J. Appl. Prob. **57,** 760–774 (2020)
*doi:*10.1017/jpr.2020.26

EXPLICIT RESULTS ON CONDITIONAL DISTRIBUTIONS OF GENERALIZED EXPONENTIAL MIXTURES

CLAUDIA KLÜPPELBERG,* *Technical University of Munich*
MIRIAM ISABEL SEIFERT,** *Ruhr University Bochum*

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 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

2010 Mathematics Subject Classification: Primary 62E10
Secondary 60E05; 60K10; 90B25

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 of both high theoretical and high 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) 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 $\sum_{i\in\mathcal{N}} \pi_i = 1$. Note that this distribution class is also known in the literature under other names, such as *generalized Erlang* [9], *hypoexponential*, as its coefficient of variation is smaller than with the exponential distribution [19], or *generalized hyperexponential* [13].

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., [7, 13, 25]. More recently, research has been more concentrated on probabilistic or statistical properties of exponential mixture distributions; see, e.g., [2, 12, 15, 17, 23]. 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

Received 24 May 2019; revision received 31 January 2020; accepted 6 February 2020.
* Postal address: Boltzmannstraße 3, 85748 Garching, Germany. Email: cklu@tum.de
** Postal address: Universitätsstraße 150, 44801 Bochum, Germany. Email: miriam.seifert@rub.de

attention in the current literature (see, e.g., [1]); an excellent review of recent results on phase-type distributions can be found in [10].

Convolutions of exponential distributions have been proved to be relevant in various application fields: in management science, [8] provides results on admission scheduling in a clinic with respect to stable bed demand where patient stay lengths follow GEM distributions; in reliability theory, [18] provides bounds for the probability that a system of independent components will operate completely when the component failure probabilities are exponentially distributed with pairwise distinct rates, while [27] derives results on finding the optimal rate of preventive maintenance in Markov systems with GEM time-to-failure distribution; further applications are elaborated by [4] in renewal theory, [9, 26] in actuarial science, and [3] in network science; [11] shows how to apply GEMs for approximating arbitrary distribution functions with positive half-line support. Recently, [16] investigated financial risk measures for a system of asymptotically exponentially distributed losses; however, they showed that summation 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 $\mathbb{P}(S_{\mathcal{N}}{>}s \mid S_{\mathcal{A}}{>}t)$ and $\mathbb{P}(S_{\mathcal{A}}{>}t \mid S_{\mathcal{N}}{>}s)$ for $s, t > 0$. Hence, we quantify the 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.

1.1. Notation 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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

$\lambda_i > 0$, 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 $\mathbb{E}[X_i] = 1/\lambda_i$. 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

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

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 1. *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 1 with pairwise distinct λ_i makes our analysis more challenging; it is posed in the current literature by, e.g., [9, 18, 20, 25]. 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 also be 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 $\lambda_i := \tilde{\lambda}_i/\theta_i$.

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

Proposition 1. ([14], Theorem 1.) *Under Assumption 1, the sum $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i$ has density*

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

with mixing proportions

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

The sum $S_{\mathcal{N}}$ with density (2) follows a 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 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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

