A general family of overdispersed probability laws

J.F. Walhin[1] and **J. Paris**[2]

Abstract. Count data reported over one period of time often show overdispersion and infinite divisibility compared to the Poisson distribution. We propose a methodology which can be extended to pure birth processes derived from Poisson processes. Our model reveals itself as the most general of this type. Our family encompasses many distributions analysed separately and used in bonus-malus systems.

Keywords: Panjer's algorithm, mixed process, infinite divisibility, completely monotonic function, Poisson distribution, bonus-malus system, maximum likelihood, heterogeneity.

1 Introduction

Count data, as for example the number of claims reported to an insurance company during a period of time, occur in many practical problems. When only randomness looks present, they seem to be described by a Poisson distribution which is the first choice for non-negative integer valued random variables. In many cases, for example in motor third party liability, this choice is rejected by the usual Chi Square test for goodness of fit. It is immediately observed that the Poisson model underestimates the variance because overdispersion occurs (the variance is larger than the mean), indicating that the population heterogeneity of the drivers has not been taken into account by the Poisson model and its single parameter. This suggests that more parameters are needed to describe the distribution of the data. As the sequence of signs of the difference between observed and expected under the Poisson distribution is $+, -, +$, the result of Shaked (1980) reveals that it is natural to try to use a mixed Poisson distribution.

Moreover, indication of infinite divisibility of the data is suggested and especially when the frequency is low, an excess of zeroes relative to the Poisson distribution often arises.

Many models have been built to try to solve each part of the problem separately. Here we propose a methodology and a model to solve the problem as a whole. It is possible to build a model which is the more general of this type.

In many cases, data are collected in a single period of time which is taken as unity and so time is eliminated. Nevertheless, the model has to be extended to stochastic processes. Our model is not only adapted to this situation but also to data reported over several consecutive periods and, moreover, to the problem met in the industry sector where defaults are reported over periods of different lenghts and to the situation where several zero default are observed but not reported.

[1] Institut des Sciences Actuarielles, Université Catholique de Louvain and Secura Belgian Re, corresponding author, e-mail: jfw@secura-re.com

[2] Institut de Statistique, Université Catholique de Louvain

2 The motivation

The following kind of data are frequent in insurance, marketing, biometry, financial problems. As an example, we consider a typical Swiss motor portfolio (see table 1), published by Bühlmann (1970) and used by, among others, Lemaire (1985), Tremblay (1992), Denuit (1997), Walhin and Paris (1999b). As usual, the data set gives the number N_k of policyholders having reported k claims to the company during the period $(k = 0, 1, 2, \dots)$.

Number of accidents k	Number N_k of policyholders	Expected frequency under Poisson	Sign difference
0	103704	102629.55	$+$
1	14075	15921.95	$-$
2	1766	1235.07	$+$
3	255	63.87	
4	45	2.48	
5	6	0.07	
6	2	0.00	

Table 1. Reference portfolio

If we try to fit a Poisson distribution with parameter λ estimated by maximum likelihood : $\hat{\lambda} = \bar{N} = 0.15514$, the usual χ^2 statistic of goodness of fit, with 3 degrees of freedom is 2550.93 and leads to the rejection of the Poisson distribution. Note: for the grouping rule, in order to find the χ^2 statistic, we try to obey the rule B in Lemaire (1995), i.e. each theoretical frequency is at least 1 and 80% of the theoretical frequencies are at least 5.

Moreover, the sequence of signs of the differences between observed and expected is the Shaked one. This suggests a Poisson mixture.

The estimated variance $S^2 = 0.17931$ and the value of the statistic of the usual asymptotic Poisson overdispersion test based on

$$\sqrt{\frac{n-1}{2}}\left(\frac{S^2}{\bar{N}} - 1\right) \sim N(0,1) \tag{1}$$

(see Gart (1975) and Böhning (1994)), equivalent to the Fisher index of overdispersion, is 28.14, indicating a strong presence of overdispersion which is present in a mixed Poisson probability law and also in a compound Poisson probability law corresponding to an infinitely divisible probability law in the discrete case (see Feller (1968), Tome 1, CH XII, Section 3, p271).

Thus it is natural to analyse the presence of infinite divisibility in the data. In the discrete case, the expression $k_2k_4 - k_3^2$, where k_j is the cumulant of order j of the random variable, is zero in the Poisson case and strictly positive for any other infinitely divisible random variable. Using the k-statistic of Fisher as an estimation of the corresponding cumulant, we observe a value of 0.0149 indicating a possible strictly positive value. Using the corrected version of the statistic (see Gupta, Móri, Székely (1994))

$$k_2 k_4 - k_3^2 \sim N(0, \frac{8\lambda^4}{n}(1 + 12\lambda + 3\lambda^2)) \tag{2}$$

where the parameter λ of the Poisson is estimated by maximum likelihood, gives for our example an observed reduced value of 44.32 indicating a possible presence of inifinite divisibility as in a compound Poisson probability law.

