Датум пријема рада: 15.04.2014.
Датум прихватања рада: 15.06.2014
Оригинални научни рад
УДК: 347.426.6:368
DOI: 10.7251/NOE1614016R

MODELIRANJE ŠTETA U OSIGURANJU PRIMENOM MEŠAVINE LINEARNIH EKSPONENCIJALNIH RASPODELA

MODELLING INSURANCE LOSSES USING MIXED LINEAR EXPONENTIAL DISTRIBUTION

Prof. dr Vesna Rajić, Prof. dr Tatjana Rakonjac-Antić
Katedra za statistiku i matematiku i Katedra za ekonomsku politiku i razvoj
Ekonomski fakultet Beograd, Srbija
vesnac@ekof.bg.ac.rs, rakonjac@ekof.bg.ac.rs

Rezime: *Određivanje raspodela verovatnoća šteta je veoma značajno za osiguravajuće kompanije, jer se na osnovu njih određuju premije osiguranja, rezerve, solventnosni kapital i samopridržaj kompanije. U radu se predlaže primena mešavine linearnih eksponencijalnih raspodela za modeliranje šteta u osiguranju. Ocene po metodu maksimalne verodostojnosti parametara ove raspodele se dobijaju korišćenjem EM algoritma. Izvedeni su obrasci za izračunavanje VaR i TVaR.*

Ključne riječi: *štete u osiguranju, linearna eksponencijalna raspodela, mešavina linearnih eksponencijalnih raspodela, funkcija maksimalne verodostojnosti, mere rizika.*

Abstract: Determining distribution losses is very important for insurance companies, because distribution losses determine insurance premiums, reserves, solvency capital and company's retention. In this paper we suggest the use of a mixed linear exponential distribution for modeling insurance losses. The maximum likelihood estimators of the parameters can be obtained by using EM Algorithm. We derived forms for calculating VaR and TvaR.

Key Words: *insurance losses, linear exponential distribution, mixed linear exponential distribution, maximum likelihood function, risk measures*

UVOD

Linearna eskponencijalna raspodela je raspodela koja je do sada uglavnom korišćena za modeliranje veka trajanja uređaja u teoriji pouzdanosti ili za opisivanje životnog veka pacijenata u medicinskim istraživanjima.

Njena funkcija gustine je oblika:

$$f(x) = (\theta x + \lambda) e^{-(\frac{\theta}{2}x^2 + \lambda x)}, \; x > 0,$$

gde su $\theta > 0$ i $\lambda \geq 0$ parametri raspodele.

U ovom radu se analizira mešavina linearnih eksponencijalnih raspodela, koju su uveli Rajić i Stanojević (videti [10]) i mogućnosti njene primene u osiguranju imovine. Svrha osiguranja imovine jeste nadoknada štete koja nastaje na toj imovini (videti [7]).

Za slučajnu promenljivu X se kaže da ima mešavinu linearnih eksponencijalnih raspodela (skraćeno MLED) sa parametrima $0 < a_1 < 1$, $0 < a_2 < 1$, $\theta_1 > 0$, $\lambda_1 \geq 0$, $\theta_2 > 0$, $\lambda_2 \geq 0$, ako je njena funkcija gustine oblika:

$$f(x;\Theta) = a_1(\theta_1 x + \lambda_1) e^{-(\frac{\theta_1}{2}x^2 + \lambda_1 x)} + a_2(\theta_2 x + \lambda_2) e^{-(\frac{\theta_2}{2}x^2 + \lambda_2 x)}, \quad (1)$$

gde je: $a_1 + a_2 = 1$ i $\Theta = (a_1, a_2, \theta_1, \theta_2, \lambda_1, \lambda_2)$.

Funkcija raspodele je oblika:

$$F(x;\Theta)=a_1[1-e^{-(\frac{\theta_1}{2}x^2+\lambda_1 x)}]+a_2[1-e^{-(\frac{\theta_2}{2}x^2+\lambda_2 x)}] \quad (2)$$

za sve $x>0$, dok su r-ti momenat i varijansa (videti [10]):

$$\mu_r(\Theta)=\Gamma(\frac{r-i}{2}-j+1)\cdot$$
$$\left[a_1\sum_{i=0}^{r}\sum_{j=0}^{\infty}\binom{r}{i}\binom{\frac{r-i}{2}}{j}2^{\frac{r-i}{2}-j}\frac{(-1)^i\lambda_1^{i+2j}}{\theta_1^{\frac{r+i}{2}+j}}+\right. \quad (3)$$
$$\left.a_2\sum_{i=0}^{r}\sum_{j=0}^{\infty}\binom{r}{i}\binom{\frac{r-i}{2}}{j}2^{\frac{r-i}{2}-j}\frac{(-1)^i\lambda_2^{i+2j}}{\theta_2^{\frac{r+i}{2}+j}}\right],$$

