LOSS MODELING WITH MIXTURES DISTRIBUTIONS IN R PACKAGE

Grzegorz Sitek

University of Economics in Katowice
e-mail: grzegors12@wp.pl

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY
Nr 15(21)

ISSN 1644-6739
e-ISSN 2449-9765

DOI: 10.15611/sps.2017.15.09
JEL Classification: C46

Summary: Finite mixtures of probability distributions may be successfully used in the modeling of probability distributions of losses. These distributions are typically heavy tailed and positively skewed. Finding the distribution that fits loss data well is often difficult. The paper shows that the use of mixed models can significantly improve the goodness-of-fit of the loss data. The paper also presents an algorithm to find estimates of parameters of mixture distribution and gives an illustrative example. The analytical approach is probably the most often used in practice and certainly the most frequently adopted in the actuarial literature. It is reduced to finding a suitable analytical expression which fits the observed data well. For parameters estimation we use the maximum likelihood method applying the Newton-Raphson and EM algorithm. Computations of goodness-of-fit can be judged using the Akaike information criterion.

Keywords: finite mixture of distributions, loss distribution, maximum likelihood estimate, EM algorithm.

1. Introduction

Many problems in actuarial science involve the building of a mathematical model that can be used to forecast or predict insurance costs. Hence modeling is an important procedure for actuaries so that they can estimate the degree of uncertainty as to when a claim will be made and how much will be paid. In particular, the modeling of claims and outstanding claims lead to the pricing of insurance premiums and an estimation of claim reserving, respectively. The most useful approach to uncertainty representation is through probability, so we will concentrate on probability models. Loss distribution, as any other probability distribution, is completely determined by the cumulative distribution F(x) or density functions f(x).

Fitting an adequate loss distribution to real insurance data sets is a relevant problem and not an easy task in actuarial literature. Since it is well recognized that the loss distribution is strongly skewed with

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 15(21)

heavy tails, different candidates for claim severity distribution have been considered in the applications: the log-Normal, the Burr, the Weibull, the Gamma and the Pareto distribution in the context of Extreme Value Theory, see for example [McNeil 1997] and [Embrechts, Klüppelberg, Mikosch 1997].

Finite mixtures of probability distributions may be successfully used in the modeling of probability distributions of loss. [Bernardi, Maruotti, Lea 2012] propose a mixture of Skew Normal densities for modeling the loss distribution to deal with data displaying large and positive skewness as well as a wide right tail.

2. Loss distribution

The lognormal, Pareto, Burr, Weibull, and gamma distributions are typical candidates for claim size distributions to be considered in this paper.

2.1. Lognormal distribution

Two parameter lognormal distribution is given by the density

$$f(y) = \frac{1}{\sqrt{2\pi}\, y\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (\ln y - \mu)^2 \right\} \quad y > 0 . \tag{1}$$

If a random variable Y has a lognormal distribution, LN(μ, σ^2), then a variable log(Y) has a normal distribution with expectation μ and variance σ^2. In empirical studies of wage and income, distributions are considered as three-parameter distribution, in which in addition to μ and σ there is a third parameter τ. The parameter τ is the theoretical minimal value of Y. The k-th raw moment m_k of the log-normal variate can be easily derived using the results for normal random variables:

$$m_k = \exp(k\mu + \frac{1}{2} k^2 \sigma^2) . \tag{2}$$

The maximum likelihood estimates m and s^2 for the parameter μ and the parameter σ are

$$\hat{m} = \frac{1}{n} \sum_{i=1}^{n} \ln y_i , \quad \hat{s}^2 = \frac{1}{n} \sum_{i=1}^{n} (\ln y_i - m)^2 . \tag{3}$$

The log-normal distribution is very useful in the modeling of claim sizes. It is right-skewed, has a thick tail and fits many situations well.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 15(21)

2.2. Gamma distribution

The gamma (more precisely, the Pearson type III) distribution is certainly among the five most popular distributions in applied statistics when unimodal and positive data are available.