The test for deviation in a single cell (see Rao (1973) p395) can be applied. In the particular case of the Poisson distribution, the test statistic for the deviation in the zero cell is based on

$$\frac{N_0 - ne^{-\hat{\lambda}}}{\sqrt{ne^{-\hat{\lambda}}(1 - e^{-\hat{\lambda}} - \hat{\lambda}e^{-\hat{\lambda}})}} \sim N(0, 1). \tag{3}$$

The observed value, 32.18, indicates a strong deviation in the zero cell.
This observation is in agreement with the preceding ones because, by Jensen's inequality, the mixed Poisson distribution, with mixing distribution $U(\lambda)$, has the following property

$$\int_0^\infty e^{-\lambda} dU(\lambda) \geq e^{-\int_0^\infty \lambda dU(\lambda)}. \tag{4}$$

3 The model

Let $N(t)$ be a counting (pure birth) process on the interval $(0, t]$ $(N(0) = 0)$. As the data are counts of accident insurance policies reporting a number of claims during a particular year, it is natural from the preceding considerations to suppose that $N(t)$ is an infinitely divisible mixed Poisson process (see Grandell (1997)) for which

$$\begin{aligned} \Pi(n, t) &= \mathbb{P}[N(t) = n] \\ &= \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} dU(\lambda) \quad , \quad n = 0, 1, \ldots \end{aligned} \tag{5}$$

Therefore

$$\Pi(0, t) = \mathbb{P}[N(t) = 0] = \int_0^\infty e^{-\lambda t} dU(\lambda) \tag{6}$$

is a completely monotonic function ($\Pi(0, t)$ possesses derivatives $\Pi^{(n)}(0, t)$ of all orders, and $(-1)^n \Pi^{(n)}(0, t) \geq 0 \quad \forall t > 0$) as a Laplace transform (see Feller (1971), Tome 2, CH XIII, Section 4, Theorem 1, page 439) and determines entirely the distribution of $N(t)$, for fixed t, because

$$\Pi(n, t) = (-1)^n \frac{t^n}{n!} \Pi^{(n)}(0, t). \tag{7}$$

For $N(t)$ to be infinitely divisible, it is sufficient that $U(\lambda)$ is infinitely divisible (see Maceda (1948)). Moreover Feller (1971), Tome 2, CH XIII, Section 7, Theorem 1, page 450, showed that $U(\lambda)$ is infinitely divisible if its Laplace transform,

$$\Pi(0, t) = \int_0^\infty e^{-\lambda t} dU(\lambda),$$

may be written as

$$\psi_{N(t)}(u) = e^{-\theta(t - tu)}, \tag{8}$$

where θ is a Bernstein function ($\theta \geq 0, \theta(0) = 0, \theta'$ is completely monotonic), which restricts the choice of completely monotonic functions $\Pi(0, t)$.

Due to the mixed Poisson representation of the model, the probability generating function of $N(t)$ is

$$\psi_{N(t)}(u) = e^{-\theta(t - tu)}. \tag{9}$$

Taking the derivative of the logarithm of ψ, we obtain

$$\frac{d}{du} \ln \psi_{N(t)}(u) = t\theta'(t - tu), \tag{10}$$

where the function $\theta'(t - tu)$, taken as a function of u, is absolutely monotonic (all derivatives of the function exist and are positive). Thus, it can be expanded in series as

$$\theta'(t - tu) = \sum_{n=0}^\infty r_n(t) u^n, \tag{11}$$

where the $r_n(t) \geq 0$. By equating terms with identical powers of u in the two members of

$$\frac{d}{du} \psi_{N(t)}(u) = \psi_{N(t)}(u) t\theta'(t - tu), \tag{12}$$

we obtain the recursive relationship

$$\begin{aligned} \Pi(0, t) &= e^{-\theta(t)}, \\ (n + 1)\Pi(n + 1, t) &= t \sum_{j=0}^n r_j(t) \Pi(n - j, t). \end{aligned} \tag{13}$$

These relationships are identical to those of Steutel (1973) (see also Katti (1967)) which characterize infinitely divisible probability distributions for discrete random variables.
The representation of $N(t)$ as a compound Poisson process of the type

$$N(t) = \Xi_1 + \cdots + \Xi_{L(t)}, \tag{14}$$

where the Ξ_i are iid (independent and identically distributed) integer valued random variables and $L(t)$ is Poisson distributed and independent of the Ξ_i is easily obtained from the probability generating function of $N(t)$ (see (9)).
As $1 - \frac{\theta(t - tu)}{\theta(t)}$ is an absolutely monotonic function taking the value 1 for $u = 1$, it is the probability generating function of an integer random variable (see Feller (1971), Tome 2, CH VII, Theorem 2, p223):

$$\psi_{\Xi(t)}(u) = 1 - \frac{\theta(t - tu)}{\theta(t)} = \sum_{j=1}^\infty p_{\Xi(t)}(j) u^j, \tag{15}$$

where

$$p_{\Xi(t)}(j) = \mathbb{P}[\Xi(t) = j].$$

We immediately see that

$$\begin{aligned} p_{\Xi(t)}(0) &= 0 \\ p_{\Xi(t)}(j) &= (-1)^{j-1} \frac{t^j}{j!} \frac{\frac{d^j}{dt^j}\theta(t)}{\theta(t)} \quad , \quad j = 1, 2, \ldots \end{aligned} \tag{16}$$

The probability distribution of the Ξ_i is immediately deduced from the knowledge of the function $\theta(t)$. The probability generating function (9) can also be written as

$$\psi_{N(t)}(u) = e^{-\theta(t)[1-\psi_{\Xi(t)}(u)]}. \tag{17}$$

$N(t)$ is distributed as the sum of a random number $L(t)$ (which is Poisson with mean $\theta(t)$) of iid random variables Ξ_i independent of $L(t)$ and distributed as $\Xi(t)$.

