The actual claim amount and frequency distributions within a bonus-malus system

J.F. Walhin*• and J. Paris*

∗ *Université Catholique de Louvain*

• *Secura Belgian Re*

Abstract

We have applied the algorithm of Lemaire and the non-parametric mixed Poisson fit to a motor portfolio in order to find the actual claim amount and frequency distributions that are affected by the presence of a bonus-malus system due to the desire developed by drivers for bonus.

Keywords

Mixed Poisson distribution, non-parametric fit, bonus-malus system, hunger for bonus, claim amount distribution, frequency distribution, censoring, maximum likelihood.

1. Introduction

When an insurer uses a bonus-malus system independent of the claim amounts, he will notice a tendency with the insureds not to report the smaller claims. Indeed it is in some cases more interesting for the policyholder to bear himself the cost of Third Party than to report the claim and to pay higher premiums in the future because of the malus. Lemaire (1977) called this fact the hunger for bonus. See also Lemaire (1995).

The hunger for bonus induces that the introduction of a (new) bonus-malus system creates a censored view of the claim amount and frequency distributions. Indeed some of the lower claim amounts will not be reported to the Insurance Companies. Of course, for the policyholder, the natural question is : "up to which level of claim amount is it interesting for me to bear the cost myself ?"

Lemaire (1977) answers this question by using an algorithm related to dynamic programming.

In this paper, we will apply Lemaire's (1977) algorithm and the non-parametric Mixed Poisson fit to a motor portfolio (see Walhin and Paris (1999)) in order to redefine the actual claim amount and frequency distributions. This problem was already implicitely posed by Lemaire (1977) in his paper where he states that he had to use old claim amount data because of recent data being influenced by the introduction of the bonus-malus system.

Throughout the paper we will use a numerical example. The data associated with this example are

a) observed frequency distribution

Number of accidents	Number of policyholders
0	103704
1	14075
2	1766
3	255
4	45
5	6
6	2

Table 1: Observed frequency distribution

This reference portfolio has already been used in Walhin and Paris (1999) where we were looking for parametric and non-parametric mixed Poisson fits.

b) observed claim amount distribution

6	6	10	11	17	18	20	26	27	34
42	44	47	54	59	60	61	61	61	61
64	64	65	66	67	68	71	71	73	75
76	81	85	87	93	94	101	103	105	109
110	110	113	116	116	129	134	134	141	141
151	154	156	159	167	171	172	173	174	179
181	183	185	187	195	195	203	226	235	240
251	255	273	340						

Table 2: Observed claim amount distribution

This is a small hypothetical data set we use for pedagogical purpose. The whole numerical application of the paper can be performed with the data sets given in the introduction.

c) bonus-malus system used by the company (or by the market as is still the case in Belgium).
The bonus-malus system we use is the one derived in Walhin and Paris (1999) :

- $s = 9$: 9 classes numbered $0, 1, \cdots, 8$. 0 is the minimum class. 8 is the maximum class.
- Entry of the system is in class 4.
- In case of a claims free year, the policyholder comes down one class
- In case of claim(s), the policyholder climbs up 3 classes per claim.
- The bonuses and maluses (in percentage) are given in the following table :

s	0	1	2	3	4	5	6	7	8
C_s	75	80	90	95	100	150	170	185	250

Table 3: Percentage premiums

The paper is structured as follows : section 2 briefly describes Lemaire's algorithm. Section 3 recalls the non-parametric mixed Poisson fit while section 4 recalls an efficient way to find the stationary distribution of the policyholders within a bonus-malus system. In section 5 we formulate our problem and a solution is given in section 6 in the form of an algorithm. Section 7 is devoted to the numerical solution of the problem with the data sets presented in the introduction. The conclusion is given in section 8.

2. Lemaire's algorithm

The algorithm of Lemaire (1977) needs the following hypothesis:

Let a bonus-malus system be with s classes : $i = 0, \cdots, s-1$
the claims frequency of a policyholder be Poisson distributed
the claim amount distribution be X, with cumulative density function (cdf) $F_X(x)$
β be the actualisation rate forecast for the future
P be the commercial premium, i.e. the base premium at level 100%, including security loading, administration expenses, profit and taxes.
$1 - t$ with $0 \leq t < 1$ be the time remaining until the next premium payment
m be the number of claims reported to the Company in $[0, t)$

With this hypothesis an iterative algorithm can be performed to find the optimal policy of the driver as a function of his bonus-malus level. The optimal policy is merely the optimal retention of the driver in function of his bonus-malus level. It is the level of claim amount up to which it is interesting for the policyholder to bear the cost himself and not to report the claim to the company. Of course, the optimal policy is also a function of t, the time at which the claim occurs and m, the number of claims reported before t unless one works at time $t = 0$. Optimal frequencies of the driver are also given by the algorithm.

