The compound Poisson random variable's approximation to the individual risk model

Jingping Yang
Department of Financial Mathematics, Peking University

Joint work with Prof. Shulin Zhou of Peking Univeristy
and Mr. Zhenyong Zhang of Peking University

1. Introduction

The advantage of Compound Poisson model:

(1). Recursive calculation (Panjer (1981))

(2). Combination and decomposition (Panjer and Willmot (1992, Chapter 6) or Kaas, et al (2001, Chapter 3)).

(3). The approximation to the individual risk models in distribution by compound Poisson models.

Some references:

(1) Bühlmann et al (1977) illustrated why a cautious insurer should prefer the compound Poisson model to the individual risk model in the sense of stop-loss order.

(2) Gerber (1979, Chapter 4) gave a description of the choice of the Poisson parameter and introduced two cases which are often used in the later discussions.

(3) Gerber (1984), Hipp (1985), Hipp (1986), Michel (1987), De Pril and Dhaene (1992), Sundt (1993), Dhaene and Sundt (1997) investigated the error bounds for approximation in terms of distribution or stop-loss premium.

(4) Kaas, Van Heerwaarden and Goovaerts (1988b) discussed the approximation of the aggregate claims and the stop-loss premiums by approximating the aggregate claims by the sum of a compound Poisson random variable (r.v.) and another r.v. determined by stop-loss order.

(5) Kuon, Radtke and Reich (1993) studied the approximation quality when the portfolio keeps growing.

The former papers are mainly focused on approximation in the aggregate claims distribution and related functions, such as stop-loss premiums.

Motivation of our paper:

Given the observation data from the individual risk model, how should one determine the r.v.'s in the corresponding compound Poisson model?

The aim of this paper is to develop a method for carrying out such an approximation.

The outline of our paper

(1). By minimizing the expectation of the absolute deviation of compound Poisson r.v.'s from the total loss associated with the individual risk model, we present an optimal approximation model.

(2). We also give a numerical method to evaluate the approximation error.

(3). Finally we discuss the influence of the Poisson parameter on the approximation error.

(4). It is assumed that the individual risks are independent. We first consider the case that the individual risks are homogenous, then apply the homogenous results to approximate the heterogenous risk model.

2. Our approximation prinple

Consider a portfolio containing n homogenous insurance risks. Let X_i denote the loss associated with the i-th risk, $i = 1, 2, \cdots, n$. Assume that $X_1, X_2, \cdots, X_n$ are independent identically distributed with common distribution F, where the variance $Var(X_1)$ is finite and $0 < F(0) < 1$.The number of claims for the portfolio is denoted as N_n, i.e.,

$$N_n = \#\{i : X_i > 0, i \leq n\}.$$

The total claim amount equals

$$S_n = \sum_{i=1}^{n} X_i.$$

An auxiliary sequence of risks $\{X_{n+1}, X_{n+2}, \cdots, \}$ is introduced, where $X_1, X_2, \cdots$ are independent identically distributed.

Denote

$$M_1 = \inf\{i \geq 1 : X_i > 0\};$$

$$M_n = \inf\{i > M_{n-1} : X_i > 0\}, \quad n \geq 2.$$

And we define a claim sequence

$$Y_i = X_{M_i}, \quad i \geq 1$$

which is a subsequence of $\{X_i, i \geq 1\}$.

The total claim amount $S_n = \sum_{i=1}^{n} X_i$ can be expressed as $S_n = \sum_{i=1}^{N_n} Y_i$.

Note that

(1). $Y_i, i \geq 1$ are independent and identically distributed (Li and Yang (2001));

(2). N_n is the number of claims in the portfolio $\{X_1, X_2, \cdots, X_n\}$;

(3). The independence between N_n and Y_i is fulfilled (Li and Yang (2001)).

(4). Summing over the first N_n claims of the sequence $Y_i, i \geq 1$, we thus obtain the total loss S_n associated with the portfolio $\{X_1, \cdots, X_n\}$.

Our approximation methodology:

(1). It is assumed that the approximation model has the same claim sequence as the individual risk model.

(2). The number of claims $N(\theta)$, as a Poisson r.v., should be estimated from the observation data.

(3). The independence between $N(\theta)$ and Y_i, $i \geq 1$ is assumed.

An approximation to S_n is defined by

$$S^* = \sum_{i=1}^{N(\theta)} Y_i.$$

Let $F_{poi(\theta)}$ denote Poisson distribution with mean θ, and $F_{bin(n,q)}$ denote binomial distribution with parameters (n, q) where $q = 1 - F(0)$.

