TWO LOGNORMAL MODELS FOR REAL DATA

Raluca Vernic, Sandra Teodorescu and Elena Pelican

Abstract

Based on the statistical analysis of a data sample from property insurance, in this paper we consider two lognormal mixture models that we fitted to our data sample. The first model is a usual two components lognormal mixture for which we used the EM algorithm to estimate the parameters. The second one, called a composite lognormal-lognormal model, can in fact be reduced to a particular two components mixture model having truncated lognormals as mixture distributions. This composite model is studied in some detail and we present some specific parameters estimation methods. We also discuss and compare both models fit to the data.

1 Introduction

This paper is motivated by the statistical study of a real data set consisting of claims from property insurance, to which we tried to fit an adequate probability distribution. Property insurance provides protection against most property risks, such as fire, theft and some water damage, including also specialized forms of insurance like flood insurance, earthquake insurance or home insurance. On the general insurance scale, our data set comes from *fire and allied perils insurance* (i.e. property insurance against fire, earthquake, flood etc.).

Between the classical theoretical distributions, the lognormal and Pareto ones are frequently used to model property loss amounts. For example, fire

Key Words: distribution, insurance.
Mathematics Subject Classification: 60E05, 91B30.
Received: April 2009
Accepted: October 2009

insurance data were previously studied among others by Shpilberg (1997), McNeil (1997) or Resnick (1977), who fitted several distributions to such data (like truncated lognormal, ordinary Pareto or generalized Pareto), and, more recently, by Cooray and Ananda (2005) and Scollnik (2007), who applied several composite lognormal-Pareto models.

In this paper we consider two lognormal models to model the claims distribution of a data set from fire and allied perils insurance. The data were kindly provided by a Romanian insurance company and consist of the entire property loss amount settled during year 2007 for its own portfolio, more precisely of $n = 1039$ settled claims.

Since the histogram of the log-data shows a bi-modality, and the best fit among the classical distributions was given by the lognormal, we decided to consider two lognormal mixture models for these data. The first model, presented in Section 2, is a usual two components lognormal mixture for which we used the EM (Expectation-Maximization) algorithm to estimate the parameters. The second one, called a composite lognormal-lognormal model, can in fact be reduced to a particular two components mixture model having truncated lognormals as mixture distributions. This composite model is studied in some detail in Section 3.

In order to simplify our study, we used the relation between the normal and the lognormal distributions, and worked mainly on the corresponding normal mixture models. This is why we start Section 3 by first presenting some properties of the truncated normal distribution (Section 3.1). Then the composite normal-normal model is derived in Section 3.2, where we also present the relation with the composite lognormal-lognormal model and some specific parameters estimation methods. In Section 3.3, we also discuss and compare the fitting of both models to the data.

In the following, we will denote by φ the standard normal $N(0,1)$ density and by Φ its cumulative distribution function (cdf). We use $N\left(\mu,\sigma^2\right)$ to denote the univariate normal distribution with parameters $\mu \in \mathbb{R}, \sigma > 0$, while $\varphi(\cdot;\mu,\sigma)$ denotes its density, i.e.

$$\varphi(x;\mu,\sigma) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ x \in \mathbb{R}.$$

2 A two components mixture

Based on the histogram of the log-data from Figure 1 that shows a bi-modality, we decided to try to fit to the log-data a two components normal mixture. Therefore, if X denotes the random variable (r.v.) under study and $Y = \ln X$, then we assume that

$$Y = IY_1 + (1 - I)\, Y_2,$$

where $Y_i \sim N\left(\mu_i, \sigma_i^2\right), i = 1, 2$, and I is a Bernoulli r.v. taking the values 1 and 0 with probabilities $r \in [0, 1]$ and $1 - r$, respectively. This mixture model is explicit: generate a value of I and then, depending on the outcome, deliver either Y_1 or Y_2. The density of Y is

$$f_Y(x) = r\varphi(x;\mu_1, \sigma_1) + (1 - r)\, \varphi(x;\mu_2, \sigma_2)\,,\ x \in \mathbb{R}, \tag{1}$$

while a straightforward calculation gives the density of X as

$$f_X(x) = r\frac{1}{x}\varphi(\ln x;\mu_1, \sigma_1) + (1 - r)\, \frac{1}{x}\varphi(\ln x;\mu_2, \sigma_2)\,,\ x > 0.$$

Hence, the density of the r.v. X under study results also as a two components mixture, but the mixing densities are lognormals.

In order to fit this model to a data sample, we must first estimate the parameters. This can be done by the EM algorithm.

