Insurance: Mathematics and Economics 79 (2018) 184–193

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

On the evaluation of some multivariate compound distributions with Sarmanov's counting distribution

Raluca Vernic *

Faculty of Mathematics and Computer Science, Ovidius University of Constanta, 124 Mamaia Blvd., 900527 Constanta, Romania
Institute for Mathematical Statistics and Applied Mathematics, Calea 13 Septembrie 13, 050711 Bucharest, Romania

ARTICLE INFO

Article history:
Received September 2016
Received in revised form November 2017
Accepted 17 January 2018
Available online 2 February 2018

Keywords:
Multivariate compound model
Sarmanov's multivariate discrete distribution
Recursive evaluation
Panjer's class of distributions
Fast Fourier Transform method

ABSTRACT

In this paper, we consider Sarmanov's multivariate discrete distribution as counting distribution in two multivariate compound models: the first model assumes different types of independent claim sizes (corresponding to, e.g., different types of insurance policies), while in the second model, we introduce some dependency between the claims (motivated by the events that can simultaneously affect several types of policies). Since Sarmanov's distribution can join different types of marginals, we also assume that these marginals belong to Panjer's class of distributions and discuss the evaluation of the resulting compound distribution based on recursions. Alternatively, the evaluation of the same distribution using the Fast Fourier Transform method is also presented, with the purpose to significantly reduce the computing time, especially in the dependency case. Both methods are numerically illustrated and compared from the point of view of speed and accuracy.

1. Introduction

Because it allows for a flexible dependency structure that can join different types of marginals, Sarmanov's (1966) family of distributions has been recently used in applied studies in various fields. In insurance and finance, this distribution has been considered in connection with premiums calculation by Hernández-Bastida et al. (2009) and Hernández-Bastida and Fernández-Sánchez (2013), with ruin theory by Yang and Hashorva (2013), to model bivariate motor claims by Bahraoui et al. (2015) or to approach the capital allocation problem by Bargès et al. (2009), Cossette et al. (2013), Hashorva and Ratovomirija (2015), Vernic (2016, 2017), Ratovomirija (2016) and Ratovomirija et al. (2017). Even if the just mentioned papers deal with the continuous form of Sarmanov's distribution, the flexibility of this class also allows for a discrete form, see, e.g., Danaher and Smith (2011), Hofer and Leitner (2012). In this paper, we shall use the discrete Sarmanov distribution to model the random vector of number of claims in a multivariate collective model. The motivation of this choice is based on its more flexible dependence structure compared with previous choices like, e.g., the classical multivariate Poisson distribution, and on the fact that it also allows for recursive evaluation procedures of the resulting compound distribution. Moreover, in Bolance and Vernic (2017), Sarmanov's trivariate distribution provided a good fit to a real trivariate count data set.

Recall the univariate collective model used for the aggregate claims of a portfolio,

$$S=\sum_{j=0}^{N}X_j, \tag{1}$$

where N is the number of claims with probability mass function (pmf) denoted by p, $X_0=0$, while $X_1, X_2, \ldots$ are independent, identically distributed (i.i.d.) nonnegative discrete claim amounts with pmf h, and independent of N. The distribution of S is called compound with counting distribution p and severity distribution h. This univariate collective model can be extended to a multivariate setting generated by, e.g., $m \geq 2$ different types of claims (corresponding to, e.g., different types of insurance policies, like home and auto insurance; see, e.g., Bolance and Vernic, 2017) as

$$(S_1,\ldots,S_m)=\left(\sum_{l=0}^{N_1}X_{1l},\ldots,\sum_{l=0}^{N_m}X_{ml}\right), \tag{2}$$

where the pmf of the random vector consisting of the (dependent) claim numbers $(N_1,\ldots,N_m)$ is still denoted by p, $\left(X_{jl}\right)_{l\geq 1}$ are i.i.d. nonnegative discrete claim amounts of type j with pmf h_j, independent of the claim numbers; by convention, $X_{j0}=0$, $0\leq j\leq m$. A usual simplifying assumption is that the different types of claim amounts are also independent of each other; however, we shall also introduce a certain form of dependency between them.