Remark: as driving abilities vary from individual to individual, the Poisson parameter λ changes and the mixing distribution U reflects the heterogeneity of the portfolio. Many models have been built with a particular choice of the function U but it is difficult to validate the choice because there are so many particularities of the drivers which have to be taken into account that the function U has to be continuous. As the observation bears on the number of claims during a single period, the basic paper of Simar (1976) indicates that the estimation of U is always purely discrete indicating only that the portfolio can be divided into a low number of homogeneous classes and consequently this procedure is not adapted to the situation. Denuit and Lambert (2000) try to solve this problem by a smoothed version of the Simar non-parametric maximum likelihood estimator.

The fact that the model is not rejected is not a proof of an appropriate choice because the model taken into account can have properties which are in contradiction with the natural properties of the real model.

The basic problem is the choice of the function θ (or θ') but to be able to estimate such a function with an infinite number of restrictions (a completely monotonic function has an infinity of derivatives with alternate sign) we need an infinite number of observations which is impossible. So to be efficient with the kind of data at our disposal, we need to restrict the number of difficulties and adopt a parametric form for θ. The choice is dictated by the following consideration: the main problem in risk theory is the determination of the probability distribution function of the random sum

$$S_{N(t)} = X_1 + X_2 + \cdots + X_{N(t)}, \tag{18}$$

where the X_i are iid and represent the cost of claims, $N(t)$ is the number of claims in $(0, t]$ and is supposed to be independent of the X_i.

The two main objectives are:

- an easy evaluation of the probability distribution function of $S_{N(t)}$ through a procedure which avoids the use of convolutions.
- an easy evaluation of the intensity of the process

$$\mathbb{E}[N(t+1) - N(t)|N(t)], \tag{19}$$

on which the premium for the period [t,t+1] is based.

The next section proposes a solution.

4 A remarkable family of probability distributions

In order to build an infinitely divisible mixed Poisson process, it is sufficient to choose the function $\theta'(t)$. The family characterized by

$$\theta'(t) = \frac{p}{(1+ct)^a} \quad , \quad p > 0\,,\ c > 0\,,\ a \geq 0, \tag{20}$$

has been introduced by Hofmann (1955) and used by Thyrion (1961), Kestemont and Paris (1985), Walhin and Paris (1999b).

By integration we immediately find

$$\begin{aligned} \theta(t) &= pt \quad \text{if } a = 0 \\ &= \frac{p}{c}\ln(1+ct) \quad \text{if } a = 1 \\ &= \frac{p}{c(1-a)}[(1+ct)^{1-a} - 1] \quad \text{if } a \neq 1. \end{aligned} \tag{21}$$

This formulation encompasses many probability distributions. The parameter a distinguishes between the different distributions. If t is fixed, for $a = 0$ we find the ordinary Poisson distribution, for $a = \frac{1}{2}$ the Poisson Inverse Gaussian distribution, for $a = 1$ the Negative Binomial distribution, for $a = 2$ the Polya-Aeppli distribution. The limiting case $a \to \infty$, $c \to 0$ such that $ac \to b$ gives the Neyman Type A distribution for which

$$\theta(t) = \frac{p}{b}(1 - e^{-bt}). \tag{22}$$

For notation purposes, we will write $Ho(p, c, a)$ for the distribution of this family with parameters p, c, a. This family can be divided into two parts:

1 the distributions with $0 \leq a \leq 1$. In this case

$$\lim_{t\to\infty} \theta(t) = +\infty. \tag{23}$$

From Tauberian results (see Feller (1971), Tome 2, CH XIII, Section 5), this leads to

$$\lim_{t\to\infty} \Pi(0,t) = 0 = \lim_{\lambda\to 0} U(\lambda), \tag{24}$$

which is the general situation in insurance. This result explains why the limiting cases, the Poisson distribution ($a = 0$) and the Negative Binomial distribution ($a = 1$) have received so much attention in counting distributions.

2 the distributions with $a > 1$. In this case

$$\lim_{t\to\infty} \theta(t) = d > 0, \tag{25}$$

where $d = \frac{p}{c(a-1)}$ and

$$\lim_{t\to\infty} \Pi(0,t) = e^{-d} = \lim_{\lambda\to 0} U(\lambda). \tag{26}$$

In this case the random variable Λ has a point mass at the origin.
Letting in this case

$$\theta(t) = d\left[1 - (1 - \frac{\theta(t)}{d})\right], \tag{27}$$

and defining

$$Q(0,t) = 1 - \frac{\theta(t)}{d}, \tag{28}$$

$Q(0,0) = 0$, $Q(0,t) \geq 0$, $Q(0,t)$ completely monotonic, we obtain

$$\Pi(0,t) = e^{-d[1-Q(0,t)]}, \tag{29}$$

and also another procedure of construction.
In the particular case $a > 1$ in the Hofmann process,

$$Q(0,t) = (1+ct)^{1-a} \tag{30}$$

is the Laplace transform of a Gamma random variable.

