Lecture notes on phase–type distributions for 02407 Stochastic Processes

Bo Friis Nielsen

October 2022

Springer

Contents

Chapter 1

Phase–type distributions

1.1 Notation

We will use $\mathbf{1}$ to denote a column vector of ones of appropriate dimension.

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} .$$

Correspondingly we will let $\mathbf{0}$ denote the vector of zeros.

Matrices is represented by capital primarily roman letters like $\boldsymbol{T}$ and $\boldsymbol{A}$. The symbol $\boldsymbol{I}$ will be used for a unity matrix of appropriate dimension, while $\mathbf{0}$ is a matrix of zero's of appropriate dimension.

1.2 Matrix results

In this section we have collected a few slightly specialized matrix results.

Lemma 1. *The inverse of the block Matrix*

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \mathbf{0} & \boldsymbol{C} \end{bmatrix} \tag{1.1}$$

can be written as

$$\begin{bmatrix} \boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}^{-1} \\ \mathbf{0} & \boldsymbol{C}^{-1} \end{bmatrix} \tag{1.2}$$

whenever $\boldsymbol{A}$ and $\boldsymbol{C}$ are invertible.

Proof. By direct verification.

1.2.1 The Kronecker product

Many operations with phase–distributions are conveniently expressed using the Kronecker product. For two matrices $\boldsymbol{A}$ with dimension $\ell \times k$ and $\boldsymbol{B}$ with dimension $n \times m$ we define the Kroneckerproduct $\otimes$ by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1k}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2k}\boldsymbol{B} \\ \vdots & \dots & \vdots & \vdots \\ a_{\ell 1}\boldsymbol{B} & a_{\ell 2}\boldsymbol{B} & \dots & a_{\ell k}\boldsymbol{B} \end{bmatrix} \tag{1.3}$$

Example 1. Consider the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{I}$ given by:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 7 & 1 \\ 3 & 5 & 11 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 13 & 4 \\ 0 & 17 \end{bmatrix}, \quad \boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad .$$

Then $\boldsymbol{A} \otimes \boldsymbol{B}$, $\boldsymbol{A} \otimes \boldsymbol{I}$ and $\boldsymbol{I} \otimes \boldsymbol{B}$ is given by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} 2\cdot 13 & 2\cdot 4 & 7\cdot 13 & 7\cdot 4 & 1\cdot 13 & 1\cdot 4 \\ 2\cdot 0 & 2\cdot 17 & 7\cdot 0 & 7\cdot 17 & 1\cdot 0 & 1\cdot 17 \\ 3\cdot 13 & 3\cdot 4 & 5\cdot 13 & 5\cdot 4 & 11\cdot 13 & 11\cdot 4 \\ 3\cdot 0 & 3\cdot 17 & 5\cdot 0 & 5\cdot 17 & 11\cdot 0 & 11\cdot 17 \end{bmatrix}$$

$$\boldsymbol{A} \otimes \boldsymbol{I} = \begin{bmatrix} 2 & 0 & 0 & 7 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 7 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 7 & 0 & 0 & 1 \\ 3 & 0 & 0 & 5 & 0 & 0 & 11 & 0 & 0 \\ 0 & 3 & 0 & 0 & 5 & 0 & 0 & 11 & 0 \\ 0 & 0 & 3 & 0 & 0 & 5 & 0 & 0 & 11 \end{bmatrix}$$

$$\boldsymbol{I} \otimes \boldsymbol{B} = \begin{bmatrix} 13 & 4 & 0 & 0 & 0 & 0 \\ 0 & 17 & 0 & 0 & 0 & 0 \\ 0 & 0 & 13 & 4 & 0 & 0 \\ 0 & 0 & 0 & 17 & 0 & 0 \\ 0 & 0 & 0 & 0 & 13 & 4 \\ 0 & 0 & 0 & 0 & 0 & 17 \end{bmatrix} \quad .$$

The following rule is very convenient. If the usual matrix products $\boldsymbol{LU}$ and $\boldsymbol{MV}$ exist, then

$$(\boldsymbol{L} \otimes \boldsymbol{M})(\mathbf{U} \otimes \mathbf{V}) = \boldsymbol{LU} \otimes \boldsymbol{MV} \quad . \tag{1.4}$$

A natural operation for continuous time phase–type distributions is $\boldsymbol{A} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}$, thus motivating the definition of the Kronecker sum defined by the symbol $\oplus$.

$$\boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B} \tag{1.5}$$

1.3 Discrete phase–type distributions

Definition 1. A discrete phase–type distribution is the distribution of the time to absorption in a finite discrete time Markov chain with transition matrix $\boldsymbol{P}$ of dimension $m+1$ as given by 1.6. The Markov chain has m transient and 1 absorbing state.

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{s} \\ \boldsymbol{0} & 1 \end{bmatrix} \quad . \tag{1.6}$$

The initial probability vector is denoted by $(\boldsymbol{\alpha}, \alpha_{m+1})$. The pair $(\boldsymbol{\alpha}, \boldsymbol{S})$ is called a representation for the phase–type distribution.

