in E} p_{kj} \mathbb{P}_i (X_{n-1} = k, T_i > n-1) \\
&= \sum_{k \in E} p_{kj} \sum_{n=1}^{\infty} \mathbb{P}_i (X_{n-1} = k, T_i > n-1) \\
&= \sum_{k \in E} p_{kj} \nu_k,
\end{aligned}
$$

by definition of ν_j. Then $\boldsymbol{\nu} = \boldsymbol{P}\boldsymbol{\nu}$. If j is not in the same recurrence class as i, then $\nu_j = 0 < \infty$. If j is contained in the same recurrence class as i then $i \sim j$ and there exists an m such that $p_{ji}^{(m)} > 0$. Thus, as $\boldsymbol{\nu} = \boldsymbol{\nu}\boldsymbol{P} = \ldots = \boldsymbol{\nu}\boldsymbol{P}^m$,

$$
\infty > 1 = \nu_i = \sum_{k \in E} \nu_k p_{ki}^{(m)} \geq \nu_j p_{ji}^{(m)},
$$

from which we conclude that $\nu_j < \infty$. It is also clear that $\boldsymbol{\nu} \neq \mathbf{0}$ and $\nu_k \geq 0$ for all $k \in E$. □

Corollary 2.2. *If a Markov chain is irreducible and recurrent then for any stationary measure $\boldsymbol{\nu}$ we have that $\nu_i > 0$ for all $i \in E$.*

Proof. This follows from $\boldsymbol{\nu} = \boldsymbol{\nu}\boldsymbol{P}^m$ where m is chosen such that $p_{ji}^{(m)} > 0$, and since

$$
\nu_i = \sum_{k \in E} \nu_k p_{ki}^{(m)} \geq \nu_j p_{ji}^{(m)}.
$$

Since a stationary measure satisfies that $\boldsymbol{\nu} \neq 0$ we may choose a state j such that $\nu_j > 0$. By the inequality above we then conclude that all other $\nu_i > 0$ as well. □

We define $\boldsymbol{\nu}^{(i)}$ as the stationary measure given by

$$\nu_j^{(i)} = \mathbb{E}_i \left[\sum_{n=0}^{T_i - 1} I\{X_n = j\} \right]. \tag{2.2}$$

The superscript (i) indicates the dependence on the choice of recurrent state i. If we consider $j = i$, then only one term in the sum is non–zero so we conclude that $\nu_i^{(i)} = 1$. We call the measure $\boldsymbol{\nu}^{(i)}$ the canonical stationary measure.

Corollary 2.3.

$$\nu_j^{(i)} = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n).$$

Proof. Follows immediately from the Beppo–Levi theorem (which states that we may interchange integration and summation of non–negative measurable functions). □

Lemma 2.1. *Let $\boldsymbol{\nu}$ be a stationary measure with $\nu_i = 1$. Then $\boldsymbol{\nu} = \boldsymbol{\nu}^{(i)}$.*

Proof. Recall that

$$\nu_j^{(i)} = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n).$$

Now, $\mathbb{P}_k(X_n = j, T_i > n)$ is the probability of going from k a j without visiting state i in between. This is a so–called taboo probability. Let us assume that $j \neq k$. For $n = 1$ the taboo probability is the usual transition probability. For $n = 2$,

$$\mathbb{P}_k(X_n = j, T_i > n) = \sum_{k \neq i} p_{ik} p_{kj}.$$

Define $\tilde{\boldsymbol{P}}$ as the transition matrix $\boldsymbol{P}$ but with the i'th column replaced by zeros. Then

$$\mathbb{P}_k(X_n = j, T_i > n) = \sum_{\ell \in E} \tilde{p}_{k\ell} \tilde{p}_{\ell j}.$$

By induction it is clear that

$$(\mathbb{P}_k(X_n = j, T_i > n))_{k,j \in E} = \tilde{\boldsymbol{P}}^n.$$

Thus $\mathbb{P}_i(X_n = j, T_i > n)$ is the ij 'th element of $\tilde{\boldsymbol{P}}^n$, i.e. $\mathbb{P}_i(X_n = j, T_i > n) = \boldsymbol{e}_i' \tilde{\boldsymbol{P}}^n \boldsymbol{e}_j$, where $\boldsymbol{e}_i'$ is the i'th Euclidian row vector. From Corollary 2.3 follows

$$\boldsymbol{\nu}^{(i)} = (\nu_j^{(i)})_{j \in E} = \boldsymbol{e}_i' \sum_{n=0}^{\infty} \tilde{\boldsymbol{P}}^n.$$

As $\boldsymbol{\nu}$ is stationary with $\nu_i = 1$ we have that

$$\nu_j = \delta_{ij} + \left(\boldsymbol{\nu} \tilde{\boldsymbol{P}} \right)_j,$$

because if $i = j$ then $\left(\boldsymbol{\nu}\tilde{\boldsymbol{P}}\right)_j = 0$ by construction of $\tilde{\boldsymbol{P}}$ and $\delta_{ij} = 0$ otherwise, and $\left(\boldsymbol{\nu}\tilde{\boldsymbol{P}}\right)_j = \nu_j$ being stationary. Thus

$$\begin{aligned}\boldsymbol{\nu} &= \boldsymbol{e}_i' + \boldsymbol{\nu}\tilde{\boldsymbol{P}} \\ &= \boldsymbol{e}_i' + \left(\boldsymbol{e}_i' + \boldsymbol{\nu}\tilde{\boldsymbol{P}}\right)\tilde{\boldsymbol{P}} \\ &= \boldsymbol{e}_i'(\boldsymbol{I} + \tilde{\boldsymbol{P}}) + \boldsymbol{\nu}\tilde{\boldsymbol{P}}^2 \\ &= \ldots \\ &= \boldsymbol{e}_i' \sum_{n=0}^{N} \tilde{\boldsymbol{P}}^n + \boldsymbol{\nu}\tilde{\boldsymbol{P}}^{N+1},\end{aligned}$$

where $\boldsymbol{I}$ is the identity matrix . Since $\tilde{\boldsymbol{P}}^n \to 0$ (its elements are being dominated by a tail probability) as $n \to \infty$ and hence $N \to \infty$,

$$\boldsymbol{\nu} = \boldsymbol{e}_i' \sum_{n=0}^{N} \tilde{\boldsymbol{P}}^n + \boldsymbol{\nu}\tilde{\boldsymbol{P}}^{N+1} \to \boldsymbol{e}_i' \sum_{n=0}^{\infty} \tilde{\boldsymbol{P}}^n = \boldsymbol{\nu}^{(i)}.$$

□

Corollary 2.4. *If a Markov chain is irreducible and recurrent, there exists a stationary measure. All stationary measures are proportional.*

Proof. If $\boldsymbol{\nu}$ is stationary with $\nu_i = c$, then $\boldsymbol{\mu} = \boldsymbol{\nu}/c$ is stationary with $\mu_i = \nu_i^{(i)}$ and hence $\boldsymbol{\mu} = \boldsymbol{\nu}^{(i)}$, i.e. $\boldsymbol{\nu} = c\boldsymbol{\nu}^{(i)}$. □

Definition 2.11. A recurrent state i positively postively recurrent if $\mathbb{E}_i(T_i) < \infty$ and null recurrent if $\mathbb{E}_i(T_i) = \infty$.

Corollary 2.5. *If a Markov chain is irreducible and recurrent, then either all states are positively postively recurrent or null recurrent. That is, positive recurrence and null recurrence are class properties.*

Proof. All stationary measures are proportional to $\boldsymbol{\nu}^{(j)}$, $j \in E$. Also, $|\boldsymbol{\nu}^{(j)}| = \sum_k \nu_k^{(j)}$ is either finite or infinite. Since $\sum_k \nu_k^{(j)} = \mathbb{E}_j T_j$ this implies that either all states are either positively postively recurrent or null recurrent. □

Hence for irreducible and recurrent Markov chains a stationary measure always exists and it is unique up to multiplication by a constant. The question concerning the existence of stationary distributions is hence equivalent to the question when it is possible to normalize a stationary measure.

Corollary 2.6. *If a Markov chain is irreducible and positively recurrent, then there exists a unique stationary distributions $\boldsymbol{\pi}$ given by*

$$\pi_j = \frac{1}{\mathbb{E}_i T_i} \mathbb{E}_i \sum_{n=0}^{T_i - 1} I\{X_n = j\} = \frac{1}{\mathbb{E}_j T_j} > 0.$$

Proof. If the chain is postively recurrent then,

$$|\nu^{(i)}| = \sum_j \nu_j^{(i)} = \sum_j \mathbb{E}_i \sum_{n=0}^{T_i-1} I\{X_n = j\} = \mathbb{E}_i T_i < \infty$$

and hence

$$\boldsymbol{\pi} = \boldsymbol{\nu}^{(i)}/|\boldsymbol{\nu}^{(i)}|$$

by uniqueness of multiplication up to a constant. □

Corollary 2.7. *An irreducible Markov chain with a finite state–space is positively recurrent.*

Proof. It is clear that

$$\sum_{k\in E}\sum_{n=0}^{\infty} I\{X_n = k\} = \infty$$

and since $|E| < \infty$ there exists a $k \in E$ such that

$$\sum_{n=0}^{\infty} I\{X_n = k\} = \infty.$$

Thus k is recurrent and hence all states are recurrent. A stationary measure then exists. The total mass of

$$|\boldsymbol{\nu}^{(i)}| = \mathbb{E}_i T_i < \infty$$

by $|E| < \infty$ and hence postively recurrent. □

Definition 2.12. The period of a state i is the largest integer $d(i)$ such that

$$\mathbb{P}_i(T_i \in L_{d(i)}) = 1$$

where $L_d = \{d, 2d, 3d, 4d, \ldots\}$. If the period is one the state is called aperiodic.

Theorem 2.11. *Periodicity is a class property: if i and j are in the same recurrence class, then they have the same period.*

Proof. Let i be a recurrent state with period $d(i)$. Let j be another state in the same recurrence class. Then $i \sim j$ and consequently there exists $m, n > 0$ such that $p_{ij}^{(n)} > 0$ y $p_{ji}^{(m)} > 0$. Thus

$$p_{ii}^{(n+m)} = \sum_{k\in E} p_{ik}^{(n)} p_{ki}^{(m)} \geq p_{ij}^{(n)} p_{ji}^{(m)} > 0$$

so we see that $n+m \in L_{d(i)}$. Now take p: $p_{jj}^{(p)} > 0$. Then

$$p_{ii}^{(m+m+p)} \geq p_{ij}^{(n)} p_{jj}^{(p)} p_{ji}^{(m)} > 0$$

so we also have that $n+m+p \in L_{d(i)}$. Hence $p \in L_{d(i)}$ and $d(j) \geq d(i)$ By symmetry we obtain that $d(j) \leq d(i)$. □

Theorem 2.12. *For an aperiodic Markov chain with transition probabilities p_{ij} there exists an N such that the n–step transition probabilities $p_{ii}^{(n)} > 0$ for all $n \geq N$.*

Proof. Let $C = \{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$. Then there exists and n such that $n, n+1 \in C$ since otherwise the period of the chain would be larger or equal to 2 not aperiodic. Now take any $m > n$. Divide m by n writing $m = an + b$, where $0 \leq b \leq n-1$ and a, b are integers. But then

$$m = an + b = an + b(n+1-n) = (a-b)n + b(n+1).$$

Now let m be so large that $a - b > 0$. Let N be the smallest such number. Then $p_{ii}^{(m)} > 0$ for all such $m \geq N$ since m is a multiple of n and $n+1$ both of which have positive step transition probabilities. □

It follows immediately that

Corollary 2.8. *If $\{X_n\}_{n\in\mathbb{N}}$ is irreducible and aperiodic, then for all $i, j \in E$ there exists an N such that if $n \geq N$ then $p_{ij}^{(n)} > 0$.*

2.4 Convergence of transition probabilities

Next we study the asymptotic properties of transition probabilties and the Markov chains themselves. First we restrict out attention to so–called *ergodic* chains.

Definition 2.13. A Markov chain is called ergodic if is irreducible, aperiodic and recurrent.

Theorem 2.13. *Consider an irreducible, positive recurrent and aperiodic Markov chain $\{X_n\}_{n\in\mathbb{N}}$ with state–space E, n–step transition probabilities $p_{ij}^{(n)}$ and stationary distribution $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$. Then*

$$p_{ij}^{(n)} \to \pi_j > 0 \ \ as \ \ n \to \infty.$$

We shall prove this theorem by using a technique referred to as the coupling method. The basic idea is to start two independent Markov chains, one from a fixed state i and another according to its stationary distribution. If we can the prove that the time until the two chains coincide for the first time is finite with probability one, then by the strong Markov property both chains will probabilistically have the same behaviour after this event and we conclude that the chain which was initiated in a fixed point i is now in a stationary mode. This in turn implies that the probability of the fixed initated chain being in some state j must converge to the stationary distribution.

Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be two Markov chains defined on the same probability space. Let

$$T = \inf\{n \in \mathbb{N} | X_n = Y_n\}$$

be the time where the two chains coincide for the first time. Then define a third process $\{Z_n\}_{n\in\mathbb{N}}$ by

$$Z_n = \begin{cases} X_n \text{ if } n \leq T \\ Y_n \text{ if } n > T \end{cases}$$

Then Z_n evolves like the X_n process until the latter meets with the Y_n process and from then onwards it will coincide with the latter. The X_n and Z_n process "couples" at time T. Coupling can be defined more broadly and extended to numerous situation but for the moment the above intuitive formulation is sufficient.

Lemma 2.2. *(Coupling inequality)*

$$|\mathbb{P}(X_n = i) - \mathbb{P}(Z_n = i)| \leq \mathbb{P}(T > n).$$

Proof.

$$\begin{aligned} \mathbb{P}(X_n = i) &= \mathbb{P}(X_n = i, T \leq n) + \mathbb{P}(X_n = i, T > n) \\ &= \mathbb{P}(Z_n = i, T \leq n) + \mathbb{P}(X_n = i, T > n) \\ &\leq \mathbb{P}(Z_n = i) + \mathbb{P}(T > n). \end{aligned}$$

By symmetry, we then conclude the desired result. □

We now turn to the proof of Theorem 2.13. Let $\{X_n\}_{n\in N}$ be as in the theorem and let $\{Z_n\}_{n\in\mathbb{N}}$ be an independent stationary version of the Markov chain, i.e. it is initiated according to the stationary distribution $\boldsymbol{\pi}$. The bivariate process defined by $W_n = (X_n, Y_n)$ is a Markov chain on the state–space $E \times E$ and with transition probabilities $r_{(i_1,i_2),(j_1,j_2)} = p_{i_1 j_1} p_{i_2 j_2}$. The n–step transition probabilities are likewise given by

$$r^{(n)}_{(i_1,i_2),(j_1,j_2)} = p^{(n)}_{i_1 j_1} p^{(n)}_{i_2 j_2}.$$

Since the chains are aperiodic, there is an N such both $p^{(n)}_{i_1 j_1} > 0$ and $p^{(n)}_{i_2 j_2} > 0$ for all $n \geq N$ (see Corollary 2.8). Thus for $n \geq N$, $r^{(n)}_{(i_1,i_2),(j_1,j_2)} > 0$ and hence the chains $\{W_n\}_{n\in\mathbb{N}}$ must be irreducible. It is also clear, that $\nu_{ij} = \pi_i \pi_j$ is a stationary distribution for $\{W_n\}_{n\in\mathbb{N}}$ (arranged as a row vector). In vector notation, $\boldsymbol{\nu} = \boldsymbol{\pi} \otimes \boldsymbol{\pi}$, where $\otimes$ is the Kronecker product between two vectors (or matrices). Now we use the positive recurrence to conclude that $\pi_i > 0$ for all i and hence that $\nu_{ij} = \pi_i \pi_j > 0$ for all i, j. Thus $\{W_n\}_{n\in\mathbb{N}}$ is positively recurrent as well. Let $i \in E$. Then $T \leq T_{ii} = \inf\{n \in \mathbb{N} : W_n = (i,i)\}$. Since $\{W_n\}_{n\in\mathbb{N}}$ is positively recurrent, then T_{ii} is finite a.s. and then so is T. Then the process defined by

$$Z_n = \begin{cases} X_n \text{ if } n \leq T \\ Y_n \text{ if } n > T \end{cases}$$

is well defined, and T being a stopping time implies by appealing to the strong Markov property that the chains $\{X_n\}_{n\in\mathbb{N}}$ and $\{Z_n\}_{n\in\mathbb{N}}$ have the same joint distri-

butions. Thus by the coupling inequality,

$$\begin{aligned}|\mathbb{P}(Z_n = i) - \pi_i| &= |\mathbb{P}(Z_n = i) - \mathbb{P}(Y_n = i)| \\ &\leq \mathbb{P}(T > n)\end{aligned}$$

which is valid for all i. In particular,

$$\sup_i |\mathbb{P}(Z_n = i) - \pi_i| \to 0$$

as $n \to \infty$. Replacing $\mathbb{P}$ with $\mathbb{P}_k$ then implies that

$$\sup_i |p_{ki}^{(n)} - \pi_i| \to 0$$

as $n \to \infty$. This completes the proof of Theorem 2.13.

2.5 Markov chain Monte Carlo algorithms

Consider a (possibly infinite) distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ...) = (\pi_i)_{i \in E}$, where E is the index set. We shall now construct a Markov chain which has $\boldsymbol{\pi}$ as stationary distribution. Let $\boldsymbol{Q}$ be any transition matrix defined on E. Then define numbers

$$\alpha_{ij} = \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right). \tag{2.3}$$

Then define a Markov chain $\{Y_n\}_{n \geq 0}$ with state space E and transition probabilities

$$p_{ij} = q_{ij}\alpha_{ij},\ i \neq j, \quad p_{ii} = 1 - \sum_{j \neq i} q_{ij}\alpha_{ij}.$$

Theorem 2.14. *(Metropolis–Hastings)* $\{Y_n\}_{n \geq 0}$ *has stationary distribution* $\boldsymbol{\pi}$.

Proof.

$$\begin{aligned}\sum_i \pi_i p_{ij} &= \sum_{i \neq j} \pi_i q_{ij}\alpha_{ij} + \pi_j \left(1 - \sum_{k \neq j} q_{jk}\alpha_{jk}\right) \\ &= \pi_j + \sum_{k \neq j} \left(\pi_k q_{kj}\alpha_{kj} - \pi_j q_{jk}\alpha_{jk}\right),\end{aligned}$$

and it is easy to see that all $\pi_k q_{kj}\alpha_{kj} - \pi_j q_{jk}\alpha_{jk} = 0$. □

The Markov chain $\{Y_n\}_{n \geq 0}$ can be simulated in the following way. If $Y_0 = i$, then draw a j according to $(q_{ij})_{j \in E}$, which is accepted with probability α_{ij} in which case $Y_1 = j$ and rejected with probability $1 - \alpha_{ij}$ in which case $Y_1 = i$ (remain in i). The transition matrix $\boldsymbol{Q}$ may be anything and is often referred to as the proposal distribution, while the α_{ij} are known as the acceptance ratios. The above simulation

scheme is also known as Metropolis–Hastings algorithm. If the Markov chain settles into stationary mode, then any draw from the Metropolis–Hastings algorithm yields a realization of $\boldsymbol{\pi}$.

The Gibbs sampler is a special case of the Metropolis–Hastings algorithm, but require that we specify a two–dimensional distribution $\pi(\cdot,\cdot)$ with marginals $\boldsymbol{\pi}$. In this case $\alpha_{ij} = \min(1, \pi_j \pi(j \mid i)/(\pi_i \pi(i \mid j))) = 1$.

Much of the Markov chain Monte Carlo methods applied in practice rely on the fact that we may construct a Markov chain which has a stationary distribution that is our distribution of interest. Whether we apply the Gibbs sampler or a more general Metropolis–Hastings algorithm, in the end our distribution of interest is obtained through a stationary sample of the chain. This is ususally obtained by letting the chain run a certain anmount of time (burn–in period) until it enters a stationary mode, and then using a certain amount of data points (draws) from the stationary distribution (our distribution of interest) to perform a statistical inference.

2.6 Spectral properties of transition and sub–transition matrices

2.7 Markov jump processes

Markov processes in continuous time are various. In the following we shall focus on the ones which have a discrete (finite or infinitely countable) state–space. By nature, such processes are piecewise constant and transitions occur via jumps. They are often referred to as Markov jump process or continuous time Markov chains.

Definition 2.14. A Markov jump process $\{X_t\}_{t\geq 0}$ ($t \in \mathbb{R}$) with values in the discrete state–space E is a stochastic process with the following property:

$$\mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, ..., X_{t_1} = i_1, X(0) = i_0) = \mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}).$$

The process is called time–homogenous if also $\mathbb{P}(X_{t+h} = j | X_t) = P_{ij}(h)$ only depends on h. We call $p_{ij}(h)$ for transition probabilities and define the corresponding transition matrix $\boldsymbol{P}(h) = \{p_{ij}(h)\}_{i,j\in E}$.

Theorem 2.15. *Chapman-Kolmogorov:*

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

Proof. By the law of total probability,

$$\begin{aligned} p_{ij}(t+s) &= \mathbb{P}(X_{t+s} = j | X_0 = i) \\ &= \sum_k \mathbb{P}(X_{t+s} = j, X_s = k | X_0 = i) \\ &= \sum_k \mathbb{P}(X_{t+s} = j | X_s = k, X_0 = i) \mathbb{P}(X_s = k | X_0 = i) \end{aligned}$$

$$= \sum_k \mathbb{P}(X_{t+s} = j | X_s = k)\mathbb{P}(X_s = k | X_0 = i)$$
$$= \sum_k p_{ik}(s)p_{kj}(t).$$

□

Let $\mathscr{F}_t = \sigma(X_s : s \leq t)$.

Definition 2.15. A non–negative random variable τ is called a stopping time for $\{X_t\}_{t\geq 0}$ if $\{\tau \leq t\} \in \mathscr{F}_t$ for all t.The σ–algebra $\mathscr{F}_\tau$ consists of the measurable sets A such that

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

Theorem 2.16. *Any Markov jump process $\{X_t\}_{t\geq 0}$ satisfies the strong Markov property*

$$\mathbb{P}(X_{t+h_1} = i_1, ..., X_{t+h_n} = i_n | \mathscr{F}_\tau) = \mathbb{P}(X_{h_1} = i_1, ..., X_{h_n} = i_n | X_0 = X_\tau).$$

Proof. pending[1]. □

Often τ is defined in terms of the values of X_t such that X_τ is deterministic. For example τ could be the dirst time $\{X_t\}_{t\geq 0}$ is in state i. In this case $X_\tau = i$ and consequently $\{X_0 = X_\tau\} = \{X_0 = i\}$.

Let $S_0 = 0 < S_1 < S_2 < ...$ denote the times of the jumps. The differences $T_n = S_{n+1} - S_n$ are the times between jumps. Hence T_0 is the time until the first jump, T_1 the time between the first and the second jump etc. Finally, let $Y_0, Y_1, ...$ denote the sequence of states visited, i.e. $Y_i = X(S_i)$. In case there is a last S_n (for example if there is an absorbing state) we then define $T_n = T_{n+1} = ... = \infty$ and $Y_n = Y_{n+1} = ... = i$ if $X(S_n) = i$.

It is clear that we can reconstruct the sample paths $t \to X_t$ from the knowledge of $\{(Y_n, T_n)\}_{n\geq 0}$. Hence the joint distribution of (Y_n, T_n) if of principal interest. Let

$$p_n = \mathbb{P}_i(Y_k = i_k, T_{k-1} > t_k, k = 1, ..., n).$$

Theorem 2.17. *There exists numbers $\lambda_i \geq 0$ and a transition matrix $\boldsymbol{Q}$ such that*

$$p_n = \prod_{k=1}^{n} q_{i_{k-1}i_k} \exp(-\lambda(i_{k-1})t_k).$$

Proof. Define $f(t) = \mathbb{P}_i(T_0 > t)$. Then

$$\begin{aligned} f(t+s) &= \mathbb{P}_i(T_0 > t+s) \\ &= \mathbb{E}_i(\mathbb{P}_i(T_0 > t+s | \mathscr{F}_t)) \\ &= \mathbb{E}_i(\mathbb{P}_i(T_0 > t+s, T_0 > t | \mathscr{F}_t)) \\ &= \mathbb{E}_i(I\{T_0 > t\}\mathbb{P}_i(T_0 > t+s | \mathscr{F}_t)) \ \text{(measurability)} \\ &= \mathbb{E}_i(I\{T_0 > t\}\mathbb{P}_i(T_0 > s)) \ \text{(Markov and } X_t = i) \end{aligned}$$

[1] pending

$$= f(t)f(s).$$

Since f non–increasing (it is a tail probability) we have that f is on the form

$$f(t) = \exp(-\lambda_i t)$$

for some $\lambda_i \geq 0$. Then

$$\begin{aligned}\mathbb{P}_i(Y_1 = j, T_0 > t) &= \mathbb{E}_i\left(\mathbb{P}_i(Y_1 = j, T_0 > t|\mathscr{F}_t)\right)\\ &= \mathbb{E}_i\left(I\{T_0 > t\}\mathbb{P}_i(Y_1 = j)\right)\\ &= q_{ij}\exp(-\lambda_i t),\end{aligned}$$

where $q_{ij} = \mathbb{P}_i(Y_1 = j)$.

The general formula follows by induction and the strong Markov property. Suppose that p_m is valid for $m = n-1$. Then

$$\begin{aligned}p_n &= \mathbb{P}_i(Y_k = i_k, T_{k-1} > t_k, k = 1,...,n)\\ &= \mathbb{P}_i(Y_n = i_n, T_n > t_n, Y_k = i_k, T_{k-1} > t_k, k = 1,...,n-1)\\ &= \mathbb{E}_i\left(\mathbb{P}_i(Y_n = i_n, T_n > t_n, Y_k = i_k, T_{k-1} > t_k, k = 1,...,n-1|\mathscr{F}_{S_{n-1}})\right)\\ &= \mathbb{E}_i\left(I\{Y_k = i_k, T_{k-1} > t_k, k = 1,...,n-1\}\mathbb{P}_i(Y_n = i_n, T_{n-1} > t_n|\mathscr{F}_{S_{n-1}})\right)\\ &= \mathbb{P}_{i_{n-1}}(Y_1 = i_n, T_0 > t_n)p_{n-1} \text{ (measurable and strong Markov)}\\ &= q_{i_{n-1}i_n}\exp(-\lambda_{i_{n-1}}t_n)p_{n-1},\end{aligned}$$

and the result follows by induction. □

Corollary 2.9. *The following holds true:*

(i) $\{Y_n\}$ is a Markov chain (called the embedded chain).
(ii) There are numbers $\lambda_i \geq$ such that $\{T_\ell\}$ are conditionally independent with $T_\ell \sim \exp(\lambda_{Y_\ell})$.

Proof. (i) above. Concerning (ii) we note that

$$\begin{aligned}&\mathbb{P}_i(T_0 > t_0,...,T_{n-1} > t_{n-1}|Y_1 = i_1,...,Y_n = i_n)\\ &\quad= \frac{\mathbb{P}_i(T_0 > t_0,...,T_{n-1} > t_{n-1}, Y_1 = i_1,...,Y_n = i_n)}{\mathbb{P}_i(Y_1 = i_1,...,Y_n = i_n)}\\ &\quad= \frac{\prod_{k=1}^n q_{i_{k-1}i_k}\exp(-\lambda_{i_{k-1}t_k})}{\prod_{k=1}^n q_{i_{k-1}i_k}} \quad (i_0 = i)\\ &\quad= \prod_{k=1}^n \exp(-\lambda_{i_{k-1}}t_k),\end{aligned}$$

which implies conditional independence and that the the distributions are exponetials. □

We have proved that if $\{X_t\}$ is a Markov jump process then there exists numbers $\lambda_i \geq 0$ and probabilities q_{ij} with the above properties. Then opposite result is also valid and is essentially an exercise in the use of Kolmogorov's existency theorem.

In the following we suppose that $\lambda_i > 0$ for all i and that the Markov chain $\{Y_n\}$ has transition matrix $\boldsymbol{Q} = \{q_{ij}\}$. In particular, $q_{ii} = 0$ by definition of $\{Y_n\}$.

Definition 2.16. The intensity matrix (or infinitesimal generator) of $\{X_t\}_{t\geq 0}$ $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j\in E}$ is defined by

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

We notice that the row elements of $\boldsymbol{\Lambda}$ sums to zero. Since the holding times in each state i are exponentially distributed $\exp(\lambda_i)$, the probability of a jump out of state i during the time interval $[t, t+dt)$ is $\lambda_i dt$ (or formally $\lambda_i h + o(h)$,where $o(h)$ is a function such that $o(h)/h \to 0$ when $h \to 0$). Conditionally on the event that there is a jump in $[t, t+dt)$ and $X_{t-} = i$ (left limit), the probability that the next state to which the process jumps is j is q_{ij}. Hence the probability of jumping from state i to state j during $[t, t+dt)$ is $\lambda_i dt q_{ij} = \lambda_{ij} dt$.

Theorem 2.18. *(Kolmogorov's differential equations)*
If $\boldsymbol{P}^t = \{p_{ij}(t)\}$ denotes the transition probabilities of a Markov jump process,

$$p_{ij}(t) = \mathbb{P}_i(X_t = j) = \mathbb{P}(X_{s+t} = j | X_s = i),$$

then

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

Proof. Conditioning on the time of the first jump T_0, we get that

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

The integrant is bounded on $[0,T]$ and hence $t \to p_{ij}^t$ is differentiable. Thus

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

$$= \sum_k \lambda_{ik} p^t_{kj}$$

which is the same as

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

The other equation is more difficult to prove and will be proved subject to the additional restriction that the intensities are bounded, i.e. $\sup_i \lambda_i < \infty$ though the result holds in general. We recall that

$$p^h_{ij} = \lambda_{ij} h + o(h)$$

as $h \to 0$, and when $i = j$,

$$p^h_{ii} = 1 - \lambda_i h + o(h).$$

Hence

$$\frac{p_{ij} - \delta_{ij}}{h} \to \lambda_{ij} \text{ cuando } \quad h \to 0.$$

We shall need the expression to be uniformlly bounded in i, j and h. First,

$$\begin{aligned} 0 \le p^s_{ij} &\le \text{probability of an arrival in } [0,s] \\ &= \int_0^s \lambda_i e^{-\lambda_i u} du \\ &= \lambda_i \int_0^s e^{-\lambda_i u} du \\ &\le \lambda_i \int_0^s 1 du \\ &= \lambda_i s \end{aligned}$$

which implies

$$0 \le \frac{p^s_{ij}}{s} \le \lambda_i \le \sup_j \lambda_j.$$

Similarly,

$$\begin{aligned} 0 \le 1 - p^s_{ii} &\le \text{the probability of a jump in } [0,s] \\ &\le \lambda_i s \end{aligned}$$

implying the same as above. Hence by dominated convergence,

$$\begin{aligned} \frac{p^{s+h}_{ij} - p^s_{ij}}{h} &= \frac{1}{h}\left(\sum_k p^s_{ik} p^h_{kj} - p^s_{ij}\right) \\ &= \sum_k p^s_{ik} \frac{p^h_{kj} - \delta_{kj}}{h} \\ &\to \sum_k p^s_{ik} \lambda_{kj} \end{aligned}$$

which implies that

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

□

Corollary 2.10. *If E is finite, then*

$$\boldsymbol{P}^t = \exp(\boldsymbol{\Lambda}t) = \sum_{n=0}^{\infty} \frac{\boldsymbol{\Lambda}^n t^n}{n!}.$$

Proof. Follows from standard theory of linear systems of differential equations and from $\boldsymbol{P}^0 = \boldsymbol{I}$. □

The classification of states in a Markov jump process follows from

Theorem 2.19. *The following statements are equaivalent:*

(i) $\{Y_n\}_{n\in E}$ *is irreducible.*
(ii) $\forall i,j \in \exists\, t > 0 : p_{ij}^t > 0.$
(iii) $\forall i,j \in \forall t > 0 : p_{ij}^t > 0.$

Proof. (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i) is obvious. (i) $\Longrightarrow$ (iii): Take $i,j \in E$. Then i,j communicates (with respect to $\{Y_n\}$) and hence there exists a path from i to j, say, $i,i_1,i_2,...,i_n,j$ such that $q_{ii_1} > 0$, $q_{i_1i_2} > 0$,...,$q_{i_nj} > 0$. Now the probability

$$p_{ij}^t \geq \text{prob. of going } \; i \to i_1 \to ... \to i_n \to j \text{ in time } t.$$

The right hand expression is a convolution of expoential distributions (hence Gamma distributed) which has a strictly positive density at all points. This implies that $p_{ij}^t > 0$ for all $t > 0$. □

Definition 2.17. $\{X_t\}_{t\geq 0}$ is irreducible if there exists a $t > 0$ such that $p_{ij}^t > 0$ for all $i,j \in E$.

Hence $\{X_t\}_{t\geq 0}$ isirreducible if and only if $\{Y_n\}$ is irreducible and if and only if $p_{ij}^t > 0$for all $t > 0$.

Recurrence and trasience is based on the corresponding properties of $\{Y_n\}$:

Definition 2.18. $\{X_t\}$ is recurrent (transient) if $\{Y_n\}$ is recurrent (transient).

Definition 2.19. A row vector $\boldsymbol{\nu} = \{\nu_i\}_{i\in E}$ is called a stationary measure for a Markov jump process with transition probability matrix $\boldsymbol{P}^t$ if $\boldsymbol{\nu}$ is finite, non–zero, non–negative and satisfying

$$\boldsymbol{\nu} = \boldsymbol{\nu}\boldsymbol{P}^t$$

for all $t \geq 0$.

Theorem 2.20. *Suppose that a Markov jump process $\{X_t\}_{t\geq 0}$ is irreducible and recurrent. Let k be an arbitrary state and define the vector $\boldsymbol{\nu} = \{\nu_i\}_{i\in E}$ by*

$$\nu_i = \mathbb{E}_k\left(\int_0^{U_k} 1_{\{X_s=i\}} ds\right),$$

where $U_k = \inf\{t > 0 : X_t = k, X_{t-} \neq k\}$ *is the time of first return to state* k *and* $X_{t-} = \lim_{s\uparrow t} X_s$ *the limit from the left. Then* $\boldsymbol{\nu}$ *is a unique, up to multiplication with a constant, stationary measure for the process* $\{X_t\}_{t\geq 0}$.

Proof.

$$\begin{aligned}
\nu_i &= \mathbb{E}_k\left(\int_0^{U_k} 1_{\{X_s=i\}} ds\right) \\
&= \mathbb{E}_k\left(\int_0^{\infty} 1_{\{X_s=i, U_k>s\}} ds\right) \\
&= \int_0^{\infty} \mathbb{P}_k(X_s = i, U_k > s) ds. \quad \text{(by Fubini)}
\end{aligned}$$

U_k is a stopping time and hence $\{U_k > s\}$ is $\sigma(X_u : 0 \leq t \leq s)$–measurable, i.e. the event can be determined whether it has taken place or not knowing the evolution of the process X_t up to time s. Hence X_{t+s} is conditional independent of $\{U_k > s\}$ given X_s and we conclude that

$$p_{ij}^t = \mathbb{P}(X_{s+t} = j | X_s = i, U_k > s)$$

and hence

$$\begin{aligned}
\nu_i p_{ij}^t &= \mathbb{P}(X_{s+t} = j | X_s = i, U_k > s) \int_0^{\infty} \mathbb{P}_k(X_s = i, U_k > s) ds \\
&= \int_0^{\infty} \mathbb{P}(X_{s+t} = j, X_s = i, U_k > s) ds.
\end{aligned}$$

Summing over i, on the left hand side yields the j'th entry of the vector $\boldsymbol{\nu}\boldsymbol{P}^t$ giving

$$\left(\boldsymbol{\nu}\boldsymbol{P}^t\right)_j = \int_0^{\infty} \mathbb{P}(X_{s+t} = j, U_k > s) ds.$$

Interchanging integration and expectation again yields

$$\begin{aligned}
\left(\boldsymbol{\nu}\boldsymbol{P}^t\right)_j &= \mathbb{E}_k\left(\int_0^{U_k} 1_{\{X_{t+s}=j\}} ds\right) \\
&= \mathbb{E}_k\left(\int_t^{t+U_k} 1_{\{X_s=j\}} ds\right) \\
&= \mathbb{E}_k\left(\int_t^{U_k} 1_{\{X_s=j\}} ds\right) + \mathbb{E}_k\left(\int_{U_k}^{U_k+t} 1_{\{X_s=j\}} ds\right) \\
&= \mathbb{E}_k\left(\int_t^{U_k} 1_{\{X_s=j\}} ds\right) + \mathbb{E}_k\left(\int_0^{t} 1_{\{X_s=j\}} ds\right) \\
&= \mathbb{E}_k\left(\int_0^{U_k} 1_{\{X_s=j\}} ds\right)
\end{aligned}$$

$$= \nu_j.$$

Hence $\boldsymbol{\nu} = \boldsymbol{\nu P}$. Uniqueness follows from the discrete case. Indeed, $X_0, X_1, ..$ is a Markov chain and since it is irreducible, then $p_{ij}^t > 0$ for all t and hence in particular for $t = 1$. Concerning recurrence, we have assumed that $\{X_t\}_{t\geq 0}$ is recurrent, so i will be visited infinitely often (i.o.) by X_t. We do not know, however, whether these point of visit will coincide with any of the time points $0, 1, 2, ...$ from which the corresponding chain is conformed. Let $T_1^i, T_2^i, ...$ be the time spent in state i every time the process X_t enters that state. They are exponentially distributed and hence $\mathbb{P}(T_n^i > 1) > 0$. Thus there will an infinite number of points where $X_t = i$ for some t and the corresponding holding time will be larger than 1. This will ensure that the continuos process is in state i also at integer times for these events. Hence the chain will also visit state i i.o. and hence the chain is recurrent. Then by Corollary 2.4 we conclude that $\boldsymbol{\nu}$, which is also a stationary measure for the chain, must be unique up to multiplication with a constant. □

Next we consider the embedded chain, $\{Y_n\}_{n\in\mathbb{N}}$. It is clear, that if i is a recurrent state for $\{X_t\}_{t\geq 0}$ then so it is for the embedded chain $\{Y_n\}_{n\in\mathbb{N}}$, and according to Theorem 2.10 there exists a stationary measure $\boldsymbol{\mu}$ for $\{Y_n\}_{n\in\mathbb{N}}$ defined by

$$\mu_j = \mathbb{E}_i \left(\sum_{n=0}^{\tau_i - 1} I\{Y_n = j\} \right), \tag{2.4}$$

where τ_i is the number jumps of $\{X_t\}_{t\geq 0}$ until it returns to state i. The relation between the stationary measure of the Markov jump process and its embedded chain is given by the following theorem.

Theorem 2.21. *Let $\{X_t\}_{t\geq 0}$ is irreducible and recurrent with intensity matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j\in E}$. Let k be an arbitrary point in the state–space. Then the relation between the stationary measures $\boldsymbol{\nu}$ of Theorem 2.20 and (2.4) both defined in terms of the state k, is given by $\mu_j = \lambda_j \nu_j$, where $\lambda_j = -\lambda_{jj}$.*

Proof. The Markov jump process is piecewise constant with holding times in the differents states $T_0, T_1, ...$ as defined previously. Then

$$\begin{aligned}
\nu_j &= \mathbb{E}_k \left(\int_0^{U_k} 1_{\{X_t = j\}} dt \right) \\
&= \mathbb{E}_k \left(\sum_{n=0}^{\tau_k - 1} T_n 1_{\{Y_n = j\}} \right) \\
&= \sum_{n=0}^{\infty} \mathbb{E}_k \left(T_n 1_{\{Y_n = j, \tau_k > n\}} \right) \\
&= \sum_{n=0}^{\infty} \mathbb{E}_k \left[\mathbb{E}_k \left(T_n 1_{\{Y_n = j, \tau_k > n\}} \,\middle|\, \{Y_n\}_{n\in\mathbb{N}} \right) \right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}_k \left[1_{\{Y_n = j, \tau_k > n\}} \mathbb{E}_k \left(T_n \mid \{Y_n\}_{n\in\mathbb{N}} \right) \right] \quad \text{(measurability)}
\end{aligned}$$

$$= \sum_{n=0}^{\infty} \mathbb{E}_k \left[1_{\{Y_n=j, \tau_k>n\}} \mathbb{E}_k \left(T_n \,|\, Y_n = j \right) \right]$$
$$= \lambda_j^{-1} \mu_j.$$

□

Theorem 2.22. *A stationary measure $\boldsymbol{\nu}$ of a Markov jump process $\{X_t\}_{t\geq 0}$ with intensity matrix $\boldsymbol{\Lambda}$ satisfies and can by found as solution to*

$$\boldsymbol{\nu}\boldsymbol{\Lambda} = \mathbf{0}.$$

Proof. Differentiating $\boldsymbol{\nu}\boldsymbol{P}^t = \boldsymbol{\nu}$ implies that $\boldsymbol{\nu}\boldsymbol{\Lambda}\boldsymbol{P}^t = \mathbf{0}$ and the result follows letting $t \downarrow 0$. Alternatively, use the relation from Theorem 2.21 between the two measures $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Since $\boldsymbol{\mu}$ is stationary for the embedded chain and the embedded chain has transition probabilities $r_{ij} = \lambda_{ij}/\lambda_i = -\lambda_{ij}/\lambda_{ii}$ for $i \neq j$ and $r_{ii} = 0$ we have that

$$\sum_{i \in E} \mu_i r_{ij} = \mu_j$$

which then is equivalent to

$$\sum_{i \neq j} \lambda_i \nu_i \frac{\lambda_{ij}}{\lambda_i} = \lambda_j \nu_j = -\lambda_{jj} \nu_j$$

from which the result follows immediately. □

2.8 Explosion

A Markov jump process is called non–explosive if it has a finite number of trabsitions in finite intervals. This is usually the way we think about them but for Markov jump processes with an infininte state–space there may be situation where this property does not hold. For example in a pure birth process the jump intensities may increase (and equivalenty the expected holding times in each state decrease) so quickly that there may an infinite number of jumps within a finite time, a property similar to convergence of an infinite series. The formal definition is as follows.

Definition 2.20. Consider a Markov jump process $\{X_t\}_{t\geq 0}$ with times of transition (jumps) $S_0 = 0, S_1, S_2, \ldots$. Define $S_{max} = \sup_n S_n$. Then we say that the process is explosive if $\mathbb{P}(S_{max} < \infty) > 0$.

Hence a process is non–explosive if $\mathbb{P}(S_{\max} = +\infty) = 1$. A simple condition for non–explosiveness is that $\sup_i \lambda_i < \infty$, where λ_i are the jump rates in the different states. If the rates are bounded, then there will be a maximum rate and jumps will not be able to accumulate in finite. As a special case of bounded intensities is obtained when the state–space is finite. Also if the processes is recurrent, then it will return to

a state i with a certain rate $\lambda_i > 0$ infinitely often and hence only contrinutions from this state will make $S_{\max}$ infinite. In particular, a stationary Markov jump process cannot be explosive.

A less obvious criterium is the following.

Theorem 2.23. *(Reuter's criterium) A Markov jump process with intensity matrix $\boldsymbol{\Lambda}$ is non–explosive if and only if the only non–negative bounded solution to $\boldsymbol{\Lambda x} = \boldsymbol{x}$ is $\boldsymbol{x} = \boldsymbol{0}$.*

Proof. First we prove the "if" direction by proving its negation. Hence assume that the Markov jump process is explosive. Hence S_{max} is finite with positive probability, and for some $i \in E$

$$x_i = \mathbb{E}_i \left(e^{-S_{\max}} \right) > 0.$$

The explosion time is the sum of all times between jumps,

$$S_{\max} = T_0 + T_1 + T_2 + ...$$

which are exponentially distributed. Conditioning on $T_0 = y$ and the state j to which the process jumps, we get that with q_{ij} being the transition probabilities of the embedded chain, that

$$\begin{aligned} x_i &= \int_0^\infty \lambda_i e^{-\lambda_i y} \sum_{j \neq i} q_{ij} \mathbb{E}_j \left(e^{-(S_{\max} + y)} \right) dy \\ &= \int_0^\infty \sum_{j \neq i} \lambda_{ij} x_j e^{-y - \lambda_i y} dy \\ &= \sum_{j \neq i} \frac{x_j \lambda_{ij}}{1 + \lambda_i}, \end{aligned}$$

that is,

$$(1 + \lambda_i) x_i = (1 - \lambda_{ii}) x_i = \sum_{j \neq i} \lambda_{ij} x_j$$

which implies that $\boldsymbol{x} = \boldsymbol{\Lambda x}$. Hence $\boldsymbol{x}$ is a bounded (by one) non–negative, non–zero solution to $\boldsymbol{x} = \boldsymbol{\Lambda x}$.

To prove the "only if", assume non–explosiveness of the process. Hence $S_{\max} = +\infty$ with probability one and

$$T_0 + T_1 + ... + T_n \uparrow S_{max} = +\infty \quad a.s.$$

Consider $x_i^{(0)} = 1$ and

$$x_i^{(n+1)} = \mathbb{E}_i \left(e^{-T_0 - T_1 - .... - T_n} \right).$$

Condition on $T_0 = y$ and the first jump to $j \neq i$ again to obtain

$$x_i^{(n+1)} = \int_0^\infty \lambda_i e^{-\lambda_i y} \sum_{j \neq i} q_{ij} e^{-y} x_j^{(n)} dy$$

$$= \sum_{j \neq i} \frac{x_j^{(n)} \lambda_i}{1+\lambda_i}.$$

Now assume that there is an $\boldsymbol{x}$ which is bounded, non–negative and satisfies $\boldsymbol{\Lambda x} = \boldsymbol{x}$. We must then prove it to be the zero vector. First we notice, that since $\boldsymbol{x}$ is bounded we may assume that $x_i \leq 1 = x_i^{(0)}$ for all i. Secondly we notice that $\boldsymbol{\Lambda x} = \boldsymbol{x}$ is equivalent to

$$x_i = \sum_{j \neq i} \frac{\lambda_{ij} x_j}{1+\lambda_i}.$$

Now assume (induction hypothesis) that $x_i \leq x_i^{(n)}$ for all i. Then $x_j \lambda_{ij} \leq x_j^{(n)} \lambda_{ij}$ for all $j \neq i$, thus

$$\frac{x_j \lambda_{ij}}{1+\lambda_i} \leq \frac{x_j^{(n)} \lambda_{ij}}{1+\lambda_i}$$

and hence

$$x_i = \sum_{j \neq i} \frac{x_j \lambda_{ij}}{1+\lambda_i} \leq \sum_{j \neq i} \frac{x_j^{(n)} \lambda_{ij}}{1+\lambda_i} = x_i^{(n+1)}.$$

Hence it follows by induction that $x_i \leq x_i^{(n)}$ for all i and n. By monotone convergence, however, $x_i^{(n)} \to 0$ as $n \to \infty$ so $\boldsymbol{x} = \boldsymbol{0}$. □

2.9 Ergodicity for Markov jump processes

Definition 2.21. A Markov jump process is called ergodic if it is irreducible, recurrent and has a stationary distribution.

Thus all stationary measures of an ergodic Markov jump process have finite mass.

2.10 Time–reversibility

Time reversion plays an important role in the following chapters and we shall here provide an executive account of the basic construction and properties. First we consider the case of a time–homogeneous Markov chain $\{X_n\}_{n \geq 0}$ with some discrete state–space E and transition matrix $\boldsymbol{P} = \{p_{ij}\}_{i,j \in E}$. Fix a number $N \in \mathbb{N}$, and define the time–reversed Markov chain $\{\tilde{X}_n\}_{n \geq 0}$ by

$$\tilde{X}_i = X_{N-i}, \;\; i = 0, 1, ..., N.$$

Consider the transition probabilities of the reversed chain. If all $\mathbb{P}(X_n = i) > 0$, then

$$\begin{aligned}\mathbb{P}(\tilde{X}_{n+1}=j|\tilde{X}_n=i) &= \mathbb{P}(X_{N-n-1}=j|X_{N-n}=i)\\ &= \frac{\mathbb{P}(X_{N-n}=i|X_{N-n-1}=j)\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i)}\\ &= p_{ji}\frac{\mathbb{P}(X_{N-n-1}=j)}{\mathbb{P}(X_{N-n}=i)}. \end{aligned} \tag{2.5}$$

The latter expression does not depend on n if and only if the terms $\mathbb{P}(X_n=i)$ do not depend on n, i.e. if $\{X_n\}_{n\geq 0}$ is stationary. If $\{X_n\}_{n\geq 0}$ is stationary with stationary distribution $\boldsymbol{\pi}=(\pi_i)_{i\in E}$ and $\pi_i>0$ for all $i\in E$, then

$$\mathbb{P}(\tilde{X}_{n+1}=j|\tilde{X}_n=i)=\frac{p_{ji}\pi_j}{\pi_i}.$$

and

$$\begin{aligned}\mathbb{P}(\tilde{X}_0=i_0,\tilde{X}_1=i_1,...,\tilde{X}_n=i_n) &= \mathbb{P}(X_0=i_N,X_1=i_{N-1},...,X_N=i_0)\\ &= \pi_{i_N}p_{i_N,i_{N-1}}p_{i_{N-1},i_{N-2}}\cdots p_{i_1,i_0}\\ &= \pi_{i_N}\frac{\mathbb{P}(\tilde{X}_N=i_N|\tilde{X}_{N-1}=i_{N-1})\pi_{i_{N-1}}}{\pi_{i_N}}\cdots\frac{\mathbb{P}(\tilde{X}_1=i_1|\tilde{X}_0=i_0)\pi_{i_0}}{\pi_{i_1}}\\ &= \pi_{i_0}\prod_{k=1}^{N}\mathbb{P}(\tilde{X}_k=i_k|\tilde{X}_{k-1}=i_{k-1})\\ &= \mathbb{P}(\tilde{X}_0=i_0)\prod_{k=1}^{N}\mathbb{P}(\tilde{X}_k=i_k|\tilde{X}_{k-1}=i_{k-1})\end{aligned}$$

since $\pi_{i_0}=\mathbb{P}(X_0=i_0)=\mathbb{P}(X_N=i_0)=\mathbb{P}(\tilde{X}_0=i_0)$ by stationarity. Thus by Theorem 2.2 $\{\tilde{X}_n\}_{n\geq 0}$ is a time–homogeneous Markov chain with state–space E and transition matrix $\tilde{\boldsymbol{P}}=\{\tilde{p}_{ij}\}_{i,j\in E}$ given by the transition probabilities

$$\tilde{p}_{ij}=\frac{p_{ji}\pi_i}{\pi_j}.$$

Also, the these transition probabilities satisfy

$$\left(\boldsymbol{\pi}\tilde{\boldsymbol{P}}\right)_j=\sum_{i\in E}\pi_i\tilde{p}_{ij}=\sum_{i\in E}\pi_i\frac{\pi_j p_{ji}}{\pi_i}=\pi_j$$

so $\boldsymbol{\pi}$ is again a stationary distribution for the time–reversed process. Hence we have proved

Theorem 2.24. *If $\{X_n\}_{n\geq 0}$ is a stationary Markov chain with stationary distribution $\boldsymbol{\pi}=\{\pi_i\}_{i\in E}$ and $\pi_i>0$ for all $i\in E$ and transition probabilities p_{ij}, then the time reversed process $\tilde{X}=0=X_N,\tilde{X}_1=X_{N-1},...,\tilde{X}_N=X_0$ for some fixed $N\in\mathbb{N}$ is a time–homogenous Markov chain with transition probabilities*

$$\tilde{p}_{ij}=\frac{p_{ji}\pi_j}{\pi_i}.$$

The transition matrix $\tilde{\boldsymbol{P}}$ may be written as

$$\tilde{\boldsymbol{P}} = \boldsymbol{\Delta}^{-1}(\boldsymbol{\pi})\boldsymbol{P}'\boldsymbol{\Delta}(\boldsymbol{\pi}),$$

where $\boldsymbol{\Delta}(\boldsymbol{\pi})$ is the diagonal matrix with $\boldsymbol{\pi}$ vector in the diagonal. The vector $\boldsymbol{\pi}$ is a stationary distribution for the time reversed transition matrix $\tilde{\boldsymbol{P}}$.

We shall now consider whether the opposite is also the case. Hence assume that the time reversed chain is time–homogenous. We shall also assume that all $\mathbb{P}(X_n = i) > 0$ for all n and i and that $p_{ij} > 0$ for i, j. Then (2.5) yields

$$\tilde{p}_{ij} = p_{ji}\frac{\mathbb{P}(X_{N-n-1} = j)}{\mathbb{P}(X_{N-n} = i)}$$

which for $i = j$ and with $\rho_i = p_{ii}/\tilde{p}_{ii}$ implies that

$$\begin{aligned}\mathbb{P}(X_n = i) &= \frac{p_{ji}}{\tilde{p}_{ij}}\mathbb{P}(X_{n-1} = i)\\ &= \rho_i\mathbb{P}(X_{n-1} = i)\\ &= \ldots\\ &= \rho_i^n\mathbb{P}(X_0 = i).\end{aligned}$$

Then inserting this expression into (2.5) and re–arranging we obtain that

$$\left(\frac{\rho_i}{\rho_j}\right)^{N-n-1} = \rho_i\frac{\mathbb{P}(X_0 = i)}{\mathbb{P}(X_0 = j)}$$

which is valid for all n and which consequently implies that $\rho_i = \rho_j$ for all i, j.

Then

$$\mathbb{P}(X_n = i) = \rho^n\mathbb{P}(x_0 = i)$$

and summing over i we see that $\rho = 1$. But this means that $\mathbb{P}(X_n = i) = \mathbb{P}(x_0 = i)$ for all n and hence stationary. Thus we have proved the reverse under additional conditions.

Theorem 2.25. *If the time–reversed chain is time–homogeneous, if $\mathbb{P}(X_n = i) > 0$ for all i and n and if $p_{ij} > 0$ for all i, j, then the original chain must be stationary.*

Corollary 2.11. *Under the conditions of Theorem 2.25 we have that the forward and backward chains have the same distribution if and only if $\pi_i p_{ij} = \pi_j p_{ji}$.*

Proof. We know from Theorem 2.25 that the original chain has a stationary distribution $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$, and since the transition probabilities characterize the distribution of a Markov chain (Theorem 2.2), the equivalence in distributions amounts to $\tilde{p}_{ij} = p_{ij}$ for all i, j, but since $\tilde{p}_{ij} = p_{ji}\pi_j/\pi_i$, the result immediatly follows. □

The equation $\pi_i p_{ij} = \pi_j p_{ji}$ is commonly referred to as "detailed balance" and is used frequently in a more general setting in Markov chain Monte Carlo methods and Bayesian analysis.

We now consider time reversibility for Markov jump processes $\{X_t\}_{t\geq 0}$ with discrete state–space E and intensity matrix $\boldsymbol{\Lambda}$. Since $\lambda_{ij}dt$ is the probability of a transition from state i to state j in a small time interval $[t,t+dt)$, then we might suspect that time reversion is obtained by letting the intensities for the time reversed process be $\tilde{\lambda}_{ij} = \lambda_{ji}\pi_j/\pi_i$, where $\boldsymbol{\pi} = \{\pi_i\}_{i\in E}$ is the stationary distribution of $\{X_t\}_{t\geq 0}$.

Definition 2.22. Consider a Markov jump process $\{X_t\}_{t\in\mathbb{R}}$ in doubly infinite time. Then define the time–reversed process $\{\tilde{X}_t\}_{t\in\mathbb{R}}$ by

$$\tilde{X}_t = X_{-t-} = \lim_{s\uparrow -t} X_s.$$

The reason for defining the reversed process in term of the left limits is in order to preserve the cádlág (continuous from the right and limits from the left) property. Doubly infinite time is convenient in order to avoid problems by choosing a "reversion" point.

We proceed as follows. Make a discretization of length h and consider the corresponding Markov chain $\{X_{nh}\}_{n\in\mathbb{N}}$. If the Markov jump process is irreducible and recurrent, then the corresponding Markov chain $\{X_{nh}\}_{n\in\mathbb{N}}$ is irreducible, recurrent and aperiodic. The latter is easily seen from occupation times in the different states are exponentially distributed and hence can take any value with positive probabilities. Hence $p^t_{ij} > 0$ for all $t > 0$.

Theorem 2.26. *Let $\{X_t\}_{t\in\mathbb{R}}$ be an irreducible, recurrent and stationary Markov jump process with stationary distribution $\boldsymbol{\pi}$. Then the time reversed process $\{\tilde{X}_t\}_{t\in\mathbb{R}}$ is a well defined Markov jump process with intensity matrix*

$$\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Delta}^{-1}(\boldsymbol{\pi})\boldsymbol{\Lambda}'\boldsymbol{\Delta}(\boldsymbol{\pi}).$$

Furthermore, $\{\tilde{X}_t\}_{t\in\mathbb{R}}$ is stationary with stationary distribution $\boldsymbol{\pi}$.

Proof. Since $\{X_t\}_{t\in\mathbb{R}}$ is stationary, it is non–explosive and hence the time–reversed process is again a Markov jump process. We now perform a discretization argument in order to show that the process is time–homogenous and that the transition intensities are on the form described in the theorem.

Consider the discretized process with step length h, $\{X_{nh}\}_{n\in\mathbb{N}}$. This process is an irreducible and recurrent Markov chain (see the end of proof of Theorem 2.20 for a careful argument). Furthermore it is aperiodic due to the exponentially distributed occupation times in the different states implies that the discretized chain may be in the same state consecutively with positive probability (there is a path from a certain state i and back to i (recurrence) and the probability that the Markov jump process will undertake this path in time h is a convolution of exponential distributions and hence positive). Hence the chain is irreducible, recurrent and aperiodic. Since $\pi_i > 0$ for all i (irreducibility) we have by Theorem 2.24 that the reversed Markov chain is time–homogenous. This is hence also be the case for the corresponding reversed Markov jump process. Also, the reversed process will again be irreducible, recurrent and with stationary distribution $\boldsymbol{\pi}$. This follows from the continuity of the transition

probabilities (in fact, they are differentiable by Kolmogorov's forward and backward equations) where we may approximate a transition probability p_{ij}^x arbitrarily close by a discretized Markov chain.

Then consider its intensity matrix. By definition,

$$\begin{aligned}\tilde{\lambda}_{ij} &= \lim_{h\downarrow 0} \frac{\tilde{p}_{ij}}{h} \\ &= \frac{\pi_j p_{ji}^h/\pi_i}{h} \\ &= \frac{\pi_j \lambda_{ji}}{\pi_i}\end{aligned}$$

as was to be proved. □

2.11 Spectral properties of intensity and sub–intensity matrices

2.12 Jumps and occupations in Markov processes

In this section we are interested in the distribution of the total times a Markov process spends in different states during a fixed time interval $[0,t]$ as well as the distribution of the number of jumps between the different states. We shall provide a transform which in principle characterizes the joint distribution of both occupation times and jumps and from which we may calculate moments and conditional moments of these.

Before treating the general case, we shall turn our attention to the case of a two dimensional Markov jump process in which case we may derive the density of the total time spent in a state using a simple probabilistic argument. Consider a two–dimensional Markov jump process $\{X_t\}_{t\geq 0}$ with state space $E = \{1,2\}$ and intensity matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}.$$

Let $R_i(t)$ denote the total time the process has spent in state i, $i = 1,2$ during $[0,t]$, i.e.

$$R_i(t) = \int_0^t 1\{X_u = i\}du.$$

Notice that $R_1(t) + R_2(t) = t$ so that it is only necessary to find the marginal density of $R_1(t)$, say. Assume that $X(0) = 1$. Define

$$\begin{aligned} g_n(x;\lambda) &= \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \\ p_n(x;\lambda) &= \frac{(\lambda x)^n}{n!} e^{-\lambda x}. \end{aligned}$$

Thus $g_n(x;\lambda)$ is the density of an Erlang $\mathrm{Er}(\lambda;n)$ distribution, which is the distribution of the convolution of n independent exponential distributions with intensities λ, and $p_n(\lambda;x)$ are the Poisson probabilities with intensity λ.

