Multivariate matrix Mittag–Leffler distributions

Hansjörg Albrecher[1] · Martin Bladt[2] · Mogens Bladt[3]

Abstract
We extend the construction principle of multivariate phase-type distributions to establish an analytically tractable class of heavy-tailed multivariate random variables whose marginal distributions are of Mittag–Leffler type with arbitrary index of regular variation. The construction can essentially be seen as allowing a scalar parameter to become matrix-valued. The class of distributions is shown to be dense among all multivariate positive random variables and hence provides a versatile candidate for the modelling of heavy-tailed, but tail-independent, risks in various fields of application.

Keywords Multivariate distribution · Heavy tails · Markov process · Mittag–Leffler distribution · Phase-type · Matrix distribution · Extremes · Laplace transforms

1 Introduction

The joint modelling of dependent risks is a crucial task in many areas of applied probability and quantitative risk management, see, for example, McNeil et al. (2015). While in many situations, there is a reasonable amount of data available for the fitting procedure of univariate risks, the identification of multivariate models is much more delicate. A frequent approach proposed in applications is to use the

✉ Hansjörg Albrecher
hansjoerg.albrecher@unil.ch

Martin Bladt
martin.bladt@unil.ch

Mogens Bladt
bladt@math.ku.dk

1 Department of Actuarial Science, Faculty of Business and Economics and Swiss Finance Institute, University of Lausanne, 1015 Lausanne, Switzerland

2 Department of Actuarial Science, Faculty of Business and Economics, University of Lausanne, 1015 Lausanne, Switzerland

3 Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark

available data for univariate fitting and choose a parametric copula to combine the margins, where the parameters of that copula are then either assumed a priori or estimated from the joint data. The choice of such a copula is of course crucial for the resulting joint distribution and the conclusions one draws from it, cf. Mai and Scherer (2017) and Mikosch (2006). In multivariate extremes, which is currently a very active research topic, one typically uses less restrictive assumptions for the quantification of joint exceedances, see, for example, Falk et al. (2019) and Kiriliouk et al. (2019). Some specific families, like multivariate regular variation, are considered particularly attractive in this context, as they have a natural interpretation in terms of how to extend univariate behaviour into higher dimensions (Ho and Dombry 2019; Joe and Li 2011; Resnick 2002). These results focus, however, on the asymptotic behaviour, so that for a concrete application with an available data set, one typically has to choose thresholds above which this respective behaviour is assumed (Wan and Davis 2019), and the bulk of the distribution is then to be modelled by a different distribution (see, for example, Beirlant et al. 2004; Albrecher et al. 2017, Ch.IV.5).

In this paper, we would like to establish a family of multivariate distributions that can be applied for modelling across the entire positive orthant, so that no threshold selection is needed. In particular, we are interested in a family that leads to explicit and tractable expressions for the model fitting and interpretation. While such a family already exists for marginally light (exponentially bounded) tails in the form of multivariate phase-type (MVPH) distributions, our goal here is to develop a related family with heavy-tailed marginal distributions. The univariate starting point for this procedure is the matrix Mittag–Leffler (MML) distribution, which is a heavy-tailed distribution that was recently studied in Albrecher et al. (2019), and which proved to be very tractable, with excellent fitting properties. While in principle, there are many possible ways of defining a vector of random variables with given marginals, we want to consider here the natural concept of multivariate families that can be characterized by the property that any linear combination of the components of such a vector is again of the same marginal type. This is exactly one possible definition of MVPH distributions (so any linear combination of the coordinates of a random vector is again (univariate) phase-type), and it is also a characterizing property of multivariate regular variation in a random vector, namely that any linear combination of the coordinates of such a vector is again (univariate) regularly varying, see Basrak et al. (2008).

The goal is hence to study the class of multivariate random vectors for which such a property applies with MML marginal distributions. It will turn out that for this approach to work, we first need to consider a slightly more general class, which we will refer to as generalized MML distributions. We will show that the analysis developed for the MVPH case can then be extended to our more general situation. In particular, we will establish some properties of this class and work out explicit expressions for a number of concrete cases. The analysis is considerably simpler for the symmetric situation where all marginal distributions share the same index of regular variation, but the general case can be handled as well. The resulting multivariate MML distribution is asymptotically independent, i.e. there is tail independence for each bivariate pair of components. In the case of multivariate regular variation, the sub-class of random

vectors with asymptotic independence was studied and characterized in terms of second-order conditions in Resnick (2002), where also concrete application areas for such heavy-tailed but asymptotically independent risks are given. In a sense, the multivariate MML family of distributions we introduce here is another candidate for models in this domain, with the advantage of being explicit and tractable across the entire range $\mathbb{R}_+^n$. In that respect, this family is also an interesting alternative to multivariate Linnik distributions (see, for example, Anderson 2019; Lim and Teo 2010), which can be conveniently defined in terms of their characteristic function, have the range $\mathbb{R}^n$ (rather than $\mathbb{R}_+^n$) and also have heavy-tailed marginals, but which do not lead to explicit expressions for the multivariate density.

The remainder of the paper is organized as follows. Section 2 recapitulates the construction principle of univariate and multivariate PH distributions and provides the available background on MML distributions. Section 3 introduces generalized MML distributions. In Sect. 4, we then develop the necessary theoretical background for our definition of the multivariate MML family and establish some of its properties. We also consider power transforms, which will provide useful flexibility for modelling applications, and we derive denseness properties of the resulting multivariate family. In Sect. 5, we work out a concrete simple example in detail and illustrate resulting dependence properties for this case. Section 6 concludes.

2 Phase-type distributions

2.1 Notation

We shall apply a common convention from phase-type theory that matrices are expressed in bold capital letters (e.g. $\boldsymbol{T}, \boldsymbol{\Lambda}$), and row vectors are bold minuscular Greek letters (e.g. $\boldsymbol{\pi}, \boldsymbol{\alpha}$), while column vectors are bold minuscular roman letters (e.g. $\boldsymbol{t}$, $\boldsymbol{x}$). Elements of matrices and vectors are denoted by their corresponding minuscular unbold letters with indices, e.g. $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{a} = (a_i)$. If $\boldsymbol{a} = (a_1, \dots, a_n)$ is a vector, then by $\boldsymbol{\Delta}(\boldsymbol{a})$, we shall denote the diagonal matrix with $\boldsymbol{a}$ as diagonal.

2.2 Univariate phase-type distributions

Phase-type distributions are defined as the distribution of the time until absorption of a finite state-space Markov jump process with one absorbing state and the other states being transient.

Let p be a positive integer and $\{X_t\}_{t \geq 0}$ denote a Markov jump process on $E = \{1, \dots, p, p+1\}$, where states $1, 2, \dots, p$ are transient and state $p+1$ is absorbing. Let $\pi_i = \mathbb{P}(X_0 = i)$ and assume that $\pi_1 + \cdots \pi_p = 1$, i.e. initiation in the absorbing state is not possible. The intensity matrix of $\{X_t\}_{t \geq 0}$ can be written as

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \boldsymbol{0} & 0 \end{pmatrix}, \tag{1}$$

where T is the $p \times p$ *sub-intensity* matrix whose off-diagonal elements consist of transition rates between the transient states, t is a p-dimensional column vector, and $\mathbf{0}$ is a p-dimensional row vector. The diagonal elements of T are given by $t_{ii} = -\sum_{j \neq i} t_{ij} + t_i$, since the row sums of Λ must be zero.

Let e denote the vector of ones and $\pi = (\pi_1, \ldots, \pi_p)$. Dimensions are usually suppressed, and e may then have any adequate dimension depending on the context.

Then, the time until absorption,

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

is said to have a phase-type (PH) distribution with representation (π, T) and we write $\mathrm{PH}_p(\pi, T)$. Since rows of Λ sum to zero, we get $t = -Te$. Note that the case $p = 1$ leads to an exponential distribution.

If $\tau \sim \mathrm{PH}_p(\pi, T)$, then a number of relevant formulas can be written compactly in matrix notation, like, e.g.

$$f(x; \pi, T) = \pi e^{Tx} t, \quad x > 0,$$
$$F(x; \pi, T) = 1 - \pi e^{Tx} t, \quad x > 0,$$
$$L(s; \pi, T) = \pi (sI - T)^{-1} t, \quad s > \mathrm{Re}\,(\eta_{\max}),$$
$$\mathbb{E}(\tau^\alpha) = \Gamma(\alpha + 1) \pi (-T)^{-\alpha} e, \quad \alpha > 0,$$

for the density, c.d.f., Laplace transform and (fractional) moments, respectively. Here, $\eta_{\max}$ denotes the eigenvalue with maximum real part of T, and this real part is strictly negative. In particular, the Laplace transform is well defined for all $s \geq 0$ and in a neighbourhood around zero.

Remark 1 Representations (π, T) of phase-type distributions are not unique. In fact, one can construct an infinite number of different representations, which may even be of different orders p. Hence, phase-type representations may also suffer from over-parametrization, and it is not possible to attach a specific significance to individual elements of an intensity matrix. While one can typically construct a certain behaviour by means of structuring the sub-intensity matrix T, the opposite task of deducing such a behaviour from a given matrix is typically not possible. Some simple cases, however, may be described. For instance, $p = 1$ means one phase and the resulting distribution is exponential and hence unimodal. For $p = 2$, bimodality cannot be achieved either, as one could at most aim for a mixture of exponentials. For $p = 3$, it is possible to have a mixture of an exponential with an Erlang(2) which is bimodal.

For further details on phase-type expressions, we refer to Albrecher et al. (2019) and Bladt and Nielsen (2017).