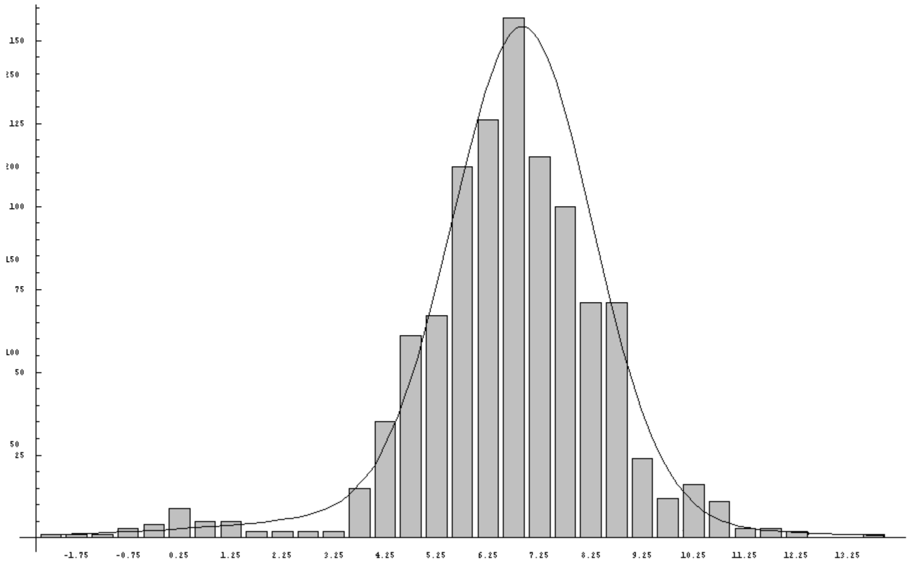

Figure 1: Histogram of the log-data with estimated normal curve

2.1 The EM algorithm

The EM algorithm is a popular tool for simplifying difficult maximum likelihood problems, see e.g. Dempster et al. (1977). This algorithm was originally designed for mixtures of normal distributions, and therefore it can be used when we notice that the distribution graph (e.g. histogram) presents a multimodality. Then the number of modes should give an idea on the number of mixed distributions. In the following, we will describe this algorithm for a two components mixture of normal distributions.

Let $(y_1, ..., y_n)$ be a data sample from the random variable Y. From Dempster et al. (1977), we have the following EM algorithm used to estimate the parameters $r, \mu_1, \sigma_1, \mu_2, \sigma_2$, for the two components normal mixture (1):

1. Take initial guesses for the parameters, e.g. $\tilde{r} = 0.5$, $\tilde{\mu}_1, \tilde{\mu}_2$ taken at random from the observed data, and $\tilde{\sigma}_1 = \tilde{\sigma}_2$ equal to the overall sample standard deviation.

2. *Expectation step*: compute the "responsibilities",

$$\tilde{\gamma}_i = \frac{\tilde{r}\,\varphi\left(y_i;\tilde{\mu}_1, \tilde{\sigma}_1\right)}{\tilde{r}\,\varphi\left(y_i;\tilde{\mu}_1, \tilde{\sigma}_1\right) + \left(1-\tilde{r}\right)\varphi\left(y_i;\tilde{\mu}_2, \tilde{\sigma}_2\right)},\ i = 1, ..., n.$$

3. *Maximization step*: compute the weighted means and variances,

$$\begin{aligned} \tilde{\mu}_1 &= \frac{\sum_{i=1}^n \tilde{\gamma}_i y_i}{\sum_{i=1}^n \tilde{\gamma}_i},\ \tilde{\sigma}_1^2 = \frac{\sum_{i=1}^n \tilde{\gamma}_i \left(y_i - \tilde{\mu}_1\right)^2}{\sum_{i=1}^n \tilde{\gamma}_i}, \\ \tilde{\mu}_2 &= \frac{\sum_{i=1}^n \left(1-\tilde{\gamma}_i\right) y_i}{\sum_{i=1}^n \left(1-\tilde{\gamma}_i\right)},\ \tilde{\sigma}_2^2 = \frac{\sum_{i=1}^n \left(1-\tilde{\gamma}_i\right)\left(y_i - \tilde{\mu}_2\right)^2}{\sum_{i=1}^n \left(1-\tilde{\gamma}_i\right)}, \end{aligned}$$

   and the mixing probability $\tilde{r} = \frac{1}{n}\sum_{i=1}^n \tilde{\gamma}_i$.

4. Iterate Steps 2 and 3 until convergence.

2.2 Numerical results

The main characteristics of our data sample are $n = 1039$,

Minimum	Maximum	Mean	Variance	Standard Deviation
0.12	746261.11	4297.1759	693372009.4	26331.95795

Also, for the log-data we have

Minimum	Maximum	Mean	Variance	Standard Deviation
-2.12026	13.52283	6.64286	3.61212	1.90056

We applied the EM algorithm described before for the log-data, starting with different initial values. More precisely, we varied the initial guesses for σ_1, σ_2 and r, while μ_1 and μ_2 where fixed, equal to the two modes. Each time, the EM algorithm (that we implemented in TurboPascal) gave one the following two solutions:

	$\tilde{\mu}_1$	$\tilde{\mu}_2$	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$	$\tilde{r}$	Kolmogorov dist.	LogLikelihood
Sol. I	0.188	6.839	0.903	2.381	0.029	0.036	-2044.825
Sol. II	5.404	6.814	14.120	1.910	0.121	0.024	-2052.091

In order to check the fitting to the log-data of both resulting mixture distributions, we evaluated the log-likelihood values and the Kolmogorov distances, also given in the above table. Kolmogorov's goodness-of-fit test accepts both distributions, therefore we compared the log-likelihood values, and even if the race is very close, based on this criteria the first solution seems better.

For comparison reasons, in Figure 2 we show the shapes of both normal mixtures (denoted EM I for Solution I and EM II for Solution II), together with the classical normal fitted curve and the data histogram. Note that, while the EM I curve tends to capture the small left hump, EM II fits better to the right higher hump, while the classical normal gives the poorest fit between these three distributions. Therefore, we conclude that our initial (not transformed) data are well modeled by a two components lognormal mixture. However, in next section we will introduce another model that fits these data.

