

CONTINUOUS SCALED PHASE-TYPE DISTRIBUTIONS

HANSJÖRG ALBRECHER, MARTIN BLADT, MOGENS BLADT, AND JORGE YSLAS

ABSTRACT. Products between phase-type distributed random variables and any independent, positive and continuous random variable are studied. Their asymptotic properties are established, and an expectation-maximization algorithm for their effective statistical inference is derived and implemented using real-world datasets. In contrast to discrete scaling studied in earlier literature, in the present continuous case closed-form formulas for various functionals of the resulting distributions are obtained, which facilitates both their analysis and implementation. The resulting mixture distributions are very often heavy-tailed and yet retain various properties of phase-type distributions, such as being dense (in weak convergence) on the set of distributions with positive support.

1. INTRODUCTION

A phase-type (PH) distributed random variable is defined as the time until absorption of a time-homogeneous pure-jump Markov process with a finite state-space having one absorbing state and all others transient. PH distributions are particularly attractive within applied probability and statistics since its members are explicitly described through functionals involving matrix exponentials, making this class both versatile and tractable, see [Bladt and Nielsen \(2017\)](#) for a comprehensive account of PH distributions. Furthermore, the PH class is known to be dense (in the sense of weak convergence) among the distributions concentrated on the positive half-line, so that one can approximate any non-negative random variable with arbitrary precision by a PH random variable. However, PH random variables are always light-tailed (have an exponentially bounded tail), and for many applications, this is too restrictive. In particular, in various application areas, there is a focus on modeling the tail, which is often heavier than exponential.

As an alternative, a class of discrete scaled PH distributions (NPH) was introduced in [Bladt et al. \(2015\)](#). The NPH class consists of distributions that can be expressed as the distribution of a product VY , where V is a discrete random variable, and Y is PH distributed. Parameter estimation for the NPH class was subsequently treated in [Bladt and Rojas-Nandayapa \(2018\)](#) via an expectation-maximization (EM) algorithm (see also [Asmussen et al. \(1996\)](#) for the unscaled case). Finally, in [Rojas-Nandayapa and Xie \(2018\)](#), the tail behavior in the general case, when the scale component is

Date: June 10, 2022.

1991 Mathematics Subject Classification. Primary: 60E05 Secondary: 60G70; 60J22; 62F10; 62N01; 62P05.

Key words and phrases. heavy tails; parameter estimation; phase-type; scale mixtures.

also allowed to be continuous, was studied in detail. In particular, the authors showed that a scaled PH distribution is heavy-tailed if and only if the scaling distribution is unbounded. Thus, the NPH class is of particular interest for modeling purposes due to its inherited denseness in the class of distributions on the positive half-line and its genuinely heavy tails (for a recent discussion of actuarial applications of discrete mixtures of exponential distributions, see also [Cossette et al. \(2021\)](#)). However, one downside of NPH distributions is that expressions such as the density and distribution functions are given in terms of infinite matrices, which in practice can only be computed up to a finite number of terms.

In this paper, we study the class of scaled PH distributions for the case when the scaling component is continuous. It turns out that under this construction, one can obtain closed-form formulas for different functionals. We provide new results on the tail behavior of these distributions, which complements the work in [Rojas-Nandayapa and Xie \(2018\)](#) and answers some open questions posed there. We then proceed to develop an EM algorithm for maximum-likelihood estimation of these models, which enables the use of this class of models on real-life data. We adapt the algorithm to the case of censored data, and we show how the EM algorithm can be employed to approximate a known given theoretical distribution. A crucial difference to the NPH arises as a by-product of the closed-form formulas that we obtain for the mixed distributions: the latter can be evaluated using functional calculus tools, which avoids the truncation of the infinite series of the NPH case. Recently, [Furman et al. \(2021\)](#) considered the scaling of PH distributions in a multivariate context with an interpretation of background risk models, in the light of which one might also see the present contribution as a theoretical underpinning of marginal properties of such models together with the development of an estimation algorithm for them when faced with real data.

However, although we are exclusively interested in products of random variables where at least one of them is PH distributed, much work has been done in more general probabilistic settings. We give a brief overview here. Breiman's lemma (cf. [Breiman \(1965\)](#)) and then extended by [Cline and Samorodnitsky \(1994\)](#)) established the regular variation of the product of two independent random variables where one is regularly varying and the other one has a moment condition. In [Embrechts and Goldie \(1980\)](#), the closure property was shown, namely that the product of regularly varying variables is again regularly varying with the same tail index, although no explicit asymptotics were provided. Subsequently, [Cline \(1986\)](#) (see also [Cline \(1987\)](#)) linked the asymptotics of products of independent random variables in terms of survival function ratios. More generally, [Cline and Samorodnitsky \(1994\)](#) studied the closure property of product convolutions within the subexponential class of distributions. Under more relaxed conditions, [Tang \(2006\)](#) established the closure of product convolutions for a slightly smaller class than the subexponential one. The latter author also considered necessary and sufficient conditions of the product distribution to be long-tailed when one of the component variables satisfies a generalization of lattice and long-tailed distributions. The multivariate subexponential case was studied in [Samorodnitsky and Sun \(2016\)](#).

Recently, [Xu et al. \(2017\)](#) found necessary and sufficient conditions for the subexponentiality of the product convolution, and provided a sufficient condition for the reverse problem: establishing the subexponentiality of a component given that of the product convolution. Finally, we would like to remark that the approach in the present paper is less general than the one in some of the papers above, but our focus is on a statistically tractable class that still allows for a fairly broad body and tail behavior of the distribution.

The rest of the paper is organized as follows. In [Section 2](#) we present an overview of the PH class and its most important properties for our purposes. In [Section 3](#) we introduce the class of continuous scaled PH distributions, provide new insights on their tail behavior ([Section 3.1](#)), and derive an EM algorithm for parameter estimation ([Section 3.2](#)) as well as an extension to the case of censored observations ([Section 3.3](#)). In [Section 4](#) we provide some numerical illustrations with real-life data and concerning the approximation of given distributions. Finally, [Section 5](#) concludes.

2. PRELIMINARIES ON PH DISTRIBUTIONS

This section presents the relevant preliminaries on PH distributions. For a random variable X , the notation $X \sim F$ for F being a distribution function, density, or acronym, is understood as X following the distribution uniquely associated with F . Unless stated otherwise, equalities between random objects hold almost surely. For two real-valued functions, g, h the terminology $g(t) \sim h(t)$, as $t \rightarrow \infty$ means that $\lim_{t \rightarrow \infty} g(t)/h(t) = 1$.

Let $(J_t)_{t \geq 0}$ denote a time-homogeneous Markov jump process on a state space $\{1, \dots, p, p+1\}$, where states $1, \dots, p$ are transient and state $p+1$ is absorbing. Then $(J_t)_{t \geq 0}$ has an intensity matrix of the form

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

where $\mathbf{T} = (t_{kl})_{k,l=1,\dots,p}$, is a $p \times p$ sub-intensity matrix, $\mathbf{t} = (t_1, \dots, t_p)^\top$ is a p -dimensional column vector, and $\mathbf{0}$ is the p -dimensional row vector of zeroes. Since rows of $\mathbf{\Lambda}$ sum to zero, we have that $\mathbf{t} = -\mathbf{T}\mathbf{e}$, where \mathbf{e} is the p -dimensional column vector of ones. Let $\pi_k = \mathbb{P}(J_0 = k)$, $k = 1, \dots, p$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)$ be the initial distribution of the chain, and assume that $\mathbb{P}(J_0 = p+1) = 0$, that is, the time until absorption is necessarily positive. Then we say that the time until absorption

$$Y = \inf\{t \geq 0 \mid J_t = p+1\}$$

has a phase-type distribution with representation $(\boldsymbol{\alpha}, \mathbf{T})$ and we write $Y \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{T})$. It can be shown that the density f_Y and distribution function F_Y for $Y \sim \text{PH}(\boldsymbol{\alpha}, \mathbf{T})$ are given by the closed-form expressions

$$\begin{aligned} f_Y(y) &= \boldsymbol{\pi} \exp(\mathbf{T}y)\mathbf{t}, & y > 0, \\ F_Y(y) &= 1 - \boldsymbol{\pi} \exp(\mathbf{T}y)\mathbf{e}, & y > 0 \end{aligned}$$

in terms of the matrix exponential of a matrix, defined as

$$\exp(\mathbf{M}) = \sum_{n=0}^{\infty} \frac{\mathbf{M}^n}{n!}.$$

More generally, if g is any analytic function and \mathbf{M} is a matrix, we may define $g(\mathbf{M})$ by Cauchy's formula, given by

$$g(\mathbf{M}) = \frac{1}{2\pi i} \oint_{\Gamma} g(z)(z\mathbf{I} - \mathbf{M})dz,$$

where Γ is a simple closed path in \mathbb{C} which encloses the eigenvalues of \mathbf{M} , see (Bladt and Nielsen, 2017, Section 3.4.) for details. More generally, we refer the reader to Doolittle (1998) for a condensed treatment of functions of matrices through the Cauchy integral, and Higham (2008) for other possible and equivalent methods.

The evaluation of such expressions can be done in several ways but is not always a straightforward task, especially for high-dimensional matrices, and complex functional analysis serves mostly as a mathematical tool. When performing estimation on real-world data, the resulting matrix is often diagonalizable. Hence, alternative methods such as diagonalization (and, more generally, the Jordan decomposition) of matrices are powerful tools in this context.

The fact that the underlying Markov chain is time-homogeneous has the consequence that the sojourn times, that is, the time spent in each state at each visit, are necessarily exponentially distributed. Furthermore, the tail $\bar{F} = 1 - F$ of a PH distribution is asymptotically exponential and has the following analytic expression:

$$(2.1) \quad \bar{F}(y) = \sum_{j=1}^m \sum_{k=0}^{\kappa_j-1} y^k \exp(\operatorname{Re}(-\lambda_j)y) [a_{jk} \sin(\operatorname{Im}(-\lambda_j)y) + b_{jk} \cos(\operatorname{Im}(-\lambda_j)y)],$$

where $-\lambda_j$ are the eigenvalues of the Jordan blocks \mathbf{J}_j of \mathbf{T} , with corresponding dimensions κ_j , $j = 1, \dots, m$, and a_{jk} and b_{jk} are constants depending on $\boldsymbol{\pi}$ and \mathbf{T} . If $-\lambda$ is the largest real eigenvalue of \mathbf{T} and n is the dimension of the Jordan block of λ , then it is easy to see from (2.1) that

$$(2.2) \quad \bar{F}(y) \sim cy^{n-1} \exp(-\lambda y), \quad y \rightarrow \infty,$$

where c is a positive constant. That is, all PH distributions have exponential tails with Erlang-like second-order bias terms. Consequently, the practical modeling of heavy tails using PH distributions can be problematic when the tail behavior is of interest. Nonetheless, the above formulas serve as a building block for the analysis that will follow in the sequel.

