SFB 823

Explicit results on conditional distributions of generalized exponential mixtures

Claudia Klüppelberg, Miriam Isabel Seifert

Discussion Paper

Nr. 2/2020

Explicit results on conditional distributions of generalized exponential mixtures

Claudia Klüppelberg [1] and Miriam Isabel Seifert [2]

January 23, 2020

Abstract

For independent exponentially distributed random variables X_i, $i \in \mathcal{N}$ with distinct rates λ_i we consider sums $\sum_{i\in\mathcal{A}} X_i$ for $\mathcal{A} \subseteq \mathcal{N}$ which follow generalized exponential mixture (GEM) distributions. We provide novel explicit results on the conditional distribution of the total sum $\sum_{i\in\mathcal{N}} X_i$ given that a subset sum $\sum_{j\in\mathcal{A}} X_j$ exceeds a certain threshold value $t > 0$, and vice versa. Moreover, we investigate the characteristic tail behavior of these conditional distributions for $t \to \infty$. Finally, we illustrate how our probabilistic results can be applied in practice by providing examples from both reliability theory and risk management.

Keywords: Conditional distribution, Convolution, Exponential distribution, Generalized exponential mixture distribution, Phase-type distributions, Reliability theory, Risk management, Tail behavior

MSC(2010): 60E05, 60K10, 62E10, 90B25

[1] Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany. Email: cklu@tum.de

[2] Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany. Email: miriam.seifert@rub.de

1 Introduction

Stochastic properties of sums of independent exponentially distributed random variables X_i for $i \in \mathcal{N}$, where $\mathcal{N} \subset \mathbb{N}$ is some finite set, are both of high theoretical and practical relevance. Under the assumption that the rate parameters $\lambda_i > 0$ are pairwise distinct, the distribution of the sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ can be represented as a generalized exponential mixture (GEM)[1] with distribution function $F_{S_{\mathcal{N}}}(x) = 1 - \sum_{i\in\mathcal{N}} \pi_i \,\mathrm{e}^{-\lambda_i x}$, $x > 0$, with real-valued mixing proportions π_i which satisfy $\sum_{i\in\mathcal{N}} \pi_i = 1$.

The early theoretical literature on exponential mixtures is mainly focused on necessary or sufficient conditions for the mixing proportions π_i to ensure that for $x > 0$ the expression $1 - \sum_{i\in\mathcal{N}} \pi_i \,\mathrm{e}^{-\lambda_i x}$ defines a valid distribution function; see e.g. Bartholomew [6], Harris et al. [12], Steutel [23]. More recently, research has been more concentrated on probabilistic or statistical properties of exponential mixture distributions; see e.g. Amari and Misra [2], Favaro and Walker [11], Jewell [14], Kochar and Xu [16], Navarro et al. [22]. Mixtures and convolutions of exponential distributions constitute important subclasses of so-called phase-type distributions which are defined in terms of an underlying Markov jump process. Phase-type distributions attract much attention in the current literature (see e.g. Albrecher and Bladt [1]); an excellent review of recent results on phase-type distributions can be found in the book of Bladt and Nielsen [9].

Convolutions of exponential distributions have been proved to be relevant in various application fields: in management science Bekker and Koeleman [7] provide results on admission scheduling in a clinic with respect to stable bed demand where patient stay lengths follow GEM distributions; in reliability theory Kordecki [17] provides bounds for the probability that a system of independent components will completely operate when the component failure probabilities are exponentially distributed with pairwise distinct rates, while Yin et al. [25] derive results on finding the optimal rate of preventive maintenance in Markov systems with GEM time-to-failure distribution; further applications are elaborated by Asmussen [4] in renewal theory, by Bergel and Egídio dos Reis [8] and Willmot and Woo [24] in actuarial science, and by Anjum and Perros [3] in network science; Dufresne [10] shows how to apply GEMs for approximating arbitrary distribution functions with positive half-line support. Klüppelberg and Seifert [15] investigate financial risk measures for a system of asymptotically exponentially distributed losses, however, they show that summation

[1] This distribution class is known in the literature under various names, e.g. *generalized Erlang* (cf. Bergel and Egídio dos Reis [8]), *hypoexponential* as its coefficient of variation is smaller than one as by the exponential distribution (cf. Li and Li [18]), as well as *generalized hyperexponential* distribution (cf. Harris et al. [12]).

of such losses does not lead to GEM distributions.

Although the above-mentioned literature on unconditional GEM distributions is rather substantial, to the best of our knowledge results on conditional distributions for sums of independent exponentially distributed random variables are not yet available. In this paper we close this gap and deduce explicit results on conditional distributions for such sums. The assumption of pairwise distinct rate parameters makes the analysis more challenging compared to those for the sum of independent identically exponentially distributed random variables which follows an Erlang distribution. Nevertheless, we also handle the case where we relax the restriction that the rate parameters are all pairwise distinct.

In the present paper we consider the total sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ as well as subset sums $S_{\mathcal{A}} = \sum_{i\in\mathcal{A}} X_i$ based on subsets $\mathcal{A} \subseteq \mathcal{N}$ of independent exponentially distributed X_i, and deduce the conditional distributions for both $P(S_{\mathcal{N}} > s \mid S_{\mathcal{A}} > t)$ and $P(S_{\mathcal{A}} > t \mid S_{\mathcal{N}} > s)$ for $s, t > 0$. Hence, we quantify interdependence between the total sum $S_{\mathcal{N}}$ and subset sums $S_{\mathcal{A}}$ of independent exponential random variables. Besides providing results for finite thresholds $s, t > 0$, we also present statements quantifying the tail behavior when conditioning on extreme events $\{S_{\mathcal{N}} > s\}$ for $s \to \infty$ and $\{S_{\mathcal{A}} > t\}$ for $t \to \infty$. Our novel probabilistic results are essential when only partial information on a subset is available, but the quantity of interest is the total sum $S_{\mathcal{N}}$, or, vice versa, when there is information about the total sum $S_{\mathcal{N}}$ but the distribution of some subset sum $S_{\mathcal{A}}$ is of interest.

This paper is organized as follows. In section 2 we give expressions for the distribution of sums $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ of independent exponentially distributed random variables X_i, and deduce the characteristic tail behavior of their survival functions. In section 3 we investigate the conditional distributions for the sum $S_{\mathcal{N}}$ given that some X_j, $j \in \mathcal{N}$ exceeds a certain threshold value, as well as, conversely, those of X_j given that $S_{\mathcal{N}}$ exceeds some threshold. Next, in section 4 we generalize these results by providing the conditional distributions for the total sum $S_{\mathcal{N}}$ and subset sums $S_{\mathcal{A}} = \sum_{j\in\mathcal{A}} X_j$ for some $\mathcal{A} \subseteq \mathcal{N}$. In section 5 we illustrate our theoretical results by presenting relevant examples where our results may provide support for decision making in practice. The proofs are placed in section 6.

Notations and conventions: Two functions f and g are said to be (i) *asymptotically equivalent* $f \sim g$ if $f(x)/g(x) \to 1$ for $x \to \infty$ and (ii) *proportional* $f \propto g$ if $f(x)/g(x) = c$ for all x and some constant $c > 0$. For some $|\mathcal{N}| \in \{2, 3, 4, \ldots\}$ we use the notation $\mathcal{N} := \{1, 2, \ldots, |\mathcal{N}|\}$ and denote by $|\mathcal{A}|$ the cardinality of a set $\mathcal{A} \subseteq \mathcal{N}$. Further, we write f_{X_i} for the density of the random variable X_i. We denote by $\mathrm{Exp}(\lambda_i)$ the class of exponentially distributed random variables with rate parameter $\lambda_i > 0$, such that $X_i \in \mathrm{Exp}(\lambda_i)$ has the density $f_{X_i}(x) = \lambda_i \, \mathrm{e}^{-\lambda_i x}$ for

$x > 0$, and the expectation $E[X_i] = 1/\lambda_i$. Given the rate parameters $\lambda_1, \ldots, \lambda_{|\mathcal{N}|}$ with $\lambda_i \neq \lambda_j$ for $i \neq j$, the minimal rates for the sets $\mathcal{N}$ and $\mathcal{A} \subseteq \mathcal{N}$ are denoted as:

$$\lambda_{\mathfrak{n}} = \min\{\lambda_i \mid i \in \mathcal{N}\}\,, \quad \lambda_{\mathfrak{a}} = \min\{\lambda_j \mid j \in \mathcal{A}\}\,, \tag{1}$$

where $\mathfrak{n}$ (or $\mathfrak{a}$) denotes the index of the random variable $X_{\mathfrak{n}}$ (or $X_{\mathfrak{a}}$) with minimal rate parameter in set $\mathcal{N}$ (or $\mathcal{A}$). As the usual convention, we set $\prod_{i\in\mathcal{A}} c_i := 1$ for arbitrary c_i when $\mathcal{A}$ is the empty set.

