Информационные процессы, Том 2, № 2, 2002, стр. 216–219.
© *2002 Kolev.*

Volodya, I Miss You
(Two Correlated Collective Risk Models)

N. Kolev

Department of Statistics, University of Sao Paulo, Brazil
email: nkolev@ime.usp.br
Received October 14, 2002

This paper is dedicated to the memory of Vladimir Kalashnikov.
He stimulated me to study random sums under dependence,
suggesting as a first step just Model 1 presented below.

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 $Y_1, Y_2, \ldots Y_N$ 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 $(Y_1, Y_2, \ldots)$ are independent random variables (r.v.'s); (ii) $Y_1, Y_2, \ldots$ are independent and (iii) $Y_1, Y_2, \ldots$ 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. This kind of correlation can be presented by the following model: For $i = 1, 2, \ldots$, define $Y_i = J(U_i, V)$, where $U_1, U_2, \ldots$ and V are independent r.v.'s, and $U_i \sim F(u),\ V \sim G(v)$, and $J(u, v)$ is some correlated link function. Therefore, the distribution of Y_i's is determined by the link function $J(u, v)$ and two distribution functions $F(u)$ and $G(v)$. A simple example of the correlated link function is the linear function $J(u, v) = au + bv + c$. It can be seen that the r.v.'s $Y_1, Y_2, \ldots$ are independent if V is a constant.

Our analysis is based on the correlation coefficient, despite its pitfalls. A basic object is to investigate 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 $P(Y = r) = \pi_r,\ r = 0, 1, 2, \ldots, k,\ \sum_{r=0}^{k} \pi_r = 1$. At first, we consider the following model.

Model 1 (equicorrelated claims): We assume that Y_i's are equicorrelated, i.e.

$$Corr(Y_i, Y_j) = \rho_1, \quad \rho_1 \in (-1, 1), \quad i \neq j.$$

Accordingly, the sequence of claim amounts Y_i's are no more independent, but equicorrelated with $Corr(Y_i, Y_j) = \rho_1$ for $i \neq j$. The joint PGF of the random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ may now be written as

$$G_{\mathbf{Y}}(\mathbf{t}) = \rho_1 \left[\sum_{r=0}^{k} \pi_r \left(\prod_{j=1}^{n} t_j \right)^r \right] + (1 - \rho_1) \prod_{j=1}^{n} G_Y(t_j), \quad |t_j| \leq 1, \tag{2}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_k)$. The marginal distribution of Y_i is obtained immediately from (2) by setting $t_s = 1$ for $s \neq i,\ \ s, i = 1, 2, \ldots, n$. Suitable differentiation of (2) gives $Cov(Y_i, Y_j) = E(Y_i Y_j) - E^2(Y) = \rho_1 Var(Y)$. This verifies the fact that the parameter ρ_1 appearing in (2) is just the correlation coefficient between Y_i and Y_j for $i \neq j$, as stated by our Model 1.

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

$$G_{S_n}(t) = E(t^{S_n}) = \rho_1 \left(\sum_{r=0}^{k} \pi_r t^{nr} \right) + (1 - \rho_1) \left(G_Y(t) \right)^n \tag{3}$$

and it is obtained from (2) by setting $t_j = t,\ j = 1, 2, \ldots, n$. Straightforward calculations show that $E(S_n) = nE(Y)$ and $Var(S_n) = nVar(Y)[1 + (n-1)\rho_1]$, and therefore ρ_1 can be negative, provided $\rho_1 > -(n-1)^{-1}$.

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

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

If Y_i's are independent, substituting $\rho_1 = 0$ 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 $p = P(Y_i > M)$. The r.v.'s $Z_i = \mathbf{I}_{\{Y_i > M\}}$ have the same Bernoulli distribution with parameter $p \in [0, 1]$ and PGF given by $G_Z(t) = q + pt,\ q = 1 - p$.

Note that $Y_1, \ldots, Y_n$ are equally correlated with an identical distribution according to Model 1. Therefore, $Z_1, \ldots, Z_n$ are exchangeable Bernoulli variables. Under the above conditions the following statement is true.

Lemma 1. *Let $Corr(Y_i, Y_j) = \rho_1, i \neq j$. Then $Corr(Z_i, Z_j) = \rho_1$ iff*

$$P(Z_i = 1 \mid Z_j = 1) = p + q\rho_1 \quad \text{and} \quad P(Z_i = 0 \mid Z_j = 0) = q + p\rho_1. \tag{5}$$

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 p and correlation coefficient ρ_1.

Now, using (4) with $k = 1,\ \pi_0 = q$ and $\pi_1 = p$ we obtain the PGF of N^R

$$G_{N^R}(t) = q\rho_1 + p\rho_1 G_N(t) + (1 - \rho_1) G_N(q + pt). \tag{6}$$

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 p. Let us note that if $N = n = const$, then $G_N(t) = t^n$ and (6) takes the form

$$G_X(t) = \rho_1(q + pt^n) + (1 - \rho_1)(q + pt)^n,$$

which is a PGF of the correlated binomial r.v. X with parameters n, p and ρ_1. In our case, X gives the number of n equicorrelated claims that have size greater than M.

Consider now the following **Model 2 (adjacent correlated claims):** We replace (ii) by the following more realistic assumption

$$\rho_2 = Corr(Y_{i+1}, Y_i), \quad \rho_2 \in (-1, 1), \quad i = 1, 2, \ldots.$$

In this case the indicator r.v.'s $Z_i = \mathbf{I}_{\{Y_i > M\}}$ also have the same Bernoulli distribution with parameter p and Lemma 1 is valid for $i = j+1$. The following result is true.

