Electronic Journal of Statistics
Vol. 8 (2014) 894–926
ISSN: 1935-7524
DOI: 10.1214/14-EJS914

A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions*

Shai Covo

Department of Management, Bar-Ilan University, 52900 Ramat-Gan, Israel
e-mail: green355@netvision.net.il

and

Amir Elalouf

Department of Management, Bar-Ilan University, 52900 Ramat-Gan, Israel
e-mail: amir.elalouf@biu.ac.il

Abstract: It is well known that the sum S of n independent gamma variables—which occurs often, in particular in practical applications—can typically be well approximated by a single gamma variable with the same mean and variance (the distribution of S being quite complicated in general). In this paper, we propose an alternative (and apparently at least as good) single-gamma approximation to S. The methodology used to derive it is based on the observation that the jump density of S bears an evident similarity to that of a generic gamma variable, S being viewed as a sum of n independent gamma processes evaluated at time 1. This observation motivates the idea of a gamma approximation to S in the first place, and, in principle, a variety of such approximations can be made based on it. The same methodology can be applied to obtain gamma approximations to a wide variety of important infinitely divisible distributions on $\mathbb{R}_+$ or at least predict/confirm the appropriateness of the moment-matching method (where the first two moments are matched); this is demonstrated neatly in the cases of negative binomial and generalized Dickman distributions, thus highlighting the paper's contribution to the overall topic.

MSC 2010 subject classifications: Primary 62E17; secondary 60G50, 60G51.
Keywords and phrases: Sum of independent gamma variables, approximation, moment-matching method, gamma process, Lévy measure, infinitely divisible distributions, negative binomial distribution, generalized Dickman distribution.

Received July 2013.

*This is an original survey paper.

1. Introduction

Throughout this paper, $\mathrm{Gamma}(\alpha, \beta)$ denotes the gamma distribution with density

$$f(x) = \frac{x^{\alpha-1}\mathrm{e}^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$$

for $x > 0$. In the first part, we wish to approximate the sum

$$S = X_1 + \cdots + X_n, \tag{1.1}$$

where X_i ($i = 1, \ldots, n$) are independent $\mathrm{Gamma}(\alpha_i, \beta_i)$ random variables (RV's), by a single gamma RV. More precisely, we consider two approximation methods, as indicated in the abstract. For convenience, it is assumed throughout that $n \geq 2$. When all the β_i are equal, S is gamma distributed, and no approximation is required. However, this case is important from a mathematical/theoretical point of view, and is therefore not excluded.

Convolutions of gamma distributions (or sums of independent gamma variables) occur often, in particular in practical applications. See e.g. the brief overview in [15]. Particularly worth noticing in this respect is the fact that a weighted sum of independent chi-square RV's can be written as a sum of independent gamma RV's (indeed, $a\chi^2_\nu \sim \mathrm{Gamma}(\nu/2, 2a)$, $a > 0$, where χ^2_ν denotes a chi-square RV with ν degrees of freedom); see e.g. the brief survey in [9]. However, exact expressions for the probability density function of S are quite complicated in general (see below). To avoid analytical or computational difficulties, it is useful in certain applications to approximate the exact convolution by a single gamma distribution. While the most natural candidate is the gamma distribution with the correct mean and variance, it is not necessarily the best one to use; hence the importance of the topic at hand.

The second part of this paper, written as a substantial complement to the first, is devoted to developing a gamma approximation to *infinitely divisible* (ID) distributions on $\mathbb{R}_+$. It makes a significant theoretical and practical contribution.

The rest of the paper is organized as follows. The exact density function of S is considered in Section 1.1. The approximation of S by a gamma variable with the same mean and variance, $X_{\mathrm{m}} \sim \mathrm{Gamma}(\alpha_{\mathrm{m}}, \beta_{\mathrm{m}})$ in our notation, is considered in Section 1.2. Our proposed approximation is fully developed in Section 2. First, Section 2.1 provides some preliminaries from the theory of Lévy processes, a key point being that S can be viewed as a sum of n independent gamma processes evaluated at time 1. Then, in Section 2.2, the approximation scheme is established. The approximating RV, $X_* \sim \mathrm{Gamma}(\alpha_*, \beta_*)$, has by construction the same mean as S. Based on a practical heuristic, the parameter β_* is chosen to minimize, over $\beta > 0$, the squared distance $\psi(\beta)$ given in (2.8) (where $\mu = \sum_{i=1}^n \alpha_i\beta_i$). In Section 2.3, the main results for the sum-of-gammas case are presented. Theorem 2.1 expresses β_* as the solution of some equation (the solution is readily available numerically); then α_* is given by $\alpha_* = (\sum_{i=1}^n \alpha_i\beta_i)/\beta_*$. Theorem 2.1 further gives lower and upper bounds for α_* and β_*, which are the same as those given for α_{m} and β_{m} in Proposition 1.1.

Next, Proposition 2.1 states that $\beta_* \le \beta_{\rm m}$, with strict inequality unless all the β_i are equal (in which case X_* and $X_{\rm m}$ are identically distributed as S). The proof of Proposition 2.1 relies on Lemma A.1, which states some general moment inequality. An immediate corollary of Proposition 2.1 (namely, Corollary 2.1) is that $\mathrm{Var}(X_*) \le \mathrm{Var}(S)$, with strict inequality unless all the β_i are equal. Section 3 performs a brief numerical study of the approximations $X_{\rm m}$ and X_* to S. The quality of the approximations has been tested numerically by comparing the gamma densities of the approximating RV's with the exact density of S. The results suggest that the approximation X_* to S is, in general, slightly better than $X_{\rm m}$. After some preliminaries in Section 4.1, the proposed methodology is generalized in Section 4.2 with S replaced by an integrable ID RV on $\mathbb{R}_+$. The theoretical basis of the general methodology is justified in Remark 4.1. Various examples demonstrate its applicability to ID distributions other than convolutions of gammas or at least its good agreement with the moment-matching method. Particularly interesting is Example 4.3, which provides new insights into the gamma approximation to the negative binomial. Gamma approximation to the generalized Dickman distribution is considered in detail in Section 4.3, where three gamma approximations are proposed as alternatives to the one with the same mean and variance. (A brief account of this distribution is included as well.) Appendix A is devoted to proofs.

1.1. The exact density function

Various expressions for the exact density of S are available in the literature. Two of them are given below ((1.2)–(1.3)).

Let f_S denote the density of S. A classical expression for f_S is given by

$$f_S(x) = \prod_{i=1}^{n}\left(\frac{\beta_1}{\beta_i}\right)^{\alpha_i}\sum_{k=0}^{\infty}\frac{\delta_k x^{\sum_{i=1}^n \alpha_i + k - 1}\exp(-x/\beta_1)}{\beta_1^{\sum_{i=1}^n \alpha_i + k}\Gamma(\sum_{i=1}^n \alpha_i + k)} \tag{1.2}$$

for $x > 0$, where $\beta_1 = \min_i(\beta_i)$ and the coefficients δ_k satisfy the recurrence relation

$$\delta_{k+1} = \frac{1}{k+1}\sum_{i=1}^{k+1}\left[\sum_{j=1}^{n}\alpha_j\left(1-\frac{\beta_1}{\beta_j}\right)^i\right]\delta_{k+1-i},$$

with initial condition $\delta_0 = 1$. See [11, Eq. (3)]; the result is due to [10]. A simple-looking expression for f_S is given by

$$f_S(x) = \frac{1}{\pi}\int_0^\infty \frac{\cos\left(\sum_{k=1}^n \alpha_k \arctan(\beta_k t) - xt\right)}{\prod_{k=1}^n (1+t^2\beta_k^2)^{\alpha_k/2}}\,\mathrm{d}t \tag{1.3}$$

for $x > 0$. See [11, p. 134]; the result is due to [6]. A classical result not noted in the review paper [11] is Sim's expression for f_S [14, p. 140].

Remark 1.1. When $X_1, \dots, X_n$ are all exponential (i.e., $\alpha_i = 1$ for all $i = 1, \dots, n$) with distinct means (i.e., $i \neq j \Rightarrow \beta_i \neq \beta_j$), f_S is given by

$$f_S(x) = \prod_{i=1}^{n} \lambda_i \sum_{j=1}^{n} \frac{\exp(-\lambda_j x)}{\prod_{k=1, k\neq j}^{n} (\lambda_k - \lambda_j)}$$

for $x > 0$, where $\lambda_i = 1/\beta_i$, $i = 1, \dots, n$. See e.g. [11, Eq. (1)].

1.2. The common approximation

The common approach is to approximate S by a gamma RV with the same first and second moments (moment-matching method, henceforth abbreviated as MMM). See e.g. [1, 8], and, in particular, [15] and Section 4.1 in the unpublished notes of Massey, available at http://www-personal.umd.umich.edu/~fmassey/gammaRV. Nevertheless, it might not be easy to motivate the idea of a gamma approximation to S in the first place (in a self-contained manner); in this context, see [1] and reference 14 therein, and Massey's notes. A straightforward motivation for this idea is offered at the beginning of Section 2.2.

Let $X_{\mathrm{m}} \sim \mathrm{Gamma}(\alpha_{\mathrm{m}}, \beta_{\mathrm{m}})$ denote the approximating RV associated with the MMM (the subscripts "m" here mean "moments"). By

$$\mathrm{E}(X_{\mathrm{m}}) = \alpha_{\mathrm{m}}\beta_{\mathrm{m}}, \ \mathrm{E}(S) = \sum_{i=1}^{n} \alpha_i \beta_i,$$

$$\mathrm{Var}(X_{\mathrm{m}}) = \alpha_{\mathrm{m}}\beta_{\mathrm{m}}^2, \ \mathrm{Var}(S) = \sum_{i=1}^{n} \alpha_i \beta_i^2,$$

the MMM yields

$$\alpha_{\mathrm{m}} = \frac{\mu^2}{\sum_{i=1}^{n} \alpha_i \beta_i^2}, \ \beta_{\mathrm{m}} = \frac{\sum_{i=1}^{n} \alpha_i \beta_i^2}{\mu}, \tag{1.4}$$

where

$$\mu = \sum_{i=1}^{n} \alpha_i \beta_i.$$

Define

$$\beta_{\min} = \min(\beta_1, \dots, \beta_n), \ \beta_{\max} = \max(\beta_1, \dots, \beta_n),$$

and

$$\alpha_{\min} = \min(\alpha_1, \dots, \alpha_n).$$

The following result is easy to establish (see Section 4.1, Propositions 3, 4, in the unpublished notes of Massey).

Proposition 1.1. *The parameter β_{m} has the following lower and upper bounds:*

$$\beta_{\min} \leq \frac{\mu}{\sum_{i=1}^{n} \alpha_i} \leq \beta_{\mathrm{m}} \leq \beta_{\max}. \tag{1.5}$$

These inequalities are strict unless all the β_i are equal. The parameter α_{m} has the following lower and upper bounds:

$$\alpha_{\min} < \alpha_{\mathrm{m}} \le \sum_{i=1}^{n} \alpha_i. \tag{1.6}$$

The right inequality is strict unless all the β_i are equal.

The only non-trivial part in proving Proposition 1.1 is to show that

$$\mu^2 \le \sum_{i=1}^{n} \alpha_i \sum_{i=1}^{n} \alpha_i \beta_i^2,$$

with strict inequality unless all the β_i are equal. This result follows directly from Cauchy-Schwarz inequality.

The MMM yields the right parameters for the case when all the β_i are equal. Namely, if $X_i \sim \mathrm{Gamma}(\alpha_i, \beta)$ for all $i = 1, \ldots, n$, then X_{m}, as S, is $\mathrm{Gamma}(\sum_{i=1}^{n} \alpha_i, \beta)$ distributed.

Remark 1.2. An even simpler approximation to S is given in [17, Theorem 16]; it is stated that to approximate S in the sense of relative entropy, $\mathrm{Gamma}(\alpha_+, \mu/\alpha_+)$, which has the same mean as S, is no worse than $\mathrm{Gamma}(a, b)$ whenever $a \ge \alpha_+$, where $b > 0$ and $\alpha_+ \equiv \sum_{i=1}^{n} \alpha_i$. However, as can be verified numerically, this approximation is inferior, to say the least (cf. (1.6)).

2. A novel approximation

2.1. *Preliminaries*

This section provides some basic facts from the theory of Lévy processes concerning gamma distributions and their convolutions. A classical reference on Lévy processes is the comprehensive book [13].

Gamma distributions and their convolutions can be characterized in terms of the associated *Lévy densities*. Let X be a $\mathrm{Gamma}(\alpha, \beta)$ RV. Its Laplace transform (given explicitly by $\mathrm{E}[\mathrm{e}^{-uX}] = (1+\beta u)^{-\alpha}$, $u \ge 0$) admits the following representation (see e.g. [13, Example 8.10]):

$$\mathrm{E}\big[\mathrm{e}^{-uX}\big] = \exp\left[\int_0^\infty (\mathrm{e}^{-ux} - 1)\frac{\alpha \mathrm{e}^{-x/\beta}}{x}\,\mathrm{d}x\right]$$

for $u \ge 0$. It follows readily (using (1.1)) that

$$\mathrm{E}\big[\mathrm{e}^{-uS}\big] = \exp\left[\int_0^\infty (\mathrm{e}^{-ux} - 1)\frac{\sum_{i=1}^{n} \alpha_i \mathrm{e}^{-x/\beta_i}}{x}\,\mathrm{d}x\right].$$

The functions ρ, ρ_S defined for $x > 0$ by

$$\rho(x) = \frac{\alpha \mathrm{e}^{-x/\beta}}{x},\ \rho_S(x) = \frac{\sum_{i=1}^{n} \alpha_i \mathrm{e}^{-x/\beta_i}}{x} \tag{2.1}$$

are thus the Lévy densities of X and S, respectively (cf. [13, Eq. (21.1)]; the measures ν, ν_S on $(0,\infty)$ given by $\nu(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$, $\nu_S(\mathrm{d}x) = \rho_S(x)\,\mathrm{d}x$ are the Lévy measures of X and S, respectively). They can be interpreted as follows. Let $\{X(t) : t \geq 0\}$ be a *gamma process* such that $X(1)$, as X, is Gamma(α,β) distributed (then $X(t) \sim$ Gamma$(\alpha t,\beta)$, for any $t > 0$), and let $\{X_i(t) : t \geq 0\}$, $i = 1,\dots,n$, be independent gamma processes such that $X_i(1)$, as X_i, is Gamma(α_i,β_i) distributed. Define the process $\{S(t) : t \geq 0\}$ by $S(t) = \sum_{i=1}^n X_i(t)$, $t \geq 0$, so that $S(1)$ and S are identically distributed. Being Lévy processes, the above processes have stationary independent increments and are zero at $t = 0$. Further, they are pure jump processes with strictly increasing sample paths; then their value at time t equals the accumulated sum of jumps up to time t. The number of jumps of the process $\{X(\cdot)\}$ (respectively, $\{S(\cdot)\}$) up to time $t\,(> 0)$ with size in the interval $[a,b] \subset (0,\infty)$ is Poisson distributed with parameter $t\int_a^b \rho(x)\,\mathrm{d}x$ (respectively, $t\int_a^b \rho_S(x)\,\mathrm{d}x$). Consequently, both processes have infinitely many jumps (of very small size) in any finite time interval (note that ρ and ρ_S integrate to ∞ over the positive half-line, due to their behavior near 0). In conclusion, the RV's X and S can be represented as an infinite sum of random jumps, which is characterized by the corresponding Lévy density as indicated above for the RV's $X(t)$ and $S(t)$ at time $t = 1$.