2 Generalized exponential mixtures

Throughout this paper we consider distributions of sums $S_{\mathcal{N}} := \sum_{i\in\mathcal{N}} X_i$ and $S_{\mathcal{A}} := \sum_{i\in\mathcal{A}} X_i$ with some $\mathcal{A} \subseteq \mathcal{N}$ for random variables X_i which satisfy the following
Assumption (A): The random variables $X_i \in \mathrm{Exp}(\lambda_i)$, $i \in \mathcal{N}$, are stochastically independent with pairwise distinct rate parameters $\lambda_i \neq \lambda_j$ for all $i \neq j$.

Note that the setting with $\lambda_i = \lambda > 0$ for all $i \in \mathcal{N}$ would result in an Erlang distribution for the sum of independent random variables. Assumption (A) with pairwise distinct λ_i makes our analysis more challenging; it is posed in the current literature by e.g. Bergel and Egídio dos Reis [8], Kordecki [17], McLachlan [20] and Steutel [23]. In Remark 3(ii) at the end of this section we indicate how to handle the case where the restriction for pairwise distinct parameters is relaxed; i.e., when some (or even all) rate parameters coincide.

Remark 1. Note that all results on sums $S_{\mathcal{N}}$ would be also valid for linear combinations $\sum_{i\in\mathcal{N}} \theta_i Y_i$ of independent $Y_i \in \mathrm{Exp}(\tilde{\lambda}_i)$ and coefficients $\theta_i > 0$, $i \in \mathcal{N}$, when $\tilde{\lambda}_i \theta_j \neq \tilde{\lambda}_j \theta_i$ for all $i \neq j$. The statements on linear combinations can be obtained by the linear transformation $X_i := \theta_i Y_i \in \mathrm{Exp}(\lambda_i)$ with $\lambda_i := \tilde{\lambda}_i/\theta_i$. $\diamondsuit$

Before we present our findings, we summarize the established results on the convolution of exponentially distributed random variables.

Proposition 1 (Jasiulewicz and Kordecki [13], Theorem 1)**.** *Under Assumption (A) the sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ has density:*

$$f_{S_{\mathcal{N}}}(x) = \sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \lambda_i\, \mathrm{e}^{-\lambda_i x}\,, \quad x > 0\,, \tag{2}$$

with mixing proportions

$$\pi_{i(\mathcal{N})} := \prod_{j\in\mathcal{N}\setminus\{i\}} \frac{\lambda_j}{\lambda_j - \lambda_i} \;\in (-\infty, \infty)\,, \quad i \in \mathcal{N}\,. \tag{3}$$

The sum $S_{\mathcal{N}}$ with density (2) follows a *generalized exponential mixture* (GEM) distribution as it allows – in contrast to a classical mixture – for mixing proportions of both positive and negative signs. As the mixing proportions depend on the underlying set $\mathcal{N}$ we denote them by $\pi_{i(\mathcal{N})}$. Note that a change of the underlying set could lead to a change of all mixing proportions; i.e., in general we have $\pi_{i(\mathcal{N})} \neq \pi_{i(\mathcal{A})}$ for all $i \in \mathcal{A} \subset \mathcal{N}$.

Remark 2. Due to the density representation in (2) it follows that the mixing proportions sum up to one:

$$\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} = 1 \,, \tag{4}$$

where exactly $\lfloor |\mathcal{N}|/2 \rfloor$ of the mixing proportions $\pi_{i(\mathcal{N})}$ are negative. More precisely, the mixing proportions alternate in sign when the rate parameters are ordered (which can be done without loss of generality) as $\lambda_1 < \lambda_2 < \ldots < \lambda_{|\mathcal{N}|}$, then $\pi_{i(\mathcal{N})}$ is positive for odd and negative for even indices $i \in \mathcal{N}$, as there are exactly $(i-1)$ negative denominators in (3). $\diamondsuit$

After providing results for finite $x > 0$, we establish the characteristic tail behavior of GEM distributions for $x \to \infty$. In the next corollary we show that the random variable with the smallest parameter $\lambda_{\mathfrak{n}}$, cf. (1), determines the asymptotics:

Corollary 1. *Under Assumption (A), the survival function of $S_{\mathcal{N}} = \sum_{i \in \mathcal{N}} X_i$ satisfies:*

$$\begin{aligned} P(S_{\mathcal{N}} > x) \quad &= \quad \sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} \, \mathrm{e}^{-\lambda_i x} \,, \quad x > 0 \,, \\ &\sim \quad \pi_{\mathfrak{n}(\mathcal{N})} \, \mathrm{e}^{-\lambda_{\mathfrak{n}} x} \, \propto \, P(X_{\mathfrak{n}} > x) \;\; \textit{for} \;\; x \to \infty \,. \end{aligned}$$

The statements of Proposition 1 and Corollary 1 are illustrated in Figure 1 where we show, first, the different shapes of the GEM and exponential survival functions for non-asymptotic regions and, second, how good the exponential distribution with the smallest tail parameter $\lambda_{\mathfrak{n}}$ is for the asymptotic approximation of GEM.

Similarly to the established result that phase-type distributions have asymptotic tails of Erlang distributions (cf. Asmussen et al. [5, sect. 5.7]), our result in Corollary 1 states that the subclass of GEM distributions leads to asymptotic tails of exponential distributions. However, if we allow the tail parameters to coincide for different indices, then the distribution of $S_{\mathcal{N}}$ has the asymptotic tail of an Erlang distribution, as we discuss it in the following remark.

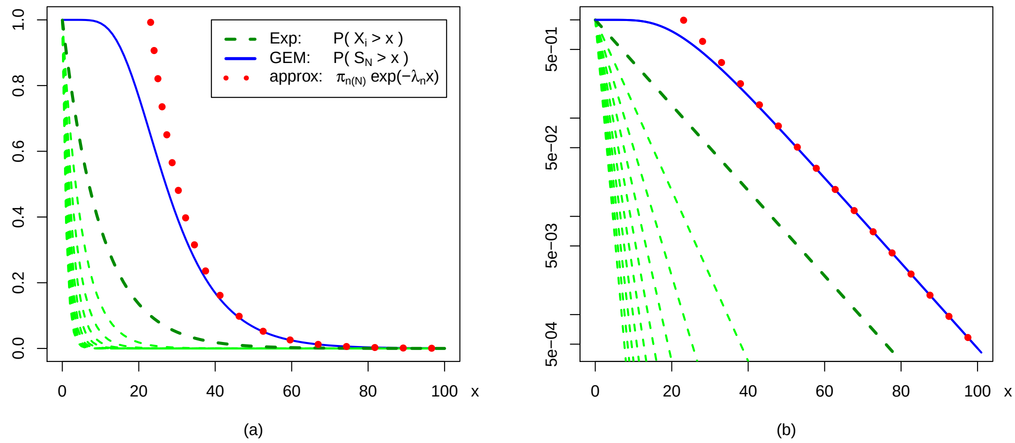

Figure 1: (a): Survival functions of exponentially distributed X_i, $i = 1, \ldots, 10$ with distinct rate parameters $\lambda_1 = 0.1, \lambda_2 = 0.2, \ldots, \lambda_{10} = 1.0$ (dashed lines), and the function $\pi_{\mathfrak{n}(\mathcal{N})}\, \mathrm{e}^{-\lambda_{\mathfrak{n}} x}$ (dotted line) with $\mathfrak{n} = 1$ as approximation of $P(S_{\mathcal{N}} > x)$ (solid line); cf. Corollary 1.
(b): Same functions as in (a) on a logarithmic vertical axis to illustrate that the approximating function (dotted line) has the same slope as the function of dominant $X_{\mathfrak{n}}$ (bold dashed line), both determined by rate parameter $\lambda_{\mathfrak{n}}$.