The matrix $(\boldsymbol{I} - \boldsymbol{S})$ is non–singular (i.e. the only solution to $\boldsymbol{x} = \boldsymbol{x}\boldsymbol{S}$ is $\boldsymbol{x} = \boldsymbol{0}$). One consequence is, that at least one of the row sums of $\boldsymbol{S}$ is strictly less than 1. As $\boldsymbol{P}$ is a stochastic matrix it satisfies equation 1.7.

$$\boldsymbol{P}\boldsymbol{1} = \boldsymbol{1} \tag{1.7}$$

and we have

$$\boldsymbol{S}\boldsymbol{1} + \boldsymbol{s} = \boldsymbol{1} \;\text{ or }\; \boldsymbol{s} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{1} \quad . \tag{1.8}$$

It follows from

$$\boldsymbol{P}^n = \begin{bmatrix} \boldsymbol{S}^n & (\boldsymbol{I} - \boldsymbol{S}^n)\boldsymbol{1} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

that

$$P(X > n) = \boldsymbol{\alpha}\boldsymbol{S}^n\boldsymbol{1} \quad .$$

Thus

$$P(X \leq n) = 1 - \boldsymbol{\alpha}\boldsymbol{S}^n\boldsymbol{1} \quad . \tag{1.9}$$

Example 2. The simplest possible discrete phase–type distribution is obtained, when the dimension of $\boldsymbol{S}$ is m=1. In this case we have

$$\boldsymbol{P} = \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix}$$

$\boldsymbol{\alpha} = (1)$. As will be clear later this phase–type distribution is simply a geometric distribution with parameter $1-p$. A sum of geometrically distributed random variables has a negative binomial distribution. The negative binomial distribution can be expressed as a phase–type distribution by

$$\boldsymbol{P} = \left[\begin{array}{ccccccc|c} p & 1-p & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & p & 1-p & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & p & 1-p & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & p & 1-p & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & p & 1-p \\ \hline \multicolumn{7}{c|}{\boldsymbol{0}} & 1 \end{array}\right]$$

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

1.3.1 Cumulative distribution and probability mass function

We will now give the argument leading to formula (1.9) in more detail. We first consider the probabilities for the transient states after n transitions. That is $p_i^{(n)} = \text{Prob}(X_n = i)$. We collect these probabilities in a vector to get $\boldsymbol{p}^{(n)} = (p_1^{(n)}, \ldots, p_m^{(n)})$. Using standard arguments for discrete time Markov chains we get

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)}\boldsymbol{S} = \boldsymbol{\alpha}\boldsymbol{S}^n \ . \tag{1.10}$$

The event that absorption occurs at time x can be partitioned using the union of the events that the chain is in state i $(i = 1, \ldots, m)$ at time $x-1$, and that absorption happens at i at time x. The probability of the former is $p_i^{(x-1)}$ while the probability of the latter event is s_i. Thus the probability mass function can be expressed as

$$f(x) = \sum_{i=1}^{m} p_i^{(x-1)} s_i = \boldsymbol{p}^{(x-1)}\boldsymbol{s} = \boldsymbol{\alpha}\boldsymbol{S}^{x-1}\boldsymbol{s}, \qquad x > 0 \ . \tag{1.11}$$

The cumulative probability function can now be found by summation of $f(x)$ or by noting that absorption has occurred if the process is no longer in one of the transient states at time x. The probability of being in one of the transient states is $\sum_{i=1}^{m} p_i^{(x)} = \boldsymbol{p}^{(x)}\mathbf{1} = \boldsymbol{\alpha}\boldsymbol{S}^x\mathbf{1}$. Thus, the cumulative distribution function is given by

$$F(x) = 1 - \boldsymbol{\alpha}\boldsymbol{S}^x\mathbf{1}, \qquad x \geq 0 \tag{1.12}$$

Example 3. For the geometric distribution in Example 2 we find using (1.11) and (1.12):

$$f(x) = 1 \cdot p^{x-1}(1-p) \tag{1.13}$$

$$F(x) = 1 - 1 \cdot p^x \tag{1.14}$$

$$\tag{1.15}$$

1.3.2 The generating function

The generating function for a non–negative discrete random variable X is given by (1.16)

$$H(z) = \mathbb{E}\left(z^X\right) = \sum_{x=0}^{\infty} z^x f(x) \tag{1.16}$$

For a discrete phase–type random variable we find

$$H(z) = \sum_{x=0}^{\infty} z^x f(x) = \alpha_{m+1} + \sum_{x=1}^{\infty} z^x \boldsymbol{\alpha}\boldsymbol{S}^{x-1}\boldsymbol{s} = \alpha_{m+1} + z\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s} \tag{1.17}$$

Here we have used the geometric series $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$ for matrices. (see e.g. [2] Theorem 28.1). The result is $\sum_{i=0}^{\infty} \boldsymbol{A}^i = (\boldsymbol{I} - \boldsymbol{A})^{-1}$ whenever $|\lambda_i| < 1$ for all i, where λ_i are the eigenvalues of $\boldsymbol{A}$.

Example 4. For the geometric distribution we get the well–known result

$$H(z) = z(1-pz)^{-1}(1-p) = \frac{z(1-p)}{1-zp} \tag{1.18}$$

1.3.3 Moments

The factorial moments for a discrete random variable can be obtained by successive differentiation of the generating function.

$$\mathbb{E}(X(X-1)\dots(X-(k-1)) = \left.\frac{d^{H(z)}k}{dz^k}\right|_{z=1} .$$

Thus for a discrete phase–type variable with representation $(\boldsymbol{\alpha}, \boldsymbol{S})$ we get

$$\mathbb{E}(X(X-1)\dots(X-(k-1))) = k!\boldsymbol{\alpha}(\boldsymbol{I}-\boldsymbol{S})^{-k}\boldsymbol{S}^{k-1}\mathbf{1} \ .$$

The matrix $\boldsymbol{U} = (\boldsymbol{I}-\boldsymbol{S})^{-1}$ is of special importance as the i,j'th element has a probabilistic interpretation as the expected time spent in state j before absorption conditioned on starting in state i. By a little bit of matrix calculation we can express the generating function using $\boldsymbol{U}$ rather than $\boldsymbol{S}$ to get

$$H(z) = \alpha_{m+1} + \boldsymbol{\alpha}\left(\boldsymbol{U}\frac{1-z}{z} + \boldsymbol{I}\right)^{-1}\mathbf{1} \ ,$$

(where the latter expression obviously is not defined for $z = 0$).