2.3 Multivariate phase-type distributions

A nonnegative random vector $X = (X_1, \ldots, X_n)$ is phase-type distributed (MVPH) if all nonnegative, non-vanishing linear combinations of its coordinates X_i,

$i = 1, \dots, n$ have a (univariate) phase-type distribution. This is the most general definition of a multivariate phase-type distribution which, however, lacks practicality since it does not suggest how to construct such distributions. It contains a sub-class of multivariate distributions, MPH*, which have multidimensional Laplace transforms of the form

$$L_X(\boldsymbol{u}; \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R}) = E(\mathrm{e}^{-<u,X>}) = \boldsymbol{\pi}(\boldsymbol{\Delta}(\boldsymbol{Ru}) - \boldsymbol{T})^{-1}\boldsymbol{t}, \tag{2}$$

and we write that $X \sim \mathrm{MPH}^*(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$. Here, $(\boldsymbol{\pi}, \boldsymbol{T})$ is a phase-type representation of dimension p, say, $\boldsymbol{R}$ is a $p \times n$ matrix and $\boldsymbol{u} = (u_1, \dots, u_n) \in \mathbb{R}_+^n$. Furthermore, the joint Laplace transform exists in a neighbourhood around zero (Bladt and Nielsen 2017, Thm.8.1.2).

The form (2) is established from the following probabilistic construction (cf. Kulkarni 1989). Consider the Markov jump process $\{X_t\}_{t \geq 0}$ underlying the phase-type distribution with representation $(\boldsymbol{\pi}, \boldsymbol{T})$. The n columns of $\boldsymbol{R} = \{r_{ik}\}$ are p-dimensional vectors which contain nonnegative numbers. These numbers are "rewards" to be earned during sojourns in state i. If τ denotes the time until absorption of the underlying Markov jump process, then

$$X_k = \sum_{i=1}^{p} \int_0^\tau 1\{X_t = i\} r_{ik} \, \mathrm{d}t, \quad k = 1, \dots, n \tag{3}$$

is the total reward earned according to column k of $\boldsymbol{R}$ until absorption. The structure matrix $\boldsymbol{R}$ hence picks scaled sojourns out of the underlying Markov jump process. Correlation between different total rewards, X_i and X_j say, will then depend on the structure of $\boldsymbol{R}$ and on the underlying stochastic process. If there are common states in which reward is earned for both X_i and X_j, then this will contribute to a positive correlation between them. If there are no common states, the correlation will be entirely generated by the structure of the $\boldsymbol{T}$ matrix. Negative correlation between X_i and X_j is achieved if large rewards earned in one reduces the one earned in the other and vice versa. Specific constructions of dependencies between phase-type distributed random variables with given marginals are non-trivial, see, for example, Bladt and Nielsen (2010), for an example with exponentially distributed marginals.

The random variables X_k defined in (3) are again phase-type distributed and in general dependent since different variables may be generated through earning positive rewards on certain common states (while in other states, there may be zero reward for one variable whenever the other has positive reward). If all $r_{ik} > 0$, $i = 1, \dots, p$, then X_k is phase-type distributed with initial distribution $\boldsymbol{\pi}$ and sub-intensity matrix $\boldsymbol{\Delta}^{-1}(\boldsymbol{r}_{\cdot k})\boldsymbol{T}$. This follows easily from a sample path argument: if reward r_{ik} is earned during a sojourn in state i, then the distribution of the reward during a sojourn is exponentially distributed with intensity $-t_{ii}/r_{ik}$.

If some $r_{ik} = 0$, then finding a representation for X_k is more involved. Let $\boldsymbol{w} \geq \boldsymbol{0}$ denote a nonzero vector. For obtaining the k'th marginal distribution, we would choose $\boldsymbol{w} = \boldsymbol{e}_k'$, the k'th Euclidean unit vector, while for a more general projection, we may choose $\boldsymbol{w} = c_1 \boldsymbol{e}_1 + \cdots + c_n \boldsymbol{e}_k$ for some constants c_i, $i = 1, \dots, n$. For this given $\boldsymbol{w}$, decompose the set of transient states $E = \{1, \dots, p\}$ into $E = E_+ \cup E_0$,

where E_+ denotes states $i \in E$ for which $(Rw)_i > 0$ and E_0 states $i \in E$ for which $(Rw)_i = 0$. Decompose $\boldsymbol{\pi} = (\boldsymbol{\pi}_+, \boldsymbol{\pi}_0)$ and

$$T = \begin{pmatrix} T_{++} & T_{+0} \\ T_{0+} & T_{00} \end{pmatrix} \qquad (4)$$

accordingly. Then, we have the following theorem which is proved in Bladt and Nielsen (2017, p. 441).

Theorem 1 *The distribution of $\langle X, w \rangle$ is given by an atom at zero of size $q = \boldsymbol{\pi}_0 \left(I - \left(-T_{00} \right)^{-1} T_{0+} \right) e$ and an absolute continuous part given by a possibly defective phase-type distribution with representation $(\boldsymbol{\pi}_w, T_w)$, where*

$$\boldsymbol{\pi}_w = \boldsymbol{\pi}_+ + \boldsymbol{\pi}_0 \left(-T_{00} \right)^{-1} T_{0+} \text{ and } T_w = \boldsymbol{\Delta} \left((Rw)_+ \right)^{-1} \left(T_{++} + T_{+0} \left(-T_{00} \right)^{-1} T_{0+} \right)$$

This means that

$$\begin{aligned} \boldsymbol{\pi} (\boldsymbol{\Delta}(Ruw) - T)^{-1} t &= \mathbb{E} \left(e^{-\langle X, uw \rangle} \right) \\ &= \mathbb{E} \left(e^{-u\langle X, w \rangle} \right) \\ &= q + \boldsymbol{\pi}_w (uI - T_w)^{-1} t_w, \end{aligned} \qquad (5)$$

where $t_w = -T_w e$.

Remark 2 It is still an open question whether $\text{MPH}^* \subset \text{MVPH}$ or whether $\text{MPH}^* = \text{MVPH}$.

Remark 3 As for univariate phase-type distributions, representations $(\boldsymbol{\pi}, T, R)$ of MPH^* are not uniquely determined by their distributions, and they may be over-parametrized as well. In particular, the interplay between T and R introduces further ambiguity.

While both MPH^* and MPVH distributions lack explicit formulas for distribution and density functions, there is a sub-class of MPH^* distributions that does allow explicit forms. The latter is the one where the structure of the underlying Markov chain is of so-called *feed-forward* type.

Let $C_1, \ldots, C_n$ be sub-intensity matrices, and let $D_1, \ldots, D_n$ denote nonnegative matrices such that $-C_i e = D_i e$. The matrices D_i are not necessarily square matrices, with the number of rows being equal to the number of rows in C_i and the number of columns equal to the number of rows (and columns) of C_{i+1}. Define

$$\boldsymbol{\beta} = (\boldsymbol{\pi}, \boldsymbol{0}, \ldots, \boldsymbol{0}) \text{ and } T = \begin{pmatrix} C_1 & D_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & C_2 & D_2 & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & C_3 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & C_n \end{pmatrix}, \qquad (6)$$

and let the reward matrix be

$$R = \begin{pmatrix} e & 0 & 0 & \cdots & 0 \\ 0 & e & 0 & \cdots & 0 \\ 0 & 0 & e & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & e \end{pmatrix}. \tag{7}$$

The structure of the R matrix implies that the i'th total reward, X_i, then equals the inter-arrival time between arrivals $i-1$ and i. Positive correlation between two consecutive inter-arrivals $i-1$ and i can then be obtained by choosing the matrix D_i in such a way that a long (short) duration of the Markov chain in block $i-1$ will imply a long (short) duration in block i as well. For a negative correlation, we have to choose the matrix D_1 such that the implications are reversed. The joint density of the MPH* distribution is then given by

$$f(x_1, \ldots, x_n; \pi, T, R) = \pi e^{C_1 x_1} D_1 e^{C_2 x_2} D_2 \cdots D_{n-1} e^{C_n x_n} D_n e. \tag{8}$$

Remark 4 The matrices C_i are sub-intensity matrices, providing a phase-type distributed time until arrival i. The matrices D_i are nonnegative matrices containing intensities for initiating a new inter-arrival time for arrival $i+1$ at the time of the arrival i. Hence, the matrices D_i create the dependence between the inter-arrivals. In particular, if $D_i = c_i \pi_{i+1}$, where $c_i = -C_i e$ is the exit rate (column) vector corresponding to C_i and π_{i+1} is some probability (row) vector on $\{1, 2, \ldots, p_i\}$, then the inter-arrivals are independent.

Remark 5 The (full) matrix D_n is not really needed for our purposes, but only the exit vector $c_n = -C_n e = D_n e$. Thus, we may rewrite (8) in the form

$$f(x_1, \ldots, x_n; \pi, T, R) = \pi e^{C_1 x_1} D_1 e^{C_2 x_2} D_2 \cdots D_{n-1} e^{C_n x_n} c_n. \tag{9}$$

We shall, however, maintain the notation with D_n for notational reasons. Since $-C_i e = D_i e$ for all i, this also implies the exit vector

$$t = -Te = (0, 0, \ldots, 0, c_n)',$$

so $D_n e$, which is not part of T, is part of t [see (1)].

Remark 6 Note that the restriction $-C_i e = D_i e$ reduces the effective number of parameters contributed from those matrices from $2p_i^2$ to $2p_i^2 - p_i$. In particular, the model of (9), and therefore also (8), has $p_1 - 1 + \sum_{i=1}^{n-1} p_i (2p_i - 1) + p_n^2$ effective degrees of freedom.

Remark 7 If $C_i = C$ and $D_i = D$ for all i, then (8) is the joint density function for the first n inter-arrival times of a Markovian arrival process (MAP) (see, for example, Neuts 1979; Bladt and Nielsen 2017). This class of point processes is dense in the class of point processes on $\mathbb{R}_+$ (see Asmussen and Koole 1993), and therefore,

the class of distributions given by (8) is also dense—in the sense of weak convergence and with flexible dimension of the matrices C and D—in the class of multivariate distributions on $\mathbb{R}_+^n$.

Later, we shall need the joint fractional moments for such distributions, which are given in the following lemma.