Remark 3. Here we comment on the case of possibly equal rate parameters λ_i for some (or all) random variables X_i, $i \in \mathcal{N}$, where the following results hold:

(i) If the parameters λ_i coincide for different values of i, then the distribution of $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ is not GEM any longer. If – as a special case – it holds that $\lambda_i = \lambda$ for all $i \in \mathcal{N}$, then $S_{\mathcal{N}}$ is Erlang distributed. In general, if only some λ_i coincide, $S_{\mathcal{N}}$ follows a *generalized Erlang mixture* distribution, cf. Mathai [19] and Moschopoulos [21].

(ii) The corresponding results of Proposition 1 if $\lambda_j = \lambda_i$ for some $j \neq i$ can be obtained as limits for $\lambda_j \to \lambda_i$. Consider e.g. the case of two variables X_1 and X_2 where taking such limit leads to an Erlang distribution as follows:

$$\begin{aligned}\lim_{\lambda_2\to\lambda_1} P\big(X_1 + X_2 > x\big) \quad &= \quad \lim_{\lambda_2\to\lambda_1} \frac{\lambda_2\,\mathrm{e}^{-\lambda_1 x} - \lambda_1\,\mathrm{e}^{-\lambda_2 x}}{\lambda_2 - \lambda_1} \\ = \lim_{\lambda_2\to\lambda_1} \big(\mathrm{e}^{-\lambda_1 x} + \lambda_1 x\,\mathrm{e}^{-\lambda_2 x}\big) \;&=\; (1 + \lambda_1 x)\,\mathrm{e}^{-\lambda_1 x}\,, \quad x > 0\,. \end{aligned} \tag{5}$$

The asymptotic results in Corollary 1 for $x \to \infty$ give asymptotic tails of Erlang distribution with shape parameter $q = |Q|$, where the set $Q := \{i \in \mathcal{N} \mid \lambda_i = \min_{k\in\mathcal{N}} \lambda_k\}$ contains the indices of all variables with the smallest rate parameter. Hence, the value q is the number of asymptotically dominant random variables X_i with the smallest rate parameter. $\diamondsuit$

3 Conditional distributions for GEM sums and a selected exponential random variable

Now we provide our novel results on the conditional distribution for the sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ given that some element X_j exceeds a certain threshold value, as well as, conversely, on the conditional distribution for X_j given that $S_{\mathcal{N}}$ exceeds some threshold. In the next theorem we deduce the conditional probabilities which illustrate both finite and asymptotic influence of X_j on the sum $S_{\mathcal{N}}$.

Theorem 1. *Under Assumption (A), the conditional probabilities for $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ satisfy given that $X_j > x > 0$ for some $j \in \mathcal{N}$:*

(i) for finite $s > x$:

$$P(S_{\mathcal{N}} > s \mid X_j > x) \;=\; P(S_{\mathcal{N}} > s - x) \;=\; \sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i(s-x)}\,;$$

(ii) asymptotically for $x \to \infty$ and some positive function $s(x)$ with $s(x)-x \to \infty$:

$$P(S_{\mathcal{N}} > s(x) \mid X_j > x) \;\sim\; \pi_{\mathfrak{n}(\mathcal{N})}\, \mathrm{e}^{-\lambda_{\mathfrak{n}}(s(x)-x)} \;\propto\; P(X_{\mathfrak{n}} > s(x) - x)\,,$$

with $\mathfrak{n}$ as in (1).

Throughout this paper, we exclude trivial cases where the corresponding conditional probability is equal to 1. For example, in Theorem 1 we only consider $s > x$. Note also that the asymptotically dominant random variable $X_{\mathfrak{n}}$ has the largest expectation of all X_i for $i \in \mathcal{N}$.

Note that in Theorem 1(ii) one can use any function $s(x)$ which increases faster than the identity function $s(x) = x$. E.g., for the important case of a linear function we obtain asymptotically for $x \to \infty$ and some $\alpha > 1$ that:

$$P(S_{\mathcal{N}} > \alpha x \mid X_j > x) \;\sim\; \pi_{\mathfrak{n}(\mathcal{N})}\, \mathrm{e}^{-(\alpha-1)\lambda_{\mathfrak{n}} x} \;\propto\; P(X_{\mathfrak{n}} > (\alpha - 1)x)\,.$$

Remark 4. The results of Theorem 1(i) reveal the following interesting features: The conditional distribution of the sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ given some X_j, $j \in \mathcal{N}$

(a) is equal to the shifted *unconditional* distribution of $S_{\mathcal{N}}$;

(b) is independent of the specific index j of the variable we condition on;

(c) depends *only* on the difference $s - x$ of the threshold values.

In particular, points (a) and (c) display a certain "no-memory" property of the GEM distributions by conditioning on a single exponential random variable. $\diamondsuit$

Next we present the complementary result to Theorem 1 with conditioning on the sum $S_{\mathcal{N}} > s$.

Proposition 2. *Under Assumption (A), the conditional probabilities for X_j, $j \in \mathcal{N}$ satisfy given that $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i > s > 0$:*

(i) for finite $s > x > 0$:

$$\begin{aligned} P(X_j > x \mid S_{\mathcal{N}} > s) &= \frac{P(X_j > x)\, P(S_{\mathcal{N}} > s - x)}{P(S_{\mathcal{N}} > s)} \\ &= \mathrm{e}^{-\lambda_j x} \frac{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i (s-x)}}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i s}}, \end{aligned}$$

and for finite $x \geq s$:

$$P(X_j > x \mid S_{\mathcal{N}} > s) = \frac{P(X_j > x)}{P(S_{\mathcal{N}} > s)} = \frac{\mathrm{e}^{-\lambda_j x}}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i s}};$$

(ii) asymptotically for $s \to \infty$ and some positive function $x(s)$ with $x(s) \to \infty$ it holds:

- *if $s - x(s) \to \infty$:*

$$P(X_j > x(s) \mid S_{\mathcal{N}} > s) \sim \mathrm{e}^{-(\lambda_j - \lambda_{\mathfrak{n}}) x(s)} = \frac{P(X_j > x(s))}{P(X_{\mathfrak{n}} > x(s))},$$

- *if $s - x(s) \to -\infty$:*

$$P(X_j > x(s) \mid S_{\mathcal{N}} > s) \sim \frac{\mathrm{e}^{-(\lambda_j x(s) - \lambda_{\mathfrak{n}} s)}}{\pi_{\mathfrak{n}(\mathcal{N})}} \propto \frac{P(X_j > x(s))}{P(X_{\mathfrak{n}} > s)},$$

- *if $s - x(s) \to c \in (-\infty, \infty)$:*

$$P(X_j > x(s) \mid S_{\mathcal{N}} > s) \sim K_c\, \mathrm{e}^{-(\lambda_j - \lambda_{\mathfrak{n}}) s} \propto \frac{P(X_j > s)}{P(X_{\mathfrak{n}} > s)},$$

with $K_c := \mathrm{e}^{\lambda_j c}/\pi_{\mathfrak{n}(\mathcal{N})}$ for $c \leq 0$ and $K_c := \sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-(\lambda_i - \lambda_j)c}/\pi_{\mathfrak{n}(\mathcal{N})}$ for $c > 0$ and with $\mathfrak{n}$ as in (1).

In Proposition 2 we show that $P(X_j > x \mid S_{\mathcal{N}} > s)$ depends (even asymptotically) on the distribution of the particular random variable X_j in contrast to the counterpart $P(S_{\mathcal{N}} > s \mid X_j > x)$ investigated in Theorem 1.

