

Approximation of Ruin Probabilities via Erlangized Scale Mixtures

Oscar Peralta¹, Leonardo Rojas-Nandayapa², Wangyue Xie³ and Hui Yao³

¹Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark, osgu@dtu.dk

²Mathematical Sciences, University of Liverpool, UK, leorojas@liverpool.ac.uk

³School 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) and 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 $\Pi \star G$ to approximate 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 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. We are doing so throughout this manuscript.

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* 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:

$$\Pi \star G(u) := \int_0^\infty G(u/s) d\Pi(s) = \int_0^\infty \Pi(u/s) dG(s).$$

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 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), thus 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 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 $\Pi_m \rightarrow \widehat{F}$ (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 increasing and unbounded, then $\Pi_m \star G_m \rightarrow \widehat{F}$. Then we can 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 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 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 binomial coefficients instead of computationally expensive matrix operations. In practice, the infinite series can be computed only up to a fine number of terms, but as we will show, this numerical error can be controlled by selecting an appropriate distribution Π . Such a 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 (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 emphasise 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 discretization and the Erlangization part 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, given that the tail behavior of $\Pi * G_m$ is the same for each $\xi(m) \geq 1$.

The rest of the paper is organized as the 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 $\Pi_m * G_m$ approximating the integrated tail of a general claim size distribution F . Based on the results of Bladt et al. (2015), we introduce a simplified infinite series representation of the ruin probability $\psi_{\Pi_m * G_m}$. In Section 4, we construct the bound for the error of approximation $|\psi_{\hat{F}} - \psi_{\Pi * G}|$. In Section 5, we provide a bound for the numerical error of approximation induced by truncating the infinite series representation of $\psi_{\Pi_m * G_m}$. A numerical example illustrating the sharpness of our result is given in Section 6. 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 section 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*. The resulting approximating distribution will be more 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 $E = \{0, 1, 2, \dots, p\}$. The states $\{1, 2, \dots, p\}$ are transient while the state 0 is an absorbing state. Hence, phase-type distributions are characterized by a p -dimensional row vector $\beta = (\beta_1, \dots, \beta_p)$, corresponding to the initial probabilities of each of the transient states of the Markov jump process, and an intensity matrix

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

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

$$g(y) = \beta e^{\mathbf{\Lambda} y} \lambda, \quad G(y) = 1 - \beta e^{\mathbf{\Lambda} y} \mathbf{e}, \quad \int_0^\infty y dG(y) = -\beta \mathbf{\Lambda}^{-1} \mathbf{e}.$$

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 m -dimensional vector $\boldsymbol{\beta} = (1, 0, \dots, 0)$ and the $m \times m$ dimensional matrix

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

We denote $\text{Erlang}(\lambda, m)$. In this paper we will be particularly interested in the sequence of $G_m \sim \text{Erlang}(\xi(m), \xi(m))$ distributions with $\xi(m) \rightarrow \infty$. These type of 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 $G_m(y) \rightarrow \mathbb{I}_{[1, \infty)}(y)$ weakly, where \mathbb{I} is the indicator function.

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*

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

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 *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.

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

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

where $dH_\Pi(s) = s d\Pi(s) / \mu_\Pi$ is called the moment distribution of Π and \widehat{G} is the integrated tail of G . We use μ to denote the expectation.

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

Observe that

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

□

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 point mass at one probability measure, however, a self-contained proof is provided.

Proposition 2.3. *Let $dH_F(s) := s dF(s) / \mu_F$ be the moment distribution of F and $U \sim U(0, 1)$. Then*

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

Proof.

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

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

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 $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots)$ such that $\boldsymbol{\pi} \mathbf{e}_\infty = 1$, where \mathbf{e}_∞ is an infinite dimensional column vector with all elements to be 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 $G(u/s_i) \sim \text{PH}(\boldsymbol{\beta}, \boldsymbol{\Lambda}/s_i)$ 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; 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 infinite number of transient states, having representation $(\boldsymbol{\alpha}, \mathbf{T})$ 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 infinite dimensional phase-type distributions contains heavy-tailed distributions but it is mathematically intractable. The rest of the classes defined above remain dense, contain both light and heavy-tailed distributions and are more tractable 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.

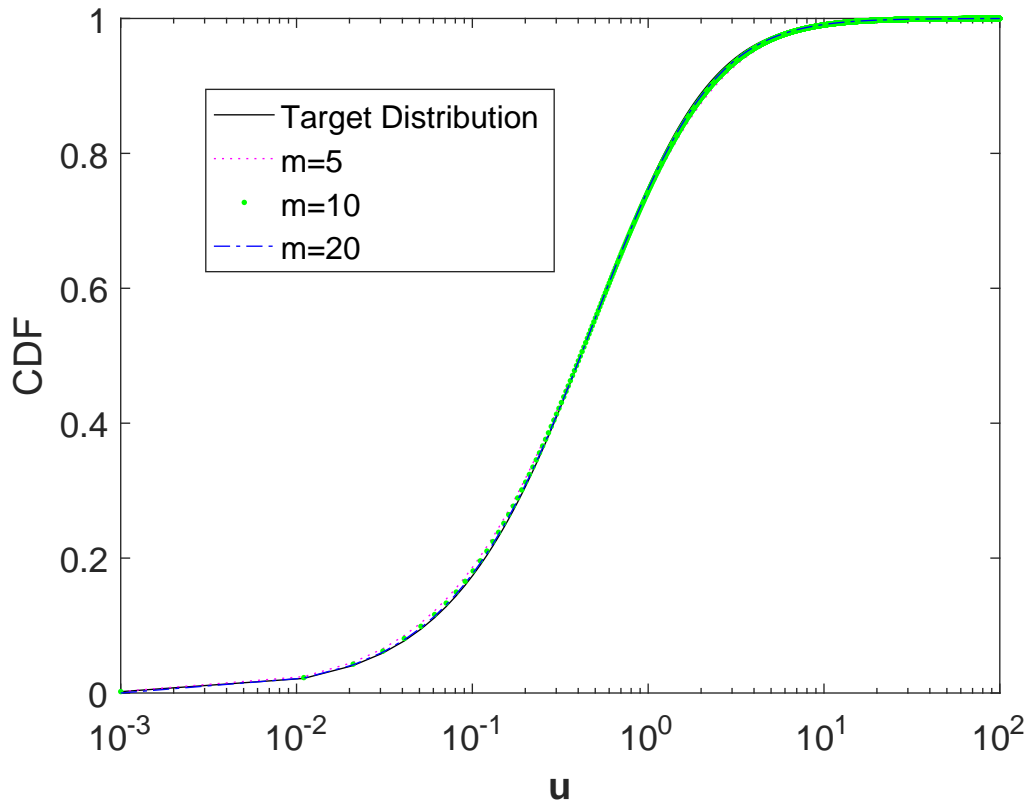
Proposition 2.4. Let Π_m be a sequence of nonnegative discrete distributions such that $\Pi_m \rightarrow \Pi$ and $G_m \sim \text{Erlang}(\xi(m), \xi(m))$. Then

$$\Pi_m \star G_m \rightarrow \Pi.$$

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

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 (see Figure 1).

Figure 1: Comparison between a target cumulative distribution function \widehat{F} and its *Erlangized scale mixture* approximations $\Pi_m \star G_m$. In this example we have considered a Pareto distribution supported over $[0, \infty)$. The approximating distributions Π_m were taken as discretizations of the target distribution \widehat{F} over some geometric progression $\{e^{t_k} : k \in \mathbb{Z}, t_k = k/m\}$, while we take $\xi(m) = m$, so $G_m \sim \text{Erlang}(m, m)$.



3 Ruin probabilities

In this section we introduce a method of approximation for the ruin probability in the Cramér–Lundberg risk model using Erlangized scale mixtures. We apply the results of Bladt et al. (2015) to obtain expressions for the ruin probability in terms of infinite series involving operations with finite dimensional arrays, and exploit the simple structure of the Erlang distribution to obtain explicit formulas which will be free of matrix operations.

