

# BETWEEN INDIVIDUAL AND COLLECTIVE MODEL FOR THE TOTAL CLAIMS

BY R. KAAS

*University of Amsterdam*

A. E. VAN HEERWAARDEN

*University of Amsterdam*

AND

M. J. GOOVAERTS

*K. U. Leuven and University of Amsterdam*

## ABSTRACT

This article studies random variables whose stop-loss rank falls between a certain risk (assumed to be integer-valued and non-negative, but not necessarily of life-insurance type) and the compound Poisson approximation to this risk. They consist of a compound Poisson part to which some independent Bernoulli-type variables are added.

Replacing each term in an individual model with such a random variable leads to an approximating model for the total claims on a portfolio of contracts that is computationally almost as attractive as the compound Poisson approximation used in the standard collective model. The resulting stop-loss premiums are much closer to the real values.

## 1. INTRODUCTION

Suppose we are interested in the distribution function of the total claims  $S$  on a certain insurance portfolio. If  $X_i$  is the random claim of contract  $i$ , assumed to be non-negative and integer-valued,  $S$  can be written as

$$(1) \quad S = \sum_{i=1}^n X_i.$$

Unless  $n$  is small it is not advisable to use convolution to compute the distribution of  $S$  directly. Even if the contracts are of life-insurance type and one uses, like DE PRIL (1986), the additional structure present in the problem (integer contract sizes in a small range, only few different claim probabilities), already for moderately large  $n$  the computing time involved is prohibitive.

Recently much progress has been made in constructing algorithms to approximate the distribution of the total claims with a controllable error bound. We mention KORNIA (1983), HIPP (1986), and DE PRIL (1988); see also KUON, REICH and REIMERS (1987) and REIMERS (1988).

In this paper we try to improve the standard collective model approximation in a transparent way, and at the cost of only a small increase in computing time. We give an extension of the method introduced in KAAS, VAN HEERWAARDEN and GOOVAERTS (1988) for the case of a life portfolio.

One obtains good approximations by replacing the individual model (1) with a collective model  $S''$  having a compound Poisson distribution. See BOWERS *et al.* (1987, Chapter 13), or KAAS (1987, Chapter 1). One way to obtain  $S''$  is to replace each summand  $X_i$  of  $S$  in (1) by a compound Poisson sum (with parameter 1) of iid random variables with the same distribution as the original term, as follows:

$$(2) \quad S'' = \sum_{i=1}^n \sum_{j=1}^{N_i} X_i^{(j)}.$$

Being a sum of compound Poisson random variables, the random variable  $S''$  is also a compound Poisson distribution with as claims distribution the arithmetic average of the distributions of the individual contracts, and with Poisson parameter  $n$ .

A more standard way to derive a good fitting collective model is the following. One assumes the contracts to have claim probability  $q_i$  and approximates the aggregate claims with a compound Poisson distribution with Poisson parameter  $\lambda$  and as claims distribution a weighted average of the conditional claims distributions, given that a claim occurs. Using  $\lambda = \sum q_i$ , and weights  $q_i/\lambda$ , we obtain the same distribution as  $S''$ .

KAAS *et al.* (1988) assume the term  $X_i$  to be of Bernoulli type: either  $X_i = 0$ , or  $X_i$  equals some known amount  $M_i$ , as is common in life-insurance applications. They replace most terms of (1) by compound Poisson sums, but leave risks with high risk premiums unchanged. If  $V$  is the set of indices of risks with high risk premium, the following random variable  $S'$  results:

$$(3) \quad S' = \sum_{i \notin V} \sum_{j=1}^{N_i} X_i^{(j)} + \sum_{i \in V} X_i.$$

The distribution of the first term of (3) can be computed using Panjer's recursion or an algorithm based on the Fast Fourier Transform. The large risks in the set  $V$  are added using straightforward convolution. If the size of  $V$  is small, this second step costs only little extra time.

The stop-loss premiums with  $S'$  are larger than those of  $S$ , but not as large as those of  $S''$ ; since both  $S'$  and  $S''$  are integer-valued and non-negative, the difference for  $S'$  and  $S''$  of the total error of the stop-loss premiums, summed over all integer retentions, equals  $\frac{1}{2}(\text{Var}[S''] - \text{Var}[S'])$ . This property is proven in KAAS *et al.* (1988).