By a general property of increasing Lévy processes (see (4.2) and (4.4)), the mean and variance of X and S can be simply expressed in terms of the corresponding Lévy density. Indeed, note that

$$\int_0^\infty x\rho(x)\,\mathrm{d}x = \alpha\beta,\ \int_0^\infty x\rho_S(x)\,\mathrm{d}x = \sum_{i=1}^n \alpha_i\beta_i, \tag{2.2}$$

$$\int_0^\infty x^2\rho(x)\,\mathrm{d}x = \alpha\beta^2,\ \int_0^\infty x^2\rho_S(x)\,\mathrm{d}x = \sum_{i=1}^n \alpha_i\beta_i^2. \tag{2.3}$$

Further, consider the functions H and H_S defined, for $x \geq 0$, by

$$H(x) \equiv \int_x^\infty u\rho(u)\,\mathrm{d}u = \alpha\beta\mathrm{e}^{-x/\beta}$$

and

$$H_S(x) \equiv \int_x^\infty u\rho_S(u)\,\mathrm{d}u = \sum_{i=1}^n \alpha_i\beta_i\mathrm{e}^{-x/\beta_i}.$$

For $x = 0$, it holds $H(0) = \mathrm{E}(X)$ and $H_S(0) = \mathrm{E}(S)$. For $x > 0$, consider the representation of X and S as a sum of jumps, as indicated at the end of the previous paragraph. Truncating the jumps smaller than x results in (compound Poisson) RV's, $X^{x\uparrow}$ and $S^{x\uparrow}$, such that $\mathrm{E}(X^{x\uparrow}) = H(x)$ and $\mathrm{E}(S^{x\uparrow}) = H_S(x)$. The functions H and H_S will play a fundamental role in the sequel.

2.2. The approximation scheme

The basic observation is that a single-gamma approximation to S is appropriate due to the evident similarity between the corresponding Lévy densities, ρ and

ρ_S, as given in (2.1). In any approximation scheme, it is desirable that the approximating gamma RV satisfies the following requirements: it is identically distributed as S in the trivial case where all the β_i are equal, and it has the same mean as S. The RV $X_{\rm m}$ of Section 1.2 satisfies both requirements, and the additional requirement that $\mathrm{Var}(X_{\rm m}) = \mathrm{Var}(S)$. By (2.2)–(2.3), the requirements $\mathrm{E}(X_{\rm m}) = \mathrm{E}(S)$, $\mathrm{Var}(X_{\rm m}) = \mathrm{Var}(S)$ (yielding (1.4)) can be expressed in terms of the Lévy densities $\rho_{\rm m}$ and ρ_S of $X_{\rm m}$ and S as

$$\int_0^\infty x\rho_{\rm m}(x)\,\mathrm{d}x = \int_0^\infty x\rho_S(x)\,\mathrm{d}x, \tag{2.4}$$

$$\int_0^\infty x^2\rho_{\rm m}(x)\,\mathrm{d}x = \int_0^\infty x^2\rho_S(x)\,\mathrm{d}x, \tag{2.5}$$

respectively. We now turn to consider the proposed approximation.

Motivated by the observation made at the beginning of this section, we propose the following approximation scheme. Let $X_* \sim \mathrm{Gamma}(\alpha_*, \beta_*)$ denote the approximating RV, and ρ_* its Lévy density. Define the function H_* by

$$H_*(x) \equiv \int_x^\infty u\rho_*(u)\,\mathrm{d}u = \alpha_*\beta_*\mathrm{e}^{-x/\beta_*} \tag{2.6}$$

for $x \geq 0$. Then, the desired condition $\mathrm{E}(X_*) = \mathrm{E}(S)$, i.e.

$$\alpha_*\beta_* = \mu, \tag{2.7}$$

can be expressed as $H_*(0) = H_S(0)$, which, as written, is the same as condition (2.4) with $\rho_{\rm m}$ replaced by ρ_*. The counterpart of condition (2.5) for the proposed approximation is based on a practical heuristic, as follows (see further Remark 4.1 for clarification and generalization of the underlying idea). Let X_* and $X_*^{x\uparrow}$, for $x > 0$, play the role of X and $X^{x\uparrow}$ at the end of Section 2.1, respectively. Then $H_*(x) = \mathrm{E}(X_*^{x\uparrow})$. Now, suppose that, in some sense, $\{\mathrm{E}(X_*^{x\uparrow}) : x > 0\}$ well approximates $\{\mathrm{E}(S^{x\uparrow}) : x > 0\}$; then, one may intuitively expect that X_* appropriately approximates S. Noting that, under (2.7),

$$\big|\mathrm{E}\big(X_*^{x\uparrow}\big) - \mathrm{E}\big(S^{x\uparrow}\big)\big| = |H_*(x) - H_S(x)| = \left|\mu\mathrm{e}^{-x/\beta_*} - \sum_{i=1}^n \alpha_i\beta_i\mathrm{e}^{-x/\beta_i}\right|,$$

it is thus natural to choose H_* to minimize $||H - H_S||_2^2$ over H of the form $H(x) = \mu\mathrm{e}^{-x/\beta}$, i.e. define the parameter β_* as the minimizer, over $\beta > 0$, of the squared distance $\psi(\beta)$ given by

$$\psi(\beta) = \int_0^\infty \left[\mu\mathrm{e}^{-x/\beta} - \sum_{i=1}^n \alpha_i\beta_i\mathrm{e}^{-x/\beta_i}\right]^2 \mathrm{d}x. \tag{2.8}$$

The parameter β_* is derived in Theorem 2.1 below. Then α_* is determined from (2.7).

We conclude this section with a few remarks. It would have been essentially the same, but somewhat less convenient, to consider the counterparts $\tilde{H}_*$ and

$\tilde{H}_S$ of H_* and H_S defined, for $0 < x \leq \infty$, by $\tilde{H}_*(x) = \int_0^x u\rho_*(u)\,\mathrm{d}u$ and $\tilde{H}_S(x) = \int_0^x u\rho_S(u)\,\mathrm{d}u$. On the other hand, one can consider substantially different approximation schemes based on the observation made at the beginning of this section. However, the present approximation proved to be quite satisfactory: both from an accuracy point of view (see Section 3) and from the point of view of mathematical tractability (as reflected in Section 2.3). The fact that this approximation is based on a global condition (consider (2.8)) may account for its good performance. In this context, note that, using the notation of Section 2.1, the conditions (1) $\mathrm{E}(X) = \mathrm{E}(S)$ and (2) $x\rho(x)|_{0+} = x\rho_S(x)|_{0+}$ (or, equivalently, $H'(0+) = H'_S(0+)$) yield $\alpha = \sum_{i=1}^n \alpha_i$, $\beta = \mu / \sum_{i=1}^n \alpha_i$, i.e. the Gamma$(\alpha_+, \mu/\alpha_+)$ approximation in Remark 1.2. However, condition (2) is local, and hence it is not surprising that the Gamma$(\alpha_+, \mu/\alpha_+)$ approximation is inferior. Finally, note that analogous approximations can be established based on the following counterparts of (2.8), where $H(x) = \mu \mathrm{e}^{-x/\beta}$:

$$\psi_1(\beta) = ||H - H_S||_1 = \int_0^\infty \left| \mu \mathrm{e}^{-x/\beta} - \sum_{i=1}^n \alpha_i \beta_i \mathrm{e}^{-x/\beta_i} \right| \mathrm{d}x,$$

$$\psi_\infty(\beta) = ||H - H_S||_\infty = \max_{x>0} \left| \mu \mathrm{e}^{-x/\beta} - \sum_{i=1}^n \alpha_i \beta_i \mathrm{e}^{-x/\beta_i} \right|.$$

The prominent advantage of (2.8) lies in its mathematical tractability.

2.3. The main results

We now state the main results for the sum-of-gammas case. The proofs are given in the appendix. Note that Proposition 2.1 refines the upper bound (respectively, lower bound) on β_* (respectively, α_*) given in Theorem 2.1.

Theorem 2.1. *The parameter β_* is the solution $\beta > 0$ of the equation*

$$\frac{\mu}{2} - 2\sum_{i=1}^n \frac{\alpha_i \beta_i^3}{(\beta_i + \beta)^2} = 0. \tag{2.9}$$

It has the following lower and upper bounds:

$$\beta_{\min} \leq \frac{\mu}{\sum_{i=1}^n \alpha_i} \leq \beta_* \leq \beta_{\max}. \tag{2.10}$$

These inequalities are strict unless all the β_i are equal. It follows that

$$\alpha_{\min} < \alpha_* \leq \sum_{i=1}^n \alpha_i, \tag{2.11}$$

where the right inequality is strict unless all the β_i are equal.

Remark 2.1. The parameter β_* is readily available numerically. In particular, the bisection method can be applied, since the left-hand side of (2.9) strictly increases from negative to positive as β increases from $\beta_{\min}$ to $\beta_{\max}$.

Proposition 2.1. *It holds that* $\beta_* \le \beta_\mathrm{m}$ *(and hence* $\alpha_* \ge \alpha_\mathrm{m}$*). The inequality is strict unless all the* β_i *are equal.*

Corollary 2.1. *It holds that* $\mathrm{Var}(X_*) \le \mathrm{Var}(S)$*. The inequality is strict unless all the* β_i *are equal.*

Remark 2.2. For the non-trivial case where $\beta_{\min} < \beta_{\max}$, we can assume without loss of generality that the β_i are all distinct. Indeed, note that if we partition $\{1, \ldots, n\}$ into sets A_j, $j = 1, \ldots, k$, such that, for each j, $i \in A_j \Rightarrow \beta_i = \tilde{\beta}_j$, where $\tilde{\beta}_1, \ldots, \tilde{\beta}_k$ are distinct ($k \ge 2$), then the same approximations ($\mathrm{Gamma}(\alpha_\mathrm{m}, \beta_\mathrm{m})$ and $\mathrm{Gamma}(\alpha_*, \beta_*)$) would be obtained for the sum $X_1 + \cdots + X_n$ and the sum $\sum_{i \in A_1} X_i + \cdots + \sum_{i \in A_k} X_i$ $(= X_1 + \cdots + X_n)$ of k independent $\mathrm{Gamma}\big(\sum_{i \in A_j} \alpha_i, \tilde{\beta}_j\big)$ RV's.

3. Numerical study

This section performs a brief numerical study of the approximations X_* and X_m to S. The quality of the approximations has been tested by comparing the gamma densities of the approximating RV's, denoted by f_* and f_m respectively, with the exact density of S, namely f_S. The density f_S has been evaluated using (1.2), by truncating the infinite sum at a suitably large value. It should be stressed, however, that the study is very limited in scope, as there are numerous combinations $(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)$ that are worth considering. Here we highlight only a few prominent points, based on results obtained for $n = 2, 3, 4$, with $\alpha_i \in (0.5, 20)$ and $\beta_i \in (0.1, 20)$. More specifically, let $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$ be independent uniform$(0.5, 20)$ and uniform$(0.1, 20)$ RV's, respectively. Suppose without loss of generality that the β_i are sorted in increasing order, so that, in particular, $\beta_1 = \min_i(\beta_i)$ (and hence (1.2) can be applied). For each fixed $n = 2, 3, 4$, we generated 10000 realizations of the parameters α_i and β_i, and computed the corresponding parameters α_m, β_m, α_*, and β_*, as well as the ratio β_*/β_m $(= \alpha_\mathrm{m}/\alpha_*)$. The generated data provided the basis for most of our conclusions presented below.

Our first observation is that the approximating distributions $\mathrm{Gamma}(\alpha_*, \beta_*)$ and $\mathrm{Gamma}(\alpha_\mathrm{m}, \beta_\mathrm{m})$, though very different in construction, are typically very close to each other (while both have the same mean, by Corollary 2.1 the former has smaller variance, unless in the trivial case when S itself is gamma distributed). We demonstrate this by showing that the ratio β_*/β_m is typically very close to 1. For $N = 10000$, a fixed $n = 2, 3, 4$, and an interval $I \subset (0, 1)$, let $Q_N(I; n) = \sum_{k=1}^N \mathbf{1}((\beta_*/\beta_\mathrm{m})_k \in I)$, where $\mathbf{1}$ is the indicator function and $(\beta_*/\beta_\mathrm{m})_k$ denotes the kth realization of β_*/β_m as indicated above. Tabulated values of $Q_N(I; n)$ for various intervals $I \subset (0, 1)$ are presented in Table 1, confirming our claim.

