Approximation of Ruin Probabilities via Erlangized Scale Mixtures

Oscar Peralta[1], Leonardo Rojas-Nandayapa[2], Wangyue Xie[3] and Hui Yao[3]

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark, osgu@dtu.dk
[2]Mathematical Sciences, University of Liverpool, UK, leorojas@liverpool.ac.uk
[3]School of Mathematics and Physics, The University of Queensland, Australia, w.xie1@uq.edu.au, h.yao@uq.edu.au

Abstract

In this paper, we extend an existing scheme for numerically calculating the probability of ruin of a classical Cramér–Lundberg reserve process having absolutely continuous but otherwise general claim size distributions. We employ a dense class of distributions that we denominate *Erlangized scale mixtures* (ESM) that correspond to nonnegative and absolutely continuous distributions which can be written as a Mellin–Stieltjes convolution $\Pi \star G$ of a nonnegative distribution Π with an Erlang distribution G. A distinctive feature of such a class is that it contains heavy-tailed distributions.

We suggest a simple methodology for constructing a sequence of distributions having the form $\Pi \star G$ with the purpose of approximating the integrated tail distribution of the claim sizes. Then we adapt a recent result which delivers an explicit expression for the probability of ruin in the case that the claim size distribution is modelled as an Erlangized scale mixture. We provide simplified expressions for the approximation of the probability of ruin and construct explicit bounds for the error of approximation. We complement our results with a classical example where the claim sizes are heavy-tailed.

Keywords: phase-type; Erlang; scale mixtures; infinite mixtures; heavy-tailed; ruin probability.

1 Introduction

In this paper, we propose a new numerical scheme for the approximation of ruin probabilities in the classical compound Poisson risk model — also known as Cramér–Lundberg risk model (cf. Asmussen and Albrecher, 2010). In such a risk model, the surplus process is modelled as a compound Poisson process with negative linear drift and a nonnegative jump distribution F, the later corresponding to the claim size distribution. The ruin probability within infinite horizon and initial capital u, denoted $\psi(u)$, is the probability that the supremum of the surplus process is larger than u. The Pollaczek–Khinchine formula (see Equation (3.1)) provides the exact value of $\psi(u)$, though it can be explicitly computed in very few cases. Such a formula is a functional of $\widehat{F}$, the integrated tail distribution of F; from here on, we will use $\psi_{\widehat{F}}(u)$ instead of $\psi(u)$ to denote this dependence. A useful fact is that the Pollaczek–Khinchine formula can be naturally extended in order to define $\psi_G(u)$ even if G does not correspond to an integrated tail distribution; in this case, $\psi_G(\cdot)$ corresponds to the survival probability of certain terminating renewal process.

The approach advocated in this paper is to approximate the integrated claim size distribution $\widehat{F}$ by using the family of *phase-type scale mixture distributions* introduced in Bladt et al. (2015), but we also consider the more common approach of approximating the claim size distribution F. The family of phase-type scale mixture distributions is dense within the class of nonnegative distributions, and it is formed by distributions which can be expressed as a Mellin–Stieltjes convolution, denoted $\Pi \star G$, of an arbitrary nonnegative distribution Π and a phase-type distribution G (cf. Bingham et al., 1987). The Mellin–Stieltjes convolution corresponds to the distribution of the product between two independent random variables having distributions Π and G, respectively.

In particular, if Π is a nonnegative discrete distribution and $\Pi \star G$ is itself the integrated tail of a phase-type scale mixture distribution, then an explicit computable formula for the ruin probability $\psi_{\Pi\star G}(u)$ of the Cramér–Lundberg process with claims having integrated tail distribution $\Pi \star G$ is given in Bladt et al. (2015). Hence, it is plausible that if $\Pi \star G$ is *close enough* to the integrated tail distribution $\widehat{F}$ of the claim sizes, then we can use $\psi_{\Pi\star G}(u)$ as an approximation for $\psi_{\widehat{F}}(u)$, the ruin probability of a Cramér–Lundberg process having claim size distribution F. One of the key features of the class of phase-type scale mixtures is that if Π has unbounded support, then $\Pi \star G$ is a heavy-tailed distribution (Rojas-Nandayapa and Xie, 2015; Su and Chen, 2006; Tang, 2008), confirming the hypothesis that the class of phase-type scale mixtures is more appropriate for approximating tail-dependent quantities involving heavy-tailed distributions. In contrast, the class of classical phase-type distributions is light-tailed and approximations derived from this approach may be inaccurate in the tails (see also Vatamidou et al., 2014, for an extended discussion).

Our contribution is to propose a systematic methodology to approximate any absolutely continuous integrated tail distribution $\widehat{F}$ using a particular subclass of phase-type scale mixtures called *Erlangized scale mixtures* (ESM). The proposed approximation is particularly precise in the tails and the number of parameters remains controlled. Our construction requires a sequence $\{\Pi_m : m \in \mathbb{N}\}$ of nonnegative discrete distributions having the property $\Pi_m \to \widehat{F}$ (often taken as a discretization of the target distribution over some countable subset of the support of $\widehat{F}$), and a sequence of Erlang distributions with equal shape and rate parameters, denoted $G_m \sim \text{Erlang}(\xi(m), \xi(m))$. If the sequence $\xi(m) \in \mathbb{N}$ is unbounded, then $\Pi_m \star G_m \to \widehat{F}$. We adapt the results in Bladt et al. (2015) to compute $\psi_{\Pi_m\star G_m}(u)$, and use this as an approximation of the ruin probability of interest.

To assess the quality of $\psi_{\Pi_m\star G_m}(u)$ as an approximation of $\psi_{\widehat{F}}(u)$ we identify two sources of theoretical error. The first source of error comes from approximating $\widehat{F}$ via Π_m, so we refer to this as the *discretization error*. The second source of error is due to the Mellin-Stieltjes convolution with G_m, so this will be called the *Erlangization error*. The two errors are closely intertwined so it is difficult to make a precise assessment of the effect of each of them in the general approximation. Instead, we use the triangle inequality to separate these as follows

$$\underbrace{\left|\psi_{\widehat{F}}(u) - \psi_{\Pi_m\star G_m}(u)\right|}_{\text{Approximation error}} \leq \underbrace{\left|\psi_{\widehat{F}}(u) - \psi_{\widehat{F}\star G_m}(u)\right|}_{\text{Erlangization error}} + \underbrace{\left|\psi_{\widehat{F}\star G_m}(u) - \psi_{\Pi_m\star G_m}(u)\right|}_{\text{Discretization error}}.$$

Therefore, the error of approximating $\psi_{\widehat{F}}(u)$ with $\psi_{\Pi_m\star G_m}(u)$ can be bounded from above with the aggregation of the Erlangization error and the discretization error. In our developments below, we construct explicit tight bounds for each source of error.

We remark that the general formula for $\psi_{\Pi\star G}(u)$ in Bladt et al. (2015) is computational intensive and can be difficult or even infeasible to implement since it is given as an infinite series with terms involving products of finite dimensional matrices. We show that for our particular model, $\psi_{\Pi\star G_m}(u)$ can be simplified down to a manageable formula involving the probability density function (pdf) and cumulative distribution function (cdf) of the negative binomial distribution instead of computationally expensive matrix operations. In practice, the infinite series can be computed only up to a finite number of terms, but as we will show, this numerical error can be controlled by selecting an appropriate distribution Π. This truncated approximation of $\psi_{\Pi\star G}(u)$ will be denoted $\widetilde{\psi}_{\Pi\star G}(u)$. We provide explicit bounds for the numerical error induced by truncating the infinite series.

All things considered, we contribute to the existing literature for computing ruin probabilities for the classical Cramér–Lundberg model by proposing a new practical numerical scheme. Our method, coupled with the bounds for the error of approximation, provides an attractive alternative for computing ruin probabilities based on a simple, yet effective idea.

The approach described above is a further extension to the use of phase-type distributions for approximating general claim size distributions (cf. Asmussen, 2003; Latouche and Ramaswami, 1999; Neuts, 1975). Several attempts to approximate the probability of ruin for Cramér–Lundberg model have been made before (see Vatamidou et al. (2013) and references therein). A recent and similar approach can be found in Santana et al. (2016) which uses discretization and Erlangizations argument as its backbone. We emphasize here that we address the problem of finding the probability of ruin differently. Firstly, we propose to directly approximate the integrated tail distribution

instead of the claim size distribution. This will yield far more accurate approximations of the probability of ruin. Secondly, since we investigate the Erlangization and the discretization errors separately, we are able to provide tight error bounds for our approximation method. This will prove to be helpful in challenging examples such as the one presented here: the heavy-traffic Cramér–Lundberg model with Pareto distributed claims. Lastly, each approximation of ours is based on a mixture of Erlang distributions of fixed order, while the approach in Santana et al. (2016) is based on a mixture of Erlang distributions of increasing order. By keeping the order of the Erlang distribution in the mixture fixed, we can smartly allocate more computational resources in the discretization part, yielding an overall better approximation. More importantly, we find the use of ESM more natural because increasing the order of the Erlang distributions in the mixture translates in having different levels of accuracy of Erlangization at different points. The choice of having sharper Erlangization in the tail of the distribution than in the body seems arbitrary and is actually not useful tail-wise because the tail behavior of $\Pi * G_m$ is the same for each $\xi(m) \geq 1$.

The rest of the paper is organized as follows. Section 2 provides an overview of the main concepts and methods. In Section 3, we present the methodology for constructing a sequence of distributions of the form $\Pi_m \star G_m$ approximating a nonnegative continuous distribution. Based on the results of Bladt et al. (2015), we introduce two simplified infinite series representations of the ruin probability. In Section 4, we construct the bounds for the error of each approximation. In Section 5, we provide a bound for the numerical errors of approximation induced by truncating the infinite series representation. A numerical example illustrating the sharpness of our results is given in Section 6.2. Some conclusions are drawn in Section 7.

2 Preliminaries

In this section, we provide a summary of basic concepts needed for this paper. In Subsection 2.1 we introduce the family of classical phase-type (PH) distributions and their extensions to phase-type scale mixtures and infinite dimensional phase-type (IDPH) distributions. We will refer to the former class of distributions as *classical* in order to make a clear distinction from the two later classes of distributions. In Subsection 2.2, we introduce a systematic method for approximating nonnegative distributions within the class of phase-type scale mixtures; such a method will be called approximation via *Erlangized scale mixtures* (ESM). The resulting approximating distribution will be considerably tractable due to the special structure of the Erlang distribution.

2.1 Phase-type scale mixtures

A phase-type (PH) distribution corresponds to the distribution of the absorption time of a Markov jump process $\{X_t\}_{t\geq 0}$ with a finite state space $E = \{0, 1, 2, \cdots, p\}$. The states $\{1, 2, \cdots, p\}$ are transient while the state 0 is an absorbing state. Phase-type distributions are characterized by a p-dimensional row vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$, corresponding to the initial probabilities of each of the transient states of the Markov jump process, and an intensity matrix

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

The subintensity matrix $\boldsymbol{\Lambda}$ corresponds to the transition rates among the transient states while the column vector $\boldsymbol{\lambda}$ corresponds to the exit probabilities to the absorption state. Since $\boldsymbol{\lambda} = -\boldsymbol{\Lambda}\mathbf{e}$, where $\boldsymbol{e}$ is a column vector with all elements to be 1, then the pair $(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ completely characterizes the absorption distribution; the notation $\mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ is reserved for such a distribution. The cdf, pdf and expectation of $G \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ are given by the following closed-form expressions:

$$G(y) = 1 - \boldsymbol{\beta}\mathrm{e}^{\boldsymbol{\Lambda} y}\boldsymbol{e}, \quad G'(y) = g(y) = \boldsymbol{\beta}\mathrm{e}^{\boldsymbol{\Lambda} y}\boldsymbol{\lambda}, \quad \int_0^\infty y \mathrm{d}G(y) = -\boldsymbol{\beta}\boldsymbol{\Lambda}^{-1}\boldsymbol{e}.$$

A particular example of PH distribution which is of interest in our later developments is that of an Erlang distribution. It is simple to deduce that the Erlang distribution with parameters (λ, m) has a PH-representation given by the the m-dimensional vector $\boldsymbol{\beta} = (1, 0, \cdots, 0)$ and the $m \times m$

dimensional matrix

$$\mathbf{\Lambda} = \begin{pmatrix} -\lambda & \lambda & & \\ & \ddots & \ddots & \\ & & -\lambda & \lambda \\ & & & -\lambda \end{pmatrix}.$$

We employ the notation Erlang(λ, m). In this paper, we will be particularly interested in the sequence of $G_m \sim \text{Erlang}(\xi(m), \xi(m))$ distributions with $\xi(m) \to \infty$ as $m \to \infty$. These sequences are associated to a methodology often known as Erlangization (approximation of a constant via Erlang random variables). Using Chebyshev inequality, it is simple to prove that $G_m \to \mathbb{I}_{[1,\infty)}$ weakly, where $\mathbb{I}_A$ denotes the indicator function over the set A.

Next, we turn our attention to the class of phase-type scale mixture distributions (Bladt et al., 2015). In this paper, we introduce such a class via *Mellin–Stieltjes convolution*

$$\Pi \star G(u) := \int_0^\infty G(u/s) d\Pi(s) = \int_0^\infty \Pi(u/s) dG(s), \tag{2.1}$$

where $G \sim \text{PH}(\boldsymbol{\beta}, \mathbf{\Lambda})$ and Π is a proper nonnegative distribution.

Mellin–Stieltjes convolutions can be interpreted in two equivalent ways. The most common one is to interpret the distribution $\Pi \star G$ as a *scaled mixture distribution*; for instance, $\int G(u/s) d\Pi(s)$ can be seen as a mixture of the scaled distributions $G_s(u) = G(u/s)$ with scaling distribution $\Pi(s)$ (and vice versa). However, it is often more practical to see that $\Pi \star G$ corresponds to the distribution of the product of two independent random variables having distributions Π and G. Furthermore, the integrated tail of $\Pi \star G$ is given in the following proposition. Throughout the current manuscript, μ_H will denote the first moment of any given distribution function H.

Proposition 2.1. *Let Π and G be independent and nonnegative distributions with finite first moments. Then, the integrated tail of $\Pi \star G$ is given by*

$$\widehat{\Pi \star G} = H_\Pi \star \widehat{G},$$

where $\mathrm{d}H_\Pi(s) = s\mathrm{d}\Pi(s)/\mu_\Pi$ is called the moment distribution of Π *and $\widehat{G}$ is the integrated tail of G.*

Proof. Since the Mellin–Stieltjes convolution of Π and G can be seen as the distribution of the product of two independent random variables having distribution Π and G, then $\mu_{\Pi\star G} = \mu_\Pi \cdot \mu_G$.

Observe that

$$\begin{aligned} \widehat{\Pi \star G}(u) &= \frac{1}{\mu_\Pi \cdot \mu_G} \int_0^u (1 - \Pi \star G(t))\, \mathrm{d}t \\ &= \frac{1}{\mu_\Pi} \int_0^u \int_0^\infty \frac{1 - G(t/s)}{\mu_G} \mathrm{d}\Pi(s) \mathrm{d}t \\ &= \int_0^\infty \widehat{G}(u/s) \frac{s\mathrm{d}\Pi(s)}{\mu_\Pi} \\ &= \int_0^\infty \widehat{G}(u/s) \mathrm{d}H_\Pi(s) = H_\Pi \star \widehat{G}(u). \end{aligned}$$

□

Remark 2.2. If G is a PH distribution $G \sim \text{PH}(\boldsymbol{\beta}, \mathbf{\Lambda})$, then $\widehat{G} \sim \text{PH}(-\boldsymbol{\beta}\mathbf{\Lambda}^{-1}/\mu_G, \mathbf{\Lambda})$ is also a PH distribution (cf. Asmussen and Albrecher, 2010, Corollary 2.3.(b), Chapter IX).

The following can be seen as a particular case of Proposition 2.1 when G corresponds to the Dirac measure with point mass at one δ_1 (notice that if $G = \delta_1$, then $\Pi \star G = \Pi$ and $\widehat{G} = U$). However, a self-contained proof is provided in the proposition below.

Proposition 2.3. *Let F be a nonnegative distribution with finite mean μ_F, $\mathrm{d}H_F(s) := s\mathrm{d}F(s)/\mu_F$ be the moment distribution of F, and U the uniform distribution $U(0,1)$. Then the integrated tail distribution of F satisfies*

$$\widehat{F} = H_F \star U.$$

Proof.

$$\begin{aligned}\widehat{F}(u) = \frac{1}{\mu_F}\int_0^u (1-F(t))\,\mathrm{d}t &= \frac{1}{\mu_F}\int_0^u \int_0^\infty \mathbb{I}_{(t,\infty)}(s)\mathrm{d}F(s)\mathrm{d}t \\ &= \frac{1}{\mu_F}\int_0^\infty \left\{\int_0^u \mathbb{I}_{[0,s)}(t)\mathrm{d}t\right\}\mathrm{d}F(s) \\ &= \frac{1}{\mu_F}\int_0^\infty \{u\wedge s\}\,\mathrm{d}F(s) \\ &= \int_0^\infty \{(u/s)\wedge 1\}\,\frac{s\mathrm{d}F(s)}{\mu_F} = H_F \star U(u),\end{aligned}$$

where the second equality follows from Tonelli's theorem and from the fact that for $s,t \geq 0$, $\mathbb{I}_{(t,\infty)}(s) = \mathbb{I}_{[0,s)}(t)$. □

In this paper, we are particularly interested in the case where Π is a discrete distribution having support $\{s_i : i \in \mathbb{N}\}$ with $0 < s_1 < s_2 < \dots$ and vector of probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots)$ such that $\boldsymbol{\pi}\mathbf{e}_{\boldsymbol{\infty}} = 1$, where $\mathbf{e}_{\boldsymbol{\infty}}$ is an infinite dimensional column vector with all elements equal to 1. In such a case, the distribution of $\Pi \star G$ can be written as

$$(\Pi \star G)(u) = \sum_{i=1}^{\infty} G(u/s_i)\pi_i, \quad u \geq 0.$$

Since the scaled phase-type distributions $G(u/s_i) \sim \mathrm{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda}/s_i)$ are PH distributions again, we choose to call $\Pi \star G$ a *phase-type scale mixture* distribution. The class of phase-type scale mixtures was first introduced in Bladt et al. (2015), though they restricted themselves to distributions Π supported over the natural numbers. One of the main features of the class of *phase-type scale mixtures* having a nonnegative discrete scaling distribution Π is that it forms a subclass of the so called *infinite dimensional phase-type* (IDPH) distributions (see Shi et al. (1996)). Indeed, in such a case $\Pi \star G$ can be interpreted as the distribution of absorption time of a Markov jump process with one absorbing state and an infinite number of transient states, having representation $(\boldsymbol{\alpha}, \boldsymbol{T})$ where $\boldsymbol{\alpha} = (\boldsymbol{\pi} \otimes \boldsymbol{\beta})$, the Kronecker product of $\boldsymbol{\pi}$ and $\boldsymbol{\beta}$, and

$$\mathbf{T} = \begin{pmatrix} \boldsymbol{\Lambda}/s_1 & 0 & 0 & \cdots \\ 0 & \boldsymbol{\Lambda}/s_2 & 0 & \cdots \\ 0 & 0 & \boldsymbol{\Lambda}/s_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Finally, if the underlying phase-type distribution G is Erlang and Π is any nonnegative discrete distribution, then we say that the distribution $\Pi \star G$ is an *Erlangized scale mixture*. We will discuss more properties of this distribution in later sections.

