МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2015
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2015
Proceedings of the Forty Fourth Spring Conference
of the Union of Bulgarian Mathematicians
SOK "Kamchia", April 2–6, 2015

SOME APPROACHES FOR MODELING CLAIMS PROCESS IN GENERAL INSURANCE*

Elitsa Raeva, Velizar Pavlov

This work presents a brief overview of some proper statistical distributions for modeling of claims in general insurance. Applications of normal approximation, normal power transformation, power transformation and Panjer's recursion for modeling the number and the size of claims have been considered. There are examples for the errors which occur when using these methods. Some comparisons were made.

1. Introduction. In economic theory, there are two main types of risks – speculative and clean. In general insurance the object of study are clean risks. It can be determined as the possibility of loss or damage. The insurer's risks are related to the realization of loss if the premium is less than or equal to the expected payments in the claim process. This process can be divided into two stages. First – an incident which causes the damage and second – a claim that has been evaluated and then paid. It should be borne in mind that not every incident becomes a claim and that the amount of the claim may differ from the price of the damage. In this regard, a special research interest is the modeling of claims by various mathematical methods for analysis and evaluation.

2. Modeling of claims. One of the most important problems in the theory and practice of insurance is related to modeling of the observed empirical statistical distributions of the number and size of claims. For this purpose we usually seek appropriate probability distribution, which can be approximately close to the respectively empirical distribution.

2.1. Modeling of the number of claims. For modeling the number of claims with success the Poisson's distribution and Negative binomial distribution may be used. Poisson distribution has some important characteristics that make it a suitable for theoretical model. It has only one parameter that is to be evaluated. Also, it is additive, i.e. the sum of independent random variables which follow Poisson's distribution is a random variable with a Poisson distribution.

It is important to note that the distribution of the number of claims would be Poisson's, if following conditions are met simultaneously:

– actions occur independently of each other;
– only one claim can occur at a given time;
– the probability for a claim in some sub-period (for example a month) is proportional of the length of that sub-period.

*2010 Mathematics Subject Classification: 62C46, 62G22.
Key words: statistical distribution, models, insurance, claim, risk.

If we ignore the third condition and assume that the Poisson parameter q is a random variable with Gamma distribution, then it can be shown, that the resulting distribution is negative binomial, which also holds useful additive properties [2].

If the expected number of claims is large enough the Central Limit Theorem is applicable and their distribution can be approximated with a normal distribution. We should note that in order to use this approximation it is sufficient to have 50 or more requests available. Otherwise, in homogeneous distribution of the number of claims (coefficient of variation to 30%) as a theoretical model Poisson distribution can be used, while for disposition with a greater degree of variation – negative binomial distribution.

2.2. Modeling the size of claims. Size distribution of the claims can rarely be presented in a simple form. In general, these distributions are: unilateral (negative claims can not occur); highly skewed (have a strong asymmetry); concentrated around the most popular value (moda).

Given these characteristics, for such theoretical models Log-normal distribution, Pareto distribution and Gamma distribution can be used.

Log-normal distribution is logarithmic transformation of the normal distribution. It covers very well the main part of the cost of claims, but too quickly falls at the large values in the tails. At the higher end of the distribution – for example, when assessing the excess of the lost insurance premium it is preferable to use the Pareto distribution. It should be borne in mind that it does not give a sufficient description of the entire field of the data. The Gamma distribution is characterized by flexibility in the center area of the size of the claims, but has a longer tail.

2.3. Methods for modeling the total cost of a random number of claims. Let's consider the distribution of the total value of a set of claims, the condition that their number adopt random values. The resulting stochastic process can be modeled by compounding and convolution of probability distributions. In case of compounding of equally distributed values of the claims we have:

$$g(C) = \sum p(n) f^{n*}(C), \quad \text{or} \quad G(C) = \sum p(n) F^{n*}(C),$$

where

C is the total value of the claims,

$g(C)$ is the probability that the total value of the claims would be exactly C,

$G(C)$ is the probability that the total value of the claims would be less or equal to C,

$p(n)$ is the probability that there will be exactly n claims during the year,

$f^{n*}(C)$ is the probability that if exactly n claims occur during the year, their total value to be exactly C,

$F^{n*}(C)$ is the probability that if exactly n claims occur during the year, their total value to be less or equal to C.

In a similar way a convolution of two discrete probability distributions, like a counting of all possible combinations of exactly n claims whose total value is C gives:

$$f^{n*}(C) = f * f^{(n-1)*}(C) = \sum_{X=0}^{C} f(X) f^{(n-1)*}(C - X),$$

$$F^{n*}(C) = F * F^{(n-1)*}(C) = \sum_{X=0}^{C} F(X) F^{(n-1)*}(C - X),$$

where

X is the value of individual claim,

$f(X)$ is the probability that this claim costs exactly X,

$F(X)$ is the probability that this claim costs X or less.