Having concluded that the approximating distributions are typically very close to each other, we proceed to consider briefly the quality of the approximations. In analyzing the quality of the $\mathrm{Gamma}(\alpha_\mathrm{m}, \beta_\mathrm{m})$ approximation, Stewart et al. [15] considered the eight parameter combinations $(\alpha_1, \alpha_2, \beta_2)$ presented in

TABLE 1
Tabulated values of $Q_N(I;n)$, $N = 10000$, for $n = 2, 3, 4$ and various intervals $I \subset (0, 1)$

I	n		
	2	3	4
$(0.99, 1.00)$	5588	3594	2519
$(0.98, 0.99)$	1832	2627	2957
$(0.97, 0.98)$	1080	1501	1914
$(0.96, 0.97)$	567	905	1114
$(0.95, 0.96)$	303	471	585
$(0.94, 0.95)$	172	276	349
$(0.93, 0.94)$	100	198	189
$(0.92, 0.93)$	96	111	117
$(0.91, 0.92)$	54	92	70
$(0.90, 0.91)$	40	54	45
$(0.60, 0.90)$	168	171	141

TABLE 2
Tabulated rounded values of $\alpha_{\rm m}$, $\beta_{\rm m}$, α_*, β_*, and $\beta_*/\beta_{\rm m}$, for the eight parameter combinations $(\alpha_1, \alpha_2, \beta_2)$ considered in [15]

α_1	α_2	β_1	β_2	$\alpha_{\rm m}$	$\beta_{\rm m}$	α_*	β_*	$\beta_*/\beta_{\rm m}$
2	2	1	2	3.6000	1.6667	3.6724	1.6338	0.9803
5	2	1	2	6.2308	1.4444	6.4222	1.4014	0.9702
2	5	1	2	6.5455	1.8333	6.6099	1.8155	0.9902
5	5	1	2	9.0000	1.6667	9.1809	1.6338	0.9803
2	2	1	10	2.3960	9.1818	2.4167	9.1033	0.9914
5	2	1	10	3.0488	8.2000	3.1336	7.9780	0.9729
2	5	1	10	5.3865	9.6538	5.4026	9.6250	0.9970
5	5	1	10	5.9901	9.1818	6.0418	9.1033	0.9914

Table 2 of the present paper ($\beta_1 = 1$). They concentrated on the case $n = 2$, arguing that the approximation generally improves with increasing n. Figure 2 of Stewart et al. [15], as does the bottom plot in Figure 1 of the present paper, shows the corresponding eight exact density functions f_S and their approximations $f_{\rm m}$; the overall agreement is evidently good. Figure 3 of Stewart et al. relates to the cumulative distribution functions. Note that, for each of the parameter combinations in Table 2, $\beta_*/\beta_{\rm m}$ is very close to 1 (note the agreement with Table 1). Accordingly, and as the top plot in Figure 1 confirms, f_S is also well approximated by f_* (apparently at least as good as by $f_{\rm m}$).

Clearly, it is reasonable to expect that a ratio $\beta_*/\beta_{\rm m}$ significantly smaller than 1 (say, smaller than 0.8) generally corresponds to the case of significantly different approximations f_* and $f_{\rm m}$ to f_S, so that at least one of them might not be satisfactory enough. While in this (non-typical) case ${\rm Var}(X_*) = (\beta_*/\beta_{\rm m}){\rm Var}(S) << {\rm Var}(S)$ and ${\rm Var}(X_{\rm m}) = {\rm Var}(S)$, the former approximation may be preferable, as we see next. For each $n = 2, 3, 4$, four examples were selected (from the data mentioned in the first paragraph of this section) in which $(\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n)$ yield a $\beta_*/\beta_{\rm m}$ value significantly smaller than 1; see Tables 3–5. Figures 2–4 show the corresponding exact density functions f_S and their approximations f_* and $f_{\rm m}$, confirming our claim.

Despite the limited scope of our numerical study, the overall results (including others not presented here) suggest that the approximation f_* to f_S is, in general,

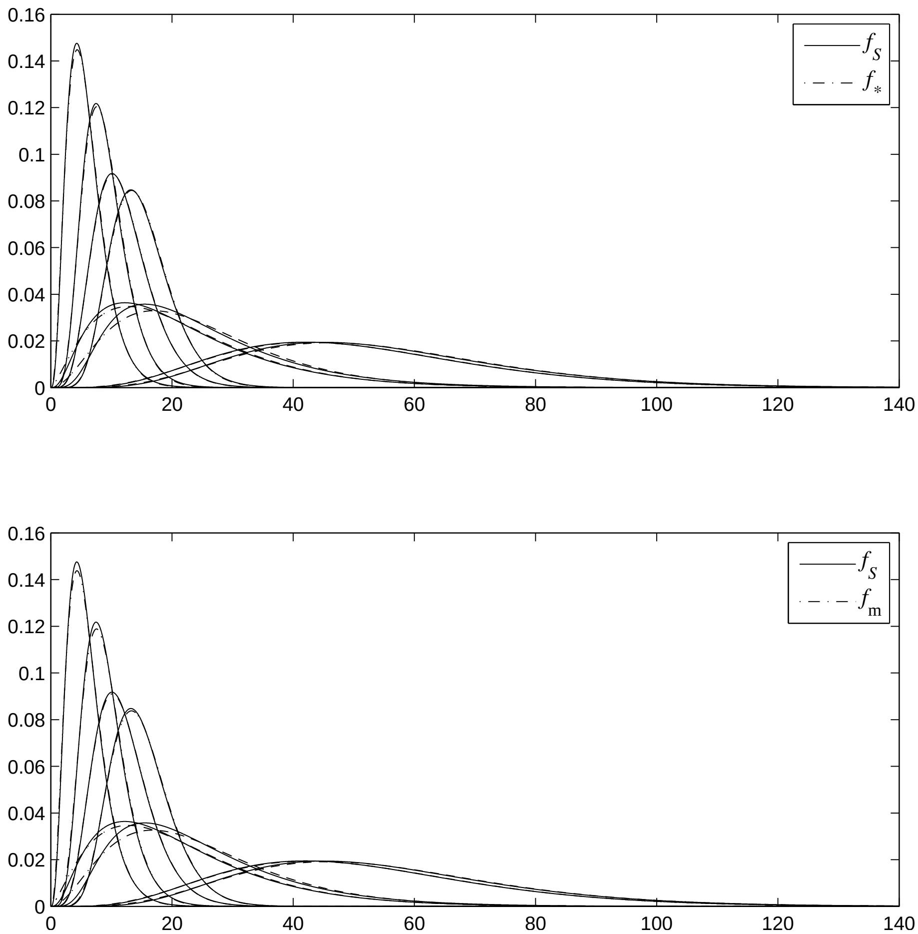

FIG 1. *Plots of f_S and f_* (top) and of f_S and f_{m} (bottom), corresponding to the eight parameter combinations in Table 2.*

TABLE 3
Examples where $(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)$, $n = 2$, yield a ratio $\beta_*/\beta_{\mathrm{m}}$ significantly smaller than 1 (rounded values)

Example	i	α_i	β_i	α_{m}	β_{m}	α_*	β_*	$\beta_*/\beta_{\mathrm{m}}$
1	1	15.368	1.594	4.699	8.053	6.917	5.471	0.679
	2	0.671	19.901					
2	1	19.391	1.691	11.993	3.263	16.014	2.444	0.749
	2	0.556	11.397					
3	1	7.832	0.424	2.322	2.568	2.988	1.995	0.777
	2	0.501	5.265					
4	1	19.395	2.217	9.929	6.714	12.637	5.275	0.786
	2	1.590	14.883					

TABLE 4
Counterpart of Table 3 for the case $n = 3$

Example	i	α_i	β_i	$\alpha_{\rm m}$	$\beta_{\rm m}$	α_*	β_*	$\beta_*/\beta_{\rm m}$
1	1	17.718	0.400	7.135	5.257	10.821	3.466	0.659
	2	10.316	2.009					
	3	0.615	15.761					
2	1	6.238	1.824	7.378	6.425	9.553	4.962	0.772
	2	6.937	3.814					
	3	0.501	19.110					
3	1	5.523	0.516	12.717	4.145	16.238	3.246	0.783
	2	19.050	1.852					
	3	1.401	10.405					
4	1	14.198	1.701	6.725	7.558	8.455	6.011	0.795
	2	1.765	3.682					
	3	1.276	15.817					

TABLE 5
Counterpart of Tables 3 and 4 for the case $n = 4$

Example	i	α_i	β_i	$\alpha_{\rm m}$	$\beta_{\rm m}$	α_*	β_*	$\beta_*/\beta_{\rm m}$
1	1	6.002	1.734	10.162	5.267	14.522	3.686	0.700
	2	14.365	1.874					
	3	0.980	5.899					
	4	0.606	17.211					
2	1	15.763	0.591	10.677	5.813	13.520	4.591	0.790
	2	10.882	1.335					
	3	5.440	2.679					
	4	1.884	12.553					
3	1	19.531	1.038	8.292	6.514	10.496	5.146	0.790
	2	1.065	2.816					
	3	2.996	4.574					
	4	1.118	15.240					
4	1	4.471	0.434	14.391	2.186	18.205	1.728	0.791
	2	17.847	0.594					
	3	7.332	1.206					
	4	1.995	5.056					

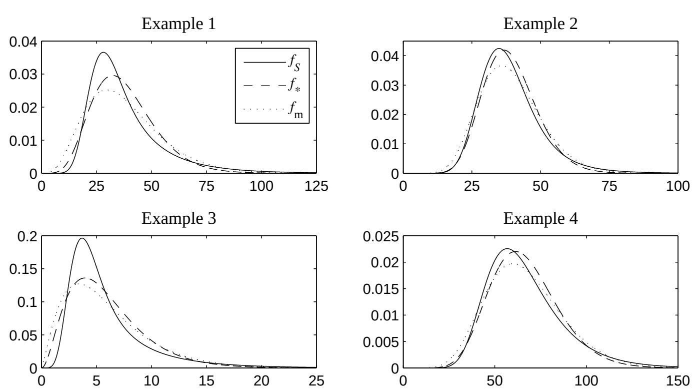

FIG 2. *Plots of f_S, f_*, and $f_{\rm m}$ corresponding to Examples 1–4 of Table 3.*

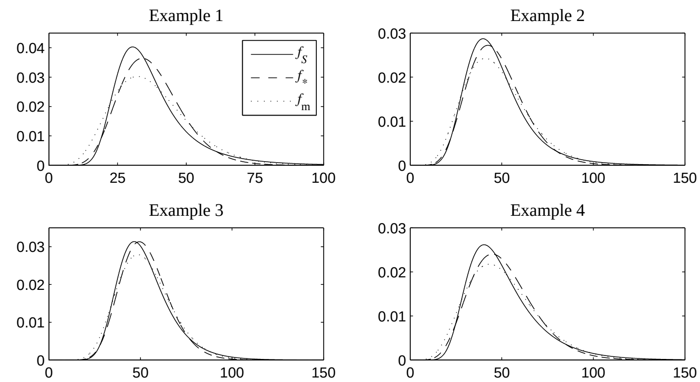

FIG 3. *Counterpart of Figure 2 for Table 4.*

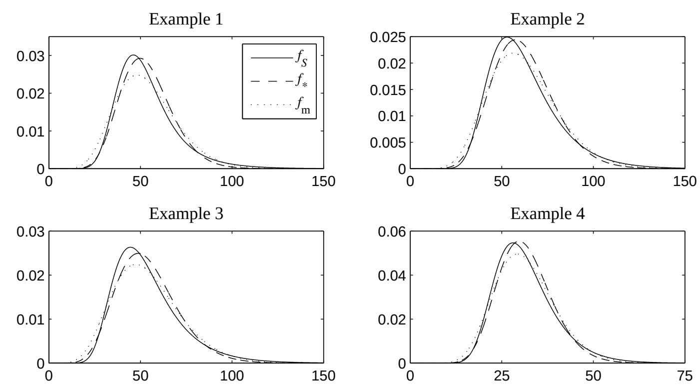

FIG 4. *Counterpart of Figures 2 and 3 for Table 5.*

slightly better than f_{m} (still, these approximations are typically very close to each other). Here we assume that the underlying parameters α_i and β_i are such that the approximations are appropriate in the first place. While this is typically the case, with suitably chosen parameter values the approximations error might be very large.

4. Gamma approximation to infinitely divisible (ID) distributions on $\mathbb{R}_+$

4.1. Preliminaries

This section provides preliminaries on ID distributions on $\mathbb{R}_+$ and the associated *subordinators*. Additional details can be found in [13] or elsewhere.

The gamma distribution is ID; i.e., for any $n \in \mathbb{N}$, it is the n-fold convolution of a probability measure μ_n on $\mathbb{R}$ (specifically, $\mathrm{Gamma}(\alpha, \beta) = \mu_n^{*n}$, where

$\mu_n = \text{Gamma}(\alpha/n, \beta)$). The distribution of S is ID, being a convolution of gamma distributions. The class of ID distributions includes surprisingly many important distributions as special cases; see e.g. [13, Chapter 2, Section 8]. It is a basic fact in the theory of Lévy processes that there is a one-to-one correspondence between ID distributions and distributions of Lévy processes at time 1 [13, Theorem 7.10]. Lévy processes with nondecreasing paths, or equivalently (by [13, Theorem 24.11]) with one-dimensional distributions on $\mathbb{R}_+$, are called subordinators. Thus, there is a one-to-one correspondence between ID distributions on $\mathbb{R}_+$ and distributions of subordinators at time 1. Clearly, addition of a positive constant, say $\gamma_0 > 0$, preserves infinite divisibility on $\mathbb{R}_+$; this corresponds to addition of a drift term $\{\gamma_0 t : t \geq 0\}$ to the associated subordinator. Hence, in the present context, it suffices to consider driftless subordinators.

The methodology used to derive the approximation X_* to S (including its variants indicated at the end of Section 2.2) can be adapted to any other *integrable* ID random variable on $\mathbb{R}_+$, as long as the underlying Lévy measure (or density, if it exists) is convenient for the calculations involved. In fact, as discussed in Section 4.2, it is often the case in general that an ID distribution of interest has a simple Lévy measure yet a complicated distribution or density function (two specific examples being the convolution of gamma distributions with arbitrary parameters and the generalized Dickman distribution considered in Section 4.3 below); hence the importance of the proposed methodology. As a particularly interesting opposite example, the lognormal distribution, although ID with a simple density function, has unknown Lévy measure [2].