The solution of the algorithm is shown to be unique if $\beta < 1$, which is always the case if the interest rate is positive.

In short, the algorithm of Lemaire gives the optimal frequency and the optimal retention of a driver based on the actual claim amount and frequency distributions of the driver.

3. Non-parametric Mixed Poisson fit

The Mixed Poisson distribution is often used to model the number of claims in a motor portfolio.
Let $\Pi(k,t)$ be the probability that a risk causes k accidents in t years. We have

$$\Pi(k,t) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dU(\lambda) \quad , \quad k \geq 0$$

Classical distributions are the Negative Binomial and the Poisson Inverse Gaussian.
In Walhin and Paris (1999), a non-parametric fit for $\Pi(k,1)$ is discussed. We recall that the maximum likelihood of the non-parametric fit is attained by a mixture of some Poisson distributions, depending on the form of the portfolio.

We have

$$\Pi(k,1) = \sum_{j=1}^{r} p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!} \quad , \quad k \geq 0$$
$$0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_r$$

If we define u as the maximum number of claims per risk and v as the number of classes for which the observation is different from 0 then the maximum likelihood is unique under the following conditions :

$$\begin{aligned} r &\leq \min\left(v, \left\lfloor \frac{u+2}{2} \right\rfloor\right) \quad \text{if } \lambda_1 = 0 \\ r &\leq \min\left(v, \left\lfloor \frac{u+1}{2} \right\rfloor\right) \quad \text{if } \lambda_1 > 0 \end{aligned}$$

For the example described in the introduction, we find

$\lambda_1 = 0.05461$	$p_1 = 0.56189$
$\lambda_2 = 0.24599$	$p_2 = 0.41463$
$\lambda_3 = 0.95618$	$p_3 = 0.02348$

Table 4: Maximum likelihood estimation for the reference portfolio

This fit gives an interesting interpretation of the portfolio : there are 3 types of risks λ_j , $j = 1, 2, 3$ with probabilities p_j , $j = 1, 2, 3$.

4. The stationary distribution of the policyholders within a bonus-malus system

In order to perform our calculations we will assume that the bonus-malus has existed for a long time and that it has reached its stationary distribution.
Let $p(x)$ be the probability that a driver with average frequency λ causes x claims during a given year. The transition probability matrix (Q) of this driver within the bonus-malus system described in the introduction is thus

s	0	1	2	3	4	5	6	7	8
0	$p(0)$	0	0	$p(1)$	0	0	$p(2)$	0	$1-p(0)-p(1)-p(2)$
1	$p(0)$	0	0	0	$p(1)$	0	0	$p(2)$	$1-p(0)-p(1)-p(2)$
2	0	$p(0)$	0	0	0	$p(1)$	0	0	$1-p(0)-p(1)$
3	0	0	$p(0)$	0	0	0	$p(1)$	0	$1-p(0)-p(1)$
4	0	0	0	$p(0)$	0	0	0	$p(1)$	$1-p(0)-p(1)$
5	0	0	0	0	$p(0)$	0	0	0	$1-p(0)$
6	0	0	0	0	0	$p(0)$	0	0	$1-p(0)$
7	0	0	0	0	0	0	$p(0)$	0	$1-p(0)$
8	0	0	0	0	0	0	0	$p(0)$	$1-p(0)$

Table 5: Transition probability matrix

As is the case for each bonus-malus system, we have an irreducible (there are no cycles) Markov chain whose states are all ergodic (each state can be attained from another state). Under those conditions, there is a stationary probability distribution that is given by :

$$\mathbf{e}_\infty(\lambda) = \lim_{n\to\infty} Q^n \mathbf{e_0}(\lambda)$$

where $\mathbf{e_0}(\lambda)$ denotes any initial distribution of the drivers in the bonus-malus system. The stationary probability distribution is independent of $\mathbf{e_0}(\lambda)$.
The stationary probability distribution is also given by solving

$$\mathbf{e}_\infty(\lambda) = \mathbf{e}_\infty(\lambda) Q$$

with the norming condition

$$\sum_{i=0}^{s-1} e_\infty(i;\lambda) = 1$$

where $e_\infty(i;\lambda)$ is the i^{th} element of the vector $\mathbf{e}_\infty(\lambda)$.
If one is interested in the stationary distribution of the portfolio ($\mathbf{e}_\infty$), we only have to take

the weighted average of the stationary distributions for the different types of policyholders (see Walhin and Paris (1999) for details). With our non-parametric mixed Poisson fit, we have