The density of the gamma distribution is

$$f(y) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \text{ for } y > 0,\ \beta,\ \alpha > 0, \tag{4}$$

where α, $\beta > 0$, with α being a shape and β a scale parameter. The likelihood equations for a simple random sample of size n are

$$\begin{cases} \sum_{i=1}^{n} \log X_i - n \log \hat{\beta} - n\Psi(\hat{\alpha}) = 0 \\ \sum_{i=1}^{n} X_i - n\hat{\alpha}\hat{\beta} = 0 \end{cases}. \tag{5}$$

These can be solved iteratively, and indeed procedures for estimation in the gamma distribution are nowadays available in many statistical software packages. The k-th raw moment can be easily derived from the Laplace transform:

$$m_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\beta^k}. \tag{6}$$

2.3. Pareto distribution

The density of a Pareto distribution is

$$f(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \quad x > 0. \tag{7}$$

The Pareto distribution is defined in terms of its c.d.f.

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha} \qquad x > 0. \tag{8}$$

Clearly, the shape parameter α and the scale parameter λ are both positive. The k-th raw moment:.

$$m_k = \lambda^k k! \frac{\Gamma(\alpha - k)}{\Gamma(\alpha)}. \tag{9}$$

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 15(21)

The expected value of a Pareto random variable distribution is

$$E(X) = \frac{\lambda}{\alpha - 1} \quad \alpha > 1. \tag{10}$$

The variance of a Pareto random variable distribution is

$$D^2(X) = \frac{\lambda^2 \alpha}{(\alpha - 2)(\alpha - 1)^2} \quad \alpha > 2. \tag{11}$$

Hence the Pareto distribution has very heavy tails. The method of moments estimators are given by:

$$\hat{\alpha} = 2\frac{\hat{m}_2 - \hat{m}_1^2}{\hat{m}_2 - 2\hat{m}_1^2}. \tag{12}$$

$$\hat{\lambda} = \frac{\hat{m}_2 \hat{m}_1}{\hat{m}_2 - 2\hat{m}_1^2}. \tag{13}$$

Note that the estimators are well defined only when $\hat{m}_2 - 2\hat{m}_1^2 > 0$. Unfortunately there are no closed form expressions for the maximum likelihood estimators and they can only be evaluated numerically.

Suppose that a variable X has (conditional on β) an exponential distribution with mean β^{-1}. Further, suppose that β itself has a gamma distribution. The unconditional distribution of X is a mixture and is called the Pareto distribution. Moreover, it can be shown that if X is an exponential random variable and Y is a gamma random variable, then X/Y is a Pareto random variable.

As in the log-normal distribution, the Laplace transform does not have a closed form representation and the moment generating function does not exist. Moreover, like the exponential pdf, the Pareto density is monotone decreasing, which may not be adequate in some practical situations.

2.4. Burr distribution

The c.d.f.'s of all Burr distributions satisfy the differential equation

$$F'(x) = F(x)[1 - F(x)]g(x). \tag{14}$$

where F is distribution and g is some non-negative function. The most widely known of the (non-uniform) Burr distributions is the Burr XII distribution, frequently just called the Burr distribution. The density is

$$f(x) = \frac{\alpha\lambda^{\alpha}\theta x^{\theta-1}}{[\lambda + x^{\theta}]^{\alpha+1}} \text{ for } x > 0,\ \alpha > 0,\ \lambda > 0,\ \theta > 0. \qquad (15)$$

The *k*-th raw moment

$$m_k = \frac{1}{\Gamma(\alpha)}\lambda^{\frac{k}{\theta}}\Gamma\left(1+\frac{k}{\theta}\right)\Gamma\left(1-\frac{k}{\theta}\right). \qquad (16)$$

exists only for $k < \alpha\theta$.The maximum likelihood and method of moments estimators for the Burr distribution can only be evaluated numerically.

Experience has shown that the Pareto formula is often an appropriate model for the claim size distribution, particularly where exceptionally large claims may occur. However, there is sometimes a need to find heavy tailed distributions which offer greater flexibility than the Pareto law, including a non-monotone pdf. Such flexibility is provided by the Burr distribution and its additional shape parameter $\theta > 0$. If Y has the Pareto distribution, then the distribution of $Y = X^{1/\theta}$ is known as the Burr distribution.

2.5. Weibull distribution

If V is an exponential variable, then the distribution of $X = V^{1/\tau}, \tau > 0$, is called the Weibull (or Frechet) distribution. The density is given by:

$$f(x) = \tau\beta x^{\tau-1}e^{-\beta x^{\tau}}, x > 0. \qquad (17)$$

The *k*-th raw moment can be shown to be

$$m_k = \lambda^{-\frac{k}{\tau}}\Gamma\left(1+\frac{k}{\tau}\right). \qquad (18)$$

As in the Burr distribution, the maximum likelihood and method of moments estimators can only be evaluated numerically.

