B-KUL-D0N55A, Advanced Non-Life Insurance Mathematics, Home Assignment 1 – Part I

Dieter Verbeke

November 2016

1 Introduction

This text covers the first part of the first assignment for the course Advanced Non-life Insurance Mathematics. It addresses the task of fitting various parametric models to censored and truncated loss data, and compares results.

2 Data

The file `SeverityCensoring.txt` contains information about 9062 claims paid by an insurance company over some observation period. It comprises five columns. The first two, `policyID` and `claimID`, are classifiers that are of no particular interest to us. The column `claimAmount` lists the amount the insurance company paid on each claim. The column `deductible` reveals that there is a fixed deductible of 100 EUR for each policy. As a result, all observed claim amounts are truncated from the left at 100. The column `rc` indicates whether right-censoring is present or not: `NA` expresses that the claim is fully settled by the end of the observation period and therefore the observed claim amount is the full (uncensored) loss associated with it. On the other hand, a number in the `rc` column implies that the claim is not yet fully settled, and consequently the observed claim amount is right-censored.

3 Maximum likelihood estimation of parametric loss models

3.1 Likelihood of right-censored and left-truncated data

In [1] a generic framework is presented that accounts for left truncation and right censoring of the observations. For independent samples $\{(y_1, t_1, \delta_1), \ldots, (y_n, t_n, \delta_n)\}$, the likelihood is

$$\prod_{\delta_i=0} \frac{f(y_i)}{1 - F(t_i)} \cdot \prod_{\delta_i=1} \frac{1 - F(y_i)}{1 - F(t_i)}. \tag{1}$$

In the above $f(x)$ denotes the probability density function (pdf) and $F(x)$ the cumulative density function (cdf) of the underlying (actual) severities. Furthermore, y_i is the registered claim, t_i the treshold or deductible, and δ_i the indicator function, i.e.

$$\delta_i = \begin{cases} 1: & y_i \text{ is censored} \\ 0: & \text{otherwise} \end{cases} \tag{2}$$

3.2 Parametric models

Various parametric models were fitted to the censored and truncated loss data following a maximum likelihood approach. With the general expression for the likelihood (Eq. 1) given, computing the likelihood for each parametric model is straightforward. Five different distributions were examined, namely, an exponential, a log-normal, an inverse Gaussian, a Burr type XII, and a five component Erlang mixture distribution. For the sake of brevity, the reader is referred to [2] (Burr type XII) and [1] (others) for details on their particular parameterization.
In all cases direct optimization of the log-likelihood function was applied, with the notable exception of the Erlang mixture distribution. For the latter, the expectation-maximization (EM) algorithm developed in [4] was employed.

4 Non-parametric inference

A convenient non-parametric estimator for the so-called survival function $1 - F(x)$ of truncated and censored severities is the Kaplan-Meier estimator (see [3], Chapter 12).

5 Technicalities of the implementation

5.1 Numerical optimization

For all but the Erlang mixture distribution direct optimization was carried out, using a general-purpose numerical optimization routine. To be precise, the native `optim()` function based on the Nelder-Mead algorithm was used. This choice of algorithm refutes the need for derivative information. Numerical optimization requires (good) initial values for the decision variables.
The method of moments (MoM) was used to obtain a reasonable initialization. In short, the method of moments involves equating the sample moments of the raw loss data (y_i) [1] with the theoretical moments of the respective distributions. This principle could be applied easily, except for the Burr distribution and the Erlang mixture distribution. Table 1 gives the expressions used in the MoM for the exponential, log-normal and inverse Gaussian distribution.

[1] In other words, when intializing the parameters, the data were treated as if no censoring or truncation occured.

Table 1: Expressions used for the method of moments.

Distribution	# par.	Relations
Exponential	1	$\mathrm{E}[y] = 1/\lambda$
Log-normal	2	$\mathrm{E}[y] = \mu,\ \mathrm{Var}[y] = \mu^3/\lambda$
Inverse gamma	2	$\mathrm{E}[y] = \exp(\mu + \frac{\sigma^2}{2}),\ \mathrm{Var}[y] = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$

The MoM for the Burr distribution can only be evaluated numerically. Instead of pursuing this direction a more pragmatic approach was taken. To remediate the influence of the initialization a range of starting values was explored. Eventually, the optimization algorithm was initialized with $\alpha_0 = 1$, $\gamma_0 = 1$, $\theta_0 = 10^{-12}$. Although the outcome of this procedure was satisfactory, I must admit no sound mathematical guarantees can be given. A more sophisticated approach is definitely possible.
The EM-algorithm is an iterative method and requires initial values. Finding good starting values is more intricate than in the previous cases. However, the initialization is taken care of by the routines developed in [4].
Table 2 contains a summary of the starting values for each parametric model.

