МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2017
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2017
Proceedings of the Forty-sixth Spring Conference
of the Union of Bulgarian Mathematicians
Borovets, April 9–13, 2017

MIXTURES, COMPOUNDS AND INFLATIONS OF NEGATIVE MULTINOMIAL DISTRIBUTION*

Pavlina Jordanova

The paper considers some relations between mixed, compound and inflated geometric and negative multinomial distributions. The dependence structure of multivariate random sums (compounds), discussed in this paper, is born to two factors, the equal number of summands in the coordinates and the multivariate distribution of the coordinates with equal indexes.

1. Introduction. Throughout the paper we suppose that $(\Omega, \mathcal{A}, P)$ is a given probability space and all considered random variables (r.vs.) and random vectors are defined on it. First we clarify the meaning of the concepts that we use. There are different notations and definitions for mixed, compound, generalized or inflated distributions and random sums in the scientific literature. In their book, Gnedenko et al. [1] consider limit theorems in the univariate case of "random sums". Willmot [10], [11] considers the compound geometric case. Sundt et al. [9] works on recursive relations of multivariate random sums, with equal number of summands, and call them "compounds", or more precisely "multivariate compounds of Type I". Smith [8] uses the name of the distribution of the number of summands, followed by "–", and then the name of the distribution of the summands. Here we use the following definitions and notations.

Definition 1.1. *Let N be a non-negative integer valued random variable (r.v.) and $(Y_{1i}, Y_{2i}, \dots, Y_{ki})$, be independent identically distributed (i.i.d.) random vectors. A random vector $(X_{1N}, X_{2N}, \dots, X_{kN})$, with coordinates*

$$X_{sN} = I_{N>0} \sum_{i=1}^{N} Y_{si}, \quad s = 1, 2, \dots, k \tag{1}$$

is called **compound N(N-stopped sum, random sum) with equal number of Y summands**. *In any particular case the letters N and Y are replaced with the name of the corresponding distribution. Briefly we will denote this in the following way*

$$(X_{1N}, X_{2N}, \dots, X_{kN}) \sim C\,N\,Y(\vec{a}_N; \vec{b}_Y).$$

Here the letter C comes from "compound". In any particular case, N is replaced by the abbreviation of the distribution of the number of summands, with parameters $\vec{a}_N$ and

*This research is partially supported by Project RD-08-96/06.02.2017 from the Scientific Research Fund in Konstantin Preslavsky University of Shumen, Bulgaria.

2010 Mathematics Subject Classification: 62E15.

Key words: mixed distributions, compound distributions, inflated distributions, negative multinomial distribution.

Y is replaced by the abbreviation of the distribution of the vector of summands, with parameters $\vec{b}_Y$.

Some properties of these distributions are summarised e.g. in [5].

In the definition of "mixed distributions" we follow the notations of Johnson et al. [3] but the approach of Grandell [2].

Definition 1.2. Mixed F_1 distribution with F_M mixing, *or briefly*

$$F_1(M, \vec{a}) \bigwedge_M F_M,$$

means that the parameter M in the distribution F_1 is replaced by a random variable with distribution F_M. Here $\vec{a}$ is the vector of parameters of the distribution F_1.

Let us note that some authors, see e.g. [3], call the last distributions "compounds".

With respect to the concept about "zero inflated distributions" we follow Johnson et al. [3]. Let us note that, frequently these distributions are particular cases of appropriate mixed or compound distributions. For example, if we consider a mixed r.v.

$$\xi \sim F_1(M, \vec{a}_F) \bigwedge_M F_M$$

and M is a strictly positive r.v., then $I_A\xi$ has the corresponding inflated distribution

$$F_1(I_A M, \vec{a}_F) \bigwedge_M F_M \equiv F_{1,A}(M, \vec{a}_F) \bigwedge_M F_M,$$

which is again mixed. Here and further on, I_A means the indicator of the event A or, which is the same, Bernoulli r.v., taking value 1 with probability $P(A)$ and value 0 with probability $1 - P(A)$, for some fixed event A.

Definition 1.3. *Let Z be a r.v. with probability law F, and the event A is such that its indicator, or I_A is independent on Z. We call the distribution of their product $I_A Z$,* **zero Inflated** *F or $ZIF(P(A); \vec{a}_F)$ distribution, where in any particular case, F will be replaced by the name or abbreviation of the corresponding distribution F with parameters $\vec{a}_F$.*

The properties of negative multinomial (NMn) distribution can be found e.g. in [3]. Let us now just remind the definition.