Lemma 1 *Suppose that* $X = (X_1, X_2, \ldots, X_n)$ *has a joint phase-type distribution with density* (8). *Then, for* $\theta_i > 0, i = 1, \ldots, n,$

$$\mathbb{E}\big(X_1^{\theta_1} X_2^{\theta_2} \cdots X_n^{\theta_n}\big) = \left(\prod_{i=1}^n \Gamma(\theta_i + 1)\right) \boldsymbol{\pi} \left(\prod_{i=1}^n (-\boldsymbol{C}_i)^{-\theta_i - 1} \boldsymbol{D}_i \right) \boldsymbol{e}$$

Proof It is sufficient to prove the lemma for $n = 2$.

$$\begin{aligned}
\mathbb{E}\big(Z_1^{\theta_1} Z_2^{\theta_2}\big) &= \int_0^\infty \int_0^\infty z_1^{\theta_1} z_2^{\theta_2} \boldsymbol{\pi} e^{\boldsymbol{C}_1 z_1} \boldsymbol{D}_1 e^{\boldsymbol{C}_2 z_2} \boldsymbol{D}_2 \boldsymbol{e} \, \mathrm{d}z_1 \, \mathrm{d}z_2 \\
&= \boldsymbol{\pi} \int_0^\infty z_1^{\theta_1} e^{\boldsymbol{C}_1 z_1} \, \mathrm{d}z_1 \boldsymbol{D}_1 \int_0^\infty z_2^{\theta_2} e^{\boldsymbol{C}_2 z_2} \, \mathrm{d}z_2 \boldsymbol{D}_2 \boldsymbol{e} \\
&= \boldsymbol{\pi} L_{z^{\theta_1}}(-\boldsymbol{C}_1) \boldsymbol{D}_1 L_{z^{\theta_2}}(-\boldsymbol{C}_2) \boldsymbol{D}_2 \boldsymbol{e},
\end{aligned}$$

where $L_{z^\theta}(u) = \Gamma(u + 1)/u^{\theta+1}$ is the Laplace transform for $z \to z^\theta$. Since the Laplace transforms are analytic (where they are defined), the result follows by a functional calculus argument [see Theorem 3.4.4 of Bladt and Nielsen (2017)]. $\qquad \square$

2.4 Matrix Mittag–Leffler distributions

Let $(\boldsymbol{\pi}, \boldsymbol{T})$ be a phase-type representation. Then, a random variable X has a matrix Mittag–Leffler (MML) distribution with representation $(\alpha, \boldsymbol{\pi}, \boldsymbol{T})$, if it has Laplace transform

$$L_X(u; \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = \boldsymbol{\pi}(u^\alpha \boldsymbol{I} - \boldsymbol{T})^{-1} \boldsymbol{t}, \quad u \geq 0,$$

where $0 < \alpha \leq 1$. We write $X \sim$ MML $(\alpha, \boldsymbol{\pi}, \boldsymbol{T})$. Let

$$E_{\alpha,\beta}(z) = \sum_{k=0}^\infty \frac{z^k}{\Gamma(\alpha k + \beta)}, \quad z \in \mathbb{C},$$

denote the Mittag–Leffler (ML) function. Then (see Albrecher et al. 2019), the density of X is given by

$$f(x; \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = x^{\alpha-1} \boldsymbol{\pi} \, E_{\alpha,\alpha}(\boldsymbol{T} x^\alpha) \boldsymbol{t}, \quad x > 0,$$

and the corresponding c.d.f. is

$$F(x; \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = 1 - \boldsymbol{\pi} E_{\alpha,1}(\boldsymbol{T}x^{\alpha})\boldsymbol{e}, \quad x > 0.$$

The ML function with (complex) matrix argument $\boldsymbol{A}$ is defined as

$$E_{\alpha,\beta}(\boldsymbol{A}) = \sum_{k=0}^{\infty} \frac{\boldsymbol{A}^k}{\Gamma(\alpha k + \beta)}.$$

For $\beta > 0$, one can express the (then entire) ML function of a matrix $\boldsymbol{A}$ by Cauchy's formula

$$E_{\alpha,\beta}(\boldsymbol{A}) = \frac{1}{2\pi i} \int_{\gamma} E_{\alpha,\beta}(z)(z\boldsymbol{I} - \boldsymbol{A})^{-1} \, \mathrm{d}z,$$

where γ is a simple path enclosing the eigenvalues of $\boldsymbol{A}$. Invoking the residue theorem, for each entry of the matrix $E_{\alpha,\beta}(z)(z\boldsymbol{I} - \boldsymbol{A})^{-1}$, then provides a simple method for calculating $E_{\alpha,\beta}(\boldsymbol{A})$.

As outlined in Albrecher et al. (2019), MML distributions with $0 < \alpha < 1$ are heavy-tailed with tail indices less than one, so that their mean does not exist. This may be too restrictive in many situations, and one way to obtain a closely related class of distributions is by considering power transformations of the original MML distributed random variables. Indeed, if $X \sim$ MML $(\alpha, \boldsymbol{\pi}, \boldsymbol{T})$, then $X^{1/\nu}$ has density

$$f(x; \nu, \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = \nu x^{\nu\alpha-1} \boldsymbol{\pi} E_{\alpha,\alpha}(\boldsymbol{T}x^{\nu\alpha})\boldsymbol{t}, \quad x > 0,$$

and distribution function

$$F(x; \nu, \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = 1 - \boldsymbol{\pi} E_{\alpha,1}(\boldsymbol{T}x^{\alpha\nu})\boldsymbol{e}, \quad x > 0,$$

for $\nu > 0$ (cf. Albrecher et al. 2019). Rewriting $\beta = \nu\alpha$ leads to the reparametrization

$$f(x; \beta, \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = \frac{\beta}{\alpha} x^{\beta-1} \boldsymbol{\pi} E_{\alpha,\alpha}(\boldsymbol{T}x^{\beta})\boldsymbol{t}, \quad x > 0, \tag{10}$$

and

$$F(x; \beta, \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = 1 - \boldsymbol{\pi} E_{\alpha,1}(\boldsymbol{T}x^{\beta})\boldsymbol{e}, \quad x > 0. \tag{11}$$

Thus, for any $0 < \alpha \leq 1$ and $\beta > 0$, (10) and (11) define densities and their corresponding distribution functions, with tail index β instead of α. We shall refer to distributions with densities of the form (10) as power MML and write $X \sim$ MML$^{1/\nu}(\alpha, \boldsymbol{\pi}, \boldsymbol{T})$. Their Laplace transforms are somewhat more involved. Indeed, the Laplace transform for $X \sim$ MML$^{1/\nu}(\alpha, \boldsymbol{\pi}, \boldsymbol{T})$ is given by [see Formula (5.1.30) in Gorenflo et al. (2014) and compare to Gorenflo et al. (2014, p. 364)]

$$L_X(s; \nu, \alpha, \boldsymbol{\pi}, \boldsymbol{T}) = s^{-\nu\alpha} \boldsymbol{\pi} \left(\sum_{k=0}^{\infty} \frac{\Gamma(\nu\alpha(k+1))}{\Gamma(\alpha(k+1))} (s^{-\nu\alpha}\boldsymbol{T})^k \right) \boldsymbol{t}, \quad s \geq 0, \tag{12}$$

where the series expansion relates to a generalized Wright hypergeometric function. The similarity with the Laplace transform for $Y \sim \text{MML}(\alpha, \pi, T)$ may be appreciated by rewriting

$$L_Y(s; \alpha, \pi, T) = \pi(s^\alpha I - T)^{-1} t = s^{-\alpha} \pi(I - s^{-\alpha} T)^{-1} t, \quad s \geq 0, \qquad (13)$$

where we also notice that (12) reduces to (13) for $\nu = 1$.

3 Generalized matrix Mittag–Leffler distributions

The convolution of Mittag–Leffler distributions is not a Mittag–Leffler distribution. However, if the components in the convolution have the same tail index, then the resulting distribution is a MML.

Theorem 2 *Suppose that $X \sim \text{MML}(\alpha, \pi_1, T_1)$ and $Y \sim \text{MML}(\alpha, \pi_2, T_2)$. Then,*

$$X + Y \sim \text{MML}(\alpha, \pi, T),$$

with

$$\pi = (\pi_1, 0) \quad \text{and} \quad T = \begin{pmatrix} T_1 & t_1 \pi_2 \\ 0 & T_2 \end{pmatrix}.$$

Proof This result follows from the Laplace transform of $X + Y$ being

$$
\begin{aligned}
L_{X+Y}(u; \alpha, \pi, T) &= \pi_1 (u^\alpha I - T_1)^{-1} t_1 \pi_2 (u^\alpha I - T_2)^{-1} t_2 \\
&= (\pi_1, 0) \begin{pmatrix} (u^\alpha I - T_1)^{-1} & -(u^\alpha I - T_1)^{-1}(-t_1 \pi_2)(u^\alpha I - T_2)^{-1} \\ 0 & (u^\alpha I - T_2)^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ t_2 \end{pmatrix} \\
&= (\pi_1, 0) \left(u^\alpha I - \begin{pmatrix} T_1 & t_1 \pi_2 \\ 0 & T_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} 0 \\ t_2 \end{pmatrix}.
\end{aligned}
$$

$\square$

Since $X \sim \text{MML}(\alpha, \pi_1, T_1)$ implies that $cX \sim \text{MML}(\alpha, \pi, T)$ for any constant $c > 0$, where

$$\pi = \pi_1 \quad \text{and} \quad T = c^{-\alpha} T_1,$$

we conclude that if $X_1, X_2, \ldots, X_n$ are independent MML with the *same* tail index α, then any linear combination $c_1 X_1 + \cdots + c_n X_n$ with $c_1, c_2, \ldots, c_n \geq 0$ is again MML with tail index α.

