Research Article

Applied Stochastic Models in Business and Industry

Received 18 December 2009, Revised 19 January 2010, Accepted 19 January 2010 Published online 24 March 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.838

Risk modelling with the mixed Erlang distribution

Gordon E. Willmot[a] and X. Sheldon Lin[b*†]

A review of analytical and computational properties of the mixed Erlang distribution is given in the context of risk analysis. Basic distributional properties are discussed, and examples of members of the class are provided. Its use in aggregate claims, stop-loss analysis, and risk measures is considered, as are applications in ruin theoretic analysis of the surplus process. Statistical estimation of model parameters is then discussed. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: failure rate; mean residual lifetime; excess loss; equilibrium distribution; stop-loss; VaR; TVaR; total claims distribution; random sum; compound distribution; ruin theory; deficit at ruin; finite time ruin; M/G/1 queue; EM algorithm

1. Introduction

In this paper we review many of the useful analytic and computational properties of the class of countable mixtures of Erlang distributions in the context of risk analysis and management.

Motivation for the use of the mixed Erlang class is threefold. The class is extremely flexible in terms of possible shapes of the probability density function (pdf) of its members, and is capable of multimodality as well as a wide variety of degrees of skewness in the right tail, often a region of particular interest for risk management purposes. In fact, any positive continuous distribution may be approximated to an arbitrary degree of accuracy by a member of the mixed Erlang class [1, pp. 163–164], i.e. the mixed Erlang class is dense in the set of positive continuous probability measures. Second, the mixed Erlang class is much larger than was previously thought, and includes many distributions whose membership in the class is not immediately obvious (such as the generalized Erlang distribution, used as an interclaim time distribution by Gerber and Shiu [2]). See Willmot and Woo [3] for further details. In fact, Shanthikumar [4] demonstrated that the class of phase-type distributions (e.g. [5]) is included in the mixed Erlang class. See also Stanford and Zadeh [6]. Third, a wide variety of quantities of interest for aggregate claims, stop-loss, and ruin theoretic analysis may be computed in a straightforward manner using infinite series, which invariably involves elementary mathematical quantities only.

The class of phase-type distributions and the class of combinations of exponentials (e.g. [7, 8]) share with the mixed Erlang class many of the advantages of flexibility of shape as well as the availability of tractable computational formulae in many of these situations. A notable advantage of the mixed Erlang class is the availability of computational formulae for finite time ruin probabilities (e.g. [9]). Moreover, the infinite series methodology available for the mixed Erlang class typically avoids the root finding approach needed for partial fraction expansion of Laplace transforms as well as for identification of eigenvalues required for evaluation of matrix exponentials.

For $\beta > 0$ and $j \varepsilon \{1, 2, \ldots\}$, let

$$e_j(y) = \frac{\beta(\beta y)^{j-1} \mathrm{e}^{-\beta y}}{(j-1)!}, \quad y > 0, \tag{1}$$

[a]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1
[b]Department of Statistics, University of Toronto, Toronto, Ont., Canada M5S 3G3
*Correspondence to: X. Sheldon Lin, Department of Statistics, University of Toronto, Toronto, Ont., Canada M5S 3G3.
†E-mail: sheldon@utstat.utoronto.ca

Contract/grant sponsor: Natural Sciences and Engineering Research Council of Canada
Contract/grant sponsor: Munich Reinsurance Company of Canada

be the pdf of an Erlang$-j$ distribution, and $\{q_1, q_2, \ldots\}$ be a discrete probability distribution. Then

$$f(y)=\sum_{j=1}^{\infty} q_j \frac{\beta(\beta y)^{j-1}\mathrm{e}^{-\beta y}}{(j-1)!}=\sum_{j=1}^{\infty} q_j e_j(y), \quad y>0, \tag{2}$$

is the pdf of an Erlang mixture, or mixture of Erlangs. If $q_1=1$ then (2) is an exponential pdf, whereas if $q_j=1$ then (2) reduces to the Erlang$-j$ pdf (1). Other examples with simple q_j's include the $E_{1,k}$ random variable with $q_k=1-q_1$ and the $E_{k,k+1}$ random variable with $q_{k+1}=1-q_k$ (both discussed by Tijms [1]), and the noncentral chi-squared random variable (e.g. [10, Chapter 29]) with 2 m degrees of freedom and noncentrality parameter λ, in which case $\beta=\frac{1}{2}$ and $q_j=(\lambda/2)^{j-m}\mathrm{e}^{-\lambda/2}/(j-m)!$ for $j=m, m+1, \ldots$, and $q_j=0$ otherwise. Other members of the mixed Erlang class are discussed in Section 2.

The Laplace transform of (2) is

$$\tilde{f}(s)=\int_0^{\infty} \mathrm{e}^{-sy} f(y)\,\mathrm{d}y=Q\left(\frac{\beta}{\beta+s}\right) \tag{3}$$

where

$$Q(z)=\sum_{j=1}^{\infty} q_j z^j \tag{4}$$

is the probability generating function (pgf) of $\{q_1, q_2, \ldots\}$. Clearly, (3) implies that the mixed Erlang distribution may be viewed as the compound distribution of a random sum of exponential (with mean $1/\beta$) random variables. Furthermore, the key argument employed in Section 2 to demonstrate that a distribution is of mixed Erlang form is the representation of its Laplace transform in the form (3) for $\beta>0$ judicially chosen and $Q(z)$ a pgf.

In Section 3 distributional properties associated with loss analysis are discussed. In particular, excess loss (or equivalently residual lifetime) distributions and their means are discussed due to their relevance in the presence of a deductible. Similarly, aggregate claims and stop-loss moments are considered. Equilibrium distributions are also considered due to their relationships to residual lifetimes and stop-loss moments, and also because of their importance in ruin theory.

Ruin theory is considered in Section 4, where infinite time ruin probabilities and the deficit at ruin are considered in the Sparre Andersen (renewal risk) model. In the important special case involving the classical compound Poisson risk model, finite time ruin probabilities are also briefly discussed.

Finally, in Section 5, fitting of Erlang mixtures to data using maximum likelihood estimation via the EM algorithm is discussed (e.g. [11]). In this case, it is assumed that the mixing distribution is finite, i.e. $q_k=0$ for $k>r$ where $r\varepsilon\{1, 2, \ldots\}$, and the parameters to be estimated are $\beta, q_1, q_2, \ldots, q_r$.

2. Nontrivial examples of Erlang mixtures

In this section we demonstrate that many distributions are of mixed Erlang form, essentially by changing the scale parameter. We begin with the following algebraic identity:

$$\frac{\beta_1}{\beta_1+s}=\frac{\beta}{\beta+s}\left\{\frac{\frac{\beta_1}{\beta}}{1-\left(1-\frac{\beta_1}{\beta}\right)\frac{\beta}{\beta+s}}\right\} \tag{5}$$

which is of interest when $0<\beta_1\leqslant\beta<\infty$. The left-hand side of (5) is the Laplace transform of an exponential (with mean $1/\beta_1$) random variable, and the right-hand side may be expressed as $Q(\beta/(\beta+s))$ where $Q(z)=z(\beta_1/\beta)/\{1-(1-\beta_1/\beta)z\}$. For $\beta_1<\beta$ this expresses the well-known result that a zero-truncated geometric sum of exponential random variables is again exponential (e.g. [12, pp. 74–75]). The following example provides an immediate application of (5).

