Asymptotic tail behavior of phase-type scale mixture distributions

Leonardo Rojas-Nandayapa and Wangyue Xie

School of Mathematics and Physics, The University of Queensland, Brisbane, QLD, Australia,
l.rojas@uq.edu.au, w.xie1@uq.edu.au

Abstract

We consider *phase-type scale mixture* distributions which correspond to distributions of a product of two independent random variables: a phase-type random variable Y and a nonnegative but otherwise arbitrary random variable S called the *scaling random variable*. We investigate conditions for such a class of distributions to be either light- or heavy-tailed, we explore subexponentiality and determine their maximum domains. Particular focus is given to phase-type scale mixture distributions where the scaling random variable S has discrete support — such a class of distributions have been recently used in risk applications to approximate heavy-tailed distributions. Our results are complemented with several examples.

Keywords: phase-type; Erlang; discrete scale mixtures; infinite mixtures; heavy-tailed; subexponential; maximum domain of attraction; products; ruin probability.

1 Introduction

In this paper we consider the class of nonnegative distributions defined by the *Mellin–Stieltjes convolution* (Bingham et al., 1987) of two nonnegative distributions G and H, given by

$$F(x) = \int_0^\infty G(x/s)\mathrm{d}H(s), \qquad x \geq 0. \tag{1.1}$$

A distribution of the form (1.1) will be called a *phase-type scale mixture* if G is a (classical) phase-type (PH) distribution (cf. Latouche and Ramaswami, 1999) and H is a proper nonnegative distribution that we shall call the *scaling distribution*. A phase-type scale mixture distribution may be seen as the distribution of a random variable X of the form $X := S \cdot Y$ where $S \sim H$ and $Y \sim G$; accordingly, we refer to S as the *scaling random variable*. This terminology can be explained by using the following conditional argument: we observe that $(X|S = s) \sim G_s$ where $G_s(x) := G(x/s)$ corresponds to the distribution of the (scaled) random variable $s \cdot Y$ which is itself a PH distribution, hence the distribution F can be thought as a mixture of the scaled PH distributions in $\{G_s : s > 0\}$ with respect to the scaling distribution H.

Our motivation for studying the tail behavior of phase-type scale mixtures is their use for approximating heavy-tailed distributions in risk applications (Bladt et al., 2015). We shall recall that the (classical) PH class corresponds to distributions of absorption times of Markov jump processes with one absorbing state and a finite number of transient states (the exponential, Erlang and hyperexponential distributions are examples of PH distributions). The PH class possesses many properties: expressions for densities, cumulative distributions, moments and integral transforms have closed-form expressions which are given in terms of matrix exponentials; it is closed class under scaling, finite mixtures and finite convolutions (Assaf and Levikson, 1982; Maier and O'Cinneide, 1992). The PH class is particularly attractive for modelling purposes because it is dense in the nonnegative distributions (cf. Asmussen, 2003), so one could in principle approximate any nonnegative distribution with an arbitrary precision. This classical approach has been widely studied and reliable methodologies for approximating nonnegative distributions are already available (cf. Asmussen et al., 1996).

However, distributions in the PH class are light-tailed and belong to the Gumbel domain of attraction exclusively (Kang and Serfozo, 1999). Therefore, the PH class cannot correctly capture the characteristic behavior of a heavy-tailed distribution in spite of its denseness. In fact, this approach may deliver unreliable approximations for important quantities of interest, such as the ruin probability of a Cramér–Lundberg risk process with heavy-tailed claim size distributions (Vatamidou et al., 2012, 2014). As an alternative, the PH class has been extended to distributions of absorption times having a countable number of transient states (this approach is attributed to Neuts, 1981). The later class, which goes under the name of infinite dimensional phase-type distributions (IDPH), is known to contain heavy-tailed distributions. Nevertheless, the IDPH class is no longer mathematically tractable and it is not fully documented yet (to the best of the authors' knowledge, one of the few published references available outlining its mathematical properties is Shi et al. (1996); another reference of interest is Greiner et al. (1999), who considered infinite mixtures of exponential distributions to approximate power-tailed distributions).

To address this issue, Bladt et al. (2015) proposed the use of phase-type scale mixtures having discrete scaling distributions to approximate heavy-tailed distributions. Such a class of distributions is a structured subfamily of the IDPH class and contains the PH class so it is trivially dense in the nonnegative distributions, but it is mathematically tractable and it contains a rich variety of heavy-tailed distributions. In particular, Bladt et al. (2015) provided renewal results that can be applied to obtain exact expressions for the ruin probability of a classical Cramér–Lundberg risk process having claim sizes distributed according to a phase-type scale mixture with discrete scaling distribution. This approach was further explored in Peralta et al. (2016) where a systematic methodology to approximate arbitrary heavy-tailed distributions via phase-type scale mixtures was provided. Furthermore, Bladt and Rojas-Nandayapa (2017) provided statistical estimation procedures based on the EM algorithm to adjust phase-type scale mixtures to heavy-tailed data/distributions.

In spite of the flexibility for modelling heavy-tailed distributions and the mathematically tractability of the class of phase-type scale mixtures, the tail properties of the proposed class are not fully understood yet; this paper concentrates on this issue. We collect and adapt some known results which are available in different contexts and we prove new results that will allow us to provide a characterization of the tail behavior of phase-type scale mixtures as well as a classification of their maximum domains of attraction. We expect our results to be useful for modelling purposes by providing a better understanding of the advantages and limitations of such an approach as well as providing criteria for selecting appropriate scaling distributions for approximating general heavy-tailed distributions. Our results are summarized below.

Firstly, we concentrate on classifying light and heavy-tailed distributions. We recall that PH distributions are inherently *light-tailed*; in contrast, a phase-type scale mixture is heavy-tailed if and only if its scaling distribution has unbounded support. An interesting heuristic interpretation of this result is as follows: a PH random variable multiplied with a random variable S is heavy-tailed if S has unbounded support. We provide a simple proof of this fact but we remark that this result was recently proven in a different context (cf. Su and Chen, 2006; Tang, 2008).

Secondly, we focus on the maximum domains of attraction and subexponential properties of the class of phase-type scale mixtures. A classical result for the Fréchet case is Breiman's lemma (Breiman, 1965), which applied to the current context implies that a phase-type scale mixture with a regularly varying scaling distribution remains regularly varying with the same index (hence subexponential). In addition, we investigate analogue results for scaling distributions in the Gumbel domain of attraction. We show that if a certain higher order derivative of the Laplace–Stieltjes transform of the reciprocal of the scaling random variable $\mathcal{L}_{1/S}(\theta)$ is a von Mises function, then $F \in \mathrm{MDA}(\Lambda)$; in addition, we provide a verifiable condition for subexponentiality.

We then specialize in phase-type scale mixture distributions having discrete support. Such a class of distributions is of critical importance in applications due to its mathematical tractability as these correspond to distributions of the absorption time of a Markov jump process having an infinite number of transient states. We outline a simple methodology which allows to determine their asymptotic behaviour by constructing a phase-type scale mixture distributions having asymptotically proportional tail probabilities. This methodology can be *reverse–engineered* so we can construct discrete scaling distributions for approximating the tail probability of some arbitrary target distribution.

The rest of the paper is organized as follows. In Section 2 we set up notation and summarize some of the standard facts on heavy-tailed, phase-type and related distributions. Then we introduce the class of phase-type scale mixtures and examine some of its asymptotic properties. Our main results are presented in Section 3 and 4. Section 3 is devoted to the general case while 4 is specialized in discrete scaling distributions. In Section 5 we present our conclusions and discuss further directions of research.

2 Preliminaries

In this section we provide a summary of the concepts needed for this paper. Most results in this section are fundamental and well known. In subsection 2.1 we define *phase-type* (PH) distributions and their extension to phase-type scale mixtures. We will refer to the former class of distributions as *classical* in order to make a clear distinction from the later class. In subsection 2.2 we establish the definition of heavy-tailed distributions that will be used in this paper, then we discuss various important subfamilies of heavy-tailed distributions. In that section, we also provide a brief summary of results connecting extreme value theory with heavy-tailed distributions and subexponentiality.