The distribution of $R_1(t)$ has an atom at t of size $\exp(-\lambda_1 t)$ since this is the probability that $\{X(t)\}_{t\geq 0}$ will never jump during $[0,t]$ and hence $R_1(t) = t$ with this same probability. We now consider the absolute continuous part. In this case there will be at least one jump from state 1 to state 2. In general, if there are exactly n jumps from state 1 to state 2 during $[0,t]$, there must be either n or $n-1$ jumps from state 2 to state 1 during the same period depending on the state of $X(t)$. If $X(t) = 1$ then the number of state transitions from state 2 to 1 will also be n while it will only be $n-1$ if $X(t) = 2$. Let $A_{n,m}$ denote the event that there are exactly n transitions from 1 to 2 and m transitions from 2 to 1 during $[0,t]$, $m = n-1, n$. Let $f_{R_1(t)}(x)$ denote the absolute continuous part of the density of $R_1(t)$. Then

$$\begin{aligned} f_{R_1(t)}(x)dx = &\sum_{n=1}^{\infty} \mathbb{P}(R_1(t) \in [x, x+dx), A_{n,n}) \\ &+ \sum_{n=1}^{\infty} \mathbb{P}(R_1(t) \in [x, x+dx), A_{n,n-1}). \end{aligned}$$

If $R_1(t) = x$ then $R_2(t) = t - x$ and vice versa. The event $\{R_1(t) \in [x, x+dx), A_{n,n}\}$ is hence equivalent to that during $[0,x]$ there must be exactly n Poisson arrivals (in state 1) and that the convolution of the n exponential distributions (in state 2) exactly amounts to $t-x$. Similarly, the event $\{R_1(t) \in [x, x+dx), A_{n,n-1}\}$ is equivalent to the convolution of exactly n exponential distributions (of state 1) amounts to x and that there are exactly $n-1$ Poisson arrivals (in state 2) during $t-x$. Since the distributions are conditionally independent given their states, we have that

$$\begin{aligned} f_{R_1(t)}(x)dx = &\sum_{n=1}^{\infty} p_n(x;\lambda_1) g_n(t-x;\lambda_2) dx \\ &+ \sum_{n=1}^{\infty} g_n(x;\lambda_1) dx \; p_{n-1}(t-x;\lambda_2)..\end{aligned}$$

Inserting the analytical expressions for $g_n(x;\lambda)$ and $p_n(x;\lambda)$ respectively we get that

Theorem 2.27. *In a two–dimensional Markov jump process that starts in state 1, the density of the total time $R_1(t)$ spent in state 1 up to time t is given by*

$$\begin{aligned} f_{R_1(t)}(x) = e^{-\lambda_1 x} e^{-\lambda_2 (t-x)} &\left[\sqrt{\frac{\lambda_1 \lambda_2 x}{t-x}} I_1\left(2\sqrt{\lambda_1 \lambda_2 x (t-x)}\right) \right. \\ &\left. + \lambda_1 I_0(2\sqrt{\lambda_1 \lambda_2 x (t-x)}) \right], \end{aligned} \tag{2.6}$$

where $I_n(x)$ is the modified Bessel function of order n defined by

$$I_n(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+n}}{k!(n+k)!}.$$

The density of total time spent in state 2, $R_2(t) = t - R_1(t)$, under the same conditions has the density $f_{R_2}(x) = f_{R_1(t)}(t-x)$.

Corollary 2.12. *If $\boldsymbol{\pi} = (\pi_1, \pi_2)$ denotes the initial condition of the Markov jump process, $\pi_i = \mathbb{P}(X_0 = i)$, $i = 1,2$, then the density of $R_1(t)$ is given by*

$$f_{R_1(t)}(x) = \sqrt{\lambda_1\lambda_2}e^{-\lambda_1 x}e^{-\lambda_2(t-x)}\left(\pi_1\sqrt{\frac{x}{t-x}} + \pi_2\sqrt{\frac{t-x}{x}}\right) I_1(2\sqrt{\lambda_1\lambda_2 x(t-x)})$$
$$+(\lambda_1\pi_1 + \lambda_2\pi_2)e^{-\lambda_1 x}e^{-\lambda_2(t-x)}I_0(2\sqrt{\lambda_1\lambda_2 x(t-x)}).$$

Symmetrically, the density of $R_2(t)$ is given by the above expression interchanging x and $t-x$.

Proof. Given that the process initiates in state 1, the density is given by (2.6). Conditional on initiation in state 2, the density is given by the formula (2.6) interchanging

$$\lambda_1 x e^{-\lambda_2(t-x)}\left[\sqrt{\frac{\lambda_1\lambda_2 x}{t-x}}I_1\left(2\sqrt{\lambda_1\lambda_2 x(t-x)}\right) + \lambda_1 I_0(2\sqrt{\lambda_1\lambda_2 x(t-x)})\right]$$
$$+\pi_2 e^{-\lambda_1 x}e^{-\lambda_2(t-x)}\left[\sqrt{\frac{\lambda_1\lambda_2(t-x)}{x}}I_1\left(2\sqrt{\lambda_1\lambda_2 x(t-x)}\right) + \lambda_2 I_0(2\sqrt{\lambda_1\lambda_2 x(t-x)})\right].$$

Collecting the terms yields the result. □

Example 2.1. Consider a two–state Markov jump process with initial distribution $\boldsymbol{\pi} = (\frac{3}{4}, \frac{1}{4})$ and intensity matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}.$$

Then $R_1(10)$ has the density given in Figure 2.2.

In the following we now present the case of a general finite–dimensional Markov jump process. Let $\{X_t\}_{t\geq 0}$ be an p–state irreducible Markov jump process with intensity matrix $\boldsymbol{Q} = \{q_{ij}\}$. Let $N_{ij}(t)$ denote the number of jumps from state i to state j during $[0,t]$. We make the convention that $N_{ii}(t) = 0$ for i. Concerning the occupation times we make a slight generalization to continuously earned rewards. Let $\boldsymbol{a} = (a_1, a_2, ..., a_p)$ denote a real vector of rewards. When the process is in state i a constant reward at rate a_i is earned. Then the total reward earned when the process is in state i is

$$R_i(t) = \int_0^t a_i 1\{X(s) = i\}ds. \tag{2.7}$$

The total reward earned in $[0,t]$ is hence

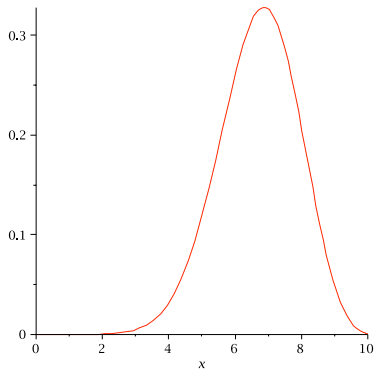

Fig. 2.2 The density of time spent in state one in a two dimensional Markov jump process

$$R(t) = \sum_{i=1}^{p} R_t(t).$$

If $a_i = 1$, $R_i(t)$ is the total time $\{X_t\}_{t\geq 0}$ has spent in state i during $[0,t]$.

We shall consider a transform which is a Laplace–Stieltjes transform in the continuous variables $R_i(t)$ and a generating function in the discrete variables $N_{ij}(t)$. Let $\boldsymbol{s} = (s_1, s_2, ..., s_p) \geq \mathbf{0}$ and $\boldsymbol{Z} = \{z_{ij}\}_{i,j=1,...,p}$. Define the transform

$$V^*_{ij}(\boldsymbol{s}, \boldsymbol{Z}; t) = \mathbb{E}\left(\exp\left(-\sum_{j=1}^{p} s_i R_i(t) \right) \prod_{h,k} z_{hk}^{N_{hk}(t)} 1\{X_t = j\} \middle| X_t = i \right).$$

This expectation is essentially an integral with respect to the conditional probability measure

$$V_{ij}(\boldsymbol{s}, \boldsymbol{z}; t) = \mathbb{P}(\boldsymbol{R}(t) \leq \boldsymbol{s}, \boldsymbol{Z} = \boldsymbol{z}, X(t) = j | X(i) = i),$$

where we understand that $\boldsymbol{R}(t) \leq \boldsymbol{s}$ if and only if $R_i(t) \leq s_i$ for all i. Conditioning on whether there is an jump or not during $[0,t]$, and if there is, to which state it jumps, we get that

$$\begin{aligned} V_{ij}(\boldsymbol{s}, \boldsymbol{z}; t) = {} & \delta_{ij} \exp(\boldsymbol{q}_{ii} t) \boldsymbol{U}(\boldsymbol{s} - a_i t \boldsymbol{e}_i) 1\{\boldsymbol{z} = \mathbf{0}\} + \\ & \sum_{k \neq i} \int_0^t \exp(q_{ii} u) q_{ik} V_{kj}(\boldsymbol{s} - a_i u \boldsymbol{e}_i, \boldsymbol{z} - \boldsymbol{J}(i,k); t - u) du, \end{aligned} \tag{2.8}$$

where $\boldsymbol{J}(i,k)$ is the matrix with element 1 at (i,k) and zero elsewhere and $\boldsymbol{U}(\boldsymbol{x} - \boldsymbol{b})$ is the multivariate distributions function for the distribution concentrated (degenerated) at $\boldsymbol{b}$. If there are no jumps, which happens with probability $\exp(q_{ii})$, then reward can only be earned in state i and no jumps have occurred. This explains the first term in (2.8). The second term is obvious conditioning on a jump from state i to state k.

Denote the first term by $V_{ij}^1(\boldsymbol{s},\boldsymbol{z};t)$ and the second integral by $V_{ij}^2(\boldsymbol{s},\boldsymbol{z};t)$. Obviously we have that

$$V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};t)=\int_0^\infty\cdots\int_0^\infty\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\exp\left(-\sum_{j=1}^p s_jx_j\right)dV_{ij}(\boldsymbol{x},\boldsymbol{K};t),$$

which is a kind of mixed generating function and *Laplace–Stieltjes* transform.

For a moment we shall consider an alternative formula which is the generating function mixed with the corresponding *Laplace* transform:

$$V_{ij}^0(\boldsymbol{s},\boldsymbol{Z};t)=\int_0^\infty\cdots\int_0^\infty\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\exp\left(-\sum_{j=1}^p s_jx_j\right)V_{ij}(\boldsymbol{x},\boldsymbol{K};t)dx_1\cdots dx_p.$$

We now insert (2.8) in this formula. Now $V_{ij}(\boldsymbol{s},\boldsymbol{z};t)=V_{ij}^1(\boldsymbol{s},\boldsymbol{z};t)+V_{ij}^2(\boldsymbol{s},\boldsymbol{z};t)$ and

$$\begin{aligned}
&\int_0^\infty\cdots\int_0^\infty\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\exp\left(-\sum_{j=1}^p s_jx_j\right)V_{ij}^1(\boldsymbol{x},\boldsymbol{K};t)dx_1\cdots dx_p\\
&=\int_0^\infty\cdots\int_0^\infty\exp\left(-\sum_{j=1}^p s_jx_j\right)\delta_{ij}e^{q_{ii}t}\boldsymbol{U}(\boldsymbol{s}-a_it\boldsymbol{e}_i)dx_1\cdots dx_p\\
&=\delta_{ij}e^{q_{ii}t}\int_0^\infty e^{-s_1x_1}dx_1\cdots\int_{a_it}^\infty e^{-s_ix_i}dx_i\cdots\int_0^\infty e^{-s_px_p}dx_p\\
&=\delta_{ij}e^{(q_{ii}-a_is_i)t}\frac{1}{s_1s_2\cdots s_p}.
\end{aligned}$$

The second terms is

$$\begin{aligned}
&\int_0^\infty\cdots\int_0^\infty\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\exp\left(-\sum_{j=1}^p s_jx_j\right)V_{ij}^2(\boldsymbol{x},\boldsymbol{K};t)dx_1\cdots dx_p\\
&=\int_0^\infty\cdots\int_0^\infty\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\exp\left(-\sum_{j=1}^p s_jx_j\right)\\
&\quad\cdot\left(\sum_{r\neq i}\int_0^t\exp(q_{ii}u)q_{ir}V_{rj}(\boldsymbol{s}-a_iu\boldsymbol{e}_i,\boldsymbol{z}-\boldsymbol{J}(i,r);t-u)du\right)dx_1\cdots dx_p\\
&=\sum_{r\neq i}\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\int_0^t\exp(q_{ii}u)q_{ir}\int_0^\infty\cdots\int_0^\infty\exp\left(-\sum_{j=1}^p s_jx_j\right)\\
&\quad\cdot V_{rj}(\boldsymbol{s}-a_iu\boldsymbol{e}_i,\boldsymbol{z}-\boldsymbol{J}(i,r);t-u)dx_1\cdots dx_pdu\\
&=\sum_{r\neq i}\sum_{\boldsymbol{K}}\prod_{k,\ell}z_{k\ell}^{K_{k\ell}}\int_0^t\exp(q_{ii}u)q_{ir}\int_0^\infty\cdots\int_{a_iu}^\infty\cdots\int_0^\infty\exp\left(-\sum_{j=1}^p s_jx_j\right)\\
&\quad\cdot V_{rj}(\boldsymbol{s}-a_iu\boldsymbol{e}_i,\boldsymbol{z}-\boldsymbol{J}(i,r);t-u)dx_1\cdots dx_pdu
\end{aligned}$$

We now make two change of variables. The first is $s' = s - a_i u$ and the second $k'_{ij} = k_{ij} - 1$. Then we get that

$$\begin{aligned}
&\int_0^\infty \cdots \int_0^\infty \sum_{\boldsymbol{K}} \prod_{k,\ell} z_{k\ell}^{K_{k\ell}} \exp\left(-\sum_{j=1}^{p} s_j x_j\right) V_{ij}^2(\boldsymbol{x},\boldsymbol{K};t)dx_1 \cdots dx_p \\
&= \sum_{r\neq i} \int_0^t \exp\left((q_{ii} - a_i s_i)u\right) q_{ir} z_{ir} V_{rj}^0(\boldsymbol{s},\boldsymbol{z};t-u)du.
\end{aligned}$$

Hence we have that

$$\begin{aligned}
V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) &= \delta_{ij} e^{(q_{ii}-a_i s_i)t} \frac{1}{s_1 s_2 \cdots s_p} \\
&\quad + \sum_{r\neq i} \int_0^t \exp\left((q_{ii} - a_i s_i)u\right) q_{ir} z_{ir} V_{rj}^0(\boldsymbol{s},\boldsymbol{z};t-u)du
\end{aligned}$$

and hence

$$\begin{aligned}
V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) e^{-(q_{ii}-a_i s_i)t} &= \delta_{ij} \frac{1}{s_1 s_2 \cdots s_p} \\
&\quad + \sum_{r\neq i} \int_0^t \exp\left(-(q_{ii} - a_i s_i)(t-u)\right) q_{ir} z_{ir} V_{rj}^0(\boldsymbol{s},\boldsymbol{z};t-u)du \\
&= \delta_{ij} \frac{1}{s_1 s_2 \cdots s_p} + \sum_{r\neq i} \int_0^t \exp\left(-(q_{ii} - a_i s_i)u\right) q_{ir} z_{ir} V_{rj}^0(\boldsymbol{s},\boldsymbol{z};u)du.
\end{aligned}$$

Differentiating both side of this equation yields

$$\begin{aligned}
\frac{d}{dt} V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) e^{-(q_{ii}-a_i s_i)t} - V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t)(q_{ii} - a_i s_i) e^{-(q_{ii}-a_i s_i)t} \\
= \sum_{r\neq i} e^{-(q_{ii}-a_i s_i)t} q_{ir} z_{ir} V_{rj}^0(\boldsymbol{s},\boldsymbol{z};t)
\end{aligned}$$

Dividing through by the exponential and using that $z_{ii} = 1$ we get that

$$\frac{d}{dt} V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) = -a_i s_i V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) + \left[(\boldsymbol{Q} \bullet \boldsymbol{z})\boldsymbol{V} q^0(\boldsymbol{s},\boldsymbol{z};t)\right]_{ij},$$

where $\bullet$ denotes the Schur–product of matrix multiplication, $\boldsymbol{V}^0(\boldsymbol{s},\boldsymbol{z};t) = \{V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t)\}_{ij}$ and with the initial condition that $V^0(\boldsymbol{s},\boldsymbol{z};0) = \delta_{ij}(s_1 \cdots s_p)^{-1}$. If we let $\boldsymbol{\Delta}(\boldsymbol{s})$ denote the diagonal matrix with $\boldsymbol{s} = (s_1, ..., s_p)$ as diagonal values, then we can write

$$\frac{d}{dt} V_{ij}^0(\boldsymbol{s},\boldsymbol{z};t) = \left[\left((\boldsymbol{Q} \bullet \boldsymbol{z}) - \boldsymbol{\Delta}(\boldsymbol{s})\boldsymbol{\Delta}(\boldsymbol{a})\right)\boldsymbol{V}^0(\boldsymbol{s},\boldsymbol{z};t)\right]_{ij}.$$

The Laplace transform $V_{ij}^0(\boldsymbol{s},\boldsymbol{Z};t)$ is related the corresponding Laplace–Stieltjes transform $V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};t)$ by $V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};t) = s_1 s_2 \cdots c_p V_{ij}^0(\boldsymbol{s},\boldsymbol{Z};t)$ as can be seen directly

using integration by parts. $V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};t)$ obviously satisfy the same differential equations, but the initial conditions now amounts to $V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};0)=\delta_{ij}$. The solution to this system of linear differential equations is simply the matrix–exponential of the coefficient matrix times the matrix of initial conditions, the latter being the identity matrix.

Thus we have proved that

Theorem 2.28. *The joint Laplace–Stieltjes transform and generating function $V_{ij}^*(\boldsymbol{s},\boldsymbol{Z};t)$ is given by*

$$\boldsymbol{V}^*(\boldsymbol{s},\boldsymbol{Z};t)=\exp\left[(\boldsymbol{Q}\bullet\boldsymbol{Z}-\boldsymbol{\Delta}(\boldsymbol{a})\boldsymbol{\Delta}(\boldsymbol{s}))t\right],$$

where $\boldsymbol{V}^*(\boldsymbol{s},\boldsymbol{Z};t)=\{V_{ij}^*(\boldsymbol{s},\boldsymbol{z};t)\}_{i,j=1,\ldots p}$.

This theorem gives the joint transform of the continuously earned rewards and the number of jumps up to time t. We can make slight generalization by allowing instantaneous rewards earned at jumps from a state k to a state ℓ. Let $c_{k\ell}$ denote the instataneous reward at a transition from state k to state ℓ. Let $S_{ij}(t)=c_{ij}N_{ij}(t)$ denote the total instantaneous reward earned by transition from state i to state j. Define matrix $\boldsymbol{C}=\{c_{k\ell}\}_{k,\ell}$ with $c_{kk}=0$. Finally we consider the total reward up to time t is given by

$$R^T(t)=\sum_i R_t(t)+\sum_{k\neq\ell}S_{k\ell}(t).$$

Letting $s_i=\xi_{k\ell}=s$ for all i,k,ℓ we define

$$\begin{aligned}\boldsymbol{U}^*(s,\boldsymbol{x}i(s);t)&=\left\{\mathbb{E}\left(e^{-sR^T(t)}1\{X(t)=j\}\middle|X(0)=i\right)\right\}_{i,j}\\&=\exp\left((\boldsymbol{Q}\bullet\boldsymbol{x}i(s)-s\boldsymbol{\Delta}(\boldsymbol{a}))t\right),\end{aligned}$$

where $\boldsymbol{x}i(s)=\{\exp(-sc_{k\ell})\}_{k,\ell}$. Of particular interest is to calculate the expected value and variance of the total reward. Define first two moment matrices $\boldsymbol{M}_1(t)$ and $\boldsymbol{M}_2(t)$ by

$$\begin{aligned}\boldsymbol{M}_1(t)_{ij}&=\mathbb{E}\left(R^T(t)1\{X(t)=j\}|X(0)=i\right)\\\boldsymbol{M}_2(t)_{ij}&=\mathbb{E}\left(\left[R^T(t)\right]^2 1\{X(t)=j\}\middle|X(0)=i\right)\end{aligned}$$

Theorem 2.29. *$\boldsymbol{M}_1(t)$ satisfies the differential equation*

$$\frac{d}{dt}\boldsymbol{M}_1(t)=\boldsymbol{M}_1(t)\boldsymbol{Q}+\exp(\boldsymbol{Q}t)\left(\boldsymbol{Q}\bullet\boldsymbol{C}+\boldsymbol{\Delta}(\boldsymbol{a})\right),$$

with initial condition $\boldsymbol{M}_1(0)=\boldsymbol{0}$, the matrix of zeros.

Proof. It is clear that

$$\boldsymbol{M}_1(t)=-\frac{\partial}{\partial s}\boldsymbol{U}^*(s,\boldsymbol{x}i(s);t)|_{s=0}.$$

First differentiate $\boldsymbol{U}^*(s;t)$ with respect to t obtaining

$$\frac{\partial}{\partial t}\boldsymbol{U}^*(s,\boldsymbol{xi}(s);t) = \boldsymbol{U}^*(s,\boldsymbol{xi}(s);t)\left[\boldsymbol{Q}\bullet\boldsymbol{xi}(s) - s\boldsymbol{\Delta}(\boldsymbol{a})\right]. \tag{2.9}$$

Next notice that $\boldsymbol{xi}(0)$ equals the matrix of 1's and $\frac{\partial}{\partial s}\boldsymbol{xi}(s)\Big|_{s=0} = -\boldsymbol{C}$. Thus in particular, $\boldsymbol{U}^*(0,\boldsymbol{xi}(0);t) = \exp(\boldsymbol{Q}t)$. Then applying the differential operator $-\frac{\partial}{\partial s}$ evaluated at 0 to (2.9) we get that

$$\frac{d}{dt}\boldsymbol{M}_1(t) = \boldsymbol{M}_1(t)\boldsymbol{Q} + e^{\boldsymbol{Q}t}\left(\boldsymbol{Q}\bullet\boldsymbol{C} + \boldsymbol{\Delta}(\boldsymbol{a})\right).$$

The initial condition $\boldsymbol{M}_1(0) = \boldsymbol{0}$ is obvious. □

Theorem 2.30. *$\boldsymbol{M}_2(t)$ satisfies the differential equation*

$$\frac{d}{dt}\boldsymbol{M}_1(t) = \boldsymbol{M}_1(t)\boldsymbol{Q} + 2\boldsymbol{M}_1(t)\left(\boldsymbol{Q}\bullet\boldsymbol{C} + \boldsymbol{\Delta}(\boldsymbol{a})\right) + \exp(\boldsymbol{Q}t)\left(\boldsymbol{Q}\bullet\boldsymbol{C}\bullet\boldsymbol{C}\right),$$

with initial condition $\boldsymbol{M}_2(0) = \boldsymbol{0}$.

Proof. Differentiate equation (2.9) twice with respect to s at zero. □

The following "marginal" expected rewards are of interest in statistical applications. Let

$$\begin{aligned}
M_{ij}^k(t) &= \mathbb{E}\left(R_k(t)1\{X(t)=j\}|X(0)=i\right)\\
F_{ij}^{k\ell}(t) &= \mathbb{E}\left(N_{k\ell}(t)1\{X(t)=j\}|X(0)=i\right)\\
\tilde{M}_{ij}(t) &= \mathbb{E}\left(R_k(t)|X(0)=i,X(t)=j\right)\\
\tilde{F}_{ij}^{k\ell}(t) &= \mathbb{E}\left(N_{k\ell}(t)|X(0)=i,X(t)=j\right)
\end{aligned}$$

Theorem 2.31. *The function $t \to M_{ij}^k(t)$ solves the differential equation*

$$\frac{d}{dt}M_{ij}^k(t) = \sum_\ell M_{i\ell}^l(t)q_{\ell j} + \exp(\boldsymbol{Q}t)_{ij}\delta_{jk},$$

and with the initial condition that $M_{ij}^k(0) = 0$.

Proof. Follows from Theorem 2.29 with $c_{k\ell} = 0$ for all k,ℓ and $c_i = 1$ if $i = k$ and 0 otherwise. □

Corollary 2.13. *$\tilde{M}_{ij}(t)$ is obtained in terms of $M_{ij}^k(t)$ by*

$$\tilde{M}_{ij}(t) = \frac{M_{ij}^k(t)}{\boldsymbol{e}_i'\exp(\boldsymbol{Q}t)\boldsymbol{e}_j}.$$

Proof. Just notice that

$$\mathbb{P}(X(t)=j|X(0)=i) = \boldsymbol{e}_i'\exp(\boldsymbol{Q}t)\boldsymbol{e}_j.$$

□

Next: corollaries for each case separately: continuous rewards, occupations times, jumps. Also two dimensional explicit case with probabilistic proof. Concrete examples. Then moments deduced from the transforms. Consider extracting conditional moments as well. May article on this.

2.13 Coincidence probabilities and the Karlin–McGregor Theorem

2.14 Regenerative stochastic processe

2.15 Piecewise deterministic Markov processes

Exercises

Exercise 2.1. Let $\{N(t)\}_{t\geq 0}$ denote a Poisson process with intensity β. Let $S_1, S_2, ...$ denote the arrival times of the process. Construct a new point process in following way. With probability p the arrival time S_i is an arrival for the new process as well and with probability $1-p$ it is deleted. Prove that the corresponding counting process is a Poisson process with intensity $p\beta$.

Exercise 2.2. Let $\{M(t)\}_{t\geq 0}$ and $\{N(t)\}_{t\geq 0}$ denote two independent Poisson processes with intensities α and β. Prove that the process $\{M(t)+N(t)\}_{t\geq 0}$ is a Poisson process with intensity $\alpha+\beta$.

Exercise 2.3. Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with intensity $\lambda > 0$. Prove that if $s > t$, then $N(s)$ conditional on $N(t) = n$ has a binomial distribution $\text{bin}(n, \frac{s}{t})$.

Exercise 2.4. Let $S_1, S_2, ...$ denote the arrival times of a Poisson process with intensity β. Prove that conditional on $N(t) = n$, $S_1, ..., S_n$ are independent and uniformly distributed over the interval $[0,t]$.

Exercise 2.5. Prove that $\{X_n\}_{n\in\mathbb{N}}$ is a Markov chain if and only if

$$\begin{aligned}&\mathbb{P}(X_{n+m} = i_{n+m}, ..., X_{n+1} = i_{n+1} \mid X_n = i_n, ..., X_0 = i_0)\\ &= \mathbb{P}(X_{n+m} = i_{n+m}, ..., X_{n+1} = i_{n+1} \mid X_n = i_n)\end{aligned}$$

for all $n \in \mathbb{N}, i_0, i_1, ..., i_{n+m} \in E$.

Exercise 2.6. Let $\{X_n\}_{n\geq 0}$ be a Markov chain. Prove that $\{X_{kn}\}_{n\geq 0}$ is again a Markov chain.

Exercise 2.7. Prove Theorem 2.4 .

Exercise 2.8. Prove that communication between states $\sim$ is an equivalence relation.

Exercise 2.9. Let $\boldsymbol{\nu}$ be a stationary distribution of a Markov chain. Prove that if the distribution of X_n is $\boldsymbol{\nu}$ then so is the distribution of X_{n+1}.

Exercise 2.10. Prove the Chapman–Kolmogorov's equation.

Exercise 2.11. Prove that both for Markov chains and Markov jump processes it is true that the time–reversed of the time–reversed transition matrix (or intensity matrix) coincide with the original transition matrix (intensity matrix).

Chapter 3
Phase–type distributions

3.1 Introduction

In this chapter we introduce a class of distributions which play an important role in various branches of applied probability such as risk theory, queueing theory and telecommunications. They are generated by Markov jump processes (or chains in the discrete case) with a finite number of states. We start our treatment with continuous time phase–type distributions rather than their discrete counterpart mainly due to the fact that they traditionally have received far more attention and therefore are more important concerning applications.

3.2 Continuous phase–type distributions

Consider a Markov jump process $\{X_t\}_{t\geq 0}$ on a state–space $E = \{1,2,...,p,p+1\}$ where states $1,2,...,p$ are transient and state $p+1$ absorbing. Then the process has an intensity matrix on the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 0 \end{pmatrix}, \tag{3.1}$$

where $\boldsymbol{T}$ is a $n \times n$ *sub–intensity* matrix. Since rows in an intensity matrix must sum to 0 we also have that

$$\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e},$$

where $\boldsymbol{e}$ is the p–dimensional column vector of 1's. The vector $\boldsymbol{t}$ is called the exit rate vector since it contains the rates by which the process jumps to the absorbing state.

We need to specify an initial distribution for the process. Let $\pi_i = \mathbb{P}(X_0 = i)$ and $\boldsymbol{\pi} = (\pi_1,...,\pi_p)$ be the initial distribution of $\{X_t\}_{t\geq 0}$ defined on the first p states only. We assume that $\boldsymbol{\pi}\boldsymbol{e} = \sum_{i=1}^{p} \pi_i = 1$; otherwise there is an atom at 0 of size $1 - \boldsymbol{\pi}\boldsymbol{e}$.

For the sake of clarity and notation we shall postpone this discussion to a separate section.

Let

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

be the time until absorption. The distribution of τ only depends on $\boldsymbol{\pi}$ and $\boldsymbol{T}$ since $\boldsymbol{t}$ is given in terms of $\boldsymbol{T}$.

Definition 3.1. The distribution of τ is said to have a phase–type distribution with initial distribution $\boldsymbol{\pi}$ and sub–intensity matrix $\boldsymbol{T}$. We write

$$\tau \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T}) \quad \text{or} \quad \tau \sim \mathrm{PH}_p(\boldsymbol{\pi}, \boldsymbol{T})$$

if the dimension p needs a specification.

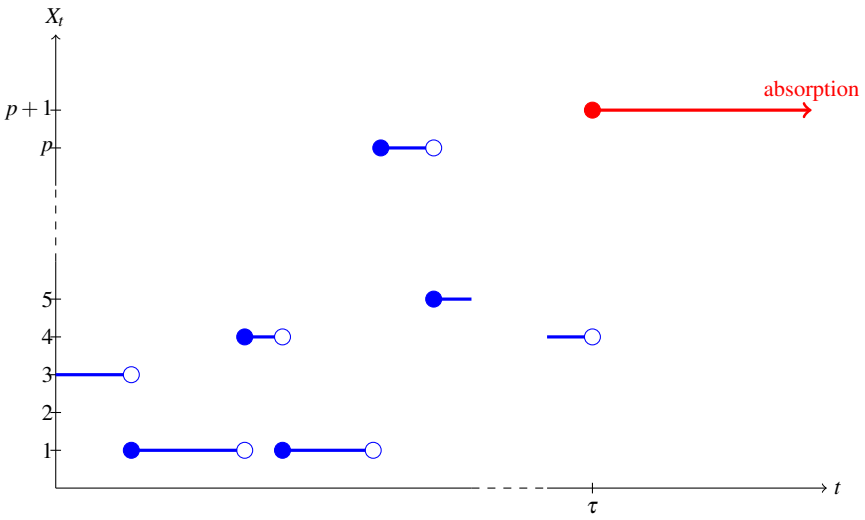

Fig. 3.1 A Markov process with p transient states (blue) and one absorbing state (red). The time τ until absorption takes place is called a Phase–type distribution. The arrow on the red line indicates that the process will remain in this state forever after. The filled and empty circles indicates that the process is continuous from the right.

Example 3.1. (Convolution of exponential distributions) Let $Y = X_1 + X_2 + \ldots + X_p$ where $X_1, X_2, \ldots$ are i.i.d. and exponentially distributed with means $1/\lambda_i$ respectively. We may think of Y being the total time it takes for a customer to pass through p nodes, where the time he has to spend in each node is random and exponentially distributed with mean $1/\lambda_i$. We may illustrate this interpretation via a flow diagram such as Figure 3.2.

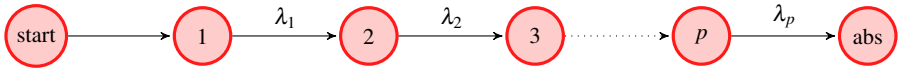

Fig. 3.2 Sums of exponential distributions. A customer moves from his initial position "start" to node 1, from node 1 he moves to node 2 etc. until he ends up in the absorbing state. In each node i he remains an exponentially distributed amount of time with mean $1/\lambda_i$.

The customer enters node 1 (from start) and remains there a random time which is exponentially distributed with mean $1/\lambda_1$. When leaving node 1 he passes immediately to node 2, in which he remains for a random time, independent of the previous, and which is exponentially distributed with mean $1/\lambda_2$. Continuing this way, he leaves the last node p at time Y. We may describe this construction by a Markov jump process, where each node is a state of the Markov process. We add an additional state, abs, which is an absorbing state, i.e. once he enters this state he will remain there for ever after.

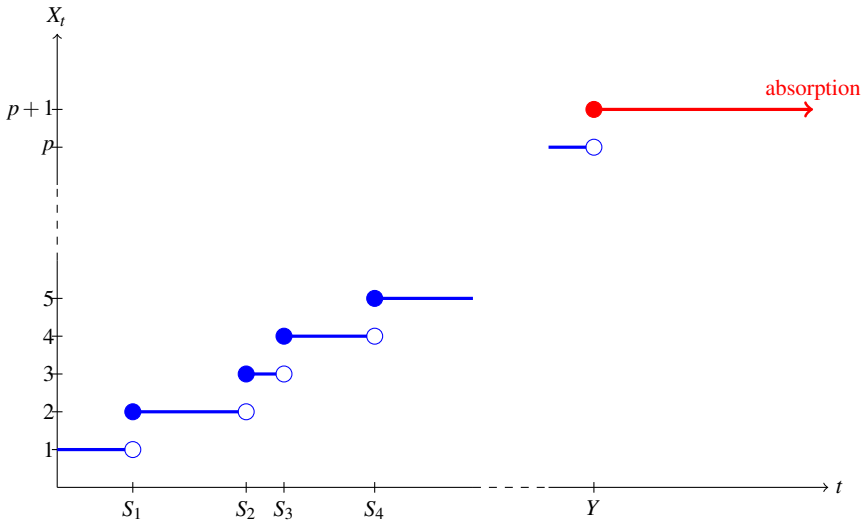

Fig. 3.3 The sum of independent exponentially distributed random variables may be represented as the time it takes the above Markov jump process to reach the absorbing state $p+1$. Here $S_i = X_1 + X_2 + \ldots + X_i$ is the time of the i'th jump. In particular, $S_p = Y$.

The Markov jump process will initiate in state 1, in which by property of Markov jump processes it will remain an exponentially distributed random time X_1. The rate for which the process will leave state 1 is λ_1. This is indicated in Figure 3.2 as the intensity of transition above the arrow. The process must then jump to state 2 with probability 1, in which it will remain an exponentially distributed random time, X_2, and hence jump to state 3 with probability 1. After repeating this construction, we

let the process get absorbed in a state "abs" (state $p+1$ in Figure 3.3), and we see that the time until absorption is exactly $Y = X_1 + ... + X_p$.

Thus we define a Markov jump process $\{X_t\}_{t\geq 0}$ with state–space

$$E = \{1, 2, 3, ..., p, p+1\},$$

where state $p+1$ is the absorbing state, initial distribution $\mathbb{P}(X_0 = 1) = 1$ and intensity matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & ... & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & ... & 0 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & -\lambda_p & \lambda_p \\ 0 & 0 & 0 & 0 & ... & 0 & 0 \end{pmatrix}.$$

Thus Y has a Phase–type distribution since it is the time until absorption in the Markov jump process just constructed (see Figure 3.3). The representation is given by $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$, where

$$\boldsymbol{\pi} = (1\ 0\ 0\ ...0), \quad \boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & ... & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & ... & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & -\lambda_p \end{pmatrix}.$$

We may notice already at this point, that we could have constructed the absorption time by letting the Markov jump process pass throthroughught the different states in any alternative order, so the construction of the Phase–type distributions is by no means unique. □

Definition 3.2. The distribution of $Y = X_1 + ... + X_p$, where $X_1, X_2, ..., X_p$ are independent and exponentially distributed with means $1/\lambda_i$ respectively, is called a generalized Erlang distribution and we write $Y \sim \mathrm{Er}_p(\lambda_1, ..., \lambda_p)$. If $\lambda_1 = ... = \lambda_p = \lambda$ then Y is said to have a p–dimensional Erlang distribution, in which case we write $Y \sim \mathrm{Er}_p(\lambda)$.

Example 3.2. (Mixtures of exponential distributions) Let f_i be the density of a exponential distribution with mean $1/\lambda_i$ and let $\pi_1, ..., \pi_p$ denote probabilities such that $\pi_1 + ... + \pi_p = 1$. Then the distribution expressed in terms of its density f given by

$$f(x) = \sum_{i=1}^{p} \pi_i f_i(x)$$

is a mixture of exponential distributions. We can draw/simulate a value for a random variable having density given by f, by first choosing an index i with probability π_i and then draw an exponentially distributed random variable with mean $1/\lambda_i$. This

procedure suggest the flow diagram as illustrated in Figure 3.4. A customer, which initiates at start, must choose a node i with probability π_i at which he receives a services which takes an exponentially distributed amount of time with mean $1/\lambda_i$. From node i he then translates to the absorbing state. The time he remains in the system has a distribution with density f. As the times at the nodes are exponentially distributed, we may represent the movements of the customer by that of a Markov jump process. Define a Markov jump process $\{X_t\}_{t\geq 0}$ with state–space

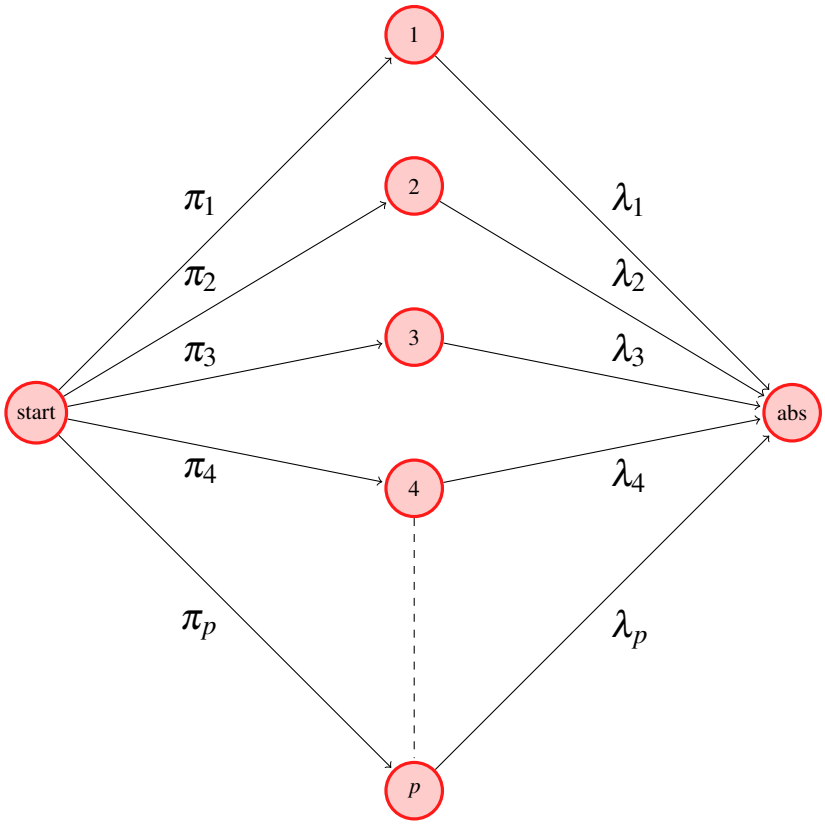

Fig. 3.4 Flow diagram for the mixtures of exponential distributions. A customer initiating at "start" goes to state i with probability π_i, where he remains for an exponentially distributed random time with mean $1/\lambda_i$. Hence the rate by which he leaves state i is λ_i as indicated. From state i he goes the an absorbing state.

$$E = \{1, 2, ..., p, p+1\},$$

initial distribution $\pi_i = \mathbb{P}(X_0 = i)$ (the Markov jump process must not start in the absorbing state) and intensity matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & 0 & 0 & 0 \ldots & 0 & \lambda_1 \\ 0 & -\lambda_2 & 0 & 0 \ldots & 0 & \lambda_2 \\ 0 & 0 & -\lambda_3 & 0 \ldots & 0 & \lambda_3 \\ \vdots & \vdots & \vdots & \vdots \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \ldots & -\lambda_p & \lambda_p \\ 0 & 0 & 0 & 0 \ldots & 0 & 0 \end{pmatrix}.$$

Hence the time until absorption is exactly that of the customer passing through the flow diagram. Hence a mixture of exponential distributions is of phase–type and the representation PH$(\boldsymbol{\pi}, \boldsymbol{T})$ which corresponds to the construction above is given by

$$\boldsymbol{\pi} = (\pi_1 \ \pi_2 \ .... \ \pi_p), \quad \boldsymbol{T} = \begin{pmatrix} -\lambda_1 & 0 & 0 & 0 \ldots & 0 \\ 0 & -\lambda_2 & 0 & 0 \ldots & 0 \\ 0 & 0 & -\lambda_3 & 0 \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \vdots & \vdots \\ 0 & 0 & 0 & 0 \ldots & -\lambda_p \end{pmatrix}.$$

Again we notice that the representation is by no means unique in that any permutation of the order of the states would result in the same distribution.

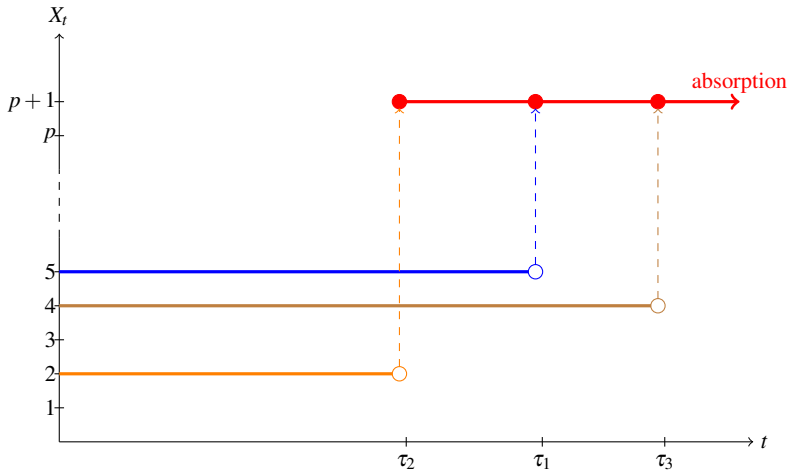

Fig. 3.5 Three realizations of a Markov jump process which initiates at random in some state $i = 1, ..., p$ according to the probabilities $\pi_1 ..., \pi_p$, remains in the chosen state for an exponentially distributed time with mean $1/\lambda_i$ for finally jumping to the absorbing state $p+1$. In this figure we shown three realizations of such a process generating random variables τ_1, τ_2 and τ_3 having a hyper–exponential distribution.

□

Definition 3.3. The distribution of the mixture of p independent exponential distributions with mean $1/\lambda_i$ is called a Hyper–exponential distribution, and we write $\text{Hexp}_p(\lambda_1,...,\lambda_p)$.

The following lemma is crucial to the following analysis.

Lemma 3.1. *The matrix exponential of (3.1) is given by*

$$\exp(\boldsymbol{\Lambda} s) = \begin{pmatrix} \exp(\boldsymbol{T}s) & \boldsymbol{e} - \exp(\boldsymbol{T}s)\boldsymbol{e} \\ 0 & 1 \end{pmatrix}$$

Proof. Using $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ we get that

$$\boldsymbol{\Lambda}^n = \begin{pmatrix} \boldsymbol{T}^n & -\boldsymbol{T}^n\boldsymbol{e} \\ \boldsymbol{0} & 0 \end{pmatrix}, \ \boldsymbol{\Lambda}^0 = \boldsymbol{I}_{p+1} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix},$$

where $\boldsymbol{I}_k$ denote the identity matrix of dimension $k \times k$. Then

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

from which the result is immediate. □

If $\boldsymbol{P}^s$ denotes the transition probability matrix at time–lag s, then we know from the differential equation of Kolmogorov that $\boldsymbol{P}^s = \exp(\boldsymbol{\Lambda} s)$. In particular, we have that the restriction of $\boldsymbol{P}^s$ to the first p transient states is simply $\exp(\boldsymbol{T}s)$, i.e.

$$\boldsymbol{P}^s|_E = \exp(\boldsymbol{T}s) \ .$$

Another way of putting this is by

$$\mathbb{P}(X(t) = j, t \le \tau | X(0) = i) = \left(e^{\boldsymbol{T}t}\right)_{ij}. \tag{3.2}$$

Theorem 3.1. *If $\tau \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ then the density f of τ is given by*

$$f(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{t}.$$

Proof. Since $f(x)dx = \mathbb{P}(\tau \in (x, x+dx])$ we get by conditioning on a initial state i and a state j at time x of the underlying Markov process $\{X_t\}_{t \ge 0}$ that

$$f(x)dx = \sum_{i,j=1}^{p} \mathbb{P}(\tau \in (x,x+dx]|X_x = j, X_0 = i)\mathbb{P}(X_x = j|X_0 = i)\mathbb{P}(X_0 = i)$$
$$= \sum_{i,j=1}^{p} \mathbb{P}(\tau \in (x,x+dx]|X_x = j)\boldsymbol{P}_{ij}^x \pi_i.$$

Since t_j, the j'th element of $\boldsymbol{t}$, is the exit rate to the absorbing state we have by (**??**) that

$$\mathbb{P}(\tau \in (x,x+dx]|X_x = j) = t_j dx.$$

Also $\boldsymbol{P}_{ij}^x = \exp(\boldsymbol{T}x)_{ij}$ for all $i, j = 1, ..., p$ so

$$f(x)dx = \sum_{i,j=1}^{p} \pi \exp(\boldsymbol{T}x)_{ij} t_j dx = \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{t} dx.$$

□

The distribution function can be found by integration of the density. However, there is a neat probabilistic argument leading to the result which we present as an alternative and we leave it to the reader to establish the result by simple integration if desired.

Theorem 3.2. *If $\tau \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ the the distribution function of τ is given by*

$$F(x) = 1 - \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{e}.$$

Proof. $\{\tau > x\} = \{X_x \in E\}$, where $\{X_t\}_{t \geq 0}$ is the underlying Markov jump process. Hence

$$1 - F(x) = \mathbb{P}(\tau > x) = \mathbb{P}(X_x \in E) = \sum_{j=1}^{p} \mathbb{P}(X_x = j)$$
$$= \sum_{i,j=1}^{p} \mathbb{P}(X_x = j|X_0 = i)\mathbb{P}(X_0 = i)$$
$$= \sum_{i,j=1}^{p} \pi_i \boldsymbol{P}_{ij}^x = \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{e}.$$

□

Before going on and inverting the sub–intensity matrix $\boldsymbol{T}$ we shall make sure that it is non–singular. This follows from the result

Theorem 3.3. *In a Markov jump process with $p+1$ states and intensity matrix*

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

the matrix $\boldsymbol{S}$ is invertible if and only if the states $1, 2, ..., p$ are transient.

Proof. Assume that all states $1, 2, ..., p$ are transient and consider the equation $\boldsymbol{v}\boldsymbol{T} = \boldsymbol{0}$. This is the same as

$$v_i\lambda_i = \sum_{j\neq i} v_j t_{ji} = \sum_{j\neq i} v_j \lambda_j q_{ji},$$

where $\lambda_i = -t_{ii} > 0$ for all i (otherwise state i would be absorbing and hence not transient) and q_{ji} are the probabilities of transition of the embedded Markov chain, i.e. $q_{ji} = t_{ji}/\lambda_j$. Let $\tilde{\boldsymbol{P}} = \{q_{ij}\}_{i,j=1,\ldots,p}$ with $q_{ii} = 0$ for all i. Then with $u_i = \lambda_i v_i$ and $\boldsymbol{u} = (u_1, \ldots, u_p)$ we have that $\boldsymbol{u} = \boldsymbol{u}\tilde{\boldsymbol{P}}$ and iterating $\boldsymbol{u} = \boldsymbol{u}\tilde{\boldsymbol{P}}^n$ for all n. The ij'th element of $\tilde{\boldsymbol{P}}^n$ is the probability that the embedded chain initiating in state i will be in state j n jumps ahead. Since states $1, 2, \ldots, p$ are transient, $\tilde{\boldsymbol{P}}^n \to \boldsymbol{0}$ as $n \to \infty$ and hence we conclude that transience implies that

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

hence that $\boldsymbol{v} = \boldsymbol{0}$ and that the columns of $\boldsymbol{T}$ are linearly independent. Hence $\boldsymbol{T}$ is invertible.

Now suppose that $\boldsymbol{T}$ is invertible. We shall prove that the states $1, 2, .., p$ are then transient. Let a_i denote the probability of eventual absorption into state $p+1$ initiating at state $i = 1, 2, \ldots, p$. Let q_{ij} denote the probability that the Markov jump process being in state i jumps to state j the next time it jumps (this is in fact the transition probabilities of the embedded Markov chain). Then conditioning on the next jump,

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

Now $q_{ij} = -t_{ij}/t_{ii}$ (see Definition 2.16) so

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

which in matrix notation amounts to

$$\boldsymbol{t} + \boldsymbol{T}\boldsymbol{a} = \boldsymbol{0}.$$

Since $\boldsymbol{T}$ is invertible, we solve for $\boldsymbol{a}$ and obtain

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

so $a_i = 1$ for all i, which in turn implies that all states $1, 2, \ldots, p$ must be transient. □

Corollary 3.1. *Let $PH_p(\boldsymbol{\alpha}, \boldsymbol{S})$ be a phase–type distribution. Then $\boldsymbol{S}$ is invertible.*

Proof. A phase–type distribution is defined as the absorption time of a Markov jump process with $p+1$ states among which $1, 2, \ldots, p$ are transient and $p+1$ is absorbing. The former is equivalent to $\boldsymbol{S}$ being invertible by Theorem 3.3. □

Theorem 3.4. *Let $\boldsymbol{U} = \{u_{ij}\} = (-\boldsymbol{T})^{-1}$. Then u_{ij} is the expected time spent in state j prior to absorption given initiation in state i.*

Proof. Let Z_j denote the time spent in state j prior to absorption. Then

$$\begin{aligned}\mathbb{E}_i(Z_j) &= \mathbb{E}_i\left(\int_0^\tau 1\{X(t)=j\}dt\right)\\ &= \int_0^\infty \mathbb{E}_i\left(1\{X(t)=j\}1\{\tau \geq t\}\right)dt\\ &= \int_0^\infty \mathbb{P}_i(X(t)=j, \tau \geq t)dt\\ &= \int_0^\infty \left(e^{\boldsymbol{T}t}\right)_{ij} dt \qquad \text{(by (3.2).}\end{aligned}$$

Define $\boldsymbol{v} = \left(\mathbb{E}_i(Z_j)\right)_{i,j}$. Then

$$\boldsymbol{v} = \int_0^\infty e^{\boldsymbol{T}t}dt = (-\boldsymbol{T})^{-1} = \boldsymbol{U}.$$

□

Corollary 3.2. *The mean of a $PH(\boldsymbol{\pi}, \boldsymbol{T})$ distributed random variable is $-\boldsymbol{\pi}\boldsymbol{T}^{-1}\boldsymbol{e}$.*

Proof. Just notice that $-\boldsymbol{T}^{-1}\boldsymbol{e}$ is the vector which i'th element is the expected time the process spends in $\{1,2,....,p\}$ prior to absorption given initiation in state i. □

HERE:[1] higher order moments, use moment distribution, show $(1-F(x))/\mu$ is PH if F is so, then use first order to derive second order.

Theorem 3.5. *The Laplace transform $L_X(s)$ of $X \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

$$L_X(s) = \mathbb{E}\left(e^{-sX}\right) = \boldsymbol{\pi}\left(s\boldsymbol{I} - \boldsymbol{T}\right)^{-1}\boldsymbol{t}.$$

Proof.

$$\begin{aligned}L_X(s) &= \int_0^\infty e^{-sx}\boldsymbol{\pi}e^{\boldsymbol{T}x}\boldsymbol{t}dx\\ &= \int_0^\infty \boldsymbol{\pi}e^{-sx\boldsymbol{I}}e^{\boldsymbol{T}x}\boldsymbol{t}dx\\ &= \int_0^\infty \boldsymbol{\pi}e^{-(s\boldsymbol{I}-\boldsymbol{T})x}\boldsymbol{t}dx\\ &= \boldsymbol{\pi}\left(s\boldsymbol{I}-\boldsymbol{T}\right)^{-1}\boldsymbol{t}.\end{aligned}$$

□

We notice that the Laplace transform of a phase–type distribution is hence a rational function in s. Hence we have proven that

[1] to be done

Theorem 3.6. *The class of phase–type distributions is a sub–class of the distributions with a rational Laplace transform, i.e. the class of matrix–exponential distributions.*

From the expression of the Laplace transform we obtain the moments by differentiation.

Corollary 3.3. *The n'th moment of $X \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

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

Phase–type distributions may be heavily over–parametrized and the dimension may not reflect the minimal possible dimension. An extreme case occurs when the exit rate vector is constant.

Theorem 3.7. *Let $X \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$. If $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e} = \lambda\boldsymbol{e}$ for some constant $\lambda > 0$, then $X \sim \exp(\lambda)$, i.e. an exponential distribution with rate λ.*

Proof.

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

where $S(x) = 1 - F(x) = \mathbb{P}(X > x)$ is the survival function of X. But since $f(x) = -S'(x)$ and $S(0) = 1$ we have that

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

which implies that $S(x) = \exp(-\lambda x)$, i.e. $X \sim \exp(\lambda)$. □

Example 3.3. Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$ be any initial distribution with $\pi_1 + \pi_2 + \pi_3 = 1$ and let

$$\boldsymbol{T} = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Then $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e} = (1,1,1)'$ and hence $X \sim \exp(1)$. This seemingly unnecessary complication for representing an exponential distribution may in fact be used for constructing correlated exponential random variables.

A higher order phase–type distribution may also collapse to a exponential distribution through the initial distribution. If $-\xi < 0$ is the eiqenvalue of maximum real part for $\boldsymbol{T}$, then (Perron–Frobenius) there is a corresponding eigenvector with purely positive elements, $\boldsymbol{v}$, say. We may assume that $\boldsymbol{v}\boldsymbol{e} = 1$. Then if $\boldsymbol{\pi} = \boldsymbol{v}$ we have that

$$1 - F(x) = \boldsymbol{v}e^{\boldsymbol{T}x}\boldsymbol{e} = 1 + \sum_{n=1}^{\infty}(-\xi)^n x^n \boldsymbol{v}\boldsymbol{e} = \exp(-\xi\, x).$$

Hence the phase–type distribution reduces to an exponential distribution with intensity ξ.

Theorem 3.8. *If $X \sim PH_p(\boldsymbol{\alpha}, \boldsymbol{S})$ and $Y \sim PH_q(\boldsymbol{\beta}, \boldsymbol{T})$ be independent phase–type distributions of dimension p and q respectively. Then $X+Y$ has a phase–type distributions with initial distribution $\boldsymbol{\pi} = (\boldsymbol{\alpha}, \mathbf{0})$, where $\mathbf{0}$ is a q–dimensional vector of zeros, and with sub–intensity matrix $\boldsymbol{U}$ given by*

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} \\ \mathbf{0} & \boldsymbol{T} \end{pmatrix},$$

where $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ and $\mathbf{0}$ is a $q \times p$–matrix of zeros.

Proof. We define a new Markov jump process $\{Z_t\}_{t \geq 0}$ with state–space

$$\{1, 2, ..., p, p+1, ..., p+q, p+q+1\}.$$

We define $\mathbb{P}(Z_0 = i) = \alpha_i$ for $i = 1, ..., p$ and zero otherwise, and its intensity matrix by

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T} & \boldsymbol{t} \\ \mathbf{0} & \mathbf{0} & 0 \end{pmatrix}.$$

This Markov jump process first evolves through a Markov jump process $\{X_t\}_{t \geq 0}$ defined on $\{1, 2, ..., p\}$, which is exactly the Markov jump process underlying X prior to absorption. It then exits states $\{1, 2, ..., p\}$ for good by a jump to one of the state $\{p+1, p+2, ..., p+q\}$. The transition rate for going from $i \in \{1, 2, ..., p\}$ to $j + p \in \{p+1, ..., p+q\}$ is $s_i \pi_j$. Hence the rate by which $\{Z_t\}_{t \geq 0}$ exits from state i to some state $j + p \in \{p+1, ..., p+q\}$ is simply s_i. Thus the time from initiation until $\{Z_t\}_{t \geq 0}$ leaves states $\{1, 2, ..., p\}$ has a Phase–type distribution $\mathrm{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$. At the time of exit from states $\{1, 2, ..., p\}$, a new state $\{p+1, ..., p+q\}$ is chosen independently of the exiting state according to the probabilities $\beta_j, j = 1, ..., q$, where β_j becomes the probability that the process jumps to state $j + p$. From hereafter, $\{Z_t\}_{t \geq 0}$ evolves according to

$$\begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \mathbf{0} & 0 \end{pmatrix},$$

and the remaining time until absorption evidently has a Phase–type distribution $\mathrm{PH}_q(\boldsymbol{\beta}, \boldsymbol{T})$ and which by construction is independent of $\mathrm{PH}_p(\boldsymbol{\alpha}, \boldsymbol{S})$. The total time until absorption has a distribution which is then exactly that of $X + Y$. See Figure 3.6 for details on the construction. □

Thus, by iteration, a convolution of a finite number of phase–type distributions is again a phase–type distribution. The resulting phase–type distribution has a dimension which is (at most) the sum of the dimensions of the terms in the convolution.

Theorem 3.9. *Let $X \sim PH_p(\boldsymbol{\alpha}, \boldsymbol{S})$, $Y \sim PH_q(\boldsymbol{\beta}, \boldsymbol{T})$ and*

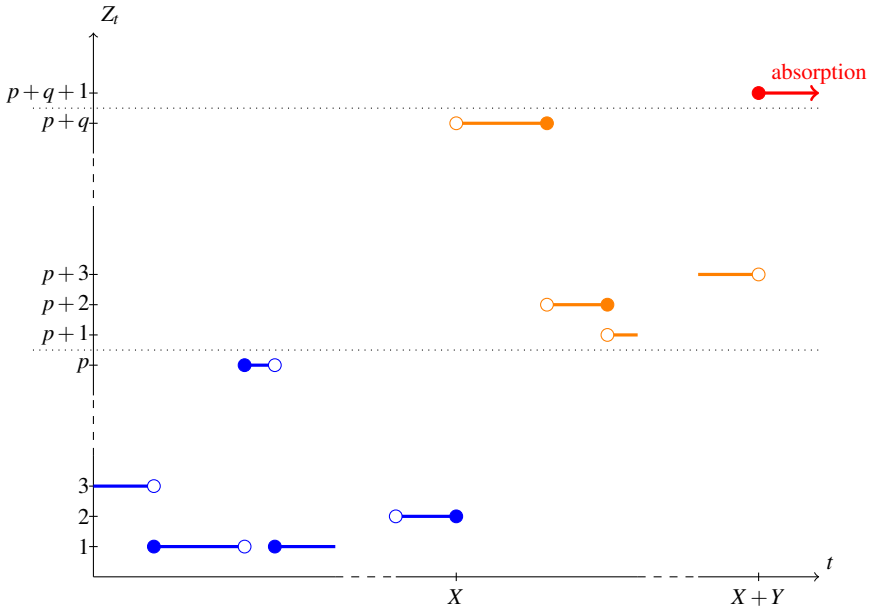

Fig. 3.6 The construction of a Markov jump process which has $X+Y$ as absorption time, where X and Y are two independent Phase–type distributed random variables. The blue process corresponds to the Markov jump process underlying the random variable X up to the time of absorption, whereas the orange corresponds to the process underlying Y. The orange process is started at the time of exit of the blue process and according to its original initial distribution making the two processes independent.

$$U = \begin{cases} X \text{ with probability } \rho_1 \\ Y \text{ with probability } \rho_2 \end{cases},$$

where $\rho_1 + \rho_2 = 1$. *Then* U *has a phase–type distribution with initial vector* $(\rho_1\boldsymbol{\alpha}, \rho_2\boldsymbol{\beta})$ *and sub–intensity matrix*

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T} \end{pmatrix}.$$

Proof. Left to the reader. □

So the class of phase–type distributions is closed under finite mixtures, but the class is not closed under arbitrary infinite mixtures. Consider a mixture of Erlang distributions where the order of the distributions is taken from a Poisson distribution. In this case we obtain a non–rational Laplace transform, see Exercise 3.1. When the mixing probabilities are given by a discrete phase–type then the infinite mixture is of phase–type too.

