STATISTICAL ANALYSIS OF MIXTURES UNDERLYING PROBABILITY OF RUIN

R. Potocký, M. Stehlík

Received: July 28, 2009

Abstract


If the hypothesis on exponentially distributed claims in a risk (or surplus) model is untenable then, in many cases, the assumption that they are mixtures of two (or more) exponentials is a suitable substitute. In the first part of the paper tests of homogeneity for exponentially distributed claims are discussed and their properties are stated. The statistical properties of parameter estimations for such claims are also mentioned. In the second part the classical Cramer-Lundberg ruin model is discussed when claims are distributed as mixtures of exponentials. Our attention is focussed primarily on assessment of accuracy of approximations obtained. Then our results are compared to those already known.

ELRH test, ELR2 test, pension pillar, ruin probability

In economics, finance and insurance, many claims are mixed from various risk sources. Therefore, the adequate statistical model may use mixture distributions. In our paper we concentrate on two-component exponential scale mixtures and discuss their influence on Cramer-Lundberg ruin model. For the classical risk model with a constant dividend barrier and claim size distribution of exponential and a mixture of exponentials type see (Scheldon, Willmont, Drekic, 2003). For heavy tailed claims see (Potocký, Stehlík, 2007).

Here we consider a homogeneous portfolio of independent, identically distributed positive claims with the distribution function \( F \) and the finite expectation (mean) \( \mu \). The claims occur in random times \( T_n \) and their number in the time interval \([0, t]\) is counted by the process \( N(t) = \text{sup} \{ n \geq 1, T_n \leq t \} \).

If the inter-arrival times are exponentially distributed \( N(t) \) is a homogeneous Poisson process with intensity, say, \( \lambda \). This is the classical Cramér-Lundberg model. If they have Erlang distribution, i.e. Gamma-distribution with \( \alpha = 2 \), \( N(t) \) is a renewal process (for details see, e.g. Potocký, Stehlík, 2007). Both models are very popular among actuaries.

The corresponding process of aggregate claims is

\[
S(t) = \sum_{n=1}^{N(t)} X_n
\]

Suppose that the insurer has an amount of money set aside for this portfolio at time 0. This amount of money is called the initial surplus or free reserves and is denoted by \( u \geq 0 \). The insurer’s surplus at any future time \( t \) is a random variable, since its value depends on the claims experience up to time \( t \). It will be denoted by \( U(t) \). So we have the model

\[
U(t) = u + ct - S(t)
\]

where \( c \) means the premium income rate in one time unit. The model is called the surplus model or risk model. It follows easily that \( EU(t)/t \to c - \lambda \mu \) for \( t \to \infty \).

So the condition \( c - \lambda \mu > 0 \) is necessary for the solvency of the insurance company. However, it can happen that \( U(t) \) falls below zero as a result of the last claim. In such a case we say that ruin has occured. Of course, the company wishes to keep the probability of such event as small as possible. Therefore we define the probability of ultimate ruin as

\[
\Psi(u) = P(U(t) < 0 \text{ for some } t \in (0, \infty]).
\]

The paper is organized as follows. In the first part of the paper we discuss the claims modelled by scale mixtures of exponentials. Also tests of homogeneity for exponentially distributed claims are discussed and their properties are stated. The statistic properties of parameter estimations for such claims are also mentioned. In the second part the classical Cramer-Lundberg ruin model is discussed when claims are distributed as mixtures of exponentials. Our at-
The main advantages of this test statistic is that under \( H_0 \) it does not depend on the unknown value of the parameter \( \theta \).

### RESULTS

#### Ruin probability in the classical Cramer-Lundberg model

In this case \( \Psi(u) \) satisfies the integro-differential equation

\[
\Psi(u) = \lambda/e \Psi(u) - \lambda/e \int_0^u f(x)Y(u-x)dx - \lambda/e \bar{F}(u), \quad u \geq 0, \tag{3}
\]

where \( f(x) \) means the density corresponding to \( F \) and \( \bar{F}(u) = 1 - F(u) \). (see, e.g. Bühlmann, 1970 and Gerber, 1979.) It is well known that for exponentially distributed claims

\[
\Psi(u) = \frac{1}{1 + \rho} \exp \left( \frac{-\rho u}{\mu(1 + \rho)} \right) \tag{4}
\]

where \( \rho = c/(\lambda \mu) - 1 \).