Lemma 2. *Let $Corr(Y_{i+1}, Y_i) = \rho_2$, $i = 1, 2, \ldots, n-1$. Then $Corr(Z_{i+1}, Z_i) = \rho_2$ iff the sequence $\{Z_1, \ldots, Z_n\}$ forms a homogeneous 0-1 Markov chain with initial distribution*

$$p = P(Z_1 = 1) = 1 - P(Z_1 = 0) \tag{7}$$

and transition probability matrix

$$\mathbf{P} = \begin{pmatrix} q + p\rho_2 & p(1-\rho_2) \\ q(1-\rho_2) & p + q\rho_2 \end{pmatrix}. \tag{8}$$

For any fixed $m = 1, \ldots, n-1$ the transition probability matrix for $m-$steps of the sequence $\{Z_1, \ldots, Z_n\}$ has the form

$$\mathbf{P}^{(m)} = \begin{pmatrix} q + p\rho_2^m & p(1-\rho_2^m) \\ q(1-\rho_2^m) & p + q\rho_2^m \end{pmatrix}$$

and

$$Corr(Z_{s+m}, Z_s) = \rho_2^m, \quad \text{for} \quad m = 1, \ldots, n-1 \quad \text{and} \quad s = 1, \ldots, n-m. \tag{9}$$

Moreover, for m even, $0 \le \rho_2^m < 1$; for m odd, $-1 < \rho_2^m < 1$.

Therefore Model 2 implies that the binary sequence $(Z_1, ..., Z_n)$ is an initial segment of a stationary 0-1 Markov chain (and hence Z_i's are not exchangeable as in Model 1). The autoregressive model can be given as an example for Model 2, but not for Model 1.

Relation (9) shows that the random variables of any finite subset of consecutive members of a sequence of correlated Bernoulli trials are not equicorrelated (as in Model 1). Also, the influence of the s-th claim amount into the following ones decreases when m increases, $s + m \le n$.

Let us note, that our dependent Bernoulli r.v's Z_i are modeled with a special choice of the transition probabilities (8), which satisfy that the unconditional probabilities are just given by (7), and Z_i's are still the r.v.'s reflecting the outcomes $\{Y_i > M\}$ (classified as a "success") or $\{Y_i \le M\}$ ("failure") in the $i-$th trail.

Let $(Z_1, \ldots, Z_n)$ denote the initial segment of of 0-1 Markov chain with transition probability matrix $\mathbf{P}$ and let (8) holds. Then the joint distribution

$$P(Z_1 = z_1, \ldots, Z_n = z_n) = D(n) C^{z_1 + z_n} A^{z_1 + \cdots + z_n} B^{z_1 z_2 + \cdots + z_{n-1} z_n},$$

where the coefficients C, A and B depend on p and ρ_2, and $D(n)$ depends on n, p and ρ_2. A recurrence relation for the joint PGF $G_{\mathbf{Z}}(\mathbf{t})$ of random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)$ is given by the following lemma.

Lemma 3. *Let us denote $G_n(t_1, \ldots, t_n) = G_{\mathbf{Z}}(\mathbf{t})$. Under suppositions of Model 2 the following recurrence relation is fulfilled for any $n \ge 1$*

$$G_{n+1}(t_1, \ldots, t_n, t_{n+1}) = \rho_2 G_n(t_1, \ldots, t_{n-1}, t_n t_{n+1}) + (1-\rho_2) G_n(t_1, \ldots, t_n)(q + p t_{n+1}).$$

We are ready now to give an explicit expression for the joint PGF $G_{\mathbf{Z}}(\mathbf{t})$. It seems that the corresponding result appears first time in literature.

Lemma 4. *Let $\{Z_i\}$ be a sequence of Bernoulli r.v.'s with parameter p, such that $\rho_2 = Corr(Z_{i+1}, Z_i)$, $(i = 1, 2, \ldots)$. Then the joint PGF of $G_{\mathbf{Z}}(\mathbf{t})$ is given by*

$$G_{\mathbf{Z}}(\mathbf{t}) = D(n) \sum_{j=1}^{n} \rho_2^{n-j} (1-\rho_2)^{j-1} \sum_{k_1, \ldots, k_j} \prod_{r=1}^{j} \left(q + p \prod_{i_r = k_{r-1}+1}^{k_{r-1}+k_r} t_{i_r} \right), \tag{10}$$

where $k_0 = 0$ *and the summation is on all non-negative integers* $k_1, \ldots, k_j$ *which satisfy the relation* $k_1 + k_2 + \cdots + k_j = n,\ j = 1, 2, \ldots, n.$

Consider a new variate $S_n = Z_1 + \cdots + Z_n$. By setting all arguments $t_{i_r} = t$ in (10) we obtain the PGF of S_n

$$G_{S_n}(t) = D(n) \sum_{j=1}^{n} \rho_2^{n-j} (1-\rho_2)^{j-1} \sum_{k_1+\cdots+k_j=n} \prod_{r=1}^{j} (q + pt^{k_r}). \tag{11}$$

The last expression represents the PGF of Markov chain binomial distribution. $E(S_n)$ and $Var(S_n)$ can be obtained from (11) after some algebra. We have $E(S_n) = np$ and

$$Var(S_n) = npq + \frac{2pq\rho_2^2}{(1-\rho_2)^2} \left[1 - n\rho_2^{n-1} + (n-1)\rho_2^n\right].$$

But np and npq are just the mean and the variance of binomial distribution with parameters n and p. From the last relation it can be seen that the correlation ρ_2 can take negative values also, provided $1 - n\rho_2^{n-1} + (n-1)\rho_2^n < 0$.

At the end, assuming that N is a r.v. (and therefore N^R is random), one can use the total probability formula and (11) to get the PGF of $G_{N_R}(t)$ which is much complicated than (6).