Theorem 3.10. *Consider a random variable* N *that is discrete phase–type distributed with representation* $(\boldsymbol{\alpha}, \boldsymbol{S})$ *and continuous random variables* X_i *that are phase–type distributed with representation* $(\boldsymbol{\beta}, \boldsymbol{T})$. *The random variable* $Z = \sum_{i=0}^{N} X_i$

follows a continuous phase–type distribution with representation $(\boldsymbol{\gamma}, \boldsymbol{L})$*, where*

$$\boldsymbol{\gamma} = (\boldsymbol{\alpha} \otimes \boldsymbol{\beta}) \tag{3.3}$$

and

$$\boldsymbol{L} = [\boldsymbol{I} \otimes \boldsymbol{T} + \boldsymbol{S} \otimes \boldsymbol{t}\boldsymbol{\beta}] \tag{3.4}$$

Proof. Consider a two–dimensional Markov chain, where the first coordinate corresponds to a state in the Markov chain of the discrete phase–type distribution and the second coordinate corresponds to a state of the continuous phase–type distribution, and let the state space be lexicographically ordered with the state of the discrete chain taking the role as the most significant digit. The probability of initiating in a state corresponding to state i of the discrete chain and state j of the continuous chain is $\alpha_i\beta_j$ which accounts for the form of $\boldsymbol{\gamma}$. Now before the first absorption related to X_1 the Markov chain evolves according to $\boldsymbol{T}$ with only chances in the second coordinate ordered as the least significant digit. An absorption related to a continuous variable occurs with intensities governed by $\mathbf{t}$. Upon such an absorption a new discrete state will be chosen according to $\boldsymbol{S}$ with the probability of exiting given by $\boldsymbol{s}$. In this way we can have changes in the first coordinate of the process corresponding to the most significant digit. Finally, the new level of the second coordinate has to be picked which governed by $\boldsymbol{\beta}$. □

Example 3.4. Suppose $N \sim geo(1-p)$ and $X_i \sim \exp(\lambda)$ then $(\boldsymbol{\alpha}, \boldsymbol{S}) = ((1), [p])$ and $(\boldsymbol{\beta}, \boldsymbol{T}) = ((1), [-\lambda])$ and we get that Z is phase–distributed with $((1), [-p\lambda])$. This is a restatement of the well–known result, that a mixture of Erlang distributions with geometric weights gives an exponential distribution. This result is used when showing that the waiting time in the M/M/1 queue is exponential.

♠ remember to relate back to this result in the Chapter on renewal theory, it can be seen as a terminating renewal process

The following theorem generalizes the result on particularly the minimum of a finite number of exponential distributions.

Theorem 3.11. *With* $X \in PH(\boldsymbol{\alpha}, \boldsymbol{S})$ *and* $Y \in PH(\boldsymbol{\beta}, \boldsymbol{T})$ *we have that* $Z_{min} = \min(X, Y)$ *is phase distributed with representation* $(\boldsymbol{\gamma}, \boldsymbol{L})$ *given by 3.5:*

$$\boldsymbol{L} = \boldsymbol{S} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{T}, \tag{3.5}$$

where $\boldsymbol{\gamma} = \boldsymbol{\alpha} \otimes \boldsymbol{\beta}$*, and that* $Z_{max} = \max(X, Y)$ *is phase type distributed with representation* $(\boldsymbol{\kappa}, \boldsymbol{K})$ *given by 3.6:*

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{S} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{T} & \boldsymbol{I} \otimes \boldsymbol{t} & \boldsymbol{s} \otimes \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T} \end{bmatrix} \tag{3.6}$$

with $\boldsymbol{\kappa} = (\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{\alpha}\alpha_{y,m+1}, \alpha_{x,k+1}\boldsymbol{\beta})$.

Proof. ♠ Her vil en figur med to "konkurrerende" fasetypefordelinger være meget fin

The sub–intensity matrix $\boldsymbol{S} \oplus \boldsymbol{T}$ of the theorem corresponds to a superposition of the two Markov chains induced by X and Y, where the initial state is chosen according to the probability vector $\boldsymbol{\alpha} \otimes \boldsymbol{\beta}$. The random variable $Z_{\min}$ corresponds to the time to absorption of this Markov chain. For the maximum we will have to wait for the second of the two processes to be absorbed. The matrices $\boldsymbol{I} \otimes \boldsymbol{t}$ and $\boldsymbol{s} \otimes \boldsymbol{I}$ corresponds to the event that one of the two chains enters the absorbing state while the state of the other remain unchanged. The Markov chain then proceeds according to $\boldsymbol{S}$ or $\boldsymbol{T}$ until absorption. □

In general any finite order statistics of a phase–type distribution will be phase–type distributed.

Theorem 3.12. *The class of phase–type distributions is dense in the class of distributions on* $\mathbb{R}_+$.

3.3 Discrete phase–type distributions

Similar to the continuous case, we shall define a discrete phase–type distribution as the time until absorption of a finite–state space Markov chain with one absorbing state and the rest being transient. The analysis of the discrete case is rather similar to the continuous one, one of the main differences being that transitions to the same state may are indeed possible for Markov chains.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a Markov chain with state–space $\{1,2,...,p,p+1\}$, where the states $1,2,...,p$ are transient and $p+1$ is absorbing. Let $\pi_i = \mathbb{P}(X_0 = i)$, $\boldsymbol{\pi} = (\pi_1,...,\pi_p)$ and assume that $\boldsymbol{\pi e} = \pi_! + ... + \pi_p = 1$. Then $\{X_n\}_{n\in\mathbb{N}}$ has a transition matrix $\boldsymbol{P}$ on the form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{pmatrix}, \tag{3.7}$$

where $\boldsymbol{T}$ is a $p \times p$ sub–transition matrix and $\boldsymbol{t}$ is a p–dimensional column vector. Since t_i is the probability of jumping to absorbing state directly from state i we shall refer to these probabilities as exit probabilities, a similar terminology as for the continuous case. Since the rows sum to 1 we have that

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

Definition 3.4. Let $\tau = \inf\{n \geq 1 | X_n = p+1\}$ be the time until absorption. Then we say that τ has a (discrete) phase–type distribution with initial distribution $\boldsymbol{\pi}$ and sub–transition matrix $\boldsymbol{T}$, and we write

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

Remark 3.1. Please notice that similarly to the continuous time definition, this definition does not allow for an atom at zero. Atoms at zero are easy to handle and shall be considered whenever needed.

Example 3.5. (Geometric distribution) Consider the geometric distribution with with parameter p, Geom(p). Its density is given by

$$f(x) = p^{x-1}(1-p), \;\; x = 1,2,....,$$

where $0 < p < 1$ is a probability often interpreted as the probability of succes or failure. Then the distribution is discrete phase–type with a representation $\boldsymbol{\pi} = (1)$, $\boldsymbol{T} = (p)$ and $\boldsymbol{t} = (1-p)$.

Lemma 3.2. *For $n \geq 1$ we have that*

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

Proof. Left to the reader. □

As for the continuous case we observe that

$$\boldsymbol{P}^{(n)} | \{1,2...,p\} = \boldsymbol{T}^n,$$

and

$$\mathbb{P}_i(X_n = j, \tau > n) = (\boldsymbol{T}^n)_{ij}. \tag{3.8}$$

Theorem 3.13. *Let $\boldsymbol{U} = \{u_{ij}\} = (\boldsymbol{I} - \boldsymbol{T})^{-1}$. Then u_{ij} is the expected time the Markov chain spends in state j prior to absorption given that it initiates in state i.*

Proof. Let Z_j denote the time spent in state j prior to absorption. Then

$$\begin{aligned} \mathbb{E}_i(Z_i) &= \mathbb{E}_i \left(\sum_{n=0}^{\tau-1} 1\{X_n = j\} \right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_i \left(X_n = j, \tau - 1 \geq n \right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_i \left(X_n = j, \tau > n \right) \\ &= \sum_{n=0}^{\infty} (T^n)_{ij}. \end{aligned}$$

Putting on matrix for $\boldsymbol{v} = (\mathbb{E}_i(Z_j))_{ij}$ then gives that

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

□

Theorem 3.14. *The density f of $\tau \sim DPH(\boldsymbol{\pi}, \boldsymbol{T})$ is given by*

$$f(n) = \boldsymbol{\pi} \boldsymbol{T}^{n-1} \boldsymbol{t}, \;\; n \geq 1.$$

Proof. The probability that the Markov chain is in (transient) state $i \in \{1,...,p\}$ at time $n-1$ is

$$\sum_{j=1}^{p} \pi_j \left(\boldsymbol{P}^{n-1}\right)_{ij} = \sum_{j=1}^{p} \pi_j \left(\boldsymbol{T}^{n}\right)_{ij} = \left[\boldsymbol{\pi}\boldsymbol{T}^{n-1}\right]_i.$$

Given that the Markov chain is in state i by time $n-1$, the probability of absorption at time n is simply t_i. Thus

$$f(n) = \sum_{i=1}^{p} \left(\boldsymbol{\pi}\boldsymbol{T}^{n-1}\right)_i t_i = \boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t}.$$

□

Theorem 3.15. *The distribution function F of $\tau \sim DPH(\boldsymbol{\pi},\boldsymbol{T})$ is given by*

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

Proof. From (3.8) we get that

$$\begin{aligned} 1 - F(n) &= \mathbb{P}(\tau > n) \\ &= \sum_{i,j=1}^{p} \pi_i \mathbb{P}_i \left(X_n = j, \tau > n\right) \\ &= \boldsymbol{\pi}\boldsymbol{T}^{n}\boldsymbol{e}. \end{aligned}$$

□

Theorem 3.16. *Let $X \sim DPH(\boldsymbol{\pi},\boldsymbol{T})$ with density f. The probability generating function $\hat{f}$ is then given by the rational function*

$$\hat{f}(z) = \mathbb{E}(z^X) = z\boldsymbol{\pi}\left(\boldsymbol{I} - z\boldsymbol{T}\right)^{-1}\boldsymbol{t} = \boldsymbol{\pi}\left(z^{-1}\boldsymbol{I} - \boldsymbol{T}\right)^{-1}\boldsymbol{t},$$

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

Proof. Let $|z| \leq 1$. Then $\hat{f}(z) = \mathbb{E}(z^X) \leq 1 < \infty$, and

$$\begin{aligned} \hat{f}(z) &= \mathbb{E}(z^X) \\ &= \sum_{n=1}^{\infty} z^n \mathbb{P}(X = n) \\ &= \sum_{n=1}^{\infty} z^n \boldsymbol{\pi}\boldsymbol{T}^{n-1}\boldsymbol{t} \\ &= z\sum_{n=0}^{\infty} \boldsymbol{\pi}\left(z\boldsymbol{T}\right)^{n}\boldsymbol{t} \\ &= z\left(\boldsymbol{I} - z\boldsymbol{T}\right)^{-1}\boldsymbol{t}. \end{aligned}$$

The last step follows from the series $\sum_{n=0}^{\infty}(z\boldsymbol{T})^n$ being convergent ($z \leq 1$) and the inverse of $\boldsymbol{I} - z\boldsymbol{T}$ exists unless z is an eigenvalue for $\boldsymbol{T}$. [2] □

Moments; factorial moments;

Let us consider a relation between continuous and discrete phase–type distributions in the following. If $\boldsymbol{S}_c = \{s_{ij}\}_{i,j=1,..,p}$ is the sub–generator of a continuous phase–type distribution with representation PH$(\alpha, \boldsymbol{S}_c)$, then there exists a constant $a > 0$ such that $\boldsymbol{S}_d := \boldsymbol{I} + \frac{1}{a}\boldsymbol{S}_c$ is sub–transition matrix, which is then the generator for a discrete phase–type distribution with representation DPH$(\alpha, \boldsymbol{S}_d)$. Indeed we may take any a larger that $\max_i(-s_{ii})$. On the other hand we may also conclude, that given sub–transition matrix $\boldsymbol{S}_d$ which generates a discrete phase–type distribution, then there exists a constant $a > 0$ such that $S_c := a(\boldsymbol{S}_d - \boldsymbol{I})$ is a sub–intensity matrix which generate phase–type distributions. If we use this relation between $\boldsymbol{S}_c$ and $\boldsymbol{S}_d$ see for example that the Laplace transform of the continuous phase–type distribution can be written as

$$\begin{aligned} L_X(s) &= \boldsymbol{\alpha}\,(s\boldsymbol{I} - \boldsymbol{S}_c)^{-1}\,(-\boldsymbol{S}_c\mathbf{e}) \\ &= \boldsymbol{\alpha}\,(s\boldsymbol{I} - a(\boldsymbol{S}_d - \boldsymbol{I}))^{-1}\,(-a(\boldsymbol{S}_d - \boldsymbol{I})\mathbf{e}) \\ &= a\boldsymbol{\alpha}\,((s+a)\boldsymbol{I} - \boldsymbol{S}_d)^{-1}\,(\boldsymbol{I} - \boldsymbol{S}_d)\mathbf{e} \\ &= \boldsymbol{\alpha}\left(\frac{s+a}{a}\boldsymbol{I} - \boldsymbol{S}_d\right)^{-1}(\boldsymbol{I} - \boldsymbol{S}_d)\mathbf{e}, \end{aligned}$$

which equals the probability generating function $\hat{f}(z)$ of DPH$(\alpha, \boldsymbol{S})$ with $z = a/(a+s)$. The other way around goes the same way. Hence we have proved

Theorem 3.17. *$\hat{f}(z)$ is the probability generating function of a discrete phase–type distribution if and only if for sufficiently large a, $\hat{f}(\frac{a}{a+s})$ is the Laplace transform of a continuous phase–type distribution.*

Exercises

3.1. Show that an infinite mixture of Erlang distributions with form parameter N and scale parameter λ has a non–rational Laplace transform $H(s) = \exp(-\mu s/(s+\lambda))$ whenever $N \sim \text{Pois}(\mu)$. Thus the distribution can neither be of phase–type nor matrix–exponential.

3.2. Consider the hyper–exponential distribution which is $PH\left((p, 1-p)\begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}\right)$ where $\lambda_1 \geq \lambda_2$. Find the similarity transformation that transforms this representation into $PH\left((1,0), \begin{bmatrix} -\lambda_1 & (1-p)*(\lambda_1 - \lambda_2) \\ 0 & -\lambda_2 \end{bmatrix}\right)$.

[2] Meget mere her eller tidligere om rationelle frembringende funktioner

3.3. Consider an experiment where we note the outcome of throwing a dice consecutively until a certain sequence of numbers appears for the first time. For example the first time two sixes will appear one after the other. Let $x, y \in \{1,2,3,4,5,6\}$ and let τ_{xy} be the number of throws with the dice until the sequence "xy" appears for the first time.

(1) Prove that τ_{xy} has a discrete phase–type distribution and find a representation.
(2) Prove that the means $\mathbb{E}(\tau_{66}) = 42$ and $\mathbb{E}(\tau_{56}) = 36$. Is it unexpected that they are different?

3.4. Prove that if $X \sim \mathrm{PH}(\alpha, \boldsymbol{S})$ then the embedded chain which originates from the underlying Markov jump process at jump times defines a discrete phase–type distribution, and find a representation.

Chapter 4
Renewal theory

4.1 General theory

A renewal process is a point process where the time between arrivals are independent and identically distributed apart from possibly the fist arrival. To model this, we consider $T_1, T_2, ...$ i.i.d. $\sim F$ where we shall assume that $F(0) = 0$, i.e. all variables T_i are strictly positive and the F is absolutely continuous. Let $S_n = T_1 + ... + T_n$. We interpret the T_i's as the distribution between arrivals. Hence multiple arrivals is impossible due to $F(0) = 0$. The S_i's are the corresponding arrival times. There are situations where it may be convenient to assume that the time of the first arrival is distinct from the rest. This occurs e.g. if the times between arrivals is not exponentially distributed and if the time initiation of observation does not coincide with the time of an arrival. We shall find such situations both in e.g. stationary renewal processes and ladder processes in random walks. In order to deal with an initial distribution distinct from zero, we define another arrival S_0 which may have distribution G, which again is assumed to be absolutely continuous or to be the degenerate distribution at zero, i.e. $S_0 = 0$ a.s. Let T_0 denote the time until the arrival S_0 (actually, $T_0 = S_0$). If $S_0 = 0$ then the renewal process is said to be pure or zero–delayed, while if S_0 is not degenerate at zero, it is called delayed. Again, the intuition for the latter term is that observation initiates delayed. This does not imply that the mean of G is less than F or is stochastically smaller; in some situation it may be large like for example in the case of ladder height distributions.

An arrival at zero is convenient in some situation (random walk ladder processes e.g.) and may not be in other. We can always get rid of the point at zero by considering a delayed renewal process with $G = F$. Manu formulas, however, simplify by having an initial arrival at zero.

4.1.1 Pure renewal process

Define the counting process $\{N(t)\}_{t\geq 0}$ by

$$N(t) = \inf\{n : S_n > t\} = \sup\{n : S_n \leq t\} + 1,$$

i.e. $N(t)$ the number of arrivals up to time t (including the one at zero which is not a "real" arrival).

Define the renewal function $U(t) = \mathbb{E}(N(t))$. There exists the following important relation between the counting process and the arrival epochs,

$$N(t) \leq n \quad \text{if and only if} \quad S_n > t. \tag{4.1}$$

Applying (4.1) to the definition of $U(t)$, we get that

$$\begin{aligned} U(t) &= \sum_{n=0}^{\infty} \mathbb{P}(N(t) > n) \\ &= \sum_{n=0}^{\infty} \mathbb{P}(S_n \leq t) \\ &= 1 + \sum_{n=1}^{\infty} F^{*n}(t) \\ &= \sum_{n=0}^{\infty} F^{*n}(t) \end{aligned}$$

where $*$ denotes the convolution operator, $F^{*0} \stackrel{\text{def}}{=} 1$ and $F^{*2} = F * F$, $F^{*n} = F * F^{*(n-1)}$.

The renewal function is the exected number of arrivals up to time t. Its derivative $u(t) = U'(t)$ is called the renewal density, and in spite of not being a density function in the classical sense of probability theory, it has the following property: $u(t)dt$ is the probability that there is an arrival in $[t, t+dt)$. To see this, we notice that $u(t)dt = U(t+dt) - U(t) = \mathbb{E}(N[t, t+dt))$ and since $F(0) = 0$ there may at most be one arrival in $[t, t+dt)$ (two or more arrivals having probability $o(dt)$), so we have that $u(t)dt = \mathbb{P}(\text{there is an arrival in } [t, t+dt))$ (up to en error of at most $o(dt)$).

Theorem 4.1. *$U(t) < \infty$ for all $t > 0$.*

Proof. First notice that since $F^{*(n+m)} = F^{*n} * F^{*m}$ we have that

$$\begin{aligned} F^{*(n+m)} &= \int_0^t F^{*m}(t-x) dF^{*n}(x) \\ &\leq F^{*m}(t) F^{*n}(t). \end{aligned}$$

Let $n = m * r + k$ for some fixed r. Then

$$F^{*n}(t) = F^{*(mr+k)}(t)$$

$$
\begin{aligned}
&= F^{*(r+((m-1)r+k))}(t) \\
&\leq F^{*r}(t)F^{*((m-1)r+k)}(t) \\
&\leq \ldots \\
&\leq [F^{*r}(t)]^m F^{*k}(t).
\end{aligned}
$$

Since the variables T_i are strictly positive, we may choose an $r \in \mathbb{N}$ such that $\mathbb{P}(T_1 + \ldots + T_r > t) > 0$. Actually, if the support of F is all $(0,\infty)$ then $r = 1$ is sufficirent. If the support is finite it may be necessary with several arrivals in order to reach t with a positive probability. We conclude that there exists an r such that $F^{*r}(t) < 1$. Hence

$$
\begin{aligned}
U(t) &= \sum_{n=0}^{\infty} F^{*n}(t) \\
&= \sum_{m=0}^{\infty} \sum_{k=0}^{r-1} F^{*(mr+k)}(t) \\
&\leq \sum_{k=0}^{r-1} F^{*k}(t) \sum_{m=0}^{\infty} [F^{*r}(t)]^m < \infty.
\end{aligned}
$$

□

A direct consequence of the above theorem is that $F^{*n}(t) \to 0$ as $n \to \infty$.

Consider the renewal function U. Conditioning on the time of the first arrival $T_1 = x$ we get that

$$
\begin{aligned}
U(t) &= \mathbb{E}(N(t)) \\
&= \int_0^{\infty} \mathbb{E}(N(t)|T_1 = x) f(x) dx \\
&= \int_0^{t} \mathbb{E}(N(t)|T_1 = x) f(x) dx \\
&= \int_0^{\infty} [1 + U(t-x)] f(x) dx \\
&= F(t) + \int_0^{t} U(t-x) f(x) dx \\
&= F(t) + \int_0^{t} U(t-x) dF(x) \\
&= F(t) + F * U(t),
\end{aligned}
$$

where f is the density of T_1. Hence U satisfies the equation

$$U = F + F * U.$$

This is an example of an renewal equation. The argument by conditioning on the first arrival time is referred to as a renewal argument. In general a renewal equation is an equation on the form

$$A(t) = a(t) + \int_0^t A(t-x)dF(x),$$

where a is a known function and A is an unknown function to be solved for.

Theorem 4.2. *Let $a(t)$ be a function which is bound on bounded intervals. Then there is one and only one solution, $A(t)$, which is bounded on bounded intervals to the equation*

$$A(t) = a(t) + \int_0^t A(t-x)dF(x),$$

and the solution is given by

$$A(t) = \int_0^t a(t-x)dU(x) = U * a(t).$$

Proof. First we prove that the proposed solution is bounded on bounded intervals. Recalling that $\sup_{0\le t\le T} f(t)$ is a norm (the uniform norm or supremum norm) we have that

$$\begin{aligned}
\sup_{0\le t\le T} |A(t)| &= \sup_{0\le t\le T} \left| \int_0^t a(t-x)dU(x) \right| \\
&\le \int_0^T \sup_{0\le s\le T} |a(s)| dU(x) \\
&= \sup_{0\le t\le T} |a(t)| U(T),
\end{aligned}$$

and since $U(T) < \infty$ the result follows. Now we prove that the proposed solution in fact satisfies the reneal equation.

$$\begin{aligned}
A(t) &= U * a(t) \\
&= a(t) + \left(\sum_{n=1}^{\infty} F^{*n} * a \right)(t) \\
&= a(t) + F * \left(\sum_{n=0}^{\infty} F^{*n} * a \right)(t) \\
&= a(t) + F * A(t).
\end{aligned}$$

Finally we probe uniqueness among solution which are bounded on bounded intervals. Let $B(t)$ be another solution which is bounded in bounded intervals. Then B satisfies $B = a + F * B$. Hence

$$\begin{aligned}
|A(t) - B(t)| &= |F * A(t) - F * B(t)| \\
&= |F * (A - B)(t)| \\
&= \ldots \\
&= |F^{*n} * (A - B)(t)|
\end{aligned}$$

$$
\begin{aligned}
&= \left| \int_0^t (A-B)(t-x)dF^{*n}(x) \right| \\
&\leq \int_0^\infty |A(t-x)-B(t-x)|\, dF^{*n}(x) \\
&\leq \sup_{0\leq s\leq t} |A(s)-B(s)| F^{**n}(t),
\end{aligned}
$$

and since A and B are bounded on bounded intervals, and since $F^{*n}(t) \to 0$ it follows that $A = B$. □

A nice application of Theorem 4.2 is the following. Consider the expected time of the first arrival after time t, i.e. $\mathbb{E}(S_{N(t)})$. First we apply a renewal argument conditioning upon the time of the first arrival. Let $A(t) = \mathbb{E}(S_{N(t)})$. Conditioning on $T_1 = x$ we see that if $x > t$, then $S_1 = x$ and so $S_{N(t)} = x$. If $x \leq t$, then $\mathbb{E}(S_{N(t)}|T_1 = x) = x + A(t-x)$. Thus

$$
\begin{aligned}
A(t) &= \int_t^\infty x dF(x) + \int_0^t (x + A(t-x)) dF(x) \\
&= \mathbb{E}(T_1) + \int_0^t A(t-x) dF(x).
\end{aligned}
$$

This is a renewal equation with $a(t) = \mathbb{E}(T_1)$ constant, and hence bounded. Thus $A(t)$ is given by the solution

$$
\begin{aligned}
A(t) &= \int_0^t \mathbb{E}(T_1) dU(x) \\
&= \mathbb{E}(T_1) U(t),
\end{aligned}
$$

that is, we have proved that

Theorem 4.3.

$$\mathbb{E}(S_{N(t)}) = \mathbb{E}(T_1) U(t).$$

Using this theorem we can calculate the limiting behaviour of $U(t)$.

Theorem 4.4. *If $\{N(t)\}_{t\geq 0}$ is a renewal process with times between arrivals having a finite mean μ, then*

$$\lim_{t\to\infty} \frac{U(t)}{t} = \frac{1}{\mu}.$$

Proof. As $t < S_{N(t)}$ we have that

$$t < \mathbb{E}(S_{N(t)}) = \mu U(t)$$

giving

$$\frac{U(t)}{t} > \frac{1}{\mu}$$

and hence

$$\liminf_{t\to\infty} \frac{U(t)}{t} \geq \frac{1}{\mu}.$$

In order to establish the opposite inequality with lim sup, we shall bound the variables T_i s and let

$$T_i^c = \begin{cases} T_i \text{ if } T_i \leq c \\ c \text{ if } T_i > c \end{cases}$$

Now consider the renewal process having inter-arrival times T_i^c, and let $N^c(t)$ denote the corresponding counting process. Since the T_i^c's are bounded by c we have that $S_{N^c(t)} \leq t + c$ and hence

$$t + c \geq \mathbb{E}(S_{N^c(t)}) = \mu^c U^c(t),$$

where μ^c is the mean of T_i^c and U^c is the corresponding renewal function. First,

$$\mu^c = \int_0^\infty \mathbb{P}(T_i^c > x)dx = \int_0^c \mathbb{P}(X_i^c > x)dx = \int_0^c (1 - F(x))dx,$$

where $T_i \sim F$ is the distribution function of the original inter-arrivals.On the other hand, $T_i^c \leq T_i$ and hence $N^c(t) \geq N(t)$ for all $t > 0$ and hence $U^c(t) \geq U(t)$. We conclude that

$$t + c \geq \mu^c U^c(t) \geq \mu^c U(t),$$

which is the same as

$$\frac{U(t)}{t} \leq \frac{1 + c/t}{\mu^c}.$$

Taking limit,

$$\limsup_{t\to\infty} \frac{U(t)}{t} \leq \frac{1}{\mu^c}.$$

This is valid for all c,and hence also in the limit

$$\limsup_{t\to\infty} \frac{U(t)}{t} \leq \lim_{c\to\infty} \frac{1}{\mu^c}.$$

But

$$\lim_{c\to\infty} \mu^c = \lim_{c\to\infty} \int_0^c (1 - F(x))dx = \int_0^\infty (1 - F(x))dx = \mu$$

so

$$\limsup_{t\to\infty} \frac{U(t)}{t} \leq \frac{1}{\mu}.$$

Hence the limit of $U(t)/t$ exists and equals $1/\mu$. □

4.1.2 Delayed and stationary renewal processes

Let $T_0 \sim G$ y $T_1, T_1, \ldots \sim F$. We assume that the T_i's are independent. Let $S_0 = T_0$ and $S_n = T_1 + \ldots + T_n$. Let $N(t)$ be the number of arrivals in the delayed process, and

$$U_D(t) = \mathbb{E}(N(t)).$$

We reserve the symbol U for the renewal function of the corresponding non–delayed process, i.e.

$$U(t) = \sum_{n=0}^{\infty} F^{*n}(t).$$

Applying a renewal argument, we get that

$$\begin{aligned} U_D(t) &= \mathbb{E}(N(t)) \\ &= \int_0^{\infty} \mathbb{E}(N(t)|T_1 = x) dG(x) \\ &= \int_0^{t} \mathbb{E}(N(t)|T_1 = x) dG(x) \\ &= \int_0^{t} U(t-x) dG(x) \\ &= \int_0^{t} G(t-x) dU(x) \quad (G * U = U * G). \end{aligned}$$

Thus U_D has exactly the form of a solution to the renewal equation

$$U_D(t) = G(t) + \int_0^t U_D(t-x) dF(x).$$

Of particular interest are the the so-called stationary renewal processes. Here we think of the processes having evolved for an infinite amount of time when the observer arrives. What is the distribution of the time until the first arrival after observation starts?

Formally we define,

Definition 4.1. A renewal process is stationary if $\{N_D(t+s) - N_D(t)\} \stackrel{D}{=} \{N_D(s)\}$.

For a stationary renewal process we have that $U_D(t) = ct$. To see this we simply take expected values of $U_D(t+s) - U_D(t) = U_D(s)$, i.e. $U_D(t+s) = U_D(t) + U_D(s)$ implying the desired form.

Now

$$U_D(t) = G(t) + \int_0^t U_D(t-x) dF(x)$$

Hence $U_D(t) = ct$ if and only if

$$ct = G(t) + \int_0^t c \cdot (t-y) F(dy)$$

or

$$G(t) = ct - \int_0^t c \cdot (t-y) F(dy).$$

Integration by parts yields

$$\int_0^t c \cdot (t-y) F(dy) = [c \cdot (t-y) F(y)]_0^t + c \int_0^t F(y) dy$$

from which

$$G(t) = c \int_0^t (1 - F(y)) dy.$$

If G is non–defective (has mass one), then $G(\infty) = 1$, and $c \int_0^\infty (1-F(y))dy = 1$. But $\int_0^\infty (1-F(y))dy = \int_0^\infty \mathbb{P}(X > y)dy = \mu$, where $X \sim F$. Hence $c = \mu^{-1}$. Thus

$$G(t) = \frac{1}{\mu} \int_0^t (1 - F(x)) dx.$$

Theorem 4.5. *A renewal process is stationary if and only if Y_0 has a density on the form*

$$f_{Y_0}(x) = \frac{1}{\mu}(1 - F(x)).$$

It follows that any stationary renewal process satisfies

$$\mathbb{E}(N(t, t+h)) = U_D(t+h) - U_D(t) = (t+h)/\mu - t/\mu = h/\mu.$$

Its derivative is given by

$$U_D'(t) = \frac{1}{\mu}.$$

4.1.3 Age and residual life–time processes

We shall consider some useful properties which are related to renewal process.

Definition 4.2. Let $\{N(t)\}_{t \geq 0}$ be a (possibly delayed) renewal process with arrival epochs $S_0, S_1, S_2, ...$ Then we define the following processes:

1. The residual life time $R_t = S_{N(t)} - t$, i.e. the time t until the next arrival.
2. The age $A_t = t - S_{N(t)-1}$, i.e. the time which has passed since the last arrival prior to time t. This property is only well defined if there actually is an arrival prior to time t.
3. The spread $S_t = A_t + R_t = S_{N(t)} - S_{N(t)-1}$, i.e. the time between the previous and coming arrival in a renewal processes observed at time t.

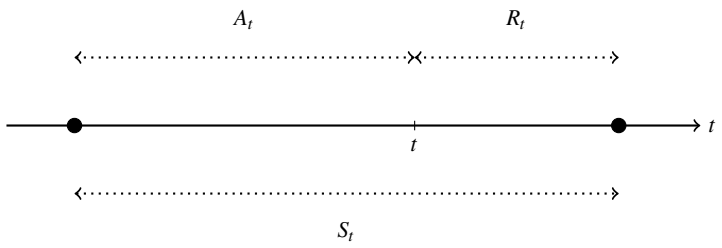

Fig. 4.1 The residual life–time R_t, the age A_t and the spread S_t in an inter–arrival interval of a renewal process which contains the time t. The two black circles are the arrivals just prior to and just after t.

Consider the age and residual life–time A_t and R_t at time t. Note that $A_t = x$ if and only if the inter–arrival time T of the interval which contains t satisfies $T > x$. Thus

$$\begin{aligned}\mathbb{P}(R_t > y|\, A_t = x) &= \mathbb{P}(T > x+y|\, T > x)\\ &= \frac{\mathbb{P}(T > x+y)}{\mathbb{P}(T > x)}\\ &= \frac{\bar{F}(x+y)}{\bar{F}(x)}.\end{aligned}$$

Now assume that the renewal process is stationary. Then time until the first arrival T_1 has distribution G. Now $\mathbb{P}(R_0 \in [t,t+dt)) = \mathbb{P}(T_1 \in [t,t+dt)) = G(dt)$. Also $\mathbb{P}(A_0 \in [t,t+dt))$ is the same probability. By stationarity, the $R_t \sim R_0$ and $A_t \sim A_0$ for all t so we conclude that for a stationary renewal process, the distribution of R_t and A_t coincide and equals G. But then

$$\begin{aligned}\mathbb{P}(A_t > x, R_t > y) &= \int_x^\infty \mathbb{P}(R_t > y|\, A_t = x) G(dx)\\ &= \int_x^\infty \frac{\bar{F}(z+y)}{\bar{F}(z)} G(dz)\\ &= \int_x^\infty \frac{\bar{F}(z+y)}{\bar{F}(z)} \frac{1}{\mu} \bar{F}(z)dz\\ &= \frac{1}{\mu}\int_x^\infty \bar{F}(z+y)dz\\ &= \frac{1}{\mu}\int_{x+y}^\infty \bar{F}(z)dz\\ &= \bar{G}(x+y).\end{aligned}$$

Then we differentiate with respect to x and y to obtain the joint density of (A_t, R_t) given by

$$f_{(A_t,R_t)}(x,y) = \frac{f(x+y)}{\mu}.$$

We have proved the following theorem.

Theorem 4.6. *In a stationary renewal process, the joint distribution of the age and residual life–time (A_t,R_t) is given by the density*

$$f_{(A_t,R_t)}(x,y) = \frac{f(x+y)}{\mu}.$$

In particular, we conclude that they have the same marginal distribution $A_t \sim R_t \sim G$.

Finally we find the distribution of the spread S_t. By convolution, we find the density $f_{S_t}(x)$ as

$$\begin{aligned} f_{S_t}(x) &= \int_0^x f_{(A_t,S_t)}(s,x-s)ds \\ &= \int_0^x \frac{f(s+(x-s)}{\mu}ds \\ &= \frac{xf(x)}{\mu}. \end{aligned}$$

This is the so–called first order moment distribution. We have proved

Theorem 4.7. *Consider a stationary renewal process. Then the distribution of the spread S_t has a first order moment distribution with a density given by*

$$f_{S_t}(x) = \frac{xf(x)}{\mu}.$$

That S_t does not have density f in spite of the assumption that all inter–arrival times will have a distribution with this density, is often referred to as the "inspection paradox". Since the process is stationary, then any time t may be considered a completely random time, so when we pick an interval according to a random time, this interval will not have density f but $xf(x)/\mu$ instead, the latter being more heavy tailed. The first moment distribution is also refereed to as the "length biased sampling" density, and it is in term of this expression that an explication for the inspection paradox may be found. Indeed, it is plausible that when sampling an interval of inter–arrival at random, it is more likely that a large interval is picked than a small. Also it seems reasonable that the longer the interval, the larger the probability, and in fact the first moment distribution sort of dictates that the probability increases linearly with the length of the interval. So if $xf(x)/\mu dx$ is the probability of an inter–arrival of size in $[x,x+dx)$ at the picked time t, then $(2x)f(2x)dx/\mu$ is the probability of seeing the double sized interval. Notice that we are not doubling the probabilities since $f(x)$ and $f(2x)$ are distinct.

4.1.4 Terminating renewal processes

Frequently we are in situations where the renewal process only exhibits a finite number of renewals. We shall consider such processes as terminating. Formally, this can occur when there exists an arrival epoch such that $S_n = \infty$ because then we will not see any further arrivals. $S_n = \infty$ happens if $T_n = \infty$ which in turn may happen if the inter-arrival distribution is defective, i.e.$||F|| = \lim_{t\to\infty} F(t) < 1$, in which case there must exists an atom at infinity in order for F to be a proper distribution.

Definition 4.3. A renewal process is terminal if its inter-arrival distributions F is defective, i.e. $||F|| < 1$. The life-time M of a terminating renewal process is defined as

$$M = \sup\{S_n |\ S_n < \infty\}.$$

Theorem 4.8. *For a zero–delayed and terminating renewal process with renewal function $U(t)$ we have that*

$$\mathbb{P}(M \leq x) = (1 - ||F||)U(x).$$

Proof. Define $Z(x) = \mathbb{P}(M \leq x)$ for $x \geq 0$. Condition on T_1. If $T_1 = +\infty$, which happens with probability $1 - ||F||$, then $M = 0$ and hence $\mathbb{P}(M \leq x) = 1$ for all $x \geq 0$. If $T_1 < \infty$ then conditioning on $T_1 = s < \infty$, $M \leq x$ if and only if the renewal process which initiates at time s has a life–time less or eual to $x - s$ which happens with probability $Z(x-s)$. Hence

$$Z(x) = (1 - ||F||) \cdot 1 + \int_0^x Z(x-s)dF(s)$$

This is a renewal equation with $z(x) = 1 - ||F||$ constant and hence bounded. By Teorem 4.2, and noting that the theory of renewal equations is also valid for defective distributions, we conclude that $Z(x)$ is given by

$$\begin{aligned} Z(x) &= U * (1 - ||F||)(x) \\ &= (1 - ||F||)U(x). \end{aligned}$$

since $1 - ||F||$ is constant. □

4.2 Matrix–exponential renewal theory

4.2.1 Phase–type renewal theory

Now consider a renewal process with inter–arrival times which are i.i.d. PH$(\boldsymbol{\pi}, \boldsymbol{T})$ phase–type distributed.

Theorem 4.9. *For a zero–delayed phase–type renewal process with inter–arrival times distributed as PH($\boldsymbol{\pi},\boldsymbol{T}$), the renewal density u is given by*

$$u(x) = \boldsymbol{\pi} e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x}\boldsymbol{t}.$$

Proof. Let the times between arrivals be $Y_1, Y_2, \ldots$ i.i.d. $\sim \mathrm{PH}(\boldsymbol{\pi},\boldsymbol{T})$. By concatenating the underlying Markov jump processes of the Y_i's up to the time of absorption we obtain a Markov jump process with intensity matrix $\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}$. To see this, consider the concatenated process, which is obviously a Markov jump process and let $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{ij}$ denote its intensity matrix.

Consider first $i \neq j$. Then $\lambda_{ij}dt$ the probability of a jump from i to j during $[t,t+dt)$ in the concatenated process. A jump from i to j can either occur as a result of a transition from i to j by one of the underlying Markov jump processes or by one of the underlying Markov jump processes exiting from state i and the following Markov jump process initiating at the time of exit starts in state j. The probability of the former is $t_{ij}dt$ while the probability of the latter is $t_i dt\ \pi_j$. Since the events are mutually exclusive we conclude that

$$\lambda_{ij}dt = t_{ij}dt + t_i\pi_j dt = (t_{ij} + t_i\pi_j)dt.$$

For $i = j$, $1+\lambda_{ii}dt$ is the probability of no jump of the concatenated process in $[t,t+dt)$. This happens if either there are no jumps in an underlying process or if an underlying process terminates and the next process in line starts in state i. The probability of the former is $1+t_{ii}dt$ while the probability of the latter is $t_i dt\pi_i$. Hence

$$1+\lambda_{ii}dt = 1+t_{ii}dt + t_i dt\pi_i,$$

resulting in $\lambda_{ii} = t_{ii} + t_i\pi_i$. Hence $\boldsymbol{\Lambda} = \boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}$ and the transition matrix is hence $\boldsymbol{P}^x = \exp((\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x)$.

Now $u(x)dx$ is the probability of a renewal in $[x,x+dx)$, which in turn is equivalent with one of the underlying Markov jump processes exiting in $[x,x+dx)$ to the absorbing state. Conditioning on the current state of the concatenated process at time x we get that

$$\begin{aligned} u(x)dx &= \sum_{i,j=1}^{p} \pi_i p_{ij}^x t_j dx \\ &= \sum_{i,j=1}^{p} \pi_i \left(e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x}\right)_{ij} t_j dx \\ &= \boldsymbol{\pi} e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x}\boldsymbol{t}dx, \end{aligned}$$

from which the result follows. □

Theorem 4.10. *Consider a delayed renewal process with initial inter–arrival time $Y_1 \sim PH(\boldsymbol{\alpha},\boldsymbol{T})$ and subsequent inter-arrival times $Y_2, Y_3, \ldots \sim PH(\boldsymbol{\pi},\boldsymbol{T})$. Then the renewal density u is given by*

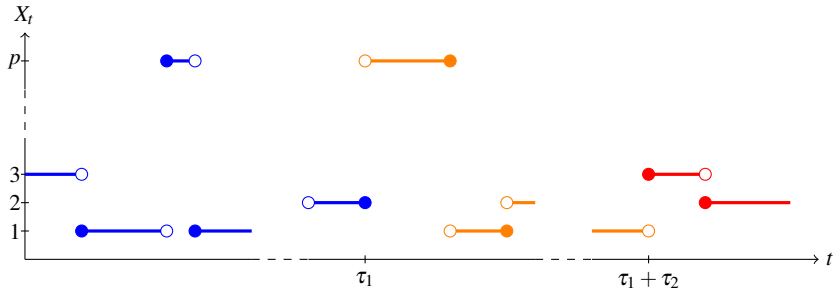

Fig. 4.2 Concatenating the Markov jump processes which generate the Phase–type distributions we obtain a new Markov jump process the state–space of which coincides with the transient states of the original Markov jump process.

$$u(x) = \boldsymbol{\alpha} e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x}\boldsymbol{t}.$$

Proof. Follows by an argument as for the zero–delayed case, the only difference being the initial distribution of the concatinated process is $\boldsymbol{\alpha}$ instead of $\boldsymbol{\pi}$. □

The next theorem proves that the time lapse from any time point x until the next arrival in a phase–type renewal process is again of phase–type.

Theorem 4.11. *The residual life–time $R_x = S_{N(x)} - x$ in a (possibly delayed) phase–type renewal process has a phase–type distribution.*

In the zero–delayed case where $Y_1, Y_2, \ldots$ are i.i.d. $\sim PH(\boldsymbol{\pi}, \boldsymbol{T})$,

$$R_x \sim PH\left(\boldsymbol{\pi}\exp\left((\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x\right), \boldsymbol{T}\right)$$

while

$$R_x \sim PH\left(\boldsymbol{\alpha}\exp\left((\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})x\right), \boldsymbol{T}\right)$$

in the delayed case where the delay distribution of Y_1 is $PH(\boldsymbol{\alpha}, \boldsymbol{T})$ and $Y_2, Y_3, \ldots$ are i.i.d. $\sim PH(\boldsymbol{\pi}, \boldsymbol{T})$.

Proof. Simply notice that the distribution of the concatinated process at time x is given by $\boldsymbol{\pi} exp\left((\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})\right)$ ($\boldsymbol{\alpha}\exp\left((\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})\right)$ respectively). This serves as the initial distribution for a phase–type process initiating at time x. □

Theorem 4.12. *A stationary renewal process, where $Y_2, Y_3, \ldots$ are i.i.d. $\sim PH(\boldsymbol{\pi}, \boldsymbol{T})$, is a delayed renewal process with an initial inter–arrival time given by*

$$Y_1 \sim PH\left(\frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}, \boldsymbol{T}\right).$$

Proof. From Theorem 4.5 we have that Y_1 has density f_1 given by

$$f_1(y) = \frac{1}{\mu}\left(1 - F(y)\right)$$

$$= \frac{1}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}} \left(1 - (1 - \boldsymbol{\pi} e^{\boldsymbol{T}y}\boldsymbol{e})\right)$$
$$= \frac{1}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}} \boldsymbol{\pi}(-\boldsymbol{T}^{-1}) e^{\boldsymbol{T}y}(-\boldsymbol{T}\boldsymbol{e})$$
$$= \left(\frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\right) e^{\boldsymbol{T}y}\boldsymbol{t}.$$

□

Finally we shall need a result for defective renewal processes. We recall that a distribution G is called defective if $||G|| = \lim_{x\to+\infty} G(x) < 1$. If the missing mass $1 - ||G||$ is replaced at zero such that the resulting distribution is hence no longer defect, it is called a zero–modified distribution.

Theorem 4.13. *The life–time M of the defective phase–type renewal process with inter–arrivals $Y_1, Y_2, ..$ i.i.d. $\sim PH\boldsymbol{\pi}, \boldsymbol{T})$ (i.e. $\boldsymbol{\pi}\boldsymbol{e} < 1$) is a zero–modified phase–type distribution with atom at zero of size $1 - \boldsymbol{\pi}\boldsymbol{e}$ and absolute continuous part being defective phase–type with representation $(\boldsymbol{\pi}, \boldsymbol{T} + \boldsymbol{t}\boldsymbol{\pi})$.*

Proof. The life–time is zero with probability $1 - \boldsymbol{\alpha}\boldsymbol{e}$ and if not the result follows by concatenating Markov jump processes underlying inter–arrivals. □

4.2.2 General Matrix–exponential renewal theory: analytic approach

For matrix–exponential distributions we lack the probabilistic interpretation of the phase–type distributions which allowed to concatinate the sample paths of the underlying Markov jump processes up to the time of absorption. In this section we proceed with an entirely analytical approach.

The following well–known lemma will play an important role.

Lemma 4.1. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ be matrices where $\boldsymbol{A}$ and $\boldsymbol{D}$ are non–singular square matrices. Then*

$$(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$$

whenever the terms make sense.

Proof. The proof is a simple application of the technique Gaussian elimination when finding the inverse to the block–matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}.$$

Consider the inversion $\boldsymbol{\Lambda}$ by Gaussian elimination starting by eliminating $\boldsymbol{C}$ (subtract from the second row $\boldsymbol{C}\boldsymbol{A}^{-1}$ times the first row),

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} \qquad \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$$
$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{I} \end{pmatrix}$$
$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}+\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{I} \end{pmatrix}$$
$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \qquad \begin{pmatrix} \boldsymbol{A}^{-1}+\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & (\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{pmatrix}.$$

By symmetry, if we start by eliminating $\boldsymbol{B}$ first instead of $\boldsymbol{C}$, we end up with $(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$ as the upper left entrance to the inverse of the block matrix. Hence the proposed formula follows. □

If we let $\boldsymbol{D}=1$, $\boldsymbol{A}=x\boldsymbol{I}-\boldsymbol{S}$, $\boldsymbol{B}=\boldsymbol{s}$ and $\boldsymbol{C}=\boldsymbol{\alpha}$, then Lemma 4.1 results in

$$\begin{aligned}(x\boldsymbol{I}-\boldsymbol{S}-\boldsymbol{s}\boldsymbol{\alpha})^{-1} &= (x\boldsymbol{I}-\boldsymbol{S})^{-1} \\ &\quad +(x\boldsymbol{I}-\boldsymbol{S})^{-1}(1-\boldsymbol{\alpha}(x\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s})^{-1}\boldsymbol{\alpha}(x\boldsymbol{I}-\boldsymbol{S})^{-1}.\end{aligned}$$

Then multiplying with a row vector $\boldsymbol{\beta}$ from the left and the column vector $\boldsymbol{s}$ on the right, we have proved that

Lemma 4.2.

$$\boldsymbol{\beta}(x\boldsymbol{I}-\boldsymbol{S}-\boldsymbol{s}\boldsymbol{\alpha})^{-1}\boldsymbol{s}=\frac{\boldsymbol{\beta}(x\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s}}{1-\boldsymbol{\alpha}(x\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s}}.$$

Now consider a delayed renewal process where $Y_1 \sim \mathrm{ME}(\boldsymbol{\beta},\boldsymbol{S},\boldsymbol{s})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim \mathrm{ME}(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$. Then we renewal density $u(x)$ is given by

$$u(x)=f_1(x)+\sum_{n=1}^{\infty} f_1 * f^{*(n)}(x),$$

where f_1 denotes the density corresponding to Y_1 and f the common density of the Y_i's, $i=2,3,\ldots$. Taking Laplace transform we get that

$$\begin{aligned}\hat{u}(x) &= \hat{f}_1(x)+\sum_{n=1}^{\infty}\hat{f}_1(x)\hat{f}(x)^n \\ &= \frac{\hat{f}_1(x)}{1-\hat{f}(x)} \\ &= \frac{\boldsymbol{\beta}(x\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s}}{1-\boldsymbol{\alpha}(x\boldsymbol{I}-\boldsymbol{S})^{-1}\boldsymbol{s}} \\ &= \beta(x\boldsymbol{I}-\boldsymbol{S}-\boldsymbol{\alpha}\boldsymbol{s})^{-1}\boldsymbol{s},\end{aligned}$$

from which is follows that

Theorem 4.14. *A delayed matrix–exponential renewal process, where $Y_1 \sim ME(\boldsymbol{\beta},\boldsymbol{S},\boldsymbol{s})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim ME(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ has a renewal density u given by*

$$u(x) = \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{s}.$$

It follows trivially that

Corollary 4.1. *A zero–delayed matrix–exponential renewal process, where $Y_1, Y_2, \ldots$ i.i.d. $\sim ME(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ has a renewal density u given by*

$$u(x) = \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{s}.$$

Next we consider the distribution of the residual life–time.

Theorem 4.15. *Consider a delayed matrix–exponential renewal process with inter–arrivals $Y_1 \sim ME(\boldsymbol{\beta}, \boldsymbol{S}, \boldsymbol{s})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim ME(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Let $R_x = S_{N(x)} - x$ denote the residual life–time. Then $R_x \sim ME(\boldsymbol{\beta}\exp((\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x), \boldsymbol{S})$*

Proof. Let $f_{R_x}(z)$ be the density of R_x. Then $f_{R_x}(z)dz$ is the probability that $R_x \in (z, z+dz]$. On the event that there has not been any arrivals in $[0,x]$, the probability that $R_x \in (z, z+dz]$ is that the first arrival must fall in that interval, the probability of which is $\mathbb{P}(Y_1 \in (z+x, z+x+dz]) = f_1(z+x)dz$. On the event that there has been at least one arrival prior to x, the probability that the last event prior to x happened at time y and that the next is at time $z+x$ is $u(y)f(x+z-y)dz$. Summing up,

$$f_{R_x}(z) = f_1(x+z) + \int_0^x u(y)f(z+x-y)dy.$$

But now

$$f_1(z) = \boldsymbol{\beta} e^{\boldsymbol{S}z}\boldsymbol{s},\ u(y) = \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{\alpha}\boldsymbol{s})y}\boldsymbol{s},\ f(y) = \boldsymbol{\alpha} e^{\boldsymbol{S}y}\boldsymbol{s}.$$

Now

$$\frac{d}{dz}e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x} = e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})$$

so

$$e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{s}\boldsymbol{\alpha} = \frac{d}{dz}e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x} - e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{S}.$$

But then

$$\begin{aligned}
\int_0^x u(y)f(z+x-y)dy &= \int_0^x \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{s}\boldsymbol{\alpha} e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy \\
&= \int_0^x \boldsymbol{\beta}\left(\frac{d}{dz}e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y} - e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}\boldsymbol{S}\right)e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy \\
&= \int_0^x \boldsymbol{\beta}\frac{d}{dz}e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy - \int_0^x \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}\boldsymbol{S}e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy \\
&= \left[\boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}\right]_0^x \\
&\quad + \int_0^x \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}\boldsymbol{S}e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy - \int_0^x \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})y}\boldsymbol{S}e^{\boldsymbol{S}(z+x-y)}\boldsymbol{s}dy \\
&= \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{\alpha}\boldsymbol{s})x}e^{\boldsymbol{S}z}\boldsymbol{s} - f_1(z+x),
\end{aligned}$$

so that

$$f_{R_x}(z) = f_1(z+x) + \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{\alpha}\boldsymbol{s})x} e^{\boldsymbol{S}z}\boldsymbol{s} - f_1(z+x) = \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{\alpha}\boldsymbol{s})x} e^{\boldsymbol{S}z}\boldsymbol{s}.$$

□

Theorem 4.16. *A delayed matrix–exponential renewal process with inter–arrivals $Y_1 \sim ME(\boldsymbol{\beta},\boldsymbol{S},\boldsymbol{s})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim ME(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ is stationary if and only if*

$$\boldsymbol{\beta} = \frac{\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}}{\boldsymbol{\alpha}\boldsymbol{S}^{-2}\boldsymbol{s}}.$$

Proof. The proof is similar to the phase–type case, cfr. Theorem 4.12 . □

Theorem 4.17. *Consider a terminating renewal process with (obviously defective) inter–arrival distribution being matrix–exponential with representation $(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$. Then its life–time M is a zero–modified matrix–exponential distribution with an atom at zero of size $1+\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s}$ and an absolute continuous given by the representation $(\boldsymbol{\alpha},\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha},(1+\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s})\boldsymbol{s})$.*

Proof. Let F_M denote the distribution function of M. Then

$$\begin{aligned} 1-F_M(x) &= \mathbb{P}(M>x) \\ &= \mathbb{P}(R_x<\infty) \\ &= \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x} e^{\boldsymbol{S}y}\boldsymbol{s}\,dy \\ &= -\boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}\boldsymbol{S}^{-1}\boldsymbol{s}. \end{aligned}$$

Hence the corresponding density f_M is given by

$$\begin{aligned} f_M(x) &= -\frac{d}{dx}(1-F_M(x)) \\ &= -\boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})\boldsymbol{S}^{-1}\boldsymbol{s} \\ &= \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x}(1+\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s})\boldsymbol{s}. \end{aligned}$$

The mass is $1-F_M(0) = -\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s}$ and hence the defect is $1+\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{s}$ which is the mass which is placed at zero. □

4.2.3 Matrix–exponential renewal theory via flows

Consider a delayed renewal process with a representation for which $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$, where $Y_1 \sim \mathrm{ME}_p(\boldsymbol{\beta},\boldsymbol{S})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim \mathrm{ME}(\boldsymbol{\alpha},\boldsymbol{S})$. Then we consider two valid flows generated by $(\boldsymbol{\beta},\boldsymbol{S})$ and $(\boldsymbol{\alpha},\boldsymbol{S})$ respectively. Let $U_1, U_2, U_3, \ldots$ be i.i.d uniformly distributed over $[0,1]$ and let $Y_1 = T_{\boldsymbol{\beta}}(U_1)$ and $Y_i = T_{\boldsymbol{\alpha}}(U_i)$ respectively where $T_{\boldsymbol{\beta}}(x)$ is

the time until the liquid level of container $p+1$ reaches level x for the flow $(\boldsymbol{\beta}, \boldsymbol{S})$ and similarly for $T_{\boldsymbol{\alpha}}$.

We now consider the above delayed renewal process in terms of the concatenating the flows such that when there is an arrival, a new flow is initiated. Let $Y_i(t)$ denote the content of container i by time t, which is now a random variable since we do not know how many arrivals (refills) there has been and when, and we let $z_i(t) = \mathbb{E}(Y_i(t))$. Let $\boldsymbol{Y}(t) = (Y_1(t), ..., Y_p(t))$ and $\boldsymbol{z}(t) = (z_1(t), ..., z_p(t))$.

Theorem 4.18. *The renewal density $u(x)$ is given by*

$$u(x) = \boldsymbol{z}(t)\boldsymbol{s}.$$

Proof. We have that

$$\begin{aligned} u(x)dx &= \mathbb{P}(\text{there is an arrival in } [t, t+dt)) \\ &= \mathbb{E}\left(\mathbb{P}(\text{there is an arrival in } [t, t+dt))|\boldsymbol{Y}(t)\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{p} Y_j(t)s_j dt\right) \\ &= \sum_{j=1}^{p} z_j(t)s_j dt \\ &= \boldsymbol{z}(t)\boldsymbol{s}dt. \end{aligned}$$

□

Theorem 4.19.

$$\boldsymbol{z}(t) = \boldsymbol{\beta} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})t}.$$

Proof. A standard renewal argument is applied. If there are no arrivals in $[0,t]$, then the expected amount of liquid at time t in the containers is given by $\boldsymbol{\beta}\exp(\boldsymbol{S}t)$. If there is a first arrival at time $x \in [0,t]$, which happens with probability $\boldsymbol{\beta}\exp(\boldsymbol{S}x)\boldsymbol{s}dx$, then the expected value of the contents at time t is $\boldsymbol{z}_\alpha(t-x)$, where $\boldsymbol{z}_\alpha(t)$ is the expected amount of liquid in the p containers at time t in the corresponding non–delayed renewal process, i.e. a renewal process with intial distribution $\boldsymbol{\alpha}$ instead of $\boldsymbol{\beta}$. Thus we have that

$$\boldsymbol{z}(t) = \boldsymbol{\beta} e^{\boldsymbol{S}t} + \int_0^t \boldsymbol{\beta} e^{\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{z}_\alpha(t-x)dx.$$

A change of variable yields

$$\boldsymbol{z}(t) = \boldsymbol{\beta} e^{\boldsymbol{S}t} + \int_0^t \boldsymbol{\beta} e^{\boldsymbol{S}(t-x)}\boldsymbol{s}\boldsymbol{z}_\alpha(x)dx.$$

Consider the matrix $\boldsymbol{V}(t)$ defined by

$$\boldsymbol{V}(t) = e^{\boldsymbol{S}t} + \int_0^t e^{\boldsymbol{S}(t-x)}\boldsymbol{s}\boldsymbol{z}_\alpha(x)dx.$$

Then $\boldsymbol{z}(t) = \boldsymbol{\beta}\boldsymbol{V}(t)$. Pre–multiply $\boldsymbol{V}(t)$ with $e^{-\boldsymbol{S}t}$,

$$e^{-\boldsymbol{S}t}\boldsymbol{V}(t) = \boldsymbol{I} + \int_0^t e^{-\boldsymbol{S}x}\boldsymbol{s}\boldsymbol{z}_\alpha(x)dx$$

and differentiate to obtain

$$-\boldsymbol{S}e^{-\boldsymbol{S}t}\boldsymbol{V}(t) + e^{\boldsymbol{S}t}\boldsymbol{s}\boldsymbol{V}'(t) = e^{-\boldsymbol{S}t}\boldsymbol{s}\boldsymbol{z}_\alpha(t)$$

from which

$$\boldsymbol{V}'(t) = \boldsymbol{S}\boldsymbol{V}(t) + \boldsymbol{s}\boldsymbol{z}_\alpha(t).$$

Now $\boldsymbol{z}_\alpha(t) = \boldsymbol{\alpha}\boldsymbol{V}(t)$ so

$$\boldsymbol{V}'(t) = \boldsymbol{S}\boldsymbol{V}(t) + \boldsymbol{s}\boldsymbol{\alpha}\boldsymbol{V}(t) = (\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha})\boldsymbol{V}(t),$$

and with $\boldsymbol{V}(0) = \boldsymbol{I}$ we have that

$$\boldsymbol{V}(t) = \exp\left((\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha})t\right),$$

so

$$\boldsymbol{z}(t) = \boldsymbol{\beta}\boldsymbol{V}(t) = \boldsymbol{\beta}e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})t}.$$

□

Theorem 4.20. *The residual life–time R_t for the delayed renewal process has a Matrix–exponential ditribution with representation*

$$ME_p(\boldsymbol{\beta}\exp\left((\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha})t\right), \boldsymbol{S}).$$

Proof. The content of the containers by time t is $\boldsymbol{Y}(t)$. If the next arrival after time t happens by time $t+u$, then there are no arrivals in the time interval from t to $t+u$ and by Theorem 1.14 the content by time $t+u$ will be $\boldsymbol{Y}(t)\exp(\boldsymbol{S}u)$. Then by Theorem 1.15 the probability of an arrival at $[t+u, t+u+du)$ is $\boldsymbol{Y}(t)\exp(\boldsymbol{S}u)\boldsymbol{s}du$. Hence

$$\begin{aligned}
\mathbb{P}(R_t \in [u, u+du)) &= \mathbb{E}\left(\mathbb{P}\left(R_t \in [u, u+du) | Y(t)\right)\right) \\
&= \mathbb{E}\left(\boldsymbol{Y}(t)e^{\boldsymbol{S}u}\boldsymbol{s}du\right) \\
&= \mathbb{E}\left(\boldsymbol{Y}(t)\right)e^{\boldsymbol{S}u}\boldsymbol{s}du \\
&= \boldsymbol{z}(t)e^{\boldsymbol{S}u}\boldsymbol{s}du,
\end{aligned}$$

and the result the follows from Theorem 4.19. □

Theorem 4.21. *A delayed matrix–exponential renewal process with inter–arrivals $Y_1 \sim ME_p(\boldsymbol{\beta}, \boldsymbol{S})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim ME_p(\boldsymbol{\alpha}, \boldsymbol{S})$ is stationary if and only if*

$$\boldsymbol{\beta} = \frac{\boldsymbol{\alpha}(-\boldsymbol{S})^{-1}}{\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{e}}.$$

Proof. The proof is similar to the phase–type case, cfr. Theorem 4.12 . □

Exercises

Exercise 4.1. Consider a delayed matrix–exponential renewal process with inter–arrivals $Y_1 \sim \mathrm{ME}_p(\boldsymbol{\beta}, \boldsymbol{S})$ and $Y_2, Y_3, \ldots$ i.i.d. $\sim \mathrm{ME}_p(\boldsymbol{\alpha}, \boldsymbol{S})$. Let A be the event that there are no arrivals in $[t, t+dt)$. Prove that

$$\mathbb{E}(Y_i(t+dt)1_A) = z_i(t) + \sum_{j \neq i} z_j(t) S_{ij} dt - \sum_{j \neq i} Z_i(t) S_{ij} dt - z_i(t) s_i dt.$$

Conclude that

$$z_i(t+dt) = z_i(t) + \sum_j z_j(t) S_{ji} dt + \sum_i z_j(t) s_j \alpha_i dt$$

and since $\boldsymbol{z}(0) = \boldsymbol{\beta}$ that

$$\boldsymbol{z}(t) = \boldsymbol{\beta} e^{(\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha})t}.$$

Chapter 5
Random Walks

5.1 Random Walks and ladder methods

By a random walk $\{S_n\}_{n\in\mathbb{N}}$ on $\mathbb{R}$ we understand a stochastic process in discrete time for which $S_0 = 0$, and for $n \geq 1$

$$S_n = X_1 + X_2 + ... + X_n,$$

where $X_1, X_2, ...$ are i.i.d. $\sim F$ for some distribution F on $\mathbb{R}$. If F is concentrated on $[0,\infty)$, then S_n is simply the n'th arrival epoch in a renewal process which we have already considered in a previous chapter. We shall henceforth assume that F does not have a support contained in $[0,\infty)$.