For constructing approximations of the ruin probability we follow two alternative approaches. In the first approach we approximate directly the integrated tail distribution \widehat{F} via Erlangized scale mixtures and is the one that we advocate in this paper, we shall call it *approximation A*. This straightforward approach delivers explicit formulas which are simple to write and implement; as we will see, the approximations obtained are very accurate. However, the approximation obtained by using this approach 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. Therefore, an approximating distribution for the claim sizes is not immediately available in this setting. It is also required to have an explicit expression for the integrated tail distribution \widehat{F} .

A second approach, which is named as *approximation B*, is also analysed where the claim size distribution is approximated with an Erlangized scale mixture. This is equivalent to approximating the integrated tail \widehat{F} with the integrated tail distribution of an Erlangized scale mixture distribution. As we will show later, 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 can be considered more natural but the resulting expressions are more complex and the approximations are less accurate. The error of approximation is bigger as a result of the *amplifying* effect of integrating the tail probability of the approximating distribution. Its implementation is more involved and the computational times are much slower when compared to the results delivered using approximation A.

We remark that approximation B is the more commonly used, like for instance in Bladt et al. (2015) and Santana et al. (2016). Thus we have included its analysis for comparison purposes.

The remaining content of this section is organised 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 u is the initial reserve of an insurance company, the premiums flow in at a rate 1 per unit time t , X_1, X_2, \dots 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. So R_t is a risk model for the time evolution of the reserve of the insurance company. We say that ruin occurs if and only if the reserve ever drops below zero; we denote $\psi_{\widehat{F}}(u) := \inf\{R_t < 0 : t > 0\}$.

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:

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

where $\rho = \gamma\mu_F < 1$ is the average claim amount per unit time, F^{*n} denotes the n th-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:

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

The calculation of 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 the the probability that the lifetime of the renewal is larger than u is given by

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

where $\boldsymbol{\alpha} = (\boldsymbol{\pi} \otimes \boldsymbol{\beta})$, $\mathbf{T} = (\mathbf{s} \times \mathbf{I}_\infty)^{-1} \otimes \boldsymbol{\Lambda}$ and $\mathbf{t} = -\mathbf{T} \mathbf{e}_\infty$. Here $\mathbf{s} = (s_1, s_2, \dots)$, \mathbf{I} is an identity matrix and \mathbf{I}_∞ is that of infinite dimension. The formula above is not of practical use because the vectors $\boldsymbol{\alpha}$, \mathbf{t} and the matrix \mathbf{T} have infinite dimensions. However, using the special structure of \mathbf{T} , they further refined the formula above and expressed $\psi_{\Pi \star G}$ as an infinite series involving matrices and vectors of finite dimension which characterize the underlying distributions Π and G .

Next, we obtain the explicit formula for $\psi_{\Pi \star G}(u)$ in terms of the parameters characterising the renewal distribution $\Pi \star G$ (equivalently 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 phase-type scale mixture distribution, so their results are given instead in terms of the parameters characterising the underlying claim size distribution. For simplicity of notation, we will write $G_m \sim \text{Erlang}(\xi, \xi)$ instead of $\text{Erlang}(\xi(m), \xi(m))$ for the rest of the paper.

Proposition 3.1 (Bladt et al. (2015)). *Let $0 < \rho < 1$,*

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

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

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

where

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

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 $-\mathbf{T}$, then from Theorem 3.1 in Bladt et al. (2015), we have

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

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

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

It is not difficult to see that

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

and

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

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 that shows that if the claim sizes are distributed according to an Erlangized scale mixture $\Pi \star G_m$, then its 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 formula of Bladt et al. (2015).

Proposition 3.2.

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

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

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

where

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

A drawback from the formulas given above is that the calculation of the quantities B_{ij} , C_{ij} and D_{ij} is computationally expensive since these involve costly matrix operations. However, these expression can be simplified in our case because because the subintensity matrix $\mathbf{\Lambda}$ of an Erlang distribution can be written as a bidiagonal matrix, while the vectors denoting the initial distribution β and the absorption rates λ 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 take relatively simple forms. These are given in the following Lemma.

Lemma 3.3. *Suppose that $G_m \sim \text{Gamma}(\xi, \xi)$, then*

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

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

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

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

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

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

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

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

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 the average claim amount per unit of time is $\rho = \gamma \mu_F$.

First, we approximate the integrated tail \widehat{F} with an Erlangized scale mixture $\Pi \star G_m$ where Π is an approximating discrete distribution of \widehat{F} , that is, the approach of approximation A. The approximation for $\psi_{\widehat{F}}(u)$ is given next:

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

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

where

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

and

$$\mathcal{B}_i = \sum_{j=1}^{\infty} \frac{\pi_j s_1}{s_j} \text{bin}(\xi - 1; i, s_1 / s_j), \quad \mathcal{C}_n = \sum_{j=1}^{\infty} \pi_j \text{Bin}(\xi - 1; n, s_1 / s_j),$$

where $\text{bin}(\cdot; n, p)$ and $\text{Bin}(\cdot; n, p)$ denote the pdf and cdf respectively of a binomial distribution with parameters n and p .

Proof. The result follows by letting $\rho = \gamma \mu_F$, $\theta = \xi$, $\lambda = \xi$, applying Proposition 3.1 and Lemma 3.3 given in the previous subsection. \square

We propose to use $\psi_{\Pi \star G_m}$ as an approximation of ruin probability $\psi_{\widehat{F}}$. One of the most attractive features of the result above is that because of the simple structure of Erlangized scale mixture it is possible to rewrite the approximation of the ruin probability in simple terms which are free of matrix operations. In particular, the simplified expressions for the values of κ_n given in terms of the binomial distribution are particularly convenient for computational purposes.

As stressed before, for approximation A we sacrifice the interpretation of the approximation $\psi_{\Pi \star G_m}$ as the ruin probability of some Cramér–Lundberg reserve process since it is not possible to easily identify a distribution whose integrated tail corresponds to the Erlangized scale mixture distribution $\Pi \star G_m$. 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 sizes.

As mentioned before, a more common and somewhat natural approach is to approximate the claim size distributions via Erlangized scale mixtures, i.e. approximation B. The following theorem provides an expression for approximation B of the probability of ruin $\psi_{\widehat{F}}$ with the ruin probability of a reserve process having claim sizes $\Pi \star G_m$. This result could be useful for instance in a situation where the integrated tail is not available and it is difficult to compute.

Note that we have modified the intensity of the Poisson process in order to match the average claim amount per unit of time $\rho = \gamma\mu_F$ of the original process. This selection will help to demonstrate uniform convergence.

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

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

where

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

and

$$\mathcal{C}_i = \sum_{j=1}^{\infty} \pi_j \text{Bin}(\xi - 1; i, s_1/s_j), \quad \mathcal{D}_n = \sum_{j=1}^{\infty} \pi_j s_j \sum_{k=0}^{\xi-1} \frac{\xi - k}{\xi} \text{bin}(k; n, s_1/s_j).$$

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

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

Then by induction, we can get for $1 \leq n \leq \xi$,

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

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

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 bounds for the error of 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

$$|\psi_{\widehat{F}}(u) - \psi_{\Pi \star G_m}(u)| \leq \left| \psi_{\widehat{F}}(u) - \psi_{\widehat{F} \star G_m}(u) \right| + \left| \psi_{\widehat{F} \star G_m}(u) - \psi_{\Pi \star G_m}(u) \right|.$$

For approximation B in Theorem 3.5 we have an analogous bound

$$\left| \psi_{\widehat{F}}(u) - \psi_{H_{\Pi} \star \widehat{G}_m}(u) \right| \leq \left| \psi_{\widehat{F}}(u) - \psi_{H_F \star \widehat{G}_m}(u) \right| + \left| \psi_{H_F \star \widehat{G}_m}(u) - \psi_{H_{\Pi} \star \widehat{G}_m}(u) \right|.$$

We will rely on the Pollaczek–Khinchine formula (3.1) for the construction of the bounds. Recall that the formula above is interpreted as the probability that a terminating renewal process having defective renewal probability $\rho \widehat{F}(\cdot)$ will reach level u before terminating. In our two approximations of $\psi_{\widehat{F}}$, we have 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 tail exclusively. For instance, the error of Erlangization in approximation A is given by

$$\left| \psi_{\widehat{F}}(u) - \psi_{\widehat{F} \star G_m}(u) \right| = \left| \sum_{n=1}^{\infty} (1 - \rho) \rho^n \left(\widehat{F}^{*n}(u) - \widehat{F} \star G_m^{*n}(u) \right) \right|. \quad (4.1)$$

Note that $n = 0$ in the above series is equal to zero.