Example 1 (*mixture of two exponentials*)
Suppose that (without loss of generality) $\beta_1<\beta_2$, $0<p<1$ and

$$f(y)=p\beta_1\mathrm{e}^{-\beta_1 y}+(1-p)\beta_2\mathrm{e}^{-\beta_2 y}, \quad y>0.$$

Then

$$\tilde{f}(s)=p\frac{\beta_1}{\beta_1+s}+(1-p)\frac{\beta_2}{\beta_2+s},$$

and using (5) with β replaced by β_2, it follows that

$$\tilde{f}(s)=\frac{\beta_2}{\beta_2+s}\left\{(1-p)+p\frac{\dfrac{\beta_1}{\beta_2}}{1-\left(1-\dfrac{\beta_1}{\beta_2}\right)\dfrac{\beta_2}{\beta_2+s}}\right\}.$$

That is, $\tilde{f}(s)=Q(\beta_2/(\beta_2+s))$ where

$$Q(z)=z\left\{(1-p)+p\frac{\dfrac{\beta_1}{\beta_2}}{1-\left(1-\dfrac{\beta_1}{\beta_2}\right)z}\right\}.$$

As

$$Q(z)=\left\{(1-p)+p\left(\frac{\beta_1}{\beta_2}\right)\right\}z+p\sum_{j=2}^{\infty}\left(\frac{\beta_1}{\beta_2}\right)\left(1-\frac{\beta_1}{\beta_2}\right)^{j-1}z^j,$$

it follows by extracting the coefficient of z^j that $f(y)$ is of the form (2) with $\beta=\beta_2$, $q_1=(1-p)+p(\beta_1/\beta_2)$, and $q_j=p(\beta_1/\beta_2)(1-\beta_1/\beta_2)^{j-1}$ for $j=2,3,\ldots$.

The idea of utilizing the same scale parameter via (5) in the previous example may be generalized substantially, as in the following example.

Example 2 (*countable mixtures of Erlangs*)
Now assume that

$$f(y)=\sum_{i=1}^{n}\sum_{k=1}^{\infty}p_{ik}\frac{\beta_i(\beta_i y)^{k-1}\mathrm{e}^{-\beta_i y}}{(k-1)!}$$

where $n\epsilon\{2,3,\ldots\}$, $p_{ik}\geqslant 0$ for all i and k, and $\sum_{i=1}^{n}\sum_{k=1}^{\infty}p_{ik}=1$. We remark that arbitrary finite mixtures may be obtained with $p_{ik}=0$ for all sufficiently large k. Assuming that $\beta_i<\beta_n$ for $i<n$, (5) may be used with β_1 replaced by β_i and β by β_n for each $i=1,2,\ldots,n$, to express the Laplace transform

$$\tilde{f}(s)=\sum_{i=1}^{n}\sum_{k=1}^{\infty}p_{ik}\left(\frac{\beta_i}{\beta_i+s}\right)^k$$

in the form (3) with $\beta=\beta_n$ and

$$Q(z)=\sum_{i=1}^{n}\sum_{k=1}^{\infty}p_{ik}z^k\left\{\frac{\dfrac{\beta_i}{\beta_n}}{1-\left(1-\dfrac{\beta_i}{\beta_n}\right)z}\right\}^k.$$

Thus, $f(y)$ may be re-expressed as

$$f(y)=\sum_{j=1}^{\infty}q_j\frac{\beta_n(\beta_n y)^{j-1}\mathrm{e}^{-\beta_n y}}{(j-1)!},\quad y>0,$$

where (see [3, p. 103], for further details)

$$q_j=\sum_{i=1}^{n}\sum_{k=1}^{j}p_{ik}\binom{j-1}{k-1}\left(\frac{\beta_i}{\beta_n}\right)^k\left(1-\frac{\beta_i}{\beta_n}\right)^{j-k},\quad j=1,2,\ldots .$$

We remark that $n=\infty$ is possible as long as $\sup_k\beta_k<\infty$.

The generalized Erlang distribution (e.g. [2]) is also an Erlang mixture, as is now demonstrated.

Example 3 (*generalized Erlang distribution*)
Let X_i have an exponential distribution with mean $1/\beta_i$ for $i=1,2,\ldots,n$, and assume that the X_i's are statistically independent. Then the Laplace transform of the sum $X_1+X_2+\cdots+X_n$ is given by $\tilde{f}(s)=\prod_{i=1}^{n}(\beta_i/(\beta_i+s))$. Again assuming that $\beta_i<\beta_n$ for $i<n$, (5) may be used with β_1 replaced by β_i and β by β_n to express $\tilde{f}(s)$ in the form (3) with β replaced by β_n and

$$Q(z)=z^n\prod_{i=1}^{n-1}\frac{\frac{\beta_i}{\beta_n}}{1-\left(1-\frac{\beta_i}{\beta_n}\right)z}.$$

Therefore the pdf $f(y)$ is given by

$$f(y)=\sum_{j=1}^{\infty}q_j\frac{\beta_n(\beta_n y)^{j-1}\mathrm{e}^{-\beta_n y}}{(j-1)!},\quad y>0,$$

where, assuming that the β_i's are all distinct (this restriction is removed in the next example), $q_j=0$ for $j<n$ and

$$q_j=\sum_{i=1}^{n-1}w_i\frac{\beta_i}{\beta_n}\left(1-\frac{\beta_i}{\beta_n}\right)^{j-n},\quad j=n,n+1,\ldots,$$

with

$$w_i=\prod_{\substack{k=1\\k\neq i}}^{n-1}\frac{\frac{\beta_k}{\beta_n-\beta_k}}{\frac{\beta_k}{\beta_n-\beta_k}-\frac{\beta_i}{\beta_n-\beta_i}},\quad i=1,2,\ldots,n-1.$$

See Willmot and Woo [3, pp. 101–102] for details.

The following example appears to be similar to that of Example 3, but with two important differences. First, the distribution is not necessarily of phase-type or a combination of exponentials. Second, there is no simple representation for the q_j's in general, but they may be obtained numerically in a straightforward manner.

Example 4 (*a sum of gammas*)
Consider the generalization of Example 3 where the Laplace transform of $f(y)$ is given by

$$\tilde{f}(s)=\prod_{i=1}^{n}\left(\frac{\beta_i}{\beta_i+s}\right)^{\alpha_i},$$

corresponding to the distribution of the sum $X_1+X_2+\cdots X_n$, with the X_i's being independent random variables, and X_i has the gamma pdf $\beta_i(\beta_i y)^{\alpha_i-1}\mathrm{e}^{-\beta_i y}/\Gamma(\alpha_i)$. We assume that the α_i's are positive (not necessarily integers), but the sum $m=\sum_{i=1}^{n}\alpha_i$ is assumed to be a positive integer (we remark that ruin probabilities in the classical compound Poisson risk model with a sum of gamma claim amounts (even if m is not an integer) may be computed using the Shiu expansion, as described in [3, pp. 113–114]. Assuming that $\beta_i<\beta_n$ for $i<n$, it follows from (5) that $\tilde{f}(s)=Q(\beta_n/(\beta_n+s))$ where

$$Q(z)=z^m\prod_{i=1}^{n-1}\left\{\frac{\frac{\beta_i}{\beta_n}}{1-\left(1-\frac{\beta_i}{\beta_n}\right)z}\right\}^{\alpha_i}.$$

The probabilities $\{q_1,q_2,\ldots\}$ correspond to convolutions of negative binomial probabilities, shifted to the right by m. Thus, simple analytic formulas for $\{q_1,q_2,\ldots\}$ may be derived in some cases, such as when $\alpha_i=1$ for all i (as given in Example 3) or when $n=2$. In general, however, it follows that $q_j=0$ for $j<m$, $q_m=\prod_{i=1}^{n-1}(\beta_i/\beta_n)^{\alpha_i}$, and $\{q_{m+1},q_{m+2},\ldots\}$ may be computed using the Panjer-type recursion (e.g. [3, p. 105])

$$q_j=\frac{1}{j-m}\sum_{k=1}^{j-m}\left\{\sum_{i=1}^{n-1}\alpha_i\left(1-\frac{\beta_i}{\beta_n}\right)^k\right\}q_{j-k},\quad j=m+1,m+2,\ldots.$$

The following example is somewhat different than the previous examples in that it does not make direct use of (5).