It is shown in (Gerber, 1979) that (4) can be rewritten in the form

\[
e^{\Psi(u)} = \lambda \left[ \int_0^u (1 - F(x))dx + \int_0^u \Psi(u-x)(1 - F(x))dx \right]. \tag{5}
\]

Consider now a mixture of 2 exponential distributions with density functions \( f_1(x) = \alpha \exp(-\alpha x) \) and \( f_2(x) = \beta \exp(-\beta x) \), respectively, where \( 0 < \alpha < \beta \), i.e. the density function of the mixture will be \( f(x) = pf_1(x) + (1 - p)f_2(x) \).

We know that the moment generating function is

\[
M(r) = \begin{cases} 
\frac{1}{\alpha - r} & (1 - p) \frac{\beta}{\beta - r} 
\end{cases}
\]

provided \( r < \alpha \).

Having in mind the result for exponential distributions we seek the solution of (3) in the form

\[
\Psi(u) = C_i \exp(-r_i u) + C_j \exp(-r_j u) \tag{7}
\]

for suitable \( C_i, r_i, i = 1, 2 \).

Substituting in (3) we obtain that are the solutions of the equation

\[
ct^2 - [(\alpha + \beta) + \lambda]t + \alpha \beta c - \lambda((1-p)\alpha + p\beta) = 0. \tag{8}
\]

The solutions are

\[
r_i = 1/2(\alpha + \beta - \lambda/c - \sqrt{(\beta - \alpha - \lambda/c)^2 + 4\rho/(\beta - \alpha)}),
\]

and

\[
r_j = 1/2(\alpha + \beta - \lambda/c + \sqrt{(\beta - \alpha - \lambda/c)^2 + 4\rho/(\beta - \alpha)}). \tag{10}
\]
It holds $r_1 < \alpha < r_2 < \beta$.

We also have

$$C_1 = \frac{r_1(r_2 - \alpha)(r_2 - \beta)}{(r_2 - r_1)(r_1 \alpha + \beta)} \quad \text{and} \quad C_2 = \frac{r_1(r_2 - \alpha)(r_2 - \beta)}{(r_1 - r_2)(r_1 \alpha + \beta)}. \quad (11)$$

It is worth investigating the behaviour of (7) as the function of $p$, the rest of parameters being fixed.

In the rest of the paper our attention will be focused on the Cramér-Lundberg approximation of probability of ruin in this case. It is well known that in general we have

$$\Psi(u) \sim K \exp(-Ru), \quad (12)$$

where the constant depends on the value of the first derivative of the moment generating function of the distribution in the Lundberg exponent $R$. It can be shown that in the case of the mixture of two exponentials considered above $R = r_1$ and $K = C_1$.

It follows that unlike exponential distribution the Cramér-Lundberg approximation is not exact in this case. Again the exactness of (12) depends on $p$.

**Example 1 (upper contamination)** Let us consider a claim distribution of the form $0.1 \exp(-x) + 0.9 \exp(-\theta x)$. We have $\alpha = 1, \beta = \theta, \lambda = 1, c = 1$. Thus, $r_1$ are solutions of equation $r_1^2 - \theta r_1 + \theta - (0.9 + 0.1 \theta) = 0$.

We obtain solutions $r_1 = 0.5 - 0.1 \sqrt{25 \theta - 900 + 90}$,

$r_2 = 0.5 + 0.1 \sqrt{25 \theta - 900 + 90}$.

The dependence of $r_2$ on $\theta$ can be seen from Figure 1.

Let us fix $\theta = 10$ for the sake of simplicity. Then we got solutions $r_1 = 0.8890390418, r_2 = 9.110960958$. Finally we have $C_1 = 0.1120276489, C_2 = 0.07797235108$. Therefore $\Psi(u) = 0.1120276489 \exp(-0.8890390418u) + 0.07797235108 \exp(-9.110960958u)$.

**Example 2 (lower contamination)** Let us consider a claim distribution of the form $0.9 \exp(-x) + 0.1 \exp(-\theta x)$. We have $\alpha = 1, \beta = \theta, \lambda = 1, c = 1$. Thus, $r_1$ are solutions of equation $r_1^2 - \theta r_1 + \theta - (0.9 + 0.1 \theta) = 0$.

We obtain solutions $r_1 = 0.50 - 0.1 \sqrt{25 \theta - 100 + 10}$,

$r_2 = 0.50 + 0.1 \sqrt{25 \theta - 100 + 10}$.