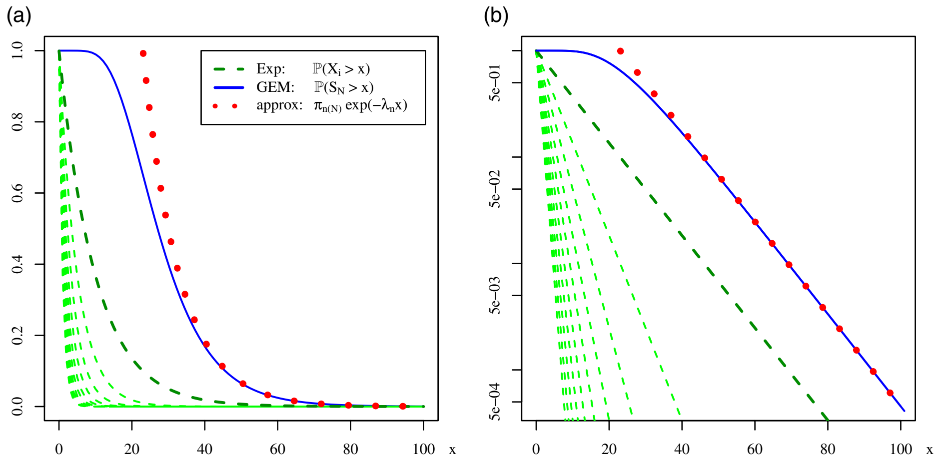

FIGURE 1: (a) Survival functions of exponentially distributed X_i, $i=1, \ldots, 10$, with distinct rate parameters $\lambda_1=0.1$, $\lambda_2=0.2, \ldots, \lambda_{10}=1.0$ (dashed lines), and the function $\pi_{\mathfrak{n}(\mathcal{N})}\,\mathrm{e}^{-\lambda_{\mathfrak{n}} x}$ (dotted line) with $\mathfrak{n}=1$ as an approximation of $\mathbb{P}(S_{\mathcal{N}}>x)$ (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 the rate parameter $\lambda_{\mathfrak{n}}$.

loss of generality) as $\lambda_1<\lambda_2<\cdots<\lambda_{|\mathcal{N}|}$, 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).

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 1, the survival function of* $S_{\mathcal{N}}=\sum_{i\in\mathcal{N}} X_i$ *satisfies*

$$\mathbb{P}(S_{\mathcal{N}}>x)=\sum_{i\in\mathcal{N}} \pi_{i(\mathcal{N})}\ \mathrm{e}^{-\lambda_i x}, \qquad x>0,$$
$$\sim\pi_{\mathfrak{n}(\mathcal{N})}\ \mathrm{e}^{-\lambda_{\mathfrak{n}} x} \propto \mathbb{P}(X_{\mathfrak{n}}>x) \text{ for } x\to\infty.$$

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. [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 discussed in the following remark.

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 no longer GEM. If, as a special case, we have $\lambda_i=\lambda$ for all $i\in\mathcal{N}$, then

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

$S_{\mathcal{N}}$ is Erlang distributed. In general, if only some λ_i coincide, $S_{\mathcal{N}}$ follows a *generalized Erlang mixture* distribution, cf. [21, 22].

(ii) The results corresponding to 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 a limit leads to an Erlang distribution as follows:

$$\lim_{\lambda_2 \to \lambda_1} \mathbb{P}\big(X_1 + X_2 > x\big) = \lim_{\lambda_2 \to \lambda_1} \frac{\lambda_2\, \mathrm{e}^{-\lambda_1 x} - \lambda_1\, \mathrm{e}^{-\lambda_2 x}}{\lambda_2 - \lambda_1}$$
$$= \lim_{\lambda_2 \to \lambda_1} \big(\mathrm{e}^{-\lambda_1 x} + \lambda_1 x\, \mathrm{e}^{-\lambda_2 x}\big) = (1 + \lambda_1 x)\, \mathrm{e}^{-\lambda_1 x}, \qquad x > 0.$$

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.

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 the finite and asymptotic influence of X_j on the sum $S_{\mathcal{N}}$.

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

(i) for finite $s > x$:

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

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

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

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 $s(x) = x$. For example, for the important case of a linear function we obtain asymptotically, for $x \to \infty$ and some $\alpha > 1$,

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

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 GEM distributions by conditioning on a single exponential random variable.