Some important instruments in the analysis of random walks are the so–called associated ladder processes. Define

$$\tau_+ = \inf\{k \in \mathbb{N} \mid S_k > 0\} \tag{5.1}$$
$$\tau_- = \inf\{k \in \mathbb{N} \mid S_k \leq 0\} \tag{5.2}$$

Let $\tau_+(1) = \tau_+$, $\tau_-(1) = \tau_-$ and

$$\tau_+(n+1) = \inf\{k > \tau_+(n) \mid S_k > S_{\tau_+(n)}\} \tag{5.3}$$
$$\tau_-(n+1) = \inf\{k > \tau_-(n) \mid S_k \leq S_{\tau_-(n)}\} \tag{5.4}$$

Definition 5.1. The points $(n, S_{\tau_+(n)})$ are called the (strictly) ascending ladder points of the random walk, while $(n, S_{\tau_-(n)})$ are referred to as the (weakly) descending ladder points. We let G_+ denote the distribution of S_{τ_+} and G_- the distribution of S_{τ_-}. The random variables $S_{\tau_+(n)}$ and $S_{\tau_-(n)}$ are referred to as ladder heights while their differences $S_{\tau_+(n+1)} - S_{\tau_+(n)}$ and $S_{\tau_-(n+1)} - S_{\tau_-(n)}$ as ladder steps.

Since G_+ and G_- is only well defined if $\tau_+ < \infty$ and $\tau_- < \infty$, we have that

$$G_+(x) = \mathbb{P}\left(S_{\tau_+} \leq x, \tau_+ < \infty\right), \;\; x \geq 0$$
$$G_-(x) = \mathbb{P}\left(S_{\tau_-} \leq x, \tau_- < \infty\right), \;\; x \leq 0.$$

Recall that the total mass of a distribution G is defined as $||G|| = \lim_{x\to\infty} G(x)$. Then

$$\begin{aligned} ||G_+|| &= \lim_{x\to\infty} \mathbb{P}\left(S_{\tau_+} \le x, \tau_+ < \infty\right) \\ &= \mathbb{P}(\tau_+ < \infty) \\ ||G_-|| &= \lim_{x\to\infty} \mathbb{P}\left(S_{\tau_-} \le x, \tau_- \infty\right) \\ &= G_-(0) \\ &= \mathbb{P}(\tau_- < \infty). \end{aligned}$$

If $||G_+|| < 1$, then $1 - ||G_+||$ is the probability that $\tau_+ = +\infty$ and that there are no ascending ladder steps. In this case there is a positive probability that the random walk will remain negative always. Similar considerations apply to $||G_-|| < 1$.

We notice that the definition of ascending and descending ladder heights is not symmetric in the sense that for the ascending ladder epochs we chose the condition that a new maximum must be strictly larger than previous maxima, while for the descending ladder process only matching previous maxima heights is necessary. The difference plays no role if the F is absolutely continuous, but for the general theory we must attend the problem concerning what happens if $S_{\tau_-} = 0$ with a positive probability. To this end we introduce the weakly ascending ladder epochs by

$$\tau_+^w = \inf\{k \in \mathbb{N} \mid S_n \ge 0\} \tag{5.5}$$

and the strictly descending ladder epochs by

$$\tau_-^w = \inf\{k \in \mathbb{N} \mid S_n < 0\}. \tag{5.6}$$

The corresponding ladder heights are $S_{\tau_+^w}$ and $S_{\tau_-^s}$, and with an obvious notation we let their corresponding distributions be G_+^w and G_-^s respectively. Subsequent ladder epochs are defined in a similar fashion.

5.2 Time reversing or dual processes

Time reversing the random walk is a simple yet powerful method for calculating complicated probabilities. For $S_0 = 0, S_1, ..., S_n$ define the time reversed process S_k^* by

$$S_k^* = S_n - S_{n-k}, \;\; k = 0, 1, ..., n.$$

Then $S_k^* = X_n + X_{n-1} + ... + X_{n-k+1} \sim X_1 + X_2 + ... + X_k = S_k$ and hence $(S_0, S_1, ..., S_n) \sim (S_0^*, S_1^*, ..., S_n^*)$. Furthermore,

$$S_k^* = S_n^* - S_{n-k}^* = S_n - (S_n - S_{n-(n-k)}) = S_k,$$

i.e. the reversed of the reversed process equals the original process. Now

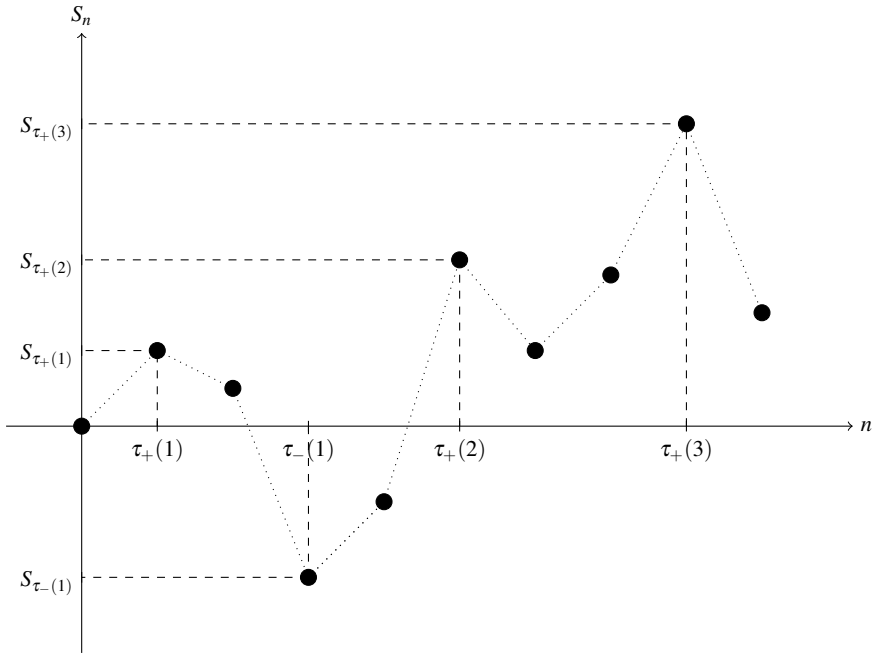

Fig. 5.1 A random walk with ladder epochs and ladder heights.

$$\begin{aligned}\mathbb{P}(S_n \in I, S_n > S_k, k = 0,1,...,n-1) &= \mathbb{P}(S_n \in I, S_n > S_{n-k}, k = 1,...,n)\\ &= \mathbb{P}(S_n^* \in I, S_n - S_{n-k} > 0, k = 1,2,...,n) \quad (S_n^* = S_n)\\ &= \mathbb{P}(S_n^* \in I, S_k^* > 0, k = 1,...,n) \quad (S_k^* = S_n - S_{n-k})\\ &= \mathbb{P}(S_n \in I, S_1 > 0, S_2 > 0, ..., S_n > 0).\end{aligned}$$

The left hand side is the probability that (n, S_n) is a strict ladder point which belongs to the set I. The final expression is the probability that the random walk at time n belongs to I and that the random walk previously have remained positive. The result is also true for $n = 0$ considering there are no previous arrivals and hence both sides coincide by definition. We have proved the simple but useful result:

Theorem 5.1. *For $n = 0,1,...$ we have that*

$$\mathbb{P}(S_n \in I, S_n > S_k, k = 0,1,...,n-1) = \mathbb{P}(S_n \in I, S_1 > 0, S_2 > 0, ..., S_n > 0). \quad (5.7)$$

Next consider (the possibly defective) renewal process with inter–arrival distribution G_+. This process is visualized in Figure 5.1 on the ordinate axis, where the ladder steps $S_{\tau_+}, S_{\tau_+(2)} - S_{\tau_+}, ...$ are the inter–arrival times of the renewal process which ascends the axis as a result of projecting the overshoots of the ladder processes.

Now

$$G_+(I) = \mathbb{P}(S_{\tau_+} \in I) = \int_I dG_+(x),\ I \subset \mathbb{R}_+ \text{ measurable},$$

and let U_+ denote the corresponding renewal measure

$$U_+ = \sum_{k=0}^{\infty} G_+^{*k}.$$

Summing the left hand side of (5.7) over $n = 0, 1, ...$ we get

$$\begin{aligned}\sum_{n=0}^{\infty} \mathbb{P}(S_n \in I, S_n > S_k, k = 0, 1, ..., n-1) &= \mathbb{E}\left(\sum_{n=0}^{\infty} 1\{S_n \in I, S_n > S_k, k = 0, 1, ..., n\}\right) \\ &= \text{expected number of ladder height falling in } I \\ &= U_+(I).\end{aligned}$$

On the other hand, the sum of the right hand side of (5.7) over $n = 0, 1, ...$ amounts to

$$\sum_{n=0}^{\infty} \mathbb{P}(S_n \in I, S_1 > 0, S_2 > 0, ..., S_n > 0) = \mathbb{E}\left(\sum_{n=0}^{\infty} 1\{S_n \in I, S_k > 0, k = 1, ..., n\}\right),$$

i.e. the expected amount of points from the random walk which falls in I and such that all previous points are positive. This result is referred to as the duality theorem for random walks which we collect in the following result.

Theorem 5.2. *(duality) The renewal measure $U_+(I)$, $I \subseteq \mathbb{R}_+$ measurable, for the strictly ascending ladder height distribution may be interpreted as (a) the expected number of ladder heights falling in I and (b) the expected number of points from the random walk falling in I and such that all previous points are positive.*

The difference between weak and strict ladder heights is the content of the following theorem.

Theorem 5.3. *If $\rho = \mathbb{P}(S_{\tau_+} = 0)$, then $\rho < 1$ and*

$$G_+^w(x) = \rho\,\delta_0 + (1-\rho)G_+(x) \tag{5.8}$$

where δ_0 is the Dirac delta function (i.e. the distribution degenerate at 0).

Proof. An atom at zero will take place if there no further positive ladder heights after zero or if a ladder height hits zero exactly, i.e.

$$\rho = \sum_{n=1}^{\infty} \mathbb{P}(S_k < 0, k = 1, ..., n-1, S_n = 0).$$

If $X_1 > 0$, then it is impossible that $S_n = 0$ with all previous negative and since $\mathbb{P}(X_1 > 0) > 0$ we have that $\rho < 1$. The probability that $S_{\tau_+} > 0$ is $1 - \rho$, in which case the weak and strict ladder height distributions coincide. Hence the result. □

Let us consider the negative ladder heights for a moment. Let $\xi = \mathbb{P}(S_{\tau_-} = 0)$. Then

$$\xi = \sum_{n=1}^{\infty} \mathbb{P}(S_k > 0, k = 1, ..., n-1, S_n = 0).$$

But

$$\begin{aligned}\mathbb{P}(S_k > 0, k = 1, ..., n-1, S_n = 0) &= \mathbb{P}(S_k^* > 0, k = 1, ..., n-1, S_n^* = 0) \\ &= \mathbb{P}(S_n - S_{n-k} > 0, k = 1, ..., n-1, S_n = 0) \quad (S_n^* = S_n = 0) \\ &= \mathbb{P}(S_k < 0, k = 1, ..., n-1, S_n = 0)\end{aligned}$$

and summing over n we get that $\xi = \rho$. Therefore, by symmetry,

$$G_-(x) = \rho\delta_0 + (1-\rho)G_-^s(x). \tag{5.9}$$

Theorem 5.4. *We can express the renewal measures U_+ and U_- of the corresponding ascending and descending ladder renewal processes as*

$$\begin{aligned}U_+(I) &= \mathbb{E}\left(\sum_{k=0}^{\tau_- - 1} 1\{S_k \in I\}\right) \\ U_-(I) &= \mathbb{E}\left(\sum_{k=0}^{\tau_+ - 1} 1\{S_k \in I\}\right).\end{aligned}$$

Proof. By the duality result,

$$U_+(I) = \mathbb{E}\left(\sum_{n=0}^{\infty} 1\{S_n \in I, S_k > 0, k = 1, ..., n\}\right)$$

and the observation that

$$\{S_1 > 0, ..., S_n > 0\} = \{\tau_- > n\}$$

we get that

$$\begin{aligned}U_+(I) &= \mathbb{E}\left(\sum_{n=0}^{\infty} 1\{S_n \in I, \tau_- > n\}\right) \\ &= \mathbb{E}\left(\sum_{n=0}^{\tau_- - 1} 1\{S_n \in I\}\right).\end{aligned}$$

The result for U_- is similar. □

Remark 5.1. The measures U_- and U_+ are important tools in the analysis of stochastic models and are used in a number of different situations. They are often referred to as the pre–τ_+ and pre–τ_- occupation measures respectively and denoted by R_+

and R_-. The reason for this terminology is due to Theorem 5.4. Hence $R_+(I)$ is the expected amount of time the random walk spends in the set $I \subseteq (-\infty, 0)$ prior to the first up–crossing, and R_- is the expected amount of time the random walk spends in the set $I \subseteq (0, \infty)$ prior to the first down–crossing.

Inserting $I = [0, \infty)$ for U_+ we get that

$$||U_+|| = \mathbb{E}(\tau_-)$$

and since $U_+ = \sum_{n=0}^{\infty} G_+^{*n}$ we get that

$$\mathbb{E}(\tau_-) = ||U_+|| = \frac{1}{1 - ||G_+||}. \tag{5.10}$$

Similarly,

$$\mathbb{E}(\tau_+) = \frac{1}{1 - ||G_-||}. \tag{5.11}$$

By (5.8) we have that the renewal measure which corresponds to the weak ascending ladder process, U_+^w say, is given by

$$\begin{aligned} U_+^w(x) &= \sum_{n=0}^{\infty} (G_+^w)^{*n}(x) \\ &= \sum_{n=0}^{\infty} (\rho \delta_0(x) + (1-\rho) G_+(x))^{*n}. \end{aligned}$$

Taking Laplace transforms,

$$\begin{aligned} \hat{U}_+^w(x) &= \sum_{n=0}^{\infty} (\rho + (1-\rho)\hat{G}_+(x))^n \\ &= \frac{1}{1 - (\rho + (1-\rho)\hat{G}_+(x))} \\ &= \frac{1}{1-\rho} \frac{1}{1 - \hat{G}_+(x)} \\ &= \frac{1}{1-\rho} \hat{U}_+(x). \end{aligned}$$

Inverting the Laplace transform results in

$$U_+ = (1-\rho) U_+^w. \tag{5.12}$$

Hence also

$$U_+^w = \frac{1}{1-\rho} U_+ \tag{5.13}$$

and in particular

$$U_+^w(0) = \frac{1}{1-\rho},$$

which coincides with mean in the distribution which is one plus the geometric distribution $\rho^k(1-\rho)$, $k = 1,2,....$ This geometric distribution is the number of returns to zero before generating the first non–trivial (non–zero) ladder epoch.

The formula (5.7) also holds for $<$ replaced by $\leq$. The proof of this fact is entirely identical. Thus

$$\mathbb{P}(S_n \in I, S_n \geq S_k, k = 0,1,...,n-1) = \mathbb{P}(S_n \in I, S_1 \geq 0, S_2 \geq 0, ..., S_n \geq 0).$$

Let $I = [0,\infty]$ and we obtain that

$$\begin{aligned}\mathbb{P}(S_n \geq S_k, k = 0,1,...,n-1) &= \mathbb{P}(S_1 \geq 0, S_2 \geq 0, ..., S_n \geq 0)\\ &= \mathbb{P}(\tau_- > n).\end{aligned}$$

The left hand side add up to the weak ascending renewal measure $U_+^w(\infty)$ when summing from 1 to ∞. If the descending ladder distribution is defective, there is a positive probability that no more ladder points will occur and $\mathbb{P}(\tau_- = \infty) > 0$. But then $\{\mathbb{P}(\tau_- > n)\}_{n\in\mathbb{N}}$ is bounded away from zero and the series

$$\sum_{n=0}^{\infty} \mathbb{P}(\tau_- > n) = \infty.$$

By (5.12) this implies that $U_+(\infty) = U_+^w(\infty) = \infty$. Hence if one ladder renewal process is terminating, the other is not.

Lemma 5.1. *$\mathbb{E}(\tau_-) < \infty$ if and only if $||G_+|| < \infty$. By symmetry, $\mathbb{E}(\tau_+) < \infty$ if and only if $||G_-|| < \infty$.*

Proof. First recall that

$$||G_+|| = \sum_{n=0}^{\infty} G_+^{*n}(\infty).$$

But

$$G_+^{*2}(\infty) = \lim_{x\to\infty} \int_0^x G_+(x-y)dG_+(y) = ||G_+||^2$$

and, by induction, in general we have that $G_+^{*n}(\infty) = ||G_+||^n$. Now $\mathbb{E}(\tau_-) = U_+^w(\infty)$ and by (5.13),

$$\mathbb{E}(\tau_-) = \frac{1}{1-\rho} U_+(\infty) = \frac{1}{1-\rho}\frac{1}{1-||G_+||},$$

from which the result follows. □

Consider the maximum of the random walk

$$M = \sup_{n\geq 0} S_n.$$

Theorem 5.5. *(Pollaczek–Khintchine) If $||G_+|| < \infty$, then M is finite almost surely with a distribution which consists of an atom at zero of probability $1 - ||G_+||$ and an absolute continuous part given by*

$$\mathbb{P}(M \le x) = (1 - ||G_+||) \sum_{n=1}^{\infty} G_+^n(x).$$

Proof. The number of ascending ladder heights is a random variable which has a geometric distribution with parameter $||G_+||$. The probability that the maximum is attained in exactly n ladder heights and does not exceed x is $(1 - ||G_+||)G_+^n(x)$. Summing over n yields the result. □

Theorem 5.6. *(Classification) A random walk $S_n = X_1 + ... + X_n$ is either oscillating (if $\mathbb{E}(X_i) = 0$), drifting to $+\infty$ (if $\mathbb{E}(X_i) > 0$) or $-\infty$ (if $\mathbb{E}(X_i) < 0$).*

If the random walk is oscillating, then $||G_+|| = ||G_-|| = 1$, $\mathbb{E}(\tau_+) < \infty$ and $\mathbb{E}(\tau_-) < \infty$.

If $S_n \to +\infty$, then $||G_+|| = 1$, $||G_-|| < 1$, $\mathbb{E}(\tau_+) < \infty$ and $\mathbb{E}(\tau_-) = \infty$.

If $S_n \to -\infty$, then $||G_+|| < 1$, $||G_-|| = 1$, $\mathbb{E}(\tau_+) = \infty$ and $\mathbb{E}(\tau_-) < \infty$.

Proof. First consider the case $\mathbb{E}(X_i) = 0$. The random variable $\text{limsup}_{n\to\infty} S_n = \lim_{n\to\infty} \sup_{m\ge n} S_m$ is not affected by interchanging a finite number of X_i's. Hence the Hewitt–Savage 0-1 law states, that the event $\{\text{limsup}_{n\to\infty} S_n \le x\}$ has either probability 0 or 1. Thus there is an $x \in [-\infty, \infty]$ such that $\text{limsup}_{n\to\infty} S_n = x$ a.s. This x cannot be finite. Indeed, if $x < \infty$, then

$$x = \text{limsup}_{n\to\infty} S_n = \text{limsup}_{n\to\infty}(S_n - X_1)$$

implying $x + X_1 \sim x$ which is impossible unless X_1 and hence S_n is degenerated at zero. Thus $\text{limsup}_{n\to\infty} S_n = \infty$ a.s. or $\text{limsup}_{n\to\infty} S_n = -\infty$. But if the latter were the case, then $S_n \to -\infty$ with probability one. But this is impossible since this would also apply to the random walk $-S_n$ by symmetry. Hence $\text{limsup}_{n\to\infty} S_n = \infty$. Similarly, $\text{liminf}_{n\to\infty} S_n = -\infty$. Now

$$\text{limsup}_{n\to\infty} S_n = \text{limsup}_{n\to\infty} S_{\tau_+(n)}$$

so if the former equals $+\infty$ a.s., then with probability one, $S_{\tau_+(n)} \to \infty$ as $n \to \infty$ and hence $||G_+|| = 1$. A similar argument shows that $||G_-|| = 1$. By (5.1) both $\mathbb{E}(\tau_+) < \infty$ and $\mathbb{E}(\tau_-) < \infty$.

If $\mathbb{E}(X_i) > 0$, then by the law of large numbers, $S_n/n \to \mathbb{E}(X_1)$ implying that $S_n \to +\infty$. Then $||G_+|| = 1$, $\mathbb{E}(\tau_+) < \infty$ □

We next prove some important results concerning random walks where X_i is on the form $U_i - T_i$, and where U_i are i.i.d. and matrix–exponentially distributed and T_i are i.i.d. with an arbitrary distribution A, say, which is absolutely continuous and concentrated on $\mathbb{R}_+$.

The case where the U_i's have a phase–type distribution is particularly simple to derive from probabilistic arguments. Assume that

$$U_i \sim \text{PH}(\boldsymbol{\alpha}, \boldsymbol{S}).$$

We are going to calculate the positive ladder height distribution G_+. Since $S_{n+1} = S_n + U_{n+1} - T_{n+1}$, the we may think of generating S_{n+1} from S_n by first subtracting T_{n+1} and then realizing a Markov jump process (of the phase–type distribution U_{n+1}) vertically upwards from the point $S_n - T_n$. At the time of absorption U_{n+1}, we then mark the point $S_{n+1} = S_n - T_{n+1} + U_{n+1}$. See Figure 5.2 for details. At the time of τ_+ of first entrance to $(0,\infty)$, there is a Markov jump process underlying a phase–type distribution which crosses the axis and ends at the first ladder height value. Let $\boldsymbol{\alpha}_+$ denote the vector which i'th elements is the probability that the Markov jump process crosses the axis in state i. Then

$$\boldsymbol{\alpha e} = \sum_i (\alpha_+)_i = ||G_+||.$$

After crossing the axis, the Markov jump process evolves according to its sub–intensity matrix $\boldsymbol{S}$. Hence the first ladder height has a (possibly defective if $||G_+|| < 1$) phase–type distribution with representation PH$(\boldsymbol{\alpha}_+, \boldsymbol{S})$. We summarize these considerations in the following theorem.

Theorem 5.7. *If $S_n = X_1 + ... + X_n$ is a random walk with $X_i = U_i - T_i$ where $T_i \sim A$ i.i.d. has a general absolutely continuous distribution concentrated on $(0,\infty)$ and U_i are i.i.d., independent of $\{T_i\}$ and phase–type distributed PH$(\boldsymbol{\alpha}, \boldsymbol{S})$, then G_+ is a (possibly defective) phase–type distribution with representation PH$(\boldsymbol{\alpha}_+, \boldsymbol{S})$ for some sub–probability vector $\boldsymbol{\alpha}_+$ which satisfies $||G_+|| = \boldsymbol{\alpha}_+\boldsymbol{e}$.*

By a simple renewal argument it is clear that all ladder heights $S_{\tau_+(n+1)} - S_{\tau_+(n)}$ are i.i.d. $\sim$ PH$(\boldsymbol{\alpha}_+, \boldsymbol{S})$. Then the ladder heights $S_{\tau_+(n)}$ are the arrivals in a phase–type renewal process. By concatenating the underlying Markov jump process we obtain a Markov jump process $\{m_t\}_{t\geq 0}$ with intensity matrix $\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+$ (see Figure 5.2). The initial distribution $m_0 \sim \boldsymbol{\alpha}_+$.

If the random walk has negative drift, then $||G_+|| < 1$ and the renewal process is terminating. In this case $\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+$ will be a sub–intensity matrix and the Markov jump process $\{m_t\}_{t\geq 0}$ will be terminating. The number of ladder increments $S_{\tau_+(n+1)} - S_{\tau_+(n)}$ is geometrically distributed with probability of a new ladder increment occurring begin $||G_+||$. The life–time of $\{m_t\}_{t\geq 0}$ coincides with the maximum M^+. We proved the following theorem.

Theorem 5.8. *If $||G_+|| < 1$ (which happens exactly when the mean of T_1 is strictly larger than the mean of U_1), the M^+ is finite almost surely and has a distribution consisting of an atom at 0 of size $1 - ||G_+||$ and an absolute continuous part being phase–type distributed with representation PH$(\boldsymbol{\alpha}_+, \boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+)$.*

Finally we shall address the question of how to calculate $\boldsymbol{\alpha}_+$.

Theorem 5.9. *The initial distribution $\boldsymbol{\alpha}_+$ satisfies the following equation*

$$\boldsymbol{\alpha}_+ = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+)u} dA(u).$$

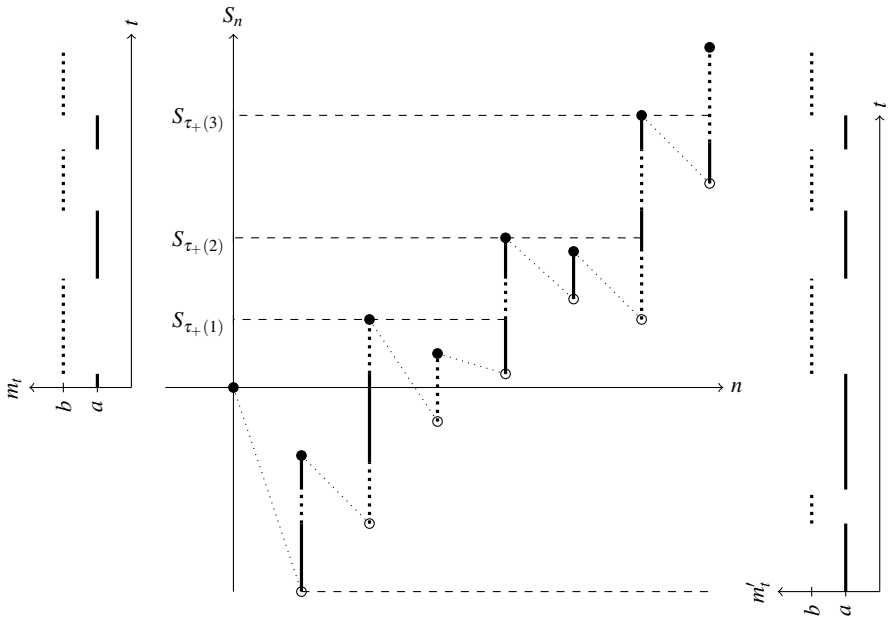

Fig. 5.2 A random walk $S_n = X_1 + ... + X_n$ with $X_i = U_i - T_i$ an with U_i being phase–type distributed with two phases a (thick line) and b (dotted line). The random walk starts in 0. Then $S_0 - T_1$ is constructed (hollow circle). Then $S_1 = S_0 - T_1 + U_1$ is constructed (solid circle) by letting a Markov jump process which underlies the phase–type distribution run from the hollow point and vertically up until absorption. This determines S_1. The first ladder height is obtained through a Markov jump process crossing level zero. Similarly, subsequent ladder heights are also obtained through remaining Markov jump processes. Theses processes are concatenated into a process $\{m_t\}_{t\geq 0}$. On the right hand side we consider the delayed phase–type renewal process constructed from the first arrival time begin the original phase–type distribution and the sub–sequent ones being the ladder height increments. Concatenating the underlying Markov jump processes in the delayed renewal process defines the process $\{m'_t\}_{t\geq 0}$

Proof. We define a new Markov jump process $\{m'_t\}_{t\geq 0}$ in the following way. The first point in the random walk S_1 may be generated by starting a Markov jump process underlying the phase–type distribution at $-T_1$ and let it run vertically upwards until absorption takes place. Let $-T_1$ be the time of zero in a delayed phase–type renewal process, where the first arrival time is U_1 and the next ones being the sub–sequent ladder increments, see Figure 5.2.

Concatenating the underlying Markov jump process in this delayed renewal process defines a (possibly terminating) Markov jump process $\{m'_t\}_{t\geq 0}$ with initial distribution $m'_0 \sim \boldsymbol{\alpha}$ and intensity matrix $\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+$.

Given that $T_1 = s$, the process of $\{m'_t\}_{t\geq 0}$ crosses zero at time s. Thus we must have that $m'_s \sim m_0$. Hence we must have that

$$\boldsymbol{\alpha}_+ = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}_+)u} dA(u).$$

□

Define the function

$$\phi(\boldsymbol{\beta}) = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\beta})u} dA(u).$$

Lemma 5.2. *$\phi(\boldsymbol{\beta})$ is an increasing function in $\boldsymbol{\beta}$ meaning that if $\boldsymbol{\beta} \leq \boldsymbol{\gamma}$ (i.e. $\beta_i \leq \gamma_i$ for all i) then $\phi(\boldsymbol{\beta}) \leq \phi(\boldsymbol{\gamma})$.*

Proof. Since $\boldsymbol{s} \geq \mathbf{0}$, $\boldsymbol{\beta} \to \boldsymbol{s}\boldsymbol{\beta}$ is increasing, hence so is $\boldsymbol{\beta} \to \boldsymbol{S}+\boldsymbol{s}\boldsymbol{\beta}$ and consequently $\boldsymbol{\beta} \to \exp((\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\beta})x)$ for any $x \geq 0$. Since $\boldsymbol{\alpha} \geq \mathbf{0}$ and $\boldsymbol{\alpha}\exp((\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\beta})x) \geq \mathbf{0}$, then $\boldsymbol{\beta} \to \phi(\boldsymbol{\beta})$ is also increasing. □

Theorem 5.10. *Define $\boldsymbol{\alpha}_+^{(0)} = \mathbf{0}$ and recursively*

$$\boldsymbol{\alpha}_+^{(n+1)} = \phi\left(\boldsymbol{\alpha}_+^{(n)}\right).$$

Then

$$\boldsymbol{\alpha}_+ = \lim_{n\to\infty} \alpha_+^{(n)}.$$

Proof. Since $\phi(\cdot)$ is an increasing function, then $\boldsymbol{\alpha}_+^{(n)}$ is an increasing sequence. On the other hand, $\mathbf{0} = \boldsymbol{\alpha}_+^{(0)} \leq \boldsymbol{\alpha}_+$ trivially so $\boldsymbol{\alpha}_+^{(1)} = \phi(\boldsymbol{\alpha}_+^{(0)}) \leq \phi(\boldsymbol{\alpha}_+) = \boldsymbol{\alpha}_+$. Continuing this way, we have that $\boldsymbol{\alpha}_+^{(n)} \leq \boldsymbol{\alpha}_+$ for all n. Hence $\{\boldsymbol{\alpha}_+^{(n)}\}_{n\in\mathbb{N}}$ is an bounded and increasing sequence, which is hence convergent. Moreover, it is immediate that $\lim_{n\to\infty} \boldsymbol{\alpha}_+^{(n)} \leq \boldsymbol{\alpha}_+$.

To prove the opposite inequality we define

$$\tilde{\alpha}_i^{(n)} = \mathbb{P}(m'_{T_1} = i, \tau_+ \leq n),$$

i.e. the probability that the process $\{m'_t\}_{t\geq 0}$ crosses the axis in state i and there has at most been $n-1$ negative random walk values prior to this event. Letting $n \to \infty$ we get that

$$\tilde{\alpha}_i^{(n)} \to \mathbb{P}(m_{T_1} = i, \tau_+ < \infty) = (\boldsymbol{\alpha}_+)_i.$$

It is then enough to prove that

$$\tilde{\boldsymbol{\alpha}}^{(n)} \leq \boldsymbol{\alpha}_+^{(n)}.$$

We do this by induction. For $n = 0$ it is trivial and assume that the inequality holds for n. Notice that $\tilde{\boldsymbol{\alpha}}^{(n)}$ is increasing in n.

At each ladder point below zero, if there are any, starting with S_1, there can be at most n points of the random walk below the S_1 level. The same is the case for any other ladder point below zero which is not S_1. Therefore the probability that the Markov jump process crosses a certain ladder height level in state i and on the event that $\tau_+ \leq n+1$ is at most $\tilde{\alpha}_i^{(n)}$. Hence by the induction hypothesis,

$$\begin{aligned}\tilde{\boldsymbol{\alpha}}^{(n+1)} &\leq \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\tilde{\boldsymbol{\alpha}}^{(n)})x} dA(x) \\ &\leq \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}^{(n)})x} dA(x) \\ &= \phi\left(\boldsymbol{\alpha}_+^{(n)}\right) \\ &= \boldsymbol{\alpha}^{(n+1)}.\end{aligned}$$

□

For the extension to the general matrix–exponential it is convenient to use the Wiener–factorization which will be treated in the next section.

5.3 Wiener–Hopf factorization

Theorem 5.11. *Consider the random walk $S_n = X_1 + ... + X_m$ where $X_1, X_2, ...$ i.i.d. $\sim F$ and $S_0 = 0$. Then*

$$F = G_+ + G_- - G_+ * G_-.$$

Proof. The identity follows from calculating the expression $\mathbb{E}\left(\sum_{n=0}^{\tau_+} 1\{S_n \in I\}\right)$ in two ways. First we separate out $n = \tau_+$. This amounts to

$$\begin{aligned}\mathbb{E}\left(\sum_{n=0}^{\tau_+} 1\{S_n \in I\}\right) &= \mathbb{E}\left(\sum_{n=0}^{\tau_+-1} 1\{S_n \in I\}\right) + \mathbb{P}(S_{\tau_+} \in I) \\ &= U_-(I) + G_+(I)\end{aligned}$$

by Theorem 5.4. Next we separate out $n = 0$. This gives

$$\begin{aligned}\mathbb{E}\left(\sum_{n=0}^{\tau_+} 1\{S_n \in I\}\right) &= \delta_0(I) + \mathbb{E}\left(\sum_{n=1}^{\tau_+} 1\{S_n \in I\}\right) \\ &= \delta_0(I) + \mathbb{E}\left(\sum_{n=1}^{\tau_+} 1\{X + S_{n-1} \in I\}\right) \quad (\text{for some } X \sim F) \\ &= \delta_0(I) + \int_{\mathbb{R}} \mathbb{E}\left(\sum_{n=0}^{\tau_+-1} 1\{S_n \in I - x\}\right) dF(x) \\ &= \delta_0(I) + U_- * F(I).\end{aligned}$$

Thus we have proved that

$$U_- + G_+ = \delta_0 + U_- * F. \tag{5.14}$$

Multiply (convolution) both sides of the equation by G_-. This results in

$$U_- * G_- + G_+ * G_- = G_- + U_- * G_- * F.$$

But $U_- * G_- = \sum_{n=0}^{\infty} G_-^{*n} * G_- = \sum_{n=1}^{\infty} G_-^{*n} = U_- - \delta_0$. Inserting this expression gives

$$U_- - \delta_0 + G_+ * G_- = G_- + (U_- - \delta_0) * F = G_- + U_- * F - F. \tag{5.15}$$

Now from (5.3) we see that $U_- - \delta_0 = U_- * F - G_+$ and substituting the first two terms of (5.15) with this expression results in

$$U_- * F - G_+ + G_+ * G_- = G_- + U_- * F - F$$

and hence

$$F = G_+ + G_- - G_+ * G_-.$$

□

Corollary 5.1. *If $I \subseteq (0,\infty)$ then*

$$G_+(I) = U_- * F(I). \tag{5.16}$$

If $I \subseteq (-\infty,0)$, then

$$U_-(I) = U_- * F(I).$$

Proof. This follows straightaway from . □

Now consider the random walk where $X_i = U_i - T_i$, $U_i \sim \mathrm{ME}(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ i.i.d. and $T_i \sim A$ i.i.d. and independent of each other. Now $F = F_U * F_{-T}$, where F_U and F_{-T} are the distribution functions of U_1 and $-T_1$ respectively. Then from 5.16 we get that $G_+ = F * U_- = F_U * F_{-T} * U_-$ and hence

$$G_+(x) = \int_{-\infty}^{0} F_U(x-y)d(F_{-T} * U_-(y))$$

which implies that

$$g_+(x) = \int_{-\infty}^{0} f_u(x-y)d(F_{-T} * U_-(y)) = \left[\int_{-\infty}^{0} \boldsymbol{\alpha} e^{-\boldsymbol{S}y} d(F_{-T} * U_-(y))\right] e^{\boldsymbol{S}x}\boldsymbol{s},$$

where g_+ and f_U are the densities corresponding to G_+ and F_U respectively. Hence there exists a vector $\boldsymbol{\alpha}_+$ such that the distribution G_+ is matrix–exponential with representation $(\boldsymbol{\alpha}_+,\boldsymbol{S},\boldsymbol{s})$. If $(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ has property (R) (see Definition 1.4), then it follows from Theorem 1.8 that $\boldsymbol{\alpha}_+$ is unique.

It follows from Corollary 4.1 together with Remark **??** if the drift of the random walk is negative, then the maximum $M = \sup_{n\geq 0} S_n$ is finite. Then the maximum of the random walk corresponds to the life–time of the defective renewal process of the ascending ladder heights. Hence by Theorem 4.17 M must have a zero–modified matrix–exponential distribution with representation $(\boldsymbol{\alpha}_+,\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}_+,(1+\boldsymbol{\alpha}_+\boldsymbol{S}^{-1}\boldsymbol{s})\boldsymbol{s})$. We summarize the results in the following theorem.

Theorem 5.12. *For the random walk $S_n = X_1 + ... + X_n$ with $X_i = U_i - T_i$, $U_i \sim ME(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ i.i.d. and $T_i \sim A$ i.i.d. and independent of each other, the ascending ladder height distribution G_+ is a matrix–exponential with a representation*

$(\boldsymbol{\alpha}_+, \boldsymbol{S}, \boldsymbol{s})$ *for some vector* $\boldsymbol{\alpha}_+$. *If* $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$ *has property (R), then* $\boldsymbol{\alpha}_+$ *is unique. If the random walk has negative drift, the maximum M of the random walk has a zero–modified matrix–exponential distribution with and atom at zero of size* $1 + \boldsymbol{\alpha}_+ \boldsymbol{S}^{-1} \boldsymbol{s}$ *and an absolute continuous and defective part with representation* $(\boldsymbol{\alpha}_+, \boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+, (1 + \boldsymbol{\alpha}_+ \boldsymbol{S}^{-1} \boldsymbol{s})\boldsymbol{s})$.

The next result states that we may find $\boldsymbol{\alpha}_+$ in a similar way as for the phase–type case.

Theorem 5.13. *If* $(\boldsymbol{\alpha}_+, \boldsymbol{S}, \boldsymbol{s})$ *is a p dimensional representation which has property (R), then* $\boldsymbol{\alpha}_+$ *satisfies*

$$\boldsymbol{\alpha}_+ = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+)x} dA(x).$$

Furthermore, if $\alpha_+^{(0)} = \boldsymbol{0}$ *and*

$$\boldsymbol{\alpha}_+^{(n+1)} = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S} + \boldsymbol{s}\boldsymbol{\alpha}_+^{(n)})x} dA(x),$$

then $\boldsymbol{\alpha}_+^{(n)} \to \boldsymbol{\alpha}_+$ *as* $n \to \infty$.

Proof. Define for fixed $\boldsymbol{S}$ and $\boldsymbol{s}$ the norm of a vector $\boldsymbol{\beta}$ by

$$||\boldsymbol{\beta}|| = \int_0^\infty \left| \boldsymbol{\beta} e^{\boldsymbol{S}x} \boldsymbol{s} \right| dx.$$

To see that it is a norm we use the property (R). By Theorem 1.7 then $\dim(R_e) = p$. Obviously $||\cdot||$ satisfies the triangle inequality and scaling property. If $||\boldsymbol{\beta}|| = 0$, then $\boldsymbol{\beta} e^{\boldsymbol{S}x} \boldsymbol{s}$ for almost all x, and by continuity hence for all x. But since $\dim(R_e) = p$, where we recall $R_e = \{e^{\boldsymbol{S}x}\boldsymbol{s} | x \geq 0\}$, then $\boldsymbol{\beta} = \boldsymbol{0}$. Thus $||\cdot||$ is indeed a norm. Since this a norm on $\mathbb{C}^p$, a finite dimensional vector–space, this norm is equivalent to the Euclidean norm on $\mathbb{C}^p$. Hence in order to prove that $\boldsymbol{\alpha}_+^{(n)} \to \boldsymbol{\alpha}_+$ we may do it in the $||\cdot||$ norm instead of the Euclidean norm, i.e. prove that

$$||\boldsymbol{\alpha}_+^{(n)} - \boldsymbol{\alpha}_+|| \to 0$$

as $n \to \infty$.

We shall also introduce a partial ordering $\preceq$ defined by

$$\boldsymbol{\alpha} \preceq \boldsymbol{\beta} \iff \forall\, x \geq 0 :\ \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} \leq \boldsymbol{\beta} e^{\boldsymbol{S}x} \boldsymbol{s}.$$

Now suppose that $\boldsymbol{\alpha} \preceq \boldsymbol{\beta} \preceq \boldsymbol{\gamma}$. Then

$$\forall x \geq 0 :\ \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} \leq \boldsymbol{\beta} e^{\boldsymbol{S}x} \boldsymbol{s} \leq \boldsymbol{\gamma} e^{\boldsymbol{S}x} \boldsymbol{s}.$$

Then also

$$\forall x \geq 0 :\ 0 \leq (\boldsymbol{\gamma} - \boldsymbol{\beta})\, e^{\boldsymbol{S}x} \boldsymbol{s} \leq (\boldsymbol{\gamma} - \boldsymbol{\alpha})\, e^{\boldsymbol{S}x} \boldsymbol{s}$$

from which

$$\int_0^\infty \left|(\boldsymbol{\gamma}-\boldsymbol{\beta})\, e^{\boldsymbol{S}x}\boldsymbol{s}\right| dx \leq \int_0^\infty \left|(\boldsymbol{\gamma}-\boldsymbol{\alpha})\, e^{\boldsymbol{S}x}\boldsymbol{s}\right| dx$$

and hence

$$||\boldsymbol{\gamma}-\boldsymbol{\beta}|| \leq ||\boldsymbol{\gamma}-\boldsymbol{\alpha}||.$$

We shall use this fact to construct another sequence $\boldsymbol{\beta}_+^{(n)}$ for which $\boldsymbol{\beta}_+^{(n)} \preceq \boldsymbol{\alpha}_+^{(n)} \leq \boldsymbol{\alpha}_+$ and such that $||\boldsymbol{\beta}_+^{(n)} - \boldsymbol{\alpha}_+|| \to 0$ as $n \to \infty$ and from this we can then conclude that $||\boldsymbol{\alpha}_+^{(n)} - \boldsymbol{\alpha}_+|| \to 0$ as $n \to \infty$.

Let

$$\varphi(\boldsymbol{\alpha}) = \int_0^\infty \boldsymbol{\alpha} e^{(\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha})x} dA(x).$$

This is exactly the function which defines de sequence $\{\boldsymbol{\alpha}_+^{(n)}\}_{n\in\mathbb{N}}$, i.e. $\boldsymbol{\alpha}_+^{(n+1)} = \varphi(\boldsymbol{\alpha}_+^{(n)})$. We shall now prove that

$$\boldsymbol{\alpha} \preceq \boldsymbol{\beta} \implies \varphi(\boldsymbol{\alpha}) \preceq \varphi(\boldsymbol{\beta}). \tag{5.17}$$

□

5.4 Partial eigenfunction methods

5.5 Bilateral matrix–exponential distributions

Exercises

Exercise 5.1. Prove, that if we include an arrival at zero in the definition of renewal process, then (5.12) must be replaced by

$$U_+^w = \frac{1}{1-\rho} U_+.$$

Chapter 6
Representations of phase–type distributions

6.1 Irreducible representations

Definition 6.1. A phase–type representation $\mathrm{PH}_p(\boldsymbol{\alpha},\boldsymbol{S})$ is called irreducible if there is a positive probability of visiting any of the transient states starting from the initial distribution $\boldsymbol{\alpha}$, i.e. if

$$\left(\boldsymbol{\alpha}e^{\boldsymbol{S}x}\right)_i > 0$$

for $i = 1,...,p$ and for some $x > 0$.

Note that in the definition above we may replace "for some $x > 0$" by "for all $x > 0$".

Theorem 6.1. *A phase–type representation $PH_p(\boldsymbol{\alpha},\boldsymbol{S})$ is irreducible if and only if the matrix $\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}$ is an irreducible intensity matrix, i.e. an intensity matrix of an irreducible Markov jump proces. Here, as usual, $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$.*

Proof. If the representation is irreducible, then we can reach any state starting according to the initial distribution $\boldsymbol{\alpha}$. Now let $\{X_t\}_{t\geq 0}$ be a Markov jump process with intensity matrix $\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}$ and assume that $X_t = i$. We must then prove that $p_{ij}^s = \mathbb{P}(X_{t+s} = j|X_t = i) > 0$ for all j and $s > 0$. But X_t is obtained by concatenation of the Markov jump processes generating phase–type inter–arrivals in a phase–type renewal process. Hence we can reach the state j at least after a renewal has occurred. Hence $p_{ij}^s > 0$ for some $s > 0$.

On the other hand, assume that $\boldsymbol{S}+\boldsymbol{s}\boldsymbol{\alpha}$ is irreducible. Again, the structure of the intensity matrix is that of a phase–type renewal process. Since all states in the concatenated process communicate, then also after a renewal and initiating with $\boldsymbol{\alpha}$ we can reach any state. Hence the phase–type representation is irreducible. □

6.1.1 Time–reversed representations

Consider an irreducible representation $\mathrm{PH}_p(\boldsymbol{\pi},\boldsymbol{T})$ of a phase–type distribution. Then the Markov jump process $\{X_t\}_{t\geq 0}$ with intensity matrix $\boldsymbol{\Lambda}=\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}$ is irreducible, and since the state–space is finite, it is also (positive) recurrent. Then by Theorems 2.20 and 2.21 we conclude that the process has a stationary distribution $\boldsymbol{\nu}$. Then by Theorem 2.26 we conclude that $\{X_t\}$ is time reversible with intensity matrix

$$\tilde{\boldsymbol{\Lambda}}=\boldsymbol{\Delta}^{-1}(\boldsymbol{\nu})\boldsymbol{\Lambda}'\boldsymbol{\Delta}(\boldsymbol{\nu})$$

and with the same stationary distribution $\boldsymbol{\nu}$ (Theorem 2.26).

Theorem 6.2. *Consider an irreducible phase–type representation $PH_p(\boldsymbol{\pi},\boldsymbol{T})$. A stationary measure $\tilde{\boldsymbol{\nu}}$ of the process $\{X_t\}_{t\geq 0}$ with intensity matrix $\boldsymbol{\Lambda}=\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}$ is given by*

$$\tilde{\boldsymbol{\nu}}=\boldsymbol{\pi}(-\boldsymbol{T})^{-1}. \tag{6.1}$$

Furthermore, $\tilde{\boldsymbol{\nu}}>\boldsymbol{0}$. The interpretation of $\tilde{\nu}_i$ is the expected time the phase–type process spends in state i prior to absorption. The corresponding stationary distribution $\boldsymbol{\nu}$ is given by

$$\boldsymbol{\nu}=\frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}. \tag{6.2}$$

Proof. $\tilde{\boldsymbol{\nu}}=\boldsymbol{\pi}(-\boldsymbol{T})^{-1}$ obviously satisfies that $\tilde{\boldsymbol{\nu}}(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi})=\boldsymbol{0}$. Furthermore, $(-\boldsymbol{T})^{-1}$ is non–negative since the ij'th element being the expected time the Markov jump process underlying a phase–type distribution will spend in state j given that it initiates in state i. By irreducibility, initiating with distribution $\boldsymbol{\pi}$, the expected time the process spends in each state must be strictly positive since there is a strictly positive probability that it will visit the state prior to absorption. Hence $\tilde{\boldsymbol{\nu}}>\boldsymbol{0}$. Then the result follows from Theorem 2.22.

We notice that $\tilde{\boldsymbol{\nu}}=\boldsymbol{\pi}(-\boldsymbol{T})^{-1}$ is not a distribution unless the mean of the phase–type distribution is one. The corresponding stationary distribution $\boldsymbol{\nu}$ is obviously given by

$$\boldsymbol{\nu}=\frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}.$$

□

Now consider the time reversed process $\{\tilde{X}_t\}_{t\geq 0}$. Then

$$\tilde{\boldsymbol{\Lambda}}=\boldsymbol{\Delta}^{-1}(\boldsymbol{\nu})\boldsymbol{\Lambda}'\boldsymbol{\Delta}(\boldsymbol{\nu})=\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{\Lambda}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}}).$$

Inspired by this equation we now propose

Theorem 6.3. *(Time–reversed phase–type representation) Let $\tilde{\boldsymbol{\pi}}=\boldsymbol{t}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})$, $\tilde{\boldsymbol{T}}=\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})$ and $\tilde{\boldsymbol{t}}=\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{\pi}'$. Then $(\tilde{\boldsymbol{\pi}},\tilde{\boldsymbol{T}},\tilde{\boldsymbol{t}})$ is a phase–type representation of the same distribution which as representation $(\boldsymbol{\pi},\boldsymbol{T},\boldsymbol{t})$.*

Proof. First we prove that $(\tilde{\boldsymbol{\pi}}, \tilde{\boldsymbol{T}}, \tilde{\boldsymbol{t}})$ is indeed a phase–type representation. To this end we must prove that $\tilde{\boldsymbol{\pi}}$ is a probability vector, that $\tilde{\boldsymbol{T}}$ is a sub–intensity matrix and that $\tilde{\boldsymbol{t}} = -\tilde{\boldsymbol{T}}\boldsymbol{e}$.

Concerning $\boldsymbol{\pi}$, it is clear that the elements of $\boldsymbol{t}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})$ are all non–negative. Using that $\tilde{\boldsymbol{\nu}} = -\boldsymbol{\pi}\boldsymbol{T}^{-1}$, or which is the same, $-\tilde{\boldsymbol{\nu}}\boldsymbol{T} = \boldsymbol{\pi}$, we get that the sum of the elements is given by that

$$\begin{aligned}
\boldsymbol{t}'\Delta(\tilde{\boldsymbol{\nu}})\boldsymbol{e} &= (-\Delta(\tilde{\boldsymbol{\nu}})\boldsymbol{T}\boldsymbol{e})'\boldsymbol{e} \\
&= -\boldsymbol{e}'\Delta(\tilde{\boldsymbol{\nu}})\boldsymbol{T}\boldsymbol{e} \\
&= -\boldsymbol{\nu}\boldsymbol{T}\boldsymbol{e} \\
&= \boldsymbol{\pi}\boldsymbol{e} \\
&= 1.
\end{aligned}$$

Similarly, the exit rate vector

$$\begin{aligned}
\Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{\pi}' &= \Delta^{-1}(\tilde{\boldsymbol{\nu}})(-\tilde{\boldsymbol{\nu}}\boldsymbol{T})' \\
&= -\Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\Delta(\tilde{\boldsymbol{\nu}})\boldsymbol{e} \\
&= -\tilde{\boldsymbol{T}}\boldsymbol{e}.
\end{aligned}$$

That $\tilde{\boldsymbol{T}}$ is an sub–intensity matrix is also clear since the diagonal elements are still negative, all other elements positive and since $\tilde{\boldsymbol{t}}$ is obviously non–negative, then the sums of the rows of $\tilde{\boldsymbol{T}}$ must be non–positive.

To show that the reversed representation gives rise to the original phase–type distribution follows trivially from

$$\begin{aligned}
f(x) &= \boldsymbol{\pi}e^{\boldsymbol{T}x}\boldsymbol{t} \\
&= \boldsymbol{\pi}\Delta(\tilde{\boldsymbol{\nu}})\Delta^{-1}(\tilde{\boldsymbol{\nu}})\ \exp[\boldsymbol{T}x]\ \Delta(\tilde{\boldsymbol{\nu}})\Delta^{-1}(\tilde{\boldsymbol{\nu}}) \\
&= \boldsymbol{\pi}\Delta(\tilde{\boldsymbol{\nu}})\ \exp\left[\Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}\Delta(\tilde{\boldsymbol{\nu}})x\right]\ \Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{t} \\
&= \boldsymbol{t}'\Delta^{-1}(\tilde{\boldsymbol{\nu}})\ \exp\left[\Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\Delta(\tilde{\boldsymbol{\nu}})x\right]\ \Delta^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{\pi}' \\
&= \tilde{\boldsymbol{\pi}}e^{\tilde{\boldsymbol{T}}x}\,\tilde{\boldsymbol{t}},
\end{aligned}$$

where second to the last equation is simply obtained by transposition. □

It is clear from the construction that the Markov jump process obtained through concatenating the Markov jump processes underlying the phase–type distributions using the time reversed representations coincide with the time–reversed Markov jump process obtained through time reversing the concatenated Markov jump processes of the phase–type distributions of the original representation.
[1]

[1] Here: Given an irr. and recurrent Markov jump process, construct a renewal process which has this a concatenated process

6.1.2 Similarity transformations

The time reversed phase–type construction is an example of a similarity transformation. If $(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{t})$ is a p–dimensional representation of either a phase–type distribution or more generally a matrix–exponential distribution, then for any non–singular $p \times p$–matrix $\boldsymbol{M}$ we have that

$$f(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{t} = \boldsymbol{\pi} \boldsymbol{M} \exp\left[\boldsymbol{M}^{-1} \boldsymbol{T} \boldsymbol{M} x\right] \boldsymbol{M}^{-1} \boldsymbol{t},$$

so also $(\boldsymbol{\pi}\boldsymbol{M}, \boldsymbol{M}^{-1}\boldsymbol{T}\boldsymbol{M}, \boldsymbol{M}^{-1}\boldsymbol{t})$ is a representation. We say that the latter is obtained through a similarity transformation of the former.

Example of similarity transformations...

As mentioned previously, a phase–type distribution will in general have many distint representations.

- Minimal order known from Cinneide
- Similarity transformations, particularly the one giving the reversed representation
- Expansion of order - to be used in multivariate chapter
- Special discussion on APH/TPH/Coxian subclass

Chapter 7
Characterization of phase–type distributions

7.1 Some concepts from convex analysis

If C is a subset of $\mathbb{R}^n$, then the smallest linear subspace which contains C is called the affine hull of C. We are going to consider C relative to this subspace and define properties like closure and interior of C relative to this.

Definition 7.1. Let $C \subset \mathbb{R}^n$ and $B(\boldsymbol{x},r) = \{\boldsymbol{y} \in \mathbb{R}^n | \ \|\boldsymbol{y}-\boldsymbol{x}\| \leq r\}$ be the closed ball about $\boldsymbol{x}$ with radius r. Then the affine hull, $\text{aff}(C)$, of C is defined by

$$\text{aff}(C) = \{\lambda_1 \boldsymbol{x}_1 + ... + \lambda_n \boldsymbol{x}_n \mid \boldsymbol{x}_i \in C, \lambda_i \in \mathbb{R}, i = 1,2,...,n, \lambda_1 + ... + \lambda_n = 1\}.$$

Definition 7.2. The relative interior of a set $C \subset \mathbb{R}^n$ is given by

$$\text{ri}(C) = \{\boldsymbol{x} \in \text{aff}(C) \mid \exists \varepsilon > 0 : B(\boldsymbol{x},\varepsilon) \cap \text{aff}(C) \subset C\}.$$

Example 7.1. Consider the disc D in $\mathbb{R}^3$ given by

$$D = \{(x,y,z) \in \mathbb{R}^3 \ | \text{Ł} \ z = 1, (x-2)^2 + (y-2)^2 \leq 1\}.$$

Seen as a set in $\mathbb{R}^3$, D does not have an interior since we cannot place 3–dimensional balls about points in D which will still remain in D. However, the affine hull of D is the plane $\{(x,y,z) \in \mathbb{R}^3 \mid z = 1\}$ and relative to this plane in 2 dimensions the interior of the disc is well defined. This is exactly the relative interior of the disc (see Fig. **??** for details). Indeed, the relative interior of the disc consists in all those points of the plane for which there exists a ball (in $\mathbb{R}^3$) about the the points such that the intersection of the ball with the plane (which is a circle then) is contained in the disc. □

If C is a n–dimensional convex subset of $\mathbb{R}^n$, then $\text{aff}(C) = \mathbb{R}^n$ since the smallest linear space in $\mathbb{R}^n$ which contains C is $\mathbb{R}^n$. In this case the relative interior of C coincides with the interior of the set C in the usual sense.

We recall that a set C is convex if for all $x, y \in C$ and $0 \leq \lambda \leq 1$ we have that $\lambda x + (1-\lambda)y \in C$. Similarly as for the affine hull we shall consider the convex hull of a set C as the smallest convex set which contains C. This amounts to

Definition 7.3.

$$\mathrm{co}(C) = \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \;\middle|\; \boldsymbol{x}_i \in C, \lambda_i \geq 0, i = 1, ..., k, \lambda_1 + ... + \lambda_k = 1, k = 1, 2, ... \right\}.$$

This definition is quite similar to the affine hull apart from the restrictions on the non–negativity of the λ_i's. In particular, the convex hull of a finite number of points/vectors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ is given by

$$\mathrm{co}(\boldsymbol{x}_1, ..., \boldsymbol{x}_n) = \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \;\middle|\; \lambda_i \geq 0, i = 1, ..., n, \lambda_1 + ... + \lambda_n = 1 \right\}.$$

Theorem 7.1. *Let $C \subseteq \mathbb{R}^n$ be a convex set. Then*

$$\boldsymbol{x} \in ri(C) \text{ and } \boldsymbol{y} \in cl(C) \implies (1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y} \in ri(C)$$

for all $0 \leq \lambda < 1$, and where $cl(C)$ denotes the closure of the set C.

Proof. We may assume without loss of generality that C has dimension n since all vectors we shall consider are in the set C or on its boundary. Then the relative interior is the same as the interior of the set. Now take $\boldsymbol{x}, \boldsymbol{y}$ and λ as stated. Since $\boldsymbol{y} \in \mathrm{cl}(C)$ then for all $\varepsilon > 0$ we must have that $y \in B(0, \varepsilon) + C$ (any ball about $\boldsymbol{y}$ intersects with C). Then

$$\begin{aligned} B((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}, \varepsilon) &= (1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y} + \varepsilon B(0,1) \\ &\subseteq (1-\lambda)\boldsymbol{x} + \lambda (B(0,\varepsilon) + C) + \varepsilon B(0,1) \\ &= (1-\lambda)\boldsymbol{x} + \lambda C + (\lambda + 1)\varepsilon B(0,1) \\ &= (1-\lambda)\left(\boldsymbol{x} + \varepsilon \frac{1+\lambda}{1-\lambda} B(0,1) \right) + \lambda C. \end{aligned}$$

Now $\boldsymbol{x}$ belongs to the interior of C, so for ε sufficiently small $\boldsymbol{x} + \varepsilon \frac{1+\lambda}{1-\lambda} B(0,1)$ will be contained in C, so for ε sufficiently small we have that

$$Ł B((1-\lambda)x + \lambda y, \varepsilon) \subseteq (1-\lambda)C + \lambda C = C$$

which proves that indeed $(1-\lambda)x + \lambda y$ is in the interior of C. □

From the above it is clear, that if C is a convex set and both $\boldsymbol{x}$ and $\boldsymbol{y}$ belongs to $\mathrm{ri}(C)$ then for $0 \leq \lambda \leq 1$ also $(1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y} \in \mathrm{ri}(C)$. Hence $\mathrm{ri}(C)$ is again a convex set. Also, if both $\boldsymbol{x}$ and $\boldsymbol{y}$ belongs to $\mathrm{cl}(C)$, then we take a sequence $\{\boldsymbol{x}_n\} \subseteq \mathrm{ri}(C)$ such that $\boldsymbol{x}_n \to \boldsymbol{x}$. Then all $(1-\lambda)\boldsymbol{x}_n + \lambda \boldsymbol{y} \in \mathrm{ri}(C)$, and since it converges to $(1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}$ the latter must be in the closure by definition. The affine hull is obviously the same for both $\mathrm{ri}(C)$ and $\mathrm{cl}(C)$ and they have the same dimension as C. This implies that if C is convex and non–empty, then $\mathrm{ri}(C)$ is non–empty as well.