The convolution of MML distributions with different tail indices is not an MML distribution, but naturally leads to an extended class of MML which we refer to as *Generalized* MML, as we will define in the sequel. If $X \sim \text{MML}(\alpha, \pi_1, T_1)$ and $Y \sim \text{MML}(\beta, \pi_2, T_2)$ with $\alpha \neq \beta$, then calculations similar to the proof of Theorem 2 lead to $X + Y$ having Laplace transform

$$L_{X+Y}(u) = (\boldsymbol{\pi}_1, \mathbf{0})\left(\boldsymbol{\Delta}(u^\alpha \boldsymbol{I}, u^\beta \boldsymbol{I}) - \begin{pmatrix} \boldsymbol{T}_1 & \boldsymbol{t}_1 \boldsymbol{\pi}_2 \\ \mathbf{0} & \boldsymbol{T}_2 \end{pmatrix}\right)^{-1}\begin{pmatrix} \mathbf{0} \\ \boldsymbol{t}_2 \end{pmatrix}, \qquad (14)$$

where $\boldsymbol{\Delta}(\boldsymbol{A}, \boldsymbol{B})$ denotes the block diagonal matrix

$$\boldsymbol{\Delta}(\boldsymbol{A}, \boldsymbol{B}) = \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{pmatrix}$$

for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The linear combination $c_1 X + c_2 Y$ will then have a Laplace transform on the form:

$$L_{c_1 X + c_2 Y}(u) = (\boldsymbol{\pi}_1, \mathbf{0})\left(\boldsymbol{\Delta}(u^\alpha \boldsymbol{I}, u^\beta \boldsymbol{I}) - \begin{pmatrix} c_1^{-\alpha} \boldsymbol{T}_1 & c_1^{-\alpha} \boldsymbol{t}_1 \boldsymbol{\pi}_2 \\ \mathbf{0} & c_2^{-\beta} \boldsymbol{T}_2 \end{pmatrix}\right)^{-1}\begin{pmatrix} \mathbf{0} \\ c_2^{-\beta} \boldsymbol{t}_2 \end{pmatrix}.$$

This motivates the following definition.

Definition 1 A random variable X is said to have a (univariate) generalized matrix Mittag–Leffler distribution, if there exist $\alpha_1, \dots, \alpha_n$ with $0 < \alpha_i \le 1$, and a phase-type representation $(\boldsymbol{\pi}, \boldsymbol{T})$ for which the absolutely continuous part of its Laplace transform is given by

$$L_X^{\text{cont}}(u; \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}) = \boldsymbol{\pi}(\boldsymbol{\Delta}(u^{\alpha_1} \boldsymbol{I}_1, \dots, u^{\alpha_n} \boldsymbol{I}_n) - \boldsymbol{T})^{-1}\boldsymbol{t}, \quad u \ge 0,$$

where $\boldsymbol{I}_k$ are identity matrices and $\dim(\boldsymbol{I}_1) + \cdots + \dim(\boldsymbol{I}_n) = \dim(\boldsymbol{T})$. We write

$$X \sim \text{GMML}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}),$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}_+^n$.

Then, if $X_1, \dots, X_n$ are independent with

$$X_i \sim \text{GMML}(\boldsymbol{\alpha}_i, \boldsymbol{\pi}_i, \boldsymbol{T}_i),$$

we get

$$X_1 + \cdots + X_n \sim \text{GMML}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T})$$

where

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n),$$
$$\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \mathbf{0}, \dots, \mathbf{0}),$$

and

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_1 & \boldsymbol{t}_1 \boldsymbol{\pi}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_2 & \boldsymbol{t}_2 \boldsymbol{\pi}_3 & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_3 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \vdots \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots \vdots \vdots & \boldsymbol{T}_n \end{pmatrix}.$$

By scaling, any nonnegative nonzero linear combination of GMML distributed random variables will again follow a GMML distribution.

4 The multivariate matrix Mittag–Leffler distribution

Motivated by Sect. 3, we proceed now to define the multivariate MML in a similar way as their underlying multivariate phase-type distributions.

Definition 2 A random vector $X = (X_1, \dots, X_n)$ has a multivariate GMML distribution in the wide sense, if all nonnegative non-vanishing linear combinations $c_1 X_1 + \cdots + c_n X_n$ have a GMML distribution.

As for MVPH distributions, this definition is not very practical from a constructive point of view, and we shall introduce a sub-class inspired by (2). To this end, we first notice the following result.

Lemma 2 *Let $\phi(s_1, \dots, s_k)$ be a multidimensional Laplace transform, and let $g_1(x)$, ..., $g_k(x)$ denote functions for which $-g_i$ are completely monotone. Then, it follows that*

$$L(s_1, \dots, s_k) = \phi(g_1(s_1), \dots, g_k(s_k))$$

is again a Laplace transform.

Proof This follows immediately from the multidimensional Bernstein–Widder theorem, see Bochner (2005, p. 87), which states that a multivariate function $\phi(s_1, \dots, s_k)$ is a multidimensional Laplace transform if and only if it is infinitely often differentiable and

$$(-1)^{n_1 + \cdots + n_k} \frac{\partial^{n_1 + \dots + n_k} \phi}{\partial s_1^{n_1} \dots \partial s_k^{n_k}} \geq 0$$

for all $n_1 \geq 0, \dots, n_k \geq 0$. □

From this we immediately get the following important result.

Theorem 3 *Let $(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ be a representation for a multivariate PH distribution* (2). *Then, the multidimensional function*

$$\phi(\boldsymbol{u}) = \boldsymbol{\pi}(\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{u}^{\boldsymbol{\alpha}}) - \boldsymbol{T})^{-1}\boldsymbol{t}, \quad \boldsymbol{u} \in \mathbb{R}_+^n, \tag{15}$$

with $\boldsymbol{u}^{\boldsymbol{\alpha}} = (u_1^{\alpha_1}, \dots, u_n^{\alpha_n})$, is a multidimensional Laplace transform.

From Theorem 1, we now obtain the following.

Theorem 4 *Let $w \geq 0$ denote a nonzero vector, and let $X = (X_1, \ldots, X_n)$ have a distribution given by the joint Laplace transform* (15) *with all $\alpha_i = \alpha$. Decompose (π, T) as in* (4) *according to Rw^α. Then, the distribution of $\langle X, w \rangle$ has an atom at zero of size $q = \pi_0 \left(I - \left(-T_{00} \right)^{-1} T_{0+} \right) e$ and a possibly defective absolute continuous part which is* $\mathrm{MML}\,(\alpha, \pi_{w^\alpha}, T_{w^\alpha})$, *where $(\pi_{w^\alpha}, T_{w^\alpha})$ is given in Theorem* 1.

Proof The result follows from

$$
\begin{aligned}
\mathbb{E}\left(\mathrm{e}^{-u\langle X, w\rangle}\right) &= \mathbb{E}\left(\mathrm{e}^{-\langle X, uw\rangle}\right) \\
&\overset{(15)}{=} \pi(\Delta(Ru^\alpha w^\alpha) - T)^{-1} t \\
&\overset{(5)}{=} q + \pi_{w^\alpha}(u^\alpha I - T_{w^\alpha})^{-1} t_{w^\alpha}.
\end{aligned}
$$

$\square$

For possibly distinct α_i, we proceed as follows.

Theorem 5 *Let $w \geq 0$ denote a nonzero vector, and let $X = (X_1, \ldots, X_n)$ be a random vector with joint Laplace transform* (15). *Decompose (π, T) as in* (4) *according to Rw^α. Then, the distribution of $\langle X, w \rangle$ has an atom at zero of size $p = \pi_0 \left(I - \left(-T_{00} \right)^{-1} T_{0+} \right) e$ and a possibly defective absolute continuous part which is* $\mathrm{GMML}\,(\alpha, \pi_{w^\alpha}, T_{w^\alpha})$, *where $(\pi_{w^\alpha}, T_{w^\alpha})$ is given in Theorem* 1.

Proof We have that

$$
\begin{aligned}
\mathbb{E}\left(\mathrm{e}^{-u\langle X, w\rangle}\right) &= \mathbb{E}\left(\mathrm{e}^{-\langle X, uw\rangle}\right) \\
&= \pi(\Delta(R(uw)^\alpha) - T)^{-1} t \\
&= \pi(\Delta(Rw^\alpha)\Delta(u^\alpha) - T)^{-1} t,
\end{aligned}
$$

where $\Delta(u^\alpha) = \mathrm{diag}\,(u^{\alpha_1}, \ldots, u^{\alpha_n})$. Now splitting into blocks according to E_+ and E_0, we see that

$$
\begin{aligned}
\pi(\Delta(Rw^\alpha)\Delta(u^\alpha) - T)^{-1} t &= \pi \begin{pmatrix} \Delta(Rw^\alpha)_+ \Delta(u^\alpha)_+ - T_{++} & -T_{+0} \\ -T_{0+} & -T_{00} \end{pmatrix}^{-1} t \\
&= (\pi_+, \pi_0) \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} t_+ \\ t_0 \end{pmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
A_{11} &= \left(\Delta(Rw^\alpha)_+ \Delta(u^\alpha)_+ - T_{++} - T_{+0}(-T_{00})^{-1}T_{0+} \right)^{-1} \\
&= \left(\Delta(u^\alpha)_+ - \left(\Delta(Rw^\alpha)_+\right)^{-1}\left[T_{++} + T_{+0}(-T_{00})^{-1}T_{0+} \right] \right)^{-1} \Delta(Rw^\alpha)_+^{-1} \\
&= \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1},
\end{aligned}
$$

$$
A_{12} = \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1} T_{+0}(-T_{00})^{-1},
$$

$$
A_{21} = (-T_{00})^{-1} T_{0+} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1},
$$

$$
A_{22} = (-T_{00})^{-1} \left(I + T_{0+} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1} T_{+0}(-T_{00})^{-1} \right).
$$

Then,

$$
\boldsymbol{\pi}_+ A_{11} + \boldsymbol{\pi}_0 A_{21} = \boldsymbol{\pi}_{w^\alpha} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1},
$$
$$
\boldsymbol{\pi}_+ A_{12} + \boldsymbol{\pi}_0 A_{22} = \boldsymbol{\pi}_0(-T_{00})^{-1} + \boldsymbol{\pi}_{w^\alpha} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} \Delta(Rw^\alpha)_+^{-1} T_{+0}(-T_{00})^{-1}.
$$

Now inserting

$$
\begin{pmatrix} t_+ \\ t_0 \end{pmatrix} = -Te = \begin{pmatrix} -T_{++}e - T_{+0}e \\ -T_{0+}e - T_{00}e \end{pmatrix},
$$

we get

$$
\begin{aligned}
&\left(\boldsymbol{\pi}_+ A_{11} + \boldsymbol{\pi}_0 A_{21} \right) t_+ + \left(\boldsymbol{\pi}_+ A_{12} + \boldsymbol{\pi}_0 A_{22} \right) t_0 \\
&= \boldsymbol{\pi}_0 (I - (-T_{00})^{-1}T_{0+})e + \boldsymbol{\pi}_{w^\alpha} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} t_{w^\alpha} \\
&= p + \boldsymbol{\pi}_{w^\alpha} \left(\Delta(u^\alpha)_+ - T_{w^\alpha} \right)^{-1} t_{w^\alpha}
\end{aligned}
$$

with

$$
t_{w^\alpha} = -T_{w^\alpha}e.
$$

$\square$

From the previous results, we see that we have found a sub-class of multivariate matrix Mittag–Leffler distributions with explicit Laplace transform. This allows us to concentrate on this class, and to make the following definition.