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

Proposition 2. *Under Assumption 1, given that* $S_{\mathcal{N}} = \sum_{i\in\mathcal{N}} X_i > s > 0$*, the conditional probabilities for* X_j*,* $j \in \mathcal{N}$*, satisfy*

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

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

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

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

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

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

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

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

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

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

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

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 $\mathbb{P}(X_j > x \mid S_{\mathcal{N}} > s)$ depends (even asymptotically) on the distribution of the particular random variable X_j, in contrast to the counterpart $\mathbb{P}(S_{\mathcal{N}} > s \mid X_j > x)$ 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 $0 < \beta < 1$,

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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

and, for $\beta \geq 1$,

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

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

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

  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 $\mathbb{P}(S_{\mathcal{N}} > s \mid X_j > x)$ and its counterpart $\mathbb{P}(X_j > x \mid S_{\mathcal{N}} > s)$ 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 the realizations is large but the sum $S_{\mathcal{N}}$ exceeds a high threshold merely by a cumulation effect.

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 1, given that the subset sum $S_{\mathcal{A}} = \sum_{j \in \mathcal{A}} X_j > t > 0$, 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:*

(i) for finite $s > t$:

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

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

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

with $\mathcal{A}_j^{\star} := (\mathcal{N} \setminus \mathcal{A}) \cup \{j\}$ *and* $\mathfrak{n}$, $\mathfrak{a}$ *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 $\mathcal{A}_1$, $\mathcal{A}_2 \subseteq \mathcal{N}$ be such that $\mathcal{A}_1 \cap \mathcal{A}_2 = \{j\}$ for some $j \in \mathcal{N}$. Then

$$\pi_{j(\mathcal{A}_1)} \cdot \pi_{j(\mathcal{A}_2)} = \pi_{j(\mathcal{A}_1 \cup \mathcal{A}_2)}. \tag{5}$$

In particular, for the mixing proportions in Theorem 2 we have

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

The complementary result to Theorem 2 is given next.

Proposition 3. *Under Assumption 1, given that the total sum* $S_{\mathcal{N}} = \sum_{i \in \mathcal{N}} X_i > s > 0$, *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:*

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

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

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

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

with $K_c := \pi_{\mathfrak{a}(\mathcal{A})} \mathrm{e}^{\lambda_{\mathfrak{a}} c} / \pi_{\mathfrak{n}(\mathcal{N})}$ *for* $c \leq 0$ *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:

Corollary 2. *Under Assumption 1:*

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

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

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

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

Differently to Theorem 1(i), where the no-memory property for GEM holds given that a single component exceeds some finite threshold, in Corollary 2(ii) the no-memory feature holds only asymptotically. Moreover, this corollary proves that the GEM distribution of sum $S_{\mathcal{N}}$ belongs to the class of type-Γ distributions which satisfy the gamma variation property:

$$\lim_{t \to \infty} \frac{\mathbb{P}(S_{\mathcal{N}} > t + z\psi(t))}{\mathbb{P}(S_{\mathcal{N}} > t)} = \mathrm{e}^{-z}$$

for all $0 < z < \infty$ and some positive auxiliary function ψ. More precisely, the result in Corollary 2(i) implies that the auxiliary function can be the constant $\psi \equiv 1/\lambda_{\mathfrak{n}}$; for a detailed analysis of type-Γ distributions and their relevance in the field of extreme value theory see, e.g., [6, 24].

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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

Theorems 1 and 2 help to gain interesting results concerning conditional distributions of these averages.

Proposition 4. *Under Assumption 1, the conditional probabilities of the total average given the subset average are asymptotically proportional for all subsets $\mathcal{A} \subseteq \mathcal{N}$ of the same cardinality:*

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

with positive constants $C(\mathcal{A}) = \prod_{k\in\mathcal{A}^{\star}_{\mathfrak{a},\mathfrak{n}}} \lambda_k/(\lambda_k - \lambda_{\mathfrak{n}})$, where $\mathcal{A}^{\star}_{\mathfrak{a},\mathfrak{n}} := ((\mathcal{N} \setminus \mathcal{A}) \cup \{\mathfrak{a}\}) \setminus \{\mathfrak{n}\}$. 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 monotonically 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 of 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 the probabilities that either the total system average or a subset average exceeds a high threshold βs, given that the total system average $\bar{S}_{\mathcal{N}}$ exceeds a 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 a smaller asymptotic order $o(\,\cdot\,)$ than the other.