Definition 1.4. *Let $n \in \mathbb{N}$, $0 < p_i$, $i = 1, 2, \ldots, k$ and $p_1 + p_2 + \cdots + p_k < 1$. A vector $(Y_1, Y_2, \ldots, Y_k)$ is called* **negative multinomial distributed with parameters** *$n, p_1, p_2, \ldots, p_k$, if its probability mass function (p.m.f.) is*

$$P(\xi_1 = i_1, \xi_2 = i_2, \ldots, \xi_k = i_k) = \\ = \binom{n + i_1 + i_2 + \cdots + i_k - 1}{i_1, i_2, \ldots, i_k, n-1} p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k} (1 - p_1 - p_2 - \cdots - p_k)^n,$$

$$i_s = 0, 1, \ldots, \; s = 1, 2, \ldots, k.$$

Briefly $(\xi_1, \xi_2, \ldots, \xi_k) \sim NMn(n; p_1, p_2, \ldots, p_k)$.

It is well known that if $A_1, A_2, \ldots, A_k$ describe all possible mutually exclusive "successes" and the event $\overline{A}_1 \cap \overline{A}_2 \cap \cdots \cap \overline{A}_k$ presents the "failure" in one trial, then the i-th coordinates ξ_i of the above vector, can be interpreted as the number of "successes" of type A_i $i = 1, 2, \ldots, k$ until n-th "failure" happens. For $k = 1$ this distribution is negative Binomial (NBi) with parameters n and $1 - p_1$ to be denoted briefly by $NBi(n; 1 - p_1)$. In particular $NBi(1; 1 - p_1)$ distribution is $Ge(1 - p_1)$, i.e. geometric with parameter $1 - p_1$. If $\xi_1 \sim Ge(1 - p_1)$, then $\xi_1 + 1$ has shifted geometric distribution $(Ge_1(1 - p_1))$.

We call the distribution of sum of n independent, $Ge_1(1-p_1)$ distributed r.vs. shifted negative Binomial with parameters n and $1-p_1$. Briefly $ShNBi(n, 1-p_1)$.

We use the following notations: $\stackrel{d}{=}$ means coincidence in distribution; $\equiv$ is for equivalence between classes of distributions.

2. Univariate case. In this section, we suppose that $k = 1$, $p, p_N \in (0,1)$ and $\beta > 0$. Here $Po(\lambda)$ means Poisson distribution with parameter $\lambda > 0$, $Exp(\beta)$ means exponential distribution with parameter β and $Gamma(n, \beta)$ is for Gamma distribution with parameters n and β. It is well known and easy to prove, that

$$Po(M)\underset{M}{\wedge} Gamma(n,\beta) \equiv NBi(n, \frac{\beta}{1+\beta}),\ NBi(n, p_N) \equiv Po(M)\underset{M}{\wedge} Gamma(n, \frac{p_N}{1-p_N}),$$

$$Po(M)\underset{M}{\wedge} Exp(\beta) \equiv Ge(\frac{\beta}{1+\beta}),\ Ge(p_N) \equiv Po(M)\underset{M}{\wedge} Exp(\frac{p_N}{1-p_N}),$$

$$C\,ShNBiGe(n, p_N; p) \equiv NBi\left(n, \frac{pp_N}{1-p(1-p_N)}\right).$$

In particular $C\,Ge_1Ge(p_N; p) \equiv Ge\left(\frac{pp_N}{1-p(1-p_N)}\right)$. See e.g. [2]. Now it is easy to see that

$$C\,Ge_1Ge(p_N; p) \equiv Po(N)\underset{N}{\wedge} Exp\left(\frac{pp_N}{1-p}\right),$$

$$C\,ShNBiGe(n, p_N; p) \equiv Po(N)\underset{N}{\wedge} Gamma\left(n, \frac{pp_N}{1-p}\right),$$

$$C\,Ge_1Ge_1(p_N; p) \equiv Ge_1(pp_N), \quad C\,ShNBi\,Ge_1(n, p_N; p) \equiv ShNBi(n, pp_N).$$

In the next theorem we consider the distribution of a random sum with geometrically distributed number and shifted geometric summands, i.e. $C\,Ge\,Ge_1$ distribution. It seems to be introduced before Iwunor [6], who estimates its parameters and shows some applications in rural out-migration. This distribution is widely used in risk theory. See e.g. Willmot [10], [11] and Jordanova [4]. Some properties of this distribution can be seen e.g. in Minkova [7]. Here we show its relation with Bernoulli distribution.

Theorem 2.1. *Let $Y_i \sim Ge_1(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim Ge(p_N)$ be independent on them. Denote by I_A the indicator of the event A with $P(A) = 1 - p_N$. Then the r.v. X_N, defined in* (1) *for $k = 1$, satisfy the following equalities.*

i) *$X_N \stackrel{d}{=} I_A\xi$, where $\xi \sim Ge_1(pp_N)$ and I_A are independent, i.e.*

$$CGeGe_1(p_N; p) \equiv ZIGe_1(1-p_N; pp_N).$$

ii) $(X_N | X_N > 0) \sim Ge_1(pp_N)$.