All the classes of distributions defined above are particularly attractive for modelling purposes, in part because they are dense in the nonnegative distributions (both the class of infinite dimensional phase-type distributions and the class of phase-type scale mixtures trivially inherit the dense property from classical phase-type distributions, while the proof that the class of Erlangized scale mixtures being dense is simple and given in the next subsection). The class of IDPH distributions contains heavy-tailed distributions, but its infinite-dimensionality makes it computationally intractable. Both phase-type scale mixtures and Erlangized scale mixtures remain dense, contain both light and heavy-tailed distributions and are more tractable than general IDPH distributions, from both theoretical and computational perspectives. Here, we concentrate on a particular subclass of the phase-type scale mixtures defined in Bladt et al. (2015) by narrowing such a class to Erlangized scale mixtures having scaling distribution Π with general discrete support.

2.2 Approximations via Erlangized scale mixtures

Next, we present a methodology for approximating an arbitrary nonnegative distribution Π within the class of Erlangized scale mixtures. The construction is simple and based on the following straightforward result.

Proposition 2.4. *Let Π_m be a sequence of nonnegative discrete distributions such that $\Pi_m \to \Pi$ and $G_m \sim$ Erlang$(\xi(m), \xi(m))$, where $\xi : \mathbb{N} \to \mathbb{N}$ is such that $\xi(m) \to \infty$ as $m \to \infty$. Then*

$$\Pi_m \star G_m \longrightarrow \Pi.$$

Proof. Since Π_m converges weakly to Π and G_m converges weakly to the distribution $\mathbb{I}_{[1,\infty)}$ (the distribution of a constant), then the result follows directly from an application of Slutsky's theorem (cf. Theorem 7.7.1 Ash and Doléans-Dade, 2000). □

For convenience, we refer to this method of approximation as *approximation via Erlangized scale mixtures*. The sequence of discrete distributions Π_m can be seen as rough approximations of the nonnegative distribution Π. Since G_m is an absolutely continuous distribution with respect to the Lebesgue measure, then the Mellin–Stieltjes convolution has a *smoothing* effect over the rough approximating distributions Π_m. Indeed, $\Pi_m \star G_m$ is an absolutely continuous distribution with respect to the Lebesgue measure. The example below illustrates the approximation method described above.

Example 2.5. As target distribution function, we consider the following cdf

$$\Pi(x) = 1 - \left(1 + \frac{x}{\phi - 1}\right)^{-\phi}, \qquad x > 0. \tag{2.2}$$

The distribution in (2.2) corresponds to a scaling transformation of a classical Pareto distribution having a single parameter $\phi > 1$ and supported over the positive real axis. This distribution will be denoted as Pareto(ϕ) from here on. We remark that the scaling is chosen so the mean of Π is 1. This parametrization is often selected for practical purposes but other parametrizations of the Pareto distribution are common as well.

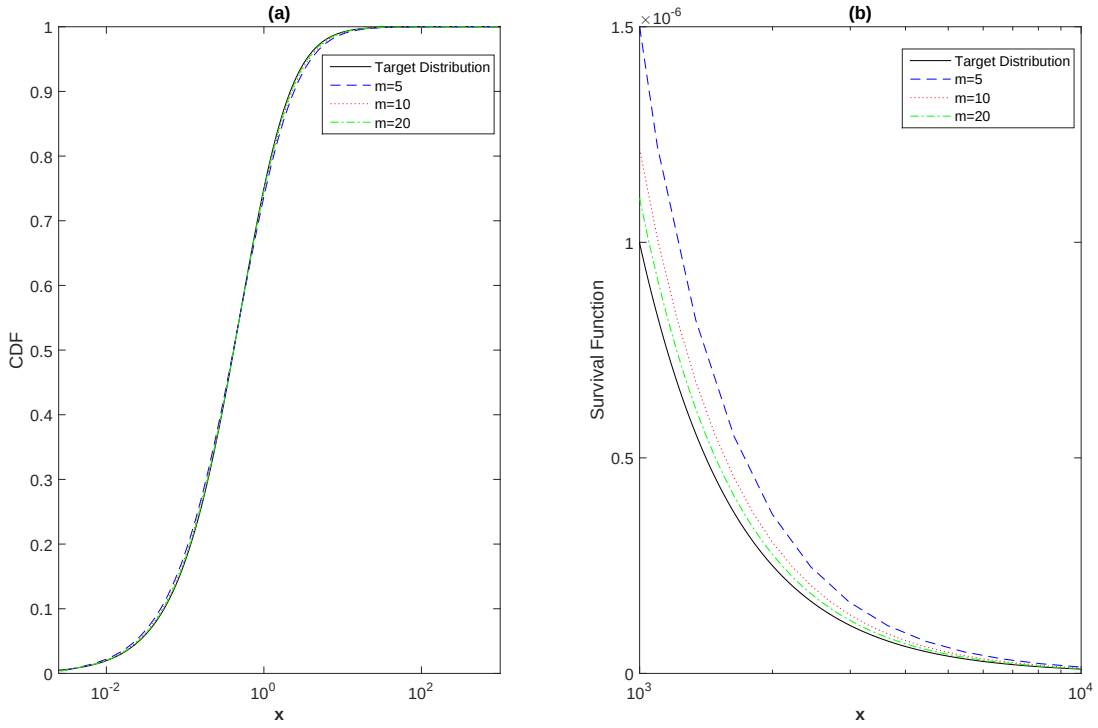
For this example, we take $\xi(m) = m$, so that $G_m \sim \text{Erlang}(m, m)$ and we consider the following sequence Π_m of approximating distributions of Π

$$\Pi_m(x) = \sum_{k=1}^{\infty} \Pi\Big(s_k^{(m)}\Big) \mathbb{I}_{\left[s_k^{(m)}, s_{k+1}^{(m)}\right)}(x). \tag{2.3}$$

The function Π_m corresponds to the discrete cumulative distribution function that matches the target distribution function Π on the set $\mathcal{S}_m := \{s_k^{(m)} : k \in \mathbb{Z}^+\}$. For this particular example, we select $\mathcal{S}_m$ corresponding to a geometric sequence; that is, $s_k^{(m)} := s_1^{(m)} \mathrm{e}^{k/m}$ with $s_1^{(m)} > 0$. Moreover, if $s_1^{(m)} \to 0$ as $m \to \infty$, then $\Pi_m(x) \to \Pi(x)$ pointwise for all $x \in \mathbb{R}$, so that $\Pi_m \to \Pi$ weakly. The selection of the sets $\mathcal{S}_m$ as a geometric sequence obeys a practical purpose: note that the approximating distributions in (2.3) is given as an infinite series and in practice this can only be computed up numerically to a finite number of terms. By selecting a geometric sequence, we can obtain better numerical approximations of Π in the tail region.

In the Figure 1 below, we plot the approximations of a Pareto(2) distribution for various different values of m (we choose $s_1^{(m)} = \mathrm{e}^{-6}$). We remark that the proposed approximating distribution provides accurate approximations of the target distribution Π for rather small values of m. Moreover, the selection of $\mathcal{S}_m$ as a geometric sequence provides a sharp approximation in the tail regions.

Figure 1: Comparison between a target distribution $\Pi \sim$ Pareto(2) and its *Erlangized scale mixture* approximations $\Pi_m \star G_m$, where Π_m is given in (2.3) and $G_m \sim \text{Erlang}(m, m)$. Panel (a) shows the target cdf and its approximations. Panel (b) shows the survival functions of the approximating distributions and the target distribution plotted in the tail region.



3 Ruin probabilities

In this section, we introduce two methods for approximating the ruin probability in the Cramér–Lundberg risk model using Erlangized scale mixtures. Both methods are similar in the sense that the results of Bladt et al. (2015) are adapted to obtain computable expressions for the ruin probability in terms of infinite series involving operations with finite dimensional arrays. The simple structure of the Erlang distribution is exploited to obtain formulas which are free of matrix operations. The two methods differ in the way the integrated claim size distribution $\widehat{F}$ is approximated.

The first method consists in approximating the integrated claim size distribution $\widehat{F}$ via Erlangized scale mixtures. This method is the one that we advocate in this paper and we shall call it *approximation A*. This straightforward approach delivers explicit formulas which are simple to write and implement. Moreover, we will verify empirically that the approximation obtained tends to be more accurate than the one delivered by the second method described below. We remark that this approach has two minor disadvantages. The first disadvantage is that the approximation obtained cannot be easily related to the probability of ruin of some reserve processes because we cannot identify an Erlangized scale mixture as the integrated tail of a phase-type scale mixture; hence an approximating distribution for the claim sizes is not immediately available in this setting. The second disadvantage is that an explicit expression of the integrated tail distribution $\widehat{F}$ is required. In cases where the integrated claim size distribution $\widehat{F}$ is not available explicitly, this can be easily approximated numerically; nevertheless, this numerical approximation will introduce a new source of error so the alternative method described below may be preferred.

The second method, labeled *approximation B*, consists in approximating the claim size distribution F via Erlangized scale mixtures. This approximation is indirect because it is equivalent to approximating the integrated tail distribution $\widehat{F}$ with the integrated tail of an Erlangized scale

mixture distribution. Such an integrated tail distribution is in the class of phase-type distributions, so similar explicit formulas for the ruin probability are obtained. This approach is considered more natural and is the most commonly used; see for instance Bladt et al. (2015) and Santana et al. (2016). Moreover, an explicit expression for the integrated tail distribution $\widehat{F}$ is not necessary for its implementation. However, this method delivers approximations having more complex expressions, and its error of approximation tends to be larger and more difficult to assess due to the *amplifying* effect of integrating the tail probability of the approximating distribution. Also, its implementation is more involved and the computational times are much slower when compared to the results delivered using approximation A.

The remaining content of this section is organized as follows: in Subsection, 3.1 we introduce some basic concepts of ruin probabilities in the classical Cramér–Lundberg risk model. The two approximations of the ruin probability via Erlangized scale mixtures are presented in Subsection 3.2.

3.1 Ruin probability in the Cramér–Lundberg risk model

We consider the classical compound Poisson risk model (cf. Asmussen and Albrecher, 2010):

$$R_t = u + t - \sum_{k=1}^{N_t} X_k.$$

Here t represents time, u is the initial reserve of an insurance company, premiums flow in at a rate 1 per unit of time, $X_1, X_2, \ldots$ are i.i.d. claim sizes with common distribution F and mean μ_F, $\{N_t\}_{t\geq 0}$ is a Poisson process with rate γ denoting the arrival of claims. Thus, $\{R_t\}$ models the evolution in time of an insurance company's reserve. We say that ruin occurs if the reserve ever drops below zero. The time of ruin is denoted $\tau := \inf\{R_t < 0 : t > 0\}$, while the ruin probability is denoted $\psi_{\widehat{F}}(u) := \mathbb{P}(\tau < \infty \mid R_0 = u)$.

For such a model, the well-known Pollaczek–Khinchine formula (cf. Asmussen and Albrecher, 2010) implies that the ruin probability can be expressed in terms of convolutions:

$$\psi_{\widehat{F}}(u) = (1-\rho)\sum_{n=1}^{\infty} \rho^n \overline{\widehat{F}^{*n}}(u), \tag{3.1}$$

where $\rho = \gamma\mu_F < 1$ is the average claim amount per unit of time, F^{*n} denotes the nth-fold convolution of F, $\overline{F} := 1 - F$ denotes the tail probability of F, and $\widehat{F}$ is the integrated tail distribution also known as the stationary excess distribution:

$$\widehat{F}(u) = \frac{1}{\mu_F}\int_0^u \overline{F}\,(t)\mathrm{d}t.$$

The calculation of the ruin probability is conveniently approached via renewal theory. The ruin probability $\psi_{\widehat{F}}(u)$ of the classical Cramér–Lundberg process can be written as the probability that a terminating renewal process reaches level u. In such a model, the distribution of the renewals is defective, and given by $\rho\widehat{F}(u)$. In particular, if the renewals follow a defective phase-type scale mixture distribution with distribution $\rho\Pi\star G$ with $0 < \rho < 1$, then Bladt et al. (2015) derived that the probability that the lifetime of the renewal is larger than u is given by

$$\psi_{\Pi\star G}(u) = \rho\boldsymbol{\alpha}\mathrm{e}^{(\boldsymbol{T}+\rho\boldsymbol{t}\boldsymbol{\alpha})u}\boldsymbol{e}_\infty, \tag{3.2}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\pi}\otimes\boldsymbol{\beta})$, $\boldsymbol{T} = \mathrm{diag}(\boldsymbol{s})^{-1}\otimes\boldsymbol{\Lambda}$, $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}_\infty$, and $\otimes$ denotes the Kronecker product between matrices/vectors. Here $\boldsymbol{s} = (s_1, s_2, \ldots)$, and $\mathrm{diag}(\boldsymbol{s})$ is an infinite dimensional diagonal matrix with diagonal elements equal to $(s_1, s_2, \ldots)$. Formula (3.2) is not of practical use because the vectors $\boldsymbol{\alpha}$, $\boldsymbol{t}$ and the matrix $\boldsymbol{T}$ have infinite dimensions. However, using the special structure of $\boldsymbol{T}$, they further refined the formula above and expressed $\psi_{\Pi\star G}$ as an infinite series involving matrices and vectors of finite dimensions which characterize the underlying distributions Π and G.

In what follows, we obtain an explicit formula for $\psi_{\Pi\star G}(u)$ in terms of the parameters characterising the renewal distribution $\Pi\star G$ (the integrated tail distribution). This is a slight generalization

of the results given in Bladt et al. (2015) who implicitly assumed that $\Pi \star G$ is the integrated tail of a phase-type scale mixture distribution, so their results are given instead in terms of the parameters characterising the underlying claim size distribution. For simplicity, we will write $G_m \sim \text{Erlang}(\xi, \xi)$ instead of $\text{Erlang}(\xi(m), \xi(m))$ for the rest of the paper. Also, $\boldsymbol{I}$ will denote an identity matrix of appropiate dimension and $\boldsymbol{I_\infty}$ its infinite dimensional analogue. To simply our notation, we suppress the index m from the distribution Π_m; this convention will be adopted for the rest of the paper.

Proposition 3.1. *Let $0 < \rho < 1$. Then,*

$$\psi_{\Pi \star G_m}(u) = \sum_{n=0}^{\infty} \kappa_n \frac{(\theta u/s_1)^n \mathrm{e}^{-\theta u/s_1}}{n!}, \tag{3.3}$$

where θ is the largest diagonal element of $-\boldsymbol{\Lambda}$, and

$$\kappa_n = \begin{cases} \rho, & n = 0, \\ \rho \left[\dfrac{s_1}{\theta} \left(\displaystyle\sum_{i=0}^{n-1} \kappa_{n-1-i} \sum_{j=1}^{\infty} \frac{\pi_j}{s_j} B_{ij} \right) + \sum_{j=1}^{\infty} \pi_j C_{nj} \right], & n > 0, \end{cases}$$

where

$$B_{ij} := \boldsymbol{\beta}(\boldsymbol{I} + (s_j\theta/s_1)^{-1}\boldsymbol{\Lambda})^i \boldsymbol{\lambda}, \qquad C_{nj} := \boldsymbol{\beta}(\boldsymbol{I} + (s_j\theta/s_1)^{-1}\boldsymbol{\Lambda})^n \boldsymbol{e}.$$

Proof. Since θ is the largest diagonal element of $-\boldsymbol{\Lambda}$ and $\{s_i\}$ is an increasing sequence, then θ/s_1 is the largest diagonal element of $-\boldsymbol{T}$. From Theorem 3.1 in Bladt et al. (2015), we have that

$$\psi_{\Pi \star G_m}(u) = \sum_{n=0}^{\infty} \kappa_n \frac{(\theta u/s_1)^n \mathrm{e}^{-\theta u/s_1}}{n!},$$

where $\kappa_0 = \rho(\boldsymbol{\pi} \otimes \boldsymbol{\beta})\mathbf{e_\infty} = \rho \sum_{i=0}^{\infty} \pi_i = \rho$, and

$$\kappa_n = \rho \left[\sum_{i=0}^{n-1} \frac{s_1}{\theta} (\boldsymbol{\pi} \otimes \boldsymbol{\beta}) \left(\boldsymbol{I_\infty} + \frac{s_1}{\theta}\boldsymbol{T} \right)^i \boldsymbol{t} \kappa_{n-1-i} + (\boldsymbol{\pi} \otimes \boldsymbol{\beta}) \left(\boldsymbol{I_\infty} + \frac{s_1}{\theta}\boldsymbol{T} \right)^n \boldsymbol{e_\infty} \right].$$

It is not difficult to see that

$$(\boldsymbol{\pi} \otimes \boldsymbol{\beta}) \left(\boldsymbol{I_\infty} + \frac{s_1}{\theta}\boldsymbol{T} \right)^i \boldsymbol{t} = \sum_{j=1}^{\infty} \pi_j \boldsymbol{\beta} \left(\boldsymbol{I} + \frac{s_1}{s_j\theta}\boldsymbol{\Lambda} \right)^i \left(-\frac{\boldsymbol{\Lambda e}}{s_j} \right) = \sum_{j=1}^{\infty} \frac{\pi_j}{s_j} B_{ij}$$

and

$$(\boldsymbol{\pi} \otimes \boldsymbol{\beta}) \left(\boldsymbol{I_\infty} + \frac{s_1}{\theta}\boldsymbol{T} \right)^n \boldsymbol{e_\infty} = \sum_{j=1}^{\infty} \pi_j \boldsymbol{\beta} \left(\boldsymbol{I} + \frac{s_1}{s_j\theta}\boldsymbol{\Lambda} \right)^n \boldsymbol{e} = \sum_{j=1}^{\infty} \pi_j C_{nj},$$

where B_{ij} and C_{nj} are defined as above. □

Proposition 3.1 is to be interpreted as the probability that the lifetime of a defective renewal process exceeds level u. An interpretation in terms of the risk process is not always possible since we may not be able to identify a claim size distribution having integrated tail $\Pi \star G_m$. The result above can be seen as a (slight) generalization of Theorem 3.1 of Bladt et al. (2015). This can be seen from Proposition 2.1, which shows that if the claim sizes are distributed according to an Erlangized scale mixture $\Pi \star G_m$, then the integrated tail of $\Pi \star G_m$ remains in the family of phase-type scale mixtures. Using the results of Proposition 2.1 and Remark 2.2, we recover the formulas found at the beginning of p. 12 in Bladt et al. (2015).

