Modelling and Simulation in Non-life Insurance

Viera Pacáková

***Abstract*—** Modelling and simulations are necessary and useful in actuarial analysis in non-life insurance mainly in cases where we cannot find a sufficient number of real data. A typical example is the collective risk, where we know only one value for the calendar year. Using goodness of fit tests we can often find distributions of the number and of the amount of claims and Monte Carlo simulations enable us to simulate a sufficient number of total claims amount. Simulated values then we will use in determining the collective risk model, calculating the premium and risk measures. The article presents a theoretical approach of these probability models and simulations, together with examples of applications.

***Keyword*—** Loss distributions, collective risk model, Monte Carlo simulation, premium calculation, risk measures.

I. Introduction

The insurance risk theory is the analysis of the stochastic features of non-life insurance business. The field of risk theory has grown rapidly. There are now many papers and textbooks, which study the foundations of risk processes along strictly theoretical lines. On the other hand there is a need to develop the theories into forms suitable for practical purposes and to demonstrate their application. Modern computer simulation techniques open up a wide field of practical applications for risk theory concepts, without requiring the restrictive assumptions and sophisticated mathematics, of many traditional aspect of insurance risk theory [5, xiii]. The analyses of insurance risks are an important part of the project Solvency II preparing by European Commission.

While the risk assessment of insurance company in connection with its solvency is a rather complex and comprehensible problem, its solution starts with statistical modelling of number and amount of individual claims. Successful solution of these fundamental problems enables solving of curtail problems of insurance such as modelling and simulation of collective risk, premium and reinsurance premium calculation, estimation of probability of ruin etc.

Application of Monte Carlo simulation allows finds the approximate probability model of the collective risk in non-life insurance portfolio. Simulation of the compound distribution function of the aggregate claim amount can be carried out, if the distribution functions of the claim number process and the claim size are assumed given.

Manuscript received May 14, 2011.This work was supported in part by the project GA ČR 402/09/1866 titled Modelling, Simulation and Management of Insurance Risks.

Author is with the Institute of Mathematics, University of Pardubice, Faculty of Economics and Administration, Studentská 95, 532 10 Pardubice, Czech Republic (e-mail: vpacakova@gmail.com).

II. Modelling of Collective Risk

We shall consider a short term insurance contract covering a risk. By a risk we mean either a single policy or a specified group of policies. The random variable S denotes the aggregate claims paid by the insurer in the year in respect of this risk. We are going to construct the models for the random variable S, so called the *collective risk models* as in [1], [3], [6 or [11].]

A first step in the construction of a collective risk model is to write S in terms of the number of claims arising in the year, denoted by the random variable N, and the amount of each individual claim. Let the random variable X_i denote the amount of the i-th claim. Then

$$S = X_1 + X_2 + ... + X_N \tag{1}$$

where $X_1, X_2, ..., X_N$ are independent identically distributed variables, $N, X_1, X_2, ..., X_N$ are mutually independent, and if $N = 0$ than $S = 0$ too.

The problems we will be solving are the derivation of the moments and distribution of S in terms of the moments and distributions of N and the X_i 's.

We will assume that the moments and the distributions of N and X_i 's known with certainty. In practice these would probably be estimated from some relevant data using methods of parameters' estimation and goodness of fit tests.

We shall denote by $G(s)$ distribution function of S and $F(x)$ the distribution function of X_i, so that $G(s) = P(S \le s) = F_S(s)$ and $F(x) = P(X_i \le x)$. The k-th moment of X_i about zero, $k = 1, 2, 3, ...$, will be denoted as $m_k = E(X_i^k)$.