3. RANDOM SCALINGS OF PH RANDOM VARIABLES

In this section we consider a univariate PH random variable $Y \sim \text{PH}(\boldsymbol{\pi}, \mathbf{T})$ and define its randomly scaled counterpart

$$(3.1) \quad X := \frac{1}{\Theta} Y,$$

where Θ is some positive real-valued random variable, independent of Y . From the probabilistic construction of a PH random variable, it is clear that such a scaling can be conditionally realized by a transformation of the time axis, so that it can be subsumed through a modified sub-intensity matrix, that is,

$$(3.2) \quad X \mid \Theta = \theta \sim \text{PH}(\boldsymbol{\pi}, \theta \mathbf{T}).$$

for every realization of Θ . This simple observation allows us to now obtain the basic properties of random variables satisfying (3.1), stated in the following result. Here, and in what follows, we denote by

$$\mathcal{L}_\Theta(s) = \mathbb{E}[\exp(-s\Theta)], \quad s > 0,$$

the Laplace transform of Θ and by $\mathcal{L}'_\Theta(s)$ its corresponding derivative with respect to s .

In what follows, all considered functions are analytic in the region where the eigenvalues of $-\mathbf{T}$ lie, such that we may consider evaluating them at $-\mathbf{T}$, and more generally at $-\mathbf{T}x$ for any $x > 0$, without any concern. Recall also that for Laplace transforms, the analytic property in the domain of absolute convergence follows from Morera's theorem (cf Rudin (1987)).

Proposition 3.1. *Let X be given by (3.1). Then*

- (1) $F_X(x) = 1 - \boldsymbol{\pi} \mathcal{L}_\Theta(-\mathbf{T}x) \mathbf{e}, \quad x > 0.$
- (2) $f_X(x) = -\boldsymbol{\pi} \mathcal{L}'_\Theta(-\mathbf{T}x) \mathbf{t}, \quad x > 0.$
- (3) $\mathcal{L}_X(s) = \boldsymbol{\pi} \mathbb{E} \left[\left(\frac{s}{\Theta} \mathbf{I} - \mathbf{T} \right)^{-1} \right] \mathbf{t}, \quad s > 0.$
- (4) $\mathbb{E}(X^\nu) = \mathbb{E}(1/\Theta^\nu) \Gamma(\nu+1) \boldsymbol{\pi} (-\mathbf{T})^{-\nu} \mathbf{e},$ for $\nu \geq 0$, provided that it is well-defined.

Proof. (1) follows from (3.2), since one can obtain the tail of X as follows:

$$\begin{aligned} \overline{F}_X(x) &= \mathbb{P}(X > x) \\ &= \int_0^\infty \mathbb{P}(X > x \mid \Theta = \theta) dF_\Theta(\theta) \\ &= \boldsymbol{\pi} \int_0^\infty \exp(\theta \mathbf{T}x) dF_\Theta(\theta) \mathbf{e}. \end{aligned}$$

Taking derivatives in the above expression yields

$$f_X(x) = -\boldsymbol{\pi} \int_0^\infty \theta \mathbf{T} \exp(\theta \mathbf{T}x) dF_\Theta(\theta) \mathbf{e},$$

from which (2) easily follows. (3) follows from

$$\begin{aligned}\mathcal{L}_X(s) &= \mathbb{E}[\exp(-sX)] \\ &= \int_0^\infty \mathbb{E}[\exp(-(s/\theta)Y)]dF_\Theta(\theta) \\ &= \int_0^\infty \boldsymbol{\pi}((s/\theta)\mathbf{I} - \mathbf{T})^{-1}\mathbf{t}dF_\Theta(\theta).\end{aligned}$$

Finally (4) is a consequence of $\mathbb{E}(Y^\nu) = \Gamma(\nu + 1)\boldsymbol{\pi}(-\mathbf{T})^{-\nu}\mathbf{e}$, see [Bladt and Nielsen \(2017\)](#). \square

Remark 3.1. If Θ has an infinitely divisible distribution F_Θ , then

$$\mathcal{L}_\Theta(x) = \exp(-B(x))$$

for a unique Bernstein function B , see for instance [Schilling et al. \(2012\)](#), Prop.3.12. In this case, F_X has representation

$$F_X(x) = 1 - \boldsymbol{\pi} \exp(-B(-\mathbf{T}x))\mathbf{e}.$$

Here $-B(-\mathbf{T}x)$ is a sub-intensity matrix for all $x > 0$ (cf. [Berg et al. \(1993\)](#)). The corresponding density is then

$$f_X(x) = \boldsymbol{\pi} \exp(-B(-\mathbf{T}x))B'(-\mathbf{T}x)\mathbf{T}\mathbf{e} = -\boldsymbol{\pi} \exp(-B(-\mathbf{T}x))B'(-\mathbf{T}x)\mathbf{t}.$$

The above formulas show that for a fairly broad class of mixing distributions we obtain a closed-form formula in terms of Bernstein functions for the resulting CPH distribution, which can be computationally advantageous.

Remark 3.2. Scaling a PH distribution with a variable whose law is of unbounded support always results in a heavy-tailed distribution (see the next subsection for a formal definition of heavy-tailedness). However, the precise nature of the tail asymptotics, including higher-order expansions, can be calculated by applying a Jordan normal form expansion in Property (1) above. Note that the resulting behavior is solely determined by the functional form of the Laplace transform of Θ and the eigenvalues of \mathbf{T} .

Thus, we make the following formal definition of such an X in terms of scaling.

Definition 3.2. A random variable X is said to have a *continuous scaled phase-type* (CPH) distribution with representation $(\boldsymbol{\pi}, \mathbf{T})$ and scaling variable Θ if its distribution function is given by

$$F_X(x) = 1 - \boldsymbol{\pi}\mathcal{L}_\Theta(-\mathbf{T}x)\mathbf{e}, \quad x > 0,$$

in which case X is the product of $1/\Theta$ and an independent $\text{PH}(\boldsymbol{\pi}, \mathbf{T})$ variable. We write $X \sim \text{CPH}(\boldsymbol{\pi}, \mathbf{T}, \Theta)$.

Theorem 3.3. Let \mathcal{C} be a family of positive random variables in which we may find a sequence which degenerates weakly into a positive constant $c > 0$. Then the class

$$\mathcal{D} = \{Y/\Theta \mid \Theta \in \mathcal{C}, Y \sim \text{PH}\}$$

of CPH distributions is weakly dense in the class of distributions on the positive half-line. In other words, for any given positive random variable Z , we may find a sequence $\{X_n\}_{n=1}^\infty \subset \mathcal{D}$ such that

$$X_n \xrightarrow{d} Z.$$

Proof. The proof is a simple application of convergence through the diagonal of an array. For instance, by choosing a sequence of scaling random variables Θ_n with constant mean k and variances shrinking to zero, we may then use the corresponding denseness property of regular PH distributions to find a suitable sequence $\{Y_n\}_{n=1}^\infty$ of PH distributed random variables with $Y_n \rightarrow Z$, which in particular establishes the required convergence of the $X_n = Y_n/\Theta_n$. Alternatively, one can adapt along the same lines the elementary proof of the denseness of mixtures of Erlang distributions (which are a special case of PH distributions) to the continuously scaled case. \square

We now present some examples where the distribution of X has an explicit expression. This is not only of mathematical interest, but also of practical relevance when considering their evaluation and estimation.

Example 3.1 (Gamma mixing). Consider Gamma mixing with $\Theta \sim \text{Gamma}(\alpha, 1)$ for any positive shape parameter α . Then

$$(3.3) \quad \overline{F}_X(x) = \boldsymbol{\pi}(\mathbf{I} - x\mathbf{T})^{-\alpha} \mathbf{e}, \quad x > 0,$$

which is well-defined since $x \cdot \text{Re}(\lambda) < 1$ holds for any $x > 0$, and λ any eigenvalue of \mathbf{T} . Thus $\mathbf{I} - x\mathbf{T}$ is invertible. We also have that

$$f_X(x) = \alpha \boldsymbol{\pi}(\mathbf{I} - x\mathbf{T})^{-\alpha-1} \mathbf{t}, \quad x > 0.$$

We call this distribution *matrix-Pareto type II* to distinguish it from the matrix-Pareto distribution introduced in [Albrecher and Bladt \(2019\)](#). Note that there is a fundamental difference to the latter in that the tail behavior is specified by a scalar (shape) parameter α , and the scale is determined by the matrix \mathbf{T} . In contrast, the matrix-Pareto distribution in [Albrecher and Bladt \(2019\)](#) has a tail behavior that depends on the eigenvalues of the matrix \mathbf{T} (arising from its matrix-valued shape parameter) and a scalar scale parameter.

Observe that mixing with more general $\Theta \sim \text{Gamma}(\alpha, \beta)$ results in the same class, since $\beta > 0$ is a scale parameter and PH distributions are closed with respect to deterministic scalings. Consequently, from this CPH class, we may obtain the classical PH class when degenerating the Gamma mixing distribution into any positive point mass (by keeping α/β constant and letting $\alpha, \beta \rightarrow \infty$), i.e., the so-called Erlangization. The latter property is of particular interest when using this class of distributions for modeling purposes. \square

Example 3.2 (Positive stable mixing). Consider Lévy mixing with parameter $\eta > 0$ and density

$$f_\Theta(\theta) = \frac{\eta}{2\sqrt{\pi}\theta^3} \exp(-\eta^2/(4\theta)), \quad \theta > 0,$$

which is a positive stable distribution with stability parameter $1/2$, then

$$\begin{aligned}\bar{F}_X(x) &= \boldsymbol{\pi} \mathcal{L}_\Theta(-\mathbf{T}x) \mathbf{e} \\ &= \int_0^\infty \boldsymbol{\pi} \exp(\theta \mathbf{T}x) \mathbf{e} \frac{\eta}{2\sqrt{\pi\theta^3}} \exp(-\eta^2/(4\theta)) d\theta \\ &= \boldsymbol{\pi} \exp\left(\eta \left(-\sqrt{-\mathbf{T}}\right) \sqrt{x}\right) \mathbf{e}.\end{aligned}$$

The matrix $-\sqrt{-\mathbf{T}}$ is again a sub-intensity matrix, see (Higham, 2008, Page 160), so $(\boldsymbol{\pi}, -\sqrt{-\mathbf{T}})$ can be seen as the parameters of another PH distribution.