Example 5 (*an inflation model*)
If claims are subject to inflation at (net) force of interest δ in a compound mixed Poisson process, the effective claim size pdf over $(0, t)$ may be expressed (e.g. [13, p. 213] or [14]) as

$$f(y) = \frac{P(e^{\delta t} y) - P(y)}{\delta t y}, \quad y > 0,$$

where $P(y)$ is the claim size distribution function (df) in the absence of inflation. If $P(y) = 1 - e^{-\mu y}$, then

$$f(y) = \frac{e^{-\beta(1-\phi)y} - e^{-\beta y}}{-y \ln(1-\phi)}, \quad y > 0,$$

where $\beta = \mu e^{\delta t}$ and $\phi = 1 - e^{-\delta t}$ if $\delta > 0$, and $\beta = \mu$ and $\phi = 1 - e^{\delta t}$ if $\delta < 0$. Using the usual Taylor series expansion for the exponential function results in

$$f(y) = \sum_{j=1}^{\infty} q_j \frac{\beta(\beta y)^{j-1} e^{-\beta y}}{(j-1)!}, \quad y > 0,$$

where

$$q_j = \frac{\phi^j}{-j \ln(1-\phi)}, \quad j = 1, 2, \ldots,$$

is of logarithmic series form.

Distributional properties of mixed Erlangs as they pertain to models for losses are considered next.

3. Loss model properties and risk measures

The df $F(y) = 1 - \bar{F}(y) = \int_0^y f(x)\,dx$ is

$$\bar{F}(y) = e^{-\beta y} \sum_{k=0}^{\infty} \bar{Q}_k \frac{(\beta y)^k}{k!}, \quad y > 0, \tag{6}$$

where $\bar{Q}_k = \sum_{j=k+1}^{\infty} q_j$ for $k = 0, 1, 2, \ldots$.

We remark that value at risk (VaR) is a commonly used risk measure. At confidence level p it is the $100p$th percentile of the loss distribution. In other words, VaR equals v_p, where $\bar{F}(v_p) = 1 - p$. Clearly, from (6) v_p can be easily computed numerically.

As the mixed Erlang distribution is a compound distribution with exponential claims, the asymptotic estimate of Embrechts *et al.* [15] applies. We use the notation $a(x) \sim b(x)$, $x \to \infty$, to mean $\lim_{x\to\infty} a(x)/b(x) = 1$. The function $C(x)$ varies slowly at infinity if $C(tx) \sim C(x)$, $x \to \infty$, for all $t > 0$. If $q_j \sim C(j) j^{\alpha} \phi^j$, $j \to \infty$, where $C(x)$ varies slowly at infinity, $-\infty < \alpha < \infty$, and $0 < \phi < 1$, then

$$\bar{F}(y) \sim \frac{\beta^{\alpha} \phi^{\alpha+1}}{1-\phi} C(y) y^{\alpha} e^{-\beta(1-\phi)y}, \quad y \to \infty. \tag{7}$$

As (1) yields $\int_0^{\infty} y^k e_j(y)\,dy = \beta^{-k}(k+j-1)!/(j-1)!$ for $k = 1, 2, 3, \ldots$, it follows that the kth moment of the mixed Erlang is

$$\int_0^{\infty} y^k f(y)\,dy = \beta^{-k} \sum_{j=1}^{\infty} q_j \frac{(k+j-1)!}{(j-1)!}.$$

If losses occur according to the mixed Erlang law $F(y)$, and there is a deductible of x, then payments occur according to the excess loss (on residual lifetime) df $F_x(y) = 1 - \bar{F}_x(y) = 1 - \bar{F}(x+y)/\bar{F}(x)$, whose density $f_x(y) = F_x'(y)$ may be expressed as [16, pp. 20–21]

$$f_x(y) = \sum_{j=1}^{\infty} q_{j,x} \frac{\beta(\beta y)^{j-1} e^{-\beta y}}{(j-1)!}, \quad y > 0, \tag{8}$$

where

$$q_{j,x} = \frac{\sum_{i=j}^{\infty} q_i \dfrac{(\beta x)^{i-j}}{(i-j)!}}{\sum_{m=0}^{\infty} \bar{Q}_m \dfrac{(\beta x)^m}{m!}}, \tag{9}$$

and thus the excess loss distribution is a different mixture of the same Erlang distributions. The excess loss distribution is of importance in connection with the analysis of the deficit at ruin, as discussed in the next section.

The failure rate (or hazard rate, force of mortality) is of interest in analysis and classification of right tail behaviour, and from (2) and (6) may be expressed as

$$\mu(y) = \frac{f(y)}{\bar{F}(y)} = \beta \frac{\sum_{j=0}^{\infty} q_{j+1} \dfrac{(\beta y)^j}{j!}}{\sum_{j=0}^{\infty} \bar{Q}_j \dfrac{(\beta y)^j}{j!}}. \tag{10}$$

As $q_{j+1} = \bar{Q}_j - \bar{Q}_{j+1}$, it is clear from (10) that $\mu(y) \leqslant \beta$, implying further that $\bar{F}(y) \geqslant \mathrm{e}^{-\beta y}$. Also, $\mu(0) = \beta q_1$. If $Q(z)$ has radius of convergence $z_0 \geqslant 1$, then from (3), $\tilde{f}(s) = Q(\beta/(\beta+s))$ has left abscissa of convergence $-s_0$ where $z_0 = \beta/(\beta - s_0)$, or equivalently $s_0 = \beta(1 - z_0^{-1})$. Thus, by Widder [17], $\mu(\infty) = \lim_{y\to\infty} \mu(y) = s_0 = \beta(1 - z_0^{-1})$. For finite mixtures, $z_0 = \infty$ and $\mu(\infty) = \beta$.

Another distribution useful for examination of right tail behaviour (as well as for ruin theoretic analysis) is the equilibrium or integrated tail distribution, with df $F_e(y) = 1 - \bar{F}_e(y)$ and pdf $f_e(y) = F_e'(y)$ given in the mixed Erlang case by (e.g. [16, pp. 15–16])

$$f_e(y) = \frac{\bar{F}(y)}{\int_0^{\infty} \bar{F}(x)\,\mathrm{d}x} = \sum_{j=1}^{\infty} q_j^* \frac{\beta(\beta y)^{j-1} \mathrm{e}^{-\beta y}}{(j-1)!}, \quad y > 0, \tag{11}$$

where

$$q_j^* = \frac{\bar{Q}_{j-1}}{\sum_{k=1}^{\infty} k q_k}, \quad j = 1, 2, \ldots, \tag{12}$$

again a different mixture of the same Erlang distributions.

The mean excess loss (mean residual lifetime) $r(x) = \int_0^{\infty} y f_x(y)\,\mathrm{d}y$ is useful not only as the mean in the presence of a deductible but also for right tail analysis. Its reciprocal is well known to be the failure rate of the equilibrium distribution (e.g. [16, p. 19]), i.e. $1/r(y) = f_e(y)/\bar{F}_e(y)$. Thus, by analogy with (10),

$$r(y) = \frac{\sum_{j=0}^{\infty} \bar{Q}_j^* \dfrac{(\beta y)^j}{j!}}{\beta \sum_{j=0}^{\infty} q_{j+1}^* \dfrac{(\beta y)^j}{j!}}, \tag{13}$$

where $\bar{Q}_j^* = \sum_{k=j+1}^{\infty} q_k^*$ for $j = 0, 1, 2, \ldots$. One has $r(0) = \int_0^{\infty} y f(y)\,\mathrm{d}y = (\sum_{j=1}^{\infty} j q_j)/\beta$, and $r(\infty) = 1/\mu(\infty) = z_0/\{\beta(z_0 - 1)\}$, and $r(\infty) = 1/\beta$ if $z_0 = \infty$. Also, $\mu(y) \leqslant \beta$ implies that $r(y) \geqslant 1/\beta$ (e.g. [16, pp. 21–23]).