In case that the claim size distribution is continuous, the convolution of two probability density functions $f(X)$ and $h(X)$ is defined by Stieltjes integration

$$f * h(C) = \int_0^C h(C - X) dF(X).$$

Compounding and convolution of probability distributions can rarely be presented in a convenient form for practical work. There are number of approximations, which lead to a satisfactory numerical results and are generally preferred.

Normal approximation. If the number of claims n is large enough and the distribution of the size of individual claim has final moments, it follows by the Central Limit Theorem:

$$G(C) = \Phi(C;\, n\mu, n\sigma^2), \quad \text{where} \quad \Phi(x; \mu, \sigma^2) = \frac{1}{2\pi\sigma} \int e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

μ is the mathematical expectation for one claim and σ^2 is the dispersion.

Consider the probability that the size of the claims exceeds the value C $(1 - G(C))$. If the shape of the distribution of claims is characterized with positive asymmetry and has for example, Gamma distribution with cumulative distribution function

$$G(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} \int_0^x t^{k-1} e^{\frac{-t}{\theta}} dt, \; k, \theta > 0,$$

then modeling the claim process with normal approximation, errors occur which can be calculated by the formula

$$Error = \left| \frac{1 - G(x)}{1 - \Phi(x)} - 1 \right|.$$

Table 1. Errors in the case of Normal approximation

Skewness	Error							
γ/x	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5
0.30	2e-17	3.6e-11	5.9e-11	9.7e-10	1.6e-10	2.5e-10	6.5e-10	1.6e-09
0.50	5e-05	8.8e-05	1.4e-04	2.3e-04	3.7e-04	5.8e-04	0.001	0.003
0.75	0.007	0.011	0.017	0.024	0.031	0.034	0.022	0.013
1.00	0.039	0.051	0.045	0.019	0.021	0.064	0.133	0.141
2.00	0.123	0.264	0.277	0.147	0.229	1.19	12.57	211.8
3.00	0.398	0.328	0.250	2.9	19.9	185.9	790e+02	10e+07

Some values of these errors for different coefficients of skewness γ and sizes of claims x are presented in Table 1. To calculate the Gamma distribution function the following basic relationships between parameters have been used

$$\gamma = \frac{2}{\sqrt{k}}, \quad k(\gamma) = \frac{4}{\gamma^2}, \quad \mu(\gamma) = k\theta, \quad \sigma^2(\gamma) = k\theta^2.$$

The values of the normal distribution function are calculated with the parameters μ, σ^2, obtained depending on the asymmetry coefficient and by substituting $\theta = 1$. For example, when $\gamma = 2$ and total cost of claims $x = 4$, the error from the approximation in the assessment of insurance premiums is

$$Error = \left| \frac{1 - G(x; \frac{4}{\gamma^2}, 1)}{1 - \Phi(x; \frac{4}{\gamma^2}, \frac{4}{\gamma^2})} \right| = 12.57.$$

Table 1 shows also that with increasing coefficient of skewness and the values of claims, the errors are also increasing significantly. In these cases, some transformations of the values of the random variable could be useful – Normal power transformation and Power transformation [1].

Table 2. Errors in the case of Normal power transformation

Skewness	Error							
γ/x	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5
0.30	2.2e-11	3.6e-11	5.6e-11	8.5e-11	1.2e-10	1.9e-10	3.9e-10	7.7e-10
0.50	5.6e-05	8.8e-05	1.3e-04	1.9e-04	2.8e-04	3.8e-04	6.9e-04	0.001
0.75	0.007	0.011	0.015	0.019	0.018	0.009	0.043	0.147
1.00	0.045	0.051	0.030	0.021	0.010	0.195	0.398	0.581
2.00	0.031	0.264	0.425	0.540	0.624	0.686	0.771	0.824
3.00	0.156	0.328	0.397	0.419	0.410	0.377	0.236	0.030

Normal power transformation. For normal power transformation the results from the first row of the decomposition of Cornish-Fisher are using

$$G\left(x + \frac{\gamma}{6}(x^2 - 1) + \ldots\right) = \Phi(x),$$

then

$$G(x) = \Phi\left(\frac{-3}{\gamma} + \sqrt{1 + \frac{6x}{\gamma} + \frac{9}{\gamma^2}}\right).$$

In these cases the errors from the approximation of Gamma distribution using normal power transformation have the following expression:

$$Error = \left| \frac{1 - G(x; k(\gamma), \theta)}{1 - \Phi\left(\frac{-3}{\gamma} + \sqrt{(1 + \frac{6x}{\gamma} + \frac{9}{\gamma^2}}; \mu(\gamma), \sigma^2(\gamma)\right)} - 1 \right|.$$

Table 2 shows that the speed of growth of the approximation errors with Normal power transformation is not so large as when using an ordinary Normal approximation.