Proposition 3.2 (Bladt et al. (2015))**.**

$$\psi_{H_\Pi \star \widehat{G}}(u) = \sum_{n=0}^{\infty} \kappa_n \frac{(\theta u/s_1)^n \mathrm{e}^{-\theta u/s_1}}{n!}, \tag{3.4}$$

where θ is the largest diagonal element of $-\boldsymbol{\Lambda}$ and

$$\kappa_n = \begin{cases} \rho, & n = 0, \\ \dfrac{\rho}{\mu_\Pi \cdot \mu_G} \left[\dfrac{s_1}{\theta} \left(\sum\limits_{i=0}^{n-1} \kappa_{n-1-i} \sum\limits_{j=1}^{\infty} \pi_j C_{ij} \right) + \sum\limits_{j=1}^{\infty} \pi_j s_j D_{nj} \right], & n > 0, \end{cases}$$

where

$$C_{ij} := \boldsymbol{\beta}(\boldsymbol{I} + (s_j\theta/s_1)^{-1}\boldsymbol{\Lambda})^i \boldsymbol{e}, \qquad D_{nj} := \boldsymbol{\beta}(-\boldsymbol{\Lambda})^{-1}(\boldsymbol{I} + (s_j\theta/s_1)^{-1}\boldsymbol{\Lambda})^n \boldsymbol{e}.$$

A drawback of the formulas given above is that the calculation of the quantities B_{ij}, C_{ij} and D_{ij} is computationally expensive since it involves costly matrix operations. However, these expressions can be simplified in our case because the subintensity matrix $\boldsymbol{\Lambda}$ of an Erlang distribution can be written as a bidiagonal matrix, while the vectors denoting the initial distribution $\boldsymbol{\beta}$ and the absorption rates $\boldsymbol{\lambda}$ are proportional to canonical vectors. Hence, the resulting expressions for the terms B_{ij}, C_{ij} and D_{ij} in Proposition 3.1 and Proposition 3.2 will take relatively simple forms. These are given in the following lemma.

Lemma 3.3. *Suppose that $G_m \sim$ Erlang(ξ, ξ), then*

$$B_{ij} = \begin{cases} 0, & i < \xi - 1, \\ \xi \binom{i}{\xi - 1} \left(1 - \dfrac{s_1}{s_j}\right)^{i-\xi+1} \left(\dfrac{s_1}{s_j}\right)^{\xi-1}, & i \geq \xi - 1, \end{cases}$$

$$C_{ij} = \begin{cases} 1, & i \leq \xi - 1, \\ \sum\limits_{k=0}^{\xi-1} \binom{i}{k} \left(1 - \dfrac{s_1}{s_j}\right)^{i-k} \left(\dfrac{s_1}{s_j}\right)^{k}, & i \geq \xi - 1, \end{cases}$$

$$D_{ij} = \begin{cases} 1 - \dfrac{i}{\xi}\dfrac{s_1}{s_j}, & i \leq \xi, \\ \sum\limits_{k=0}^{\xi-1} \dfrac{\xi - k}{\xi} \binom{i}{k} \left(1 - \dfrac{s_1}{s_j}\right)^{i-k} \left(\dfrac{s_1}{s_j}\right)^{k}, & i > \xi. \end{cases}$$

Proof. Let $(\boldsymbol{\beta}, \boldsymbol{\Lambda})$ be the canonical parameters of the phase-type representation of an Erlang(ξ, ξ) distribution (see Subsection 2.1), so that $\theta = \xi$. Recall that

$$\begin{aligned} B_{ij} &:= \boldsymbol{\beta}(\boldsymbol{I} + (s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})^i \boldsymbol{\lambda}, \\ C_{ij} &:= \boldsymbol{\beta}(\boldsymbol{I} + (s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})^i \boldsymbol{e}, \\ D_{ij} &:= \boldsymbol{\beta}(-\boldsymbol{\Lambda})^{-1}(\boldsymbol{I} + (s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})^i \boldsymbol{e}. \end{aligned}$$

Observe that the matrix $(\boldsymbol{I} + (s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})$ is bidiagonal with all the elements in the diagonal being equal to $1 - s_1/s_j$. In particular, the (k, ℓ)-th entry of the i-th power of such a matrix is given by

$$(\boldsymbol{I} + (s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})^i_{k\ell} = \begin{cases} \binom{i}{\ell - k} \left(1 - \dfrac{s_1}{s_j}\right)^{i-\ell+k} \left(\dfrac{s_1}{s_j}\right)^{\ell-k} & 1 \leq k \leq \ell \leq i+1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, B_{ij} corresponds to the $(1,\xi)$-entry of the matrix $(\boldsymbol{I}+(s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})$ multiplied by ξ. C_{ij} corresponds to the sum of the elements of the first row of $(\boldsymbol{I}+(s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})$. For the last case, observe that $\boldsymbol{\Lambda}^{-1}=-\lambda^{-1}\boldsymbol{U}$ where $\boldsymbol{U}$ is an upper triangular matrix of ones. Thus, D_{ij} corresponds to the sum of the elements of $(\boldsymbol{I}+(s_j\xi/s_1)^{-1}\boldsymbol{\Lambda})$ and divided by ξ. D_{ij} is written as the sum of all the elements in the upper diagonals divided by ξ. □

3.2 Ruin probability for Erlangized scale mixtures

In this subsection, we specialize in approximating the ruin probability $\psi_{\widehat{F}}(u)$ using Erlangized scale mixtures. We assume that the target Cramér–Lundberg risk process has Poisson intensity γ and claim size distribution F, so that the average claim amount per unit of time is $\rho=\gamma\mu_F$.

The first method consists in approximating the integrated tail $\widehat{F}$ with an Erlangized scale mixture $\Pi\star G_m$ where Π is an approximating discrete distribution of $\widehat{F}$. Hence, approximation A of the ruin probability $\psi_{\widehat{F}}$ is obtained by using $\psi_{\Pi\star G_m}$. We remark that under this approach, we sacrifice the interpretation of the approximation $\psi_{\Pi\star G_m}$ as the ruin probability of some Cramér–Lundberg reserve process since it might not be possible to identify a distribution whose integrated tail corresponds to the Erlangized scale mixture distribution $\Pi\star G_m$. Moreover, we also lose the interpretation of the value ρ as the average claim amount per unit of time (in the original risk process, the value of ρ is selected as the product of the expected value of an individual claim multiplied by the intensity of the Poisson process), but for practical computations this is easily fixed by simply letting $\rho=\gamma\mu_F$ where μ_F is the mean value of the original claim size distribution. An explicit expression for the approximation $\psi_{\Pi\star G_m}(u)$ is given next.

Theorem 3.4 (Approximation A)**.** *Let Π be a nonnegative discrete distribution supported over $\{s_i : i\in\mathbb{N}\}$, $G_m\sim$ Erlang(ξ,ξ) and $\rho=\gamma\mu_F<1$. The lifetime of a terminating renewal process with $\rho(\Pi\star G_m)$-distributed renewal intervals is given by*

$$\psi_{\Pi\star G_m}(u)=\sum_{n=0}^{\infty}\kappa_n^A\frac{(\xi u/s_1)^n\mathrm{e}^{-\xi u/s_1}}{n!}, \tag{3.5}$$

where

$$\kappa_n^A=\begin{cases}\gamma\mu_F, & 0\le n\le\xi-1,\\ \gamma\mu_F\left[\sum\limits_{i=\xi-1}^{n-1}\kappa_{n-1-i}^A\mathcal{B}_i+\mathcal{C}_n\right], & \xi\le n,\end{cases}$$

and

$$\mathcal{B}_i=\sum_{j=1}^{\infty}\frac{\pi_j s_1}{s_j(1-s_1/s_j)}\mathrm{nbin}(\xi-1;i-\xi+2,s_1/s_j),$$

$$\mathcal{C}_n=\sum_{j=1}^{\infty}\frac{\pi_j}{1-s_1/s_j}\mathrm{NBin}(\xi-1,n-\xi+2,s_1/s_j).$$

$\mathrm{NBin}(\cdot;r,p)$ *and* $\mathrm{nbin}(\cdot;r,p)$ *denote the cdf and pdf, respectively, of a negative binomial distribution with parameters r and p.*

Proof. The result follows by letting $\rho=\gamma\mu_F$, $\theta=\xi$, $\lambda=\xi$, applying Proposition 3.1 and Lemma 3.3. □

Next, we look at approximation B, which consists in approximating the claim size distribution F via Erlangized scale mixture distributions $\Pi\star G_m$, that is, Π is now a direct discrete approximation of F rather than of $\widehat{F}$. Approximation B corresponds to the ruin probability of an alternative reserve process having claim sizes $\Pi\star G_m$, with Poisson process' intensity selected in such a way that the average claim amount per unit of time matches the value $\rho=\gamma\mu_F$ of the original reserve process. Since this is equivalent to approximating the integrated tail distribution $\widehat{F}$ with the distribution $H_\Pi\star\widehat{G}_m$ (see Proposition 2.1), then Approximation B will be denoted by $\psi_{H_\Pi\star\widehat{G}_m}(u)$. The following theorem provides an explicit expression for approximation B:

Theorem 3.5 (Approximation B)**.** *Let* Π *be a nonnegative discrete distribution supported over* $\{s_i : i \in \mathbb{N}\}$ *and* $G_m \sim Erlang(\xi, \xi)$*. The probability of ruin in the Cramér–Lundberg model having intensity* $\gamma\mu_F/\mu_\Pi$ *and claim size distribution* $\Pi \star G_m$ *is given by*

$$\psi_{H_\Pi \star \widehat{G}_m}(u) = \sum_{n=0}^{\infty} \kappa_n^B \frac{(\xi u/s_1)^n \mathrm{e}^{-\xi u/s_1}}{n!}, \tag{3.6}$$

where

$$\kappa_n^B = \begin{cases} \gamma\mu_F, & n = 0, \\ (\gamma\mu_F - 1)\left(1 + \dfrac{\gamma\mu_F s_1}{\mu_\Pi \xi}\right)^n + 1, & 1 \le n \le \xi, \\ \dfrac{\gamma\mu_F s_1}{\mu_\Pi \xi} \displaystyle\sum_{i=0}^{n-1} \kappa_{n-1-i}^B \mathcal{C}_i + \dfrac{\gamma\mu_F}{\mu_\Pi}\mathcal{D}_n, & \xi < n, \end{cases}$$

and

$$\mathcal{C}_i = \sum_{j=1}^{\infty} \frac{\pi_j}{1 - s_1/s_j} \mathrm{NBin}(\xi - 1; i - \xi + 2, s_1/s_j),$$

$$\mathcal{D}_n = \sum_{j=1}^{\infty} \frac{\pi_j s_j}{1 - s_1/s_j} \sum_{k=0}^{\xi-1} \frac{\xi - k}{\xi} \mathrm{nbin}(k; n - k + 1, s_1/s_j).$$

Proof. Let $\theta = \xi$ and $\lambda = \xi$. If $1 \le n \le \xi$, then from Proposition 3.2 and Lemma 3.3, we have that

$$\begin{aligned} \kappa_n^B &= \frac{\gamma\mu_F s_1}{\mu_\Pi \xi} \sum_{i=0}^{n-1} \kappa_{n-1-i}^B + \frac{\gamma\mu_F}{\mu_\Pi} \sum_{j=1}^{\infty} \pi_j s_j \left(1 - \frac{n}{\xi}\frac{s_1}{s_j}\right), \\ &= \frac{\gamma\mu_F s_1}{\mu_\Pi \xi} \sum_{i=0}^{n-1} \kappa_{n-1-i}^B + \gamma\mu_F - \frac{\gamma\mu_F}{\mu_\Pi \xi} \sum_{j=1}^{\infty} s_j \pi_j n \frac{s_1}{s_j} \\ &= \frac{\gamma\mu_F s_1}{\mu_\Pi \xi} \left(\sum_{i=0}^{n-1} \kappa_i^B - n\right) + \gamma\mu_F. \end{aligned}$$

By induction, we can get for $1 \le n \le \xi$,

$$\kappa_n^B = (\gamma\mu_F - 1)\left(1 + \frac{\gamma\mu_F s_1}{\mu_\Pi \xi}\right)^n + 1.$$

The cases $n = 0$ and $\xi < n$ follow directly from applying Proposition 3.2 and Lemma 3.3. □

An attractive feature of the two approximations presented above is that because of the simple structure of the phase-type representation of the Erlang distribution G_m, it is possible to rewrite each approximation in simple terms free of matrix operations. In particular, the simplified expressions for the values of κ_n^A and κ_n^B are given in terms of the negative binomial distribution which are particularly convenient for computational purposes. Approximation A has a simpler expression which tends to be more accurate, easier to implement and faster to compute. On the other hand, approximation B can be computed even when the integrated claim size distribution is not available.

4 Error bounds for the ruin probability

In this section, we will assess the accuracy of the two proposed approximations for the ruin probability. We will do so by providing upper bounds for the errors of approximation. For each approximation we identify two sources of error. The first source is due to the Mellin–Stieltjes convolution with the Erlang distribution; we will call this the *Erlangization error*. The second source of error is due to the approximation of the integrated tail $\widehat{F}$ (via Π in the first case, and via H_Π in the

second case); we will refer to this as the *discretization error*. For the case of approximation A in Theorem 3.4, we can use the triangle inequality to bound the overall error with the aggregation of the two types of errors, that is

$$\left|\psi_{\widehat{F}}(u)-\psi_{\Pi\star G_m}(u)\right| \leq \underbrace{\left|\psi_{\widehat{F}}(u)-\psi_{\widehat{F}\star G_m}(u)\right|}_{\text{Erlangization error A}} + \underbrace{\left|\psi_{\widehat{F}\star G_m}(u)-\psi_{\Pi\star G_m}(u)\right|}_{\text{Discretization error A}}. \tag{4.1}$$

For approximation B in Theorem 3.5, we obtain an analogous bound

$$\left|\psi_{\widehat{F}}(u)-\psi_{H_\Pi\star\widehat{G}_m}(u)\right| \leq \underbrace{\left|\psi_{H_F\star U}(u)-\psi_{H_F\star\widehat{G}_m}(u)\right|}_{\text{Erlangization error B}} + \underbrace{\left|\psi_{H_F\star\widehat{G}_m}(u)-\psi_{H_\Pi\star\widehat{G}_m}(u)\right|}_{\text{Discretization error B}}. \tag{4.2}$$

In the last inequality, we used Proposition 2.3, which states that $\widehat{F}=H_F\star U$.

We will rely on the Pollaczek–Khinchine formula (3.1) for the construction of the error bounds. Recall that the Pollaczek–Khinchine formula for $\psi_{\widehat{F}}(u)$ can be interpreted as the probability that a terminating renewal process having defective renewal probability $\rho\widehat{F}(\cdot)$ will ever reach level u before terminating. In our two approximations of $\psi_{\widehat{F}}$, we selected the value of $\rho=\gamma\mu_F$ so we can write the errors of approximation in terms of the differences between the convolutions of the integrated tails exclusively. For instance, the error of Erlangization in approximation A is given by

$$\left|\psi_{\widehat{F}}(u)-\psi_{\widehat{F}\star G_m}(u)\right| = \left|\sum_{n=1}^{\infty}(1-\rho)\rho^n\left(\overline{\widehat{F}^{*n}}(u)-\overline{\left(\widehat{F}\star G_m\right)^{*n}}(u)\right)\right|. \tag{4.3}$$

For our approximation B, it is noted that setting the parameter $\rho=\gamma\mu_F$ is equivalent to calculating the ruin probability for a risk process having integrated claim sizes distributed according to $H_\Pi\star\widehat{G}_m$ while the intensity of the Poisson process is changed to $\gamma\mu_F/\mu_\Pi$. With such an adjustment, it is possible to write both the Erlangization and discretization errors in terms of differences of higher order convolutions as the ones given in (4.3).

We divide the current section in three parts. In Subsection 4.1, we refine an existing upper bound introduced in Vatamidou et al. (2014) for the distance between two ruin probabilities $|\psi_{\widehat{F_1}}(u)-\psi_{\widehat{F_2}}(u)|$. This refinement will be used in Subsections 4.2 and 4.3, where we provide upper bounds for each individual error in equations (4.1) and (4.2).

4.1 General bounds for the error of approximation

The following theorem provides a refinement for the upper bound provided in Vatamidou et al. (2014). As mentioned before, this result will be used later in the construction of upper bounds for each individual error.

Theorem 4.1. *Let $\widehat{F_1}$ and $\widehat{F_2}$ be two nonnegative distributions. For any fixed $u>0$ it holds that*

$$\left|\psi_{\widehat{F_1}}(u)-\psi_{\widehat{F_2}}(u)\right| \leq \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\frac{(1-\rho)\rho}{\left(1-\rho\widehat{F_1}(u)\right)\left(1-\rho\widehat{F_2}(u)\right)}. \tag{4.4}$$

Proof. We claim that for any $n\geq 1$,

$$\sup_{s<u}\left\{\left|\widehat{F_1}^{*n}(s)-\widehat{F_2}^{*n}(s)\right|\right\} \leq \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\sum_{i=0}^{n-1}\widehat{F_1}^{i}(u)\widehat{F_2}^{n-1-i}(u). \tag{4.5}$$

Let us prove it by induction. It is clearly valid for $n=1$. Let us assume that it is valid for some $n\geq 1$. Then

$$\begin{aligned}
&\sup_{s<u}\left\{\left|\widehat{F_1}^{*n+1}(s)-\widehat{F_2}^{*n+1}(s)\right|\right\}\\
&\quad=\sup_{s<u}\left\{\left|\widehat{F_1}^{*n+1}(s)-\widehat{F_1}^{*n}*\widehat{F_2}(s)+\widehat{F_1}^{*n}*\widehat{F_2}(s)-\widehat{F_2}^{*n+1}(s)\right|\right\}\\
&\quad\leq\sup_{s<u}\left\{\left|\widehat{F_1}^{*n+1}(s)-\widehat{F_1}^{*n}*\widehat{F_2}(s)\right|\right\}+\sup_{s<u}\left\{\left|\widehat{F_1}^{*n}*\widehat{F_2}(s)-\widehat{F_2}^{*n+1}(s)\right|\right\}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
\sup_{s<u}\left\{\left|\widehat{F_1}^{*n+1}(s)-\widehat{F_1}^{*n}*\widehat{F_2}(s)\right|\right\} &\leq \sup_{s<u}\left\{\int_0^s\left|\widehat{F_1}(r)-\widehat{F_2}(r)\right|\mathrm{d}\widehat{F_1}^{*n}(r)\right\}\\
&\leq \sup_{s<u}\left\{\int_0^s \sup_{l<u}\left\{\left|\widehat{F_1}(l)-\widehat{F_2}(l)\right|\right\}\mathrm{d}\widehat{F_1}^{*n}(r)\right\}\\
&= \sup_{l<u}\left\{\left|\widehat{F_1}(l)-\widehat{F_2}(l)\right|\right\}\sup_{s<u}\left\{\int_0^s \mathrm{d}\widehat{F_1}^{*n}(r)\right\}\\
&= \sup_{l<u}\left\{\left|\widehat{F_1}(l)-\widehat{F_2}(l)\right|\right\}\widehat{F_1}^{*n}(u)\\
&\leq \sup_{l<u}\left\{\left|\widehat{F_1}(l)-\widehat{F_2}(l)\right|\right\}\widehat{F_1}^{n}(u). \qquad (4.6)
\end{aligned}$$