Theorem 7.2. *Let $C \neq \emptyset$ be a convex set in $\mathbb{R}^n$. Then*

$$\boldsymbol{z} \in ri(C) \iff \forall \boldsymbol{x} \in C \; \exists \mu > 1 : \mu \boldsymbol{z} + (1-\mu)\boldsymbol{x} \in C.$$

Proof. If $\boldsymbol{z} \in \mathrm{ri}(C)$ then there is a ball in $\mathbb{R}^n$ about $\boldsymbol{z}$ such that its intersection with the affine hull of C is contained in C. Hence a line from any point in C to $\boldsymbol{z}$ may be continued further at least within this neighborhood (see Fig. **??**).

Conversely assume that for a given $\boldsymbol{x} \in C$ there is a $\mu > 1$ such that $\mu \boldsymbol{z} + (1-\mu)\boldsymbol{x} \in C$. Since C is non–empty, so is $\mathrm{ri}(C)$ and we may then choose $\boldsymbol{x} \in \mathrm{ri}(C)$. So we know that $\boldsymbol{x} \in \mathrm{ri}(C)$ and $\boldsymbol{y} := \mu \boldsymbol{z} + (1-\mu)\boldsymbol{x} \in C$ for some μ. Then

$$\boldsymbol{z} = (1 - \frac{1}{\mu})\boldsymbol{x} + \frac{1}{\mu}\boldsymbol{y}$$

and $\boldsymbol{z} \in \mathrm{ri}(C)$ then follows by Theorem 7.1. □

It follows immediately that

Corollary 7.1. *Let C be an non–empty convex set in $\mathbb{R}^n$. Then*

$$\boldsymbol{z} \in int(C) \iff \forall \boldsymbol{y} \in \mathbb{R}^n \; \exists \varepsilon > 0 : \boldsymbol{z} + \varepsilon \boldsymbol{x} \in C,$$

where $int(C)$ denotes the interior of the set C.

Lemma 7.1. *If C_1 and C_2 are non–empty convex sets in a finite–dimensional vector space, then*

$$ri(co(C_1 \cup C_2)) = \{pc_1 + (1-p)c_2 | c_1 \in C_1, c_2 \in C_2, 0 < p < 1\}.$$

7.2 Characterization of phase–type distributions via their densities

While the problem of finding a minimal representation of a matrix–exponential distribution is easily solved by reducing its rational Laplace transform to an irreducible fraction (the numerator and denominator not having common roots) and then using the canonical representation for this fraction, the equivalent problem of finding a minimal representation in the class of phase–type distributions is a difficult and still open problem. Since the phase–type distributions have rational Laplace transforms we can of course always find a minimal matrix–exponential representation, but we may be interested in finding a minimal representation where the special structure of the initial distribution and sub–intensity matrix is preserved, that is, a representation by which the probabilistic interpretation is preserved.

We shall not be able to characterize minimal representations for phase–type distributions, however, in the course to of a characterization of the phase–type distributions themselves we shall analyze the reducibility of phase–type distributions from a probabilistic point of view.

It is clear that by adding states to the state–space of a Markov jump process underlying a phase–type distribution which cannot be reached from the original states or initiated in, the time until absorption will have the same distributions as originally. Such states are superfluous states which may be deleted.

Theorem 7.3. *Let $X \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ be a phase–type distribution. Then its density $f(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t}$ is strictly positive for all $x > 0$.*

Proof. Since the density of a phase–type distribution is invariant under different representation we may assume without loss of generality that the $\mathrm{PH}_p(\boldsymbol{\pi}, \boldsymbol{T})$ is an irreducible representation. Then

$$\left(\boldsymbol{\pi} e^{\boldsymbol{T}x}\right)_i > 0$$

for all $x > 0$ and $i = 1, ..., p$. Since one of the exit rates t_i, $i = 1, ..., p$ must be strictly positive it then follows that the density f must satisfy

$$f(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x}\boldsymbol{t} = \sum_{i=1}^{p} \left(\boldsymbol{\pi} e^{\boldsymbol{T}x}\right)_i t_i > 0.$$

□

Definition 7.4. Let $C \subseteq \mathscr{S}$ be a subset of the space of signed measures. Then we say that C is R–invariant if for all $\boldsymbol{\nu} \in C$ we have that $R_t\boldsymbol{\nu} \in C$ for all t.

The following lemma is of key importance when characterizing the class of phase–type distributions later on.

Lemma 7.2. *If $co(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d)$ is R–invariant, then $\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d$ have rational Laplace–Stieltjes transforms. Furthermore, there exists a matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j=1....,d}$ with $\lambda_{ij} \geq 0$ for $i \neq j$ and $\lambda_{ii} = -\sum_{j\neq i}\lambda_{ij}$ such that*

$$\Gamma\boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij}\boldsymbol{\nu}_j.$$

Proof. $\mathrm{co}(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d)$ is obviously a finite–dimensional vector space (of dimension at most $d-1$ since the λ's sum up to one) and R–invariance implies that $R_t\boldsymbol{\nu}_i \in \mathrm{co}(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d)$ for all t. Hence $V_R(\boldsymbol{\nu}_i) = \mathrm{span}\{R_t\boldsymbol{\nu}_i : t \geq 0\} \subseteq \mathrm{co}(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d)$ and $V_R(\boldsymbol{\nu}_i)$ is hence of finite dimension. Thus by Theorem 1.26, ν_i has a rational Laplace–Stieltjes transform.

Concerning the existence of the matrix $\boldsymbol{\Lambda}$ we proceed as follows. Since $R_t\boldsymbol{\nu}_i \in \mathrm{co}(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d)$ for all t, then for each t there are numbers $c_{ij}(t)$, $i, j = 1, ..., d$ (i corresponds to the index on $\boldsymbol{\nu}_i$) such that $c_{i1}(t) + ... + c_{id} = 1$ for all $i = 1, ..., d$ and

$$R_t\boldsymbol{\nu}_i = \sum_{j=1}^{d} c_{ij}(t)\boldsymbol{\nu}_j$$

for $i = 1, ..., d$. Thus we have that

$$\frac{R_t \boldsymbol{\nu}_i - \boldsymbol{\nu}_i}{t} = \frac{1}{t}\left(\sum_{j=1}^{d} c_{ij}(t)\boldsymbol{\nu}_i - \boldsymbol{\nu}_i\right) = \sum_{j=1}^{d} \frac{c_{ij}(t)}{t}(\boldsymbol{\nu}_j - \boldsymbol{\nu}_i)$$
$$\in C_i := \left\{ \sum_{i=1}^{d} \lambda_j(\boldsymbol{\nu}_j - \boldsymbol{\nu}_i) \middle| \lambda_j \geq 0, j = 1, ..., d, \lambda_1 + ... + \lambda_d = 1 \right\}$$

for all t. The sets C_i are known as a convex cones and are obviously a closed sets. Hence

$$\Gamma \boldsymbol{\nu}_i = \lim_{t \downarrow 0} \frac{R_t \boldsymbol{\nu}_i - \boldsymbol{\nu}_i}{t} \in C_i$$

for $i = 1, ..., d$. Hence there exists λ_{ij} such that

$$\Gamma \boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij}(\boldsymbol{\nu}_j - \boldsymbol{\nu}_i).$$

Since $\boldsymbol{\nu}_i - \boldsymbol{\nu}_j = 0$ for $j = i$, we can choose λ_{ii} as we please, and we define

$$\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}.$$

Then

$$\Gamma \boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij}(\boldsymbol{\nu}_j - \boldsymbol{\nu}_i) = \sum_{\substack{j=1 \\ j \neq i}}^{d} \lambda_{ij}(\boldsymbol{\nu}_j - \boldsymbol{\nu}_i)$$
$$= \sum_{\substack{j=1 \\ j \neq i}}^{d} \lambda_{ij}\boldsymbol{\nu}_j - \sum_{\substack{j=1 \\ j \neq i}}^{d} \lambda_{ij}\boldsymbol{\nu}_i = \sum_{\substack{j=1 \\ j \neq i}}^{d} \lambda_{ij}\boldsymbol{\nu}_j + \lambda_{ii}\boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij}\boldsymbol{\nu}_j.$$

□

The matrix $\boldsymbol{\Lambda}$ in the theorem above is obviously an intensity matrix for a Markov jump process. In the next theorem we shall specialize to the situation where the structure of the Markov jump process is such that there is an absorbing state.

Theorem 7.4. *Let $\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d$ be signed measures with no atom at zero. Consider the convex hull*

$$P = co(\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_d, \delta_0)$$

where δ_0 is the measure which density is the Dirac delta function. Then P is R–invariant if and only if there exists a matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j=1....,d+1}$ with $\lambda_{ij} \geq 0$ for $i \neq j$, $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$ and $\lambda_{d+1,i} = 0$ for $i = 1, ..., d+1$ such that

$$\Gamma \boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij}\boldsymbol{\nu}_j + \lambda_{i,d+1}\delta_0,$$

and $\boldsymbol{S} = \{\lambda_{ij}\}_{i,j=1,...,d}$ *is a sub–intensity matrix which generate phase–type distributions of the form* $PH(\boldsymbol{\alpha}, \boldsymbol{S})$.

Proof. Assume the existence of $\boldsymbol{\Lambda}$, the restriction of which to the states $1,2,...,d$, $\boldsymbol{S}$, is a sub–intensity matrix which generates phase–type distributions with representations $(\boldsymbol{\alpha}, \boldsymbol{S})$ for some stochastic vector $\boldsymbol{\alpha}$. The density of such phase–type distributions are given by $f(x) = \boldsymbol{\alpha}\exp(\boldsymbol{S}x)\boldsymbol{s}$. Hence

$$R_t f(x) = f(x+t) = \boldsymbol{\alpha} e^{\boldsymbol{S}t} e^{\boldsymbol{S}x} \boldsymbol{s}$$

which is defective phase–type with representation $(\boldsymbol{\alpha}\exp(\boldsymbol{S}t), \boldsymbol{S})$. This proves the R–invariance and the "if" part of the theorem.

Now assume that P as stated is R–invariant. Then from Lemma 7.2 we have that there exists an intensity matrix $\boldsymbol{\Lambda}$ such that

$$\Gamma \boldsymbol{\nu}_i = \sum_{j=1}^{d} \lambda_{ij} \boldsymbol{\nu}_j + \lambda_{i,d+1} \delta_0, \;\; i = 1,2,...,d \tag{7.1}$$

and since the infinitesimal generator of Dirac delta function is zero we assign values zero to $\lambda_{d+1,i}$ for $i = 1,2,...,d+1$. Hence the structure of $\boldsymbol{\Lambda}$ is

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{S} & \boldsymbol{s} \\ \boldsymbol{0} & 0 \end{pmatrix},$$

where $\boldsymbol{S}$ is the restriction of $\boldsymbol{\Lambda}$ to the indices $1,2,...,d$. In order for $\boldsymbol{S}$ to be a sub–intensity matrix which generate phase–type distributions which must show that $\boldsymbol{S}$ is invertible since this, in turn, is equivalent to the existence of states $1,2,...,d$ which are transient (see Theorem 3.3). To this end we prove that the eigenvalue of $\boldsymbol{S}$ which has the largest real part is strictly negative. This implies that all eigenvalues are different from 0 and hence that the matrix is invertible.

We now apply (7.1) to the survival function $\bar{F}_i : x \to \nu_i([x,\infty))$. First notice that since ν_i have rational Laplace transforms (P being R–invariant), and hence analytic, the functions $\bar{F}_i$ must be differentiable. The generator Γ applied to $\bar{F}_i$ is simply the derivative from the right of $\bar{F}_i$, which in turn coincides with the derivative of $\bar{F}_i$ the latter being differentiable. Hence (7.1) amounts to

$$\bar{F}_i'(x) = \sum_{i=1}^{d} \lambda_{ij} \bar{F}_j(x).$$

Since we have assumed all μ_i not having any atom at zero, $\bar{F}_i(0) = 1$ and the system of linear differential equations has the unique solution

$$\bar{\boldsymbol{F}}(x) = \begin{pmatrix} \bar{F}_1(x) \\ \bar{F}_2(x) \\ \vdots \\ \bar{F}_d(x) \end{pmatrix} = e^{\boldsymbol{S}x}\boldsymbol{e}.$$

Only the diagonal elements of $\boldsymbol{S}$ are negative, so there exists a constant a such that $\boldsymbol{M} = a\boldsymbol{I} + \boldsymbol{S}$ is a matrix with strictly positive entries. By the Perron–Frobenius theory, the matrix $\boldsymbol{M}$ has a real eigenvalue r which dominate the real parts of all other eigenvalues. Furthermore, there is a strictly positive eigenvector $\boldsymbol{\nu}$ which corresponds to r. By possibly normalizing the vector $\boldsymbol{\nu}$, we may assume that $\boldsymbol{\nu e} = 1$. Now consider $g(x) = \boldsymbol{\nu}\bar{\boldsymbol{F}}(x)$. By the Hahn–Jordan decomposition (Theorem 1.22) is follows that $\bar{F}_i = \nu_i([0,\infty)) \to 0$ as $x \to \infty$ (since this is a known property for non–negative measures) and hence $g(x) \to 0$ as $x \to \infty$. But since ν is eigenvector for $\boldsymbol{M}$ and hence for $\boldsymbol{S}$, we have that

$$g(x) = \boldsymbol{\nu} e^{\boldsymbol{S}x}\boldsymbol{e} = \boldsymbol{\nu e} e^{\kappa x},$$

where κ is the eigenvalue for $\boldsymbol{S}$ which corresponds $\boldsymbol{\nu}$. Since $g(x) \to 0$ as $x \to \infty$ we must have that $\kappa < 0$. Furthermore, the eigenvalue which corresponds to $\boldsymbol{M} = a\boldsymbol{I} + \boldsymbol{S}$ is $a + \kappa$. Since $a + \kappa$ is the eigenvalue with the largest real part among all eigenvalues for $\boldsymbol{M}$, we conclude that κ is the eigenvalue with the largest real part among all eigenvalues of $\boldsymbol{S}$. □

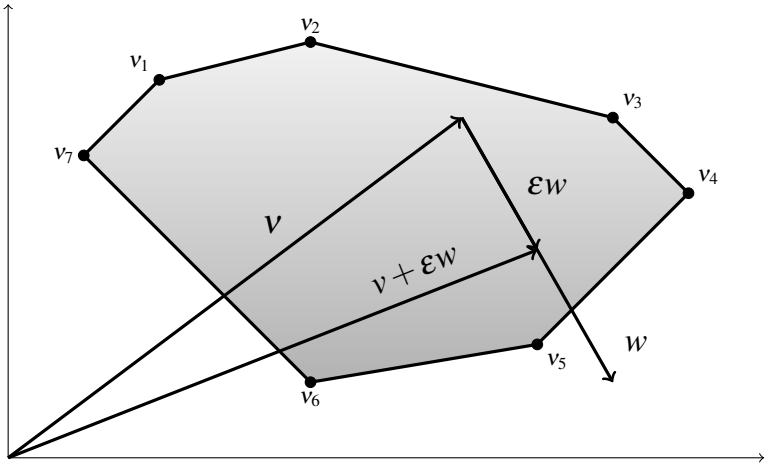

Fig. 7.1 The vector $\boldsymbol{w}$ points inwards to the set $P = \text{co}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, ..., \boldsymbol{\nu}_7)$ from the point $\boldsymbol{\nu}$ since by scaling it down to $\varepsilon\boldsymbol{w}$ and adding to $\boldsymbol{\nu}$ it remains in P.

Definition 7.5. Let P be a subset of the space of signed measures. Then we say that a vector $\boldsymbol{w}$ points inwards to P from a vector $\boldsymbol{\nu}$ if there exits a $\delta > 0$ so that $\boldsymbol{\nu} + \varepsilon\boldsymbol{w} \in P$ for all $0 < \varepsilon < \delta$.

Remark 7.1. Notice that if P is convex, then $\boldsymbol{\nu} + \varepsilon\boldsymbol{w} \in P$ for some ε implies that $\boldsymbol{\nu} + \varepsilon'\boldsymbol{w} \in P$ for all $0 < \varepsilon' < \varepsilon$ (see Figure 7.1, where the whole blue segment will be in P).

Remark 7.2. On a finite dimensional vector–space, the operator Γ may be represented by a matrix so the exponential operator $\exp(\Gamma x)$ is essentially a ma-

trix exponential function. Hence we have that the usual approximation formula $\lim_{n\to\infty}(1+ax/n)^n = \exp(ax)$ is valid also for matrices and hence for the operator Γ as well, and the convergence is understood to be in the $\|\cdot\|$–norm of the vector space under consideration.

Theorem 7.5. *Let $\boldsymbol{\nu}_1,...,\boldsymbol{\nu}_m$ be signed measures and let $P = co(\boldsymbol{\nu}_1,...,\boldsymbol{\nu}_m)$. Then P is R–invariant if and only if $\Gamma\boldsymbol{\nu}_i$ point inwards to P at $\boldsymbol{\nu}_i$ for $i = 1,2,...,m$.*

Proof. Suppose that P is R–invariant. Then there exists an intensity matrix $\boldsymbol{\Lambda} = \{\lambda_{ij}\}_{i,j=1,...,m}$ such that

$$\Gamma\boldsymbol{\nu}_i = \sum_{j=1}^{m} \lambda_{ij}\boldsymbol{\nu}_j$$

for $i = 1,2,...,m$. Let $\varepsilon = -\lambda_{ii}^{-1}$. Then $|\varepsilon\lambda_{ij}| \le 1$ and $\sum_{j=1}^{m} \varepsilon\lambda_{ij} = 0$ so

$$\boldsymbol{\nu}_i + \varepsilon \sum_{j=1}^{m} \lambda_{ij}\boldsymbol{\nu}_j$$

is a convex combination of $\boldsymbol{\nu}_1,...,\boldsymbol{\nu}_m$ and hence it belongs to P. This proves the first implication.

Conversely we assume that $\Gamma\boldsymbol{\nu}_i$ points inwards to P at $\boldsymbol{\nu}_i$ for $i = 1,2,...,m$. Since $\boldsymbol{\nu}_i + \varepsilon_i\Gamma\boldsymbol{\nu}_i \in P$ for some $\varepsilon_i > 0$, $i = 1,2,...,m$, then

$$\boldsymbol{\nu}_i + \varepsilon\Gamma\boldsymbol{\nu}_i \in P$$

for all $\varepsilon < \varepsilon_0 = \min(\varepsilon_1,...,\varepsilon_m)$ since P is convex (see Remark 7.1).

Now let $\boldsymbol{\nu} \in P$. Since P is a convex hull, there exists $c_1,...,c_m$ constants such that $c_1 + ... + c_m = 1$ and

$$\boldsymbol{\nu} = \sum_{i=1}^{m} c_i\boldsymbol{\nu}_i.$$

Then

$$\begin{aligned}\boldsymbol{\nu} + \varepsilon\Gamma\boldsymbol{\nu} &= \sum_{i=1}^{m} c_i\boldsymbol{\nu}_i + \varepsilon\sum_{i=1}^{m} c_i\Gamma\boldsymbol{\nu}_i \\ &= \sum_{i=1}^{m} c_i\left(\boldsymbol{\nu}_i + \varepsilon\Gamma\boldsymbol{\nu}_i\right),\end{aligned}$$

and since $\boldsymbol{\nu}_i + \varepsilon\Gamma\boldsymbol{\nu}_i \in P$ for all $0 < \varepsilon < \varepsilon_0$, $i = 1,2,...,m$ and since $c_1 + ... + c_m = 1$, then also $\boldsymbol{\nu} + \varepsilon\Gamma\boldsymbol{\nu} \in P$ for $0 < \varepsilon < \varepsilon_0$. This means that for a fixed $t \ge 0$, there exists an $n_0 \in \mathbb{N}$ such that

$$\left(\boldsymbol{I} + \frac{t}{n}\Gamma\right)\boldsymbol{\nu} = \boldsymbol{\nu} + \frac{t}{n}\Gamma\boldsymbol{\nu} \in P$$

for all $n \ge n_0$. Repeating the argument, we that for sufficiently large n that

$$\left(\boldsymbol{I} + \frac{t}{n}\Gamma\right)^n \boldsymbol{\nu} \in P.$$

This holds for all n. Thus in the $\|\cdot\|$–limit,

$$P \ni \lim_{n\to\infty} \left(\boldsymbol{I} + \frac{t}{n}\Gamma\right)\boldsymbol{\nu} = e^{\Gamma t}\boldsymbol{\nu} = R_t\boldsymbol{\nu}.$$

Hence P is R–invariant. □

We need quite a few lemma's and to review some material from convex analysis in order to be able to prove the characterization theorem.

Lemma 7.3. *Let $f, g : [0,\infty) \to \mathbb{R}$ be continuous functions such that*

$$\liminf_{x\to\infty} \frac{f(x)}{|g(x)|} > 0$$

and f being strictly positive. Then there exists an $\varepsilon > 0$ such that

$$f(x) \geq \varepsilon |g(x)|$$

for all $x \geq 0$.

Proof. Since $\liminf_{x\to\infty} \frac{f(x)}{|g(x)|} > 0$ there exists $\delta > 0$ such that

$$0 < \delta < \liminf_{x\to\infty} \frac{f(x)}{|g(x)|} = \lim_{x_0\to\infty} \inf_{x\geq x_0} \frac{f(x)}{|g(x)|},$$

so there exists a $x_0 \geq 0$ such that $f(x) \geq \delta|g(x)|$ when $x \geq x_0$.

Concerning $x \in [0,x_0]$ we proceed as follows. Let

$$m = \min_{x\in[0,x_0]} f(x),\ M = \min_{x\in[0,x_0]} |g(x)|$$

which do exist since f and g are continuous functions. Then $0 < m < \infty$ (f being strictly positive) and $0 \leq M < \infty$.

If $g(x) = 0$ for all $x \in [0,x_0]$, then $M = 0$, but then for any $M' > M = 0$ we have that $f(x) \geq f(x)/M'|g(x)| = 0$ on $[0,x_0]$. If $g(x) \neq 0$, $x \in [0,x_0]$ then $M > 0$ and

$$f(x) = \frac{f(x)}{|g(x)|}|g(x)| \geq \frac{m}{M}|g(x)| \geq \frac{m}{M'}|g(x)|$$

for any $M' > M$. Thus it holds in any circumstance that

$$f(x) \geq \frac{m}{M'}|g(x)|$$

for all $x \in [0,x_0]$ for some $M' > M$. Letting $\varepsilon = \min(\delta, m/M')$ we see that $f(x) \geq \varepsilon|g(x)|$ for $x \geq 0$. □

Lemma 7.4. *Let U be a finite–dimensional vector space of real–valued continuous functions on $[0,\infty)$ such that there is a strictly positive $f \in U$ with the property that*

$$\liminf_{x\to\infty} \frac{f(x)}{|u(x)|} > 0 \tag{7.2}$$

for all $u \in U$. Let $C = \{g \in U | g(x) \geq 0\}$. Then

(a) C is a convex cone and

$$g \in int(C) \iff \exists \varepsilon > 0 : g(x) \geq \varepsilon f(x)$$

for all $x \geq 0$.

(b) If ϕ is a linear function on U with $\phi(f)$, $\phi(g) \geq 0$ for $g \in C$, then with $V = \{u \in U : \phi(u) = 1\}$ we have that

$$g \in ri(C \cap V) \iff \exists \varepsilon > 0 : g - \varepsilon f \in C.$$

Proof. Let $g_1, g_2 \in C$. Then for any $\alpha, \beta > 0$ we have that $\alpha g_1 + \beta g_2 \in C$ as well. Hence C is a convex cone. Now assume that $g \in \text{int}(C)$. Then by Corollary 7.1 with $u = -f$ we have that there exists an $\varepsilon > 0$ such that $g + \varepsilon u = g - \varepsilon f \in C$, i.e. $g(x) \geq \varepsilon f(x)$ for all $x \geq 0$.

Conversely, assume that there is an $\varepsilon > 0$ such that $g(x) \geq \varepsilon f(x)$ for all $x \geq 0$. Now f satisfies that

$$\liminf_{x\to\infty} \frac{f(x)}{|u(x)|} > 0$$

for any $u \in U$. Since $f > 0$ and both f and u are assumed to be continuous, we have from Lemma 7.3, that there is an $\varepsilon_1 > 0$ such that $f(x) \geq \varepsilon_1 |u(x)|$ for all $x \geq 0$. In particular, $f(x) \geq \varepsilon_1 u(x)$ for all $x \geq 0$. Now let $u \in U$. Then

$$g(x) + \varepsilon_1 \varepsilon u(x) = (g(x) - \varepsilon f(x)) + \varepsilon \left(f(x) + \varepsilon_1 u(x)\right).$$

Now both terms of the right hand side are non–negative, so we conclude that there is an $\varepsilon_2 = \varepsilon\varepsilon_1 > 0$ such that

$$g(x) + \varepsilon_2 u(x) \in C$$

for all $u \in C$. By Corollary 7.1 this is equivalent to $g \in \text{int}(C)$.

Concerning the last part of the Lemma, first we notice that V is an non–empty ($f \in V$) affine set (ϕ being linear), so we have that $\text{ri}(V) = V$. Secondly, $\text{ri}(C) = \text{int}(C)$, C being an n–dimensional convex set. Thus $\text{ri}(C \cap V) = \text{ri}(C) \cap \text{ri}(V) = \text{int}(C) \cap V$. But an element of V is in $\text{int}(C)$ if and only if there exists an $\varepsilon > 0$ such that $g(x) \geq \varepsilon f(x)$ for all $x \geq 0$ by (a) above. □

We now consider a μ such that $V_R(\mu) = \text{span}\{R_t\mu : t \geq 0\}$ is of finite dimension. Then from Theorem 1.26 we know that the Laplace(–Stieltjes) transform of μ is a rational function. Let β denote the pole (root of the denominator polynomial) which has the largest real part among all the poles and let the order be d. Assume that there are b other poles β_i of orders d_i.

Then the density of μ (not necessarily a probability density) is given by

$$f_\mu(x) = r(x)e^{\beta x} + \sum_{i=1}^{b} r_i(x)e^{\beta_i x}$$

where $r(x)$ is a polynomial of order $d-1$, $r_i(X)$ polynomials of orders $d_i - 1$, β is real (negative) and larger than the real parts of all other β_i.

Lemma 7.5. *The generator $\Gamma\mu$ of the signed measure μ may be represented by*

$$\Gamma\mu = f_\mu(0)\delta_0 + \nu',$$

where f_μ is the density of the absolute continuous part of μ and ν' is an absolute continuous signed measure with density f'_μ, the right derivative of f_μ.

Proof. An atom at zero for a non–negative random variable is represented by its distribution function F being positive at 0. Now define $F(x) = \mu((-\infty, x])$. Then

$$\begin{aligned}\Gamma F(0) &= \lim_{x\downarrow\infty} \frac{R_t\mu((-\infty,0]) - \mu((-\infty,0])}{t} \\ &= \lim_{x\downarrow\infty} \frac{\mu((-\infty,t]) - \mu((-\infty,0])}{t} \\ &= f_\mu(0).\end{aligned}$$

Hence $\Gamma\mu$ has an atom at zero of size $f_\mu(0)$. Its absolute continuous part is given by a measure the density of which is

$$\lim_{t\downarrow 0} \frac{f_\mu(t+x) - f_\mu(x)}{t} = f'_\mu(x),$$

the right derivative of f_μ. □

Lemma 7.6. *The exponential distribution with intensity $-\beta$ is contained in $V_R(\mu)$.*

Proof. If we perform a partial fraction expansion of the Laplace–Stieltjes transform $\hat{\mu}(s)$ it is clear that a term proportional to $1/(s+\beta)$ will appear, and we conclude that the exponential distribution with intensity $-\beta$ is contained in $V_R(\mu)$. □

Definition 7.6. Let W be the sub–space of $V_R(\mu)$ given by

$$W = \{\nu \in V_R(\mu) \mid \hat{\nu}(s) \text{ has a pole at } \beta \text{ of order 1 or no pole at } \beta\}.$$

Lemma 7.7. *The exponential distribution with intensity $-\beta$ is contained in W.*

Proof. By a partial fraction expansion of $\hat{\nu}(s)$ for a measure ν which do have a pole at β, we conclude similarly as above that the exponential distribution with intensity $-\beta$ is contained in W. □

Definition 7.7. For a measure $\nu \in V_R(\mu)$ we denote by $\bar{\nu}$ the measure

$$\bar{\nu} = \frac{\nu - \nu(\{0\})\delta_0}{1 - \nu(\{0\})},$$

i.e. the measure obtained by removing a possible atom at zero an renormalize the measure to assign the total original mass to the absolute continuous part only.

Definition 7.8. We let $\mathscr{S}_+$ denote the subspace of $\mathscr{S}$ of the signed measures on $[0,\infty)$ which are probability measures and we let $\mathscr{S}_{\text{abs}}$ denote the subspace of signed measures which are absolute continuous and has no atom at zero.

Since β is the pole with maximum real part in $V_R(\mu)$ we then have that for all $\nu \in W \cap \mathscr{S}_{\text{abs}}$ that

$$\liminf_{x\to\infty} \frac{e^{\beta x}}{|f_\nu(x)|} > 0$$

where $f_\nu(x)$ denotes the absolute continuous density corresponding to ν.

Now we can let $W \cap \mathscr{S}_{\text{abs}}$ play the role of U in Lemma 7.4. Furthermore, let $f(x) = -\beta e^{\beta x}$ be the exponential distribution and the linear functional

$$\phi(g) = \int_0^\infty g(x)dx,$$

the total mass of the density g. Then V is the affine set of all densities which integrate to 1, and $C \cap V$ is the set of all non–negative densities which integrate to 1, i.e. probability densities. Hence $C \cap V$ corresponds to the set of absolute continuous probability measures within $U = W \cap \mathscr{S}_{\text{abs}}$. Then Lemma 7.4 (b) states that for a $g \in C \cap V \cap \mathscr{S}_{\text{abs}}$ (i.e. g being an absolute continuous probability density) satisfies

$$g \in \text{ri}(C \cap V \cap W \cap \mathscr{S}_{\text{abs}})$$

if and only if there is an $\varepsilon > 0$ such that

$$g(x) \geq \varepsilon e^{\beta x}$$

for all $x \geq 0$. We summarize this result in the following

Theorem 7.6. *$g \in ri(W \cap \mathscr{S}_+ \cap \mathscr{S}_{abs})$ if and only if there exists and $\varepsilon > 0$ such that*

$$g(x) \geq \varepsilon e^{\beta x}$$

for all $x \geq 0$.

We next move on to investigate what happens if we allow for an atom at zero as well. If $\nu \in W$ has an atom at zero and an absolute continuous part (normalized) $\bar{\nu}$, then we can write

$$\nu = p\delta_0 + (1-p)\bar{\nu}$$

for some $0 < p < 1$. But then $\nu \in \text{co}(\delta_0, \bar{\nu})$ by definition of convex hull. Since ν has an atom at 0, we conclude that $\{\delta_0\}$ is contained in $V_R(\mu)$ which is only possible

if μ itself has an atom part at 0, i.e. $0 < \mu(\{0\}) < 1$. Hence by Lemma 7.1 and $\nu \in \mathrm{co}(\delta_0, \bar{\nu})$ we have that a probability measure ν with an atom part at 0 satisfies $\nu \in \mathrm{ri}(W)$ if and only if $0 < \mu(\{0\}) < 1$ and its density g corresponding to its absolute (normalized) continuous parts satisfies that there exists an $\varepsilon > 0$ such that

$$g(x) \geq \varepsilon e^{\beta x}$$

for all $x \geq 0$. This is summarized in

Theorem 7.7. *A probability measure ν with an atom part at zero satisfies that $\nu \in ri(W \cap \mathscr{S}_+)$ if and only if $0 < \mu(\{0\}) < 1$ and its density g corresponding to its absolute (normalized) continuous parts satisfies that there exists an $\varepsilon > 0$ such that*

$$g(x) \geq \varepsilon e^{\beta x}$$

for all $x \geq 0$.

Corollary 7.2. *If g is the density corresponding to the absolute continuous part of a probability measure ν which belongs to the relative interior of W, then*

(1) $g(x) > 0$.
(2) The Laplace–Stieltjes transform of ν has pole of order 1 at β
(3) $\overline{R_t \nu}$ converges in supremum norm to the distribution (measure) of an exponential distribution with intensity $-\beta$.
(4) ν satisfies that its Laplace transform has a pole of maximum real part among all poles and that the density corresponding to its absolut continuous part is strictly positive everywhere on $[0, \infty)$.

Proof. (1) is obvious since $g(x) \geq \varepsilon e^{\beta x}$ for all $x \geq 0$.

That ν belongs to (in fact the interior) of W means that its density of the absolute continuous part is given by

$$g(x) = r(x)e^{\beta x} + \sum_{i=1}^{b} r_i(x)e^{\beta_i x},$$

where $r_i(x)$ are polynomials but $r(x)$ is the constant polynomial, $r(x) = r_0$ say. Now since β is the eigenvalue with largest real part and since it is negative, all other terms $r_i(x)e^{\beta_i x}$ goes to zero faster than $r_0 e^{\beta x}$ as $x \to \infty$. Hence by $g(x) \geq \varepsilon e^{\beta x}$ we must have that r_0 is positive since otherwise the density will be negative eventually. Thus ν will have a pole at β which is obviously of order 1 since $\nu \in W$. This proves (2).

Concerning (3), the density of the absolute continuous part of $R_t \nu$ is $g(x+t)$ which amounts to

$$g(x+t) = r_0 e^{\beta t} e^{\beta x} + \sum_{i=1}^{b} r_i(x+t)e^{\beta_i (x+t)}.$$

Normalizing so it integrates to 1, for large t the normalizing factor is approximately $ce^{\beta t}$ for some constant, since this term dominates the rest of the terms as $t \to \infty$.

Hence the normalized density of the absolute continuous part of $\overline{R_t\nu}$ converges point–wise to $-\beta e^{\beta x}$ (since the limit must integrate to 1 as well, the constant in front of $e^{\beta x}$ must be $-\beta$). Point–wise convergence implies convergence in norm in a finite–dimensional vector space, so we have that $R_t\nu$ converges in supremum norm to the exponential distribution with rate $-\beta$.

Concerning (4), strict positivity follows from Theorems 7.6 and 7.7. Any pole of the Laplace transform of ν is also a pole for μ and vice versa. □

Now consider the density of the absolute continuous part of μ, $f_\mu(x)$. Then it might happen that $f_\mu(0) = 0$. Thus if we make a Taylor expansion of f_μ about 0, then the constant term will be zero, i.e.

$$f_\mu(x) = \sum_{j=\omega}^{\infty} a_i x^i, \;\; x \geq 0,$$

for some $\omega \geq 1$.

Definition 7.9. The largest ω for which

$$f_\mu(x) = \sum_{j=\omega}^{\infty} a_i x^i, \;\; x \geq 0,$$

is called the order of the zero of f_μ at zero.

Lemma 7.8. *Let $\mu \in V_R(\mu)$ be a probability distribution on $[0,\infty)$ such that the vector–space is finite–dimensional, and suppose that the density of the absolute continuous part of μ, f_μ, is strictly positive on $(0,\infty)$ and that the pole of maximal real part at β is real. Let ω be the order of the zero of f_μ at zero. Then*

1. *If $\omega > 0$ then the density of the absolute continuous part corresponding to $\mu + \varepsilon\Gamma\mu$ has a zero of order $\omega - 1$ at zero for any $\varepsilon \neq 0$ (not necessarily positive).*
2. *For all ε sufficiently small,*

$$0 < \lim_{x\to\infty} \frac{f_\mu(x)}{f_\mu(x) + \varepsilon f'_\mu(x)} < \infty.$$

3. *If $f_\mu(0) > 0$, then density of the absolute continuous part of $\mu + \varepsilon\Gamma\mu$ is also positive at zero for all sufficiently small $\varepsilon > 0$.*
4. *For all sufficiently small ε, $\mu + \varepsilon\Gamma\mu$ is a probability measure.*
5. *For all sufficiently small ε, the density of the absolute continuous part of $\mu + \varepsilon\Gamma\mu$ is strictly positive on $(0,\infty)$, and its Laplace transform has a pole of maximum real part at β.*

Proof. 1. First we may split the density of μ into $f_\mu(0)\delta_0$ (atom at zero) and the absolute continuous part $f_\mu(x)$. Recall that $\Gamma\mu = \lim_{t\to 0} \frac{R_t\mu - \mu}{t}$, so in terms of densities, the density corresponding to $\Gamma\mu$ is evidently

$$\lim_{t\downarrow 0} \frac{f_\mu(t+x) - f_\mu(x)}{t} = f'_\mu(x),$$

where f' denote the right derivative of f_μ. But

$$f'_\mu(x) = \sum_{j=\omega-1}^{\infty} (j+1)a_{j+1}x^j$$

so the density for the absolute continuous part of $\mu + \varepsilon\Gamma\mu$, $f_\mu(x) + \varepsilon f'_\mu(x)$, has a zero at 0 of order $\omega - 1$ at

2. Since

$$f_\mu(x) = r(x)e^{\beta x} + \sum_{i=1}^{b} r_i(x)e^{\beta_i x}$$

where $r(x)$ is a polynomial of order d, then

$$f'_\mu(x) = r'(x)e^{\beta x} + r(x)\beta e^{\beta x} + \sum_{i=1}^{b} r'_i(x)e^{\beta_i x} + \sum_{i=1}^{b} \beta_i r_i(x)e^{\beta_i x}$$

and hence

$$\frac{f'_\mu(x)}{f_\mu(x)} \to \beta$$

as $x \to \infty$ since $0 > \beta > \beta_i$ is the dominating exponent and since the order of r' is one less than that of r. Hence

$$\frac{f_\mu(x)}{f_\mu(x) + \varepsilon f'_\mu(x)} = \frac{1}{1 + \varepsilon f'_\mu(x)/f_\mu(x)} \to \frac{1}{1+\varepsilon\beta},$$

and the latter is finite whenever $\varepsilon\beta \neq 1$ and finite and positive whenever $|\varepsilon| < -1/\beta$. This proves 2.

3. If $f_\mu(0) > 0$, then $f_\mu(0) + \varepsilon f_\mu(0) > 0$ for ε sufficiently small regardless of the values of $f'_\mu(0)$.

4. First we use that

$$\frac{f'_\mu(x)}{f_\mu(x)} \to \beta$$

as $x \to \infty$, and since β is negative

$$\frac{f_\mu(x)}{|f'_\mu(x)|} \to -\frac{1}{\beta},$$

as $x \to \infty$. Thus if we take a $0 < \delta < -1/\beta$, then there exists an x_0 such that

$$f_\mu(x) \geq \delta |f'_\mu(x)|$$

for all $x \geq x_0$.

If $f_\mu(0) = 0$, then

$$f_\mu(x) = \sum_{j=\omega}^{\infty} a_j x^j$$

for some $\omega > 0$. Since $f_\mu(x)$ is non–negative, then $a_\omega > 0$ since otherwise the density would become negative at very small values for x since x^ω is the term which approaches 0 slowest among the potences. Now the expansion for f'_μ is given by

$$f'_\mu(x) = \sum_{j=\omega}^{\infty} j a_j x^{j-1}.$$

The first non–zero term is $\omega a_\omega > 0$. Hence there exists an $x_1 > 0$ such that $f'_\mu(x) \geq 0$ for $0 \leq x \leq x_1$ with a similar argument as to which $a_\omega > 0$. If $f_\mu(0) > 0$ we let $x_1 = 0$.

Now we consider the behavior of f_μ in the interval $[x_1, x_0]$. If $f_\mu(0) = 0$, then $x_1 > 0$ and since f_μ is strictly positive on $(0,\infty)$ then the minimum m of f_μ over $[x_1, x_0]$ is strictly positive as well. If $f_\mu(0) > 0$ then $x_1 = 0$ and since f_μ is strictly positive on $(0,\infty)$, then the minimum m of f_μ over $[x_1, x_0]$ is again positive.

Let $M = \max\{|f'_\mu(x)| : x \in [x_1, x_0]\}$. Take any $M_1 > M$. This ensures that $M_1 > 0$ since M might be zero. Then on $[x_1, x_0]$ we have that

$$\frac{f_\mu(x)}{f'_\mu(x)} \geq \frac{m}{M_1}$$

or

$$f_\mu(x) \geq \frac{m}{M_1} |f'_\mu(x)|$$

for $x \in [x_1, x_0]$.

Since $f'_\mu(x) \geq 0$ on $[0, x_1]$ and since f_μ is non–negative over all $[0,\infty)$, we have that $f_\mu + \varepsilon f'_\mu$ is non–negative over $[0, x_1]$. If we take $0 < \varepsilon < \min(\delta, m/M_1)$, then

$$f_\mu(x) \geq \varepsilon |f'_\mu(x)|$$

for all $x \geq x_1$, but this in turn implies that

$$f_\mu + \varepsilon f'_\mu(x) \geq 0$$

for all $x \geq 0$. Also,

$$\int_0^\infty \left(f_\mu(x) + \varepsilon f'_\mu(x)\right) dx = \int_0^\infty f_\mu(x) dx - \varepsilon f_\mu(0)$$

since $f_\mu(+\infty) = 0$. So if $\varepsilon < \min(\delta, m/M_1)$ is sufficiently small, then we may subtract $\varepsilon f_\mu(0)$ from the absolute continuous parts of $\mu + \varepsilon\Gamma\mu$ and add to the atom at zero. Hence $\mu + \varepsilon\Gamma\mu$ is a probability measure.

5. The density of the absolute continuous part of $\mu + \varepsilon\Gamma\mu$, $f_\mu(x) + \varepsilon f'_\mu(x)$, is a probability density with a possible atom at zero. This means that $f_\mu(x) + \varepsilon f'_\mu(x) \geq 0$ for all $x \in [0,\infty)$ and since $f_\mu(x) > 0$ for all $x \in (0,\infty)$, we have that $f_\mu(x) + \varepsilon_1 f'_\mu(x) > 0$ for $\varepsilon_1 < \varepsilon$.

The poles of the Laplace transforms of f_μ and f'_μ coincide since $\mathscr{L}(f'_\mu,s) = s\mathscr{L}(f_\mu,s) - f_\mu(0)$. Hence also the Laplace transform of $f_\mu + \varepsilon f'_\mu$ and f_μ have the same poles. This proves 5. □

We now extend Theorem 7.6 from the case of characterizing the relative interior of the signed measures which have Laplace transforms with poles of order at most 1 at β to the general case where the root of the Laplace transform at β is a most d.

Theorem 7.8. *Let $\nu \in V_R(\mu)$. Then $\nu \in ri(V_R(\mu) \cap \mathscr{S}_+ \cap \mathscr{S}_{abs})$ if and only if there exists an $\varepsilon > 0$ such that*

$$f_\nu(x) \geq \varepsilon(1 + x^{d-1})e^{\beta x}$$

for all $x \geq 0$.

Proof. The proof is similar to that of Theorem 7.6, replacing $f(x) = e^{\beta x}$ with $f(x) = (1 + x^{d-1})e^{\beta x}$. The reason for not using $x^{d-1}e^{\beta x}$ instead, is that we obtain a density which is positive at zero which will be used later on. □

The corresponding result to Theorem 7.7 allowing for an atom at zero is

Theorem 7.9. *Let $\nu \in V_R(\mu)$. Then ν is contained in the relative interior of the probability measures of $V_R(\nu)$ if and only if $0 < \mu(\{0\}) < 1$ and its density f_ν of the absolute continuous part satisfies for some $\varepsilon > 0$,*

$$f_\nu(x) \geq \varepsilon(1 + x^{d-1})e^{\beta x}$$

for all $x \geq 0$.

Proof. Similar to that of Theorem 7.7 and Theorem 7.8 □

R–invariance is the main vehicle for proving the characterization theorem, so we shall note the following sets are *R*–invariant.

Theorem 7.10. *The sets*

1. $V_R(\mu)$
2. W
3. $\mathscr{S}_+$
4. $ri(V_R(\mu) \cap \mathscr{S}_+)$
5. $ri(W \cap \mathscr{S}_+)$

are all R–invariant.

Proof. $V_R(\mu)$ is invariant by definition. If $\nu \in W$, then μ has a Laplace transform with a pole at β of order at most 1, and so does $R_t\mu$. Hence W is also *R*–invariant. The case of $\mathscr{S}_+$ is obvious. Concerning 4., if $\nu \in \mathrm{ri}(V_R(\mu) \cap \mathscr{S}_+)$ if and only if the absolute continuous part f_ν of ν satisfies that

$$f_\nu(x) \geq \varepsilon(1 + x^{d-1})e^{\beta x}$$

for some $\varepsilon > 0$ and all $x \geq 0$, but then in particular,

$$R_t f_\nu(x) \geq \varepsilon(1 + (x+t)^{d-1}e^{\beta(x+t)} \geq \varepsilon e^{\beta t}(1 + x^{d-1})e^{\beta x} = \varepsilon'(1 + x^{d-1})e^{\beta x}$$

for all $x \geq 0$, where $\varepsilon' = \varepsilon e^{\beta t}$, but this in turn is equivalent to $R_t \nu \in \mathrm{ri}(V_R(\mu) \cap \mathscr{S}_+)$. Hence $\mathrm{ri}(V_R(\mu) \cap \mathscr{S}_+)$ is R–invariant. The proof of $\mathrm{ri}(W \cap \mathscr{S}_+)$ is similar. □

The affine hull of $W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}}$ consists of all linear combination $\sum_i a_i x_i$ where $a_1 + ... + a_\ell = 1$ and $x_1, ..., x_\ell \in W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}}$, and it is obvious that if no x_i has an atom at zero, then no affine combination will, and if all x_i have an Laplace transform with a pole at β of maximum order 1, then so will any affine combination. However, the affine combination of probability measures is not a probability measure but a signed measure with a unit mass. Thus

$$\mathrm{aff}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}}) = W \cap \mathscr{S}_{\mathrm{abs}} \cap S_1$$

where S_1 is the set of signed measures on $[0, \infty)$ with unit mass, i.e.

$$S_1 = \{\nu \in \mathscr{S} : \nu([0, \infty)) = 1.$$

The exponential distribution with intensity $-\beta$ trivially satisfies

$$-\beta e^{\beta x} \geq \varepsilon e^{\beta x}$$

for some $\varepsilon > 0$ and all x, so by Theorem 7.6 we have that the exponential distribution with intensity $-\beta$ is contained in $\mathrm{ri}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$.

Recall that μ has a pole of maximum real part at $\beta < 0$ of multiplicity d and b other poles at β_i, $i = 1, 2, ..., b$ with dimensionality d_i. Thus

$$f_\mu(x) = r(x) e^{\beta x} + \sum_{i=1}^{b} r_i(x) e^{\beta_i x},$$

where $r(x)$ is a polynomial of order $d-1$ and $r_i(x)$ are polynomials of orders $d_i - 1$, $i = 1, 2, ..., b$. The dimension of $V_R(\mu)$ is $k = d + d_1 + ... + d_b$, and so is the dimension of $V_R(\mu) \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}}$. The dimension of $\mathrm{aff}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$ is hence $(k - d + 1) - 1 = k - d$.

Choose

$$\mu_1, ..., \mu_\ell \in \mathrm{ri}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}}) \;\; \ell \geq k - d \tag{7.3}$$

such that $\mathrm{ri}(\mathrm{co}(\mu_1, ..., \mu_\ell))$ is an open subset of $\mathrm{aff}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$ (implying that $\mathrm{ri}(\mathrm{co}(\mu_1, ..., \mu_\ell))$ is having full dimension relative to the affine hull) and such that it also contains the exponential distribution with intensity $-\beta$. This is indeed possible as we may just let one of the μ_is correspond to the exponential distribution.

Now $\mu_i \in \mathrm{ri}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$ implies by Corollary 7.2, item (4), that μ_i has a pole at β of maximum real part and a strictly positive density. By Lemma 7.8, 5., for all sufficiently small ε, we have that also $\mu + \varepsilon \Gamma \mu$ has a pole at β of maximum real part and a density for its absolute continuous part which is strictly positive. Now choose an $\varepsilon > 0$ sufficiently small so that all

$$\mu_i^* := \mu_i + \varepsilon \Gamma \mu_i, \; i = 1, 2, ..., \ell$$

defines measures with poles of maximum real part at β and densities f_i^* being positive everywhere on $[0,\infty)$ (also at zero, see Corollary 7.2). By Lemma 7.8, the densities f_i^* are asymptotically equivalent to $e^{\beta x}$ ($0 < \lim_{x\to\infty} f_i^*(x)/e^{\beta x} < \infty$). In particular,

$$\liminf_{x\to\infty} \frac{f_i^*(x)}{e^{\beta x}} > 0$$

so by Lemma 7.3 there exists an ε_1 such that

$$f_i^*(x) \geq \varepsilon_1 e^{\beta x}$$

for all $x \geq 0$. Since $\mu_i \in \mathrm{ri}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$then the by Theorem 7.6 their densities are positive at zero and hence by Lemma 7.5 the $\Gamma\mu$ must have atoms at zero and since μ_is are probability measures these atoms must lie between 0 and 1, both limits excluded. By Lemma 7.8, 3. and 5., it the follows that μ_i^* is also positive at 0 and its density strictly positive everywhere. Thus its atom must be between zero and 1 as well and from Theorem 7.7 we then conclude that $\mu_i^* \in \mathrm{ri}(W \cap \mathscr{S}_+)$.

Now $\mathrm{ri}(W \cap \mathscr{S}_+)$ is R–invariant so $R_t\mu_i \in \mathrm{ri}(W \cap \mathscr{S}_+)$ for all $t \geq 0$. Then the absolute continuous part $\overline{R_t\mu_i} \in \mathrm{ri}(W \cap \mathscr{S}_+ \cap \mathscr{S}_{\mathrm{abs}})$. But then by Corollary 7.2 $\overline{R_t\mu_i}^*$ converges in supremum norm to the exponential distribution with intensity $-\beta$ as $t \to \infty$, and since $\mathrm{ri}(\mathrm{co}(\mu_1,...,\mu_\ell))$ contains the exponential distribution with intensity $-\beta$ and since it is an open subset of the affine space $\mathrm{aff}(W \cap \mathscr{S}_+ \cap \mathscr{S}_1)$ in which all $R_t\mu_i^*$ belongs, we conclude that there exists a T such that $\overline{R_T\mu_i}^* \in \mathrm{ri}(\mathrm{co}(\mu_1,...,\mu_\ell))$.

Since the μ_i^* have atoms at zero then so do $R_T\mu_i^*$. The operator preserves that $R_T\mu_i^*$ is a again a probability measure, so the atom must again be between 0 and 1, so by Theorem 7.7 we conclude that $R_T^*\mu_i^* \in \mathrm{ri}(\mathrm{co}(\delta_0,\mu_1,...,\mu_\ell))$. This holds for all $i = 1,...,\ell$ and also for all $R_t\mu_i^*$ with $t \geq T$. We collect these results in

Theorem 7.11. *There exists an $\varepsilon > 0$ such that*

$$\mu_i^* = \mu_i + \varepsilon\Gamma\mu_i,\ i = 1,2,...,\ell$$

are all contained in $ri(W \cap \mathscr{S}_+)$. Furthermore, for t sufficiently large, $t \geq T$ say, we have that

$$R_t\mu_i^* \in ri(co(\delta_0,\mu_1,...,\mu_\ell))$$

for all $t \geq T$ and $i = 1,2,...,\ell$.

Next we consider what happens for $t \leq T$. Since $R_t\mu_i^* \in \mathrm{ri}(W \cap \mathscr{S}_+)$ for all t and since $t \to R_t\mu_i^*$ are continuous (follows from the continuity of the (signed) measures and the definition of R_t), then the images of the compact set $[0,T]$, $\{R_t\mu_i^* : t \in [0,T]\}$, are again compact sets. Furthermore, since all $R_t\mu_i^* \in \mathrm{ri}(W \cap \mathscr{S}_+)$, then they are also subsets of $\mathrm{ri}(W \cap \mathscr{S}_+)$. We have proved

Lemma 7.9. *$\{R_t\mu_i^* : t \in [0,T]\}$ are compact subsets of $ri(W \cap \mathscr{S}_+)$ for all $i = 1,2,...,\ell$.*

Lemma 7.10. *Let ν be a signed measure. Then in terms of right derivatives, we have that*

$$\frac{d}{dt}(R_t\nu) = \Gamma(R_t\mu) \quad t \geq 0.$$

Proof.

$$\begin{aligned}\frac{d}{dt}R_t\nu &= \lim_{h\downarrow 0}\frac{R_{t+h}\nu - R_t\nu}{h}\\ &= \lim_{h\downarrow 0}\frac{R_hR_t\nu - R_t\nu}{h}\\ &= \Gamma(R_t\mu).\end{aligned}$$

□

In studying the trajectories of $R_t\mu$ as a function of t we could use the direction by which they move according to derivatives $\Gamma R_t\mu$. We shall now prove that it is not necessary to keep track of all derivatives but only a finite amount of Euler approximants.

We define approximant,

$$\mu_{ij}^* = \left(\boldsymbol{I} + \frac{T}{n}\Gamma\right)^j \mu_i^* = \left(\boldsymbol{I} + \frac{\frac{T}{n}j\Gamma}{j}\right)^j \mu_i^*$$

for $i = 1,...,\ell$ and $j = 1,2,...,n$ for some fixed n. Hence for n and j large, μ_{ij}^* is close to $\exp(\Gamma jT/n\mu_i^*) = R_{jT/n}\mu_i^*$. In particular, with $j = n$ we approximate $R_T\mu_i^*$. We suppress the dependency on T and n by μ_{ij}^* as they shall be considered fixed.

Lemma 7.11. *There exists a number n such that*

$$\mu_{ij}^* \in ri(W \cap \mathscr{S}_+)$$

for all $i = 1,2,...,\ell, j = 0,1,...,n$ *and*

$$\mu_{in}^* \in ri(co(\delta_0, \mu_1, ..., \mu_\ell))$$

for all $i = 1,2,...,\ell$.

Proof. First we notice that $\mu_{ij} \in \mathscr{S}_1$, i.e. they have total mass $\nu_{ij}([0,\infty)) = 1$ (but they are no necessarily probability measures). To this end notice that since μ_i^* are probability measures and $\Gamma\mu_i^*([0,\infty)) = 0$ by definition of Γ, then

$$\mu_{ij}^*([0,\infty)) = \left(\boldsymbol{I} + \frac{T}{n}\Gamma\right)^j \mu_i^*([0,\infty)) = \mu_i^*([0,\infty)) = 1$$

since by expanding $\left(\boldsymbol{I} + \frac{T}{n}\Gamma\right)^j$, all terms but the $\boldsymbol{I}$ involve potences of Γ and hence their contributions are zero. Hence $\mu_{ij}^* \in \mathscr{S}_1$.

All μ_i^* are probability measures in W (see Lemma 7.11). Since W is a closed and R–invariant vector space, then μ_{ij}^* are contained in W as well. Thus $\mu_{ij}^* \in W \cap \mathscr{S}_1$.

The μ_{ij}^* may approximate arbitrarily close $R_{jT/n}\mu_i^*$. Now by Lemma 7.9 $\{R_t\mu_i^* : t \in [0,T]\}$ is a compact set in $\mathrm{ri}(W \cap \mathscr{S}_+) \subset W \cap \mathscr{S}_1$, and since $jT/n \in [0,T]$ for $j = 1,2,...,n$, we may choose n large enough to ensure that

$$\mu_{ij}^* \in W \cap \mathscr{S}_+$$

for all $i = 1,2,...,\ell$ and $j = 0,1,...,n$.

As mentioned earlier, when we choose $j = n$, then μ_{in}^* approximates $R_T\mu_i^*$. Since the latter is contained in $\mathrm{ri}(\mathrm{co}(\delta_0,\mu_1,\mu_2,...,\mu_\ell))$ (see Theorem 7.11) and the latter being a subset of $W \cap \mathscr{S}_1$, we conclude that for a sufficiently large n, μ_{in} will also be in $\mathrm{ri}(\mathrm{co}(\delta_0,\mu_1,\mu_2,...,\mu_\ell))$. □

Definition 7.10. With the n chosen large enough to satisfy Lemma 7.11, we define the following set

$$\mathscr{P}^* = \mathrm{co}\left(\left\{\bar{\mu}_{ij}^*, i = 1,2,...,\ell, j = 0,1,...,n-1\right\} \cup \{\delta_0,\mu_1,...,\mu_\ell\}\right).$$

Theorem 7.12. *$\mathscr{P}^*$ is R–invariant.*

Proof. We shall prove the theorem by means of Theorem 7.5. First notice that $\mu_{ij}^* \in \mathscr{P}^*$ since the μ_{ij}^*s are convex combinations of δ_0 and $\bar{\mu}_{ij}^*$. Also $\mu_{in}^* \in \mathscr{P}^*$ by the second result of Lemma 7.11.

Next we prove that Γ point inwards at $\mu_1,...,\mu_\ell$. But this follows from the construction of μ_{ij}^* since

$$\mu_i + \varepsilon\Gamma\mu_i = \mu_i^* = \mu_{i0}^* \in \mathscr{P}^*.$$

Now consider the points $\bar{\mu}_{ij}^*$, $i = 1,2,...,\ell$, $j = 0,1,...,n-1$. Since

$$\mu_{i,j+1}^* = \left(\boldsymbol{I} + \frac{T}{n}\Gamma\right)\mu_{ij}^*$$

all are in $\mathscr{P}^*$, then we see that Γ point inwards to $\mathscr{P}^*$ at μ_{ij}^* for all $i = 1,2,..,\ell$, $j = 0,1,...,n-1$.

We next prove that then also Γ point inwards to $\mathscr{P}^*$ at $\bar{\mu}_{ij}^*$. First we notice that $\mathscr{P}^*$ is closed under the $\bar{}$ –operation on the measures (removing atom and normalizing) since all elements but δ_0 are atom–free. Define atom weights $a_{ij} = \mu_{ij}^*(0)$ and $b_{ij} = \Gamma\mu_{ij}^*(0)$ for $i = 1,2,...,\ell$, $j = 0,1,...,n-1$. Then $0 < a_{ij} < 1$ ($\mu_{ij}^* \in \mathscr{S}_+$ so they are probability measures and their atom weights are less than one consequently) and $b_{ij} \geq 0$ since by Lemma 7.5 b_{ij} coincide with the value of the density of $\Gamma\mu_{ij}^*$ at zero. Choose ε_0 such that $a_{ij} + \varepsilon_0 b_{ij} < 1$. Then using that $\mu_{ij}^* = (1-a_{ij})\bar{\mu}_{ij}^* + a_{ij}\delta_0$ we get

$$\begin{aligned}\overline{\mu_{ij}^* + \varepsilon_0\Gamma\mu_{ij}^*} &= \frac{\mu_{ij}^* + \varepsilon_0\Gamma\mu_{ij}^* - (a_{ij} + \varepsilon_0 b_{ij})}{1 - a_{ij} - \varepsilon_0 b_{ij}} \\ &= \frac{(1-a_{ij})\bar{\mu}_{ij}^* + a_{ij}\delta_0 + \Gamma\left((1-a_{ij})\bar{\mu}_{ij}^* + a_{ij}\delta_0\right) - (a_{ij} + \varepsilon_0 b_{ij})}{1 - a_{ij} - \varepsilon_0 b_{ij}}\end{aligned}$$

$$= \frac{(1-a_{ij})\left(\bar{\mu}^*_{ij} + \varepsilon_0 \Gamma \bar{\mu}^*_{ij}\right) - b_{ij}\varepsilon_0 \delta_0}{1 - a_{ij} - \varepsilon_0 b_{ij}}. \quad (\Gamma \delta_0 = 0)$$

Thus

$$\bar{\mu}^*_{ij} + \varepsilon_0 \Gamma \bar{\mu}^*_{ij} = \frac{1 - a_{ij} - \varepsilon_0 b_{ij}}{(1-a_{ij})} \overline{\mu^*_{ij} + \varepsilon_0 \Gamma \mu^*_{ij}} + \frac{b_{ij}\varepsilon_0}{1-a_{ij}} \delta_0$$

and since Γ point inwards to $\mathscr{P}^*$ at μ^*_{ij}, then $\mu^*_{ij} + \varepsilon_0 \Gamma \mu^*_{ij} \in \mathscr{P}^*$ and since $\mathscr{P}^*$ is closed under $\bar{}$–operations then also $\overline{\mu^*_{ij} + \varepsilon_0 \Gamma \mu^*_{ij}} \in \mathscr{P}^*$. Since $a_{ij} + \varepsilon_0 b_{ij} < 1$, then $\varepsilon_0 b_{ij}/(1-a_{ij}) \in (0,1)$, so the above linear combination is in fact a convex combination and hence we conclude that

$$\bar{\mu}^*_{ij} + \varepsilon_0 \Gamma \bar{\mu}^*_{ij} \in \mathscr{P}^*.$$

Hence Γ point inwards to $\mathscr{P}^*$ at $\bar{\mu}^*_{ij}$. That Γ point inwards to $\mathscr{P}^*$ from δ_0 is trivial. Hence we have proved that Γ point inwards from all points of the convex hull, and hence it follows that the convex hull is indeed R–invariant. □

In the construction above it was necessary to augment the set $\mathrm{co}(\delta_0, \mu_1, ..., \mu_\ell)$ with the $\bar{\mu}^*_{ij}$ in order to make the resulting space point inwards at all its points which generate the convex hull since this, in turn, is equivalent to the set being R–invariant. The R–invariance is exactly the key to phase–type property.