More generally, let us consider a positive stable random variable Θ with corresponding Laplace transform $\mathcal{L}_\Theta(s) = \exp(-s^\alpha)$, where $\alpha \in (0, 1]$. Then,

$$\bar{F}_X(x) = \boldsymbol{\pi} \exp(-(-\mathbf{T})^\alpha x^\alpha) \mathbf{e}.$$

Again, $(\boldsymbol{\pi}, -(-\mathbf{T})^\alpha)$ are parameters corresponding to another PH distribution. Such distributions have a Weibull-type tail behavior, and they span the same class as matrix-Weibull laws for $\alpha \in (0, 1]$ as discussed in (Albrecher and Bladt, 2019, Section 4.1). Note, however, that the two parametrizations differ, and for a general α , the underlying sub-intensity structure is not maintained. \square

3.1. Tail behavior of scaled PH distributions. The tail behavior of the CPH class can be studied by exploiting its representation (3.1). In this section we use as before the notation $V = 1/\Theta$. Some asymptotics of this type of construction were treated in Rojas-Nandayapa and Xie (2018). We now recall some of these results, extend them, and provide a counter-example for a conjecture posed there.

One of the main observations in Rojas-Nandayapa and Xie (2018) is that if V is unbounded, then X is heavy-tailed in the sense that $\limsup_{x \rightarrow \infty} \bar{F}_X(x) \exp(\epsilon x) = \infty$ for all $\epsilon > 0$. Perhaps the most well-known class of heavy-tailed distributions is the regularly varying class denoted by \mathcal{R} . Recall that a random variable Z and its distribution function F_Z are called regularly varying with index $\alpha > 0$ (we write $F_Z \in \mathcal{R}_{-\alpha}$) if

$$\bar{F}_Z(x) = x^{-\alpha} L(x), \quad x > 0,$$

where L is a slowly varying function, that is, $L(cx)/L(x) \rightarrow 1$ as $x \rightarrow \infty$ for all $c > 0$.

A standard result to obtain the tail behavior of a product of independent random variables, when one of the components is regularly varying, is Breiman's lemma (Breiman, 1965):

Lemma 3.4. *Let $F_Z \in \mathcal{R}_{-\alpha}$, and assume that $\mathbb{E}[Y^{\alpha+\epsilon}] < \infty$, for some $\epsilon > 0$. Then $F_{ZY} \in \mathcal{R}_{-\alpha}$, and*

$$\bar{F}_{ZY}(x) \sim \mathbb{E}[Y^\alpha] \bar{F}_Z(x), \quad x \rightarrow \infty.$$

Example 3.3. For the Matrix-Pareto type II distribution with tail (3.3) obtained by Gamma mixing, we have that $1/\Theta \in \mathcal{R}_{-\alpha}$, so that $\boldsymbol{\pi}(\mathbf{I} - x\mathbf{T})^{-\alpha} \mathbf{e} \sim Cx^{-\alpha}$ as $x \rightarrow \infty$. One can, of course, also prove this directly by decomposing the matrix into Jordan blocks. \square

Example 3.4. If $\Theta \sim \text{PH}(\boldsymbol{\pi}_0, \mathbf{T}_0)$, we can show from straightforward calculations and using (2.1) that $1/\Theta$ has tail behavior

$$\bar{F}_{1/\Theta}(x) \sim \sum_{j=1}^m \sum_{k=0}^{\kappa_j-1} \frac{1}{x^k} b_{jk}^0, \quad x \rightarrow \infty,$$

where b_{jk}^0 are real-valued constants. This shows that $1/\Theta \in \mathcal{R}_{-\alpha}$ for some non-negative integer α . In particular, the ratio of any two PH distributed random variables is always regularly varying of integer order. \square

We say that Y is α -regular variation determining (α -rvd) if $F_Z \in \mathcal{R}_{-\alpha}$ whenever $F_{YZ} \in \mathcal{R}_{-\alpha}$. It has been noted in [Rojas-Nandayapa and Xie \(2018\)](#) that special cases of PH distributions are α -rvd. A converse of Breiman's lemma was given in [Jacobsen et al. \(2009\)](#) as follows.

Proposition 3.5. *Let Y be a positive random variable with $\mathbb{E}[Y^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$. Then Y is α -rvd if and only if*

$$\mathbb{E}[Y^{\alpha+i\eta}] \neq 0, \quad \eta \in \mathbb{R},$$

where $\alpha + i\eta \in \mathbb{C}$, that is, $i^2 = -1$.

We can translate the above condition into the PH setting as follows.

Proposition 3.6. *Let $Y \sim \text{PH}(\boldsymbol{\pi}, \mathbf{T})$. Then Y is α -rvd for $\alpha > 0$ if and only if*

$$(3.4) \quad \boldsymbol{\pi}(-\mathbf{T}^{-1})^{\alpha+i\eta} \mathbf{e} \neq 0, \quad \eta \in \mathbb{R}.$$

Proof. First observe that the function

$$w(x) = x^{\alpha+i\eta}, \quad x > 0,$$

has a well-defined Laplace transform for any $\alpha > 0$, given by

$$\mathcal{L}_w(s) = \int_0^\infty \exp(-sx) x^{(\alpha+i\eta+1)-1} dx = \frac{\Gamma(\alpha + i\eta + 1)}{s^{\alpha+i\eta+1}}, \quad s > 0.$$

Consequently, by functional calculus,

$$\mathbb{E}[w(Y)] = \boldsymbol{\pi} \mathcal{L}_w(-\mathbf{T}) \mathbf{t} = \Gamma(\alpha + i\eta + 1) \boldsymbol{\pi}(-\mathbf{T}^{-1})^{\alpha+i\eta} \mathbf{e}.$$

Since the Γ -function never vanishes, the result follows. \square

It was conjectured in [Rojas-Nandayapa and Xie \(2018\)](#) that any PH distribution is α -rvd for any $\alpha > 0$. Here we provide a counterexample.

Example 3.5 (Hyperexponential). Consider the following sub-intensity matrix corresponding to a hyperexponential PH distribution

$$\mathbf{T} = \begin{pmatrix} -1 & 0 \\ 0 & -\exp(-\pi) \end{pmatrix}.$$

Then its associated Green matrix is given by

$$\mathbf{U} = -\mathbf{T}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \exp(\pi) \end{pmatrix},$$

and thus we may easily calculate its complex power as (cf. [Doolittle \(1998\)](#))

$$\mathbf{U}^{\alpha+i} = \begin{pmatrix} 1 & 0 \\ 0 & \exp((\alpha+i)\pi) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\exp(\alpha\pi) \end{pmatrix}.$$

Therefore

$$\mathbf{U}^{\alpha+i}\mathbf{e} = \begin{pmatrix} 1 \\ -\exp(\alpha\pi) \end{pmatrix},$$

and defining the initial vector as

$$\boldsymbol{\pi} = \left(\frac{\exp(\alpha\pi)}{1 + \exp(\alpha\pi)}, 1 - \frac{\exp(\alpha\pi)}{1 + \exp(\alpha\pi)} \right),$$

we then have that

$$\boldsymbol{\pi}\mathbf{U}^{\alpha+i}\mathbf{e} = 0.$$

In fact, we may always construct a PH distribution such that the property (3.4) fails for a pre-specified $\eta \in \mathbb{R}$. This is achieved by considering

$$\mathbf{T} = \begin{pmatrix} -1 & 0 \\ 0 & -\exp(-\pi/\eta) \end{pmatrix}$$

and

$$\boldsymbol{\pi} = \left(\frac{\exp(\alpha\pi/\eta)}{1 + \exp(\alpha\pi/\eta)}, 1 - \frac{\exp(\alpha\pi/\eta)}{1 + \exp(\alpha\pi/\eta)} \right),$$

and again $\boldsymbol{\pi}\mathbf{U}^{\alpha+i\eta}\mathbf{e} = 0$. □

Example 3.6. We now provide a concrete example of an explicit distribution which multiplied by a hyperexponential is regularly varying but itself is not. Consider Y following a hyperexponential PH distribution with parameters

$$\boldsymbol{\pi} = \left(\frac{\exp(\pi)}{1 + \exp(\pi)}, 1 - \frac{\exp(\pi)}{1 + \exp(\pi)} \right),$$

and

$$\mathbf{T} = \begin{pmatrix} -1 & 0 \\ 0 & -\exp(-\pi) \end{pmatrix}.$$

Now, take V with tail function given by

$$\bar{F}_V(x) = \left(1 + \frac{1}{2} \sin(\log(x)) \right) x^{-1}.$$

Note that this distribution is not regularly varying. Then the product VY has tail distribution

$$\bar{F}_{VY}(x) = \mathbb{P}(VY > x) = x^{-1} \int_0^\infty y f_Y(y) dy = x^{-1} \boldsymbol{\pi}(-\mathbf{T})^{-1}\mathbf{e},$$

meaning that \bar{F}_{VY} is regularly varying with index -1 . In fact, the above construction can be seen as a special case of Example 5 in [Maulik and Resnick \(2004\)](#). Note that

this construction hinges on having at least two mixing components, in line with the previous example.

3.1.1. *Subexponentiality and other classes of heavy-tailed distributions.* We now study other types of tail behavior, such as Weibull-type and lognormal-type, which fall into the Gumbel max-domain of attraction. We consider the following two examples.

Definition 3.7 (Weibull-type tails). A distribution function F is in the Weibull-type class if

$$\bar{F}(x) \sim cx^\beta \exp(-\lambda x^\tau), \quad x \rightarrow \infty,$$

for some constants $\beta \in \mathbb{R}$ and $\tau, \lambda, c > 0$. A Weibull-type distribution is heavy-tailed if $\tau \in (0, 1)$ and light-tailed otherwise. If V is Weibull-type with parameter $\tau > 0$, then $X = VY$ is Weibull-type with parameter $\tau/(\tau + 1)$, see [Arendarczyk et al. \(2011\)](#). In particular, and in contrast to the regularly varying case, random scaling results in a distribution with a heavier tail than each of the two components (but still of Weibull-type).

Definition 3.8 (Lognormal-type tails). A distribution function F is in the lognormal-type class if

$$\bar{F}(x) \sim cx^\beta (\log x)^\xi \exp(-\lambda (\log x)^\gamma), \quad x \rightarrow \infty,$$

for some constants $\beta, \xi \in \mathbb{R}$, $\gamma > 1$ and $\lambda, c > 0$, and we write $F \in \text{LN}(\gamma)$. In particular, the lognormal distribution belongs to $\text{LN}(2)$. In [Rojas-Nandayapa and Xie \(2018\)](#), it was shown that if V has a standard lognormal tail, then

$$\frac{\mathbb{P}(VY > x)}{\mathbb{P}(V > x)} \rightarrow \infty, \quad x \rightarrow \infty,$$

and $X = VY$ is subexponential (see below).