Definition 3 Let $X = (X_1, \dots, X_n)$ be a random vector. Then, we say that X has a multivariate matrix generalized Mittag–Leffler distribution if it has joint Laplace transform given by (15) and write

$$
X \sim \text{GMML}\,(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R}).
$$

The following result generalizes Theorem 3.6 of Albrecher et al. (2019) to the multivariate case. In particular, it gives the probabilistic interpretation of the GMML class as a family of random vectors whose marginals are absorption times of randomly scaled, time-inhomogeneous Markov processes. The dependence of the corresponding Markov processes arises from the fact that they are all generated according to a reward structure on an underlying common Markov jump process.

Theorem 6 *Let $X = (X_1, \ldots, X_n) \sim$ GMML $(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$. Then,*

$$X \stackrel{d}{=} \boldsymbol{W^{1/\alpha}} \bullet \boldsymbol{S_\alpha}, \tag{16}$$

where $\boldsymbol{W^{1/\alpha}} = (W_1^{1/\alpha_1}, \ldots, W_n^{1/\alpha_n})$ with $\boldsymbol{W} = (W_1, \ldots, W_n) \sim$ MPH $^*(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ [see (2)], and where $\boldsymbol{S_\alpha} = (S_{\alpha_1}, \ldots, S_{\alpha_n})$ is a vector of independent stable random variables, each with Laplace transform $\exp(-u^{\alpha_i})$. Here, $\bullet$ refers to component-wise multiplication of vectors.

Proof We observe that

$$
\begin{aligned}
\mathbb{E}(\exp(-\langle \boldsymbol{u}, \boldsymbol{W^{1/\alpha}} \bullet \boldsymbol{S_\alpha} \rangle)) &= \int_{\mathbb{R}_+^n} \mathbb{E}(\exp(-\langle \boldsymbol{u}, \boldsymbol{w^{1/\alpha}} \bullet \boldsymbol{S_\alpha} \rangle)) \, dF_W(\boldsymbol{w}) \\
&= \int_{\mathbb{R}_+^n} \exp(-[u_1^{\alpha_1} w_1 + \cdots + u_n^{\alpha_n} w_n]) \, dF_W(\boldsymbol{w}) \\
&= \int_{\mathbb{R}_+^n} \exp(-\langle \boldsymbol{u^\alpha}, \boldsymbol{w} \rangle) \, dF_W(\boldsymbol{w}) \\
&= \boldsymbol{\pi}(\boldsymbol{\Delta}(\boldsymbol{R}\boldsymbol{u^\alpha}) - \boldsymbol{T})^{-1}\boldsymbol{t},
\end{aligned}
$$

which implies the desired representation. □

Remark 8 From representation (16), we have that the marginals of any multivariate GMML distribution are regularly varying with indices $\alpha_1, \ldots, \alpha_n$, all smaller than 1. Moreover, by the multivariate version of Breiman's lemma (cf. Basrak et al. 2008) and the fact that multivariate phase-type distributions have moments of all orders, it follows that the tail-independence structure of the vector $\boldsymbol{S_\alpha}$ carries over to X. That is, the multivariate GMML family introduced in this paper has (very) heavy-tailed GMML marginals, but is tail-independent. As mentioned in the introduction, application areas for such models are, for example, given in Resnick (2002).

A consequence of $\alpha_i < 1$ is that the mean does not exist. To alleviate this potential practical drawback, it was proposed in Albrecher et al. (2019) to consider power-transformed variables in the univariate case. In the same way, we propose the following definition.

Definition 4 Let $X \sim$ GMML$(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$. For $\nu > 0$, we define

$$Y = X^{1/v} \sim \text{GMML}^{1/v}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R}),$$

and refer to it as the class of power multivariate MML distributions.

Under the power transform, the class is in general no longer closed under linear combinations. For fixed $\boldsymbol{\alpha}$, however, it possesses the following denseness property [in contrast to distributions with Laplace transform (15)]. Here, 'dense on $\mathbb{R}^n_+$' means dense in the sense of weak convergence among all distributions on $\mathbb{R}^n_+$.

Theorem 7 *(i) The class of* $\text{GMML}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ *variables is dense on* $\mathbb{R}^n_+$.
(ii) For any fixed $\boldsymbol{\alpha}$, *the class of* $\text{GMML}^{1/v}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ *variables is dense on* $\mathbb{R}^n_+$.
(iii) For any fixed marginal tail indices $\boldsymbol{\alpha} \bullet \boldsymbol{v} = \boldsymbol{\gamma}^{-1} > 0$, *the class of*
$\text{GMML}^{1/v}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ *variables is dense on* $\mathbb{R}^n_+$.

Proof (i) The statement is evident by noticing that we may choose $\boldsymbol{\alpha} \equiv \mathbf{1}$ and recalling that the class of variables with Laplace transform (2) is dense on $\mathbb{R}^n_+$.

(ii) Let $\mathbf{0} < \boldsymbol{v}_1 < \boldsymbol{v}_2 < \cdots$ be any increasing and (entry-wise) diverging sequence of vectors and Y be an arbitrary random vector on $\mathbb{R}^n_+$. Let $\boldsymbol{S}_{\boldsymbol{\alpha}}$ be as in Theorem 6 and notice that $\boldsymbol{S}_{\boldsymbol{\alpha}}^{1/v_n} \to \mathbf{1}$. In particular, $\boldsymbol{S}_{\boldsymbol{\alpha}}^{1/v_n} \xrightarrow{d} \mathbf{1}$. Moreover, we may choose an independent sequence of vectors $\boldsymbol{W}_n$ with Laplace transforms of the form (2) such that $\boldsymbol{W}_n^{1/v_n} \xrightarrow{d} Y$. Applying the continuous mapping theorem, and by the characterization of Theorem 6, the statement follows.

(iii) Similar to the previous case, let $\mathbf{0} < \boldsymbol{\alpha}_1 < \boldsymbol{\alpha}_2 < \cdots$ be an increasing sequence of vectors, converging to $\mathbf{1}$, and set $\boldsymbol{v}_n = (\boldsymbol{\gamma} \bullet \boldsymbol{\alpha}_n)^{-1}$. With $\boldsymbol{S}_{\boldsymbol{\alpha}}$ as in Theorem 6, we have that $\boldsymbol{S}_{\boldsymbol{\alpha}_n}^{1/v_n} \xrightarrow{d} \mathbf{1}$. Choosing an independent sequence of vectors $\boldsymbol{W}_n$ with Laplace transforms of the form (2) and with $\boldsymbol{W}_n^{1/v_n} \xrightarrow{d} Y$, the proof is finished as before. $\square$

Remark 9 The above result shows how several classes of multivariate Mittag–Leffler distributions and their power transforms are dense in the set of all distributions of the n-dimensional positive orthant. However, since we are dealing with a tail-independent model, the number of phases increases drastically when faced with the need to capture dependence above high thresholds. Heuristically, the tail dependence is only correctly modelled in the limit. This is in some way analogous to the fact that phase-type distributions are dense on all distributions on the positive real line, but they are all light-tailed (of exponential decay), and very large dimensions are needed for approximations of heavy-tailed distributions, cf. Bladt and Nielsen (2017).

5 Special structures and examples

From the previous sections, it becomes clear that the tail behaviour of the GMML class is determined by the parameters α_i (cf. Remark 8) and the dependence structure is mainly triggered by the parameters of the reward matrix $\boldsymbol{R}$, as these determine joint contributions to the size of each component. The marginal behaviour and overall shape in the body of the distribution are then finally implied by the

structure of the phase-type components $(\boldsymbol{\pi}, \boldsymbol{T})$. In particular, the dimension p of the latter also determines the potential for possible multimodalities of the components. In fact, Theorem 7 on the denseness of $\text{GMML}^{1/\nu}$ distributions on $\mathbb{R}_+^n$ relies [implicitly in part (i)] on the possibility of having arbitrarily large dimension p, a flexibility that is needed for modelling multiple modes, as the latter can require many phases. However, due to the possibly complex interaction of all parameters, one cannot uniquely assign the role of each of the parameters to achieve a particular distributional behaviour or shape. Moreover, for arbitrary combinations of parameters, it is not always possible to get an explicit expression for the density of a GMML distribution (a complication inherited from the phase-type distributions).

We now proceed to give an example of a sub-class that, however, does allow an explicit form. To that end, consider the special structure (6) and (7) for $(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$, which in the exponential case led to the density (8),

$$f(x_1, \ldots, x_n; \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R}) = \boldsymbol{\pi} e^{C_1 x_1} \boldsymbol{D}_1 e^{C_2 x_2} \boldsymbol{D}_2 \cdots \boldsymbol{D}_{n-1} e^{C_n x_n} \boldsymbol{D}_n \boldsymbol{e}.$$

This choice of $(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$, when plugged into (15), results in the joint Laplace transform of $X \sim \text{GMML}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$

$$L_X(\boldsymbol{u}; \theta) = \boldsymbol{\pi} \begin{pmatrix} u_1^{\alpha_1} \boldsymbol{I} - \boldsymbol{C}_1 & -\boldsymbol{D}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & u_2^{\alpha_2} \boldsymbol{I} - \boldsymbol{C}_2 & -\boldsymbol{D}_2 & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & u_3^{\alpha_3} \boldsymbol{I} - \boldsymbol{C}_3 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \vdots \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & u_n^{\alpha_n} \boldsymbol{I} - \boldsymbol{C}_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{D}_n \boldsymbol{e} \end{pmatrix},$$

(17)

where we now use the shorthand notation $\theta = (\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$. For the resulting class of GMML distributions, we can derive joint and marginal density functions, but first we notice the following lemma.

