RT-MAE 2002-08

MULTINOMIAL LATENT MODEL FOR RANDOM SUMS

by

Nikolai Kolev
And
Delhi Paiva

Palavras-Chave: Cluster, Correlation coefficient; Maximum likelihood; Multinomial distribution; Random sums; Simulation.
Classificação AMS: Primary 62E15, 60E05, 60E10
Secondary 62A10, 65U05
(AMS Classification)

- Maio de 2002 -

Multinomial Latent Model for Random Sums

Nikolai Kolev and Delhi Paiva
Department of Statistics, University of São Paulo, Brazil

Abstract

In this paper we study random sums relaxing the classical assumption of mutual independence between the variables involved. The possible dependence is measured by the correlation coefficient. We assume that the variables are assigned to independent clusters with equal correlation between them within the cluster. The multinomial distribution is an appropriate to model such a situation. As an application we obtain the distribution of the accumulated claim that reinsurer have to pay. The corresponding result is an extension of the existing ones. A simulation study is performed and estimation of the parameters in the case of two clusters is presented.

Key words and Phrases: *Cluster, Correlation coefficient; Maximum likelihood; Multinomial distribution; Random sums; Simulation.*
AMS (MOS) subject classification: **Primary** 62E15, 60E05, 60E10
Secondary 62A10, 65U05

1 Introduction

The basic object of the classical risk model is to evaluate random sums which represent the accumulated sum of the claims in some fixed time period $(0, T]$. It is assumed that the involved random variables are mutually independent, see e.g. Bowers et *al.* (1997). Only during the last years appeared more than 100 articles where one can find studies about random sums under presence of some type of dependence between variables involved, see e.g. Kass et *al.* (2001).

In this paper we present a new model which generalizes the existing and classical results related to random sums. We consider claims distributed in independent clusters assuming a presence of the dependence (equicorrelation) between the claim amounts that belong to the same cluster. We measure that dependence with the correlation coefficient. The multinomial distribution is used to model this behavior.

The paper is organized as follows. In Section 2 we give some notations and preliminary results about equicorrelated random sums. In Section 3 we present our multinomial correlated model which is a natural extension of the existing in the literature models, and which is closer to the reality. In section 4, we obtain estimates of the involved parameters in the particular case of two clusters using the method of maximum likelihood. Simulation results are given and discussed.

2 Sums of equally correlated random variables

In this section we present some basic results about sums of random variables (r.v.'s) which are equally correlated.

Let us consider a collective insurance contract in some fixed time period $(0, T]$. Let N denote the number of claims in $(0, T]$ and $Y_1, Y_2, \dots Y_N$ the corresponding claims. Then

$$S_N = \sum_{i=1}^{N} Y_i \tag{1}$$

is the total claim amount. In the classical theory it is assumed that (i) N and $(Y_1, Y_2, \dots)$ are independent r.v.'s; (ii) $Y_1, Y_2, \dots$ are independent and (iii) $Y_1, Y_2, \dots$ have the same distribution. The assumptions (ii) and (iii) of mutual independence between the components of the sum is very convenient, mainly because the mathematics is easier.

In many situations, the individual claims (risks) are dependent since they are influenced by the same economic environment. In this study we relax the condition (ii) supposing that the claim amounts are dependent random variables. The observed claims are usually correlated because they contain a common random factor. Our analysis is based on the correlation coefficient, nevertheless its pitfalls. A basic object is to study the effect of correlation on the random sum (1).

For simplicity, let $\{Y_i\}$ be a sequence of non-negative integer-valued r.v.'s with distribution $P(Y = r) = \pi_r$, $r = 0, 1, 2, \dots, k$, $\sum_{r=0}^{k} \pi_r = 1$.

