DTU Library

On the use of functional calculus for phase-type and related distributions

Bladt, Mogens; Campillo Navarro, Azucena; Nielsen, Bo Friis

Link back to DTU Orbit

ON THE USE OF FUNCTIONAL CALCULUS FOR PHASE–TYPE AND RELATED DISTRIBUTIONS

MOGENS BLADT, AZUCENA CAMPILLO NAVARRO & BO FRIIS NIELSEN

DTU COMPUTE TECHNICAL REPORT-2014-18

ABSTRACT. The area of phase–type distributions is renowned for its ability to obtain closed form formulas or algorithmically exact solutions to many complex stochastic models. The method of functional calculus will provide an additional tool along these lines for establishing results in terms of functions of matrices. Functional calculus, which is a branch of operator theory frequently associated with complex analysis, can be applied to phase–type and matrix–exponential distributions in a rather straightforward way. In this paper we provide a number of examples on how to execute the formal arguments.

1. INTRODUCTION

In the area of phase–type or matrix–exponential distributions we are often able to obtain closed form formulas for many complex stochastic models. Examples include closure properties of phase-type distributions such as convolutions, moment distributions, maximum or minimum, or expressions for renewal densities, ruin probabilities and stationary distributions of certain queues. Often the problem involves establishing a representation, or computing some functional of a phase–type distributed random variable X on the form $\mathbb{E}(w(X))$ for some function w.

In this paper we introduce the method of functional calculus to show how to obtain closed form formulas in a number of situations for expressions like $\mathbb{E}(w(X))$, and we further provide an example from moment distributions that establishes a matrix–exponential representation of an

---

n'th order moment distribution which is intriguingly similar to a block–Erlang form. We also provide a short proof for the canonical representation of a matrix-exponential distribution, which is a representation parametrised in terms of the coefficients of its Laplace transform.

Expectations on the form $\mathbb{E}(w(X))$ appear naturally in a variety of situations. For particular choices of w like

$$w(x) = e^{-sx}, \quad w(x) = e^{ix} \ \text{ and } \ w(x) = x^{\alpha-1}$$

the expectation produces the Laplace transform, characteristic function and Mellin transform respectively, which all characterise the distribution. In size–biased sampling, we have densities on the form

$$g(x) = \frac{w(x)f(x)}{\mu_w}, \quad \mu_w = \mathbb{E}(w(X)),$$

where f is a given density, w is a non-negative function, and X a random variable with density f. If $w(x) = x^n$ we are talking about moment distributions of order n, but examples with e.g. fractional powers of X also appear in practice (see e.g. [16]). Of course it is of interest to be able to express the normalising factor μ_w in closed form whenever possible.

The rest of the paper is organised as follows. In section 2 we review some basic concepts regarding functions of matrices and (holomorphic) functional calculus. In section 3 we setup the basic notation and facts regarding phase–type distributions, and in section 4 we express the expected value of $w(X)$ in terms of the Laplace transform of w, and apply the result to some particular choices of w. We also provide an alternative derivation for the representation of the n'th order moment distributions using functional calculus. As it turns out, the calculations we perform are not particularly linked to the probabilistic structure of phase–type distributions, so in section 5 we consider various extensions including matrix–exponential distributions, multivariate phase–type, and a class of heavy tailed infinite–dimentional phase–type distributions, where the Mellin transform are of particular interest.

2. Functional Calculus

Functional calculus can most easily be understood as a method for replacing real and complex variables of functions with matrices, or more generally, operators. A well known example is the matrix exponential $\exp(\boldsymbol{A})$ of a finite square matrix $\boldsymbol{A}$. Less known examples are $\cos(\boldsymbol{A})$ or $\boldsymbol{A}^\alpha$ for non–integer values of α. In this paper we shall show how holomorphic functional calculus may be used in the area of phase–type and matrix–exponential distributions.

If $\boldsymbol{A}$ is a finite square matrix and f a function, we may define $f(\boldsymbol{A})$ in several equivalent ways. If f is the polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, the operator

$$f(\boldsymbol{A}) := a_0 \boldsymbol{I} + a_1 \boldsymbol{A} + \cdots + a_n \boldsymbol{A}^n$$

is well defined. More generally, if f is an analytic function

$$f(z) = \sum_{n=0}^{\infty} a_n (z - a)^n, \ a, z \in \mathbb{C},$$

with radius of convergence R, we may then define

$$f(\boldsymbol{A}) = \sum_{n=0}^{\infty} a_n (\boldsymbol{A} - a\boldsymbol{I})^n$$

if $|s_i - a| < R$ for all eigenvalues s_i of $\boldsymbol{A}$.

A more general definition which is applicable also for non-analytic functions uses the Jordan decomposition. Here the function f is required to be sufficiently smooth with derivatives existing of sufficiently high order. The matrix $\boldsymbol{A}$ can be written on Jordan canonical form as

$$\boldsymbol{A} = \boldsymbol{MJM}^{-1}$$

for some matrix $\boldsymbol{M}$ and

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{J}_2 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{J}_3 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{J}_k \end{pmatrix},$$

where

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

is the i'th Jordan block. Let $\mu_1, ..., \mu_m$ denote distinct eigenvalues of $\boldsymbol{A}$ and $n_1, ..., n_m$ the highest orders among the Jordan blocks associated with the eigenvalues. We suppose that f is defined on the spectrum of $\boldsymbol{A}$ (sp($\boldsymbol{A}$)) with $n_i - 1$ derivatives at the point μ_i for $i = 1, 2, \ldots, m$.

Then we define

$$f(\boldsymbol{A}) = \boldsymbol{M} \begin{pmatrix} f(\boldsymbol{J}_1) & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & f(\boldsymbol{J}_2) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & f(\boldsymbol{J}_3) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & f(\boldsymbol{J}_k) \end{pmatrix} \boldsymbol{M}^{-1}$$

with

$$f(\boldsymbol{J}_i) = \begin{pmatrix} f(s_i) & f'(s_i) & \frac{f''(s_i)}{2!} & \cdots & \frac{f^{(p_i-1)}(s_i)}{(p_i-1)!} \\ 0 & f(s_i) & f'(s_i) & \cdots & \frac{f^{(p_i-2)}(s_i)}{(p_i-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f(s_i) \end{pmatrix}$$

where $s_i = \mu_j$ for some $1 \le j \le m$ and $p_i \le n_j$.

For analytic functions f we may provide an alternative definition in terms of the Cauchy integral which has the advantage that it can be easily generalised to more general operators like e.g. infinite dimensional matrices or general bounded operators.

Let A be a bounded operator on a Banach space $(X, \|\cdot\|)$. The term $(z\boldsymbol{I} - A)^{-1}$ is called the resolvent of the operator and defines the spectrum. The spectrum consists of values $z \in \mathbb{C}$ for which the resolvent is not a linear operator on the Banach space. Hence all eigenvalues for $\boldsymbol{A}$ are contained in the spectrum, but in general spectra for operators on infinite dimensional Banach spaces may be larger than the set of their eigenvalues. The function of the operator A can now be defined in the following way.

Definition 2.1. *Let A be a bounded operator on a Banach space $(X, \|\cdot\|)$ and let γ be a simple closed path which encloses the spectrum of A. If f is a function which is holomorphic (analytic) on and inside the path γ, then we define*

$$f(A) = \frac{1}{2\pi i} \oint_\gamma f(z)(z\boldsymbol{I} - A)^{-1} dz.$$

All definitions discussed in this section are equivalent when the function is analytic. The theory of (holomorphic) functional calculus is treated extensively in [10], whereas [11] deals with methods for defining functions of matrices different to the Cauchy integral.

3. Phase–type distributions

Consider a Markov jump processes $\{X_t\}_{t\ge 0}$ with a finite state space $E = \{1, 2, \ldots, p+1\}$, where $\{1, 2, \ldots, p\}$ are transient states and the

state $p+1$ is absorbing. Then the process $\{X_t\}_{t\geq0}$ has an intensity matrix on the form

$$\begin{pmatrix} \boldsymbol{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\boldsymbol{T}$ is a $p\times p$ dimensional sub-intensity matrix, $\mathbf{t}$ is a p-dimensional column vector of exit intensities (exit to the absorbing state), and $\mathbf{0}$ is a p-dimensional row vector of zeros. The initial distribution of $\{X_t\}_{t\geq0}$ is denoted by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)$, where $\pi_i = \mathbb{P}(X_0 = i)$, and it is concentrated on the first p states, i.e. $\mathbb{P}(X_0 = p+1) = 0$.