$$\sigma^2=\frac{2a_1^2}{\theta_1}\left\{1-\left[\left(\sqrt{\frac{\theta_1}{2}}\mu_1+\frac{\lambda_1}{2}\sqrt{\frac{2}{\theta_1}}\right)^2-\frac{\lambda_1^2}{2\theta_1}\right]\right\}$$
$$+\frac{2a_2^2}{\theta_2}\left\{1-\left[\left(\sqrt{\frac{\theta_2}{2}}\mu_2+\frac{\lambda_2}{2}\sqrt{\frac{2}{\theta_2}}\right)^2-\frac{\lambda_2^2}{2\theta_2}\right]\right\}. \quad (4)$$

Cilj ovog rada je modeliranje realnog skupa podataka o štetama primenom mešavine linearnih eksponencijalnih raspodela istovremeno obezbeđujući premiju koja je sa visokom sigurnošću dovoljna za pokriće očekivanih šteta. Struktura rada je sledeća. U odeljku 2 se određuju ocene parametara koristeći EM algoritam (skraćeno od eng. Expectation-Maximization). Izvedene su, u odeljku 3, odgovarajuće mere rizika za individualne štete koje slede posmatranu mešavinu raspodela. U odeljku 4 se na realnom skupu podataka testira adekvatnost primene predložene mešavine raspodela i određuje pouzdanost rezultujuće premije osiguranja. Zaključna razmatranja data su u odeljku 5.

2. METOD MAKSIMALNE VERODOSTOJNOSTI KORIŠĆENJEM EM ALGORITMA

Ocene po metodu maksimalne verodostojnosti parametara uvedene raspodele se mogu dobiti korišćenjem EM algoritma (pogledati [1,2,3,6,8,9,11]). Kod mešavine raspodela razmatra se funkcija oblika:

$$\ln L(\Theta\,|\,X,Y)=\sum_{i=1}^{N}\ln(a_{y_i}f_{y_i}(x_i\,|\,\Theta_{y_i}))\,, \quad (5)$$

gde je $X=(x_1,\dots,x_N)$, $Y=(y_1,...,y_N)$, $y_i\in\{1,2\}$ za svako i i $y_i=k$ ako je i-ti uzorak generisan iz k-te komponente mešavine raspodela. Takođe je:

$$\Theta=(a_1,a_2,\theta_1,\theta_2,\lambda_1,\lambda_2)\,,$$
$$\Theta_1=(a_1,\theta_1,\lambda_1)\,,$$
$$\Theta_2=(a_2,\theta_2,\lambda_2)\,,$$
$$a_1+a_2=1 \text{ i}$$
$$f_i(x\,|\,\Theta_i)=(\theta_i x+\lambda_i)e^{-(\frac{\theta_i}{2}x^2+\lambda_i x)},\ i=1,2.$$

Uslovna očekivana vrednost od (5) je oblika:

$$Q(\Theta,\Theta^g)=\sum_{j=1}^{2}\sum_{i=1}^{N}\ln(a_j)f(j\,|\,x_i,\Theta^q)$$
$$+\sum_{j=1}^{2}\sum_{i=1}^{N}\ln(f_j(x_i\,|\,\Theta_j))f(j\,|\,x_i,\Theta^q)\,, \quad (6)$$
$$=Q_1(\Theta,\Theta^g)+Q_2(\Theta,\Theta^g).$$

gde su $\Theta^g=(a_1^g,a_2^g,\Theta_1^g,\Theta_2^g)$ trenutne ocene parametara. Sledeći korak je maksimizacija očekivanja (5), odakle se dobija:

$$a_j=\frac{1}{N}\sum_{i=1}^{N}f(j\,|\,x_i,\Theta^g)\,,$$

i:

$$\sum_{i=1}^{N}\frac{x_i f(1\,|\,x_i,\Theta^g)}{\theta_1 x_i+\lambda_1}-\frac{1}{2}\sum_{i=1}^{N}x_i^2 f(1\,|\,x_i,\Theta^g)=0, \quad (7)$$