3 A composite Normal-Normal model

3.1 The truncated Normal distribution

A r.v. Y has a *doubly truncated* normal distribution with parameters $\mu \in \mathbb{R}, \sigma > 0$, if its density function is

$$f_Y(x) = \frac{\varphi(x;\mu,\sigma)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \ a \leq x \leq b. \tag{2}$$

Here, a and b are the *lower* and *upper truncation points*, respectively, while $\Phi\left(\frac{a-\mu}{\sigma}\right)$ and $1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$ are the *degrees of truncation* from *below* and, respectively, from *above.* If a is replaced by $-\infty$, or b by ∞, the distribution is *singly truncated* from *above (right)*, or *below (left)*, respectively. The case $a = \mu, b = \infty$ produces the *half-normal* distribution. For details on the truncated normal distribution see e.g. Johnson et al. (1994).

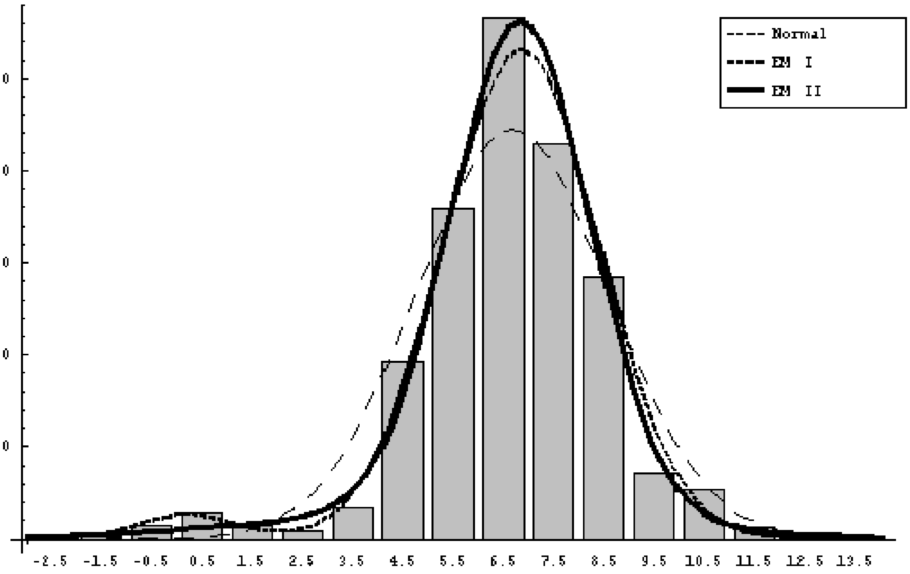

Figure 2: Fitted Normal, EM I and EM II curves to the log-data

The moment generating function (mgf) of the doubly truncated normal distribution is

$$M_Y(t)=\frac{\Phi\left(\frac{b-\mu}{\sigma}-\sigma t\right)-\Phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}e^{\mu t+\sigma^2t^2/2}, \tag{3}$$

from where the mean results as

$$\mathbb{E}Y=\mu-\sigma\frac{\varphi\left(\frac{b-\mu}{\sigma}\right)-\varphi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)},$$

while

$$\mathbb{E}Y^2=\mu^2+\sigma^2+\sigma^2\frac{\varphi'\left(\frac{b-\mu}{\sigma}\right)-\varphi'\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}-2\mu\sigma\frac{\varphi\left(\frac{b-\mu}{\sigma}\right)-\varphi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)},$$

with φ' denoting the first derivative of φ. We will now separately consider singly truncated distributions.

3.1.1 Upper (right) truncated Normal distribution

As mentioned before, this distribution is obtained by taking in (2) $a = -\infty$. Let us denote by Y_1 a r.v. having this distribution. Then its density is

$$f_{Y_1}(x) = \frac{\varphi(x;\mu,\sigma)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}, \ -\infty < x \leq b, \tag{4}$$

while the mgf (3) reduces to

$$M_{Y_1}(t) = \frac{\Phi\left(\frac{b-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)} e^{\mu t + \sigma^2 t^2/2}. \tag{5}$$

For this distribution, the first two moments are

$$\mathbb{E}Y_1 = \mu - \sigma \frac{\varphi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}, \ \mathbb{E}Y_1^2 = \mu^2 + \sigma^2 - \sigma(b+\mu)\frac{\varphi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)}.$$

3.1.2 Lower (left) truncated Normal distribution

This distribution results by taking $b = \infty$ in (2). Denoting by Y_2 a r.v. with this distribution, its density becomes

$$f_{Y_2}(x) = \frac{\varphi(x;\mu,\sigma)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \ a \leq x < \infty. \tag{6}$$

Its mgf results from (3) as

$$M_{Y_2}(t) = \frac{1 - \Phi\left(\frac{a-\mu}{\sigma} - \sigma t\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)} e^{\mu t + \sigma^2 t^2/2}, \tag{7}$$

and the first two moments are

$$\mathbb{E}Y_2 = \mu + \sigma \frac{\varphi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \ \mathbb{E}Y_2^2 = \mu^2 + \sigma^2 + \sigma(a+\mu)\frac{\varphi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}.$$

3.2 The composite Normal-Normal model

3.2.1 The general model

To model statistical data coming from two different distributions, Cooray and Ananda (2005) introduced a composite lognormal-Pareto model, that was further developed by Scollnik (2007). Furthermore, Teodorescu and Vernic (2009) studied a general composite model, whose density is defined as

$$f(x)=\begin{cases} rf_1^*(x), & -\infty<x\leq\theta \\ (1-r)f_2^*(x), & \theta<x<\infty \end{cases}, \tag{8}$$

where $0\leq r\leq 1$, while f_1^* and f_2^* are singly truncated densities from above and, respectively, below, with truncation point in θ. This density function can be also interpreted as a two component mixture model with mixing weights r and $1-r$, i.e.

$$f(x)=rf_1^*(x)+(1-r)f_2^*(x). \tag{9}$$

In the following, we consider the particular case when f_1^* is the right truncated normal density, while f_2^* is the left truncated normal density. We call this a *composite normal-normal density.* Notice that even if the density (9) seems very similar with (1), the main difference is that in (9) the densities involved in the right hand side are truncated, while in (1) they are not. So the two mixture models are basically different.