5 General properties

The factorial cumulant generating function for $N(t)$,

$$\begin{aligned} \ln \mathbb{E}[(1+u)^{N(t)}] &= \ln \Pi(0,-tu) \\ &= -\theta(-tu), \end{aligned} \tag{31}$$

yields the factorial cumulants

$$\begin{aligned} \kappa_1 &= pt, \\ \kappa_2 &= pcat^2, \\ \kappa_3 &= pc^2a(a+1)t^3, \\ \kappa_4 &= pc^3a(a+1)(a+2)t^4, \end{aligned} \tag{32}$$

from which we derive

$$\begin{aligned} \mathbb{E}N(t) &= pt, \\ \mathbb{V}arN(t) &= pt + pcat^2. \end{aligned} \tag{33}$$

Except for the case $a = 0$, we always have

$$\mathbb{V}arN(t) > \mathbb{E}N(t), \tag{34}$$

and for fixed c, the difference increases with a.
Different formulations of skewness and kurtosis can be obtained (see Walhin (2000)).
The ordinary Poisson distribution appears as a limiting case. If we consider a Poisson distribution with the same expected value, pt, as introduced above, that is

$$p(n,t) = e^{-pt}\frac{(pt)^n}{n!} \quad , \quad n = 0, 1, \ldots \tag{35}$$

we have the Jensen inequality

$$p(0,t) \geq \Pi(0,t), \tag{36}$$

which explains the excess of zeroes. In the same way (see Feller (1943)) we have

$$\frac{p(1,t)}{p(0,t)} = pt \geq \frac{\Pi(1,t)}{\Pi(0,t)} = t\theta^{'}(t), \tag{37}$$

and $\Pi(0,t)$ is an increasing function of a for which

$$\lim_{a\to 0} \Pi(0,t) = p(0,t). \tag{38}$$

From Chebyshev's inequality for integrals of decreasing functions (see Hardy et al. (1964)), we have

$$\Pi(0,t+s) \geq \Pi(0,t)\Pi(0,s), \tag{39}$$

which is an evident result in insurance. Indeed one expects that the probability of no claim for a policyholder over a long period $(t+s)$ is larger than the probability of no claim for two policyholders over periods t and s respectively.
The Cauchy-Schwartz inequality gives

$$\frac{n\Pi(n,t)}{\Pi(n-1,t)} \leq \frac{(n+1)\Pi(n+1,t)}{\Pi(n,t)}, \tag{40}$$

whereas these relationships are constant for the Poisson distribution.

Let us assume that $N \sim Ho(p,c,a)$. Let us introduce the random variable

$$\begin{aligned} Y_i &= 1 \text{ if } X_i > D, \\ &= 0 \text{ if } X_i \leq D. \end{aligned}$$

If we count the number of independent events with characteristic $X_i > D$ amongst the N, that is

$$N^{'} = Y_1 + Y_2 + \cdots + Y_N, \tag{41}$$

we have

$$\begin{aligned} \psi_{N'}(u) &= \psi_N(\psi_Y(u)) \\ &= e^{-\theta(t(1-F_X(D))(1-u)}, \end{aligned} \tag{42}$$

which shows that

$$N^{'} \sim Ho(p(1-F_X(D)), c(1-F_X(D)), a). \tag{43}$$

From (16) and

$$\begin{aligned} \frac{d^n}{dt^n}\theta(t) &= (-1)^{n-1}pc^{n-1}\frac{\Gamma(a+n-1)}{\Gamma(a)}(1+ct)^{1-a-n} \quad , \\ & a > 0, \end{aligned} \tag{44}$$

we can deduce the probability distribution of the Ξ_i. It is summarized in table 2.

Parameter a	Distribution of $\Xi(t)$
$a = 0$	Degenerate $\mathbb{P}[\Xi(t) = 1] = 1$
$0 < a < 1$	Extended Truncated Negative Binomial
$a = 1$	Logarithmic Distribution
$a > 1$	Truncated Negative Binomial
$a \to \infty$	Truncated Poisson

Table 2. Candidates for the distribution of $\Xi(t)$

We can verify that

$$\frac{p_{\Xi(t)}(n)}{p_{\Xi(t)}(n-1)} = \frac{ct}{1+ct}(1 + \frac{a-2}{n}) \quad , \quad n > 1, \tag{45}$$

and so the probability distribution of the Ξ_i belongs to the $(\alpha, \beta, 1)$ class to which an extended Panjer algorithm applies (see Sundt and Jewell (1981)). Note that $(\alpha, \beta, 1)$ is a reparameterization for the classical notation $(a, b, 1)$ due to the parameter a already being used within the Hofmann distribution.

