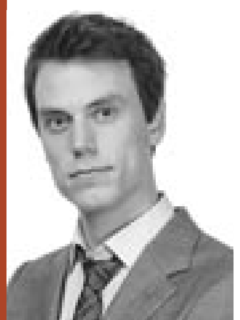
By Katrien Antonio and Roel Verbelen



Dr. K. Antonio is Associate Professor in Actuarial Science at KU Leuven and Assistent Professor in Actuarial Science at Univeristy of Amsterdam.

R. Verbelen is PhD Student in Statistics and Actuarial Science at KU Leuven.

LOSS MODELING USING MIXTURES OF ERLANGS

At the 2nd R in insurance conference on July 14, 2014 at Cass Business School in London, Roel Verbelen presented his research findings on a class of flexible distributions called mixtures of Erlangs. In several applications in the context of loss modelling, he demonstrated his implemented fitting procedure and graphical tools built in R.

Modeling data on claim sizes is crucial when pricing insurance products. Insurance data are often modeled using a parametric distribution such as a gamma, lognormal or Pareto distribution. The usual way to proceed in loss modeling, pricing and reserving is to calibrate the data using several of these distributions and then select, among these, the most appropriate model based on an information criterion. These classes of distributions may however not always be flexible enough in terms of the possible shapes of their members in order to obtain a satisfying fit (e.g. multimodality). Furthermore, using these distributions in aggregate loss models does not lead to an analytical form for the corresponding aggregate loss distribution. Evaluation of the model (e.g. the calculation of risk measures) is therefore based on simulation algorithms. Ideally, loss models require on the one hand the flexibility of nonparametric density estimation techniques to describe the insurance losses and on the other hand the feasibility to analytically quantify the risk, which is exactly what mixtures of Erlangs have to offer.

A mixture of M Erlang distributions with common scale parameter $\theta > 0$ has density

$$f(x;\boldsymbol{\alpha},\boldsymbol{r},\theta) = \sum_{j=1}^{M} \alpha_j \frac{x^{r_j-1} e^{-\frac{x}{\theta}}}{\theta^{r_j}(r_j-1)!} = \sum_{j=1}^{M} \alpha_j f(x;r_j,\theta) \qquad \text{for } x > 0,$$

where the positive integers $\boldsymbol{r} = (r_1, \dots, r_M)$ with $r_1 < \dots < r_M$ are the shape parameters of the Erlang distributions and $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_M)$ with $\alpha_j > 0$ and $\sum_{j=1}^{M} \alpha_j = 1$ are the weights used in the mixture. The class of mixtures of Erlang distributions is very flexible in terms of the possible shapes of its members. It has been shown that this class is *dense* in the space of positive distributions. At the same time, it is possible to work analytically with this class leading to explicit expressions of many quantities of interest (see Klugman et al., 2013; Willmot and Lin, 2011; Willmot and Woo, 2007).

In actuarial science, data are often incomplete due to censoring and truncation. Data are censored in case you only observe an interval in which a data point is lying without knowing its exact value. Truncation entails that it is only possible to observe the data of which the values lie in a certain range. Policy modifications such as deductibles lead to left truncated losses and policy limits to right censored claim sizes. Left truncation is also present in life insurance where members of pension schemes and holders of insurance contracts only enter a portfolio at a certain adult age. Censored and truncated data occur in the context of claim reserving as well (see Antonio and Plat, 2014). Indeed, the reserving actuary wants to predict the future development of claims when setting aside reserves at the present moment and has to deal with claims being reported but not yet settled (RBNS) and claims being incurred but not yet reported (IBNR). In operational risk, data are left truncated as they are only recorded in case they exceed a certain threshold.

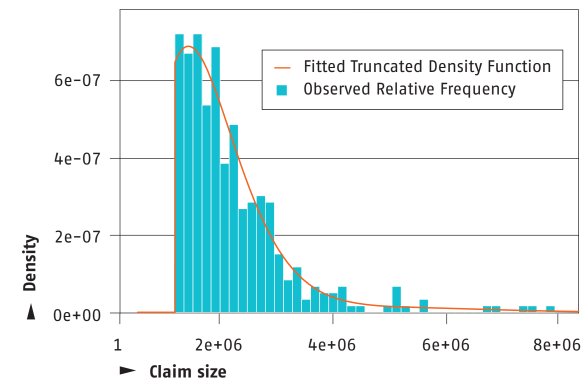
Interest however is in the underlying distribution of the uncensored and untruncated data instead of the observed censored and/or truncated data. Hence the censoring and truncation has to be accounted for in the analysis. In Verbelen et al. (2015b), we develop an extension of the fitting mixtures of Erlangs to censored and truncated data. We implement the fitting procedure using R[1] and show how mixtures of Erlangs can be used to adequately represent any univariate distribution in a wide variety of applications where data are allowed to be censored and truncated.