Let Y denote a r.v. having a composite normal-normal distribution. We will now apply some properties deduced by Teodorescu and Vernic (2009), to the particular case of this composite model. We start with its density, and impose a continuity condition at θ. Then we obtain the following result.

Proposition 1. The composite normal-normal density function is given by

$$f(x)=\begin{cases} r\dfrac{\varphi(x;\mu_1,\sigma_1)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}, & -\infty<x\leq\theta \\ (1-r)\dfrac{\varphi(x;\mu_2,\sigma_2)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}, & \theta<x<\infty \end{cases}, \tag{10}$$

where the parameters $\mu_1,\mu_2\in\mathbb{R},\sigma_1,\sigma_2>0$ and $0\leq r\leq 1$ satisfy the continuity condition

$$r=\frac{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)\varphi(\theta;\mu_2,\sigma_2)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)\varphi(\theta;\mu_2,\sigma_2)+\left(1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)\right)\varphi(\theta;\mu_1,\sigma_1)}. \tag{11}$$

Proof. The density function (10) results immediately by inserting (4) and (6) in (8). Then the continuity condition at the threshold $\theta, f(\theta-0)=f(\theta+0)$, easily gives the formula (11) of r.□

Remark 1. In order to obtain a smooth density, we usually impose a differentiability condition at θ, i.e. $f'(\theta-0)=f'(\theta+0)$, that gives

$$r=\frac{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)\varphi'(\theta;\mu_2,\sigma_2)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)\varphi'(\theta;\mu_2,\sigma_2)+\left(1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)\right)\varphi'(\theta;\mu_1,\sigma_1)}.$$

Using $\varphi'(x;\mu,\sigma) = -\varphi(x;\mu,\sigma)\frac{x-\mu}{\sigma^2}$ in this last expression of r and combining the result with (11), we obtain the supplementary condition

$$\frac{\theta-\mu_1}{\sigma_1^2} = \frac{\theta-\mu_2}{\sigma_2^2}, \tag{12}$$

or, equivalently,

$$\theta = \frac{\sigma_1^2\mu_2 - \sigma_2^2\mu_1}{\sigma_1^2 - \sigma_2^2}.$$

Unfortunately, though this supplementary condition reduces the number of unknown parameters from 5 to 4, it is of no help for our real data. This is because from (12), it results that in this case we must necessarily have θ smaller or equal to both μ_1, μ_2, or both μ_1, μ_2 smaller or equal to θ, while our data shows that we rather have $\mu_1 \leq \theta \leq \mu_2$.

Proposition 2. The cdf of the composite normal-normal distribution is

$$F(x) = \begin{cases} r\dfrac{\Phi\left(\frac{x-\mu_1}{\sigma_1}\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}, & -\infty < x \leq \theta \\ 1 + (1-r)\dfrac{\Phi\left(\frac{x-\mu_2}{\sigma_2}\right) - 1}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}, & \theta < x < \infty \end{cases}. \tag{13}$$

Proof. From Teodorescu and Vernic (2009), we have that

$$F(x) = \begin{cases} r\dfrac{\Phi\left(\frac{x-\mu_1}{\sigma_1}\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}, & -\infty < x \leq \theta \\ r + (1-r)\dfrac{\Phi\left(\frac{x-\mu_2}{\sigma_2}\right) - \Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}, & \theta < x < \infty \end{cases},$$

which easily gives (13).□

Proposition 3. The mgf of the composite normal-normal distribution is

$$M(t) = re^{\mu_1 t+\sigma_1^2t^2/2}\frac{\Phi\left(\frac{\theta-\mu_1}{\sigma_1} - \sigma_1 t\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)} + (1-r)\,e^{\mu_2 t+\sigma_2^2t^2/2}\frac{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2} - \sigma_2 t\right)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}.$$

Proof. We have that

$$M(t) = \mathbb{E}\left(e^{tY}\right) = r\int_{-\infty}^{\theta} e^{tx}\frac{\varphi(x;\mu_1,\sigma_1)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}dx + (1-r)\int_{\theta}^{\infty} e^{tx}\frac{\varphi(x;\mu_2,\sigma_2)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}dx.$$

We recognize the two integrals to be the mgf-s of the right and, respectively, left truncated normal distributions. Using formulas (5) and (7), the result is immediate.□

Remark 2. By differentiating the mgf, or by using the expressions of the truncated distributions moments, we obtain the following formulas for the first two moments of the composite normal-normal distribution

$$\begin{aligned}\mathbb{E}Y &= r\left(\mu_1-\sigma_1\frac{\varphi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}\right)+(1-r)\left(\mu_2+\sigma_2\frac{\varphi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}\right),\\ \mathbb{E}\left(Y^2\right) &= r\left(\mu_1^2+\sigma_1^2-\sigma_1\left(\theta+\mu_1\right)\frac{\varphi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}\right)+\\ &\quad+(1-r)\left(\mu_2^2+\sigma_2^2+\sigma_2\left(\theta+\mu_2\right)\frac{\varphi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}\right).\end{aligned}$$

Proposition 4. If Y follows a composite normal-normal distribution, then $X=e^Y$ is composite lognormal-lognormal distributed.