2.1 Phase-type scale mixtures and related distributions

First we introduce classical phase-type distributions. A phase-type distribution corresponds to the distribution of the absorption time of a Markov jump process $\{X_t\}_{t\geq 0}$ with a finite state space $E = \{0, 1, 2, \cdots, p\}$. The states $\{1, 2, \cdots, p\}$ are transient while the state 0 is an absorbing one. Let $Y := \inf\{t \geq 0 : X_t = 0\}$ be the random variable of the time until absorption in state 0. The distribution of Y is called *phase-type distribution* (cf. Asmussen, 2003; Latouche and Ramaswami, 1999). We concentrate on proper distributions so the Markov jump process cannot be started in the absorbing state. Hence, phase-type distributions are characterized by a p-dimensional row vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$ (corresponding to the probabilities of starting the Markov jump process in each of the transient states), and an intensity matrix

$$\mathbf{Q} = \begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{\lambda} & \boldsymbol{\Lambda} \end{pmatrix},$$

where $\boldsymbol{\Lambda}$ is a $p \times p$ sub-intensity matrix. Since rows in a intensity matrix must sum to 0 we also have $\boldsymbol{\lambda} = -\boldsymbol{\Lambda}\boldsymbol{e}$, where $\boldsymbol{e}$ is the $p-$dimensional column vector of 1s. Phase-type distributions are denoted $\mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ and its density and cumulative distribution functions are given by

$$g(x) = \boldsymbol{\beta}\mathrm{e}^{\boldsymbol{\Lambda}x}\boldsymbol{\lambda}, \qquad G(x) = 1 - \boldsymbol{\beta}\mathrm{e}^{\boldsymbol{\Lambda}x}\mathbf{e}, \qquad \forall x > 0.$$

In this paper we are particularly interested in distributions of scaled phase-type random variables $s \cdot Y$ where $Y \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ and $s > 0$. From the expressions above, it follows easily that $s \cdot Y \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda}/s)$, so the class of phase-type distributions is closed under scaling transformations. The following is a well known result describing the tail behavior of phase-type distributions:

Proposition 2.1. *Let $G_s \sim \mathrm{PH}(\boldsymbol{\beta},\, \boldsymbol{\Lambda}/s)$. The tail probability of G_s can be written as*

$$\overline{G}_s(x) = \sum_{j=1}^{m} \sum_{k=0}^{\eta_j - 1} \left(\frac{x}{s}\right)^k \mathrm{e}^{\Re(-\lambda_j)x/s} \left[c_{jk}^{(1)} \sin(\Im(-\lambda_j)x/s) + c_{jk}^{(2)} \cos(\Im(-\lambda_j)x/s) \right].$$

Here m is the number of Jordan blocks of the matrix $\boldsymbol{\Lambda}$, $\{-\lambda_j : j = 1, \ldots, m\}$ are the corresponding eigenvalues and $\{\eta_j : j = 1, \ldots, m\}$ the dimensions of the Jordan blocks. The values $c_{jk}^{(1)}$, $c_{jk}^{(2)}$ are constants depending on the initial distribution $\boldsymbol{\beta}$, the dimension of the j-th Jordan block η_j and the generalized eigenvectors of $\boldsymbol{\Lambda}$.

All eigenvalues of a sub-intensity matrix $\boldsymbol{\Lambda}$ have negative real parts and the one with the largest absolute value is always real. Therefore the asymptotic behavior of a scaled phase-type distribution is determined by the largest eigenvalue and the largest dimension among the Jordan blocks associated to the largest eigenvalue (see also Asimit and Jones, 2006; Asmussen, 2003).

It is also well known that if the subintensity matrix Λ is irreducible, then the tail probabilities of phase-type distributions decay exponentially (cf. Proposition IX.1.8 Asmussen and Albrecher, 2000), that is if $G_s \sim \mathrm{PH}(\beta, \Lambda/s)$, then $\overline{G}_s(x) \sim C\mathrm{e}^{-\lambda x}$ for some constant C and λ begin the largest eigenvalue. The assumption that the matrix $\mathbf{\Lambda}$ is not irreducible can be further relaxed if all eigenvalues are real; in such a case the following result can be easily deduced from the previous Proposition:

Corollary 2.2. *Let $G_s \sim \mathrm{PH}(\boldsymbol{\beta}, \Lambda/s)$. If Λ has real eigenvalues, then*

$$\overline{G}_s(x) = \frac{\gamma x^{\eta-1}\mathrm{e}^{-\lambda x/s}}{s^{\eta-1}}\left(1+o(1)\right), \qquad x \to \infty,$$

where $-\lambda$ is the largest real eigenvalue of $\mathbf{\Lambda}$, η is the largest dimension of the Jordan block associated to $-\lambda$ and γ is a positive constant.

Notice that if all the eigenvalues of $\mathbf{\Lambda}$ are real ($\Im(-\lambda_j) = 0$), then

$$\overline{G}_s(x) = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk}\left(\frac{x}{s}\right)^k \mathrm{e}^{-\lambda_j x/s}.$$

Next we introduce the class of phase-type scale mixture distributions which is a central for this paper.

Definition 2.3. *(Phase-type scale mixture distributions).*
We say a distribution $F(x)$ is a phase-type scale mixture with scaling distribution H and phase-type distribution $G \sim \mathrm{PH}(\boldsymbol{\beta}, \mathbf{\Lambda})$ if the distribution F can be written as a Mellin–Stieltjes convolution

$$F(x) = \int_0^\infty G(x/s)\mathrm{d}H(s). \tag{2.1}$$

Observe that in the definition above it is implicit that H must be nonnegative without an atom in 0. Particularly, when the scaling distribution H is discrete and supported over a countable set of nonnegative numbers $\{s_i : i \in \mathbb{N}\}$, then the Mellin–Stieltjes convolution above reduces to the following infinite series:

$$F(x) = \sum_{i=1}^{\infty} p(i)G(x/s_i),$$

where $p(i) := H(s_i) - H(s_{i-1})$ is the probability mass function of H and $s_0 = 0$. Phase-type scale mixture distributions are absolutely continuous:

Proposition 2.4. *A phase-type scale mixture distribution F is absolutely continuous and its density function can be written as*

$$f(x) = \int_0^\infty \frac{g(x/s)}{s}\mathrm{d}H(s),$$

where g is the density of the phase-type distribution G.

Proof. Since H is a finite measure and $G_s(x) := G(x/s)$ is bounded and continuous at x, then F is an absolutely continuous distribution with

$$f(x) = F'(x) = \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}x}G(x/s)\mathrm{d}H(s) = \int_0^\infty \frac{g(x/s)}{s}\mathrm{d}H(s).$$

□

The tail probability of a phase-type scale mixture $\overline{F} := 1 - F$ can also be written as a Mellin–Stiltjes convolution of H and $\overline{G}$:

$$\overline{F}(x) = 1 - \int_0^\infty G(x/s)\mathrm{d}H(s) = \int_0^\infty (1 - G(x/s))\mathrm{d}H(s) = \int_0^\infty \overline{G}(x/s)\mathrm{d}H(s).$$

Therefore, using proposition 2.1 it is straightforward to see that there exist constants c'_{jk} and c'_k, such that

$$\overline{F}(x) \le \sum_{j=1}^{m} \sum_{k=0}^{\eta_j-1} c'_{jk} \int_0^\infty \left(\frac{x}{s}\right)^k \mathrm{e}^{\Re(-\lambda_j)x/s} \mathrm{d}H(s) \le \sum_{k=0}^{\eta-1} c'_k \int_0^\infty \left(\frac{x}{s}\right)^k \mathrm{e}^{-\lambda x/s} \mathrm{d}H(s).$$

In the expression above, $-\lambda$ is the largest eigenvalue of the sub-intensity matrix $\mathbf{\Lambda}$ and η the largest dimension of the Jordan blocks having associated eigenvalue $-\lambda$. Hence, only the largest real eigenvalue determines the asymptotic behavior of a phase-type scale mixture distribution.

2.2 Heavy-tailed distributions, Extremes and Subexponentiality

We say that a nonnegative distribution H is *heavy-tailed* if

$$\limsup_{s\to\infty} \overline{H}(s)\mathrm{e}^{\theta s} = \infty, \quad \forall \theta > 0,$$

where $\overline{H}(s) = 1 - H(s)$ is the *tail probability of the distribution* H. Otherwise, we say that H is a *light-tailed* distribution. The Pareto and Lognormal distributions are examples of heavy-tailed distributions while the class of phase-type distributions is light-tailed. An example of a class of distributions containing both light and heavy-tailed is that of Weibullian distributions: a nonnegative distribution H is said to be Weibullian with shape parameter $p > 0$ (Arendarczyk and Dębicki, 2011) if

$$\overline{H}(s) = Cs^\delta \exp(-\lambda s^p)(1 + o(1)), \quad \lambda, C > 0, \delta \in \mathbb{R}.$$

Clearly, a Weibullian distribution with parameter p is heavy-tailed if $0 < p < 1$ while it is light-tailed if $p \ge 1$.

The definition of light/heavy-tailed distributions is too general for most practical purposes and it is more common to consider certain subfamilies of distributions. For instance, the so-called *Embrechts–Goldie* class of distributions, denoted $\mathcal{L}(\lambda)$, consist of nonnegative distributions H having the property

$$\lim_{s\to\infty} \frac{\overline{H}(s-u)}{\overline{H}(s)} = \mathrm{e}^{\lambda u}, \quad \lambda \ge 0$$

holds for all u (Embrechts and Goldie, 1980). Distributions in the class $\mathcal{L}(0)$ are heavy-tailed and these are known as long-tailed distributions. In contrast, if $\lambda > 0$ then a distribution in the class $\mathcal{L}(\lambda)$ is light-tailed. As an example, the class of Weibullian distributions with parameters $p = 1$ is contained in $\mathcal{L}(1)$. From Proposition 2.1, is clear that a PH distribution is in $\mathcal{L}(\lambda)$ where $-\lambda$ is the largest eigenvalue of the sub-intensity matrix $\mathbf{\Lambda}$.