Let $\{Z(t) : t \geq 0\}$ be a pure-jump subordinator, i.e. a nondecreasing Lévy process with no drift, so that $Z(t)$ equals the accumulated sum of jumps (if any) up to time t. Any such process is completely characterized by a measure, ν_Z, on $(0, \infty)$ such that $\int_{(0,\infty)} \min(x, 1)\nu_Z(\mathrm{d}x) < \infty$ (the Lévy measure of $\{Z(\cdot)\}$). Specifically, the number of jumps of the process $\{Z(\cdot)\}$ up to time t with size in the interval $(a, b] \subset (0, \infty)$ is Poisson distributed with parameter $t\nu_Z((a, b])$ (where Poisson(0) and Poisson(∞) mean 0 and ∞, respectively). Consequently, if $\nu_Z((0, \infty)) = \infty$ (equivalently, $\nu_Z((0, \varepsilon]) = \infty$ for any $\varepsilon > 0$), the process has infinitely many jumps (of very small size) in any finite time interval; if, on the other hand, $\lambda := \nu_Z((0, \infty)) \in (0, \infty)$, then $\{Z(\cdot)\}$ is a compound Poisson process (CPP) with rate λ and jump distribution $F = \lambda^{-1}\nu_Z$, so that it can be represented as $Z(t) = \sum_{i=1}^{N(t)} Y_i$, where $\{N(t) : t \geq 0\}$ is a Poisson process with rate λ independent of a sequence $Y_1, Y_2, \ldots$ of i.i.d. RV's with distribution F. In either case, $Z(t)$ is ID with Laplace transform

$$\mathrm{E}\left[\mathrm{e}^{-uZ(t)}\right] = \exp\left[t \int_{(0,\infty)} (\mathrm{e}^{-ux} - 1)\nu_Z(\mathrm{d}x)\right] \tag{4.1}$$

for $u \geq 0$; see e.g. [13, Eq. (21.1)]. (This simplifies easily in the CPP case, in terms of the Laplace transform of F.) Suppose in the sequel that Z is equal in distribution to $Z(1)$, and actually identify the two RV's. (Z will play the same role as S before, i.e. the RV to be approximated.) Then, Z is integrable if and only if $\int_{(1,\infty)} x\nu_Z(\mathrm{d}x) < \infty$, in which case

$$\mathrm{E}(Z) = \int_{(0,\infty)} x\nu_Z(\mathrm{d}x). \tag{4.2}$$

Assuming that Z is integrable, consider the function H_Z defined, for $x \geq 0$, by

$$H_Z(x) = \int_{(x,\infty)} u\nu_Z(\mathrm{d}u). \tag{4.3}$$

For $x = 0$, it holds $H_Z(0) = \mathrm{E}(Z)$. For $x > 0$, consider the representation of Z as a sum of jumps, as indicated above for the RV $Z(t)$ at time $t = 1$. Truncating the jumps not exceeding x results in a (compound Poisson) RV, $Z^{x\uparrow}$, such that $\mathrm{E}(Z^{x\uparrow}) = H_Z(x)$. (The function H_Z will play the same role as H_S before.)

For completeness and in view of (4.11) below, it is worth noting the following. Z has finite variance if and only if $\int_{(1,\infty)} x^2\nu_Z(\mathrm{d}x) < \infty$, in which case

$$\mathrm{Var}(Z) = \int_{(0,\infty)} x^2\nu_Z(\mathrm{d}x). \tag{4.4}$$

Thus, with the above notation for the CPP case, if $\nu_Z((0,\infty)) \in (0,\infty)$, then $\mathrm{Var}(Z) = \lambda \int_{(0,\infty)} x^2 F(\mathrm{d}x) = \lambda\mathrm{E}(Y_1^2)$, as required. In fact, both (4.2) and (4.4) are special cases of the following result (see [16, Proposition 1.2] or, more specifically, [3, p. 93]). Z has finite nth moment if and only if $\int_{(1,\infty)} x^n\nu_Z(\mathrm{d}x) < \infty$, in which case Z has *cumulants* κ_j, $j = 1,\ldots,n$, given by

$$\kappa_j = \int_{(0,\infty)} x^j\nu_Z(\mathrm{d}x). \tag{4.5}$$

In particular, if Z has finite fourth moment, its skewness and kurtosis are given by $\kappa_3/{\kappa_2}^{3/2}$ and $\kappa_4/{\kappa_2}^2$, respectively. For the Gamma(α,β) distribution, (4.5) yields $\kappa_n = (n-1)!\alpha\beta^n$ for every n, and thus skewness $2/\sqrt{\alpha}$ and kurtosis $6/\alpha$.

4.2. Description and illustration of the general methodology

Let Z be an integrable ID RV as above, to be approximated by a gamma RV. Denote its mean by μ_Z. As before, let $X_* \sim \mathrm{Gamma}(\alpha_*,\beta_*)$ denote the approximating RV, and define H_* as in (2.6). Then, the desired condition $\mathrm{E}(X_*) = \mathrm{E}(Z)$, i.e.

$$\alpha_*\beta_* = \mu_Z, \tag{4.6}$$

can be expressed as $H_*(0) = H_Z(0)$. Proceeding analogously to Section 2.2, and with the same notation, note that, under (4.6),

$$\left|\mathrm{E}\big(X_*^{x\uparrow}\big) - \mathrm{E}\big(Z^{x\uparrow}\big)\right| = |H_*(x) - H_Z(x)| = \left|\mu_Z \mathrm{e}^{-x/\beta_*} - \int_{(x,\infty)} u\nu_Z(\mathrm{d}u)\right|.$$

Then, depending on the complexity of the calculations involved, etc., H_* is to be chosen to minimize $||H - H_Z||_1$ or $||H - H_Z||_2^2$ or $||H - H_Z||_\infty$, respectively, over H of the form

$$H(x) = \mu_Z \mathrm{e}^{-x/\beta}, \tag{4.7}$$

i.e. β_* is to be defined as the minimizer, over $\beta > 0$, of

$$\psi_1(\beta) = \int_0^\infty \big|H(x) - H_Z(x)\big|\,\mathrm{d}x \tag{4.8}$$

or

$$\psi_2(\beta) = \int_0^\infty \big[H(x) - H_Z(x)\big]^2\,\mathrm{d}x \tag{4.9}$$

or

$$\psi_\infty(\beta) = \max_{x>0} \big|H(x) - H_Z(x)\big|, \tag{4.10}$$

respectively. Once β_* is evaluated, α_* is determined from (4.6). (For convenience, we may omit the distinction between the different β_*'s.)

The following remark is fundamental from a theoretical point of view.

Remark 4.1. Let ρ be the Lévy density of a Gamma(α, β) RV X such that $\alpha\beta = \mu_Z$. The functions H and H_Z defined in (4.7) and (4.3) are the tail measures of M and M_Z, respectively, where M and M_Z are the measures on $(0, \infty)$ given by $M(\mathrm{d}x) = x\rho(x)\,\mathrm{d}x$ and $M_Z(\mathrm{d}x) = x\nu_Z(\mathrm{d}x)$. Under the sum-of-jumps representation of X and Z, $M(\cdot)$ and $M_Z(\cdot)$, respectively, give the mean *sum of jumps* with size in $\cdot$ (rather than the mean *number of jumps* with size in $\cdot$, as do the Lévy measures alone). This fact justifies the theoretical basis of the proposed methodology. Moreover, by virtue of M_Z (and M) being finite with total measure μ_Z, $|{\mu_Z}^{-1}H - {\mu_Z}^{-1}H_Z|$ is merely the absolute difference between two tail distribution functions (DF's), ${\mu_Z}^{-1}H$ being the tail DF of the exponential distribution with mean β.

In view of (4.8)–(4.10), it is essential from a practical point of view that $H_Z(x)$ admits a simple expression. Aiming towards a good approximation, it is desirable that this expression be in notable agreement with $H(x)$. (Note that both H_Z and H are monotone decreasing from μ_Z at $x = 0$ to 0 as $x \to \infty$; further, H_Z is continuous if and only if ν_Z is continuous.) This is fulfilled particularly well in the case $Z = S$, where $H_Z(x) = \sum_{i=1}^n \alpha_i\beta_i \mathrm{e}^{-x/\beta_i}$ and $H(x) = \mu \mathrm{e}^{-x/\beta}$ ($\mu = \sum_{i=1}^n \alpha_i\beta_i$), thus accounting for the high quality of the approximation X_* to S (where (4.9) was used to derive β_*). Plots of $H_S(x)$, $H_*(x) := \mu \mathrm{e}^{-x/\beta_*}$, and $H_\mathrm{m}(x) := \mu \mathrm{e}^{-x/\beta_\mathrm{m}}$ corresponding to the eight parameter combinations $(\alpha_1, \alpha_2, \beta_2)$ of Table 2 are shown in Figures 5–6. Note the agreement between Figures 5–6 and the approximations shown in Figure 1 for the corresponding density functions, f_S, f_*, and f_m. The applicability of the proposed methodology to ID distributions other than convolutions of gammas will be considered and exemplified throughout the rest of this paper.

The following heuristic remark is of particular importance from a practical point of view.

Remark 4.2. When $H_Z(x)$ admits a simple expression which is nonetheless not convenient for the derivation of β_*, it can still be used to indicate the

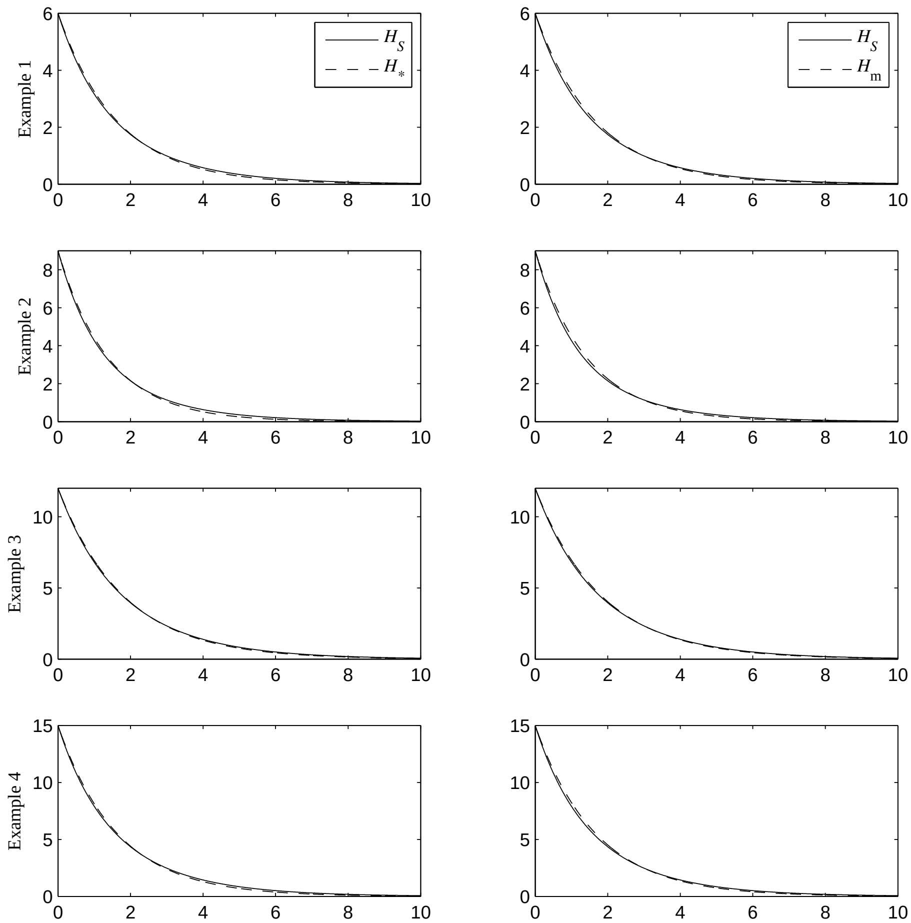

Fig 5. *Plots of* $H_S(x) = \alpha_1\beta_1 e^{-x/\beta_1} + \alpha_2\beta_2 e^{-x/\beta_2}$, $H_*(x) = (\alpha_1\beta_1 + \alpha_2\beta_2)e^{-x/\beta_*}$, *and* $H_{\mathrm{m}}(x) = (\alpha_1\beta_1 + \alpha_2\beta_2)e^{-x/\beta_{\mathrm{m}}}$ *corresponding to the first four parameter combinations in Table 2 (Examples 1–4, respectively).*

appropriateness of the MMM. This assumes, of course, that Z has finite variance, in which case the MMM yields

$$\alpha_{\mathrm{m}} = \frac{\mu_Z^2}{\mathrm{Var}(Z)},\ \beta_{\mathrm{m}} = \frac{\mathrm{Var}(Z)}{\mu_Z} \tag{4.11}$$

as the parameters of the approximating gamma distribution. Suppose that H_Z has a shape similar enough to a decreasing exponential function, so that it agrees well enough with H for suitably chosen β-values (and thus in particular with the theoretical H_*). If it turns out to be the case for $\beta = \beta_{\mathrm{m}}$, then it may be expected that $\beta_* \approx \beta_{\mathrm{m}}$ and that the corresponding gamma approximations are indeed appropriate. This idea will be illustrated repeatedly in the sequel.