The regularly varying class, the Weibull-type class (with $\tau \in (0, 1)$) and the lognormal-type class are subclasses of the so-called subexponential class \mathcal{S} . Recall that $F \in \mathcal{S}$ if $\lim_{x \rightarrow \infty} \bar{F}^{*2}(x)/\bar{F}(x) = 2$, where $F^{*2} = F * F$ denotes the 2-fold convolution of F . The property can be shown to extend to n -fold convolutions, $n \in \mathbb{N}$. We now focus on finding more general conditions under which a CPH distribution has subexponential tail behavior. The following result provides sufficient conditions for subexponentiality of the product of independent random variables.

Theorem 3.9 (Theorem 2.1 in [Cline and Samorodnitsky \(1994\)](#)). *Let V and Y be independent non-negative random variables with distribution functions F_V and F_Y , respectively, and with Y not degenerate at zero. Let F_{VY} be the distribution of the product VY . Assume that $F_V \in \mathcal{S}$. If there is a function $a : (0, \infty) \rightarrow (0, \infty)$ such that:*

- (1) $a(x) \uparrow \infty$ as $x \rightarrow \infty$;
- (2) $x/a(x) \uparrow \infty$ as $x \rightarrow \infty$;
- (3) $\lim_{x \rightarrow \infty} \bar{F}_V(x - a(x))/\bar{F}_V(x) = 1$;

$$(4) \lim_{x \rightarrow \infty} \overline{F}_Y(a(x)) / \overline{F}_{VY}(x) = 0;$$

then $F_{VY} \in \mathcal{S}$.

Remark 3.3. A distribution function F that satisfies $\overline{F}(x \pm h(x)) \sim \overline{F}(x)$ as $x \rightarrow \infty$ is called h -insensitive. Thus, Condition (3) in Theorem 3.9 translates to F_V being a -insensitive, where $a(\cdot)$ satisfies Conditions (1) and (2). We refer to Foss et al. (2011) for further reading on h -insensitive distribution functions.

The following Corollary provides sufficient conditions for subexponential behavior of scaled PH distributions.

Corollary 3.10. *Let V be a non-negative random variable with distribution function $F_V \in \mathcal{S}$, and let $Y \sim PH(\boldsymbol{\pi}, \mathbf{T})$ independent of V . Let $a : (0, \infty) \rightarrow (0, \infty)$ be a function such that:*

- (1) $a(x) \uparrow \infty$ as $x \rightarrow \infty$;
- (2) $x/a(x) \uparrow \infty$ as $x \rightarrow \infty$;
- (3) $\sqrt{x}/a(x) \rightarrow 0$ as $x \rightarrow \infty$;
- (4) $\lim_{x \rightarrow \infty} \overline{F}_V(x - a(x)) / \overline{F}_V(x) = 1$.

Then $X = VY$ has subexponential tail behavior.

Proof. Since $a(\cdot)$ satisfies Conditions (1)-(3) of Theorem 3.9, it remains to show that Condition (4) is also satisfied. Using $\overline{F}_V(x^{1/2}) \overline{F}_Y(x^{1/2}) \leq \overline{F}_{VY}(x)$ for all $x > 0$, (2.2), and $x^{1/2}/a(x) \rightarrow 0$ as $x \rightarrow \infty$, we have that

$$\lim_{x \rightarrow \infty} \frac{\overline{F}_Y(a(x))}{\overline{F}_{VY}(x)} \leq \lim_{x \rightarrow \infty} \frac{1}{cx^{(n-1)/2}} \frac{\exp(-\lambda_0 x^{1/2})}{\overline{F}_V(x^{1/2})}$$

for any $0 < \lambda_0 < \lambda$. Since $F_V \in \mathcal{S}$, then $\exp(\epsilon x) \overline{F}_V(x) \rightarrow \infty$ as $x \rightarrow \infty$ for all $\epsilon > 0$, see (Foss et al., 2011, Lemma 2.17). Thus, we conclude that

$$\lim_{x \rightarrow \infty} \frac{\overline{F}_Y(a(x))}{\overline{F}_{VY}(x)} = 0.$$

It follows then, by Theorem 3.9, that the product VY has subexponential tail behavior. \square

Remark 3.4. Note that the conditions on V above are solely in terms of its tail behavior, and even though they seem rather restrictive, they are satisfied by several classes of distributions, including: intermediate regularly varying distributions, see (Foss et al., 2011, Theorem 2.47), Weibull-type distributions with $\tau \in (0, 0.5)$ and lognormal-type distributions with $\gamma > 1$. Note also that $a(x) = M\sqrt{x}$ for some $M > 1$ is possible, and then the class of distributions that satisfies the conditions of Corollary 3.10 is closely related to the \sqrt{x} -insensitive class of distributions, which can be characterized by means of a convergence on probability, see (Foss et al., 2011, Theorem 2.49).

We now address different subclasses of \mathcal{S} , and we start with the \mathcal{A} class of distributions. Recall that F belongs to the \mathcal{A} class if $F \in \mathcal{S}$ and

$$(3.5) \quad \limsup_{x \rightarrow \infty} \frac{\overline{F}(cx)}{\overline{F}(x)} < 1$$

for some $c > 1$.

Remark 3.5. Condition (3.5) is a mild restriction, and it is satisfied by a vast number of distributions, including those in Table 1.2.6 in Embrechts et al. (2013), making this a relevant subclass of \mathcal{S} .

Note that, in particular, $\mathcal{A} \subset \mathcal{S}$, and both are subclasses of the more general class of long-tailed distributions \mathcal{L} (see the appendix for a definition). The following proposition, which characterizes the tail behavior of a scaled phase-type distribution when the scaling is long-tailed, is key for deriving more explicit tail asymptotics when dealing with different subclasses of \mathcal{S} .

Proposition 3.11. *Let V be a non-negative random variable with distribution function $F_V \in \mathcal{L}$ and let $Y \sim PH(\boldsymbol{\pi}, \mathbf{T})$, independent of V . Then $\overline{F}_Y(x) = o(\overline{F}_{VY}(bx))$ for all $b > 0$.*

Proof. Let $b > 0$. Since $Y \sim PH(\boldsymbol{\pi}, \mathbf{T})$, we have that (2.2) holds. Using $\overline{F}_V(\lambda_0 x) \overline{F}_Y(b/\lambda_0) \leq \overline{F}_{VY}(bx)$ for all $x > 0$, we have that

$$\lim_{x \rightarrow \infty} \frac{\overline{F}_Y(x)}{\overline{F}_{VY}(bx)} \leq \lim_{x \rightarrow \infty} \frac{\exp(-\lambda_0 x)}{\overline{F}_V(\lambda_0 x) \overline{F}_Y(b/\lambda_0)}$$

for any $0 < \lambda_0 < \lambda$. Given that $F_V \in \mathcal{L}$, then $\exp(\epsilon x) \overline{F}_V(x) \rightarrow \infty$ as $x \rightarrow \infty$ for all $\epsilon > 0$, which implies the result. \square

The next result provides sufficient conditions for the tail behavior in the \mathcal{A} class of the product of independent random variables.

Theorem 3.12 (Theorem 2.1 in Tang (2006)). *Let V and Y be independent non-negative random variables with distribution functions F_V and F_Y , respectively, and let F_{VY} be the distribution function of the product VY . If $F_V \in \mathcal{A}$ and $\overline{F}_Y(x) = o(\overline{F}_{VY}(bx))$ for all $b > 0$, then $F_{VY} \in \mathcal{A}$.*

As an immediate consequence, we obtain the following result for scaled phase-type distributions.

Corollary 3.13. *Let V be a non-negative random variable with distribution function F_V and let $Y \sim PH(\boldsymbol{\pi}, \mathbf{T})$, independent of V . If $F_V \in \mathcal{A}$, then $F_{VY} \in \mathcal{A}$.*

Proof. Follows directly from Theorem 3.12, Proposition 3.11 and using that $\mathcal{A} \subset \mathcal{L}$. \square

A necessary and sufficient condition for subexponentiality of the product was more recently derived in Xu et al. (2017). Denote by $D[F]$ the set of all positive discontinuities of F .

Theorem 3.14 (Theorem 1.2 in [Xu et al. \(2017\)](#)). *Let V and Y be independent non-negative random variables with distribution functions F_V and F_Y , respectively, and let F_{VY} be the distribution of the product VY . Then $F_{VY} \in \mathcal{S}$ if and only if $F_V \in \mathcal{S}$ and either $D[F_V] = \emptyset$ or $D[F_V] \neq \emptyset$ and*

$$(3.6) \quad \overline{F}_Y(x/d) - \overline{F}_Y((x+1)/d) = o(\overline{F}_{VY}(x)),$$

for all $d \in D[F_V]$.

Remark 3.6. In particular, the result above implies that when the scaling component in a scaled phase-type random variable is a continuous random variable with subexponential tail (thus, falling in the CPH class), the resulting distribution is subexponential as well. However, note that when dealing with discrete scaling, one should proceed more carefully since Condition (3.6) is, in general, not easy to verify. Thus, the different criteria provided here are useful tools to determine the subexponentiality of the product, and the use of each of them depends on each particular case.

We now consider other subclasses of \mathcal{S} . More specifically, we consider the classes of *Extended regularly varying* (\mathcal{E}), *Intermediate regularly varying* (\mathcal{I}), and *Dominated varying* (\mathcal{D}) distributions. These classes satisfy $\mathcal{R} \subset \mathcal{E} \subset \mathcal{I} \subset \mathcal{D}$ and $\mathcal{I} \subset (\mathcal{D} \cap \mathcal{L}) \subset \mathcal{S} \subset \mathcal{L}$. We refer to the appendix for their definition.

The specific tail behavior of a PH distribution now yields the following results.

Corollary 3.15. *Let V be a non-negative random variable with distribution function F_V and let $Y \sim PH(\boldsymbol{\pi}, \mathbf{T})$ independent of V . Then we have that*

- (1) *If $F_V \in \mathcal{R}$ then $F_{VY} \in \mathcal{R}$ and their indices of regular variation are the same.*
- (2) *If $F_V \in \mathcal{E}$ then $F_{VY} \in \mathcal{E}$.*
- (3) *If $F_V \in \mathcal{I}$ then $F_{VY} \in \mathcal{I}$.*
- (4) *If $F_V \in \mathcal{D} \cap \mathcal{L}$ then $F_{VY} \in \mathcal{D} \cap \mathcal{L}$.*
- (5) *If $F_V \in \mathcal{L}$ then $F_{VY} \in \mathcal{L}$.*

Proof. Note that for all cases $F_V \in \mathcal{L}$, then Proposition 3.11 implies that $\overline{F}_Y(x) = o(\overline{F}_{VY}(bx))$ for all $b > 0$. Thus,

- (1) Follows from ([Cline and Samorodnitsky, 1994](#), Corollary 3.6 (ii)) (and is in fact also a direct consequence of Breiman's lemma).
- (2) Follows from ([Cline and Samorodnitsky, 1994](#), Theorem 3.5 (iii)).
- (3) Follows from ([Cline and Samorodnitsky, 1994](#), Theorem 3.4 (ii)).
- (4) Follows from ([Cline and Samorodnitsky, 1994](#), Theorem 3.3 (ii) and Theorem 2.2 (iii)).
- (5) Follows from ([Cline and Samorodnitsky, 1994](#), Theorem 2.2 (iii)).