In this paper we assume the terms  $X_i$  of  $S$  to be more general risks. Of course one might still exclude the large risks when applying the Poisson approximation, as was proposed in KAAS *et al.* (1988). The convolution process then costs time proportional to the total number of mass points of the large risks, many of which are not large at all. We will replace each term of  $S$  by a random variable with stop-loss premiums between those of the terms of  $S$  and  $S''$ , and with a form

similar to (3). It will be obtained by replacing most mass points by a compound Poisson part, and only a few very large mass points by Bernoulli-type random variables. This way the same accuracy is achieved using less computational effort. Compared to the approximate methods mentioned above our method is much more intuitive and easier to understand. A further point we wish to make is the following. One should not spend too much energy on the exact calculation of this distribution, because there are other sources of error with much greater influence. For instance if the probabilities of a mortality table are multiplied by a factor  $1 + \varepsilon$ , the relative change in the stop-loss premiums is of the same order of magnitude  $\varepsilon$ .

2. ON STOP-LOSS ORDER

If risk  $X$  (with  $E[X] < \infty$ ) has stop-loss premiums lower than risk  $Y$  for all retentions uniformly, we say that  $X$  precedes  $Y$  in stop-loss order and write  $X \ll Y$ . Stop-loss order is transitive:  $X \ll Y$  and  $Y \ll Z$  implies  $X \ll Z$ . It is not a complete order: for many random variables  $X$  and  $Y$  we have neither  $X \ll Y$ , nor  $Y \ll X$ .

We quote a theorem from GOOVAERTS, HAEZENDONCK and DE VYLDER (1984, Chapter 4) on the invariance properties of stop-loss order.

**THEOREM 1.** *If  $N_1, X_1, X_2, \dots$  and  $N_2, Y_1, Y_2, \dots$  are independent random variables with counting variables  $N_1$  and  $N_2$  satisfying  $N_1 \ll N_2$ , and with  $X_i \ll Y_i$  for all  $i$ , we have*

$$(4) \quad \sum_{j=1}^{N_1} X_j \ll \sum_{j=1}^{N_2} Y_j.$$

If  $E[X] = E[Y]$ , a sufficient condition for  $X \ll Y$  is the following: there exists an  $\alpha \in \mathbb{R}$  with the property that  $F_X(x) \leq F_Y(x)$  for  $x < \alpha$ ,  $F_X(x) \geq F_Y(x)$  for  $x \geq \alpha$ . We say then that risk  $Y$  is *more dangerous* than  $X$ .

Some special cases of Theorem 1: first, taking  $N_1 \equiv N_2 \equiv 2$ , we see  $X + Y \ll X + Z$  if  $Y$  and  $Z$  are independent of  $X$  and  $Y \ll Z$ . Second, if  $N_1 \equiv 1$ ,  $N_2$  is Poisson(1) and  $X_1, X_2, \dots$  are iid random variables distributed as  $X$ , we have  $N_1 \equiv N_2$  since  $N_2$  is more dangerous. It follows that  $X$  precedes its compound Poisson approximation:

$$(5) \quad X \ll \sum_{j=1}^{N_2} X_j.$$

Taking  $N_1 \equiv n$  and using (5), we may conclude that the individual model (1) precedes the collective model (2).

3. LESS CONSERVATIVE APPROXIMATIONS

Consider one of the terms  $X$  of (1), omitting the index  $i$  for notational convenience. Suppose it either has range  $\{0, x_1, x_2, \dots, x_m\}$  or  $\{x_1, x_2, \dots, x_m\}$  with

$0 \leq x_1 \leq x_2 \leq \dots \leq x_m$ . Introducing indicator variables  $I_j$ ,  $j = 1, 2, \dots, m$ , defined as  $I_j = 1$  if  $X = x_j$ , and 0 otherwise, we may write

$$(6) \quad X = \sum_{j=1}^m I_j \cdot x_j.$$

The marginal distribution of the random variables  $I_j$  is 'Bernoulli( $p_j$ ), with  $p_j = P[X = x_j]$ . They are dependent: we have  $\sum I_j = 0$  if  $X = 0$ , 1 otherwise.