Theorem 3. *Under Assumption 1, the conditional probabilities for concentrated (C) and diversified (D) subset averages, given that the total average $\bar{S}_{\mathcal{N}} > s$, namely*

$$\mathbb{P}_{\mathrm{C}}(s) := \mathbb{P}\big(\bar{S}_{\mathcal{A}} > \beta s \big| \bar{S}_{\mathcal{N}} > s\big), \qquad \mathbb{P}_{\mathrm{D}}(s) := \mathbb{P}\big(\bar{S}_{\mathcal{N}} > \beta s \big| \bar{S}_{\mathcal{N}} > s\big)$$

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

$$\begin{aligned} \mathbb{P}_{\mathrm{C}}(s) &= o(\mathbb{P}_{\mathrm{D}}(s)) \Leftrightarrow \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > R_\beta, \\ \mathbb{P}_{\mathrm{C}}(s) &\sim k_\beta \mathbb{P}_{\mathrm{D}}(s) \Leftrightarrow \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} = R_\beta, \\ \mathbb{P}_{\mathrm{D}}(s) &= o(\mathbb{P}_{\mathrm{C}}(s)) \Leftrightarrow \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} < R_\beta, \end{aligned}$$

whereby the boundary R_β is given as

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

and the corresponding constants $k_\beta := 1$ for $0 < \beta \le 1$, $k_\beta := \prod_{k\in\mathcal{A}\setminus\{\mathfrak{a}\}} (\lambda_k - \lambda_{\mathfrak{a}})/(\lambda_k - \lambda_{\mathfrak{n}})$ for $1 < \beta < |\mathcal{N}|/|\mathcal{A}|$ and $k_\beta := \pi_{\mathfrak{n}(\mathcal{N})}/\pi_{\mathfrak{a}(\mathcal{A})}$ for $\beta \ge |\mathcal{N}|/|\mathcal{A}|$.

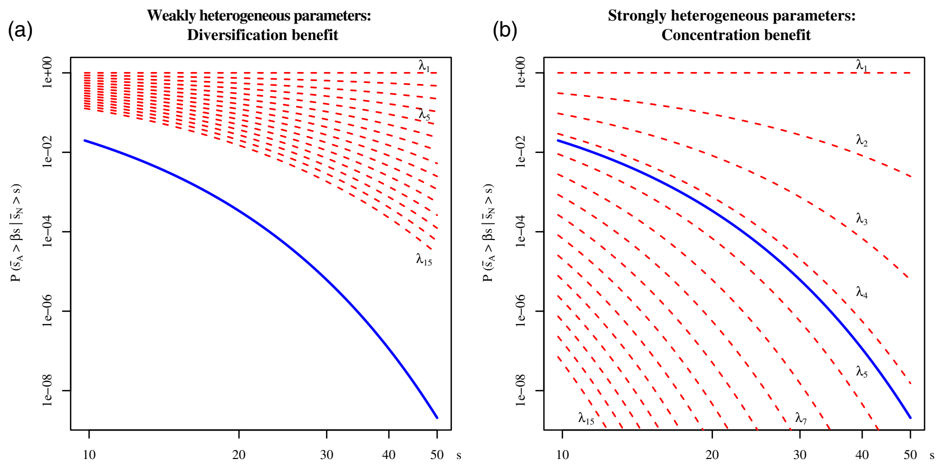

FIGURE 2: Log–log comparison of conditional probabilities for 15 totally concentrated subsets based on single elements $\mathcal{A} \in \{\{1\}, \ldots, \{15\}\}$ (dashed lines) and for the system average based on the set $\mathcal{N} = \{1, \ldots, 15\}$ (solid line) with $\beta = 9$. (a) Scenario with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

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 condition in Theorem 3 is satisfied.

For example, 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 a 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 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 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).