Theorem 7.13. *If $\nu \in \mathscr{P}^*$, then ν is the distribution of a phase–type distribution.*

Proof. This follows straightaway from Theorem 7.4. □

Corollary 7.3. *A probability measure $\mu \in ri(W \cap \mathscr{S}_+ \cap \mathscr{S}_{abs})$ is of phase–type.*

Proof. When choosing the μ_i's (in (7.3)) we could let μ be any of the μ_is and hence μ_i is of phase–type since it belongs to $\mathscr{P}^*$. □

Though being an important start, $\mathscr{P}^*$ is however not large enough for a general characterization since we are restricted to the sub–space W. The idea is now to augment this space in further step in order to construct a convex hull of a finite number of measures which indeed do cover the general situation.

We now consider a general $\mu \in V_R(\mu)$ with a pole at β of order d. Assume that the density of the absolute continuous part of μ, f_μ, has a zero of order ω at zero, i.e. f_μ has a Taylor expansion about 0 and the first non–zero term in the expansion is of order ω (see Definition 7.9). Next recall, that the density of the absolute continuous part of $\mu + \varepsilon \Gamma \mu$ is given by $f_\mu + \varepsilon f'_\mu$. By Lemma 7.8, (2.) and (5.), we may choose an $\varepsilon_1 > 0$ such that $\mu^{(1)} = \mu + \varepsilon_1 \Gamma \mu$ satisfies

$$0 < \lim_{x \to \infty} \frac{f_\mu(x)}{f_\mu(x) + \varepsilon_1 f'_\mu(x)} < \infty$$

and the density of $\mu + \varepsilon_1 \Gamma \mu$, $f_\mu + \varepsilon_1 f'_\mu$, is strictly positive on $(0,\infty)$ and its Laplace transform has a pole of maximum real part at β. The density $f_\mu + \varepsilon_1 f'_\mu$ obviously

has a rational Laplace transform and we may apply Lemma 7.8 again to the measure $\mu + \varepsilon_1 \Gamma\mu$. We do this $\omega + 1$ times. Thus formally we let

$$\begin{aligned} \mu^{(0)} &= \mu \\ \mu^{(i)} &= \mu^{(i-1)} + \varepsilon_i \Gamma \mu^{(i-1)} \quad i = 1, 2, ..., \omega + 1, \end{aligned} \tag{7.4}$$

where the ε_is are chosen such that the conditions and hence results of Lemma 7.8 (2.) and (5.) are satisfied. We define

$$\mu^* = \mu^{(\omega+1)}.$$

We furthermore choose $\varepsilon_{\omega+1} > 0$ so small that also Lemma 7.8 (3.) is invoked. Indeed, since by construction $\mu^{(i)}$ has a zero of order $\omega - i$, $i = 1, 2, .., \omega$, then $\mu^{(i)}$ has a density which is positive at zero (by definition of the order of a zero) and hence by Lemma 7.8 (3.) we may choose ε_ω so small that apart from properties (2.) and (5.) of the same Lemma, we also have that μ^* has an atom at zero (strictly positive point mass).

We let f^* denote the density of μ^*. All the densities of the $\mu^{(i)}$'s have equivalent tails by Lemma 7.8 (2.), so they all have tails equivalent to μ itself and the order of the tail of μ is $x^{d-1}e^{\beta x}$, which is the dominant term. Iterating Lemma 7.8 (2.) we conclude that

$$0 < \lim_{x\to\infty} \frac{f_\mu(x)}{f^*(x)} < \infty$$

so in particular

$$\liminf_{x\to\infty} \frac{f^*(x)}{f_\mu(x)} > 0$$

which implies that

$$\liminf_{x\to\infty} \frac{f^*(x)}{(x^{d-1}+1)e^{\beta x}} > 0$$

implying that there exists an $\varepsilon > 0$ (Lemma 7.3) such that

$$f^*(x) \geq \varepsilon(x^{d-1}+1)e^{\beta x}.$$

Since f^* is positive ar zero, and hence implies an point mass there, then by Theorem 7.9 we conclude that $\mu^* \in \mathrm{ri}(V_R(\mu) \cap \mathscr{S}_+)$. We have proved

Theorem 7.14.

$$\mu^* \in ri(V_R(\mu) \cap \mathscr{S}_+).$$

Definition 7.11.

$$\mathscr{P}_0 = \mathrm{co}\left(\mathscr{P}^* \cup \{\tau_2, \tau_3, ..., \tau_d\}\right),$$

where τ_i are the distributions (measures) corresponding to the Erlang$(i, -\beta)$ distributions, i.e. the density of τ_i is given by $x^{i-1}e^{\beta x}/(i-1)!$.

Theorem 7.15. *There exists a N such that $R_t\mu^* \in ri(\mathscr{P}_0)$ for $t \geq N$.*

Proof. f^* is strictly positive and since it has pole of maximum real part at β is must be on the form

$$f^*(x) = r^*(x)e^{\beta x} + h(x),$$

where $r^*(x)$ is a polynomial of order $d-1$, and

$$h(x) = \sum_{i=1}^{b} r_i^*(x)e^{\beta_i x}$$

for some polynomials $r_i^*(x)$, $i = 1,2,...,d$. The latter expression for h follows easily from the fact that the densities of some ν and $\nu + \varepsilon\Gamma\nu$ are both rational functions with the same poles. Since h does not have a pole at β, then the corresponding measure belongs to W. Since β is pole of maximum real part, h decays slower to 0 than $e^{\beta x}$, so there exists a $\delta > 0$ such that

$$|h(x)| \leq Ae^{(\beta-\delta)x}$$

for some constant $A > 0$.

On the other hand the polynomial r^* of the dominating term can be written as

$$r^*(x) = \sum_{i=0}^{d-1} \rho_i x^i,$$

for some constants ρ_i, where $\rho_{d-1} > 0$ since μ^* is constructed in such a way that it has a pole of maximum real part of order d at β. We now calculate density of the absolute continuous part of $R_t\mu^*$, $f_{R_t\mu^*}(x) = f^*(x+t)$, which amounts to

$$\begin{aligned}
f_{R_t\mu^*}(x) &= f^*(x+t) \\
&= r^*(x+t)e^{\beta(x+t)} + h(x+t) \\
&= \sum_{i=0}^{d-1} \rho_i (x+t)^i + h(x+t) \\
&= \sum_{i=0}^{d-1} \rho_i \sum_{j=0}^{i} \binom{i}{j} x^j t^{i-j} e^{\beta(x+t)} + h(x+t) \\
&= \sum_{j=0}^{d-1} \sum_{i=j}^{d-1} \rho_i \binom{i}{j} x^j t^{i-j} e^{\beta(x+t)} + h(x+t) \\
&= r^*(t)e^{\beta(x+t)} + \sum_{j=1}^{d-1} \sum_{i=j}^{d-1} \rho_i \binom{i}{j} x^j t^{i-j} e^{\beta(x+t)} + h(x+t) \\
&= \sum_{j=1}^{d-1} c_j(t) \frac{(-\beta)^{j+1}}{j!} x^j e^{\beta x} + e^{\beta t} r^*(t) e^{\beta x} + h(x+t) \\
&= f_t(x) + \left(e^{\beta t} r^*(t) e^{\beta x} + h(x+t) \right)
\end{aligned}$$

where

$$c_j(t) = \frac{j!}{(-\beta)^{j+1}} \sum_{i=j}^{d-1} \binom{i}{j} \rho_i t^{i-j}$$

and

$$f_t(x) = \sum_{j=1}^{d-1} c_j(t) \frac{(-\beta)^{j+1}}{j!} x^j e^{\beta x}.$$

The dominant term in the $c_j(t)$ formula is of order $t^{d-1-j}e^{\beta t}$ as $t \to \infty$. Since $\rho_{d-1} > 0$, then there exists a t_1 such that if $t \geq t_1$ then $c_j(t)$ is positive because the non–dominant terms will be fading out.

Let $\alpha_{1,t} = \int_0^\infty f_t(x)dx$. Then since f_t is a linear combination of Erlang densities with weights $c_j(t)$, we have that $\alpha_{1,t} = \sum_{j=1}^{d-1} c_j(t)$. Also, if $t \geq t_1$ then each $c_j(t)$ will be positive and hence $\alpha_{1,t} > 0$. Then $f_t/\alpha_{1,t}$ is probability density and it is obviously a linear combination of the Erlang distribution $\tau_2, ..., \tau_d$ with weights $c_j(t)/\alpha_{1,t}$ on the τ_j measure. Let $\theta_{1,t}$ denote measure which corresponds to the density $f_t/\alpha_{1,t}$. Then in terms of measures,

$$\theta_{1,t} = \sum_{j=2} d \frac{c_j(t)}{\alpha_{1,t}} \tau_j.$$

Since the coefficients to the τ_i are positive for $t \geq t_1$, then this is a convex combination of the τ_i for $t \geq t_1$ and hence for $t \geq t_1$ we have that

$$\theta_{1,t} \in \text{co}(\tau_2, .., \tau_d).$$

In fact,

$$\theta_{1,t} \in \text{ri}\,(\text{co}(\tau_2, .., \tau_d))$$

for $t \geq t_1$ since the weights are strictly positive.

Now we consider the other part of the function $f_{R_t\mu^*}(x)$, which is

$$g_t(x) := e^{\beta t} r^*(t) e^{\beta x} + h(x+t).$$

Thus $f_{R_t\mu^*}(x) = f_t(x) + g_t(x)$. Again we wish to normalize to a density, so we consider

$$\begin{aligned}
\alpha_{2,t} : &= \int_0^\infty g_t(x)dx \\
&= \int_0^\infty f_{R_t\mu^*}(x)dx - \int_0^\infty f_t(x)dx \\
&= \int_0^\infty f_{\mu^*}(x+t)dx - \alpha_{1,t} \\
&= \mu^*(t,\infty) - \alpha_{1,t}.
\end{aligned}$$

Since $g_t(x) = e^{\beta x} r^*(t) e^{\beta x} + h(x+t)$, the leading term of $r^*(t)$, $\rho_{d-1}t^{d-1}$ is positive, and since $|h(x)| \leq Ae^{(\beta-\delta)x}$ for some $\delta > 0$, then the over all leading term as $t \to \infty$

is $\rho_{d-1}t^{d-1}e^{\beta t} > 0$. This also hold for its integral. Thus as $t \to \infty$, $\alpha_{2,t}$ must be positive (going to $+\infty$). Now normalize, to obtain distribution with density $g_t/\alpha_{2,t}$ for sufficiently large t in order for $\alpha_{2,t} > 0$, say $t \geq t_2$. It is clear that

$$\frac{g_t(x)}{\alpha_{2,t}} = \frac{e^{\beta t}r^*(t)e^{\beta x} + h(x+t)}{-\beta^{-1}e^{\beta t}r^*(t) + \int_0^\infty h(x+t)dx},$$

and since $\int_0^\infty h(x+t)dx = \int_t^\infty h(x)dx \to 0$ as $t \to \infty$ by $|h(x)| \leq Ae^{(\beta-\delta)x}$, then it is clear that as $t \to \infty$,

$$\frac{g_t(x)}{\alpha_{2,t}} \to -\beta e^{\beta x}$$

point–wise. Since $\tau_1 \in V_R(\mu)$ and the latter begin finite–dimensional, this implies also convergence in $||\cdot||$–norm of $\theta_{2,t}$ to τ_1, where $\theta_{2,t}$ is the measure with density $g_t/\alpha_{2,1}$. We recall that $\mu_1, ..., \mu_\ell$ have been chosen in such a way by formula 7.3 that $\tau_1 \in \text{ri}\,(\text{co}(\mu_1, ..., \mu_\ell)$. Hence there exists some $N > t_2$ such that $\theta_{2,t} \in \text{ri}\,(\text{co}(\mu_1, ..., \mu_\ell)$. Thus for sufficiently large t, $\theta_{1,t} \in \text{ri}\,(\text{co}(\tau_2, .., \tau_d))$ and $\theta_{2,t} \in \text{ri}\,(\text{co}(\mu_1, ..., \mu_\ell)$, so for such a t,

$$R_t\mu^* = \mu^*[0,t)\delta_0 + p_1\theta_{1,t} + p_2\theta_{2,1}$$

and where $p_1, p_2 \geq 0$ and $\mu[0,t) + p_1 + p_2 = 1$. Hence for such a t,

$$R_t\mu^* \in \text{ri}\,(\text{co}\,(\delta_0, \tau_2, ..., \tau_d, \mu_1, ..., \mu_\ell))$$

and hence for large t, $R_t\mu^* \in \mathscr{P}_0$ since the latter contains the former. □

Now the final step is an augmentation in order to make $\mathscr{P}_0$ R–invariant, or equivalently, pointing inwards from all the generating points of the convex hull.

Lemma 7.12. *$\{R_t\mu^* : t \in [0,N]\}$ is a compact subset of ri$(co(V_R(\mu) \cap \mathscr{S}_+)$.*

Proof. We have proved that $\mu^* \in \text{ri}(V_R(\mu) \cap \mathscr{S}_+)$ and since $\text{ri}(V_R(\mu) \cap \mathscr{S}_+)$ it follows that $R_t\mu^* \in \text{ri}(V_R(\mu) \cap \mathscr{S}_+)$ for $t \in [0,N]$. Since $t \to R_t\mu^*$ is continuous in the $||\cdot||$–norm, then $\{R_t\mu^* : t \in [0,N]\}$ is compact being a continuous image of a compact set, and is contained in $\text{ri}(V_R(\mu) \cap \mathscr{S}_+)$. □

Define approximants

$$\mu^*_{(j)} := \left(\boldsymbol{I} + \frac{N}{n}\Gamma\right)^j \mu^* = \left(\boldsymbol{I} + \frac{jN/n}{j}\Gamma\right)^j \mu^*$$

which are close to $\exp(\frac{jN/n}{j}\Gamma\mu^*) = R_{jN/n}\mu^*$. Now all $R_{jN/n}\mu^* \in \text{ri}(V_R(\mu) \cap \mathscr{S}_+)$ for $j = 0, 1, ..., n$ and hence we may choose n sufficiently large so that all $\mu^*_{(j)} \in \text{ri}(V_R(\mu) \cap \mathscr{S}_+)$ for $j = 0, 1, ..., n$. Also, $\mu^*_{(n)}$ approximates $R_N\mu^*$ and since $R_N\mu^* \in \text{ri}(\mathscr{P}_0)$, then we may find an n sufficiently large so that $\mu^*_{(n)} \in \text{ri}(\mathscr{P}_0)$. Now let n be so large that is satisfies both former requirements.

Definition 7.12.

$$\mathscr{P} = \mathrm{co}\left(\mathscr{P}_0 \cup \{\mu, \mu^{(1)}, ..., \mu^{(\omega)} \cup \{\bar{\mu}^*_{(j)} : j = 0, 1, ..., n\}\right),$$

where the $\mu^{(i)}$s are defined in formula (7.4).

We now prove that $\mathscr{P}$ is R–invariant by proving that Γ point inwards at all points in the generator of the convex hull for $\mathscr{P}$. Since $\mathscr{P}_0$ itself is R–invariant, then for any point in the generator of $\mathscr{P}_0$ Γ point inwards towards $\mathscr{P}_0$ and hence towards $\mathscr{P}$ as well.

Next take a $\mu^{(i)}$, $i = 0, 1, ..., \omega - 1$. Then since

$$\mu^{(i+1)} = (\boldsymbol{I} + \varepsilon_{i+1}\Gamma)\mu^{(i)} \in \mathscr{P}$$

then Γ point inwards toward $\mathscr{P}$ at $\mu^{(i)}$, $i = 0, ..., \omega - 1$.

Next we go on to $\mu^*_{(j)}$, $j = 0, 1, ..., n - 1$. But $\mu^*_{(j)}$ is a convex combination of $\delta_0 \in \mathscr{P}$ and $\bar{\mu}^*_{(j)} \in \mathscr{P}$ so $\mu^*_{(j)} \in \mathscr{P}$. Also, n was chosen so large that $\mu_{(n)} \in \mathscr{P}_0$ so $\mu_{(n)} \in \mathscr{P}$ as well. That Γ point inwards at $\bar{\mu}^*_{(j)}$ when it does at $\mu^*_{(j)}$ now follows from the $\mathscr{P}$ being closed under the $\bar{}$ –operation as in the proof of Theorem 7.12.

Thus $\mathscr{P}$ is R–invariant and we have proved our main theorem.

Theorem 7.16. *(Analytic characterization of phase–type distributions) A distribution μ with a rational Laplace transform is a phase–type distribution if and only if it has a possible point mass at zero, the density of its absolute continuous part is strictly positive on $(0, \infty)$ and its Laplace transform has a pole of maximum real part.*

Proof. If μ has a rational Laplace transform, then $V_R(\mu)$ is finite–dimensional (Theorem 1.26) and the result follows from the construction above. The opposite result is that of Theorems 7.3 and 1.4 . □

Example 7.2. Consider the distribution with density

$$f(x) = 2e^{-x}(1 - \cos(x)).$$

Then its Laplace transform is given by

$$\hat{f}(s) = \frac{2}{s^3 + 4s^2 + 5s + 2},$$

which is clearly a rational function, and hence the distribution is matrix–exponential, however, since f is not strictly positive on $(0, \infty)$ ($f(p2\pi) = 0$, $p = 0, 1, ...$) it cannot be of phase–type.

7.3 Characterization of discrete phase–type distributions

Needed for the Schutzenberger application of the closure characterization to obtain the above result.

7.4 Characterization of phase–type distributions via closure properties

Theorem 7.17. *The class of phase–type distributions is the smallest class of distributions on* $\mathbb{R}_+ = [0,\infty)$ *which*

1. *is closed under finite mixtures and finite convolutions,*
2. *is closed under of the convolution of a geometrically distributed number of terms, and*
3. *which contains an atom at zero and all exponential distributions.*

Proof. By Theorem 3.9 the class of phase–type distributions is closed under finite mixtures, by Theorem 3.8 closed under finite convolutions and by Theorem 3.10 it is closed under the convolution of a geometrically distributed number of phase–type distributions, since the geometric distributions is a discrete phase–type distribution. Also, the class may be assigned a mass at zero when we allow for initiation of the underlying Markov jump process at zero, and the class obviously contains all exponential distributions.

Then the smallest class, C, which satisfies 1.,2. and 3. of the theorem is contained in the class of phase–type distributions. In order to show, that the class of phase–type distributions indeed coincide with C, we must prove that any phase–type distribution is contained in C.

To this end we perform a proof by induction on the order of the phase–type distributions.

A phase–type distribution of order 1 is a mixture of an atom at zero and an exponential distribution. Both are by definition contained in C, and their mixtures as well, so first order phase–type distributions belongs to C.

Assume that all n'th order phase–type distributions belongs to C. Let τ be a phase–type distribution of order $n+1$. Let $\{X_t\}_{t\geq 0}$ denote the underlying Markov jump process which generates τ. This process has $n+2$ states, the last being absorbing and the previous ones being transient. We may assume that there is no atom at zero by the closure property of mixtures of C. Hence the density of τ is given by

$$f(x) = \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s},$$

for some $n+1$–dimensional vector $\boldsymbol{\alpha}$, $(n+1)\times(n+1)$–matrix $\boldsymbol{S}$ and $n+$–dimensional column vector $\boldsymbol{s}$. We may also write

$$f(x) = \sum_{i=1}^{n+1} \alpha_i \boldsymbol{e}_i e^{\boldsymbol{S}x}\boldsymbol{s}.$$

Hence by the closure property under mixtures of C, it is sufficient to prove that the phase–type distributions of the form $(\boldsymbol{e}_i, \boldsymbol{S})$ are contained in S, and by relabeling, it boils down to proving that $(\boldsymbol{e}_1, \boldsymbol{S})$ belongs to C.

In order to use the induction assumption, we single out the state 1, and write the Λ on the form

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \boldsymbol{\lambda} & \lambda_{1,n+2} \\ \boldsymbol{t}^{(1)} & \boldsymbol{T} & \boldsymbol{t}^{(n+2)} \\ 0 & \mathbf{0} & 0 \end{pmatrix}.$$

Thus the process starts in state 1, remains there an exponentially distributed amount of time, then jumps either to the absorbing state (with intensity $\lambda_{1,n+2}$ or with probability $-\lambda_{1,n+2}/\lambda_{11}$) or to a state in $\{2,3,...,n+1\}$ according to the probability distribution $-\boldsymbol{\lambda}/\lambda_{11}$. The time it remains in states $\{2,3,...,n+1\}$ before leaving the subset again is phase–type distributed with initial distribution $\boldsymbol{\lambda}$ and sub–intensity matrix $\boldsymbol{T}$. When it leaves states $\{2,3,...,n+1\}$, it either jumps to state 1 or to the absorbing state. Given that is does not jump to the absorbing state, the time until it jumps to state is phase–type distributed with initial vector $\boldsymbol{\lambda}$ and sub–intensity matrix $\boldsymbol{T} + \boldsymbol{\Delta}(\boldsymbol{t}^{(n+2)})$, which can easily be seen by an argument using taboo probabilities. Here $\boldsymbol{\Delta}(\boldsymbol{a})$) is the diagonal matrix with $\boldsymbol{a} = (a_1,...,a_n)$ in the diagonal.

Similarly, the time until absorption from states $\{2,3,...,n+1\}$ given that it does not jump to state 1 first, is phase–type distributed with initial vector $\boldsymbol{t}^{(1)}$ and sub–intensity matrix $\boldsymbol{T} + \boldsymbol{\Delta}(\boldsymbol{t}^{(1)})$.

Then define the following variables. Let N be the number of times the process $\{X_t\}_{t\geq 0}$ are in state 1 before absorption. Let U_i, $i = 1,2,...,N$ denote the times spent in states 1. Let V_i, $i = 1,2,...,N$ denote the times spent in states $\{2,3,...,n+1\}$.

Then V_N is phase–type distributed, since it is the conditional distribution of the time until leaving states $\{2,3,...,n+1\}$ given that it does not leave for state 1, and $V_1,....,V_{N-1}$ are phase–type distributed being the conditional distributions of the time until the process leaves states $\{2,3,...,n+1\}$ given that it does not leave for state $n+2$.

Clearly N is geometrically distributed and given N, all U_i and V_i are independent (strong Markov property) with either exponential or phase–type distributions of order n, The time until absorption τ is

$$\tau = \sum_{i=1}^{N} U_i + \sum_{i=1}^{N} V_i.$$

Since C is closed under finite convolutions, it now follows that $U_i + V_i$ must be in C for all i, and then since it is closed under geometric convolutions as well, we conclude that τ has a distribution in C. □

7.5 Some background on formal power series in one variable

Consider an infinite sequence of complex numbers $\boldsymbol{c} = (c_0, c_1, c_2, ...)$. Let

$$\mathbb{C}^\infty = \{(c_0, c_1, c_2, ...) \mid c_i \in \mathbb{C}, i = 0,1,2,...\}$$

be the set of all such sequences. Let $\mathbb{R}^\infty$ denote the subset of all real sequences.

Definition 7.13. We define summation and multiplication of $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^\infty$ by

$$\boldsymbol{a}+\boldsymbol{b} = (a_0+b_0, a_1+b_1, ..., a_n+b_n, ...)$$
$$\boldsymbol{ab} = (a_0b_0, a_1b_0+a_0b_1, ..., \sum_{j=0}^{n} a_j b_{n-j}, ...)$$

We recognize the n'th term of $\boldsymbol{ab}$, $(\boldsymbol{ab})_n$, as the coefficient to the n'th power in the product of two power series $\sum_{i=0}^{\infty} a_i z^i$ and $\sum_{i=0}^{\infty} b_i z^i$, i.e. if

$$\sum_{i=0}^{\infty} c_i z^i = \left(\sum_{i=0}^{\infty} a_i z^i\right)\left(\sum_{i=0}^{\infty} b_i z^i\right)$$

then

$$c_n = \sum_{j=0}^{n} a_j b_{n-j}.$$

Similarly, $(\boldsymbol{a}+\boldsymbol{b})_n = a_n + b_n$, is the n'th coefficient of the sum of the two power series.

We have that for $\boldsymbol{e}_0 = (1,0,0,....)$, $\boldsymbol{e}_0^n = \boldsymbol{e}_0$. Also, with $\boldsymbol{e}_i = (0,...,0,1,0,...)$ (the 1 is placed on the i'th place) we have that

$$\boldsymbol{e}_1^2 = (0,0,1,0,....) = \boldsymbol{e}_2, ..., \boldsymbol{e}_1^n = \boldsymbol{e}_n, ...$$

If we define scalar multiplication of a constant λ and $\boldsymbol{c} \in \mathbb{C}^{\infty}$ by $\lambda \boldsymbol{c} = (\lambda c_0, \lambda c_1, ...)$, then we can write

$$\boldsymbol{c} = \sum_{i=0}^{\infty} c_i \boldsymbol{e}_1^n$$

with the convention that $\boldsymbol{e}_1^0 = \boldsymbol{e}_0 = (1,0,0,....)$. Hence any complex sequence of numbers may be written in this way.

Definition 7.14. We say that $\boldsymbol{a} = \boldsymbol{b}$ if $a_i = b_i$ for all $i = 0,1,....$ Also, way say that $\boldsymbol{b}$ is an inverse of $\boldsymbol{a}$ if $\boldsymbol{ab} = \boldsymbol{ba} = \boldsymbol{e}_0 = (1,0,0,...)$. In this case we write $\boldsymbol{b} = \boldsymbol{a}^{-1}$. Finally, we let $\lambda = \boldsymbol{e}_1 = (0,1,0,...)$ be the basic variable for which

$$\boldsymbol{c} = \sum_{i=0}^{\infty} c_i \boldsymbol{\lambda}^n.$$

Example 7.3. Consider $\sum_{i=0}^{\infty} \lambda^n = (1,1,1,...)$. Then by definition of the product,

$$(1,-1,0,0,....)(1,1,1,1,...) = (1,0,0,0,....)$$

and

$$(1,1,1,1,...)(1,-1,0,0,....) = (1,0,0,0,...)$$

which is the same as

$$(1-\boldsymbol{\lambda})(1+\lambda+\lambda^2+....) = (1+\lambda+\lambda^2+....)(1-\lambda) = (1,0,0,...).$$

Thus by definition of inverse,

$$(1-\lambda)^{-1} = 1+\lambda+\lambda^2+....$$

Here no convergence considerations are involved and yet the formulas are rigorous and valid. □

Theorem 7.18. $\boldsymbol{a} \in \mathbb{C}^\infty$ *has an inverse if and only if* $a_0 \neq 0$.

Proof. If $\boldsymbol{a}$ has an inverse, $\boldsymbol{b}$ say, then

$$\boldsymbol{ab} = (a_0b_0, a_1b_0+a_0b_1, ...) = (1,0,0,....)$$

and hence a_0 cannot be 0. On the other hand, if $a_0 \neq 0$, then from the equation above, $b_0 = 1/a_0$, $a_1b_0+a_0b_1 = 0$ implying that $b_1 = -a_1b_0/a_0$, $a_2b_0+a_1b_1+a_0b_2 = 0$ implying that $b_2 = -(a_2b_0+a_1b_1)/a_0$, and continuing this way we see that we get a unique solution $\boldsymbol{b}$ which exists if a_0. □

We shall now consider sequences of formal power series. Let

$$\boldsymbol{c}_i = \sum_{k=0}^{\infty} c_{ki}\lambda^k.$$

Definition 7.15. The sequence $\{\boldsymbol{c}_i\}_{i=1,2,...}$ is said to be summable or to admit addition if

$$\forall r \geq 0\ \exists N \in \mathbb{N}:\ n \geq N \Longrightarrow a_{0n} = ... = a_{rn} = 0.$$

Now recall that $\boldsymbol{c}_i$ are sequences of numbers

$$\boldsymbol{c}_i = (c_{0i}, c_{1i}, c_{2i}, ...).$$

If the sequence is summable, then $\{c_{ki}\}_i$ will be zero from a certain point onwards. Then the infinite sum of sequences $\sum_{j=1}^{\infty} \boldsymbol{c}_j$ converges element-wise since each coordinate reduces to a finte sum. Thus the term summability.

Example 7.4. Consider the power series $S = \sum_{i=1}^{\infty} c_i\lambda^i$, i.e. the constant term equals zero. Let

$$S^n = \sum_{j=0}^{\infty} c_{jn}\lambda^j.$$

Then the lowest potence of S^n is λ^n. Then the coefficients $c_{0n} = ... = c_{n-1,n} = 0$, and since all coefficients to S^m, $m > n$, also satisfies this property we have that $a_{0m} = ... = a_{n-1,m} = 0$ for all $m \geq n$. Hence the sequence

$$\boldsymbol{c}_n = S^n$$

is summable.

Definition 7.16. Consider a formal power series with constant term 0. Then we define the star of the series as the sum

$$S^* = \sum_{n=0}^{\infty} S^n$$

with the convention $S^0 = 1$, and plus of the series as

$$S^+ = \sum_{n=1}^{\infty} S^n.$$

□

Then we obviously have that

$$S^* = 1 + S^+$$

and

$$S^*S = SS^* = S\sum_{n=0}^{\infty} S^n = \sum_{n=0}^{\infty} S^{n+1} = S^+.$$

Thus $1-S$ is the inverse of S^* since

$$(1-S)S^* = S^* - S^+ = 1 = S^*(1-S).$$

Assume that $S = \sum_{i=1}^{\infty} a_i \lambda^i$ and that $S^+ = \sum_{n\geq 1} S^n = \sum_{i=1}^{\infty} c_i \lambda^i$. Since the constant term of S is zero, then potences in S^+ of size n originates from $S, S^2, S^3, ..., S^n$ only, and after a short consideration we conclude that

$$c_n = \sum_{m=1}^{n} \sum_{i_1+...+i_m=n} a_{i_1} a_{i_2} \cdots a_{i_m}. \tag{7.5}$$

Definition 7.17. Let S and T be formal power series. The rational operations of formal power series are:

1. Summation $S+T$.
2. Product ST.
4. Quasi-inversion, S^* and S^+.

Scaling with a constant c is a particular case of the product operation, since we may define a constant by the trivial formal power series $S = (c, 0, 0, ...)$.

Definition 7.18. Let P_{pol} denote the class of finite formal power series, i.e. formal power series on the form

$$S = \sum_{i=0}^{n} c_i \lambda^i$$

for some finite number n (i.e. a polynomial in λ).

Let P_{rat} denote the class of formal power series obtained by applying repeatedly the rational operations to the elements of P_{pol}. Then we say that a formal power series S is rational if $S \in P_{rat}$.

It is clear, that any finite polynomial may be constructed by repeated rational operations on constants and λ. Hence constants and λ generated the finite polynomials and hence the rational power series. We have proved

Lemma 7.13. *The rational power series is the class of power series which*

1 contains the function 1 *and* λ*,*
2 is closed under sum, product and quasi–inverses.

It is important to restrict the formal power series to certain sub–classes of $\mathbb{C}$, for example $\mathbb{R}_+$ in our case. To this end we shall formulate the possible target set in terms of semi–rings.

Definition 7.19. A semi–ring $R = (C, +, \cdot, 1, 0)$ is an algebraic structure, where

1. C is a non–empty set (the elements of the semi–ring),
2. summation $+$ is associative and commutative,
3. multiplication $\cdot$ is associative,
4. multiplication is distributive with respect tot addition, i.e. $a \cdot (b+c) = a \cdot b + a \cdot c$ and $(b+c) \cdot a = b \cdot a + c \cdot a$,
5. 1 is a neutral element for multiplication, i.e. $a \cdot 1 = 1 \cdot a = a$,
6. 0 is a neutral element for summation, i.e. $a + 0 = 0 + a = a$,
7. $0 \cdot 1 = 1 \cdot 0 = 0$.

Both neutral elements are necessarily unique (if 1 and $1'$ are neutral elements for multiplication, then $1 = 1 \cdot 1' = 1'$ and if 0 and $0'$ are neutral elements for addition, $0 = 0 + 0' = 0'$). We shall use the usual notation that

$$\underbrace{a + a + \ldots + a}_{n \text{ summands}} = na \quad \text{and} \quad \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ factors}} = a^n.$$

Definition 7.20. The set of formal power series in one variable z with coefficients in a set C which belongs to a semi–ring $R = (C, +, \cdot, 1, 0)$ is denoted by $R\langle\langle z \rangle\rangle$.

We now introduce the concept of recognizable power series. This is going to provide the link to the phase–type characterization.

Definition 7.21. Consider a formal power series $S \in R\langle\langle z \rangle\rangle$ for some semi–ring $R = (C, +, \cdot, 1, 0)$. Then S is called R–recognizable, or just recognizable if confusion it not possible, if there exists a matrix representation of its coefficients $c_i \in C$, that is, there exists a number $p \geq 1$, vector of dimension p, $\boldsymbol{\pi}$ (row), $\boldsymbol{t}$ (column) and a $p \times p$ matrix $\boldsymbol{T}$ such that all elements of the vectors and matrix belong to C, and

$$c_i = \boldsymbol{\pi} \mu(s^i) \boldsymbol{t}$$

where μ is a matrix–valued function with the property

$$\mu(s^i) = \boldsymbol{T}^i \quad (\boldsymbol{T}^0 = \boldsymbol{I}).$$

The reason for writing the definition in this slightly cumbersome way instead of pluggin in the $\boldsymbol{T}^n$ straightaway is partly to maintain a link to automaton theory and partly in order to make its extension to higher dimensions more transparent.

A similar definition is needed for rational power series.

Definition 7.22. An R–rational power series is obtained from power series in $R\langle\langle z \rangle\rangle$ by applying the rational operations summation, product and quasi–inverses.

Theorem 7.19. *Let S_1 and S_2 be two recognizable formal power series in $R\langle\langle\lambda\rangle\rangle$. Then S_1+S_2 and S_1S_2 are recognizable formal power series. If S is a recognizable formal power series with constant term zero, then S^* and S^+ are recognizable.*

Proof. Let $S_1=\sum_{i=0}^{\infty}a_1\lambda^i$ and $S_2=\sum_{i=0}^{\infty}b_i\lambda^i$. Since S_1 and S_2 are recognizable, then there exists dimensions p_1, p_2 and vectors $\boldsymbol{\alpha}$, $\boldsymbol{s}$ (dimensions p_1), $\boldsymbol{\beta}$ and $\boldsymbol{t}$ (dimensions p_2) and matrices $\boldsymbol{T}$ $(p_1\times p_1)$, $\boldsymbol{S}$ $(p_2\times p_2)$ such that

$$a_i=\boldsymbol{\alpha}\boldsymbol{S}^i\boldsymbol{s},\ b_i=\boldsymbol{\beta}\boldsymbol{T}^i\boldsymbol{t},$$

and the elements of the vectors and matrices belong to the set C of the semi–ring $R=(C,+,\cdot,1,0)$. Now consider

$$S_1+S_2=\sum_{i=0}^{\infty}(a_i+b_i)\lambda^i.$$

Define

$$\boldsymbol{\gamma}=(\boldsymbol{\alpha}\ \ \boldsymbol{\beta})\ \ \boldsymbol{U}=\begin{pmatrix}\boldsymbol{S}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{T}\end{pmatrix},\ \ \boldsymbol{u}=\begin{pmatrix}\boldsymbol{s}\\ \boldsymbol{t}\end{pmatrix}.$$

Then all elements of $\boldsymbol{\gamma}$, $\boldsymbol{U}$ and $\boldsymbol{u}$ are also elements of C, and $\boldsymbol{\gamma}\boldsymbol{U}^i\boldsymbol{u}=a_i+b_i$. Hence the formal power series S_1+S_2 is recognizable.

Concerning the product, we choose

$$\boldsymbol{\gamma}=(\boldsymbol{\alpha}\ \ \boldsymbol{0})\ \ \boldsymbol{U}=\begin{pmatrix}\boldsymbol{S}&\boldsymbol{s}\boldsymbol{\beta}\\ \boldsymbol{0}&\boldsymbol{T}\end{pmatrix},\ \ \boldsymbol{u}=\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{t}\end{pmatrix}.$$

Then the elements of the matrix and vectors just defined belong to C, and

$$\boldsymbol{U}^2=\begin{pmatrix}\boldsymbol{S}^2&\boldsymbol{S}\boldsymbol{s}\boldsymbol{\beta}+\boldsymbol{s}\boldsymbol{\beta}\boldsymbol{T}\\ \boldsymbol{0}&\boldsymbol{T}^2\end{pmatrix},$$

so

$$\boldsymbol{\gamma}\boldsymbol{U}^2\boldsymbol{u}=\boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{s}\boldsymbol{\beta}\boldsymbol{t}+\boldsymbol{\alpha}\boldsymbol{s}\boldsymbol{\beta}\boldsymbol{T}\boldsymbol{t}=a_1b_0+a_0b_1$$

which indeed coincide with the coefficient to λ^2 of the power series S_1S_2. The general result now follows by induction. Hence S_1S_2 is recognizable.

If $S=\sum_{i=0}^{\infty}a_i\lambda^i=\sum_{i=1}^{\infty}a_i\lambda^i$ is a recognizable formal power series with constant term $a_0=0$, then there exists vectors $\boldsymbol{\alpha}$, $\boldsymbol{t}$ and matrix $\boldsymbol{T}$ such that

$$a_i=\boldsymbol{\alpha}\boldsymbol{T}^i\boldsymbol{t},i\geq 1,\ a_0=\boldsymbol{\alpha}\boldsymbol{t}=0.$$

Define

$$\boldsymbol{\gamma}=\boldsymbol{\alpha},\ \ \boldsymbol{U}=\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\alpha}\boldsymbol{T}\ \ \boldsymbol{u}=\boldsymbol{t}.$$

Then

$$\begin{aligned}\boldsymbol{\gamma}\boldsymbol{U}\boldsymbol{u}&=\boldsymbol{\alpha}(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\alpha}\boldsymbol{T})\boldsymbol{t}\\ &=\boldsymbol{\alpha}\boldsymbol{T}\boldsymbol{t}\end{aligned}$$

$$\begin{aligned}\boldsymbol{\gamma U}^2\boldsymbol{u} &= \boldsymbol{\alpha}\left(\boldsymbol{T}^2+\boldsymbol{Tt\alpha T}+\boldsymbol{\alpha tT}^2+\boldsymbol{t\alpha Tt\alpha T}\right)\boldsymbol{t}\\ &= \boldsymbol{\alpha T}^2\boldsymbol{t}+(\boldsymbol{\alpha Tt})^2\\ \boldsymbol{\gamma U}^3\boldsymbol{u} &= \ldots\\ &= \boldsymbol{\alpha}^3\boldsymbol{T}^3\boldsymbol{t}+(\boldsymbol{\alpha T}^2\boldsymbol{t})\boldsymbol{\alpha Tt}+\boldsymbol{\alpha Tt}(\boldsymbol{\alpha T}^2\boldsymbol{t})+(\boldsymbol{\alpha Tt})^3\end{aligned}$$

etc. Hence by induction we see that the terms $\boldsymbol{\gamma U}^i\boldsymbol{u}$ coincide with the coefficients c_i of S^+. Hence S^+ is recognizable. In order to prove that S^* is recognizable it is enough to prove that the constant term $1=(1,0,0,....)$ is recognizable and appealing to the result that the sum of two recognizable power series is again recognizable. For the constant series,

$$S=1=\sum_{i=0}^{\infty}a_i\lambda^i$$

with $a_0=1$ and $a_i=0$, $i\geq 1$, we simply choose $\boldsymbol{\gamma}=1$, $\boldsymbol{U}=\{0\}$ and $\boldsymbol{u}=1$. Then $\boldsymbol{\gamma U}^0\boldsymbol{u}=1$ and $\boldsymbol{\gamma U}^i\boldsymbol{u}=0$ for $i\geq 1$. □

It follows immediately from the definition of rationality, that

Corollary 7.4. *A R–rational formal power series is R–recognizable.*

The opposite result also holds true and establishes the final equivalence result (the theorem of Schützenberger). In order to prove the opposite implication and hence the equivalence we shall need a lemma. We shall here consider formal power series where the coefficients are $p\times p$ matrices. The result is a matrix where the elements are formal power series in one dimension.

Theorem 7.20. *Let $\boldsymbol{S}$ denote a formal power series with matrix ($p\times p$) coefficients and where the constant term is the zero matrix. Let $\boldsymbol{q}$ and $\boldsymbol{z}$ be p–dimensional vectors. Then the equation*

$$\boldsymbol{z}=\boldsymbol{q}+\boldsymbol{zS} \tag{7.6}$$

has a unique $\boldsymbol{z}=\mathbf{q}\boldsymbol{S}^*$. Furthermore, if $\boldsymbol{q}$ and S are rational, then so is $\boldsymbol{z}$.

Proof. Let $\boldsymbol{A}_i$ be a sequence of $p\times p$ matrices, and let

$$S=\sum_{i=1}^{\infty}\boldsymbol{A}_i\lambda^i.$$

Then we may write

$$S=\begin{pmatrix}S_{11} & S_{12} & \ldots & S_{1p}\\ S_{21} & S_{22} & \ldots & S_{2p}\\ \vdots & \vdots & \vdots & \vdots\\ S_{p1} & S_{p2} & \ldots & S_{pp}\end{pmatrix}$$

where S_{ij}, $i,j=1,...,p$ are formal power series. Then

$$S^*=\sum_{n=0}^{\infty}S^n=\boldsymbol{I}+\sum_{n=1}^{\infty}S^n=\boldsymbol{I}+S^*S.$$

Pre–multiplying this equation by the vector $\boldsymbol{q}$ yields

$$\boldsymbol{q}\boldsymbol{S}^* = \boldsymbol{q} = (\boldsymbol{q}\boldsymbol{S}^*)\,S$$

so $\boldsymbol{q}\boldsymbol{S}^*$ indeed solves the equation $\boldsymbol{z} = \boldsymbol{q} + \boldsymbol{z}S$.

Let $\boldsymbol{z} = \boldsymbol{q}S^*$. Then iterating with $S^* = \boldsymbol{I} + S^*S$ we get that

$$\begin{aligned}
\boldsymbol{z} &= qS^* \\
&= q\,(\boldsymbol{I} + S^*S) \\
&= q + qS^*S \\
&= q + q(\boldsymbol{I} + S^*S)S \\
&= q + qS + qS^*S^2 \\
&= \ldots \\
&= q\sum_{i=0}^{n-1} S^i + qS^*S^n \\
&= q\sum_{i=0}^{n-1} S^i + \boldsymbol{z}S^n,
\end{aligned}$$

and since S has constant term zero, then all coefficients of potences less than n in S^n are zero, so $\boldsymbol{z}S^n$ only contains terms of potences larger than or equal to n. Then the first $n-1$ coordinates in the sequences which correspond to the power series of $\boldsymbol{q}S^*$ will all remain fixed after n iterations. Hence the solution is unique.

To prove the rationality, assume that $\boldsymbol{q}$ and $\boldsymbol{S}$ are rational. Then each term of S, S_{ij}, are rational. In one dimension it is easy to see that a constant time the rational S_{11}^* (solution to the equation) is indeed rational. Now we make induction on the dimensionality. Assume that it holds for dimension $\ell - 1$. Then consider

$$(z_1,\ldots,z_\ell) = (q_1,\ldots,q_\ell) + (z_1,\ldots,z_\ell)\begin{pmatrix} S_{11} & S_{12} & \ldots & S_{1\ell} \\ S_{21} & S_{22} & \ldots & S_{2\ell} \\ \vdots & \vdots & \ldots & \vdots \\ S_{\ell 1} & S_{\ell 2} & \ldots & s_{\ell\ell} \end{pmatrix}.$$

Isolating z_ℓ we get that

$$\begin{aligned}
z_\ell &= q_\ell + \sum_{i=1}^{\ell} z_i S_{i\ell} \\
&= q_\ell + \sum_{i=1}^{\ell-1} z_i S_{i\ell} + z_\ell S_{\ell\ell} \\
&= = r_\ell + z_\ell S_{\ell\ell},
\end{aligned}$$

where $r_\ell = q_\ell + \sum_{i=1}^{\ell-1} z_i S_{i\ell}$. This an equation in one dimension and z_ℓ is obviously rational if r_ℓ is rational. But by induction hypothesis, $z_1, ..., z_\ell$ are rational and so is r_ℓ. □

Remark 7.3. The equation (7.6) has the structure of a renewal equation and S^* then serves a the renewal function. Of course we are here in a more general setting and we no further boundedness condition in order to get a unique solution in terms of formal power series since no convergence considerations play a role.

Theorem 7.21. *(Schützenberger)*
A formal power series is R–recognizable if and only if it is R–rational.

Proof. Corollary 7.4 provides the "if" part. For the "only if" part we proceed as follows. Suppose that S is R–recognizable. Then there exists a dimension p, vectors $\boldsymbol{\alpha}, \boldsymbol{t}$ and matrix $\boldsymbol{T}$ such that

$$S = \sum_{n=0}^{\infty} \boldsymbol{\alpha} \boldsymbol{T}^n \boldsymbol{t} \lambda^n.$$

Consider the matrix polynomial,

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

This is a formal power series with constant term zero. Furthermore,

$$\boldsymbol{P}^* = \sum_{n=0}^{\infty} \boldsymbol{P}^n = \sum_{n=0}^{\infty} \boldsymbol{T}^n \lambda^n$$

and since

$$S = \sum_{n=0}^{\infty} \boldsymbol{\alpha} \boldsymbol{T}^n \boldsymbol{t} \lambda^n = \boldsymbol{\alpha} \boldsymbol{P}^* \boldsymbol{t}.$$

Now the unique solution to

$$\boldsymbol{z} = \boldsymbol{\alpha} + \boldsymbol{z}\boldsymbol{P}$$

is exactly that of $\boldsymbol{z} = \boldsymbol{\alpha}\boldsymbol{P}^*$. But $\boldsymbol{\alpha}$ and $\boldsymbol{P}$ are both R–rational (being a constant and polynomial) so we conclude that $\boldsymbol{z} = \boldsymbol{\alpha}\boldsymbol{P}^*$ is also R–rational. Hence $\boldsymbol{\alpha}\boldsymbol{P}^*\boldsymbol{t}$ (linear combination) is also R–rational. □

7.6 Characterization of discrete phase–type distributions

We now apply the Schützenberger theorem to prove a closure characterization of the discrete phase–type distributions. Let τ be a discrete phase–type distribution with representation $\mathrm{DPH}_d(\boldsymbol{\alpha}, \boldsymbol{T})$. Then its probability generating function f^* is given by

$$f^*(z) = \sum_{n=1}^{\infty} \boldsymbol{\alpha} \boldsymbol{T}^{n-1} \boldsymbol{t} z^n.$$

Then f^* is obviously a $\mathbb{R}_+$–recognizable formal power series. Thus the class generating functions of discrete phase–type distributions, DPH, is a subclass of the recognizable formal power series.

Lemma 7.14. *Suppose that a $\mathbb{R}_+$–recognizable power series $\hat{f}(z) = \sum_{n=0}^{\infty} f_n z^n$ satisfies that $0 < \hat{f}(1) < \infty$. Then $\hat{f}(z)/\hat{f}(1)$ is the probability generating function of a discrete phase–type distribution.*

Proof. The power series being $\mathbb{R}_+$–recognizable implies that there exists vectors $\boldsymbol{\alpha}$ (row), $\boldsymbol{s}$ (column) and matrix $\boldsymbol{S}$ such that $f_i = \boldsymbol{\alpha}\boldsymbol{S}^i\boldsymbol{s}$, where all elements of the vectors and matrices belongs to $\mathbb{R}_+$.

We shall assume a certain irreducibility property, namely that

$$\forall i \; \exists k \; : \left(\boldsymbol{\alpha}\boldsymbol{S}^k\right)_i > 0. \tag{7.7}$$

If this condition is not satisfied, then there is an i such that $(\boldsymbol{\alpha}\boldsymbol{S}^k)_i = 0$ for all $k = 0, 1, \ldots$. But then we may remove entrances i from the vectors and the row and column i from $\boldsymbol{S}$ to obtain vectors $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{s}}$ and matrix $\tilde{\boldsymbol{S}}$ such that

$$f_i = \boldsymbol{\alpha}\boldsymbol{S}^i\boldsymbol{s} = \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{S}}^i\tilde{\boldsymbol{s}}.$$

Hence continuing this way we may assume irreducibility as defined by (7.7).

Since $0 < \hat{f}(1) < \infty$, then not all f_i can be zero and hence $\boldsymbol{s}$ must be non–zero, since $\boldsymbol{s} = \boldsymbol{0}$ implies that $f_i = 0$ for all i. But then by irreducibility, $f_i > 0$ for all i.

A similar argument show that we may also assume that

$$\forall i \; \exists k \; : \left(\boldsymbol{S}^k\boldsymbol{s}\right)_i > 0 \tag{7.8}$$

since otherwise we may reduce the vectors and matrices by entrances corresponding to some index i.

Now let i, j be arbitrary indices of the state–space. Let k_1 and k_2 be numbers such that

$$\begin{aligned}
\infty &> \sum_{n=0}^{\infty} \boldsymbol{\alpha}\boldsymbol{S}^n\boldsymbol{s} \\
&\geq \sum_{n=0}^{\infty} \boldsymbol{\alpha}\boldsymbol{S}^{k_1}\boldsymbol{S}^n\boldsymbol{S}^{k_2}\boldsymbol{s} \\
&\geq \sum_{n=0}^{\infty} \left(\boldsymbol{\alpha}\boldsymbol{S}^{k_1}\right)_i s_{ij}^{(n)} \left(\boldsymbol{S}^{k_2}\boldsymbol{s}\right)_j
\end{aligned}$$

so therefore $\sum_{n=0}^{\infty} s_{ij}^{(n)}$ must be finite as well. This proves that under irreducibility (7.7) and (7.8) we must have that $\sum_{n=0}^{\infty} \boldsymbol{S}^n$ converges element–wise, and necessarily,

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

By the comparison principle for power series, also $\sum_{n=0}^{\infty} \boldsymbol{S}^n z^n$ must be convergent for $|z| \leq 1$, and

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

Thus for $|z| \leq 1$ we have that

$$\sum_{i=0}^{\infty} f_i z^i = \boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s}.$$

We must then prove that

$$\hat{g}(z) := \frac{\hat{f}(z)}{\hat{f}(1)} = \frac{\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}}$$

is the probability generating function of a discrete phase–type distribution.

To this end we first consider the expression $\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s}$. Since $\boldsymbol{I} = \boldsymbol{I} - z\boldsymbol{S} + z\boldsymbol{S}$ we get by right multiplication of $(\boldsymbol{I} - z\boldsymbol{S})^{-1}$ that

$$(\boldsymbol{I} - z\boldsymbol{S})^{-1} = \boldsymbol{I} + z\boldsymbol{S}(\boldsymbol{I} - z\boldsymbol{S})^{-1}$$

and hence

$$\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s} = \boldsymbol{\alpha}\boldsymbol{s} + z\boldsymbol{\alpha}\boldsymbol{S}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s}. \tag{7.9}$$

If Δ is any invertible matrix of order compatible to $\boldsymbol{S}$, then

$$\begin{aligned} \boldsymbol{S}(\boldsymbol{I} - z\boldsymbol{S})^{-1} &= \boldsymbol{S}(\boldsymbol{\Delta}\boldsymbol{\Delta}^{-1} - z\boldsymbol{S})^{-1} \\ &= \boldsymbol{S}\boldsymbol{\Delta}(\boldsymbol{\Delta} - z\boldsymbol{S}\boldsymbol{\Delta})^{-1} \\ &= \boldsymbol{S}\boldsymbol{\Delta}(\boldsymbol{I} - z\boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^{-1} \end{aligned}$$

so

$$\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s} = \boldsymbol{\alpha}\boldsymbol{s} + \boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{\Delta}(\boldsymbol{I} - z\boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{s}. \tag{7.10}$$

We now have to choose $\boldsymbol{\Delta}$ carefully in order to obtain a discrete phase–type representation. It turn out that a good choice is

$$\boldsymbol{\Delta} = \mathrm{diag}(\xi_1, ..., \xi_p),$$

where p is the dimension of the matrix $\boldsymbol{S}$ and

$$\boldsymbol{\xi} = (\xi_1, ..., \xi_d) = (\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}.$$

By (7.8) we have that $\boldsymbol{\xi} > \boldsymbol{0}$, i.e. all entrances are strictly positive. Hence $\boldsymbol{\Delta}$ is non–negative with a strictly positive diagonal and hence invertible. Now

$$\boldsymbol{\Delta}\boldsymbol{e} = \boldsymbol{\xi} = (\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s} \tag{7.11}$$

so

$$\boldsymbol{s} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{\Delta}\boldsymbol{e}$$

and hence

$$\boldsymbol{\Delta}^{-1}\boldsymbol{s} = (\boldsymbol{I} - \boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})\boldsymbol{e}. \tag{7.12}$$

Thus (7.10) amounts to

$$\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s} = \boldsymbol{\alpha}\boldsymbol{s} + \boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{\Delta}(\boldsymbol{I} - z\boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})^{-1}(\boldsymbol{I} - \boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})\boldsymbol{e}.$$

Thus

$$\hat{g}(z) = \frac{\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} = \frac{\boldsymbol{\alpha}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} + \frac{\boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{\Delta}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}}(\boldsymbol{I} - z\boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})^{-1}(\boldsymbol{I} - \boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta})\boldsymbol{e}.$$

Now define

$$\begin{aligned} \pi_{p+1} &= \frac{\boldsymbol{\alpha}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} \\ \boldsymbol{\pi} &= \frac{\boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{\Delta}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} \\ \boldsymbol{T} &= \boldsymbol{\Delta}^{-1}\boldsymbol{S}\boldsymbol{\Delta}. \end{aligned}$$

Then

$$\hat{g}(z) = \pi_{p+1} + \boldsymbol{\pi}(\boldsymbol{I} - z\boldsymbol{T})^{-1}(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e}$$

is on the "right form" for being the probability generating function of a discrete phase–type distribution, and all that is left to show is that $\boldsymbol{T}$ is a sub–stochastic matrix and that $\boldsymbol{\pi}$ is a non–negative vector vector the elements of which together with π_{p+1} sum to one.

By (7.12) we have that

$$(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{e} = \boldsymbol{\Delta}^{-1}\boldsymbol{s} \geq \boldsymbol{0}$$

so $\boldsymbol{T}$ is sub–stochastic. It is cleat that $\boldsymbol{\pi} \geq \boldsymbol{0}$ since all entries in the numerator are non–negative and since the denominator is $\hat{f}(1)$ which is supposed to be in $(0,\infty)$. Furthermore,

$$\begin{aligned} \boldsymbol{\pi}\boldsymbol{e} + \pi_{p+1} &= \frac{\boldsymbol{\alpha}\boldsymbol{S}\boldsymbol{\Delta}\boldsymbol{e} + \boldsymbol{\alpha}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} \\ &= \frac{\boldsymbol{\alpha}\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s} + \boldsymbol{\alpha}\boldsymbol{s}}{\boldsymbol{\alpha}(\boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}} \quad \text{(by (7.11))} \\ &= 1 \quad \text{(by (7.9))}. \end{aligned}$$

□

Theorem 7.22. *The class of discrete phase–type DPH distributions is the smallest class of distributions on* $\{0,1,2,....\}$ *for which the corresponding class of probability generating functions,*

$$DPH^* = \{\hat{f}(z) | f \in DPH\},$$

where $\hat{f}$ denotes the corresponding probability generating function, satisfies that

1. $1, z \in DPH^*$,
2. *is closed under sums and products,*
3. *and is closed under quasi–inversion.*

Proof. A distribution which satisfies 1.-3. above is $\mathbb{R}_+$–rational since all its coefficients in the formal power expansion must be probabilities and hence non–negative. This is turn is equivalent to it being $\mathbb{R}_+$–recognizable. Then by Lemma 7.14 the distribution must be of discrete phase–type. □

7.7 Characterization of continuous phase–type distributions revisited

Apply discrete characterization + schutzenberger to get original continuous characterization.

Exercises

Exercise 7.1. Find the normalisation constant c such that

$$f(x) = ce^{-x}(2 + \cos(2\pi x) + \sin(4\pi x))$$

is a density, calculate its Laplace transform and prove that it is a density of a phase–type distribution. Plot the density. Prove, that if we instead consider f proportional to $e^{-x}(2 + \cos(2\pi x) + \sin(3\pi x))$, then f will be a matrix–exponential distribution which is not of phase–type.

Chapter 8
Characterization of matrix–exponential distributions

Chapter 9
Further properties of matrix–exponential and phase–type distributions

9.1 Moment distributions

Let f be a density function of a random variable X with support on $[0,\infty)$. Let $\mu_n = \mathbb{E}(X^n)$ be the n'th moment. If $\mu_n < \infty$ we define the n'th order moment distribution with density f_n by

$$f_n(x) = \frac{x^n f(x)}{\mu_n}.$$

We shall consider moment distribution of either phase–type or matrix–exponential distributions, an we shall treat the two cases separately though the latter class contains the forms because of the issues concerning the representation. For both classes we know that $\mu_n < \infty$ is satisfied for all orders n.

Moment distributions have their origin in size–biased sampling. In one dimension we talk about length biased sampling. Suppose that the failures of a component in a machine occur according to a renewal process, and that the times between failures have distribution given by a density f. Now consider the component of the machine at a random time, T say. What is the failure time of the sampled component? By considering the component which works a time T is equivalent to sampling an interval in the renewal process in which the same component is working. Since we sample randomly, the probability that the component has failure time x must be proportional to x. Hence the distribution of the failure time of the component must have density proportional to $xf(x)$. This implies the desired form f_1 of the density.

Likewise, in higher dimensions the sampling random objects may be biased by their size. For example, random sampling of mineral grains depends on the volume of these, and with an argument similar to the one performed for length biasing, if f is the density of the distribution which gives us the volumes of the grains, the volume of a certain sampled grain, if sampled randomly, will have a density proportional to $x^3 f(x)$, implying the exact form f_3.

9.1.1 Moment distributions in a matrix–exponential setting

We now continue with the formal development of the subject. First consider the Laplace transform of f, $\mathscr{L}(f,s) = \int_0^\infty e^{-sx} f(x)dx$. Differentiating n times with respect to s yields

$$\mathscr{L}^{(n)}(f,s) = \frac{d^n}{ds^n}\mathscr{L}(f,s) = (-1)^n \int_0^\infty x^n e^{-sx} f(x)dx$$

so we conclude that the Laplace transform of the n'th moment distribution is given by

$$\mathscr{L}(f_n,s) = \int_0^\infty \frac{x^n f(x)}{\mu_n} e^{-sx} dx = \frac{\mathscr{L}^{(n)}(f,s)}{\mu_n}.$$

Now assume that f is matrix–exponential. Then its Laplace transform is a rational function. But the Laplace transform of f_n is proportional to the n'th derivative of the Laplace transform of f and is hence also rational, proving that f_n is again matrix–exponential.

Assume the f is phase–type. By O'cinneide's characterization theorem (Theorem 7.16) we then known that f is strictly positive everywhere on $(0,\infty)$ and that $\mathscr{L}(f,s)$ has a pole of maximum real part. Since $f(x) > 0$ for $x > 0$ then also $f_n(x) > 0$ for $x > 0$. Furthermore, the poles remain the same under differentiation of the Laplace transform since we only get powers of the denominator for the new denominator, and hence the pole which has the maximum real part is again real. Thus Theorem 7.16 implies that f_n must the density of a phase–type distribution. We have proved the following qualitative result.

Theorem 9.1. *The n'th order moment distribution of a phase–type distribution is again phase–type. Similarly, the n'th order moment distribution of matrix–exponential distribution is again matrix–exponential.*

Corollary 9.1. *If f is the density of a matrix–exponential distribution which is not phase–type, then its n'th order moment distribution is again matrix–exponential but not phase–type.*

Proof. If there exists a $x_0 \in (0,\infty)$ for which $f(x_0) = 0$, then also $f_n(x_0) = 0$ for all n. Hence f_n cannot be of phase–type (Theorem 7.16). Since the poles remain intact by differentiating the Laplace transform, also if f fails to be of phase–type due to the pole criterium, so do f_n for all n. □

Hence the class of phase–type distributions, the class of matrix–exponential distributions and the class of matrix–exponential distributions which are not phase–type are all closed under formation of moment distributions. These are qualitative results which are interesting in their own right but of limited practical use since we have not provided representations of any of them.