1.3.4 Closure properties

One appealing feature of phase–type distributions is that the class is closed under a number of operations. The closure properties are a main contributing factor to the popularity of phase–type distributions in probabilistic modeling of technical systems. In particular we will see the that the class is closed under addition, finite mixtures, and finite order statistics.

Theorem 1. *[Sum of two independent PH variables] Consider two discrete random variables X and Y with representation $(\boldsymbol{\alpha}, \boldsymbol{S})$ and $(\boldsymbol{\beta}, \boldsymbol{T})$ respectively. Then the random variable $Z = X + Y$ follows a discrete phase–type distribution with representation $(\boldsymbol{\gamma}, \mathbf{L})$ given by 1.19.*

$$\begin{bmatrix} \boldsymbol{L} & \boldsymbol{l} \\ \mathbf{0} & 1 \end{bmatrix} = \left[\begin{array}{cccc|cccc|c} S_{11} & S_{12} & \dots & S_{1m} & s_1\beta_1 & s_1\beta_2 & \dots & s_1\beta_k & s_1\beta_{k+1} \\ S_{21} & S_{22} & \dots & S_{2m} & s_2\beta_1 & s_2\beta_2 & \dots & s_2\beta_k & s_2\beta_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mm} & s_m\beta_1 & s_m\beta_2 & \dots & s_m\beta_k & s_m\beta_{k+1} \\ \hline 0 & 0 & \dots & 0 & T_{11} & T_{12} & \dots & T_{1k} & t_1 \\ 0 & 0 & \dots & 0 & T_{21} & T_{22} & \dots & T_{2k} & t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & T_{k1} & T_{k2} & \dots & T_{kk} & t_k \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \end{array}\right] \ , \tag{1.19}$$

$\boldsymbol{\gamma} = (\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}\beta_1, \alpha_{m+1}\beta_2, \dots, \alpha_{m+1}\beta_k)$. *In matrix notation*

$$\begin{bmatrix} \boldsymbol{L} & \boldsymbol{l} \\ \boldsymbol{0} & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} & \beta_{k+1}\boldsymbol{s} \\ \boldsymbol{0} & \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & & 1 \end{bmatrix} \tag{1.20}$$

$\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \alpha_{m+1}\boldsymbol{\beta}), \gamma_{m+k+1} = \alpha_{m+1}\beta_{k+1}$.

Proof. The probabilistic proof is done by concatenating the transition matrices for the transient states of the Markov chains related to X and Y. We interpret the random variables Z, X, and Y as time variables. In order to get the random variable Z we first start a Markov chain related to X given by $(\boldsymbol{\alpha}, \boldsymbol{S})$. Immediately upon absorption from X we start the Markov chain related to Y given by $(\boldsymbol{\beta}, \boldsymbol{T})$. The terms $s_i\beta_j$ ensures that the initial probability distribution of the Y chain is indeed $\boldsymbol{\beta}$.

One can alternatively proceed entirely analytically by manipulations with generating functions.

$$H_Z(z) = H_X(z)H_Y(z) = (\alpha_{m+1} + z\boldsymbol{\alpha}(\boldsymbol{I} - z\boldsymbol{S})^{-1}\boldsymbol{s})(\beta_{k+1} + z\boldsymbol{\beta}(\boldsymbol{I} - z\boldsymbol{T})^{-1}\boldsymbol{t}) \tag{1.21}$$

After some straightforward but tedious calculations, we omit the details for now, one obtains

$$H_Z(z) = \gamma_{m+1} + z\boldsymbol{\gamma}(\boldsymbol{I} - z\mathbf{L})^{-1}\boldsymbol{l} \tag{1.22}$$

Remark 1. An important implication is that the representation for a phase–type distribution can not be unique as the representation for $Z = X + Y$ is not in general symmetric in the parameters of the X and Y chains. Thus, typically the $\mathbf{L}$ matrix will be different depending on which chain we choose to represent X while the expressions for the distribution like $F_Z(x)$, $f_Z(x)$ or $H_Z(z)$ will be identical. See Section 1.3.5 for a somewhat deeper discussion of this and some supplementary examples.

1.3.4.1 Finite mixtures of phase–type distributions

Given X_i phase–type distributed with representation $(\boldsymbol{\alpha_i}, \boldsymbol{S}_i)$ we have $Z = I_i X_i$ with $\sum_{i=1}^{k} I_i = 1$ and $P(I_i = 1) = p_i$). We see that the random variable Z is itself phase–type distributed with representation $(\boldsymbol{\gamma}, \mathbf{L})$ given by 1.23:

$$\mathbf{L} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{S}_k \end{bmatrix} \tag{1.23}$$

$\boldsymbol{\gamma} = (p_1\boldsymbol{\alpha_1}, p_2\boldsymbol{\alpha_2}, \dots, p_k\boldsymbol{\alpha_k})$.

Example 5. For two geometric distributions with parameters p_x and p_y we have the representation

$$\mathbf{L} = \begin{bmatrix} p_x & 0 \\ 0 & p_y \end{bmatrix}$$

with $\boldsymbol{\gamma} = (p_{0x}, p_{0y}) = (p_{0x}, 1 - p_{0x})$, where p_{0x} is the probability of choosing the first respectively the second geometric distribution.

1.3.4.2 Order statistics

Initially we focus on the distribution of the smallest Z_1 and the largest Z_2 of two independent variables X and Y. Thus $Z_1 = \min(X,Y)$ and $Z_2 = \max(X,Y)$. First we will consider the an example.