$$\sum_{i=1}^{N}\frac{f(1\,|\,x_i,\Theta^g)}{\theta_1 x_i+\lambda_1}-\sum_{i=1}^{N}x_i f(1\,|\,x_i,\Theta^g)=0, \quad (8)$$

$$\sum_{i=1}^{N}\frac{x_i f(2\,|\,x_i,\Theta^g)}{\theta_2 x_i+\lambda_2}-\frac{1}{2}\sum_{i=1}^{N}x_i^2 f(2\,|\,x_i,\Theta^g)=0, \quad (9)$$

$$\sum_{i=1}^{N}\frac{f(2\,|\,x_i,\Theta^g)}{\theta_2 x_i+\lambda_2}-\sum_{i=1}^{N}x_i f(2\,|\,x_i,\Theta^g)=0. \quad (10)$$

Rešavanjem jednačina (7), (8), (9), (10) korišćenjem nekog programa (npr. Maple 11 ili Fortran) dobijaju se ocene nepoznatih parametara. Najbolje je primeniti EM algoritam nekoliko puta sa različitim skupovima slučajno izabranih početnih vrednosti i dobiti konzistentne ocene (ova preporuka je data u [5]).

3. IZRAČUNAVANJE MERA RIZIKA

Izračunavanje mera rizika je od izuzetnog značaja za određivanje premija osiguranja, solventnosnog kapitala, tehničkih rezervi kompanije i sprovođenje alokacije kapitala. U cilju upravljanja rizikom važne mere su vrednost pod rizikom (VaR) i uslovno očekivanje repa raspodele (TVaR). Više detalja o primeni ovih mera u cilju određivanja premije osiguranja može se pronaći u [4].

Neka je X slučajna promenljiva koja opisuje štete i $F(x)$ njena funkcija raspodele data jednačinom (2). Vrednost pod rizikom (VaR) na nivou poverenja p (u oznaci V_p) je 100p-ti percentil raspodele, tj. $F(V_p) = p$. Može se pokazati da je V_p rešenje jednačine:

$$a_1[1-e^{-(\frac{\theta_1}{2}V_p^2+\lambda_1 V_p)}]+a_2[1-e^{-(\frac{\theta_2}{2}V_p^2+\lambda_2 V_p)}]=p. \quad (11)$$

U cilju rešavanja jednačine (11) i određivanja V_p može se koristiti bilo koji kompjuterski program.

Uslovno očekivanje repa raspodele na nivou poverenja p (engl. tail value at risk ili conditional tail expectation) se definiše kao $\mathrm{TVaR} = E(X \mid X > V_p)$ i izračunava na sledeći način:

$$\mathrm{TVaR} = E(X \mid X > V_p) =$$
$$= a_1 V_p e^{-\left[\frac{\theta_1}{2}V_p^2+\lambda_1 V_p\right]} + a_1 e^{\frac{\lambda_1^2}{2\theta_1}} \frac{1}{\sqrt{\theta_1}} \sqrt{2\pi}\Phi\left(-\sqrt{\theta_1}V_p - \frac{\lambda_1}{\sqrt{\theta_1}}\right) +$$
$$+ a_2 V_p e^{-\left[\frac{\theta_2}{2}V_p^2+\lambda_2 V_p\right]} + a_2 e^{\frac{\lambda_2^2}{2\theta_2}} \frac{1}{\sqrt{\theta_2}} \sqrt{2\pi}\Phi\left(-\sqrt{\theta_2}V_p - \frac{\lambda_2}{\sqrt{\theta_2}}\right),$$

GDE JE $\Phi(\cdot)$ FUNKCIJA NORMALNE RASPODELE.

4. ANALIZA PODATAKA

Osiguranje imovine obuhvata opšte osiguranje imovine i transportno imovinsko osiguranje. U opšte osiguranje imovine svrstavaju se osiguranje stvari, osiguranje od odgovornosti i osiguranje kredita. Osiguranje mašina od loma predstavlja deo osiguranja stvari (videti [7]). U ovom delu rada upravo se analiziraju podaci o štetama u osiguranju mašina od loma koji su zabeleženi u jednom gradu u Srbiji 2003. godine. Ukupno je bilo 339 opservacija, a zbog poverljivosti podataka navode se samo neke od osnovnih karakteristika:

- Srednja vrednost podataka je 20323,22 novčanih jedinica;
- Standardna devijacija je 54175,6 novčanih jedinica;
- Koeficijent asimetrije je 8,78, što sugeriše da je u pitanju znatno asimetrična raspodela;
- Interval varijacije, odnosno raspon je 754942 novčanih jedinica.