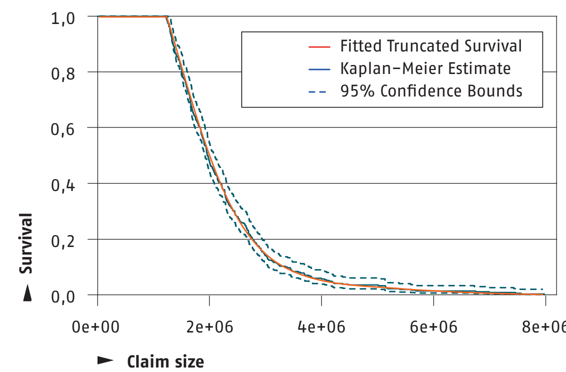
One of the real world datasets we analyze is the Secura Re dataset discussed in Beirlant et al. (2004). The data contain information on 371 automobile claims from 1988 until 2001 gathered from several European insurance companies. The data are uncensored, but left truncated at 1200000 since a claim is only reported to the reinsurer if the claim size is at least as large as 1200000 euro. The use of mixtures of Erlangs provides a smooth fit of the insurance losses (see Figure 1) as well as an analytical price for an excess-of-loss reinsurance contract.

1 – Available online at www.feb.kuleuven.be/roel.verbelen.

(a) Parameter estimates

r_j	α_j	θ
5	0.971	360 096.1
15	0.029	





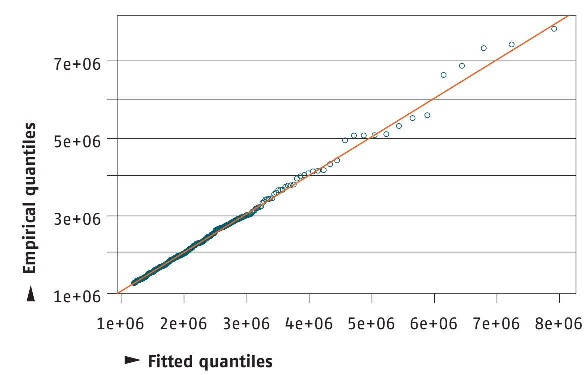


Figure 1: Parameter estimates and graphical goodness-of-fit plots of the fitted mixture of 2 Erlangs for the Secura Re dataset.

Lee and Lin (2012) introduce the class of multivariate mixtures of Erlang distributions (MME) as a generalization of the class of univariate mixture of Erlang distributions to model dependent losses. The authors show how this multivariate extension retains the desirable properties of flexibility (denseness) on the one hand and tractability at the other hand (see also Willmot and Woo, 2014).

A d-variate Erlang mixture is defined as a mixture such that each mixture component is the joint distribution of d independent Erlang distributions with a common scale parameter $\theta > 0$. The dependence structure is captured by the combination of the positive integer shape parameters of the Erlangs in each dimension.

We denote the positive integer shape parameters of the jointly independent Erlang distributions in a mixture component by the vector $\boldsymbol{r} = (r_1, \dots, r_d)$ and the set of all shape vectors with non-zero weight by R. The mixture weights are denoted by $\boldsymbol{\alpha} = \{\alpha_r \mid \boldsymbol{r} \in R\}$ and must satisfy $\alpha_r \geq 0$ and $\sum_{r \in R} \alpha_r = 1$. The density of a d-variate Erlang mixture evaluated in $\boldsymbol{x} = (x_1, \dots, x_d)$ with $x_j > 0$ for $j = 1, \dots, d$ can then be written as

$$f(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{r},\theta) = \sum_{r \in R} \alpha_r f(\boldsymbol{x};\boldsymbol{r},\theta) = \sum_{r \in R} \alpha_r \prod_{j=1}^{d} f(x_j;r_j,\theta) = \sum_{r \in R} \alpha_r \prod_{j=1}^{d} \frac{x_j^{r_j-1} e^{-x_j/\theta}}{\theta^{r_j}(r_j-1)!}.$$

The two main novelties we present in Verbelen et al. (2015a) are (i) an extension of the fitting procedure of MME to be able to deal with censored and truncated data and (ii) a computationally more efficient initialization and adjustment strategy for the shape parameter vectors in order to make the estimation procedure more flexible and effective. The improvements (i) and (ii) allow us to analyze realistic data with diverse forms of dependence.

Here, we illustrate the developed fitting technique with an additional actuarial science example, not discussed in the paper. We consider an insurance dataset, collected by the US Insurance Services Office (ISO), comprising of 1500 non-life insurance claims of which both the indemnity payment or loss as well as the allocated loss adjustment expense (ALAE) are observed, both in USD. ALAE is the additional expense associated with the settlement of the claim, e.g. lawyers' fees , experts' opinions, and claims investigation expenses. For each claim, we also recorded the policy limit of the contract, due to which 34 claims are right censored. Even though only 34 of the 1500 observations are right censored, the censoring cannot be neglected and has to be taken into account when estimating the joint distribution. This dataset is also studied in e.g. Frees and Valdez (1998); Klugman and Parsa (1999); Beirlant et al. (2004); Denuit et al. (2006a,b).