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 given above.

We will divide this section in three parts. In subsection 4.1 we refine an existing bound introduced in Vatamidou et al. (2014) for the error of approximation of the ruin probability. This refined result will be used in the construction of bounds for the error of discretization. In subsections 4.2 and 4.3 we provide bounds for the errors for each of the two approximations proposed.

4.1 General bounds for the error of approximation

The following Theorem provides a refined bound for the error of approximation for the ruin probability provided by Vatamidou et al. (2014).

Theorem 4.1. *For any distributions with positive support \widehat{F}_1 and \widehat{F}_2 and fixed $u > 0$, we have that*

$$|\psi_{\widehat{F}_1}(u) - \psi_{\widehat{F}_2}(u)| \leq \sup_{s < u} \{ |\widehat{F}_1(s) - \widehat{F}_2(s)| \} \frac{(1 - \rho) \rho}{(1 - \rho \widehat{F}_1(u))(1 - \rho \widehat{F}_2(u))}.$$

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

$$\sup_{s < u} \{ |\widehat{F}_1^{*n}(s) - \widehat{F}_2^{*n}(s)| \} \leq \sup_{s < u} \{ |\widehat{F}_1(s) - \widehat{F}_2(s)| \} \sum_{i=0}^{n-1} \widehat{F}_1^i(u) \widehat{F}_2^{n-1-i}(u). \quad (4.2)$$

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

$$\begin{aligned} & \sup_{s < u} \{ |\widehat{F}_1^{*(n+1)}(s) - \widehat{F}_2^{*(n+1)}(s)| \} \\ &= \sup_{s < u} \{ |\widehat{F}_1^{*n+1}(s) - \widehat{F}_1^{*n} * \widehat{F}_2(s) + \widehat{F}_1^{*n} * \widehat{F}_2(s) - \widehat{F}_2^{*n+1}(s)| \} \\ &\leq \sup_{s < u} \{ |\widehat{F}_1^{*n+1}(s) - \widehat{F}_1^{*n} * \widehat{F}_2(s)| \} + \sup_{s < u} \{ |\widehat{F}_1^{*n} * \widehat{F}_2(s) - \widehat{F}_2^{*n+1}(s)| \}. \end{aligned}$$

Clearly,

$$\begin{aligned}
\sup_{s < u} \{|\widehat{F}_1^{*n+1}(s) - \widehat{F}_1^{*n} * \widehat{F}_2(s)|\} &\leq \sup_{s < u} \left\{ \int_0^s |\widehat{F}_1(r) - \widehat{F}_2(r)| d\widehat{F}_1^{*n}(r) \right\} \\
&\leq \sup_{s < u} \left\{ \int_0^s \sup_{l < u} \{|\widehat{F}_1(l) - \widehat{F}_2(l)|\} d\widehat{F}_1^{*n}(r) \right\} \\
&= \sup_{l < u} \{|\widehat{F}_1(l) - \widehat{F}_2(l)|\} \sup_{s < u} \left\{ \int_0^s d\widehat{F}_1^{*n}(r) \right\} \\
&= \sup_{l < u} \{|\widehat{F}_1(l) - \widehat{F}_2(l)|\} \widehat{F}_1^{*n}(u) \\
&\leq \sup_{l < u} \{|\widehat{F}_1(l) - \widehat{F}_2(l)|\} \widehat{F}_1^n(u). \tag{4.3}
\end{aligned}$$

In the last step we have used that $\widehat{F}_1^{*n}(u)$ corresponds to the probability of an event where the sum of n i.i.d. random variables is smaller equal than u while $\widehat{F}_1^n(u)$ corresponds to the probability of the maximum of i.i.d. random variables is smaller 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

$$\begin{aligned}
\sup_{s < u} \{|\widehat{F}_1^{*n} * \widehat{F}_2(s) - \widehat{F}_2^{*n+1}(s)|\} &\leq \sup_{s < u} \left\{ \int_0^s |\widehat{F}_1^{*n}(r) - \widehat{F}_2^{*n}(r)| d\widehat{F}_2(r) \right\} \\
&\leq \sup_{s < u} \left\{ \int_0^s \sup_{l < u} \{|\widehat{F}_1^{*n}(l) - \widehat{F}_2^{*n}(l)|\} d\widehat{F}_2(r) \right\} \\
&= \sup_{l < u} \{|\widehat{F}_1^{*n}(l) - \widehat{F}_2^{*n}(l)|\} \sup_{s < u} \left\{ \int_0^s d\widehat{F}_2(r) \right\} \\
&\leq \left(\sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} \sum_{i=0}^{n-1} \widehat{F}_1^i(u) \widehat{F}_2^{n-1-i}(u) \right) \widehat{F}_2(u) \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} \sum_{i=0}^{n-1} \widehat{F}_1^i(u) \widehat{F}_2^{n-i}(u). \tag{4.4}
\end{aligned}$$

Summing (4.3) and (4.4), we get that

$$\sup_{s < u} \{|\widehat{F}_1^{*n+1}(s) - \widehat{F}_2^{*n+1}(s)|\} \leq \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} \sum_{i=0}^n \widehat{F}_1^i(u) \widehat{F}_2^{n-i}(u),$$

so that formula (4.2) is valid for all $n \geq 1$. Finally,

$$\begin{aligned}
|\psi_{\widehat{F}_1}(u) - \psi_{\widehat{F}_2}(u)| &\leq \sum_{n=1}^{\infty} (1 - \rho) \rho^n |\widehat{F}_1^{*n}(u) - \widehat{F}_2^{*n}(u)| \\
&\leq \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} (1 - \rho) \sum_{n=1}^{\infty} \rho^n \sum_{i=0}^{n-1} \widehat{F}_1^i(u) \widehat{F}_2^{n-1-i}(u) \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} (1 - \rho) \sum_{i=0}^{\infty} \sum_{n=i+1}^{\infty} \rho^n \widehat{F}_1^i(u) \widehat{F}_2^{n-1-i}(u) \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} (1 - \rho) \sum_{i=0}^{\infty} \sum_{n=0}^{\infty} \rho^{n+i+1} \widehat{F}_1^i(u) \widehat{F}_2^n(u) \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} (1 - \rho) \rho \sum_{i=0}^{\infty} \rho^i \widehat{F}_1^i(u) \sum_{n=0}^{\infty} \rho^n \widehat{F}_2^n(u) \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} (1 - \rho) \rho \frac{1}{1 - \rho \widehat{F}_1(u)} \frac{1}{1 - \rho \widehat{F}_2(u)} \\
&= \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} \frac{(1 - \rho) \rho}{(1 - \rho \widehat{F}_1(u))(1 - \rho \widehat{F}_2(u))}.
\end{aligned}$$

□

We remark that the bound given above is a refinement of the result obtained in Vatamidou et al. (2014): The construction of our bound is based on the inequality (4.4) and given by

$$\sup_{s < u} \{|\widehat{F}_1^{*n} * \widehat{F}_2(s) - \widehat{F}_2^{*n+1}(s)|\} \leq \sup_{s < u} \{|\widehat{F}_1(s) - \widehat{F}_2(s)|\} \sum_{i=0}^{n-1} \widehat{F}_1^i(u) \widehat{F}_2^{n-i}(u).$$

The expression on the right hand side takes values in $(0, 1)$ for all values of n . In contrast, the quantity used in Vatamidou et al. (2014) to bound the expression in the left hand side is $n\widehat{F}(u)$, which goes to infinity as $n \rightarrow \infty$. 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

$$\frac{(1 - \rho)^2}{(1 - \rho\widehat{F}_1(u))(1 - \rho\widehat{F}_2(u))} \leq 1.$$

4.2 Error bounds for $\psi_{\Pi \star G_m}$

This subsection is dedicated to the construction of the bounds for approximation A suggested in Theorem 3.4.

4.2.1 Bounds for the Erlangization error of $\psi_{\Pi \star G}$

A bound for the Erlangization error is constructed throughout the following results.

Lemma 4.2. *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*

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

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

Proof.

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

□

An upper bound for the Erlangization error is given next.