Another commonly used risk measure is tail value at risk (TVaR). The TVaR of a loss model at confidence level p is the conditional expected loss given that the loss exceeds VaR at the same confidence level. Based on the earlier definition of VaR as v_p, the TVaR can be computed as $r(v_p) + v_p$ explicitly using (13).

Monotonicity of $\mu(y)$ and $r(y)$ and other risk classification properties based on the excess loss df $F_x(y)$ and the equilibrium df $F_e(y)$ are closely connected with (and essentially reproduce) those of the discrete mixing distribution $\{q_1, q_2, \ldots\}$. See Esary *et al.* [18] for details.

Turning next to aggregate claims analysis under the assumption that the individual claims distribution is of mixed Erlang form, let the pgf of the number of claims distribution be $P(z) = \sum_{n=0}^{\infty} p_n z^n$. The total claims df $G(x) = 1 - \bar{G}(x)$ then has Laplace–Stieltjes transform $\tilde{g}(s) = P\{\tilde{f}(s)\}$ with $\tilde{f}(s)$ the mixed Erlang Laplace transform (3). Thus

$$\tilde{g}(s) = C\left(\frac{\beta}{\beta+s}\right), \tag{14}$$

with

$$C(z)=\sum_{n=0}^{\infty} c_n z^n = P\{Q(z)\} \tag{15}$$

the pgf of a discrete compound distribution. Then $G(0)=p_0$, and $G(x)$ has pdf $g(x)=G'(x)$ for $x>0$ given by

$$g(x)=\sum_{n=1}^{\infty} c_n \frac{\beta(\beta x)^{n-1}\mathrm{e}^{-\beta x}}{(n-1)!}, \quad x>0, \tag{16}$$

again of mixed Erlang from (but $\sum_{n=1}^{\infty} c_n = 1-c_0$ in this case). Thus

$$\bar{G}(x)=\int_x^{\infty} g(y)\,\mathrm{d}y=\mathrm{e}^{-\beta x}\sum_{j=0}^{\infty}\bar{C}_j\frac{(\beta x)^j}{j!}, \quad x\geqslant 0, \tag{17}$$

with $\bar{C}_j=\sum_{n=j+1}^{\infty} c_n$ for $j=0,1,2,\ldots$. Evaluation of (16) and (17) typically requires evaluation of the discrete distribution $\{c_0, c_1, \ldots\}$ (apart from simple special cases where simple Laplace transform inversion is possible). We remark that recursive techniques may sometimes be used because $\{c_0, c_1, \ldots\}$ is a discrete compound distribution (for example, Panjer-type recursions apply when $\{p_0, p_1, \ldots\}$ is of Poisson or negative binomial form; see [19], for instance). There is also some asymptotic help available. If $p_n \sim K n^{\alpha}\phi^n$, $n\to\infty$, where $K>0$, $-\infty<\alpha<\infty$, and $0<\phi<1$, and there exists $\tau>1$ satisfying $Q(\tau)=\phi^{-1}$, then (e.g. [20]) $c_n \sim K_1 n^{\alpha}\tau^{-n}$, where $K_1 = K\{\phi\tau Q'(\tau)\}^{-\alpha-1}$. In turn, the result of Embrechts *et al.* [15] may then be applied, yielding the estimate

$$\bar{G}(x)\sim C_0^* x^{\alpha}\mathrm{e}^{-\beta(1-1/\tau)x}, \quad x\to\infty, \tag{18}$$

with $C_0^*=K_1\beta^{\alpha}\tau^{-\alpha}/(\tau-1)$. We remark that a similar formula to (18) holds if $q_n\sim K_2 n^{\alpha}\phi^n$, where $K_2>0$, $\alpha<-1$, $0<\phi<1$, and $P'(Q(1/\phi))<\infty$. See Willmot [20] for details.

It is not difficult to show that (16) satisfies

$$g(x+y)=\beta^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} c_{j+k+1} e_{j+1}(x) e_{k+1}(y) \tag{19}$$

for $x\geqslant 0$ and $y\geqslant 0$ (e.g. [21]). Using (19), stop-loss moments of any positive order may be obtained from the damped exponential series [3, p. 106]

$$\int_x^{\infty}(y-x)^{\alpha}\,\mathrm{d}G(y)=\mathrm{e}^{-\beta x}\sum_{j=0}^{\infty} a_{j,\alpha}\frac{(\beta x)^j}{j!} \tag{20}$$

where

$$a_{j,\alpha}=\beta^{-\alpha}\sum_{k=1}^{\infty} c_{j+k}\frac{\Gamma(\alpha+k)}{(k-1)!}. \tag{21}$$

Evidently, (20) reduces to (17) when $\alpha=0$. When $\alpha=1$, the stop-loss premium results, and in this case one can easily obtain $a_{j,1}=\beta^{-1}\sum_{i=j}^{\infty}\bar{C}_i$. That is, using integration by parts on the right-hand side of (20), the stop-loss premium may be expressed as

$$\int_x^{\infty}\bar{G}(y)\,\mathrm{d}y=\beta^{-1}\mathrm{e}^{-\beta x}\sum_{j=0}^{\infty}\left(\sum_{i=j}^{\infty}\bar{C}_i\right)\frac{(\beta x)^j}{j!}, \tag{22}$$

and using (18) and Grandell [13, p. 181] yields the asymptotic estimate

$$\int_x^{\infty}\bar{G}(y)\,\mathrm{d}y\sim C_1^* x^{\alpha}\mathrm{e}^{-\beta(1-1/\tau)x}, \quad x\to\infty, \tag{23}$$

with $C_1^*=\tau C_0^*/\{\beta(\tau-1)\}$.

Applications involving ruin and the surplus are considered next.

4. Ruin and surplus analysis

In this section we will consider the Sparre Andersen risk model (see [21], and references therein, for a complete description of the model) under the assumption that claim sizes follow the mixed Erlang distribution. Let $\delta \geqslant 0$, and denote the 'discounted density of the surplus immediately prior to ruin with zero initial reserve' by $h_\delta(x)$ (this function is known for some special cases of the Sparre Andersen model (as is discussed later), including the classical compound Poisson risk model). Define $\phi_\delta \varepsilon(0, 1)$ by

$$\phi_\delta = \int_0^\infty h_\delta(x)\,\mathrm{d}x, \tag{24}$$

and the 'ladder height' pdf by

$$b_\delta(y) = \int_0^\infty f_x(y) \left\{ \frac{h_\delta(x)}{\phi_\delta} \right\} \mathrm{d}x, \tag{25}$$

where $f_x(y)$ is the excess loss pdf given in the mixed Erlang case by (8). Let $B_\delta(y) = 1 - \bar{B}_\delta(y) = \int_0^y b_\delta(x)\,\mathrm{d}x$ be the ladder height df. Of interest is the 'time of ruin' T if ruin occurs (i.e. the surplus becomes negative) and we write '$T=\infty$' if ruin does not occur. Let $\bar{G}_\delta(x) = 1 - G_\delta(x) = E\{\mathrm{e}^{-\delta T} I(T<\infty)\}$ where $x \geqslant 0$ is the initial surplus, and $I(A) = 1$ if A occurs and $I(A) = 0$ otherwise. The infinite time ruin probability $\psi(x) = \Pr(T<\infty)$ is the special case $\delta = 0$, so that $\psi(x) = \bar{G}_0(x)$. Then (e.g. [21]), $G_\delta(x)$ is a compound geometric df with geometric parameter ϕ_δ given by (24) and 'secondary' distribution $B_\delta(x)$; that is

$$\bar{G}_\delta(x) = \sum_{n=1}^\infty (1-\phi_\delta)\phi_\delta^n \bar{B}_\delta^{*n}(x), \quad x \geqslant 0, \tag{26}$$

with $1 - \bar{B}_\delta^{*n}(x)$ being the df of the n-fold convolution of the df $B_\delta(x)$ with itself. The Laplace–Stieltjes transform is

$$\int_0^\infty \mathrm{e}^{-sx}\,\mathrm{d}G_\delta(x) = \frac{1-\phi_\delta}{1-\phi_\delta \tilde{b}_\delta(s)}, \tag{27}$$

where

$$\tilde{b}_\delta(s) = \int_0^\infty \mathrm{e}^{-sx} b_\delta(x)\,\mathrm{d}x. \tag{28}$$