Proof. We will first find the density of X, using its cdf and the relation with Y's cdf. For $x>0$, we have

$$F_X\left(x\right)=\Pr\left(X<x\right)=\Pr\left(e^Y<x\right)=\Pr\left(Y<\ln x\right)=F_Y\left(\ln x\right).$$

Differentiating gives

$$f_X\left(x\right)=\frac{1}{x}f_Y\left(\ln x\right),$$

and using (10), we obtain

$$f_X\left(x\right)=\begin{cases} r\dfrac{\frac{1}{x}\varphi\left(\ln x;\mu_1,\sigma_1\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)}, & 0<x\le e^\theta\\ \left(1-r\right)\dfrac{\frac{1}{x}\varphi\left(\ln x;\mu_2,\sigma_2\right)}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)}, & e^\theta<x<\infty\end{cases}.$$

We recall that $\frac{1}{x}\varphi\left(\ln x;\mu,\sigma\right), x>0$, is the density of a lognormal distribution with parameters $\mu\in\mathbb{R},\sigma>0$, while $\Phi\left(\frac{\ln x-\mu}{\sigma}\right)$ is its cdf. Note that f_X is in the form (8) with f_1^* a right truncated lognormal density with parameters μ_1,σ_1, and f_2^* a left truncated lognormal density with parameters μ_2,σ_2, while in this case the threshold parameter is $\theta'=e^\theta$. Therefore, $X=e^Y$ has indeed a composite lognormal-lognormal distribution that inherits the parameters $\mu_i,\sigma_i,i=1,2$, and r from its composite normal-normal correspondent.□

3.2.2 Statistical inference: parameters estimation

The composite normal-normal distribution has five unknown parameters μ_i, σ_i, $i = 1, 2$, and θ, while r results from (11). In Vernic and Teodorescu (2009) we presented two algorithms for estimating the parameters of a general composite distribution. We will now particularize them for our situation, then we will adapt an EM algorithm to this complex estimation problem.

An estimation method based on moments and quantiles.

For $(y_1, ..., y_n)$ a random data sample, we denote by q_2 and q_3 its second and, respectively, third empirical quartiles, by q_α its α-quantile, $0 < \alpha < 1$, by $\bar{y}$ its empirical mean, while $\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$. Note that, when writing the equations based on the empirical quantiles and the cdf F given in (13), we have several situations depending on the position of θ between q_α, q_2 and q_3. Since θ is unknown, we should consider all possible situations and see which one gives the best solution. But for our particular data, based on the histogram and on the particular values of the quartiles, we assumed that $\theta < q_2$, and that α is chosen such that $q_\alpha < \theta$. Therefore, the system we used is

$$\begin{cases} \bar{y} = \mathbb{E}Y \\ \overline{y^2} = \mathbb{E}\left(Y^2\right) \\ \alpha = r\dfrac{\Phi\left(\frac{q_\alpha-\mu_1}{\sigma_1}\right)}{\Phi\left(\frac{\theta-\mu_1}{\sigma_1}\right)} \\ 0.5 = 1 + (1-r)\dfrac{\Phi\left(\frac{q_2-\mu_2}{\sigma_2}\right)-1}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)} \\ 0.75 = 1 + (1-r)\dfrac{\Phi\left(\frac{q_3-\mu_2}{\sigma_2}\right)-1}{1-\Phi\left(\frac{\theta-\mu_2}{\sigma_2}\right)} \end{cases}.$$

Introducing $\alpha_i = (\theta - \mu_i)/\sigma_i, i = 1, 2$, we reparameterized the system. Hence, using $\sigma_i = (\theta - \mu_i)/\alpha_i$ and Remark 2, the system becomes

$$\begin{cases} \bar{y} = r\left(\mu_1 - \frac{\theta-\mu_1}{\alpha_1}\frac{\varphi(\alpha_1)}{\Phi(\alpha_1)}\right) + (1-r)\left(\mu_2 + \frac{\theta-\mu_2}{\alpha_2}\frac{\varphi(\alpha_2)}{1-\Phi(\alpha_2)}\right) \\ \overline{y^2} = r\left(\mu_1^2 + \left(\frac{\theta-\mu_1}{\alpha_1}\right)^2 - \frac{\theta^2-\mu_1^2}{\alpha_1}\frac{\varphi(\alpha_1)}{\Phi(\alpha_1)}\right) + (1-r)\times \\ \times\left(\mu_2^2 + \left(\frac{\theta-\mu_2}{\alpha_2}\right)^2 + \frac{\theta^2-\mu_2^2}{\alpha_2}\frac{\varphi(\alpha_2)}{1-\Phi(\alpha_2)}\right) \\ \alpha = \dfrac{r}{\Phi(\alpha_1)}\Phi\left(\frac{(q_\alpha-\mu_1)\alpha_1}{\theta-\mu_1}\right) \\ 0.5 = \frac{1-r}{1-\Phi(\alpha_2)}\left(1-\Phi\left(\frac{(q_2-\mu_2)\alpha_2}{\theta-\mu_2}\right)\right) \\ 0.25 = \frac{1-r}{1-\Phi(\alpha_2)}\left(1-\Phi\left(\frac{(q_3-\mu_2)\alpha_2}{\theta-\mu_2}\right)\right) \end{cases}, \qquad (14)$$