The probability distribution of $N(t)$ is easily available as

$$\begin{array}{rcl} \Pi(0,t) & = & e^{-\theta(t)}, \\ \Pi(x+1,t) & = & \frac{pt}{(1+ct)^a} \sum_{i=0}^{x} \left(\frac{ct}{1+ct}\right)^i \frac{\Gamma(a+i)\Pi(x-i,t)}{i!(x+1)\Gamma(a)} \quad , \\ & & x \geq 0 \end{array} \tag{46}$$

Unfortunately, this formula requires knowledge of all the probabilities. In the particular cases $a = 0, \frac{1}{2}, 1$, simplified stable recursions are known. For $a = 2$, a simple formula is due to Evans (1953). However this formula is not stable. It does not seem possible to find a simple (stable) formula for the general case.

Although the Hofmann distribution (with $a \notin \{0, 1\}$) does not belong to the (α, β, m) class, the probability distribution of the random sum

$$S_{N(t)} = X_1 + X_2 + \cdots + X_{N(t)}, \tag{47}$$

may be computed recursively by introducing the auxiliary random variable

$$V(t) = X_1 + X_2 + \cdots + X_{\Xi(t)}, \tag{48}$$

whose probability distribution can be obtained by the extended Panjer algorithm. In a second step, we obtain the probability distribution of $S_{N(t)}$ as

$$S_{N(t)} = V_1(t) + V_2(t) + \cdots + V_{L(t)}(t), \tag{49}$$

where the random variables $V_i(t)$ are independent and have the same distribution as $V(t)$. The distribution of $S(t)$ is easy to obtain because $L(t)$ is Poisson distributed and so $S_{N(t)}$ can be evaluated by the Panjer algorithm. Whenever the X_i are integer random variables, we have

$$\begin{array}{rcl} f_{V(t)}(0) & = & 1 - \frac{\theta(t - t f_X(0))}{\theta(t)} \\ f_{V(t)}(x) & = & \frac{1}{1 - \frac{ct}{1+ct} f_X(0)} \Big[p_{\Xi(t)}(1) f_X(x) \\ & & + \sum_{i=1}^{x} (1 + (a-2)\frac{i}{x}) f_X(i) f_{V(t)}(x-i) \Big] \quad , \\ & & x \geq 1 \end{array} \tag{50}$$

where

$$\begin{array}{rcl} f_{V(t)}(x) & = & \mathbb{P}[V(t) = x], \\ f_{S(t)}(x) & = & \mathbb{P}[S(t) = x]. \end{array}$$

6 Bayesian analysis of the model

As $\Pi(0, t)$ determines completely the distribution function U of the random variable Λ, the cumulant generating function of Λ is

$$\ln \mathbb{E}(e^{u\Lambda}) = \ln \Pi(0, -u) = -\theta(-u), \tag{51}$$

from which we deduce the cumulants of Λ in the Hofmann process:

$$\begin{array}{rcl} \kappa_1 & = & p, \\ \kappa_j & = & p(ac)^{j-1}(1+\frac{1}{a})(1+\frac{2}{a})\ldots(1+\frac{j-2}{a}) \quad , \quad j \geq 2. \end{array} \tag{52}$$

In particular we find

$$\begin{array}{rcl} \mathbb{E}\Lambda & = & p, \\ \mathbb{V}ar\Lambda & = & pac, \end{array}$$

and the coefficient of variation of Λ, $\sqrt{\frac{ac}{p}}$, is an a priori measure of the heterogeneity of the portfolio.
If we observe the stochastic process $N(t)$ during the period, we can, a priori, deduce the intensity of the process

$$\begin{array}{rcl} \mathbb{E}[N(t+1) - N(t) | N(t) = k] & = & \frac{k+1}{t} \frac{\Pi(k+1,t)}{\Pi(k,t)} \\ & = & \mathbb{E}[\Lambda | N(t) = k]. \end{array} \tag{53}$$

Comparing this expression with $\mathbb{E}\Lambda = \mathbb{E}N(1)$ gives us the possibility to set up a well-balanced bonus-malus system. Alternatives exist (see Walhin and Paris (1999a) or Denuit and Dhaene (2000)).
The following result is surprising

$$\mathbb{E}[\Lambda | N(t) = 0] = \theta'(t). \tag{54}$$

Indeed, knowing the history of policyholders without claims allows to determine the distribution of $N(t)$, as well as the distribution of Λ, with the help of a single regression function. This regression function characterizes the probability distribution of the two random variables $N(t)$ and Λ and solves many problems of characterization.
In the same way, we have

$$\mathbb{V}ar[\Lambda | N(t) = 0] = -\theta''(t), \tag{55}$$

and thus, it is possible to define the a posteriori coefficient of variation of Λ for that particular case and to compare both.
Remark: formula (54) indicates that in a stable portfolio it is sufficient to know the behaviour over a long period of drivers who do not report claims to be able to know the model.

7 Extensions

Other parameterizations exist for the Hofmann distribution: the Generalized Poisson Pascal distribution (see Panjer and Willmot (1992)) and those proposed by Hougaard et al. (1997). Our modelization has the advantage of simplicity and it easily allows extensions.