□

3.2. An EM algorithm for CPH distributions. We present an EM algorithm for estimating the CPH class of distributions when Θ is any positive, continuous random variable. Assume that Θ belongs to a parametric family depending on the vector $\boldsymbol{\alpha}$ and denote by $f_{\Theta}(\cdot; \boldsymbol{\alpha})$ its corresponding density function.

As is usual, we will exploit the path representation of PH distributions seen as absorption times of a finite-state Markov jump process. Thus, the complete data are the entire paths of the Markov jump processes and the scaling component Θ . In this setting, not only do we not observe the state sojourn times and transitions, but the realizations of Θ are also not observed.

Consider x_1, \dots, x_M an iid sample from a CPH distributed random variable, and let $L_c(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\alpha}; \mathbf{x})$ denote the corresponding complete data likelihood function. In order to write out the latter explicitly, we need to make the following definitions. Let B_k be the number of times the underlying Markov jump processes start in state k , N_{kl} the total number of transitions from state k to l until absorption, N_k the number of times that k was the last state to be visited before absorption, and finally let Z_k be the cumulated time that the Markov jump processes spent in state k .

Then, using $f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$ and conditioning on the paths of the Markov jump processes

$$\begin{aligned} L_c(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\alpha}; \mathbf{x}) &= f_{\Theta}(\theta; \boldsymbol{\alpha}) \prod_{k=1}^p \pi_k^{B_k} \prod_{k=1}^p \prod_{l=1, l \neq k}^p (\theta t_{kl})^{N_{kl}} \exp(-\theta t_{kl} Z_k) \\ &\quad \times \prod_{k=1}^p (\theta t_k)^{N_k} \exp(-\theta t_k Z_k). \end{aligned}$$

Consequently, the corresponding log-likelihood (disregarding the terms which do not depend on any parameters) is given by

$$\begin{aligned} l_c(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\alpha}; \mathbf{x}) &= \sum_{k=1}^p B_k \log(\pi_k) + \sum_{k=1}^p \sum_{l=1, l \neq k}^p N_{kl} \log(t_{kl}) - \sum_{k=1}^p \sum_{l=1, l \neq k}^p t_{kl} \theta Z_k \\ &\quad + \sum_{k=1}^p N_k \log(t_k) - \sum_{k=1}^p t_k \theta Z_k + \log(f_{\Theta}(\theta; \boldsymbol{\alpha})). \end{aligned}$$

With this decomposition of the full likelihood at hand, we now outline the necessary computations that are needed in each of the two steps of the EM algorithm, and then collect the main formulas at the end.

E-Step

This step consists of computing the conditional expectation of the complete data log-likelihood given the observed data, and given some fixed parameters. We consider one

(generic) data point ($M = 1$) and let $x = x_1$. Then

$$\begin{aligned}
\mathbb{E} [\log(f_{\Theta}(\Theta; \boldsymbol{\alpha})) \mid X = x] &= \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) f_{\Theta|X}(\theta|x) d\theta \\
&= \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} d\theta \\
&= \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{f_X(x)} d\theta \\
&= \frac{\int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) \boldsymbol{\pi} \exp(\theta \mathbf{T} x) \theta \mathbf{t} f_{\Theta}(\theta) d\theta}{f_X(x)}.
\end{aligned}$$

Regarding B_k ,

$$\begin{aligned}
\mathbb{E} [B_k \mid X = x] &= \mathbb{P}(J_0 = k \mid X = x) \\
&= \int_0^{\infty} \mathbb{P}(J_0 = k, \Theta \in d\theta \mid X = x) d\theta \\
&= \int_0^{\infty} \frac{\mathbb{P}(J_0 = k, \Theta \in d\theta, X \in dx)}{\mathbb{P}(X \in dx)} d\theta \\
&= \frac{\int_0^{\infty} \mathbb{P}(X \in dx \mid J_0 = k, \Theta = \theta) \mathbb{P}(J_0 = k \mid \Theta = \theta) f_{\Theta}(\theta) d\theta}{\mathbb{P}(X \in dx)} \\
&= \frac{\int_0^{\infty} \pi_k \mathbf{e}_k^{\top} \exp(\theta \mathbf{T} x) \theta \mathbf{t} f_{\Theta}(\theta) d\theta}{f_X(x)},
\end{aligned}$$

where \mathbf{e}_k denotes a p -dimensional column vector with all entries equal to zero except the k -th entry, which equals one.

For the term which involves the product of Θ and Z_k we will make use of the tower property of conditional expectations, namely that

$$\mathbb{E} [\Theta Z_k \mid X = x] = \mathbb{E} (\Theta \mathbb{E} (Z_k \mid \Theta, X = x) \mid X = x).$$

Since

$$\begin{aligned}
\mathbb{E} [Z_k \mid \Theta = \theta, X = x] &= \mathbb{E} \left[\int_0^{\infty} 1_{\{J_u = k\}} du \mid \Theta = \theta, X = x \right] \\
&= \int_0^{\infty} \mathbb{P}(J_u = k \mid \Theta = \theta, X = x) du \\
&= \frac{\int_0^{\infty} \mathbb{P}(J_u = k, \Theta \in d\theta, X \in dx) du}{\mathbb{P}(\Theta \in d\theta, X \in dx)},
\end{aligned}$$

then

$$\begin{aligned}
\mathbb{E} [\Theta Z_k \mid X = x] &= \int_0^{\infty} \theta \frac{\int \mathbb{P}(J_u = k, \Theta \in d\theta, X \in dx) du}{\mathbb{P}(\Theta \in d\theta, X \in dx)} \mathbb{P}(\Theta \in d\theta \mid X = x) d\theta \\
&= \int_0^{\infty} \theta \frac{\int \mathbb{P}(J_u = k, \Theta \in d\theta, X \in dx) du}{\mathbb{P}(X \in dx)} d\theta
\end{aligned}$$

$$\begin{aligned}
&= \frac{\int_0^\infty \theta \int_0^\infty \mathbb{P}(X \in dx \mid J_u = k, \Theta = \theta) \mathbb{P}(J_u = k \mid \Theta = \theta) f_\Theta(\theta) du d\theta}{\mathbb{P}(X \in dx)} \\
&= \frac{\int_0^\infty \theta \int_0^x \mathbf{e}_k^\top \exp(\theta \mathbf{T}(x-u)) \theta \mathbf{t} \boldsymbol{\pi} \exp(\theta \mathbf{T}u) \mathbf{e}_k du f_\Theta(\theta) d\theta}{f_X(x)}.
\end{aligned}$$

Similarly, one finds that

$$\mathbb{E}[N_{kl} \mid X = x] = \frac{\int \theta t_{kl} \int_0^x \mathbf{e}_l^\top \exp(\theta \mathbf{T}(x-u)) \theta \mathbf{t} \boldsymbol{\pi} \exp(\theta \mathbf{T}u) \mathbf{e}_k du f_\Theta(\theta) d\theta}{f_X(x)},$$

and

$$\mathbb{E}[N_k \mid X = x] = \frac{\int_0^\infty \theta t_k \boldsymbol{\pi} \exp(\theta \mathbf{T}x) \mathbf{e}_k f_\Theta(\theta) d\theta}{f_X(x)}.$$

For $M > 1$, we simply sum over x_i , $i = 1, \dots, M$, in the formulas above.

M-Step

Having found the required expectations, the maximization of the conditional expected log-likelihood in terms of the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\pi}$ and \mathbf{T} is done separately. Some of the expressions will in general not have an explicit solution.

In full generality, for the parameter $\boldsymbol{\alpha}$ of the scaling distribution we write

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\boldsymbol{\alpha}} \mathbb{E}[\log(f_\Theta(\Theta; \boldsymbol{\alpha})) \mid \mathbf{X} = \mathbf{x}].$$

For the parameters of the PH component, we first address the estimation of $\boldsymbol{\pi}$. Consider the Lagrange function

$$\zeta(\boldsymbol{\pi}) = \sum_{k=1}^p B_k \log(\pi_k) + \mu \left(1 - \sum_{k=1}^p \pi_k \right),$$

where μ is a Lagrange multiplier. Then

$$\frac{\partial \zeta(\boldsymbol{\pi})}{\partial \pi_k} = \frac{B_k}{\pi_k} - \mu = 0 \iff \mu \pi_k = B_k.$$

Summing over k

$$\mu = \sum_{k=1}^p B_k = M,$$

thus

$$\hat{\pi}_k = \frac{B_k}{M}.$$

Now, we consider the non-diagonal elements of \mathbf{T} , t_{kl} , $k \neq l$. We have that

$$\frac{\partial l_c}{\partial t_{kl}} = N_{kl} \frac{1}{t_{kl}} - \theta Z_k = 0,$$

implying

$$\hat{t}_{kl} = \frac{N_{kl}}{\theta Z_k}.$$

In a similar way, we can show that for t_k

$$\hat{t}_k = \frac{N_k}{\theta Z_k}.$$

Finally, we obtain the estimators for the diagonal elements of \mathbf{T} , by

$$\hat{t}_{kk} = - \left(\hat{t}_k + \sum_{l \neq k} \hat{t}_{kl} \right).$$

We summarize the two steps into a completed EM algorithm.

Algorithm 3.1 (EM algorithm for CPH distributions).