Table 2: Starting values for the various parametric models.

Distribution	# par.	Initial values
Exponential	1	$\lambda_0 = 5.4458 \cdot 10^{-4}$
Log-normal	2	$\mu_0 = 6.7751,\ \sigma_0 = 1.4807$
Inverse gamma	2	$\mu_0 = 1836.0269,\ \lambda_0 = 540.6223$
Burr type XII	3	$\alpha_0 = 1,\ \gamma_0 = 1,\ \theta_0 = 10^{-12}$
Erlang mixture	11	$\{\alpha_i\}_0 = \{0.8301, 1.0159 \cdot 10^{-9}, 6.6816 \cdot 10^{-3}, 0.1274, 0.0357\}$
		$\{r_i\}_0 = \{4, 6, 9, 14, 248\}$
		$\theta_0 = 163.8607$

5.2 Dependencies

The implementation relies on a number of code libraries. A list of all used packages is given here.

`actuar:` Burr distribution.

`survival:` Inverse Gaussian distribution.

`statmod:` Kaplan-Meier estimate.

`2014-12-16_ME.R:` Erlang mixture library.

`plotly`: Interactive charting library.

5.3 Code

The code for the second part of the assignment is provided in the supplementary material. `ANLIM1_part1.R` is the main file, while `define_custom_functions.R` contains some supporting functions.

6 Results

Table 3 provides the ML parameter estimates, and the value of the Akaike information criterion (AIC) for each distribution. Based on the AIC criterion the Erlang mixture delivers the best fit, followed by the Burr, inverse Gaussian, the log-normal, and eventually the exponential distribution. This ordering is also reflected in a graphical display. Figure 1 shows the Kaplan-Meier estimate of the survival function for the loss data, as well as the best-fitting parametric survival functions. As the AIC value decreases amongst the various models, the corresponding survival function matches the Kaplan-Meier estimate more closely.

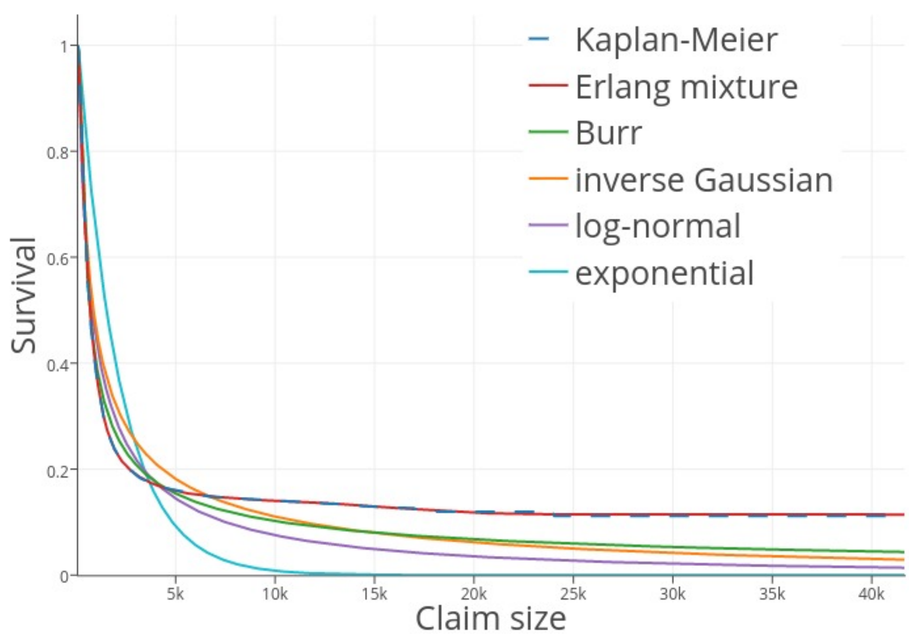

Figure 1: Kaplan-Meier estimate of the survival function for the loss data, as well as the best-fitting Erlang mixture, Burr type XII, inverse Gaussian, log-normal and exponential survival functions.

Table 3: Maximum likelihood parameter estimates for various distributions.