3. Maximum likelihood method

To use the method of maximum likelihood, one first specifies the joint density function for all observations. For an independent and identically distributed sample, this joint density function is

$$L = \prod_{i=1}^{n} f(x_i|\theta). \qquad (19)$$

The maximum likelihood estimate (MLE) of θ is the value of θ that maximises (19) it is the value that makes the observed data the "most probable". Rather than maximising this product which can be quite tedious, one often uses the fact that the logarithm is an increasing function so it will be equivalent to maximise the log likelihood. As the sample size increases to infinity, sequences of maximum-likelihood estimators have the following properties [Fisz 1969]: consistency, asymptotic normality, efficiency, it achieves the Cramér–Rao lower bound when the sample size tends to infinity.

4. Finite mixture models

Suppose X to be a non-negative random variable with continuous distribution. The density function f is given as a weighted average of K component densities $f_j(x|\theta)$ with mixing proportions π_j [Jajuga 1990]

$$f(x|\theta) = \sum_{j=1}^{K} \pi_j f_j(x|\theta_j), \tag{20}$$

where $\pi_j \geq 0,\ j = 1,2,...,K\ \sum_{j=1}^{K} \pi_j = 1$.

and component densities $f_j(x|\theta_j)$ depend on p-dimensional (in general unknown) vector parameters θ_j. For the estimation of unknown parameters (from a random sample x_i, $i = 1,\ldots,n$) the maximum likelihood estimation is usually used in order to obtain the estimate of the parameter. From (19) it follows that the likelihood function is equal to

$$l(x_1, x_2, ..., x_n|\theta, \pi) = \prod_{i=1}^{n} \left[\sum_{j=1}^{K} \pi_j f_j(x|\theta_j) \right]. \tag{21}$$

Now, we want to maximize the complete data log likelihood

$$\sum_{i=1}^{n} \ln f(x_i|\theta, \pi) \tag{22}$$

with respect to $\sum_{j=1}^{K} \pi_j = 1$. For this, we actually need to introduce a Lagrange multiplier. This gives us the likelihood function of:

$$L_0 = \sum_{i=1}^{n} \ln f(x_i | \theta, \pi) - \lambda \left[\sum_{j=1}^{K} \pi_j - 1 \right]. \quad (23)$$

First, we have to find the partial first derivatives of L_0 and set them equal to zero.

The task of maximizing the likelihood function can be solved using the EM algorithm. This is a numeric procedure that consists of two steps. The first step is called Expectation (probabilities π_j are estimated) and the second one Maximization, where estimated values from the first step are used in order to establish new approximations of parameters θ. These two steps are repeated until a solution is found. Generally, the EM algorithm does not guarantee the absolute maximum of the logarithmic likelihood function, but only the local extreme [Titterington, Smith, Makov 1985]. In the model for complete data associated with the model, each random vector $C_i = (X_i; Z_i)$, where $Z_i = (Z_{ij}, j = 1,2,...K)$ and $Z_{ij} \in \{0,1\}$ is a Bernoulli random variable indicating that individual i comes from component j. Since each individual comes from exactly one component, this implies $\sum_{j=1}^{K} Z_{ij} = 1$ and $P(Z_{ij} = 1) = \pi_j$. The complete-data density for one observation is thus

$$h_\theta(c_i) = h_\theta(x_i, z_i) = \sum_{j=1}^{K} I_{z_{ij}} \pi_j f_j(x_i). \quad (24)$$

Instead of the observed log-likelihood, the EM algorithm iteratively maximizes the operator

$$Q(\theta | \theta^{(t)}) = E\left[\log h_\theta(C) | x, \theta^{(t)})\right]. \quad (25)$$

where $\theta^{(t)}$ is the current value θ at iteration t, and the expectation is with respect to the distribution $k_\theta(c \mid x)$ of c given x, for the value $\theta^{(t)}$ of the parameter.

E-step: compute $Q(\theta | \theta^{(t)})$

M-step: set $\theta^{(t+1)} = \arg\max_{\theta \in \psi} Q(\theta | \theta^{(t)})$.

E-step: Calculate the "posterior" probabilities (conditional on the data and $\theta^{(t)}$) of component inclusion,

$$p_{ij}^{(t)} \overset{def}{=} P_{\theta^{(t)}}(Z_{ij}=1|x_i) = \frac{\pi_j^{(t)} f_j^{(t)}(x_i)}{\sum_{j'=1}^{K} \pi_{j'}^{(t)} f_{j'}^{(t)}(x_i)} \tag{26}$$

for $i = 1,2,...,n$ and $j = 1,2,...,K$.