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

In Figure 2 we plot for $\beta = 9$ the survival functions $\mathbb{P}(\bar{S}_{\mathcal{A}} > \beta x \mid \bar{S}_{\mathcal{N}} > x)$ 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). This 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) we have $\lambda_j/\lambda_{\mathfrak{n}} < 43/3$ only for $j \le 4$, which implies that concentration on the single objects $j \in \{5, 6, \ldots, 15\}$ is more advantageous compared to holding all objects in the system.

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 the 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 $\mathbb{P}(\bar{S}_{\mathcal{A}} > s) = o(\mathbb{P}(\bar{S}_{\mathcal{N}} > s))$ 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 by, for example, comparing results on Value at Risk and conditional Value at Risk.

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 we have $f_{(X_j - x | X_j > x)} = f_{X_j}$. 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 $X_j > x$,

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

This implies that

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

which proves statement (i) of Theorem 1.

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

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

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

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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

distributions with parameters λ_j, $\pi_{j(\mathcal{A})}$, $j \in \mathcal{A}$, or λ_k, $\pi_{k(\mathcal{N}\setminus\mathcal{A})}$, $k \in \mathcal{N} \setminus \mathcal{A}$, respectively. For $s > t$ we obtain

$$\begin{aligned}
\mathbb{P}\big(S_{\mathcal{N}} > s, S_{\mathcal{A}} > t\big) &= \mathbb{P}\Bigg(\sum_{i\in\mathcal{N}} X_i > s, \sum_{j\in\mathcal{A}} X_j > t\Bigg) \\
&= \int_t^s \mathbb{P}\Bigg(\sum_{k\in\mathcal{N}\setminus\mathcal{A}} X_k > s-u\Bigg) f_{\sum_{j\in\mathcal{A}} X_j}(u)\,\mathrm{d}u + \mathbb{P}\Bigg(\sum_{j\in\mathcal{A}} X_j > s\Bigg) \\
&= \sum_{j\in\mathcal{A}} \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \lambda_j \pi_{j(\mathcal{A})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\, \mathrm{e}^{-\lambda_k s} \int_t^s \mathrm{e}^{(\lambda_k-\lambda_j)u}\,\mathrm{d}u + \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathrm{e}^{-\lambda_j s} \\
&= \sum_{j\in\mathcal{A}} \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \frac{\lambda_j \pi_{j(\mathcal{A})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}}{\lambda_k-\lambda_j} \Big[\mathrm{e}^{-\lambda_j s} - \mathrm{e}^{-(\lambda_j t + \lambda_k(s-t))}\Big] + \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\; \mathrm{e}^{-\lambda_j s} \\
&= \sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})} \Bigg[\Bigg(1 - \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k(\{j\}\cup\{k\})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\Bigg) \mathrm{e}^{-\lambda_j s} \\
&\quad + \sum_{k\in\mathcal{N}\setminus\mathcal{A}} \pi_{k(\{j\}\cup\{k\})}\, \pi_{k(\mathcal{N}\setminus\mathcal{A})}\, \mathrm{e}^{-(\lambda_j t+\lambda_k(s-t))}\Bigg].
\end{aligned}$$

Next, we use the properties from Eqs. (4) and (5), which together imply that

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

and obtain

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

Consequently,

$$\mathbb{P}\Bigg(\sum_{i\in\mathcal{N}} X_i > s \mid \sum_{j\in\mathcal{A}} X_j > t\Bigg) = \frac{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathbb{P}(X_j > t)\, \mathbb{P}\big(\sum_{k\in(\mathcal{N}\setminus\mathcal{A})\cup\{j\}} X_k > s-t\big)}{\sum_{j\in\mathcal{A}} \pi_{j(\mathcal{A})}\, \mathbb{P}(X_j > t)}. \tag{6}$$

Asymptotically for $t \to \infty$ and $s(t) - t \to \infty$ the summands in (6) with $j = \mathfrak{a}$ and $k = \mathfrak{n}$ dominate, which implies that