For the special case of linear functional dependence between the lower thresholds for sum $S_{\mathcal{N}}$ and variable X_j we obtain asymptotically for $s \to \infty$ and $0 < \beta < 1$:

$$P(X_j > \beta s \mid S_{\mathcal{N}} > s) \sim \mathrm{e}^{-(\lambda_j - \lambda_{\mathfrak{n}}) \beta s} = \frac{P(X_j > \beta s)}{P(X_{\mathfrak{n}} > \beta s)},$$

and for $\beta \geq 1$:

$$P(X_j > \beta s \mid S_{\mathcal{N}} > s) \sim \frac{\mathrm{e}^{-(\beta\lambda_j - \lambda_{\mathfrak{n}}) s}}{\pi_{\mathfrak{n}(\mathcal{N})}} \propto \frac{P(X_j > \beta s)}{P(X_{\mathfrak{n}} > s)}.$$

Remark 5. Theorem 1 and Proposition 2 point out the following characteristic properties of the conditional distributions under consideration:

- The distribution of $S_{\mathcal{N}}$ contains *all* information to quantify the influence of the random variables X_j on the sum. This holds not only asymptotically for large X_j, but also exactly for all $s, x > 0$ as $P(S_{\mathcal{N}} > s \mid X_j > x) = P(S_{\mathcal{N}} > s - x)$.

- The above-mentioned "no-memory" property allows for a simple quantification of the influence of a single random variable on the aggregated sum, as it is given immediately by the distribution of the sum. Thereby, it is irrelevant which particular random variable X_j we condition on.

- By conditioning on $\{S_{\mathcal{N}} > s\}$, Proposition 2 states that

$$P(X_j > x \mid S_{\mathcal{N}} > s) = P(X_j > x) \, \frac{P(S_{\mathcal{N}} > s - x)}{P(S_{\mathcal{N}} > s)}\,, \quad s > x > 0\,. \tag{6}$$

  Hence, this statement involves only marginal distributions of $S_{\mathcal{N}}$ and X_j but not their joint distribution.

- The qualitative difference between the probability $P(S_{\mathcal{N}} > s \mid X_j > x)$ and its counterpart $P(X_j > x \mid S_{\mathcal{N}} > s)$ is based on the following intuition: the event $\{S_{\mathcal{N}} > s\}$ does not specify which random variables X_j cause a threshold exceedance. Such events may comprise very different scenarios for possible realizations of random variables $X_1, \ldots, X_{|\mathcal{N}|}$, for example scenarios with a few large realizations as well as scenarios where none of realizations is large but the sum $S_{\mathcal{N}}$ exceeds a high threshold merely by a cumulation effect. $\diamondsuit$

4 Conditional distributions for GEM sums

We generalize the results from the previous section by providing expressions for conditional distributions of the total sum $S_{\mathcal{N}} = \sum_{i \in \mathcal{N}} X_i$ and the subset sum $S_{\mathcal{A}} = \sum_{j \in \mathcal{A}} X_j$ for some $\mathcal{A} \subseteq \mathcal{N}$.

Theorem 2. *Under Assumption (A), for each subset $\mathcal{A} \subseteq \mathcal{N}$ the conditional probabilities for the total sum $S_{\mathcal{N}} = \sum_{i \in \mathcal{N}} X_i$ satisfy given that the subset sum $S_{\mathcal{A}} = \sum_{j \in \mathcal{A}} X_j > t > 0$:*

(i) for finite $s > t$:

$$\begin{aligned} P\big(S_{\mathcal{N}} > s \mid S_{\mathcal{A}} > t\big) &= \frac{\sum_{j \in \mathcal{A}} \pi_{j(\mathcal{A})} \, P(X_j > t) P(\sum_{k \in \mathcal{A}_j^\star} X_k > s - t)}{\sum_{j \in \mathcal{A}} \pi_{j(\mathcal{A})} \, P(X_j > t)} \\ &= \frac{\sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{A}_j^\star} \pi_{j(\mathcal{A})} \, \pi_{k(\mathcal{A}_j^\star)} \, \mathrm{e}^{-(\lambda_j - \lambda_k)t - \lambda_k s}}{\sum_{j \in \mathcal{A}} \pi_{j(\mathcal{A})} \, \mathrm{e}^{-\lambda_j t}}\,, \end{aligned}$$

with $\mathcal{A}_j^\star := (\mathcal{N} \setminus \mathcal{A}) \cup \{j\}$;

(ii) asymptotically for $t \to \infty$ *and some positive function* $s(t)$ *with* $s(t) - t \to \infty$*:*

$$P(S_\mathcal{N} > s(t) \mid S_\mathcal{A} > t) \sim \pi_{\mathfrak{n}(\mathcal{A}_\mathfrak{a}^\star)}\, \mathrm{e}^{-\lambda_\mathfrak{n}(s(t)-t)} \propto \frac{P(X_\mathfrak{n} > s(t))}{P(X_\mathfrak{n} > t)},$$

with $\mathcal{A}_j^\star := (\mathcal{N} \setminus \mathcal{A}) \cup \{j\}$ *and* $\mathfrak{n}$ *as in* (1).

In the results of Theorem 2 we use mixing proportions $\pi_{j(\mathcal{A})}$ or $\pi_{k(\mathcal{A}_j^\star)}$ based on subsets $\mathcal{A} \subseteq \mathcal{N}$ or $\mathcal{A}_j^\star \subseteq \mathcal{N}$, respectively. They are defined analogously to those in (3) based on $\mathcal{N}$. Moreover, these mixing proportions based on two sets with a single element in common are related to each other as stated in the next remark.

Remark 6. Let $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{N}$ such that $\mathcal{A}_1 \cap \mathcal{A}_2 = \{j\}$ for some $j \in \mathcal{N}$. Then it follows:

$$\pi_{j(\mathcal{A}_1)} \cdot \pi_{j(\mathcal{A}_2)} = \pi_{j(\mathcal{A}_1 \cup \mathcal{A}_2)} . \tag{7}$$

In particular, for the mixing proportions in Theorem 2 it holds that:

$$\pi_{j(\mathcal{A})} \cdot \pi_{j(\mathcal{A}_j^\star)} = \pi_{j(\mathcal{A})} \cdot \pi_{j((\mathcal{N}\setminus\mathcal{A})\cup\{j\})} = \pi_{j(\mathcal{N})} \quad \text{for all } j \in \mathcal{A} . \tag{8}$$

$\diamondsuit$

The complementary result to Theorem 2 is as follows:

Proposition 3. *Under Assumption (A), for each* $\mathcal{A} \subseteq \mathcal{N}$ *the conditional probabilities for the subset sum* $S_\mathcal{A} = \sum_{j\in\mathcal{A}} X_j$ *satisfy given that the total sum* $S_\mathcal{N} = \sum_{i\in\mathcal{N}} X_i > s > 0$*:*

(i) for finite $s > t > 0$*:*

$$\begin{aligned} P\big(S_\mathcal{A} > t \mid S_\mathcal{N} > s\big) &= \frac{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, P\left(X_j > t\right) P\left(\sum_{k\in\mathcal{A}_j^\star} X_k > s - t\right)}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, P(X_i > s)} \\ &= \frac{\sum_{j\in\mathcal{A}} \sum_{k\in\mathcal{A}_j^\star} \pi_{j(\mathcal{A})}\, \pi_{k(\mathcal{A}_j^\star)}\, \mathrm{e}^{-(\lambda_j - \lambda_k)t - \lambda_k s}}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i s}}, \end{aligned}$$

and for finite $t \geq s$*:*

$$P(S_\mathcal{A} > t \mid S_\mathcal{N} > s) = \frac{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, P\left(X_j > t\right)}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, P(X_i > s)} = \frac{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j t}}{\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i s}},$$

with $\mathcal{A}_j^\star := (\mathcal{N} \setminus \mathcal{A}) \cup \{j\}$;

(ii) asymptotically for $s \to \infty$ and some positive function $t(s)$ with $t(s) \to \infty$ it holds:

- *if $s - t(s) \to \infty$:*

$$P\big(S_{\mathcal{A}} > t(s) \mid S_{\mathcal{N}} > s\big) \sim \prod_{k \in \mathcal{A} \setminus \{\mathfrak{a}\}} \frac{\lambda_k - \lambda_{\mathfrak{n}}}{\lambda_k - \lambda_{\mathfrak{a}}} \,\mathrm{e}^{-(\lambda_{\mathfrak{a}} - \lambda_{\mathfrak{n}}) t(s)} \,\propto\, \frac{P(X_{\mathfrak{a}} > t(s))}{P(X_{\mathfrak{n}} > t(s))},$$

which reduces in the special case $\mathfrak{a} = \mathfrak{n}$ to $P\big(S_{\mathcal{A}} > t(s) \mid S_{\mathcal{N}} > s\big) \to 1$;