Lemma 3
$$\int_0^\infty x^{\alpha-1} E_{\alpha,\alpha}(\boldsymbol{T} x^\alpha) \, dx = -\boldsymbol{T}^{-1}.$$

Proof Since $\lambda \to \lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha)$ is an analytic function, and a density as a function of x, we get that

$$\int_0^\infty x^{\alpha-1} E_{\alpha,\alpha}(\boldsymbol{T} x^\alpha) \, dx = \int_0^\infty x^{\alpha-1} \frac{1}{2\pi i} \int_\gamma E_{\alpha,\alpha}(s x^\alpha)(s\boldsymbol{I} - \boldsymbol{T})^{-1} \, ds \, dx$$

$$= \frac{1}{2\pi i} \int_\gamma \left(\int_0^\infty x^{\alpha-1} E_{\alpha,\alpha}(s x^\alpha) \, dx \right) (s\boldsymbol{I} - \boldsymbol{T})^{-1} \, ds$$

$$= \frac{1}{2\pi i} \int_\gamma (-s^{-1})(s\boldsymbol{I} - \boldsymbol{T})^{-1} \, ds$$

$$= -\boldsymbol{T}^{-1}.$$

$\square$

Remark 10 The matrix $U = -T^{-1}$ is the so-called Green matrix which has the following probabilistic interpretation: the element (i, j) of U is the expected time that the Markov jump process underlying a phase-type distribution with generator T spends in state j (prior to absorption) given that it starts in state i.

The main result of this section is as follows.

Theorem 8 *The Laplace transform* (17) *can equivalently be written as*

$$L_X(\boldsymbol{u};\boldsymbol{\theta}) = \boldsymbol{\pi} \left(\prod_{i=1}^{n} (u_i^{\alpha_1} \boldsymbol{I} - \boldsymbol{C}_i)^{-1} \boldsymbol{D}_i \right) \boldsymbol{e}, \quad \boldsymbol{u} \in \mathbb{R}_+^n. \tag{18}$$

The corresponding joint density is given by

$$f_X(x_1, \ldots, x_n; \boldsymbol{\theta}) = \boldsymbol{\pi} \left(\prod_{i=1}^{n} x_i^{\alpha_i - 1} E_{\alpha_i, \alpha_i}(\boldsymbol{C}_i x_i^{\alpha_i}) \boldsymbol{D}_i \right) \boldsymbol{e}, \quad x_i > 0, \ i = 1, \ldots, n. \tag{19}$$

For the i'th marginal distribution of X_i, we have

$$X_i \sim \text{MML}\,(\alpha_i, \boldsymbol{\beta}_i, \boldsymbol{C}_i)$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\pi} \prod_{j=1}^{i-1} (-\boldsymbol{C}_j)^{-1} \boldsymbol{D}_j.$$

Proof It is sufficient to prove the result for $n = 2$. (18) follows from the general block diagonal inversion formula

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

Concerning (19), we have that

$$\int_0^\infty \int_0^\infty e^{-s_1 x_1 - s_2 x_2} \boldsymbol{\pi} x_1^{\alpha_1} E_{\alpha_1, \alpha_1}(\boldsymbol{C}_1 x_1^{\alpha_1}) \boldsymbol{D}_1 x_2^{\alpha_2} E_{\alpha_2, \alpha_2}(\boldsymbol{C}_2 x_2^{\alpha_2}) \boldsymbol{D}_2 \boldsymbol{e} \, dx_1 \, dx_2$$

$$= \int_0^\infty e^{-s_1 x_1} x_1^{\alpha_1} \boldsymbol{\pi} E_{\alpha_1, \alpha_1}(\boldsymbol{C}_1 x_1^{\alpha_1}) \, dx_1 \boldsymbol{D}_1 \int_0^\infty e^{-s_2 x_2} x_2^{\alpha_2} E_{\alpha_2, \alpha_2}(\boldsymbol{C}_2 x_2^{\alpha_2}) \boldsymbol{D}_2 \boldsymbol{e} \, dx_2$$

$$= \boldsymbol{\pi}(u_1^{\alpha_1} \boldsymbol{I} - \boldsymbol{C}_1)^{-1} \boldsymbol{D}_1 (u_2^{\alpha_2} \boldsymbol{I} - \boldsymbol{C}_2)^{-1} \boldsymbol{D}_2 \boldsymbol{e}$$

$$= (\boldsymbol{\pi}, \boldsymbol{0}) \begin{pmatrix} u_1^{\alpha_1} \boldsymbol{I} - \boldsymbol{C}_1 & -\boldsymbol{D}_1 \\ \boldsymbol{0} & u_2^{\alpha_2} \boldsymbol{I} - \boldsymbol{C}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{D}_2 \boldsymbol{e} \end{pmatrix},$$

which is of the form (15).

The result on the marginal distributions follows from Lemma 3 and by using that $(\boldsymbol{C}_i + \boldsymbol{D}_i)\boldsymbol{e} = \boldsymbol{0}$, implying that $(-\boldsymbol{C}_i)^{-1}\boldsymbol{D}_i\boldsymbol{e} = \boldsymbol{e}$. $\qquad\square$

The previous result can be used in the construction of bivariate (or multivariate) Mittag–Leffler distributions of a reasonably general type.

Example 1 (Bivariate Mittag–Leffler distribution) In this example, we construct a class of bivariate distributions with Mittag–Leffler distributed marginals. The starting point is the construction of a bivariate exponential distribution underlying the MML. For details on this construction, we refer to Sect. 8.3.2 of Bladt and Nielsen (2017). Let m be a positive integer and

$$S = \begin{pmatrix} -m\lambda & (m-1)\lambda & 0 & \ldots & 0 & 0 \\ 0 & -(m-1)\lambda & (m-2)\lambda & \ldots & 0 & 0 \\ 0 & 0 & -(m-2)\lambda & \ldots & 0 & 0 \\ \vdots & \vdots & v\vdots & \ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -2\lambda & \lambda \\ 0 & 0 & 0 & \cdots & 0 & -\lambda \end{pmatrix}.$$

Then, for any initial distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$, the phase-type distribution PH $(\boldsymbol{\pi}, S)$ is simply an exponential distribution with intensity λ. Similarly, if we let

$$\tilde{S} = \begin{pmatrix} -\mu & \mu & 0 & \ldots & 0 & 0 \\ 0 & -2\mu & 2\mu & \ldots & 0 & 0 \\ 0 & 0 & -3\mu & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(m-1)\mu & (m-1)\mu \\ 0 & 0 & 0 & \ldots & 0 & -m\mu \end{pmatrix}$$

and $\tilde{\boldsymbol{\pi}} = \frac{1}{m}e = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$, then PH $(\tilde{\boldsymbol{\pi}}, \tilde{S})$ is again exponentially distributed with intensity μ. Let P be a doubly stochastic matrix, i.e. its elements are nonnegative and

$$Pe = e \quad \text{and} \quad e'P = e',$$

and define

$$T = \begin{pmatrix} S & \lambda P \\ 0 & \tilde{S} \end{pmatrix}.$$

Consider the reward matrix

$$R = \begin{pmatrix} e & 0 \\ 0 & e \end{pmatrix}.$$

Then, MPH $^*(e_1', T, R)$ is a bivariate exponential distribution. This class of bivariate exponential distributions is capable of achieving any feasible correlation (ranging from $1 - \pi^2/6$ to 1) by choosing m sufficiently large and P adequately (see Bladt and Nielsen 2010). Independence is achieved for

$$P = \frac{1}{m}E,$$

where $E = \{1\}_{i,j=1,\ldots,m}$ is the matrix of ones, maximum negative (minimum) correlation (up to order m) by

$$P = I$$

and maximum positive correlation for order up to m by

$$P = \{\delta_{i,m-i+1}\},$$

which is the anti-diagonal unit matrix, cf. He et al. (2012).

The corresponding GMML $(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ then has a density f of the form

$$f(x_1, x_2; \boldsymbol{\theta}) = m\lambda\mu x_1^{\alpha_1-1} x_2^{\alpha_2-1} \boldsymbol{e}_1' E_{\alpha_1,\alpha_1}(Sx_1^{\alpha_1}) P E_{\alpha_2,\alpha_2}(\tilde{S}x_2^{\alpha_2}) \boldsymbol{e}_n, \quad x_1, x_2 > 0, \quad (20)$$

where as usual $\boldsymbol{e}_i$ denotes the i'th Euclidian unit vector. The marginals are Mittag–Leffler distributions with densities

$$f_{X_1}(x; \alpha_1, \lambda) = \lambda x^{\alpha_1-1} E_{\alpha_1,\alpha_1}(-\lambda x^{\alpha_1-1}) \quad \text{and} \quad f_{X_2}(x; \alpha_2, \mu) = \mu x^{\alpha_2-1} E_{\alpha_2,\alpha_2}(-\mu x^{\alpha_2-1}),$$

for $x > 0$, which follows directly from the invariance under different representations (parametrizations), or by simple integration and using Lemma 3. Note that the present dependence structure has a very natural interpretation as a copula constructed in terms of combining marginal order statistics, cf. Baker (2008) and Bladt and Nielsen (2017 Sec. 8.3.2), here for Mittag–Leffler marginals.