In the last step, we used that $\widehat{F}^{*n}(u)$ corresponds to the probability of an event where the sum of n i.i.d. random variables is smaller or equal than u while $\widehat{F}^n(u)$ corresponds to the probability of the maximum of i.i.d. random variables is smaller or equal than u: if the random variables are nonnegative then the probability of the sum is clearly smaller than the probability of the maximum. Using the hypothesis induction, we have that

$$\begin{aligned}
\sup_{s<u}\left\{\left|\widehat{F_1}^{*n}*\widehat{F_2}(s)-\widehat{F_2}^{*n+1}(s)\right|\right\} &\leq \sup_{s<u}\left\{\int_0^s\left|\widehat{F_1}^{*n}(r)-\widehat{F_2}^{*n}(r)\right|\mathrm{d}\widehat{F_2}(r)\right\}\\
&\leq \sup_{s<u}\left\{\int_0^s \sup_{l<u}\left\{\left|\widehat{F_1}^{*n}(l)-\widehat{F_2}^{*n}(l)\right|\right\}\mathrm{d}\widehat{F_2}(r)\right\}\\
&= \sup_{l<u}\left\{\left|\widehat{F_1}^{*n}(l)-\widehat{F_2}^{*n}(l)\right|\right\}\sup_{s<u}\left\{\int_0^s \mathrm{d}\widehat{F_2}(r)\right\}\\
&\leq \left(\sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\sum_{i=0}^{n-1}\widehat{F_1}^{i}(u)\widehat{F_2}^{n-1-i}(u)\right)\widehat{F_2}(u)\\
&= \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\sum_{i=0}^{n-1}\widehat{F_1}^{i}(u)\widehat{F_2}^{n-i}(u). \qquad (4.7)
\end{aligned}$$

Summing (4.6) and (4.7), we get that

$$\sup_{s<u}\left\{\left|\widehat{F_1}^{*n+1}(s)-\widehat{F_2}^{*n+1}(s)\right|\right\} \leq \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\sum_{i=0}^{n}\widehat{F_1}^{i}(u)\widehat{F_2}^{n-i}(u),$$

so that formula (4.5) is valid for all $n \geq 1$. Finally,

$$\begin{aligned}
\left|\psi_{\widehat{F_1}}(u)-\psi_{\widehat{F_2}}(u)\right| &\leq \sum_{n=1}^{\infty}(1-\rho)\rho^n\left|\widehat{F_1}^{*n}(u)-\widehat{F_2}^{*n}(u)\right|\\
&\leq \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}(1-\rho)\sum_{n=1}^{\infty}\rho^n\sum_{i=0}^{n-1}\widehat{F_1}^{i}(u)\widehat{F_2}^{n-1-i}(u)\\
&= \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}(1-\rho)\sum_{i=0}^{\infty}\sum_{n=i+1}^{\infty}\rho^n\widehat{F_1}^{i}(u)\widehat{F_2}^{n-1-i}(u)\\
&= \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}(1-\rho)\sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\rho^{n+i+1}\widehat{F_1}^{i}(u)\widehat{F_2}^{n}(u)\\
&= \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}(1-\rho)\rho\sum_{i=0}^{\infty}\rho^i\widehat{F_1}^{i}(u)\sum_{n=0}^{\infty}\rho^n\widehat{F_2}^{n}(u)\\
&= \sup_{s<u}\left\{\left|\widehat{F_1}(s)-\widehat{F_2}(s)\right|\right\}\frac{(1-\rho)\rho}{\left(1-\rho\widehat{F_1}(u)\right)\left(1-\rho\widehat{F_2}(u)\right)}.
\end{aligned}$$

□

The construction of our bound differs from the one proposed by Vatamidou et al. (2014) on the inequality (4.7) which is given by

$$\sup_{s<u}\left\{\left|\widehat{F_1}^{*n} * \widehat{F_2}(s) - \widehat{F_2}^{*n+1}(s)\right|\right\} \leq \sup_{s<u}\left\{\left|\widehat{F_1}(s) - \widehat{F_2}(s)\right|\right\} \sum_{i=0}^{n-1} \widehat{F_1}^{i}(u)\widehat{F_2}^{n-i}(u).$$

Each term in the summation on the right hand side takes values in $(0, 1)$. In contrast, the terms in the analogue summation used in Vatamidou et al. (2014) to bound the expression in the left hand side are $n\widehat{F}(u)$, which go to infinity as $n \to \infty$. We remark however, that the final bound for the error term proposed there remains bounded. A comparison of the two bounds reveals that the one suggested above improves Vatamidou et al. (2014)'s bound by a factor of

$$\frac{(1-\rho)^2}{\left(1-\rho\widehat{F_1}(u)\right)\left(1-\rho\widehat{F_2}(u)\right)} \leq 1.$$

Remark 4.2. The bound given in (4.4) is formed by the product of two terms. The computation of the second term is the simplest one since its expression is given explicitly. However, the first term is a functional of the distance between the integrated tail distributions, so in general, this quantity can be very difficult to compute with precision. In our developments below, we will need to construct computable upper bounds for this term in each case considered.

4.2 Error bounds for approximation A

Next, we concentrate on the construction of upper bounds for the error of approximation A proposed in Theorem 3.4. As indicated in (4.1), an upper bound for the theoretical error of approximation A will be constructed as the aggregation of an upper bound for the Erlangization error (see Subsection 4.2.1) plus an upper bound for the discretization error (see Subsection 4.2.2).

4.2.1 Bounds for the Erlangization error of approximation A

For the construction on an upper bound for the Erlangization error A, we will employ Theorem 4.1 above; this result requires the quantity $\sup_{\ell\leq u}\left|\widehat{F}(\ell) - \widehat{F} \star G_m(\ell)\right|$. Lemma 4.3 below provides a computable upper bound for such a quantity. The resulting upper bound for the Erlangization error of approximation A is given in Theorem 4.4.

Lemma 4.3. *Let $\{\mathcal{A}_k : k \in \mathbb{N}\}$ be an decreasing collection of closed intervals in $\mathbb{R}^+$, so $\mathcal{A}_k = [a_k, b_k]$ and $\mathcal{A}_{k+1} \subset \mathcal{A}_k$. If $\mathcal{A}_0 = [0, \infty]$ and $\mathcal{A}_k \searrow \{1\}$ then*

$$\sup_{\ell\leq u}\left\{\left|\widehat{F}(\ell) - \widehat{F} \star G_m(\ell)\right|\right\} \leq \sum_{k=0}^{\infty} \sup_{\ell<u}\left\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\right\}(G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})),$$

where $G_m(\mathcal{A}_k) := G_m(b_k) - G_m(a_k)$.

Proof.

$$\begin{aligned}
\sup_{\ell\leq u}\left\{\left|\widehat{F}(\ell) - \widehat{F} \star G_m(\ell)\right|\right\} &\leq \sup_{\ell<u}\left\{\left|\sum_{k=0}^{\infty}\left[\widehat{F}(\ell)\int_{\mathcal{A}_k/\mathcal{A}_{k+1}} \mathrm{d}G_m(s) - \int_{\mathcal{A}_k/\mathcal{A}_{k+1}} \widehat{F}(\ell/s)\mathrm{d}G_m(s)\right]\right|\right\} \\
&= \sum_{k=0}^{\infty}\sup_{\ell<u}\left\{\left|\int_{\mathcal{A}_k/\mathcal{A}_{k+1}}\left[\widehat{F}(\ell) - \widehat{F}(\ell/s)\right]\mathrm{d}G_m(s)\right|\right\} \\
&\leq \sum_{k=0}^{\infty}\sup_{\ell<u}\left\{\widehat{F}(\ell/b_k) - \widehat{F}(\ell/a_k)\right\}(G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})) \\
&\leq \sum_{k=0}^{\infty}\sup_{\ell<u}\left\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\right\}(G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})).
\end{aligned}$$

□

Applying Theorem 4.1 and Lemma 4.3, we obtain the upper bound (4.8) for the Erlangization error.

Theorem 4.4. *Let $\{\mathcal{A}_k : k \in \mathbb{N}\}$ be a sequence as defined in Lemma 4.3. Then*

$$\left|\psi_{\widehat{F}}(u) - \psi_{\widehat{F}\star G_m}(u)\right| \leq \frac{\rho}{\left(1-\rho\widehat{F}(u)\right)} \sum_{k=0}^{\infty} \sup_{\ell<u} \left\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\right\} (G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})). \quad (4.8)$$

Moreover, if $\widehat{F}$ is absolutely continuous with bounded density then $\psi_{\widehat{F}} \to \psi_{\widehat{F}\star G_m}$ uniformly as $\xi(m) \to \infty$.

Proof. All that is left to prove is uniform convergence. Notice that (4.8) can be further bounded above by

$$\left|\psi_{\widehat{F}}(u) - \psi_{\widehat{F}\star G_m}(u)\right| \leq \frac{\rho}{1-\rho} \sum_{k=0}^{\infty} \sup_{\ell>0} \left\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\right\} (G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})). \quad (4.9)$$

Notice that if $\widehat{F}$ is an absolutely continuous distribution with a bounded density, then for any sequence of nonempty sets such that $\mathcal{A}_k \searrow \{1\}$, it holds that for every $\epsilon > 0$, we can find $k_0 \in \mathbb{N}$ such that $\sup_{\ell>0}\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\} < \epsilon(1-\rho)/2\rho$ for all $k > k_0$. Similarly, we can find $\xi(m_0) \in \mathbb{N}$ large enough such that $1 - G_m(A_{k+1}) \leq \epsilon(1-\rho)/2\rho$. Putting together this results, we obtain that for all $k \geq k_0$ and $m \geq m_0$

$$\left|\psi_{\widehat{F}}(u) - \psi_{\widehat{F}\star G_m}(u)\right| \leq \frac{\rho}{1-\rho}\left[\sup_{\ell>0}\left\{\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)\right\} + (1 - G_m(\mathcal{A}_{k+1}))\right] = \epsilon.$$

Hence, uniform convergence follows. □

In our numerical experiments, we found that in order to obtain a sharp numerical upper bound it is enough to take a small number of summands in (4.8), say, up to the K-th term. This is equivalent to take $\mathcal{A}_k = \{1\}$ for all $k \geq K$.

4.2.2 Bounds for the discretization error of approximation A

Next, we address the construction of a bound for the discretization error $\left|\psi_{\widehat{F}\star G_m}(u) - \psi_{\Pi\star G_m}(u)\right|$. The following theorem makes use of (4.4) for the construction of an upper bound for the discretization error.

Theorem 4.5. *Let*

$$\eta^A := \sup_{0\leq s\leq u} \left\{\left|\widehat{F}\star G_m(s) - \Pi\star G_m(s)\right|\right\}.$$

Then for all $0 < \delta < \infty$ it holds that

$$\left|\psi_{\widehat{F}\star G_m}(u) - \psi_{\Pi\star G_m}(u)\right| \leq \frac{\eta^A(1-\rho)\rho}{\left(1-\rho\left(\widehat{F}(u/\delta) + G_m(\delta)\right)\right)(1-\rho(\Pi(u/\delta) + G_m(\delta)))}. \quad (4.10)$$

Proof. Apply the bound (4.4) provided in Theorem 4.1, so that

$$\left|\psi_{\widehat{F}\star G_m}(u) - \psi_{\Pi\star G_m}(u)\right| \leq \frac{\eta^A(1-\rho)\rho}{\left(1-\rho\widehat{F}\star G_m(u)\right)(1-\rho\Pi\star G_m(u))}.$$

The result follows from observing that

$$\widehat{F}\star G_m(u) = \int_0^{\delta} \widehat{F}(u/s)\mathrm{d}G_m(s) + \int_{\delta}^{\infty} \widehat{F}(u/s)\mathrm{d}G_m(s) \leq G_m(\delta) + \widehat{F}(u/\delta).$$

An upper bound for $\Pi\star G_m(u)$ can be found in an analogous way. □

The last step in the construction of an upper bound for the discretization error is finding an upper bound for η^A. We suggest the bound in Proposition 4.6 below.

Proposition 4.6. *For $0 < \delta < \infty$, we have that $\eta^A \leq \eta^A(\delta)$ where*

$$\eta^A(\delta) = \sup_{u/\delta \leq s < \infty} \left\{ \left| \widehat{F}(s) - \Pi(s) \right| \right\} G_m(\delta) + \sup_{0 < s \leq u/\delta} \left\{ \left| \widehat{F}(s) - \Pi(s) \right| \right\} \overline{G}_m(\delta). \tag{4.11}$$

Proof.

$$\begin{aligned}
\left| \widehat{F} \star G_m(u) - \Pi \star G_m(u) \right| &= \left| \int_0^\infty \widehat{F}(u/s) \mathrm{d}G_m(s) - \int_0^\infty \Pi(u/s) \mathrm{d}G_m(s) \right| \\
&\leq \int_0^\infty \left| \widehat{F}(u/s) - \Pi(u/s) \right| \mathrm{d}G_m(s) \\
&\leq \int_0^\delta \left| \widehat{F}(u/s) - \Pi(u/s) \right| \mathrm{d}G_m(s) + \int_\delta^\infty \left| \widehat{F}(u/s) - \Pi(u/s) \right| \mathrm{d}G_m(s) \\
&\leq \sup_{u/\delta \leq s < \infty} \left\{ \left| \widehat{F}(s) - \Pi(s) \right| \right\} G_m(\delta) + \sup_{0 < s \leq u/\delta} \left\{ \left| \widehat{F}(s) - \Pi(s) \right| \right\} \overline{G}_m(\delta).
\end{aligned}$$

□

The bound $\eta^A(\delta)$ given in Equation (4.11) is easily computed for most cases of practical interest. In particular, if Π is a discretization of $\widehat{F}$, then it is only necessary to search for the supremum in the support of the distribution Π.

Note as well that the upper bound (4.10) in Theorem 4.5 decreases as Π gets closer to $\widehat{F}$ which is reflected in the value of $\eta^A(\delta)$. In addition, the bound (4.10) will become smaller as long as the terms $\widehat{F}(u/\delta) + G_m(\delta)$ and $\Pi(u/\delta) + G_m(\delta)$ in the denominator become bigger while the value of $\eta^A(\delta)$ becomes smaller. However, there is a trade-off since the quantities involved in these terms are inversely related to each other. In fact, the value of δ minimizing this overall bound can be easily found numerically with most pre-built optimization algorithms in standard packages. Notice as well that if the tail probability of $\widehat{F}$ is well approximated by Π, then the error bound will in general decrease. This suggests that Π should provide a good approximation of $\widehat{F}$ particularly in the tail in order to reduce effectively the error of approximation.

4.3 Error bounds for approximation B

Next, we turn our attention to approximation B of the ruin probability when the claim size distribution F is approximated via Erlangized scale mixtures. As with approximation A, an upper bound for the theoretical error of approximation B is found as the aggregation of the Erlangization error B (see Subsection 4.3.1) plus the discretization error B (see Subsection 4.3.2). We remark that the bounds presented in this section are simple and sufficient to show uniform convergence. However, these bounds can be too rough when the value of ρ is close to 1. An alternative bound is given in Subsection 8.1 in the appendix (condensed in Theorem 8.3); the construction of this alternative bound is more complicated and it delivers a significant improvement only for small values of ρ.

4.3.1 Bounds for the Erlangization error of approximation B

The construction of a tight bound for the Erlangization error for approximation B turns out to be more involved. Theorem 4.7 below provides a first bound for the Erlangization error of approximation B. We remark that in comparable cases, the bound proposed here is tighter than the equivalent bound found in Santana et al. (2016).

Theorem 4.7.

$$\left| \psi_{H_F \star U}(u) - \psi_{H_F \star \widehat{G}_m}(u) \right| \leq \frac{\rho \left(1 - \widehat{G}_m(1) \right)}{1 - \rho \widehat{G}_m(1)} \leq \frac{\rho}{1-\rho} \sqrt{\frac{1}{2\pi\xi}}$$

where $\xi := \xi(m)$ is the parameter of the Erlang distribution G_m. Moreover $\psi_{H_F \star U}$ converges uniformly to $\psi_{H_F \star \widehat{G}_m}$ as $m \to \infty$.

Proof. Since $G_m(s) \to \mathbb{I}_{[1,\infty)}(s)$ for all $s \neq 1$ so

$$\widehat{g}_m(s) := \frac{\mathrm{d}}{\mathrm{d}s}\widehat{G}_m(s) = 1 - G_m(s) \to \mathbb{I}_{[0,1)}(s), \qquad \forall s \neq 1.$$

Let $\{X_n'\}$ be a sequence of independent and identically H_F distributed random variables. Then, by Propositions 2.1 and 2.3,

$$\begin{aligned}
\left|\widehat{F}^{*n}(u) - \widehat{F \star G}^{*n}(u)\right| &= \left|(H_F \star U)^{*n}(u) - \left(H_F \star \widehat{G}_m\right)^{*n}(u)\right| \\
&= \left| \int_{\mathbb{R}^n} \cdots \int \mathbb{P}(s_1X_1' + \cdots + s_nX_n' \leq u) U(\mathrm{d}s_1) \ldots U(\mathrm{d}s_n) \right. \\
&\qquad \left. - \int_{\mathbb{R}^n} \cdots \int \mathbb{P}(s_1X_1' + \cdots + s_nX_n' \leq u) \widehat{G}_m(\mathrm{d}s_1) \ldots \widehat{G}_m(\mathrm{d}s_n) \right| \\
&\leq \left| \int_{\mathbb{R}^n} \cdots \int \mathbb{P}(s_1X_1' + \cdots + s_nX_n' \leq u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right| \\
&= \left| \int_{[0,1]^n} \cdots \int \mathbb{P}(s_1X_1' + \cdots + s_nX_n' \leq u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right. \\
&\qquad \left. - \int_{\mathbb{R}^n \setminus [0,1]^n} \cdots \int \mathbb{P}(s_1X_1' + \cdots + s_nX_n' \leq u) \left(\prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right| \\
&\leq \sup \left\{ \int_{[0,1]^n} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \ , \ \int_{\mathbb{R}^n \setminus [0,1]^n} \cdots \int \prod_{i=1}^n \widehat{g}_m(s_i) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right\}
\end{aligned}$$

That the last is equal to $1 - \widehat{G}_m^n(1)$ follows from Lemma 8.1 in the appendix. Therefore, we have that

$$\begin{aligned}
\left|\psi_{H_F \star U}(u) - \psi_{H_F \star \widehat{G}_m}(u)\right| &\leq \sum_{n=1}^{\infty} (1-\rho)\rho^n \left|\widehat{F}^{*n}(u) - (\widehat{F \star G_m})^{*n}(u)\right| \\
&\leq \sum_{n=1}^{\infty} (1-\rho)\rho^n \left(1 - \widehat{G}_m^n(1)\right) \\
&= \rho \left(1 - \frac{(1-\rho)\widehat{G}_m(1)}{1 - \rho\widehat{G}_m(1)}\right) = \frac{\rho\left(1 - \widehat{G}_m(1)\right)}{1 - \rho\widehat{G}_m(1)} \leq \frac{\rho}{1-\rho}\sqrt{\frac{1}{2\pi\xi}}.
\end{aligned}$$

The last inequality follows from Lemma 4.8 below, which provides an explicit expression as well as an upper bound for $1 - \widehat{G}_m(1)$. Uniform convergence follows by noting that the last bound converges to 0 as $\xi \to \infty$. □

Notice that the bound proposed depends on the parameter of the Erlang distribution ξ and the average claim amount per unit of time ρ only. It does not depend on the initial reserve u, nor the underlying claim size distribution F. However, this bound is too rough for values of ρ which are close to 1, and in such cases it is of little practical use. We attempted to construct various alternative bounds with the aim of obtaining a tighter bound. One of these bounds can be found in Theorem 8.3 in the appendix. Such an alternative bound can provide significant improvements for values of ρ close to 0, but it remains rough for values of ρ close to 1. This notorious difficulty in implementing an upper bound for the Erlangization error of approximation B further underlines the superiority of approximation A.