We will use approximate, not exact methods for valuating $G(s)$. For the approximate methods we need to know the moments of S. Basic expressions, known in actuarial literature, for example in [6], [8], [10], [11], [12] [15], we can write as

$$\begin{aligned} E(S) &= E(N)m_1 \\ D(S) &= E(N)\left(m_2 - m_1^2\right) + D(N)\, m_1^2 \\ M_S(z) &= M_N\left(\ln M_X(z)\right) \end{aligned} \tag{2}$$

The distribution of S is an example of a compound distribution. We consider the most important case when N is Poisson with parameter λ. We say that S is compound Poisson distribution with parameters λ and $F(x)$. In this case results (2) we can express in simple forms:

$$E(S) = \lambda m_1 \tag{3}$$

$$D(S) = \lambda\left(m_2 - m_1^2\right) + \lambda m_1^2 = \lambda m_2 \tag{4}$$

$$\mu_3(S) = \lambda m_3 \tag{5}$$

$$\gamma = \frac{\mu_3(S)}{[D(S)]^{\frac{3}{2}}} = \frac{\lambda m_3}{(\lambda m_2)^{\frac{3}{2}}} > 0 \tag{6}$$

The coefficient of skewness γ shows that the distribution of S is positively skewed and for large values of λ the distribution of S is almost symmetric.

We suppose that all we know or can confidently estimate about S are its mean and variance. Bearing in mind the Central Limit Theorem, this suggests assuming S is approximately normally distributed. An important drawback of this approximation may be that normal density is symmetric, i.e. has zero skewness, and has a right hand tail which goes to zero very quickly. For many types of insurance the distribution of S is positively skewed with a fairly heavy right hand tail and so normal approximation will tend to underestimate $P(S > x)$ for large values of x.

Suppose we know or can estimate with reasonable confidence the first three moments of S. One way of avoiding or at least reducing the problem of underestimating tail probabilities is to approximate the distribution of S by a translated gamma distribution. Let μ, σ^2 and γ denote the mean, variance and coefficient of skewness of S. We assume S has approximately the same distribution as the random variable $k+Y$, where k is a constant and Y has a gamma distribution $G(\alpha, \beta)$. The parameters k, α and β are chosen so that $k+Y$ has the same first three moments as S.

Equating the coefficients of skewness, variance and means of S with the same characteristics of $k+Y$ gives the following three formulae

$$\gamma = \frac{2}{\sqrt{\alpha}} \qquad \sigma^2 = \frac{\alpha}{\beta^2} \qquad \mu = k + \frac{\alpha}{\beta} \tag{7}$$

from which α, β and k can be calculated.

III. Monte Carlo Simulation of Collective Risk

The Monte Carlo simulation of the values of S, presented for example in [8], [12] and [13], consists of the following steps:

1. Generate the number of claims n_1 from the known distribution of variable N (Poisson, negative binomial) using the random number generator.
2. Generate from the known distribution of the individual claim amount X just n_1 values of the individual losses $x_1, x_2, ..., x_{n_1}$.
3. The sum $s_1 = x_1 + x_2 + ... + x_{n_1}$ gives the first random number s_1 of the aggregate claim amount (collective risk) S.
4. The steps 1 to 3 repeat n-times to get generated random numbers $s_1, s_2, ..., s_n$ from unknown distribution of S.

Simulated values $s_1, s_2, ..., s_n$ enable us to solve two important tasks:

1. To verify suitability of probability models of S, those have been found by other actuarial methods.
2. To find the probability model of S by application of Goodness of Fit Tests using sampling data generated by Monte Carlo simulation procedure.

IV. Using Collective Risk Model

A. Premium Calculations

Given a risk S, we refer its expected value $E(S)$ as the *pure premium*. An insurance company must charge more than the pure premium to cover expenses, allow for variability in the number and amount of claims, and make a profit [3, p.99]. In a simple model for determining premiums, assume that we use a loading (safety or security) factor θ, whereby the risk premium RP is of the form

$$RP = (1+\theta)E(S) \tag{8}$$

Basic of premium calculation as we know of the collective risk S probability model is equitation

$$P(S \le s_{0,95}) = P(S \le (1+\theta)E(S)) = 0{,}95 \tag{9}$$

from which we can calculate loading factor θ as

$$\theta = \frac{s_{0,95} - E(S)}{E(S)} \tag{10}$$

where $s_{0,95}$ is 95$^{\text{th}}$ percentile of distribution of S, defined as value for which is valid $G(s_{0,95}) = F_S(s_{0,95}) = 0{,}95$.