Another important subclass of heavy-tailed distributions is that of subexponential distributions. Such a class of distributions contains practically all the heavy-tailed distributions commonly used. We say that H belongs to the class of subexponential distributions, denoted $H \in \mathcal{S}$, if

$$\limsup_{s\to\infty} \frac{\overline{H}^{*n}(s)}{\overline{H}(s)} = n,$$

where $\overline{H}^{*n}$ is the tail probability of the n-fold convolution of H. The defining property of the class of subexponential distributions has an extensive range of theoretical and practical implications (cf. Foss et al., 2011). Among these a very important one is the *principle of the single large jump*, which in heuristic terms says that the event where a sum of n iid subexponential random variables exceeds a large threshold is *most* likely a consequence of having a single random variable taking a very large value (jump).

In this paper, we are particularly interested in providing sufficient conditions for a phase-type scale mixture to be subexponential. However, the task of determining whether a given heavy-tailed distributions is subexponential or not, can be very challenging in some cases. We will resort to extreme value theory to address this issue since there exist a variety of results relating the subexponential property with maximum domains of attractions. First, we state the central theorem in extreme value theory (Fisher and Tippett, 1928; Gnedenko, 1943).

Theorem 2.5 (Fisher-Tippett-Gnedenko Theorem). *Let $\{S_1, \cdots S_n\}$ be a sequence of i.i.d. random variables with common distribution H. If there exist some proper sequences of (norming) constants $c_n > 0$ and $d_n \in \mathbb{R}$ and non-degenerate distribution M such that*

$$\frac{\max\{S_1, \cdots, S_n\} - d_n}{c_n} \xrightarrow{d} S, \quad \text{where } S \sim M,$$

then M can only belong to one of three distributions: Fréchet (Φ), Weibull (Ψ) or Gumbel (Λ). We say that H belongs to the maximum domain attraction of M: $H \in$MDA(M) and we call c_n and d_n the sequence of norming constants.

The Weibull domain of attraction is composed of distributions with support bounded above so a phase-type scale mixture cannot belong to such domain. The Fréchet domain of attraction is characterized by *Regular Variation* (de Haan, 1970). A distribution H is regularly varying with index $\alpha > 0$ if

$$\lim_{s\to\infty} \frac{\overline{H}(st)}{\overline{H}(s)} = t^{-\alpha}, \quad t > 0, \tag{2.2}$$

and denote $H \in \mathcal{R}_{-\alpha}$. Otherwise, if the limit above is 0 for all $t > 1$, then we say that H is a distribution of *rapid variation* and denote it $H \in \mathcal{R}_{-\infty}$ (cf. Bingham et al., 1987). Regular variation characterizes the Fréchet domain of attraction via the following relation

$$\overline{H} \in \mathcal{R}_{-\alpha} \quad \Longleftrightarrow \quad H \in \text{MDA}(\Phi_\alpha).$$

This characterisation is important because regularly varying distributions are subexponential.

The Gumbel domain of attraction is more involved. It contains both light and heavy tailed distributions. A main result in extreme value theory indicates that a distribution H belongs to the Gumbel domain of attraction iff $\overline{H}$ is tail-equivalent to a von Mises function. The following provides sufficient conditions for a distribution to be a von Mises function.

Theorem 2.6 (de Haan (1970)). *Let $\overline{H}$ be a twice differentiable nonnegative distribution with unbounded support. Then $\overline{H}$ is a von Mises function iff there exists s_0 such that $\overline{H}''(s) < 0$ for all $s > s_0$, and*

$$\lim_{s\to\infty} \frac{\overline{H}(s)\overline{H}''(s)}{\left(\overline{H}'(s)\right)^2} = -1. \tag{2.3}$$

Moreover, von Mises functions are functions of rapid variation (cf. Bingham et al., 1987).

Goldie and Resnick (1988) provide a sufficient condition for a distribution $H \in \text{MDA}(\Lambda)$ to be subexponential:

Theorem 2.7 (Goldie and Resnick (1988)). *Let $H \in \text{MDA}(\Lambda)$ be an absolutely continuous function with density h, then $H \in \mathcal{S}$ if*

$$\liminf_{s\to\infty} \frac{\overline{H}(ts)}{h(ts)} \frac{h(s)}{\overline{H}(s)} > 1, \qquad \forall t > 1. \tag{2.4}$$

Therefore, since a phase-type scale mixture distribution is not only absolutely continuous but twice differentiable and its second derivative is negative then we can verify if it belongs to the Gumbel domain of attraction by just checking the condition 2.3 in Theorem 2.6. Subexponentiality can be checked via Theorem 2.7. This will be the material of the following section.

3 Tail behavior of scaled random variables

This section is devoted to characterising the tail properties of the class of phase-type scale mixture distributions. Firstly, we collect some relevant results about the asymptotic tail behavior of products of random variables which provide sufficient conditions on the scaling random variable S such that its associated phase-type scale mixture distribution is either light or heavy-tailed. We extend this result to allow more general distributions and provide a simplified proof (Theorem 3.1).

Secondly, in Subsection 3.2 we focus on determining the maximum domain of attraction of a phase-type scale mixture distribution according to its scaling distribution. In the Fréchet case, Breiman's lemma implies that a phase-type scale mixture distribution remains in the Fréchet domain of attraction (hence regularly varying) if the scaling distribution is in the same domain. The converse of Breiman's lemma does not hold true in general, and finding sufficient conditions and counterexamples is considered challenging (cf. Damen et al., 2014; Denisov and Zwart, 2007; Jacobsen et al., 2009; Jessen and Mikosch, 2006). For the Gumbel case we provide conditions on the Laplace transform of reciprocal of the scaling random variable $1/S$ so the associated phase-type scale mixture distribution belongs to the Gumbel domain of attraction as well as to further determine if it is subexponential. We illustrate with examples that such conditions are verifiable in some important cases of interest.

3.1 Asymptotic tail behavior

The tail behavior of the distribution of a product of nonnegative random variables has attracted a considerable amount of research interest. For instance, Su and Chen (2006) show that if the distribution of a random variable S_1 is in $\mathcal{L}(\lambda)$ for $\lambda > 0$ and S_2 has unbounded support, then the distribution of $S_1 \cdot S_2$ is in $\mathcal{L}(0)$ (long-tailed), hence it is heavy-tailed (see also Tang, 2008). If one further assumes that S_2 is Weibullian with parameter $0 < p \leq 1$, then Liu and Tang (2010) show that the product $S_1 \cdot S_2$ is subexponential. A somewhat related result which extends beyond the class $\mathcal{L}(\gamma)$ is in Arendarczyk and Dębicki (2011) where it is shown that the product of two Weibullian random variables with parameters p_1 and p_2 is Weibullian with parameter $p_1p_2/(p_1+p_2)$. Therefore, the product of Weibullians can be either light- or heavy-tailed; furthermore, if the scaling distribution is Weibullian, then the phase-type scale mixture distribution is subexponential.

These results imply that a phase-type scale mixture distribution is heavy-tailed if and only if the scaling distribution has unbounded support. This conclusion can also be obtained from our Theorem 3.1 below where we provide sufficient conditions under which a product of two general random variables can be classified either as light- or heavy-tailed. The simplified proof provided here is elementary.

Theorem 3.1. *Consider S_1 and S_2 two nonnegative independent random variables with unbounded support, where $S_1 \sim H_1$ and $S_2 \sim H_2$. Let H be the distribution of the product $S_1 \cdot S_2$.*

1. *If there exist $\theta > 0$ and $\xi(x)$ a nonnegative function such that*
$$\limsup_{x\to\infty} \mathrm{e}^{\theta x}\Big(\overline{H}_1(x/\xi(x)) + \overline{H}_2(\xi(x))\Big) = 0, \tag{3.1}$$
then H is a light-tailed distribution.
2. *If there exists $\xi(x)$ a nonnegative function such that for all $\theta > 0$ it holds that*
$$\limsup_{x\to\infty}\ \mathrm{e}^{\theta x}\,\overline{H}_1(x/\xi(x)) \cdot \overline{H}_2(\xi(x)) = \infty, \tag{3.2}$$
then H is a heavy-tailed distribution.