with r from (11) given by

$$r=\frac{\alpha_2\left(\theta-\mu_1\right)\varphi\left(\alpha_2\right)\Phi\left(\alpha_1\right)}{\alpha_2\left(\theta-\mu_1\right)\varphi\left(\alpha_2\right)\Phi\left(\alpha_1\right)+\alpha_1\left(\theta-\mu_2\right)\varphi\left(\alpha_1\right)\left(1-\Phi\left(\alpha_2\right)\right)}. \tag{15}$$

Since it involves the cdf Φ of the standard normal distribution, this system must be solved by numerically methods. We denote the resulting solution by $\breve{\mu}_i, \breve{\sigma}_i, i=1,2$, and $\breve{\theta}$.

Remark 3. Initially, instead of q_α we wanted to use the first empirical quartile q_1, but from the data we have that $\theta < q_1$, and the resulting equation is

$$0.75=\frac{1-r}{1-\Phi\left(\alpha_2\right)}\left(1-\Phi\left(\frac{\left(q_1-\mu_2\right)\alpha_2}{\theta-\mu_2}\right)\right).$$

Unfortunately, this equation is of no help because considered together with the last two equations in system (14) leads to an underdetermined subsystem, that cannot be solved uniquely or cannot be solved at all. This is why we had to look for a quantile such that $q_\alpha < \theta$.

An algorithm based on the method of maximum likelihood (ML). Without loss of generality, we assume that the data sample is ordered, i.e. $y_1 \leq ... \leq y_n$. In order to apply the ML method, we must know the integer value m such that the unknown parameter θ is in between the m-th and $(m+1)$-th observations, i.e. $x_m \leq \theta < x_{m+1}$. Assuming that somehow we know this m, using again the notation α_i and (15), the likelihood function is

$$L\left(y_1,...,y_n\right)=\left(\frac{r}{\Phi\left(\alpha_1\right)}\right)^m\left(\frac{1-r}{1-\Phi\left(\alpha_2\right)}\right)^{n-m}\prod_{i=1}^{m}\varphi\left(y_i;\mu_1,\sigma_1\right)$$

$$\prod_{j=m+1}^{n}\varphi\left(y_j;\mu_2,\sigma_2\right)=$$

$$=\left(\frac{\alpha_1\alpha_2}{\alpha_2\left(\theta-\mu_1\right)\varphi\left(\alpha_2\right)\Phi\left(\alpha_1\right)+\alpha_1\left(\theta-\mu_2\right)\varphi\left(\alpha_1\right)\left(1-\Phi\left(\alpha_2\right)\right)}\right)^n\times$$
$$\varphi^{n-m}\left(\alpha_1\right)\varphi^m\left(\alpha_2\right)\prod_{i=1}^{m}\varphi\left(\frac{\left(y_i-\mu_1\right)\alpha_1}{\theta-\mu_1}\right)\prod_{j=m+1}^{n}\varphi\left(\frac{\left(y_j-\mu_2\right)\alpha_2}{\theta-\mu_2}\right).$$

Denoting by

$$u\left(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta\right)=\alpha_2\left(\theta-\mu_1\right)\varphi\left(\alpha_2\right)\Phi\left(\alpha_1\right)+\alpha_1\left(\theta-\mu_2\right)\varphi\left(\alpha_1\right)\left(1-\Phi\left(\alpha_2\right)\right),$$

the likelihood system becomes

$$\left\{\begin{array}{l}
(1^\circ)\; 0 = \frac{\partial \ln L}{\partial \mu_1} = \frac{n\alpha_2\varphi(\alpha_2)\Phi(\alpha_1)}{u(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta)} - \frac{\alpha_1^2}{(\theta-\mu_1)^3}\sum_{i=1}^{m}\left(y_i-\mu_1\right)\left(y_i-\theta\right)\\
(2^\circ)\; 0 = \frac{\partial \ln L}{\partial \mu_2} = \frac{n\alpha_1\varphi(\alpha_1)(1-\Phi(\alpha_2))}{u(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta)} - \frac{\alpha_2^2}{(\theta-\mu_2)^3}\sum_{j=m+1}^{n}\left(y_j-\mu_2\right)\left(y_j-\theta\right)\\
(3^\circ)\; 0 = \frac{\partial \ln L}{\partial \alpha_1} = \frac{n}{\alpha_1} - (n-m)\,\alpha_1 - \frac{n\varphi(\alpha_1)\left(\alpha_2(\theta-\mu_1)\varphi(\alpha_2)+\left(1-\alpha_1^2\right)(\theta-\mu_2)(1-\Phi(\alpha_2))\right)}{u(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta)} -\\
\qquad\qquad - \frac{\alpha_1}{(\theta-\mu_1)^2}\sum_{i=1}^{m}\left(y_i-\mu_1\right)^2\\
(4^\circ)\; 0 = \frac{\partial \ln L}{\partial \alpha_2} = \frac{n}{\alpha_2} - m\alpha_2 - \frac{n\varphi(\alpha_2)\left(\left(1-\alpha_2^2\right)(\theta-\mu_1)\Phi(\alpha_1)-\alpha_1(\theta-\mu_2)\varphi(\alpha_1)\right)}{u(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta)} -\\
\qquad\qquad - \frac{\alpha_2}{(\theta-\mu_2)^2}\sum_{j=m+1}^{n}\left(y_j-\mu_2\right)^2\\
(5^\circ)\; 0 = \frac{\partial \ln L}{\partial \theta} = -\frac{n(\alpha_2\varphi(\alpha_2)\Phi(\alpha_1)+\alpha_1\varphi(\alpha_1)(1-\Phi(\alpha_2)))}{u(\mu_1,\mu_2,\alpha_1,\alpha_2,\theta)} + \frac{\alpha_1^2}{(\theta-\mu_1)^3}\sum_{i=1}^{m}\left(y_i-\mu_1\right)^2 +\\
\qquad\qquad + \frac{\alpha_2^2}{(\theta-\mu_2)^3}\sum_{j=m+1}^{n}\left(y_j-\mu_2\right)^2
\end{array}\right. . \tag{16}$$