B. Determining the Value of Risk

A risk measure, which summarizes the overall risk exposures of the company, helps the company evaluate if there is sufficient capital to overcome adverse events. Risk measures for blocks of policies can also be used to assess the adequacy of the premium charged [15, p. 115].

Value at Risk (*VaR*) is probably one of the most widely used measures of risk in financial sector. Simply speaking, the *VaR* of a loss variable X is the minimum value of the distribution such that the probability of the loss larger than this value is not more than a given probability [15, p. 120-121].

Let X be a continuous random variable with distribution function $F_X(x)$ (*df*) and probability density function (*pdf*)

ISBN: 978-1-61804-016-9

$f_X(x)$ defined as $f_X(x) = \frac{dF_X(x)}{dx}$. The quantile function (*qf*) is the inverse of the *df*. Thus if $F_X(x_\delta) = \delta$ then $x_\delta = F_X^{-1}(\delta)$.

In statistical terms, *VaR* at probability level δ, denoted by $VaR_\delta(X)$, is a quantile as defined formally as follows

$$VaR_\delta(X) = F_X^{-1}(\delta) = x_\delta \qquad (11)$$

The quantile x_δ indicates the loss which will be exceeded with probability 1-δ, but does not provide information about how bad the loss might be if loss exceeds this threshold. To address this issue, we may compute the conditional tail expectation (*CTE*) with tolerance 1-δ, which is defined as in (15, p. 123)

$$CTE_\delta(X) = E\left(X \middle| X > VaR_\delta(X)\right) \qquad (12)$$

We put the risk premium *RP* equal to the quantile $S_{1-\alpha} = S_{0,95}$ and we will define loss function *X* as $X = S - RP$. We will calculate VaR_δ for this loss function in case that $S > RP$.

According to the above assumptions we can write:

$$P(S - RP < VaR_\delta | S > RP) = \delta$$
$$P(S < RP + VaR_\delta | S > RP) = \delta$$

The above equation can be written as

$$\frac{P(RP < S < RP + VaR_\delta)}{P(S > RP)} = \delta$$

Using *df* of *S* it can be expressed as follows:

$$\frac{F_S(RP + VaR_\delta) - (1-\alpha)}{\alpha} = \delta \qquad (13)$$

Results from (13):

$$F_S(RP + VaR) = \alpha \cdot \delta + (1-\alpha) = \vartheta$$

where we have put

$$\vartheta = \alpha \cdot \delta + (1-\alpha) \qquad (14)$$

Then

$$P(S < RP + VaR_\delta) = \vartheta$$
$$RP + VaR_\delta = S_\vartheta$$
$$VaR_\delta = S_\vartheta - RP \qquad (15)$$

V. Example of Application

Suppose that the number of claims *N* incurred in time period of one year follows a Poisson distribution with parameter λ=10 000. We know the values of 91 individual claims made on an insurance portfolio. We will assume that these individual claim amounts are drawn from a particular distribution, called a loss distribution. Using maximum likelihood estimation and goodness of fit tests in statistical analytical system Statgraphics Centurion XV we have verified that lognormal distribution with parameters μ=9,741 and σ^2=2,165 give a very good fit to the empirical data of individual claims amounts.

The first three moments of lognormal distribution can be calculated by the formula

$$E\left(X^k\right) = e^{k\mu + \frac{\sigma^2}{2}k^2}$$

Then $m_1 = 50\,171{,}2;\ m_2 = 2193613;\ m_3 = 8{,}35827 \cdot 10^{16}$

Using formulas (3) to (6) we have calculated the descriptive measures of the collective risk *S*:

$E(S)$ = 501 712 000, $D(S)$ = 2,19361E+14, $\gamma(S)$ = 0,257.

Except of the normal approximation of *S* with parameters μ = 501 712 00 and σ^2 = 2,19361E+14 we can approximate the distribution of *S* by a translated gamma with parameters α = 60,4376; β = 5,24896E-07 and k = 386 570 080,3, calculated by (7).

