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

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

for $x > 0$ with

Mixtures of Erlangs with common scale

Probability density function

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

for $x > 0$ with

- number of Erlangs M;

Mixtures of Erlangs with common scale

Probability density function

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

for $x > 0$ with

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

Mixtures of Erlangs with common scale

Probability density function

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

for $x > 0$ with

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

Mixtures of Erlangs with common scale

Probability density function

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

for $x>0$ with

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

Mixtures of Erlangs with common scale

Probability density function

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

for $x > 0$ with

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

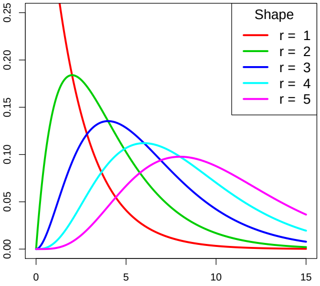


Figure: Varying the shape r with scale $\theta = 2$.

Mixtures of Erlangs with common scale

Probability density function

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

for $x>0$ with

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

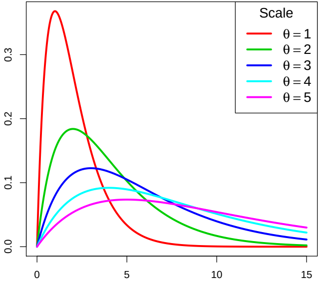


Figure: Varying the scale θ with shape $r=2$.

Why mixtures of Erlangs?

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

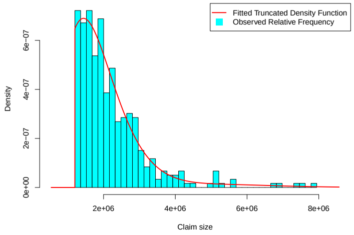
- versatile classes of distributions;
- mathematically tractable allowing analytical expressions of quantities of interest;
- fitting procedure based on the EM algorithm;
- able to deal with censored and/or truncated data;
- implemented in R, making use of the package `doParallel`.

Secura Re data I

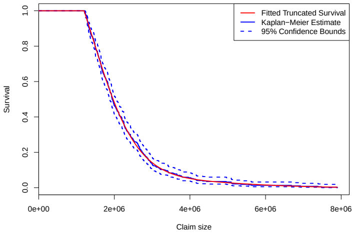
- Price an unlimited excess-loss layer above an operational priority R.
- 371 automobile claims from 1988 until 2001 from several European insurers, corrected, among others, for inflation.
- Left truncated at 1 200 000 euro, since the claims are only reported to the reinsurer if they are larger.

Secura Re data II

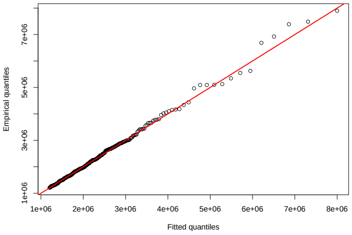
Parameter estimates		
r_j	α_j	θ
5	0.971	360 096.1
15	0.029	



(a) Fitted density function and histogram.



(b) Fitted survival and Kaplan-Meier.



(c) QQ plot.

Secura Re data III

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 θ but with different weights.

References

Verbelen, R., Gong, L., Antonio, K., Badescu, A., and Sheldon, L. [2014]
Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm.
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 $\Pi(R)$.

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

Secura Re data V

- Fitted Erlang mixture estimates the net premium using intrinsically all data points, but postulate a ligher tail.
- Resulting net premiums are lower and differ strongly at the high-end of the sample range.
- Reinsurer should carefully investigate the tail behavior.

Secura Re data VI

- 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 $\Pi(R)$.

R	Non-Parametric	Exp-Par	Mixture of Erlangs
1 250 000	981 238.0	944 217.8	981 483.1
1 500 000	760 637.6	734 371.6	760 912.9
1 750 000	583 403.6	571 314.1	582 920.1
2 000 000	445 329.8	444 275.5	444 466.6
2 250 000	340 853.2	344 965.2	339 821.4
2 500 000	263 052.7	267 000.7	262 314.6

Censored and truncated data

Censored sample $\mathcal{X} = \{ (l_i, u_i) | \, i = 1, \dots, n \}$, 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 $i = 1, \dots, n$.
- $t^l = 0$ 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 $\mathcal{Y} = \{(x_i, \boldsymbol{z_i}) | i = 1 \ldots n\}$ containing all uncensored observations x_i and their corresponding component-indicator vector $\boldsymbol{z}_i$ 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

Complete data log-likelihood

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

with

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

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

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

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

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

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

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

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

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

M step

M-step $$\mathbf{\Theta}^{(k)} = \arg\max_{\mathbf{\Theta}} Q(\mathbf{\Theta}; \mathbf{\Theta}^{(k-1)})$$

leading to

$$\begin{cases} \beta_j^{(k)} = \dfrac{\sum_{i \in U} {}^{u}z_{ij}^{(k)} + \sum_{i \in C} {}^{c}z_{ij}^{(k)}}{n} & \text{for } j = 1, \ldots, M\,, \\ \theta^{(k)} = \dfrac{\left(\sum_{i \in U} x_i + \sum_{i \in C} E\left(X_i \,\middle|\, l_i, u_i, t^l, t^u; \theta^{(k-1)}\right)\right)/n - T^{(k)}}{\sum_{j=1}^{M} \beta_j^{(k)} r_j} , \end{cases}$$

with

$$T^{(k)} = \sum_{j=1}^{M} \beta_j^{(k)} \frac{\left(t^l\right)^{r_j} e^{-t^l/\theta} - \left(t^u\right)^{r_j} e^{-t^u/\theta}}{\theta^{r_j-1}(r_j-1)!\left(F(t^u; r_j, \theta) - F(t^l; r_j, \theta)\right)}\Bigg|_{\theta=\theta^{(k)}} .$$

Choice of the shape parameters and of the number of Erlangs in the mixture

- Initial choice of M and shape parameters $\boldsymbol{r} = s \cdot (1, \dots, M)$ with s a spread factor.
- Initialization of $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \theta)$ based on Tijms's proof of denseness (Tijms [1994]):

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

  for $j = 1, \dots, M$.
- Apply EM algorithm, adjust the shapes $\boldsymbol{r}$ 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 α_i, refit and readjust shapes.