By adding the first two equations of this system with the last one we obtain a simpler equation

$$0 = \frac{\alpha_1^2}{\left(\theta-\mu_1\right)^2}\left(\sum_{i=1}^{m} y_i - m\mu_1\right) + \frac{\alpha_2^2}{\left(\theta-\mu_2\right)^2}\left(\sum_{j=m+1}^{n} y_j - (n-m)\,\mu_2\right), \tag{17}$$

that can be used to replace any of the added equations. Even so, the resulting system is very complex and it requires again numerical methods. Moreover, we must check that the solution for θ satisfies the condition $x_m \le \theta < x_{m+1}$. Therefore, we adopted the following algorithm:

Step 1. Take as initial values the ones resulting from the previous method (based on moments and quantiles) and find the corresponding m.

Step 2. For the current m, find the solution $\hat{\mu}_i, \hat{\sigma}_i, i = 1, 2$, and $\hat{\theta}$ of system (16), eventually with one equation replaced with (17). If the resulting $\hat{\theta}$ is situated in the interval $[x_m, x_{m+1})$ then keep the new found solution; otherwise, go to Step 3.

Step 3. Find the new m such that $x_m \le \hat{\theta} < x_{m+1}$, then resume Step 2 with this m and initial values $\hat{\mu}_i, \hat{\sigma}_i, i = 1, 2$, and $\hat{\theta}$.

Remark 4. The original algorithm described in Teodorescu and Vernic (2009) executes Step 2 for each $m = 1, 2, ..., n-1$, solving system (16) until a correct solution is found. Because we have a large amount of data and this could take too long, we preferred to transform the original algorithm as above, including thus the available data information and the results from the first method.

Remark 5. To execute Step 2, we needed to specify starting values to the mathematical software we used. This is why we indicated how to take these values at Steps 1 and 3.

An ECM (Expectation Conditional Maximization) algorithm. Dempster et al. (1977) provided a good description of the EM method, and since 1977 a number a papers suggested various other ways to perform the computations involved, enlarging the applicability of the EM algorithm to different kinds of models and distributions. For example, Meng and Rubin (1993) introduced the ECM algorithm in which they suggested a component-wise maximization of the loglikelihood function, that sometimes simplifies the maximization problem.

We adapted the ECM algorithm to our composite model as follows: let $\Theta = (\mu_1, \mu_2, \alpha_1, \alpha_2)$, and, being an iterative algorithm, we denote by $\left(\theta^{(0)}, \Theta^{(0)}\right)$, $\left(\theta^{(1)}, \Theta^{(1)}\right)$, ... the sequence of provisional values that converges to the optimal solution. Then the algorithm repeats the following step:

Step k. Determine $\Theta^{(k-1)}$ that maximizes $\ln L$ subject to the constraint $\theta = \theta^{(k-1)}$ by solving equations (1°-4°) from system (16); then determine $\theta^{(k)}$ that maximizes $\ln L$ subject to the constraint $\Theta = \Theta^{(k-1)}$ by solving equation (5°) from the same system.

The algorithm starts with an initial value for $\theta^{(0)}$ and stops when the differences between $\left(\theta^{(k-1)}, \Theta^{(k-1)}\right)$ and $\left(\theta^{(k)}, \Theta^{(k)}\right)$ are small enough.

This algorithm simplifies the solving of system (16), but still needs numerical methods.

3.3 Numerical results

We applied all three methods described above on our log-data set. For the first method, the needed empirical values are

$$\bar{y} = 6.64286,\ \overline{y^2} = 47.73630,\ q_{0.03} = 2.35213,\ q_2 = 6.74524,\ q_3 = 7.75043.$$

As already mentioned, we chose $\alpha = 0.03$ because $q_{0.03}$ is smaller than the initial value of θ, i.e. smaller than 3. To solve the complex system (14), we implemented in SciLab 5.1.1 (an open source platform for numerical computation, for details see http://www.scilab.org/), the trust-region-dogleg algorithm (see Conn et al. 2000, Nocedal and Wright 1999, Powell 1970). The starting values $\mu_i, \sigma_i, i = 1, 2$, were taken from the estimated two components mixture model, and a particular choice of $\theta = 3$. The resulting solution is

$$\breve{\mu}_1 = 0.1357,\ \breve{\mu}_2 = 6.9223,\ \breve{\sigma}_1 = 1.1700,\ \breve{\sigma}_2 = 1.3467,\ \breve{\theta} = 2.5896,\ \breve{r} = 0.0417.$$