7.A. Changing frequency

As the process is stationary, we have

$$\mathbb{E}[N(t+1) - N(t)] = \mathbb{E}N(1) = p, \tag{56}$$

but the introduction of a bonus-malus system may lead to a reduction of the frequency of the claims reported and so we can have

$$\mathbb{E}[N^*(t+1) - N^*(t)] = pv^t \quad , \quad 0 < v < 1. \tag{57}$$

A new model can be built and the new parameter v can be estimated if we have at our disposal data reported over a sufficient number of successive years. This type of extension has been proposed by Besson and Partrat (1992) within the Negative Binomial and Poisson Inverse Gaussian models. It is easy to show that we have

$$\mathbb{P}[N^*(t) = n] = \Pi(n, \frac{1-v^t}{1-v}). \tag{58}$$

7.B. The most general model

In this subsection, we work with a Hofmann process and we exclude the Poisson case, i.e. $a > 0$.
As $\theta^{'}(t)$ is a decreasing function of t, we have

$$\lim_{t\to\infty} \theta^{'}(t) = 0. \tag{59}$$

In this case, from formula (54) the drivers who reported zero claim during a long period will pay a very low premium for the period $[t, t+1]$. This is unacceptable. As it happens, even the best drivers may cause an accident. So we must apply a basic minimum premium and replace $\theta^{'}(t)$ by

$$\theta_1^{'}(t) = \delta + \theta^{'}(t). \tag{60}$$

The function $\theta_1^{'}(t)$ is also completely monotonic and from a general result on completely monotonic function (see Berg and Forst (1975)) this is the most general situation of this type. The corresponding counting process $N_1(t)$ is the sum of two independent components: the first $N^*(t)$ is a simple Poisson process with mean δt which describes the purely random part and the second is a mixed infinitely divisible Poisson process related to the behaviour of the drivers.
Let $\Pi_1(n,t)$ be the probability law of $N_1(t)$. It is easy to evaluate this distribution as well as the compound distribution of

$$X_1 + \cdots + X_{N_1(t)}$$

(see Walhin (2000) for details).
The extended Hofmann process has four parameters. The new parameter δ is identifiable because if data are available over a long period, we have

$$\begin{array}{rcl} \lim\limits_{t\to\infty} \frac{\Pi_1(0,t+1)}{\Pi_1(0,t)} & = & e^{-\delta}, \\ \lim\limits_{t\to\infty} \frac{1}{t}\frac{\Pi_1(1,t)}{\Pi_1(0,t)} & = & \delta. \end{array} \tag{61}$$

In the general model we always have

$$\lim_{t\to\infty} \Pi(0,t) = 0. \tag{62}$$

Remark: when $N(t)$ is Negative Binomial distributed, one speaks of the Lüders distribution (see Lüders (1959)). In the case $N(t)$ is Neyman Type A distributed, the resulting distribution is called the Short distribution (introduced by Cresswell and Frogatt (1963)).

8 Estimation of the parameters

The Hofmann process has three parameters, one of them distinguishes between the different probability laws. Several methods can be proposed for the estimation of the parameters. The moment method is inadequate for non-negative integer valued random variables. So the first method we apply is the maximum likelihood. From Hürliman (1990) we know that the sample mean is the maximum likelihood estimator of the parameter p but it is not possible to obtain explicit formulae for the estimators of a and c. They can be obtained numerically and their efficiency can be estimated.
We fix the time to $t = 1$ as the observation is on a one year basis.
We will denote by $as\sigma_x$ the asymptotic standard deviation of the estimate of x and by $as\rho(x,y)$ the asymptotic correlation between the estimates of x and y.
Results for our reference portfolio are given in table 3.

	Obs	Ho				
0	103704	103704.60	l	-54609.59	$as\sigma_p$	0.0012
1	14075	14072.52	χ^2	0.434	$as\sigma_c$	0.0687
2	1766	1769.26	df	2	$as\sigma_a$	0.0824
3	255	255.23	$p-value$	0.574	$as\rho(p,c)$	0.0463
4	45	41.98	p	0.15514	$as\rho(p,a)$	0.0002
5	6	7.58	c	0.3480	$as\rho(c,a)$	-0.9718
6	2	1.46	a	0.4483		
7		0.29				
8		0.06				

Table 3. Hofmann fit

We note that the asymptotic correlation between the estimates of p and a is almost 0. This property has been observed in every used data set. It is a good property saying that the estimates of p and a are almost independent, which strenghtens our opinion that a chooses the distribution whereas p is merely the average claims frequency. The asymptotic correlation between the estimates of c and a is almost -1, which is not surprising as we know that the product ac is part of the variance. So we expect the estimates of these two parameters to move in opposite directions.
Note that the asymptotic variance-covariance matrix has been obtained as the inverse of the observed Information matrix, i.e. the negative of the second partial derivative of the log-likelihood. A better way to estimate this variance-covariance matrix is to compute the Fisher Information matrix, which is based only on the first partial derivatives of the log of the $\Pi(n,1)$. The elements of the Fisher Information matrix write

$$a_{ij}(\theta) = n\sum_{k=0}^{\infty} \Pi(k,1)(\frac{\partial}{\partial\theta_i}\ln\Pi(k,1))(\frac{\partial}{\partial\theta_j}\ln\Pi(k,1)), \tag{63}$$

where θ is the vector of parameters to be estimated.
Special attention should be paid here as the infinite sum will have to be truncated. Results for with this method are given in table 4.