Example 6. Let X be negative binomially distributed with parameters $k_x = 2, p_x$ and let Y be negative binomially distributed with parameters $k_y = 2, p_y$. The matrices $\boldsymbol{S}_x$ and $\boldsymbol{S}_y$ are given by

$$\boldsymbol{S}_x = \begin{bmatrix} p_x & 1-p_x \\ 0 & p_x \end{bmatrix} \qquad \boldsymbol{S}_y = \begin{bmatrix} p_y & 1-p_y \\ 0 & p_y \end{bmatrix}$$

with $\boldsymbol{\alpha}_x = (1,0)$ and $\boldsymbol{\alpha}_y = (1,0)$. We define $Z_1 = \min(X,Y)$ and proceed by creating a Markov chain that describes the simultaneous evolution of the X and the Y chains and thus the evolution of the Z chain. The process will have 4 transient states corresponding to all possible combinations of the X and Y states. We denote the four states by $(1,1),(1,2),(2,1),(2,2)$. The transition from (1,1) to (1,2) occurs whenever we have no state change in the X chain (probability p_x) and the Y chain changes state (probability $1-p_y$). In summary the final Markov chain tracks the time until the first of the two original chains reaches the absorbing state. Thus,

$$\boldsymbol{S}_{\min(X,Y)} = \begin{bmatrix} p_x p_y & p_x(1-p_y) & (1-p_x)p_y & (1-p_x)(1-p_y) \\ 0 & p_x p_y & 0 & (1-p_x)p_y \\ 0 & 0 & p_x p_y & p_x(1-p_y) \\ 0 & 0 & 0 & p_x p_y \end{bmatrix}$$

and we have that $\boldsymbol{\alpha}_{\min(X,Y)} = (1,0,0,0)$. Further, we see that we can write $\boldsymbol{S}_{\min(X,Y)} = \boldsymbol{S}_x \otimes \boldsymbol{S}_y$ and $\boldsymbol{\alpha}_{\min(X,Y)} = \boldsymbol{\alpha}_x \otimes \boldsymbol{\alpha}_y$.
With respect to the distribution of $Z_2 = \max(X,Y)$ we need to include 4 more states $(1,3),(2,3),(3,1),(3,2)$ corresponding to the possibility that one of the two chains survives the absorption of the other. It is convenient to order the state space as $(1,1),(1,2),(2,1),(2,2),(1,3),(2,3),(3,1),(3,2)$. And we obtain the phase type generator $\boldsymbol{S}_{\max(X,Y)}$

$$\boldsymbol{S}_{\max(X,Y)} =$$

$$\begin{bmatrix} p_x p_y & p_x(1-p_y) & (1-p_x)p_y & (1-p_x)(1-p_y) & 0 & 0 & 0 & 0 \\ 0 & p_x p_y & 0 & (1-p_x)p_y & p_x(1-p_y) & (1-p_x)(1-p_y) & 0 & 0 \\ 0 & 0 & p_x p_y & p_x(1-p_y) & 0 & 0 & (1-p_x)p_y & (1-p_x)(1-p_y) \\ 0 & 0 & 0 & p_x p_y & 0 & p_x(1-p_y) & 0 & (1-p_x)p_y \\ 0 & 0 & 0 & 0 & p_x & 1-p_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_y & 1-p_y \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_y \end{bmatrix}$$

The general result is that for X phase–type distributed with $(\boldsymbol{S}_x, \boldsymbol{\alpha}_x)$ and Y phase–type distributed with $(\boldsymbol{S}_y, \boldsymbol{\alpha}_y)$ $\min(X,Y)$ is phase–type distributed with representation $(\boldsymbol{L}, \boldsymbol{\gamma})$ given by 1.24:

$$\boldsymbol{L} = \boldsymbol{S}_x \otimes \boldsymbol{S}_y \ , \tag{1.24}$$

where $\boldsymbol{\gamma} = \boldsymbol{\alpha}_x \otimes \boldsymbol{\alpha}_y$. Further $\max(X,Y)$ is phase–type distributed with representation $(\mathbf{L}, \boldsymbol{\gamma})$ given by 1.25:

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{S}_x \otimes \boldsymbol{S}_y & \boldsymbol{S}_x \otimes \boldsymbol{t}_y & \boldsymbol{t}_x \otimes \boldsymbol{S}_y \\ \mathbf{0} & \boldsymbol{S}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{S}_y \end{bmatrix} \tag{1.25}$$

with $\boldsymbol{\gamma} = (\boldsymbol{\alpha}_x \otimes \boldsymbol{\alpha}_y, \boldsymbol{\alpha}_x \alpha_{y,m+1}, \alpha_{x,k+1} \boldsymbol{\alpha}_y)$. Here the dimension of $\boldsymbol{S}_x$ is k and the dimension of $\boldsymbol{S}_y$ is m. We write $\boldsymbol{l}$ explicitly:

$$\boldsymbol{l} = \begin{bmatrix} \boldsymbol{t}_x \otimes \boldsymbol{t}_y \\ \boldsymbol{t}_x \\ \boldsymbol{t}_y \end{bmatrix} \tag{1.26}$$

1.3.4.3 Other properties

We briefly mention that all discrete probability distribution functions with finite support (i.e. $f(x) = 0$ for all $x \geq x_0$) are of phase–type.

Random sums of independent discrete phase–type variables where the number of terms in the random sum is itself phase–type distributed is phase type distributed with representation.

$$(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \boldsymbol{S} \otimes \boldsymbol{I} + \boldsymbol{s}\boldsymbol{\alpha} \otimes \boldsymbol{T})$$

1.3.5 Non–uniqueness of representations

A main drawback when modeling with phase–type distributions is the non–uniqueness of their representations. Thus in most cases a number of different representations will give rise to the same distribution. Thus, only in very special cases will a representation be unique.

Example 7. The distribution with representation $((1,0), \boldsymbol{S})$ with $\boldsymbol{S}$ given by

$$\boldsymbol{S} = \begin{bmatrix} p_1 & p_2 - p_1 \\ 0 & p_2 \end{bmatrix} \tag{1.27}$$

is simply a geometric distribution with $f(x) = p_2^{x-1}(1-p_2)$.