In the mixed Erlang case, substitution of (8) into (25) results in

$$\begin{aligned} b_\delta(y) &= \int_0^\infty \left\{ \sum_{j=1}^\infty q_{j,x} \frac{\beta(\beta y)^{j-1}\mathrm{e}^{-\beta y}}{(j-1)!} \right\} \left\{ \frac{h_\delta(x)}{\phi_\delta} \right\} \mathrm{d}x \\ &= \sum_{j=1}^\infty \left\{ \int_0^\infty q_{j,x} \frac{h_\delta(x)}{\phi_\delta}\,\mathrm{d}x \right\} \frac{\beta(\beta y)^{j-1}\mathrm{e}^{-\beta y}}{(j-1)!}. \end{aligned}$$

That is,

$$b_\delta(y) = \sum_{j=1}^\infty q_j(\delta) \frac{\beta(\beta y)^{j-1}\mathrm{e}^{\beta y}}{(j-1)!}, \tag{29}$$

where

$$q_j(\delta) = \int_0^\infty q_{j,x} \frac{h_\delta(x)}{\phi_\delta}\,\mathrm{d}x. \tag{30}$$

Obviously, $\{q_1(\delta), q_2(\delta), \ldots\}$ is a discrete probability distribution, and (29) implies that the ladder height distribution is again a (different) mixture of Erlangs. Then (28) may be expressed as

$$\tilde{b}_\delta(s) = Q_\delta\left(\frac{\beta}{\beta+s}\right) \tag{31}$$

where, again,

$$Q_\delta(z)=\sum_{j=1}^{\infty} q_j(\delta) z^j \tag{32}$$

is a pgf. In turn, substitution of (31) into (27) results in

$$\int_0^{\infty} \mathrm{e}^{-sx}\, \mathrm{d}G_\delta(x)=C_\delta\left(\frac{\beta}{\beta+s}\right) \tag{33}$$

where

$$C_\delta(z)=\sum_{n=0}^{\infty} c_n(\delta) z^n=\frac{1-\phi_\delta}{1-\phi_\delta Q_\delta(z)} \tag{34}$$

is the pgf of a discrete compound geometric distribution. It follows immediately from (17) and (33) that (26) becomes

$$\bar{G}_\delta(x)=\mathrm{e}^{-\beta x} \sum_{j=0}^{\infty} \bar{C}_j(\delta) \frac{(\beta x)^j}{j!}, \quad x \geqslant 0, \tag{35}$$

where

$$\bar{C}_j(\delta)=\sum_{n=j+1}^{\infty} c_n(\delta), \quad j=0,1,2, \ldots. \tag{36}$$

Note that using Feller [22, p. 265], it follows from (34) that the exponential damping weights (36) in (35) satisfy

$$\sum_{j=0}^{\infty} \bar{C}_j(\delta) z^j=\frac{1-C_\delta(z)}{1-z}=\frac{\phi_\delta}{1-\phi_\delta Q_\delta(z)} \frac{1-Q_\delta(z)}{1-z}. \tag{37}$$

Therefore, again using (34) as well as (32),

$$\bar{C}_j(\delta)=\frac{\phi_\delta}{1-\phi_\delta} \sum_{n=0}^{j} c_n(\delta) \bar{Q}_{j-n}(\delta) \tag{38}$$

where $\bar{Q}_i(\delta)=\sum_{j=i+1}^{\infty} q_j(\delta)$.

In addition to the explicit formula (38), a recursive formula may be derived. A rearrangement of (37) yields

$$\sum_{j=0}^{\infty} \bar{C}_j(\delta) z^j=\phi_\delta Q_\delta(z)\left\{\sum_{j=0}^{\infty} \bar{C}_j(\delta) z^j\right\}+\phi_\delta \frac{1-Q_\delta(z)}{1-z},$$

from which it follows that for $j=1,2, \ldots$,

$$\bar{C}_j(\delta)=\phi_\delta \sum_{k=1}^{j} q_k(\delta) \bar{C}_{j-k}(\delta)+\phi_\delta \bar{Q}_j(\delta). \tag{39}$$

Beginning with $\bar{C}_0(\delta)=1-c_0(\delta)=\phi_\delta$, (39) may be used to compute the $\bar{C}_j(\delta)$'s recursively. As mentioned earlier, the ruin probability may be computed in this manner as it is the special case of $\bar{G}_\delta(x)$ with $\delta=0$.

Of particular interest in connection with risk management is the distribution of the deficit given that ruin occurs. Its pdf may be expressed for an initial surplus of x as (e.g. [23])

$$k_x(y)=\frac{\{1-\psi(0)\} b_{x,0}(y) \bar{B}_0(x)+\int_0^x b_{x-t,0}(y) \bar{B}_0(x-t) g_0(t)\, \mathrm{d}t}{\{1-\psi(0)\} \bar{B}_0(x)+\int_0^x \bar{B}_0(x-t) g_0(t)\, \mathrm{d}t}, \tag{40}$$

where $\bar{B}_0$ is the ladder height tail (when $\delta=0$),

$$b_{x,0}(y)=\frac{b_0(x+y)}{\bar{B}_0(x)}, \quad x>0, \tag{41}$$

is the excess loss pdf associated with $B_0(y)$ and a deductible of x, and $g_0(t)$ is the compound geometric density

$$g_0(t)=G_0^{\prime}(t)=\sum_{i=1}^{\infty} c_i(0) \frac{\beta(\beta t)^{i-1} \mathrm{e}^{-\beta t}}{(i-1)!}. \tag{42}$$

As $b_0(y)$ is a mixture of Erlangs from (29), we may write

$$b_{x,0}(y)=\sum_{m=1}^{\infty} q_{m,x}(0)\frac{\beta(\beta y)^{m-1}\mathrm{e}^{-\beta y}}{(m-1)!} \tag{43}$$

where, using (6) and (9)

$$q_{m,x}(0)=\frac{\sum_{j=m}^{\infty} q_j(0)\dfrac{(\beta x)^{j-m}}{(j-m)!}}{\mathrm{e}^{\beta x}\bar{B}_0(x)}. \tag{44}$$

Substitution of (43) into (40) results in

$$k_x(y)=\sum_{m=1}^{\infty} \theta_{m,x}\frac{\beta(\beta y)^{m-1}\mathrm{e}^{-\beta y}}{(m-1)!}, \tag{45}$$

where

$$\theta_{m,x}=\frac{\{1-\psi(0)\}q_{m,x}(0)\bar{B}_0(x)+\int_0^x q_{m,x-t}(0)\bar{B}_0(x-t)g_0(t)\,\mathrm{d}t}{\{1-\psi(0)\}\bar{B}_0(x)+\int_0^x \bar{B}_0(x-t)g_0(t)\,\mathrm{d}t}. \tag{46}$$