0. Initialize with some “arbitrary” $(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\alpha})$.

1. (*E-step*) Calculate

$$\mathbb{E}(B_k | \mathbf{X} = \mathbf{x}) = \sum_{n=1}^M \int_0^\infty \frac{\pi_k \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} x_n) \boldsymbol{\theta} \mathbf{t}}{f_X(x_n)} f_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

$$\mathbb{E}(\Theta Z_k | \mathbf{X} = \mathbf{x}) = \sum_{n=1}^M \int_0^\infty \theta \frac{\int_0^{x_n} \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} (x_n - u)) \boldsymbol{\theta} \mathbf{t} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{f_X(x_n)} f_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

$$\mathbb{E}(N_{kl} | \mathbf{X} = \mathbf{x}) = \sum_{n=1}^M \int_0^\infty \theta t_{kl} \frac{\int_0^{x_n} \mathbf{e}_l^\top \exp(\boldsymbol{\theta} \mathbf{T} (x_n - u)) \boldsymbol{\theta} \mathbf{t} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{f_X(x_n)} f_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

$$\mathbb{E}(N_k | \mathbf{X} = \mathbf{x}) = \sum_{n=1}^M \int_0^\infty \theta t_k \frac{\boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} x_n) \mathbf{e}_k}{f_X(x_n)} f_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

2. (*M-step*) Let

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\boldsymbol{\alpha}} \mathbb{E}(\log(f_\Theta(\boldsymbol{\theta}; \boldsymbol{\alpha})) | \mathbf{X} = \mathbf{x})$$

$$= \arg \max_{\boldsymbol{\alpha}} \sum_{n=1}^M \int_0^\infty \log(f_\Theta(\boldsymbol{\theta}; \boldsymbol{\alpha})) \frac{\boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} x_n) \boldsymbol{\theta} \mathbf{t}}{f_X(x_n)} f_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

$$\hat{\pi}_k = \frac{\mathbb{E}(B_k | \mathbf{X} = \mathbf{x})}{M}, \quad \hat{t}_{kl} = \frac{\mathbb{E}(N_{kl} | \mathbf{X} = \mathbf{x})}{\mathbb{E}(\Theta Z_k | \mathbf{X} = \mathbf{x})}, \quad \hat{t}_k = \frac{\mathbb{E}(N_k | \mathbf{X} = \mathbf{x})}{\mathbb{E}(\Theta Z_k | \mathbf{X} = \mathbf{x})},$$

$$\hat{t}_{kk} = - \sum_{l \neq k} \hat{t}_{kl} - \hat{t}_k.$$

Let $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \dots, \hat{\pi}_p)$, $\hat{\mathbf{T}} = \{\hat{t}_{kl}\}_{k,l=1,\dots,p}$, and $\hat{\mathbf{t}} = (\hat{t}_1, \dots, \hat{t}_p)^\top$.

3. Assign $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$, $\boldsymbol{\pi} := \hat{\boldsymbol{\pi}}$, $\mathbf{T} := \hat{\mathbf{T}}$, $\mathbf{t} := \hat{\mathbf{t}}$ and GOTO 1.

Remark 3.7. In general, none of the integrals will have explicit solutions, and approximations will have to be employed. Namely, one may discretize the continuous distribution and approximate the integrals by numerical methods, such as Simpson's rule, or use diagonalization for the integration of matrix functions. In practice, a combination of the two approaches yields good results.

Remark 3.8. The discrete case follows in a similar way, where sums replace integrals and are much easier to handle. Such an EM algorithm for NPH distributions was introduced in [Bladt and Rojas-Nandayapa \(2018\)](#), and the above formulas can be seen as the limit when the discretization becomes infinitely fine. In [Bladt and Rojas-Nandayapa \(2018\)](#), in fact, several of the illustrations arise as discretized continuous random variables.

3.3. Censored data. We call a data point *right-censored* at v if it takes an unknown value above v , and *left-censored* at w if it takes an unknown value below w . It is called *interval-censored* if it takes an unknown value within the interval $(v, w]$. Left-censoring is a particular case of interval-censoring with $v = 0$, while right-censoring is obtained by fixing v and letting $w \rightarrow \infty$.

The EM algorithm for censored data works in much the same way as for uncensored data, and the only change in Algorithm 3.1 is in the E-step. The derivation follows the approach taken for the PH case, see [Olsson \(1996\)](#). We consider first

$$\begin{aligned} \mathbb{E}[\log(f_{\Theta}(\Theta; \boldsymbol{\alpha}))\mathbf{1}(X > v)] &= \int_v^{\infty} \mathbb{E}[\log(f_{\Theta}(\Theta; \boldsymbol{\alpha})) \mid X = x] f_X(x) dx \\ &= \int_v^{\infty} \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) f_{\Theta|X}(\theta|x) d\theta f_X(x) dx \\ &= \int_v^{\infty} \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) \boldsymbol{\pi} \exp(\theta \mathbf{T} x) \boldsymbol{\theta} \mathbf{t} f_{\Theta}(\theta) d\theta dx \\ &= \int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) \boldsymbol{\pi} \exp(\theta \mathbf{T} v) \mathbf{e} f_{\Theta}(\theta) d\theta. \end{aligned}$$

Then, for interval-censored data, we obtain

$$\begin{aligned} \mathbb{E}[\log(f_{\Theta}(\Theta; \boldsymbol{\alpha})) \mid X \in (v, w)] &= \frac{\mathbb{E}[\log(f_{\Theta}(\Theta; \boldsymbol{\alpha}))\mathbf{1}(X > v)] - \mathbb{E}[\log(f_{\Theta}(\Theta; \boldsymbol{\alpha}))\mathbf{1}(X > w)]}{\mathbb{P}(X \in (v, w])} \\ &= \frac{\int_0^{\infty} \log(f_{\Theta}(\theta; \boldsymbol{\alpha})) (\boldsymbol{\pi} \exp(\theta \mathbf{T} v) \mathbf{e} - \boldsymbol{\pi} \exp(\theta \mathbf{T} w) \mathbf{e}) f_{\Theta}(\theta) d\theta}{\mathbb{P}(X \in (v, w])}. \end{aligned}$$

In the following we present the resulting remaining formulas (their derivations are similar to the above and are thus omitted for brevity).

$$\mathbb{E}(B_k \mid X \in (v, w]) = \int_0^{\infty} \frac{\pi_k \mathbf{e}_k^{\top} \exp(\theta \mathbf{T} v) \mathbf{e} - \pi_k \mathbf{e}_k^{\top} \exp(\theta \mathbf{T} w) \mathbf{e}}{\mathbb{P}(X \in (v, w])} f_{\Theta}(\theta) d\theta,$$

$$\mathbb{E}(\Theta Z_k \mid X \in (v, w])$$

$$\begin{aligned}
&= \int_0^\infty \frac{\theta}{\mathbb{P}(X \in (v, w])} \left[\int_v^w \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du - \int_0^w \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} (w - u)) \mathbf{e} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du \right. \\
&\quad \left. + \int_0^v \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} (v - u)) \mathbf{e} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du \right] f_\Theta(\theta) d\theta,
\end{aligned}$$

$$\mathbb{E}(N_{kl} \mid X \in (v, w])$$

$$\begin{aligned}
&= \int_0^\infty \frac{\theta t_{kl}}{\mathbb{P}(X \in (v, w])} \left[\int_v^w \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du - \int_0^w \mathbf{e}_l^\top \exp(\boldsymbol{\theta} \mathbf{T} (w - u)) \mathbf{e} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du \right. \\
&\quad \left. + \int_0^v \mathbf{e}_l^\top \exp(\boldsymbol{\theta} \mathbf{T} (v - u)) \mathbf{e} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du \right] f_\Theta(\theta) d\theta,
\end{aligned}$$

$$\mathbb{E}(N_k \mid X \in (v, w]) = \int_0^\infty \theta t_k \frac{\int_v^w \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{\mathbb{P}(X \in (v, w])} f_\Theta(\theta) d\theta.$$

Remark 3.9. The main challenge when implementing the above formulas is obtaining numerical estimates for integrals defined with respect to matrix exponentials, which are usually slow to evaluate. A suite of fast and useful routines for calculating matrix exponentials (and more generally, estimation tools for PH distributions) can be found, for instance, in the R package `matrixdist`, see [Bladt and Yslas \(2021a,b\)](#).

4. EXAMPLES

In this section we present three detailed numerical illustrations of the estimation of CPH distributions via the EM algorithm from the previous section. In the first two examples we fit CPH distributions to real data sets, the second one containing censored observations, while in the last example we consider the estimation of a CPH distribution to a theoretical given distribution. In all cases, we ran the algorithms until the changes in the successive log-likelihoods became negligible. Note that the purpose of this section is not to thoroughly compare to and outperform other models for the given data (which with criteria like AIC or BIC would in any case not be evident given the non-identifiability and overparametrization issues of PH distributions in general). Instead, our aim is to illustrate how the algorithms developed in this paper can, in fact, be implemented in a straightforward manner. As a consequence, we showcase CPH distributions as interesting alternatives in the statistical modeling toolkit of the respective application areas. Note that the concrete purpose may then decide which model one might want to use.

4.1. Dutch fire insurance data. We consider claims above 1 million (Euro) from the Dutch fire insurance claim data set studied in [Albrecher et al. \(2017\)](#), subtract 1 million to all data points (to shift them to the origin), and scale by a factor of 10^{-6} . Subsequently, we fit a matrix-Pareto type II to the resulting sample. To reduce the number of parameters, we consider a Coxian structure (which is often sufficient)

of dimension 3 in the PH component, obtaining in this way the following estimated parameters

$$\begin{aligned} \hat{\boldsymbol{\pi}} &= (1, 0, 0) , \\ \hat{\mathbf{T}} &= \begin{pmatrix} -0.8620 & 0.8079 & 0 \\ 0 & -2.4341 & 1.1014 \\ 0 & 0 & -1.5808 \end{pmatrix} , \\ \hat{\alpha} &= 1.3792 , \end{aligned}$$

and corresponding log-likelihood of $-2,657.666$. For reference, we also consider a matrix-Pareto (type I) model (Albrecher et al., 2020) defined by

$$\bar{F}(x) = \boldsymbol{\pi} \left(\frac{x}{\beta} + 1 \right)^{\mathbf{T}} \mathbf{e}, \quad x > 0.$$

We again use a Coxian structure of dimension 3 in the PH component. The resulting estimated parameters are

$$\begin{aligned} \hat{\boldsymbol{\pi}} &= (1, 0, 0) , \\ \hat{\mathbf{T}} &= \begin{pmatrix} -1.4370 & 1.3183 & 0 \\ 0 & -9.0824 & 2.3177 \\ 0 & 0 & -5.8812 \end{pmatrix} , \\ \hat{\beta} &= 1.5821 , \end{aligned}$$

with corresponding log-likelihood $-2,657.702$.