In this paper, we assume that Y_i's are equicorrelated, i.e.

$$Corr(Y_i, Y_j) = \rho, \quad \rho \in (-1, 1), \quad i \neq j.$$

Accordingly, the sequence of claim amounts Y_i's are no more independent, but equicorrelated with $Corr(Y_i, Y_j) = \rho$ for $i \neq j$. The joint PGF of the random vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ may now be written as

$$G_{\mathbf{Y}}(\mathbf{t}) = \rho \left[\sum_{r=0}^{k} \pi_r \left(\prod_{j=1}^{n} t_j \right)^r \right] + (1-\rho) \prod_{j=1}^{n} G_Y(t_j), \quad |t_j| \leq 1, \tag{2}$$

where $\mathbf{t} = (t_1, t_2, \dots, t_k)$, e.g. Tallis (1962). The marginal distribution of Y_i is obtained immediately from (2) by setting $t_s = 1$ for $s \neq i$, $s, i = 1, 2, \dots, n$. Suitable differentiation of (2) gives $Cov(Y_i, Y_j) = E(Y_i Y_j) - E^2(Y) = \rho Var(Y)$. This verifies the fact that the parameter ρ appearing in (2) is just the correlation coefficient between Y_i and Y_j for $i \neq j$.

Consider now the new variate $S_n = Y_1 + \dots + Y_n$. The PGF for S_n is given by

$$G_{S_n}(t) = E(t^{S_n}) = \rho \left(\sum_{r=0}^{k} \pi_r t^{nr} \right) + (1-\rho) \left(G_Y(t) \right)^n \tag{3}$$

and it is obtained from (2) by setting $t_j = t$, $j = 1, 2, \dots, n$. Straightforward calculations show that $E(S_n) = nE(Y)$ and $Var(S_n) = nVar(Y)[1 + (n-1)\rho]$, and therefore ρ can be negative, provided $\rho > -(n-1)^{-1}$.

Let the PGF of N be $G_N(t)$. Using the total probability formula and (3) we obtain

$$G_{S_N}(t) = \rho_1 \sum_{r=0}^{k} \pi_r G_N(t^r) + (1-\rho_1) G_N(G_Y(t)), \tag{4}$$

see Kolev and Paiva (2000). Application of the above model in discrete time series and a correlated extension of Stuetel and van Harn operator, e.g. Stuetel and van Harn (1979), are given in Kolev and Paiva (2000) and Paiva and Kolev (2001).

From (4) we obtain the mean and the variance of S_N:

$$\begin{aligned} E(S_N) &= E(N)E(Y), \\ Var(S_N) &= E(N)Var(Y) + Var(N)[E(Y)]^2 + \rho Var(Y) E[N(N-1)]. \end{aligned}$$

If Y_i's are independent, substituting $\rho = 0$ in (4), we get the well-known classical result: $G_{S_N}(t) = G_N(G_Y(t))$.

Consider a stop-loss reinsurance contract with retention M. Let $N_R = \sum_{i=1}^{N} \mathbf{I}_{\{Y_i > M\}}$ denote the number of claims the reinsurer has to pay for and let $\alpha = P(Y_i > M)$. The r.v.'s $Z_i = \mathbf{I}_{\{Y_i > M\}}$ have the same Bernoulli distribution with parameter $\alpha \in [0,1]$ and a PGF given by $G_Z(t) = 1 - \alpha + \alpha t$.

Note that $Y_1, \ldots, Y_n$ are equally correlated with an identical distribution. Therefore, $Z_1, \ldots, Z_n$ are exchangeable Bernoulli variables. Under the above conditions the following statement is true, see Dimitrov and Kolev (2002).

