EVALUATING COMPOUND GENERALIZED POISSON DISTRIBUTIONS RECURSIVELY

BY M. J. GOOVAERTS AND R. KAAS

Amsterdam/Leuven

ABSTRACT

In this paper we give a recursive scheme, involving Panjer's recursion, to compute the distribution of a compound sum of integer claims, when the number of summands follows a Generalized Poisson distribution. Also, an elegant derivation is given for some basic properties of this counting distribution.

1. THE GENERALIZED POISSON DISTRIBUTION

The Generalized Poisson distribution, see CONSUL (1989), is an integer-valued, non-negative distribution with two parameters θ and λ. A random variable N having this distribution with parameters θ and λ is also denoted as a GP (θ, λ) random variable. In the first section we repeat the mathematical properties of this distribution, giving a short and elegant derivation. The second section contains a recursive algorithm to compute the probabilities of a compound Generalized Poisson distribution. This algorithm is obtained by the well-known technique of differentiating the generating function and comparing coefficients of resulting power series. This function, however, is known only in an implicit form, so the process is not as trivial as usual.

An actuarial application of the Generalized Poisson distribution, linking it to the ruin model, can be found in GERBER (1990). Other chance mechanisms generating this distribution are described in CONSUL (1989). One of these is the Galton-Watson branching process, which is a model with many conceivable actuarial applications. In this process, the spreading of a certain disease is modeled as follows. Suppose M individuals are originally infected. Each of these infects L_i other individuals, $i = 1, \ldots, M$. These in turn infect L_{ij} new victims, $j = 1, \ldots, L_i$, and so on. Now if M is a Poisson (θ) distributed random variable, and the L_i, L_{ij}, $\ldots$ are independent Poisson (λ) random variables, the total number N of people infected has a Generalized Poisson distribution with parameters θ and λ.

The parameters θ and λ are non-negative; the Poisson distribution is the special case with $\lambda = 0$. Assume $\lambda < 1$ to ensure that N remains finite with probability one.

Consider the total number of individuals B_i infected by the ith person, including this person himself, and define B_{ij} analogously for the jth person

infected by $i, j = 1, \ldots, L_i$. Obviously B_i and B_{ij} are random variables with the same distribution. We can write B_i as:

$$(1) \qquad\qquad B_i = 1 + \sum_{j=1}^{L_i} B_{ij}.$$

Let B be distributed as B_i and B_{ij}. From relation (1), and using some well-known properties of compound Poisson (λ) distributions, we can directly derive expressions for the mean, variance and generating function of B. The mean can be computed as follows:

$$(2) \qquad\qquad E[B] = 1 + \lambda E[B] \Rightarrow E[B] = \frac{1}{1-\lambda}.$$

The variance and the second moment can be computed from:

$$(3) \qquad\qquad \mathrm{Var}\,[B] = \lambda E[B^2] \Rightarrow E[B^2] = \frac{1}{(1-\lambda)^3}.$$

If $G_B(u) = E[u^B]$ is the generating function of the B_i and B_{ij} random variables, it must satisfy the following relation:

$$(4) \qquad\qquad G_B(u) = u G_{L_i}(G_B(u)) = u\, e^{\lambda(G_B(u)-1)}.$$

Writing $t = t(u) = G_B(u)$, we obtain from (4):

$$(5) \qquad\qquad u = t\, e^{-\lambda(t-1)}.$$

The probabilities $P[B = i]$ are the coefficients of the power series representation of $t(u)$. To determine them from relation (4), we use a slightly simplified form of Relation 3.6.7 in ABRAMOWITZ and STEGUN (1965; Lagrange's expansion): if $u = f(t)$, $f(0) = 0$, $f'(0) \neq 0$, and g is any function infinitely differentiable, then

$$(6) \qquad\qquad g(t) = g(0) + \sum_{k=1}^{\infty} \frac{u^k}{k!} \left[\frac{d^{k-1}}{dt^{k-1}} \frac{g'(t)t^k}{f(t)^k} \right]_{t=0}.$$

The distribution of B is found by taking $g(t) = t$ and using $u = f(t)$ as in (5), resulting in the Borel distribution:

$$(7) \qquad\qquad P[B = i] = \frac{(i\lambda)^{i-1}\, e^{-i\lambda}}{i!}, \qquad i = 1, 2, \ldots.$$

Since a GP (θ, λ) random variable is a compound Poisson (θ) sum of Borel (λ) random variables, its generating function equals $e^{\theta(G_B(u)-1)}$, so the density of a GP (θ, λ) random variable N is found by taking $g(t) = e^{\theta(t-1)}$ in (6), leading to:

$$(8) \qquad\qquad P[N = n] = \frac{\theta(\theta+n\lambda)^{n-1}\, e^{-\theta-n\lambda}}{n!}, \qquad n = 0, 1, \ldots.$$

To compute mean and variance of N directly from (8) involves rather tricky mathematics. Using (2) and (3), however, it is trivial exercise; note that for $\lambda > 0$ the mean exceeds the variance:

$$(9) \qquad E[N] = \frac{\theta}{1-\lambda}; \qquad \text{Var}\,[N] = \frac{\theta}{(1-\lambda)^3}.$$

Being a compound Poisson (θ) sum (of Borel (λ) distributions), a GP (θ, λ) random variable is easily seen to be infinitely divisible, as for any $n = 1, 2, \ldots$ it can be written as the convolution of n GP $\left(\dfrac{\theta}{n}, \lambda\right)$ variables.