Clearly, (46) implies that $\sum_{m=1}^{\infty}\theta_{m,x}=1$, and thus (45) implies that the deficit pdf $k_x(y)$ is a mixture of Erlangs. We will now evaluate the integrals in (46) to get $\theta_{m,x}$. As the denominator of (46) is simply a normalizing constant, we focus on the numerator. It follows from (44) that

$$q_{m,x-t}(0)\bar{B}_0(x-t)=\mathrm{e}^{-\beta(x-t)}\sum_{j=m}^{\infty} q_j(0)\frac{\beta^{j-m}(x-t)^{j-m}}{(j-m)!}. \tag{47}$$

Thus, from (42) and (47),

$$\begin{aligned}\int_0^x q_{m,x-t}(0)\bar{B}_0(x-t)g_0(t)\,\mathrm{d}t &= \int_0^x \mathrm{e}^{-\beta(x-t)}\left\{\sum_{j=m}^{\infty} q_j(0)\frac{\beta^{j-m}(x-t)^{j-m}}{(j-m)!}\right\}\left\{\sum_{i=1}^{\infty} c_i(0)\frac{\beta(\beta t)^{i-1}\mathrm{e}^{-\beta t}}{(i-1)!}\right\}\mathrm{d}t\\ &= \mathrm{e}^{-\beta x}\sum_{i=1}^{\infty}\sum_{j=m}^{\infty} c_i(0)q_j(0)\frac{\beta^{i+j-m}}{(i-1)!(j-m)!}\int_0^x t^{i-1}(x-t)^{j-m}\,\mathrm{d}t.\end{aligned}$$

But

$$\int_0^x t^{i-1}(x-t)^{j-m}\,\mathrm{d}t=x^{i+j-m}\frac{(i-1)!(j-m)!}{(i+j-m)!},$$

implying that

$$\int_0^x q_{m,x-t}(0)\bar{B}_0(x-t)g_0(t)\,\mathrm{d}t=\mathrm{e}^{-\beta x}\sum_{i=1}^{\infty}\sum_{j=m}^{\infty} c_i(0)q_j(0)\frac{(\beta x)^{i+j-m}}{(i+j-m)!}.$$

Noting that $1-\psi(0)=1-\bar{G}_0(0)=1-\bar{C}_0(0)=c_0(0)$ from (35) and (36), it follows from (47) with $t=0$ that the numerator of (46) may be expressed as

$$\begin{aligned}\{1-\psi(0)\}q_{m,x}(0)\bar{B}_0(x)+\int_0^x q_{m,x-t}(0)\bar{B}_0(x-t)g_0(t)\,\mathrm{d}t &= \mathrm{e}^{-\beta x}c_0(0)\sum_{j=m}^{\infty} q_j(0)\frac{(\beta x)^{j-m}}{(j-m)!}\\ &\quad +\mathrm{e}^{-\beta x}\sum_{i=1}^{\infty}\sum_{j=m}^{\infty} c_i(0)q_j(0)\frac{(\beta x)^{i+j-m}}{(i+j-m)!}\\ &= \mathrm{e}^{-\beta x}\sum_{i=0}^{\infty}\sum_{j=m}^{\infty} c_i(0)\frac{(\beta x)^{i+j-m}}{(i+j-m)!}.\end{aligned} \tag{48}$$

For notational convenience, let the reciprocal of the denominator of (46) be $M(x)$, and define the sequence of functions

$$\tau_n(x)=\sum_{i=0}^{\infty} c_i(0)\frac{x^{i+n}}{(i+n)!}, \quad n=0,1,2,\ldots \tag{49}$$

and (46), (48), and (49) result in

$$\theta_{m,x}=M(x)\mathrm{e}^{-\beta x}\sum_{j=m}^{\infty} q_j(0)\tau_{j-m}(\beta x).$$

But

$$\begin{aligned}\sum_{m=1}^{\infty}\theta_{m,x} &= M(x)\mathrm{e}^{-\beta x}\sum_{m=1}^{\infty}\sum_{j=m}^{\infty} q_j(0)\tau_{j-m}(\beta x)\\ &= M(x)\mathrm{e}^{-\beta x}\sum_{j=1}^{\infty} q_j(0)\sum_{m=1}^{j}\tau_{j-m}(\beta x)\\ &= M(x)\mathrm{e}^{-\beta x}\sum_{j=1}^{\infty} q_j(0)\sum_{i=0}^{j-1}\tau_i(\beta x).\end{aligned}$$

Since this sum is 1, it follows that

$$\theta_{m,x}=\frac{\sum_{j=m}^{\infty} q_j(0)\tau_{j-m}(\beta x)}{\sum_{j=1}^{\infty} q_j(0)\sum_{i=0}^{j-1}\tau_i(\beta x)}. \tag{50}$$

To summarize, in the Sparre Andersen model with mixed Erlang claim amounts, the deficit distribution is also a mixed Erlang with pdf (45), and the mixing weights are given by (50) with $q_j(0)$ given by (30) and $\delta=0$ and $\tau_n(x)$ by (49).

As mentioned earlier, the discounted surplus density $h_\delta(x)$ is not known in general for the Sparre Andersen model, but is known for some special cases. It is needed, however, for evaluation of $q_j(\delta)$ in (30) and other quantities discussed here. With Coxian interclaim times, $h_\delta(x)=\bar{F}(x)\sum_{j=1}^{n} a_j^*(\delta)\mathrm{e}^{-\rho_j(\delta)x}$ where the a_j^*'s are constants and $\{\rho_1(\delta),\rho_2(\delta),\ldots,\rho_n(\delta)\}$ are n distinct roots of Lundberg's fundamental equation (e.g. [24]). Then (24) becomes, using (3),

$$\phi_\delta=\sum_{j=1}^{n} a_j^*(\delta)\int_0^{\infty}\mathrm{e}^{-\rho_j(\delta)x}\bar{F}(x)\,\mathrm{d}x=\sum_{j=1}^{n}\frac{a_j^*(\delta)}{\rho_j(\delta)}\left\{1-Q\left(\frac{\beta}{\beta+\rho_j(\delta)}\right)\right\}. \tag{51}$$

Using (6) and (9), (30) becomes

$$\begin{aligned}q_j(\delta) &= \int_0^{\infty}\frac{\mathrm{e}^{-\beta x}}{\bar{F}(x)}\left\{\sum_{i=j}^{\infty} q_i\frac{(\beta x)^{i-j}}{(i-j)!}\right\}\bar{F}(x)\left\{\sum_{k=1}^{n} a_k^*(\delta)\mathrm{e}^{-\rho_k(\delta)x}\right\}\mathrm{d}x\\ &= \frac{1}{\phi_\delta}\sum_{k=1}^{n} a_k^*(\delta)\sum_{i=j}^{\infty}\frac{q_i}{(i-j)!}\int_0^{\infty}(\beta x)^{i-j}\mathrm{e}^{-\{\beta+\rho_k(\delta)\}x}\,\mathrm{d}x\\ &= \frac{1}{\beta\phi_\delta}\sum_{k=1}^{n} a_k^*(\delta)\sum_{i=j}^{\infty} q_i\left\{\frac{\beta}{\beta+\rho_k(\delta)}\right\}^{i-j+1}.\end{aligned} \tag{52}$$