Lemma 1. *Let $Corr(Y_i, Y_j) = \rho$, $i \neq j$. Then $Corr(Z_i, Z_j) = \rho$ iff*

$$P(Z_i = 1 \mid Z_j = 1) = \alpha + (1-\alpha)\rho \quad \text{and} \quad P(Z_i = 0 \mid Z_j = 0) = (1-\alpha) + \alpha\rho. \tag{5}$$

Proof: The r.v.'s $Z_i, i = 1, 2, \ldots$, are identically Bernoulli distributed with parameter α and this provides $E(Z) = \alpha = P(Y_i > M)$ and $Var(Z) = \alpha(1-\alpha)$. Then $Corr(Z_i, Z_j) = \rho$ for $i \neq j$ iff $Cov(Z_i, Z_j) = \rho_1 Var(Z)$, i.e. when

$$E(Z_i Z_j) - \alpha^2 = \alpha(1-\alpha)\rho.$$

On the other hand, we have

$$E(Z_i Z_j) = 1.1.P\{Z_i = 1, Z_j = 1\} = P\{Z_i = 1, Z_j = 1\}.$$

Therefore, $P\{Z_i = 1, Z_j = 1\} = \alpha^2 + \alpha(1-\alpha)\rho$ and by conditional probability formula we obtain the first relation in (5). The second equality in (5) follows by similar arguments considering the variables $Z_i^* = 1 - Z_i$, $i = 1, 2, \ldots$. □

Thus, the indicator r.v.'s Z_i are equicorrelated. The conditional probabilities (5) determine the law of appearance of the claims with larger (less) than **M** size, given the values of α and correlation coefficient ρ.

Now, using (4) with $k = 1$, $\pi_0 = 1 - \alpha$ and $\pi_1 = \alpha$ we obtain the PGF of N_R

$$G_{N_R}(t) = \rho(1-\alpha) + \rho\alpha G_N(t) + (1-\rho) G_N(1 - \alpha + \alpha t). \tag{6}$$

After that one can replace $G_N(t)$ with $G_{N_R}(t)$ in (4) to get the distribution of the accumulated clam that the reinsurer has to pay for. The relation (6) shows that N_R can be represented as a mixture of point mass concentrated at zero, the distributions of N and $\sum_{i=1}^{N} W_i$, where W_i's are independent identically distributed Bernoulli r.v.'s with parameter α. Let us note that if $N = n = const$, then $G_N(t) = t^n$ and (6) takes the form

$$G_X(t) = \rho(1 - \alpha + \alpha t^n) + (1 - \rho)(1 - \alpha + \alpha t)^n, \tag{7}$$

which is a PGF of the correlated binomial r.v. X with parameters n, α and ρ. Let us denote this fact by $X \sim CBi(n, p, \rho)$. In our case, X gives the number of n equicorrelated claims that have size greater than M.

As a consequence of (4) we can represent S_n as a mixture

$$S_n \stackrel{d}{=} (1 - \rho)X_1 + \rho X_2,$$

where the r.v. X_1 have binomial distribution with parameters n e α, i.e., $X_1 \sim Bi(n, \alpha)$, and the distribution of X_2 is given by $P(X_2 = 0) = 1 - \alpha$ and $P(X_2 = n) = \alpha$. Therefore, for the probability mass function of S_n one obtains

$$\begin{aligned} &P(S_n = 0) = (1 - \rho)(1 - \alpha)^n + \rho(1 - \alpha); \\ &P(S_n = j) = (1 - \rho)\tbinom{n}{j}\alpha^j(1 - \alpha)^{n-j}, \quad \text{for} \quad j = 1, 2, \ldots, n - 1; \\ &P(S_n = n) = (1 - \rho)\alpha^n + \rho\alpha. \end{aligned}$$

3 Multinomial correlated model

The model in the previous section implicitly assumes that all claims belong to a single homogeneous cluster. In practice one can observe that the claims are grouped in independent clusters with a possible dependence between claim amounts within the cluster. To model such a situation we assume that our n claims are distributed into K independent clusters ($n \geq K$).