Figures 4.1 and 4.2 show that both fitted distributions provide adequate and very similar models for the sample. Regarding the tail behavior, we have that the index of regular variation of the matrix-Pareto type II model is given by $\hat{\alpha} = 1.3792$, while for the matrix-Pareto (type I) model, it is given by the negative of the largest real eigenvalue, viz. $\lambda = 1.4370$. Note that both these two estimates are comparable with results obtained in previous studies. For instance, in (Albrecher et al., 2017, Page 107), a value of 1.255 is proposed in their splicing model. In contrast to traditional extreme value techniques, the present matrix models are global. Hence, the tail index is only one aspect of the distribution, not the calibration focus, yet the fit is very satisfactory. As previously noted, when comparing matrix-Pareto type I and type II distributions, the latter has a tail parameter that does not depend on the underlying Markov structure. The advantage, in this case, is that it may also estimate the tail index in a first step separately with extreme value techniques and then estimate the other parameters in a subsequent step (similar to the approach in Bladt and Rojas-Nandayapa (2018) for the NPH case). On the other hand, the matrix-Pareto type II is computationally more demanding. The running times and resulting estimators depend on the quality of the approximations in the EM algorithm formulas. For the Dutch fire insurance data, in both cases, increasing the dimension of the PH component does not improve the fit significantly. For instance, for a Coxian structure of dimension 5 in the PH component, we obtain a log-likelihood of $-2,657.603$ in the matrix-Pareto

type I model and $-2,657.653$ in the type II. Thus, the choice of dimension 3 seems eventually very reasonable here.

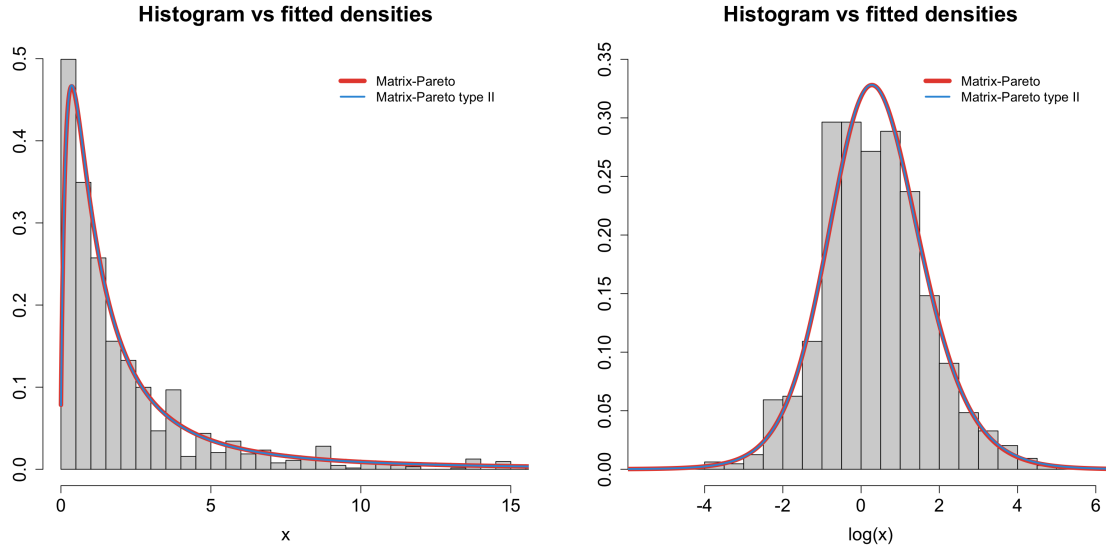


FIGURE 4.1. Histogram of Dutch fire insurance data versus fitted matrix-Pareto type I and II distributions (left), and histogram of the log-data (right)

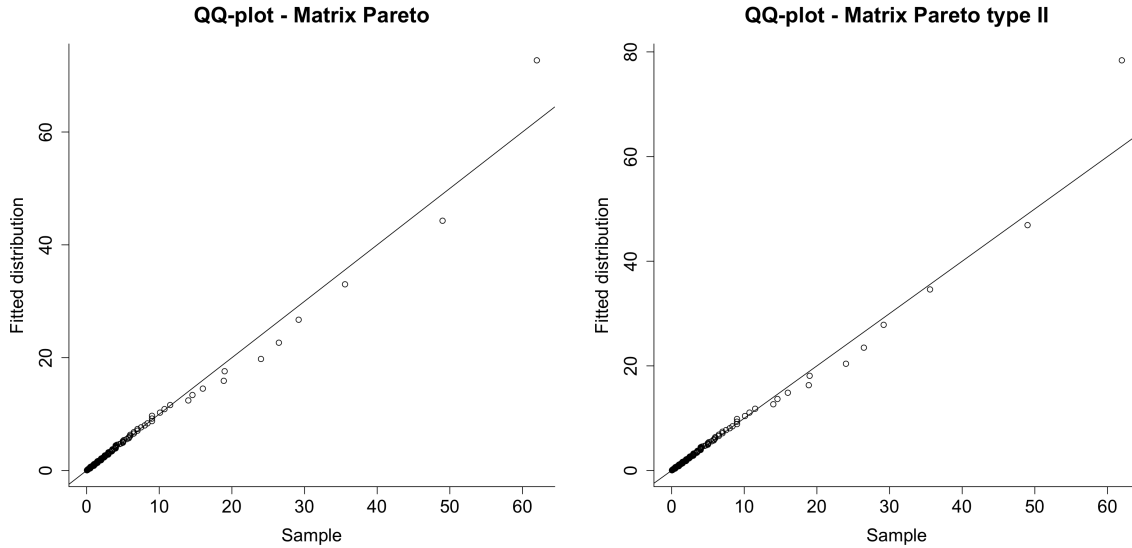


FIGURE 4.2. QQ-plot of Dutch fire insurance data versus fitted matrix-Pareto type I distribution (left), and QQ-plot of Dutch fire insurance data versus fitted matrix-Pareto type II distribution (right)

Note that in [Albrecher and Bladt \(2019\)](#), a previous analysis of the same data set employing the matrix-Pareto (type I) distribution was considered with a specific and

sparse matrix structure of dimension 20, whereas here, we are interested in the best fit of a low-dimensional structure.

4.2. Loss insurance data. We consider the loss insurance claim data set from [Frees and Valdez \(1998\)](#). The data comprises 1500 insurance claims from a real-life insurance company, where each data point consists of an indemnity payment (loss) and an allocated loss adjustment expense (ALAE). For the present analysis, we only use the loss component (scaled by a factor of 10^{-4}), for which 34 observations are right-censored, and fit a matrix-Pareto type II of 4 phases. The resulting fitted parameters are

$$\hat{\boldsymbol{\pi}} = (0.0476, 0.0289, 0.1412, 0.7823) ,$$

$$\hat{\boldsymbol{T}} = \begin{pmatrix} -2.9587 & 0.1886 & 1.2395 & 0.6833 \\ 0.5585 & -3.5859 & 0.6233 & 0.0364 \\ 0.1152 & 0.0650 & -0.5554 & 0.2892 \\ 0.5079 & 1.9315 & 0.4666 & -3.0784 \end{pmatrix} ,$$

$$\hat{\alpha} = 1.3744 .$$

The quality of the fit is supported by [Figure 4.3](#), where we see that the cumulative hazard of the matrix-Pareto type II model is close to the one implied by the non-parametric Kaplan-Meier estimator. A recent analysis of the same data set employing the matrix-Pareto (type I) distribution can be found in [Bladt and Yslas \(2021a\)](#).

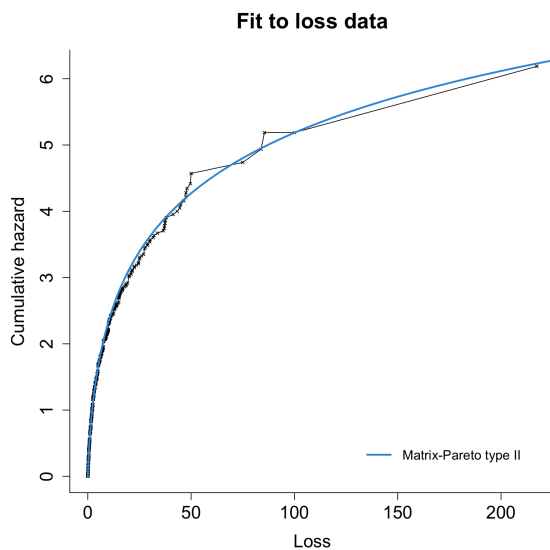


FIGURE 4.3. Cumulative hazard function of the fitted matrix-Pareto type II versus the non-parametric Nelson-Aalen estimator of the sample.

4.3. Fitting to a known distribution. Algorithm [3.1](#) can be modified to fit a CPH distribution to a theoretical given distribution. This is done in a similar way as in the PH case, and the idea consists of considering sequences of empirical distributions with

increasing sample size. Let h be an absolutely continuous density. Then, Algorithm 3.1 and dominated convergence yield:

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &\rightarrow \arg \max_{\boldsymbol{\alpha}} \int_0^\infty \int_0^\infty \log(f_\Theta(\theta; \boldsymbol{\alpha})) \frac{\boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} x) \boldsymbol{\theta} \mathbf{t}}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx, \\ \hat{\pi}_k &\rightarrow \int_0^\infty \int_0^\infty \frac{\pi_k \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} x) \boldsymbol{\theta} \mathbf{t}}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx, \\ \hat{t}_{kl} &\rightarrow \frac{\int_0^\infty \int_0^\infty \theta t_{kl} \frac{\int_0^x \mathbf{e}_l^\top \exp(\boldsymbol{\theta} \mathbf{T} (x-u)) \boldsymbol{\theta} \mathbf{t} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx}{\int_0^\infty \int_0^\infty \theta \frac{\int_0^x \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} (x-u)) \boldsymbol{\theta} \mathbf{t} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx}, \\ \hat{t}_k &\rightarrow \frac{\int_0^\infty \int_0^\infty \theta t_k \frac{\boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} x) \mathbf{e}_k}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx}{\int_0^\infty \int_0^\infty \theta \frac{\int_0^x \mathbf{e}_k^\top \exp(\boldsymbol{\theta} \mathbf{T} (x-u)) \boldsymbol{\theta} \mathbf{t} \boldsymbol{\pi} \exp(\boldsymbol{\theta} \mathbf{T} u) \mathbf{e}_k du}{f_X(x)} f_\Theta(\theta) d\theta h(x) dx}.\end{aligned}$$

As an example, we consider a matrix-Weibull distribution (Albrecher and Bladt, 2019), whose density function is given by

$$h(x) = \boldsymbol{\pi} \exp(\mathbf{S} x^\beta) \mathbf{s} \beta x^{\beta-1}, \quad x > 0, .$$

where \mathbf{S} is a sub-intensity matrix, and $\mathbf{s} = -\mathbf{S}\mathbf{e}$. For the present illustration, we take parameters

$$\begin{aligned}\boldsymbol{\pi} &= (0.5, 0.3, 0.2) , \\ \mathbf{S} &= \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 0 & -5 \end{pmatrix} , \\ \beta &= 0.75 .\end{aligned}$$

Then, we approximate this distribution with a CPH distribution of 3 phases with a general Coxian structure in the PH component and positive stable mixing. The fitted distribution has parameters

$$\begin{aligned}\hat{\boldsymbol{\pi}} &= (0.2326, 0.4098, 0.3576) , \\ \hat{\mathbf{T}} &= \begin{pmatrix} -1.3436 & 1.2722 & 0 \\ 0 & -11.2165 & 7.2335 \\ 0 & 0 & -0.8981 \end{pmatrix} , \\ \hat{\alpha} &= 0.7932 .\end{aligned}$$

Figure 4.4 shows that we recover the shape of the original distribution. Moreover, the parameter $\hat{\alpha} = 0.7932$, which determines the heaviness of the tail, is close to the

corresponding one of $\beta = 0.75$ for the given theoretical model. This is no surprise since stable mixing generates the same class as matrix-Weibull distributions. In fact, $-(\hat{\mathbf{T}})^{\beta}$ approximates \mathbf{S} , and although the former matrix is a valid sub-intensity matrix, in the present case it falls outside of the Coxian structure (but it is still upper triangular).