Using the own computer program of Monte Carlo simulation in SAS system we have generated 10 000 values of aggregate claim amount *S*.

We use these simulated values of *S* to verify the suitability of translated gamma distribution with above calculated parameters by goodness of fit tests in SAS Enterprise Guide 3.0. Because of *p*-value > 0, 05, all three tests in these system confirm the translated gamma distribution gives a good fit to collective risk model of *S*.

Using sampling data generated by Monte Carlo simulation we have found 3-parameters lognormal distribution using *Distribution Fitting* procedure of statistical system Statgraphics Centurion XV (Table 1) with parameters estimated by maximum likelihood method (Table 2) as a model that give a good fit to simulated data of *S*.

TABLE 1
RESULTS OF KOLMOGOROV-SMIRNOV TEST

	Lognormal (3-Parameter)
DPLUS	0,00589260
DMINUS	0,00776702
DN	0,00776702
P-Value	0,58247

Source: Own calculation, Statgraphics Centurion XV

TABLE 2
PARAMETERS OF FITTED DISTRIBUTIONS

Lognormal (3-Parameter)
mean = 5,01935E8
standard deviation = 1,47275E7
lower threshold = 3,02948E8

Source: Own calculation, Statgraphics Centurion XV

Histogram of the simulated values of *S* and *pdf* of lognormal distribution with parameters from Table 2 we can see at Fig 1.

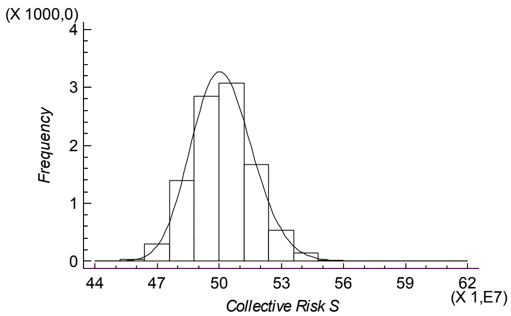

Fig 1 Histogram and fitted 3-parameters lognormal model

Procedure *Critical values* of Statgraphics Centurion XV allows finding quantiles S_p of lognormal distribution of S for any probability p. Table 3 includes selected quantiles that we will use in the following.

TABLA 3
QUANTILES FOR S

Lower Tail Area (<=)	*Lognormal (3-Parameter)*
0,95	5,2703E8
0,995	5,42993E8
0,99975	5,59599E8

Source: Own calculation, Statgraphics Centurion XV

If we put the risk premium RP equal to the quantile $S_{0,95}$=5,2703E8 (Table 3) and we define the loss function X as $X=S-RP$, we can calculate $VaR_\delta = VaR_{0,995}$ for this loss function in case that $S > RP$.

According to (14) we can calculate value:

$$\vartheta = \alpha \cdot \delta + (1-\alpha) = 0{,}05 \cdot 0{,}995 + 0{,}95 = 0{,}99975$$

Using Statgraphics Centurion XV, procedure *Critical value* (Table 3) refer to (15) we get result

$$VaR_\delta = VaR_{0,995} = S_{0,99975} - RP =$$
$$= 5{,}59599E8 - 5{,}2703E8 = 32\ 569\ 000$$

The minimum value that exceed the difference between aggregate claim amount S and risk premium RP with probability 0,995 (if collective risk S is greater than RP) is 32 569 000 monetary units.

Using procedure *Distribution Fitting* of statistical package Statgraphics Centurion XV we can find 2-parameters Pareto distribution that is fit on simulated values of collective risk S which are higher than $RP = S_{0,95} = 5{,}2703E8$.

Probability density function of this distribution has the form

$$f(x) = ba^B x^{-b-1}, \quad b > 0,\ a \geq 0 \tag{16}$$

Mean value is expressed as follows:

$$E(X_a) = \frac{a \cdot b}{b-1}, \qquad b>1 \tag{17}$$

Table 4 includes the parameters of the Pareto distribution with density function (16) estimated by maximum likelihood method and table 5 includes the results of Kolmogorov-Smirnov test, that verified good fit with Pareto probability model.