This can be seen by deriving the generating function for the distribution. Alternatively it is seen that $\boldsymbol{t} = \begin{bmatrix} 1-p_2 \\ 1-p_2 \end{bmatrix}$, thus there is a constant probability $1-p_2$ of absorption not dependent on the state of the chain and thus independent of the time elapsed.

1.4 Continuous phase–type distributions

Many definitions and results regarding discrete time phase–type distributions carry over verbatim to the continuous time case, other need minor modifications. We first extend Definition 1 in.

Definition 2. A phase–type distribution is the distribution of the time to absorption in a finite Markov jump process (continuous time Markov chain) of dimension $m+1$, where one state is absorbing and the remaining m states are transient. A phase type distribution is uniquely given by an m dimensional row vector $\boldsymbol{\alpha}$ and an $m \times m$ matrix $\boldsymbol{S}$. We call the the pair $(\boldsymbol{\alpha}, \boldsymbol{S})$a representation for the phase type distribution. The vector $\boldsymbol{\alpha}$ can be interpreted as the initial probability vector among the m transient states, while the the matrix $\boldsymbol{S}$ can be interpreted as the one step transition probability matrix among the transient states in the discrete case and as the infinitesimal generator matrix among the transient states in the continuous case. A phase–type distribution is uniquely given by any representation. However, several representations can lead to the same phase–type distribution. We will elaborate a little bit on this in Section 1.4.7.

The generator matrix for the Markov jump process in the continuous case for a given representation is given by (1.28)

$$\mathbf{Q} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{s} \\ \mathbf{0} & 0 \end{bmatrix} \tag{1.28}$$

This section will be quite repetitive restating a number of results now for the continuous case. In some cases the exact formulation of results and properties will vary slightly.

1.4.1 Probability functions

Once again the most apparent result regards the survival function.

As in Section 1.3.1 we will consider the probabilities $p_i(t)(i = 1, \ldots, m)$ of the Markov jump process $\{J(t); t \geq 0\}$ being in transient state i at time t. We collect these probabilities in the vector $\boldsymbol{p}(t)$. Further we define the vector $\boldsymbol{p_+}(t) = (\boldsymbol{p}(t), p_{m+1}(t))$, which can be found as the standard solution to the Chapman-Kolmogorov equations

$$\boldsymbol{p_+}'(t) = \boldsymbol{p_+}(t)\mathbf{Q} \ , \tag{1.29}$$

such that $\boldsymbol{p}(t)$ satisfies:

$$\boldsymbol{p}'(t) = \boldsymbol{p}(t)\boldsymbol{S} \ . \tag{1.30}$$

The solution to this system is $\boldsymbol{p}(t) = \boldsymbol{\alpha} e^{t\boldsymbol{S}}$ ([1] page 182). Thus the probability that the jump process is not yet absorbed at time t is $\boldsymbol{p}(t)\mathbf{1} = \boldsymbol{\alpha} e^{t\boldsymbol{S}} = P(X > t)$, where X is the time to absorption. We get

$$P(X \leq x) = F(x) = 1 - \boldsymbol{\alpha} e^{\boldsymbol{S}x}\mathbf{1} \ . \tag{1.31}$$

Now using $e^{x\boldsymbol{S}} = \sum_{i=0}^{\infty} \frac{(x\boldsymbol{S})^i}{i!}$ we find $f(x) = F'(x) = -\boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{S}\mathbf{1}$. Finally using $\boldsymbol{S}\mathbf{1} + \boldsymbol{s} = \mathbf{0}$ we get

$$f(x) = \boldsymbol{\alpha} e^{\boldsymbol{S}x}\boldsymbol{s} \ . \tag{1.32}$$

Example 8. Choosing the dimension of $\boldsymbol{S}$ to be 1 we find the

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ 0 & 0 \end{bmatrix}$$

corresponding to the phase–type representation $((1),[-\lambda])$. We find $F(t) = 1 - e^{-\lambda t}$ and $f(t) = \lambda e^{-\lambda t}$ - an exponential distribution.

1.4.2 The Laplace-transform

The Laplace transform of a continuous probability distribution for the random variable X is defined by $\mathbb{E}\left(e^{-\theta X}\right)$. For the continuous part of a phase type distribution we need to evaluate $\int_0^\infty e^{-\theta t} f(t)dt = \boldsymbol{\alpha}(\theta \boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s}$, and we get

$$\mathbb{E}\left(e^{-\theta X}\right) = H(\theta) = \alpha_{m+1} + \boldsymbol{\alpha}(\theta \boldsymbol{I} - \boldsymbol{S})^{-1}\boldsymbol{s} \tag{1.33}$$

Theorem 2. *Let $\boldsymbol{U} = (-\boldsymbol{S})^{-1}$, then the (i,j)th element u_{ij} is the expected time spent in state j given initiation in state i prior to absorption.*

Proof. Let Z_j denote the time spent in state j prior to absorption. Then

$$\begin{aligned}
\mathbb{E}(Z_j|J(0) = i) &= \mathbb{E}\left(\int_0^\tau \delta_{X(t)=j} dt | J(0) = i\right) \\
&= \int_0^\infty \mathbb{E}\left(\delta_{X(t)=j}\delta_{\tau \geq t}|_{J(0)=i}\right) dt \\
&= \int_0^\infty P\left(\delta_{X(t)=j}\delta_{\tau \geq t}|J(0) = i\right) dt \\
&= \int_0^\infty \left(e^{\boldsymbol{S}t}\right)_{ij} dt \\
&= (-\boldsymbol{S})^{-1}
\end{aligned}$$

1.4.3 Moments

We now have

Corollary 1. *The mean of a PH$(\boldsymbol{\alpha},\boldsymbol{S})$ distributed random variable is $\boldsymbol{\alpha}\boldsymbol{U}\mathbf{1}$.*

Proof. Just notice that $\boldsymbol{U}\mathbf{1}$ is the vector which i'th element is the expected time the process spends in any state prior to absorption given initiation in state i.