Numerically, it can be dangerous to implement Equation 26 exactly as written due to the possibility of the indeterminate form 0/0 in cases where x_i is so far from any of the components that all $f_{j'}^{(t)}(x_i)$ values result in a numerical underflow to zero. Thus, many of the routines in mix-tools [Benaglia et al. 2009] actually use the equivalent expression

$$p_{ij}^{(t)} = \left[1 + \sum_{j' \neq j}^{K} \frac{\pi_{j'}^{(t)} f_{j'}^{(t)}(x_i)}{\pi_j^{(t)} f_j^{(t)}(x_i)}\right]^{-1}. \tag{27}$$

M-step for π

$$\pi_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} p_{ij}^{(t)} \quad \text{for} \quad j = 1,2,...,K. \tag{28}$$

5. Fitting loss distributions

Fitting distributions to data is a common task in statistics and consists in choosing a probability distribution modeling the random variable, as well as finding the parameter estimates for that distribution. The fitdistr function estimates distribution parameters by maximizing the likelihood function. For the log-Normal, exponential and Poisson distributions fitdistr function used the closed-form MLEs. For all other distributions, direct optimization of the log-likelihood is performed using optim. The estimated standard errors are taken from the observed information matrix, calculated by a numerical approximation. For one-dimensional problems the Nelder-Mead method is used, and for multi-dimensional problems the BFGS method. The default method is an implementation of that of [Nelder, Mead 1965], that uses only function values and is robust but relatively slow. It will work reasonably well for non-differentiable functions. BFGS is a quasi-Newton method (also known as a variable metric algorithm), specifically as published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno. This uses function values and gradients to build up a picture of the surface to be optimized.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 15(21)

We conducted the analysis for two datasets. The first is the Danish fire losses dataset. These data describe large fire insurance claims in Denmark from Thursday 3rd January 1980 to Monday 31st December 1990. The data are contained in a numeric vector. The dates of each observation are contained in a times attribute. The second data (AutoBi) are from the Insurance Research Council (IRC), a division of the American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America. The data, collected in 2002, contains information on demographic information about the claimant, attorney involvement and the economic loss (LOSS, in thousands), among other variables. We consider here a sample of n = 1340 losses from a single state. Computations of goodness-of-fit can be judged using the Akaike information criterion.

$$\mathrm{AIC} = -2*l(\boldsymbol{\psi}) + 2*k. \tag{29}$$

where $l(\boldsymbol{\psi})$ – log-likelihood function, k-number of parameters.

If different models are compared, the smaller the value of AIC, the better fit. When fitting continuous distributions, two goodness-of-fit statistics: Cramer-von Mises and Kolmogorov-Smirnov are classically considered.

On the basis of the table above, we conclude that the best quality of the fit to the empirical data was obtained in the case of distribution Burr XII, and the worst for the Weibull and gamma distribution.

The more points that are close to the line, the better fit the distribution. The pp-plot reveals that only the Burr distribution sufficiently fits the data.

We repeat the previously described procedures for data AutoBi.

Table 1. Goodness-of-fit statistics. Data: Danish

Theoretical distribution	AIC	Kolmogorov-Smirnov statistic	Cramer-von Mises statistic
Lognormal	8119.795	0.137	14.791
Gamma	9538.921	0.201	37.075
Burr XII	**6859.522**	**0.026**	**0.294**
Pareto	9249.666	0.312	37.717
Weibull	9611.243	0.273	36.261

Source: own calculations.

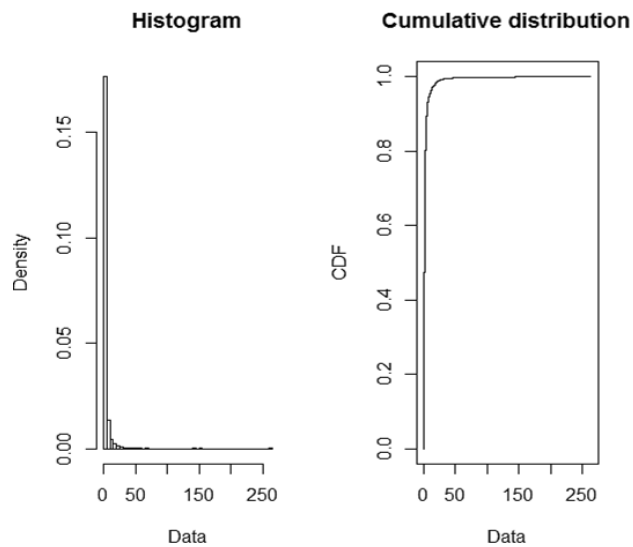

Fig. 1. Histogram and empirical cumulative distribution of Danish Loss
Source: own elaboration.