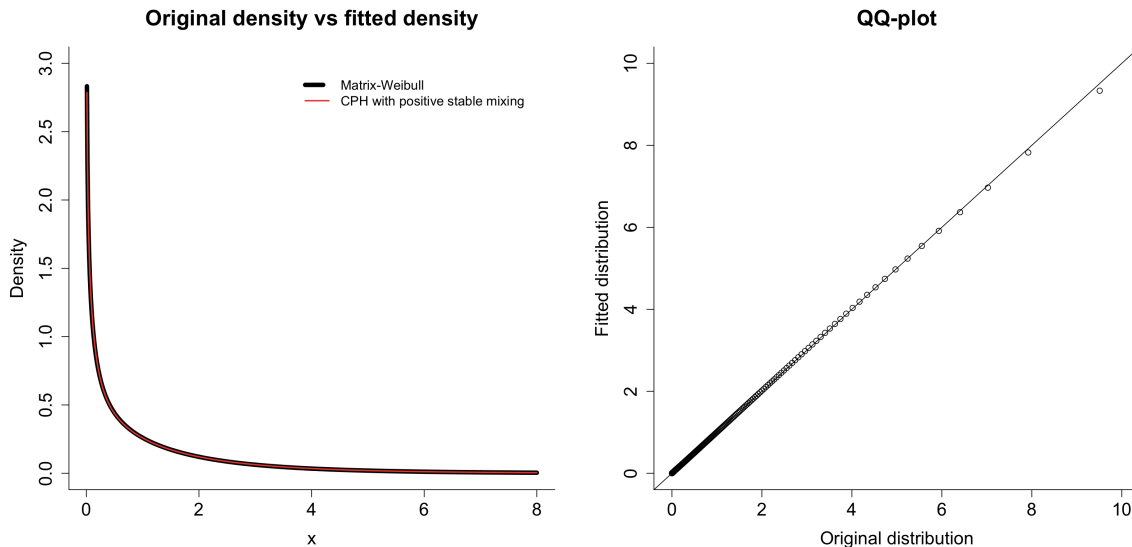


FIGURE 4.4. Density of the original matrix-Weibull versus density of the fitted CPH (left), and corresponding QQ-plot (right)

5. CONCLUSION

We studied scaled phase-type distributions when the scaling component is continuous. Particular emphasis was given to the closed-form expressions involved in different functionals, and we adapted and extended some recent results regarding the tail behavior of these distributions. We derived an EM algorithm for maximum likelihood estimation for fully observed data and outlined extensions for censored data. The case of fitting to a given theoretical distribution function was illustrated using a modified version of the latter algorithms. The performance of the proposed algorithm was exemplified in various numerical examples. The results suggest that these global models correctly capture the data's overall shape while remaining in agreement with existing tail behavior analysis. In addition to the results and insight gained for the one-dimensional setup considered in this paper, which is of interest on its own, one can also view the present analysis as a starting point for a multivariate framework when scaling a multivariate random vector with respect to the same continuous random variable. It will be interesting to develop the corresponding implementation and examine the performance on data as well as interpret the implied near-Archimedean dependence structure between the different dimensions from an applied perspective. That approach has potential

for parsimonious multivariate modeling, but its concrete development leads to several challenges along the way, which we intend to address in future work.

Acknowledgement. Hansjörg Albrecher and Martin Bladt would like to acknowledge financial support from the Swiss National Science Foundation Project 200021_191984.

Jorge Yslas would like to acknowledge financial support from the Swiss National Science Foundation Project IZHRZ0_180549.

A. DEFINITIONS

Definition A.1. A distribution function F on $\mathbb{R}_+ = [0, \infty)$ is called:

- (1) *Extended regularly varying* if

$$\liminf_{x \rightarrow \infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} \geq \lambda^{-\phi}$$

for some $\phi \geq 0$ and all $\lambda \geq 1$. The class of extended regularly varying distributions is denoted by \mathcal{E} .

- (2) *Intermediate regularly varying* if

$$\lim_{\lambda \downarrow 1} \liminf_{x \rightarrow \infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = 1.$$

The class of intermediate regularly varying distributions is denoted by \mathcal{I} .

- (3) *Dominated varying* if

$$\liminf_{x \rightarrow \infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} > 0$$

for some $\lambda > 1$. The class of dominated varying distributions is denoted by \mathcal{D} .

- (4) *Long-tailed* if

$$\lim_{x \rightarrow \infty} \frac{\overline{F}(x - y)}{\overline{F}(x)} = 1$$

for any $y \in \mathbb{R}$. The class of long tailed distributions is denoted by \mathcal{L} .

REFERENCES

- Albrecher, H., Beirlant, J., and Teugels, J. L. (2017). *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons, Chichester.
- Albrecher, H. and Bladt, M. (2019). Inhomogeneous phase-type distributions and heavy tails. *Journal of Applied Probability*, 56(4):1044–1064.
- Albrecher, H., Bladt, M., and Bladt, M. (2020). Matrix Mittag–Leffler distributions and modeling heavy-tailed risks. *Extremes*, 23:425–450.
- Arendarczyk, M., Debicki, K., et al. (2011). Asymptotics of supremum distribution of a Gaussian process over a Weibullian time. *Bernoulli*, 17(1):194–210.
- Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441.

- Berg, C., Boyadzhiev, K., and Delaubenfels, R. (1993). Generation of generators of holomorphic semigroups. *Journal of the Australian Mathematical Society*, 55(2):246–269.
- Bladt, M. and Nielsen, B. F. (2017). *Matrix-Exponential Distributions in Applied Probability*. Springer.
- Bladt, M., Nielsen, B. F., and Samorodnitsky, G. (2015). Calculation of ruin probabilities for a dense class of heavy tailed distributions. *Scandinavian Actuarial Journal*, 2015(7):573–591.
- Bladt, M. and Rojas-Nandayapa, L. (2018). Fitting phase-type scale mixtures to heavy-tailed data and distributions. *Extremes*, 21:285–313.
- Bladt, M. and Yslas, J. (2021a). matrixdist: An R package for inhomogeneous phase-type distributions. *arXiv preprint arXiv:2101.07987*.
- Bladt, M. and Yslas, J. (2021b). *matrixdist: Statistics for Matrix Distributions*. R package version 1.1.2.
- Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability & Its Applications*, 10(2):323–331.
- Cline, D. B. (1986). Convolution tails, product tails and domains of attraction. *Probability Theory and Related Fields*, 72(4):529–557.
- Cline, D. B. (1987). Convolutions of distributions with exponential and subexponential tails. *Journal of the Australian Mathematical Society (Series A)*, 43(03):347–365.
- Cline, D. B. and Samorodnitsky, G. (1994). Subexponentiality of the product of independent random variables. *Stochastic Processes and their Applications*, 49(1):75–98.
- Cossette, H., Marceau, E., Mtalai, I., and Veilleux, D. (2021). Univariate and multivariate mixtures of exponential distributions, with applications in risk modeling. *Applied Stochastic Models in Business and Industry*.
- Doolittle, E. (1998). Analytic functions of matrices. *Lecture Note*.
- Embrechts, P. and Goldie, C. M. (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Series A)*, 29(02):243–256.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling Extremal Events: For Insurance and Finance*, volume 33. Springer Science & Business Media.
- Foss, S., Korshunov, D., Zachary, S., et al. (2011). *An Introduction to Heavy-Tailed and Subexponential Distributions*, volume 6. Springer.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.
- Furman, E., Kye, Y., and Su, J. (2021). Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. *Insurance: Mathematics and Economics*, 96:153–167.
- Higham, N. J. (2008). *Functions of Matrices: Theory and computation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Jacobsen, M., Mikosch, T., Rosiński, J., Samorodnitsky, G., et al. (2009). Inverse problems for regular variation of linear filters, a cancellation property for σ -finite measures and identification of stable laws. *The Annals of Applied Probability*, 19(1):210–242.

- Maulik, K. and Resnick, S. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.
- Olsson, M. (1996). Estimation of phase-type distributions from censored data. *Scandinavian Journal of Statistics*, 23(4):443–460.
- Rojas-Nandayapa, L. and Xie, W. (2018). Asymptotic tail behaviour of phase-type scale mixture distributions. *Annals of Actuarial Science*, 12(2):412–432.
- Rudin, W. (1987). *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA.
- Samorodnitsky, G. and Sun, J. (2016). Multivariate subexponential distributions and their applications. *Extremes*, 19(2):171–196.
- Schilling, R. L., Song, R., and Vondracek, Z. (2012). *Bernstein functions.* de Gruyter.
- Tang, Q. (2006). The subexponentiality of products revisited. *Extremes*, 9(3-4):231–241.
- Xu, H., Cheng, F., Wang, Y., and Cheng, D. (2017). A necessary and sufficient condition for the subexponentiality of the product convolution. *Advances in Applied Probability*, 50(1):57–73.

FACULTY OF BUSINESS AND ECONOMICS, UNIVERSITY OF LAUSANNE, QUARTIER DE CHAMBRONNE,
1015 LAUSANNE, AND SWISS FINANCE INSTITUTE, LAUSANNE, SWITZERLAND

Email address: `hansjoerg.albrecher@unil.ch`

FACULTY OF BUSINESS AND ECONOMICS, UNIVERSITY OF LAUSANNE, QUARTIER DE CHAMBRONNE,
1015 LAUSANNE, SWITZERLAND

Email address: `martin.bladt@unil.ch`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COPENHAGEN, UNIVERSITETSPARKEN 5, DK-
2100 COPENHAGEN, DENMARK

Email address: `bladt@math.ku.dk`

INSTITUTE OF MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE, UNIVERSITY OF BERN,
ALPENEGGSTRASSE 22, CH-3012 BERN, SWITZERLAND

Email address: `jorge.yslas@stat.unibe.ch`