3.1 General Case (K clusters)

Let each claim be randomly assigned to the k-th cluster with probability p_k, $\sum_{k=1}^{K} p_k = 1$. Accordingly, the sizes of the clusters form a random vector $\mathbf{n} = (n_1, n_2, \ldots, n_K)$ that may be considered as a latent variable having multinomial distribution with parameters $n = \sum_{k=1}^{K} n_k$ and $\mathbf{p} = (p_1, \ldots, p_k)$, i.e. $\mathbf{n} \sim Mn(n, \mathbf{p})$. Let us assume that the claims Y_{kj} belonging to the k-th cluster ($k = 1, \ldots, K$, $j = 1, \ldots, n_k$), are equicorrelated with $\rho_k = Corr(Y_{kj}, Y_{kj'})$. Then the indicator r.v.'s $Z_{kj} = \mathbf{I}_{\{Y_{kj} > M\}}$ have Bernoulli distribution with parameter $\alpha_k = P(Y_{kj} > M)$ and are also equicorrelated with $\rho_k = Corr(Z_{kj}, Z_{kj'})$ according to Lemma 1. According to the conclusions in Section 2, the number of claims $S_k = \sum_{j=1}^{n_k} Z_{kj}$ the re-insurer has to pay for a k-th cluster has a $CBi(n_k, \alpha_k, \rho_k)$ distribution with a PGF $G_{S_k}(t)$ given by (7). Since the K clusters are assumed to be independent we have the following result

Lemma 2. *The PGF of the sum $\sum_{k=1}^{K} S_k$ can be represented by*

$$G_{\sum_{k=1}^{K} S_k}(t) = E_{\mathbf{n}}\left[\prod_{k=1}^{K} G_{S_k}(t)\right], \tag{8}$$

where the expectation $E_{\mathbf{n}}(\cdot)$ is taken with respect to the multinomial random vector $\mathbf{n} \sim Mn(n, \mathbf{p})$ and $G_{S_k}(t)$ is the PGF given by (7).

After that one can use the total probability formula and (8) to obtain distribution of S_N, assuming K to be a r.v. with a given distribution.

3.2 Particular Case ($K = 2$ clusters)

When the number of the clusters is $K > 3$, it is difficult to obtain from (8) an explicit and applicable representation for the distribution of S_n. Here, we consider a particular case where the claims are grouped in two independent clusters. Thus, the vector $\mathbf{n} = (n_1, n_2)$ has a binomial distribution with parameters n and p_1, i.e. $\mathbf{n} \sim Bi(n, p_1)$ with $n = n_1 + n_2$.

Under the above assumption we will obtain the exact distribution of $S_n = S_{n_1} + S_{n_2}$ when $S_{n_k} \sim CBi(n_k, \alpha_k, \rho_k)$, $k = 1, 2$, with $\rho_1 \neq \rho_2$ and $\alpha_1 \neq \alpha_2$. From (8) we obtain that the PGF of S_n is given by

$$G_{S_n}(t) = E_{\mathbf{n}}\left[G_{S_{n_1}}(t) G_{S_{n_2}}(t)\right]. \tag{9}$$

Using (9) and the total probability formula for $\mathbf{m} = (m_1, n - m_1)$, we obtain

$$\begin{aligned} G_{S_n}(t) &= \sum_{n}\left[G_{S_{n_1}}(t) G_{S_{n_2}}(t)\middle| \mathbf{n} = \mathbf{m}\right] P(\mathbf{n} = \mathbf{m}) \\ &= \sum_{m_1=0}^{n}\left\{\prod_{k=1}^{2}\left[\rho_k(1 - \alpha_k + \alpha_k t^n) + (1 - \rho_k)(1 - \alpha_k + \alpha_k t)^n\right]\right\}\binom{n}{m_1} p_1^{m_1}(1 - p_1^{n-m_1}). \end{aligned}$$