- *if $s - t(s) \to -\infty$:*

$$P\big(S_{\mathcal{A}} > t(s) \mid S_{\mathcal{N}} > s\big) \sim \frac{\pi_{\mathfrak{a}(\mathcal{A})}}{\pi_{\mathfrak{n}(\mathcal{N})}} \,\mathrm{e}^{-(\lambda_{\mathfrak{a}} t(s) - \lambda_{\mathfrak{n}} s)} \,\propto\, \frac{P(X_{\mathfrak{a}} > t(s))}{P(X_{\mathfrak{n}} > s)},$$

- *if $s - t(s) \to c \in (-\infty, \infty)$:*

$$P\big(S_{\mathcal{A}} > t(s) \mid S_{\mathcal{N}} > s\big) \sim K_c \,\mathrm{e}^{-(\lambda_{\mathfrak{a}} - \lambda_{\mathfrak{n}}) s} \,\propto\, \frac{P(X_{\mathfrak{a}} > s)}{P(X_{\mathfrak{n}} > s)},$$

with $K_c := \pi_{\mathfrak{a}(\mathcal{A})} \,\mathrm{e}^{\lambda_{\mathfrak{a}} c} / \pi_{\mathfrak{n}(\mathcal{N})}$ for $c \le 0$ and $K_c := \pi_{\mathfrak{a}(\mathcal{A})} \sum_{k \in \mathcal{A}^\star_{\mathfrak{a}}} \pi_{k(\mathcal{A}^\star_{\mathfrak{a}})} \,\mathrm{e}^{-(\lambda_k - \lambda_{\mathfrak{a}}) c} / \pi_{\mathfrak{n}(\mathcal{N})}$ for $c > 0$ and with $\mathcal{A}^\star_j := (\mathcal{N} \setminus \mathcal{A}) \cup \{j\}$, $\mathfrak{n}$, $\mathfrak{a}$ as in (1).

For the special case that $\mathcal{A} = \mathcal{N}$, the results in Theorem 2 can be simplified by applying the no-memory property of the exponential distribution as follows:

Corollary 2. *Under Assumption (A), it holds that:*

(i) for finite $s > t > 0$:

$$\begin{aligned} P\big(S_{\mathcal{N}} > s \mid S_{\mathcal{N}} > t\big) &= \frac{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} P(X_i > t) P(X_i > s - t)}{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} P(X_i > t)} \\ &= \frac{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} P(X_i > s)}{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} P(X_i > t)} = \frac{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} \,\mathrm{e}^{-\lambda_i s}}{\sum_{i \in \mathcal{N}} \pi_{i(\mathcal{N})} \,\mathrm{e}^{-\lambda_i t}}; \end{aligned}$$

(ii) asymptotically for $t \to \infty$ and some positive function $s(t)$ with $s(t) - t \to \infty$:

$$P(S_{\mathcal{N}} > s(t) \mid S_{\mathcal{N}} > t) \sim \mathrm{e}^{-\lambda_{\mathfrak{n}}(s(t) - t)} \,\propto\, \frac{P(X_{\mathfrak{n}} > s(t))}{P(X_{\mathfrak{n}} > t)}.$$

Differently to Theorem 1(i) where given that a single component exceeds some threshold the no-memory property for GEM holds for finite thresholds, in Corollary 2(ii) the no-memory feature holds only asymptotically.

Our results on conditional distributions for sums and subset sums of exponential variables might also be of interest for the analysis of phase-type distributions as they are important representatives of this class.

5 Illustrative examples

Now we illustrate the practical relevance of the results presented in the previous sections. In the following we concentrate on the total average $\bar{S}_{\mathcal{N}} := (1/|\mathcal{N}|)\sum_{i\in\mathcal{N}} X_i$ taken over all elements in the system $\mathcal{N}$, and the subset average $\bar{S}_{\mathcal{A}} := (1/|\mathcal{A}|)\sum_{j\in\mathcal{A}} X_j$ taken over some subset $\mathcal{A} \subset \mathcal{N}$. Our setting is determined by the cardinality $|\mathcal{A}|$ of the subset compared to the total number $|\mathcal{N}|$ of system elements, as well as by the rate parameters λ_j for $j \in \mathcal{A}$ and λ_i for $i \in \mathcal{N}$. In particular, the smallest rate parameter $\lambda_{\mathfrak{a}}$ in the subset is of interest, more precisely, whether $\mathfrak{a} = \mathfrak{n}$ or $\mathfrak{a} \neq \mathfrak{n}$, with $\mathfrak{a}$ and $\mathfrak{n}$ defined in (1). First we show how the statements in Theorems 1 and 2 help to gain interesting results concerning conditional distributions of these averages.

Proposition 4. *Under Assumption (A), the conditional probabilities of the total average given the subset average are asymptotically proportional for all subsets* $\mathcal{A} \subseteq \mathcal{N}$*:*

$$P(\bar{S}_{\mathcal{N}} > \alpha t \mid \bar{S}_{\mathcal{A}} > t) \sim C(\mathcal{A})\, \mathrm{e}^{-(\alpha|\mathcal{N}|/|\mathcal{A}|-1)\lambda_{\mathfrak{n}} t} \quad \text{for } \alpha > |\mathcal{A}|/|\mathcal{N}|\,, \ t \to \infty\,,$$

with positive constants $C(\mathcal{A}) = \prod_{k\in\widetilde{\mathcal{A}}} \lambda_k/(\lambda_k - \lambda_{\mathfrak{n}})$*, where* $\widetilde{\mathcal{A}} := ((\mathcal{N}\setminus\mathcal{A})\cup\{\mathfrak{a}\})\setminus\{\mathfrak{n}\}$*. Moreover, for* $\mathcal{A}_1 \subset \mathcal{A}_2$ *we have* $C(\mathcal{A}_1) > C(\mathcal{A}_2)$*. Hence, the constants* $C(\mathcal{A})$ *decrease strictly monotone from the value* $C(\{j\}) = \pi_{\mathfrak{n}(\mathcal{N})} > 1$ *for one-element subsets down to* $C(\mathcal{N}) = 1$ *for the total system average.*

The asymptotic result in Proposition 4 applies, for instance, in a situation where only partial information is available. Let a manufacturing company exploit $|\mathcal{N}|$ different machines with (yearly) preventive maintenance times X_i for $i = 1, \ldots, |\mathcal{N}|$, so that these maintenance times can be modeled as independent $\mathrm{Exp}(\lambda_i)$-random variables with $\lambda_i \neq \lambda_j$ for $i \neq j$. Assume that in some subsidiary with $|\mathcal{A}|$ machines the subset average maintenance time $(1/|\mathcal{A}|)\sum_{j\in\mathcal{A}} X_j$ exceeds a high threshold t in the current year. The statement in Proposition 4 allows us to quantify the conditional probability whether the total average maintenance time $(1/|\mathcal{N}|)\sum_{i\in\mathcal{N}} X_i$ for the whole company exceeds the value αt. Such statements are important for optimizing the maintenance schedule with respect to the most efficient allocation of the company's resources.

The asymptotic statements in the next theorem allow us to compare probabilities that either the total system average or a subset average exceed a high threshold βs, given that the total system average $\bar{S}_{\mathcal{N}}$ exceeds threshold s. This is the complementary result to Proposition 4. In particular, we contrast the conditional probabilities for averages based either on a concentrated (C) subset $\mathcal{A} \subset \mathcal{N}$ or on the diversified (D) total set $\mathcal{N}$, and establish conditions when one of these probabilities is of smaller asymptotic order than the other one.