Applying our proposed methodology allows us to describe the joint distribution of the losses and the expenses using an MME, either on the original or on the log scale. In Figure 2 we show the results on the log scale for the best-fitting MME with 8 mixture components. The marginals as well as the dependence structure seem to be captured appropriately. We further use this fit to estimate the conditional distribution of ALAE given the value of the claim loss. Besides, we use the tractability of MME to derive explicit expression of the net reinsurance premiums of an excess-of-loss reinsurance layer C *xs* R for various levels of the retention R and the limit L = R + C.

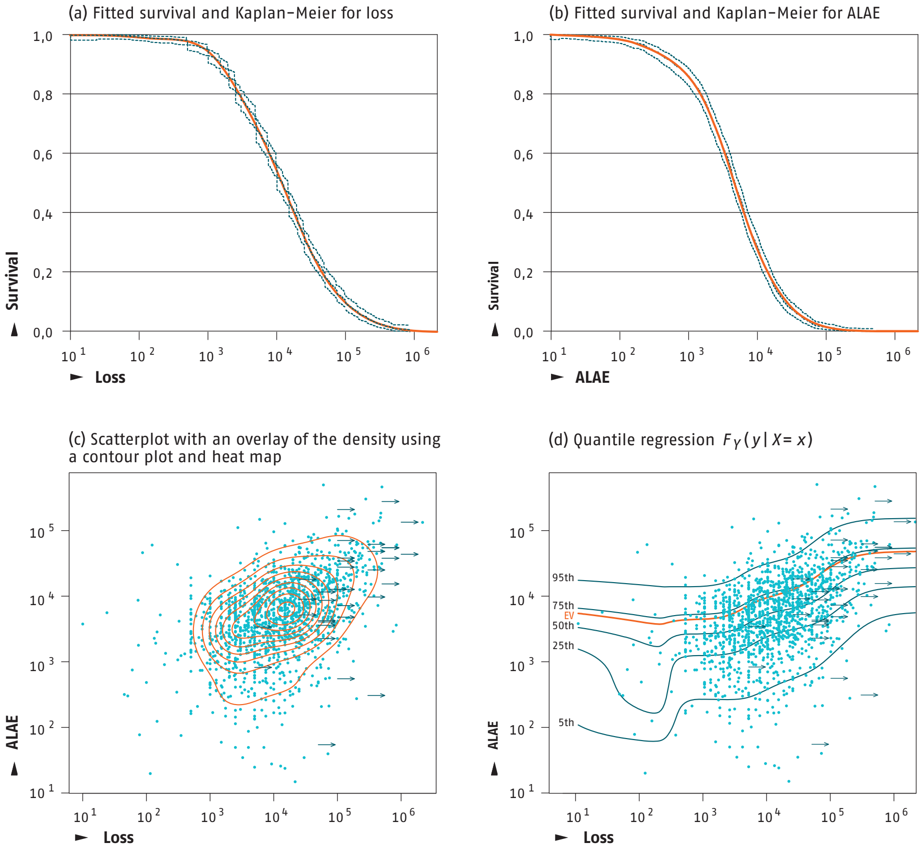


Figure 2: Graphical goodness-of-fit plots of the fitted MME with 8 mixture components for the Loss ALAE dataset on the log scale.

References

- Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. Scandinavian Actuarial Journal, 2014(7):649–669.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., and Ferro, C. (2004). Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics. Wiley.
- Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006a). Actuarial Theory for Dependent Risks: Measures, Orders and Models. Wiley.
- Denuit, M., Purcaru, O., and Van Keilegom, I. (2006b). Bivariate archimedean copula models for censored data in non-life insurance. Journal of Actuarial Practice, 13:5–32.
- Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. North American actuarial journal, 2(1):1–25.
- Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2013). Loss models: Further topics. John Wiley & Sons.
- Klugman, S. A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. Insurance: Mathematics and Economics, 24(1-2):139–148.
- Lee, S. C. and Lin, X. S. (2012). Modeling dependent risks with multivariate Erlang mixtures. ASTIN Bulletin, 42(1):153–180.
- Verbelen, R., Antonio, K., and Claeskens, G. (2015a). Multivariate mixtures of Erlangs for density estimation under censoring and truncation. Submitted for publication.
- Verbelen, R., Gong, L., Antonio, K., Badescu, A., and Lin, X. S. (2015b). Fitting mixtures of Erlangs to censored and truncated data using the EM algorithm. Submitted for publication.
- Willmot, G. E. and Lin, X. S. (2011). Risk modelling with the mixed Erlang distribution. Applied Stochastic Models in Business and Industry, 27(1):2–16.
- Willmot, G. E. and Woo, J.-K. (2007). On the class of Erlang mixtures with risk theoretic applications. North American Actuarial Journal, 11(2):99–115.
- Willmot, G. E. and Woo, J.-K. (2014). On some properties of a class of multivariate Erlang mixtures with insurance applications. ASTIN Bulletin, FirstView:1–23.