TABLE 4
MAXIMUM LIKELIHOOD PARAMETER ESTIMATION

Pareto (2-Parameter)
shape = 74,8166
lower threshold = 5,27027E8

Source: Own calculation, Statgraphics Centurion XV

TABLE 5
RESULTS OF KOLMOGOROV-SMIRNOV TEST

	Pareto (2-Parameter)
DPLUS	0,0143693
DMINUS	0,0278264
DN	0,0278264
P-Value	0,837141

Source: Own calculation, Statgraphics Centurion XV

If we denote the shape of 2-paremetres Pareto model as b and the lower threshold as a, we can calculate the mean of the values of collective risk S that exceed of the RP refer to (17) as follow:

$$E(S|S > RP) = \frac{c \cdot \theta}{c-1} = 534\,166\,681{,}3$$

REFERENCES

[1] R. E. Beard, *Risk Theory, the Stochastic Basis of Insurance*. London: Chapman and Hall, 1994.

[2] R. E. Beard, T. Pentikäinen, E. Pesonen, *Risk Theory*. Cambridge: Chapman and Hall, 1990.

[3] P. J. Boland, *Statistical and Probabilistic Methods in Actuarial Science*. London: Chapman&Hall/CRC, 2007.

[4] I. D. Currie, *Loss Distributions*. London and Edinburgh: Institute of Actuaries and Faculty of Actuaries, 1993.

[5] C. D. Daykin, T. Pentikäinen, M. Pesonen, *Practical Risk Theory for Actuaries*. London: Chapman and Hall, 1996.

[6] D. C. M. Dickson, H. R. Waters, *Risk models*. London and Edinburgh: Institute of Actuaries and Faculty of Actuaries, 1993.

[7] R. V. Hogg, S. A. Klugman, *Loss Distributions*. New York: John Wiley & Sons, 1984.

[8] G. Horáková, V. Mucha, *Teória rizika v poistení* (Risk Theory in Insurance). Bratislava: Vyda-vateľstvo Ekonóm, 2006.

[9] R. Kaas, M. Goovaerts, J. Dhaene, M. Denuit, *Modern Actuarial Risk Theory*. Boston: Kluwer Academic Publishers, 2001.

[10] B. Linda, *Stochastické modely operačního výzkumu* (Stochastic Models of Operation Research). Bratislava: STATIS, 2004.

[11] V. Pacáková, Aplikovaná poistná štatistika (Applied Insurance Statistics). Bratislava: Iura Edition, 2004.

[12] V. Pacáková, "Modelovanie a simulácia rizík v neživotnom poistení," *E+M – Ekonomie a Management*, Technická univerzita v Liberci, 3/2007, ročník X, pp. 109-121.

[13] R. Y. Rubinstein, *Simulation and the Monte Carlo Method*. New York: John Wiley &Sons, Inc., 1981.

[14] Ľ. Sipková, E. Sodomová, *Modelovanie kvantilovými funkciami* (Modelling with quantile function). Bratislava: Vydavateľstvo Ekonóm, 2007.

[15] Y. K. TSE, *Nonlife Actuarial Models*. Cambridge: University Press, 2009.

Prof. RNDr. Viera Pacáková, PhD. graduated in Econometrics (1970) at Comenius University Bratislava, 1978 - RNDr. in Probability and Mathematical Statistics, degree CSc. (PhD) at University of Economics (EU) Bratislava in 1986, assoc prof. in Quantitative Methods in Economics in 1998 and professor in Econometrics and Operation Research in 2006 at University of Economics in Bratislava. She was working at Department of Statistics, Faculty of Economic Informatics, EU Bratislava since 1970 to January 2011 and she has been working at Faculty of economics and administration in Pardubice since 2005. She has been concentrated on actuarial science since 1994 in connection with the actuarial education in the Slovak republic and she has been achieved considerable results in this area.