A PROCEDURE TO COMPUTE VALUES OF THE GENERALISED POISSON FUNCTION*

ESA HOVINEN
Helsinki

The purpose of this paper is to describe a technical procedure, which enables one to compute values of the generalised Poisson distribution function, with an accuracy considered sufficient for insurance companies and with satisfactory speed. The procedure requires a fast medium sized computer.

The computation of values of the generalised Poisson distribution function has become a timely problem in Finland, because of the introduction by the Supervisory Service of more stringent requirements in determining limits of the so called equalisation reserves, which have their theoretical basis in the random fluctuations of claims amounts. The question has also been discussed in papers submitted by Dr. Pentikäinen [3] and Dr. Pesonen [4] to this Colloquium. Because the practical computation is a further problem, the Federation of Finnish Insurance Companies set up a committee in 1962 to gather and work up the necessary statistics from various branches of insurance and to develop the computational methods ready for use in practice. The committee has almost completed its work, and one of the results, a procedure to compute values of the generalised Poisson function with a mixed method, is described below. The method is referred to briefly in [4]. The programming and further planning work has been done by Dr. Loimaranta and M. Sc. Pörn.

On the principles

The computation is based to a partitioning of the distribution function of one claim,

$$S(x) = p_1 S_1(x) + p_2 S_2(x) + p_3 S_3(x) \qquad (1)$$

*) Those interested in the program written in ALGOL for an Elliott 503 computer should write to "Vahinkovakuutusyhtiöiden Tilastokeskus", Bulevardi 28, Helsinki, Finland.

https://doi.org/10.1017/S0515036100010709 Published online by Cambridge University Press

where

$$\begin{aligned} p_1 &= S(x_1) \\ p_2 &= S(x_2) - S(x_1) \text{ and} \\ p_3 &= 1 - S(x_2);\ x_2 \geqslant x_1. \end{aligned} \tag{2}$$

The functions $S_v(x)$ are

$$S_1(x) = \begin{cases} \dfrac{1}{p_1} S(x) \text{ when } x \leqslant x_1, \\ 1 \text{ elsewhere} \end{cases}$$

$$S_2(x) = \begin{cases} 0 \text{ when } x \leqslant x_1 \\ \dfrac{1}{p_2} (S(x) - S(x_1)) \text{ when } x_1 < x \leqslant x_2 \\ 1 \text{ elsewhere} \end{cases} \tag{3}$$

$$S_3(x) = \begin{cases} 0 \text{ when } x \leqslant x_2 \\ \dfrac{1}{p_3} (S(x) - S(x_2)) \text{ elsewhere.} \end{cases}$$

According to a lemma in /4/, the generalised Poisson distribution $F(x)$, which under certain assumptions is the distribution of total claims, can be represented as a convolution of three generalised Poisson distributions:

$$F(x) = \sum_0^\infty \frac{e^{-n} n^k}{k!} S^{k*}(x) \tag{4}$$

$$= F_1(x) * F_2(x) * F_3(x), \tag{5}$$

where

$$F_v(x) = \sum_0^\infty \frac{e^{-p_v n} (p_v n)^k}{k!} S_v^{k*}(x) \tag{6}$$

are separately generalised Poisson functions. The function $F_1(x)$ is computed by normal approximation and the functions F_2 and F_3 by Monte-Carlo method. The final convolutions (5) are performed in the same simulations.

https://doi.org/10.1017/S0515036100010709 Published online by Cambridge University Press

The point x_1 is so selected, that $F_1(x)$ is approximated by a normal approximation with sufficient accuracy. This is done by a subsidiary program which prepares the input data for the main program. The theory for the computation of x_1 is presented in /4/. At the same time, the mean value and standard deviation of $F_1(x)$ are computed. The point x_2 is determined so, that only a few most dangerous points remain in S_3, e.g. by setting $p_3 = 1/100\ p_2$.