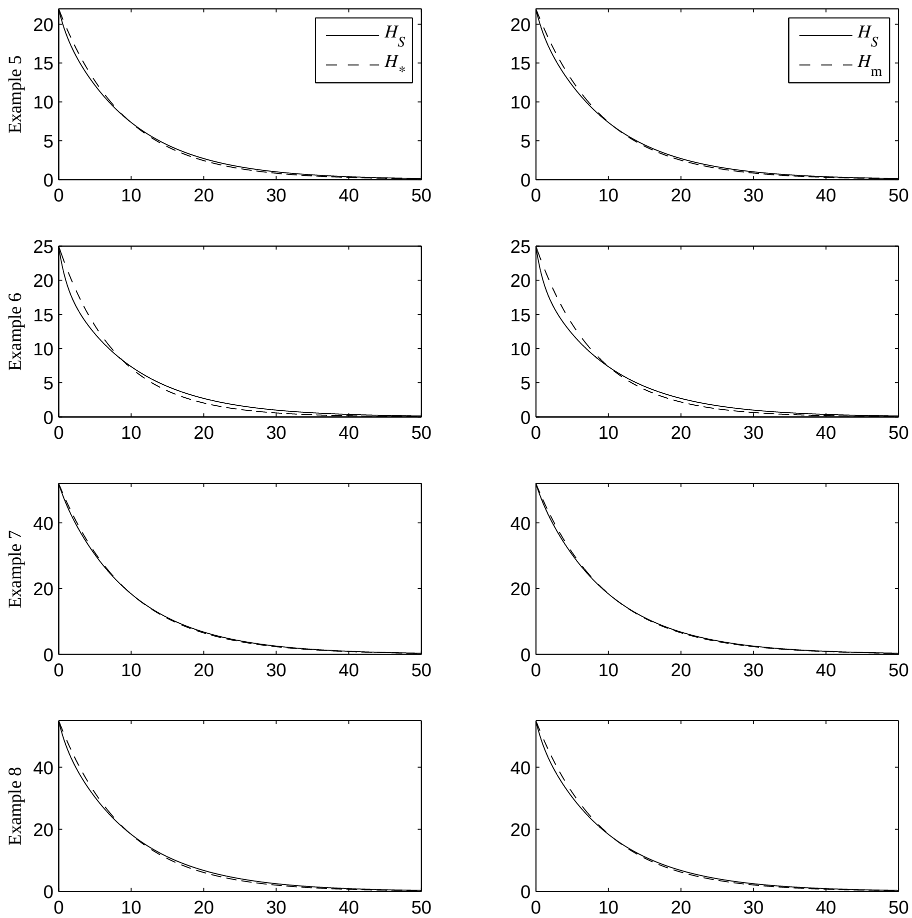

FIG 6. *Counterpart of Figure 5 for the last four parameter combinations in Table 2 (Examples 5–8, respectively).*

The proposed methodology may be suitable for a variety of ID distributions. For a start, suppose that Z has a compound Poisson distribution, corresponding to a CPP (at time 1) with rate λ and absolutely continuous jump distribution F on $(0, \infty)$ with finite mean (so that Z be integrable). The distribution of Z is quite complicated in general. Indeed, by the law of total probability,

$$\mathrm{P}(Z \leq x) = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} F^{*k}([0, x]), \tag{4.12}$$

where F^{*k} is the k-fold convolution of F ($F^{*0} := \delta_0$ is the delta distribution concentrated at 0, $F^{*1} := F$). (Note that Z has mass $\mathrm{e}^{-\lambda}$ at 0.) However, if F is simple, then so is ν_Z, as $\nu_Z = \lambda F$. Two specific examples are given next.

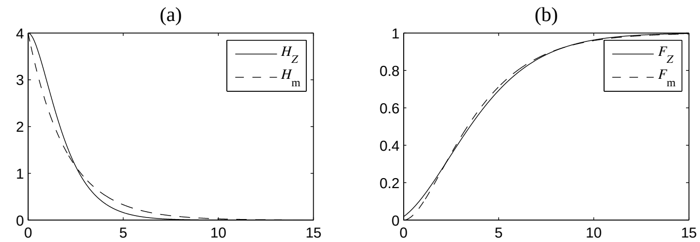

FIG 7. *Plots relating to Example 4.1:* (a) *The functions* $H_Z(x) = \lambda\theta(1 + x/\theta)\mathrm{e}^{-x/\theta}$ *and* $H_{\mathrm{m}}(x) = \lambda\theta\mathrm{e}^{-x/(2\theta)}$, $\lambda = 4$, $\theta = 1$; (b) *The DF's* F_Z *and* F_{m} *of* Z *and the Gamma*$(\lambda/2, 2\theta)$ *distribution, respectively.*

Example 4.1. When F above is the exponential distribution with mean θ, so that $\nu_Z(\mathrm{d}x) = \lambda\theta^{-1}\mathrm{e}^{-x/\theta}\,\mathrm{d}x$, $x > 0$, it holds

$$H_Z(x) = \int_x^\infty u\lambda\theta^{-1}\mathrm{e}^{-u/\theta}\,\mathrm{d}u = \lambda\theta\left(1 + \frac{x}{\theta}\right)\mathrm{e}^{-x/\theta}.$$

Here $\mu_Z = \lambda\theta$, and hence $H(x) = \lambda\theta\mathrm{e}^{-x/\beta}$. It can be checked graphically that the functions H_Z and H agree fairly well for β values around 2θ. In view of Remark 4.2, note that the MMM yields $\beta_{\mathrm{m}} = 2\theta$ (and thus $\alpha_{\mathrm{m}} = \lambda/2$), thus further confirming the proposed methodology. The quality of the Gamma$(\lambda/2, 2\theta)$ approximation can be easily assessed, by comparing the respective DF's. Following [3, p. 98], it follows from (4.12) (where $F^{*k} = \mathrm{Gamma}(k, \theta)$, $k \geq 1$) that

$$\mathrm{P}(Z \leq x) = 1 - \mathrm{e}^{-x/\theta}\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}\sum_{i=0}^{k-1}\frac{1}{i!}\left(\frac{x}{\theta}\right)^i$$

for $x > 0$ (note that $\mathrm{P}(Z = 0) = \mathrm{e}^{-\lambda}$). The DF's of Z and the Gamma$(\lambda/2, 2\theta)$ distribution are plotted in Figure 7(b) for $\lambda = 4$, $\theta = 1$ (the agreement is fairly good, taking into account the mass of Z at 0); the associated functions $H_Z(x) = \lambda\theta(1 + x/\theta)\mathrm{e}^{-x/\theta}$ and $H_{\mathrm{m}}(x) := \lambda\theta\mathrm{e}^{-x/(2\theta)}$ are plotted in Figure 7(a).

Example 4.2. When F is the uniform distribution on $(0, 2\theta)$, so that $\nu_Z(\mathrm{d}x) = \lambda(2\theta)^{-1}\mathbf{1}_{(0,2\theta)}(x)\,\mathrm{d}x$, it holds

$$H_Z(x) = \int_x^\infty u\lambda(2\theta)^{-1}\mathbf{1}_{(0,2\theta)}(u)\,\mathrm{d}u = \lambda\theta\left[1 - \left(\frac{x}{2\theta}\right)^2\right]\mathbf{1}_{[0,2\theta]}(x).$$

Here again, $H(x) = \lambda\theta\mathrm{e}^{-x/\beta}$. However, contrary to the previous example, H_Z (concave on $[0, 2\theta]$) and H (convex) do not agree well, for any $\beta > 0$.

Remark 4.3. Clearly, the assumption made above that F is absolutely continuous is not essential from a theoretical point of view. As the simplest discrete example, let $F = \delta_1$ (the delta distribution concentrated at 1), so that $Z \sim \text{Poisson}(\lambda)$; then $\nu_Z = \lambda\delta_1$, and it holds

$$\begin{aligned} H_Z(x) &= \int_{(x,\infty)} u\lambda\delta_1(\mathrm{d}u) \\ &= \lambda\mathbf{1}_{[0,1)}(x). \end{aligned} \tag{4.13}$$

Here $H(x) = \lambda \mathrm{e}^{-x/\beta}$. Not surprisingly in view of the obvious difference between H_Z and H here, (4.8), (4.9), and (4.10) lead to very different β_* values, namely, ≈ 0.596, ≈ 1.04, and ≈ 1.44, respectively. The MMM yields $\beta_{\mathrm{m}} = 1$ (and thus $\alpha_{\mathrm{m}} = \lambda$). However, this example, along with Examples 4.3 and 4.4 below (as well as the results for the sum-of-gammas case), should not suggest that (4.9) better agrees with the MMM, as indicated by the results of Section 4.3.2 below.

The point made in Remark 4.3 is further illustrated in the following example concerning the negative binomial distribution. Here, despite the fact that the associated function H_Z (denoted below by H_{Z_t}) has infinitely many jump discontinuities, the parameter β_* corresponding to (4.9) is readily obtainable numerically, and turns out to agree very well with β_{m}. While a gamma approximation to the negative binomial is well known ((4.14) below can be found e.g. in [7, p. 386]), Example 4.3 provides new insights to it.

Example 4.3. Let $0 < p < 1$, and set $q = 1 - p$, $\lambda = -\log(p)$. Suppose that $\{Z(\cdot)\}$ is a CPP with rate λ and jump distribution F on the positive integers such that $F(\{k\}) = \lambda^{-1}k^{-1}q^k$, $k \in \mathbb{N}$ (logarithmic distribution). Fix $t > 0$ (real). Then $Z(t)$ is negative binomial with parameters t and p, denoted as $Z(t) \sim \text{NB}(t,p)$, meaning that its distribution is concentrated on $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ with

$$\mathrm{P}(Z(t) = k) = \frac{(-t)(-t-1)\cdots(-t-k+1)}{k!} p^t(-q)^k$$

for $k \in \mathbb{Z}_+$; see [13, Example 4.6]. In particular, $Z(1)$ is geometric with parameter p: $\mathrm{P}(Z(1) = k) = pq^k$, $k \in \mathbb{Z}_+$. Let ν_{Z_t} denote the Lévy measure of $Z(t)$. Then, $\nu_{Z_t} = t\nu_{Z_1} = t\lambda F$, and hence ν_{Z_t} is concentrated on $\mathbb{N}$ with $\nu_{Z_t}(\{k\}) = tk^{-1}q^k$, $k \in \mathbb{N}$. Therefore,

$$\begin{aligned} H_{Z_t}(x) &:= \int_{(x,\infty)} u\nu_{Z_t}(\mathrm{d}u) \\ &= \sum_{k=\lfloor x \rfloor+1}^{\infty} ktk^{-1}q^k \\ &= \frac{tq}{p} q^{\lfloor x \rfloor}. \end{aligned}$$

Being $\text{NB}(t,p)$, the mean and variance of $Z(t)$ are given by

$$\mathrm{E}(Z(t)) = \frac{tq}{p},\ \mathrm{Var}(Z(t)) = \frac{tq}{p^2}.$$

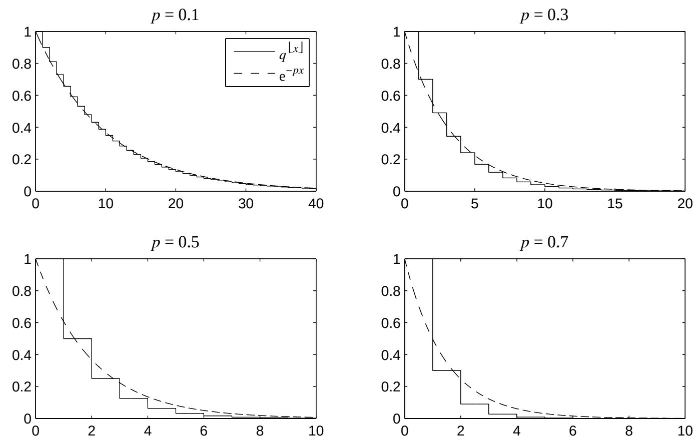

Fig 8. *Plots of* $q^{\lfloor x \rfloor}$ *and* e^{-px}, $p = 0.1, 0.3, 0.5, 0.7$, *relating to Example 4.3.*

Table 6
Tabulated rounded values of $\beta_\mathrm{m} = 1/p$ *and* β_* *(under (4.9)) relating to Example 4.3*

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β_m	10.0000	5.0000	3.3333	2.5000	2.0000	1.6667	1.4286	1.2500	1.1111
β_*	10.0023	5.0052	3.3418	2.5122	2.0164	1.6874	1.4540	1.2803	1.1464

(For any square integrable Lévy process $\{X(t) : t \geq 0\}$, it holds that $\mathrm{E}(X(t)) = t\mathrm{E}(X(1))$ and $\mathrm{Var}(X(t)) = t\mathrm{Var}(X(1))$.) By (4.11), the MMM yields

$$\alpha_\mathrm{m} = tq,\ \beta_\mathrm{m} = 1/p. \tag{4.14}$$

Thus, in view of Remark 4.2, it is instructive to compare the functions $H_{Z_t}(x) = (tq/p)q^{\lfloor x \rfloor}$ and $H_\mathrm{m}(x) := (tq/p)\mathrm{e}^{-px}$, or just $q^{\lfloor x \rfloor} = \mathrm{e}^{\log(q)\lfloor x \rfloor}$ and e^{-px}. This is done in Figure 8 for $p = 0.1, 0.3, 0.5, 0.7$. For small p, $\log(q) \approx -p$, hence the very good agreement in this case. Moreover, under (4.9), it can be obtained that β_* is the minimizer, over $\beta > 0$, of $\beta[1 - 4(1 - \mathrm{e}^{-1/\beta})/(1 - q\mathrm{e}^{-1/\beta})]$. As Table 6 shows, for a wide range of p-values, β_* is very close to $\beta_\mathrm{m} = 1/p$, and the approximation improves as p decreases. A novel heuristic justification of gamma approximation to the negative binomial (at least for small p) is thus established.

Returning to the general case of Z being an integrable ID RV on $\mathbb{R}_+$ (corresponding to (4.1) with $t = 1$), the following fact accounts for the wide applicability of the proposed methodology (at least from a theoretical point of view). Let Z_i, $i = 1, \ldots, n$, be independent integrable ID RV's on $\mathbb{R}_+$ with respective Lévy measures ν_{Z_i}. Then, $Z := \sum_{i=1}^n Z_i$ is an integrable ID RV on $\mathbb{R}_+$ with Lévy measure $\nu_Z = \sum_{i=1}^n \nu_{Z_i}$. (In particular, if the Z_i have respective Lévy densities ρ_{Z_i}, then Z has Lévy density $\rho_Z = \sum_{i=1}^n \rho_{Z_i}$.) It follows that

$$H_Z(x) = \sum_{i=1}^n H_{Z_i}(x), \tag{4.15}$$

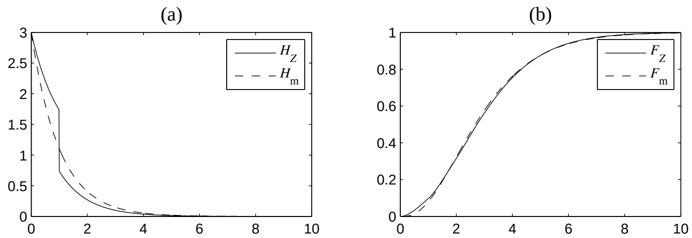

FIG 9. *Plots relating to Example 4.4:* (a) *The functions* $H_Z(x) = ab\mathrm{e}^{-x/b} + \lambda \mathbf{1}_{[0,1)}(x)$ *and* $H_{\mathrm{m}}(x) = (ab+\lambda)\mathrm{e}^{-x/\beta_{\mathrm{m}}}$, $a = 2$, $b = 1$, $\lambda = 1$; (b) *The DF's* F_Z *and* F_{m} *of* Z *and the* Gamma$(\alpha_{\mathrm{m}}, \beta_{\mathrm{m}})$ *distribution, respectively.*

where H_{Z_i} is defined according to (4.3). The key point here is that although the distribution of Z above is complicated in general, being an n-fold *convolution*, the corresponding function H_Z is just the *sum* of the respective functions H_{Z_i}.

