AN ESTIMATION OF CLAIMS DISTRIBUTION

Nawojiro Eshita

Foreword

There are two phases of difficulties in estimating a claims distribution. If we are going to estimate the claims distribution as accurately as possible, we should gather considerably long terms statistics. While economical and social environment will change. As a result the statistics gathered should be amended by a kind of trend value. One of difficulties here is the estimation of that trend value. Another difficulty is the estimation of claims distribution as being the stochastic distribution. In the case of considering claim size, the estimation becomes more difficult.

The intention of this paper is to propose an actual way of estimating stochastic claims distribution considering various kind of claim size by the use of a computer. Regarding the problem of amending claims distribution by a trend Value I will discuss at another time.

1. The Model of Claims Distribution

(1) *The logical claims distribution*

The claims distribution is the distribution of claim amount which a insurer paid for a definite period, for example, for one year.

Therefore, the claims distribution should be analized by two factors.

The distribution of claims frequency.

The distribution of claim size.

a. *The distribution of claims frequency*

Since the claims frequency is a number which is calculated from the stand point whether or not claim occur in the risk group which an insurer is retaining in a definite term, the logical distribution of claims frequency is considered to be a binomial distribution.

Assuming that numbers of risks and the average claims occurrence rate in a risk group are n and p respectively, the distribution is expressed by the following binomial expansion formula.

$$b(k; n \cdot p) = (nk) p^k \cdot q^{n-k}$$

where

$$q = 1 - p$$

k is probable claims number occurring.

The probability of claims occurrence in each risk in a risk collective which a insurer is retaining is not always the same. Accordingly an risk collective is separated into many kinds of risk groups with different number of risks and claims occurrence rate. Assume the number of risks and claim occurrence rate of m number of risk groups are $n_1, n_2, \ldots, n_m$ and $p_1, p_2, \ldots, p_m$ respectively the following formula holds.

$$p = \frac{n_1 p_1 + n_2 p_2 + \ldots + n_m p_m}{N}$$

The distribution of claims frequency of a risk collective is accordingly expressed by the following formula

$$\binom{n_1}{k} p_1^k q_1^{n_1-k} * \binom{n_2}{k} p_2^k q_2^{n_2-k} * \ldots * \binom{n_m}{k} p_m^k q_m^{n_m-k}$$

where, $*$ show convolution.

b. The distribution with claim size distribution

In considering claim size, the distribution becomes more and more complex. The calculation is almost unrealistic, even if a computer is used. Assuming that a risk collective is constructed by risks being $p_1, p_2, \ldots, p_m$ of claim occurrence rate and $s_1, s_2, s_3, \ldots, s_l$ of claim size and $n_{11}, n_{12}, n_{13}, \ldots, n_{ml}$ of risk number which are the case of $p_1 s_1, p_2 s_2, p_3 s_3, \ldots, p_m s_l$ respectively, the risk collective is expressed by the array of following risk groups.

$$R_{11}[n_{11}, p_1, s_1], \quad R_{12}[n_{12}, p_1, s_2] \ldots \ldots R_{1l}[n_{1l}, p_1, s_l]$$
$$R_{21}[n_{21}, p_2, s_1], \quad R_{22}[n_{22}, p_2, s_2] \ldots \ldots R_{2l}[n_{2l}, p_2, s_l]$$
$$\cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot$$
$$\cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot$$
$$\cdot \qquad\qquad\qquad \cdot \qquad\qquad\qquad \cdot$$
$$R_{m1}[n_{m1}, p_m, s_1], \; R_{m2}[n_{m2}, p_m, s_2] \ldots \ldots R_{ml}[n_{ml}, p_m, s_l].$$

In the case of each risk being independent stochastically, the claims distributions of each group are shown by the following formula.

$$b_{11}(s_1 k;\, n_{11},\, p_1) \;=\; \binom{n_{11}}{k}\, p_1^k \cdot q_1^{n_{11}-k}$$

$$b_{12}(s_2 k;\, n_{12},\, p_1) \;=\; \binom{n_{12}}{k}\, p_1^k \cdot q_1^{n_{12}-k}$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$b_{ml}(s_l k;\, n_{ml},\, p_m) \;=\; \binom{n_{ml}}{k}\, p_m^k \cdot q_m^{n_{ml}-k}$$

Since a claims distribution of a risk collective is a compound function of claims distribution of these risk groups, while compound function of binomial distributions is not always a binomial distribution, a claims distribution of a risk collective is not always a binomial distribution. Accordingly we describe the distribution function of a risk collective as $\varrho(sk;\, n,\, p)$.

$$\varrho(sk;\, n,\, p) = b_{11}(s_1 k;\, n_{11},\, p_1) * b_{12}(s_2 k;\, n_{12},\, p_1) * \ldots$$
$$\ldots * b_{ml}(s_l k;\, n_{ml},\, p_m).$$

This is the logical model for claims distribution which may occur in a risk collective.