Distribution	**# par.**	**MLE**[a]	**AIC**[b]
Exponential	1	$\lambda = 4.8651 \cdot 10^{-4}$	127,996
Log-normal	2	$\mu = 6.1544$, $\sigma = 1.9553$	121,207
Inverse gamma	2	$\mu = 5508.2863$, $\lambda = 427.9284$	120,652
Burr type XII	3	$\alpha = 3.0419$, $\gamma = 0.1577$, $\theta = 33.7171$	119,870
Erlang mixture	11	$\{\alpha_i\}$={0.8228, 0.0462, 0.01073, 0.0200, 0.1002}	119,843
		$\{r_i\} = \{1, 6, 14, 27, 96\}$	
		$\theta = 612.2609$	

[a] Maximum likelihood estimate.
[b] Akaike information criterion.

References

[1] K. Antonio and R. Verbelen. Claim severity modeling, 2016. Unpublished lecture notes, Advanced Non-Life Insurance Mathematics, B-KUL-D0N55A.

[2] Christophe Dutang, Vincent Goulet, and Mathieu Pigeon. actuar: An r package for actuarial science. *Journal of Statistical Software*, 25(7):38, 2008.

[3] S.A. Klugman, H.H. Panjer, and G.E. Willmot. *Loss Models: From Data to Decisions*. Wiley Series in Probability and Statistics. Wiley & Sons, 2012.

[4] Roel Verbelen, Lan Gong, Katrien Antonio, Andrei Badescu, and Sheldon Lin. Fitting mixtures of erlangs to censored and truncated data using the em algorithm. *ASTIN Bulletin: The Journal of the International Actuarial Association*, 45(03):729–758, 2015.

B-KUL-D0N55A, Advanced Non-Life Insurance Mathematics, Home Assignment 1 – Part II

Dieter Verbeke

November 2016

1 Introduction

This text covers the second part of the first assignment for the course Advanced Non-life Insurance Mathematics. It examines a splicing technique for fitting a body-tail combination model to a set of loss data, using a shifted exponential distribution for the body and a Pareto distribution with unit scale for the tail.

2 Data

The file `SecuraRe.txt` contains information about 371 claims reported over some observation period. It comprises two columns. The first column specifies the year in which the loss was incurred, the second one lists the claim amount.

3 A body-tail combination with splicing

In this part of the assignment, splicing is used to fit a body-tail combination to a data set. The combination consists of a shifted exponential distribution for the body, and a Pareto distribution with unit scale for the tale ([1]). The probability density function of the spliced distribution is

$$f(x) = \begin{cases} \frac{n-k}{n} \cdot \frac{\lambda \exp(-\lambda(x-1{,}200{,}000))}{1-\exp(-\lambda(X_{n-k,n}-1{,}200{,}000))} & x \leq X_{n-k,n} \\ \frac{k}{n} \cdot \frac{\alpha(x+1)^{-(\alpha+1)}}{(X_{n-k,n}+1)^{-\alpha}} & x > X_{n-k,n} \end{cases}, \tag{1}$$

where n is the sample size and $X_{n-k,n}$ is denotes the $(k+1)^{\text{th}}$ largest observation in the data set. The cumulative density function for this distribution is

$$F(x) = \begin{cases} \frac{n-k}{n} \cdot \frac{1-\exp(-\lambda(x-1{,}200{,}000))}{1-\exp(-\lambda(X_{n-k,n}-1{,}200{,}000))} & x \leq X_{n-k,n} \\ 1 - \frac{k}{n} \cdot \left(\frac{x+1}{X_{n-k,n}+1}\right)^{\alpha} & x > X_{n-k,n} \end{cases} \tag{2}$$

Expression (2) straightforwardly follows from computing the integrals

$$F(x) = \int_0^x f(t)\mathrm{d}t \quad \text{for} \quad x \leq X_{n-k,n}, \tag{3}$$

and

$$F(x) = \frac{n-k}{n} + \int_{X_{n-k,n}}^{x} f(t)\mathrm{d}t \quad \text{for} \quad x > X_{n-k,n}, \tag{4}$$

with $f(x)$ given by (1). Moreover, one can easily verify that $F(x)$ is continuous in $x = X_{n-k,n}$, and that it is a non-decreasing function with $\lim_{x\to\infty} F(x) = 1$. Therefore, the functions $f(x)$ and $F(x)$ define a probability distribution.