After some algebra we get

$$\begin{aligned} G_{S_n}(t) = \; & \rho_1\rho_2(1 - \alpha_1)(1 - \alpha_2) + \rho_1\rho_2\alpha_1\alpha_2 t^n \\ & + \rho_1\rho_2(1 - \alpha_1)\alpha_2[p_1 + (1 - p_1)t]^n + \rho_1\rho_2\alpha_1(1 - \alpha_2)[1 - p_1 + p_1 t]^n \\ & + (1 - \rho_1)(1 - \rho_2)[1 - \alpha_2 + \alpha_2 p_1 - \alpha_1 p_1 + (\alpha_2 - \alpha_2 p_1 + \alpha_1 p_1)t]^n \\ & + (1 - \rho_1)\rho_2(1 - \alpha_2)[1 - \alpha_1 p + \alpha_1 p_1 t]^n \\ & + \rho_1(1 - \rho_2)(1 - \alpha_1)(1 - \alpha_2 + \alpha_2 p_1 + (\alpha_2 - \alpha_2 p_1)t]^n \\ & + (1 - \rho_1)\rho_2\alpha_2(p_1[\alpha_1 - \alpha_1 p_1 + (1 - \alpha_1 + \alpha_1 p_1)t]^n \\ & + \rho_1(1 - \rho_2)\alpha_1[1 - \alpha_2 - p_1 + \alpha_2 p_1 + (\alpha_2 + p_1 - \alpha_2 p_1)t]^n. \end{aligned}$$

The last expression can be represented as a mixture

$$G_{S_n}(t) = \sum_{l=1}^{9} a_l G_{Z_l}(t), \tag{10}$$

where $\sum_{i=1}^{9} a_i = 1$, Z_1 is the constant 0, Z_2 is the constant n and r.v.'s Z_l, ($l = 3, 4, \ldots, 9$) have binomial distribution with parameters n e β_l, respectively. The values of the coefficients a_i and parameters β_l are given in Table 1. From (10) we conclude that the random the sum S_n is given by the following mixture

$$S_n \stackrel{d}{=} \sum_{l=1}^{9} a_l Z_l. \tag{11}$$

Let us only note, that if we consider K clusters the PGF of S_n would contain $(K+1)^K$ terms.

Table 1. *Coefficients of the mixture (10) and parameters of r.v.'s Z_l.*

Variable	a_l	β_l
Z_1	$\rho_1\rho_2(1-\alpha_1)(1-\alpha_2)$	
Z_2	$\rho_1\rho_2\alpha_1\alpha_2$	
Z_3	$\rho_1\rho_2(1-\alpha_1)\alpha_2$	$1-p_1$
Z_4	$\rho_1\rho_2\alpha_1(1-\alpha_2)$	p_1
Z_5	$(1-\rho_1)(1-\rho_2)$	$\alpha_2 - \alpha_2 p_1 + \alpha_1 p_1$
Z_6	$(1-\rho_1)\rho_2(1-\alpha_2)$	$\alpha_1 p_1$
Z_7	$\rho_1(1-\rho_2)(1-\alpha_1)$	$\alpha_2 - \alpha_2 p_1$
Z_8	$(1-\rho_1)\rho_2\alpha_2$	$1-\alpha_1+\alpha_1 p_1$
Z_9	$\rho_1(1-\rho_2)\alpha_1$	$\alpha_2 + p_1 - \alpha_2 p_1$

Let us denote the model (10) as Model B (with parameters $n, p_1, \alpha_1, \alpha_2, \rho_1, \rho_2$). We consider the following particular cases:

- Model $B1$ (with parameters n, p_1, α, ρ): when $\alpha_1 = \alpha_2 = \alpha$ and $\rho_1 = \rho_2 = \rho$;
- Model $B2$ (with parameters $n, p_1, \alpha_1, \alpha_2, \rho$): when $\rho_1 = \rho_2 = \rho$;
- Model $B3$ (with parameters $n, p_1, \alpha, \rho,_1 \rho_2$): when $\alpha_1 = \alpha_2 = \alpha$.

In the next tables we show the values of the coefficients a_i, $i = 1, \ldots, 9$ of the mixture (11) for the models $B1$, $B2$ and $B3$. One can obtain these values after corresponding substitution in (10).