Definition 3.1. *The time until absorption*

$$\tau = \inf\{t \geq 0 \mid X_t = p+1\}$$

is said to have a phase–type distribution and we write

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

The set of parameters $(\boldsymbol{\pi}, \boldsymbol{T})$ is said to be a representation of the phase–type distribution, and the dimension of $\boldsymbol{\pi}$, which is p, is said to be the dimension of the phase-type representation. We also write $\tau \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ to specify the dimension of the phase-type representation.

For the present article it is important to mention some basic properties of phase-type distributions. If $\tau \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$, the density f of τ is given by

$$f(s) = \boldsymbol{\pi} \exp(\boldsymbol{T}s) \mathbf{t},$$

where $\mathbf{t} = -\boldsymbol{T}\mathbf{e}$ and $\mathbf{e}$ is a p-dimensional column vector of ones. The corresponding distribution function is

$$F(s) = 1 - \boldsymbol{\pi} \exp(\boldsymbol{T}s) \mathbf{e}.$$

The sub-intensity matrix $\boldsymbol{T}$ has eigenvalues with strictly negative real parts. In particular, $\boldsymbol{T}$ is invertible and we have the following integration rule

$$\int_0^\infty \exp(\boldsymbol{T}s)\,ds = \boldsymbol{T}^{-1}\exp(\boldsymbol{T}s)\Big|_0^\infty = -\boldsymbol{T}^{-1}.$$

Finally we shall make optimal use of the n-th order $(n \geq 1)$ moment of τ, which is given by

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

For further properties on phase–type distributions, see e.g. [14], [1] or [5].

4. Functional calculus for phase–type distributions

Let $X \sim \mathrm{PH}_p(\boldsymbol{\pi}, \boldsymbol{T})$, $p < \infty$, and suppose that the weight function w has a Taylor expansion

$$w(x) = \sum_{n=0}^{\infty} a_n (x - a)^n$$

with radius of convergence $R > 0$. For specific choices of w we may then calculate $\mu_w = \mathbb{E}(w(X))$.

As a motivation we provide the following example where μ_w can be calculated in a straightforward manner by a power series expansion.

Example 4.1 (cosine and sine). Let $w(x) = \cos(x)$. Then

$$w(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!},$$

which is convergent in all $\mathbb{R}$. Since the density of a phase–type distribution is bounded, $w(x)\boldsymbol{\pi} \exp(\boldsymbol{T}x)\mathbf{t}$ has a power series expansion with the same radius of convergence, which in this case is infinity. Hence we may exchange summation and integration and we get

$$
\begin{aligned}
\mathbb{E}(w(X)) &= \int_0^\infty w(x)\boldsymbol{\pi} e^{\boldsymbol{T}x}\mathbf{t}\, dx \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \int_0^\infty x^{2n}\boldsymbol{\pi} e^{\boldsymbol{T}x}\mathbf{t}\, dx \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} (2n)!\boldsymbol{\pi}(-\boldsymbol{T}^{-1})^{2n}\mathbf{e} \\
&= \sum_{n=0}^{\infty} \boldsymbol{\pi}\left(-\boldsymbol{T}^{-2}\right)^n \mathbf{e} \\
&= \boldsymbol{\pi}\left(\boldsymbol{I} + \boldsymbol{T}^{-2}\right)^{-1}\mathbf{e} \\
&= \boldsymbol{\pi}(-\boldsymbol{T})\left(\boldsymbol{I} + \boldsymbol{T}^2\right)^{-1}\mathbf{t}.
\end{aligned}
$$

Similarly, for $w(x) = \sin(x)$ we have that

$$w(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{(n+1)!},$$

which has an infinite spectral radius of convergence, and proceeding as for cosine, we get that

$$\mathbb{E}(\sin(X)) = \boldsymbol{\pi}(\boldsymbol{I} + \mathbf{T}^2)^{-1}\mathbf{t}.$$

$\square$

We have now seen some examples of expressing μ_w in terms of $\boldsymbol{\pi}$, $\boldsymbol{T}$ and $\mathbf{t}$. The question rises if it is possible to express $\mathbb{E}(w(X))$ in terms of these parameters for more general (analytic) functions. The answer to this question is surprisingly simple.

Theorem 4.2. *Let w be a function with Laplace transform*

$$L_w(s) = \int_0^\infty e^{-sx} w(x) dx$$

that exists for all $s \geq 0$. Then

$$\mathbb{E}(w(X)) = \boldsymbol{\pi} L_w(-\boldsymbol{T})\mathbf{t}.$$

Proof. The Laplace transform is analytic in its domain of convergence, hence it is analytic in the positive half–plane. The spectrum of $-\boldsymbol{T}$ consists of a finite number of eigenvalues which all have strictly positive real parts, so we can find a simple closed path that encloses the eigenvalues. It is located within the positive half–plane. Therefore $L_w(-\boldsymbol{T})$ can be defined in accordance with Definition 2.1, and it follows that

$$\mu_w = \boldsymbol{\pi} \int_0^\infty e^{\boldsymbol{T}x} w(x) dx \mathbf{t} = \boldsymbol{\pi} L_w(-\boldsymbol{T})\mathbf{t}.$$

□

Theorem 4.3 (sine and cosine transforms). *Let $X \sim PH_p(\boldsymbol{\pi}, \boldsymbol{T})$ with density f. The sine and cosine transforms for X (or equivalently f), defined by*

$$
\begin{aligned}
\hat{f}_c(u) &= \mathbb{E}(\cos(uX)) = \int_0^\infty f(x)\cos(ux)dx \\
\hat{f}_s(u) &= \mathbb{E}(\sin(uX)) = \int_0^\infty f(x)\sin(ux)dx
\end{aligned}
$$

respectively, are given by

$$\hat{f}_c(u) = \boldsymbol{\pi}(-\boldsymbol{T})(u^2\boldsymbol{I} + \mathbf{T}^2)^{-1}\mathbf{t} \quad and \quad \hat{f}_s(u) = u\boldsymbol{\pi}(u^2\boldsymbol{I} + \mathbf{T}^2)^{-1}\mathbf{t}.$$

Remark 4.4. *Please note that we recover the results of Example 4.1 with $u = 1$.*

Proof. The Laplace transforms of sine and cosine are respectively given by

$$L_{\sin}(t) = \frac{t}{1 + t^2}, \quad \text{and} \quad L_{\cos}(t) = \frac{1}{1 + t^2}.$$

Hence

$$
\begin{aligned}
\hat{f}_s(u) &= \int_0^\infty \boldsymbol{\pi} e^{\boldsymbol{T}x} \mathbf{t} \sin(ux) dx \\
&= \int_0^\infty \boldsymbol{\pi} e^{\boldsymbol{T}x/u} \mathbf{t} \frac{1}{u} \sin(x) dx \\
&= \frac{1}{u} \boldsymbol{\pi} L_{\sin}(-\boldsymbol{T}/u) \mathbf{t} \\
&= u \boldsymbol{\pi} (u^2 \boldsymbol{I} + \boldsymbol{T}^2)^{-1} \mathbf{t}.
\end{aligned}
$$

The argument regarding $\hat{f}_c$ is entirely similar. $\qquad\square$

Another important weight function is the fractional power, and is closely related to the Mellin transform.

Theorem 4.5. *Let* $X \sim PH(\boldsymbol{\pi}, \boldsymbol{T})$. *Then the Mellin transform of* X,

$$
M_X(\alpha) = \mathbb{E}(X^{\alpha-1}) = \int_0^\infty s^{\alpha-1} \boldsymbol{\pi} e^{\boldsymbol{T}s} \mathbf{t} ds,
$$

which exists at least for $\alpha > 0$, *is given by*

$$
M_X(\alpha) = \Gamma(\alpha) \boldsymbol{\pi} (-\boldsymbol{T})^{-\alpha} \mathbf{t} = \Gamma(\alpha) \boldsymbol{\pi} (-\boldsymbol{T})^{-\alpha+1} \mathbf{e}.
$$

Proof. Consider the Laplace transform

$$
L(\alpha) = \int_0^\infty x^{\alpha-1} e^{-sx} dx
$$

with $s > 0$. This integral is convergent for $\alpha > 0$ so we conclude that L is analytic in the half–plane $\{z \in \mathbb{C} : Re(z) > 0\}$. Therefore the result follows from Theorem 4.2 and the fact that the Laplace transform for x^α is given by $\Gamma(\alpha + 1)/s^{\alpha+1}$.