$as\sigma_p$	$as\sigma_c$	$as\sigma_a$	$as\rho(p,c)$	$as\rho(p,a)$	$as\rho(c,a)$
0.001223	0.068904	0.082642	0.046447	0	-0.971996

Table 4. Asymptotic variance-covariance by the Fisher Information

Note that it is necessary to truncate the sum in k far enough in order to get convergence. In particular this is true for the $as\rho(p, a)$. Moreover, derivating the $\Pi(k, 1)$ with respect to the parameters p, c and a is a tedious task. Fortunately, formulae giving these derivatives are given in Panjer and Willmot (1992) for the parameterization under the form of the Generalized Poisson Pascal Distribution (see section 7). Having the variance-covariance matrix within this parameterization , it is not difficult to find the variance-covariance matrix under the Hofmann parameterization (see also Panjer and Willmot (1992) page 313).

It turns out that $\hat{p}$ and $\hat{a}$ are asymptotically uncorrelated. This fact seems to be true even for very small samples and for any values of the parameters. Nevertheless this is only a conjecture. It remains to be proved.

Another method of estimation is the minimum χ^2 method presented by Berkson (1980) as an alternative to maximum likelihood. We will not use this method because of the thickness of the right tail of the distribution and the difficulties related to the grouping of the classes with low frequency. We therefore prefer the alternative method proposed by Kestemont and Paris (1985) with equivalent properties as the maximum likelihood: estimate p by the sample mean and the parameters a and c with the relations

$$\begin{aligned} \frac{n_0}{n} &= e^{-\theta(1)}, \\ \frac{n_1}{n} &= \frac{d}{dt}\theta(t)\Big|_{t=1} e^{-\theta(1)}, \end{aligned} \tag{64}$$

where n_i is the number of observations in the class i.

This methodology attaches the greatest importance to the two claims with the most important strength.

It is possible to analyse the quality of this procedure with the two statistics

$$\begin{aligned} T &= k_2 - p(1 + ac) \\ V &= k_3 - (pc^2 a(a + 1) + 3pca + p) \end{aligned} \tag{65}$$

where k_2 and k_3 are the Fisher k statistics and so are $\sim N\left(0, o(\frac{1}{n})\right)$.

Results for our numerical example are given in table 5.

	Obs	Ho		
0	103704	103704.00	l	-54609.60
1	14075	14075.00	χ^2	0.438
2	1766	1766.78	df	2
3	255	255.39	$p - value$	0.803
4	45	42.26	T	-0.00006
5	6	7.69	V	-0.00076
6	2	1.50	p	0.15514
7		0.30	c	0.3546
8		0.06	a	0.4406

Table 5. Hofmann fit with the proportion estimation method

We study in table 6 the case of the Hofmann + Poisson Distribution.

	Obs	Ho + Po				
0	103704	103703.67	l	-54609.53	$as\sigma_\delta$	0.1244
1	14075	14076.17	χ^2	1.16	$as\sigma_p$	0.1244
2	1766	1763.51	df	1	$as\sigma_c$	0.3319
3	255	258.60	$p - value$	0.280	$as\sigma_a$	0.2600
4	45	42.18	δ	0.0524	$as\rho(\delta, p)$	-0.9999
5	6	7.28	p	0.1027	$as\rho(\delta, c)$	-0.9805
6	2	1.30	c	0.2581	$as\rho(\delta, a)$	0.9952
7		0.24	a	0.9119	$as\rho(p, c)$	0.9805
8		0.04			$as\rho(p, a)$	-0.9952
					$as\rho(c, a)$	-0.9947

Table 6. Hofmann+Poisson fit

We note that the results are very bad. The asymptotic standard deviations are high and the asymptotic correlations are near 1 in absolute value. This is due to the fact that the model finds it difficult to separate the effect of δ from p. At the limit, in model $Po + Po$, we would have a problem of identification between the two parameters. Another reason is perhaps that introducing parameter δ makes the model overparameterized and leads to instability of the estimates.

We are not surprised that the asymptotic correlation between δ and a is near 1. Indeed, as δ represents the pure Poisson part of the process, it is clear that if we give a great importance to the Poisson part, the model will tend to a great a in order to compensate the effect of δ.

To conclude, we can say that the extended model has to be interpreted with care due to the high sampling errors. Nevertheless, in the absence of data over a longer period, maximum likelihood estimation is the best we can do.

9 Conclusion

This paper presents a model for the number of claims which encompasses a large number of probability laws. The model is tractable and has a lot of attractive properties which make it suitable for application not only in the insurance domain but also in many other fields where overdispersion relative to the Poisson model is observed.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referee for the various remarks which have led to a better presentation of the paper.