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

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

Hence, statements (i) and (ii) in Theorem 2 for $s > t$ are proven. For $s \leq t$ we obtain the trivial case that $\mathbb{P}(\sum_{i\in\mathcal{N}} X_i > s, \sum_{j\in\mathcal{A}} X_j > t) = \mathbb{P}(\sum_{j\in\mathcal{A}} X_j > t)$, and consequently $\mathbb{P}(\sum_{i\in\mathcal{N}} X_i > s \mid \sum_{j\in\mathcal{A}} X_j > t) = 1$. The results in Proposition 3 follow by Bayes' theorem. □

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

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

for $t \to \infty$, with constants

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

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

Proof of Theorem 3. Proposition 3(ii) with $t(s) = |\mathcal{A}|\beta s/|\mathcal{N}|$ and Corollary 2(ii) imply that, for $\mathbb{P}_{\mathrm{C}}(s) := \mathbb{P}(\bar{S}_{\mathcal{A}} > \beta s \mid \bar{S}_{\mathcal{N}} > s)$ and $\mathbb{P}_{\mathrm{D}}(s) := \mathbb{P}(\bar{S}_{\mathcal{N}} > \beta s \mid \bar{S}_{\mathcal{N}} > s)$, we have, asymptotically for $s \to \infty$,

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

$$\mathbb{P}_{\mathrm{D}}(s) \sim \begin{cases} 1 & \text{for } 0 < \beta \leq 1, \\ \mathrm{e}^{-(\beta-1)\lambda_{\mathfrak{n}} s} & \text{for } \beta > 1, \end{cases}$$

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 $0 < \beta \leq 1$:

$$\mathbb{P}_{\mathrm{C}}(s) = o(\mathbb{P}_{\mathrm{D}}(s)) \;\Leftrightarrow\; \mathfrak{a} \neq \mathfrak{n}, \text{ i.e. } \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > 1, \text{ and } \mathbb{P}_{\mathrm{C}}(x) \sim \mathbb{P}_{\mathrm{D}}(x) \;\Leftrightarrow\; \mathfrak{a} = \mathfrak{n};$$

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

$$\begin{aligned} \mathbb{P}_{\mathrm{C}}(s) = o(\mathbb{P}_{\mathrm{D}}(s)) &\Leftrightarrow (\lambda_{\mathfrak{a}}-\lambda_{\mathfrak{n}})|\mathcal{A}|\beta/|\mathcal{N}| > (\beta-1)\lambda_{\mathfrak{n}} \\ &\Leftrightarrow \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > 1 + (\beta-1)|\mathcal{N}|/(\beta|\mathcal{A}|); \end{aligned}$$

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

$$\mathbb{P}_{\mathrm{C}}(s) = o(\mathbb{P}_{\mathrm{D}}(s)) \;\Leftrightarrow\; \beta|\mathcal{A}|\lambda_{\mathfrak{a}}/|\mathcal{N}| - \lambda_{\mathfrak{n}} > (\beta-1)\lambda_{\mathfrak{n}} \;\Leftrightarrow\; \lambda_{\mathfrak{a}}/\lambda_{\mathfrak{n}} > |\mathcal{N}|/|\mathcal{A}|. \qquad \square$$

Acknowledgements

We would like to thank two anonymous reviewers for their constructive comments which led to improvements in the paper. This research has been financially supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

https://doi.org/10.1017/jpr.2020.26 Published online by Cambridge University Press

References

[1] ALBRECHER, H. AND BLADT, M. (2019). Inhomogeneous phase-type distributions and heavy tails. *J. Appl. Prob.* **56**, 1044–1064.

[2] AMARI, S. V. AND MISRA, R. B. (1997). Closed-form expressions for distribution of sum of exponential random variables. *IEEE Trans. Reliab.* **46**, 519–522.

[3] ANJUM, B. AND PERROS, H. (2011). Adding percentiles of Erlangian distributions. *IEEE Commun. Lett.* **15**, 346–348.