We can write the expression (20) slightly more explicit. The eigenvalues of S are $-m\lambda, -(m-1)\lambda, \ldots, -\lambda$. To the eigenvalue $-\lambda k$, there corresponds an eigenvector $\boldsymbol{v}^{(k)} = (v_1^{(k)}, \ldots, v_n^{(k)})$ with

$$v_1^{(k)} = 1$$
$$v_{i+1}^{(k)} = \left(1 - \frac{k-1}{m-i}\right) v_i^{(k)}, \quad i = 1, \ldots, m-1.$$

Similarly, $\tilde{S}$ has eigenvalues $-\mu m, -\mu(m-1), \ldots, -\mu$, and to the eigenvalue $-k\mu$, there corresponds an eigenvector $\boldsymbol{w}^{(k)}$ with

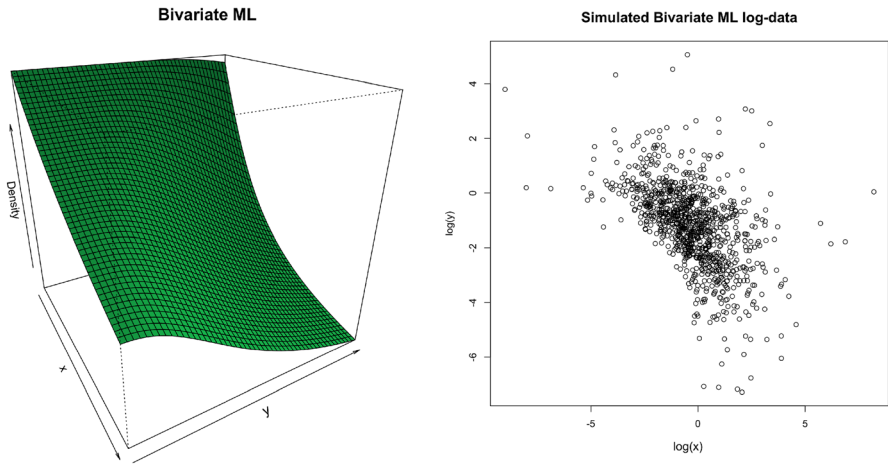$$w_1^{(k)} = 1$$
$$w_{i+1}^{(k)} = \left(1 - \frac{k}{i}\right) w_i^{(k)}, \quad i = 1, \ldots, m-1.$$

Considering $\boldsymbol{v}^{(k)}$ and $\boldsymbol{w}^{(k)}$ as column vectors, we form the matrices $V = (\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)})$ and $W = (\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(m)})$. Then, we may write

$$E_{\alpha_1,\alpha_1}(Sx^{\alpha_1}) = V\Delta\left(E_{\alpha_1,\alpha_1}(-m\lambda x^{\lambda_1}), \ldots, E_{\alpha_1,\alpha_1}(-\lambda x^{\alpha_1})\right) V^{-1},$$
$$E_{\alpha_2,\alpha_2}(\tilde{S}x^{\alpha_2}) = W\Delta\left(E_{\alpha_2,\alpha_2}(-m\mu x^{\alpha_2}), \ldots, E_{\alpha_1,\alpha_1}(-\mu x^{\alpha_2})\right) W^{-1}.$$

Bivariate ML

Simulated Bivariate ML log-data

Fig. 1 Density and 1000 simulated data points from a bivariate ML distribution with negative log-correlation (empirical correlation of −0.53)

Bivariate ML

Simulated Bivariate ML log-data
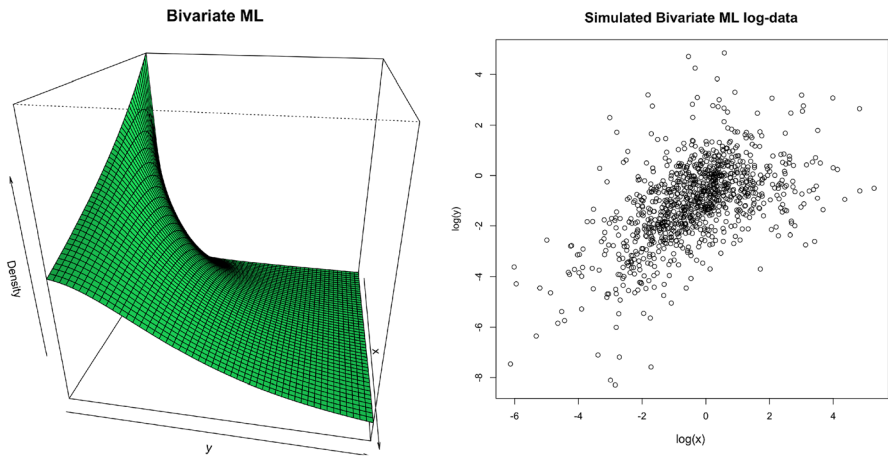
Fig. 2 Density and 1000 simulated data points from a bivariate ML distribution with positive correlation (empirical correlation of 0.55)

Though the correlation between the Mittag–Leffler marginals is not defined (since moments of orders larger than α do not exist), some notion of dependence may be appreciated from the correlation structure of the underlying phase-type distribution.

In Fig. 1, we depict a bivariate Mittag–Leffler density along with simulated data for the parameters $\alpha = (0.6, 0.7)$, $m = 20$, $\lambda = 1$, $\mu = 2$, and P is the identity matrix.

In Fig. 2, we use the same parameters but with P being the counter-identity matrix. As expected, the sign of the log-correlation is determined by the structure

of the latter matrix. Notice that the number of effective parameters corresponding to each of the two proposed structures is five. □

Concerning the power MML with this structure, we have the following result.

Theorem 9 *Assume that X has joint density* (19). *Then,* $Y = X^{1/\nu}$ *has the joint density*

$$f_Y(x_1, \ldots, x_n; \boldsymbol{\nu}, \boldsymbol{\theta}) = \boldsymbol{\pi}\left(\prod_{i=1}^{n} \nu_i x_i^{\alpha_i \nu_i - 1} E_{\alpha_i, \alpha_i}(C_i x_i^{\alpha_i \nu_i}) D_i\right)\boldsymbol{e}, \quad x_i > 0, \ i = 1, \ldots, n,$$

and joint moments

$$\mathbb{E}\left(Y_1^{\theta_1} Y_2^{\theta_2} \cdots Y_n^{\theta_n}\right)$$

$$= \prod_{i=1}^{n}\left(\frac{\Gamma(1 - \theta_i/(\nu_i\alpha_i))\Gamma(1 + \theta_i/(\nu_i\alpha_i))}{\Gamma(1 - \theta_i/\nu_i)}\right)\boldsymbol{\pi}\left(\prod_{i=1}^{n}(-C_i)^{-\theta_i/\nu_i\alpha_i - 1} D_i\right)\boldsymbol{e},$$

where $\nu_i\alpha_i > \theta_i > 0,$ *for* $i = 1, 2, \ldots, n.$

Proof The form of the joint density is immediate. Concerning the form of the moments, it suffices to consider the case $n = 2$. Using the decomposition (6), we get

$$\mathbb{E}(Y_1^{\theta_1} Y_2^{\theta_2}) = \mathbb{E}\left(W_{\alpha_1}^{\frac{\theta_1}{\alpha_1\nu_1}} W_{\alpha_2}^{\frac{\theta_2}{\alpha_2\nu_2}} S_{\alpha_1}^{\frac{\theta_1}{\nu_1}} S_{\alpha_2}^{\frac{\theta_2}{\nu_2}}\right)$$

$$= \mathbb{E}\left(W_1^{\frac{\theta_1}{\alpha_1\nu_1}} W_2^{\frac{\theta_2}{\alpha_2\nu_2}}\right)\mathbb{E}\left(S_{\alpha_1}^{\frac{\theta_1}{\nu_1}}\right)\mathbb{E}\left(S_{\alpha_2}^{\frac{\theta_2}{\nu_2}}\right),$$

where (W_1, W_2) has a bivariate phase-type distribution with joint density (8). Since

$$\mathbb{E}\left(S_{\alpha_i}^{\frac{\theta_i}{\nu_i}}\right) = \frac{\Gamma\left(1 - \frac{\theta_i}{\alpha_i\nu_1}\right)}{\Gamma\left(1 - \frac{\theta_i}{\nu_1}\right)},$$

the result then follows from Lemma 1.

□

Example 2 Consider the case of a bivariate MML distribution, $\theta_1 = \theta_2 = 1$, $\nu_i\alpha_i > 1$ and that C_1 and C_2 have the same dimension. (The latter can always be achieved by augmenting the smaller one.) Using the abbreviation

$$c_i = \frac{\Gamma(1 - 1/(\nu_i\alpha_i))\Gamma(1 + 1/(\nu_i\alpha_i))}{\Gamma(1 - 1/\nu_i)}, \quad i = 1, 2,$$

we get

$$\mathbb{E}(Y_1) = c_1\boldsymbol{\pi}(-\boldsymbol{C}_1)^{-1/(\alpha_1 v_1)-1}\boldsymbol{D}_1\boldsymbol{e},$$
$$\mathbb{E}(Y_2) = c_2\boldsymbol{\pi}(-\boldsymbol{C}_1)^{-1}\boldsymbol{D}_1(-\boldsymbol{C}_2)^{-1/(\alpha_1 v_1)-1}\boldsymbol{D}_2\boldsymbol{e},$$
$$\mathbb{E}(Y_1 Y_2) = c_1 c_2\boldsymbol{\pi}(-\boldsymbol{C}_1)^{-1/(\alpha_1 v_1)-1}\boldsymbol{D}_1(-\boldsymbol{C}_2)^{-1/(\alpha_2 v_2)-1}\boldsymbol{D}_2\boldsymbol{e}.$$

If $v_i\alpha_i > 2$, we can calculate variances and correlation. Indeed, with

$$c_i' = \frac{\Gamma(1 - 2/(v_i\alpha_i))\Gamma(1 + 2/(v_i\alpha_i))}{\Gamma(1 - 2/v_i)}, \quad i = 1, 2,$$

one has

$$\mathbb{E}(Y_1^2) = c_1'\boldsymbol{\pi}(-\boldsymbol{C}_1)^{-2/(\alpha_1 v_1)-1}\boldsymbol{D}_1\boldsymbol{e}$$
$$\mathbb{E}(Y_2^2) = c_2'\boldsymbol{\pi}(-\boldsymbol{C}_1^{-1}\boldsymbol{D}_1)(-\boldsymbol{C}_1)^{-2/(\alpha_2 v_2)-1}\boldsymbol{D}_2\boldsymbol{e}$$