Theorem 4.3. *Let $\{\mathcal{A}_k : k \in \mathbb{N}\}$ be a sequence as defined in Lemma 4.2. Then*

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

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

In our numerical experiments we found that it is enough to take a finite number K of sets $\mathcal{A}_1, \dots, \mathcal{A}_K$ to obtain a usable numerical bound. This is equivalent to take $\mathcal{A}_k = \{1\}$ for all $k \geq K$ in the Theorem above.

Proof. The proof follows from Theorem 4.1, Lemma 4.2 and the following observation

$$\frac{1}{1 - \rho \widehat{F} \star G_m(u)} \leq \frac{1}{1 - \rho}, \quad \forall u > 0.$$

To prove uniform convergence we simply note that the expression above can be further bounded above by

$$\left| \psi_{\widehat{F}}(u) - \psi_{\widehat{F} \star G_m}(u) \right| \leq \frac{\rho}{1 - \rho} \sum_{k=0}^{\infty} \sup_{\ell > 0} (\widehat{F}_{b_k}(\ell) - \widehat{F}_{a_k}(\ell)) (G_m(\mathcal{A}_k) - G_m(\mathcal{A}_{k+1})). \quad (4.5)$$

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 $1 - G_m(\mathcal{A}_{k+1}) \leq \epsilon(1 - \rho)/2\rho$. Putting together this results we obtain that for all $k \geq k_0$ and $m \geq m_0$

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

Hence, uniform convergence follows. \square

4.2.2 Bounds for the discretization error of $\psi_{\Pi \star G}$

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 our refinement of Vatamidou et al. (2014)'s bound for the construction of an upper bound for the discretization error.

Theorem 4.4. *Let*

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

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

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

The bound above decreases as Π gets close to \widehat{F} ; this is reflected in the value of η . The bound will become smaller as long as terms $\widehat{F}(u/\delta) + G_m(\delta)$ and $\Pi(u/\delta) + G_m(\delta)$ in the denominator become bigger. The value of δ minimizing this bound can be easily found numerically.

Proof. The result follows from observing that

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

We just apply our refinement of Vatamidou et al. (2014)'s bound provided in Theorem 4.1.

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

A lower bound for $\Pi \star G_m(u)$ can be found in an analogous way. \square

The last step in the construction of an upper bound for the discretization error is finding an upper bound for $\eta = \sup_{0 \leq s \leq u} |\widehat{F} \star G_m(s) - \Pi \star G_m(s)|$. We suggest a bound in the following Proposition.

Proposition 4.5. *Let $0 < \delta < \infty$, then*

$$\sup_{0 \leq s \leq u/\delta} |\widehat{F} \star G_m(s) - \Pi \star G_m(s)| \leq \eta(\delta),$$

where

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

Proof.

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

□

In practice, we would select a value of δ which minimizes the upper bound $\eta(\delta)$. Notice, that if the tail probability of 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 $\psi_{H_\Pi \star \widehat{G}_m}$

Next we turn our attention to approximation B of the ruin probability when the claim size distribution F is approximated via Erlangized scale mixtures. We remark that the bounds presented in this section are simple and sufficient to show uniform convergence. However, these bounds are too rough for practical purposes. A set of more refined bounds can be obtained but their construction and expressions are more complicated, so these have been relegated to the appendix.

4.3.1 Bounds for the Erlangization error of $\psi_{H_\Pi \star \widehat{G}}$

The following theorem provides a first bound for the Erlangization error of the approximation $\psi_{H_F \star \widehat{G}_m}$. A tighter bound for the Erlangization error can be found in the Appendix.

Theorem 4.6.

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

where ϵ_m is defined as in Lemma 4.7.

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

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

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

$$\begin{aligned} \left| \widehat{F}^{*n}(u) - \widehat{F \star G}^{*n}(u) \right| &= \left| (H_F \star U)^{*n}(u) - (H_F \star \widehat{G}_m)^{*n}(u) \right| \\ &\leq \int \cdots \int_{\mathbb{R}^n} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n \\ &\leq \int \cdots \int_{\mathbb{R}^n} \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n. \end{aligned}$$

That the last integral is bounded by $2(1 - (1 - \epsilon_m)^n)$ follows from Corollary 8.4 in the Appendix. Therefore, we have that

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

Lemma 4.7 provides an explicit bound for ϵ_m . \square

The following result provides with an explicit expression useful for obtaining the integrated distance between the survival function $1 - G_m$ and the density of a $U(0, 1)$ distribution. That is

$$\int_0^{\infty} \left| (1 - G_m(s)) - \mathbb{I}_{0,1}(s) \right| ds.$$

Lemma 4.7.

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

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,

$$\int_1^{\infty} (1 - G_m(s)) ds = 1 - \int_0^1 (1 - G_m(s)) ds = \int_0^1 G_m(s) ds = \epsilon_m,$$

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

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

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

Note that the bound for the error provided above only depends on the parameter of the Erlang distribution ξ and the average claim amount per unit of time ρ . This bound does not depend on the initial reserve u , nor the underlying claim size distribution F , so $\psi_{\widehat{F \star \widehat{G}_m}}$ converges uniformly to $\psi_{\widehat{F}}$. However, in practice this bound is too rough and not useful for practical purposes. In Theorem 8.7 we provide a refinement of the bound above. The refined bound proposed in there no longer has a simple form but in return it is much sharper and more useful for practical purposes.

4.3.2 Bounds for the discretization error of $\psi_{H_{\Pi} \star \widehat{G}}$

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

Theorem 4.8. *Let*

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

then

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

An upper bound for $\sup_{0 \leq s \leq u} |H_F \star \widehat{G}_m(s) - H_{\Pi} \star \widehat{G}_m(s)|$, is suggested in the next Proposition.

Proposition 4.9. *For $\delta > 1$ we have that*

$$\sup_{0 \leq s \leq u} |H_F \star \widehat{G}_m(s) - H_{\Pi} \star \widehat{G}_m(s)| \leq \eta(\delta),$$

where

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

The construction of the previous bounds depends on the availability of the distance between moment distributions $|H_F - H_{\Pi}|$, but the later might not always be available. For such a case we suggest a bound for such a quantity in Lemma 8.8 for a specific type of approximating distributions Π . The bound presented in there 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 probability of ruin of a reserve process as given in Theorems 3.4 and 3.5 is not computable in exact form since the expression is 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 compute error bounds for the approximation of the ruin probabilities occurred by truncating those series.

A close inspection of Theorems 3.4 and 3.5 reveals that there will exist two sources of error due to truncation. The ruin probability can be seen as the expected value of κ_N where $N \sim \text{Poisson}(\xi u/s_1)$, so the first error of truncation is $\mathbb{E}[\kappa_N | N \geq N_1]$, we call N_1 the level of truncation for the ruin probability. Since the values of κ_n are bounded above by 1, then it is possible to bound this error term with $\mathbb{P}(N \geq N_1)$ and use Chernoff's bound (cf. Theorem 9.3 Billingsley, 1995) to obtain an explicit expression

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

The second source of numerical error comes from truncating the infinite series induced by the scaling distribution Π ; that is, we need to truncate the series defining the terms \mathcal{B}_i , \mathcal{C}_i and \mathcal{D}_i . The following Lemma shows these truncated series can be bounded by quantities depending on the tail probability of Π and the level of truncation s_{N_2} , where N_2 is the level of truncation for the scaling.

Lemma 5.1. *Let $S \sim \Pi$ and define $\varepsilon_1 = \mathbb{P}(S > s_{N_2})$ and $\varepsilon_2 = \mathbb{E}[S; S > s_{N_2}]$. 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 $\widetilde{\mathcal{B}}_i$, $\widetilde{\mathcal{C}}_i$ and $\widetilde{\mathcal{D}}_i$ denote to the truncated series at N_2 terms.