We can alternatively get the mean and the all non–central moments by successive differentiation of the Laplace transform. Doing this yields

$$\mu_i = i!\boldsymbol{\alpha}(-\boldsymbol{S})^{-i}\mathbf{1} \tag{1.34}$$

1.4.4 Evaluation of continuous phase–type distributions

Phase–type distributions have a rational Laplace transform. The cumulative distribution function and the probability density function will thus consist of terms of the form

$$t^i cos(\omega t + \phi) e^{-\lambda t} \quad ,$$

where i, ω, or ϕ could be 0 leading to simpler expressions.

In order to get such explicit (scalar) expressions one can use the following approach.

- Calculate $e^{\boldsymbol{S}t}$ by deriving the first terms in the series $\sum_{i=0}^{\infty} \frac{(\boldsymbol{S}t)^i}{i!}$ and then prove a general result by induction. However, this approach is usually quite cumbersome and difficult. Generally the $\boldsymbol{S}$ matrix should be upper or lower diagonal for this approach to be viable.

- Alternatively one can determine the Laplace transform, find the roots of the denominator, and then use a partial fraction expansion. Inversion of each term in the fractional expansion is now straightforward. However, to find the roots of the denominator is equivalent to finding the root of an m'th order polynomial which is non–trivial in general.

The numerical evaluation is straightforward if one of the two above mentioned methods works. In general one must resort to numerical solution of the linear equation system governing the probabilities $\boldsymbol{p}(t)$, i.e. solving the Chapman Kolmogorov equations numerically. There is a very efficient method called uniformization for calculating this solution. Introducing the quantity $\eta = -\min(T_{ii})$ one rewrites $\boldsymbol{S} = \eta(\boldsymbol{K} - \boldsymbol{I})$. The matrix $\boldsymbol{K}$ is a sub-stochastic matrix such that $\boldsymbol{K} = \boldsymbol{I} + \eta^{-1}\boldsymbol{S}$. Now

$$\boldsymbol{\alpha} e^{\boldsymbol{S}t} \mathbf{1} = e^{-\eta t} \sum_{i=0}^{\infty} \frac{\boldsymbol{\alpha}(\eta t)^i \boldsymbol{K}^i \mathbf{1}}{i!} = \sum_{i=0}^{\infty} \boldsymbol{\alpha} \boldsymbol{K}^i \mathbf{1} \frac{(\eta t)^i}{i!} e^{-\eta t} \quad .$$

This formula is very well suited for numerical evaluation as all terms in the series are non–negative and since an appropriate level for truncation of the sum can be derived from the Poisson distribution.

1.4.5 Properties of continuous phase–type distributions

As for discrete phase–type distributions the class of continuous phase–type distributions is closed under a number of standard operations occurring frequently in probability theory.

1.4.5.1 Addition of two random variables

Theorem 3. *For $X \in PH(\boldsymbol{\alpha}, \boldsymbol{S})$ and $Y \in PH(\boldsymbol{\beta}, \boldsymbol{T})$ and independent $Z = X + Y \in PH(\boldsymbol{\gamma}, \boldsymbol{L})$ with $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \alpha_{m+1}\boldsymbol{\beta})$ and*

$$\begin{bmatrix} \boldsymbol{L} & \boldsymbol{l} \\ \mathbf{0} & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{S} & \boldsymbol{s}\boldsymbol{\beta} & \beta_{k+1}\boldsymbol{s} \\ \mathbf{0} & \boldsymbol{T} & \boldsymbol{t} \\ \mathbf{0} & & 0 \end{bmatrix} \tag{1.35}$$

$$\begin{bmatrix} \boldsymbol{L} & \boldsymbol{l} \\ \boldsymbol{0} & 0 \end{bmatrix} = \left[\begin{array}{cccc|cccc|c} S_{11} & S_{12} & \dots & S_{1m} & s_1\beta_1 & s_1\beta_2 & \dots & s_1\beta_k & s_1\beta_{k+1} \\ S_{21} & S_{22} & \dots & S_{2m} & s_2\beta_1 & s_2\beta_2 & \dots & s_2\beta_k & s_2\beta_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{m1} & S_{m2} & \dots & S_{mm} & s_m\beta_1 & s_m\beta_2 & \dots & s_m\beta_k & s_m\beta_{k+1} \\ \hline 0 & 0 & \dots & 0 & T_{11} & T_{12} & \dots & T_{1k} & t_1 \\ 0 & 0 & \dots & 0 & T_{21} & T_{22} & \dots & T_{2k} & t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & T_{k1} & T_{k2} & \dots & T_{kk} & t_k \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \end{array}\right] \tag{1.36}$$

$\boldsymbol{\gamma} = (\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}\beta_1, \alpha_{m+1}\beta_2, \dots, \alpha_{m+1}\beta_k)$. In matrix notation $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \alpha_{m+1}\boldsymbol{\beta})$.

Example 9. Consider the sum $Z = \sum_{i=1}^{k} X_i$ with $X_i \in exp(\lambda_i)$. Using 1.35 we get

$$\boldsymbol{S} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_4 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\lambda_{k-2} & \lambda_{k-2} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\lambda_{k-1} & \lambda_{k-1} \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & -\lambda_k \end{bmatrix}$$