Theorem 3. *Under Assumption (A), the conditional probabilities for concentrated (C) and diversified (D) subset averages given that the total average $\bar{S}_{\mathcal{N}} > s$, namely*

$$P_C(s) := P(\bar{S}_{\mathcal{A}} > \beta s \mid \bar{S}_{\mathcal{N}} > s) \quad \text{and} \quad P_D(s) := P(\bar{S}_{\mathcal{N}} > \beta s \mid \bar{S}_{\mathcal{N}} > s)$$

for some $\mathcal{A} \subset \mathcal{N}$ and $\mathfrak{n}$, $\mathfrak{a}$ as in (1) *fulfill for $s \to \infty$ the following statements:*

$$\begin{aligned}
P_C(s) &= o(P_D(s)) &\Leftrightarrow\quad \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} &> R_\beta, \\
P_C(s) &\sim k_\beta P_D(s) &\Leftrightarrow\quad \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} &= R_\beta, \\
P_D(s) &= o(P_C(s)) &\Leftrightarrow\quad \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} &< R_\beta,
\end{aligned}$$

with

$$R_\beta = \begin{cases} 1, & \text{if } 0 < \beta \leq 1, \\ 1 + (\beta - 1)|\mathcal{N}|/(\beta|\mathcal{A}|), & \text{if } 1 < \beta < |\mathcal{N}|/|\mathcal{A}|, \\ |\mathcal{N}|/|\mathcal{A}| & \text{if } \beta \geq |\mathcal{N}|/|\mathcal{A}| \end{cases}$$

and with constant $k_\beta := 1$ for $0 < \beta \leq 1$, $k_\beta := \prod_{k \in \mathcal{A} \setminus \{\mathfrak{a}\}} (\lambda_k - \lambda_{\mathfrak{a}})/(\lambda_k - \lambda_{\mathfrak{n}})$ for $1 < \beta < |\mathcal{N}|/|\mathcal{A}|$ and $k_\beta := \pi_{\mathfrak{n}(\mathcal{N})}/\pi_{\mathfrak{a}(\mathcal{A})}$ for $\beta \geq |\mathcal{N}|/|\mathcal{A}|$.

Theorem 3 quantifies in terms of the ratio $\lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}}$ whether the asymptotic conditional probability for a subset average becomes negligible compared to that for the total average. More precisely, concentration on subset $\mathcal{A}$ leads to a conditional probability of a smaller order compared to taking the average over all random variables if and only if the corresponding conditions in Theorem 3 are satisfied.

E.g., this result proves to be useful in a system with $|\mathcal{N}|$ risky investments, where the investor should decide whether to build a diversified (D) portfolio with a large number of investments included, or to concentrate (C) on a portfolio based on subset $\mathcal{A} \subset \mathcal{N}$ of carefully selected investment opportunities. The relations between the conditional probabilities for the average concentrated (C) and diversified (D) portfolio losses in a financial stress situation with a large system loss are stated in Theorem 3.

The results of Theorem 3 indicate that a construction of a diversified portfolio should be preferred for investments X_i from similar risk classes characterized by numerically similar rate parameters λ_i, $i \in \mathcal{N}$. However, in a system where the investments X_i have strongly heterogeneous rate parameters, concentrating on a few objects identified by the criterion in Theorem 3 is advantageous in view of minimizing the probability of a large portfolio loss given a high system loss. We demonstrate this effect in the following example which is visualized in Figure 2.

Example 1. For independent exponential risk variables X_i, $i \in \mathcal{N} = \{1, \ldots, 15\}$, we compare the conditional survival functions based on the 15 one-element risks and

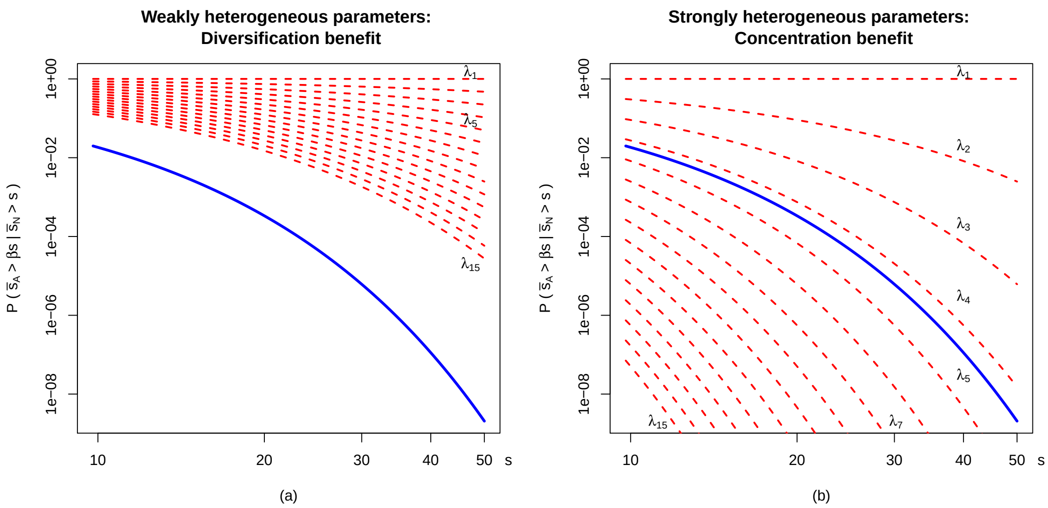

Figure 2: Log-log-plot: comparison of conditional probabilities for 15 totally concentrated subsets based on single elements $\mathcal{A} \in \{\{1\}, \ldots, \{15\}\}$ (dashed lines) and for the system average based on set $\mathcal{N} = \{1, \ldots, 15\}$ (solid line) with $\beta = 9$: plot (a) for scenario with weakly heterogeneous rate parameters λ_i and plot (b) for scenario with strongly heterogeneous rate parameters λ_i; cf. Example 1.

that for the total risk average $\bar{S}_{\mathcal{N}}$ for two different scenarios:

(a) weakly heterogeneous rate parameters: $(\lambda_1, \lambda_2, \ldots, \lambda_{15}) = (0.05, 0.075, \ldots 0.4)$,

(b) strongly heterogeneous rate parameters: $(\lambda_1, \lambda_2, \ldots, \lambda_{15}) = (0.05, 0.25, \ldots 2.85)$.

Both scenarios are comparable as the asymptotically dominant (i.e. the smallest) rate parameter takes the same value $\lambda_{\mathfrak{n}} = 0.05$ in (a) and (b).
In Figure 2 we plot for $\beta = 9$ the survival functions $P(\bar{S}_{\mathcal{A}} > \beta x \mid \bar{S}_{\mathcal{N}} > x)$ for 15 concentrated single object subsets $\bar{S}_{\mathcal{A}} = X_j$ for $\mathcal{A} = \{j\}$, $j = 1, \ldots, 15$ (dashed lines) and for the system average $\bar{S}_{\mathcal{A}} = (1/|\mathcal{N}|) \sum_{i \in \mathcal{N}} X_i$ for $\mathcal{A} = \mathcal{N}$ (solid line). It illustrates the criterion given in Theorem 3: in scenario (a) the ratio is $\lambda_j/\lambda_{\mathfrak{n}} < 1 + (\beta - 1)|\mathcal{N}|/\beta = 43/3$ for all $j \in \mathcal{N}$, which shows that the whole system average is most beneficial here. In contrast, in scenario (b) it holds $\lambda_j/\lambda_{\mathfrak{n}} < 43/3$ only for $j \leq 4$, which implies that concentration on single objects $j \in \{5, 6, \ldots, 15\}$ is more advantageous compared to holding all objects in the system. $\diamondsuit$

Remark 7. The criterion presented in Theorem 3 leads to qualitatively different recommendations with respect to concentration or diversification strategies compared to those for unconditional probabilities. In the latter case, the criterion to minimize the probability of large subset average value $\bar{S}_{\mathcal{A}}$ is as follows: Concentration on subset $\mathcal{A}$ is beneficial in contrast to diversification on the whole system $\mathcal{N}$ in the sense that $P(\bar{S}_{\mathcal{A}} > s) = o(P(\bar{S}_{\mathcal{N}} > s))$ as $s \to \infty$ if and only if the ratio of the smallest rate parameters satisfies $\lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > |\mathcal{N}|/|\mathcal{A}|$. These differences should be taken into account e.g. by comparing results on Value-at-Risk and Conditional Value-at-Risk. $\diamondsuit$

6 Proofs

Proof of Theorem 1 and Proposition 2. Due to the "no-memory" property of the exponential distribution we have that the shifted random variable $X_j - x$ given $X_j > x$ follows an $\mathrm{Exp}(\lambda_j)$ distribution for $j \in \mathcal{N}$, i.e. for the conditional density it holds that $f_{(X_j - x | X_j > x)} = f_{X_j}$. We further partition the sum $S_{\mathcal{N}}$ as follows: $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i = \sum_{i\neq j} X_i + (X_j - x) + x$. Using that the variables X_i for $i \neq j$ are independent from X_j, we obtain for the conditional density of $S_{\mathcal{N}} - x$ given $X_j > x$ that:

$$
\begin{aligned}
f_{(S_{\mathcal{N}}-x|X_j>x)}(z) &= \int_0^z f_{(\sum_{i\neq j} X_i | X_j > x)}(z-u) f_{(X_j - x|X_j>x)}(u)\mathrm{d}u \\
&= \int_0^z f_{\sum_{i\neq j} X_i}(z-u) f_{X_j}(u)\mathrm{d}u \;=\; f_{S_{\mathcal{N}}}(z)\,.
\end{aligned}
$$

This implies that:

$$
P(S_{\mathcal{N}} > s \mid X_j > x) \;=\; P(S_{\mathcal{N}} - x > s - x \mid X_j > x) \;=\; P(S_{\mathcal{N}} > s - x)\,,
$$

which proves statement (i) of Theorem 1.