(2) *The actual claims distribution*

The number of actual risks in a risk group is relatively large (for example more than 50) and the occurrence rate of claim is relatively small (for example less than 0.1). The previous logical distribution (which is the compound function of binomial distributions) of risk groups is, accordingly, replaced by the Poisson distribution. And the previous array of logical risk groups is replaced by the following array of claims distribution.

$$D_{11}[m_{11},\, s_1], \quad D_{12}[m_{12},\, s_2] \quad \ldots D_{1l}[m_{1l},\, s_l]$$
$$D_{21}[m_{21},\, s_1], \quad D_{22}[m_{22},\, s_2] \quad \ldots D_{2l}[m_{2l},\, s_l]$$
$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$
$$D_{m1}[m_{m1},\, s_1],\, D_{m2}[m_{m2},\, s_2] \ldots D_{ml}[m_{ml},\, s_l]$$

8

where

$$m_{11} = n_{11} \times p_1, \ m_{21} = n_{21} \times p_2, \ \ldots \ m_{ml} = n_{ml} \times p_m.$$

In the Poisson distribution, the compound distribution between the distribution with the value of m_1 and the value of m_2 is the same to the Poisson distribution with the value of $m_1 + m_2$. As a result, similar risk groups with a same claim size are totalled up to one. And the previous array of claim distribution is expressed by the more simple array as follows.

$$D_1[m_1, s_1], \ D_2[m_2, s_2], \ \ldots \ D_l[m_l, s_l]$$

where

$$m_1 = m_{11} + m_{21} + \ldots + m_{m1}$$
$$m_2 = m_{21} + m_{22} + \ldots + m_{m2}$$
$$\cdot \qquad \cdot$$
$$\cdot \qquad \cdot$$
$$\cdot \qquad \cdot$$
$$m_l = m_{1l} + m_{2l} + \ldots + m_{ml}.$$

Although a compound poisson distribution of some Poisson distributions is a compound Poisson distribution, the compound function of actual claims distributions is not the Poisson distribution, because of another element $(s_1, s_2, s_3, \ldots, s_l)$ are contained. Assume the actual claims distribution of a risk collective to be described as $f(sk; m)$

$$f(sk; m) = p_1(s_1 k; m_1) * p_2(s_2 k; m_2) * \ldots p_l(s_l k; m_l).$$

This is the actual model for a claims distribution which may occur in a risk collective.

2. THE ACTUAL ESTIMATION OF CLAIMS DISTRIBUTION

(1) *Simplification of model*

Since a risk collective contains varying risk groups, the actual claim distribution is constructed by various kinds of claims distributions. Assuming that claim amount occurring in a risk collective distribute from $ 10,000 to $ 1,000,000, the claims distribution may have 100 various claims distribution and therefore, we may have 100 different calculations. Calculations, however, need not be so multitudinous, for example,

a. Some distributions with approximately the same number of claims are totalled up to a single distribution.

Assume, for example, the numbers of claims with size s_1, s_6, s_{15}, s_{21}, s_{55} are m_1, m_6, m_{15}, m_{21}, m_{55} respectively and then

$$m_1 = m_6 = m_{15} = m_{21} = m_{55}.$$

The distribution is replaced by the next simplified formula.

$$p_0(s_0 k; m_0) = e^{-m_0} \frac{m_0^k}{k!}$$

where

$$s_0 = \frac{m_1 s_1 + m_6 s_6 + m_{15} s_{15} + m_{21} s_{21} + m_{55} s_{55}}{m_1 + m_6 + m_{15} + m_{21} + m_{55}}$$

$$m_0 = m_1 + m_6 + m_{15} + m_{21} + m_{55}.$$

b. Some distributions with approximately the same size are also totalled up to a single distribution. Assume, for example, the claim size of claim numbers m_1, m_6, m_{15}, m_{21}, m_{55} are s_1, s_6, s_{15}, s_{21}, s_{55} respectively and then

$$s_1 = s_6 = s_{15} = s_{21} = s_{55} = s_0$$

$$m_1 + m_6 + m_{15} + m_{21} + m_{55} = m_0.$$

The distribution is replaced by the next simplified formula.

$$p_0(s_0 k; m_0) = e^{-m_0} \frac{m_0^k}{k!}$$

c. Especially when the numbers of claim is very large, the difference of size could be ignored in the actual calculation. And then calculations need not be so multitudinous.