In the collective model (2),  $X$  is replaced by a random variable  $Z$  having a compound Poisson(1) distribution with terms distributed as  $X$ . See also (5). By Theorem 11.2 of BOWERS *et al.* (1987)  $Z$  may be written as

$$(7) \quad Z = \sum_{j=1}^m N_j \cdot x_j.$$

where the  $N_j$  are independent Poisson( $p_j$ ) random variables, counting the number of occurrences of claim size  $x_j$ . Now define the following random variable:

$$(8) \quad Y = \sum_{j=1}^m B_j \cdot x_j.$$

where the random variables  $B_j$  have the same marginal distribution as  $I_j$ , but are independent. Then we may prove

**THEOREM 2.** *For  $X$ ,  $Y$  and  $Z$  as defined in (6), (8) and (7), we have*

$$X \leq Y \leq Z.$$

**PROOF.** We will show that  $X \leq Y$  by induction on the number of terms in (6). If  $m = 1$ ,  $X$  has the same distribution as  $Y$ , so certainly  $X \leq Y$ . Now suppose the statement proven if the number of non-zero mass points is  $m - 1$  or less. Let  $X^m$  be defined as  $X - I_m \cdot x_m$ , so  $X^m$  is zero when  $X = x_m$ , and equal to  $X$  otherwise. We will compare  $X = I_m \cdot x_m + X^m$  to  $T = B_m \cdot x_m + X^m$ .  $T$  is more dangerous than  $X$ , since  $E[T] = E[X]$ , and  $F_X(x) \geq F_T(x)$  for  $x \geq x_m$  trivially since  $F_X(x_m) = 1$ ; for  $x < x_m$  we have

$$(9) \quad F_T(x) = P[B_m \cdot x_m + X^m \leq x] = P[B_m = 0] \cdot P[X^m \leq x] \\ = (1 - p_m) \cdot \{p_m + F_X(x)\} = F_X(x) + p_m \{1 - p_m - F_X(x)\} \geq F_X(x).$$

So, we have established  $X \leq T$ . By the induction hypothesis,  $X^m$  precedes the sum of the first  $m - 1$  terms in (8) in stop-loss order. Using Theorem 1 we immediately see that  $T \leq Y$ , and so, by the transitivity of stop-loss order, we have  $X \leq Y$ .

The second part of the theorem is easy: since  $B_j \leq N_j$  for each  $j$ , by twice applying Theorem 1 we have  $Y \leq Z$ . ■

In fact, we have  $X \leq Y \leq Z' \leq Z$  with  $Z'$  defined as

$$(10) \quad Z' = \sum_{j=1}^{m-1} N_j \cdot x_j + \sum_{j=m-t+1}^m B_j \cdot x_j,$$

for some  $t = 1, 2, \dots$ . So we have found random variables  $Z'$ , that are the convolution of a compound Poisson part and  $t$  Bernoulli parts, and such that  $Z'$  is between  $X$  and  $Z$  in stop-loss order.

Replacing each term  $X_i$  of  $S$  in (1) by a random variable  $Z'_i$  as in (10) we obtain  $S'$ , which is a sum of a compound Poisson distribution (the convolution of the first parts of (10)) and all the Bernoulli-type random variables that constitute the second parts of (10).

As remarked in the Introduction, the total increase in the error of the resulting stop-loss premiums equals half the difference in variances between the different models (see KAAS *et al.*, 1988). Term  $X$  contributes  $\text{Var}[X]$  to the variance of the total claims,  $Z$  as in (7) contributes  $\text{Var}[Z] = E[X^2]$  and  $Z'$  as in (10) contributes

$$(11) \quad \text{Var}[Z'] = E[X^2] - \sum_{j=m-t+1}^m p_j^2 \cdot x_j^2.$$

If, as in KAAS *et al.* (1988), we assume that the number of terms to be convoluted is fixed in advance, the mass points with the highest contributions to  $E[S]$  should be treated separately.

REMARK. There are many other random variables with stop-loss premiums between those of a certain random variable  $X$  and its compound Poisson approximation  $Y$  as in (8). Let  $I$  be any indicator variable, not necessarily a function of  $X$ . The random variable  $I$  might indicate whether  $X$  came from a special source (e.g. material damage versus personal injury) or whether  $X$  is a large claim or not, and so on. Introduce independent random variables  $X_0$  and  $X_1$  with  $X_j$  distributed as  $X | I = j$ . Then  $X$  has the same distribution as  $I \cdot X_1 + (1 - I) \cdot X_0$ . If  $B_1$  and  $B_0$  are independent random variables distributed as  $I$  and  $1 - I$  respectively, we have  $X \ll B_1 \cdot X_1 + B_0 \cdot X_0$ .