Nakon primene EM algoritma ocenjene vrednosti parametara su:

$\theta_1 = 15 \cdot 10^{-11}$, $\lambda_1 = 5 \cdot 10^{-7}$, $\theta_2 = 21 \cdot 10^{-9}$, $\lambda_2 = 75 \cdot 10^{-6}$,

$a_1 = 0,2$ i $a_2 = 0,8$. Vrednost K-S statistike iznosi 0,05751, što znači da se nulta hipoteza o saglasnosti sa mešavinom linearnih eksponencijalnih raspodela ne odbacuje. To govori da je primena mešavine linearnih eksponencijalnih raspodela na ovim podacima opravdana.

Očekivana vrednost ove raspodele za ocenjene vrednosti parametara iznosi 24549,39 (što se može izračunati koristeći izraz (3) za r=1). Poznato je da neto premija odgovara očekivanoj vrednosti raspodele šteta u osiguranju, tj.

$$P = E(X).$$

Verovatnoća da štete budu manje od neto premije se određuje na osnovu formule:

$$p = P(X<E(X))$$

i predstavlja nivo poverenja, odnosno pouzdanost sa kojom se određuje premija osiguranja. Koristeći formulu (11) i zamenom premije umesto V_p, dolazi se do odgovarajuće pouzdanosti od 0,811. Dakle, sa relativno visokom pouzdanošću može da se tvrdi da su u slučaju primene mešavine linearnih eksponencijalnih raspodela premije dovoljne za pokriće očekivanih šteta.

Prilikom analiziranja podataka o štetama iz prošlosti pored modela koji najbolje opisuje podatke treba razmatrati i pouzdanost obračuna premije. Jedino u slučaju visoke pouzdanosti (koju bi trebalo da procene osiguravajuće kompanije za konkretne slučajeve) neće postojati opasnost nedovoljnosti premije osiguranja.

ZAKLJUČAK

U ovom radu je analizirana jedna nova vrsta raspodele, mešavina lineranih eksponencijalnih raspodela. Pošto ova raspodela nije poznata u oblasti aktuarstva, predlaže se njena primena u osiguranju imovine.

U radu su izvedene korisne mere rizika, a na konkretnim podacima je izračunata adekvatna premija sa zadovoljavajućom pouzdanošću. Može se zaključiti da je veoma značajno odrediti odgovarajuću raspodelu, odnosno adekvatno modelirati podatke o štetama, što predstavlja osnov za determinisanje samopridržaja osiguravača, njegovih rezervi, itd.

LITERATURA

[1] J. A. Bilmes, A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, International Computer Science Institute, Berkeley, California, 1998.

[2] D. Bohning, Computer Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others, Champan & Hall, New York, 1999.

[3] A. P. Dempster, N. M. Laird i D. B. Rubin, "Maximum-likelihood from incomplete date via the em algorith", J. Royal Statist. Soc. Ser. B., 39 (1977), pp. 1-38.

[4] M. Hardy, An Introduction to Risk Measures for Actuarial Applications. CAS Exam Study Note, Arlington, Virginia: Casualty Actuarial Society, 2006, pp. 1-31.

[5] S-H. Kim, "Calibrated initials for an EM applied to recursive models of cstegorical variable", Computational Statistics & Data Analysis, 40, 2002, pp. 97-110.

[6] J. Kočović, V. Ć. Rajić i M. Jovanović, "Estimating a Tail of the Mixture of Log-normal and Inverse Gaussian Distribution“, Scandinavian Actuarial Journal, 2013, http://dx.doi.org/10.1080/03461238.2013.775665, 10 pages.

[7] J. Kočović, P. Šulejić, T. Rakonjac-Antić, Osiguranje, Centar za izdvačku delatnost, Ekonomski fakultet, Beograd, 2010.

[8] B. G. Linsday, "Mixture Models: Theory, Geometry and Application", NSF-CBMS Regional Conference Series in Probability and Statistics, California, 1995.

[9] G. J. McLachlan i D. A. Peel, Finite Mixture Models, Wiley, New York, 2000.

[10] V. Rajić i J. Stanojević, "Mešavina linearnih eksponencijalnih raspodela“, Zbornik radova XXXVII SYMOPIS, 2010, pp. 739-742

[11] W. Woodward, W. Parr, W., R. Schucany i H. Lindsey, "A comparision of minimum distance and maximum likelihood estimation of mixture proportio", J. Amer. Statist. Assoc., 79, 1984, pp. 590-598