Example 4.4. Let Z_1 and Z_2 be independent Gamma(a, b) and Poisson(λ) RV's, respectively. The DF of $Z = Z_1 + Z_2$ is given by the law of total probability as

$$\mathrm{P}(Z \le x) = \sum_{k=0}^{\lfloor x \rfloor} \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \int_0^{x-k} \frac{u^{a-1}\mathrm{e}^{-u/b}}{b^a \Gamma(a)}\,\mathrm{d}u$$

for $x \ge 0$. Here $H_{Z_1}(x) = ab\mathrm{e}^{-x/b}$, $x \ge 0$, and (by (4.13)) $H_{Z_2}(x) = \lambda \mathbf{1}_{[0,1)}(x)$. Hence, by (4.15),

$$H_Z(x) = ab\mathrm{e}^{-x/b} + \lambda \mathbf{1}_{[0,1)}(x),$$

whereas $H(x) = (ab+\lambda)\mathrm{e}^{-x/\beta}$. By (4.11), the MMM yields

$$\alpha_{\mathrm{m}} = \frac{(ab+\lambda)^2}{ab^2+\lambda},\ \beta_{\mathrm{m}} = \frac{ab^2+\lambda}{ab+\lambda}.$$

The DF's of Z and the Gamma$(\alpha_{\mathrm{m}}, \beta_{\mathrm{m}})$ distribution are plotted in Figure 9(b) for $a = 2$, $b = 1$, $\lambda = 1$. The agreement is quite good. The associated functions $H_Z(x) = ab\mathrm{e}^{-x/b} + \lambda \mathbf{1}_{[0,1)}(x)$ and $H_{\mathrm{m}}(x) := (ab+\lambda)\mathrm{e}^{-x/\beta_{\mathrm{m}}}$ are plotted in Figure 9(a), suggesting visually that, under (4.9), $\beta_* \approx \beta_{\mathrm{m}}\,(= 1)$. Indeed, with the selected parameters, it can be easily obtained that β_* is the minimizer, over $\beta > 0$, of $2\beta\mathrm{e}^{-1/\beta} - \beta/2 - 4\beta/(\beta+1)$, yielding $\beta_* \approx 1.0166$. This confirms once again the proposed methodology. [Under (4.10), on the other hand, β_* is the solution $\beta > 0$ of $H(1) = (H_Z(1-) + H_Z(1))/2$, yielding $\beta_* \approx 1.1275$.]

Example 4.5. Anticipating Section 4.3 (where all the details are presented), let Z_1 and Z_2 be independent Gamma(a, b) and GD(θ) RV's, respectively, so that $H_{Z_1}(x) = ab\mathrm{e}^{-x/b}$, $x \ge 0$, and $H_{Z_2}(x) = \theta(1-x)\mathbf{1}_{[0,1]}(x)$. Let $Z = Z_1 + Z_2$. Then, by (4.15),

$$H_Z(x) = ab\mathrm{e}^{-x/b} + \theta(1-x)\mathbf{1}_{[0,1]}(x),$$

whereas $H(x) = (ab+\theta)\mathrm{e}^{-x/\beta}$. A gamma approximation to Z is particularly appropriate here, the distribution of Z_2 being quite complicated by itself. By

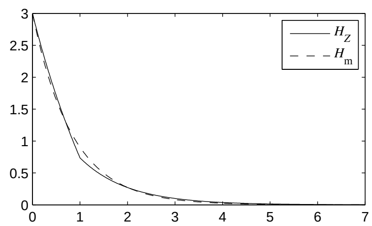

Fig 10. *Plots of* $H_Z(x) = ab\mathrm{e}^{-x/b} + \theta(1-x)\mathbf{1}_{[0,1]}(x)$ *and* $H_{\mathrm{m}}(x) = (ab+\theta)\mathrm{e}^{-x/\beta_{\mathrm{m}}}$, $a = 2$, $b = 1$, $\theta = 1$, *relating to Example 4.5.*

(4.11), the MMM yields

$$\alpha_{\mathrm{m}} = \frac{(ab+\theta)^2}{ab^2+\theta/2}, \ \beta_{\mathrm{m}} = \frac{ab^2+\theta/2}{ab+\theta}.$$

The functions $H_Z(x) = ab\mathrm{e}^{-x/b} + \theta(1-x)\mathbf{1}_{[0,1]}(x)$ and $H_{\mathrm{m}}(x) := (ab+\theta)\mathrm{e}^{-x/\beta_{\mathrm{m}}}$ are plotted in Figure 10 for $a = 2$, $b = 1$, $\theta = 1$. The good agreement confirms the proposed methodology once again.

The gamma and GD distributions share the following property: the corresponding Lévy measure is absolutely continuous with density of the form $(k(x)/x)\mathbf{1}_{(0,\infty)}(x)$, where $k(x)$, nonnegative and satisfying the integrability condition $\int_0^\infty \min(x,1)(k(x)/x)\,\mathrm{d}x < \infty$, is monotone decreasing on $(0,\infty)$ with $k(0+) > 0$ (in the Gamma(α,β) case, $k(x) = \alpha\mathrm{e}^{-x/\beta}$, $x > 0$, whereas in the GD(θ) case, $k(x) = \theta\mathbf{1}_{(0,1]}(x)$). Allowing $k(0+) = \infty$, the following statement holds: a non-delta ID distribution on $\mathbb{R}_+$ is *self-decomposable* if and only if it has Lévy measure as above (cf. [13, Corollary 15.11]). Suppose that Z is an integrable self-decomposable RV on $\mathbb{R}_+$, to be approximated by gamma. Then, the associated function H_Z is given by $H_Z(x) = \int_x^\infty k(u)\,\mathrm{d}u$. The integrability requirement on Z is equivalent to $\int_1^\infty k(x)\,\mathrm{d}x < \infty$ (a condition which is not satisfied in the α-*stable* case, $0 < \alpha < 1$, where $k(x) = bx^{-\alpha}\mathbf{1}_{(0,\infty)}(x)$ for some $b > 0$). As an example where $k(0+) = \infty$, let $k(x) = bx^{-\alpha}\mathbf{1}_{(0,1]}(x)$, $0 < \alpha < 1$, $b > 0$. Then $H_Z(x) = \mu_Z(1-x^{1-\alpha})\mathbf{1}_{[0,1]}(x)$, with $\mu_Z = b/(1-\alpha)$ (the mean of Z). Being convex on $[0,1]$, H_Z is not too far from a decreasing exponential function, i.e. from $\mu_Z\mathrm{e}^{-x/\beta}$ for suitable $\beta = \beta(\alpha) > 0$. (However, this depends largely on the value of α.) By (4.4), $\mathrm{Var}(Z) = \int_0^1 xbx^{-\alpha}\,\mathrm{d}x = b/(2-\alpha)$. The MMM then yields $\beta_{\mathrm{m}} = (1-\alpha)/(2-\alpha)$. It can be checked graphically that this typically agrees with $\beta(\alpha)$ above, and hence with the proposed methodology. Another example highlighting the point made in Remark 4.2 is that corresponding to $k(x) = a\mathrm{e}^{-x/b}\mathbf{1}_{(0,1]}(x)$, i.e. to a gamma process with the jumps greater than 1 removed. In this example, $H_Z(x) = ab(\mathrm{e}^{-x/b} - \mathrm{e}^{-1/b})\mathbf{1}_{[0,1]}(x)$ and $\beta_{\mathrm{m}} = b - (\mathrm{e}^{1/b} - 1)^{-1}$. The normalized functions $H_Z(x)/H_Z(0)$ and $H_{\mathrm{m}}(x)/H_Z(0) := \mathrm{e}^{-x/\beta_{\mathrm{m}}}$ are plotted in Figure 11 for $b = 0.1, 0.2, 0.3, 0.4$, indicating that $H_Z \approx H_{\mathrm{m}}$ for small b, with the approximation getting better and better as b decreases. This suggests that a gamma approximation to Z may be particularly appropriate here.

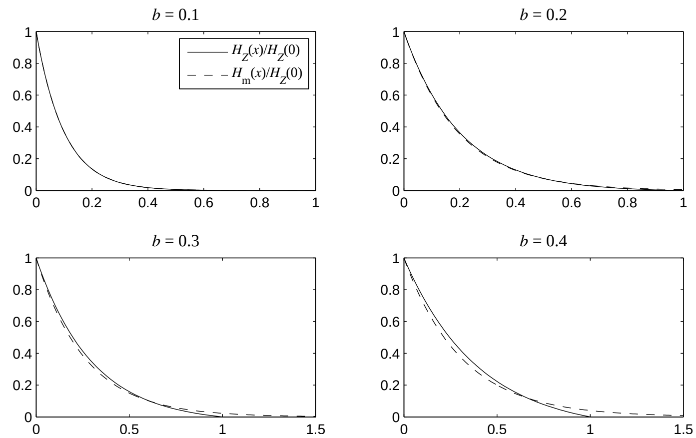

FIG 11. *Plots relating to the example above Remark 4.4.*

The last example is closely related to the paper [4]. Let $\{Z(t) : t \geq 0\}$ be a pure-jump subordinator with continuous Lévy measure ν_Z. Given $s > 0$, denote by ν_Z^s the restriction of ν_Z to $(0, s]$ and let $\{Z_s(t) : t \geq 0\}$ be the associated pure-jump subordinator. This corresponds to the original process with the jumps greater than s removed. The DF of $Z_s(t)$ is given in Theorem 2.1 of [4] (where different notation is used), in terms of ν_Z and the DF of $Z(t)$. Calculating it may be too computationally expensive, because of the multiple integrals involved. However, applying the proposed methodology to the approximation of $Z := Z_s(t)$ remains conceptually simple, since the corresponding function H_Z is given by $H_Z(x) = \int_{(x,\infty)} ut\nu_Z^s(\mathrm{d}u) = t\int_{(x,s]} u\nu_Z(\mathrm{d}u)$ (thus vanishing for $x \geq s$).

Remark 4.4. An advantage of the proposed methodology over the MMM is that the approximated distribution is not required to have finite variance. In particular, contrary to the MMM, it may be suitable for compound Poisson distributions with associated jump distribution F (on $(0, \infty)$) having infinite variance (but finite mean). An advantage of the MMM is indicated next.

Remark 4.5. Consider the example in Remark 4.3. Despite the obvious difference between H_Z and H there and the fact that the parameter λ plays no role in the minimization of $|H - H_Z|$, a gamma approximation to Z may be appropriate for large λ-values, by virtue of the central limit theorem (CLT). Specifically, this is the Gamma$(\lambda, 1)$ approximation, naturally obtained from the MMM. For the general case, suppose that $\{Z(t) : t \geq 0\}$ is a non-zero, square integrable pure jump subordinator, and let $\mu_Z = \mathrm{E}(Z(1))$, $\sigma_Z^2 = \mathrm{Var}(Z(1))$. Further, let P_{Z_t} denote the distribution of $Z(t)$. By the CLT, for t large enough,

$$P_{Z_t} \approx \mathrm{N}(t\mu_Z, t\sigma_Z^2) \approx \mathrm{Gamma}\left(t\frac{\mu_Z^2}{\sigma_Z^2}, \frac{\sigma_Z^2}{\mu_Z}\right), \tag{4.16}$$

with the gamma parameters corresponding to the MMM. Indeed, let n be a positive integer such that $t/n \approx 1$; writing $Z(t)$ as the sum $\sum_{i=1}^{n}[Z(it/n) -$

$Z((i-1)t/n)]$ of n i.i.d. RV's, each with mean $(t/n)\mu_Z$ and variance $(t/n)\sigma_Z^2$, shows that, for t large enough, $(Z(t)-t\mu_Z)/(\sqrt{t}\sigma_Z)$ is approximately standard normal, and hence the approximation $P_{Z_t} \approx \mathrm{N}(t\mu_Z, t\sigma_Z^2)$. From this follows as a special case the second approximation in (4.16); indeed, if $\{\tilde{Z}(\cdot)\}$ is a gamma process such that $\tilde{Z}(t) \sim \mathrm{Gamma}(t\mu_Z^2/\sigma_Z^2, \sigma_Z^2/\mu_Z)$, $t>0$, then $\mathrm{E}(\tilde{Z}(1)) = \mu_Z$ and $\mathrm{Var}(\tilde{Z}(1)) = \sigma_Z^2$, and hence $P_{\tilde{Z}_t} \approx \mathrm{N}(t\mu_Z, t\sigma_Z^2)$, as required.

Remark 4.5 does not reduce from the significance of the proposed methodology, as (4.16) is only designed for large t. To illustrate this point, consider the $\mathrm{NB}(t,p)$ distribution. The justification of the gamma approximation in this case (at least for small p) has been established in Example 4.3, based on the proposed methodology, independently of t. Having, in particular, the approximation $\mathrm{Gamma}(q, 1/p) \approx \mathrm{Geometric}(p)$ (corresponding to (4.14) with $t=1$), the general approximation $\mathrm{Gamma}(tq, 1/p) \approx \mathrm{NB}(t,p)$ follows naturally (by Table 6, the MMM agrees particularly well with (4.9)). The CLT argument of Remark 4.5 may only suggest that the approximation improves as t increases.

4.3. Application to the generalized Dickman (GD) distribution

4.3.1. A brief account of the GD distribution

The GD distribution has been extensively studied in the literature. Some key references are [4, 5], and [12]. This distribution (or, more specifically, the associated subordinator) appears in [4] in the context of approximation of small jumps of a gamma process (the key result there being Proposition 4.1). This issue has been thoroughly extended in the paper [5] on "approximations of small jumps of subordinators with particular emphasis on a Dickman-type limit".