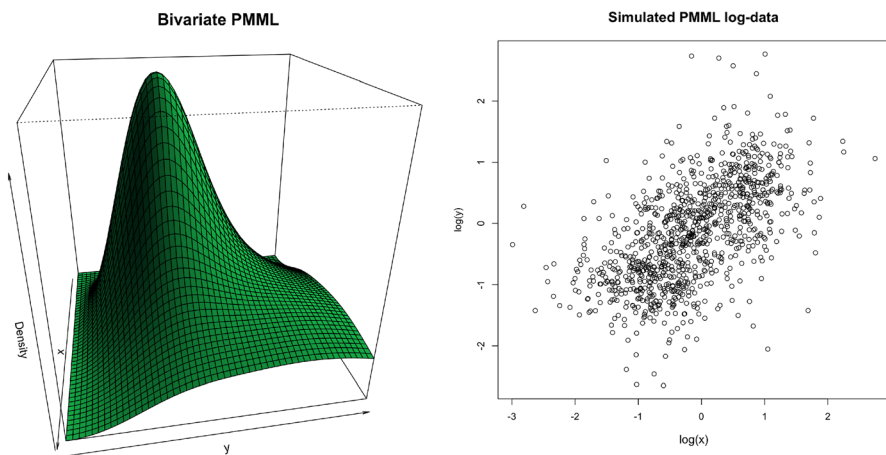from which the correlation coefficient is readily calculated.

In Fig. 3, we depict a bivariate density from a GMML $^{1/v}(\boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{R})$ distribution along with simulated data. The parameters are given by

$$\boldsymbol{\alpha} = (0.6, 0.7), \quad \boldsymbol{\beta} = \boldsymbol{v} \bullet \boldsymbol{\alpha} = (3, 3),$$

and the phase-type component being of the feed-forward structure (6) and (7), with $n = 2$, $\boldsymbol{\beta}_1 = (1/3, 1/3, 1/3)$, $\boldsymbol{\beta}_2 = \boldsymbol{0}$,

$$\boldsymbol{C}_1 = \boldsymbol{C}_2 = \begin{pmatrix} -10 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1/10 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{D}_1 = -\boldsymbol{C}_1 = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/10 \end{pmatrix}.$$

Fig. 3 Density and 1000 simulated data points from a power multivariate GMML distribution with positive correlation (true correlation of 0.35 and empirical of 0.37)

Hence, both marginals are mixtures of power Mittag–Leffler distributions. The mixing probabilities of the two distributions are also the same, (1/3, 1/3, 1/3), since the diagonal form of D_1 ensures that the second mixture draws the same component as the first. The first marginal mixture distribution has a density given by

$$f_1(x) = \frac{5}{3}x^3 \sum_{i=1}^{3} \lambda_i \, E_{0.6,0.6}(-\lambda_i x^3), \tag{21}$$

where $\lambda_1 = 10$, $\lambda_2 = 1$ and $\lambda_3 = 1/10$, while the second marginal density has the form

$$f_2(x) = \frac{10}{7}x^3 \sum_{i=1}^{3} \lambda_i \, E_{0.7,0.7}(-\lambda_i x^3). \tag{22}$$

The reward matrix is

$$R = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

and Y_1 and Y_2, simply correspond to the aforementioned mixtures. The structure of D_1 implies a strong positive correlation. For example, if Y_1 is picked from the mixture component with rate 10, then Y_2 will be picked from the same component (but then drawn independently).

In Fig. 4, we use the same parameters, except for
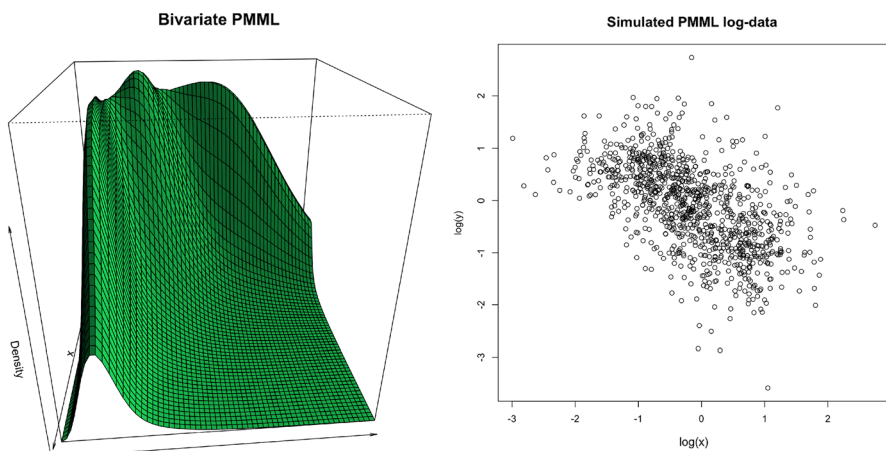
Fig. 4 Density and 1000 simulated data points from a power multivariate GMML distribution with negative correlation (true correlation of −0.32 and empirical of −0.33)

$$\boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 & 10 \\ 0 & 1 & 0 \\ 1/10 & 0 & 0 \end{pmatrix}.$$

Here, the correlation between Y_1 and Y_2 will be negative: if Y_i is drawn from the component with rate 10, then Y_j will be drawn from a component with rate 0.1, $i \neq j$. The marginal distributions are again given by (21) and (22) since the mixing probabilities are all equal. We observe how the sign of the correlation is affected by the structure of the matrix $\boldsymbol{D}_1$, and the fact that the matrices $\boldsymbol{C}_i$ are no longer of Erlang structure, the effect is qualitatively opposite to that of the bivariate ML case. One also sees that the class provides quite some flexibility in terms of the shape of the joint density function.

Remark 11 Dependence may often be constructed by introducing certain structures into the intensity matrices like in Example 1. More generally, dependence between several random variables of MPH* type may be constructed using the so-called Baker copula (Baker 2008), where order statistics are used and any feasible correlation structure can be obtained.

6 Conclusion

This paper introduces a class GMML of multivariate distributions with matrix Mittag–Leffler distributed marginals. With a construction essentially based on the multivariate phase-type distribution, the GMML class remains a flexible and tractable dense class of distributions maintaining a number of closed form properties. Two important sub-classes are considered, which lead to explicit formulas for distributional properties such as densities and fractional moments. This makes it an attractive candidate for the modelling of both theoretical and practical aspects of multivariate heavy-tailed risks, in situations with tail independence. The present construction cannot be extended to tail-dependent scenarios, so that other approaches will be needed for the latter, which will be an interesting topic for future research.

References

Albrecher, H., Beirlant, J., Teugels, J. L. (2017). *Reinsurance*: *Actuarial and statistical aspects*. Chichester: Wiley.

Albrecher, H., Bladt, M., Bladt, M. (2019). Matrix Mittag–Leffler distributions and modeling heavy-tailed risks. arXiv preprint arXiv:1906.05316.

Anderson, D. N. (1992). A multivariate Linnik distribution. *Statistics & Probability Letters*, *14*(4), 333–336.

Asmussen, S., Koole, G. (1993). Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability*, *30*(2), 365–372.

Baker, R. (2008). An order-statistics-based method for constructing multivariate distributions with fixed marginals. *Journal of Multivariate Analysis*, *99*(10), 2312–2327.

Basrak, B., Davis, R. A., Mikosch, T. (2002). A characterization of multivariate regular variation. *The Annals of Applied Probability*, *12*(3), 908–920.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. L. (2004). *Statistics of extremes*: *Theory and applications*. Chichester: Wiley.

Bladt, M., Nielsen, B. F. (2010). On the construction of bivariate exponential distributions with an arbitrary correlation coefficient. *Stochastic Models*, *26*(2), 295–308.

Bladt, M., Nielsen, B. F. (2017). *Matrix-exponential distributions in applied probability*. Berlin: Springer.

Bochner, S. (2005). *Harmonic analysis and the theory of probability*. New York: Dover Publications.

Falk, M., Padoan, S. A., Wisheckel, F. (2019). Generalized Pareto copulas: A key to multivariate extremes. *Journal of Multivariate Analysis*, *174*(104538), 17.

Gorenflo, R., Kilbas, A. A., Mainardi, F., Rogosin, S. V. (2014). *Mittag–Leffler functions, related topics and applications. Springer Monographs in Mathematics*. Berlin: Springer.

He, Q.-M., Zhang, H., Vera, J. C. (2012). On some properties of bivariate exponential distributions. *Stochastic Models*, *28*(2), 187–206.

Ho, Z. W. O., Dombry, C. (2019). Simple models for multivariate regular variation and the Hüsler–Reiß Pareto distribution. *Journal of Multivariate Analysis*, *173*, 525–550.

Joe, H., Li, H. (2011). Tail risk of multivariate regular variation. *Methodology and Computing in Applied Probability*, *13*(4), 671–693.

Kiriliouk, A., Rootzén, H., Segers, J., Wadsworth, J. L. (2019). Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics*, *61*(1), 123–135.

Kulkarni, V. G. (1989). A new class of multivariate phase type distributions. *Operations Research*, *37*, 151–158.

Lim, S. C., Teo, L. P. (2010). Analytic and asymptotic properties of multivariate generalized Linnik's probability densities. *The Journal of Fourier Analysis and Applications*, *16*(5), 715–747.

Mai, J.-F., Scherer, M. (2017). *Simulating copulas, volume 6 of Series in Quantitative Finance*. Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd.

McNeil, A. J., Frey, R., Embrechts, P. (2015). *Quantitative risk management. Princeton Series in Finance, revised edition*. Princeton, NJ: Princeton University Press.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, *9*(1), 3–20.

Neuts, M. (1979). A versatile Markovian point process. *Journal of Applied Probability*, *16*(4), 764–779.

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, *5*(4), 303–336.

Wan, P., Davis, R. A. (2019). Threshold selection for multivariate heavy-tailed data. *Extremes*, *22*(1), 131–166.