$\qquad\square$

Example 4.6 (Mellin, sine and cosine transforms of a phase–type distribution)**.** Let $X \sim \mathrm{PH}_2(\boldsymbol{\pi}, \boldsymbol{T})$, where

$$
\boldsymbol{\pi} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix}, \quad \text{and} \quad \mathbf{t} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.
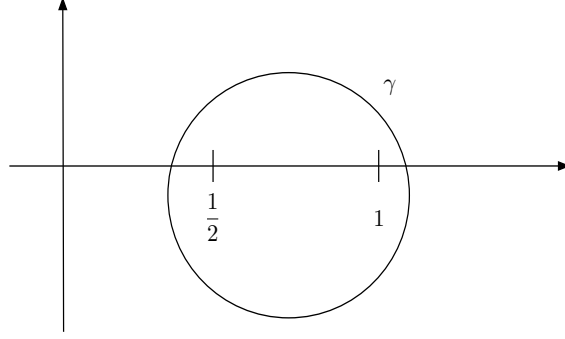$$

By the Theorem 4.5, the Mellin transform of X is given by

$$
M_X(\alpha) = \Gamma(\alpha) \boldsymbol{\pi} (\boldsymbol{U})^\alpha \mathbf{t},
$$

where $\boldsymbol{U} = -\boldsymbol{T}^{-1}$. In this case we have that

$$
\boldsymbol{U} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}
$$

with spectrum $\mathrm{sp}(\boldsymbol{U}) = \{1, \frac{1}{2}\}$.

FIGURE 1. A path surrounding the spectrum of U.

Now we calculate $\boldsymbol{U}^\alpha$ for $\alpha > 0$. For this, we choose a closed simple path γ which encloses the spectrum of $\boldsymbol{U}$ and for which the function $f(x) = x^\alpha$, $\alpha > 0$, is analytic on and inside γ (see Figure 1 for an example).

Hence by the Cauchy integral formula for the matrix $\boldsymbol{U}$ we get that

$$\boldsymbol{U}^\alpha = \frac{1}{2\pi i} \oint_\gamma z^\alpha (z\boldsymbol{I} - \boldsymbol{U})^{-1} dz.$$

Since

$$(z\boldsymbol{I} - \boldsymbol{U})^{-1} = \begin{pmatrix} z - 1 & -\frac{1}{2} \\ 0 & z - \frac{1}{2} \end{pmatrix}^{-1} = \begin{pmatrix} \dfrac{1}{z - 1} & \dfrac{1}{(z - 1)(2z - 1)} \\ 0 & \dfrac{2}{2z - 1} \end{pmatrix},$$

we have that

$$\boldsymbol{U}^\alpha = \begin{pmatrix} \dfrac{1}{2\pi i} \oint_\gamma \dfrac{z^\alpha}{z - 1} dz & \dfrac{1}{2\pi i} \oint_\gamma \dfrac{z^\alpha}{(z - 1)(2z - 1)} dz \\ 0 & \dfrac{1}{2\pi i} \oint_\gamma \dfrac{2z^\alpha}{2z - 1} dz \end{pmatrix}.$$

Then, using the residue theorem, we solve the integrals in the matrix $\boldsymbol{U}^\alpha$ obtaining

$$\frac{1}{2\pi i}\oint_\gamma \frac{z^\alpha}{z-1}dz = Res\left(\frac{z^\alpha}{z-1}, z=1\right) = 1,$$

$$\frac{1}{2\pi i}\oint_\gamma \frac{z^\alpha}{(z-1)(2z-1)}dz = Res\left(\frac{z^\alpha}{(z-1)(2z-1)}, z=1\right)$$
$$+Res\left(\frac{z^\alpha}{(z-1)(2z-1)}, z=\frac{1}{2}\right)$$
$$= 1-2^{-\alpha},$$

$$\frac{1}{2\pi i}\oint_\gamma \frac{2z^\alpha}{2z-1}dz = Res\left(\frac{2z^\alpha}{2z-1}, z=\frac{1}{2}\right)$$
$$= 2^{-\alpha},$$

where $Res(h, x=a)$ refers to the residue of the function h at a. Therefore

$$\boldsymbol{U}^\alpha = \begin{pmatrix} 1 & 1-2^{-\alpha} \\ 0 & 2^{-\alpha} \end{pmatrix},$$

and we can finally calculate the Mellin transform for X as

$$M_X(\alpha) = \Gamma(\alpha)\boldsymbol{\pi}(\boldsymbol{U})^\alpha\mathbf{t}$$
$$= \Gamma(\alpha)\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1-2^{-\alpha} \\ 0 & 2^{-\alpha} \end{pmatrix}\begin{pmatrix} 0 \\ 2 \end{pmatrix}$$
$$= 2(1-2^{-\alpha})\Gamma(\alpha).$$

Observe that if $\boldsymbol{\pi} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ then $M_X(\alpha) = \Gamma(\alpha)$. See [8] for a calculation using power series expansion.

The sine and cosine transforms are readily calculated for the present case to be

$$\hat{f}_s(u) = \frac{6u}{(u^2+1)(u^2+4)}$$
$$\hat{f}_c(u) = \frac{2(2-u^2)}{(u^2+1)(u^2+4)}.$$

A plot of $\hat{f}_s$ and $\hat{f}_c$ are given in Figure 2. The cosine transform is negative for some arguments u. It may go from positive to negative arbitrarily often. In fact, if f would be the Dirac delta function at the point a, then the cosine transform would be $\cos(ua)$, which crosses the real axis infinitely often. Since we take a sequence of Erlang distributions which converges in law to the degenerate distribution at some point a, it follows that the cosine and sine transforms may change signs arbitrarily (though finitely) often.
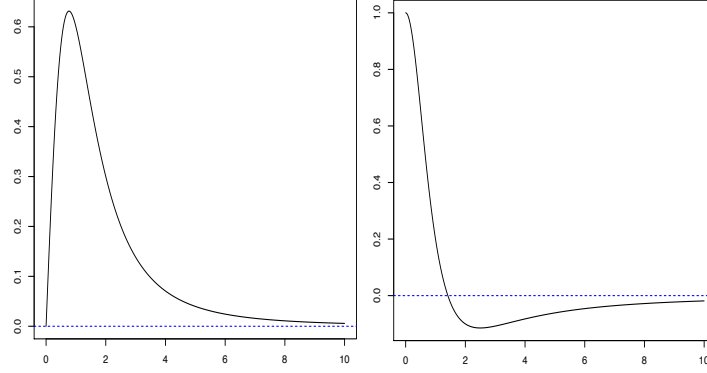
FIGURE 2.   Sine (left) and cosine (right) transform of
the phase–type distribution of Example 4.6.