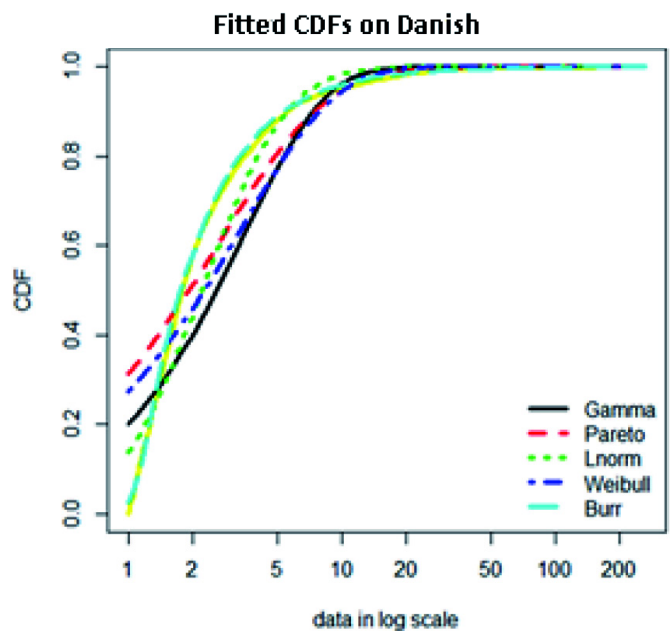

Fig. 2. Fitted CDFs on Danish. (yellow color: empirical cumulative distribution of Danish Loss)

Source: own elaboration.

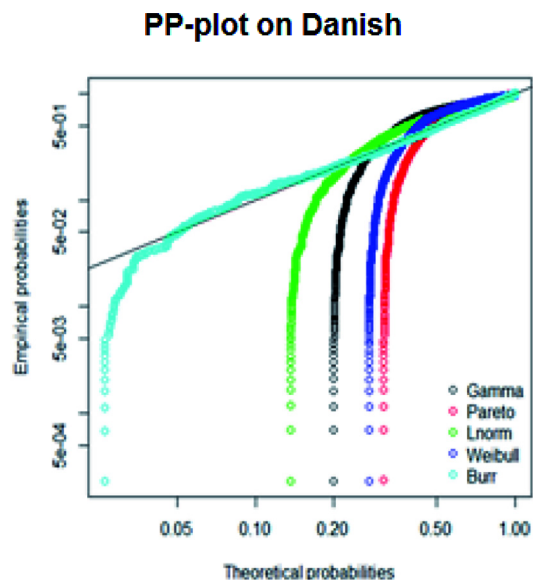

Fig. 3. Theoretical probabilities

Source: own elaboration.

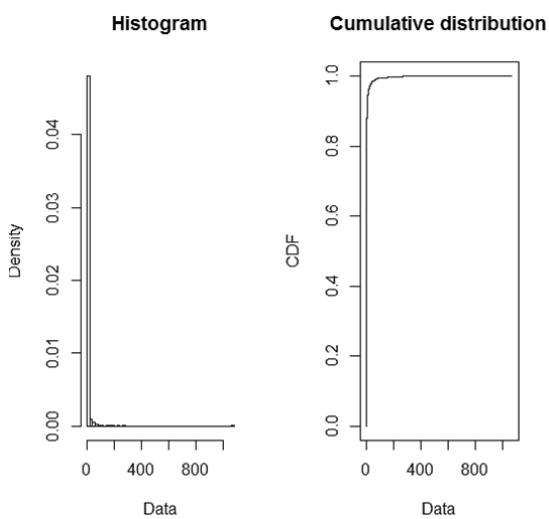

Fig. 4. Histogram and empirical cumulative distribution of AutoBi Loss

Source: own elaboration.