Let us fix $\theta = 10$ for the sake of simplicity. Then we got solutions $r_1 = 0.909028417, r_2 = 9.909175083$. Finally $C_1 = 0.9092514700, C_2 = 0.007485296299$ and $\Psi(u)$ has the form $0.90925147 \exp(-0.909824917u) + 0.007485296299 \exp(-9.909175083u)$.

For both contaminations we have computed the Cramer Lundberg approximation and graphically compared with the exact value of $\Psi(u)$ in Figures 2 and 3. As we can see form Figures, the tightness of curves in Figure 3 is so high, that we cannot distinguish individual curves. Therefore we can conclude, that Cramer Lundberg approximation works relatively better for lower contamination.

**Real Data Example:** 1st pension pillar in Slovakia

The problem that assets of a pension fund are not sufficient to cover its liabilities is of extreme importance. Such a situation may arise in some countries in connection with the so-called non-funded 1st pension pillar based on pay-as-you-go principle.
Here we consider the illustrative example of claims for the mandatory, non-funded 1st (pay-as-you-go) pillar given by Potocký, Stehlík, 2005. Therein is considered a closed group of Slovakian people, all aged 50 in the year 1998, and interest is in the estimation of the total claim amount for this group in the year 2010 when the members are supposed to retire. Table I contains the average salaries given from Statistical Yearbook, 2004, Labour Market, III.3-10, Structure of average gross nominal monthly wage of employees in the economy of the SR. In Potocký, Stehlík, 2005, the authors are interested in estimation of the probabilities $P\left(\sum_{k=1}^{N} X_k > C\right)$, where $X_i$ are individual monthly claims of the members of the above-mentioned group and $C$ is a critical (limiting) value of the fund representing the amount the fund has gathered from the contributions of the active members or from other sources. It is possible to consider $N$ as a constant or a random variable as it was treated in Potocký, Stehlík, 2005. In Potocký, Stehlík, 2007, the case that $N$ is a random variable was considered. Then it is quite natural to choose a binomial model for $N$ namely $N \sim bi(n, p)$ with $n = 130 000$ and $p$ representing the probability of surviving a 50-year person from the group to the age 62 years (such probabilities are regularly published by Slovak Statistical Office). Then one is looking for the largest $C$ such that $P\left(\sum_{k=1}^{N} X_k > C\right)$ with $p$ given in advance, e.g. 0.1 or 0.05. It can be seen from Stehlík, Střelec 2009 that one can find the tests which are close to rejection of normality at certain size. The interpretation can be that for larger samples of wages normality is violated and we can consider light-tailed claims as given in Potocký, Stehlík, 2007.

Typically it is possible to model salaries as normal variables in short-terms and lognormal at long-term. In Potocký, Stehlík, 2005, we have used the normal distribution which led to the following upper bound

$$p = 1 - \Phi \left( \frac{C/[kN]}{\mu} \right)$$

(13)

Here $\Phi$ is cdf of standardized normal distribution, $C$ is a critical level as given above, $\mu$ and $\sigma^2$ are parameters of normal distribution of salaries, $k = \frac{\sigma^2}{\mu}$ and $N$ is the number of claims. In the case of Table I we have $\mu = 29396.4$ and $\sigma = 3903.35$. Now let us consider data according to SLOVSTAT (on-line), see Table II.

### DISCUSSION AND SUMMARY

As we can see from this contribution, one can apply the Cramer-Lundberg approach also for scale mixtures of exponentials. By using the real data, one should be careful about distribution and one possibility is to use exact likelihood ratio tests, Elr2 or Elrh. We can conclude that difference is observed for case of upper or lower contamination, which should be carefully recognized. Cramer-Lundberg approximation works relatively better for the lower contamination. This topic is worth further investigation.

### SOUHRN

Štatistická analýza zmesí v pravdepodobnosti ruinovania

REFERENCES

SCHMIDLI, H., 2000: Lecture notes on Risk Theory. University of Aarhus
STEHLÍK, M., 2003: Distributions of exact tests in the exponential family, Metrika 57, 145–164

Address
prof. RNDr. Rastislav Potocký, Ph.D., Ústav aplikovanej matematiky a statistiky, Univerzita Komenského, Mlynská Dolina, 842 48 Bratislava, Slovenská republika, e-mail: potocky@fmph.uniba.sk, RNDr. Milan Stehlík, Ph.D., Ústav aplikované statistiky, JU Linz, Freistädter Straße 315, A-4040 Linz a. D., Rakousko, e-mail: Milan.Stehlik@jku.at