$\square$

Example 4.7 (Mellin, cosine and sine transforms of an Erlang distribution)**.** Let $X \sim \mathrm{PH}_n\left(\boldsymbol{\pi}, \boldsymbol{T}\right)$, where

$$\boldsymbol{\pi} = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}, \quad \boldsymbol{T} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & \cdots & 0 \\ 0 & 0 & -\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -\lambda \end{pmatrix},$$

$$\text{and} \quad \mathbf{t} = \begin{pmatrix} 0 \\ \vdots \\ \lambda \end{pmatrix}.$$

The density for X is given by $\frac{x^{n-1}}{\Gamma(n)}\lambda^n \exp(-\lambda x)$, and is known as an Erlang distribution. It is the convolution of n independent exponential distributions with the same intensity λ. The Mellin transform for X is readily obtained by

$$\begin{aligned}
\mathbb{E}(X^{\alpha-1}) &= \int_0^\infty \frac{x^{\alpha-1+n-1}}{\Gamma(n)}\lambda^n \exp(-\lambda x)dx \\
&= \lambda^{-\alpha+1}\frac{\Gamma(\alpha+n-1)}{\Gamma(n)}.
\end{aligned}$$

The calculation via the phase–type representation is somewhat longer, but instructive for general purposes. We present the argument in the

following. To calculate the Mellin transform we need to obtain $\boldsymbol{U}^\alpha$.
Recall that

$$\boldsymbol{U}^\alpha = \frac{1}{2\pi i} \oint_\gamma z^\alpha (z\boldsymbol{I} - \boldsymbol{U})^{-1} dz.$$

Since

$$(z\boldsymbol{I} - \boldsymbol{U})^{-1} = \begin{pmatrix} \frac{\lambda}{z\lambda-1} & \frac{\lambda}{\lambda(z\lambda-1)^2} & \frac{z\lambda^2}{(z\lambda-1)^3} & \cdots & \frac{z^{n-2}\lambda^{n-1}}{(z\lambda-1)^n} \\[2ex] 0 & \frac{\lambda}{z\lambda-1} & \frac{\lambda}{(z\lambda-1)^2} & \cdots & \frac{z^{n-3}\lambda^{n-2}}{(z\lambda-1)^{n-1}} \\[2ex] 0 & 0 & \frac{\lambda}{z\lambda-1} & \cdots & \frac{z^{n-4}\lambda^{n-3}}{(z\lambda-1)^{n-2}} \\[2ex] \vdots & \vdots & \vdots & \ddots & \vdots \\[2ex] 0 & 0 & 0 & \cdots & \frac{\lambda}{z\lambda-1} \end{pmatrix}$$

we get that

$$\boldsymbol{U}^\alpha = \begin{pmatrix} \frac{1}{2\pi i} \oint_\gamma \frac{z^\alpha \lambda}{z\lambda-1} dz & \frac{1}{2\pi i} \oint_\gamma \frac{z^\alpha \lambda}{(z\lambda-1)^2} dz & \cdots & \frac{1}{2\pi i} \oint_\gamma \frac{z^{\alpha+n-2}\lambda^{n-1}}{(z\lambda-1)^n} dz \\[2ex] 0 & \frac{1}{2\pi i} \oint_\gamma \frac{z^\alpha \lambda}{z\lambda-1} dz & \cdots & \frac{1}{2\pi i} \oint_\gamma \frac{z^{\alpha+n-3}\lambda^{n-2}}{(z\lambda-1)^{n-1}} dz \\[2ex] 0 & 0 & \cdots & \frac{1}{2\pi i} \oint_\gamma \frac{z^{\alpha+n-4}\lambda^{n-3}}{(z\lambda-1)^{n-2}} dz \\[2ex] \vdots & \vdots & \ddots & \vdots \\[2ex] 0 & 0 & \cdots & \frac{1}{2\pi i} \oint_\gamma \frac{z^\alpha \lambda}{z\lambda-1} dz \end{pmatrix}.$$

Now, to simplify the last matrix we will start solving the integral which
is in the diagonal, that is

$$\begin{aligned} \frac{1}{2\pi i} \oint_\gamma \frac{z^\alpha \lambda}{z\lambda - 1} dz &= \lim_{z \to \frac{1}{\lambda}} \left(z - \lambda^{-1} \right) \frac{\lambda z^\alpha}{(\lambda z - 1)} \\ &= \lim_{z \to \frac{1}{\lambda}} z^\alpha \\ &= \lambda^{-\alpha}. \end{aligned}$$

The rest of the integrals can be written like

$$\frac{1}{2\pi i} \oint_\gamma z^\alpha \frac{z^{j-1}\lambda^j}{(z\lambda - 1)^{j+1}} dz, \quad j = 1, 2, \ldots, n-1.$$

Let $f_j(z) = z^\alpha \frac{z^{j-1}\lambda^j}{(z\lambda-1)^{j+1}}$. Then with

$$\varphi(z) = \left(z - \frac{1}{\lambda}\right)^{j+1} f_j(z) = \frac{1}{\lambda} z^{\alpha+j-1},$$

we obtain that

$$Res\left(f_j, z = \frac{1}{\lambda}\right) = \frac{\varphi^{(j)}(\frac{1}{\lambda})}{j!}$$

$$= \frac{(\alpha + j - 1)(\alpha + j - 2)\cdots\alpha}{j!\lambda^\alpha},$$

where $\varphi^{(j)}(\cdot)$ denotes the j'th derivative of φ.

Finally, we have that

$$\boldsymbol{U}^\alpha = \frac{1}{\lambda^\alpha}\begin{pmatrix} 1 & \alpha & \frac{1}{2}\alpha(\alpha+1) & \cdots & \frac{1}{(n-1)!}[\alpha(\alpha+1)\cdots(\alpha+n-2)] \\[2mm] 0 & 1 & \alpha & \cdots & \frac{1}{(n-2)!}[\alpha(\alpha+1)\cdots(\alpha+n-3)] \\[2mm] 0 & 0 & 1 & \cdots & \frac{1}{(n-3)!}[\alpha(\alpha+1)\cdots(\alpha+n-4)] \\[2mm] \vdots & \vdots & \vdots & \ddots & \vdots \\[2mm] 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Therefore

$$\begin{aligned} M_X(\alpha) &= \Gamma(\alpha)\lambda^{-(\alpha-1)}\frac{\alpha(\alpha+1)\cdots(\alpha+n-2)}{(n-1)!} \\ &= \lambda^{-\alpha+1}\frac{\Gamma(\alpha+n-1)}{\Gamma(n)}. \end{aligned}$$

Next we consider the sine and the cosine transforms. Regarding the sine transform (given in Theorem 4.3), we need to calculate

$$(u^2\boldsymbol{I} + \boldsymbol{T})^{-1} = (u\boldsymbol{I} - i\boldsymbol{T})(u\boldsymbol{I} + i\boldsymbol{T})$$

where $i^2 = -1$. Since

$$u\boldsymbol{I} + i\boldsymbol{T} = \begin{pmatrix} u - \lambda i & \lambda i & 0 & \cdots & 0 \\ 0 & u - \lambda i & \lambda i & \cdots & 0 \\ 0 & 0 & u - \lambda i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u - \lambda i \end{pmatrix}$$

we have that

$$(u\mathbf{I} + i\mathbf{T})^{-1} = \begin{pmatrix} \frac{1}{u-\lambda i} & \frac{-\lambda i}{(u-\lambda i)^2} & \frac{(\lambda i)^2}{(u-\lambda i)^3} & \cdots & (-1)^{n-1}\frac{(\lambda i)^{n-1}}{(u-\lambda i)^n} \\ \\ 0 & \frac{1}{u-\lambda i} & \frac{-\lambda i}{(u-\lambda i)^2} & \cdots & (-1)^{n-2}\frac{(\lambda i)^{n-2}}{(u-\lambda i)^{n-1}} \\ \\ 0 & 0 & \frac{1}{u-\lambda i} & \cdots & (-1)^{n-3}\frac{(\lambda i)^{n-3}}{(u-\lambda i)^{n-2}} \\ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \\ 0 & 0 & 0 & \cdots & \frac{1}{u-\lambda i} \end{pmatrix}.$$

Using $(u\mathbf{I} - i\mathbf{T})^{-1} = -(-u\mathbf{I} + i\mathbf{T})^{-1}$ we get that

$$(u\mathbf{I} - i\mathbf{T})^{-1} = \begin{pmatrix} \frac{1}{u+\lambda i} & \frac{\lambda i}{(u+\lambda i)^2} & \frac{(\lambda i)^2}{(u+\lambda i)^3} & \cdots & \frac{(\lambda i)^{n-1}}{(u+\lambda i)^n} \\ \\ 0 & \frac{1}{u+\lambda i} & \frac{\lambda i}{(u+\lambda i)^2} & \cdots & \frac{(\lambda i)^{n-2}}{(u+\lambda i)^{n-1}} \\ \\ 0 & 0 & \frac{1}{u+\lambda i} & \cdots & \frac{(\lambda i)^{n-3}}{(u+\lambda i)^{n-2}} \\ \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \\ 0 & 0 & 0 & \cdots & \frac{1}{u+\lambda i} \end{pmatrix}.$$

