Loss modelling with mixtures of Erlang distributions

Roel Verbelen

Faculty of Economics and Business
KU Leuven, Belgium
roel.verbelen@kuleuven.be

R in Insurance
Cass Business School, London, UK

July 14, 2014
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1}e^{-x/\theta}}{\theta^{r_j}(r_j-1)!} \]

for \( x > 0 \) with
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1} e^{-x/\theta}}{\theta^{r_j} (r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M \);
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1} e^{-x/\theta}}{\theta^{r_j}(r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M \);
- weights \( \alpha = (\alpha_1, \ldots, \alpha_M) \) with \( \alpha_j > 0 \)
  and \( \sum_{j=1}^{M} \alpha_j = 1 \);
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1} e^{-x/\theta}}{\theta^{r_j} (r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M; \)
- weights \( \alpha = (\alpha_1, \ldots, \alpha_M) \) with \( \alpha_j > 0 \) and \( \sum_{j=1}^{M} \alpha_j = 1; \)
- positive integer shape parameters \( r = (r_1, \ldots, r_M); \)
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j - 1} e^{-x/\theta}}{\theta^{r_j} (r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M \);
- weights \( \alpha = (\alpha_1, \ldots, \alpha_M) \) with \( \alpha_j > 0 \) and \( \sum_{j=1}^{M} \alpha_j = 1 \);
- positive integer shape parameters \( r = (r_1, \ldots, r_M) \);
- common scale parameter \( \theta \).
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j x^{r_j-1} e^{-x/\theta} \frac{\theta^{r_j}}{r_j (r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M \);
- weights \( \alpha = (\alpha_1, \ldots, \alpha_M) \) with \( \alpha_j > 0 \) and \( \sum_{j=1}^{M} \alpha_j = 1 \);
- positive integer shape parameters \( r = (r_1, \ldots, r_M) \);
- common scale parameter \( \theta \).

Figure: Varying the shape \( r \) with scale \( \theta = 2 \).
Mixtures of Erlangs with common scale

Probability density function

\[ f_X(x; \alpha, r, \theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1} e^{-x/\theta}}{\theta^{r_j} (r_j - 1)!} \]

for \( x > 0 \) with

- number of Erlangs \( M \);
- weights \( \alpha = (\alpha_1, \ldots, \alpha_M) \) with \( \alpha_j > 0 \) and \( \sum_{j=1}^{M} \alpha_j = 1 \);
- positive integer shape parameters \( r = (r_1, \ldots, r_M) \);
- common scale parameter \( \theta \).

Figure: Varying the scale \( \theta \) with shape \( r = 2 \).
Why mixtures of Erlangs?

Suitable for loss modeling since

• versatile classes of distributions;
• mathematically tractable allowing analytical expressions of quantities of interest;
• fitting procedure based on the EM algorithm;
• able to deal with censored and/or truncated data;
• implemented in R, making use of the package doParallel.
Why mixtures of Erlangs?

Suitable for loss modeling since

- versatile classes of distributions;
Why mixtures of Erlangs?

Suitable for loss modeling since

- versatile classes of distributions;
- mathematically tractable allowing analytical expressions of quantities of interest;
Why mixtures of Erlangs?

Suitable for loss modeling since

- versatile classes of distributions;
- mathematically tractable allowing analytical expressions of quantities of interest;
- fitting procedure based on the EM algorithm;
Why mixtures of Erlangs?

Suitable for loss modeling since

- versatile classes of distributions;
- mathematically tractable allowing analytical expressions of quantities of interest;
- fitting procedure based on the EM algorithm;
- able to deal with censored and/or truncated data.
Why mixtures of Erlangs?

Suitable for loss modeling since

- versatile classes of distributions;
- mathematically tractable allowing analytical expressions of quantities of interest;
- fitting procedure based on the EM algorithm;
- able to deal with censored and/or truncated data;
- implemented in R, making use of the package doParallel.
Secura Re data I

- Price an unlimited excess-loss layer above an operational priority $R$.

- 371 automobile claims from 1988 until 2001 from several European insurers, corrected, among others, for inflation.

- Left truncated at 1,200,000 euro, since the claims are only reported to the reinsurer if they are larger.
Secura Re data II

Parameter estimates

<table>
<thead>
<tr>
<th>$r_j$</th>
<th>$\alpha_j$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.971</td>
<td>360 096.1</td>
</tr>
<tr>
<td>15</td>
<td>0.029</td>
<td></td>
</tr>
</tbody>
</table>

(a) Fitted density function and histogram.

(b) Fitted survival and Kaplan-Meier.