REFERENCES

[1] Berg, C. and Forst, G. (1975), Potential Theory on Locally Compact Abelian Groups, Springer-Verlag, Berlin

[2] Berkson, J. (1980), Minimum Chi-Square, not Maximum Likelihood, The Annals of Statistics, 8, 457-487

[3] Besson, J.L. and Partrat, C. (1992), Trend et Systèmes de Bonus-Malus, Astin Bulletin, 22, 11-31

[4] Böhning, D. (1994), A Note on a Test for Poisson Overdispersion, Biometrika, 81, 418-419

[5] Bühlmann, H. (1970), Mathematical Models in Risk Theory, Springer-Verlag, Berlin

[6] Cresswell, W.L. and Frogatt, P. (1963), The Causation of Bus Driver Accidents: An Epidemiological Study, Oxford University Press, London

[7] Delaporte, P. (1959), Quelques Problèmes de Statistique Mathématique Posés par l'Assurance et le Bonus pour Non Sinistre, Bulletin Trimestriel de l'Institut des Actuaires Français, 227, 87-102

[8] Denuit, M. (1997), A New Distribution of Poisson-Type for the Number of Claims, Astin Bulletin, 27, 229-242

[9] Denuit, M. and Dhaene, J. (2001), Bonus-Malus Scales Using Exponential Loss Functions, German Actuarial Bulletin, 25, 13-27

[10] Denuit, M. and Lambert, P. (2001), Smoothed NPML Estimation of the Risk Distribution Underlying Bonus-Malus Systems, Proceedings of the Casualty Actuarial Society, to appear

[11] Evans, D.A. (1953), Experimental Evidence Concerning Contagious Distributions in Ecology, Biometrika, 40, 186-211

[12] Feller, W. (1943), On a General Class of Contagious Distributions, Annals of Mathematical Statistics, 14, 389-400

[13] Feller, W. (1968), An Introduction to Probability Theory and its Applications Vol I (3ed), Wiley, New-York

[14] Feller, W. (1971), An Introduction to Probability Theory and its Applications Vol II (2ed), Wiley, New-York

[15] Gart, J.J. (1975), The Poisson Distribution: The Theory and Applications of some Conditional Tests, in Patil, G.P. et al (1975), 2, 125-140

[16] Grandell, J. (1997), Mixed Poisson Processes, Chapman and Hall

[17] Gupta, A.K., Móri, T.F. and Székely, G.J. (1994), Testing for Poissonity-Normality versus Other Infinite Divisibility, Statistic and Probability Letters, 19, 245-248

[18] Hardy, G.H., Littlewood, J.E. and Polya, G. (1964), Inequalities, Cambridge University Press

[19] Hofmann, M. (1955), Über zusammengesetzte Poisson-Prozesse und ihre Anwendungen in der Unfallversicherung, Bulletin of the Swiss Association of Actuaries, 55, 499-575

[20] Hougaard, P., Ting Lee, M.L. and Whitmore, G.A. (1997), Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes, Biometrics, 53, 1225-1238

[21] Hürlimann, W. (1990), On Maximum Likelihood Estimation for Count Data Models, Insurance : Mathematics and Economics, 9, 39-49

[22] Katti, S.K. (1967), Infinite Divisibility of Integer Valued Random Variables, Annals of Mathematical Statistics, 38, 1306-1308

[23] Kestemont, R.M. and Paris, J. (1985), Sur l'Ajustement du Nombre de Sinistres, Bulletin of the Swiss Association of Actuaries, 85, 157-163

[24] Lemaire, J. (1985), Automobile Insurance: Actuarial Models, Kluwer-Nijhoff, Netherlands

[25] Lemaire, J. (1995), Bonus-Malus Systems in Automobile Insurance, Kluwer-Boston

[26] Lüders, R. (1934), Die Statistik der seltenen Ereignisse, Biometrika, 26, 108-128

[27] Maceda, E.C. (1948), On the Compound and Generalized Poisson Distributions, Annals of Mathematical Statistics, 19, 414-416

[28] Patil, G.P., Kotz, S. and Ord, J.K. (1975), Statistical Distributions in Scientific Work, D. Reidel Publishing Company

[29] Panjer, H.H. and Willmot, G.E. (1992), Insurance Risk Models, Society of Actuaries

[30] Rao, C. (1973), Linear Statistical Inference and Its Applications, Wiley

[31] Shaked, M. (1980), On Mixtures from Exponential Families, Journal of the Royal Statistical Society B, 42, 192-198

[32] Simar, L. (1976), Maximum Likelihood Estimation of a Compound Poisson Process, The Annals of Statistics, 4, 1200-1209

[33] Steutel, F.W. (1973), Some Recent Results in Infinite Divisibility, Stochastic Processes and their Applications, 1, 123-143

[34] Sundt, B. and Jewell, W.S. (1981), Further Results on Recursive Evaluation of Compound Distributions, Astin Bulletin, 12, 27-39

[35] Thyrion, P. (1961), Contribution à l'Etude du Bonus pour non Sinistre en Assurance Automobile, Astin Bulletin, 1, 142-162

[36] Tremblay, L. (1992), Using the Poisson Inverse Gaussian Distribution in Bonus-Malus Systems, Astin Bulletin, 22, 97-106

[37] Walhin, J.F. and Paris, J. (1999a), Processus de Poisson Mélange et Formules Unifiées Pour Systèmes Bonus-Malus, Bulletin Français d'Actuariat, 3, 35-43

[38] Walhin, J.F. and Paris, J. (1999b), Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems, Astin Bulletin, 29, 81-99

[39] Walhin, J.F. (2000), Recursions for Actuaries with Applications in the Field of Reinsurance and Bonus-Malus Systems, PhD thesis, Institut de Statistique, Université Catholique de Louvain