Table 2. *Coefficients of the mixture.*

Variable	Model $B1$	Model $B2$	Model $B3$
Z_1	$\rho^2(1-\alpha)^2$	$\rho^2(1-\alpha_1)(1-\alpha_2)$	$\rho_1\rho_2(1-\alpha)^2$
Z_2	$\rho^2\alpha^2$	$\rho^2\alpha_1\alpha_2$	$\rho_1\rho_2\alpha^2$
Z_3	$\rho^2\alpha(1-\alpha)$	$\rho^2(1-\alpha_1)\alpha_2$	$\rho_1\rho_2\alpha(1-\alpha)$
Z_4	$\rho^2\alpha(1-\alpha)$	$\rho^2\alpha_1(1-\alpha_2)$	$\rho_1\rho_2\alpha(1-\alpha)$
Z_5	$(1-\rho)^2$	$(1-\rho)^2$	$(1-\rho_1)(1-\rho_2)$
Z_6	$\rho(1-\rho)(1-\alpha)$	$\rho(1-\rho)(1-\alpha_2)$	$\rho_1(1-\rho_2)(1-\alpha)$
Z_7	$\rho(1-\rho)(1-\alpha)$	$\rho(1-\rho)(1-\alpha_1)$	$\rho_1(1-\rho_2)(1-\alpha)$
Z_8	$\rho(1-\rho)\alpha$	$\rho(1-\rho)\alpha_2$	$\rho_1(1-\rho_2)\alpha$
Z_9	$\rho(1-\rho)\alpha$	$\rho(1-\rho)\alpha_1$	$\rho_1(1-\rho_2)\alpha$

For the models $B1$, $B2$ and $B3$, in the next Table 3 are given the values of the parameters β_l of r.v.'s $Z_l \sim Bi(n, \beta_l)$, $l = 3, 4, \ldots, 9$ (remember that $Z_1 = 0$ and $Z_2 = n$).

Table 3. *Parameters β_l of r.v.'s Z_l, $l = 3, 4, \ldots, 9$.*

Variable	Model $B1$	Model $B2$	Model $B3$
Z_3	$1 - p_1$	$1 - p_1$	$1 - p_1$
Z_4	p_1	p_1	p_1
Z_5	α	$\alpha_2 - \alpha_2 p_1 + \alpha_1 p_1$	α
Z_6	αp_1	$\alpha_1 p_1$	αp_1
Z_7	$\alpha - \alpha p_1$	$\alpha_2 - \alpha_2 p_1$	$\alpha - \alpha p_1$
Z_8	$1 - \alpha + \alpha p_1$	$1 - \alpha_1 + \alpha_1 p_1$	$1 - \alpha - \alpha p_1$
Z_9	$\alpha + p_1 - \alpha p_1$	$\alpha_2 + p_1 - \alpha_2 p_1$	$\alpha + p_1 - \alpha p_1$

4 Estimation and simulation

In the literature a variety of methods for estimation of the parameters of the mixture have been used. For example, Blishke (1962) used the method of factorial moments for binomial mixtures. It is impossible to apply that method directly in our case, since the mixture (11) contains two constants ($Z_1 = 0$ and $Z_2 = n$). We use the method of the maximum likelihood for estimation of the parameters of the models B, $B1$, $B2$ and $B3$. From (10) we obtain the following distribution for S_n for the most complex model B:

$$
\begin{aligned}
P(S_n = m_1) \;=\;& \rho_1\rho_2(1-\alpha_1)(1-\alpha_2)\mathbf{I}\{m_1 = 0\} + \rho_1\rho_2\alpha_1\alpha_2\mathbf{I}\{m_1 = n\} \\
&+\rho_1\rho_2(1-\alpha_1)\alpha_2\binom{n}{m_1}(1-p_1)^{m_1}p_1^{n-m_1} \\
&+\rho_1\rho_2\alpha_1(1-\alpha_2)\binom{n}{m_1}p_1^{m_1}(1-p_1)^{n-m_1} \\
&+(1-\rho_1)(1-\rho_2)\binom{n}{m_1}(\alpha_2 - \alpha_2 p_1 + \alpha_1 p_1)^{m_1}(1 - \alpha_2 + \alpha_2 p_1 - \alpha_1 p_1)^{n-m_1} \\
&+(1-\rho_1)\rho_2(1-\alpha_2)\binom{n}{m_1}(\alpha_1 p_1)^{m_1}(1-\alpha_1 p_1)^{n-m_1} \\
&+\rho_1(1-\rho_2)(1-\alpha_1)\binom{n}{m_1}(\alpha_2 - \alpha_2 p_1)^{m_1}(1 - \alpha_2 + \alpha_2 p_1)^{n-m_1} \\
&+(1-\rho_1)\rho_2\alpha_2\binom{n}{m_1}(1 - \alpha_1 + \alpha_1 p_1)^{m_1}(\alpha_1 - \alpha_1 p_1)^{n-m_1} \\
&+\rho_1(1-\rho_2)\alpha_1\binom{n}{m_1}(\alpha_2 + p_1 - \alpha_2 p_1)^{m_1}(1 - \alpha_2 - p_1 + \alpha_2 p_1)^{n-m_1},
\end{aligned}
$$

where $\mathbf{I}\{\cdot\}$ represents the indicator function and e $m_1 = 0, 1, \ldots, n$.

Let $\Theta = (p_1, \alpha_1, \alpha_2, \rho_1, \rho_2)$ be the parameter vector in this case. Denote by $\hat{\Theta} = (\hat{p_1}, \hat{\alpha_1}, \hat{\alpha_2}, \hat{\rho_1}, \hat{\rho_2})$ the maximum likelihood estimates obtained by the solution of the following system of equations

$$\frac{\partial l_n}{\partial p_1} = 0, \quad \frac{\partial l_n}{\partial \alpha_1} = 0, \quad \frac{\partial l_n}{\partial \alpha_2} = 0, \quad \frac{\partial l_n}{\partial \rho_1} = 0 \text{ and } \frac{\partial l_n}{\partial \rho_2} = 0,$$

where $l_n = \log P(S_n = m_1)$. The last system is non-linear and it does not have an explicit solution, so we use the method of Newton-Raphson which needs the initial values of the parameters. These initial values we obtain trough the method of moments.

Simulations have been done using the program OX, e.g. Doornik (1996). The aim is to obtain estimation of the parameters of the models B, $B1$, $B2$ and $B3$ defined in Section 3.

Two sets of observations grouped in two independent clusters have been generated with sizes n_1 (80 e 30) and n_2 (90 e 140), $n = n_1 + n_2 = 170$, considering equally correlated observations belonging to the same cluster for the models B, $B1$, $B2$ and $B3$. For each simulation we use the same initial values of the parameters $\alpha_1, \alpha_2, \rho_1, \rho_2$ (obtained by method of moments), fixing $p_1 = 0,5$.

In the next four tables we show the results of the simulations made. We were able to get the corresponding estimates for about 1000 iterations of the procedure proposed by OX program.

One can observe that when the number of the claims n_1 and n_2 in both clusters are close (80 and 90), the estimates of the parameters $\hat{\rho}_1$, $\hat{\rho}_2$ e $\hat{\alpha}_1$, $\hat{\alpha}_2$ are equal for the model B. When we have clusters with significantly different number of claims ($n_1 = 30$) and ($n_2 = 140$), then the corresponding parameter estimates also have different values for the model B, (see Table 4).