[4] ASMUSSEN, S. (2000). Matrix-analytic models and their analysis. *Scand. J. Statist.* **27**, 193–226.

[5] ASMUSSEN, S., NERMAN, O. AND OLSSON, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scand. J. Statist.* **23**, 419–441.

[6] BARBE, P. AND SEIFERT, M. I. (2016). A conditional limit theorem for a bivariate representation of a univariate random variable and conditional extreme values. *Extremes* **19**, 351–370.

[7] BARTHOLOMEW, D. J. (1969). Sufficient conditions for a mixture of exponentials to be a probability density function. *Ann. Math. Statist.* **40**, 2183–2188.

[8] BEKKER, R. AND KOELEMAN, P. M. (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Manage. Sci.* **14**, 237–249.

[9] BERGEL, A. I. AND EGÍDIO DOS REIS, A. D. (2016). Ruin problems in the generalized Erlang(n) risk model. *Europ. Actuarial J.* **6**, 257–275.

[10] BLADT, M. AND NIELSEN, B. F. (2017). *Matrix-Exponential Distributions in Applied Probability*. Springer, New York.

[11] DUFRESNE, D. (2007). Fitting combinations of exponentials to probability distributions. *Appl. Stoch. Models Business Industry* **23**, 23–48.

[12] FAVARO, S. AND WALKER, S. G. (2010). On the distribution of sums of independent exponential random variables via Wilks' integral representation. *Acta Appl. Math.* **109**, 1035–1042.

[13] HARRIS, C. M., MARCHAL, W. G. AND BOTTA, R. F. (1992). A note on generalized hyperexponential distributions. *Commun. Statist. Stoch. Models* **8**, 179–191.

[14] JASIULEWICZ, H. AND KORDECKI, W. (2003). Convolutions of Erlang and of Pascal distributions with applications to reliability. *Demonstr. Math.* **36**, 231–238.

[15] JEWELL, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479–484.

[16] KLÜPPELBERG, C. AND SEIFERT, M. I. (2019). Financial risk measures for a network of individual agents holding portfolios of light-tailed objects. *Finance Stoch.* **23**, 795–826.

[17] KOCHAR, S. AND XU, M. (2010). On the right spread order of convolutions of heterogeneous exponential random variables. *J. Multivar. Anal.* **101**, 165–176.

[18] KORDECKI, W. (1997). Reliability bounds for multistage structures with independent components. *Statist. Prob. Lett.* **34**, 43–51.

[19] LI, K.-H. AND LI, C. T. (2019). Linear combination of independent exponential random variables. *Methodology Comput. Appl. Prob.* **21**, 253–277.

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

[21] MATHAI, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Ann. Inst. Statist. Math.* **34**, 591–597.

[22] MOSCHOPOULOS, P. G. (1985). The distribution of the sum of independent Gamma random variables. *Ann. Inst. Statist. Math.* **37**, 541–544.

[23] NAVARRO, J., BALAKRISHNAN, N. AND SAMANIEGO, F. J. (2008). Mixture representations of residual lifetimes of used systems. *J. Appl. Prob.* **45**, 1097–1112.

[24] SEIFERT, M. I. (2016). Weakening the independence assumption on polar components: Limit theorems for generalized elliptical distributions. *J. Appl. Prob.* **53**, 130–145.

[25] STEUTEL, F. W. (1967). Note on the infinite divisibility of exponential mixtures. *Ann. Inst. Statist. Math.* **38**, 1303–1305.

[26] WILLMOT, G. E. AND WOO, J.-K. (2007). On the class of Erlang mixtures with risk theoretic applications. *N. Amer. Actuarial J.* **11**, 99–115.

[27] YIN, M.-L., ANGUS, J. E. AND TRIVEDI, K. S. (2013). Optimal preventive maintenance rate for best availability with hypo-exponential failure distribution. *IEEE Trans. Reliab.* **62**, 351–361.