The family of all Poisson r.v.'s, which have common mean θ and are independent of the claim sequence Y_i, $i \geq 1$, is denoted as $R(F_{poi(\theta)})$. The optimal Poisson r.v. $N_n^0(\theta) \in R(F_{poi(\theta)})$ is determined by the following minimizing principle:

$$H_n(\theta) =: E|S_n - \sum_{i=1}^{N_n^0(\theta)} Y_i| = \inf_{N(\theta) \in R(F_{poi(\theta)})} E|S_n - S^*|,$$

where $S^* = \sum_{i=1}^{N(\theta)} Y_i$.

3. The existence of the optimal Poisson r.v.

Given the risk sequence $\{X_1, X_2, \cdots\}$, a r.v. U can be defined as below:

(a)Construct r.v.'s $U_m, m = 0, 1, \cdots, n$ such that $U_m, m = 0, 1, \cdots, n$ and $X_i, i \geq 1$ are independent, and U_m is uniformly distributed over $(F_{bin(n,q)}(m-1)), F_{bin(n,q)}(m)]$ with probability density function $\frac{1}{F_{bin(n,q)}(m)-F_{bin(n,q)}(m-1)}$.

(b)Define

$$U = \sum_{m=0}^{n} U_m I_{\{N_n=m\}}.$$

The inverse function $F^{-1}_{bin(n,q)}$ of $F_{bin(n,q)}$ is defined as

$$F^{-1}_{bin(n,q)}(y) = \inf\{x : F_{bin(n,q)}(x) \geq y\}, y \in [0, 1]$$

The inverse function $F^{-1}_{poi(\theta)}$ of $F_{poi(\theta)}$ is defined as

$$F^{-1}_{poi(\theta)}(y) = \inf\{x : F_{poi(\theta)}(x) \geq y\}.$$

Then it can be proved that the r.v. U is uniformly distributed over [0, 1], and

$$N_n = F^{-1}_{bin(n,q)}(U).$$

Theorem 1 $F^{-1}_{poi(\theta)}(U)$ is a possible choice for the optimal Poisson r.v. with mean θ. Moreover, the approximation error $H_n(\theta)$ satisfies

$$\begin{aligned} H_n(\theta) &= E(Y_1)E|F^{-1}_{bin(n,q)}(U) - F^{-1}_{poi(\theta)}(U)| \\ &= E(Y_1)\int_0^\infty |F_{bin(n,q)}(x) - F_{poi(\theta)}(x)|dx. \end{aligned}$$

In this paper we choose the optimal Poisson r.v. as

$$N^0_n(\theta) = F^{-1}_{poi(\theta)}(U).$$

The two risks P and Q are comonotonic if there exist two nondecreasing real-valued functions u, v and a risk Z such that

$$P = u(Z), \quad Q = v(Z)$$

(see Wang, Young and Panjer(1997)).

$N_n^0(\theta)$ and N_n are both nondecreasing functions of the r.v. U. Thus they are comonotonic.

4. Some results on $N_n^0(\theta)$

The joint distribution $F_{N_n,N(\theta)}$ of $(N_n, N(\theta))$ satisfies

$$\max\{F_{bin(n,q)}(x) + F_{poi(\theta)}(y) - 1, 0\} \le F_{N_n,N(\theta)}(x, y) \le \min\{F_{bin(n,q)}(x), F_{poi(\theta)}(y)\}.$$

Theorem 2 The r.v. $N_n^0(\theta)$ is optimal in the following senses:

$$E(N_n - N_n^0(\theta))^2 = \inf_{N(\theta)\in R(F_{poi(\theta)})} E(N_n - N(\theta))^2$$

and

$$Var(N_n - N_n^0(\theta)) = \inf_{N(\theta)\in R(F_{poi(\theta)})} Var(N_n - N(\theta)).$$

5. Evaluating the approximation error

Theorem 3

(a) If $\theta \geq -n\log(1-q)$, then

$$H_n(\theta) = (\theta - nq)EY_1.$$

(b) If $0 \leq \theta < -n\log(1-q)$, then

$$H_n(\theta) = \{2\sum_{i=0}^{n}\sum_{j=0}^{i}(i-j) \times$$

$$P(N_n = i, N_n^0(\theta) = j) + \theta - nq\}EY_1.$$

The Poisson parameter θ is often chosen as $\theta = nq$ or $\theta = -n\log(1-q)$ (Gerber (1979, Chapter 4)).