Proof. For the first part consider

$$\begin{aligned}
\limsup_{x\to\infty} \overline{H}(x)\mathrm{e}^{\theta x} &= \limsup_{x\to\infty} \mathrm{e}^{\theta x}\int_0^\infty \overline{H}_1(x/s)\mathrm{d}H_2(s)\\
&= \limsup_{x\to\infty}\left[\mathrm{e}^{\theta x}\int_0^{\xi(x)} \overline{H}_1(x/s)\mathrm{d}H_2(s) + \mathrm{e}^{\theta x}\int_{\xi(x)}^\infty \overline{H}_1(x/s)\mathrm{d}H_2(s)\right]\\
&\leq \limsup_{x\to\infty}\left[\mathrm{e}^{\theta x}\overline{H}_1(x/\xi(x)) + \mathrm{e}^{\theta x}\overline{H}_2(\xi(x))\right] = 0.
\end{aligned}$$

The last equality holds by the hypothesis (3.1). Hence H is light-tailed. For the second part consider

$$\begin{aligned}
\limsup_{x\to\infty} \overline{H}(x)\mathrm{e}^{\theta x} &= \limsup_{x\to\infty}\left[\mathrm{e}^{\theta x}\int_0^{\xi(x)} \overline{H}_1(x/s)\mathrm{d}H_2(s) + \mathrm{e}^{\theta x}\int_{\xi(x)}^\infty \overline{H}_1(x/s)\mathrm{d}H_2(s)\right]\\
&\geq \limsup_{x\to\infty}\left[\mathrm{e}^{\theta x}\overline{H}_1(x/\xi(x))\overline{H}_2(\xi(x))\right] = \infty.
\end{aligned}$$

The last equality holds by hypothesis (3.2). Hence H is heavy-tailed. □

The conditions in Theorem 3.1 can be easily verified and enables us to provide a classification of the asymptotic tail behaviour of products of random variables with more general distributions. . Notice that the distributions considered in Su and Chen (2006) correspond to distributions with log-tail probabilities decaying at a linear rate, i.e. $-\log \overline{H_1}(s) = O(s)$, while the distributions in Arendarczyk and Dębicki (2011) have log-tail probabilities decaying at a power rate, i.e. $-\log \overline{H_i}(s) = O(s^{p_i})$, $i = 1, 2$. The following example considers distributions with log-tail probabilities decaying at an exponential rate, i.e. $-\log \overline{H_i}(s) = O(\mathrm{e}^s)$.

Example 3.2 (Gumbellian Products)**.** Let $H_i(x) = 1-\exp\{-\mathrm{e}^x+1\}$, $x > 0$. We choose $\xi(x) = x^\gamma$, with $0 < \gamma < 1$. Then

$$\lim_{x\to\infty} \overline{H}(x)\mathrm{e}^{\theta x} = \lim_{x\to\infty} \mathrm{e}^{\theta x+1}\left(\exp\{-\mathrm{e}^{x^{1-\gamma}}\} + \exp\{-\mathrm{e}^{x^\gamma}\}\right) = 0, \quad \forall\theta > 0.$$

Then the product of two random variables with Gumbellian-type distributions is always light-tailed. The same holds true if we replace H_2 with a Weibullian distribution with shape parameter $p > 1$. Choose $\xi(x) = x^\gamma$, with $\frac{1}{q} \le \gamma < 1$ and observe that

$$\lim_{x\to\infty} \overline{H}(x)\mathrm{e}^{\theta x} = \lim_{x\to\infty} \mathrm{e}^{\theta x}\left(\exp\{-\mathrm{e}^{x^{1-\gamma}} + 1\} + x^\delta \mathrm{e}^{-x^{\gamma p}}\right) = 0, \quad \text{for } \theta \in (0, 1).$$

3.2 Maximum domains of attraction and Subexponentiality

The scenario in the Frechét domain of attraction is well understood. Breiman's lemma (Breiman, 1965) implies that a phase-type scale mixture distribution is in the Frechét domain of attraction if its scaling distribution is in the same domain:

Lemma 3.3 (Breiman (1965))**.** *If $H \in \mathcal{R}_{-\alpha}$ and $M_G(\alpha+\epsilon) < \infty$ for some $\epsilon > 0$, then $F \in \mathcal{R}_{-\alpha}$ and*

$$\overline{F}\,(x) = M_G(\alpha)\overline{H}(x)(1 + o(1)), \qquad x \to \infty, \tag{3.3}$$

where $M_G(\alpha)$ is the α-moment of G.

Phase-type distributions are light-tailed so all their moments are finite. Therefore, a phase-type scale mixture distribution with a scaling distribution in the Fréchet domain of attraction remains in the same domain. Furthermore, the norming constants for a phase-type scale mixture distribution F can be chosen as the norming constants of H divided by the α-moment of the phase-type distribution G, that is

$$d_n = 0, \quad c_n = \frac{1}{M_G(\alpha)}\left(\frac{1}{\overline{H}}\right)^{\leftarrow}(n).$$

Moreover, when the condition of Breiman's lemma are satisfied, then the two distributions are regularly varying with the same parameter of regular variation thus implying that the tail probabilities of both distributions are asymptotically proportional (with the reciprocal of the α–moment of the phase-type distribution being the proportionality constant). This suggests that the class of phase-type scale distributions can provide good approximations of Regularly Varying distributions.

It is interesting to note that the converse of Breiman's lemma does not hold true in general. Such a problem is considered to be challenging and has attracted considerable research interest thus resulting in a rich variety of results proving sufficient conditions and counterexamples; for instance, Jessen and Mikosch (2006) provide a comprehensive list of earlier references; the most general results are given in Jacobsen et al. (2009) and Denisov and Zwart (2007) (see also Damen et al. (2014) for a multivariate version). It is not difficult to verify that some subclasses (for instance, exponential distribution, Erlang distribution, hyperexponential distribution) of PH distributions satisfy the sufficient conditions for the converse of Breiman's Lemma provided in previous results Jacobsen et al. (2009). We also conjecture that in general PH distributions satisfy the above conditions.

The situation is less understood in the Gumbel domain of attraction. We start by noting that in the Gumbel case, a phase-type scale mixture F and its scaling distribution H will have very different tail behaviors (this is contrast to the Frechét case where Breiman's lemma implies that these have asymptotically proportional tail behaviour). In particular, the tail probability of a scaling distribution in the Gumbel domain of attraction is tail equivalent to a von Mises functions, hence rapidly varying. In such a case the tail distribution of the phase-type scale mixture will be much heavier than its scaling distribution:

Proposition 3.4. *If $H \in \mathcal{R}_{-\infty}$, then*

$$\limsup_{x\to\infty} \frac{\overline{H}(x)}{\overline{F}(x)} = 0. \tag{3.4}$$

Proof. To show this we take $t > 1$ and observe that there exists a constant C such that

$$\overline{F}(x) = \mathbb{P}[SY > x] \geq \mathbb{P}[SY > x, Y \geq t] \geq \mathbb{P}[S > x/t]\mathbb{P}[Y \geq t] = \overline{H}(x/t)C,$$

Then

$$\limsup_{x\to\infty} \frac{\overline{H}(x)}{\overline{F}(x)} \leq \frac{1}{C} \limsup_{x\to\infty} \frac{\overline{H}(x)}{\overline{H}(x/t)} = 0, \quad t > 1.$$

□

The Lognormal and Weibullian distributions are rapidly varying.

Remark 3.5. This result fleshes out a limitation of the aforementioned approach for approximating distributions in the Gumbel domain of attraction since the tail probabilities will be heavier if the scaling distributions are chosen within a family of distributions asymptotically equivalent to the target distribution.

Next we look for sufficient conditions of the scaling distribution so its corresponding phase-type scale mixture will belong to the Gumbel domain of attraction and further Subexponential. We restrict our focus to phase-type distributions with sub-intensity matrices having only real eigenvalues.

Theorem 3.6. *Let $V(x) = (-1)^{\eta-1}\mathcal{L}^{(\eta-1)}_{1/S}(x)$ where η is the largest dimension among the Jordan blocks associated to the largest eigenvalue of the sub-intensity matrix. If $V(\cdot)$ is a von Mises function, then $F \in \mathrm{MDA}(\Lambda)$. Moreover, F is subexponential if*

$$\liminf_{x\to\infty} \frac{V(tx)V'(x)}{V'(tx)V(x)} > 1, \qquad \forall t > 1.$$

Proof. We can write that

$$\overline{F}(x) = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} \int_0^\infty c_{jk}\left(\frac{x}{s}\right)^k \mathrm{e}^{-\lambda_j x/s}\mathrm{d}H(s) = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk}\,\frac{(-1)^k x^k}{\lambda_j^k}\,\mathcal{L}^{(k)}_{1/S}(\lambda_j x).$$

Since $V(x) = (-1)^{\eta-1}\mathcal{L}^{(\eta-1)}_{1/S}(x)$ is a von Mises function, then $V(x)$ is of rapid variation (Bingham et al., 1987). This implies that

$$\overline{F}(x) \sim c\frac{x^{\eta-1}}{\lambda^{\eta-1}}V(\lambda x), \tag{3.5}$$

where c is some constant, $-\lambda$ is the largest eigenvalue of the sub-intensity matrix and η is the largest dimension among the Jordan blocks associated to $-\lambda$. Then it is not difficult to see that

$$\lim_{x\to\infty} \frac{\overline{F}(x)F''(x)}{(F'(x))^2} = \lim_{x\to\infty} \frac{V(\lambda x)(-V''(\lambda x))}{\big(-V'(\lambda x)\big)^2} = -1.$$

This holds true because by hypothesis $V(x) = (-1)^{\eta-1}\mathcal{L}^{(\eta-1)}_{1/S}(x)$ is a von Mises function. Hence $F \in \mathrm{MDA}(\Lambda)$ and the first part result follows. For the second part, we observe that the auxiliary function $a(x) = \overline{F}(x)/F'(x)$ obeys the following asymptotic equivalence

$$a(x) = \frac{\overline{F}(x)}{F'(x)} \sim \frac{V(\lambda x)}{-\lambda V'(\lambda x)}.$$

The distribution F is subexponential iff

$$\liminf_{x\to\infty} \frac{a(tx)}{a(x)} = \liminf_{x\to\infty} \frac{V(\lambda tx)V'(\lambda x)}{V'(\lambda tx)V(\lambda x)} > 1, \qquad \forall t > 1,$$

hence subexponentiality of F follows.