Table 2. Goodness-of-fit statistics. Data: AutoBi

Theoretical distribution	AIC	Kolmogorov-Smirnov statistic	Cramer-von Mises statistic
Lognormal	6345.768	0.091	3.017
Gamma	6942.452	0.187	11.404
Burr XII	**6292.309**	**0.066**	**1.669**
Pareto	6295.842	0.067	1.687
Weibull	6592.228	0.114	4.551

Source: own calculations.

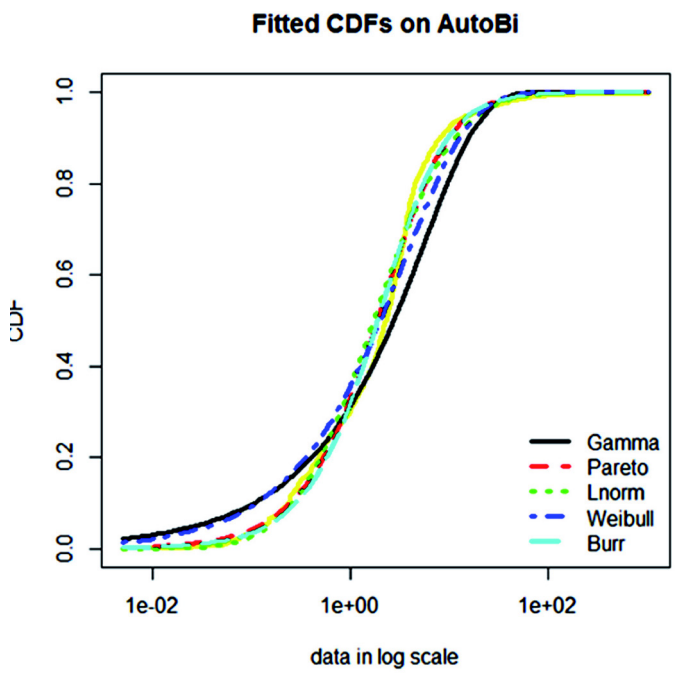

Fig. 5. Fitted CDFs on AutoBi (yellow color: empirical cumulative distribution of Danish Loss)

Source: own elaboration.

On the basis of Table 2, we conclude that the best quality of the fit to the empirical data obtained in the case of distribution Burr XII and the worst for the gamma distribution.

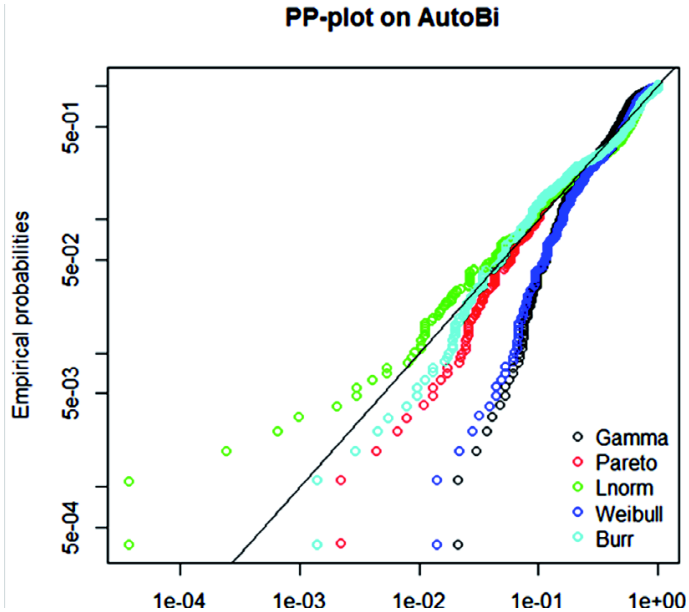

Fig. 6. Theoretical probabilities

Source: own elaboration.

6. Modeling the loss distributions using mixture densities

We first tried to use a mixture of gamma distribution and a Pareto distribution to fit the Danish Loss distribution [Charpentier 2014]. This example of more interest in actuarial science is the mixture of a light-tailed and heavy-tailed claim distribution. Let us take, for example, the mixture of a gamma distribution $G(\nu,\lambda)$ and a Pareto distribution $P(\alpha,\theta)$. The density is given by

$$f(x)=p\frac{\lambda^{\nu}x^{\nu-1}e^{-\lambda x}}{\Gamma(\nu)}+(1-p)\frac{\alpha}{\theta}\left(\frac{\theta}{\theta+x}\right)^{\theta+1},$$

where $p\in(0,1), x\in R_{+}$.