$$\mathbf{e}_\infty = \sum_{i=1}^{r} p_i \mathbf{e}_\infty(\lambda_i)$$

For our numerical example we find

	$\lambda = 0,05461$	$\lambda = 0,24600$	$\lambda = 0,95619$	Portfolio
0	$0,8278$	$0,2598$	$0,0005$	$0,5728$
1	$0,0464$	$0,0724$	$0,0008$	$0,0561$
2	$0,0490$	$0,0926$	$0,0022$	$0,0660$
3	$0,0518$	$0,1185$	$0,0057$	$0,0783$
4	$0,0095$	$0,0876$	$0,0145$	$0,0420$
5	$0,0075$	$0,0942$	$0,0369$	$0,0441$
6	$0,0052$	$0,0977$	$0,0939$	$0,0457$
7	$0,0014$	$0,0880$	$0,2386$	$0,0429$
8	$0,0009$	$0,0888$	$0,6066$	$0,0516$

Table 6: Stationary distribution of the drivers

Note that if the stationary distribution has not yet been reached, it is not a problem to work with the transient probabilities. The distribution of the drivers within the bonus-malus system after T years is given by

$$e_T(\lambda) = Q^T e_0(\lambda)$$

5. Formulation of the problem

When collecting data on a market where a bonus-malus system is in use, we do not observe the actual claim amount and frequency distributions. Indeed they are influenced by the hunger for bonus.

The actual claim amount distribution should have a lower mean while the actual claims frequency distribution should have a higher mean.

Let us assume that there is a proportion p of the driving population that reports all accidents while $(1-p)\%$ only report the claims exceeding the optimal retention given by Lemaire's algorithm.
Let us assume that a non-parametric fit for the frequency distribution has been performed on the observed portfolio (i.e. the reported claims). It reveals r types of risks λ_j with probability p_j $(j = 1, \cdots, r)$. This distribution (N) is not the distribution of the number of accidents but the distribution of the number of accidents reported to the Company.
The density function (df) of N writes

$$\Pi(k,1) = \sum_{j=1}^{r} p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!} \quad , \quad k = 0, 1, 2, \cdots$$

Let N' be the actual distribution of the number of claims. Its df writes :

$$\Pi'(k,1) = \sum_{j=1}^{r} p_j' e^{-\lambda_j'} \frac{(\lambda_j')^k}{k!} \quad , \quad k = 0, 1, 2, \cdots$$

We will assume that $p_j' = p_j \quad \forall j$ i.e. that the proportions of different risks for both distributions are the same.

Let X be the random variable representing the reported claim amounts.
Let Z be the random variable representing the actual claim amounts. This random variable is unknown while X is the observed one.
The df of X is a function of the df of Z and writes

$$f_X(x) = p f_Z(x) + (1-p) \frac{f_Z(x)}{1 - F_Z(c)} \mathbb{I}_{\{x \geq c\}} \quad , \quad x \geq 0$$

where c is the average retention limit of the portfolio.

Our aim is to find the distribution of Z and N'.

6. An algorithm to solve the problem

A solution to the problem described in section 5 will be given by means of an iterative algorithm using the non-parametric fit of the portfolio and an inversion of the algorithm of Lemaire (1977).

Step 0 : Initializing step

Have an initial guess for the parameter c. Choose a parametric distribution to fit the random variable Z.

Step 1 : Correction of the claim amount distribution

We use the average optimal retention (c) as a censor in order to find a new estimate for the parameters of the distribution of Z. Therefore we maximize the likelihood :

$$L(\theta, p|c) = \prod_{i=1}^{n} \left[p f_Z(x_i;\theta) + (1-p)\frac{f_Z(x_i;\theta)}{1-F_Z(c;\theta)}\mathbb{I}_{\{x_i \geq c\}} \right] \tag{1}$$

With the new estimate $(\hat{\theta}, \hat{p}) = argmax L(\theta, p|c)$ we move to step 2.

Step 2 : Correction of the frequency distribution

In view of the non-parametric fit, there are r types of policyholders. For each of them we repeat the following :
Let $\lambda_j \quad j = 1, \ldots, r$ be the observed frequency (found by the non-parametric fit of the reference portfolio).
Let $\lambda_j^{'} \quad j = 1, \ldots, r$ be the actual frequency.
Let $\lambda_j^{''} \quad j = 1, \ldots, r$ be the optimal frequency given by Lemaire's algorithm.
We then have :

$$\lambda_j = p\lambda_j^{'} + (1-p)\lambda_j^{''} \tag{2}$$

λ_j is our observation. $\lambda_j^{'}$ is in fact the entry of the algorithm of Lemaire (1977). $\lambda_j^{''}$ is a by-product of this algorithm.
We apply a trial-error scheme on the entry $\lambda_j^{'}$ in order to match the observation λ_i in connection with equation (2).
For this $\lambda_j^{'}$, the algorithm of Lemaire (1977) gives the optimal policy c_j.
The average optimal retention is given by

$$c = \sum_{j=1}^{r} c_j p_j$$

where p_j is the weight associated to λ_j.
With this new average optimal retention c we go back to step 1.