This solution was then considered as starting value for the algorithm based on the ML method and for the ECM algorithm, both algorithms being implemented in SciLab 5.1.1. Initially, we had $m = 34$, but after performing the described algorithms, they both ended with $m = 31$ and the improved solutions
ML based algorithm

$$\hat{\mu}_1 = 0.25984,\ \hat{\mu}_2 = 6.83586,\ \hat{\sigma}_1 = 1.00141,$$

$$\hat{\sigma}_2 = 1.54832,\ \hat{\theta} = 2.09384,\ \hat{r} = 0.02984.$$

ECM algorithm

$$\widehat{\hat{\mu}}_1 = 0.25985,\ \widehat{\hat{\mu}}_2 = 6.83586,\ \widehat{\hat{\sigma}}_1 = 1.00142,$$

$$\widehat{\hat{\sigma}}_2 = 1.54832,\ \widehat{\hat{\theta}} = 2.09385,\ \widehat{\hat{r}} = 0.02984.$$

The values of the loglikelihood function are

$$\begin{array}{rcl} \text{Moments and quantiles method} & : & -\,2068.280 \\ \text{ML based method and ECM algorithm} & : & -\,2044.378. \end{array}$$

Therefore, the solutions obtained by the ML based method and by the ECM algorithm are indeed better and there is no significant difference between them (i.e. the difference is eventually at the 5th decimal). The most important difference that we noticed when performing these two algorithms is that the ML based algorithm converges faster than the ECM one. More precisely, the ML based algorithm needed only 4 iterations to give the solution, while the ECM one needed about 40 iterations.

We mention that the Kolmogorov distance for the ML based method is 0.0357131. Hence, this distribution is also accepted by the Kolmogorov goodness-of-fit test, and based on this criterion, it is in-between the mixture models from Section 2.2. Note that the number of estimated parameters is the same for both composite and mixture models, so by comparing the loglikelihood values we can conclude that for these data, the composite model fits a bit better.

4 Conclusions

In this paper we introduced two lognormal models to model a set of real data from fire and allied perils insurance. The models were chosen based on the shape of the log-data histogram. The first model is a two components lognormal mixture, while the second one is a composite lognormal-lognormal model. Some properties of both models are given, with accent on parameters estimation. Then the models were fitted to the data. For the parameters of each model, two and respectively three sets of solutions were obtained, depending on the starting values or on the estimation method. Both models fit the data and the race is very close.

References

[1] Conn, N.R.; Gould N.I.M.; Toint Ph.L. (2000), *Trust-Region Methods*, MPS/SIAM Series on Optimization, SIAM and MPS.

[2] Cooray, K.; Ananda, M.A. (2005), *Modeling actuarial data with a composite Lognormal-Pareto model*, Scandinavian Actuarial Journal **5**, 321–334.

[3] Dempster, A.P.; Laird, M.; Rubin, D.B. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society, B **39 (1)** , 1–38.

[4] Johnson, N. L.; Kotz, S.; Balakrishnan, N. (1994), *Continuous Univariate Distributions*, vol. I. Wiley, New York.

[5] McNeil, A.J. (1997), *Estimating the tails of loss severity distributions using extreme value theory*, ASTIN Bulletin 27 (1), 117-137.

[6] Meng, X.; Rubin, D. (1993), *Maximum likelihood estimation via the ECM algorithm: A general framework*, Biometrica **80(2)**, 267-278.

[7] Nocedal, J.; Wright S.J. (1999), *Numerical Optimization*, Springer Series in Operations Research, Springer Verlag.

[8] Powell, M.J.D. (1970), *A Fortran Subroutine for Solving Systems of Nonlinear Algebraic Equations*, Numerical Methods for Nonlinear Algebraic Equations, (P. Rabinowitz, ed.), Ch.7.

[9] Resnick, S.I. (1977), *Discussion of the Danish data on large fire insurance losses*, ASTIN Bulletin **27(1)**, 139–151.

[10] Scollnik, D.P.M. (2007), *On composite Lognormal-Pareto models*, Scandinavian Actuarial Journal **1**, 20–33.

[11] Shpilberg, D.C.(1997), *The Probability Distribution of Fire Loss Amount*, Journal of Risk and Insurance, Vol. **44(1)**, 103-115.

[12] Teodorescu, S.; Vernic, R. (2009), *Some composite Exponential-Pareto models for actuarial prediction*, To appear in Romanian Journal of Economic Forecasting, 2009.

[13] Vernic, R.; Teodorescu, S. (2009),*On composite Pareto models*, Submitted.

Raluca Vernic
Faculty of Mathematics and Computer Science
“Ovidius” University, 124 Mamaia, 900527 Constanta, Romania
and ”Gheorghe Mihoc-Caius Iacob” Institute of Mathematical Statistics and Applied Mathematics
Calea 13 Septembrie No.13, Sector 5, 050711 Bucharest, ROMANIA
E-mail: rvernic@univ-ovidius.ro

Sandra Teodorescu
Faculty of Economic Sciences, Ecological University of Bucharest
E-mail: cezarina_teodorescu@yahoo.com

Elena Pelican
Faculty of Mathematics and Computer Science
“Ovidius” University, 124 Mamaia, 900527 Constanta, Romania
E-mail: epelican@univ-ovidius.ro