We start with a general matrix–exponential distribution.

Theorem 9.2. *Let f be the density of a random variable $X \sim ME_P(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Then f_n is matrix–exponential with representation $ME_{np}(\boldsymbol{\alpha}_n, \boldsymbol{S}_n, \boldsymbol{s}_n)$, where*

$$\boldsymbol{\alpha}_n = \left(\frac{\boldsymbol{\alpha}\boldsymbol{S}^{-n}}{-\boldsymbol{\alpha}\boldsymbol{S}^{-(n+1)}\boldsymbol{s}}, \mathbf{0}, ..., \mathbf{0} \right), \quad \boldsymbol{S}_n = \begin{pmatrix} \boldsymbol{S} & -\boldsymbol{S} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S} & -\boldsymbol{S} & ... & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & \boldsymbol{S} \end{pmatrix}, \quad \boldsymbol{s}_n = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \boldsymbol{s} \end{pmatrix}. \tag{9.1}$$

In particular, if $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$ the representation amounts to

$$\boldsymbol{\alpha}_n = \left(\frac{\boldsymbol{\alpha}\boldsymbol{S}^{-n}}{\boldsymbol{\alpha}\boldsymbol{S}^{-n}\boldsymbol{e}}, \mathbf{0}, ..., \mathbf{0} \right), \quad \boldsymbol{S}_n = \begin{pmatrix} \boldsymbol{S} & -\boldsymbol{S} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S} & -\boldsymbol{S} & ... & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & \boldsymbol{S} \end{pmatrix}, \quad \boldsymbol{s}_n = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \boldsymbol{s} \end{pmatrix}. \tag{9.2}$$

The matrix $\boldsymbol{S}_n$ is a block matrix of dimension $(n+1)p \times (n+1)p$, i.e. there are $(n+1) \times (n+1)$ blocks each of size $p \times p$. $\boldsymbol{\alpha}_n$, in any of its representations, is an $(n+1)p$–dimensional row vector which consists of $n+1$ blocks of p dimensional row vectors. $\boldsymbol{s}_n$ is an $(n+1)p$–dimensional column vector of n blocks of size p each.

Proof. An easy calculation show that (remember that there are $n+1$ blocks)

$$e^{\boldsymbol{S}_n x} = \begin{pmatrix} e^{\boldsymbol{S}x} & -x\boldsymbol{S}e^{\boldsymbol{S}x} & \frac{x^2\boldsymbol{S}^n}{2}e^{\boldsymbol{S}x} & ... & \frac{(-1)^n x^n \boldsymbol{S}^n}{n!}e^{\boldsymbol{S}x} \\ \mathbf{0} & e^{\boldsymbol{S}x} & -x\boldsymbol{S}e^{\boldsymbol{S}x} & ... & \frac{(-1)^{n-1} x^{n-1} \boldsymbol{S}^{n-1}}{(n-1)!}e^{\boldsymbol{S}x} \\ \mathbf{0} & \mathbf{0} & e^{\boldsymbol{S}x} & ... & \frac{(-1)^{n-2} x^{n-2} \boldsymbol{S}^{n-2}}{(n-2)!}e^{\boldsymbol{S}x} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & ... & e^{\boldsymbol{S}x} \end{pmatrix}. \tag{9.3}$$

Then we notice that the only element of the matrix $\exp(\boldsymbol{S}_n x)$ which contributes to $\boldsymbol{\alpha}_n \exp(\boldsymbol{S}_n x)\boldsymbol{s}_n$ is the upper right corner $(1 \times n)$. This element is readily calculated to be

$$\frac{1}{n!}(-1)^n x^n \boldsymbol{S}^n \exp(\boldsymbol{S}x).$$

Pre–multiplying this term with $\boldsymbol{\alpha}\boldsymbol{S}^{-n}$ and post–multiplying with $\boldsymbol{s}$ we get

$$\frac{(-1)^n}{n!} x^n \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{s} = \frac{1}{(-1)^n n!} x^n f(x).$$

The n'th moment of the general form of the matrix–exponential distribution is $n!\boldsymbol{\alpha}(-\boldsymbol{S})^{-(n+1)}\boldsymbol{s}$ from which the results follows immediately. □

It is clear that $\boldsymbol{S}_n$ is not an intensity matrix unless $\boldsymbol{S}$ is on diagonal form, which is only the case for hyper–exponential distributions.

Theorem 9.3. *The distribution function F_n for the n'th moment distribution of a general matrix–exponential distribution with representation $ME_p(\boldsymbol{\alpha},\boldsymbol{S},\boldsymbol{s})$ is given by*

$$F_n(x) = 1 - \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-n}}{\boldsymbol{\alpha}\boldsymbol{S}^{-(n+1)}\boldsymbol{s}} \sum_{i=0}^{n} \frac{(-\boldsymbol{S})^{i-1}x^i}{i!} e^{\boldsymbol{S}x}\boldsymbol{s}.$$

For a matrix–exponential representation with $\boldsymbol{s} = -\boldsymbol{S}\boldsymbol{e}$, the formula reduces to

$$F_n(x) = 1 - \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-n}}{\boldsymbol{\alpha}\boldsymbol{S}^{-n}\boldsymbol{e}} \sum_{i=0}^{n} \frac{(-\boldsymbol{S})^{i}x^i}{i!} e^{\boldsymbol{S}x}\boldsymbol{e}.$$

The dimensions of the matrices and vectors are again $(n+1)p \times (n+1)p$ and $(n+1)p$ respectively with the same block structure as the theorem above.

Proof. From Theorem 1.6 we get that

$$F_n(x) = 1 + \boldsymbol{\alpha}_n e^{\boldsymbol{S}_n x}\boldsymbol{S}_n^{-1}\boldsymbol{s}_n.$$

Now

$$\boldsymbol{S}_n^{-1} = \begin{pmatrix} \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \boldsymbol{0} & \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{S}^{-1} \end{pmatrix},$$

and so

$$\boldsymbol{S}_n^{-1}\boldsymbol{s}_n = \begin{pmatrix} \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \boldsymbol{0} & \boldsymbol{S}^{-1} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}^{-1} & \dots & \boldsymbol{S}^{-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{S}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{s} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}^{-1}\boldsymbol{s} \\ \boldsymbol{S}^{-1}\boldsymbol{s} \\ \boldsymbol{S}^{-1}\boldsymbol{s} \\ \vdots \\ \boldsymbol{S}^{-1}\boldsymbol{s} \end{pmatrix}.$$

Only the first row of (9.3) contributes to $\boldsymbol{\alpha}_n \exp(\boldsymbol{S}_n x)\boldsymbol{S}_n^{-1}\boldsymbol{s}_n$, and both results now follows immediately by inserting the relevant expressions of $\boldsymbol{\alpha}_n, \boldsymbol{S}_n$ and $\boldsymbol{s}_n$ from Theorem 9.2. □

Also the Laplace transform is of importance in the following.

Lemma 9.1. *The inverse of the matrix ($n+1$ blocks of matrices)*

$$\begin{pmatrix} u\boldsymbol{I}-\boldsymbol{S} & \boldsymbol{S} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & u\boldsymbol{I}-\boldsymbol{S} & \boldsymbol{S} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & u\boldsymbol{I}-\boldsymbol{S} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & u\boldsymbol{I}-\boldsymbol{S} \end{pmatrix}$$

is given by

$$\begin{pmatrix} (u\boldsymbol{I}-\boldsymbol{S})^{-1} & (-\boldsymbol{S})^{-1}(u\boldsymbol{I}-\boldsymbol{S})^{-2} & (-\boldsymbol{S})^{-2}(u\boldsymbol{I}-\boldsymbol{S})^{-3} & \dots & (-\boldsymbol{S})^{-n}(u\boldsymbol{I}-\boldsymbol{S})^{-(n+1)} \\ \boldsymbol{0} & (u\boldsymbol{I}-\boldsymbol{S})^{-1} & (-\boldsymbol{S})^{-1}(u\boldsymbol{I}-\boldsymbol{S})^{-2} & \dots & (-\boldsymbol{S})^{-(n-1)}(u\boldsymbol{I}-\boldsymbol{S})^{-n} \\ \boldsymbol{0} & \boldsymbol{0} & (u\boldsymbol{I}-\boldsymbol{S})^{-1} & \dots & (-\boldsymbol{S})^{-(n-2)}(u\boldsymbol{I}-\boldsymbol{S})^{-(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & (u\boldsymbol{I}-\boldsymbol{S})^{-1} \end{pmatrix}$$

Proof. Left to the reader. □

Theorem 9.4. *The Laplace transform corresponding to the representation (9.1) is given by*

$$L_X(u) = \frac{\boldsymbol{\alpha}(u\boldsymbol{I}-\boldsymbol{S})^{-(n+1)}\boldsymbol{s}}{\boldsymbol{\alpha}(-\boldsymbol{S})^{-(n+1)}\boldsymbol{s}}. \tag{9.4}$$

In the special case of (9.2), this reduces to

$$L_X(u) = \frac{\boldsymbol{\alpha}(u\boldsymbol{I}-\boldsymbol{S})^{-(n+1)}\boldsymbol{s}}{\boldsymbol{\alpha}(-\boldsymbol{S})^{-n}\boldsymbol{e}}. \tag{9.5}$$

Proof. Follows directly from Theorem 9.2 and Lemma 9.1 . □

9.1.2 Moments distributions in a phase–type setting

We now turn to the case of phase–type distributions. Of course the above representation remains valid also for this sub–class, however, as mentioned previously, the representations are not phase–type representations outside the class of hyper–exponential distributions. Therefore, and for the sake of completeness, it is desirable to derive representations of phase–type distributions which enable the possibility of probabilistic reasoning whenever they are used in modeling.

The starting point is Theorem 4.7. Hence we shall consider a stationary renewal process with inter–arrival times which are phase–type distributed with representation $\mathrm{PH}_p(\boldsymbol{\pi},\boldsymbol{T})$. We shall assume that the representation is irreducible.

According to Theorem 4.12, the delayed and initial distribution of a stationary phase–type renewal process has a distribution which is phase–type with the same sub–intensity matrix and initial distribution

$$\boldsymbol{\nu} = \frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}.$$

By (6.2), this is also the stationary distribution to the concatenated Markov jump process $\{X_t\}_{t\geq 0}$ with intensity matrix $\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}$. Hence, if $X_0 \sim \boldsymbol{\nu}$ then $X_t \sim \boldsymbol{\nu}$ for all t.

Now consider the spread $S_t = A_t + R_t$, where A_t and R_t are the age and residual life processes (see Definition 4.2). Then S_t has a distribution which is the first moment distribution of $\mathrm{PH}_p(\boldsymbol{\pi},\boldsymbol{T})$. We shall no establish a representation for the

distribution of S_t using a probabilistic argument involving A_t and R_t. Consider an arbitrary time point t. By stationarity, then $X_t \sim \boldsymbol{\nu}$ and consequently $R_t \sim \mathrm{PH}_p(\boldsymbol{\nu}, \boldsymbol{T})$. Since $A_t \sim R_t$ (Theorem 4.6) we also conclude that $A_t \sim \mathrm{PH}_p(\boldsymbol{\nu}, \boldsymbol{T})$.

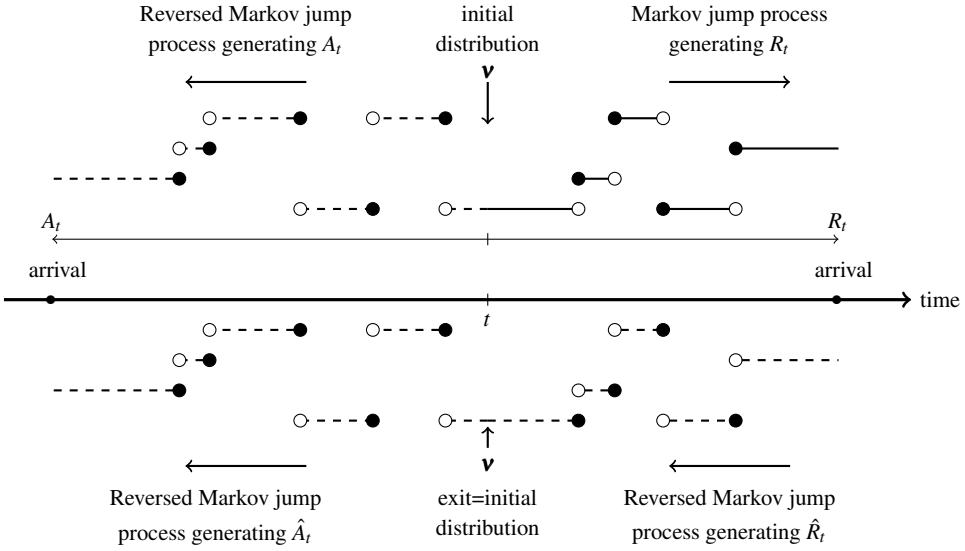

Fig. 9.1 Markov jump processes generating the spread about time t.

In Figure 9.1 we see a couple of examples of how we could generate S_t. In the upper part of the figure, at time t the Markov jump process which generates next arrival point in the renewal process (at a time distance R_t) takes off towards the right, while a Markov jump process which generates the arrival prior to t (at a time distance A_t from t) takes off towards the left. The spread is the sum of the times A_t and R_t. The times are dependent since the both initiate in the same state chosen according to $\boldsymbol{\nu}$. Conditionally on the chosen initial state, they are independent.

Because of the dependence we cannot perform a standard convolution argument for finding a representation of their sum. Instead we proceed as follows. The representations of the distributions of R_t and A_t are now of crucial importance. We could represent both A_t and R_t by either the $\mathrm{PH}_p(\boldsymbol{\nu}, \boldsymbol{T})$ representation or by its time–reversed representation $\mathrm{PH}_p(\boldsymbol{\nu}, \tilde{\boldsymbol{T}})$ (see Theorem 6.3) or by any other representation for that matter. The time–reversed representation is somehow convenient in this context because the concatenated Markov jump process obtained from the time–reversed phase–type distributions has also stationary distribution $\boldsymbol{\nu}$ (Theorem 2.26). If we time–reverse the R_t process, then we may think of running a Markov jump process from the time of the next arrival backwards to t (see Figure 9.1, lower part). The distribution of this process just prior to exit is $\boldsymbol{\nu}$ (by stationarity of the time–reversed process, Theorem 2.26) and hence we may generate the initial dis-

tribution of the A_t process, $\boldsymbol{\nu}$, by running the R_t process until absorption and using its distribution just prior to this event. From that point, we just start the A_t process running backwards according to the state from which the R_t process exited.

It is important to notice that if we use the time–reversed representation for R_t, then we must also use the time–reversed representation for A_t. We cannot mix time–reversed and original representations when concatenating these dependent processes since the state–space of the time–reversed has been flipped (notice the transposition when defining $\tilde{\boldsymbol{T}} = \boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})$, where $\tilde{\boldsymbol{\nu}}$ is defined in (6.1).

There we use A_t represented by $\mathrm{PH}_p(\boldsymbol{\nu}, \tilde{\boldsymbol{T}})$. We now need to time reverse $R_t \sim \mathrm{PH}_p(\boldsymbol{\nu}, \boldsymbol{T})$. To this end we define $\boldsymbol{m} = -\boldsymbol{\nu}\boldsymbol{T}^{-1}$ (see (6.1)). But

$$\boldsymbol{\nu} = \boldsymbol{\pi}(-\boldsymbol{T})^{-1}/\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}$$

so

$$\boldsymbol{m} = \frac{\boldsymbol{\pi}\boldsymbol{T}^{-2}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}.$$

Hence the time–reversed representation of $\mathrm{PH}_p(\boldsymbol{\nu}, \boldsymbol{T})$ has initial distribution $\boldsymbol{t}'\boldsymbol{\Delta}(\boldsymbol{m})$ and sub–intensity matrix $\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m})$. The exit rate vector of this time–reversed representation is $\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\nu}'$.

The A_t process was reversed using its original initial distribution $\boldsymbol{\pi}$ from the previous arrival up to time t so it runs backwards from time t to the previous arrival. Thus its sub–intensity matrix corresponding to the reversed A_t process is $\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})$.

Hence we have proved

Theorem 9.5. *The first moment distribution of a phase–type distribution with representation $PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ is again phase–type distributed with representation $PH_{2p}(\boldsymbol{\pi}_1, \boldsymbol{T}_1)$, where*

$$\boldsymbol{\pi}_1 = (\boldsymbol{t}'\boldsymbol{\Delta}(\boldsymbol{m}), \boldsymbol{0}), \quad \boldsymbol{T}_1 = \begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\nu) \\ \boldsymbol{0} & \boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}}) \end{pmatrix},$$

where

$$\boldsymbol{m} = \frac{\boldsymbol{\pi}\boldsymbol{T}^{-2}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}, \quad \tilde{\boldsymbol{\nu}} = \boldsymbol{\pi}(-\boldsymbol{T})^{-1}, \quad \boldsymbol{\nu} = \frac{\tilde{\boldsymbol{\nu}}}{\tilde{\boldsymbol{\nu}}\boldsymbol{e}}.$$

It is clear that we could obtain an alternative representation by time–reversing the already time–reversed A_t process and then letting it join the original R_t process in the same state (see Exercise 9.1). The following theorem provides an alternative representation in the original time direction.

Theorem 9.6. *Consider a phase-type distribution with representation $PH_p(\boldsymbol{\pi}, \boldsymbol{T})$. Then its first moment distribution can be represented by a phase-type representation $PH_{2p}(\boldsymbol{\pi}_2, \boldsymbol{T}_2)$, where*

$$\boldsymbol{\pi}_2 = \left(\frac{\boldsymbol{\pi}\boldsymbol{\Delta}(\boldsymbol{\rho}_1)}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}, \boldsymbol{0} \right), \; \boldsymbol{T}_2 = \begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{\rho}_1)\boldsymbol{T}\boldsymbol{\Delta}(\boldsymbol{\rho}_1) & \boldsymbol{\Delta}^{-1}(\boldsymbol{\rho}_1) \\ \boldsymbol{0} & \boldsymbol{T} \end{pmatrix},$$

and $\boldsymbol{\rho}_1 = (-\boldsymbol{T})^{-1}\boldsymbol{e}$.

Proof. The representation is obtained through time–reversing the representation $\mathrm{PH}_{2p}(\boldsymbol{\pi}_1, \boldsymbol{T}_1)$ of Theorem 9.5. Of central importance in the time–reversal plays the vector $-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}$. The inverse matrix of a block–matrix of the form

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{pmatrix},$$

which is easily seen by e.g. direct verification. Thus

$$\boldsymbol{T}_1^{-1} = \begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\left(\boldsymbol{T}^{-1}\right)'\boldsymbol{\Delta}(\boldsymbol{m}) & -\boldsymbol{\Delta}(\boldsymbol{m})\left(\boldsymbol{T}^{-2}\right)'\boldsymbol{\Delta}(\boldsymbol{\nu}) \\ \boldsymbol{0} & \boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\left(\boldsymbol{T}^{-1}\right)'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}}) \end{pmatrix}.$$

Thus

$$\begin{aligned} -\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1} &= \left(\boldsymbol{e}'\boldsymbol{\Delta}(\boldsymbol{m}), \boldsymbol{t}'\left(\boldsymbol{T}^{-2}\right)'\boldsymbol{\Delta}(\boldsymbol{\nu})\right) \\ &= \left(\boldsymbol{e}'\boldsymbol{\Delta}(\boldsymbol{m}), -\left(\boldsymbol{T}^{-2}\boldsymbol{T}\boldsymbol{e}\right)'\boldsymbol{\Delta}(\boldsymbol{\nu})\right) \\ &= \left(\boldsymbol{e}'\boldsymbol{\Delta}(\boldsymbol{m}), \boldsymbol{\rho}_1'\boldsymbol{\Delta}(\boldsymbol{\nu})\right) \\ &= \left(\boldsymbol{m}, \boldsymbol{\rho}_1'\boldsymbol{\Delta}(\boldsymbol{\nu})\right) \end{aligned}$$

Hence the diagonal matrix which has the vector $-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}$ as diagonal elements may be written as

$$\boldsymbol{\Delta}\left(-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}\right) = \begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu}) \end{pmatrix}$$

The time–reversed sub–intensity matrix is hence given by

$$\begin{aligned} &\boldsymbol{\Delta}^{-1}\left(-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}\right)\boldsymbol{T}_1'\boldsymbol{\Delta}\left(-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}\right) = \\ &\quad \begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu}) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\nu) \\ \boldsymbol{0} & \boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}}) \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu}) \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}^{-1}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}^{-1}(\boldsymbol{\nu}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m})\boldsymbol{T}\boldsymbol{\Delta}^{-1}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\boldsymbol{\nu}) & \boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu}) \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{T} & \boldsymbol{0} \\ \boldsymbol{\Delta}^{-1}(\boldsymbol{\rho}_1)\,\boldsymbol{\Delta}^{-1}(\boldsymbol{\rho}_1)\boldsymbol{T}\boldsymbol{\Delta}(\boldsymbol{\rho}_1) & \end{pmatrix}. \end{aligned}$$

The initial distribution of the time–reversed is given by $\boldsymbol{t}_1'\boldsymbol{\Delta}\left(-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}\right)$, where $\boldsymbol{t}_t = -\boldsymbol{T}_1\boldsymbol{e}$. The exit rate vector $\boldsymbol{t}_1$ is given by

$$\begin{aligned} \boldsymbol{t}_1 &= -\boldsymbol{T}_1\boldsymbol{e} \\ &= -\begin{pmatrix} \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\nu) \\ \boldsymbol{0} & \boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}}) \end{pmatrix}\boldsymbol{e} \end{aligned}$$

$$= \begin{pmatrix} -\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m})\boldsymbol{e} - \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\nu)\boldsymbol{e} \\ -\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})\boldsymbol{e} \end{pmatrix}$$

Now

$$\begin{aligned} -\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m})\boldsymbol{e} &= -\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}\left(\frac{\boldsymbol{\pi}\boldsymbol{T}^{-2}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\right)\boldsymbol{e} \\ &= -\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\left(\frac{\boldsymbol{\pi}\boldsymbol{T}^{-2}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\right) \\ &= -\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\left(\frac{\boldsymbol{\pi}\boldsymbol{T}^{-2}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\boldsymbol{T}\right)' \\ &= \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\nu} \end{aligned}$$

and hence

$$-\boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{m})\boldsymbol{e} - \boldsymbol{\Delta}^{-1}(\boldsymbol{m})\boldsymbol{\Delta}(\nu)\boldsymbol{e} = \boldsymbol{0},$$

i.e.

$$\boldsymbol{t}_1 = \begin{pmatrix} \boldsymbol{0} \\ -\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})\boldsymbol{e} \end{pmatrix}.$$

Then

$$\begin{aligned} \boldsymbol{t}_1'\boldsymbol{\Delta}(-\boldsymbol{\pi}_1\boldsymbol{T}_1^{-1}) &= \left(\boldsymbol{0}, -\left(\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})\boldsymbol{e}\right)'\right)\begin{pmatrix} \boldsymbol{\Delta}(\boldsymbol{m}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu}) \end{pmatrix} \\ &= \left(\boldsymbol{0}, -\boldsymbol{e}'\boldsymbol{\Delta}(\tilde{\boldsymbol{\nu}})\boldsymbol{T}\boldsymbol{\Delta}^{-1}(\tilde{\boldsymbol{\nu}})\boldsymbol{\Delta}(\boldsymbol{\rho}_1)\boldsymbol{\Delta}(\boldsymbol{\nu})\right) \\ &= (\boldsymbol{0}, -\tilde{\boldsymbol{\nu}}\boldsymbol{T}\boldsymbol{\Delta}(\boldsymbol{\rho}_1))(\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e})^{-1} \\ &= \left(\boldsymbol{0}, -\frac{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\boldsymbol{T}\boldsymbol{\Delta}(\boldsymbol{\rho}_1)\right)(\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e})^{-1} \\ &= \left(\boldsymbol{0}, \frac{\boldsymbol{\pi}\boldsymbol{\Delta}(\boldsymbol{\rho}_1)}{\boldsymbol{\pi}(-\boldsymbol{T})^{-1}\boldsymbol{e}}\right). \end{aligned}$$

Relabeling the states now yields the result. □

In order to get a phase–type representation for moment distribution of order two and above, we could in principle use the fact that $n+1$'th order moment distribution is the n'th order moment distribution of the first order moment distribution. Hence, in order to get a representation for the second order moment distribution we could simply apply Theorems 9.5 or 9.6 to the first order moment distribution. If the original phase–type distribution is of order p, the first order moment distribution would be of order $2p$ and the second order moment distribution of order $4p = 2^2 p$. Continuing this way, a n'th order moment distribution produced iteratively this way would be of order $2^n p$.

Theorem 9.7. *Consider a phase-type distribution with representation $PH_p(\boldsymbol{\alpha}, \boldsymbol{S})$. Then the n'th order moment distribution is again of phase-type with representation*

$PH_{(n+1)p}(\hat{\boldsymbol{\alpha}}_n, \hat{\boldsymbol{S}}_n)$, where

$$\hat{\boldsymbol{\alpha}}_n = \left(\frac{\rho_{n+1}}{\rho_n} \boldsymbol{s}' \boldsymbol{\Delta}(\boldsymbol{\pi}_{n+1}), \mathbf{0}, \ldots, \mathbf{0} \right)$$

$$\hat{\boldsymbol{S}}_n = \begin{bmatrix} \boldsymbol{C}_{n+1} & \boldsymbol{D}_{n+1} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_n & \boldsymbol{D}_n & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{C}_{n-1} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{C}_2 & \boldsymbol{D}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \boldsymbol{C}_1 \end{bmatrix}$$

and $\rho_i = \mu_i / i! = \boldsymbol{\alpha}(-\boldsymbol{S})^{-i}\boldsymbol{e}$ are the reduced moments,

$$\boldsymbol{\pi}_i = \rho_i^{-1} \boldsymbol{\alpha}(-\boldsymbol{S})^{-i}, \qquad \boldsymbol{C}_i = \boldsymbol{\Delta}(\boldsymbol{\pi}_i)^{-1} \boldsymbol{S}' \boldsymbol{\Delta}(\boldsymbol{\pi}_i) \qquad \textit{and} \qquad \boldsymbol{D}_i = \frac{\rho_{i-1}}{\rho_i} \boldsymbol{\Delta}(\boldsymbol{\pi}_i)^{-1} \boldsymbol{\Delta}(\boldsymbol{\pi}_{i-1}).$$

9.1.3 Lorenz curves and Gini index

One well known application of moment distributions is in the area of economics or demography, where the income distribution of a society may be expressed in terms of the first order moment distribution. Inequality in income of a society if often graphically illustrated by a function $L(t)$ where the interpretation is that the poorest $100t$ percent of the society earns $100L(t)$ percent of the total income.

We order the people in a society according to their income (in increasing order) and let F denote the income distribution in the society. If the income groups are $x_1 < x_2 < .... < x_N$ and the proportion of people earning x_i is $f(x_i)$, then the average income of the society is

$$\sum_{i=1}^{N} x_i f(x_i).$$

If the total number of individuals is M, then the total income of the society is $M\sum_{i=1}^{N} x_i f(x_i)$. The total income originating from the first m income groups is

$$M \sum_{i=1}^{m} x_i f(x_i).$$

Hence the proportion of salary which by which the poorest $\sum_{i=1}^{m} f(x_i)$ (the proportion of people in the first m income groups) contributes to the total income is

$$\frac{M\sum_{i=1}^{m} x_i f(x_i)}{M\sum_{i=1}^{N} x_i f(x_i)} = \frac{\sum_{i=1}^{m} x_i f(x_i)}{\sum_{i=1}^{N} x_i f(x_i)}.$$

In a continuous setting, this amounts to the following statement. The proportion by which the poorest $F(t)$ (the proportion of people earning less or equal to t) is given by

$$\frac{\int_0^t xdF(x)}{\int_0^\infty xdF(x)}.$$

But introducing the first moment density $f_1(x)dx = xdF(x)/\mu_1$, then this proportion equals

$$\frac{\int_0^t xdF(x)}{\int_0^\infty xdF(x)} = \int_0^t f_1(x)dx = F_1(x),$$

where F_1 is the distribution function of the first moment distribution. Define the parametric curve $t \to \gamma(t)$ by

$$\gamma : t \in [0,\infty) \to (F(t), F_1(t)). \tag{9.6}$$

Then for a given value of $F(t)$ (a quantile), the corresponding value for $F_1(t)$ is the proportion of which the poorest $F(t)$ contributes to the total income of the society.

Definition 9.1. The parametric curve defined by (9.6) is a concentration curve or Lorenz curve.

The Lorenz curve is a curve since

$$\begin{aligned}\frac{dF_1}{dF}(t) &= \frac{tf(t)dt/\mu_1}{f(t)dt}\\ &= t > 0\\ \frac{d^2F_1}{dF^2}(t) &= \frac{dt}{dF(t)}\\ &= \frac{1}{f(t)} > 0.\end{aligned}$$

If $F(0) = 0$, meaning that there are no atoms at zero, then $\gamma(0,0) = (0,0)$ and by the convexity the curve γ will lie under the line $y = x$, see Figure 9.2.

Definition 9.2. Twice the area confined by the line $y = x$, $x \in [0,1]$ and the Lorenz curve γ is called the Gini index or Gini coefficient of concentration.

While the Lorenz curve have a straight demographic or economic interpretation, the Gini index is a measure which further simplifies the measurement of inequality. Obviously, the area between the line and the Lorenz curve is at most $1/2$, the area of the triangle under the lines $y = x$, so the Gini index is a number between zero and one.

Theorem 9.8. *Consider a matrix–exponential distribution F with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Then its Lorenz curve is given by*

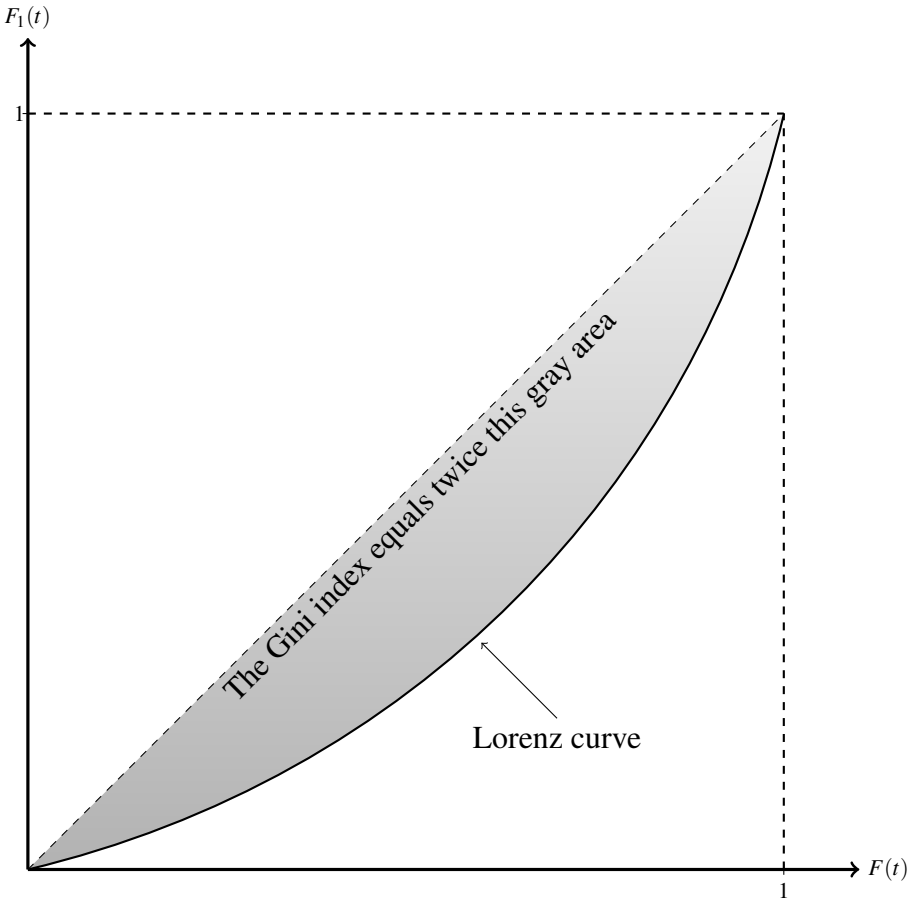

Fig. 9.2 Schematic illustration of the Lorenz curve and the Gini index. The gray area is contained in a triangle of area $\frac{1}{2}$. Complete inequality would correspond to the gray area being one half. The Gini index is hence a number between zero and one.

$$\gamma : t \to \left(1 + \boldsymbol{\alpha} e^{\boldsymbol{S}x} \boldsymbol{S}^{-1} \boldsymbol{s}, 1 + \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-2}}{\boldsymbol{\alpha}\boldsymbol{S}^{-2}\boldsymbol{s}} e^{\boldsymbol{S}x}\boldsymbol{s} - x\frac{\boldsymbol{\alpha}\boldsymbol{S}^{-1}}{\boldsymbol{\alpha}\boldsymbol{S}^{-2}\boldsymbol{s}} e^{\boldsymbol{S}x}\boldsymbol{s} \right).$$

The corresponding Gini index i given by ...

$$G = 2\left(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha}_1\right)\left(\boldsymbol{S}\oplus\boldsymbol{S}_1\right)^{-1}\left(\boldsymbol{s}\otimes\boldsymbol{S}_1^{-1}\boldsymbol{s}_1\right) - 1,$$

where $\boldsymbol{\alpha}_1$, $\boldsymbol{S}_1$ *and* $\boldsymbol{s}_1$ *are given by (9.1).*

In the case where $-\boldsymbol{S}\boldsymbol{e} = \boldsymbol{s}$ *this reduces to*

$$\gamma : t \to \left(1 - \boldsymbol{\alpha} e^{\boldsymbol{S}t}\boldsymbol{e}, 1 - \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-1}}{\boldsymbol{\alpha}\boldsymbol{S}^{-1}\boldsymbol{e}}\left(e^{\boldsymbol{S}t}\boldsymbol{e} + te^{\boldsymbol{S}t}\boldsymbol{s}\right)\right),$$

with corresponding Gini index

$$G = 2(\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}_1)\left(-(\boldsymbol{S} \oplus \boldsymbol{S}_1)\right)^{-1}(\boldsymbol{s} \otimes \boldsymbol{e}) - 1,$$

where $\boldsymbol{\alpha}_1$, $\boldsymbol{S}_1$ *and* $\boldsymbol{s}_1$ *are given by (9.2).*

Proof. The area A under the Lorenz curve is given by the standard formula

$$A = \int_0^\infty F'(t)F_1(t)dt.$$

Inserting the respectively $F_1(t) = 1 + \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 t}\boldsymbol{S}_1^{-1}\boldsymbol{s}_1$ for the general case and $F_1(t) = 1 - \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 t}\boldsymbol{e}$ for the special case immediately yields the result. For example, for the general case we have that

$$\begin{aligned} A &= \int_0^\infty \boldsymbol{\alpha} e^{\boldsymbol{S}t}\boldsymbol{s}\left(1 + \boldsymbol{\alpha}_1 e^{\boldsymbol{S}_1 t}\boldsymbol{S}_1^{-1}\boldsymbol{s}_1\right)dt \\ &= 1 + \int_0^\infty (\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}_1)\, e^{(\boldsymbol{S} \oplus \boldsymbol{S}_1)t}\left(\boldsymbol{s} \otimes \boldsymbol{S}_1^{-1}\boldsymbol{s}\right)dt \\ &= 1 - (\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}_1)\left(\boldsymbol{S} \oplus \boldsymbol{S}_1\right)^{-1}\left(\boldsymbol{s} \otimes \boldsymbol{S}_1^{-1}\boldsymbol{s}_1\right). \end{aligned}$$

The Gini index G then satisfies $G = 2(\frac{1}{2} - A)$ from which the result follows. □

9.2 Denseness

In this section we prove that the class of phase–type distributions is dense in the class of distributions on the positive reals.

Exercises

Exercise 9.1. Derive an alternative phase–type representation for the first order moment distribution of a phase–type distribution with representation $\mathrm{PH}_p(\boldsymbol{\pi}, \boldsymbol{T})$ by time–reversing the time–reversed A_t process and using the original R_t process in Figure 9.1, that is, instead of joining the two dependent processes running from the right towards the left, join two processes which runs from the left towards the right.

Exercise 9.2. Let M_n be the operator

$$M_n(f)(x) = f_n(x) = \frac{x^n f(x)}{\mu_n},$$

which produces the density of the n'th order moment distribution based on the orignal density f and where μ_n denotes the n'th moment, i.e. $\mu_n = \int_0^\infty x^n f(x)dx$. Prove,

that

$$M_{n+m}(f) = M_n(M_m(f)) = M_m(M_n(f)).$$

Exercise 9.3. Consider renewal process with inter–arrival distribution F (and density f) where we sample an interval according to if a randomly chosen point t belongs to the interval. Now assume that a given interval which was sampled is of length x. Since t was chosen at random, the conditional distribution of t, given the interval length x of the sampled interval, is assumed to be a uniform distribution over $[0,x]$. Prove, that the density of the position $z \in [0,x]$ of t within the sampled interval, $k(z)$, is given by

$$k(z) = \frac{1-F(z)}{\mu},$$

where μ is the mean of the inter–arrival times. Is the result surprising?

Exercise 9.4. Consider a general matrix–exponential distribution with representation $(\boldsymbol{\alpha}, \boldsymbol{S}, \boldsymbol{s})$. Prove that the Gini index G is given by

$$G = 2\left(\boldsymbol{\alpha} \otimes \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-2}}{\boldsymbol{\alpha}\boldsymbol{S}^{-2}\boldsymbol{e}}\right)(\boldsymbol{S} \oplus \boldsymbol{S})^{-1}(\boldsymbol{s} \otimes \boldsymbol{s}) + 2\left(\boldsymbol{\alpha} \otimes \frac{\boldsymbol{\alpha}\boldsymbol{S}^{-1}}{\boldsymbol{\alpha}\boldsymbol{S}^{-2}\boldsymbol{e}}\right)(\boldsymbol{S} \oplus \boldsymbol{S})^{-2}(\boldsymbol{s} \otimes \boldsymbol{s}) - 1.$$

Find a similar expression for the case where $\boldsymbol{s}_1 = -\boldsymbol{S}_1\boldsymbol{e}$.

Chapter 10
Markov additive processes

Chapter 11
Multivariate distributions

11.1 Special functions

♠ I guess this should go somewhere else, dependent on whether we need it at other places.

11.1.1 Bessel functions

The Bessel function of the first kind is given by

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+\alpha+1)} \left(\frac{1}{2}x\right)^{2m+\alpha} \tag{11.1}$$

11.1.2 Laguerre polynomials

♠ Something on how they appear

Several versions, in the orthonormal version we get

$$L_j^\alpha = \sqrt{\frac{\Gamma(\alpha-1)\Gamma(\alpha-1+j)}{j!}} \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{1}{\Gamma(\alpha-1+k)} x^k$$

for $\alpha = m$ (m integer)

$$L_j^m = \sqrt{\frac{(m-1)!(m-1+j)!}{j!}} \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{1}{(m-1+k)!} x^k$$

11.2 General classes

We will introduce multivariate distributions through a quite general construction based on linear reward structures on phase–type distributions. Then we will generalize this class to form a multivariate class of ME distributions. We start with the less general case as very little is known on the general multivariate ME distributions. We then continue with a sub-class which is partially of historical interest partly has some importance in its own right as a model very suitable for reliability modelling. We then continue by demonstrating how a number of multivariate exponential and gamma distributions that has been suggested fits into the framework of multivariate phase-type and matrix exponential distributions, whenever the shape parameter of the gamma distribution is integer. Not all distributions seem to belong to the class of linear reward structures and we move on to define the class of multivariate matrix–epxonential distributions as multivariate distributions on a non–negative subspace of an n-dimensional real space (R^n) that has a rational multivariate moment-generating function. These distributions can alternatively be characterized as being the distributions where all non–negative projections on the positve real line following a univariate matrix–exponential distribuion. Beyond this characterization result very little is yet known about these distributions. We conclude the chapter by demonstrating that the the theory applies to distrbutions with support on the full real space R^n.

11.2.1 MPH*

11.2.2 MPH

$$H(\boldsymbol{s}) = \mathbb{E}\left(e^{-\langle \boldsymbol{s},\boldsymbol{X}\rangle}\right) = \boldsymbol{\alpha}\left(\boldsymbol{U}\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{s}) + \boldsymbol{I}\right)^{-1}\boldsymbol{e}, \tag{11.2}$$

with $\boldsymbol{U} = (-\boldsymbol{T})^{-1}$.

11.2.3 MVME

11.3 Distribution types

11.4 MPH$*$ representations for multivariate exponential and gamma distributions

In this section we discuss a number of other bivariate and multivariate exponential and gamma distributions that have been discussed in the literature. Many of these

have rational Laplace transforms. Of these many are seen to be contained in the MPH* or even in the MPH class. In general the multivariate gamma distributions belong to MVME whenever the shape parameters are integer. The main reference is [6]. In this reference the distributions are classified as bivariate and multivariate exponential respectively gamma distributions. Here we take the approach of classifying according to the structure of the $(\boldsymbol{\alpha}, \boldsymbol{T}, \boldsymbol{R})$ representations. Specifically, we will categorize according to two different ways of writing a univariate expoential distribution as a mixture - type 1 and type 2 - and the basic result that an integer shaped gamma distributed random variable can be written as a sum of indpendent exponential random variables - type 3. In fact, most if not all multivariate distrubutions with rational Laplace transform seen so far is formed using these three basic properties. The first two types give multivariate distributions with exponential marginals. They can be combined with the third type in various ways to form distributions with gamma distributed marginals. In the following, we let Y_i, Y_{ij}, and, Z denote independent $\exp(1)$ distribtued random variables.

I. It is well-known that
$$X_i = (1-\delta)\frac{Y_i}{\lambda_i} + J\frac{Z}{\lambda_i} \ ,$$
where J is an indicator variable such that $P(J=1)=\delta$, implies that $X_i \in \exp(\lambda_i)$. For $\delta > 0$ this generates correlated random vectors in MME*. In the bivariate case an MME* representation is given by
$$\left((1,0,0), \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & \delta \\ 0 & 0 & -\frac{1}{1-\delta} \end{bmatrix}, \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \end{bmatrix} \right) \ .$$

II. This type is based on the concept of writing an exponential distribution as an infinite mixture of Erlang distributions with geometric mixing weights, i.e.
$$X_i = \frac{1-p}{\lambda_i} \sum_{j=1}^{N} Y_{ij} \ ,$$
where N is geometrically distributed with $P(N=1)=1-p$. Then we have that $X_i \in \exp(\lambda_i)$ and the corresponding random vector is in MME*. In the bivariate case an MME* representation is given by
$$\left((1,0), \begin{bmatrix} -\frac{1}{1-p} & \frac{1}{1-p} \\ \frac{p}{1-p} & -\frac{1}{1-p} \end{bmatrix}, \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \right)$$

III. As we can write an integer shaped gamma distributed random variable as a sum of independent exponential terms, it is natural to construct multivariate random variables where some exponential terms are shared among the variables in the multivariate vector. A bivariate distribution of two Erlang(2) distributed random variables can thus be formally expressed by

$$X_i = \frac{Y_i + Z}{\lambda_i} \ ,$$

which makes $X_i \in \text{Erlang}_2(\lambda_i)$ An MME* representation for this case is

$$\left((1,0,0), \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \right)$$

In the remainder of this section we first present a review of multivariate exponential and gamma distributions according to this characterization, then move on to discuss some generic constructions which also happens to be in MME*, and present some distributions that do not have a minimal order MME* representation. Finally we mention a number of multivariate exponential and gamma distributions without rational Laplace transform. These distributions are not MVME distributions.

11.4.1 Sharing of exponential terms

11.5 Specific distributions

♠ McKays bivariate gamma

The distributions in this section is constructed using the additivity property of the gamma distribution. As a gamma distributed random variable Y_i with shape parameter m_i and mean $\frac{1}{\lambda_i}$ can be written as $Y_i = \frac{1}{\lambda_i}\sum_{j=1}^{n} Z_{ij}$ where Z_{ij} are iid $\exp(1)$ distributed random variables, it is an obvious idea to construct random vectors with gamma distributed marginals sharing several component variables Z_{ij}. Here we have $Y_1 = \frac{1}{\lambda_1}\sum_{j=1}^{m_1}$, $Y_2 = \frac{1}{\lambda_1}\sum_{j=1}^{m_2}$, with $m_2 \geq m_1$. An MPH* representation for this distribution is given by

$$\left((1,0,0,\ldots,0,0), \begin{bmatrix} -1 & 1 & 0 \ldots & 0 & 0 \\ 0 & -1 & 1 \ldots & 0 & 0 \\ 0 & 0 & -1 \ldots & 0 & 0 \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ 0 & 0 & 0 \ldots & -1 & 1 \\ 0 & 0 & 0 \ldots & 0 & -1 \end{bmatrix}, \begin{bmatrix} \lambda_1^{-1} & \lambda_2^{-1} \\ \lambda_1^{-1} & \lambda_2^{-1} \\ \lambda_1^{-1} & \lambda_2^{-1} \\ \vdots & \vdots \\ 0 & \lambda_2^{-1} \\ 0 & \lambda_2^{-1} \end{bmatrix} \right)$$

or expressed more compactly

$$\left(\boldsymbol{\delta}_{1,m_1+m_2}, \boldsymbol{E}_{m_1+m_2}, \begin{bmatrix} \lambda_1^{-1}\boldsymbol{e}_{m_1} & \lambda_2^{-1}\boldsymbol{e}_{m_1} \\ \boldsymbol{0} & \lambda_2^{-1}\boldsymbol{e}_{m_2} \end{bmatrix} \right)$$

The Laplace-Stieltjes transform of this distribution is seen to be

$$\frac{1}{\left(1+\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_1}\right)^{m_1}\left(1+\frac{s_2}{\lambda_1}\right)^{m_2-m_1}}$$

with corresponding density function for $m_2 > m_1$.

$$f(y_1,y_2) = \lambda_1\lambda_2\frac{(\lambda_1 y_1)^{m_1-1}(\lambda_2 y_2-\lambda_2 y_2)^{m_2-m_1-1}}{(m_1-1)!(m_2-m_1-1)!}e^{-(\lambda_2 y_2-\lambda_1 y_1)}, \lambda_2 y_2 > \lambda_1 y_1$$

and correlation coefficient $\mathrm{Corr}(Y_1,Y_2) = \frac{m_1}{m_1+m_2}$. The distribution is known as McKays bivariate gamma when $\lambda_1 = \lambda_2$.

♠ Cheriyan and Ramabhadran's bivariate Gamma

Another construction is $Y_i = \frac{1}{\lambda_i}\left(\sum_{j=1}^{m_0} Z_{0j} + \sum_{j=1}^{m_i} Z_{ij}\right)$, with $m_i \geq 0$. The distribution has MPH* representation

$$\left(\boldsymbol{\delta}_{1,m_0+m_1+m_2}, \boldsymbol{E}_{m_0+m_1+m_2}, \begin{bmatrix} \lambda_1^{-1}\boldsymbol{e}_{m_0} & \lambda_2^{-1}\boldsymbol{e}_{m_0} \\ \lambda_1^{-1}\boldsymbol{e}_{m_1} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2^{-1}\boldsymbol{e}_{m_2} \end{bmatrix}\right)$$

and Laplace-Stieltjes transform

$$\frac{1}{\left(1+\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_2}\right)^{m_0}\left(1+\frac{s_1}{\lambda_1}\right)^{m_1}\left(1+\frac{s_2}{\lambda_2}\right)^{m_2}}$$

The density can be derived by conditioning on the sojourn time (Y_0) in the states contributing to both Y_1 and Y_2.

$$f(y_1,y_2) = \frac{e^{-\sum_{i=1}^2 x_i}}{(m_{\boldsymbol{e}}-1)!\prod_{i=1}^2(m_{\boldsymbol{\delta}_i}-1)!}\int_0^{\min(x_i)} y^{m_{\boldsymbol{e}}-1}\left(\prod_{i=1}^2(x_i-y)^{m_{\boldsymbol{\delta}_i}-1}\right)e^{(n-1)y}\mathrm{d}y$$

The distribution can be rewritten to get

$$f(y_1,y_2) = \lambda_1\frac{(\lambda_1 y_1)^{m_0+m_1}}{(m_0+m_1-1)!}e^{-\lambda_1 y_1}\lambda_2\frac{(\lambda_2 y_2)^{m_0+m_2}}{(m_0+m_2-1)!}e^{-\lambda_2 y_2}$$
$$\left\{1+\sum_{r=}^{\infty}\frac{r!m_0^{[r]}}{(m_0+m_1)^{[r]}(m_0+m_2)^{[r]}}L_r^{m_0+m_1-1}(\lambda_1 y_1)L_r^{m_0+m_2-1}(\lambda_1 y_2)\right\}$$

♠ Some tedious work needed here in order to include the λ_s maybe not that hard after all?

♠ Distributions of Type 3

The most general way of having shared phases is due to Prèkopa and Szàntai's **48.3.4**. The 2^n-1 possible subsets 2^n-1 independent gamma variables $Y_{\boldsymbol{\iota}}$, where $\boldsymbol{\iota}$ is an $n-$dimensional vector of 0's and 1's. The MME* representation is straightforward. We illustrate with the case $n = 3$.

$$\left(\boldsymbol{\delta}_1, \boldsymbol{E}(\lambda \boldsymbol{e}), \begin{bmatrix} \boldsymbol{e}_{m_{111}} & \boldsymbol{e}_{m_{111}} & \boldsymbol{e}_{m_{111}} \\ \boldsymbol{e}_{m_{110}} & \boldsymbol{e}_{m_{110}} & \mathbf{0} \\ \boldsymbol{e}_{m_{101}} & \mathbf{0} & \boldsymbol{e}_{m_{101}} \\ \mathbf{0} & \boldsymbol{e}_{m_{011}} & \boldsymbol{e}_{m_{011}} \\ \boldsymbol{e}_{m_{100}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{e}_{m_{010}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{e}_{m_{001}} \end{bmatrix}\right) \tag{11.3}$$

The $\boldsymbol{X} = \boldsymbol{AY}$ construction, where Y_i are independent gamma variables.

♠ Cheriyan and Ramabhadran's multivariate Gamma **48.3.1**

For integer valued shape parameters the distributions are in MME*. The distribution is obtained from (11.3) when the only non–zero values of $m_{\boldsymbol{\iota}}$ are $m_{\boldsymbol{e}}$ and $m_{\boldsymbol{\delta}_i}$.

$$f(x_1,\ldots,x_n) = \frac{e^{-\sum_{i=1}^n x_i}}{(m_{\boldsymbol{e}}-1)!\prod_{i=1}^n (m_{\boldsymbol{\delta}_i}-1)!} \int_0^{\min(x_i)} y^{m_{\boldsymbol{e}}-1} \left(\prod_{i=1}^n (x_i-y)^{m_{\boldsymbol{\delta}_i}-1}\right) e^{(n-1)y} \mathrm{d}y$$

$$L(\boldsymbol{s}) = \frac{1}{(1+\sum_{i=1}^n s_i)^{m_{\boldsymbol{e}}}} \frac{1}{\prod_{i=1}^n (1+s_i)^{m_{\boldsymbol{\delta}_i}}}$$
$$\left(\boldsymbol{\delta}_{1,\sum_{i=0}^n m_i}, \boldsymbol{E}_{\sum_{i=0}^n m_i}, \begin{bmatrix} \lambda_1^{-1}\boldsymbol{e}_{m_0} & \lambda_2^{-1}\boldsymbol{e}_{m_0} & \ldots & \lambda_n^{-1}\boldsymbol{e}_{m_0} \\ \lambda_1^{-1}\boldsymbol{e}_{m_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \lambda_2^{-1}\boldsymbol{e}_{m_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots\vdots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \lambda_n^{-1}\boldsymbol{e}_{m_n} \end{bmatrix}\right) .$$

The bivariate version of the distribution is described in Section 2.2 of [6] **48.2.2**. This distribution is identical to Prèkopa and Szàntai's Bivariate Gamma (**48.2.8**). The joint pdf is given in (**48.2.2**) while the CDF is given in (**48.2.8**). The bivariate distribution for $\lambda = 1$ is also mentioned as the second distribution of Section **48.2.10**.

11.5.1 Marshall Olkin's bivariate exponential

Two components that can break down either caused to individual shocks occurring according to a Poisson process with rate $\lambda_i, i = 1,2$ respectively or due to shocks that affects both types simultaneously with rate λ_{12}. An MPH* representation appears immediately as

$$\left((1,0,0), \begin{bmatrix} -(\lambda_1+\lambda_2+\lambda_{12}) & \lambda_2 & \lambda_1 \\ 0 & -(\lambda_1+\lambda_{12}) & 0 \\ 0 & 0 & -(\lambda_2+\lambda_{12}) \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \right) .$$

These distributions are also in MPH and can be seen as the bivariate distributions motivating the MPH class. Another representation of more generic form and less natural seen from an applied view is

$$\left((1,0,0),\begin{bmatrix}-1 & \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}} & \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\\ 0 & -1 & 0\\ 0 & 0 & -1\end{bmatrix},\begin{bmatrix}\frac{1}{\lambda_1+\lambda_2+\lambda_{12}} & \frac{1}{\lambda_1+\lambda_2+\lambda_{12}}\\ \frac{1}{\lambda_1+\lambda_{12}} & 0\\ 0 & \frac{1}{\lambda_2+\lambda_{12}}\end{bmatrix},\right). \tag{11.4}$$

We derive the Laplace-Stieltjes transform as

$$\frac{1}{1+\frac{s_1+s_2}{\lambda_1+\lambda_2+\lambda_{12}}}\left(\frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}}+\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\frac{1}{1+\frac{s_1}{\lambda_1+\lambda_{12}}}+\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\frac{1}{1+\frac{s_2}{\lambda_2+\lambda_{12}}}\right)$$

The marginals are exponentials with intensities $\lambda_i+\lambda_{12}$ which can be seen from the construction or by inserting $s_{3-i}=0$ into the expression for Laplace-Stieltjes transform. The distribution is not absolutely continuous as whenever $\lambda_{12}>0$ there will be a non–zero probability of observing $Y_1=Y_2$. The joint survival function is given by

$$\bar{F}(y_1,y_2)=P(Y_1>y_1,Y_2>y_2)=e^{-(\lambda_1 y_1+\lambda_2 y_2+\lambda_{12}\max(y_1,y_2)}.$$

$$\bar{F}(y_1,y_2)=\begin{cases}e^{-((\lambda_1+\lambda_{12})y_1+\lambda_2 y_2} & \text{for } 0\le y_2\le y_1\\ e^{-(\lambda_1 y_1+(\lambda_2+\lambda_{12})y_2} & \text{for } 0\le y_1\le y_2\end{cases}$$

The joint density for $Y_1\neq Y_2$ is given by

$$f(y_1,y_2)=\begin{cases}\lambda_2(\lambda_1+\lambda_{12})e^{-(\lambda_1+\lambda_{12})y_1-\lambda_2)y_2} & \text{for } 0\le y_2<y_1\\ \lambda_1(\lambda_2+\lambda_{12})e^{-\lambda_1 y_1-(\lambda_2+\lambda_{12})y_2} & \text{for } 0\le y_1<y_2\end{cases}.$$

The singularity on the line $y_1=y_2$

$$f(y,y)=\lambda_{12}e^{-\lambda_{12}y}$$

For the multivariate case we consider the joint life time of n components. Further, it is assumed that the risk that any specific subset of the components will be killed simultaneously is constant, i.e. not dependent on how many components are still functioning. These assumptions lead to exponentially distributed individual lifetimes. The joint survival function can be expressed by

$$\bar{F}_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=\exp\left(-\sum_{i=1}^{n}\lambda_i x_i-\sum_{i_1<i_2}\lambda_{i_1 i_2}\max(x_{i_1},x_{i_2})\right.$$
$$\left.-\sum_{i_1<i_2<i_3}\lambda_{i_1 i_2 i_3}\max(x_{i_1},x_{i_2},x_{i_3})\ldots-\lambda_{12\ldots n}\max(x_1,\ldots,x_n)\right).$$

To get an an MPH* formulation we first define a Markov chain with up to 2^n states. The states are denoted by $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_n)$ where $\iota_i \in \{0,1\}$, such that $\iota_i = 1$ if component i is still functioning and $\iota_i = 0$ if its life time has expired. The rate of leaving state $\boldsymbol{\iota}$ is $\mu_{\boldsymbol{\iota}} = \sum_{\boldsymbol{\kappa} \leq \boldsymbol{\iota}} \gamma_{\boldsymbol{\kappa}}$, where $\gamma_{\boldsymbol{\kappa}}$ is the intensity that kills exactly all components with a non–zero entry of $\boldsymbol{\kappa}$. Thus the probability that the chain moves from $\boldsymbol{\iota}$ to $\boldsymbol{\iota} - \boldsymbol{\kappa}$ is $p_{\boldsymbol{\iota},\boldsymbol{\iota}-\boldsymbol{\kappa}} = \frac{\gamma_{\boldsymbol{\kappa}}}{\mu_{\boldsymbol{\iota}}}$. The row in $\boldsymbol{R}$ corresponding to $\boldsymbol{\iota}$ is simply $\frac{\boldsymbol{\iota}}{\mu_{\boldsymbol{\iota}}}$.

Go from any state ι_{i_1} to ι_{i_2} whenever $\iota_{i_1} > \iota_{i_2}$ with parameter We state the joint survival function These distributions are in MPH*. A representation for $n = 3$ is

$$\left(\boldsymbol{\delta}_1, \begin{bmatrix} -1 & p_{111,110} & p_{111,110} & p_{111,110} & p_{111,110} & p_{111,110} & p_{111,110} \\ 0 & -1 & 0 & 0 & p_{110,100} & p_{110,010} & 0 \\ 0 & 0 & -1 & 0 & p_{101,100} & 0 & p_{101,001} \\ 0 & 0 & 0 & -1 & 0 & p_{011,010} & p_{011,001} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} \mu_{111}^{-1} & \mu_{111}^{-1} & \mu_{111}^{-1} \\ \mu_{110}^{-1} & \mu_{110}^{-1} & 0 \\ \mu_{101}^{-1} & 0 & \mu_{101}^{-1} \\ 0 & \mu_{011}^{-1} & \mu_{011}^{-1} \\ \mu_{100}^{-1} & 0 & 0 \\ 0 & \mu_{010}^{-1} & 0 \\ 0 & 0 & \mu_{001}^{-1} \end{bmatrix} \right). \tag{11.5}$$

$$\left(\boldsymbol{\delta}_1, \begin{bmatrix} -\lambda & \lambda_3 & \lambda_2 & \lambda_1 & \lambda_{23} & \lambda_{13} & \lambda_{12} \\ 0 & -\lambda+\lambda_3 & 0 & 0 & \lambda_2+\lambda_{23} & \lambda_1+\lambda_{13} & 0 \\ 0 & 0 & -\lambda+\lambda_2 & 0 & \lambda_3+\lambda_{23} & 0 & \lambda_1+\lambda_{12} \\ 0 & 0 & 0 & -\lambda+\lambda_1 & 0 & \lambda_3+\lambda_{13} & \lambda_2+\lambda_{12} \\ 0 & 0 & 0 & 0 & -\tau_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\tau_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\tau_3 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).$$

Here we have taken $\lambda = \lambda_1+\lambda_2+\lambda_3+\lambda_{12}+\lambda_{13}+\lambda_{23}+\lambda_{123}$, $\tau_1 = \lambda_1+\lambda_{12}+\lambda_{13}+\lambda_{123}$, $\tau 2 = \lambda_2+\lambda_{12}+\lambda_{23}+\lambda_{123}$, $\tau_3 = \lambda_3+\lambda_{13}+\lambda_{23}+\lambda_{123}$, and $\tau_4 = \lambda_2+\lambda_3+\lambda_{23}$.

Olkin and Tong's Multivariate Exponential ([6], **47.3.4**) is obtained from Marshall–Olkin's multivariate exponential distributions when $(\lambda_i, i = 1 \ldots n)$, $(\lambda_{12\ldots k_1}, \lambda_{k_1+1,\ldots,k_2}, \ldots, \lambda_{k_{r-1},\ldots,n})$, and, $\lambda_{12\ldots n}$ are the only non–zero entries of $\boldsymbol{T}$.

The representations (11.4) and (11.5) demonstrates another general principle in forming multivariate exponential and gamma distributions, apart from the sharing of exponential phases utilized in Section **??** besides the interpretation of the MO distributions as arising from applied considerations. This construction has been used in other constructions of bivariate exponentials.