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

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

Similarly, if $n \geq \xi$ then

$$\mathcal{C}_n - \tilde{\mathcal{C}}_n = \sum_{j=N_2+1}^{\infty} \pi_j \text{Bin}(\xi - 1; n, s_1/s_j) \leq \sum_{j=N_2+1}^{\infty} \pi_j = \varepsilon_1,$$

while

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

□

5.1 Truncation error for $\psi_{\Pi \star G_m}$

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

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

where

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

with

$$\tilde{\mathcal{B}}_i = \frac{\xi}{i+1} \sum_{j=1}^{N_2} \pi_j \text{bin}(\xi; i+1, s_1/s_j), \quad \tilde{\mathcal{C}}_n = \sum_{j=1}^{N_2} \pi_j \text{Bin}(\xi - 1; n, s_1/s_j).$$

Theorem 5.2. *Let $\varepsilon_1 = \mathbb{P}(S > s_{N_2})$. Then*

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

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

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

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

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

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

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

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

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

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

$$\sum_{i=\xi-1}^{n-1} (\kappa_i - \tilde{\kappa}_i)\mathcal{B}_i + (n - \xi + 1)\varepsilon_1. \quad (5.3)$$

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

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

Note that $\sum_{i=\xi}^{\infty} \mathcal{B}_i = 1$ follows from relating the formula of \mathcal{B}_i to the probability mass function of a negative binomial distribution $\text{NP}(\xi, 1 - s_1/s_j)$. Using the hypothesis that $\gamma\mu_F < 1$ and induction it is not difficult to prove that

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

Inserting the bound above into the first term of equation (5.2) and assuming that $\xi u/s_1 > 1$ we arrive at

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

□

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

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

5.2 Truncation error for $\psi_{H_{\Pi^*}\hat{G}_m}$

We write the ruin probability in Theorem 3.5 (approximation B) as a truncated series:

$$\tilde{\psi}_{H_{\Pi^*}\hat{G}_m}(u) = \sum_{n=0}^{N_1} \tilde{\kappa}_n \frac{(\xi u/s_1)^n e^{-\xi u/s_1}}{n!},$$

where

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

with

$$\tilde{\mathcal{C}}_i = \sum_{j=1}^{N_2} \pi_j \text{Bin}(\xi - 1; i, s_1/s_j), \quad \tilde{\mathcal{D}}_n = \sum_{j=1}^{N_2} \pi_j s_j \sum_{k=0}^{\xi-1} \frac{\xi - k}{\xi} \text{bin}(k; n, s_1/s_j).$$

The result and its proof are similar to the previous case.

Theorem 5.4. *Let $S \sim \Pi$ and define $\varepsilon_2 = \mathbb{E}[S; S > s_{N_2}]$. Then*

$$\psi_{H_\Pi \star \tilde{\mathcal{G}}_m}(u) - \tilde{\psi}_{H_\Pi \star \tilde{\mathcal{G}}_m}(u) \leq \varepsilon_2 e^{\frac{\xi u \gamma \mu_F}{s_1 \mu_\Pi}} \left(\frac{\gamma \mu_F}{\mu_\Pi} + 1 \right)^{-\xi} + \zeta(N_1; \xi u/s_1).$$

Building a bound for the numerical error of approximation B is more involved than for approximation A given in Theorem 5.2. The reason is that it is not simple to provide a tight bound for the $\sum_{i=1}^{\infty} \mathcal{C}_i$ as for $\sum_{i=1}^{\infty} \mathcal{B}_i$. Notice that the bound is not as tight as in the case of Theorem 5.2 and may not be of much practical use. This aspect highlights an additional advantage of our first estimator.

Proof. Observe that

$$\kappa_n - \tilde{\kappa}_n = \begin{cases} 0, & 0 \leq n \leq \xi, \\ \frac{\gamma\mu_F}{\mu_\Pi} \left[\frac{s_1}{\xi} \sum_{i=\xi+1}^{n-1} (\kappa_{n-1-i} \mathcal{C}_i - \tilde{\kappa}_{n-1-i} \tilde{\mathcal{C}}_i) + \mathcal{D}_n - \tilde{\mathcal{D}}_n \right], & \xi < n \leq N_1. \end{cases} \quad (5.4)$$

The summation in (5.4) can be rewritten as

$$\sum_{i=\xi+1}^{n-1} (\kappa_{n-1-i} \mathcal{C}_i - \tilde{\kappa}_{n-1-i} \tilde{\mathcal{C}}_i) = \sum_{i=\xi+1}^{n-1} ((\kappa_{n-1-i} - \tilde{\kappa}_{n-1-i}) \mathcal{C}_i + \tilde{\kappa}_{n-1-i} (\mathcal{C}_i - \tilde{\mathcal{C}}_i)).$$

Since $0 < \tilde{\kappa}_i \leq \kappa_i \leq 1$ and $\mathcal{C}_i \leq 1$ for $0 \leq i$ then we can use the second part of Lemma 5.1 to obtain the following bound of the expression above

$$\sum_{i=\xi+1}^{n-1} (\kappa_i - \tilde{\kappa}_i) + (n - \xi - 1)\varepsilon_2. \quad (5.5)$$

Putting (5.5) and the third part of Lemma 5.1 together we arrive at

$$\kappa_n - \tilde{\kappa}_n \leq \frac{\gamma\mu_F}{\mu_\Pi} \left(\frac{s_1}{\xi} \sum_{i=\xi+1}^{n-1} (\kappa_i - \tilde{\kappa}_i) + (n - \xi)\varepsilon_2 \right) \leq \frac{\gamma\mu_F}{\mu_\Pi} \left(\sum_{i=\xi+1}^{n-1} (\kappa_i - \tilde{\kappa}_i) + (n - \xi)\varepsilon_2 \right).$$

Induction yields that

$$\begin{aligned} \frac{\gamma\mu_F}{\mu_\Pi} \left(\sum_{i=\xi+1}^{n-1} (\kappa_i - \tilde{\kappa}_i) + (n - \xi)\varepsilon_2 \right) &= \varepsilon_2 \left(\left(\frac{\gamma\mu_F}{\mu_\Pi} + 1 \right)^{n-\xi} - 1 \right) \\ &\leq \varepsilon_2 \left(\frac{\gamma\mu_F}{\mu_\Pi} + 1 \right)^{n-\xi}. \end{aligned}$$

Hence we arrive at

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

□

6 Numerical implementations

We briefly discuss some relevant aspects of the implementation of Theorems 3.4 and 3.5.

Suppose we want to approximate a distribution \widehat{F} via Erlangized scale mixtures. The selection of the parameter $\xi \in \mathbb{N}$ of the Erlang distribution boils down to selecting a value ξ large enough so the bound provided in Theorem 4.3 and Theorem 4.6 is smaller than a preselected precision. It is however not recommended to select a value which is too large since this will require truncating at higher levels and thus resulting in a much slower algorithm (this will be further discussed below).

The most critical aspect for an efficient implementation is the selection of an appropriate approximating distribution Π . The selection can be made rather arbitrary but 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

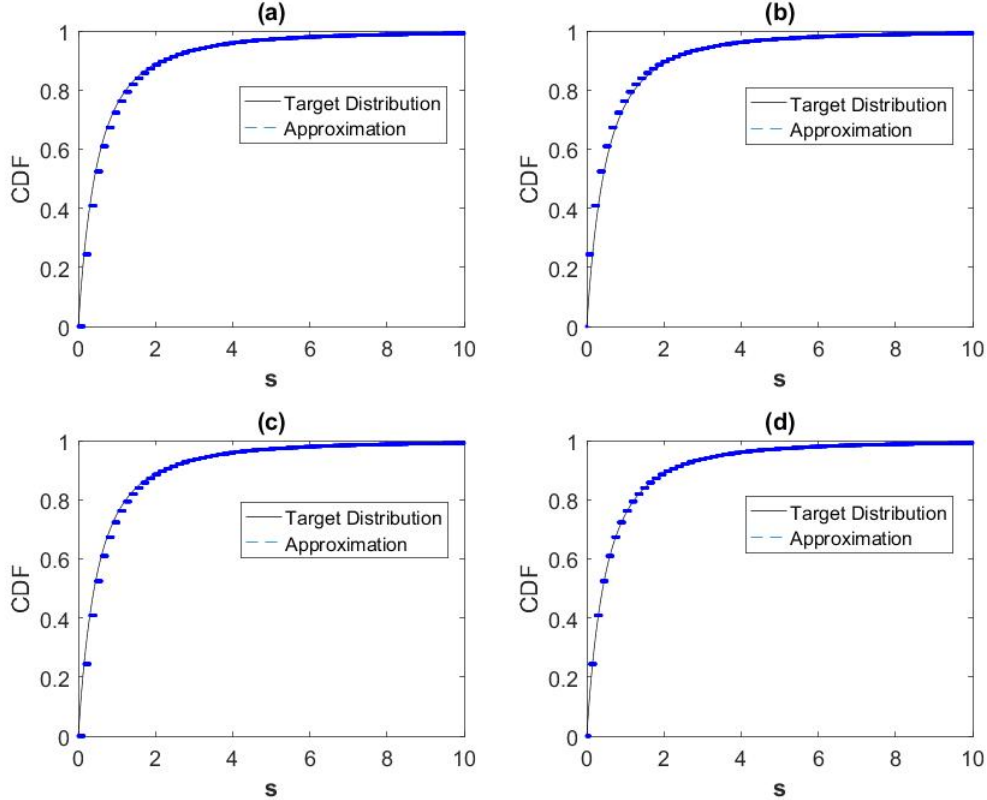
$$\Pi(s) := \sum_{k=0}^{\infty} F(w_k) \mathbb{I}_{[s_k, s_{k+1})}(s).$$