PROOF. Let  $p = P[I = 1]$  and  $q = 1 - p$ . For any retention  $d \geq 0$ , we have by taking conditional expectations depending on the values of  $I$ ,  $B_1$  and  $B_0$ :

$$\begin{aligned} E[(X - d)_+] &= p \cdot E[(X_1 - d)_+] + q \cdot E[(X_0 - d)_+] \\ &= (p^2 + pq) \cdot E[(X_1 - d)_+] + (q^2 + pq) \cdot E[(X_0 - d)_+] \\ &\leq p^2 \cdot E[(X_1 - d)_+] + q^2 \cdot E[(X_0 - d)_+] \\ &\quad + pq \cdot E[(X_1 + X_0 - d)_+] \\ &= E[(B_1 \cdot X_1 + B_0 \cdot X_0 - d)_+]. \quad \blacksquare \end{aligned}$$

The choice of  $I = 1$  if  $X = x_m$  gives another proof that  $X \ll T$  holds in Theorem 2.

#### 4. AN EXAMPLE

Consider a portfolio consisting of many small risks, such that the total claims  $S$  can be taken to be compound Poisson distributed with parameter 1 and claims

TABLE I  
STOP-LOSS PREMIUMS AT RETENTIONS  
 $d = 0, 4, 8, \dots, 32$ .

$d =$	$S + G''$	$S + G'$	$S + G$
0	3.01000	3.01000	3.01000
4	1.07603	1.06498	1.06418
8	0.44933	0.42025	0.41927
12	0.12743	0.08722	0.08672
16	0.03721	0.00829	0.00822
20	0.01143	0.00049	0.00048
24	0.00262	0.00002	0.00002
28	0.00076	0.00000	0.00000
32	0.00017	0.00000	0.00000

distribution with equal probability of claim 1, 2 and 3. If  $N_1$ ,  $N_2$  and  $N_3$  are independent Poisson(1/3) random variables,  $S$  has the same distribution as  $N_1 + 2 \cdot N_2 + 3 \cdot N_3$ . To this portfolio we add a large risk  $G$ , with probability 0.1 of a claim of 10, 0.01 of a claim of 1 and 0.89 of no claim. Define  $G'$  to be  $10 \cdot B_4 + N_5$ , and  $G''$  as  $10 \cdot N_4 + N_5$ , where  $B_4$  is Bernoulli(0.1),  $N_4$  Poisson(0.1) and  $N_5$  Poisson(0.01) distributed, all independent. We will compare the stop-loss premiums of  $S + G$ ,  $S + G'$  and  $S + G''$  at retentions  $d = 0, 4, 8, \dots, 32$ . The results are in Table 1. Notice that the stop-loss premiums for  $S + G'$  and  $S + G$  are practically the same. The fact that in  $S + G''$  more than one claim of size 10 may occur leads to considerably higher stop-loss premiums.

## REFERENCES

- BOWERS, N. L., GERBER, H. U., HICKMAN, J. C., JONES, D. A. and NESBITT, C. J. (1987); Actuarial Mathematics; *Society of Actuaries*, Itasca, IL.
- DE PRIL, N. (1986); On the exact computation of the aggregate claims distribution in the individual life model; *ASTIN Bulletin* 16, 109–112.
- DE PRIL, N. (1988); Improved approximation for the aggregate claims distribution of a life portfolio, *Scandinavian Actuarial Journal*, to be published.
- GOOVAERTS, M. J., HAEZENDONCK, J. and DE VYLDER, F. (1984); Insurance Premiums; *North-Holland*, Amsterdam.
- HIPP, CH. (1986); Improved Approximations for the Aggregate Claims Distribution in the Individual Model, *ASTIN Bulletin* 16.2, 89–100.
- KAAS, R. (1987); Bounds and approximations for some risk theoretical quantities; *University of Amsterdam*.
- KAAS, R., VAN HEERWAARDEN, A. E. and GOOVAERTS, M. J. (1988); On stop-loss premiums for the individual model, *ASTIN Bulletin* 18, 91–98.
- KORNYA, P. S. (1983); Distribution of aggregate claims in the individual risk theory model, *Transactions of the Society of Actuaries* 35, 823–836. Discussion 837–858.
- KUON, S., REICH, A. and REIMERS, L. (1987); PANJER vs KORNYA vs DE PRIL: A Comparison from a Practical Point of View, *ASTIN Bulletin* 17.2, 183–191.
- REIMERS, L. (1988); Letter to the Editor, *ASTIN Bulletin* 18.2, 113–114.

ROB KAAS, ANGELA VAN HEERWAARDEN, MARC GOOVAERTS  
*Institute for Actuarial Science and Econometrics, Jodenbreestraat 23, NL-1011 NH Amsterdam*