For fixed $\theta > 0$, let $\{Z(t) : t \geq 0\}$ be a pure-jump subordinator, characterized by the absolutely continuous Lévy measure ν_Z with density

$$\rho_Z(x) = \frac{\theta}{x}\mathbf{1}_{(0,1]}(x). \tag{4.17}$$

Then, by (4.1),

$$\mathrm{E}\big[\mathrm{e}^{-uZ(t)}\big] = \exp\left[\theta t \int_0^1 \frac{\mathrm{e}^{-ux}-1}{x}\,\mathrm{d}x\right]$$

for $u \geq 0$. Thus, for any $t > 0$, $Z(t)$ has the generalized Dickman distribution with shape parameter θt (see e.g. [12, Proposition 3(i)]). Let $Z := Z(1)$. The RV $Z \sim \mathrm{GD}(\theta)$ satisfies the distributional equation $Z \stackrel{d}{=} U^{1/\theta}(1+Z)$, where U is uniform$(0,1)$ independent of the Z on the right, and admits the representation

$$Z = {U_1}^{1/\theta} + (U_1U_2)^{1/\theta} + (U_1U_2U_3)^{1/\theta} + \cdots, \tag{4.18}$$

where $U_1, U_2, \ldots$ are i.i.d. uniform$(0,1)$ RV's; see e.g. [12, Proposition 2]. By (4.5), Z has cumulants $\kappa_n = \int_0^1 x^n(\theta/x)\,\mathrm{d}x = \theta/n$ for every n. In particular,

$$\mathrm{E}(Z) = \theta,\ \mathrm{Var}(Z) = \theta/2. \tag{4.19}$$

(The nth moment can be obtained recursively as $\mathrm{E}(Z^n) = (\theta/n)\sum_{k=0}^{n-1}\binom{n}{k}\mathrm{E}(Z^k)$; see [12, Proposition 3(v)].) The $\mathrm{GD}(\theta)$ DF is quite complicated. Denote it by F_θ. Proposition 4.2 of [4] states that F_θ is of class $C^{\lceil\theta\rceil-1}(\mathbb{R})$, its $\lceil\theta\rceil$th derivative $F_\theta^{(\lceil\theta\rceil)}(\cdot)$ of class $C^0((0,\infty))$, and, for $j = 0, 1, \ldots, \lceil\theta\rceil$,

$$F_\theta^{(j)}(x) = \frac{\mathrm{e}^{-\gamma\theta}}{\Gamma(\theta+1-j)}\left\{x^{\theta-j} + \sum_{k=1}^{\lceil x\rceil-1}(-\theta)^k\int_{B_k(x)}\left(x - \sum_{i=1}^{k}u_i\right)^{\theta-j}\frac{\mathrm{d}u_1\cdots\mathrm{d}u_k}{u_1\cdots u_k}\right\} \tag{4.20}$$

for $x > 0$, where $\gamma \approx 0.5772156649$ is Euler's constant and

$$B_k(x) = \{\mathbf{u} \in \mathbb{R}^k : 1 < u_1 < \cdots < u_k,\ u_1 + \cdots + u_k < x\}.$$

Related results are indicated in [4, pp. 386–387]. The drawback of (4.20) for large x-values is obvious. It should be stressed in this context that, while generation of $\mathrm{GD}(\theta)$ variates based on (4.18) is straightforward, it may be too computationally expensive if θ is large (because of the truncation error involved).

Gamma approximation to the GD distribution is thus of notable importance.

4.3.2. Four approximations

Below, four gamma approximations are given for the $\mathrm{GD}(\theta)$ distribution. First, by (4.11) and (4.19), the MMM yields

$$\alpha_{\mathrm{m}} = 2\theta,\ \beta_{\mathrm{m}} = 1/2$$

as the parameters of the approximating gamma distribution. Denote the corresponding density function by f_{m}. Then,

$$f_{\mathrm{m}}(x) = \frac{2^{2\theta}}{\Gamma(2\theta)}x^{2\theta-1}\mathrm{e}^{-2x} \tag{4.21}$$

for $x > 0$. The gamma approximations corresponding to (4.8), (4.9), and (4.10) are considered next. In order to distinguish between the three approximations, denote by $\beta_{*,1}$, $\beta_{*,2}$, and $\beta_{*,\infty}$ the minimizers of $\psi_1(\beta)$, $\psi_2(\beta)$, and $\psi_\infty(\beta)$, respectively; accordingly, and in accordance with (4.6), define

$$\alpha_{*,1} = \theta/\beta_{*,1},\ \alpha_{*,2} = \theta/\beta_{*,2},\ \alpha_{*,\infty} = \theta/\beta_{*,\infty}. \tag{4.22}$$

Like β_{m}, the $\beta_{*,\bullet}$'s are universal constants, independent of θ.

Proposition 4.1. *The following hold:*

(1) $\beta_{*,1} \approx 0.4845944$. *It is the minimizer over* $0 < \beta < 1$ *of*

$$-x^2(\beta) + 2x(\beta)(1-\beta) + \beta,$$

where $x(\beta)$ *is the solution* $0 < x < 1$ *of* $\mathrm{e}^{-x/\beta} = 1 - x$.

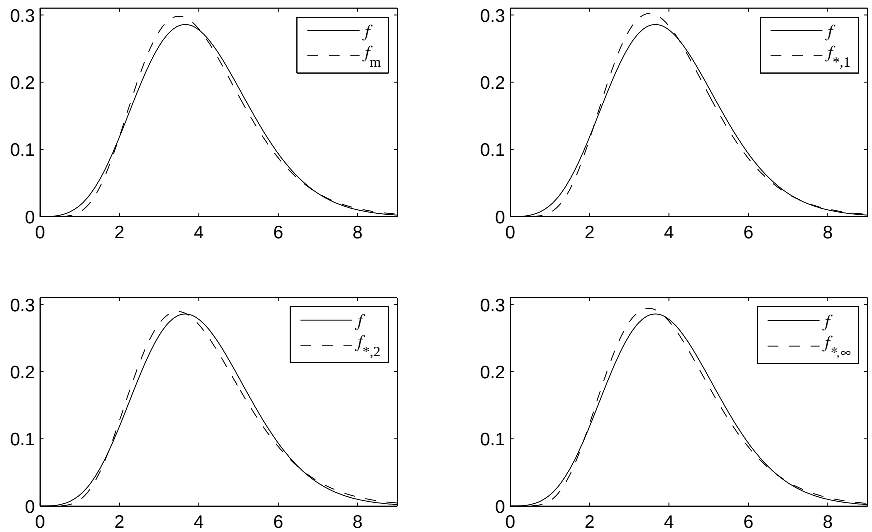

Fig 12. *Plots of the GD(4) density function, f, and the approximating gamma densities f_{m}, $f_{*,1}$, $f_{*,2}$, and $f_{*,\infty}$.*

(2) $\beta_{*,2} \approx 0.5337946$. *It is the solution $\beta > 0$ of the equation*

$$\mathrm{e}^{-1/\beta} = \frac{8\beta - 3}{8\beta + 4}.$$

(3) $\beta_{*,\infty} \approx 0.5148331$. *It is the solution $0 < \beta < 1$ of the equation*

$$\mathrm{e}^{-1/\beta} = 1 + \beta \log(\beta) - \beta.$$

The proof of Proposition 4.1 is given in the appendix.

Now, let

$$f_{*,\bullet}(x) = \frac{(\beta_{*,\bullet}^{-1})^{\alpha_{*,\bullet}}}{\Gamma(\alpha_{*,\bullet})} x^{\alpha_{*,\bullet}-1} \mathrm{e}^{-x/\beta_{*,\bullet}}, \tag{4.23}$$

for $x > 0$, be the density functions of the approximating gamma distributions, where $\bullet$ stands for 1, 2, or ∞. Since, by Proposition 4.1, the $\beta_{*,\bullet}$'s are close to $1/2$ (and hence, by (4.22), the $\alpha_{*,\bullet}$'s are close to 2θ), the $f_{*,\bullet}$'s are close to the gamma density f_{m} in (4.21). The $f_{*,\bullet}$'s and f_{m} are plotted in Figure 12 for $\theta = 4$ against the GD(θ) density function, f, as calculated from (4.20) with $j = 1$. The respective DF's, denoted by $F_{*,\bullet}$ and F_{m}, are plotted in Figure 13 against the GD(θ) DF, F, as calculated from (4.20) with $j = 0$. The approximations are quite good, taking into account the complexity of the GD distribution, on the one hand, and the simplicity of the gamma distribution, on the other. As usual, it is instructive to compare the associated functions H_Z and H. Without loss of generality, suppose that $\theta = 1$; then these functions are given by $H_Z(x) = (1 - x)\mathbf{1}_{[0,1]}(x)$ and $H(x) = \mathrm{e}^{-x/\beta}$. Because of the linearity of the former on

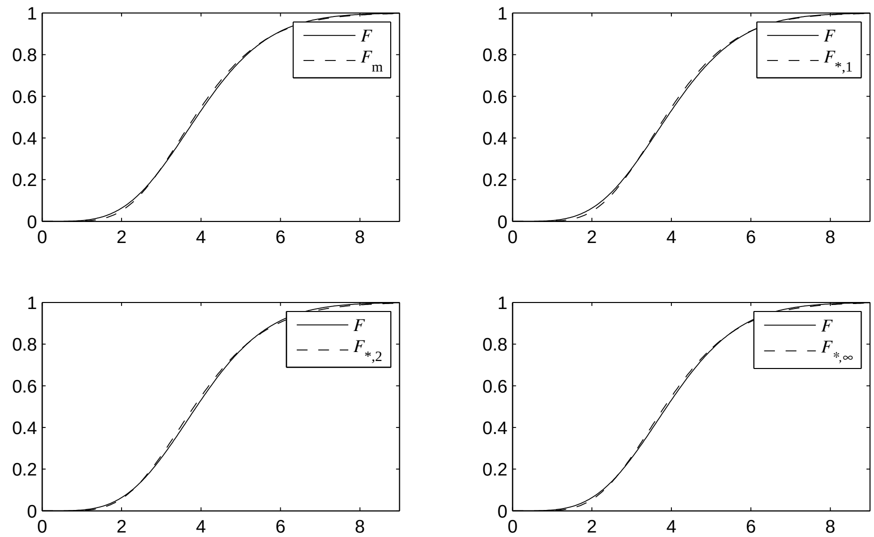

FIG 13. *Plots of the GD(4) DF, F, and the approximating gamma DF's* $F_{\rm m}$, $F_{*,1}$, $F_{*,2}$, *and* $F_{*,\infty}$.

$[0,1]$, the two functions do not agree well; yet, they are not too far for β values around $1/2$, in agreement with the four approximations above.

Appendix A: Proofs

Proof of Theorem 2.1. If all the β_i are equal, then clearly, by (2.8), β_* is equal to their common value; this value is indeed the solution $\beta > 0$ of (2.9), and the "less than or equal to" inequalities in (2.10) and (2.11) are trivially satisfied with equality (so, in particular, $\alpha_{\min} < \alpha_*$). So we suppose that $\beta_{\min} < \beta_{\max}$, and, in addition to the first statement of Theorem 2.1, need to show that (2.10) and (2.11) hold with strict inequalities. It is readily checked that

$$\psi(\beta) = \mu^2 \frac{\beta}{2} - 2\mu \sum_{i=1}^{n} \alpha_i \beta_i \frac{\beta_i \beta}{\beta_i + \beta} + \int_0^\infty \left[\sum_{i=1}^{n} \alpha_i \beta_i \mathrm{e}^{-x/\beta_i} \right]^2 \mathrm{d}x.$$

Noting that the integral term on the right is independent of β, it thus suffices to consider the minimum (over $\beta > 0$) of the function φ given by

$$\varphi(\beta) = \frac{\mu}{2}\beta - 2\sum_{i=1}^{n} \frac{\alpha_i \beta_i^2 \beta}{\beta_i + \beta}.$$

The derivative of φ is given by

$$\varphi'(\beta) = \frac{\mu}{2} - 2\sum_{i=1}^{n} \frac{\alpha_i \beta_i^3}{(\beta_i + \beta)^2}, \tag{A.1}$$

which is the left-hand side of (2.9). Since $\varphi'(\beta)$ is strictly increasing for $\beta > 0$, with $\lim_{\beta\downarrow 0}\varphi'(\beta) < 0$ and $\lim_{\beta\to\infty}\varphi'(\beta) > 0$, it thus follows that β_* is the solution $\beta > 0$ of (2.9). As is readily seen, it holds

$$\varphi'(\beta_{\min}) < 0 < \varphi'(\beta_{\max}),$$

and hence

$$\beta_{\min} < \beta_* < \beta_{\max}.$$

On the other hand, the equation $\varphi'(\beta_*) = 0$ can be brought into the form

$$\frac{1}{4} = \sum_{i=1}^{n} \frac{\alpha_i\beta_i}{\mu}\frac{1}{\left(1+\frac{\beta_*}{\beta_i}\right)^2}. \tag{A.2}$$

The function $g(x) = 1/(1+x)^2$ is strictly convex on $(0,\infty)$ (indeed, it holds $g''(x) > 0$), $\sum_{i=1}^{n}(\alpha_i\beta_i/\mu) = 1$, and, under the assumption $\beta_{\min} < \beta_{\max}$, the β_*/β_i are not all equal. Hence, by Jensen's inequality,

$$\begin{aligned}\sum_{i=1}^{n} \frac{\alpha_i\beta_i}{\mu}\frac{1}{\left(1+\frac{\beta_*}{\beta_i}\right)^2} &> \frac{1}{\left(1+\sum_{i=1}^{n}\frac{\alpha_i\beta_i}{\mu}\frac{\beta_*}{\beta_i}\right)^2}\\ &= \frac{1}{\left(1+\beta_*\frac{\sum_{i=1}^{n}\alpha_i}{\mu}\right)^2}.\end{aligned}$$

It then follows from (A.2) that $\beta_* > \mu/\sum_{i=1}^{n}\alpha_i$, and so

$$\beta_{\min} < \frac{\mu}{\sum_{i=1}^{n}\alpha_i} < \beta_* < \beta_{\max}.$$

Using (2.7), it follows from the second and third inequalities above that

$$\alpha_{\min} < \alpha_* < \sum_{i=1}^{n}\alpha_i.$$

The theorem is thus proved. □

Proof of Proposition 2.1. Since $\varphi'(\beta)$ in (A.1) is strictly increasing for $\beta > 0$, with $\varphi'(\beta_*) = 0$, the proposition will follow by showing that $\varphi'(\beta_{\rm m}) \geq 0$, with strict inequality unless all the β_i are equal. Using (1.4), the inequality $\varphi'(\beta_{\rm m}) \geq 0$ can be brought into the form

$$\sum_{i=1}^{n}\frac{\alpha_i\beta_i}{\mu}\left(\frac{\beta_i}{\beta_i+\sum_{i=1}^{n}\frac{\alpha_i\beta_i}{\mu}\beta_i}\right)^2 \leq \frac{1}{4}.$$

Noting that $\sum_{i=1}^{n}(\alpha_i\beta_i/\mu) = 1$, the last inequality can be written as

$$\mathrm{E}\left[\frac{Y}{Y+\mathrm{E}(Y)}\right]^2 \leq \frac{1}{4},$$

where Y is a discrete RV with $\mathrm{P}(Y = \beta_i) = \sum_{j,\,\beta_j=\beta_i}(\alpha_j\beta_j/\mu)$. The proposition then follows from the following general lemma, which is interesting in itself. □

Lemma A.1. *Let Y be a nonnegative random variable with finite positive expectation. Then,*

$$\mathrm{E}\left[\frac{Y}{Y+\mathrm{E}(Y)}\right]^2 \leq \frac{1}{4}.$$

The inequality is strict unless Y is a constant with probability 1.