When fitted alone, the parameters of the gamma distribution are estimated as $(\hat{\alpha},\hat{\lambda})=(1.298;0.383)$ and the parameters of the Pareto [Goulet 2016] distribution are estimated as $(\hat{\alpha},\hat{\theta})=(5.368;13.841)$. When used in the mixture, we get applying function optim in R package

$$(\hat{p},\hat{\alpha}_1,\hat{\theta}_1,\hat{\alpha}_2,\hat{\lambda}_2)=(0.684;\ 10.871;6.544;5.418;\ 30.07).$$

As only the shape parameter $\hat{\alpha}_2$ is of similar amplitude, only heavy-tailed distributions (like the Pareto) are appropriate for this dataset. The resulting goodness-of-fit is weak. (see Table 3). Optimization is based on quasi-Newton algorithms.

The parameters of the mixture of gamma distribution is estimated using the function gammmixEM. The function implements the EM algorithm in mix-tools package. Set the following parameter estimates

$$\pi_1 = 0.41 \quad \pi_2 = 0.59, \alpha_1 = 12.823, \quad \alpha_2 = 1.487, \beta_1 = 0.025, \beta_2 = 0.001.$$

The AIC criterion values lead to the conclusion that the quality of fit in the case of mixtures (AIC = 8972.02) is better than in the case of a single distribution (AIC = 9538.921).

Table 3. Goodness-of-fit statistics using mixture of distributions. Data: Danish

Theoretical distribution	AIC	Kolmogorov-Smirnov statistic	Cramer-von Mises statistic
Mixture Gamma-Pareto	7493.181	0.102	2.979
Mixture Gamma	7743.327	0.119	9.735
Burr XII	**6859.522**	**0.026**	**0.294**
Mixture Burr XII	**6758.085**	**0.053**	**1.470**
Mixture lognormal	7152.251	0.052	0.856

Source: own calculations.

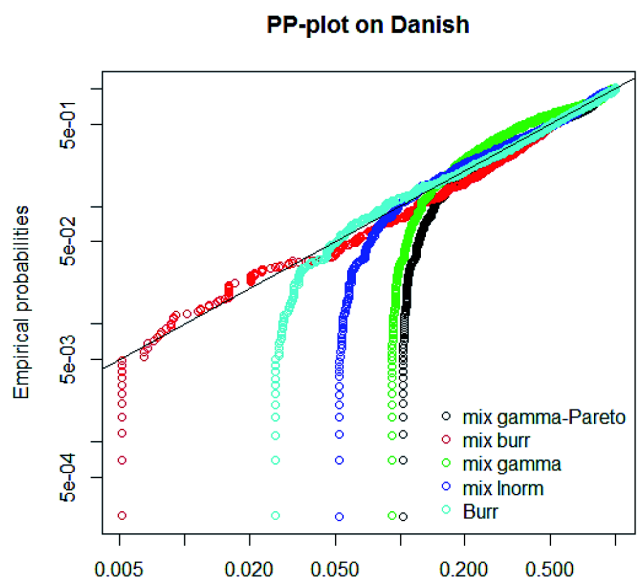

Fig. 7. Theoretical probabilities

Source: own elaboration.

ŚLĄSKI PRZEGLĄD STATYSTYCZNY

Nr 15(21)

We repeat the previously described procedures for data AutoBi. The obtained results are shown in Table 4.

Table 4. Goodness-of-fit statistics using mixture of distributions Date: AutoBi

Theoretical distribution	AIC	Kolmogorov-Smirnov statistic	Cramer-von Mises statistic
Mixture Gamma-Pareto	6161.064	0.027	0.124
Mixture Gamma	6261.004	0.052	1.028
Burr XII	6292.309	0.066	1.669
Mixture Burr XII	6156.531	0.030	0.173
Mixture lognormal	**6145.624**	**0.021**	**0.071**

Source: own calculations.

The parameters of the mixture of gamma distribution is estimated using the function gammmixEM. The function implements the EM algorithm in a mix-tools package. Set the following parameter estimates

$\pi_1 = 0.82\ \pi_2 = 0.18, \alpha_1 = 1.73\ \alpha_2 = 0.11, \beta_1 = 1.11, \beta_2 = 233.23$.

The AIC criterion values lead to the conclusion that the quality of fit in the case of mixtures (AIC = 6739.23) is better than in the case of a single distribution (AIC = 6942.452).