The function F_2 is computed by a Monte-Carlo method. For this purpose the distribution

$$P(N_2) = \sum_0^N \frac{e^{-p_2 n}(p_2 n)^k}{k!} \tag{7}$$

and the functions

$$S_2^{2^0*}(x)(= S_2(x)); S_2^{2^1*}(x); S^{2^2*}(x); \ldots; S^{2^k*}(x) \tag{8}$$

up to a sufficiently large k are needed. The convolutions (8) and the distribution (7) are determined by the main program. When a random number N_2 is generated by the help of (7), it is written as a binary number

$$N_2 = \Sigma\ a_k 2^k\ (a_k = 0,1).$$

Then, $\Sigma\ a_k$ random numbers η_k $(a_k = 1)$ between 0,1 are generated and a sample value of x from $F_2(x)$ is reached as the sum

$$\xi_2 = \Sigma\ x\ (S_2^{2^k*}(x) = \eta_k). \tag{9}$$

The function $F_3(x)$ is treated more directly. First, a random number N_3 is determined like N_2. Then, generating N_3 random values θ_k between 0,1 a sample value of F_3 is reached as

$$\xi_3 = \Sigma\ x\ (S_3(x) = \theta_k). \tag{10}$$

When a random ξ_1 is taken from the distribution $\Phi(p_1 n\ \alpha_1, \sqrt{p_1 n\ \alpha_2})$, where

$$\alpha_v = \int_0^\infty x^v dS_1(x), \tag{11}$$

the value

$$\xi = \xi_1 + \xi_2 + \xi_3 \tag{12}$$

https://doi.org/10.1017/S0515036100010709 Published online by Cambridge University Press

is a sample value of x with distribution function $F(x)$. Sufficiently many values ξ determine $F(x)$, points of which can be determined e.g. by computing the number of ξ: s exceeding given x-values. The ξ: s themselves are an additional output of the program.

The method of computing the convolutions (8) is the following, which the author presented to the meeting in Edinburgh in 1964 /2/. Let the distribution function $Q_j(x)$ be given at the points

$$\left(Q_{ji} = \frac{i}{n}\,;\, x_i = x\left(Q_j = \frac{i}{n} - \frac{i}{2n}\right)\right). \tag{13}$$

The convolution is then, approximately,

$$Q(x) = Q_1(x) * Q_2(x) \tag{14}$$

$$= \frac{1}{n} \sum_{x_{ii} < x} Q_2(x_{2j}),$$

where $x_{2j} \leqslant x - x_{1_i} < x_{2,\,j+1}$. In the computation of convolutions, the prescribed form (13) is required as input and output; first an input is made from $S_2(x)$ by the subsidiary program mentioned above, then an output for the repetitive computation of the next convolution $S_2^{2k*}(x)$ is prepared by the convolution program itself. This is done by an automatic interpolation system from certain guessed values of x, which system also accepts discontinuities. Moreover, a correction for the mean of $Q(x)$ and its tails is performed. For the accuracy of the method see /2/.

The computing program

The program to compute $F(x)$ is written in Algol for an Elliott 503 computer. A subsidiary program prepares the data needed as an input for the main program, which computes sample values of $F(x)$. In this subsidiary program, the original input $S(x)$ is assumed to be given at the points $x_i = a.b^i$. How to get the function $S(x)$ for a given company or situation is another problem which falls outside the scope of this paper. The main program is divided in six parts, which have their own function in computation.

In part 1 the numbers n, p_1 and $p_1 + p_2$ are read in, universal integers and reals are defined and the necessary storage space for $P(N_2)$ and $P(N_3)$ is determined by the procedure "store".

https://doi.org/10.1017/S0515036100010709 Published online by Cambridge University Press

In part 2, besides the necessary definitions, procedures "sum", "poisson", "norm" and "teta" are defined. With the aid of "poisson", the distributions $P(N_2)$ (see (7)) and $P(N_3)$ are determined and stored. If $p_2 n > 100$, $P(N_2)$ is computed from the approximation

$$P(N_2) = e^{-p_2 n} \sum_{v=0}^{N_2} \frac{(p_2 n)^v}{v!} \approx \Phi\left(2\left(\sqrt{N_2 + \frac{1}{2}} - \sqrt{p_2 n}\right)\right). \qquad (15)$$

The necessary storage space for the convolution functions (8) is determined and α_1, α_2, $S_2(x)$ etc. are read in. The procedure "sum" computes $\sum_{i=m}^{n} f_i$ and the procedure "poisson" computes values of the Poisson distribution. Random numbers with a normal distribution are computed by "norm", the method being developed by Dr. Loimaranta in AB Atomenergi in Sweden. Random numbers rectangularly distributed over (0,1) are generated by the multiplicative congruential method in procedure "teta".

The convolutions (8) are computed and stored in part 3. The main part of convolution program serves for the organisational interpolation work.