The distribution Π is a discretized approximating distribution which is upcrossed by F in every interval (w_{k-1}, w_k) . This type of approximation is rather general as we can consider general approximations by selecting $s_k \in (w_{k-1}, w_k)$, approximations from *below* by setting $s_k = w_k$, approximations from *above* by setting $s_k = w_{k-1}$, or the middle point (see Figure 2). Heuristically, one might expect to reduce the error of approximation by selecting the middle point (this was our selection in our experimentations below).

In practice we can just compute a finite number of terms π_k , so we end up with an improper distribution. This represents a serious issue because truncating at lower levels affects the quality of the approximation in the tail regions. Computing a larger number of terms is not often an efficient alternative since the computational times become rapidly unfeasible. Thus, our ultimate goal will be to select among the partitions of certain fixed size (we restrict the partition size since we assume we have a limited computational budget), the one that *minimizes* the distance $|F - \Pi|$, in particular in the tails. In our numerical experimentations we found that an arithmetic progression required a prohibitively large number of terms to obtain sharp approximations in the tail. We obtained better results using geometric progressions as these can provide better approximations with a reduced number of terms. Moreover, since the sequence determining the probability mass function converges faster to 0, then it is easier to compute enough terms so for practical purposes it is equivalent to work with a proper distribution.

The speed of the algorithm is heavily determined by the total number of terms of the infinite series in Theorems 3.4 and 3.5 computed. Since the probability of interest can be seen as the expected value $\mathbb{E}[\kappa_N]$ where $N \sim \text{Poisson}(\xi u/s_1)$, it is straightforward to see that the total number N_1 of terms needed to provide an accurate approximation is directly related to the value $\xi u/s_1$. Thus, large values of ξ and u combined with small values of s_1 will require longer computational times. Since smaller values of ξ and larger values of s_1 will typically result in increased errors of approximation, 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 empirically these values with the help of the error bounds found in the previous sections.

Figure 2: Four alternative discretizations of a Pareto(2) distribution. 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.



It is also worth noting that the calculation of the value κ_n for $n > \xi$ in both Theorems 3.4 and 3.5 requires the evaluation of the binomial probability mass functions $\text{bin}(\cdot; i, \cdot)$ for all $i = \xi, \dots, N_1$. 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 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 binomial probabilities.

Finally, we remark that the speed of the implementation can be significantly improved by using parallel computing. In our implementations below we have broken the series into smaller pieces and we have sent this to an HPC (high performance computing) facility to run independent units of work.

6.1 Numerical examples

In this section, we show the accuracy of our approximation A through the following Pareto example. In such an example, the claim sizes are Pareto distributed, so their integrated tails are regularly varying. The exact values of the ruin probability are given in Ramsay (2003), and are now considered a classical benchmark for comparison purposes. We have limited our numerical experiments to the Pareto with parameter 2 and net profit condition close to 0 ($\rho \rightarrow 1$) as this is one of the most challenging ruin probabilities we could find for which there are results available for comparison.

Example 6.2 (Pareto claim sizes). We consider a Cramér–Lundberg model with unit premium

rate, and claim sizes distributed according to a Pareto distribution with a single parameter $\phi > 1$ with support on the positive real axis, mean 1 and having the following cumulative distribution function

$$F(x) = 1 - \left(1 + \frac{x}{\phi - 1}\right)^{-\phi}, \text{ for } x > 0 \text{ and } \phi > 1, \quad (6.1)$$

(other parametrizations of the Pareto distribution are common as well). The integrated tail of the above distribution is regularly varying with parameter $\phi - 1$:

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

The parameters of the risk model selected were $\rho = 0.95$, $\phi = 2$. We implemented the approximation A in Theorem 3.4 and its analysis is presented next. For comparisons purposes we also included the approximation B in Theorem 3.5, but overall we found that it is less accurate, much slower and more difficult to analyze since its bounds are not tight enough.

First we analyzed the Erlangization error for approximation A. For this example, it is possible to compute the bound given by Theorem 4.1 for values of $\xi = 100, 500, 1000$. The bound appears to be tighter for smaller values of ρ while it gets loosen as long as the value of $\rho \rightarrow 1$. The bound also increases as $u \rightarrow \infty$, so probabilities of ruin with large initial reserves will be more difficult to approximate. The bound appears to decrease proportionally in ξ but in practice, we didn't noticed 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 results overall.

Table 1: Erlangization Error Bounds.

u	$\xi=100$	$\xi=500$	$\xi=1000$
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}

Next we constructed a discrete approximating distribution Π by considering a discretized Pareto supported over the geometric progression $\{e^{t_0}, e^{t_1}, e^{t_2}, \dots\}$, where $t_k = t_0 + k/K$. It is rather clear that a finer partition of the interval $[0, \infty)$ would yield a better approximation and this would be attained by letting the value of $t_0 \rightarrow -\infty$ and $K \rightarrow \infty$. However, small values of $s_1 := e^{t_0}$ affect severely the speed of the algorithm (see the discussion above) 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 $K \rightarrow \infty$. For this example we have selected these values empirically with the help of the bound in Theorem 4.4 and Proposition 4.5. We settled with $t_0 = -3$ and $K = 270$ for all the examples. The results are in the first column in the Table 2 below.

Next we selected the truncation levels. In the case of N_1 we were able to select a natural number large enough such that the truncation error was smaller than the floating point precision without increasing significantly the computational times. This selection implies that the third term in the bound for the truncation error given in Theorem 5.2 is eliminated for practical purposes. As for N_2 , we choose the smallest integer N_2 such that $\varepsilon_1 < 9.5701 \times 10^{-14}$. The error bounds are presented in the last column of Table 2. Notice that the dominant term in Theorem 5.2 is asymptotically linear in u . This pattern is also observed numerically as the error bound appears increasing linearly with respect to u , thus providing empirical evidence that suggests this bound is tight.

The numerical results for the probabilities of ruin are now summarized in Table 3. The results show that the approximated ruin probabilities are remarkably close to the true value calculated using equation (20) of Ramsay (2003).

Table 2: Error Bounds

u	Discretization Error	Truncation Error N_2
1	6.2002×10^{-4}	3.6522×10^{-9}
5	6.3557×10^{-5}	1.8261×10^{-8}
10	3.1590×10^{-5}	3.6522×10^{-8}
30	1.0617×10^{-5}	1.0957×10^{-7}
50	6.4216×10^{-6}	1.8261×10^{-7}
100	3.2491×10^{-6}	3.6522×10^{-7}
500	6.6918×10^{-7}	1.8261×10^{-6}
1000	3.3891×10^{-8}	3.6522×10^{-6}

Table 3: Approximation of ruin probabilities when claim sizes are Pareto distributed, $\rho = 0.95$ and $\phi = 2$.

u	Approximation		Ramsay
	Theorem 3.4	Theorem 3.5	
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

The numerical results above were produced with the same values of ξ , t_0 and K . We remark that as long as the value of u increases, then the numerical approximation appears to be less sharp, but this can be improved by increasing the value of K (this would make the partition finer) and to a lesser extent by reducing the value of t_0 (improving the approximation of the target distribution in a vicinity of 0). The approximation was less sensitive to increases in the value of ξ but makes it considerably slower.