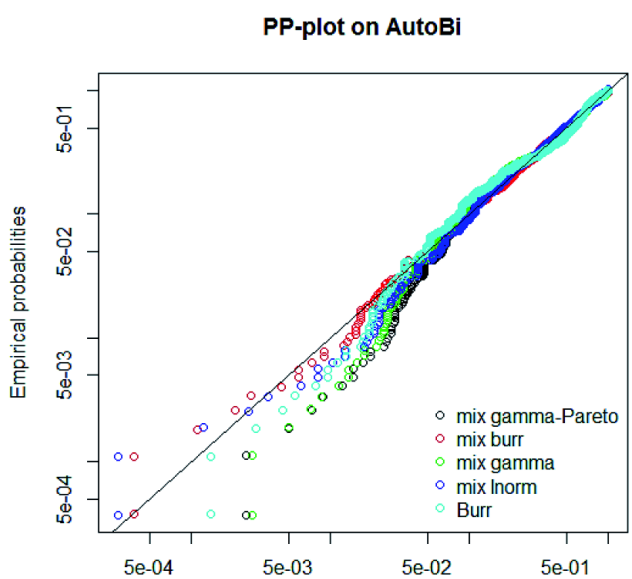

Fig. 8. Theoretical probabilities

Source: own elaboration.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 15(21)

7. Conclusions

In this paper we propose a mixture of densities for modeling the loss distribution to deal with data displaying large and positive skewness as well as a wide right tail. To see how different distributions fit the data, we compare the log likelihood from the fitdistr function and EM algorithm.

The Burr XII distribution fits best the data in both the considered examples of the empirical data. A mixture of distributions Burr XII and mixtures log-normal distributions gave a comparable or better fit to the empirical data than the distribution of Burr XII.

Mixtures of gamma distributions, whose parameters were estimated using the Newton-Raphson algorithm, better fit the empirical data than mixtures of gamma distributions whose parameters were estimated using the EM algorithm.

References

Benaglia T., Chauveau D., Hunter D.R., Young D.S., 2009, *An R package for analyzing finite mixture models*, Journal of Statistical Software.

Bernardi M., Maruotti A., Lea P., 2012, *Skew Mixture Models for Loss Distributions: A Bayesian Approach,* Munich Personal RePEc Archive.

Charpentier A., 2014, *Computational Actuarial Science with R*, CRC Press.

Embrechts P., Klüppelberg C., Mikosch, T., 1997, *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, New York.

Fisz M., 1969, *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa.

Goulet V., Auclair S., Dutang C., Milhaud X., Ouellet T., Pouliot L., Pigeon M., 2016*, Package actuary.*

Jajuga K., 1990, *Statystyczna teoria rozpoznawania obrazów*, PWN.

Klugman S.A., Panjer H.H., Willmot G.E*., Loss Models: From Data to Decisions*, John Wiley & Sons, New York.

McNeil A.J., 1997, *Estimating the tails of loss severity distributions using extreme value theory*, Austin Bulletin, 27, pp.117–137.

Nelder J.A., Mead R., 1965, *A simplex algorithm for function minimization*, Computer Journal, 7, pp. 308–313.

Titterington D.M., Smith A.F., Makov U.E., 1985, *Statistical analysis of finite mixture distributions*, Wiley.

ŚLĄSKI
PRZEGLĄD
STATYSTYCZNY

Nr 15(21)

MODELOWANIE STRAT ZA POMOCĄ MIESZANEK ROZKŁADÓW W ŚRODOWISKU R

Streszczenie: Skończone mieszanki rozkładów można wykorzystać do modelowania rozkładów szkód w ubezpieczeniach. Rozkłady te najczęściej mają grube ogony i prawostronną asymetrię. Dopasowanie rozkładu teoretycznego do empirycznego rozkładu strat jest z reguły zadaniem trudnym. W artykule pokazano, że zastosowanie mieszanek rozkładów polepsza jakość dopasowania rozkładów teoretycznych do empirycznych rozkładów strat. Ponadto przedstawiono algorytm służący do estymacji parametrów mieszanek rozkładów i przykłady zastosowań mieszanek rozkładów. Podejście analityczne jest prawdopodobnie najczęściej stosowane w praktyce i na pewno najczęściej znajduje się w literaturze aktuarialnej. Dla oszacowania parametrów używamy metody największej wiarygodności, stosując algorytmy Newtona-Raphsona i EM. Jakość dopasowania porównujemy, stosując kryterium AIC.

Słowa kluczowe: skończone mieszanki rozkładów, rozkłady strat, metoda największej wiarygodności, algorytm EM.