□

Theorem 3.6 can be applied to the Lognormal case:

Example 3.7 (Lognormal scaling)**.** Assume $H \sim \mathrm{LN}(\mu, \sigma^2)$, then F is a subexponential distribution in the Gumbel domain of attraction.

Proof. W.l.o.g. we can assume $\mu = 0$ since e^{μ} is a scaling constant. In such a case the symmetry of the normal distribution implies that the Laplace–Stieltjes transform of $1/S$ is the same as that of S, i.e.

$$\mathcal{L}_{1/S}(x) = \mathcal{L}_S(x).$$

An asymptotic approximation of the k-th derivative of the Laplace–Stieltjes transform of the lognormal distribution is given in Asmussen et al. (2016):

$$\mathcal{L}_S^{(k)}(x) = (-1)^k \mathcal{L}_S(x) \exp\{-k\omega_0(x) + \frac{1}{2}\sigma_0(x)^2 k^2\}(1 + o(1)),$$

where

$$\omega_k(x) = \mathcal{W}(x\sigma^2 \mathrm{e}^{k\sigma^2}), \quad \sigma_k(x)^2 = \frac{\sigma^2}{1 + \omega_k(x)},$$

and $\mathcal{W}(\cdot)$ is the Lambert W function. Hence We verify that

$$\begin{aligned} \lim_{x\to\infty} \frac{V(x)(-V''(x))}{(-V'(x))^2} &= \lim_{x\to\infty} \frac{\mathrm{e}^{-(\eta-1)\omega_0(x)+\frac{1}{2}\sigma_0(x)^2(\eta-1)^2} \cdot \left(-\mathrm{e}^{-(\eta+1)\omega_0(x)+\frac{1}{2}\sigma_0(x)^2(\eta+1)^2}\right)}{\mathrm{e}^{-2\eta\omega_0(x)+\sigma_0(x)^2\eta^2}} \\ &= -\lim_{x\to\infty} \exp\{\sigma_0(x)^2\} = -\lim_{x\to\infty} \exp\left\{\frac{\sigma^2}{1 + \omega_0(x)}\right\}. \end{aligned}$$

As $\omega_k(x)$ is asymptotically of order $\log(x)$ as $x \to \infty$, then $\sigma^2(1 + \omega_0(x))^{-1} \to 0$ as $x \to \infty$. Then the last limit is equal to -1 so we have shown that $F(x) \in \mathrm{MDA}(\Lambda)$. Furthermore,

$$\begin{aligned} \lim_{x\to\infty} \frac{a(tx)}{a(x)} &= \lim_{x\to\infty} \frac{(-1)^{\eta-1}\mathcal{L}_{1/S}^{(\eta-1)}(tx) \cdot (-1)^{\eta-1}\mathcal{L}_{1/S}^{(\eta)}(x)}{(-1)^{\eta-1}\mathcal{L}_{1/S}^{(\eta)}(tx) \cdot (-1)^{\eta-1}\mathcal{L}_{1/S}^{(\eta-1)}(x)} \\ &= \lim_{x\to\infty} \frac{\mathrm{e}^{-(\eta-1)\omega_0(xt)+\frac{1}{2}\sigma_0(xt)^2(\eta-1)^2} \cdot \mathrm{e}^{-\eta\omega_0(x)+\frac{1}{2}\sigma_0(x)^2\eta^2}}{\mathrm{e}^{-\eta\omega_0(xt)+\frac{1}{2}\sigma_0(xt)^2\eta^2} \cdot \mathrm{e}^{-(\eta-1)\omega_0(x)+\frac{1}{2}\sigma_0(x)^2(\eta-1)^2}} \\ &= \lim_{x\to\infty} \exp\left\{-\omega_0(x) + \omega_0(xt) + \frac{1}{2}\sigma_0(xt)^2(2\eta - 1) + \frac{1}{2}\sigma_0(x)^2(1 - 2\eta)\right\} \\ &= \lim_{x\to\infty} \exp\left\{-\omega_0(x) + \omega_0(x) + \omega_0(t) + O(\omega_0(x)^{-1})\right\} = t > 1. \end{aligned}$$

Thus F is a subexponential distribution. □

Example 3.8 (Exponential scaling)**.** Let $H \sim \exp(\beta)$. Then F is a subexponential distribution in the Gumbel domain of attraction.

Proof. Observe that $1/S$ has an inverse Gamma distribution with a Laplace–Stieltjes transform given in terms of a modified Bessel function of the second kind (Ragab, 1965):

$$\mathcal{L}_{1/S}(x) = \int_0^\infty \mathrm{e}^{-x/s}\beta\mathrm{e}^{-\beta s}\mathrm{d}s = 2\sqrt{\beta x}\mathrm{BesselK}(1, 2\sqrt{\beta x}).$$

Furthermore, its n-th derivative can be calculated explicitly also in the terms of a modified Bessel function of the second kind:

$$\mathcal{L}_{1/S}^{(n)}(x) = \int_0^\infty \left(-\frac{1}{s}\right)^n \mathrm{e}^{-x/s}\beta\mathrm{e}^{-\beta s}\mathrm{d}s = (-1)^n \cdot 2\,\beta^{\frac{n+1}{2}} x^{-\frac{n-1}{2}} \mathrm{BesselK}(n-1, 2\sqrt{\beta x}).$$

Asymptotically it holds true that

$$\mathcal{L}_{1/S}^{(n)}(x) \sim (-1)^n \sqrt{\pi}\beta^{\frac{2n+1}{4}} x^{-\frac{2n-1}{4}} \mathrm{e}^{-2\sqrt{\beta x}}, \quad x \to \infty.$$

Hence, it follows that

$$\lim_{x\to\infty} \frac{V(x)(-V''(x))}{(-V'(x))^2} = -1.$$

Therefore $V(x)$ is a von Mises function and $F \in \mathrm{MDA}(\Lambda)$. Moreover, if $t > 1$ then

$$\lim_{x\to\infty} \frac{a(tx)}{a(x)} = \lim_{x\to\infty} \frac{V(tx)V'(x)}{V'(tx)V(x)} = \sqrt{t} > 1.$$

Thus F is subexponential distribution.

□

Remark 3.9. Notice that it is possible to generalize the result of the previous example for a Gamma scaling distributions because an expression for the Laplace–Stieltjes transform of an inverse gamma distribution is known and given in terms of a modified Bessel function of the second kind. However, it involves a number of tedious calculations and therefore omitted. Note as well that in such a case it is possible to test directly if $\overline{F}$ is a von Mises function, but the calculations become cumbersome. Finally, we remark that the results of Liu and Tang (2010) imply the subexponentiality of the Exponential case.

Remark 3.10. When H is a discrete scaling distribution, we can also obtain an analogue result of Theorem 3.6. Define

$$\mathcal{DL}_{1/S}(x) = \sum_{i=1}^{\infty} \mathrm{e}^{-x/i} p(i).$$

as the discrete Laplace–Stieltjes transform of discrete scaling random variable S with probability massive function $p(i)$. Then the tail probability of the phase-type scale mixture is:

$$\overline{F}\,(x) = \sum_{j=1}^{m} \sum_{k=0}^{\eta_j-1} \sum_{i=1}^{\infty} c_{jk} \left(\frac{x}{i}\right)^k \mathrm{e}^{-\lambda_j x/i} p(i) = \sum_{j=1}^{m} \sum_{k=0}^{\eta_j-1} c_{jk}\, \frac{(-1)^k x^k}{\lambda_j^k}\, \mathcal{DL}_{1/S}^{(k)}(\lambda_j x).$$

If $V(x) = (-1)^{\eta-1}\mathcal{DL}_{1/S}^{(\eta-1)}(x)$ is a von Mises function, then $F \in$MDA(Λ).

4 Discrete Scaling Distributions

Next we focus on the case of phase-type scale mixture distributions having scaling distributions supported over countable sets of strictly positive numbers. These distributions are particularly tractable since these correspond to distributions of absorption times of Markov jump processes with an infinite number of transient states. As a consequence, these can be ideal for modelling heavy-tailed phenomena (cf. Bladt et al., 2015; Bladt and Rojas-Nandayapa, 2017; Peralta et al., 2016).

We remark however, that some of the methodologies for determining domains of attractions and subexponentiality described in the previous section are not always implementable in a straightforward way for discrete scaling distributions. One of the main difficulties is the calculation of asymptotic equivalent expressions for the infinite series defining the tail probabilities. Below we describe a simple methodology which can be used to extend results for continuous scaling distributions to their discrete scaling distributions counterparts; such a methodology provides mild conditions under which the asymptotical behavior of an infinite series is asymptotically equivalent to certain analogue integral.