Stopping rule :
We stop the process when convergence occurs. We cannot prove this convergence but in practice it is bound to happen.

Let us note that the following intuitive result is easily shown by maximum likelihood :

$$\hat{p} = \frac{F_n(c)}{F_Z(c)}$$

Proof :
Let $x_{(1)}, x_{(2)}, \cdots, x_{(k)}, c, x_{(k+1)}, \cdots x_{(n)}$ be the order statistic of our observation.
The loglikelihood of (1) writes

$$\prod_{i=1}^{k} p f_Z(x_{(i)}) \prod_{i=k+1}^{n} [p f_Z(x_{(i)}) + (1-p) \frac{f_Z(x_{(i)})}{1 - F_Z(c)}]$$

The normal equation for p gives

$$\hat{p} = \frac{\frac{k}{n}}{F_Z(c)}$$

■

Thus clearly the estimate of p, the proportion of insureds reporting all accidents will depend on the parametric distribution chosen for Z.

In short the algorithm writes

Step 0 : initialization

Do

Step 1 : correction of the claim amount distribution

Maximize (1)

Step 2 : correction of the frequency distribution

For $j = 1$ to r
Do
Try a value for $\lambda_j^{'}$ and apply Lemaire's algorithm
Until (2) is verified
Next j
Find the average optimal retention

Until convergence

7. Numerical example

For the numerical example, we use the observed data set of section 1 as well as the bonus-malus system described in section 1.

In order to use Lemaire's algorithm, we set up the following hypothesis :

1. $\beta = \frac{1}{1+0.06}$
2. $t = \frac{1}{2}$
3. $P = 35$

The commercial premium may seem very high but this is not the case. Let us look at the composition of a Third Party Liability premium in Belgium :

- If we work with the observed data sets, the pure premium is

$$\mathbb{E}N \times \mathbb{E}X = 0.155 \times 113.40 = 17.58$$

- Let us assume a 10% fluctuation loading. The premium becomes

$$17.58 \times 1.1 = 19.34$$

- Let us assume that administration expenses are 10%. The premium becomes

$$19.34 \times 1.1 = 21.27$$

- Brokerage is usually 17%. The premium then becomes

$$\frac{21.27}{1 - 17\%} = 25.63$$

- Taxes are 27.10%. The commercial premium thus amounts to

$$25.63 \times 1.2710 = 32.57$$

This shows that an average commercial premium of 35 is certainly not exaggerated.

The initializing step of the algorithm is chosen as

- $c = 30$
- Z is exponentially distributed with mean μ :

$$f_Z(x;\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \quad x \geq 0$$

We will now describe in detail the first iteration of the algorithm.
Step 1 : by maximizing (1) we get

$$
\begin{aligned}
\hat{\mu} &= 97.137 \\
p &= 0.4577
\end{aligned}
$$

Step 2 : for $j = 1, 2, 3$ we have to match equation (2). We describe in detail the trial-error scheme for $j = 1$.
Let us try an actual claim frequency $\lambda_1' = 0.075$.
An application of Lemaire's algorithm gives the following optimal retentions and frequencies :

s	0	1	2	3	4	5	6	7	8
$m = 0$	18	28	52	76	102	145	119	87	53
$m = 1$	83	108	128	91	54	0	0	0	0
$m = 2$	89	53	0	0	0	0	0	0	0

Table 7: Optimal retentions

s	0	1	2	3	4	5	6	7	8
$m = 0$	0.062	0.056	0.043	0.034	0.026	0.016	0.022	0.030	0.043
$m = 1$	0.031	0.024	0.020	0.029	0.043	0.075	0.075	0.075	0.075
$m = 2$	0.029	0.043	0.075	0.075	0.075	0.075	0.075	0.075	0.075

Table 8: Optimal frequencies

We assume that the stationary distribution is attained within our bonus-malus system. This stationary distribution has been obtained in column 1 of table 6. The parameter of the Poisson distribution in the transition probability matrix is of course 0.05461 because the drivers move in the bonus-malus system according to the frequency of reported claims.
Average values for the optimal frequencies and retentions are then easily given by the scalar product between the stationary probability vector and the optimal frequency or retention vector. We find