Proof of Lemma A.1. Without loss of generality, we can assume that $\mathrm{E}(Y) = 1$. Let F denote the distribution of Y. Note that $\int_{[0,\infty)} yF(\mathrm{d}y) = 1$ and that $\mathrm{P}(Y > 1/2) > 0$. Further, let $\tilde{F}$ denote the distribution on $(1/2, \infty)$ given by

$$\tilde{F}(\mathrm{d}y) = \frac{F(\mathrm{d}y)}{\mathrm{P}\big(Y > \frac{1}{2}\big)}.$$

Using that $y/(y+1)^2 \leq 2/9$ for $y \in [0, 1/2]$, we then have

$$\begin{aligned}
\mathrm{E}\left[\frac{Y}{Y+1}\right]^2 &= \int_{[0,\frac{1}{2}]} \left(\frac{y}{y+1}\right)^2 F(\mathrm{d}y) + \int_{(\frac{1}{2},\infty)} \left(\frac{y}{y+1}\right)^2 F(\mathrm{d}y) \\
&\leq \frac{2}{9} \int_{[0,\frac{1}{2}]} yF(\mathrm{d}y) + \mathrm{P}\Big(Y > \frac{1}{2}\Big) \int_{(\frac{1}{2},\infty)} \left(\frac{y}{y+1}\right)^2 \tilde{F}(\mathrm{d}y) \\
&= \frac{2}{9}\left[1 - \int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)\right] + \mathrm{P}\Big(Y > \frac{1}{2}\Big) \int_{(\frac{1}{2},\infty)} \frac{1}{\big(1+\frac{1}{y}\big)^2} \tilde{F}(\mathrm{d}y).
\end{aligned}$$

The function $g(y) = 1/(1 + 1/y)^2$ is strictly concave on $(1/2, \infty)$ (indeed, it holds $g''(y) < 0$ for $y > 1/2$) and the mean, $\tilde{m}$, of $\tilde{F}$ is given by

$$\tilde{m} = \frac{\int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)}{\mathrm{P}\big(Y > \frac{1}{2}\big)}.$$

Hence, by Jensen's inequality,

$$\int_{(\frac{1}{2},\infty)} \frac{1}{\big(1+\frac{1}{y}\big)^2} \tilde{F}(\mathrm{d}y) \leq \frac{1}{\left(1 + \frac{\mathrm{P}(Y>\frac{1}{2})}{\int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)}\right)^2},$$

with strict inequality unless $\tilde{F}$ is concentrated at one point. Thus,

$$\mathrm{E}\left[\frac{Y}{Y+1}\right]^2 \leq \frac{2}{9}\left[1 - \int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)\right] + \frac{\mathrm{P}\big(Y > \frac{1}{2}\big)}{\left(1 + \frac{\mathrm{P}(Y>\frac{1}{2})}{\int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)}\right)^2}, \tag{A.3}$$

with strict inequality if $\tilde{F}$ is not concentrated at one point. Define

$$H(x, a) = \frac{2}{9}(1 - a) + \frac{x}{\big(1 + \frac{x}{a}\big)^2}$$

for $x, a \in (0,1]$, where x and a play the role of $\mathrm{P}(Y > 1/2)$ and $\int_{(\frac{1}{2},\infty)} yF(\mathrm{d}y)$ in (A.3), respectively. For fixed a, H attains its maximum at $x = a$, yielding

$$H(x,a) \le a\left(\frac{1}{4} - \frac{2}{9}\right) + \frac{2}{9}.$$

Hence $H(x,a) \le 1/4$, with strict inequality unless $x = a = 1$. Now, the condition $\mathrm{P}(Y > 1/2) = 1$ implies that $\tilde{F} = F$. So by (A.3) and the line that follows it,

$$\mathrm{E}\left[\frac{Y}{Y+1}\right]^2 \le \frac{1}{4},$$

with strict inequality unless $Y = 1$ with probability 1. □

Proof of Corollary 2.1. Apply Proposition 2.1, noting that $\mathrm{Var}(X_*) = (\alpha_*\beta_*)\beta_* = \mu\beta_*$ and $\mathrm{Var}(S) = (\alpha_\mathrm{m}\beta_\mathrm{m})\beta_\mathrm{m} = \mu\beta_\mathrm{m}$. □

Proof of Proposition 4.1. For $Z \sim \mathrm{GD}(\theta)$, the associated functions H and H_Z are given by $H(x) = \theta \mathrm{e}^{-x/\beta}$ and (by (4.17))

$$\begin{aligned} H_Z(x) &= \int_x^\infty u(\theta/u)\mathbf{1}_{(0,1]}(u)\,\mathrm{d}u \\ &= \theta(1-x)\mathbf{1}_{[0,1]}(x). \end{aligned}$$

To prove (1), assume first that $0 < \beta < 1$ and note that, thus,

$$\begin{aligned} d_1 &:= \frac{1}{\theta}\int_0^\infty \left|H(x) - H_Z(x)\right|\mathrm{d}x \\ &= \int_0^{x(\beta)} \left[(1-x) - \mathrm{e}^{-x/\beta}\right]\mathrm{d}x + \int_{x(\beta)}^1 \left[\mathrm{e}^{-x/\beta} - (1-x)\right]\mathrm{d}x + \int_1^\infty \mathrm{e}^{-x/\beta}\,\mathrm{d}x \\ &= 2x(\beta) - x^2(\beta) - \beta + 2\beta\mathrm{e}^{-x(\beta)/\beta} - 1/2 \\ &= 2x(\beta) - x^2(\beta) - \beta + 2\beta(1 - x(\beta)) - 1/2 \\ &= -x^2(\beta) + 2x(\beta)(1-\beta) + \beta - 1/2, \end{aligned}$$

where $x(\beta)$ is the solution $0 < x < 1$ of $\mathrm{e}^{-x/\beta} = 1 - x$. Let

$$\varphi_1(\beta) = -x^2(\beta) + 2x(\beta)(1-\beta) + \beta.$$

Since $\lim_{\beta\downarrow 0} x(\beta) = 1$ and $\lim_{\beta\uparrow 1} x(\beta) = 0$, it holds $\lim_{\beta\downarrow 0}\varphi_1(\beta) = 1$ and $\lim_{\beta\uparrow 1}\varphi_1(\beta) = 1$. Numerical results show that $\varphi_1(\beta)$, $0 < \beta < 1$, achieves its minimum near $\tilde{\beta} = 0.4845944$, with $\varphi_1(\tilde{\beta}) \approx 0.6614857$. On the other hand, if $\beta \ge 1$, then $\mathrm{e}^{-x/\beta} > (1-x)\mathbf{1}_{[0,1]}(x)$ for all $x > 0$. In this case, $d_1 + 1/2$ is equal to β, and in particular is greater than $\varphi_1(\tilde{\beta})$. Thus, (1) is established.

To prove (2), first note that, for any $\beta > 0$,

$$\begin{aligned} \frac{1}{\theta^2}\int_0^\infty \left[H(x) - H_Z(x)\right]^2\mathrm{d}x &= \int_0^\infty \left[\mathrm{e}^{-x/\beta} - (1-x)\mathbf{1}_{[0,1]}(x)\right]^2\mathrm{d}x \\ &= \beta/2 - 2\int_0^1 \mathrm{e}^{-x/\beta}(1-x)\,\mathrm{d}x + 1/3 \\ &= -3\beta/2 + 2\beta^2(1 - \mathrm{e}^{-1/\beta}) + 1/3. \end{aligned}$$

Let

$$\varphi_2(\beta) = -3\beta/2 + 2\beta^2(1 - \mathrm{e}^{-1/\beta}).$$

Then,

$$\varphi_2'(\beta) = -3/2 + 4\beta(1 - \mathrm{e}^{-1/\beta}) - 2\mathrm{e}^{-1/\beta}. \tag{A.4}$$

It holds $\lim_{\beta\downarrow 0}\varphi_2'(\beta) = -3/2$ and $\lim_{\beta\to\infty}\varphi_2'(\beta) = 1/2$. Further, $\varphi_2'(\beta)$ is strictly increasing for $\beta > 0$. Indeed,

$$\begin{aligned}\varphi_2''(\beta) &= 4\left[1 - \mathrm{e}^{-1/\beta}\sum_{k=0}^{2}\frac{(1/\beta)^k}{k!}\right] \\ &> 4\left[1 - \mathrm{e}^{-1/\beta}\sum_{k=0}^{\infty}\frac{(1/\beta)^k}{k!}\right] \\ &= 0.\end{aligned}$$

It follows that φ_2 has a unique global minimum at the solution of $\varphi_2'(\beta) = 0$. Thus $\varphi_2'(\beta_{*,2}) = 0$, and so (A.4) yields statement (2) of the proposition (the analytical solution is readily evaluated numerically).

To prove (3), assume first that $0 < \beta < 1$ and note that, thus,

$$\begin{aligned}d_\infty &:= \frac{1}{\theta}\max_{x>0}\left|H(x) - H_Z(x)\right| \\ &= \max_{x>0}\left|\mathrm{e}^{-x/\beta} - (1-x)\mathbf{1}_{[0,1]}(x)\right| \\ &= \max\Big(\max_{0<x<x(\beta)}\big[(1-x) - \mathrm{e}^{-x/\beta}\big], \max_{x(\beta)<x\le 1}\big[\mathrm{e}^{-x/\beta} - (1-x)\big]\Big) \\ &= \max\Big(\max_{0<x<x(\beta)}\big[(1-x) - \mathrm{e}^{-x/\beta}\big], \mathrm{e}^{-1/\beta}\Big),\end{aligned}$$

where, as in (1), $x(\beta)$ is the solution $0 < x < 1$ of $\mathrm{e}^{-x/\beta} = 1 - x$. Let

$$\vartheta_\beta(x) = (1-x) - \mathrm{e}^{-x/\beta},$$

for $0 < x < x(\beta)$. Noting that $\vartheta_\beta'(x) = 0 \Leftrightarrow x = -\beta\log(\beta)$, it follows that

$$\max_{0<x<x(\beta)}\vartheta_\beta(x) = 1 + \beta\log(\beta) - \beta.$$

Thus,

$$d_\infty = \max\big(1 + \beta\log(\beta) - \beta, \mathrm{e}^{-1/\beta}\big).$$

It is readily verified that the right-hand side has a unique global minimum at the solution ($0 < \beta < 1$) of $1 + \beta\log(\beta) - \beta = \mathrm{e}^{-1/\beta}$, where its value is (obviously) smaller than e^{-1}. On the other hand, if $\beta \ge 1$, then $\theta^{-1}\left|H(1) - H_Z(1)\right| = \mathrm{e}^{-1/\beta} \ge \mathrm{e}^{-1}$. Thus, (3) is established. □

Acknowledgements

Thanks to the referee for some valuable suggestions to improve the presentation of the paper.

References

[1] Babich, F. and Lombardi, G., Statistical analysis and characterization of the indoor propagation channel. *IEEE Trans. Commun.* **48** (2000) 455–464.

[2] Bondesson, L., On the Lévy measure of the lognormal and the log-Cauchy distributions. *Methodol. Comput. Appl. Probab.* **4** (2002) 243–256. MR1965407

[3] Covo, S., The moments of a compound Poisson process with exponential or centered normal jumps. *J. Probab. Stat. Sci.* **7** (2009) 91–100. MR2743112

[4] Covo, S., One-dimensional distributions of subordinators with upper truncated Lévy measure, and applications. *Adv. Appl. Prob.* **41** (2009) 367–392. MR2541182

[5] Covo, S., On approximations of small jumps of subordinators with particular emphasis on a Dickman-type limit. *J. Appl. Prob.* **46** (2009) 732–755. MR2562319

[6] Efthymoglou, G. and Aalo, V., Performance of RAKE receivers in Nakagami fading channel with arbitrary fading parameters. *Electron. Lett.* **31** (1995) 1610–1612.

[7] Guenther, W.C., A simple approximation to the negative binomial (and regular binomial). *Technometrics* **14** (1972) 385–389. MR0383608

[8] Heath, Jr., R.W., Wu, T., Kwon, Y.H. and Soong, A.C.K., Multiuser MIMO in distributed antenna systems with out-of-cell interference. *IEEE Trans. Signal Process.* **59** (2011) 4885–4899. MR2882990

[9] Jensen, D.R. and Solomon, H., A Gaussian approximation to the distribution of a definite quadratic form. *J. Amer. Statist. Assoc.* **67** (1972) 898–902.

[10] Moschopoulos, P.G., The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* **37** (1985) 541–544. MR0818052

[11] Nadarajah, S., A review of results on sums of random variables. *Acta Appl. Math.* **103** (2008) 131–140. MR2425609

[12] Penrose, M.D. and Wade, A.R., Random minimal directed spanning trees and Dickman-type distributions. *Adv. Appl. Prob.* **36** (2004) 691–714. MR2079909

[13] Sato, K.I., *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge (1999). MR1739520

[14] Sim, C.H., Point processes with correlated gamma interarrival times. *Statist. Probab. Lett.* **15** (1992) 135–141. MR1219284

[15] Stewart, T., Strijbosch, L., Moors, J. and Batenburg, P.V., A simple approximation to the convolution of gamma distributions (revision of DP 2006-27). Discussion Paper 2007-70, Tilburg University, Center for Economic Research (2007).

[16] Tankov, P., Lévy processes in finance: inverse problems and dependence modelling. PhD thesis, École Polytechnique (2004).

[17] Yu, Y., Relative log-concavity and a pair of triangle inequalities. *Bernoulli* **16** (2010) 459–470. MR2668910