We close this subsection with the following technical result, used in the proof of Theorem 4.7, which provides an explicit expression of the integrated distance between the survival function $1-G_m(1)$ and the density of a U$(0,1)$ distribution.

Lemma 4.8. *Let $G_m \sim$ Erlang(ξ,ξ). Then*

$$1-\widehat{G}_m(1)=\int_1^\infty (1-G_m(s))\,ds=\int_0^1 G_m(s)ds=\mathrm{e}^{-\xi}\frac{\xi^{\xi-1}}{(\xi-1)!}\leq(2\pi\xi)^{-\frac{1}{2}}.$$

Proof. Firstly observe that $\mu_{G_m}=1$, it follows that $1-G_m$ is the density of the integrated tail distribution $\widehat{G}_m$. Hence,

$$1-\widehat{G}_m(1)=\int_1^\infty (1-G_m(s))\,\mathrm{d}s=1-\int_0^1 (1-G_m(s))\,\mathrm{d}s=\int_0^1 G_m(s)\mathrm{d}s,$$

and the second equality follows. For the third equality, we have that

$$\begin{aligned}
\int_0^1 G_m(s)\mathrm{d}s&=\int_0^1\left(1-\sum_{n=0}^{\xi-1}\frac{1}{n!}\mathrm{e}^{-\xi s}(\xi s)^n\right)\mathrm{d}s=1-\sum_{n=0}^{\xi-1}\frac{1}{n!}\int_0^1\mathrm{e}^{-\xi s}(\xi s)^n\mathrm{d}s\\
&=1-\sum_{n=0}^{\xi-1}\frac{1}{n!}\left(n!\xi^{-1}-\mathrm{e}^{-\xi}\sum_{k=0}^{n}\frac{n!\xi^{k-1}}{k!}\right)=\mathrm{e}^{-\xi}\sum_{n=0}^{\xi-1}\sum_{k=0}^{n}\frac{\xi^{k-1}}{k!}\\
&=\mathrm{e}^{-\xi}\sum_{k=0}^{\xi-1}(\xi-k)\frac{\xi^{k-1}}{k!}=\mathrm{e}^{-\xi}\left(\sum_{k=0}^{\xi-1}\frac{\xi^k}{k!}-\sum_{k=0}^{\xi-1}k\frac{\xi^{k-1}}{k!}\right)\\
&=\mathrm{e}^{-\xi}\left(\sum_{k=0}^{\xi-1}\frac{\xi^k}{k!}-\sum_{k=1}^{\xi-1}\frac{\xi^{k-1}}{(k-1)!}\right)=\mathrm{e}^{-\xi}\left(\sum_{k=0}^{\xi-1}\frac{\xi^k}{k!}-\sum_{k=0}^{\xi-2}\frac{\xi^k}{k!}\right)=\mathrm{e}^{-\xi}\frac{\xi^{\xi-1}}{(\xi-1)!}.
\end{aligned}$$

Finally, an application of Stirling's formula $\xi!>\sqrt{2\pi}\xi^{\xi+\frac{1}{2}}\mathrm{e}^{-\xi}$ yields $1-\widehat{G}_m(1)<(2\pi\xi)^{-\frac{1}{2}}$. □

4.3.2 Bounds for the discretization error of approximation B

Finally, we address the construction of a bound for the discretization error of approximation B. The next two results are analogous to the ones in Subsection 4.2.2 and presented without proof.

Theorem 4.9. *Let*

$$\eta^B:=\sup_{0\leq s\leq u}\left\{\left|H_F\star\widehat{G}_m(s)-H_\Pi\star\widehat{G}_m(s)\right|\right\}$$

then

$$\left|\psi_{H_F\star\widehat{G}_m}(u)-\psi_{H_\Pi\star\widehat{G}_m}(u)\right|\leq\frac{\eta^B\cdot(1-\rho)\rho}{\left(1-\rho H_F(u/\delta)\left(1-\widehat{G}_m(\delta)\right)\right)\left(1-\rho H_\Pi(u/\delta)\left(1-\widehat{G}_m(\delta)\right)\right)}.$$

An upper bound for η^B is suggested in the next Proposition.

Proposition 4.10. *For $\delta>1$, we have that $\eta^B\leq\eta^B(\delta)$ where*

$$\eta^B(\delta):=\sup_{u/\delta\leq s<\infty}\left|\overline{H}_\Pi(s)-\overline{H}_F(s)\right|\widehat{G}_m(\delta)+\sup_{0<s\leq u/\delta}\left|H_F(s)-H_\Pi(s)\right|\left(1-\widehat{G}_m(\delta)\right).$$

The calculation of the bound $\eta^B(\delta)$ requires the supremum of the distances between the moment distributions $|H_F-H_\Pi|$ in a certain compact set, but an explicit expression for the later might not always be available. For this case, we suggest a bound for such a quantity in Lemma 8.6 in the appendix which can be implemented for a specific type of approximating distributions Π described in there. This bound depends on the cdf of the distribution H_F, the restricted expected value of the claim size distribution F and its approximation Π.

5 Bounds for the numerical error of approximation

The approximations of the probability of ruin given in Theorems 3.4 and 3.5 are not computable in exact form since the expressions are given in terms of various infinite series. In practice, we can compute enough terms and then truncate the series at a level where the error of truncation is smaller than some desired precision. Since all terms involved are positive, such an approximation will provide an underestimate of the real ruin probability. In this section, we provide upper bounds for the errors incurred by truncating the infinite series, for both approximations A and B.

A close inspection of Theorems 3.4 and 3.5 reveals that there exist two sources of error due to truncation. The first source of error comes from truncating at some level N_1 in both the series from Equation (3.5) in the case of approximation A, and the series from Equation (3.6) in the case of approximation B. We call N_1 the *Poisson level of truncation*. Recall that both approximations can be seen as the expected value of κ_N where $N \sim \text{Poisson}(\xi u/s_1)$, so the associated error of truncation is $\mathbb{E}[\kappa_N : N \geq N_1]$. Since the values of κ_n are bounded from above by 1, then it is possible to bound this error with $\mathbb{P}(N \geq N_1)$ and use Chernoff's bound (cf. Theorem 9.3 Billingsley, 1995) to obtain an explicit expression. Specifically, if $N \sim \text{Poisson}(\lambda)$ and $N_1 > \lambda$, then Chernoff's bound is given by

$$1 - \zeta(N_1; \lambda) = \mathbb{P}(N > N_1) \leq \frac{\mathrm{e}^{-\lambda}(\mathrm{e} \cdot \lambda)^{N_1+1}}{(N_1+1)^{N_1+1}}, \tag{5.1}$$

where $\zeta(N_1; \lambda)$ denotes the cdf of a Poisson distribution with parameter λ evaluated at N_1. The upper bound for the numerical error associated to the Poisson level of truncation is the same for both approximations A and B.

The second source of numerical error comes from truncating the infinite series induced by the scaling distribution Π. This implies that we need to truncate the infinite series defining the terms $\mathcal{B}_i$, $\mathcal{C}_i$ and $\mathcal{D}_i$ at some level N_2 that we call the *scaling level of truncation*. The following lemma shows that the truncated series for $\mathcal{B}_i$, $\mathcal{C}_i$ and $\mathcal{D}_i$ can be bounded above by quantities depending on the tail probability $1 - \Pi(s_{N_2})$.

Lemma 5.1. *Let* $\varepsilon_1 = 1 - \Pi(s_{N_2})$ *and* $\varepsilon_2 = \int_{[s_{N_2+1}, \infty)} r \, \mathrm{d}\Pi(r)$. *Then*

$$\begin{aligned} \mathcal{B}_i - \widetilde{\mathcal{B}}_i &\leq \varepsilon_1, & 0 &\leq i\,, \\ \mathcal{C}_n - \widetilde{\mathcal{C}}_n &\leq \varepsilon_1, & \xi &\leq n\,, \\ \mathcal{D}_n - \widetilde{\mathcal{D}}_n &\leq \varepsilon_2, & \xi &\leq n\,, \end{aligned}$$

where $\widetilde{\mathcal{B}}_i$, $\widetilde{\mathcal{C}}_n$ *and* $\widetilde{\mathcal{D}}_n$ *denote to the truncated series at the level* N_2.

Proof. If $0 \leq i < \xi - 1$ then $\mathcal{B}_i = \widetilde{\mathcal{B}}_i = 0$, otherwise if $\xi - 1 \leq i \leq N_2$ then

$$\begin{aligned} \mathcal{B}_i - \widetilde{\mathcal{B}}_i &= \frac{\xi}{i+1} \sum_{j=N_2+1}^{\infty} \frac{\pi_j}{1 - s_1/s_j} \text{nbin}(\xi; i - \xi + 2, s_1/s_j) \\ &= \frac{\xi}{i+1} \sum_{j=N_2+1}^{\infty} \pi_j \text{bin}(\xi; i+1, s_1/s_j) \leq \sum_{j=N_2+1}^{\infty} \pi_j = \varepsilon_1. \end{aligned}$$

Similarly, if $n \geq \xi$ then

$$\begin{aligned} \mathcal{C}_n - \widetilde{\mathcal{C}}_n &= \sum_{j=N_2+1}^{\infty} \frac{\pi_j}{1 - s_1/s_j} \text{NBin}(\xi - 1; n - \xi + 2, s_1/s_j) \\ &= \sum_{j=N_2+1}^{\infty} \pi_j \text{Bin}(\xi - 1; n, s_1/s_j) \leq \sum_{j=N_2+1}^{\infty} \pi_j = \varepsilon_1, \end{aligned}$$

while

$$\mathcal{D}_n - \widetilde{\mathcal{D}}_n = \sum_{j=N_2+1}^{\infty} \frac{\pi_j s_j}{1-s_1/s_j} \sum_{k=0}^{\xi-1} \frac{\xi-k}{\xi} \mathrm{nbin}(k; n-k+1, s_1/s_j)$$
$$= \sum_{j=N_2+1}^{\infty} \pi_j s_j \sum_{k=0}^{\xi-1} \frac{\xi-k}{\xi} \mathrm{bin}(k; n, s_1/s_j) \leq \sum_{j=N_2+1}^{\infty} \pi_j s_j = \varepsilon_2.$$

Here, $\mathrm{bin}(\cdot; n, p)$ and $\mathrm{Bin}(\cdot; n, p)$ denote the pdf and cdf respectively of a binomial distribution with parameter n and p. □

5.1 Truncation error bounds for approximation A

We start by writing the expression for the ruin probability in Theorem 3.4 (approximation A) as

$$\widetilde{\psi}_{\Pi\star G_m}(u) = \mathrm{e}^{-\xi u/s_1} \sum_{n=0}^{N_1} \widetilde{\kappa}_n^A \frac{(\xi u)^n}{s_1^n n!},$$

where

$$\widetilde{\kappa}_n^A = \begin{cases} \gamma\mu_F, & 0 \leq n \leq \xi-1, \\ \gamma\mu_F \left(\sum_{i=\xi-1}^{n-1} \widetilde{\kappa}_{n-1-i}^A \widetilde{\mathcal{B}}_i + \widetilde{\mathcal{C}}_n \right), & \xi \leq n \leq N_1, \end{cases}$$

with

$$\widetilde{\mathcal{B}}_i = \frac{\xi}{i+1} \sum_{j=1}^{N_2} \frac{\pi_j}{1-s_1/s_j} \mathrm{nbin}(\xi; i-\xi+2, s_1/s_j),$$
$$\widetilde{\mathcal{C}}_n = \sum_{j=1}^{N_2} \frac{\pi_j}{1-s_1/s_j} \mathrm{NBin}(\xi-1; n-\xi+2, s_1/s_j).$$

Theorem 5.2. *Let $\varepsilon_1 = 1 - \Pi(s_{N_2})$. Then*

$$\psi_{\Pi\star G_m}(u) - \widetilde{\psi}_{\Pi\star G_m}(u) \leq \varepsilon_1 \left[\frac{\gamma\mu_F}{1-\gamma\mu_F} \left(\frac{\xi u}{s_1} \right) + \frac{2}{(1-\gamma\mu_F)^2} \mathrm{e}^{-(1-\gamma\mu_F)\xi u/s_1} \right] + \left(1 - \zeta(N_1; \xi u/s_1)\right),$$

where $\zeta(N_1; \xi u/s_1)$ denotes the cdf of a Poisson with parameter $\xi u/s_1$ and evaluated at N_1.

Proof. Observe that

$$\psi_{\Pi\star G_m}(u) - \widetilde{\psi}_{\Pi\star G_m}(u) = \mathrm{e}^{-\xi u/s_1} \sum_{n=0}^{N_1} \left(\kappa_n^A - \widetilde{\kappa}_n^A\right) \frac{(\xi u)^n}{s_1^n n!} + \mathrm{e}^{-\xi u/s_1} \sum_{n=N_1+1}^{\infty} \kappa_n^A \frac{(\xi u)^n}{s_1^n n!}. \tag{5.2}$$

Firstly, we consider the second term in the right hand side of (5.2). Using $\kappa_n^A \leq 1$, we obtain that if $N_1 > \xi u/s_1 - 1$, then

$$\mathrm{e}^{-\xi u/s_1} \sum_{n=N_1+1}^{\infty} \kappa_n^A \frac{(\xi u)^n}{s_1^n n!} \leq \sum_{n=N_1+1}^{\infty} \mathrm{e}^{-\xi u/s_1} \frac{(\xi u)^n}{s_1^n n!} = \left(1 - \zeta(N_1; \xi u/s_1)\right).$$

Next, we look into the first term of Equation (5.2) and observe that

$$\kappa_n^A - \widetilde{\kappa}_n^A = \begin{cases} 0, & 0 \leq n \leq \xi-1, \\ \gamma\mu_F \left[\sum_{i=\xi-1}^{n-1} \left(\kappa_{n-1-i}^A \mathcal{B}_i - \widetilde{\kappa}_{n-1-i}^A \widetilde{\mathcal{B}}_i \right) + \mathcal{C}_n - \widetilde{\mathcal{C}}_n \right], & \xi \leq n \leq N_1. \end{cases}$$

Notice that if $n \geq \xi$, we can rewrite

$$\sum_{i=\xi-1}^{n-1} \left(\kappa_{n-1-i}^A \mathcal{B}_i - \widetilde{\kappa}_{n-1-i}^A \widetilde{\mathcal{B}}_i\right) = \sum_{i=\xi-1}^{n-1} \left(\left(\kappa_{n-1-i}^A - \widetilde{\kappa}_{n-1-i}^A\right) \mathcal{B}_i + \widetilde{\kappa}_{n-1-i}^A \left(\mathcal{B}_i - \widetilde{\mathcal{B}}_i\right)\right).$$

Since $0 < \widetilde{\kappa}_i^A \leq \kappa_i^A \leq 1$ for $0 \leq i$ then we can use the first part of Lemma 5.1 to obtain the following bound of the expression above

$$\sum_{i=\xi-1}^{n-1} \left(\kappa_i^A - \widetilde{\kappa}_i^A\right) \mathcal{B}_i + (n - \xi + 1)\varepsilon_1. \tag{5.3}$$

Putting (5.3) and the second part of Lemma 5.1 together, we arrive at

$$\begin{aligned}
\kappa_n^A - \widetilde{\kappa}_n^A &\leq \gamma\mu_F \left[\sum_{i=\xi-1}^{n-1} \left(\kappa_i^A - \widetilde{\kappa}_i^A\right) \mathcal{B}_i + (n - \xi + 1)\varepsilon_1 + \varepsilon_1\right] \\
&\leq \gamma\mu_F \left[\sup_{\xi-1 \leq i < n-1} \left\{\kappa_i^A - \widetilde{\kappa}_i^A\right\} \sum_{i=\xi-1}^{\infty} \mathcal{B}_i + (n - \xi + 2)\varepsilon_1\right] \\
&\leq \gamma\mu_F \left[\sup_{\xi-1 \leq i < n-1} \left\{\kappa_{i-1}^A - \widetilde{\kappa}_{i-1}^A\right\} + (n - \xi + 2)\varepsilon_1\right].
\end{aligned}$$

Note that $\sum_{i=\xi-1}^{\infty} \mathcal{B}_i = 1$ follows from relating the formula of $\mathcal{B}_i$ to the probability mass function of a negative binomial distribution $\text{nbin}(i - \xi + 1; \xi, 1 - s_1/s_j)$. Using the hypothesis that $\gamma\mu_F < 1$ and induction, we can prove that

$$\kappa_n^A - \widetilde{\kappa}_n^A \leq \varepsilon_1 \sum_{i=2}^{n-\xi+2} i(\gamma\mu_F)^{n-\xi+3-i} \leq \varepsilon_1 \left[\frac{\gamma\mu_F}{1 - \gamma\mu_F} n + \frac{2}{(1 - \gamma\mu_F)^2} (\gamma\mu_F)^n\right]. \tag{5.4}$$

The first inequality in (5.4) is proved by induction while the second inequality is obtained in a straightforward way using geometric progressions. Details of the induction are given next.

We establish the base of induction with $n = \xi$; in such a case $\kappa_\xi^A - \widetilde{\kappa}_\xi^A = \varepsilon_1\gamma\mu_F$ and it is straightforward to check that the inequality is satisfied. Next, we assume that the inequality holds for all $\xi < n < N_1$, that is

$$\kappa_n^A - \widetilde{\kappa}_n^A \leq \varepsilon_1 \sum_{i=2}^{n-\xi+2} i(\gamma\mu_F)^{n-\xi+3-i}.$$

It remains to prove that the inequality also holds for $n + 1$:

$$\begin{aligned}
\kappa_{n+1}^A - \widetilde{\kappa}_{n+1}^A &\leq \gamma\mu_F \left[\sup_{\xi-1 \leq i < j} \left\{\kappa_{i-1}^A - \widetilde{\kappa}_{i-1}^A\right\} + (j + 1 - \xi + 2)\varepsilon_1\right] \\
&\leq \gamma\mu_F \left[\varepsilon_1 \sum_{i=2}^{j-\xi+2} i(\gamma\mu_F)^{j-\xi+3-i} + (j - \xi + 3)\varepsilon_1\right] \\
&= \varepsilon_1 \sum_{i=2}^{j-\xi+3} i(\gamma\mu_F)^{j-\xi+4-i}.
\end{aligned}$$

Thus, the first inequality holds for all $\xi \leq n \leq N_1$ by induction.