When $n=1$, the classical Poisson model results. In this case (51) and (52) yield

$$\begin{aligned}q_j(\delta) &= \frac{\dfrac{a_1^*(\delta)}{\beta}\sum_{i=j}^{\infty} q_i\left\{\dfrac{\beta}{\beta+\rho_1(\delta)}\right\}^{i-j+1}}{\dfrac{a_1^*(\delta)}{\rho_1(\delta)}\left\{1-Q\left(\dfrac{\beta}{\beta+\rho_1(\delta)}\right)\right\}}\\ &= \sum_{i=j}^{\infty} q_i\left\{\frac{\beta}{\beta+\rho_1(\delta)}\right\}^{i-j}\Bigg/\left\{\frac{1-Q\left(\dfrac{\rho}{\beta+\rho_1(\delta)}\right)}{1-\dfrac{\beta}{\beta+\rho_1(\delta)}}\right\}.\end{aligned}$$

That is,

$$q_j(\delta) = \frac{\sum_{i=j}^{\infty} q_i \left\{\frac{\beta}{\beta+\rho_1(\delta)}\right\}^{i-j}}{\sum_{i=0}^{\infty} \bar{Q}_i \left\{\frac{\beta}{\beta+\rho_1(\delta)}\right\}^{i}}. \tag{53}$$

As

$$\sum_{i=0}^{\infty} \bar{Q}_i \left\{\frac{\beta}{\beta+\rho_1(\delta)}\right\}^{i} = \sum_{j=1}^{\infty} q_j \sum_{i=0}^{j-1} \left\{\frac{\beta}{\beta+\rho_1(\delta)}\right\}^{i},$$

this agrees with Willmot and Lin [16, p. 163] and reduces to (12) when $\delta = 0$ (as $\rho_1(0) = 0$ in this situation), as it must because $b_0(y) = f_e(y)$ in this case.

We remark that the discrete compound geometric probabilities $\{c_0(\delta), c_1(\delta), \ldots\}$, which are needed when $\delta = 0$ for evaluation of $\tau_n(x)$ in (49), may be computed recursively using a Panjer-type recursion (e.g. [19]).

In the classical compound Poisson risk model with mixed Erlang claim amounts, a simple but algebraically tedious formula involving Erlang-type terms exists for the finite time ruin probabilities. In a queuing context, this is of interest in connection with the evaluation of the transient waiting time distribution in the $M/G/1$ queue (e.g. [25]). The approach to the derivation of this formula is the analytical evaluation of the density of the time of ruin by inversion of its Laplace transform $\bar{G}_\delta(x) = E\{\mathrm{e}^{-\delta T} I(T<\infty)\}$. Two complicating factors exist, however. The first involves the inversion on a term by term basis using the infinite series compound geometric tail representation for $\bar{G}_\delta(x)$. Second, as is clear from (29) and (53), δ only appears through the Lundberg root $\rho_1(\delta)$. As a practical matter, transform inversion with respect to ρ_1 may be performed, and conversion to δ typically requires the use of Lagrange's implicit function theorem. Finally, as the resulting terms are of Erlang-type, integration of the density to get the finite time ruin probabilities is straightforward. See Dickson and Willmot [9] for details.

In the next section we consider fitting of mixed Erlang distributions to data.

5. Statistical estimation

As mentioned in Section 1, phase-type distributions are commonly used for insurance claims in risk modelling. However, it is not a simple matter to fit a general phase-type distribution to insurance claim data due to both the matrix representation and its non-uniqueness. For example, the EM algorithm proposed in Asmussen *et al.* [26] is fairly complicated and its convergence can be very slow.

Recently, Ausin and Lopes [27] proposed a procedure to fit the finite mixture of Erlangs of our Example 2 to the individual claim distribution in the compound Poisson risk model, and calculated the corresponding ruin probabilities. The procedure involves the construction of a reversible jump Markov chain Monte Carlo (RJMCMC), and is similar to the earlier work done by Wiper *et al.* [28] for a finite mixture of gamma distributions. In this section, we present a simple and effective EM algorithm for a finite mixture of Erlang distributions of (2).

The EM algorithm first proposed in Dempster *et al.* [29] is an iterative algorithm with two steps: an expectation step (E-Step) and a maximization step (M-Step). It is widely used to fit a parametric model to incomplete data. Since we may treat the mixing component in a finite mixture model to be unobservable, the EM algorithm is often very effective for fitting a finite mixture model to data by estimating the model parameters including its mixing weights. Suppose that we are given a data set $y_1, y_2, \ldots, y_n$ and we wish to use the following finite mixture to fit the data:

$$f(y) = \sum_{j=1}^{J} q_j \frac{\beta(\beta y)^{j-1}\mathrm{e}^{-\beta y}}{(j-1)!} = \sum_{j=1}^{J} q_j e_j(y), \quad y>0, \tag{54}$$

where J is a preset large integer. The mixing weights q_j, $j = 1, 2, \ldots, J$ and the parameter β are to be estimated. The following iterative scheme can be obtained by the EM algorithm (see [11]):

$$q_j^{(k)} = \frac{1}{n}\sum_{i=1}^{n} p(j|y_i, \beta^{(k-1)}, q_1^{(k-1)}, \ldots, q_J^{(k-1)}), \quad j = 1, 2, \ldots, J, \tag{55}$$

where

$$p(j|y_i, \beta, q_1, \ldots, q_J) = \frac{q_j e_j(y_i)}{\sum_{k=1}^{J} q_k e_k(y_i)}, \tag{56}$$

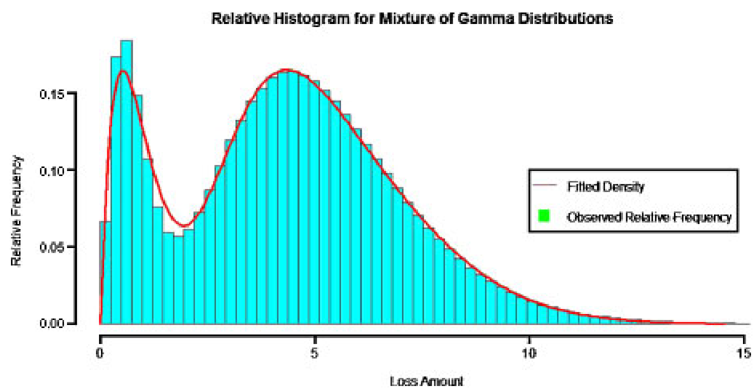

Figure 1. Histogram for a mixture of two gamma distributions and the fitted density using a mixture of 3 Erlang distributions.

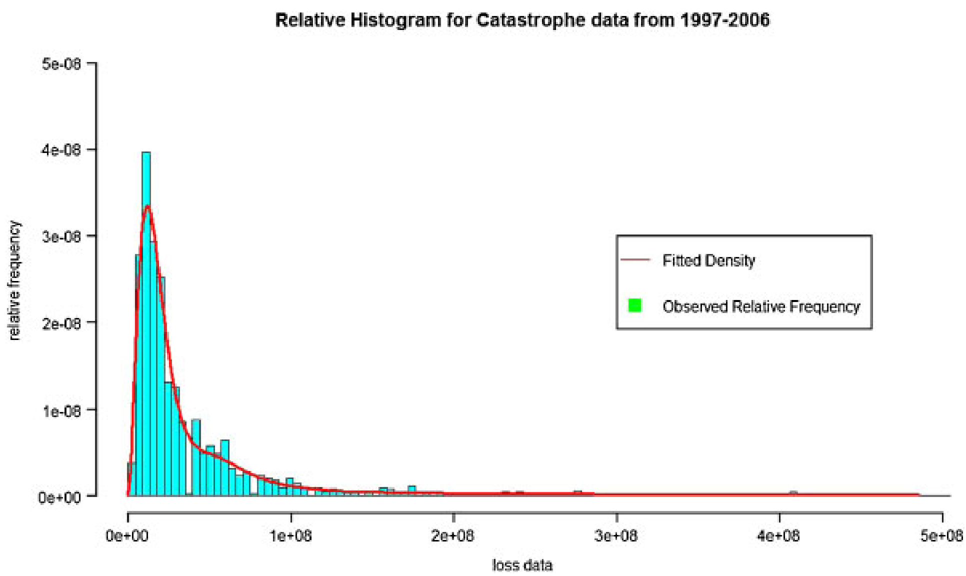

Figure 2. Histogram of observed loss and line for the fitted distribution.

and

$$\beta^{(k)} = \frac{\sum_{j=1}^{J} j q_j^{(k)}}{\sum_{i=1}^{n} y_i/n}. \tag{57}$$

In general, the M-step of the EM algorithm involves the use of an optimization method, which is often a difficult task. For example, when one implements the EM algorithm to fit a general phase-type model to data as in Asmussen *et al.* [26], the associated objective function in the maximization step could have several local maximal points that in turn may cause the algorithm failing to find the global maximum. However, when we limit candidate distributions to the class of Erlang mixtures with density (24), there is no need to employ an optimization method as shown in (55) and (57). In other words, the EM algorithm in this situation is a pure iterative algorithm, and is therefore simple and very effective.