Remark 4.1. Let $I_u : \mathbf{Z}^+ \to \mathbb{R}^+$ be collection of functions indexed by $u \in (0, \infty)$. Suppose that for each $u > 0$ there exist a measurable and bounded function $I'_u : \mathbb{R}^+ \to \mathbb{R}$ such that $I(u, k) = I'(u, k)$ for all $k \in \mathbb{Z}^+$ and

$$\int_0^\infty I'(u; y)\mathrm{d}y - M(u) \le \sum_{k=0}^\infty I(u; k) \le \int_0^\infty I'(u; y)\mathrm{d}y + M(u).$$

where $M(u) \ge \max\{I'(u, y) : y > 0\}$ is some upper bound for the function $I'(u, y)$. If

$$\lim_{u\to\infty} \frac{M(u)}{\int_0^\infty I'(u; y)\mathrm{d}y} = 0,$$

then the following asymptotic relationship holds

$$\lim_{u\to\infty} \frac{\sum_{k=0}^\infty I(u; k)}{\int_0^\infty I'(u; y)\mathrm{d}y} = 1.$$

The method provides a verifiable condition under which the infinite series can be replaced by an asymptotic integral. The next example is taken from Bladt et al. (2015). We remark that Breiman's lemma has been used to determine its asymptotic behaviour because the tail probability $\overline{H}(i)$, $i = 1, 2, \ldots$ forms a regularly varying sequence, so $\overline{H} \in \mathcal{R}_{-\alpha}$ (Bingham et al., 1987). We have included it here to illustrate the simplicity of the method proposed.

Example 4.2 (Zeta scaling)**.** Let $\alpha \ge 2$ and assume $H \sim \text{Zeta}(\alpha)$. Such a distribution is determined by $p(i) = i^{-\alpha}/\xi(\alpha)$, $i \in \mathbb{N}$ and $\xi(\cdot)$ is the Riemann zeta function. Then F is in the Fréchet domain of attraction.

Proof. The tail probability of F is given by

$$\overline{F}\,(x) = \sum_{i=1}^{\infty}\sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk}\left(\frac{x}{i}\right)^k \mathrm{e}^{-\lambda_j x/i}\frac{i^{-\alpha}}{\xi(\alpha)} = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1}\sum_{i=1}^{\infty} \frac{c_{jk}x^k}{\xi(\alpha)} i^{-(\alpha+k)}\mathrm{e}^{-\lambda_j x/i}.$$

Consider the functions $I'_{jk}(x; y) = x^k y^{-(\alpha+k)}\mathrm{e}^{-\lambda_j x/y}$ and note that each of these functions attains its single local maximum at $\hat{y} = \lambda_j x(\alpha + k)^{-1} > 0$, for all $x > 0$. Therefore

$$\int_0^\infty I'_{jk}(x; y)\mathrm{d}y - M_{jk}(x; \hat{y}) \le \sum_{i=1}^\infty x^k i^{-(\alpha+k)}\mathrm{e}^{-\lambda_j x/i} \le \int_0^\infty I'_{jk}(x; y)\mathrm{d}y + M_{jk}(x; \hat{y}).$$

Observe that

$$M_{jk}(x; \hat{y}) = x^k \mathrm{e}^{-(\alpha+k)}\left(\frac{\lambda_j}{\alpha + k}\right)^{-(\alpha+k)} x^{-(\alpha+k)} = cx^{-\alpha},$$

and

$$I'_{jk}(x) := x^k \int_0^\infty y^{-(\alpha+k)}\mathrm{e}^{-\lambda_j x/y}\mathrm{d}y = \frac{\Gamma(\alpha + k - 1)}{\lambda^{\alpha+k-1}} x^{-\alpha+1},$$

so $M_{jk}(x; \hat{y})$ is of negligible order with respect to $I'_{jk}(x)$. Then it follows that

$$\overline{F}\,(x) \sim \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} \frac{c_{jk}}{\xi(\alpha)} I_{jk}(x) = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} \frac{c_{jk}\Gamma(\alpha + k - 1)}{\xi(\alpha)\lambda^{\alpha+k-1}} x^{-\alpha+1}, \quad x \to \infty.$$

Thus $F(x) \in$MDA$(\Phi_{\alpha-1})$. Let $C = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} \frac{c_{jk}\Gamma(\alpha + k - 1)}{\xi(\alpha)\lambda^{\alpha+k-1}}$, then the norming constants can be chosen as

$$d_n = 0, \quad c_n = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n) = \left(\frac{C}{n}\right)^{\frac{1}{\alpha-1}}.$$

□

Example 4.3 (Geometric scaling)**.** Let $H \sim \text{Geo}(p)$ and G be PH distribution whose subintensity matrix has only real eigenvalues. Then F is a subexponential distribution in the Gumbel domain of attraction.

Proof. We let $p(i) = pq^i$ where $q = 1-p$. Since the geometric distribution has unbounded support, then the associated phase-type scale mixture is heavy-tailed. We next verify that it belongs to the Gumbel domain of attraction.

$$\overline{F}(x) = \sum_{i=1}^{\infty} \overline{G}\left(\frac{x}{i}\right) pq^i$$

Let $I'(x;y) = \overline{G}(x/y)p\exp\{-|\log q|y\}$ satisfies the conditions in Remark 4.1. Since the sine and cosine functions are bounded, then it is not difficult to use Proposition 2.1 to show that there exist a constant c_1 such that

$$M(x) := I(x;\hat{y}) \leq x^{\frac{k}{2}} \mathrm{e}^{-2\sqrt{x\lambda|\log q|}}(c_1 + o(1)), \qquad x \to \infty.$$

where λ is the eigenvalue in absolute value and k its largest multiplicity. If the subintensity matrix has real eigenvalues then by using Lemma 2.1 in (Arendarczyk and Dębicki, 2011) we obtain that

$$\int_0^\infty I'(x;y)\mathrm{d}y = \int_0^\infty \overline{G}(x/y)\mathrm{e}^{-y|\log q|}\mathrm{d}y = x^{k/2+1/4}\mathrm{e}^{-2\sqrt{x\lambda|\log q|}}(C_1 + o(1)), \qquad x \to \infty.$$

So, the value of $g(\hat{y};x)$ is asymptotically negligible with respect to the value of the integral and we conclude that

$$\overline{F}(x) \sim p\int_0^\infty \overline{G}(x/s)\mathrm{e}^{-y|\log q|}\mathrm{d}y = \frac{p}{|\log q|}\int_0^\infty \overline{G}(x/y)\mathrm{d}H(y)$$

where $H \sim \exp(|\log q|)$. Hence, by tail equivalence, the distribution F inherits all the asymptotic properties of its continuous counterpart, namely, a phase-type scale distribution with exponential scaling distribution with parameter $|\log q|$. □

Remark 4.4. We shall recall that the geometric version can be seen as the discrete counterpart of the exponential distribution obtained as a discretization. More precisely, the geometric distribution can be seen as a distribution supported over $\mathbb{Z}^+$ and defined by

$$H(k) = \mathcal{H}(k), \quad k = 0, 1, 2, \cdots,$$

where $\mathcal{H} \sim \exp(|\log q|)$. The probability mass function of H is given by $h(k) = \mathcal{H}(k) - \mathcal{H}(k-1)$.

This idea can be extended in order to select scaling distributions for approximating heavy-tailed distributions in the Gumbel domain of attraction. Suppose we want to approximate the tail probability of an absolutely continuous distribution $\mathcal{H}$ supported over $(0,\infty)$ via a discrete phase-type scale mixture distribution. One way to proceed is to construct a discrete distribution supported over $\mathbb{N}$ defined by $H(k) = \mathcal{H}(k) - \mathcal{H}(k-1)$; we refer to this construction as a *discretization* of $\mathcal{H}$. Moreover, the density of $\mathcal{H}$ can be used to construct a function $I'(u,k)$. In such a case the tail behavior of a phase-type scale mixture having a discretized scaling distribution inherits the asymptotic properties of its continuous counterpart.

This idea is better illustrated with the following example which suggest a methodology for approximating the tail probability of a Lognormal distribution.