Power transformation. Power transformation is set as follows

$$G(x) = \begin{cases} \Phi(x^h), & \text{if } h \neq 0 \\ \Phi(lnx), & \text{if } h = 0 \end{cases}$$

where the exponent h can be estimated using the method of moments:

$$h = 1 - \frac{\mu_3 \mu}{3\mu_2^2} = 1 - \frac{\gamma}{3\sigma^2}$$

where σ^2 is the dispersion.

For Gamma distribution $h = \frac{1}{3}$. Table 3 shows that even at a total cost of claims

$x = 5$ and $\gamma = 3$ the errors are close to zero. This means that in this case the differences in the functions of the distribution are occurring at a slower rate compared to the Normal power transformation.

It can be seen from the last two tables that the Power transformation gives more satisfactory results in a higher value of the coefficient of skewness than the Normal power transformation, but the differences are not significant in small values of γ. These characteristics may be determined by the shape of the distribution, which has been fitted (in the case of Gamma distribution).

Table 3. Errors in the case of Power transformation for $h = \frac{1}{3}$

Skewness	Error							
γ/x	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5
0.30	2.9e-11	3.6e-11	4.1e-11	4.6e-11	5.1e-11	5.6e-11	6.4e-11	7.3e-11
0.50	7.2e-05	8.8e-05	1.0e-04	1.1e-04	1.3e-04	1.4e-04	1.5e-04	1.1e-04
0.75	0.009	0.011	0.012	0.010	0.003	0.013	0.085	0.207
1.00	0.056	0.051	0.012	0.063	0.165	0.281	0.511	0.697
2.00	0.043	0.264	0.496	0.660	0.772	0.849	0.934	0.972
3.00	0.063	0.328	0.517	0.654	0.753	0.824	0.912	0.956

Panjer's recursion. When distribution of the individual claims is known, Panjer (see [4], [5]) has developed a recursive method, for mixing of Negative binomial with Poisson distribution. The joint probabilities can be calculated as follows

$$g(C) = \sum_{X=1}^{C} \left(a + \frac{bX}{C} \right) f(X) g(C - X),$$

where

$g(C)$ is the probability that the total value of the claims is C;

$f(X)$ is the probability that (possibly to be approximated with discrete distribution) the value of individual claim to be X;

a and b are constants, which depend of the claim frequency by the following recursive relation

$$p(n) = \left(a + \frac{b}{n} \right) p(n - 1),$$

where $p(n)$ is the probability that n claims occur. The initial point of the algorithm is $g(0) = p(0)$.

Computer Simulation. Another widespread method, which with the accelerated development of technical resources becomes more reliable and accessible, is the simulation of random variables. When additional factors are involved such as the size of the investment or fluctuations in the market, and the correlations between them, perhaps the only practical solution would be to use newer technologies (see [3], [6]).

3. Conclusion. The methods for modeling the process of the claim, which were described in the article are the basis, which is necessary in the conduct of mathematical analysis of insurance risk, which is the subject of our next work.

REFERENCES

[1] B. Gep, D. Cox. An analysis of transformations. *Journal of the Royal Statistical Society*, **26** (1994), 216–224.
[2] I. Hossack, J. Pollard, B. Zenhwirth. Introductory statistics with applications in general insurance. Cambridge, Cambridge University Press, 1983.
[3] T. Mikosch. Non-Life Insurance Mathematics, Springer, 2009.
[4] A. Oliveri, E. Pitacco. Introduction to Insurance Mathematics, Springer, 2011.
[5] H. Panjer. Recursive evaluation of a family of compound distribution. *ASTIN Bulletin*, **12** (1981), 73–87.
[6] J. Pollard. A handbook of numerical and statistical techniques, Cambridge, Cambridge University Press, 1977.

Elitsa Raeva
e-mail: eraeva@uni-ruse.bg
Velizar Pavlov
e-mail: vpavlov@uni-ruse.bg
Depratment of Applied Mathematics and Statistics
University of Ruse
8, Studentska Str.
7017 Ruse, Bulgaria

НЯКОИ ВЪЗМОЖНОСТИ ЗА МОДЕЛИРАНЕ НА ИСКОВЕТЕ В ОБЩОТО ЗАСТРАХОВАНЕ

Елица Раева, Велизар Павлов

В настоящата работа е направен кратък обзор и са предложени подходящи вероятностни разпределения за моделиране на исковете в общото застраховане. Разгледани са приложения на нормална апроксимация, нормална степенна трансформация, степенна трансформация и рекурсия на Панджер за моделиране на броя и големината на исковете. Представени са примери за размера на допусканите грешки. Направени са сравнения.