To begin the iterative scheme, the integer J and the initial values of the parameters, $\beta^{(0)}$ and $q_j^{(0)}$, $j=1,2,\ldots,J$, need to be specified. As mentioned in Section 1, Tijms ([1] , pp. 163–164) showed that for any positive continuous distribution with distribution function $H(y)$, $y>0$, the distribution function with density

$$f(y|\beta) = \sum_{j=1}^{\infty} [H(j/\beta) - H((j-1)/\beta)] e_j(y) \tag{58}$$

converges to $H(y)$ pointwise, as β tends to infinity. As a result, we may use a simple '80-8' rule to choose integer J and the initial scale parameter $\beta^{(0)}$, and let $q_j^{(0)}$ be the relative frequency of the data points on the interval $((j-1)/\beta^{(0)}, j/\beta^{(0)})$. See Section 3.2 of Lee and Lin [11] for more detail. The numerical experiments in Lee and Lin [11] show that the iterative scheme performs very well and always results in fast convergence when using the above choice of initial estimates. However, there is an obvious shortcoming with this approach. The larger the value of β we assign, the more the Erlangs we will use. Too many Erlangs in the mixture will result in an issue of overfitting. Thus, it is necessary to use a decision rule such as Schwartz's Bayesian Information Criterion on how many Erlangs one needs to choose. Further, since the shape parameters are pre-fixed and hence not estimated, fitting results might be sub-optimal. Adjustments to the shape parameters may be necessary. These issues are discussed in detail in Lee and Lin [11].

Finally, we present two fitting examples to illustrate how well an Erlang mixture can fit multi-modal data using the EM algorithm. The details on these examples and other fitting results including statistical analysis are given in Lee and Lin [11].

The first example involves fitting an Erlang mixture to data generated from a mixture of two gamma distributions both of which have non-integer shape parameters. Assume that the gamma distributions are gamma(2.6, 3.2) and gamma(6.3, 1.2), where the first parameter is the shape parameter and the second parameter is β. The weights are 0.2 and 0.8, respectively. Note that this distribution is not a phase-type distribution and it has two modes. Applying the EM algorithm and the parameter reduction techniques in Lee and Lin [11], we are able to use three Erlang distributions to fit the data and the fitting result is given in Figure 1.

The second example involves fitting an Erlang mixture to PCS catastrophe data which contains 1271 catastrophic losses in US from 1997 to 2005 from the Insurance Services Office (ISO). The data is multi-modal and heavy right-tailed. A mixture of 12 Erlang distributions fits the data. The fitting result is given in Figure 2.

As the graphs show, both fitting results seem satisfactory.

Acknowledgements

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada. Financial support from the Munich Reinsurance Company of Canada for Gordon E. Willmot is also gratefully acknowledged. The authors thank Dr. Fabrizio Ruggeri for his valuable comments.

References

1. Tijms HC. *Stochastic Models*: *An Algorithmic Approach*. Wiley: Chichester, 1994.
2. Gerber HU, Shiu ESW. The time value of ruin in a Sparre Andersen model. *North American Actuarial Journal* 2005; **9**(2):49–84.
3. Willmot GE, Woo JK. On the class of Erlang mixtures with risk theoretic applications. *North American Actuarial Journal* 2007; **11**(2):99–115.
4. Shanthikumar JG. Bilateral phase-type distributions. *Naval Research Logistics Quarterly* 1985; **32**:119–136.
5. Neuts MF. *Matrix-Geometric Solutions in Stochastic Models*: *An Algorithmic Approach*. Johns Hopkins University Press: Baltimore, 1981.
6. Stanford DA, Zadeh AH. Linear and Bayesian credibility theory for univariate phase-type distributions. Preprint 2010.
7. Gerber HU, Goovaerts M, Kass R. On the probability and severity of ruin. *ASTIN Bulletin* 1987; **17**:151–163.
8. Dufresne D. Fitting combinations of exponentials to probability distributions. *Applied Stochastic Models in Business and Industry* 2007; **23**:23–48.
9. Dickson DCM, Willmot GE. The density of the time to ruin in the classical Poisson risk model. *ASTIN Bulletin* 2005; **35**:45–60.
10. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions* (2nd edn), vol. 2. Wiley: New York, 1995.
11. Lee SCK, Lin XS. Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* 2010; in press.
12. Taylor HM, Karlin S. *An Introduction to Stochastic Modeling* (3rd edn). Academic Press: San Diego, 1998.
13. Grandell J. *Mixed Poisson Processes*. Chapman & Hall: London, 1997.
14. Willmot GE. The total claims distribution under inflationary conditions. *Scandinavian Actuarial Journal* 1989; 1–12.
15. Embrechts P, Maejima M, Teugels J. Asymptotic behaviour of compound distributions. *ASTIN Bulletin* 1985; **15**:45–48.
16. Willmot GE, Lin XS. *Lundberg Approximations for Compound Distributions with Insurance Applications*. Springer: New York, 2001.
17. Widder DV. *The Laplace Transform*. Princeton University Press: Princeton, 1946.
18. Esary JD, Marshall AW, Proschan F. Shock models and wear processes. *Annals of Probability* 1973; **1**:627–649.
19. Klugman SA, Panjer HH, Willmot GE. *Loss Models*: *From Data to Decisions* (3rd edn). Wiley: New York, 2008.
20. Willmot GE. Limiting tail behaviour of some discrete compound distributions. *Insurance*, *Mathematics and Economics* 1989; **8**:185–185.
21. Willmot GE. On the discounted penalty function in the renewal risk model with general interclaim times. *Insurance*: *Mathematics and Economics* 2007; **41**:17–31.
22. Feller W. *An Introduction to Probability Theory and Its Applications* (3rd edn), vol. 1. Wiley: New York, 1968.
23. Willmot GE. Compound geometric residual lifetime distributions and the deficit at ruin. *Insurance: Mathematics and Economics* 2002; **30**: 421–438.
24. Li S, Garrido J. On a general class of renewal risk process: analysis of the Gerber–Shiu function. *Advances in Applied Probability* 2005; **37**(3):836–856.
25. Asmussen S. *Ruin Probabilities*. World Scientific: Singapore, 2000.
26. Asmussen S, Nerman O, Olsson M. Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics* 1996; **23**(4): 419–441.

27. Ausin MC, Lopes HF. Bayesian estimation of ruin probabilities with heterogeneous and heavy-tailed insurance claim size distribution. *Australian and New Zealand Journal of Statistics* 2007; **49**(4):415–434.
28. Wiper M, Rios Insua D, Ruggeri F. Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics* 2001; **10**:440–454.
29. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* (*Methodological*) 1977; **39**:1–38.