$\boldsymbol{\alpha} = (1, 0, \dots, 0)$. With $\lambda_i = \lambda$ we get a sum of identically distributed exponential random variables, referred to as an Erlang–distribution. These distributions are special cases of the gamma distribution with integer shape parameter. We have

$$f(x) = \quad \lambda \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} \tag{1.37}$$

$$F(x) = \sum_{i=k}^{\infty} \frac{(\lambda x)^i}{i!} e^{-\lambda x} = 1 - \sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} \tag{1.38}$$

$$H(\theta) = \quad \left(\frac{\lambda}{\theta+\lambda}\right)^k \tag{1.39}$$

$$\mu_i = \quad \frac{(i+k-1)!}{(k-1)!\lambda^i} \tag{1.40}$$

1.4.5.2 Finite mixtures

We restate a result which is identical to the discrete case even in formulation

Theorem 4. *Any finite convex mixture of phase–type distributions is a phase type distribution. Let* $X_i \in PH(\boldsymbol{\alpha}_i, \boldsymbol{S}_i)$ $i = 1, \dots, k$ *such that* $Z = X_i$ *with probability* p_i *Then* $Z \in PH(\boldsymbol{\gamma}, \boldsymbol{L})$ *where* $\boldsymbol{\gamma} = (p_1\boldsymbol{\alpha_1}, p_2\boldsymbol{\alpha_2}, \dots, p_k\boldsymbol{\alpha_k})$ *and*

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{S}_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{S}_k \end{bmatrix} \tag{1.41}$$

Proof. Directly using the probabilistic interpretation of $(\boldsymbol{\gamma}, \boldsymbol{L})$.

Example 10. Consider the k random variables $X_i \in exp(\lambda_i)$ and assume that Z takes the value of X_i with probability p_i. The distribution of Z can be expressed as a proper mixture of the X_i's. $f_z(x) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i x}$. The distribution of Z is called a hyper exponential distribution. Using 1.41 we find a phase–type representation $(\boldsymbol{\gamma}, \boldsymbol{L})$ for Z. $\boldsymbol{\gamma} = (p_1, p_2, \dots, p_k)$.

$$\boldsymbol{L} = \begin{bmatrix} -\lambda_1 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & -\lambda_2 & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & -\lambda_k \end{bmatrix} \tag{1.42}$$

The hyper-exponential distribution is quite important and we mention its characteristics explicitly.

$$f(x) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i x} \tag{1.43}$$

$$F(x) = 1 - \sum_{i=1}^{k} p_i e^{-\lambda_i x} \tag{1.44}$$

$$H(s) = \sum_{i=1}^{k} \frac{p_i \lambda_i}{s + \lambda_i} \tag{1.45}$$

$$\mu_i = i! \sum_{i=1}^{k} \frac{p_i}{\lambda_i^i} \tag{1.46}$$

1.4.5.3 Order statistics

The order statistic of a finite number of independent discrete phase–type distributed variables is itself phase–type distributed. We will focus on the distribution of the smallest and the largest of two independent variables X with representation $(\boldsymbol{\alpha}_x, \boldsymbol{S}_x)$ and Y with representation $(\boldsymbol{\alpha}_y, \boldsymbol{S}_y)$. We motivate the derivation with a small example.

Example 11. Let X be generalized Erlang distributed with parameters $\lambda_{x,1}, \lambda_{x,2}$ and let Y be hyper-exponentially distributed with parameters $p_y, \lambda_{y,1}, \lambda_{y,2}$. We will first investigate the distribution of $\min(X, Y)$. We have

$$\boldsymbol{S}_x = \begin{bmatrix} -\lambda_{x,1} & \lambda_{x,1} \\ 0 & -\lambda_{x,2} \end{bmatrix} \qquad \boldsymbol{S}_y = \begin{bmatrix} -\lambda_{y,1} & 0 \\ 0 & -\lambda_{y,2} \end{bmatrix}$$

with $\boldsymbol{\alpha_x} = (1, 0)$ and $\boldsymbol{\alpha_y} = (p_y, 1 - p_y)$. We can now construct a Markov jump process that simultaneously describes the evolution of the two Markov jump processes related to X and Y respectively. As for Example 6 the Markov jump process will have 4 states corresponding to the all possible combinations of the states of the two original processes.

We denote the four states by $(1,1), (1,2), (2,1), (2,2)$. The transitions from (1,1) to (2,1), and from (1,2) to (2,2) occur with intensity $\lambda_{x,1}$, while no other transitions are possible from these two states. The remaining

transitions are found similarly. In summary we get a Markov jump process that describes the time until the first of the two processes gets absorbed ($\min(X,Y)$). This time is obviously then phase–type distributed with generator

$$\boldsymbol{S}_{\min(X,Y)} = \begin{bmatrix} -(\lambda_{x,1}+\lambda_{y,1}) & 0 & \lambda_{x,1} & 0 \\ 0 & -(\lambda_{x,1}+\lambda_{y,2}) & 0 & \lambda_{x,1} \\ 0 & 0 & -(\lambda_{x,2}+\lambda_{y,1}) & 0 \\ 0 & 0 & 0 & -(\lambda_{x,2}+\lambda_{y,2}) \end{bmatrix}$$

and with initial probability vector $\boldsymbol{\alpha}_{\boldsymbol{\min(X,Y)}} = (p_y, 1-p_y, 0, 0)$. Further we have $\boldsymbol{S}_{\min(X,Y)} = \boldsymbol{S}_x \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{S}_y$ and $\boldsymbol{\alpha}_{\boldsymbol{\min(X,Y)}} = \boldsymbol{\alpha}_{\boldsymbol{x}} \otimes \boldsymbol{\alpha}_{\boldsymbol{y}}$.