Example 4.5 (Lognormal discretization)**.** Let H be a discrete Lognormal distribution with parameter μ, σ and supported over $\{0, 1, 2, \cdots\}$. Assume $\mu = 0$. The tail probability $\overline{F}$ is given by

$$\overline{F}(x) = \sum_{i=1}^{\infty} \overline{G}(x/i)\left[H(i) - H(i-1)\right] = \sum_{i=1}^{\infty} \overline{G}(x/i)\int_{i-1}^{i} h(y)\mathrm{d}y,$$

where $h(\cdot)$ is the density of Lognormal distribution. Since $\overline{G}(x/y)$ is increasing in y, then we can easily find a lower bound:

$$\overline{F}(x) = \sum_{i=1}^{\infty}\int_{i-1}^{i} \overline{G}(x/i)h(y)\mathrm{d}y \geq \int_0^\infty \overline{G}(x/y)h(y)\mathrm{d}y.$$

For the upper bound, we have

$$\overline{F}(x) \leq \sum_{i=1}^{\infty} \int_{i-1}^{i} \overline{G}(x/(y+1))h(y)\mathrm{d}y = \sum_{i=1}^{\infty} \int_{i-1}^{i} \overline{G}(x/(y+1))[h(y)-h(y+1)+h(y+1)]\mathrm{d}y$$
$$\leq \int_0^{\infty} \overline{G}(x/y)h(y)\mathrm{d}y + \int_0^{\infty} \overline{G}(x/(y+1)[h(y)-h(y+1)])\mathrm{d}y.$$

For the second integral in the above, we have

$$\int_0^{\infty} \overline{G}(x/(y+1)[h(y)-h(y+1)])\mathrm{d}y$$
$$= \int_0^{1} \overline{G}(x/(y+1)[h(y)-h(y+1)])\mathrm{d}y + \int_1^{\infty} \overline{G}(x/(y+1)[h(y)-h(y+1)])\mathrm{d}y$$
$$\leq c_1\overline{G}(x/2) + c_2 \int_1^{\infty} \overline{G}(x/(y+1)\frac{h(y+1)}{(y+1)^{\beta}}\mathrm{d}y,$$

where c_1, $c_2 > 0$ are some constants and $0 < \beta < 1$.

It is not difficult to obtain this upper bound: firstly, it is easy to prove for $y \geq 1$, $\log(y+1) - \log(y) \leq 1/y$, consequently, $\log^2(y+1) - \log^2(y) \leq 2\log(y+1)/y$; then we have

$$\frac{h(y)}{h(y+1)} - 1 = \frac{y+1}{y} \exp\left\{\frac{\log^2(y+1) - \log^2(y)}{2\sigma^2}\right\} - 1$$
$$\leq \exp\left\{\frac{1}{y} + \frac{\log(y+1)}{\sigma^2 y}\right\} - 1$$
$$\leq c\left(\frac{1}{y} + \frac{\log(y+1)}{\sigma^2 y}\right), \text{where } c > 0 \text{ is some constant},$$
$$\leq \frac{c_1}{(y+1)^{\beta}}.$$

Define

$$I'_{jk}(x) := x^k \int_0^{\infty} y^{-k} \mathrm{e}^{-\lambda_j x/y} h(y)\mathrm{d}y.$$

From Example 3.7, we know that

$$\int_0^{\infty} \overline{G}(x/y)h(y)\mathrm{d}y = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk} \int_0^{\infty} (x/y)^k \mathrm{e}^{-\lambda_j x/y} h(y)\mathrm{d}y = \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk} \frac{(-1)^k x^k}{\lambda_j^k} \mathcal{L}_Y^{(k)}(\lambda_j x),$$
$$= \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk} \frac{x^k}{\lambda_j^k} \mathcal{L}_Y \exp\{-k\omega_0(\lambda_j x) + \frac{1}{2}\sigma_0(\lambda_j x)^2 k^2\}$$

So

$$I'_{jk}(x) = (-1)^k \left(\frac{x}{\lambda_j}\right)^k \mathcal{L}_Y \exp\{-k\omega_0(\lambda_j x) + \frac{1}{2}\sigma_0(\lambda_j x)^2 k^2\}$$

It is obvious that $c_1\overline{G}(x/2)$ vanishes faster than $I'_{jk}(x)$, so we can define

$$M_{jk}(x) := x^k \int_0^{\infty} y^{-k-\beta} \mathrm{e}^{-\lambda_j x/y} h(y)\mathrm{d}y$$

since

$$c_2 \int_1^{\infty} \overline{G}(x/(y+1))\frac{h(y+1)}{(y+1)^{\beta}}\mathrm{d}y = c_2 \int_2^{\infty} \overline{G}(x/y)\frac{h(y)}{y^{\beta}}\mathrm{d}y$$
$$\leq \sum_{j=1}^{m}\sum_{k=0}^{\eta_j-1} c_{jk} \int_0^{\infty} (x/y)^k y^{-\beta} \mathrm{e}^{-\lambda_j x/y} h(y)\mathrm{d}y$$

By a similar approximation as in Example 3.7, we can see

$$\begin{aligned} M_{jk}(x) &= (-1)^{k+\beta} \frac{x^k}{\lambda_j^{k+\beta}} \mathcal{L}_Y^{(k+\beta)}(\lambda_j x) \\ &= (-1)^{k+\beta} \frac{x^k}{\lambda_j^{k+\beta}} \mathcal{L}_Y \exp\{-(k+\beta)\omega_0(\lambda_j x) + \frac{1}{2}\sigma_0(\lambda_j x)^2(k+\beta)^2\} \end{aligned}$$

So $M_{jk}(x)$ is negligible compared to integral $I'_{jk}(x)$. Thus, the phase-type scale mixture with discrete Lognormal scaling distribution has the same asymptotic behavior as the one having continuous counterpart.

4.1 Non-lattice supports

The examples in the previous subsection may suggest that a phase-type scale mixture having a discretized scaling distribution will inherit the asymptotic properties of its continuous counterpart. However, such a discretization cannot be made arbitrarily. The following example illustrates this fact.

Example 4.6. Let $H \in \mathcal{R}_{-\alpha}$ be a continuous distribution and S be a discrete random variable supported over $\{s_1, s_2, \cdots\}$ satisfying

$$\mathbb{P}(S = s_i) = H(s_i) - H(s_{i-1}), \quad i = 1, 2, \cdots .$$

Suppose there exists $\epsilon > 0$, $i_0 \in \mathbb{N}$ such that $\forall i > i_0$ it holds that $s_{i+1} > s_i(1+\epsilon)$.

$$\limsup_{x\to\infty} \frac{\mathbb{P}[S > (1+\epsilon)x]}{\mathbb{P}[S > x]} = \limsup_{i\to\infty} \frac{\mathbb{P}[S > (1+\epsilon)s_i]}{\mathbb{P}[S > s_i]} = \limsup_{i\to\infty} \frac{\mathbb{P}[S > s_i]}{\mathbb{P}[S > s_i]} = 1.$$

Then S does not have a regularly varying distribution. Suppose that $Y \sim \text{Erlang}(\lambda, k)$. According to Example 4.4 in Jacobsen et al. (2009), the distribution of phase-type scale mixture random variable $S \cdot Y$ is not Regularly Varying.

Nevertheless, such a discretization will provide a *reasonable* approximation to a Regularly Distribution. The following is a continuation of our previous example and it shows that such a distribution satisfies an analogue of Breiman's lemma.

Example 4.7. Let $K > 0$ and define H_K a discrete distribution supported over $\{s_i : i \in \mathbb{Z}^+\}$, where $s_i = \exp(i/K)$, and determined by

$$H_K(s_i) = 1 - s_i^{-\alpha}, \qquad \forall i \in \mathbb{Z}^+.$$

The distribution H_k can be seen as a discretization over a geometric progression of a Pareto distribution having tail probability $H(x) = x^{-\alpha}$ supported over $[1, \infty)$. The following argument shows that H_K is no longer a regularly varying distribution. Notice that for all $t > 1$ there exist $n \in \mathbb{Z}^+$ such that $s_n < t \leq s_{n+1}$, hence

$$\liminf_{x\to\infty} \frac{\overline{H}_K(xt)}{\overline{H}_K(x)} = s_{n+1}^{-\alpha}, \qquad \limsup_{x\to\infty} \frac{\overline{H}_K(xt)}{\overline{H}_K(x)} = \begin{cases} s_n^{-\alpha} & t < s_{n+1} \\ s_{n+1}^{-\alpha} & t = s_{n+1}. \end{cases}$$

Thus, according to Example 4.4 in Jacobsen et al. (2009), the Mellin–Stieltjes convolution of an Erlang distribution G with the distribution H given above is no longer of Regular Variation (the conditions described in in Remark 4.1 are not satisfied for this example either). In spite of this, we can still analyse certain aspects of the asymptotic behaviour of such a Mellin–Stieltjes convolution. For that purpose note that the following inequalities hold for all $w > 1$

$$\mathrm{e}^{-\alpha/K}\overline{H}(w) < \overline{H}_K(w) \leq \overline{H}(w),$$

hence we obtain that

$$\mathrm{e}^{-\alpha/K} \int_0^\infty \overline{H}(x/s)dG(s) < \int_0^\infty \overline{H}_K(x/s)dG(s) \leq \int_0^\infty \overline{H}(x/s)dG(s).$$

Using Breiman's lemma we find that

$$\mathrm{e}^{-\alpha/K} < \liminf \frac{\overline{F}\,(x)}{M_G(\alpha)\overline{H}(x)} \leq \limsup \frac{\overline{F}\,(x)}{M_G(\alpha)\overline{H}(x)} \leq 1.$$

An heuristic interpretation of the inequalities above is that aysmptotically the tail probability $\overline{F}$ *oscillates* around a regularly varying tail distribution, so this example illustrates a behaviour similar to that described by Breiman's lemma. Notice that the range of oscillation collapses as $K \to \infty$ which is consistent with the fact that $H_K \to H$ weakly. A better asymptotic approximation can be found as follows which is particularly sharp for numerical purposes. Consider

$$\overline{F}\,(x) = \int_0^\infty \overline{G}\,\left(\frac{x}{s}\right) \mathrm{d}H_K(s) = (1 - \mathrm{e}^{-\alpha/K}) \sum_{i=0}^\infty \overline{G}\,(x\mathrm{e}^{-i/K})\mathrm{e}^{-\alpha i/K}$$