In order to calculate the sine transform we only need the $(1,n)$'th element of the matrix $\mathbf{A} = (u\mathbf{I} + \mathbf{T}^2)^{-1}$, which amounts to

$$\begin{aligned} a_{1n} &= \sum_{j=1}^n (-1)^{j-1}\frac{(\lambda i)^{j-1}}{(u-\lambda i)^j}\frac{(\lambda i)^{n-j}}{(u+\lambda i)^{n-j+1}} \\ \\ &= (\lambda i)^{n-1}\sum_{j=1}^n (-1)^{j-1}\frac{1}{(u-\lambda i)^j(u+\lambda i)^{n-j+1}} \\ \\ &= \frac{(\lambda i)^{n-1}}{(u^2+\lambda^2)^n}\sum_{j=1}^n (u-\lambda i)^{n-j}(-u-\lambda i)^{j-1} \\ \\ &= \frac{(\lambda i)^{n-1}}{(u^2+\lambda^2)^n}\frac{(u-\lambda i)^n - (-u-\lambda i)^n}{2u}. \end{aligned}$$

By the Binomial theorem, we then get that

$$\hat{f}_s(u) \;=\; u\boldsymbol{\pi}(u^2\boldsymbol{I}+\boldsymbol{T}^2)^{-1}\mathbf{t} = u \cdot a_{1n}\lambda$$

$$= \frac{\displaystyle\sum_{j=0}^{\left[\frac{n+1}{2}\right]-1} \binom{n}{2j+1}(-1)^j u^{2j+1}\lambda^{2n-1}}{(u^2+\lambda^2)^n}.$$

The cosine transform is obtained in a similar way noticing that

$$\hat{f}_c(u) = \boldsymbol{\pi}(u^2\boldsymbol{I}+\boldsymbol{T}^2)^{-1}\boldsymbol{T}^2\mathbf{e}$$

and that

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

so only the elements a_{1n} and $a_{1,n-1}$ are used in the formula. It is then not difficult to see that

$$\hat{f}_c(u) = \frac{\displaystyle\sum_{j=0}^{\left[\frac{n+1}{2}\right]-1} \binom{n}{2j+1}(-1)^j u^{2j}\lambda^{2n}}{(u^2+\lambda^2)^n} - \frac{\displaystyle\sum_{j=0}^{\left[\frac{n}{2}\right]-1} \binom{n-1}{2j+1}(-1)^j u^{2j}\lambda^{2(n-1)}}{(u^2+\lambda^2)^{n-1}}.$$

$\square$

The Mellin transform is often used to derive distributions for products of random variables. Indeed, if X and Y are independent random variables with Mellin transforms $M_X(s)$ and $M_Y(s)$ respectively, then the Mellin transform of XY is given by

$$M_{XY}(\alpha) = M_X(\alpha)M_Y(\alpha).$$

Theorem 4.8. *If $X \sim PH(\boldsymbol{\pi}_X,\boldsymbol{T}_X)$ and $Y \sim PH(\boldsymbol{\pi}_Y,\boldsymbol{T}_Y)$ are independent, then*

$$M_{XY}(\alpha) = \Gamma(\alpha)^2\,(\boldsymbol{\pi}_X\otimes\boldsymbol{\pi}_Y)\,(\boldsymbol{T}_X\otimes\boldsymbol{T}_Y)^{-\alpha}\,(\mathbf{t}_X\otimes\mathbf{t}_Y)\,,$$

where as usual $\mathbf{t}_X = -\boldsymbol{T}_X\mathbf{e}$ and $\mathbf{t}_Y = -\boldsymbol{T}_Y\mathbf{e}$.

Proof. By independence of the variables X and Y and the relation $(\boldsymbol{U}\boldsymbol{V})\otimes(\boldsymbol{M}\boldsymbol{N}) = (\boldsymbol{U}\otimes\boldsymbol{M})(\boldsymbol{V}\otimes\boldsymbol{N})$ we obtain

$$M_{XY}(\alpha) = \Gamma(\alpha)^2\,(\boldsymbol{\pi}_X\otimes\boldsymbol{\pi}_Y)\left((-\boldsymbol{T}_X)^{-\alpha}\otimes(-\boldsymbol{T}_Y)^{-\alpha}\right)(\mathbf{t}_X\otimes\mathbf{t}_Y)\,,$$

so it is sufficient to prove that

$$(-\boldsymbol{T}_X)^{-\alpha}\otimes(-\boldsymbol{T}_Y)^{-\alpha} = (\boldsymbol{T}_X\otimes\boldsymbol{T}_Y)^{-\alpha}\,,$$

or equivalently,

$$\boldsymbol{U}_X^\alpha \otimes \boldsymbol{U}_Y^\alpha = (\boldsymbol{U}_X \otimes \boldsymbol{U}_Y)^\alpha,$$

where $\boldsymbol{U}_X = -\boldsymbol{T}_X^{-1}$ and $\boldsymbol{U}_Y = -\boldsymbol{T}_Y^{-1}$. Since the real parts of the eigenvalues of a sub-intensity matrix $\boldsymbol{T}$ are strictly negative, it is clear that the eigenvalues of $\boldsymbol{U} = -\boldsymbol{T}^{-1}$ have strictly positive real part. Hence the sum of the arguments of any two eigenvalues of $\boldsymbol{U}_X$ and $\boldsymbol{U}_Y$ must be in $(-\pi, \pi)$, (see Appendix). This ensures that all matrix logarithms and their sums that appear in the following are principal logarithms (see [11], p. 270).

Writing

$$\boldsymbol{U}_X^\alpha = \exp\left(\alpha \log\left(\boldsymbol{U}_X\right)\right), \quad \boldsymbol{U}_Y^\alpha = \exp\left(\alpha \log\left(\boldsymbol{U}_Y\right)\right),$$

where $\log(\cdot)$ is the principal logarithm (see Appendix), we have

$$\begin{aligned}
\boldsymbol{U}_X^\alpha \otimes \boldsymbol{U}_Y^\alpha &= \exp\left(\alpha \log\left(\boldsymbol{U}_X\right)\right) \otimes \exp\left(\alpha \log\left(\boldsymbol{U}_Y\right)\right) \\
&= \exp\left(\alpha \left(\log\left(\boldsymbol{U}_X\right) \oplus \log\left(\boldsymbol{U}_Y\right)\right)\right),
\end{aligned}$$

so we only need to prove that

$$\log\left(\boldsymbol{U}_X\right) \oplus \log\left(\boldsymbol{U}_Y\right) = \log\left(\boldsymbol{U}_X \otimes \boldsymbol{U}_Y\right).$$

Since

$$\log\left(\boldsymbol{U}_X\right) \otimes \boldsymbol{I} = \log\left(\boldsymbol{U}_X \otimes \boldsymbol{I}\right),$$

and

$$\boldsymbol{I} \otimes \log\left(\boldsymbol{U}_Y\right) = \log\left(\boldsymbol{I} \otimes \boldsymbol{U}_Y\right),$$

and since the matrices are commuting, it follows that

$$\begin{aligned}
\log\left(\boldsymbol{U}_X\right) \oplus \log\left(\boldsymbol{U}_Y\right) &= \log\left(\boldsymbol{U}_X\right) \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \log\left(\boldsymbol{U}_Y\right) \\
&= \log\left(\boldsymbol{U}_X \otimes \boldsymbol{I}\right) + \log\left(\boldsymbol{I} \otimes \boldsymbol{U}_Y\right) \\
&= \log\left(\left(\boldsymbol{U}_X \otimes \boldsymbol{I}\right)\left(\boldsymbol{I} \otimes \boldsymbol{U}_Y\right)\right) \\
&= \log\left(\boldsymbol{U}_X \otimes \boldsymbol{U}_Y\right),
\end{aligned}$$

where the next to the last step is obtained from Theorem 6.6. $\square$

We now derive an expression for the representation of the n'th order moment distribution of a phase–type distribution using the method of functional calculus. The representation was originally derived in [7]. The proof we present here is constructive the original proof being an analytical verification ot the claim. The following argument shows the potential of the functional calculus.