Inserting (5.4) into the first term of Equation (5.2) and assuming that $\xi u/s_1 > 1$, we arrive at

$$\mathrm{e}^{-\xi u/s_1} \sum_{n=0}^{N_1} (\kappa_n - \widetilde{\kappa}_n) \frac{(\xi u)^n}{s_1^n n!} \leq \varepsilon_1 \left[\frac{\gamma\mu_F}{1 - \gamma\mu_F} \left(\frac{\xi u}{s_1}\right) + \frac{2}{(1 - \gamma\mu_F)^2} \mathrm{e}^{-(1-\gamma\mu_F)\xi u/s_1}\right].$$

□

Remark 5.3. The term $(1 - \zeta(N_1; \xi u/s_1))$ in Theorem 5.2 can be bounded using Chernoff's bound

$$1 - \zeta(N_1; \xi u/s_1) \leq \frac{\mathrm{e}^{-\xi u/s_1} (\mathrm{e}^1 \cdot \xi \cdot u/s_1)^{N_1+1}}{(N_1 + 1)^{N_1+1}}.$$

5.2 Truncation error bounds for approximation B

We proceed in an analogous way as in the previous section. We write the ruin probability in Theorem 3.5 (approximation B) as a truncated series:

$$\widetilde{\psi}_{H_\Pi \star \widehat{G}_m}(u) = \sum_{n=0}^{N_1} \widetilde{\kappa}_n^B \frac{(\xi u/s_1)^n \mathrm{e}^{-\xi u/s_1}}{n!},$$

where

$$\widetilde{\kappa}_n^B = \begin{cases} \gamma\mu_F, & n = 0, \\ (\gamma\mu_F - 1)\left(1 + \dfrac{\gamma\mu_F s_1}{\mu_\Pi \xi}\right)^n + 1, & 1 \le n \le \xi, \\ \dfrac{\gamma\mu_F s_1}{\mu_\Pi \xi} \displaystyle\sum_{i=0}^{n-1} \widetilde{\kappa}_{n-1-i}^B \widetilde{\mathcal{C}}_i + \dfrac{\gamma\mu_F}{\mu_\Pi} \widetilde{\mathcal{D}}_n, & \xi < n. \end{cases}$$

with

$$\widetilde{\mathcal{C}}_i = \sum_{j=1}^{N_2} \frac{\pi_j}{1 - s_1/s_j} \mathrm{NBin}(\xi - 1; i - \xi + 2, s_1/s_j),$$

$$\widetilde{\mathcal{D}}_n = \sum_{j=1}^{N_2} \frac{\pi_j s_j}{1 - s_1/s_j} \sum_{k=0}^{\xi-1} \frac{\xi - k}{\xi} \mathrm{nbin}(k; n - k + 1, s_1/s_j).$$

The following theorem provides an uppper bound for the truncation error of approximation B.

Theorem 5.4. *Let* $\varepsilon_2 = \int_{[s_{N_2+1}, \infty)} s \, \mathrm{d}\Pi(s)$ *and* N_2 *such that* $s_{N_2+1} \le 1$. *Then*

$$\psi_{H_\Pi \star \widehat{G}_m}(u) - \widetilde{\psi}_{H_\Pi \star \widehat{G}_m}(u) \le \frac{\varepsilon_2}{\mu_\Pi} \left[\frac{\gamma\mu_F}{1 - \gamma\mu_F} \left(\frac{\xi u}{s_1} \right) + \frac{(\gamma\mu_F)^{2-\xi}}{(1 - \gamma\mu_F)^2} \mathrm{e}^{-(1-\gamma\mu_F)\xi u/s_1} \right] + (1 - \zeta(N_1; \xi u/s_1)).$$

Proof. Observe that

$$\kappa_n^B - \widetilde{\kappa}_n^B = \begin{cases} 0, & 0 \le n \le \xi, \\ \dfrac{\gamma\mu_F}{\mu_\Pi} \left[\dfrac{s_1}{\xi} \displaystyle\sum_{i=\xi+1}^{n-1} \left(\kappa_{n-1-i}^B \mathcal{C}_i - \widetilde{\kappa}_{n-1-i}^B \widetilde{\mathcal{C}}_i \right) + \mathcal{D}_n - \widetilde{\mathcal{D}}_n \right], & \xi < n \le N_1. \end{cases} \tag{5.5}$$

The summation in (5.5) can be rewritten as

$$\sum_{i=\xi+1}^{n-1} \left(\kappa_{n-1-i}^B \mathcal{C}_i - \widetilde{\kappa}_{n-1-i}^B \widetilde{\mathcal{C}}_i \right) = \sum_{i=\xi+1}^{n-1} \left[\left(\kappa_{n-1-i}^B - \widetilde{\kappa}_{n-1-i}^B \right) \mathcal{C}_i + \widetilde{\kappa}_{n-1-i}^B \left(\mathcal{C}_i - \widetilde{\mathcal{C}}_i \right) \right].$$

Since $0 < \widetilde{\kappa}_i^B \le \kappa_i^B \le 1$ for $0 \le i$ then we can use the second part of Lemma 5.1 to obtain the following bound of the expression above

$$\sup_{\xi+1 \le i \le n-1} \left\{ \kappa_i^B - \widetilde{\kappa}_i^B \right\} \sum_{i=\xi+1}^{\infty} \mathcal{C}_i + (n - \xi - 1)\varepsilon_2 \le \sup_{\xi+1 \le i \le n-1} \left\{ \kappa_i^B - \widetilde{\kappa}_i^B \right\} \frac{\xi\mu_\Pi}{s_1} + (n - \xi - 1)\varepsilon_2. \tag{5.6}$$

Note that $\varepsilon_1 \le \varepsilon_2$ (because the hypothesis $s_{N_2+1} \ge 1$), while $\sum_{i=\xi+1}^{\infty} \mathcal{C}_i \le \dfrac{\xi\mu_\Pi}{s_1}$ follows from relating the formula of $\mathcal{C}_i$ to the cdf of a negative binomial distribution $\mathrm{NBin}(i - k; k + 1, 1 - s_1/s_j)$:

$$\begin{aligned}\sum_{i=\xi+1}^{\infty}\mathcal{C}_i &= \sum_{j=1}^{\infty}\pi_j\sum_{k=0}^{\xi-1}\sum_{i=\xi+1}^{\infty}\binom{i}{k}\left(1-\frac{s_1}{s_j}\right)^{i-k}\left(\frac{s_1}{s_j}\right)^{k}\\ &\leq \sum_{j=1}^{\infty}\frac{\pi_j s_j}{s_1}\sum_{k=0}^{\xi-1}\sum_{i=k}^{\infty}\binom{i}{i-k}\left(1-\frac{s_1}{s_j}\right)^{i-k}\left(\frac{s_1}{s_j}\right)^{k+1}\\ &= \sum_{j=1}^{\infty}\frac{\pi_j s_j}{s_1}\sum_{k=0}^{\xi-1}\mathrm{NBin}(i-k;k+1,1-s_1/s_j)\\ &= \frac{\xi}{s_1}\sum_{j=1}^{\infty}\pi_j s_j = \frac{\xi\mu_\Pi}{s_1}.\end{aligned}$$

Putting (5.6) and the third part of Lemma 5.1 together, we arrive at

$$\begin{aligned}\kappa_n^B-\widetilde{\kappa}_n^B &\leq \frac{\gamma\mu_F}{\mu_\Pi}\left[\frac{s_1}{\xi}\sup_{\xi+1\leq i\leq n-1}\left\{\kappa_i^B-\widetilde{\kappa}_i^B\right\}\frac{\xi\mu_\Pi}{s_1}+(n-\xi)\varepsilon_2\right]\\ &= \frac{\gamma\mu_F}{\mu_\Pi}\left[\mu_\Pi\sup_{\xi+1\leq i\leq n-1}\left\{\kappa_i^B-\widetilde{\kappa}_i^B\right\}+(n-\xi)\varepsilon_2\right].\end{aligned}$$

Next, we proceed by induction. First we establish the base of induction with $n=\xi+1$. In this case it easily holds that

$$\kappa_{\xi+1}^B-\widetilde{\kappa}_{\xi+1}^B<\frac{\varepsilon_2}{\mu_\Pi}\gamma\mu_F.$$

Then we set the hypothesis of induction and assume that the following inequality holds for any $\xi+1<n$:

$$\kappa_n^B-\widetilde{\kappa}_n^B\leq\frac{\varepsilon_2}{\mu_\Pi}\sum_{i=1}^{n-\xi}i(\gamma\mu_F)^{n-\xi+1-i}.$$

It remains to prove that the same inequality holds for $n+1$. We see that

$$\begin{aligned}\kappa_{n+1}^B-\widetilde{\kappa}_{n+1}^B &\leq \frac{\gamma\mu_F}{\mu_\Pi}\left[\mu_\Pi\left(\kappa_n^B-\widetilde{\kappa}_n^B\right)+(n+1-\xi)\varepsilon_2\right]\\ &\leq \frac{\gamma\mu_F}{\mu_\Pi}\left[\varepsilon_2\sum_{i=1}^{n-\xi}i(\gamma\mu_F)^{n-\xi+1-i}+(n+1-\xi)\varepsilon_2\right]\\ &= \frac{\varepsilon_2}{\mu_\Pi}\left[\sum_{i=1}^{n-\xi}i(\gamma\mu_F)^{n-\xi+2-i}+(n+1-\xi)\gamma\mu_F\right]\\ &= \frac{\varepsilon_2}{\mu_\Pi}\sum_{i=1}^{n-\xi+1}i(\gamma\mu_F)^{n-\xi+2-i}.\end{aligned}$$

Thus, the hypothesis holds for all $\xi+1\leq n$ by induction. Then we get that

$$\kappa_n^B-\widetilde{\kappa}_n^B<\frac{\varepsilon_2}{\mu_\Pi}\sum_{i=1}^{n-\xi}i(\gamma\mu_F)^{n-\xi+1-i}\leq\frac{\varepsilon_2}{\mu_\Pi}\left[\frac{(\gamma\mu_F)^{n+2-\xi}}{(1-\gamma\mu_F)^2}+\frac{\gamma\mu_F}{1-\gamma\mu_F}n\right],$$

concluding that

$$\mathrm{e}^{-\xi u/s_1}\sum_{n=0}^{N_1}\left(\kappa_n^B-\widetilde{\kappa}_n^B\right)\frac{(\xi u)^n}{s_1^n n!}\leq\frac{\varepsilon_2}{\mu_\Pi}\left[\frac{\gamma\mu_F}{1-\gamma\mu_F}\left(\frac{\xi u}{s_1}\right)+\frac{(\gamma\mu_F)^{2-\xi}}{(1-\gamma\mu_F)^2}\mathrm{e}^{-(1-\gamma\mu_F)\xi u/s_1}\right].$$

□

Note that the bound found for the term $\sum_{i=1}^{\infty} \mathcal{C}_i$ is not as tight as the bound for the term $\sum_{i=1}^{\infty} \mathcal{B}_i$. As a consequence, the bound for the truncation error associated to the scaling of approximation B given in the previous theorem is not as tight as the bound for the truncation error of approximation A in Theorem 5.2. This aspect further highlights an additional advantage of approximation A over approximation B.

6 Numerical implementations

Next, we address the implementation issues of the two approximations for the ruin probability proposed in this paper. In Subsection 6.1, we discuss the approximation of a nonnegative distributions via Erlangized scale mixtures; in there, we focus on the selection of appropriate scaling distributions as well as the parameter of the Erlang distribution. In Subsection 6.2, we concentrate on the implementation of approximations A and B by considering a particular example where the claim sizes follow a Pareto distribution.

6.1 General considerations

Suppose we want to approximate a distribution F via Erlangized scale mixtures. For doing so, we need to select the parameter of the Erlang distribution G_m and the scaling discrete distribution Π. A critical aspect for approximating efficiently a heavy-tailed distribution is the selection of a scaling discrete distribution Π approximating either $\widehat{F}$ in the case of approximation A, or F in the case of approximation B. The selection can be made rather arbitrary, but in this paper, we suggest the following general family of discrete distributions:

Definition 6.1. *Let $W = \{w_i : i \in \mathbb{Z}^+\}$ and $\Omega_\Pi = \{s_i : i \in \mathbb{Z}^+\}$ be sets of strictly increasing nonnegative values such that $w_0 = s_0 = \inf\{s : F(s) > 0\}$ and for all $k \in \mathbb{N}$ it holds that*

$$s_k \leq w_k \leq s_{k+1}.$$

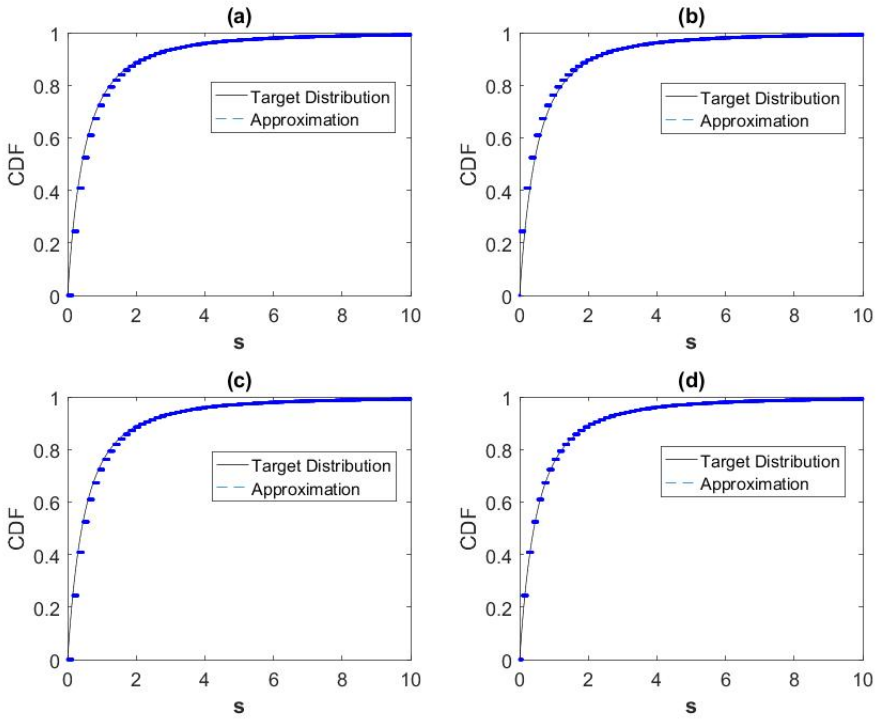
Then we define the distribution Π as

$$\Pi(s) := \sum_{k=0}^{\infty} F(w_k)\mathbb{I}_{[s_k, s_{k+1})}(s).$$

Notice that the distribution Π above corresponds to a discretized version of the target distribution F, which is supported over the set $\{s_k : k \in \mathbb{N}\}$ and upcrossed by F in every interval (s_{k-1}, s_k) exactly at the value w_k. This type of approximation is rather general since we can select various values of $w_k \in (s_{k-1}, s_k)$ to obtain different types of approximation. For instance, approximations from *below* can be obtained by setting $w_k = s_k$, while approximations from *above* are obtained by setting $w_k = s_{k+1}$; the middle point $w_k = (s_k + s_{k+1})/2$ is another reasonable possibility (see Figure 2).

Notice that if we work with distributions on the form of Definition 6.1, then due to truncation we effectively end up working with improper distributions with a finite support, that is, $\sum_{k=1}^{N_2} \pi_k < 1$. This represents an issue because selecting a low truncation level of N_2 will affect the quality of the approximation in the tail regions. Increasing the truncation level N_2 (computing a larger number of terms) is not often an efficient alternative since the computational effort increases rapidly and the algorithm becomes unfeasible. Hence, we aim at selecting collections of nonnegative numbers with finite cardinality $\{s_k : k = 1, \ldots, N_2\}$, and having *small* distances $\sup_x |\widehat{F}(x) - \Pi(x)|$ with respect to a target distribution $\widehat{F}$. Also, in order to obtain a correct description in the tails, it is desirable that the moments of Π should remain close to the moments of the original distribution. In our numerical experimentations, we found that a discretized distribution on the form of Definition 6.1 supported over an arithmetic progression will require a prohibitively large number of terms to obtain a sharp approximation in the tail. The best results were obtained by using geometric progressions as these can provide better approximations in the tails with a reduced number of terms. Also, the *lost mass* of the improper discretized distribution Π will be smaller if the finite progression is geometric, so for practical purposes this may be equivalent to work with a proper

Figure 2: In the panels below, we show the cdf $F(s)$ of the target Pareto(2) distribution together with some discretized approximating distributions. Panel (a) shows a general approximation. Panels (b) and (c) shows approximations from above and below respectively, while in the last panel (d) we have selected the middle point.



distribution. In addition, the smallest value s_1 in the support of the distribution Π will also play a role in the precision/speed of the algorithms. Further details about this particular issue will be highlighted below together with the selection of the parameters of the Erlang distribution.

The selection of the Erlang distribution G_m boils down to choose an appropriate $\xi \in \mathbb{N}$ so the bounds provided in Theorem 4.4 and Theorem 4.7 are smaller than a preselected precision. The larger the parameter ξ, the closer the distribution Π to the target distribution so we must expect better approximations of the ruin probability. However, we must recall that for implementing such approximations, we are limited to a finite computational budget, so it is not recommended to select a value ξ which is too large because this will result in a much slower algorithm. The reason for this slowdown is because larger values of ξ will increase the number of terms of the infinite series in Theorems 3.4 and 3.5 needed to achieve a desired precision. Notice that such infinite series can be seen as the expected value $\mathbb{E}[\kappa_N]$ where $N \sim \text{Poisson}(\xi u/s_1)$, so the total number N_1 of terms needed to provide an accurate approximation is directly related to the value $\xi u/s_1$ (larger values will require longer computational times). However, if we let N_2 to be fixed, then the selection of smaller values of ξ combined with larger values of s_1 the lower bound in the support of Π will result in increased errors of approximation, so there will be a natural trade-off between speed and precision in the selection of these values. In our numerical experiments below, we have selected these values with the help of the error bounds found in the previous sections.

It is also worth noting that the calculation of κ_n for $n \geq \xi$ in both Theorems 3.4 and 3.5 requires the evaluation of the probability mass function of the negative binomial distribution $\text{nbin}(\cdot; i, \cdot)$ for all $i = \xi, \ldots, N_1$. While the computation of such probabilities is relatively simple, it is not particularly efficient to compute each term separately because the computational times become very slow as n goes to infinity. Due to the recursive nature of the coefficients κ_n one may incur in significant numerical errors if the negative binomial probabilities are not calculated at a high precision. For more details, see for instance Loader (2000) for recommended strategies that can be used to increase the speed and accuracy of the negative binomial probabilities.

6.2 Numerical study

In this subsection, we illustrate the accuracy of our approximations through the following single example.

Example 6.2 (Pareto(ϕ) claim sizes)**.** We considered a Cramér–Lundberg model with unit premium rate, and claim sizes distributed according to Pareto(ϕ) having a cdf of the form (2.2). Recall the such distribution is parametrized in such a way that the expected value $\mu_F = 1$. Therefore, the average claim amount per unit of time $\rho = \gamma\mu_F$ will be simply equal to the intensity of the arrival process which is Poisson with intensity γ. In our example, the selection of γ will be made so the net profit $\rho - 1$ will be close to 0 ($\rho \to 1$), as this is one of the most challenging ruin probabilities for which there are results available for comparison. The parameters of the risk model selected were $\rho = 0.95$, $\phi = 2$, and the initial reserve $u \in \{1, 5, 10, 30, 50, 100, 500, 1000\}$. The exact values of the ruin probability are given in Ramsay (2003), and are now considered a classical benchmark for comparison purposes.

The implementation of approximation A requires the integrated tail of the Pareto(ϕ) distribution, which is regularly varying with parameter $\phi - 1$:

$$\widehat{F}(x) = \frac{1}{\mu_F}\int_0^x \overline{F}(t)\,\mathrm{d}t = \int_0^x \left(1 + \frac{t}{\phi - 1}\right)^{-\phi} \mathrm{d}t = 1 - \left(1 + \frac{x}{\phi - 1}\right)^{-(\phi-1)}.$$

Note that the distribution above corresponds to a Pareto distribution with parameter $\phi - 1$.