Asymptotically for $x \to \infty$ and $s(x) - x \to \infty$ it holds that in $P(S_{\mathcal{N}} > s(x) - x) = \sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\, \mathrm{e}^{-\lambda_i (s(x)-x)}$ the summand for $i = \mathfrak{n}$ with the smallest rate parameter $\lambda_{\mathfrak{n}}$ determines the asymptotics (cf. Corollary 1). Hence, we obtain that:

$$
P\Big(\sum_{i\in\mathcal{N}} X_i > s(x) \mid X_j > x\Big) \sim \pi_{\mathfrak{n}(\mathcal{N})}\, \mathrm{e}^{-\lambda_{\mathfrak{n}}(s(x)-x)} = \pi_{\mathfrak{n}(\mathcal{N})}\, P(X_{\mathfrak{n}} > s(x) - x)\,, \tag{9}
$$

which gives statement (ii) of Theorem 1.
The statements in Proposition 2 follow from Bayes' theorem. $\square$

Proof of Theorem 2 and Proposition 3. To analyze the joint probability of the sums $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ and $S_{\mathcal{A}} = \sum_{j\in\mathcal{A}} X_j$ we partition the sum $S_{\mathcal{N}}$ into the subset sum $\sum_{j\in\mathcal{A}} X_j$ and its complement sum $\sum_{k\in\mathcal{N}\setminus\mathcal{A}} X_k$ and use that these sums follow stochastically independent GEM distributions with parameters $\lambda_j, \pi_{j(\mathcal{A})}$, $j \in \mathcal{A}$

or $\lambda_k, \pi_{k(\mathcal{N}\setminus\mathcal{A})}$, $k \in \mathcal{N}\setminus\mathcal{A}$, respectively. For $s > t$ we obtain that:

$$P\big(S_{\mathcal{N}} > s, S_{\mathcal{A}} > t\big) \;=\; P\big(\sum_{i\in\mathcal{N}} X_i > s, \sum_{j\in\mathcal{A}} X_j > t\big)$$

$$= \int_t^s P\big(\sum_{k\in\mathcal{N}\setminus\mathcal{A}} X_k > s-u\big) f_{\sum_{j\in\mathcal{A}} X_j}(u)\mathrm{d}u \;+\; P\big(\sum_{j\in\mathcal{A}} X_j > s\big)$$

$$= \sum_{j\in\mathcal{A}} \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \lambda_j \pi_{j(\mathcal{A})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\, \mathrm{e}^{-\lambda_k s} \int_t^s \mathrm{e}^{(\lambda_k-\lambda_j)u}\mathrm{d}u \;+\; \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j s}$$

$$\begin{aligned}
&= \sum_{j\in\mathcal{A}} \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \frac{\lambda_j \pi_{j(\mathcal{A})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}}{\lambda_k - \lambda_j} \Big[\mathrm{e}^{-\lambda_j s} - \mathrm{e}^{-(\lambda_j t + \lambda_k(s-t))}\Big] \;+\; \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j s} \\
&= \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})} \Big[\big(1 - \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k(\{j\}\cup\{k\})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\big)\, \mathrm{e}^{-\lambda_j s} \\
&\qquad\qquad + \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k(\{j\}\cup\{k\})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\, \mathrm{e}^{-(\lambda_j t + \lambda_k(s-t))}\Big] .
\end{aligned}$$

Next, we use the properties from Eqs. (4) and (7) which together imply that

$$1 - \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k(\{j\}\cup\{k\})} \pi_{k(\mathcal{N}\setminus\mathcal{A})} \;=\; 1 - \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k((\mathcal{N}\setminus\mathcal{A})\cup\{j\})} \;=\; \pi_{j((\mathcal{N}\setminus\mathcal{A})\cup\{j\})} \tag{10}$$

and obtain that:

$$\begin{aligned}
&P\big(\sum_{i\in\mathcal{N}} X_i > s, \sum_{j\in\mathcal{A}} X_j > t\big) \\
&= \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j t} \Big[\pi_{j((\mathcal{N}\setminus\mathcal{A})\cup\{j\})}\, \mathrm{e}^{-\lambda_j(s-t)} + \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k((\mathcal{N}\setminus\mathcal{A})\cup\{j\})}\, \mathrm{e}^{-\lambda_k(s-t)}\Big] \\
&= \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j t} \sum_{k\in(\mathcal{N}\setminus\mathcal{A})\cup\{j\}} \pi_{k((\mathcal{N}\setminus\mathcal{A})\cup\{j\})}\, \mathrm{e}^{-\lambda_k(s-t)} \\
&= \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, P(X_j > t)\, P\big(\sum_{k\in(\mathcal{N}\setminus\mathcal{A})\cup\{j\}} X_k > s-t\big) .
\end{aligned} \tag{11}$$

Consequently, this gives:

$$P\big(\sum_{i\in\mathcal{N}} X_i > s \mid \sum_{j\in\mathcal{A}} X_j > t\big) = \frac{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, P(X_j > t)\, P\big(\sum_{k\in(\mathcal{N}\setminus\mathcal{A})\cup\{j\}} X_k > s-t\big)}{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, P(X_j > t)} . \tag{12}$$

Asymptotically for $t \to \infty$ and $s(t) - t \to \infty$ the summands in (12) with $j = \mathfrak{a}$ and $k = \mathfrak{n}$ dominate which implies that:

$$P\big(\sum_{i\in\mathcal{N}} X_i > s(t) \mid \sum_{j\in\mathcal{A}} X_j > t\big) \;\sim\; \pi_{\mathfrak{n}((\mathcal{N}\setminus\mathcal{A})\cup\{\mathfrak{a}\})}\, \mathrm{e}^{-\lambda_{\mathfrak{n}}(s(t)-t)} . \tag{13}$$

Hence, statements (i) and (ii) in Theorem 2 for $s > t$ are proven. For $s \leq t$ we obtain the trivial case that $P(\sum_{i\in\mathcal{N}} X_i > s, \sum_{j\in\mathcal{A}} X_j > t) = P(\sum_{j\in\mathcal{A}} X_j > t)$ and, consequently, $P(\sum_{i\in\mathcal{N}} X_i > s \mid \sum_{j\in\mathcal{A}} X_j > t) = 1$.