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 \hat{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 the quality of this paper. 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.
- 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.
- Su, C. and Y. Chen (2006). On the behavior of the product of independent random variables. *Sci. China Ser. A* 49(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. Vasiou, 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. Vasiou, 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 a refined bound for one of the approximations proposed in the main section. In the second subsection of the appendix we provide an auxiliary result that will be useful for the numerical computation of one of the bounds proposed.

8.1 Refinements for the Erlangization error of $\psi_{H_{\Pi^*}\widehat{G}_m}$

Through Theorem 8.7 we provide a refinement of the bound proposed in Theorem 4.6. This refined bound is much tighter although more difficult to construct and implement. The following preliminary results are needed first.

Lemma 8.1. *For any $\delta > 0$, define*

$$\epsilon_m(\delta) := \int_0^\delta G_m(s) ds.$$

Then

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

Notice that $\epsilon_m(1) = \epsilon_m = 1 - \widehat{G}_m(1)$ while $\epsilon_m(0) = 0$.

Proof. Consider

$$\begin{aligned} \int_0^\delta e^{-\xi s} (\xi s)^n ds &= -\frac{e^{-\xi\delta}}{\xi} (\xi\delta)^n + n \int_0^\delta e^{-\xi s} (\xi s)^{n-1} ds \\ &= \frac{n!}{\xi} - \frac{e^{-\xi\delta}}{\xi} \left(\sum_{i=0}^n \frac{n!}{i!} (\xi\delta)^i \right), \end{aligned}$$

so that

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

□

Lemma 8.2. *Let $0 \leq \delta_1 \leq 1$ and $1 \leq \delta_2 \leq \infty$. Define*

$$A_{\delta_1} = [\delta_1, 1]^n, \quad A^{\delta_1, \delta_2} = [\delta_1, \delta_2]^n \setminus [\delta_1, 1]^n.$$

Then

$$\begin{aligned} &\int_{A_{\delta_1}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n \\ &= (1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^n \\ &\int_{A^{\delta_1, \delta_2}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n \\ &= \left(\widehat{G}_m(\delta_2) - \widehat{G}_m(\delta_1) \right)^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_1) \right)^n. \end{aligned}$$

Proof.

$$\begin{aligned}
& \int_{A_{\delta_1}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n \\
&= \int_{A_{\delta_1}} \cdots \int 1 - \prod_{i=1}^n \widehat{g}_m(s_i) ds_1 \dots ds_n \\
&= (1 - \delta_1)^n - \left(\int_{\delta_1}^1 \widehat{g}_m(s) ds \right)^n \\
&= (1 - \delta_1)^n - \left(\int_{\delta_1}^1 1 - G_m(s) ds \right)^n \\
&= (1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m + \epsilon_m(\delta_1))^n.
\end{aligned}$$

For the second equality, notice that

$$\int_{[\delta_1, \delta_2]^n} \cdots \int \prod_{i=1}^n \widehat{g}_m(s_i) ds_1 \dots ds_n = \left(\int_{\delta_1}^{\delta_2} \widehat{g}_m(s) ds \right)^n = \left(\widehat{G}_m(\delta_2) - \widehat{G}_m(\delta_1) \right)^n,$$

so that

$$\begin{aligned}
\int_{A^{\delta_1, \delta_2}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n &= \int_{A^{\delta_1, \delta_2}} \cdots \int \prod_{i=1}^n \widehat{g}_m(s_i) ds_1 \dots ds_n \\
&= \int_{[\delta_1, \delta_2]^n \setminus [\delta_1, 1]^n} \cdots \int \prod_{i=1}^n \widehat{g}_m(s_i) ds_1 \dots ds_n \\
&= \left(\widehat{G}_m(\delta_2) - \widehat{G}_m(\delta_1) \right)^n - \left(\widehat{G}_m(1) - \widehat{G}_m(\delta_1) \right)^n.
\end{aligned}$$

□

Corollary 8.3. *Fix $\delta_2 \in (1, \infty)$. Then there exists $\delta_1 \in [0, 1]$ such that*

$$\begin{aligned}
& \int_{A_{\delta_1}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n \\
&= \int_{A^{\delta_1, \delta_2}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n,
\end{aligned}$$

where $A_{\delta_1} = (\delta_1, 1)^n$ and $A^{\delta_1, \delta_2} = (\delta_1, \delta_2)^n \setminus (\delta_1, 1)^n$.

Proof. Define the following functions with domain $[0, 1]$:

$$\begin{aligned}
p(\delta) &:= \int_{A_\delta} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n, \quad \text{and} \\
q_{\delta_2}(\delta) &:= \int_{A^{\delta, \delta_2}} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \dots ds_n.
\end{aligned}$$

By Lemma 8.2, both functions are continuous, p is non-increasing and q_{δ_2} is non-decreasing. The image of q_{δ_2} is contained in $[0, 1 - \epsilon_m(1)]$ while the image of p is exactly $[0, 1 - \epsilon_m(1)]$. All these mean that there exists a point $\delta_1 \in [0, 1]$ such that $q_{\delta_2}(\delta_1) = p(\delta_1)$, concluding the proof. □

The following Corollary follows immediately from Lemma 8.2 by setting $\delta_1 = 0$ and $\delta_2 = \infty$. This Corollary is needed in the proof of Theorem 4.6.

Corollary 8.4.

$$\int_{\mathbb{R}^n} \cdots \int \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n = 2(1 - (1 - \epsilon_m)^n).$$

The following provides a simple bound between the difference of the n -th convolution of any distribution function F with density f evaluated at two different points.

Lemma 8.5. *Let \widehat{F} be any continuous distribution function supported on $[0, \infty)$ with density function f and fix $b > a > 0$. Then*

$$|\widehat{F}^{*n}(a) - \widehat{F}^{*n}(b)| \leq \Delta_{a,b}^{\widehat{F}} \cdot \widehat{F}^{n-1}(b),$$

where $\Delta_{a,b}^{\widehat{F}} := \sup \left\{ \widehat{F}(b - a + c) - \widehat{F}(c) : c \in (0, a) \right\}$.

Proof.

$$\begin{aligned} |\widehat{F}^{*n}(a) - \widehat{F}^{*n}(b)| &= \int_a^b f^{*n}(u) du = \int_a^b \int_0^u f(u-s) f^{*(n-1)}(s) ds du \\ &\leq \int_a^b \int_0^b f(u-s) f^{*(n-1)}(s) ds du = \int_0^b f^{*(n-1)}(s) \int_a^b f(u-s) du ds \end{aligned}$$

Then there exists a constant $c \in (0, a)$ such that the previous expression is equal to

$$\left\{ \int_a^b f(u-c) du \right\} \int_0^b f^{*(n-1)}(s) ds = \left\{ \int_{a-c}^{b-c} f(u) du \right\} \widehat{F}^{*n-1}(b) \leq \left\{ \int_c^{(b-a)+c} f(u) du \right\} \widehat{F}^{n-1}(b).$$

The result follows from taking the supremum over c . \square

Lemma 8.6. *Let $\{X'_n\}$ be a sequence of i.i.d. random variables with common distribution H_F . Fix $\delta_2 \in (1, \infty)$ and let $\delta_1 \in (0, 1)$ be as in Corollary 8.3. Then*

$$\begin{aligned} &\sum_{n=0}^{\infty} (1-\rho) \rho^n \int_{A_{\delta_1} \cup A^{\delta_1, \delta_2}} \cdots \int \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{[0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \\ &= (1-\rho) \rho \Delta_{u/\delta_2, u/\delta_1}^{H_F} \left(\frac{\epsilon_m(1) - \epsilon_m(\delta_1)}{(1-\rho F(u\delta_1))(1-\rho F(u\delta_1)(1-\epsilon_m(1) + \epsilon_m(\delta_1)))} \right). \end{aligned}$$