First, we analyzed the Erlangization error. We computed the bound given by Theorem 4.1 for three values of ξ, namely $100, 500, 1000$. The results are presented in Table 1. The bound appears to be tighter for smaller values of the reserve u while it gets loosen as the value of u increases; the bound appears to have steep increases in ξ but in practice, we did not notice significant changes in the numerical approximation of the probability of ruin for values of ξ larger than 100. Nevertheless, since larger values of ξ affect the speed of the algorithm, we settled with a value of $\xi = 100$ which already gave good overall results.

Table 1: Erlagization error bounds for approximation A. $\xi = 100, 500, 1000$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	ξ=100	ξ=500	ξ=1000
1	2.5736×10^{-4}	5.1724×10^{-5}	2.5856×10^{-5}
5	1.6324×10^{-3}	3.2839×10^{-4}	1.6418×10^{-4}
10	3.8081×10^{-3}	7.6641×10^{-4}	3.8319×10^{-4}
30	1.0884×10^{-2}	2.1911×10^{-3}	1.5128×10^{-3}
50	1.5028×10^{-2}	3.0257×10^{-3}	2.6455×10^{-3}
100	2.0055×10^{-2}	4.0379×10^{-3}	2.0190×10^{-3}
500	2.6279×10^{-2}	5.2911×10^{-3}	2.6455×10^{-3}
1000	2.7265×10^{-2}	5.4896×10^{-3}	2.7448×10^{-3}

We analyzed the bounds for the Erlangization error for approximation B given in Theorem 4.7. However, for values $\rho \to 1$, these bounds turned out to be not tight enough. For comparison purposes, we employed the same set of parameters as for approximation A and this time we found that the upper bound $\rho(1 - \widehat{G}_m(1))/(1 - \rho\widehat{G}_m(1)) = 0.43096$. We attempted to use the alternative bound for the Erlangization error given by Theorem 8.3, but the gain was minimal since the value of ρ is too close to 1. This shows empirically that we can have better control when implementing approximation A; the results are presented in the second column in Table 3.

Next, we discuss the approximating distribution Π constructed as a discretized Pareto distribution on the form of Definition 6.1. We choose a distribution supported over the geometric progression

$$s_1, s_1\mathrm{e}^{1/M}, s_1\mathrm{e}^{2/M}, s_1\mathrm{e}^{3/M}, \cdots$$

where $M \in \mathbb{R}^+$ will be called the ratio of the geometric progression. We will take $w_k = s_k$, so that we are considering a distribution approximated from below. It is rather clear that a large value of $M \to \infty$ combined with a small value of $s_1 \to 0$ will deliver a finer partition of the interval $[0, \infty)$, which in turn will result in a better approximation of the target distribution. However, small values of s_1 will affect severely the speed of the algorithm (see the discussion in the previous subsection) while in practice not much precision is gained by taking it too close to 0. A similar trade-off in speed and precision occurs by letting $M \to \infty$. Using the error bounds for the discretization error, we settled with $s_1 = \mathrm{e}^{-3}$ and $M = 270$ for all the examples. The discretization error bounds for approximation A are presented in the third column in Table 2, while for approximation B these are contained in the third column in Table 3.

The theoretical error bounds, defined as the aggregation of the Erlangization error bound plus the discretization error bound, are provided in the last column of those Tables 2 and 3.

Table 2: Erlangization error bounds, discretization error bounds, and theoretical error bounds for approximation A. The theoretical error bound is the sum of Erlangization error bound and discretization error bound. $\xi = 100$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	Erlangization error bounds	Discretization error bounds	Theoretical error bounds
1	2.5736×10^{-4}	6.2002×10^{-4}	8.7738×10^{-4}
5	1.6324×10^{-3}	6.3678×10^{-5}	1.6960×10^{-3}
10	3.8081×10^{-3}	3.1590×10^{-5}	3.8397×10^{-3}
30	1.0884×10^{-2}	1.0617×10^{-5}	1.0895×10^{-2}
50	1.5028×10^{-2}	6.4216×10^{-6}	1.5034×10^{-2}
100	2.0055×10^{-2}	3.2491×10^{-6}	2.0058×10^{-2}
500	2.6279×10^{-2}	6.6918×10^{-7}	2.6280×10^{-2}
1000	2.7265×10^{-2}	3.3891×10^{-7}	2.7265×10^{-2}

Table 3: Erlangization error bounds, discretization error bounds, and theoretical error bounds for approximation B. Theoretical error bound is the sum of Erlangization error bound and discretization error bound. $\xi = 100$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	Erlangization error bounds	Discretization error bounds	Theoretical error bounds
1	4.3096×10^{-1}	6.4429×10^{-6}	4.3097×10^{-1}
5	4.3096×10^{-1}	3.0087×10^{-6}	4.3097×10^{-1}
10	4.3096×10^{-1}	2.5791×10^{-6}	4.3097×10^{-1}
30	4.3096×10^{-1}	2.2920×10^{-6}	4.3097×10^{-1}
50	4.3096×10^{-1}	2.2347×10^{-6}	4.3097×10^{-1}
100	4.3096×10^{-1}	2.1917×10^{-6}	4.3097×10^{-1}
500	4.3096×10^{-1}	2.1573×10^{-6}	4.3097×10^{-1}
1000	4.3096×10^{-1}	2.1530×10^{-6}	4.3097×10^{-1}

Finally, we selected the truncation levels. In the case of the Poisson level of truncation, we were able to select a natural number large enough N_1 such that the truncation error was smaller than the floating point precision (the smallest positive representable number in Matlab with a 64-bit operating system), without increasing significantly the computational times. This selection implies that the third term in each of the bounds for the truncation error given in Theorem 5.2 and Theorem 5.4 are eliminated for practical purposes, thus eliminating the Poisson truncation error. Similarly, the scaling level of truncation level N_2 was chosen as the smallest integer N_2 such that $\varepsilon_1 < 9.5701 \times 10^{-14}$. The first and second terms in each bound given in Theorem 5.2 are straightforward to obtain. However, calculating the value of ε_2 in the truncation error bound given in Theorem 5.4 requires some computational effort. Instead, we find an upper bound of ε_2

as follows

$$\begin{aligned}\varepsilon_2 &= \sum_{j=N_2+1}^{\infty} \pi_j\, s_j = \sum_{j=N_2+1}^{\infty} \left[(1+s_{j-1})^{-2} - (1+s_j)^{-2}\right] s_j \\ &= \frac{s_1 \mathrm{e}^{N_2/M}}{\left(1+s_1 \mathrm{e}^{(N_2-1)/M}\right)^2} + \left(\mathrm{e}^{1/M}-1\right) \sum_{j=N_2}^{\infty} \frac{s_1 \mathrm{e}^{j/M}}{\left(1+s_1 \mathrm{e}^{j/M}\right)^2} \\ &< \frac{\mathrm{e}^{1/M}}{s_1 \mathrm{e}^{(N_2-1)/M}} + \left(\mathrm{e}^{1/M}-1\right) \sum_{j=N_2}^{\infty} \frac{1}{s_1 \mathrm{e}^{j/M}} = \frac{\mathrm{e}^{1/M}+1}{s_1 \mathrm{e}^{(N_2-1)/M}}.\end{aligned}$$

Table 4: Theoretical error bounds, truncation error bounds, and total error bounds for approximation A. Total error bound is the sum of theoretical error bound and truncation error bound. $\xi = 100$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	Theoretical Error Bounds	Truncation Error Bounds	Total Error Bounds
1	8.7738×10^{-4}	3.6522×10^{-9}	8.7738×10^{-4}
5	1.6960×10^{-3}	1.8261×10^{-8}	1.6960×10^{-3}
10	3.8397×10^{-3}	3.6522×10^{-8}	3.8397×10^{-3}
30	1.0895×10^{-2}	1.0957×10^{-7}	1.0895×10^{-2}
50	1.5034×10^{-2}	1.8261×10^{-7}	1.5035×10^{-2}
100	2.0058×10^{-2}	3.6522×10^{-7}	2.0059×10^{-2}
500	2.6280×10^{-2}	1.8261×10^{-6}	2.6281×10^{-2}
1000	2.7265×10^{-2}	3.6522×10^{-6}	2.7269×10^{-2}

The truncation error bounds for approximation A are presented in the second column of Table 4. The total error, which is the aggregation of the theoretical error (Erlangization and discretization errors) plus the truncation error, is presented in the third column of the same table. In the case of the truncation error bound for approximation B, this is negligible so the total error bound is equal to the theoretical error.

The numerical results for the probabilities of ruin are now summarized in Table 5. The results show that the approximated ruin probabilities are remarkably close to the *Ramsay* value calculated using equation (20) of Ramsay (2003).

Table 5: Approximations of ruin probabilities when claim sizes are Pareto(ϕ) distributed, $\xi = 100$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	Approximation A	Approximation B	Ramsay
1	0.915506746	0.915513511	0.915525781
5	0.837217038	0.837576604	0.837251342
10	0.770595774	0.771230756	0.770605760
30	0.599128897	0.600357750	0.599042454
50	0.489803156	0.491286606	0.489654166
100	0.325521064	0.327119739	0.325305086
500	0.059229343	0.059800534	0.059131409
1000	0.024594577	0.024819606	0.024544601

Table 6 shows the distances between approximations A and B with respect to the Ramsay values. We remark that these distances are much smaller than the total error bound previously obtained. There is no clear pattern in these distances so we also calculated the relative differences,

which are defined as

$$\text{Relative difference} = \frac{\text{Approximated ruin probabilities - Ramsay}}{\text{Ramsay}}.$$

The relative differences are presented in Figure 3. Both approximations deliver small relative differences which are increasing with initial reserve u. Nevertheless, the relative differences grow much slower for approximation A which provides empirical evidence that approximation A will remain precise as u increases.

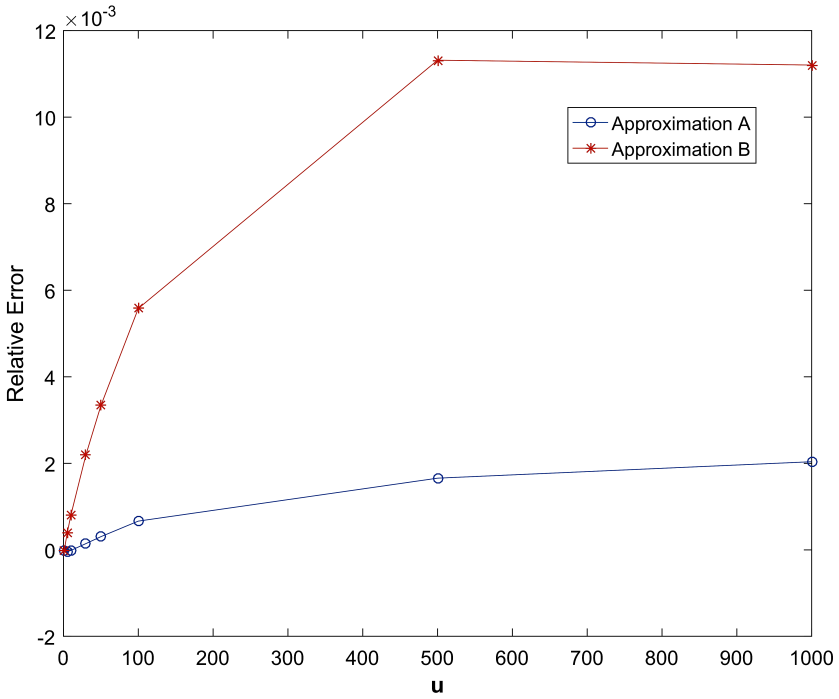
Notice that the proposed approximations can be sharpened by increasing the value of M (making the partition finer) and to a lesser extent by reducing the value of s_1 (improving the approximation of the target distribution in a vicinity of 0). The improvement obtained by increasing the value of ξ is very modest and it will slow down the algorithms considerably.

Table 6: Differences of the approximated ruin probabilities with respect to the Ramsay approximation for Pareto(ϕ) distributed claim sizes, $\xi = 100$, $\rho = 0.95$, $\phi = 2$ and initial reserve u.

u	Approximation A	Approximation B
1	-1.9035×10^{-5}	-1.2270×10^{-5}
5	-3.4304×10^{-5}	3.2526×10^{-4}
10	-9.9860×10^{-6}	6.2500×10^{-4}
30	8.6443×10^{-5}	1.3153×10^{-3}
50	1.4899×10^{-4}	1.6324×10^{-3}
100	2.1598×10^{-4}	1.8147×10^{-3}
500	9.7940×10^{-5}	6.6913×10^{-4}
1000	4.9976×10^{-5}	2.7501×10^{-4}

It is worth noticing that the approximation B appears to be more precise than approximation A for small values of the initial reserve. At first sight, this might appear quite surprising since we expected approximation A to be more precise. This could be explained by the cancelling effects of the separate errors (Erlangization, discretization, truncation). For instance, the discretized distribution Π is an approximation from below of the target distribution F, so $F \prec \Pi$ in stochastic order. With this selection it follows that $\psi_{\Pi \star G_m}(u)$ overestimates $\psi_{\widehat{F} \star G_m}(\mathrm{u})$ so the discretization error will be negative for both approximations A and B. The truncation error will be always negative since all the truncated terms have positive sign. From the analysis above, it is not possible to determine the sign of the Erlangization error.

Figure 3: Relative differences of the approximated ruin probabilities with respect to the Ramsay approximation for Pareto(ϕ) distributed claim sizes, $\rho = 0.95$, $\phi = 2$, and initial reserve u.



7 Conclusion

Bladt et al. (2015) remarked that the family of phase-type scale mixtures could be used to provide sharp approximations of heavy-tailed claim size distributions. In our work, we addressed such a remark and provided a simple systematic methodology to approximate any nonnegative continuous distribution within such a family of distributions. We employed the results of Bladt et al. (2015) and provided simplified expressions for the probability of ruin in the classical Cramér–Lundberg risk model. In particular, we opted to approximate the integrated tail distribution $\widehat{F}$ rather than the claim sizes as suggested in Bladt et al. (2015); we showed that such an alternative approach results in a more accurate and simplified approximation for the associated ruin probability. We further provided bounds for the error of approximation induced by approximating the integrated tail distribution as well as the error induced by the truncation of the infinite series. Finally, we illustrated the accuracy of our proposed method by computing the ruin probability of a Cramér-Lundberg reserve process where the claim sizes are heavy-tailed. Such an example is classical but often considered challenging due to the heavy-tailed nature of the claim size distributions and the value of the net profit condition.

Acknowledgements The authors thank Mogens Bladt for multiple discussions on the ideas which originated this paper and an anonymous referee who provided a detailed review that helped to improve its quality. OP is supported by the CONACYT PhD scholarship No. 410763 sponsored by the Mexican Government. LRN is supported by ARC grant DE130100819. WX is supported by IPRS/APA scholarship at The University of Queensland. HY is supported by APA scholarship at The University of Queensland.

References

Ash, R. and C. Doléans-Dade (2000). *Probability and Measure Theory* (Second ed.). Harcourt/Academic Press.

Asmussen, S. (2003). *Applied Probabilities and Queues* (2nd ed.). New York: Springer-Verlag.

Asmussen, S. and H. Albrecher (2010). *Ruin Probabilities (Advanced series on statistical science & applied probability; v. 14)* (2nd ed.). World Scientific Publishing Co. Pte. Ltd.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons, Inc.

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

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

Cai, J. and J. Garrido (1999). Two-sided bounds for ruin probabilities when the adjustment coefficient does not exist. *Scandinavian Actuarial Journal 1999*(1), 80–92.

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

Loader, C. (2000). Fast and accurate computation of binomial probabilities. Unpublished manuscript.

Neuts, M. F. (1975). Probability distributions of phase-type. In R. Holvoet (Ed.), *Liber amicorum Professor emeritus dr. H. Florin*, pp. 173–206. Katholieke Universiteit Leuven, Departement Wiskunde.

Ramsay, C. M. (2003). A solution to the ruin problem for Pareto distributions. *Insurance: Mathematics and Economics 33*(1), 109–116.

Rojas-Nandayapa, L. and W. Xie (2015). Asymptotic tail behaviour of phase–type scale mixture distributions. *arXiv:1502.01811v1*.

Santana, D. J., J. González-Hernández, and L. Rincón (2016). Approximation of the ultimate ruin probability in the classical risk model using Erlang mixtures. *Methodology and Computing in Applied Probability*, 1–24.

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

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

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

Vatamidou, E., I. J. B. F. Adan, M. Vlasiou, and B. Zwart (2013). Corrected phase-type approximations of heavy-tailed risk models using perturbation analysis. *Insurance: Mathematics and Economics 53*, 366–378.

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

8 Appendix: Bounds for errors of approximation

In the first subsection of this appendix, we provide an alternative bound for the Erlangization error of the approximations B proposed in Subsection 4.3.1. In the second subsection of this appendix, we provide an auxiliary result that will be useful for computing numerical bounds for the distance between the moment distributions used in Theorem 4.9.

8.1 Refinements for the Erlangization error of approximatinon A

Theorem 8.3 below provides an alternative bound for the Erlangization error of approximation B proposed in Theorem 4.7. This alternative bound is slightly tighter than the bound proposed in Theorem 4.7. However, this bound is more difficult to construct and implement. Remark 8.4 at the end of this subsection provides an argument that shows that Theorem 4.7 can be seen as a particular case of Theorem 8.3. Lemma 8.1 and 8.2 below are technical but needed in the proof of Theorem 8.3.

Lemma 8.1. *Let $0 \leq \delta \leq 1$ and $1 \leq \beta \leq \infty$. Define*

$$A_\delta = [\delta, 1]^n, \qquad A^{\delta,\beta} = [\delta, \beta]^n \setminus [\delta, 1]^n.$$

Then

$$\begin{aligned}
&\int_{A_\delta} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&\qquad = (1-\delta)^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta) \right)^n \\
&- \int_{A^{\delta,\beta}} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&\qquad = \left(\widehat{G}_m(\beta) - \widehat{G}_m(\delta) \right)^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta) \right)^n .
\end{aligned}$$

Proof. For the first inequality observe that if $(s_1, \ldots, s_n) \in A_\delta$, then $\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) = 1$. Since $\widehat{g}_m(\cdot) < 1$ then

$$\int_{A_\delta} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n$$

For the second equality, notice that if $(s_1, \ldots, s_n) \in A^{\delta,\beta}$, then $\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) = 0$, so

$$\begin{aligned}
-\int_{A^{\delta,\beta}} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n &= \int_{[\delta,\beta]^n \setminus [\delta,1]^n} \cdots \int \prod_{i=1}^n \widehat{g}_m(s_i) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&= \left(\widehat{G}_m(\beta) - \widehat{G}_m(\delta) \right)^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta) \right)^n .
\end{aligned}$$

□

Lemma 8.2. *For each $\delta \in [0, 1]$, there exists $\beta \in [1, \infty)$ such that for all $n \in \mathbb{N}$ it holds that*

$$\delta - \widehat{G}_m(\delta) = 1 - \widehat{G}_m(\beta). \tag{8.1}$$

Furthermore, for such a pair (δ, β), we have that for all $n \geq 1$,

$$\begin{aligned}
&\int_{A_\delta} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&\qquad = -\int_{A^{\delta,\beta}} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n,
\end{aligned} \tag{8.2}$$

where $A_\delta = (\delta, 1)^n$ and $A^{\delta,\beta} = (\delta, \beta)^n \setminus (\delta, 1)^n$.