Table 2-Table 4 provide some numerical values of $H_n(nq)$ and $H_n(-n\log(1-q))$ when $EY_1 = 1$.

Table 2–The expected errors when $n = 10$

q	$H_n(nq)$	$H_n(-n\log(1-q))$
0.001	0.000010	0.000005
0.005	0.000239	0.000125
0.01	0.000911	0.000503
0.05	0.015587	0.012933
0.1	0.038402	0.053605
0.5	0.524205	1.931472

6. Approximation to the total loss and related functions

Consider the total amount of the type $\sum_{i=1}^{n} g(X_i)$, where g is a non-negative measurable function and $g(0) = 0$. An approximation for $\sum_{i=1}^{n} g(X_i)$ is

$$\sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

The corresponding approximation error equals

$$h_n(\theta, g) =: \sum_{i=1}^{n} g(X_i) - \sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

Theorem 4 For the non-negative function g with $g(0) = 0$,

$$Eh_n(\theta, g) = (nq - \theta)Eg(Y_1)$$

and

$$E|h_n(\theta, g)| = \frac{Eg(Y_1)}{EY_1} H_n(\theta).$$

One interesting fact when $\theta = -nlog(1-q)$.

$$\sum_{i=1}^{n} g(X_i) = \sum_{i=1}^{N_n} g(Y_i) \leq \sum_{i=1}^{N_n^0(\theta)} g(Y_i).$$

7. The uniqueness of the Poisson parameter to minimizing $H_n(\theta)$

If there exists a unique θ_n^0, such that

$$H_n(\theta_n^0) = \min_{\theta \geq 0} H_n(\theta).$$

For $k = 0, 1, 2, \cdots, n-1$, denote $\theta^{(k)}$ to be the solution of equation

$$F_{bin(n,q)}(k) = F_{poi(\theta^{(k)})}(k).$$

Then we have the following result.

Theorem 5 It holds that

$$\theta^{(n-1)} < \theta^{(n-2)} < \cdots < \theta^{(1)} < \theta^{(0)}.$$

Theorem 6 There exists a unique θ_n^0 with $0 \leq \theta_n^0 \leq -n\log(1-q)$.

Table 5(a)–The optimal θ_n^0

n	q	nq	$-n\log(1-q)$	θ_n^0
2	0.001	0.002	0.002001	0.002001
2	0.01	0.02	0.020101	0.020101
2	0.10	0.2	0.210721	0.210721
2	0.5	1	1.386294	0.961278
10	0.001	0.01	0.010005	0.010005
10	0.01	0.1	0.100503	0.100503
10	0.10	1	1.053605	0.99907
10	0.5	5	6.931472	4.95961
100	0.001	0.1	0.100050	0.100050
100	0.01	1	1.005034	1.000
100	0.1	10	10.536051	9.9991
100	0.5	50	69.314718	49.959
1000	0.001	1	1.0005	1.000
1000	0.01	10	10.050336	10.00

Table 5(b)–The corresponding error $H_n(\theta_n^0)$

n	q	$H_n(nq)$	$H_n(-n\log(1-q))$	$H_n(\theta_n^0)$
2	0.001	1.997E-06	1.001E-06	1.001E-06
2	0.01	0.000197	0.000101	0.000101
2	0.10	0.017462	0.010721	0.010721
2	0.5	0.235797	0.386294	0.226086
10	0.001	0.00001	0.000005	0.000005
10	0.01	0.000911	0.000503	0.000503
10	0.10	0.038402	0.053605	0.038161
10	0.5	0.524205	1.931472	0.519686
100	0.001	0.000091	0.000050	0.000050
100	0.01	0.003694	0.005034	0.003694
100	0.1	0.128624	0.536051	0.128548
100	0.5	1.653039	19.314718	1.651654
1000	0.001	0.000368	0.000500	0.000368
1000	0.01	0.012545	0.050336	0.012545

8. Approximation to the heterogenous individual risk model

In practice, the individual risks $X_i, i \leq n$ of a portfolio are often independent, but not identically distributed.

We can divide the heterogenous portfolio into several independent homogenous portfolios, then approximate every homogenous portfolio separately by our method.

9. Conclusions

(1). We presented a new method to approximate the individual risk model by a compound Poisson r.v. .

(2). We discussed the calculation of the approximation error.

(3). We first focused on the homogenous individual risk models, then applied the results to the heterogenous individual risk models.

(4). Some numerical results are given.