(c) QQ plot.
Explicit expressions for

- the net premium $\Pi(R)$ of an excess-of-loss reinsurance contract with retention level $R > 1,200,000$

$$\Pi(R) = E((X - R)_+ \mid X > 1,200,000) ;$$

- the excess-loss distribution

$$X - R \mid X > R$$

which is again a mixture of Erlangs with the same scale $\theta$ but with different weights.
Submitted for publication.

Verbelen, R. [2014]
www.econ.kuleuven.be/roel.verbelen
Additional examples.
R code and illustration.
Secura Re data IV

**Table:** Non-parametric, Hill, GP and Mixture of Erlangs-based estimates for $\Pi(R)$.

<table>
<thead>
<tr>
<th>R</th>
<th>Non-Parametric</th>
<th>Hill</th>
<th>GP</th>
<th>Mixture of Erlangs</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 000 000</td>
<td>161 728.1</td>
<td>163 367.4</td>
<td>166 619.6</td>
<td>163 987.7</td>
</tr>
<tr>
<td>3 500 000</td>
<td>108 837.2</td>
<td>108 227.2</td>
<td>111 610.4</td>
<td>110 118.5</td>
</tr>
<tr>
<td>4 000 000</td>
<td>74 696.3</td>
<td>75 581.4</td>
<td>79 219.0</td>
<td>77 747.6</td>
</tr>
<tr>
<td>4 500 000</td>
<td>53 312.3</td>
<td>55 065.8</td>
<td>58 714.1</td>
<td>55 746.3</td>
</tr>
<tr>
<td>5 000 000</td>
<td>35 888.0</td>
<td>41 481.6</td>
<td>45 001.6</td>
<td>39 451.6</td>
</tr>
<tr>
<td>7 500 000</td>
<td>1074.5</td>
<td>13 944.5</td>
<td>16 393.3</td>
<td>4018.6</td>
</tr>
<tr>
<td>10 000 000</td>
<td>0.0</td>
<td>6434.0</td>
<td>8087.8</td>
<td>159.6</td>
</tr>
</tbody>
</table>
Secura Re data V

- Fitted Erlang mixture estimates the net premium using intrinsically all data points, but postulate a lighter tail.

- Resulting net premiums are lower and differ strongly at the high-end of the sample range.

- Reinsurer should carefully investigate the tail behavior.
In order to estimate $\Pi(R)$ for values of $R$ smaller than the threshold, a global statistical model is needed.

Based on the mean excess plot, Beirlant et al. (2004) propose a mixture of an exponential and a Pareto (body-tail approach).

The fitting procedure for Erlang mixtures guides us to a mixture with two components, implicitly, in a data driven way.
## Secura Re data VII

**Table:** Non-parametric, Exp-Par and Mixture of Erlangs-based estimates for $\Pi(R)$.

<table>
<thead>
<tr>
<th>R</th>
<th>Non-Parametric</th>
<th>Exp-Par</th>
<th>Mixture of Erlangs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 250 000</td>
<td>981 238.0</td>
<td>944 217.8</td>
<td>981 483.1</td>
</tr>
<tr>
<td>1 500 000</td>
<td>760 637.6</td>
<td>734 371.6</td>
<td>760 912.9</td>
</tr>
<tr>
<td>1 750 000</td>
<td>583 403.6</td>
<td>571 314.1</td>
<td>582 920.1</td>
</tr>
<tr>
<td>2 000 000</td>
<td>445 329.8</td>
<td>444 275.5</td>
<td>444 466.6</td>
</tr>
<tr>
<td>2 250 000</td>
<td>340 853.2</td>
<td>344 965.2</td>
<td>339 821.4</td>
</tr>
<tr>
<td>2 500 000</td>
<td>263 052.7</td>
<td>267 000.7</td>
<td>262 314.6</td>
</tr>
</tbody>
</table>
Censored and truncated data

Censored sample $X = \{(l_i, u_i) \mid i = 1, \ldots, n\}$, truncated to the range $[t^l, t^u]$.

- $l_i$ and $u_i$: lower and upper censoring points.
- $t^l$ and $t^u$: lower and upper truncation points.
- $t^l \leq l_i \leq u_i \leq t^u$ for $i = 1, \ldots, n$.
- $t^l = 0$ and $t^u = \infty$ mean no truncation from below and above, resp.