11.5.2 Raftery's bivariate Exponential

The construction is such that $Y_i = \frac{1-\rho}{\lambda_i} X_i + J\frac{Z}{\lambda_i}$ where J is a binary variable with $\mathbb{P}(J=1) = \rho$, and where $X_i and Z$ are mutually independent $\exp(1)$ variables. In the original model some more parameters are allowed, however these parameters are redundant and do not add genuine degrees of freedom. An MPH* representation for this distribution is

$$\left(\boldsymbol{\delta}_1, \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & \rho \\ 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} \frac{1-\rho}{\lambda_1} & 0 \\ 0 & \frac{1-\rho}{\lambda_2} \\ \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \end{bmatrix}\right),$$

with Laplace-Stieltjes transform

$$\frac{1}{1+\frac{s_1(1-\rho)}{\lambda_1}} \frac{1}{1+\frac{s_2(1-\rho)}{\lambda_2}} \left(1-\rho+\frac{\rho}{1+\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_2}}\right) = \frac{1+(1-\rho)\left(\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_2}\right)}{\left(1+\frac{s_1(1-\rho)}{\lambda_1}\right)\left(1+\frac{s_2(1-\rho)}{\lambda_2}\right)\left(1+\frac{s_1}{\lambda_1}+\frac{s_2}{\lambda_2}\right)}.$$

To derive the joint density we condition on J and for $J=0$ we further condition on $Z=z$.

$$\begin{aligned} f(y_1,y_2)) &= \mathbb{P}(J=0)f(y_1,y_2|J=0)+\mathbb{P}(J=1)f(y_1,y_2|J=1) \\ &= \mathbb{P}(J=0)f(y_1,y_2|J=0)+\mathbb{P}(J=1)\int_{z=0}^{\min(x_1,x_2)} f(y_1,y_2|J=1,Z=z)e^{-z}\mathrm{d}z. \end{aligned}$$

Now conditioned on $J=0$, Y_1 and Y_2 are independent exponentials with intensities $\frac{\lambda_i}{1-\rho}$ respectively.

$$\begin{aligned} f(y_1,y_2)) &= (1-\rho)\frac{\lambda_1}{1-\rho}e^{-\frac{\lambda_1 y_1}{1-\rho}}\frac{\lambda_2}{1-\rho}e^{-\frac{\lambda_2 y_2}{1-\rho}}+\rho\int_0^{\min(\lambda_1 y_1,\lambda_2 y_2)} e^{-z}\frac{\lambda_1\lambda_2}{(1-\rho)^2}e^{-\frac{(y_1\lambda_1-z)+(y_2\lambda_2-z)}{1-\rho}}\mathrm{d}z \\ &= \frac{\lambda_1\lambda_2}{1-\rho}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}\left(1+\frac{\rho}{1-\rho}\int_0^{\min(\lambda_1 y_1,\lambda_2 y_2)} e^{-z\left(1-\frac{2}{1-\rho}\right)}\mathrm{d}z\right) \\ &= \frac{\lambda_1\lambda_2}{1-\rho}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}\left(1+\frac{\rho}{1-\rho}\left[\frac{1-\rho}{1+\rho}e^{\frac{1+\rho}{1-\rho}z}\right]_0^{\min(\lambda_1 y_1,\lambda_2 y_2)}\right) \\ &= \frac{\lambda_1\lambda_2}{1-\rho}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}\left(\frac{1}{1+\rho}+\frac{\rho}{1+\rho}e^{\frac{1+\rho}{1-\rho}\min(\lambda_1 y_1,\lambda_2 y_2)}\right) \\ &= \begin{cases} \frac{1-\rho}{1+\rho}\frac{\lambda_1\lambda_2}{(1-\rho)^2}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}+\frac{\rho\lambda_1\lambda_2}{1-\rho^2}e^{-\frac{\lambda_2 y_2-\rho\lambda_1 y_1}{1-\rho}} & \text{for } \lambda_2 y_2 > \lambda_1 y_1 \\ \frac{1-\rho}{1+\rho}\frac{\lambda_1\lambda_2}{(1-\rho)^2}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}+\frac{\rho\lambda_1\lambda_2}{1-\rho^2}e^{-\frac{\lambda_1 y_1-\rho\lambda_2 y_2}{1-\rho}} & \text{for } \lambda_1 y_1 > \lambda_2 y_2 \end{cases} \end{aligned}$$

Now using tedious but standard calculations we obtain the joint survival function to be

$$\bar{F}(y_1,y_2) = \begin{cases} e^{-\lambda y_2}+\frac{1-\rho}{1+\rho}\left(e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}-e^{-\frac{\lambda_2 y_2-\rho\lambda_1 y_1}{1-\rho}}\right) & \text{for } \lambda_2 y_2 > \lambda_1 y_1 \\ e^{-\lambda y_1}+\frac{1-\rho}{1+\rho}\left(e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-\rho}}-e^{-\frac{\lambda_1 y_1-\rho\lambda_2 y_2}{1-\rho}}\right) & \text{for } \lambda_1 y_1 > \lambda_2 y_2, \end{cases}$$

or equivalently

$$\bar{F}(y_1,y_2) = \begin{cases} e^{-\lambda y_2} + \frac{1-\rho}{1+\rho} e^{-\frac{\lambda_2 y_2}{1-\rho}} \left(e^{-\frac{\lambda_1 y_1}{1-\rho}} - e^{\frac{\rho \lambda_1 y_1}{1-\rho}} \right) & \text{for } \lambda_2 y_2 > \lambda_1 y_1 \\ e^{-\lambda y_1} + \frac{1-\rho}{1+\rho} e^{-\frac{\lambda_1 y_1}{1-\rho}} \left(e^{-\frac{\lambda_2 y_2}{1-\rho}} - e^{\frac{\rho \lambda_2 y_2}{1-\rho}} \right) & \text{for } \lambda_1 y_1 > \lambda_2 y_2, \end{cases}$$

where the exponential marginals become obvious from the survival function.

♠ I have been cheating with the survival function, using maple, not verifying by hand - see program: /home/bfn/divtele/mvme/maple/raftery2dsurvival.maple

11.5.3 Raftery's Multivariate Exponential

The construction is such that $X_i = (1-\pi_i)Y_i + Z_{J_i}$ where the vector $(J_1,\ldots,J_n)$ takes values in $\{0,1,\ldots,m\}^n$, Y_i and Z_j are mutually independent $\exp(1)$ variables, and $P(J_i = 0) = \pi_i$. However, it is possible to give much more compact MME* representations. The representation

$$\left(\boldsymbol{\delta}_1, \begin{bmatrix} -\frac{1}{p_{00}+p_{01}} & \frac{1}{p_{00}+p_{01}} & 0 & 0 & 0 \\ 0 & -\frac{1}{p_{00}+p_{10}} & \frac{p_{10}}{p_{00}+p_{10}} & \frac{p_{01}}{p_{00}+p_{10}} & \frac{p_{11}}{p_{00}+p_{10}} \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

is valid for $n = 2$ and $m = 1$. while the construction of [2] uses 8 states. For larger values of n and m the reduction in dimension is even larger. The representation above is chosen to demonstrate the pattern for general values of n,m. However, for $n = 2$ ([6], **47.2.9**) there is a more compact representation with three phases.

$$\left(\left[\begin{array}{cccc|ccc|ccc|ccccc} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho'_{14}\rho_{23} & \rho'_{13}\rho_{24} & \rho'_{12}\rho_{34} & \rho_{123} & \rho_{124} & \rho_{134} & \rho_{234} & \rho_{1234} \\ \hline 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \rho_{34} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \rho_{24} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & \rho_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{array}\right], \left[\begin{array}{cccc} \frac{1-\rho_1}{\lambda_1} & 0 & 0 & 0 \\ 0 & \frac{1-\rho_2}{\lambda_2} & 0 & 0 \\ 0 & 0 & \frac{1-\rho_3}{\lambda_3} & 0 \\ 0 & 0 & 0 & \frac{1-\rho_4}{\lambda_4} \\ \hline \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & 0 & 0 \\ \frac{1}{\lambda_1} & 0 & \frac{1}{\lambda_3} & 0 \\ \frac{1}{\lambda_1} & 0 & 0 & \frac{1}{\lambda_4} \\ \hline 0 & \frac{1}{\lambda_2} & \frac{1}{\lambda_3} & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \frac{1}{\lambda_4} \\ 0 & 0 & \frac{1}{\lambda_3} & \frac{1}{\lambda_4} \\ \hline \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & \frac{1}{\lambda_3} & 0 \\ \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & 0 & \frac{1}{\lambda_4} \\ \frac{1}{\lambda_1} & 0 & \frac{1}{\lambda_3} & \frac{1}{\lambda_4} \\ 0 & \frac{1}{\lambda_2} & \frac{1}{\lambda_3} & \frac{1}{\lambda_4} \\ \hline \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & \frac{1}{\lambda_3} & \frac{1}{\lambda_4} \end{array}\right] \right)$$

♠ I am not really sure which way is best to present the representation

with $\boldsymbol{R}$ equal to

$$\left[\begin{array}{cccc|ccc|ccc|ccc} \frac{1-\rho_1}{\lambda_1} & 0 & 0 & 0 & \frac{1}{\lambda_1} & \frac{1}{\lambda_1} & \frac{1}{\lambda_1} & 0 & 0 & 0 & \frac{1}{\lambda_1} & \frac{1}{\lambda_1} & \frac{1}{\lambda_1} & 0 & \frac{1}{\lambda_1} \\ 0 & \frac{1-\rho_2}{\lambda_2} & 0 & 0 & \frac{1}{\lambda_2} & 0 & 0 & \frac{1}{\lambda_2} & \frac{1}{\lambda_2} & 0 & \frac{1}{\lambda_2} & \frac{1}{\lambda_2} & 0 & \frac{1}{\lambda_2} & \frac{1}{\lambda_2} \\ 0 & 0 & \frac{1-\rho_3}{\lambda_3} & 0 & 0 & \frac{1}{\lambda_3} & 0 & \frac{1}{\lambda_3} & 0 & \frac{1}{\lambda_3} & \frac{1}{\lambda_3} & 0 & \frac{1}{\lambda_3} & \frac{1}{\lambda_3} & \frac{1}{\lambda_3} \\ 0 & 0 & 0 & \frac{1-\rho_4}{\lambda_4} & 0 & 0 & \frac{1}{\lambda_4} & 0 & \frac{1}{\lambda_4} & \frac{1}{\lambda_4} & 0 & \frac{1}{\lambda_4} & \frac{1}{\lambda_4} & \frac{1}{\lambda_4} & \frac{1}{\lambda_4} \end{array}\right]$$

where $\rho'_{ij} = 1-\rho_{ij}$ and $\rho_i = \sum_{j\neq i}\rho_{ij} + \sum_{1\leq j<k\leq 4, j,k\neq i}\rho_{ijk} + \rho_{1234}$ and $\rho_{ij} = \rho_{ji}$ etc..
The joint Laplace-Stieltjes transform is

$$\prod_{i=1}^{4} \frac{1}{1+\frac{s_i(1-\rho_i)}{\lambda_i}} \left(1-\rho + \sum_{\substack{j,k,\ell \in \{2,3,4\} \\ j\neq k, j\neq \ell, k<\ell}} \frac{\rho_{1j}}{1+\frac{s_1}{\lambda_1}+\frac{s_j}{\lambda_j}} \frac{\rho_{k\ell}}{1+\frac{s_k}{\lambda_k}+\frac{s_\ell}{\lambda_\ell}} \right.$$

$$\left. + \sum_{\substack{1\leq j<k\leq 4 \\ \ell,m\in\{1,2,3,4\}\setminus\{j,k\}, \ell<m}} \frac{\rho_{jk}\rho'_{\ell m}}{1+\frac{s_j}{\lambda_j}+\frac{s_k}{\lambda_k}} + \sum_{1\leq j<k<\ell\leq 4} \frac{\rho_{jk\ell}}{1+\frac{s_j}{\lambda_j}+\frac{s_k}{\lambda_k}+\frac{s_\ell}{\lambda_\ell}} + \frac{\rho_{1234}}{1+\sum_{j=1}^{4}\frac{s_j}{\lambda_j}} \right).$$

with

$$\rho = \sum_{j=2}^{4} \rho_{1j} + \sum_{\substack{j \in \{2,3,4\} \\ k,\ell \in \{2,3,4\}\setminus\{j\}, k<\ell}} \rho'_{1j}\rho_{k\ell} + \sum_{1\le j<k<\ell\le 4} \rho_{jk\ell} + \rho_{1234}$$

From the MPH* representation and from the form of the Laplace-Stieltjes transform it should be clear that the joint densities and survival functions can be derived closely following the approach used for the bivariate case.

These distributions also have MPH representations with $(m+2)^n - 1$ phases [2] which arise from considering a shock model inspired by the shock model leading to the Marshall-Olkin distributions We illustrate the development for the case $n = 2$ and $n = 3$. Although the MPH* representations are much more compact the MPH representation could be of interest in certain applications.

♠ I am not sure I want to express the shock model in any detail. With the MPH representation all variables need to have the same rate, I don't see how that can be overcome in the MPH framework. It would probably be better to spend some time discussing the MPH* framework in light of the partition function.

11.5.4 Dussauchoy and Berlands

Dusber in the exponential version is really MO with one of the λ_i's set to equal to zero and then the bivariate gamma is an addition of a number of MO-distributed random vectors.

Dussauchoy and Berland's multivariate Gamma 48.3.8

[3] discuss how to obtain variables using an

♠ not really Gram-Schmidt

orthogonalization. Here X_i are gamma distributed with shape factors m_i with $m_i \ge m_{i-1}, \forall i$. The construction is such that X_j is independent of $X_i - \frac{\rho_{ij}\lambda_j}{\lambda_i}X_j, \qquad \forall n \ge i > j \ge 1$. The joint LST for this distribution is given by

$$\prod_{i=1}^{n} \frac{1 + \sum_{j=i+1}^{n} \rho_{ij} \frac{s_j}{\lambda_j}}{1 + \sum_{j=i}^{n} \rho_{ij} \frac{s_j}{\lambda_j}}$$

with $\rho_{ii} = 1$.

It is not clear

♠ to us

that [3] actually proof the existence of such random variables. However, we have not investigated this in detail like we have not found explicit representations in general.

♠ It really annoys me, that we haven't settled this

11.5.4.1 Bivariate case

For the bivariate case ([6],**48.2.10**) we have that $X_2 - \rho X_1$ and X_1 are independent. The simplest case is the case of a bivariate exponential. The LST is

$$\frac{1+\rho\frac{s_2}{\lambda_2}}{(1+\frac{s_2}{\lambda_2})(1+\frac{s_1}{\lambda_1}+\rho\frac{s_2}{\lambda_2})} = \frac{1}{1+\frac{s_1}{\lambda_1}+\rho\frac{s_2}{\lambda_2}}\left(\rho+\frac{1-\rho}{1+\frac{s_2}{\lambda_1}}\right)$$

An MPH* representation for this distribution is

$$\left((1,0),\begin{bmatrix}-1 & 1-\rho\\ 0 & -1\end{bmatrix},\begin{bmatrix}\frac{1}{\lambda_1} & \frac{\rho}{\lambda_2}\\ 0 & \frac{1}{\lambda_2}\end{bmatrix}\right)$$

so in this case the distribution simplifies to a Type-1 distribution of the simplest and most basic form, which can also be expressed as an e.g. Raftery distribution. For the bivariate gamma the idea is extended such that a number of bivariate exponentials is added together, further the second component can have a number of independent phases on top. Two straightforward representations appear according to whether one decides to take the exponentials serially or first have all the contributions where the first variable is involved then secondly all the contributions with only contributions to the second variable or just couple the bivariate exponentials serially. The representation is given by

$$\left((1,0,0,0),\begin{bmatrix}-1 & 1 & 0 & 0 & 0\\ 0 & -1 & (1-\rho)^2 & 2\rho(1-\rho) & \rho^2\\ 0 & 0 & -1 & 1 & 0\\ 0 & 0 & 0 & -1 & 1\\ 0 & 0 & 0 & 0 & -1\end{bmatrix},\begin{bmatrix}\frac{1}{\lambda_1} & \frac{\rho}{\lambda_2}\\ \frac{1}{\lambda_1} & \frac{\rho}{\lambda_2}\\ 0 & \frac{1}{\lambda_2}\\ 0 & \frac{1}{\lambda_2}\\ 0 & \frac{1}{\lambda_2}\end{bmatrix}\right).$$

In either case the joint density and joint survival function can be derived by conditioning on the binomial choice from each bivariate exponential. How many cases of independent contributions are involved. The joint Laplace transform is equal to

$$\left(\frac{1+\frac{\rho s_2}{\lambda_2}}{1+\frac{s_1}{\lambda_1}+\rho\frac{s_2}{\lambda_2}}\right)^{m_1}\left(\frac{1}{1+\frac{s_2}{\lambda_2}}\right)^{m_2},$$

♠ I do not think, that it is necessary to give the bivariate transform

We give an MME* representation, which is even in MPH, in the case where $m_1 = m_2 = 3$. or

$$\left((1,0,0,0), \begin{bmatrix} -1 & 1-\rho & \rho & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1-\rho & \rho \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \begin{bmatrix} \frac{1}{\lambda_1} & \frac{\rho}{\lambda_2} \\ 0 & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & \frac{\rho}{\lambda_2} \\ 0 & \frac{1}{\lambda_2} \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \right) .$$

The general bivariate construction for other values of m_1 and m_2 should be obvious from this example.

11.5.4.2 Trivariate case

$$\mathbb{E}\left(e^{-(s_1Y_1+s_2Y_2+s_3Y_3)}\right) = \frac{(1+\rho_{12}s_2+\rho_{13}s_3)(1+\rho_{23}s_3)}{(1+s_3)((1+s_1+\rho_{12}s_2+\rho_{13}s_3)(1+s_2+\rho_{23}s_3)}$$

$$\begin{aligned} Y_2 &= Z_{12}+\rho_{12}Y_1 \\ Y_3 &= Z_{13}+\rho_{13}Y_1 \\ Y_3 &= Z_{23}+\rho_{23}Y_2 \\ Y_3 &= Z_{23}++\rho_{23}Z_{12}+\rho_{12}\rho_{23}Y_1 \end{aligned}$$

with $\mathbb{E}(Z_{12}Y_1) = \mathbb{E}(Z_{12})\mathbb{E}(Y_1)$, $\mathbb{E}(Z_{13}Y_1) = \mathbb{E}(Z_{13})\mathbb{E}(Y_1)$, and $\mathbb{E}(Z_{23}Y_2) = \mathbb{E}(Z_{23})\mathbb{E}(Y_2)$. We find

$$Z_{23} = Z_{13} - \rho_{23}Z_{12} + (\rho_{13} - \rho_{12}\rho_{23})Y_1$$

or

$$(\rho_{13} - \rho_{12}\rho_{23})Y_1 = Z_{23} - Z_{13} + \rho_{23}Z_{12}$$

All the bivariate marginals are as in the previous section, with ρ equal to ρ_{12}, ρ_{13}, and ρ_{23} respectively. The joint trivariate density, however, is not straightforward. We will go along and show, that there is no MME$^*(3)$ representation.

11.5.5 Kibble type distributions

A number of distributions have been named being basically reparameterizations of the same underlying structure. The first description of these types of distributions is due to Kibble [5]. The construction is based on the fact that a geometric distribution of Erlang-distributions is an exponential distribution. This has been used previously, maybe most prominently as the prime example of discrete phase type mixtures of convolutions of phase type distributions as described in 3.1.

♠ I have introduced the type II distributions here? or maybe I shouldn't? should try to write something on the different types today (17/5/11) today also.

This is thus the prime example of Type-2 distributions. We will first consider the bivariate exponential case, as the general multivariate case follows by straightforward generalizations, yet leading to somewhat complicated expressions which might obscure the underlying quite simple idea. Let us first take a view of the MPH* representation of the bivariate exponential, which is given by

$$\left((1,0),\begin{bmatrix} -1 & 1 \\ p & -1 \end{bmatrix},\begin{bmatrix} \frac{1-p}{\lambda_1} & 0 \\ 0 & \frac{1-p}{\lambda_2} \end{bmatrix}\right). \tag{11.6}$$

As in Section 11.3 We let N denote the geometrically number of times that each state of the two–dimensional Markov chain is visited. The variable Y_1 and Y_2 are independent when conditioned on N, thus for the joint conditioned density we can write

$$f_{(Y_1,Y_2)|N=n}(y_1,y_2) = \frac{\lambda_1}{1-p}\frac{\left(\frac{\lambda_1}{1-p}y_1\right)^{n-1}}{(n-1)!}e^{-\frac{\lambda_1}{1-p}y_1}\frac{\lambda_2}{1-p}\frac{\left(\frac{\lambda_2}{1-p}y_2\right)^{n-1}}{(n-1)!}e^{-\frac{\lambda_2}{1-p}y_2}$$

leading to

$$\begin{aligned} f(y_1,y_2) &= \sum_{n=1}^{\infty} p^{n-1}(1-p)\frac{\lambda_1}{1-p}\frac{\left(\frac{\lambda_1}{1-p}y_1\right)^{n-1}}{(n-1)!}e^{-\frac{\lambda_1}{1-p}y_1}\frac{\lambda_2}{1-p}\frac{\left(\frac{\lambda_2}{1-p}y_2\right)^{n-1}}{(n-1)!}e^{-\frac{\lambda_2}{1-p}y_2} \\ &= \frac{\lambda_1\lambda_2}{1-p}e^{-\frac{\lambda_1 y_1+\lambda_2 y_2}{1-p}}\sum_{n=1}^{\infty}\frac{1}{((n-1)!)^2}\left(2\frac{p\lambda_1 y_1\lambda_2 y2}{2(1-p)^2}\right)^{n-1}, \end{aligned}$$

where the two marginals are exponential with intensities λ_1 and λ_2 respectively. The density can be expressed compactly in terms of Bessel or Laguerre functions. From 11.1 we get

$$f(y_1,y_2)$$

Here we find Correspondingly we can derive the joint survival function $G(y_1,y_2) = \mathbb{P}(Y_1 > y_1, Y_2 > y_2)$ to be

$$G(y_1,y_2) = \sum_{n=1}^{\infty} p^{n-1}(1-p)\left(\sum_{i=0}^{n-1}\frac{\left(\frac{\lambda_1}{1-p}y_1\right)^i}{i!}e^{-\frac{\lambda_1}{1-p}y_1}\right)\left(\sum_{i=0}^{n-1}\frac{\left(\frac{\lambda_2}{1-p}y_2\right)^i}{i!}e^{-\frac{\lambda_2}{1-p}y_2}\right).$$

To obtain the Laplace-Stieltjes transform we could either apply the general result 11.2 or repeat the conditioning argument. For sake of illustration we will proceed with both approaches. Using the conditioning argument we get

$$H(s_1,s_2) = \sum_{n=1}^{\infty} p^{n-1}(1-p)\left(\frac{1}{1+\frac{s_1(1-p)}{\lambda_1}}\frac{1}{1+\frac{s_2(1-p)}{\lambda_2}}\right)^n$$

$$
\begin{aligned}
&= (1-p)\frac{1}{1+\frac{s_1(1-p)}{\lambda_1}}\frac{1}{1+\frac{s_2(1-p)}{\lambda_2}}\sum_{n=1}^{\infty}p^{n-1}\left(\frac{1}{1+\frac{s_1(1-p)}{\lambda_1}}\frac{1}{1+\frac{s_2(1-p)}{\lambda_2}}\right)^{n-1}\\
&= \frac{1-p}{\left(1+\frac{s_1(1-p)}{\lambda_1}\right)\left(1+\frac{s_2(1-p)}{\lambda_2}\right)-p}\\
&= \frac{(1-p)\lambda_1\lambda_2}{(\lambda_1+s_1(1-p))(\lambda_2+s_2(1-p))-p\lambda_1\lambda_2}.
\end{aligned}
$$

Using Equation 11.2 we obtain

$$
\begin{aligned}
H(s_1,s_2) &= (1,0)\left(\begin{bmatrix}\frac{1}{1-p} & \frac{1}{1-p}\\ \frac{p}{1-p} & \frac{1}{1-p}\end{bmatrix}\begin{bmatrix}\frac{s_1(1-p)}{\lambda_1} & 0\\ 0 & \frac{s_2(1-p)}{\lambda_2}\end{bmatrix}+\begin{bmatrix}1&0\\0&1\end{bmatrix}\right)^{-1}\begin{bmatrix}1\\1\end{bmatrix}\\
&= (1,0)\begin{bmatrix}1+\frac{s_1}{\lambda_1} & \frac{s_2}{\lambda_2}\\ \frac{ps_1}{\lambda_1} & 1+\frac{s_2}{\lambda_2}\end{bmatrix}^{-1}\begin{bmatrix}1\\1\end{bmatrix}\\
&= \frac{1}{\left(1+\frac{s_1}{\lambda_1}\right)\left(1+\frac{s_2}{\lambda_2}\right)-\frac{ps_1}{\lambda_1}\frac{s_2}{\lambda_2}}(1,0)\begin{bmatrix}1+\frac{s_2}{\lambda_2} & -\frac{s_2}{\lambda_2}\\ -\frac{ps_1}{\lambda_1} & 1+\frac{s_1}{\lambda_1}\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\
&= = \frac{1}{\left(1+\frac{s_1}{\lambda_1}\right)\left(1+\frac{s_2}{\lambda_2}\right)-\frac{ps_1}{\lambda_1}\frac{s_2}{\lambda_2}}\\
&= \frac{(1-p)\lambda_1\lambda_2}{(1-p)(\lambda_1+s_1)(\lambda_2+s_2)-p(1-p)s_1s_2},
\end{aligned}
$$

and the two expressions are seen to be identical.

As MPH* representations are not unique we note that another representation that might seem more natural in some cases is

$$
\left((1,0),\begin{bmatrix}-\frac{\lambda_1}{1-p} & \frac{\lambda_1}{1-p}\\ p\frac{\lambda_2}{1-p} & -\frac{\lambda_2}{1-p}\end{bmatrix},\begin{bmatrix}1&0\\0&1\end{bmatrix}\right). \tag{11.7}
$$

11.5.5.1 Multivariate case

The basic idea applies equally well in higher dimension. An MPH* representation for an n-dimensional distribution with exponential marginals would be

$$\left((1,0,\ldots,0),\begin{bmatrix} -1 & 1 & 0\ldots & 0 & 0 \\ 0 & -1 & 1\ldots & 0 & 0 \\ 0 & 0 & -1\ldots & 0 & 0 \\ \vdots & \vdots & \vdots\ \vdots\ \vdots & \vdots & \vdots \\ 0 & 0 & 0\ldots & -1 & 1 \\ p & 0 & 0\ldots & 0 & -1 \end{bmatrix}, \begin{bmatrix} \frac{1-p}{\lambda_1} & 0 & 0 & \ldots & 0 & 0 \\ 0 & \frac{1-p}{\lambda_2} & 0 & \ldots & 0 & 0 \\ 0 & 0 & \frac{1-p}{\lambda_3} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1-p}{\lambda_{n-1}} & 0 \\ 0 & 0 & 0 & \ldots & 0 & \frac{1-p}{\lambda_n} \end{bmatrix}\right).$$

As for the bivariate case one could choose a representation, with the identity matrix as the reward matrix. The same methodology as for the bivariate case applies, and we find e.g the joint density to be

$$f(\boldsymbol{y}) = \sum_{j=1}^{\infty} p^{j-1}(1-p)\prod_{i=1}^{n}\frac{\lambda_i}{1-p}\frac{\left(\frac{\lambda_i}{1-p}y_i\right)^{j-1}}{(j-1)!}e^{-\frac{\lambda_i}{1-p}y_i},$$

and the joint survival function is

$$G(\boldsymbol{y}) = \sum_{j=1}^{\infty} p^{j-1}(1-p)\prod_{i=1}^{n}\left(\sum_{k=0}^{j-1}\frac{\left(\frac{\lambda_i}{1-p}y_i\right)^{k}}{k!}e^{-\frac{\lambda_i}{1-p}y_i}\right).$$

Now to obtain multivariate gamma distributions using this construction one adds a number m of exponentially distributed random vectors $\boldsymbol{Y}_i$ such that $\boldsymbol{Y} = \sum_{i=1}^{m}\boldsymbol{Y}_i$. The shape parameter of each component variable of $\boldsymbol{Y}$ can be adjusted by letting some of the vectors $\boldsymbol{Y}_i$ having zero components. All expressions can be derived from probabilistic arguments or using Equation 11.2, however, the expressions can become quite involved in the full generality somewhat blurring the simplicity of the basic construction.

♠ Something with see later section for some examples in the bivariate case

In the most transparent case we let all $\boldsymbol{Y}_i, i = 1,\ldots,m$ be identically distributed and obtain the density of $\boldsymbol{Y} = \sum_{i=1}^{m}\boldsymbol{Y}_i$ as

$$f(\boldsymbol{y}) = \sum_{j=m}^{\infty}\binom{j-1}{m-1}p^{j-m}(1-p)^{m}\prod_{i=1}^{n}\frac{\lambda_i}{1-p}\frac{\left(\frac{\lambda_i}{1-p}y_i\right)^{j-1}}{(j-1)!}e^{-\frac{\lambda_i}{1-p}y_i},$$

using that the number N of cycles visited now follows a negative binomial distribution (with range $m, m+1,\ldots$). The probabilistic construction has been discussed by [4].

11.5.6 Kibble and Moran's bivariate gamma

Kibble rewrote the LST to get

$$\frac{1}{((1+\frac{s_1}{\lambda_1})(1+\frac{s_1}{\lambda_1}))^\alpha}\frac{1}{\left(1+\frac{ps_1s_2}{\lambda_1\lambda_2(1+\frac{s_1}{\lambda_1})(1+\frac{s_1}{\lambda_1})}\right)^\alpha}$$

$$\frac{1}{[(1-it_1)(1-it_2)+\omega^2t_1t_2]^\alpha}$$

$$\begin{aligned} f(y_1,y_2) &= \sum_{j=0}^{\infty}\frac{\omega^{2j}\alpha^{[j]}}{j!}\left[\prod_{k=1}^{2}\left\{\sum_{\ell=0}^{j}(-1)^j\binom{j}{\ell}\frac{1}{\ell!}e^{-y_k}\right\}\right] \\ &= \left\{\prod_{k=1}^{2}\frac{1}{\Gamma(\alpha)}y_k^{\alpha-1}e^{-y_k}\right\}\left[1+\sum_{j=1}^{\infty}\rho^j L_j^{\alpha-1}(y_1)L_j^{\alpha-1}(y_2)\right] \end{aligned}$$

with $\rho=\omega^2$.

11.5.7 Jensen

Jensen made the parameter ρ in $\boldsymbol{Y}=\sum_{i=1}^{m}\boldsymbol{Y}_i$ depend on i. The MPH* representation is slightly unchanged, and the same kind of probabilistic arguments apply when deriving densities etc. These become quite complicated though but the LST which is stil reasonably easily expressed.

$$\prod_{i=1}^{m}\left(\prod_{j=1}^{n}\left(1+\frac{s_j}{\lambda_j}\right)-p_i\prod_{j=1}^{n}\lambda_j\right)^{-\frac{1}{2}}$$

In order to have these distributions in MPH* we need to have p_i such that there is an even number of pairs $\boldsymbol{Y}_i$ with distinct value of p_i.

11.6 Mixtures of type 2 and type 3, bivariate case only

11.6.1 Sarmanovs bivariate gamma

$$f(y_1,y_2)=$$

$$= \left\{ \prod_{k=1}^{2} \frac{1}{\Gamma(\alpha_k)} y_k^{\alpha_k - 1} e^{-y_k} \right\} \left[1 + \sum_{j=1}^{\infty} a_j L_j^{\alpha_1 - 1}(y_1) L_j^{\alpha_2 - 1}(y_2) \right]$$

♠ Insert a_j

11.7 Combinations of order statistics

The construction of multivariate distributions with given marginals is a topic in copula theory (see [7]). Any copula applied to marginal matrix exponential distributions will result in a multivariate distribution. The resulting joint Laplace transform is, however, not necessarily a rational function. A classical construction in producing bivariate distributions with given marginals is distributions of Morgenstern type also denote Farlie-Gumbel-Morgenstern distributions, and this will indeed lead to a multi dimensional rational Laplace transform and hence an MVME distribution.

Let $F_i(x), i = 1,2$ be univariate distributions and define a joint distribution generated by F_1 and F_2 in terms of the Farlie–Gumbel–Morgenstern copula to be

$$F(x_1,x_2) = F_1(x_1)F_2(x_2)\left[1 + \rho\left(1 - F_1(x_1)\right)\left(1 - F_2(x_2)\right)\right] \quad . \tag{11.8}$$

These distributions are in MVME if $F_1(x), F_2(x)$ have rational Laplace transforms. We show the slightly stronger result that $F(x_1,x_2)$ is contained in MME* . To see this we start with the following general result.

Lemma 11.1. *Let $F_i^{\min}(x) = 1 - (1 - F_i(x))^2$ and $F_i^{\max}(x) = F_i^2(x)$ be cumulative distribution functions of minimum respectively maximum of two independent random variables distributed according to $F_i(x)$. Then the bivariate Morgenstern distribution $F(x_1,x_2)$ generated by $F_1(x_1)$ and $F_2(x_2)$ is given by*

$$\begin{aligned} F(x_1,x_2) = {} & \frac{1+\rho}{4} F_1^{\max}(x_1)F_2^{\max}(x_2) + \frac{1-\rho}{4} F_1^{\max}(x_1)F_2^{\min}(x_2) + \\ & \frac{1-\rho}{4} F_1^{\min}(x_1)F_2^{\max}(x_2) + \frac{1+\rho}{4} F_1^{\min}(x_1)F_2^{\min}(x_2) \end{aligned}$$

.

Proof. The result is proved by eliminating $F_i(x)$ using $F_i(x) = \frac{F_i^{\min}(x) + F_i^{\max}(x)}{2}$ in (11.8). □

The lemma is useful as we can express the Morgenstern type distributions as a non–negative mixture of four bivariate distributions. Each of these four terms are joint distributions of independent variables.

♠ From here this should go in section on matrix–exponential distributions

We now relate this construction to matrix–exponential distributions. The following lemma, which is well–known for phase–type distributions, is traditionally proved by Markov chain (probabilistic) arguments.

Lemma 11.2. *Let X_1 and X_2 be two independent matrix–exponentially distributed random variables with conservative representation $(\boldsymbol{\alpha},\boldsymbol{T},\boldsymbol{t})$ such that $\boldsymbol{t}=-\boldsymbol{T}\boldsymbol{e}$ and such that there exist dual representations. Then* $\min(X_1,X_2)$ *is matrix–exponentially distributed with representation*

$$(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha},\boldsymbol{T}\oplus\boldsymbol{T},(\boldsymbol{t}\oplus\boldsymbol{t})\boldsymbol{e}) \quad .$$

and $\max(X_1,X_2)$ *is matrix–exponentially distributed with representation*

$$\left((\boldsymbol{\alpha}\otimes\boldsymbol{\alpha},\boldsymbol{0}),\begin{bmatrix}\boldsymbol{T}\oplus\boldsymbol{T} & (\boldsymbol{t}\oplus\boldsymbol{t})\\ \boldsymbol{0} & \boldsymbol{T}\end{bmatrix},\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{t}\end{bmatrix}\right)$$

Proof. We prove the representation of the minimum. The density of $\min(X_1,X_2)$ is

$$\begin{aligned} f^{\min}(x) &= 2\boldsymbol{\alpha}e^{\boldsymbol{T}x}\boldsymbol{t}\boldsymbol{\alpha}e^{\boldsymbol{T}x}\boldsymbol{e}\\ &= \left(\boldsymbol{\alpha}e^{\boldsymbol{T}x}\otimes\boldsymbol{\alpha}e^{\boldsymbol{T}x}\right)(\boldsymbol{t}\otimes\boldsymbol{e}+\boldsymbol{e}\otimes\boldsymbol{t})\\ &= (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})e^{(\boldsymbol{T}\oplus\boldsymbol{T})x}(\boldsymbol{t}\oplus\boldsymbol{t})\boldsymbol{e} \quad . \end{aligned}$$

The proof of the result for $\max(X_1,X_2)$ is similar. □

♠ Do we miss a formal rule (or rather which formal rule) to justify the last equality

Lemma 11.3. *Let $(\boldsymbol{\alpha},\boldsymbol{T},-\boldsymbol{T}\boldsymbol{e})$ be a conservative representation for a matrix exponential distribution. Further let $\boldsymbol{\pi}=\frac{1}{\mu}\boldsymbol{\alpha}(-\boldsymbol{T})^{-1}$ be non–zero solution to the equation system $\boldsymbol{\pi}(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\alpha})=\boldsymbol{0}$, with $\mu=\boldsymbol{\alpha}(-\boldsymbol{T})^{-1}\boldsymbol{e}$. Assume that all elements of $\boldsymbol{\pi}$ are different from 0. Then the distribution has an alternatively conservative representation $(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{T}},-\tilde{\boldsymbol{T}}\boldsymbol{e})$. Here*

$$\tilde{\boldsymbol{T}}=\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi}),\tilde{\boldsymbol{\alpha}}=\frac{\boldsymbol{\pi}\bullet\boldsymbol{t}}{\boldsymbol{\pi}\boldsymbol{t}} \quad ,$$

Proof. The matrix $\boldsymbol{T}$ is non–singular while $\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\alpha}=\boldsymbol{T}-\boldsymbol{T}\boldsymbol{e}\boldsymbol{\alpha}=\boldsymbol{T}(\boldsymbol{I}-\boldsymbol{e}\boldsymbol{\alpha})$ obviously has a single zero eigenvalue with corresponding left eigenvector $\boldsymbol{\pi}$. Now define $\boldsymbol{\Delta}(\boldsymbol{\pi})$ as the diagonal matrix which has the elements of $\boldsymbol{\pi}$ as entries. Then

$$\begin{aligned} \boldsymbol{\alpha}e^{\boldsymbol{T}x}\boldsymbol{t} &= \boldsymbol{\alpha}\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}e^{\boldsymbol{\Delta}(\boldsymbol{\pi})\boldsymbol{T}\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}x}\Delta(\boldsymbol{\pi})\boldsymbol{t}\\ &= \boldsymbol{t}'\boldsymbol{\Delta}(\boldsymbol{\pi})e^{\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi})x}\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{\alpha}'\\ &= \frac{\boldsymbol{t}'\boldsymbol{\Delta}(\boldsymbol{\pi})}{\boldsymbol{\pi}\boldsymbol{t}}e^{\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi})x}\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{\pi}\boldsymbol{t}\boldsymbol{\alpha}' \quad . \end{aligned}$$

Since $\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{\pi}\boldsymbol{t}\boldsymbol{\alpha}'=\boldsymbol{\Delta}(\boldsymbol{\pi})^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi})\boldsymbol{e}$ the representation is indeed of conservative form. □

The alternative representation of lemma 11.3 is known as the time reversed representation in the phase–type case. Here we define

Definition 11.1. For a conservative representation of a matrix–exponential distribution the representation of lemma 11.3 is called the dual of the representation.

Remark 11.1. Consider a conservative representation $(\boldsymbol{\alpha},\boldsymbol{T},\boldsymbol{t})$ where at least one component of $\boldsymbol{\pi} = \frac{1}{\mu}\boldsymbol{\alpha}(-\boldsymbol{T})^{-1}$ is equal to zero. Now whenever $\boldsymbol{M}$ is a regular matrix that satisfies $\boldsymbol{Me} = \boldsymbol{e}$ then $\left(\boldsymbol{\alpha M}, \boldsymbol{M}^{-1}\boldsymbol{TM}, -\boldsymbol{M}^{-1}\boldsymbol{Te}\right)$ will constitute another conservative representation. If we choose $\boldsymbol{M}$ such that $\boldsymbol{\alpha M}\left(-\boldsymbol{M}^{-1}\boldsymbol{TM}\right)^{-1} = \boldsymbol{\pi M}$ has no zero components then this representation will have a dual. Thus we can always find a dual pair representation for a matrix exponential distribution.

♠ Until here should go in section on matrix–exponential distributions

We state the following result as a lemma.

Lemma 11.4. *Let X_1 and X_2 be two independent matrix–exponentially distributed random variables with conservative representation $(\boldsymbol{\alpha},\boldsymbol{T},\boldsymbol{t})$ such that $\boldsymbol{t} = -\boldsymbol{Te}$. The dual of the conservative representations of* $\min(X_1,X_2)$ *and* $\max(X_1,X_2)$ *given in lemma 11.2 are* $\left(\tilde{\boldsymbol{\alpha}}_m, \tilde{\boldsymbol{T}}_m, \tilde{\boldsymbol{t}}_m\right)$ *and* $\left(\tilde{\boldsymbol{\alpha}}_M, \tilde{\boldsymbol{T}}_M, \tilde{\boldsymbol{t}}_M\right)$ *with*

$$\boldsymbol{\alpha}_m = \mu_m^{-1}\boldsymbol{\pi}_m \bullet (\boldsymbol{t}\oplus\boldsymbol{t})\boldsymbol{e}, \qquad \tilde{\boldsymbol{T}}_m = \boldsymbol{\Delta}(\boldsymbol{\pi}_m)^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi}_m)$$

$$\tilde{\boldsymbol{\alpha}}_M = \left(\mu_m+\mu_r\right)^{-1}\boldsymbol{\pi}_r \bullet \boldsymbol{t}, \boldsymbol{0}\right)$$

$$\tilde{\boldsymbol{T}}_M = \begin{bmatrix} \boldsymbol{\Delta}(\boldsymbol{\pi}_r)^{-1}\boldsymbol{T}'\boldsymbol{\Delta}(\boldsymbol{\pi}_r)^{-1} & \boldsymbol{\Delta}(\boldsymbol{\pi}_r)^{-1}(\boldsymbol{t}\oplus\boldsymbol{t})'\boldsymbol{\Delta}\left(\frac{\mu_m}{\mu_m+\mu_r}\boldsymbol{\pi}_m\right) \\ \boldsymbol{0} & \tilde{\boldsymbol{T}}_m \end{bmatrix} .$$

Here

$$\mu_m = (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})(\boldsymbol{T}\oplus\boldsymbol{T})^{-1}\boldsymbol{e}, \qquad \boldsymbol{\pi}_m = \mu_m^{-1}(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})(-\boldsymbol{T}\oplus\boldsymbol{T})^{-1},$$

and

$$\mu_r = (\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})(\boldsymbol{T}\oplus\boldsymbol{T})^{-1}(\boldsymbol{t}\oplus\boldsymbol{t})(\boldsymbol{T})^{-1}\boldsymbol{e},$$
$$\boldsymbol{\pi}_r = (\mu_m+\mu_r)(\boldsymbol{\alpha}\otimes\boldsymbol{\alpha})(\boldsymbol{T}\oplus\boldsymbol{T})^{-1}(\boldsymbol{t}\oplus\boldsymbol{t})(\boldsymbol{T})^{-1},$$

Proof. The result for the dual representation of $\min(X_1,X_2)$ follows from lemmas 11.2 and 11.3. To find the dual representation of $\max(X_1,X_2)$ we first appeal to a general result regarding the inversion of a block partitioned matrix

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{BC}^{-1} \\ \boldsymbol{0} & \boldsymbol{C}^{-1} \end{bmatrix} . \tag{11.9}$$

From this we calculate $\mathbb{E}(\max(X_1,X_2)) = \mu_{\mathrm{m}}+\mu_{\mathrm{r}}$, $\boldsymbol{\pi}_{\mathrm{r}}$, and $\boldsymbol{\alpha}_{\mathrm{M}}$. In the proof we will need $\boldsymbol{\pi}_{\mathrm{M}} = \left(\frac{\mu_{\mathrm{m}}}{\mu_{\mathrm{m}}+\mu_{\mathrm{r}}}\boldsymbol{\pi}_{\mathrm{m}}, \boldsymbol{\pi}_{\mathrm{r}}\right)$. lemma 11.3 applied to $\boldsymbol{T}_{\mathrm{M}}$ gives

$$\tilde{\boldsymbol{T}}\max = \boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{M}}\right)^{-1}\begin{bmatrix} \boldsymbol{T}\oplus\boldsymbol{T} & \boldsymbol{t}\oplus\boldsymbol{t} \\ \boldsymbol{0} & \boldsymbol{T} \end{bmatrix}'\boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{M}}\right)$$

$$= \begin{bmatrix} \boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{m}}\right)^{-1}(\boldsymbol{T}'\otimes\boldsymbol{T}')\boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{m}}\right) & \boldsymbol{0} \\ \boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{r}}\right)^{-1}(\boldsymbol{t}\oplus\boldsymbol{t})'\boldsymbol{\Delta}\left(\frac{\mu_{\mathrm{m}}}{\mu_{\mathrm{m}}+\mu_{\mathrm{r}}}\boldsymbol{\pi}_{\mathrm{M}}\right) & \boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{r}}\right)^{-1}\boldsymbol{T}'\boldsymbol{\Delta}\left(\boldsymbol{\pi}_{\mathrm{r}}\right) \end{bmatrix} .$$

The initial vector is similarly found to be $\left(\mathbf{0},(\mu_m+\mu_r)^{-1}\boldsymbol{\pi}_{\mathrm{r}}\bullet\boldsymbol{t}\right)$. By swapping the blocks we obtain the expression for $\tilde{\boldsymbol{\alpha}}_{\mathrm{M}}$ and $\tilde{\boldsymbol{T}}_{\mathrm{M}}$ of the lemma. □

We are now ready to state the main result of this section.

Theorem 11.1. *The bivariate Farlie-Gumbel-Morgenstern distribution generated by two matrix-exponential distribution of conservative representation* $(\boldsymbol{\alpha}_i,\boldsymbol{T}_i,-\boldsymbol{T}_i\boldsymbol{e})$ *is in MME*. An MME* representation* $(\boldsymbol{\gamma},\boldsymbol{T},\boldsymbol{R})$ *is*

$$\boldsymbol{\gamma}=(\boldsymbol{\alpha}_1\otimes\boldsymbol{\alpha}_1,\mathbf{0},\mathbf{0},\mathbf{0}) \ ,$$

$$\boldsymbol{T}=\begin{bmatrix} & \frac{1-\rho}{4}(\boldsymbol{t}_1\oplus\boldsymbol{t}_1)\boldsymbol{e}\tilde{\boldsymbol{\alpha}}_{r,2} & \frac{1+\rho}{4}(\boldsymbol{t}_1\oplus\boldsymbol{t}_1)\boldsymbol{e}\tilde{\boldsymbol{\alpha}}_{m,2} \\ \boldsymbol{T}_{M,1} & \frac{1+\rho}{2}\boldsymbol{s}_1\tilde{\boldsymbol{\alpha}}_{r,2} & \frac{1-\rho}{2}\boldsymbol{s}_1\tilde{\boldsymbol{\alpha}}_{m,2} \\ \mathbf{0} & & \tilde{\boldsymbol{T}}_{M,2} \end{bmatrix} \ ,$$

$$\boldsymbol{R}=\begin{bmatrix} \boldsymbol{e}_{m_1}\otimes\boldsymbol{e}_{m_1} & 0 \\ \boldsymbol{e}_{m_1} & 0 \\ 0 & \boldsymbol{e}_{m_2}\otimes\boldsymbol{e}_{m_2} \\ 0 & \boldsymbol{e}_{m_2} \end{bmatrix} \ .$$

Here notation is as in the preceding lemmas with obvious modifications corresponding to the introduction of the indices and $\tilde{\boldsymbol{\alpha}}_{r,2}=\left(\mu_{m,1}+\mu_{r,1}\right)^{-1}\boldsymbol{\pi}_{r,1}\bullet\boldsymbol{t}_1\Big)$.

Proof. The result can be proved by direct calculation of the Laplace transforms. Now

$$\boldsymbol{t}=-\boldsymbol{T}\boldsymbol{e}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \boldsymbol{t}_0\end{bmatrix}$$

where $\boldsymbol{t}_0=-\tilde{\boldsymbol{T}}_{\mathrm{m},2}\boldsymbol{e}$. Let $(X_1,X_2) \sim F$ where F is given by (11.8). We must now prove that

$$\tilde{L}(s)=\boldsymbol{\alpha}\left(s\boldsymbol{I}-\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{a})^{-1}\boldsymbol{T}\right)^{-1}\boldsymbol{t}$$

is indeed the the Laplace transform of $<\boldsymbol{a},\boldsymbol{X}>$. To this end we rewrite the four by four block matrix $s\boldsymbol{I}-\boldsymbol{T}$ in terms of two by two block matrices and calculate the inverse using Equation (11.9). Now with a further partitioning down to the original block structure we get

$$\boldsymbol{A}^{-1}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix},-\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}^{-1}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{bmatrix},\boldsymbol{C}^{-1}=\begin{bmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{bmatrix} \ .$$

We then use the structure of $\boldsymbol{\gamma}$ and $\boldsymbol{t}$ to see that

$$\tilde{L}(s)=(\boldsymbol{\alpha}_1\otimes\boldsymbol{\alpha}_1)\left(\boldsymbol{A}_{11}\boldsymbol{B}_{11}\boldsymbol{C}_{12}+\boldsymbol{A}_{12}\boldsymbol{B}_{21}\boldsymbol{C}_{12}+\boldsymbol{A}_{11}\boldsymbol{B}_{12}\boldsymbol{C}_{22}+\boldsymbol{A}_{12}\boldsymbol{B}_{22}\boldsymbol{C}_{22}\right)\boldsymbol{t}_0$$

from which the results follows. For instance,

$$\boldsymbol{A}_{11} = \left(s\boldsymbol{I} - \frac{1}{a_1}\boldsymbol{T}_1 \oplus \boldsymbol{T}_1 \right)^{-1}$$
$$\boldsymbol{B}_{11} = \frac{1-\rho}{4a_1}(\boldsymbol{t}_1 \oplus \boldsymbol{t}_1)\boldsymbol{e}\tilde{\boldsymbol{\alpha}}_{\mathrm{r},2}$$
$$\boldsymbol{C}_{12} = -\left(s\boldsymbol{I} - \frac{1}{a_2}\boldsymbol{\Delta}_1^{-1}\boldsymbol{T}_2'\boldsymbol{\Delta}_1 \right)^{-1} \boldsymbol{\Delta}_1^{-1}(\boldsymbol{t}_2 \oplus \boldsymbol{t}_2)'\boldsymbol{\Delta}_2 \left(s\boldsymbol{I} - \frac{1}{a_2}\tilde{\boldsymbol{T}}_{\mathrm{m},2} \right)^{-1} .$$

♠ Need to clean up the expression for $\boldsymbol{C}_{12}$ if we keep this way of doing it. The same apply to the description below.

Then substituting $s = 1$ we see that the resulting expression is the Laplace transform of the term $\frac{1-\rho}{4}F_1^{\min}(x_1)F_2^{\max}(x_2)$ with $\boldsymbol{a}$ being the transform variables. We apply lemma 11.3 to see this for $F_2^{\max}(x_2)$. □

The proposed representation of Theorem 11.1 is motivated by the following probabilistic argument in the phase–type case. First note it is straightforward to express the distribution of $< \boldsymbol{a}, \boldsymbol{X} >$ as a phase type distribution as $F(x_1,x_2)$ is a mixture of four components each of which are phase–type distributed. However, the naive construction involves twelve blocks and is not minimal while the representation of Theorem 11.1 requires only four blocks.

The underlying phase process of $a_1x_1 + a_2x_2$ can move through the system in the following four ways

1. With probability $\frac{1+\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_1}\boldsymbol{T}_1 \to \frac{1}{a_2}\boldsymbol{T}_{\mathrm{m},2} \to \frac{1}{a_2}\boldsymbol{T}_2$
2. With probability $\frac{1-\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_1}\boldsymbol{T}_1 \to \frac{1}{a_2}\boldsymbol{T}_{\mathrm{m},2}$
3. With probability $\frac{1-\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_2}\boldsymbol{T}_{\mathrm{m},2} \to \frac{1}{a_2}\boldsymbol{T}_2$
4. With probability $\frac{1+\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_2}\boldsymbol{T}_{\mathrm{m},2}$.

They depend on previous steps of the path and thus violates the Markov property. By applying the time reversed version of $F_2^{\max}$ we can avoid this problem as the four different ways through the system are now

1. With probability $\frac{1+\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_1}\boldsymbol{T}_1 \to \frac{1}{a_2}\boldsymbol{T}_2 \to \frac{1}{a_2}\tilde{\boldsymbol{T}}_{\mathrm{m},2}$
2. With probability $\frac{1-\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_1^{(m)} \to \frac{1}{a_1}\boldsymbol{T}_1 \to \frac{1}{a_2}\tilde{\boldsymbol{T}}_{\mathrm{m},2}$
3. With probability $\frac{1-\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_2}\boldsymbol{T}_2 \to \frac{1}{a_2}\tilde{\boldsymbol{T}}_{\mathrm{m},2}$
4. With probability $\frac{1+\rho}{4}$ $\frac{1}{a_1}\boldsymbol{T}_{\mathrm{m},1} \to \frac{1}{a_2}\tilde{\boldsymbol{T}}_{\mathrm{m},2}$.

The blocks of the form $\boldsymbol{T} \otimes \boldsymbol{T}$ of Theorem 11.1 can be expressed more compactly albeit less transparent.

♠ Also true for min and max

One specific example is Gumbels's Bivariate Exponentials Model II. The distribution is discussed pp. 353–354 in [6] **47.2.2**. The joint cumulative distribution function is

$$F(x_1,x_2) = \left(1 - e^{-\lambda_1 x_1}\right)\left(1 - e^{-\lambda_2 x_2}\right)\left(1 + \rho e^{-\lambda_1 x_1 - \lambda_2 x_2}\right), \qquad x_1, x_2 > 0, \qquad |\rho| \le 1 \ ,$$

where $\lambda_1 = \lambda_2 = 1$ in [6]. The joint density is

$$f(x_1,x_2) = \lambda_1\lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2}\left[1+\rho\left(2e^{-\lambda_1 x_1}-1\right)\left(2e^{-\lambda_2 x_2}-1\right)\right] \ .$$

Se [6] p.354 for results on conditional distributions, which are mixtures of exponentials. We have

$$L(s_1.s_2) = \frac{\lambda_2\,(s_1 s_2 + 2\,\lambda_1\, s_2 + 2\,\lambda_2\, s_1 + 4\,\lambda_1\,\lambda_2 + \rho\, s_1 s_2)\,\lambda_1}{(s_2+2\,\lambda_2)\,(s_2+\lambda_2)\,(s_1+2\,\lambda_1)\,(s_1+\lambda_1)} \ .$$

With MPH* representation $\boldsymbol{\gamma} = (1,0,0,0)$

$$\boldsymbol{T} = \begin{bmatrix} -2\lambda_1 & \lambda_1 & \frac{1-\rho}{2}\lambda_1 & \frac{1+\rho}{2}\lambda_1 \\ 0 & -\lambda_1 & \frac{1+\rho}{2}\lambda_1 & \frac{1-\rho}{2}\lambda_1 \\ 0 & 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 & -2\lambda_2 \end{bmatrix} \qquad \boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Another example is Farlie–Gumbel–Morgenstern–Type Bivariate Gamma **48.2.7** which is in MPH* whenever the two marginals are gamma distributions with integer shape parameter.

♠ Bladt-Nielsen bivariate exponential

♠ Convex extension of FGM using order statistics mixing, and generalized

11.7.0.1 Mixing higher order statistics

The idea can be extended. The generalization of **??** is that we can write any distribution as

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} F^{(i)}(x)$$

where $F^{(i)}(x)$ is the cumulative distribution function of the *i*th order statistics. To obtain bivariate distributions with marginals F_1 and F_2 we simply define

$$F(y_1,y_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} F_1^{(i)}(y_1) F^{(j)}(y_2)$$

with the additional requirements that

$$\sum_{i=1}^{n} p_{ij} = \frac{1}{n} \wedge \sum_{j=1}^{m} p_{ij}\frac{1}{n}.$$

If we additionally require that $p_{ij} \geq 0$ then we are certain that $F(y_1,y_2)$ is indeed a joint cumulative probability function. The idea is easily extended to dimensions more than 2, but with increasing complexity of course. This construction is general,

but whenever, the given marginals are ME distributions then the resulting distribution will be in MVME, actually in MME* where the latter is not necessarily of minimal order.

♠ I conjecture it is, here is something to prove, although it is not the most interesting result

The reversibility construction used when forming the standard FGM-s does also apply for a bivariate exponential distribution based on higher order order statistics.

A representation for the ith order statistics out of n iid exponential random variables is

$$\left((1,0,\ldots,0),\begin{bmatrix} -n\lambda & n\lambda & 0 & 0\ldots & 0 & 0 \\ 0 & -(n-1)\lambda & (n-1)\lambda & 0\ldots & 0 & 0 \\ 0 & 0 & -(n-2)\lambda & (n-2)\lambda\ \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0\ldots & -(n-i+2)\lambda & (n-i+2)\lambda \\ 0 & 0 & 0 & 0\ldots & 0 & -(n-i+1)\lambda \end{bmatrix}\right),$$

using the reversed representation we get

The exponential distribution

$$\left((1,0,\ldots,0),\begin{bmatrix} -n\lambda & (n-1)\lambda & 0 & 0\ldots & 0 & 0 \\ 0 & -(n-1)\lambda & (n-2)\lambda & 0\ldots & 0 & 0 \\ 0 & 0 & -(n-2)\lambda & (n-3)\lambda\ \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0\ldots & -2\lambda & \lambda \\ 0 & 0 & 0 & 0\ldots & 0 & -\lambda \end{bmatrix}\right),$$

with reversed representation

$$\left(\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right),\begin{bmatrix} -\lambda & \lambda & 0 & 0\ldots & 0 & 0 \\ 0 & -2\lambda & 2\lambda & 0\ldots & 0 & 0 \\ 0 & 0 & -3\lambda & 3\lambda\ \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0\ldots & -(n-1)\lambda & (n-1)\lambda \\ 0 & 0 & 0 & 0\ldots & 0 & -n\lambda \end{bmatrix}\right),$$

can be viewed either as an exponential distribution expressed as the average of the n different order statistics or as Type 2 distribution with the property applied $n-1$ times.

We can now can combine the two different representations to get exactly a bivariate distribution with exponential marginals . Minimum possible correlation $\left(1-\frac{\pi^2}{6}\right)$ and maximum possible correlation. Any correlation can thus be achieved.

11.8 Open problems

1. To which extent can gamma distributions defined through the mechanisms here be generalized to non–integer shape parameters?
2. Related can we simulate from them, using some kind of Dirichlet distribution?
3. Understand trivariate and higher order DusBers better. Can they be proved not to be in MME*?
4. Is there a Colm like characterization theorem for MVPH's?

Exercises

11.1. With

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

where (Z_1, Z_2) are independent $\exp(1)$ distributed random variables, find a (if any) MPH* representation, the joint density, joint survival function and joint Laplace-Stieltjes transform or moment generation function for the pair (Y_1, Y_2).

11.2. Suppose we have two components with failure intensities $\lambda_i, i = 1,2$ respectively. Whenever one fails the remaining has a new failure intensity λ_i'. Give a MPH* representation - if any - for the joint life time of the two components, and derive the joint density, and joint Laplace-Stieltjes transform or moment generating function. [1]

11.3. Continuation of Problem 11.2. Choose the parameters of the distribution in Problem 11.2, such that the distribution becomes the joint distribution of maximum and minimum of two independent exponential random variables.

[1] The distribution is called Freund's Bivariate Exponential and discussed p.355–362 in [6].

Chapter 12
Estimation and statistical inference for matrix–exponential distributions

Chapter 13
Estimation and statistical inference for phase–type distributions

References

1. Asmussen, S., Bladt, M.: Point processes with finite-dimensional probabilities. Stochastic Processes and their Applications **82**(1), 127–142 (1999)
2. Colm A, O., Raftery, A.E.: A continuous multivariate exponential distribution that is multivariate phase type. Statistics & Probability Letters **7**, 323–325 (1989)
3. Dussauchoy, A., Berland, R.: 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: G.P. Patil, D. Kotz, J.K. Ord (eds.) Statistical Distributions in Scientific Work, pp. 319–328. Proceedings of the NATO Advanced Study Institute, University of Calgary, Calgary, Alberta, Canada (1975)
4. Gaver, D.P.: Multivariate gamma distributions generated by mixture. Sankhy_a, Series A **32**, 123–126 (1970)
5. Kibble, W.F.: A two–variate gamma type distribution. Sankhy_a **5**, 137–150 (1941)
6. Kotz, S., Balakrishnan, N., Johnson, N.L.: Continuous Multivariate Distributions. John Wiley and Sons (2000)
7. Nelsen, R.B.: An Introduction to Copulas, seond edn. Springer Series in Statistics. Springer (2005)