* Correspondence to: Faculty of Mathematics and Computer Science, Ovidius University of Constanta, 124 Mamaia Blvd., 900527 Constanta, Romania.
E-mail address: rvernic@univ-ovidius.ro.

https://doi.org/10.1016/j.insmatheco.2018.01.006
0167-6687/

Since the evaluation of compound distributions is extremely important for insurance companies, various methods have been investigated in the literature, from which we first recall the recursive method. In the univariate setting, recursions have been intensively investigated (see Sundt and Vernic, 2009 for an overview and for references); however, in the multivariate setting, due to the complexity of the models, there is still room for research, see Part II of the book Sundt and Vernic (2009), and also the recent approaches of Jin and Ren (2014) and Robe-Voinea and Vernic (2017, 2016a). In this paper, we present recursive formulas for the model (2) under the assumptions of Sarmanov counting distribution with independent claims and, also, for an extension of the same model that includes a common shock-type dependency between claims (i.e., when some event causes claims on more than one type of policy).

Though providing exact values, the recursive evaluation of compound distributions can be time consuming, especially in the multivariate setting. This is why we shall also consider the Fast Fourier Transform (FFT) method for our models; among the alternative approximate techniques, this method became of interest due to its ability to significantly reduce the computing time (see Jin and Ren, 2014 for the bivariate case and Robe-Voinea and Vernic, 2016b for the trivariate case). The development of FFT functions into the mathematical software also contributed to the rise of interest in this technique. However, since the FFT method is an approximate one, special attention must be paid to the specific errors, the most important one being the aliasing error.

Therefore, this paper is structured as follows: in Section 2 we present some preliminaries consisting of notation, useful formulas, known recursive formulas for compound distributions, and Sarmanov's multivariate discrete distribution. Section 3 is divided in two parts: in the first part, we consider the model (2) with independent claims and Sarmanov's counting distribution, and we discuss both a recursive and a FFT based evaluation technique of the corresponding compound distribution when the distribution of each random variable (r.v.) number of claims belongs to Panjer's class. In the second part of Section 3 we introduce some dependency (of common shock-type) between the claim sizes of model (2) and discuss the same two methods under the same assumptions on the counting distribution. In Section 4 we illustrate the methods for the model with dependence and compare the results regarding the accuracy and computing time. The paper ends with a conclusions section.

2. Preliminaries

2.1. Notation, definitions and useful formulas

In the following, for simplicity, we let $\overline{1,n} = \{1, 2, \ldots, n\}$ and we denote a vector by a bold-face letter and its elements by the corresponding italic with a subscript indicating the number of the element, i.e., $\mathbf{X} = (X_1, \ldots, X_n)$ or $\mathbf{x} = (x_1, \ldots, x_n)$; moreover, $\mathbf{0}$ denotes the 0-vector, $\mathbf{1}$ the vector consisting of only ones, while $\mathbf{x} - \mathbf{y}$ and $\mathbf{x} \geq \mathbf{y}$ are considered componentwise.

For a discrete function $f : \mathbb{N} \to \mathbb{R}$, we introduce the following transforms

$$\text{Laplace transform} : \mathcal{L}_f(t) = \sum_{n=0}^{\infty} f(n) e^{-nt},$$

$$\psi_f(t) = \sum_{n=0}^{\infty} f(n) t^n, \ \varphi_f(t) = \sum_{n=0}^{\infty} f(n) e^{int}.$$

When f is a discrete pmf, the last two transforms are, respectively, the probability generating function (pgf) and the characteristic function (cf); if, instead of a distribution, we deal with a discrete r.v., say X, we shall index the corresponding transforms with the name of the r.v., hence we recall

$$\mathcal{L}_X(t) = \mathbb{E}\left[e^{-tX}\right], \ \psi_X(t) = \mathbb{E}\left[t^X\right], \ \varphi_X(t) = \mathbb{E}\left[e^{itX}\right].$$

In the multivariate setting, when $f : \mathbb{N}^m \to \mathbb{R}$, the definitions of the transforms ψ and φ become

$$\psi_f(\mathbf{t}) = \sum_{\mathbf{n}\geq\mathbf{0}} f(\mathbf{n}) \prod_{j=1}^{m} t_j^{n_j}, \ \varphi_f(\mathbf{t}) = \sum_{\mathbf{n}\geq\mathbf{0}} f(\mathbf{n}) \prod_{j=1}^{m} e^{in_jt_j}.$$