4 Maximum likelihood estimation

Maximum likelihood estimation is applied for inferring the parameters of the spliced distribution. Given $\hat{k} = 95$ from extreme value theory, the parameters λ and α are sought via direct optimization of the log-likelihood of the registered claims x_i

$$LL = \sum_{i=1}^{n} \log(f(x_i)). \tag{5}$$

5 Technicalities of the implementation

5.1 Numerical Optimization

Direct optimization using a general-purpose numerical optimization routine was carried out. To be precise, the native `optim()` function based on the Nelder-Mead algorithm was used. This choice of algorithm refutes the need for derivative information. Numerical optimization requires (good) initial values for the decision variables. The method of moments (MoM) was considered for this purpose. The theoretical mean of the data of the Pareto component was related to the mean of the observations x_i in the tail segment,

$$\mathrm{E}[x|x > X_{n-k,n}] = \int_{X_{n-k,n}}^{\infty} \frac{k}{n} \cdot \frac{\alpha(x+1)^{-(\alpha+1)}}{(X_{n-k,n}+1)^{-\alpha}} \mathrm{d}x, \tag{6}$$

leading to the following initialization

$$\alpha_0 = \frac{\frac{n}{k}\sum_i (x_i+1)}{\frac{n}{k}\sum_i (x_i+1) - (X_{n-k,k}+1)} \quad \forall i:\ x_i + 1 > X_{n-k,n} + 1 \tag{7}$$

The method of moments for the body segment, involves the numerical solution of a nonlinear equation in λ. To avoid this, again, a rather pragmatic approach was taken. Several initializations were tested, and eventually the starting value was set close to zero, i.e. $\lambda_0 = 1 \cdot 10^{-12}$. This resulted in satisfactory convergence behaviour.

5.2 Code

The code for the second part of the assignment is provided in the supplementary material. `ANLIM1_part2.R` is the main file, while `define_custom_functions.R` contains some supporting functions.

6 Results

Figure 1 compares the empirical cumulative distribution function (ecdf) with the cumulative distribution function (cdf) of the fitted composite distribution. As is evident from the graph, a good fit was obtained over the entire range of observed losses. At last, Table 1 lists the ML parameter estimates, as well as the initial values of the numerical optimization routine.

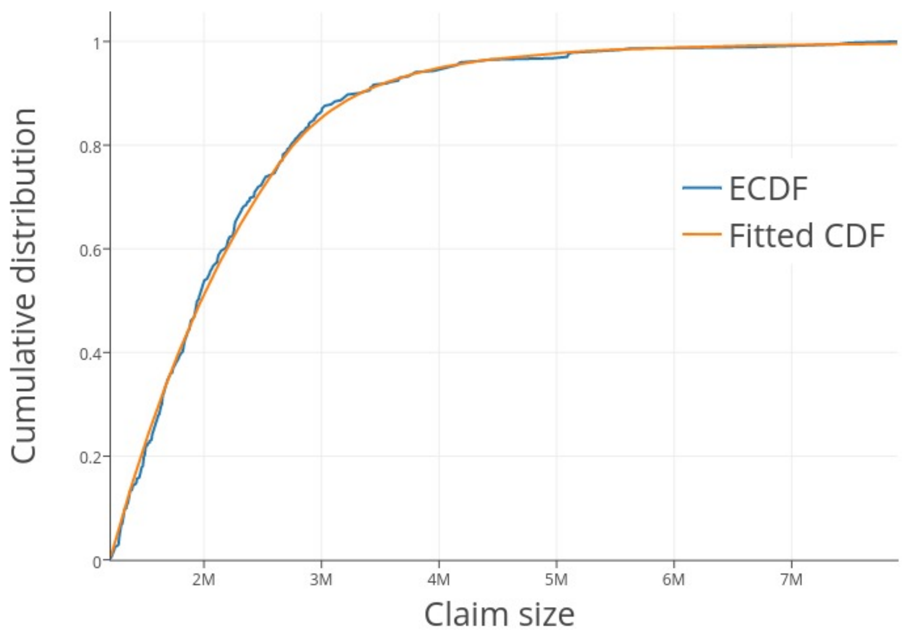

Figure 1: The empirical cumulative distribution function (ECDF) of the loss data, together with the cumulative distribution function (CDF) of the spliced distribution.

Table 1: Maximum likelihood parameter estimates.

Parameters	**Initial values**	**MLE[a]**
λ	$1 \cdot 10^{-12}$	$6.7170 \cdot 10^{-7}$
α	1.2306	3.6868

[a] Maximum likelihood estimate.

References

[1] K. Antonio and R. Verbelen. Claim severity modeling, 2016. Unpublished lecture notes, Advanced Non-Life Insurance Mathematics, B-KUL-D0N55A.