2. A RECURSIVE ALGORITHM FOR THE PROBABILITIES OF A COMPOUND GP DISTRIBUTION

To actuaries the total of the incurred claims is more relevant than their number. If the costs associated with occurrence i, $i = 1, \ldots, N$, are given by a random variable Z_i, then the total costs are given by the following compound Generalized Poisson (θ, λ) random variable:

$$(10) \qquad S = \sum_{i=1}^{N} Z_i.$$

Here the GP (θ, λ) distributed counting variable N is assumed to be independent of all Z_i, and the sequence $Z_1, Z_2, \ldots$ is i.i.d. We assume the Z_i to be integer-valued and positive. (By excluding zero-claims, we avoid problems later on, when we have to compute $P[S = 0]$ to start a recursion.)

Actuaries prefer to use counting distributions that are suitable for computations of quantities like probabilities of ruin and stop-loss premiums. Since PANJER (1981) actuaries are aware that there is a very efficient recursive algorithm to compute probabilities of S as in (10) if N is Binomial, Negative Binomial or Poisson. SUNDT and JEWELL (1981) derive similar recursions for a wider class of counting distributions. In this section we will derive recursion formulae expressing $P[S = s]$ in $P[S = j]$, $j = 0, 1, \ldots, s-1$ for the case of a Borel and a GP counting variable, too.

To this end, we will derive recursion relations for the coefficients of the generating function $G_S(u)$. Using the fact that a GP (θ, λ) distribution can be viewed as a compound Poisson (θ) sum of Borel (λ) distributions, we can rewrite S as follows:

$$(11) \qquad S = \sum_{i=1}^{M} Y_i \qquad \text{where} \qquad Y_i = \sum_{j=1}^{B_i} Z_{ij}.$$

Here M is a Poisson (θ) random variable, B_i is a Borel (λ) random variable and Z_{ij} is an i.i.d. sequence of claim amounts. Each term Y_i has a compound Borel distribution.

If N has a GP (θ, λ) distribution, and S is as in (10), by (4) and (5) we have

$$(12) \quad G_S(u) = e^{\theta(G_B(G_Z(u))-1)} = e^{\theta(t-1)} \quad \text{with } t \text{ such that } \quad t\,e^{-\lambda(t-1)} = G_Z(u).$$

This implicit description of the generating function of a compound Generalized Poisson distribution will enable us to derive relations between its probabilities. We do so in two steps. The first and most important step is to compute the coefficients of $G_B(G_Z(u))$, which amounts to computing the probability function of the compound Borel distributed Y_i random variables. The second step uses these coefficients to compute the coefficients of $G_S(u)$, simply by invoking Panjer's recursion formula.

Taking the derivative with respect to u of the logarithm of the second part of (12) provides us with the following relation:

$$(13) \quad \frac{d}{du} \log\left(t(u)\,e^{-\lambda(t(u)-1)}\right) = \frac{t'(u)}{t(u)} - \lambda t'(u) = \frac{d}{du} \log G_Z(u) = \frac{G_Z'(u)}{G_Z(u)}.$$

Rearranging leads to the following equality:

$$(14) \qquad\qquad t'(u) = \frac{t(u)}{1-\lambda t(u)} \frac{G_Z'(u)}{G_Z(u)}.$$

We introduce the following notation for the coefficients of the power series representations for the three factors appearing in (14):

$$(15) \quad t(u) = \sum_{n=1}^{\infty} \alpha_n u^n; \quad \frac{t(u)}{1-\lambda t(u)} = u\sum_{n=0}^{\infty} \beta_n u^n; \quad \frac{G_Z'(u)}{G_Z(u)} = \frac{1}{u}\sum_{n=0}^{\infty} r_n u^n.$$

Since the coefficients of u^n in (14) must be equal on both sides, we obtain the following relation:

$$(16) \qquad\qquad \alpha_{n+1}(n+1) = \sum_{j=0}^{n} \beta_j r_{n-j}.$$

The coefficients r_n depend on the known probability function of Z. We write $p_n = P[Z_i = n]$, $n = 1, 2, \ldots$; assume $p_1 > 0$. So we have

$$(17) \qquad\qquad G_Z(u) = \sum_{n=1}^{\infty} p_n u^n.$$

Using (17), rearranging the last equation of (15) and comparing coefficients of u^n leads to