The function $S(x)$, $x > x_2$ is read in and the mean values in convolutions of $S_2(x)$ are corrected in part 4.

Part 5 comprises the principal work: it computes ξ according to (12).

The output is performed in part 6, either the number of ξ' s exceeding 23 prefixed values of x or, additionally, all values of ξ.

Some general remarks

In the testing phase of programming the results were compared with some values obtained by Bohman and Esscher /1/ with $S(x)$ from non-industrial business in /1/. The results, for a sample of 10.000 from $F(x)$, were found to agree with the Swedish results. The computation took c. 8 minutes for a given distribution $F(x)$ (sample of 10.000) with the Elliott 503 computer.

The procedure to compute $F(x)$ seems to meet sufficiently well the necessary requirements of accuracy and computational speed. As a computer program it is easy to use and quite general, in practice

no information is required concerning the shape and location of the function $S(x)$ and the value of n. Moreover, the program can be modified in many directions to meet different needs. One such need is to compute the upper limit of the equalisation reserve presented in /5/. For that purpose, a sample value ξ_1 is computed for the first year, conclusions concerning the situation can then be made by a further program, then a second year in the life of the company can be simulated perhaps taking into account the result of the preceding year and so on. The situation ξ after n years can thus be reached by simulating the n years in sequence (a single ξ_v is determined for the v' th year). When this simulation series of n years is repeatedly performed sufficiently many times, the distribution of ξ is obtained. The process is fast in practice if $S(k_v x)$ can be assumed to be the same in all years (k_v = constant for each year, $k_v x$ the amount of one claim in v'th year) and n can be fixed. As additional arguments e.g. a stop loss treaty, an assumed dividend, funding, rating and reinsurance policy can be attached to the computations. The possibility of examination and comparison of quite complex situations is offered. However, such investigations (management games) are probably sufficiently realistic with a simple normal approximation.

Another generalisation of the program is reached if we can set $p_2 = 1$, by using another functions in (4) instead of

$$P(N) = \frac{e^{-n} n^k}{k!}$$

Many ways to improve the procedure can also be seen. One such way, if the possibility for further simulation can be removed, would be the following. Let the point

$$\eta_v = \xi_2 + \xi_3$$

be computed by a Monte Carlo method as in the procedure described (cf. 12). We can write

$$F(x) = F_1(x) * F'(x)$$

$$= \int_0^\infty \Phi\,(x - \eta) dF'(\eta). \tag{16}$$

With sufficient accuracy, if the simulation is repeated N times,

https://doi.org/10.1017/S0515036100010709 Published online by Cambridge University Press

the increments $dF'(\eta)$ can be regarded to lie at points η_v, with the amounts $\frac{1}{N}$. Thus

$$F(x) = \frac{1}{N} \sum \Phi (x - \eta_v) \tag{17}$$

by which $F(x)$ is directly computed at the points x. The variance of the computed value of $F(x)$ arising from the nature of the Monte Carlo method is less by this method than by the described program, and the accuracy of the result can be controlled by sequential checking during computation. The variance

$$\sigma^2 (F(x)) = \frac{1}{N-1} \left(\frac{1}{N} \sum \Phi(x - \eta_v)^2 - F(x)^2 \right)$$

$$\leqslant \frac{1}{N-1} F(x) (1 - F(x)), \tag{18}$$

where the latter expression stands for the variance for non-improved Monte Carlo procedure.

Similarly

$$\sigma^2 (1\text{-}F(x)) = \frac{1}{N-1} \left(\frac{1}{N} \sum (1 - \Phi (x - \eta_v))^2 - (1 - F(x))^2 \right)$$

$$\leqslant \frac{1}{N-1} F(x) (1 - F(x)). \tag{19}$$

The inequalities (18) and (19) reveal the improvement particularly for the tails of $F(x)$.

References

[1] Bohman, H. and Esscher, F., 1964, Studies in risk theory with numerical illustrations concerning distribution functions and stop loss risk premiums, Skand. Akt. Tidskr.

[2] Hovinen, E., 1964, A method to compute convolutions (not printed).

[3] Pentikäinen, T., 1965, On the solvency of insurance companies, ASTIN Colloquium in Lucerne.

[4] Pesonen, E., 1965, On the calculation of the generalised Poisson function, ASTIN Colloquium in Lucerne.

[5] Pesonen, E., 1965, Magnitude control of technical reserves in Finland, ASTIN Colloquium in Lucerne.