To get the distribution of $\max(X,Y)$ we need additionally to consider the states $(1,3),(2,3),(3,1),(3,2)$ corresponding to the event that one of the processes has reached the absorbing state. It is convenient to use the ordering $(1,1),(1,2),(2,1),(2,2),(1,3),(2,3),(3,1),(3,2)$. And we find $\boldsymbol{S}_{\max(X,Y)}$ to be

$$\boldsymbol{S}_{\max(X,Y)} =$$

$$\begin{bmatrix} -(\lambda_{x,1}+\lambda_{y,1}) & 0 & \lambda_{x,1} & 0 & \lambda_{y,1} & 0 & 0 & 0 \\ 0 & -(\lambda_{x,1}+\lambda_{y,2}) & 0 & \lambda_{x,1} & \lambda_{y,2} & 0 & 0 & 0 \\ 0 & 0 & -(\lambda_{x,2}+\lambda_{y,1}) & 0 & 0 & \lambda_{y,1} & \lambda_{x,2} & 0 \\ 0 & 0 & 0 & -(\lambda_{x,2}+\lambda_{y,2}) & 0 & \lambda_{y,2} & 0 & \lambda_{x,2} \\ 0 & 0 & 0 & 0 & -\lambda_{x,1} & \lambda_{x,1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\lambda_{x,2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{y,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{y,2} \end{bmatrix}$$

Theorem 5. *For $X \in PH(\boldsymbol{\alpha}_x, \boldsymbol{S}_x)$ and $Y \in PH(\boldsymbol{\alpha}_y, \boldsymbol{S}_y)$ $\min(X,Y)$ is phase distributed with representation $(\boldsymbol{\gamma}, \boldsymbol{L})$ given by 1.47:*

$$\boldsymbol{L} = \boldsymbol{S}_x \otimes \boldsymbol{I}_y + \boldsymbol{I}_x \otimes \boldsymbol{S}_y \quad , \tag{1.47}$$

where $\boldsymbol{\gamma} = \boldsymbol{\alpha}_{\boldsymbol{x}} \otimes \boldsymbol{\alpha}_{\boldsymbol{y}}$. and $\max(X,Y)$ is phase type distributed with representation $(\boldsymbol{\gamma}, \boldsymbol{L})$ given by 1.48:

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{S}_x \otimes \boldsymbol{I}_y + \boldsymbol{I}_x \otimes \boldsymbol{S}_y & \boldsymbol{I}_x \otimes \boldsymbol{s}_y & \boldsymbol{s}_x \otimes \boldsymbol{I}_y \\ \boldsymbol{0} & \boldsymbol{S}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_y \end{bmatrix} \tag{1.48}$$

med $\boldsymbol{\gamma} = (\boldsymbol{\alpha}_{\boldsymbol{x}} \otimes \boldsymbol{\alpha}_{\boldsymbol{y}}, \boldsymbol{\alpha}_{\boldsymbol{x}} \alpha_{y,m+1}, \alpha_{x,k+1} \boldsymbol{\alpha}_{\boldsymbol{y}})$. where the dimension of $\boldsymbol{S}_x$ is k and the dimension of $\boldsymbol{S}_y$ is m. We give $\boldsymbol{l}$ explicitly

$$\boldsymbol{l} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{s}_x \\ \boldsymbol{s}_y \end{bmatrix} \tag{1.49}$$

1.4.6 Phase–type renewal process

A phase–type renewal process is a point process, were the distance between two points can be described by independent and identically distributed phase–type distributed random variables. The stationary version is of special interest. The age of the process at time t is defined to be the time since the last event - or point - while the residual life time at time t is the time to the next event - or point. We have the following important result.

Theorem 6. *In a stationary phase–type renewal process the distribution of age and residual life time are phase–type distributed with representation* $(\boldsymbol{\pi}, \boldsymbol{S})$ *where* $\boldsymbol{\pi} = \frac{\boldsymbol{\alpha U}}{\boldsymbol{\alpha U 1}}$.

Proof. The phase process $J(t)$ is a finite continuous time Markov jump process with generator matrix $\boldsymbol{Q} = \boldsymbol{S} + \boldsymbol{s\alpha}$ of dimension m. The stationary probability vector $\boldsymbol{\pi}$ for this process satisfies $\boldsymbol{\pi}(\boldsymbol{S} + \boldsymbol{s\alpha}) = \mathbf{0}$ or equivalently $\boldsymbol{\pi} = \boldsymbol{\pi s\alpha U}$. By probabilistic reasoning or by post-multiplying with $\mathbf{1}$ we see that $\boldsymbol{\pi s} = \frac{1}{\boldsymbol{\alpha U 1}}$. At an arbitrary epoch the probabilistic distribution among the states in the Markov jump process is given by $\boldsymbol{\pi}$. From this point the time to absorption will be phase type distributed with the representation of the theorem. The distribution of age must be the same by symmetry.

Corollary 2. *The moments of the residual life time distribution is given by*

$$\mu_i^\star = i! \frac{\boldsymbol{\alpha U}^{i+1}\mathbf{1}}{\boldsymbol{\alpha U 1}} \quad . \tag{1.50}$$

Proof. The corollary follows directly from Theorem 6 and the equation for the moments of a phase–type distribution as given in Section 1.4.3.

Example 12. Let X be an Erlang-n distributed random variable, that is $X \in \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$ as in Example 9 with $\lambda_i = \lambda$. The matrix $\boldsymbol{S} + \boldsymbol{s\alpha}$ is given by:

$$\boldsymbol{S} + \boldsymbol{s\alpha} = \begin{bmatrix} -\lambda & \lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & \dots & 0 & -\lambda & \lambda \\ \lambda & 0 & 0 & \dots & 0 & 0 & -\lambda \end{bmatrix}$$

and we get $\boldsymbol{\pi} = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$.

1.4.7 Non–uniqueness of continuous phase–type distributions

In general there will be many phase type representations for the same distribution. Consider the following example.

Example 13. With $\boldsymbol{\alpha} = (1, 0)$ and $\boldsymbol{S}$ given by

$$\boldsymbol{S} = \begin{bmatrix} -\lambda_1 & \lambda_1 - \lambda_2 \\ 0 & -\lambda_2 \end{bmatrix} \tag{1.51}$$

where $\lambda_1 \geq \lambda_2$ we get an exponential distribution with intensity λ_2.

References

1. D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall, 1980.
2. Ole Groth Jørsboe. *Funktionalanalyse*. Matematisk Institut. Danmarks Tekniske Højskole, 1987.