Table 4. *Initial values and parameter estimates for Model B*

		Initial values					Parameter estimates				
n_1	n_2	p_1	α_1	α_2	ρ_1	ρ_2	$\hat{p}_1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$
80	90	0,50	0,52	0,39	0,62	0,29	0,38	0,47	0,47	0,73	0,73
30	140	0,50	0,52	0,39	0,62	0,29	0,81	0,83	0,07	0,92	0,30

Table 5. *Initial values and parameter estimates for Model B1*

		Initial values			Parameter estimates		
n_1	n_2	p_1	α	ρ	$\hat{p}_1$	$\hat{\alpha}$	$\hat{\rho}$
80	90	0,50	0,80	0,70	0,43	0,82	0,60
30	140	0,50	0,80	0,70	0,50	0,59	0,53

Table 6. *Initial values and parameter estimates for Model B2*

		Initial values				Parameter estimates			
n_1	n_2	p_1	α_1	α_2	ρ	$\hat{p}_1$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\rho}$
80	90	0,50	0,58	0,32	0,41	0,68	0,63	0,32	0,71
30	140	0,50	0,58	0,32	0,41	0,58	0,09	0,83	0,99

Table 7. *Initial values and parameter estimates for Model B3*

		Initial values				Parameter estimates			
n_1	n_2	p_1	α	ρ_1	ρ_2	$\hat{p}_1$	$\hat{\alpha}$	$\hat{\rho}_1$	$\hat{\rho}_2$
80	90	0,50	0,65	0,45	0,71	0,62	0,58	0,30	0,74
30	140	0,50	0,65	0,45	0,71	0,49	0,91	0,39	0,92

For the models $B1$, $B2$ and $B3$, the estimated values of the parameters have different range for the both simulations (compare Tables 5, 6 and 7).

Note that the estimates of the coefficient of correlation in the four models are relatively high, which clearly demonstrate the presence of the dependence between claims belonging to the same cluster.

References

[1] Blishke, W. R. (1962). Moment estimators for the parameters of a mixture of two binomial distributions. *Annals of Mathematical Statistics* **33**, 444-454.

[2] Bowers, N.L,, Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd ed. Schaumburg, Ill.: Society of Actuaries.

[3] Dimitrov, B. and Kolev, N. (2002). Bernoulli trials: extensions, related probability distributions and modeling powers. *Mathematics and Education in Mathematics*, 15-24.

[4] Doornik, J.A. (1996). *OX. An Object - Oriented Matrix Programming Language.* London: International Thompson Business Press.

[5] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001). *Modern Actuarial Risk Theory*, Kluwer Academic Publishers.

[6] Kolev, N. and Paiva, D. (2000). Correlated INAR(1) Process. In: *Proc. Contributed Papers COMSTAT'2000*, 337-342.

[7] Paiva, D. and Kolev, N. (2001). Extended DAR(1) Process. In: *Proc. 16th International Workshop of Statistical Modelling*, 487-490.

[8] Steutel, F.B. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability* **7**, 893-899.

[9] Tallis, G.M. (1962). The use of generalized multinomial distribution in the estimation of correlation in discrete data. *The Journal of Royal Statistical Society, Ser. B* **24**, 530-534.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2002-01 - BUENO, V.C. Asymptotic reliability for a coherent system with an infinite number of components. 2002. 10p. (RT-MAE-2002-01)

2002-02 - TAVARES, H.R., ANDRADE, D.F. Item Response Theory for Longitudinal Data: Item and Population Ability Parameters Estimation. 2002. 20p. (RT-MAE-2002-02)

2002-03 - BOLFARINE, H., VALENÇA, D.M. Score tests for Weibull-regression models with random effect. 2002. 16P. (RT-MAE-2002-03)

2002-04 - BUENO, V.C. Using a second order stochastic dominance for active redundancy allocation in a K-out-of-n system. 2002. 7p. (RT-MAE-2002-4)

2002-05 - FERRARI, S.L.P., LUCAMBIO, F., CRIBARI-NETO, F. Adjusted and Bartlett-Adjusted profile likelihood ratio tests. 2002. 19p. (RT-MAE-2002-05).

2002-06 - PEREIRA, C.A.B., STERN, J.M. Full Bayesian Significance Test: Invariant Formulation. 2002. 10p. (RT-MAE-2002-06).

2002-07 - BUENO, V.C. Optimal arrangement of components for systems of dependent components. 2002. 13p. (RT-MAE-2002-07)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil