ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS
SBORNÍK MENDELOVY ZEMĚDĚLSKÉ A LESNICKÉ UNIVERZITY V BRNĚ

Volume LVII 21 Number 6, 2009

STATISTICAL ANALYSIS OF MIXTURES UNDERLYING PROBABILITY OF RUIN

R. Potocký, M. Stehlík

Received: July 28, 2009

Abstract

POTOCKÝ, R., STEHLÍK, M.: *Statistical analysis of mixtures underlying probability of ruin.* Acta univ. agric. et silvic. Mendel. Brun., 2009, LVII, No. 6, pp. 209–214

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 assesment 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 barierr 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 X_k with the distribution function F and the finite expectation (mean) μ. The claims occur in random times T_n and their number in the time interval [0, t] is counted by the process $N(t) = \sup\{n \geq 1, T_n \leq t\}$. If the inter-arrival times are exponentially distributed $N(t)$ is a homogeneous Poisson process with intensity, say, λ. 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 $S(t) = \sum_{i=1}^{N(t)} X_i$. 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

$$U(t) = u + ct - S(t), \tag{1}$$

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

$$\Psi(u) = P\{U(t) < 0 \text{ for some } t \in (0, \infty)\}. \tag{2}$$

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 attention is focussed primarily on assesment of accuracy of approximations obtained. Then our results are compared to those already known. We also illustrate the given methods on real data from Pay as you go pillar in Slovakia.

MATERIALS AND METHODS

In many situations, claims are modeled as mixtures (so called risk-competitive model). For our purpose we will use only scale exponential mixtures. There is a vast amount of literature devoted to the exponential mixture.

For the sake of simplicity we will stay at the maximal number of components two. Natural question is what is the optimal test, having a data, whether homogeneity (one component) or heterogeneity holds (more than one component). We will concentrate on the exact LR test of homogeneity, against the 2 component alternative (so called Elr2) and against general alternative (so called ElrH). The first was provided by (Stehlík, Ososkov, 2003) and the latter by (Stehlík, 2003). The properties, comparisons to the MLRT, ADDS and other tests, and advantages of such tests has been studied in (Stehlík, Wagner, 2009). The tests of homogeneity and scale provided in this paper are asymptotically optimal in the Bahadur sense (Rublík, 1989, Part 1, Part 2) when the underlying distribution is exponential and when the alternative of the homogeneity consists of a finite sample mixture.

Because of scaling property it is enough to think about two component mixture of the form: $p\exp(-x) + (1-p)\,\theta\exp(-\theta x)$. This mixture is called to be lower contaminated for $p > 0.5$ and upper contaminated otherwise. The power of ElrH and Elr2 is relatively better for lower contamination (see (Stehlík, Wagner, 2009)). The LR statistics $-\ln\Lambda_N$ of ELRH test is derived in Theorem 3 of (Stehlík, 2006) for $y_1, \ldots, y_N$ i.i.d. from the exponential distribution. It has the form

$$N\ln\left(\sum_{i=1}^{N} y_i\right) - N\ln N - \sum_{i=1}^{N}\ln y_i.$$

A very important property of the LR test of homogeneity is its scale invariance, i.e. its distribution under H_0 is independent of the unknown scale parameter. This is an advantage in comparison to some asymptotical tests and tests depending on the true but unknown value of θ. The critical values are easy to obtain by simulation, e.g. from the standard exponential distribution or the Dirichlet distribution.

Elr2 test, the efficient testing procedure of the number of components m in the Exponential mixture for $m = 2$ was firstly introduced by (Stehlík, Ososkov, 2003). Following their results we obtain the formula

$$\Lambda_N(y) = \min_{0<K<N,\, p\in P(K)} \left\{ \frac{N^N}{K^K(N-K)^{N-K}} \, \frac{(y_{i_1} + \ldots + y_{i_K})^K (y_{i_{K+1}} + \ldots + y_{i_N})^{N-K}}{(y_1 + \ldots + y_N)^N} \right\}.$$

The main advantages of this test statistic is that under H_0 it does not depend on the unknown value of the parameter θ.

RESULTS

Ruin probability in the classical Cramer-Lundberg model

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

$$\Psi^{(1)}(u) = \lambda/c\,\Psi(u) - \lambda/c\int_0^u f(x)\Upsilon(u-x)dx - \lambda/c\,\bar{F}(u),\ u \ge 0, \quad (3)$$

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

$$\Psi(u) = \frac{1}{1+\rho}\exp\frac{-\rho u}{\mu(1+\rho)} \quad (4)$$

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

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

$$c\Psi^{(1)}(u) = \lambda\left[\int_u^{\infty}(1-F(x))dx + \int_0^u \Psi(u-x)(1-F(x))dx\right]. \quad (5)$$

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

$$M(r) = p\,\frac{\alpha}{\alpha - r} + (1-p)\,\frac{\beta}{\beta - r} \quad (6)$$

provided $r < \alpha$.

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

$$\Psi(u) = C_1\exp(-r_1 u) + C_2\exp(-r_2 u) \quad (7)$$

for suitable $C_i, r_i, i = 1,2$.

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

$$cr^2 - ((\alpha+\beta) - \lambda)r + \alpha\beta c - \lambda((1-p)\alpha + p\beta) = 0. \quad (8)$$

The solutions are

$$r_1 = 1/2(\alpha + \beta - \lambda/c - \sqrt{(\beta - \alpha - \lambda/c)^2 + 4p\lambda/c(\beta - \alpha)}) \quad (9)$$

and

$$r_2 = 1/2(\alpha + \beta - \lambda/c - \sqrt{(\beta - \alpha - \lambda/c)^2 + 4p\lambda/c(\beta - \alpha)}). \quad (10)$$

It holds $r_1 < \alpha < r_2 < \beta$.
We also have

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

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 wil be focused on the Cramér-Lundberg approximation of probability of ruin in this case. It is well known that in general we have

$$\Psi(u) \sim K\exp(-Ru), \tag{12}$$

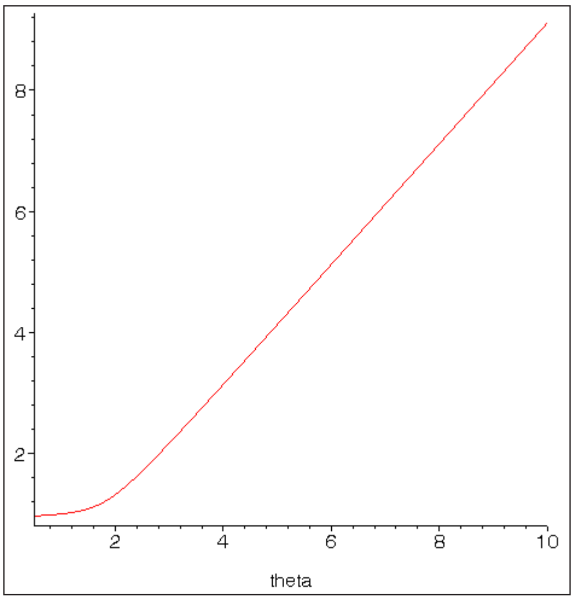
where the constant depends on the value of the first derivative of the moment generating function of the distribution in the Lundberg exponent R. It can be shown that in the case of the mixture of two exponentials considered above $R = r_1$ and $K = C_1$.

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 $0.1\exp(-x) + 0.9\theta\exp(-\theta x)$. We have $\alpha = 1, \beta = \theta, \lambda = 1, c = 1$. Thus, r_i are solutions of equation $r_2 - \theta r + \theta - (0.9 + 0.1\theta) = 0$. We obtain solutions $r_1 = 0.5\theta - 0.1\sqrt{25\theta^2 - 90\theta + 90}$, $r_2 = 0.5\theta + 0.1\sqrt{25\theta^2 - 90\theta + 90}$.

The dependence of r_2 on θ 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 $0.9\exp(-x) + 0.1\theta\exp(-\theta x)$. We have $\alpha = 1, \beta = \theta, \lambda = 1, c = 1$. Thus, r_i are solutions of equation $r^2 - \theta r + \theta - (0.1 + 0.9\theta) = 0$.

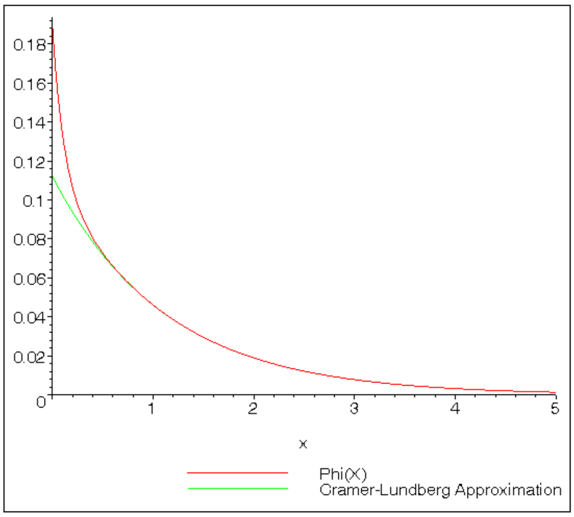


1: *Dependance of* r_2 *on* θ

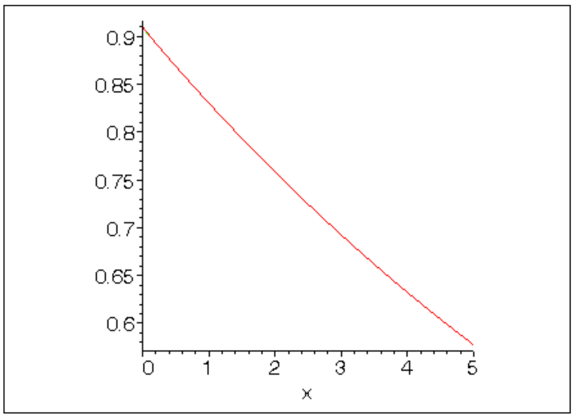
We obtain solutions $r_1 = 0.5\theta - 0.1\sqrt{25\theta^2 - 10\theta + 10}$, $r_2 = 0.5\theta + 0.1\sqrt{25\theta^2 - 10\theta + 10}$.

Let us fix $\theta = 10$ for the sake of simplicity. Then we got solutions $r_1 = 0.090824917$, $r_2 = 9.909175083$. Finally $C_1 = 0.9092514700$, $C_2 = 0{,}000748296299$ and $\Psi(u)$ has the form $0.90925147\exp(-0.090824917u) + 0.0007485296299\exp(-9.909175083u)$.

For both contaminations we have computed the Cramer Lundber 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 Lundber approximation works relatively better for lower contamination.



2: *Comparison of exact probability of ruin* $\Psi(x)$ *and its Cramer Lundberg Approximation for example 1 (Upper contamination)*



3: *Comparison of exact probability of ruin* $\Psi(x)$ *and its Cramer Lundberg Approximation for example 1 (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 $P(\Sigma_{k=1}^{N} X_k > C)$, 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 $P(\Sigma_{k=1}^{N} X_k > C)$ with p given in advance, e.g. 0.1 or 0.05.

I: *Average salaries, 1998–2002*

year	salary
1998	24 233
1999	26 862
2000	30 021
2001	31 825
2002	34 041

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

$$\bar{p} = 1 - \Phi\left(\frac{C/(kN_t) - \mu}{\sigma}\right) \qquad (13)$$

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

II: *Average salaries, 2000–2006*

year	salary
2000	29 737
2001	31 060
2002	34 262
2003	35 533
2004	34 490
2005	28 174
2006	30 077

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. Here we will use the assumption of exponential tailed claims. The ELRH test statistics for data from table II is 0.02363513. Therefore, we can accept scale homogeneity for the exponential distribution on 0.1-level. However, other reasons to use the scale non homogeneity can be found. This analysis is however out of the scope of this paper.

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

V našom príspevku sa zaoberáme aplikácou Cramer-Lundbergovej aproximácie pre škálové zmesy exponenciálnych rozdelení. Pri aplikácii uvedenej metódy na reálne dáta je potrebné otestovať vhodnosť modelu napr. testami Elr2 alebo Elrh uvedenými v príspevku. Prípad hornej a dolnej kontaminácie vedie na rozdielnu kvalitu aproximácie, preto je ich nutné starostlivo rozlišovať. Cramer-Lundbergova aproximácia je relatívne lepšia v prípade dolnej kontaminácie.

Research of Stehlík M. was partially supported by Project 50p14 "AKTION Czech Republic-Austria": Effective estimation supporting the risk assessment for Basel II".

REFERENCES

BÜHLMANN, H., 1970: Mathematical methods in risk theory. Springer-Verlag. ISSN 0072-7830.

GERBER, H. U., 1979: An introduction to mathematical risk theory. S. S. Huebner Foundation, Philadelphia, PA.

POTOCKÝ, R., STEHLÍK, M., 2005: Analysis of pensions in the 1st pilar under the expected demographic development. *Proceedings of 10th Slovak Conference on Demography, Smolenice*.

POTOCKÝ, R., STEHLÍK, M., 2007: Stochastic models in insurance and finance with respect to Basel II, JAMSI, 3, No. 2, p. 237–245

RUBLÍK, F, 1989: On optimality of the LR tests in the sense of exact slopes, Part 1, General case. Kybernetika, 13–25

RUBLÍK, F, 1989: On optimality of the LR tests in the sense of exact slopes, Part 2, Application to Individual Distributions. Kybernetika, 117–135

SHELDON, L. X. E, WILLMOT, G., DREKIC, S., 2003: The classical risk model with a constant dividend barrier: analysis of the Gerber-Shiu discounted penalty function, Insurance: Mathematics and Economics, Vol. 33, No 3, 551–566 (16)

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

STEHLÍK, M., OSOSKOV, G. A., 2003: Efficient testing of the homogeneity, scale parameters and number of components in the Rayleigh mixture. JINR Rapid Communications E-11-2003-116.

STEHLÍK, M., STŘELEC, L.,2009: On normality assumptions for claims in insurance, Acta univ. agric. et silvic. Mendel. Brun., 2009, LVII, No. 3, ISSN 1211-8516, p. 141–146

STEHLÍK, M., WAGNER, H., 2009: Exact likelihood ratio testing for homogeneity of the exponential distribution, IFAS Res. Rep. 39.

STEHLÍK, M., 2006: Exact likelihood ratio scale and homogeneity testing of some loss processes. Statistics and Probability Letters 76, 19–26.

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, JKU Linz, Freistädter Straße 315, A-4040 Linz a. D., Rakousko, e-mail: Milan.Stehlik@jku.at