The results in Proposition 3 follow by Bayes' theorem. □

Proof of Proposition 4. The statement (ii) in Theorem 2 gives that

$$P\Big((1/|\mathcal{N}|)\sum_{i\in\mathcal{N}} X_i > \alpha t \;\Big|\; (1/|\mathcal{A}|)\sum_{j\in\mathcal{A}} X_j > t\Big) \;\sim\; C(\mathcal{A})\,\mathrm{e}^{-(\alpha|\mathcal{N}|/|\mathcal{A}|\,-\,1)\lambda_{\mathfrak{n}} t},$$

for $t \to \infty$ with constants

$$C(\mathcal{A}) := \pi_{\mathfrak{n}((\mathcal{N}\setminus\mathcal{A})\cup\{\mathfrak{a}\})} \;=\; \prod_{\substack{k\in(\mathcal{N}\setminus\mathcal{A})\cup\{\mathfrak{a}\}\\ k\neq\mathfrak{n}}} \lambda_k/(\lambda_k - \lambda_{\mathfrak{n}})\,. \tag{14}$$

This form of $C(\mathcal{A})$ implies that by removing some element l from set $\mathcal{A}$, then for $\tilde{\mathcal{A}} := \mathcal{A}\setminus\{l\}$ it holds that $C(\tilde{\mathcal{A}}) = C(\mathcal{A})\cdot\lambda_i/(\lambda_i - \lambda_{\mathfrak{n}}) > C(\mathcal{A})$ with $i := \tilde{\mathfrak{a}} = \operatorname{argmin}\{\lambda_j \mid j \in \tilde{\mathcal{A}}\}$ if $l = \mathfrak{n}$ or $i := l$ if $l \neq \mathfrak{n}$. □

Proof of Theorem 3. Proposition 3(ii) with $t(s) = |\mathcal{A}|\beta s/|\mathcal{N}|$ and Corollary 2(ii) imply that for $P_C(s) := P(\bar{S}_{\mathcal{A}} > \beta s \mid \bar{S}_{\mathcal{N}} > s)$ and $P_D(s) := P(\bar{S}_{\mathcal{N}} > \beta s \mid \bar{S}_{\mathcal{N}} > s)$ it holds asymptotically for $s \to \infty$ that:

$$P_C(s) \;\sim\; \begin{cases} \mathrm{e}^{-(\lambda_{\mathfrak{a}}-\lambda_{\mathfrak{n}})|\mathcal{A}|\beta s/|\mathcal{N}|}\,/\,k_1 & \text{for } 0 < \beta < |\mathcal{N}|/|\mathcal{A}| \\ \mathrm{e}^{-(|\mathcal{A}|\beta\lambda_{\mathfrak{a}}/|\mathcal{N}|\,-\lambda_{\mathfrak{n}})s}\,/\,k_2 & \text{for } \beta \geq |\mathcal{N}|/|\mathcal{A}| \end{cases} \tag{15}$$

$$P_D(s) \;\sim\; \begin{cases} 1 & \text{for } 0 < \beta \leq 1 \\ \mathrm{e}^{-(\beta-1)\lambda_{\mathfrak{n}} s} & \text{for } \beta > 1\,, \end{cases} \tag{16}$$

with constants $k_1 = \prod_{k\in\mathcal{A}\setminus\{\mathfrak{a}\}}(\lambda_k - \lambda_{\mathfrak{a}})/(\lambda_k - \lambda_{\mathfrak{n}})$ and $k_2 = \pi_{\mathfrak{n}(\mathcal{N})}\,/\pi_{\mathfrak{a}(\mathcal{A})}\,.$
Consequently, we obtain for $s \to \infty$:

- for $0 < \beta \leq 1$:

  $P_C(s) = o(P_D(s)) \;\Leftrightarrow\; \mathfrak{a} \neq \mathfrak{n}$, i.e. $\lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > 1$, and $P_C(x) \sim P_D(x) \;\Leftrightarrow\; \mathfrak{a} = \mathfrak{n}$;

- for $1 < \beta < |\mathcal{N}|/|\mathcal{A}|$:

$$\begin{aligned} P_C(s) = o(P_D(s)) \;&\Leftrightarrow\; (\lambda_{\mathfrak{a}} - \lambda_{\mathfrak{n}})|\mathcal{A}|\beta/|\mathcal{N}| > (\beta-1)\lambda_{\mathfrak{n}} \\ &\Leftrightarrow\; \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > 1 + (\beta-1)|\mathcal{N}|/(\beta|\mathcal{A}|)\,; \end{aligned}$$

- for $\beta \geq |\mathcal{N}|/|\mathcal{A}|$:

  $$P_C(s) = o(P_D(s)) \;\Leftrightarrow\; \beta|\mathcal{A}|\lambda_{\mathfrak{a}}/|\mathcal{N}| - \lambda_{\mathfrak{n}} > (\beta-1)\lambda_{\mathfrak{n}} \;\Leftrightarrow\; \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > |\mathcal{N}|/|\mathcal{A}|\,.$$

□

Acknowledgments. This research has been in part financially supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

References

[1] Albrecher, H. and Bladt, M.: Inhomogeneous phase-type distributions and heavy tails. *Journal of Applied Probability* **56**(4), 1044-1064 (2019)

[2] Amari, S.V. and Misra, R.B.: Closed-form expressions for distribution of sum of exponential random variables. *IEEE Transactions on Reliability* **46**, 519-522 (1997)

[3] Anjum, B. and Perros, H.: Adding percentiles of Erlangian distributions. *IEEE Communications Letters* **15**(3), 346-348 (2011)

[4] Asmussen, S.: Matrix-analytic models and their analysis. *Scandinavian Journal of Statistics* **27**, 193-226 (2000)

[5] Asmussen, S., Nerman, O. and Olsson, M.: Fitting Phase-type Distributions via the EM Algorithm. *Scandinavian Journal of Statistics* **23**, 419-441 (1996)

[6] Bartholomew, D. J.: Sufficient conditions for a mixture of exponentials to be a probability density function. *The Annals of Mathematical Statistics* **40**(6), 2183-2188 (1969)

[7] Bekker, R. and Koeleman, P.M.: Scheduling admissions and reducing variability in bed demand. *Health Care Management Science* **14**, 237-249 (2011)

[8] Bergel, A.I. and Egídio dos Reis, A.D.: Ruin problems in the generalized Erlang(n) risk model *European Actuarial Journal* **6**(1), 257-275 (2016)

[9] Bladt, M. and Nielsen, B.F.: *Matrix-Exponential Distributions in Applied Probability.* Springer (2017)

[10] Dufresne, D.: Fitting combinations of exponentials to probability distributions. *Applied Stochastic Models in Business and Industry* **23**, 23-48 (2007)

[11] Favaro, S. and Walker, S.G.: On the distribution of sums of independent exponential random variables via Wilks integral representation. *Acta Applicandae Mathematicae* **109**, 1035-1042 (2010)

[12] Harris, C.M., Marchal, W.G. and Botta, R.F.: A note on generalized hyperexponential distributions. *Communications in Statistics. Stochastic Models* **8**, 179-191 (1992)

[13] Jasiulewicz, H. and Kordecki, W.: Convolutions of Erlang and of Pascal distributions with applications to reliability. *Demonstratio Mathematica* **36**(1), 231-238 (2003)

[14] Jewell, N.P.: Mixtures of exponential distributions. *The Annals of Statistics* **10**, 479-484 (1982)

[15] Klüppelberg, C. and Seifert, M.I.: Financial risk measures for a network of individual agents holding portfolios of light-tailed objects. *Finance and Stochastics* **23**(4), 795-826 (2019)

[16] Kochar, S. and Xu, M.: On the right spread order of convolutions of heterogeneous exponential random variables. *Journal of Multivariate Analysis* **101**(1), 165-176 (2010)

[17] Kordecki, W.: Reliability bounds for multistage structures with independent components. *Statistics and Probability Letters* **34**, 43-51 (1997)

[18] Li, K.-H. and Li, C.T.: Linear combination of independent exponential random variables. *Methodology and Computing in Applied Probability* **21**(1), 253-277 (2019)

[19] Mathai, A.M.: Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics* **34**, 591-597 (1982)

[20] McLachlan, G.J.: Mixtures – models and applications. In: Balakrishnan, N. and Basu, A.P. (Eds.), *The Exponential Distribution: Theory, Methods and Applications.* Gordon and Breach, Amsterdam, 307-325 (1995)

[21] Moschopoulos, P. G.: The distribution of the sum of independent Gamma random variables. *Annals of the Institute of Statistical Mathematics* **37**, 541-544 (1985)

[22] Navarro, J., Balakrishnan, N. and Samaniego, F.J.: Mixture representations of residual lifetimes of used systems. *Journal of Applied Probability* **45**(4), 1097-1112 (2008)

[23] Steutel, F.W.: Note on the infinite divisibility of exponential mixtures. *Annals of the Institute of Statistical Mathematics* **38**, 1303-1305 (1967)

[24] Willmot, G.E. and Woo, J.-K.: On the class of Erlang mixtures with risk theoretic applications. *North American Actuarial Journal* **11**(2), 99-115 (2007)

[25] Yin, M.-L., Angus, J.E. and Trivedi, K.S.: Optimal preventive maintenance rate for best availability with hypo-exponential failure distribution. *IEEE Transactions on Reliability* **62**(2), 351-361 (2013)