Proof. Clearly, for any $(r_1, \dots, r_n) \in A_{\delta_1}$ and $(s_1, \dots, s_n) \in A^{\delta_1, \delta_2}$, we have that

$$\begin{aligned} \mathbb{P}(r_1 X'_1 + \cdots + r_n X'_n \leq u) &\leq H_F^{*n}(u/\delta_1), \quad \text{and} \\ \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) &\geq H_F^{*n}(u/\delta_2), \end{aligned}$$

so that

$$\begin{aligned}
& \left| \int \cdots \int_{A_{\delta_1} \cup A^{\delta_1, \delta_2}} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left(\prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right) ds_1 \cdots ds_n \right| \\
&= \left| \int \cdots \int_{A_{\delta_1}} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \right. \\
&\quad \left. - \int \cdots \int_{A^{\delta_1, \delta_2}} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \right| \\
&\leq \left| H_F^{*n}(u/\delta_1) \int \cdots \int_{A_{\delta_1}} \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \right. \\
&\quad \left. - H_F^{*n}(u/\delta_2) \int \cdots \int_{A^{\delta_1, \delta_2}} \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \right| \\
&= (H_F^{*n}(u/\delta_1) - H_F^{*n}(u/\delta_2)) \int \cdots \int_{A_{\delta_1}} \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \\
&= (H_F^{*n}(u/\delta_1) - H_F^{*n}(u/\delta_2)) [(1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^n].
\end{aligned}$$

Using the previous results and Lemma 4.2 we get that

$$\begin{aligned}
& \sum_{n=0}^{\infty} (1 - \rho) \rho^n \int \cdots \int_{A_{\delta_1} \cup A^{\delta_1, \delta_2}} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{[0,1]}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n \\
&\leq \sum_{n=0}^{\infty} (1 - \rho) \rho^n \{H_F^{*n}(u/\delta_1) - H_F^{*n}(u/\delta_2)\} [(1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^n] \\
&\leq \sum_{n=1}^{\infty} (1 - \rho) \rho^n \{\Delta_{u/\delta_2, u/\delta_1}^{H_F} H_F^{*(n-1)}(u/\delta_1)\} [(1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^n] \\
&\leq \sum_{n=1}^{\infty} (1 - \rho) \rho^n \{\Delta_{u/\delta_2, u/\delta_1}^{H_F} H_F^{(n-1)}(u/\delta_1)\} [(1 - \delta_1)^n - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^n] \\
&= (1 - \rho) \rho \Delta_{u/\delta_2, u/\delta_1}^{H_F} \sum_{n=0}^{\infty} \rho^n H_F^n(u/\delta_1) [(1 - \delta_1)^{n+1} - (1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))^{n+1}] \\
&= (1 - \rho) \rho \Delta_{u/\delta_2, u/\delta_1}^{H_F} \left(\frac{1 - \delta_1}{1 - \rho H_F(u/\delta_1)(1 - \delta_1)} - \frac{1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1)}{1 - \rho H_F(u/\delta_1)(1 - \delta_1 - \epsilon_m(1) + \epsilon_m(\delta_1))} \right).
\end{aligned}$$

□

Theorem 8.7. For any fixed $\delta_2 \in (1, \infty)$ let δ_1 be as in Corollary 8.3. Then

$$\left| \psi_{H_F \star U}(u) - \psi_{H_F \star \widehat{G}_m}(u) \right| \leq (1 - \rho) \rho \Delta_{u/\delta_2, u/\delta_1}^{H_F} \mathcal{T}_1 + \frac{2\rho\epsilon_m}{1 - \rho(1 - \epsilon_m)} - (1 - \rho) \mathcal{T}_2.$$

where $\Delta_{a,b} := \sup_{0 \leq s \leq a} \{H_F(s + (b - a)) - H_F(s)\}$, $\epsilon_m(\delta) = \int_0^\delta G_m(s) ds$ and

$$\begin{aligned}
\mathcal{T}_1 &:= \frac{1 - \delta_1}{1 - \rho H_F(u/\delta_1)(1 - \delta_1)} - \frac{1 - \delta_1 - \epsilon_m + \epsilon_m(\delta_1)}{1 - \rho H_F(u/\delta_1)(1 - \delta_1 - \epsilon_m + \epsilon_m(\delta_1))} \\
\mathcal{T}_2 &:= \frac{1}{1 - (1 - \delta_1)\rho} - \frac{2}{1 - (1 - \delta_1 - \epsilon_m + \epsilon_m(\delta_1))\rho} + \frac{1}{(\widehat{G}_m(\delta_2) - 1 + \epsilon_m)\rho}.
\end{aligned}$$

The construction of this particular bound requires the selection of two values δ_1 and δ_2 provided in Corollary 8.3. In general, it will not be possible to write down a closed-form expression for such values but in practice this can be easily determined numerically. Recall that an explicit expression for the term $\epsilon_m(\delta)$ can be found in Lemma 8.1.

Proof. Recall that

$$\begin{aligned} \left| \widehat{F}^{*n}(u) - \widehat{F} \star \widehat{G}^{*n}(u) \right| &= \left| (H_F \star U)^{*n}(u) - (H_F \star \widehat{G}_m)^{*n}(u) \right| \\ &\leq \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} \mathbb{P}(s_1 X'_1 + \cdots + s_n X'_n \leq u) \left| \prod_{i=1}^n \mathbb{I}_{(0,1)}(s_i) - \prod_{i=1}^n \widehat{g}_m(s_i) \right| ds_1 \cdots ds_n. \end{aligned}$$

Split the last integral in two parts: over $[\delta_1, \delta_2]^n$ and over $[0, \infty)^n \setminus [\delta_1, \delta_2]^n$, bound the first one using Lemma 8.6 and the second one using Lemma 8.2 and Corollary 8.4. Then apply Pollaczek–Khinchine formula and sum the geometric series. \square

8.2 Bound for $|H_F - H_\Pi|$

As stated in subsection 4.3.2, the result of Theorem 4.8 depends on the availability of $|H_F - H_\Pi|$. 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 8.8. *Let Π be defined as in Definition 6.1 and define $\Delta_k H_F := H_F(s_k) - H_F(s_{k-1})$. Then*

$$\begin{aligned} \sup_{u \leq s < \infty} |\overline{H}_F(s) - \overline{H}_\Pi(s)| &\leq \sup_{K \leq k < \infty} \Delta_k H_F + \frac{|\mu_\Pi - \mu_F| \cdot \mathbb{E}[S; S > u]}{\mu_\Pi \mu_F} + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F}, \\ \sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| &\leq \sup_{0 \leq k \leq K} \Delta_k H_F + \frac{|\mu_\Pi - \mu_F| \cdot \mathbb{E}[S; S \leq u]}{\mu_\Pi \mu_F} + \frac{|\mathbb{E}[S; S \leq u] - \mathbb{E}[X; X \leq u]|}{\mu_F}. \end{aligned}$$

Moreover, if $\mu_F = \mu_\Pi$, then

$$\begin{aligned} \sup_{u \leq s < \infty} |\overline{H}_F(s) - \overline{H}_\Pi(s)| &\leq \sup_{K \leq k < \infty} \Delta_k H_F + \frac{|\mathbb{E}[X; X > s_K] - \mathbb{E}[S; S > s_K]|}{\mu_F}, \\ \sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| &\leq \sup_{0 \leq k \leq K} \Delta_k H_F + \frac{|\mathbb{E}[S; S \leq u] - \mathbb{E}[X; X \leq u]|}{\mu_F}. \end{aligned}$$

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 $\xi(m) \rightarrow \infty$, then $\epsilon_m \rightarrow 0$, 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 .

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

$$\begin{aligned} \left| H_F \star \widehat{G}_m(u) - H_\Pi \star \widehat{G}_m(u) \right| &= \left| \int_0^\infty H_F(u/s) d\widehat{G}(s) - \int_0^\infty H_\Pi(u/s) d\widehat{G}(s) \right| \\ &\leq \int_0^\infty |H_F(u/s) - H_\Pi(u/s)| d\widehat{G}(s) \\ &\leq \sup_{u \leq s < \infty} |H_F(s) - H_\Pi(s)| \int_0^1 d\widehat{G}(s) \\ &\quad + \sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| \int_1^\infty d\widehat{G}(s) \\ &= \sup_{u \leq s < \infty} |\overline{H}_\Pi(s) - \overline{H}_F(s)| \int_0^1 d\widehat{G}(s) \tag{8.1} \\ &\quad + \sup_{0 < s \leq u} |H_F(s) - H_\Pi(s)| \int_1^\infty d\widehat{G}(s). \tag{8.2} \end{aligned}$$

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

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

Using the previous identity we first constructing a bound for (8.1).

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

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

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

Next observe that

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

Therefore,

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

Our construction for the bound for (8.2) 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

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

Therefore,

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

□