(18)
$$\frac{\sum_{n=0}^{\infty} (n+1)p_{n+1} u^n}{u \sum_{n=0}^{\infty} p_{n+1} u^n} = \frac{1}{u} \sum_{n=0}^{\infty} r_n u^n \Rightarrow$$

$$(n+1)p_{n+1} = \sum_{j=0}^{n} r_j p_{n+1-j}, \qquad n = 0, 1, \ldots.$$

Then the r_n can be determined as follows:

(19) $$r_n = \frac{1}{p_1} \left((n+1)p_{n+1} - \sum_{j=0}^{n-1} r_j p_{n+1-j} \right) \qquad n = 0, 1, \ldots.$$

The coefficients α_n are the probabilities of Y_i to be determined. The auxiliary coefficients β_n can be expressed in $\alpha_1, \ldots, \alpha_{n+1}$, using the same technique leading to (18). Indeed the middle equation of (15) gives the result

(20)
$$\frac{u \sum_{k=0}^{\infty} \alpha_{k+1} u^k}{1 - \lambda \sum_{k=1}^{\infty} \alpha_k u^k} = u \sum_{k=0}^{\infty} \beta_k u^k \Rightarrow$$

$$\beta_n = \sum_{k=1}^{n} \beta_{n-k} \lambda \alpha_k + \alpha_{n+1}, \qquad n = 0, 1, \ldots.$$

Using (20) and the fact that $r_0 = 1$, see (19), we may write (16) as follows:

(21) $$\alpha_{n+1}(n+1) = \sum_{j=0}^{n-1} r_{n-j} \beta_j + \sum_{k=1}^{n} \beta_{n-k} \lambda \alpha_k + \alpha_{n+1}.$$

The following expression for α_{n+1} is found

(22) $$\alpha_{n+1} = \frac{1}{n} \left(\sum_{j=0}^{n-1} r_{n-j} \beta_j + \lambda \sum_{k=1}^{n} \beta_{n-k} \alpha_k \right), \qquad n = 1, 2, \ldots.$$

The probabilities α_n can now be computed successively. Indeed, if the probabilities $\alpha_1, \ldots, \alpha_n$ and the auxiliary quantities $\beta_0, \ldots, \beta_{n-2}$ are known, one computes β_{n-1} using (20), and next α_{n+1} using (22). Since $P[Z_i = 0] = 0$, the starting value α_1 can be computed as follows:

(23) $$\alpha_1 = P \left[\sum_{i=1}^{B} Z_i = 1 \right] = P[B = 1] P[Z_1 = 1] = p_1 e^{-\lambda}.$$

Note that by the requirement $p_1 > 0$ we have $\alpha_1 > 0$.

Having computed the coefficients $\alpha_1, \alpha_2, \ldots$, which are the probabilities of the random variables Y_i, we can compute the probabilities of S simply by using Panjer's recursion formula for the Poisson (θ) case, starting from $P[S = 0] = P[N = 0] = e^{-\theta}$:

$$(24) \qquad P[S = s] = \frac{\theta}{s} \sum_{j=1}^{s} j\alpha_j P[S = s-j], \qquad s = 1, 2, \ldots .$$

Remark

Taking $p_1 = 1$, $p_j = 0$ otherwise, one gets $r_j = 0$ for $j \neq 0$. Then (22) and (20) lead to a recursion for the Borel (λ) distribution (7); combining it with (23) gives a recursion for the GP (θ, λ) distribution (8).

3. CONCLUSIONS

The Generalized Poisson distribution may be a useful model when the chance mechanism used in the first section is appropriate, or any of the other models in CONSUL (1989). It can be used as an alternative to the Negative Binomial distribution when the tails of the counting distribution are thicker than those of the Poisson. It is mathematically a more complex distribution than the counting distributions usually assumed by actuaries (Binomial, Poisson or Negative Binomial), but we think that using the lines of thought given in the first section, actuaries will be able to use this distribution in their practical work.

The possible objection that this counting distribution is not suitable for actuarial calculations, which mostly involve compound sums, is removed by the recursive algorithm given in Section 2.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965) *Handbook of Mathematical Functions.* Dover Publications, New York.
CONSUL, P. C. (1989) *Generalized Poisson distributions, properties and applications.* Marcel Dekker Inc., New York, Basel.
GERBER, H. U. (1990) When does the surplus reach a given target? *Insurance: Mathematics and Economics* 9, 115–119.
PANJER, H. H. (1981) Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* 12, 22–26.
SUNDT, B. and JEWELL, W. S. (1981) Further results on recursive evaluation of compound distributions. *ASTIN Bulletin* 12, 27–39.

R. KAAS
University of Amsterdam, Institute of Actuarial Science and Econometrics, Jodenbreestraat 23, 1011 NH Amsterdam, The Netherlands.