(2) *The actual calculation of compound Poisson distribution*

The following is the calculation flow of the distribution by the use of a computer. By this flow we can easily calculate the distribution.

a. The calculation or table research of $p_1(s_1 k; m_1)$. The number of m_1 is not so large that the actual calculation or table research of $p_1(s_1 k; m_1)$ should not be difficult. At this time, we ignore the value of probability which is insignificant and therefore not pertinent.

b. The calculation or table research of $p_2(s_2k; m_2)$. The insignificant value is deleted as the previous step.

c. The calculation of compound function of $p_1(s_1k; m_1)$ and $p_2(s_2k; m_2)$.

Assume the claim size and the occurrence numbers of claims which should be convoluted as follows.

$(s_1 \times k)$	(probability)	$(s_2 \times k)$	(probability)
$s_1 \times 0$	0.00674	$s_2 \times 0$	0.00005
$s_1 \times 1$	0.03369	$s_2 \times 1$	0.00045
.	.	.	.
.	.	.	.
.	.	.	.
$s_1 \times m_1$	0.17547	$s_2 \times m_2$	0.12511
.	.	.	.
.	.	.	.
.	.	.	.
$s_1 \times 2m_1$	0.01813	$s_2 \times 2m_2$	0.00187
.	.	.	.
.	.	.	.
.	.	.	.

d. The convolution between claim amounts and probabilities of each distribution.

$s_1 \times 0$	$+ s_2 \times 0$	0.00674×0.00005
$s_1 \times 0$	$+ s_2 \times 1$	0.00674×0.00045
.	.	.
.	.	.
.	.	.
$s_1 \times 1$	$+ s_2 \times 0$	0.03369×0.00005
$s_1 \times 1$	$+ s_2 \times 1$	0.03369×0.00045
.	.	.
.	.	.
.	.	.
$s_1 \times m_1$	$+ s_2 \times 0$	0.17547×0.00005
$s_1 \times m_1$	$+ s_2 \times 1$	0.17547×0.00045
.	.	.
.	.	.
.	.	.
$s_1 \times 2m_1$	$+ s_2 \times 0$	0.01813×0.00005
$s_1 \times 2m_1$	$+ s_2 \times 1$	0.01813×0.00045

e. By the calculated value of claim amount, various values of probability are totalled and classified.

claim amount	probability
0	0.00000
s_1	0.00000
$2s_1$	0.00000
.	.
.	.
.	.
$2s_2$	0.00003
.	.
.	.
.	.
$3s_2$	0.00017
.	.
.	.
.	.

f. Delete the small value of probability and its claim amount. The distribution after the delete is described as $f_{12}(sk; m)$.

g. Calculate $f_{123}(sk; m)$ by the previous calculation step d. between $f_{12}(sk; m)$ and $p_3(s_3k; m_3)$ and proceed to the step e. and f.

h. Step g. is continued until the last.

(3) *The calculation error by deleting small value of probability*

The error is as the following, when the value of probability is counted fractionary over 0.5 as once and disregarding the rest at the below sixth.

numbers of m convoluted	error
30 * 30	0.00005
60 * 60	0.00011
120 * 120	0.00024
240 * 240	0.00052
480 * 480	0.00110

The above example illustrates that errors are small.

If we find that a more accurate calculation is necessary, it may be figured by replacing the value of probability which we neglected with a more detail value

SUMMARY

Many methods have been developed over the estimation of claims distributions This paper is one of the proposal for estimating claims distribution

In this paper, I assume that the claims distribution is a compound distribution of claim occurrence frequency distribution and claim amounts distribution And I propose a actual estimating way of claims distribution on the above reasoning

The following is the architecture of this paper

1 THE MODEL OF CLAIMS DISTRIBUTION

(1) *The logical claims distribution*

In order to make the above mentioned compound distribution, the claim amounts distribution is classified by many classes of claim amount

Claim occurrence frequency distribution of the above each class is assumed to be a binomial distribution respectively

And then the model of logical claims distribution is considered to be a compound distribution of binomial distributions which are the distribution of each class claim amount

(2) *The model of actual claim distribution*

The above logical claims distribution is difficult to estimate On the other hand, however actual claims distributions may be considered to have more numbers of risks than about 50 and be less claim occurrence rate than about 0 1

As a result, the claim occurrence distribution of each claim amount may be assumed to be a poisson distribution

2 THE ESTIMATION OF A CLAIMS DISTRIBUTION

(1) *Simplification of model*

The model of actual distribution is easier to estimate than the model of logical one In order to make estimations easier, I tried to simplify the model itself (In detail I describe it on the main paper)

(2) *Actual calculation of compound poisson distribution*

The main paper will describe the detail way of actual calculation and method to simplify calculation using a computer

(3) *The calculation error*

The main paper will describe the calculation error by the way of simplification incalculation