	Retention limit	Frequency
$m = 0$	25	0.0587
$m = 1$	85	0.0310
$m = 2$	76	0.0362
$m \geq 3$	0	0.075

Table 9: Average retention limits and frequencies in function of m

We now want figures that are independent of m. Therefore we look for an average value of the optimal retention and frequency by applying the formulae :

$$
\begin{aligned}
c_j &= \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} c_j(m) \quad j = 1, \ldots, r \\
\lambda_j^{''} &= \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} \lambda_j^{''}(m) \quad j = 1, \ldots, r
\end{aligned}
$$

Retention limit : c_1	Frequency : $\lambda_1^{''}$
27	0.0587

Table 10: Average retention limit and frequency

Equation (2) writes :

$$0.05461 \neq 0.4577\, 0.075 + (1 - 0.4577)\, 0.0587 = 0.0661$$

We then proceed by trial-error until equation (2) is matched. This happens with $\lambda_1^{'} = 0.062$. For $j = 2$ (resp. $j = 3$) we find $\lambda_2^{'} = 0.3392$ (resp. $\lambda_3^{'} = 1.0745$).

The second and subsequent iterations may now be proceeded. We find

Iteration	c	μ	p	$\lambda_1^{'}$	$\lambda_2^{'}$	$\lambda_3^{'}$
1	30	97.137	0.4577	0.062	0.3392	1.0745
2	47.7386	85.2208	0.4096	0.0637	0.3628	1.1112
3	47.6437	85.3233	0.4105	0.0637	0.3624	1.1107
4	47.6434	85.3239	0.4105	0.0637	0.3624	1.1107

Table 11: Iterations until convergence

As we see, convergence occurs. The actual claim amount distribution is then exponentially distributed with mean $\mu = 85.32$. The model shows that 41% of the policyholders report all the claims while 59% use the optimal retention. The actual frequency distribution is non-parametric mixed Poisson distributed with

$\lambda_1^{'} = 0.0637$	$p_1 = 0.56189$
$\lambda_2^{'} = 0.3628$	$p_2 = 0.41463$
$\lambda_3^{'} = 1.1107$	$p_3 = 0.02348$

Table 12: Parameters of the actual frequency distribution

With these actual distributions, the pure premium would have been

$$\mathbb{E}N \times \mathbb{E}X = 0.2122 \times 85.32 = 18.11$$

while it was

$$\mathbb{E}N \times \mathbb{E}X = 0.155 \times 113.40 = 17.58$$

with the observed distributions. As expected, the pure premium is higher with the actual distributions because in the case of the observed distribution, some claims are withheld by the policyholders which makes the aggregate claim amount distribution less important.

The frequencies now compare as

	actual frequency	frequency with bm	increase
λ_1	0.0637	0.0546	17%
λ_2	0.3622	0.2459	47%
λ_2	1.1107	0.9561	16%

Table 13: Comparison of the frequencies

This is not surprising as the bad drivers remain in the higher classes of the bonus-malus system and are less interested by the hunger for bonus because of the maximal penalty.

8. Conclusion

Changing the bonus-malus system is a task that may affect every Insurance Company. In particular, Belgian Companies will be obliged to use new bonus-malus systems due to the European directive that forbids the use of an unique bonus-malus system for all the drivers of all the Companies.
In this case it is necessary to have an idea of the actual claim amount and frequency distributions because we know that they are influenced by the hunger for bonus.
This paper gives a solution by using non-parametric mixed Poisson fits and an inversion of Lemaire's algorithm related to the hunger for bonus. The proportion of policyholders using the optimal policy of Lemaire's algorithm is also derived.

Acknowledgments and comment

We would like to thank the editor and the anonymous referees for their very constructive comments which led to a better presentation of the paper. A first version of this paper has

been presented at the International Congress on Industrial and Applied Mathematics 99 in Edinburgh, July 99.

References

Lemaire, J. (1977). La Soif du Bonus. *Astin Bulletin*, 9:181–190.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Boston : Kluwer.

Walhin, J.F. and Paris, J. (1999). Using Mixed Poisson Distributions in Connection with Bonus-Malus systems. *Astin Bulletin*, 29:81–99.

Jean-François Walhin*•
José Paris*

* Institut de Statistique
  Voie du Roman Pays, 20
  B-1348 Louvain-la-Neuve
  Belgium
  e-mail : walhin@stat.ucl.ac.be

• Secura Belgian Re
  Rue Montoyer, 12
  B-1000 Bruxelles
  Belgium