Censoring status:

- Uncensored: $t^l \leq l_i = u_i =: x_i \leq t^u$
- Left Censored: $t^l = l_i < u_i < t^u$
- Right Censored: $t^l < l_i < u_i = t^u$
- Interval Censored: $t^l < l_i < u_i < t^u$
Complete data

Complete data $\mathcal{Y} = \{(x_i, z_i) | i = 1 \ldots n\}$ containing all uncensored observations $x_i$ and their corresponding component-indicator vector $z_i$ with

$$z_{ij} = \begin{cases} 
1 & \text{if observation } x_i \text{ comes from } j\text{th component density } f(x; r_j, \theta) \\
0 & \text{otherwise}
\end{cases}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, M$. 
Complete data log-likelihood

\[ l(\Theta; Y) = \sum_{i=1}^{n} \sum_{j=1}^{M} z_{ij} \ln \left( \beta_j f(x_i; t^l, t^u, r_j, \theta) \right), \]

with \( \beta_j = \alpha_j \cdot \frac{F(t^u; r_j, \theta) - F(t^l; r_j, \theta)}{F(t^u; \Theta) - F(t^l; \Theta)} \)

and

\[ f(x_i; t^l, t^u, r_j, \theta) = \frac{f(x_i; r_j, \theta)}{F(t^u; r_j, \theta) - F(t^l; r_j, \theta)}. \]
E step

E-step

\[ Q(\Theta; \Theta^{(k-1)}) = E(l(\Theta; \mathcal{Y}) \mid \mathcal{X}; \Theta^{(k-1)}) = Q_u(\Theta; \Theta^{(k-1)}) + Q_c(\Theta; \Theta^{(k-1)}), \]

split in an uncensored and censored part. E-step boils down to computing

\[ u_{Z_{ij}}^{(k)} = P(Z_{ij} = 1 \mid x_i, t^l, t^u, \Theta^{(k-1)}) = \frac{\alpha_j^{(k-1)} f(x_i; r_j, \theta^{(k-1)})}{\sum_{m=1}^{M} \alpha_m^{(k-1)} f(x_i; r_m, \theta^{(k-1)})}, \]

for \( i \in U \) and \( j = 1, \ldots, M \).

\[ c_{Z_{ij}}^{(k)} = P(Z_{ij} = 1 \mid l_i, u_i, t^l, t^u; \Theta^{(k-1)}) = \frac{\alpha_j^{(k-1)} \left( F(u_i; r_j, \theta^{(k-1)}) - F(l_i; r_j, \theta^{(k-1)}) \right)}{\sum_{m=1}^{M} \alpha_m^{(k-1)} \left( F(u_i; r_m, \theta^{(k-1)}) - F(l_i; r_m, \theta^{(k-1)}) \right)}, \]

for \( i \in C \) and \( j = 1, \ldots, M \).

\[ E \left( X_i \mid Z_{ij} = 1, l_i, u_i, t^l, t^u; \theta^{(k-1)} \right) = \frac{r_j \theta^{(k-1)} \left( F(u_i; r_j+1, \theta^{(k-1)}) - F(l_i; r_j+1, \theta^{(k-1)}) \right)}{F(u_i; r_j, \theta^{(k-1)}) - F(l_i; r_j, \theta^{(k-1)})}, \]

for \( i \in C \) and \( j = 1, \ldots, M \).
M step

\[ \Theta^{(k)} = \arg \max_{\Theta} Q(\Theta; \Theta^{(k-1)}) \]

leading to

\[
\begin{aligned}
\beta^{(k)}_j &= \frac{\sum_{i \in U} u z^{(k)}_{ij} + \sum_{i \in C} c z^{(k)}_{ij}}{n} \\
\theta^{(k)} &= \frac{\left(\sum_{i \in U} x_i + \sum_{i \in C} E \left( X_i \mid l_i, u_i, t^l, t^u; \theta^{(k-1)} \right) \right)}{\sum_{j=1}^M \beta^{(k)}_j r_j}
\end{aligned}
\]

with

\[
T^{(k)} = \sum_{j=1}^M \beta^{(k)}_j \left( \frac{t^l}{r_j} e^{-t^l/\theta} - \frac{t^u}{r_j} e^{-t^u/\theta} \right) \left( \frac{r_j - 1}{\theta r_j - 1} \right) \left( F(t^u; r_j, \theta) - F(t^l; r_j, \theta) \right) \bigg|_{\theta=\theta^{(k)}}.
\]
Choice of the shape parameters and of the number of Erlangs in the mixture

- Initial choice of $M$ and shape parameters $r = s \cdot (1, \ldots, M)$ with $s$ a spread factor.

- Initialization of $\Theta = (\alpha, \theta)$ based on Tijms’s proof of denseness (Tijms [1994]):

  $$\theta^{(0)} = \frac{\max(x)}{r_M} \quad \text{and} \quad \alpha_j^{(0)} = \sum_{i=1}^{n} I( r_{j-1}\theta^{(0)} < x_i \leq r_j\theta^{(0)} ) \frac{n}{n},$$

  for $j = 1, \ldots, M$.

- Apply EM algorithm, adjust the shapes $r$ by shifting $r_i$ by one in a double loop-wise fashion, apply EM algorithm, repeat until likelihood no longer increases.

- Reduce $M$ based on AIC or BIC by deleting the shape $r_i$ with smallest weight $\alpha_i$, refit and readjust shapes.