Let $I(t) = \overline{G}\,(x\mathrm{e}^{-y/K})\mathrm{e}^{-\alpha y/K}$. The infinite series can be approximated via the integral

$$\int_0^\infty I(y)\mathrm{d}y = \int_0^\infty \overline{G}\,\left(x\mathrm{e}^{-y/K}\right)\mathrm{e}^{-\alpha y/K}\mathrm{d}y = K\int_1^\infty \overline{G}\,\left(\frac{x}{w}\right) w^{-(\alpha+1)}\mathrm{d}w = \frac{K}{\alpha}\int_1^\infty \overline{G}\,\left(\frac{x}{w}\right)\mathrm{d}H(w).$$

Since G is such that $M_G(\alpha+\epsilon) < \infty$ for all $\epsilon > 0$ then Breiman's lemma implies that

$$\overline{F}\,(x) \approx \frac{1 - \mathrm{e}^{-\alpha/K}}{\alpha/K} M_G(\alpha)\overline{H}(x)$$

This approximation is consistent with the bounds found above since for all $w > 0$ it holds that

$$\mathrm{e}^{-w} \leq \frac{1 - \mathrm{e}^{-w}}{w} \leq 1.$$

Hence, the asymptotic approximation suggested is contained in between the asymptotic bounds previously found.

The previous example that the tail behaviour a phase-type scale mixture distribution having a discretized scaling distribution is clearly affected by the selection of the support. Naturally, better approximation of the tail can be obtaind as long as the mesh of the partition defining the support becomes smaller.

The natural choice is to use a discretization of the target distribution over some lattice. However, this approach is not always the more suitable for numerical purposes because in practice, there is only a finite number of terms of the infinite series that can be computed, so these series are typically truncated. By selecting a discretization over a geometric progression as a scaling distribution we will obtain an infinite series which converge at a faster rate. Moreover, such a geometric progression will provide a very reasonable approximation of the tail probability as shown above. This approach has been tested successfully in Peralta et al. (2016) where they considered discretizing Pareto distributions over a geometric progression and used the corresponding phase-type scale mixture distribution to approximate Pareto claim size distributions in ruin probability calculations.

5 Conclusion

We considered the class of phase-type scale mixtures. Such distributions arise from the product of two random variables $S \cdot Y$, where $S \sim H$ is a nonnegative random variable and $Y \sim G$ is a phase-type distribution. Such a class is mathematically tractable and can be used to approximate heavy-tailed distributions.

We provided a collection of results which can be used to determine the asymptotic behavior of a distribution in the class of phase-type scale mixture distributions. For instance, if the scaling distribution H has unbounded support, then the associated phase-type scale mixture distribution is heavy-tailed. We also provided verifiable conditions which can be employed to classify the maximum domain of attraction and determine subexponentiality.

The we specialized in the case of phase-type scale mixture distributions having discrete scaling distributions since these are of critical importance in recent applications. We described a simple methodology which allows to establish the asymptotic proportionality of these distributions with their counterparts having continuous distributions. We exhibited important advantages and limitations of this approach to approximate heavy–tailed distributions and considered several important examples.

We remark that most of the results obtained here can be extended to the *matrix exponential scale mixture distributions* without to much effort. Some of our results where proven under the assumption that the phase-type distribution has an subintensity matrix has only real eigenvalues. We conjecture that such results holds for general phase-type and matrix-exponential distributions. We also conjectured that a phase-type distribution is α-regularly varying determining.

Acknowledgements

LRN is supported by Australian Research Council (ARC) grant DE130100819. WX is supported by IPRS/APA scholarship at The University of Queensland.

References

Arendarczyk, M. and K. Dębicki (2011). Asymptotics of supremum distribution of a gaussian process over a weibullian time. *Bernoulli 17*(1), 194–210.

Asimit, A. V. and B. L. Jones (2006). Extreme behavior of multivariate phase-type distributions. *Insurance: Mathematics and Economics 41*(2), 223–233.

Asmussen, S. (2003). *Applied Probabilities and Queues.* New York: Springer.

Asmussen, S. and H. Albrecher (2000). *Ruin Probabilities* (2nd ed.). Singapore: World Scientific Publishing.

Asmussen, S., J. L. Jensen, and L. Rojas-Nandayapa (2016). Exponential family techniques for the lognormal left tail. *Scandinavian Journal of Statistics 43*(3), 774–787.

Asmussen, S., O. Nerman, and M. Olsson (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics 23*, 419–441.

Assaf, D. and B. Levikson (1982). Closure of phase type distributions under operations arising in reliability theory. *The Annals of Probability 10*(1), 265–269.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation.* Cambridge: Cambridge University Press.

Bladt, M., A. Campillo Navarro, and B. Nielsen (2015). On the use of functional calculus for phase-type and related distributions. *Stochastic Models 32*(1), 1–19.

Bladt, M., B. F. Nielsen, and G. Samorodnitsky (2015). Calculation of ruin probabilities for a dense class of heavy-tailed distributions. *Scandinavian Actuarial Journal 2015*(7), 573–591.

Bladt, M. and L. Rojas-Nandayapa (2017). Estimation of infinite dimensional phasetype distributions. Submitted.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability and its Applications 10*(2), 323–331.

Damen, E., T. Mikosch, J. Rosiński, and G. Samorodnitsky (2014). General inverse problems for regular variation. *Journal of Applied Probability 51A*, 229–248.

de Haan, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre tracts. Amsterdam: Mathematisch Centrum.

Denisov, D. and B. Zwart (2007). On a theorem of Breiman and a class of random difference equations. *Journal of Applied Probability 44*(4), 1031–1046.

Embrechts, P. and C. M. Goldie (1980). On closure and factorization properties of subexponential and related distributions. *Journal of the Australian Mathematical Society (Series A) 29*, 243–256.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society 24*(2), 180–190.

Foss, S., D. Korshunov, and S. Zachary (2011). *An introduction to heavy-tailed and subexponential distributions*. Springer.

Gnedenko, B. V. (1943). On the limiting distribution of the maximum term in a random series. *Breakthroughs in Statistics, Springer Series in Statistics 1*, 195–226.

Goldie, C. M. and S. I. Resnick (1988). Distributions that are both subexponential and in the domain of attraction of an extreme value distribution. *Advances in Applied Probability 20*(4), 706–718.

Greiner, M., M. Jobmann, and L. Lipsky (1999). The importance of power-tail distributions for modelling queueing systems. *Operations Research 47*(2), 313–326.

Jacobsen, M., T. Mikosch, and J. Rosiński (2009). Inverse problems fro regular variation of linear filters, a cancellation property for σ-finite meseures and identification of stable laws. *The Annals of Applied Probability 19*(1), 210–242.

Jessen, A. H. and T. Mikosch (2006). Regularly varying functions. *Publications de L'Institut Mathématique 79*(93), 171–192.

Kang, S. and R. F. Serfozo (1999). Extreme values of phase-type and mixed random variables with parallel-processing examples. *Journal of Applied Probability 39*(1), 194–210.

Latouche, G. and V. Ramaswami (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling.* Philadelphia,USA: American Statistical and the Society for Industrial and Applied Mathematics.

Liu, Y. and Q. Tang (2010). The subexponential product convolution of two Weibull-type distributions. *Journal of the Australian Mathematical Society 89*, 277–288.

Maier, R. S. and C. A. O'Cinneide (1992). A closure characterisation of phase-type distributions. *Journal of Applied Probability 29*(1), 92–103.

Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models. An algorithmic approach.* Baltimore: John Hopkins University Press.

Peralta, O., L. Rojas-Nandayapa, W. Xie, and H. Yao (2016). Approximation of ruin probabilities via erlangized scale mixtures. Submitted.

Ragab, F. M. (1965). Multiple integrals involving product of modified Bessel functions of the second kind. *Rendiconti del Circolo Matematico di Palermo 14*(3), 367–381.

Shi, D., J. Guo, and L. Liu (1996). SPH-distributions and the rectangle-iterative algorithm. In S. Chakravarthy and A. S. Alfa (Eds.), *Matrix-Analytic Methods in Stochastic Models*, pp. 207–224. Marcel Dekker.

Su, C. and Y. Chen (2006). On the behavior of the product of independent random variables. *Sci. China Ser. A49*(3), 342–359.

Tang, Q. (2008). From light tails to heavy tails through multiplier. *Extremes 11*(4), 379–391.

Vatamidou, E., I. J. B. F. Adan, M. Vlasiou, and B. Zwart (2012). Corrected phase-type approximations of heavy-tailed risk models using perturbation analysis. *Insurance: Mathematics and Economics 53*, 366–378.

Vatamidou, E., I. J. B. F. Adan, M. Vlasiou, and B. Zwart (2014). On the accuracy of phase-type approximations of heavy-tailed risk models. *Scandinavian Actuarial Journal 2014*(6), 510–534.