Proof. Fix $\beta \in [1,\infty)$. Notice that $1-\widehat{G}_m(\beta) \in (0, 1-\widehat{G}_m(1)]$ and non-increasing. Also notice that $\delta - \widehat{G}_m(\delta)$ is non-decreasing and continuous with image $[0, 1-\widehat{G}_m(1)]$. Then clearly there exists $\beta = \beta(\delta) \in [0,1]$ such that (8.1) holds. The second part follows from Lemma 8.1 by noticing that

$$\begin{aligned}
&\int_{A_\delta} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n + \int_{A^{\delta,\beta}} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\
&\quad = (1-\delta)^n - \left(\widehat{G}_m(\beta) - \widehat{G}_m(\delta) \right)^n,
\end{aligned}$$

so that (8.2) holds if (8.1) does. □

Theorem 8.3. *Let $1 = \beta_0 \le \beta_1 < \beta_2 < \cdots \le \beta_K = \infty$ and let $1 = \delta_0 \ge \delta_1 > \delta_2 > \cdots \ge \delta_K = 0$ be such that each individual pair (δ_k, β_k) satisfies Lemma 8.2. Then,*

$$\begin{aligned}
&\left| \psi_{H_F \star U}(u) - \psi_{H_F \star \widehat{G}_m}(u) \right| \\
&\quad \le (1-\rho) \sum_{k=0}^{K-1} \left[\mathcal{T}_1\left(u/\beta_{k+1}; q_{1,k+1}\right) - \mathcal{T}_2\left(u/\beta_{k+1}; q_{2,k+1}\right) - \mathcal{T}_2\left(u/\beta_{k+1}; q_{1,k}\right) + \mathcal{T}_1\left(u/\beta_{k+1}; q_{2,k}\right) \right]
\end{aligned}$$

with

$$\mathcal{T}_\ell(s,q) := \begin{cases} \dfrac{(1-q)e^{-\ell s \kappa(s,q)} + q\overline{H}_F(s)}{(1-q)(1-qH_F(s))}, & s > 0,\ q > 0 \\ \dfrac{q}{1-q} & s = 0,\ q > 0 \\ 0 & q = 0, \end{cases}$$

where $q_{1,k} = \rho(1-\delta_k)$, $q_{2,k} = \rho\left(\widehat{G}_m(1) - \widehat{G}_m(\delta_k)\right)$ and $\kappa(s,q)$ is the solution κ of the Lundberg equation

$$\int_0^s e^{\kappa x} \overline{H}_F(\mathrm{d}x) = \frac{H_F(s)}{q}.$$

Proof. Set $\mathcal{A}_k := [\delta_k, \beta_k]$, $B_k = [\delta_{k+1}, 1]^n \setminus [\delta_k, 1]^n$ and $C_k = ([\delta_{k+1}, \beta_{k+1}]^n \setminus [\delta_{k+1}, 1]^n) \setminus ([\delta_k, \beta_k]^n \setminus [\delta_k, 1]^n)$. Let $\{X_n'\}$ be a sequence of i.i.d. random variables with common distribution H_F. Notice that

$$\begin{aligned}
&\left| \int_{\mathcal{A}_{k+1}^n \setminus \mathcal{A}_k^n} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \le u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right| \\
&\quad = \left| \int_{B_k} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \le u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right. \\
&\qquad \left. + \int_{C_k} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \le u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right| \\
&\quad = \left| \int_{B_k} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \le u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right. \\
&\qquad \left. - \int_{C_k} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \le u) \left(\prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right|. \qquad (8.3)
\end{aligned}$$

We proceed to analyze the expression inside the absolute value. Observe that for any $(r_1, \ldots, r_n) \in B_n$ we have that

$$H_F^{*n}(u) \leq \mathbb{P}(r_1 X_1' + \cdots + r_n X_n' \leq u) \leq H_F^{*n}(u/\delta_{k+1})$$

while for $(s_1, \ldots, s_n) \in C_n$, it holds that

$$H_F^{*n}(u) \geq \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \leq u) \geq H_F^{*n}(u/\beta_{k+1}).$$

Let us take the negative value of the expression inside the absolute value in (8.3). This expression is bounded by

$$H_F^{*n}(u) \left[\int_{C_k} \cdots \int \left(\prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n - \int_{B_k} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right]$$

Because of Lemma 8.1 the last is equal to 0, which implies that the expression inside the absolute value in (8.3) should be taken positive and thus bounded by

$$\begin{aligned} H_F^{*n}(u/\delta_{k+1}) \int_{B_k} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\ - H_F^{*n}(u/\beta_{k+1}) \int_{C_k} \cdots \int \left(\prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n. \end{aligned}$$

This is equal to

$$\begin{aligned} &(H_F^{*n}(u/\delta_{k+1}) - H_F^{*n}(u/\beta_{k+1})) \int_{B_k} \cdots \int \left(\prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) \mathrm{d}s_1 \ldots \mathrm{d}s_n \\ &\quad \leq \overline{H}_F^{*n}(u/\beta_{k+1}) \left[(1-\delta_{k+1})^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_{k+1}) \right)^n - (1-\delta_k)^n + \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_k) \right)^n \right]. \end{aligned}$$

If $u/\beta_{k+1} = 0$, then $\overline{H}(u/\beta_{k+1}) = 1$ and the expression above is equal to a linear combination of powers. Summing over n yields a convergent geometric series. If $u/\beta_{k+1} > 0$, then we proceed as follows. According to Cai and Garrido (1999), if $q \in (0,1)$ and $0 < s < \infty$ then

$$\frac{(1-q)e^{-2s\kappa(s,q)} + q\overline{H}_F(s)}{(1-q)(1-qH_F(s))} \leq \sum_{n=1}^{\infty} q^n \overline{H}_F^{*n}(s) \leq \frac{(1-q)e^{-s\kappa(s,q)} + q\overline{H}_F(s)}{(1-q)(1-qH_F(s))}, \tag{8.4}$$

where $\kappa(s,q)$ is referred to as the *adjustment coefficient* and defined as the solution κ of the Lundberg equation given by

$$\int_0^s e^{\kappa x} H_F(\mathrm{d}x) = \frac{H_F(s)}{q}.$$

Applying this result to our bound, we obtain that

$$\begin{aligned} &\left| \sum_{n=1}^{\infty} (1-\rho)\rho^n \int_{\mathcal{A}_{k+1}^n \setminus \mathcal{A}_k^n} \cdots \int \mathbb{P}(s_1 X_1' + \cdots + s_n X_n' \leq u) \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \mathrm{d}s_1 \ldots \mathrm{d}s_n \right| \\ &\quad \leq \sum_{n=1}^{\infty} (1-\rho)\rho^n \overline{H}_F^{*n}(u/\beta_{k+1}) \left[(1-\delta_{k+1})^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_{k+1}) \right)^n \right. \\ &\qquad \left. -(1-\delta_k)^n + \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_k) \right)^n \right] \\ &\quad \leq (1-\rho) \left[\mathcal{T}_1(u/\beta_{k+1}; q_{1,k+1}) - \mathcal{T}_2(u/\beta_{k+1}; q_{2,k+1}) - \mathcal{T}_2(u/\beta_{k+1}; q_{1,k}) + \mathcal{T}_1(u/\beta_{k+1}; q_{2,k}) \right]. \end{aligned}$$

The result follows after summing all the terms corresponding to the sets $\mathcal{A}_{k+1} \setminus \mathcal{A}_k$. □

Remark 8.4. Notice that when $\delta_1 = 0$ and $\beta_1 = \infty$, then the bound in the previous theorem reduces to the bound obtained in Theorem 4.7.

The construction of this particular bound requires the selection of pairs (δ_k, β_k) satisfying Equation (8.1) in Corollary 8.2. In most cases these values can be determined numerically by using a standard root-finding algorithm. Also, an explicit expression for the term $\widehat{G}_m(\delta)$ can be found in Lemma 8.5 below.

Lemma 8.5. *For any $\delta \geq 0$, define*

$$\epsilon_m(\delta) := \int_0^\delta G_m(s)\mathrm{d}s.$$

Then

$$\epsilon_m(\delta) = \delta - 1 + \sum_{k=0}^{\xi-1} \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!} - \delta \sum_{k=0}^{\xi-2} \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!}.$$

Proof. Recall that

$$G_m(s) = 1 - \sum_{n=0}^{\xi-1} \frac{\mathrm{e}^{-\xi s}(\xi s)^n}{n!},$$

so that

$$\begin{aligned}
\epsilon_m(\delta) &= \int_0^\delta \left(1 - \sum_{n=0}^{\xi-1} \frac{\mathrm{e}^{-\xi s}(\xi s)^n}{n!}\right) \mathrm{d}s \\
&= \delta - \sum_{n=0}^{\xi-1} \int_0^\delta \frac{\mathrm{e}^{-\xi s}(\xi s)^n}{n!}\mathrm{d}s \\
&= \delta - \frac{1}{\xi} \sum_{n=0}^{\xi-1} \left(1 - \sum_{k=0}^{n} \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!}\right) \\
&= \delta - 1 + \frac{1}{\xi} \sum_{k=0}^{\xi-1} (\xi - k) \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!} \\
&= \delta - 1 + \sum_{k=0}^{\xi-1} \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!} - \delta \sum_{k=0}^{\xi-2} \frac{\mathrm{e}^{-\xi\delta}(\xi\delta)^k}{k!}.
\end{aligned}$$

□

8.2 Error bound for the distance between moment distributions

As stated in Subsection 4.3.2, the result of Theorem 4.9 depends on the availability of the distance between moment distributions $|H_F - H_\Pi|$. In the following, we state a bound for such a quantity in the case where an explicit expression for $|H_F - H_\Pi|$ is not available or too difficult to compute.

Lemma 8.6. *Let Π be defined as in Definition 6.1 and $\Delta_k H_F := H_F(s_k) - H_F(s_{k-1})$. Then, for all $K \in \mathbb{N}$ it holds that*

$$\sup_{u \leq s < \infty} \left|\overline{H}_F(s) - \overline{H}_\Pi(s)\right| \leq \sup_{K \leq k < \infty} \Delta_k H_F + \frac{|\mu_\Pi - \mu_F| \cdot \mathbb{E}[S; S > u]}{\mu_\Pi \cdot \mu_F} + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F},$$

$$\sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| \leq \sup_{0 \leq k \leq K} \Delta_k H_F + \frac{|\mu_\Pi - \mu_F| \cdot \mathbb{E}[S; S \leq u]}{\mu_\Pi \cdot \mu_F} + \frac{|\mathbb{E}[S; S \leq u] - \mathbb{E}[X; X \leq u]|}{\mu_F},$$

where $X \sim F$, $S \sim \Pi$ and $\mu_F = \mathbb{E}[X]$ and $\mu_\Pi = \mathbb{E}S$. Moreover, if $\mu_F = \mu_\Pi$, then

$$\sup_{u \leq s < \infty} \left|\overline{H}_F(s) - \overline{H}_\Pi(s)\right| \leq \sup_{K \leq k < \infty} \Delta_k H_F + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F},$$

$$\sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| \leq \sup_{0 \leq k \leq K} \Delta_k H_F + \frac{|\mathbb{E}[S; S \leq u] - \mathbb{E}[X; X \leq u]|}{\mu_F}.$$

Proof. Since $K \in \mathbb{N}$ is such that $s_K = u$ then

$$\begin{aligned}
\left|H_F \star \widehat{G}_m(u) - H_\Pi \star \widehat{G}_m(u)\right| &= \left|\int_0^\infty H_F(u/s)d\widehat{G}(s) - \int_0^\infty H_\Pi(u/s)d\widehat{G}(s)\right| \\
&\leq \int_0^\infty |H_F(u/s) - H_\Pi(u/s)|\, d\widehat{G}(s) \\
&\leq \sup_{u\leq s<\infty} |H_F(s) - H_\Pi(s)| \int_0^1 d\widehat{G}(s) \\
&\qquad + \sup_{0<s\leq u} |H_F(s) - H_\Pi(s)| \int_1^\infty \mathrm{d}\widehat{G}(s) \\
&= \sup_{u\leq s<\infty} \left|\overline{H}_\Pi(s) - \overline{H}_F(s)\right| \int_0^1 d\widehat{G}(s) \qquad (8.5) \\
&\qquad + \sup_{0<s\leq u} |H_F(s) - H_\Pi(s)| \int_1^\infty d\widehat{G}(s). \qquad (8.6)
\end{aligned}$$

Observe that for all $0 < s < \infty$ there exists $k \geq 1$ such that $t_k \leq s < t_{k+1}$, so that

$$\left|\overline{H}_\Pi(s) - \overline{H}_F(s)\right| \leq \max\left\{\left|\overline{H}_F(s_k) - \overline{H}_\Pi(s_k)\right|, \left|\overline{H}_F(s_{k+1}) - \overline{H}_\Pi(s_k)\right|\right\}.$$

Using the previous identity, we construct a bound for (8.5):

$$\begin{aligned}
\left|\overline{H}_\Pi(s_k) - \overline{H}_F(s_k)\right| &\leq \left|\overline{H}_\Pi(s_k) - \overline{H}_F(s_{k+1})\right| + \left|\overline{H}_F(s_k) - \overline{H}_F(s_{k+1})\right| \\
&\leq \left|\overline{H}_\Pi(s_k) - \overline{H}_F(s_{k+1})\right| + \Delta_k H_F,
\end{aligned}$$

where $\Delta_k H_F := H_F(s_{k+1}) - H_F(s_k)$. In consequence,

$$\sup_{u\leq s<\infty}\left\{\left|\overline{H}_\Pi(s) - \overline{H}_F(s)\right|\right\} \leq \sup_{K\leq k<\infty}\left\{\left|\overline{H}_\Pi(s_k) - \overline{H}_F(s_{k+1})\right|\right\} + \sup_{K\leq k<\infty}\left\{\Delta_k H_F\right\}.$$

Next, observe that

$$\begin{aligned}
\sup_{K\leq k<\infty}\left\{\left|\overline{H}_\Pi(s_k) - \overline{H}_F(s_{k+1})\right|\right\} &= \sup_{K\leq k<\infty}\left\{\left|\sum_{i=k+1}^\infty \frac{s_i\pi_i}{\mu_\Pi} - \int_{s_{k+1}}^\infty \frac{tdF(t)}{\mu_F}\right|\right\} \\
&= \sup_{K\leq k<\infty}\left\{\left|\sum_{i=k+1}^\infty \int_{s_i}^{s_{i+1}} \left(\frac{s_i}{\mu_\Pi} - \frac{t}{\mu_F}\right) dF(t)\right|\right\} \\
&\leq \frac{1}{\mu_\Pi\cdot\mu_F}\sum_{i=K+1}^\infty \int_{s_i}^{s_{i+1}} |s_i\mu_F - t\mu_\Pi| dF(t) \\
&\leq \frac{1}{\mu_\Pi\cdot\mu_F}\sum_{i=K+1}^\infty \int_{s_i}^{s_{i+1}} \left(|s_i\mu_F - s_i\mu_\Pi| + |s_i\mu_\Pi - t\mu_\Pi|\right) dF(t) \\
&\leq \frac{|\mu_F - \mu_\Pi|}{\mu_\Pi\cdot\mu_F}\sum_{i=K+1}^\infty s_i\int_{s_i}^{s_{i+1}} dF(t) + \frac{1}{\mu_F}\sum_{i=K+1}^\infty \int_{s_i}^{s_{i+1}} |s_i - t| dF(t) \\
&\leq \frac{|\mu_\Pi - \mu_F|\mathbb{E}[S; S > s_K]}{\mu_\Pi\cdot\mu_F} + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{u\leq s<\infty}\left\{\left|\overline{H}_\Pi(s) - \overline{H}_F(s)\right|\right\} \\
&\qquad \leq \sup_{K\leq k<\infty}\left\{\Delta_k H_F\right\} + \frac{|\mu_\Pi - \mu_F|\cdot\mathbb{E}[S; S > u]}{\mu_\Pi\cdot\mu_F} + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F}.
\end{aligned}$$

Our construction for a bound for (8.6) is analogous. Note that

$$\begin{aligned}
|H_F(s_{k+1}) - H_\Pi(s_k)| &\leq |H_F(s_{k+1}) - H_F(s_k)| + |H_F(s_k) - H_\Pi(s_k)| \\
&\leq \Delta_k H_F + |H_F(s_k) - H_\Pi(s_k)|,
\end{aligned}$$

so

$$\sup_{0<s\leq u}\{|H_F(s)-H_\Pi(s)|\}\leq \sup_{0\leq k\leq K}\{\Delta_k H_F\}+\sup_{0\leq k\leq K}\{|H_F(s_k)-H_\Pi(s_k)|\},$$

where $s_0=\inf\{s:F(s)>0\}$. Next, observe that

$$\begin{aligned}
\sup_{0\leq k\leq K}\{|H_F(s_k)-H_\Pi(s_k)|\} &= \sup_{0\leq k\leq K}\left\{\left|\int_0^{s_k}\frac{tdF(t)}{\mu_F}-\sum_{i=1}^{k}\frac{s_i\pi_i}{\mu_\Pi}\right|\right\}\\
&= \sup_{0\leq k\leq K}\left\{\left|\sum_{i=1}^{k}\int_{s_i-1}^{s_i}\left(\frac{t}{\mu_F}-\frac{s_i}{\mu_\Pi}\right)dF(t)\right|\right\}\\
&\leq \frac{1}{\mu_\Pi\cdot\mu_F}\sum_{i=1}^{K}\int_{s_{i-1}}^{s_i}|t\mu_\Pi-s_i\mu_F|dF(t)\\
&\leq \frac{1}{\mu_\Pi\cdot\mu_F}\sum_{i=1}^{K}\int_{s_{i-1}}^{s_i}|t\mu_\Pi-s_i\mu_\Pi|+(|s_i\mu_\Pi-s_i\mu_F|)\,dF(t)\\
&\leq \frac{1}{\mu_F}\sum_{i=1}^{K}\int_{s_{i-1}}^{s_i}|s_i-t|dF(t)+\frac{|\mu_\Pi-\mu_F|}{\mu_\Pi\cdot\mu_F}\sum_{i=1}^{K}s_i\int_{s_{i-1}}^{s_i}dF(t)\\
&\leq \frac{|\mathbb{E}[S;S\leq s_K]-\mathbb{E}[X;X\leq s_K]|}{\mu_F}+\frac{|\mu_\Pi-\mu_F|\cdot\mathbb{E}[S;S\leq s_K]}{\mu_\Pi\cdot\mu_F}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{0\leq s\leq u/\delta}\{|H_F(s)-H_\Pi(s)|\}\\
&\quad\leq \sup_{0\leq k\leq K}\{\Delta_k H_F\}+\frac{|\mathbb{E}[S;S\leq u]-\mathbb{E}[X;X\leq u]|}{\mu_F}+\frac{|\mu_\Pi-\mu_F|\cdot\mathbb{E}[S;S\leq u]}{\mu_\Pi\cdot\mu_F}.
\end{aligned}$$

□

Notice that the particular selection of Π implies that it is possible to select partitions for which $\mu_\Pi=\mu_F$. Also, recall that when $\xi(m)\to\infty$, then $1-G_m(1)\to 0$, so for $\xi(m)$ sufficiently large, the bound decreases as $|\mathbb{E}[X;X>s_K]-\mathbb{E}[S;S>s_K]|$ becomes smaller. The last is achieved if the tail probability of H_Π gets *closer* to the tail probability of H_F. Finally, the value of K can be chosen so it minimizes the bound above; this optimization can be carried numerically for various examples.