For two discrete functions $f, h : \mathbb{N} \to \mathbb{R}$, we recall the definition of their convolution $f * h$, as

$$(f * h)(x) = \sum_{y=0}^{x} f(y) h(x-y), x \in \mathbb{N}.$$

If, in particular, f and h are pmfs, $f * h$ represents the pmf of a sum of two independent discrete r.v.s having distributions f and h. The n-fold convolution of the function h is recursively defined by $h^{*n} = h^{*(n-1)} * h, n \geq 1$, with $h^{*1} = h$ and, by convention, $h^{*0}(x) = 1$ if $x = 0$ and 0 otherwise. Moreover, it is easy to see that the property $\psi_{f*h} = \psi_f \psi_h$ well known for distributions, also holds for discrete functions, hence it follows that $\psi_{h^{*n}} = (\psi_h)^n$.

The convolution can also be defined in the multivariate setting for two functions $f, h : \mathbb{N}^m \to \mathbb{R}$ by

$$(f * h)(\mathbf{x}) = \sum_{\mathbf{y}=\mathbf{0}}^{\mathbf{x}} f(\mathbf{y}) h(\mathbf{x}-\mathbf{y}), \mathbf{x} \in \mathbb{N}^m.$$

We also define the compounding operation of two discrete functions $p, h : \mathbb{N} \to \mathbb{R}$ by

$$(p \vee h)(x) = \sum_{n=0}^{\infty} p(n) h^{*n}(x), x \in \mathbb{N}. \tag{3}$$

An easy calculation yields $\psi_{p\vee h}(t) = \psi_p(\psi_h(t))$. In the particular case when p and h are pmfs, $p \vee h$ represents the compound distribution with counting distribution p and severity distribution h, corresponding to model (1). Similarly, in the multivariate case, under the assumption that the $(X_{jl})_{l\geq 1}$ are independent of the $(X_{kl})_{l\geq 1}, \forall j \neq k$, we can write the compound distribution of model (2) as

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{n_1=0}^{\infty} \ldots \sum_{n_m=0}^{\infty} p(\mathbf{n}) \prod_{j=1}^{m} h_j^{*n_j}(x_j), \ \mathbf{x} \geq \mathbf{0}, \tag{4}$$

which can also be used as the definition of the compounding operation of the discrete functions $p : \mathbb{N}^m \to \mathbb{R}$ and $\mathbf{h} : \mathbb{N}^m \to \mathbb{R}^m$. In this case also, it can be proved that $\psi_{p\vee\mathbf{h}}(\mathbf{t}) = \psi_p\left(\psi_{h_1}(t_1), \ldots, \psi_{h_m}(t_m)\right)$.

We consider the following definition of the discrete Fourier transform (DFT) $\tilde{f}$ of the m-variate function f defined on the integer values $x_j = 0, 1, \ldots, r_j - 1$, with $r_j \in \mathbb{N}^*, j = \overline{1,m}$,

$$\tilde{f}(\mathbf{z}) = \sum_{x_1=0}^{r_1-1} \ldots \sum_{x_m=0}^{r_m-1} f(\mathbf{x}) \exp\left\{-2\pi i \sum_{j=1}^{m} \frac{x_j z_j}{r_j}\right\},$$
$$z_j = 0, \ldots, r_j - 1, j = \overline{1,m},$$

its inverse being given by

$$f(\mathbf{x}) = \frac{1}{\prod_{j=1}^{m} r_j} \sum_{z_1=0}^{r_1-1} \ldots \sum_{z_m=0}^{r_m-1} \tilde{f}(\mathbf{z}) \exp\left\{2\pi i \sum_{j=1}^{m} \frac{x_j z_j}{r_j}\right\},$$
$$x_j = 0, \ldots, r_j - 1, \ j = \overline{1,m}.$$

The DFT is the base of the FFT method that computes extremely fast the DFT of a discrete function, therefore having important