Let $X \sim \mathrm{PH}_p(\boldsymbol{\pi}, \boldsymbol{T})$. Consider the density

$$f_n(x) = \frac{x^n \boldsymbol{\pi} e^{\boldsymbol{T}x} \mathbf{t}}{\mu_n},$$

of the n'th order moment distribution of X with $\mu_n = \mathbb{E}(X^n) = n!\boldsymbol{\pi}(-\boldsymbol{T})^{-n}\mathbf{e}$ being the n'th moment. Let

$$\mathrm{er}_k(x;\lambda) = \frac{\lambda^k}{(k-1)!}x^{k-1}e^{-\lambda x}$$

denote the density of the k'th order Erlang distribution, i.e. the density of the distribution which is the convolution of k independent exponential distributions. Then we can write $\mathrm{er}_k(x;s)$ as

$$\mathrm{er}_k(x;s) = (1,0,\cdots,0)\exp\left(\left(\begin{array}{ccccc} -s & s & 0 & \cdots & 0 \\ 0 & -s & s & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & -s \end{array}\right)x\right)\left(\begin{array}{c} 0 \\ 0 \\ \vdots \\ s \end{array}\right),$$

where the dimensions of the vectors and matrix involved are k. Since $z \to \mathrm{er}_k(x;z)$ is an analytic function in positive half–plane, we then get

$$
\begin{aligned}
f_n(x) &= \frac{x^n\boldsymbol{\pi}e^{\boldsymbol{T}x}\mathbf{t}}{\mu_n} \\
&= \frac{\boldsymbol{\pi}}{\mu_n}x^n e^{\boldsymbol{T}x}\mathbf{t} \\
&= \frac{\boldsymbol{\pi}}{\mu_n}(-\boldsymbol{T})^{-n-1}n!\mathrm{er}_{n+1}(x;-\boldsymbol{T})\mathbf{t} \\
&= \frac{\boldsymbol{\pi}}{\mu_n}(-\boldsymbol{T})^{-n-1}(\boldsymbol{I},\boldsymbol{0},...,\boldsymbol{0})\times \\
&\qquad \exp\left(\left(\begin{array}{cccc} \boldsymbol{T} & -\boldsymbol{T} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{T} \end{array}\right)x\right)\left(\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ -\boldsymbol{T} \end{array}\right)\boldsymbol{t} \\
&= \left(\frac{\boldsymbol{\pi}\boldsymbol{T}^{-n}}{\boldsymbol{\pi}\boldsymbol{T}^{-n}\mathbf{e}},0,\cdots,0\right)\exp\left(\left(\begin{array}{cccc} \boldsymbol{T} & -\boldsymbol{T} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{T} \end{array}\right)x\right)\left(\begin{array}{c} \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \mathbf{t} \end{array}\right),
\end{aligned}
$$

where the vectors and matrix involved are $(n+1)p$–dimensional ($n+1$ blocks of size p each). The representation is a matrix–exponential one, which only in special cases is also a phase–type representation (in terms of a sub–intensity matrix). For phase–type representations in terms of a sub–intensity matrix, see [7].

5. Some extensions

By proving Theorem 4.2, we did not make use of the underlying probabilistic structure of the phase–type distribution, but only the fact that the density can be written as $\boldsymbol{\pi}\exp(\boldsymbol{T}x)\mathbf{t}$ and that the eigenvalues of $\boldsymbol{T}$ have strictly negative real parts. Hence Theorem 4.2 can be proved for any distribution which have a density on the form

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

for some row vector $\boldsymbol{\alpha}$, column vector $\mathbf{s}$ and matrix $\boldsymbol{S}$ such that all eigenvalues for $\boldsymbol{S}$ have strictly negative real parts. Such distributions are referred to as matrix–exponential distributions. The class of matrix–exponential distributions contains the class of phase–type distributions but is strictly larger.

A non–negative random variable X has a matrix–exponential distribution if and only if it has a rational Laplace transform on the form

$$L(s) = \frac{p_1 s^{n-1} + p_2 s^{n-2} + \cdots + p_n}{s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_n}.$$

The Companion matrix for the monic polynomial (i.e. one where the leading coefficient is 1) $q(s) = s^n + q_1 s^{n-1} + q_2 s^{n-2} + \cdots + q_n$ is defined by

$$\boldsymbol{C}(q) = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ -q_n & -q_{n-1} & -q_{n-2} & -q_{n-3} & \cdots & -q_1 \end{pmatrix}$$

and it satisfies that $p(s) = \det(s\boldsymbol{I}-\boldsymbol{C}(q))$, where $\boldsymbol{I}$ denotes the identity matrix. In [2] it was proved that X has a density on the form

(1) $$f_X(x) = \mathbf{p}e^{C(q)x}\mathbf{e}_n,$$

where $\mathbf{p} = (p_n, p_{n-1}, ..., p_1)$ and $\mathbf{e}_n = (0, 0, ..., 0, 1)'$. Here we give a short proof which makes use of functional calculus in the conclusion. First we prove that

$$L_X(s) = \mathbf{p}(s\boldsymbol{I} - \boldsymbol{C}(q))^{-1}\mathbf{e}_n.$$

By Cramérs rule,

$$(s\boldsymbol{I} - \boldsymbol{C}(q))^{-1} = \frac{\text{adj}((s\boldsymbol{I} - \boldsymbol{C}(q))}{\det(s\boldsymbol{I} - \boldsymbol{C}(q))},$$

where adj denote the adjoint of the matrix, which is the transpose of the co–factor matrix $\boldsymbol{K} = \{k_{ij}\}$ defined by

$$k_{ij} = (-1)^{i+j}\det((s\boldsymbol{I} - \boldsymbol{C}(q))_{-i,-j}),$$

and where $(s\boldsymbol{I} - \boldsymbol{C}(q))_{-i,-j}$ denotes the matrix that results by removing row i and column j from the corresponding matrix. We now calculate the n'th column in the adjoint matrix, which is the n'th row of the co–factor matrix. This co–factor matrix has the form

$$\begin{pmatrix} s & -1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & s & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & s & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

Its determinant is $s^{i-1}(-1)^{n-i} = s^{i-1}(-1)^{n+i}$ so the (i,n)'th element of the adjoint is given by

$$\mathrm{adj}(s\boldsymbol{I} - \boldsymbol{C}(q))_{i,n} = (-1)^{n+i}s^{i-1}(-1)^{n+i} = s^{i-1}.$$

Then the n'th column of $\mathrm{adj}(s\boldsymbol{I} - \boldsymbol{C}(q))$ is $(1, s, ..., s^{n-1})'$. Hence, by the form of $\mathbf{p}$ and $\mathbf{e}_n$, we conclude that

$$p(s) = \mathbf{p} \ \mathrm{adj}(s\boldsymbol{I} - \boldsymbol{C}(q)) \ \mathbf{e}_n$$

so that

$$L_X(s) = \frac{p(s)}{q(s)} = \frac{\mathbf{p} \ \mathrm{adj}(s\boldsymbol{I} - \boldsymbol{C}(q)) \ \mathbf{e}_n}{\det(s\boldsymbol{I} - \boldsymbol{C}(q))} = \mathbf{p}(s\boldsymbol{I} - \boldsymbol{C}(q))^{-1}\mathbf{e}_n.$$

All singularities for $(s\boldsymbol{I} - \boldsymbol{C}(q))^{-1}$ (the eigenvalues of $\boldsymbol{C}(q)$) are poles for $L_X(s)$, so the inverse Laplace transform may be calculated as a contour integral enclosing the poles. Let γ be closed simple contour enclosing the poles. Then the density for X is given by the inverse Laplace transform

$$f_X(x) = \frac{1}{2\pi i} \oint_{\gamma} e^{zx} L_X(z)dz = \mathbf{p}\left(\frac{1}{2\pi i} \oint_{\gamma} e^{zx}(z\boldsymbol{I} - \boldsymbol{C}(q))^{-1}dz\right)\mathbf{e}_n,$$

and by Definition 2.1, the integral is simply $\exp(\boldsymbol{C}(q)x)$ so

$$f_X(x) = \mathbf{p}e^{\boldsymbol{C}(q)x}\mathbf{e}_n$$

as was to be proved. This representation is referred to as the canonical form a matrix–exponential distribution.

Many results which are valid for phase–type distributions also hold true for matrix–exponential distributions (see e.g. [2] and [6]). Certainly all results proved in Section 4 are valid also for matrix–exponential distribution, and with unaltered proofs.

Example 5.1 (Mellin, sine and cosine transforms of a Matrix–exponential distribution)**.** Consider the distribution with density

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

This is a matrix–exponential distribution which is not of phase–type since the density fails to be strictly positive everywhere (see [15] for a characterisation of phase–type distributions). Its Laplace transform is given by

$$L(s) = \frac{2}{s^3 + 3s^2 + 4s + 2},$$

and by (1) we may write f on the form $f(x) = \boldsymbol{\pi} e^{\boldsymbol{C}x} \mathbf{e}_3$ with

$$\boldsymbol{\pi} = \begin{pmatrix} 2 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & -4 & -3 \end{pmatrix} \quad \text{and} \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then

$$\boldsymbol{U} := -\boldsymbol{C}^{-1} = \begin{pmatrix} 2 & 3/2 & 1/2 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

has eigenvalues $1, \frac{1}{2} - \frac{1}{2}i$ and $\frac{1}{2} + \frac{1}{2}i$, and the resolvent $R(z) = (z\boldsymbol{I} - \boldsymbol{U})^{-1}$ is given by

$$R(z) = \frac{1}{2z^3 - 4z^2 + 3z - 1} \begin{pmatrix} 2z^2 & 3z - 1 & z \\ -2z & 2z^2 - 2z & -1 \\ 2 & 4 - 2z & 2z^2 - 4z + 3 \end{pmatrix}.$$

Then for any suitable path γ enclosing the eigenvalues and such that x^α is analytic on and inside γ (similar to Figure 1),

$$\boldsymbol{U}^\alpha = \frac{1}{2\pi i} \oint_\gamma z^\alpha R(z) dz.$$

Since the Mellin transform is given by $\Gamma(\alpha)\boldsymbol{\pi}\boldsymbol{U}^\alpha \mathbf{e}_3$, we see that only the entrance $(3,3)$ of the matrix $\boldsymbol{U}^\alpha$ matters. We have that

$$\boldsymbol{U}_{3,3}^\alpha = \frac{1}{2\pi i} \oint_\gamma z^\alpha \frac{z}{2z^3 - 4z^2 + 3z - 1} dz,$$

and by means of the residue theorem we finally obtain that

$$M(\alpha) = 5 \left(1 - 2^{-\alpha/2} \cos\left(\frac{\pi}{4}\alpha\right) \right) \Gamma(\alpha).$$

The sine and cosine transforms for the matrix–exponential distribution in this case are given by

$$
\begin{aligned}
\hat{f}_s(u) &= \frac{2u(4 - u^2)}{u^6 + u^4 + 4u^2 + 4}, \\
\hat{f}_c(u) &= \frac{2(2 - 3u^2)}{u^6 + u^4 + 4u^2 + 4}.
\end{aligned}
$$

Example 5.2 (Multivariate phase–type distributions)**.** Consider a Markovian Arrival Process (MAP) with generator $\boldsymbol{Q} = \boldsymbol{C} + \boldsymbol{D}$ and initial distribution $\boldsymbol{\pi}$, where $\boldsymbol{C}$ denotes sub–intensity matrix of jump rates which are not accompanied by arrivals while $\boldsymbol{D}$ contains the jump rates accompanied by arrivals (see [12, 13]). The MAPs are dense in the class of point processes on the real axis ([4]) and the joint density of the first n inter–arrival times $X_1, ..., X_n$ is given by

$$
f(x_1, ..., x_n) = \boldsymbol{\pi} e^{\boldsymbol{C}x} \boldsymbol{D} e^{\boldsymbol{C}x_2} \boldsymbol{D} \cdots \boldsymbol{D} e^{\boldsymbol{C}x_n} \boldsymbol{D}\mathbf{e}.
$$

The joint fractional moments may now be calculated as follows. For simplicity we let $n = 2$. Then

$$
\begin{aligned}
\mathbb{E}\left(X_1^{\alpha_1} X_2^{\alpha_2}\right) &= \int_0^\infty \int_0^\infty x^{\alpha_1} y^{\alpha_2} f(x, y) dx dy \\
&= \boldsymbol{\pi} \int_0^\infty \left(\int_0^\infty x^{\alpha_1} e^{\boldsymbol{C}x} dx \right) \boldsymbol{D} y^{\alpha_2} e^{\boldsymbol{C}y} dy \boldsymbol{D}\mathbf{e} \\
&= \boldsymbol{\pi} \left(\int_0^\infty x^{\alpha_1} e^{\boldsymbol{C}x} dx \right) \boldsymbol{D} \left(\int_0^\infty y^{\alpha_2} e^{\boldsymbol{C}y} dy \right) \boldsymbol{D}\mathbf{e}.
\end{aligned}
$$

Since all eigenvalues of $\boldsymbol{C}$ have strictly negative real parts (the arrivals are easily seen to be marginally phase–type distributed) we then get from Theorem 4.5 that

$$
\mathbb{E}\left(X_1^{\alpha_1} X_2^{\alpha_2}\right) = \boldsymbol{\pi}(-\boldsymbol{C})^{-\alpha_1 - 1} \boldsymbol{D}(-\boldsymbol{C})^{-\alpha_2 - 1} \boldsymbol{D}\mathbf{e}.
$$

This formula immediately generalises to n dimensions as

$$
\mathbb{E}\left(X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_n^{\alpha_n}\right) = \boldsymbol{\pi}(-\boldsymbol{C})^{-\alpha_1 - 1} \boldsymbol{D} \cdots \boldsymbol{D}(-\boldsymbol{C})^{-\alpha_n - 1} \boldsymbol{D}\mathbf{e}.
$$

A further generalisation to Rational Arrival processes (see [3]) is immediate, which provides a class of multivariate matrix–exponential distributions. For both classes of multivariate phase–type or matrix–exponential distributions, the multivariate Mellin transform is hence defined and given by

$$
\begin{aligned}
M(\alpha_1, ..., \alpha_n) &= \mathbb{E}(X_1^{\alpha_1 - 1} \cdots X_n^{\alpha_n - 1}) \\
&= \boldsymbol{\pi}(-\boldsymbol{C})^{-\alpha_1} \boldsymbol{D} \cdots \boldsymbol{D}(-\boldsymbol{C})^{-\alpha_n} \boldsymbol{D}\mathbf{e}.
\end{aligned}
$$

Example 5.3 (Infinite dimensional phase–type distributions). Let Z be a random variable with a distribution $\{\pi_i\}_{i\in\mathbb{N}}$ concentrated on $\mathbb{N}$ and let $X \sim \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{T})$ be independent of Z. Define $Y = ZX$. For $j \in \mathbb{N}$, $jX \sim \mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{T}/j)$ and the mean of Xj is of course j times the mean of X. This means that by realising Y we essentially obtain a tail which is dictated by $\{\pi_i\}$ and a main body of the distribution given by the phase–type distribution. For details on the asymptotic behavior we refer to [8]. It is clear that the density of Y can be written on the form

$$f_Y(y) = \sum_{i=1}^{\infty} \pi_i \boldsymbol{\alpha} e^{\boldsymbol{T}/iy} \frac{\mathbf{t}}{i}.$$

In [8] it was further shown that $f_Y(y)$ may be seen as an infinite dimensional phase–type distribution

$$f_Y(y) = (\boldsymbol{\pi} \otimes \boldsymbol{\alpha}) e^{\boldsymbol{S}y} \mathbf{s},$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots)$, $\boldsymbol{S}$ is the block diagonal matrix with the i'th block being $\boldsymbol{T}/i$ and $\mathbf{s} = -\boldsymbol{S}\mathbf{e}$, and where $\mathbf{e}$ is an infinite dimensional column vector of ones. The Mellin transform of $Y = ZX$ is given by

$$
\begin{aligned}
M_Y(\theta) &= M_Z(\theta) M_X(\theta) \\
&= \left(\sum_{i=1}^{\infty} \pi_i i^{\theta-1} \right) \int_0^{\infty} x^{\theta-1} \boldsymbol{\alpha} e^{\boldsymbol{T}x} \mathbf{t}\, dx \\
&= \left(\sum_{i=1}^{\infty} \pi_i i^{\theta-1} \right) \boldsymbol{\alpha} (-\boldsymbol{T})^{-\theta} \mathbf{t}.
\end{aligned}
$$

For example if $\{\pi_i\}$ is the Riemann–Zeta distribution (also called the discrete Pareto distribution) given by

$$\pi_i = \frac{i^{-\beta}}{\zeta(\beta)}$$

for some $\beta > 0$, then the Mellin transform of Y exists only for arguments $\theta < \beta$. This was of course to be expected due to the heavy tailed character of the Riemann–Zeta distribution. $\qquad\square$

References

[1] Søren Asmussen. *Applied Probability and Queues (Stochastic Modelling and Applied Probability)*. Springer, 2003.

[2] Søren Asmussen and Mogens Bladt. Renewal theory and queueing algorithms for matrix-exponential distributions. In *Matrix-analytic methods in stochastic models (Flint, MI)*, pages 313–341. Dekker, New York, 1997.

[3] Søren Asmussen and Mogens Bladt. Point processes with finite-dimensional probabilities. *Stochastic Processes and their Applications*, 82(1):127–142, 1999.

[4] Søren Asmussen and Ger Koole. Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, 30:365–372, 1993.

[5] Mogens Bladt. A review on phase-type distributions and their use in risk theory. *Astin Bulletin*, 35(1):145–161, 2005.

[6] Mogens Bladt and Marcel F Neuts. Matrix Exponential Distributions: Calculus and Interpretations via Flows. *Stochastic Models*, 19(1):113–124, 2003.

[7] Mogens Bladt and Bo Friis Nielsen. Moment Distributions of Phase Type. *Stochastic Models*, 27(4):651–663, 2011.

[8] Mogens Bladt, Bo Friis Nielsen, and Gennady Samorodnitsky. Calculation of ruin probabilities for a dense class of heavy tailed distributions. *Scandinavian Actuarial Journal*, pages 1–19, 2014.

[9] Alexander Graham. *Kronecker Products and Matrix Calculus: With Applications.* John Wiley & Sons, 1982.

[10] Markus Haase. *The Functional Calculus for Sectorial Operators.* Springer, 2006.

[11] Nicholas J Higham. *Functions of matrices.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2008.

[12] David M Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Communications in Statistics Stochastic Models*, 7(1):1–46, 1991.

[13] Marcel F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.

[14] Marcel F. Neuts. *Matrix-Geometric Solutions in Stochastic Models an Algorithmic Approach*, volume 2 of *John Hopkins Series in Mathematical Sciences.* The John Hopkins University Press, 1981.

[15] Colm Art O'cinneide. Characterization of phase-type distributions. *Stochastic Models*, 6(1):1–57, 1990.

[16] C R Rao. On Discrete Distributions Arising out of Methods of Ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, 27:311–324, 1965.

[17] Huamin Zhang and Feng Ding. On the Kronecker Products and Their Applications. *Journal of Applied Mathematics*, 2013(7):1–8, 2013.

6. APPENDIX

Definition 6.1. *Let $\mathbb{F}$ be a field, such as $\mathbb{R}$ or $\mathbb{C}$. For any matrix $\boldsymbol{A} = (a_{ij}) \in \mathbb{F}^{n \times m}$ and $\boldsymbol{B} \in \mathbb{F}^{p \times q}$, their Kronecker product, denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$, is defined by*

$$
\begin{aligned}
\boldsymbol{A} \otimes \boldsymbol{B} &= (a_{ij}\boldsymbol{B}) \\
&= \begin{pmatrix}
a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\
a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\
\vdots & \vdots & & \vdots \\
a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B}
\end{pmatrix} \in \mathbb{F}^{(mp) \times (nq)}.
\end{aligned}
$$

Some basic properties of Kronecker product are as follows

$$
\boldsymbol{I}_{m \times m} \otimes \boldsymbol{A} = \mathrm{diag}\,(\boldsymbol{A}, \boldsymbol{A}, \ldots, \boldsymbol{A})\,.
$$

If $\boldsymbol{\pi} = (a_1, a_2, \ldots, a_m)$ and $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, then

$$\boldsymbol{\pi}^T \mathbf{t} = \boldsymbol{\pi} \otimes \mathbf{t}.$$

The next two theorems establish important commutative properties of the Kronecker product.

Theorem 6.2. *Let $\boldsymbol{A} \in \mathbb{F}^{m \times n}$ and $\mathbb{F}^{p \times q}$ then*

$$\boldsymbol{A} \otimes \boldsymbol{B} = (\boldsymbol{A} \otimes \boldsymbol{I}_{p \times p})(\boldsymbol{I}_{m \times m} \otimes \boldsymbol{B}) = (\boldsymbol{I}_{m \times m} \otimes \boldsymbol{B})(\boldsymbol{A} \otimes \boldsymbol{I}_{q \times q}).$$

This means that $\boldsymbol{I}_{m \times m} \otimes \boldsymbol{B}$ and $\boldsymbol{A} \otimes \boldsymbol{I}_{n \times n}$ are commutative for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

Theorem 6.3. *Let $\boldsymbol{A} \in \mathbb{F}^{m \times m}$. If $f(z)$ is analytic function and $f(\boldsymbol{A})$ exists, then*

$$\begin{aligned} f(\boldsymbol{I}_{n \times n} \otimes \boldsymbol{A}) &= \boldsymbol{I}_{n \times n} \otimes f(\boldsymbol{A}), \\ f(\boldsymbol{A} \otimes \boldsymbol{I}_{n \times n}) &= f(\boldsymbol{A}) \otimes \boldsymbol{I}_{n \times n}. \end{aligned}$$

Definition 6.4. *The Kronecker sum of $\boldsymbol{A} \in \mathbb{F}^{m \times m}$ and $\boldsymbol{B} \in \mathbb{F}^{n \times n}$, denoted as $\boldsymbol{A} \oplus \boldsymbol{B}$, is defined by*

$$\boldsymbol{A} \oplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I}_{n \times n} + \boldsymbol{I}_{m \times m} \otimes \boldsymbol{B}.$$

Theorem 6.5. *Let $\boldsymbol{A} \in \mathbb{F}^{m \times m}$, and $\boldsymbol{B} \in \mathbb{F}^{n \times n}$. Then*

$$\exp(\boldsymbol{A} \oplus \boldsymbol{B}) = \exp(\boldsymbol{A}) \otimes \exp(\boldsymbol{B}).$$

For any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ which commute it is known that for eigenvalue λ for $\boldsymbol{A}$ there is an eigenvalue μ for $\boldsymbol{B}$ such that $\lambda + \mu$ is an eigenvalue for $\boldsymbol{A} + \boldsymbol{B}$ (see [11]). The eigenvalue μ is called the eigenvalue corresponding to λ.

The principal logarithm of a matrix $\boldsymbol{A}$, $\boldsymbol{X} = \log(\boldsymbol{A})$, is defined as the unique matrix $\boldsymbol{X}$ for which $\exp(\boldsymbol{X}) = \boldsymbol{A}$ and $-\pi < \operatorname{Im}(\lambda(\boldsymbol{X})) < \pi$.

Theorem 6.6. *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n \times n}$ commute and have no eigenvalues in $\mathbb{R}^-$. If the eigenvalues λ_i of $\boldsymbol{A}$ and the corresponding eigenvalue μ_i of $\boldsymbol{B}$ satisfies that*

$$|\arg \lambda_i + \arg \mu_i| < \pi,$$

then

$$\log(\boldsymbol{AB}) = \log(\boldsymbol{A}) + \log(\boldsymbol{B}).$$

Proof. Since $\boldsymbol{A}$ and $\boldsymbol{B}$ commute then $\log(\boldsymbol{A})$ and $\log(\boldsymbol{B})$ are also commutative, and

$$\exp(\log(\boldsymbol{A}) + \log(\boldsymbol{B})) = \exp(\log \boldsymbol{A}) \exp(\log \boldsymbol{B}) = \boldsymbol{AB},$$

so $\log(\boldsymbol{A}) + \log(\boldsymbol{B})$ is some logarithm of $\boldsymbol{AB}$. Since

$$\operatorname{Im}(\log(\lambda_j) + \log(\mu_j)) = \arg \lambda_i + \arg \mu_i \in (-\pi, \pi),$$

then $\log(\boldsymbol{A}) + \log(\boldsymbol{B})$ is the principal logarithm of $\boldsymbol{AB}$.   □

For further details on Kronecker products and sums we refer to [9] or [17].