Proof. i.) Using consecutively: the formula for probability generating function (p.g.f.) of a random sum with equal number of summands (see e.g. [8]), the form of p.g.f. of $Ge(p_N)$ distribution and p.g.f. $Ge_1(p)$, we obtain:

$$E(z^{X_N}) = \frac{p_N}{1-(1-p_N)G(z^{Y_1})} = \frac{p_N}{1-(1-p_N)\frac{pz}{1-(1-p)z}} = \frac{p_N[1-(1-p)z]}{1-(1-p)z-(1-p_N)pz} =$$

$$= \frac{p_N[1-(1-p)z]}{1-z(1-pp_N)} = \frac{p_N - zp_N + pp_N z}{1-z(1-pp_N)} = \frac{p_N - zp_N + pp_N z \pm pp_N^2 z}{1-z(1-pp_N)} =$$

$$= p_N + \frac{pp_N z - pp_N^2 z}{1-z(1-pp_N)} = p_N + (1-p_N)\frac{zpp_N}{1-(1-pp_N)z} = E(z^{I_A \xi}).$$

Now the uniqueness of p.g.f. implies i.).

The assertion ii.) follows by the definitions of conditional probability, i.) and the definition of Ge_1 distribution. □

Corollary 2.1. *Let $Y_i \sim Ge_1(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim NBi(n, p_N)$ be independent on them. Denote by $I_{A_1}, \dots, I_{A_n}$ the independent indicators correspondingly of the events A_i with $P(A_i) = 1 - p_N$, $i = 1, 2, \dots, n$. Then the r.v. X_N, defined in (1) for $k = 1$, satisfy the following equality in distribution*

$$X_N \stackrel{d}{=} I_{A_1}\xi_1 + \cdots + I_{A_n}\xi_n, \tag{2}$$

where $\xi_i \sim Ge_1(pp_N)$, $i = 1, 2, \dots, n$ are i.i.d., and independent on $I_{A_1}, \dots, I_{A_n}$.

Note. The latter means that $CNBiGe_1(n, p_N; p)$ distribution coincides with the n-th convolution of $ZIGe_1(1 - p_N; pp_N)$ distribution.

Theorem 2.2. *Let $Y_i \sim Ge(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim NBi(n, p_N)$ be independent on them. Then for the r.v. X_N, defined in (1) and $k = 1$,*

$$X_N \sim NBi(M, p) \bigwedge_M NBi(n, p_N). \tag{3}$$

That is equivalent to the equality $C\,NBi\,Ge(n, p_N; p) \equiv NBi(M, p) \bigwedge_M NBi(n, p_N)$.

Corollary 2.2. *Let $Y_i \sim Ge(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim Ge(p_N)$ be independent on them. Then for the r.v. X_N, defined in (1) for $k = 1$,*

$$X_N \sim NBi(M, p) \bigwedge_M Ge(p_N). \tag{4}$$

In other words $C\,Ge\,Ge(p_N; p) \equiv NBi(M, p) \bigwedge_M Ge(p_N)$.

Note. The relation $\bigwedge_M$ in (3) is not symmetric with respect to p and p_N. More precisely, there exist $p \neq p_N$, such that (3) is true and

$$NBi(M, p) \bigwedge_M NBi(n, p_N) \not\equiv NBi(M, p_N) \bigwedge_M NBi(n, p).$$

Theorem 2.3. *Let $Y_i \sim Ge(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim Ge(p_N)$ be independent on them. Denote by I_A the indicator of the event A with $P(A) = 1 - p_N$. Then the r.v. X_N, defined in (1) for $k = 1$, satisfy the following equalities.*

i) $X_N \stackrel{d}{=} I_A \xi$, *where* $\xi \sim Ge\left(\frac{pp_N}{1-p(1-p_N)}\right)$ *and I_A are independent, i.e.*

$$CGeGe(p_N; p) \equiv ZIGe\left(1 - p_N; \frac{pp_N}{1-p(1-p_N)}\right).$$

ii) $(X_N | X_N > 0) \sim \left(\frac{pp_N}{1-p(1-p_N)}\right)$.

Corrolary 2.3. *Let $Y_i \sim Ge(p)$, $i = 1, 2, \dots$ be i.i.d. and the r.v. $N \sim NBi(n, p_N)$ be independent on them. Denote by $I_{A_1}, \dots, I_{A_n}$ the independent indicators correspondingly of the events A_i with $P(A_i) = 1 - p_N$, $i = 1, 2, \dots, n$. Then the r.v. X_N, defined in* (1) *for $k = 1$, satisfy the equality* (2) *where $\xi_i \sim Ge\left(\frac{pp_N}{1-p(1-p_N)}\right)$, $i = 1, 2, \dots, n$ are i.i.d., and independent on $I_{A_1}, \dots, I_{A_n}$.*

Note The last means that $CNBiGe(n, p_N; p)$ distribution coincides with the n-th convolution of $ZIGe\left(1 - p_N; \frac{pp_N}{1-p(1-p_N)}\right)$ distribution.

3. Multivariate case. Here we consider the case with geometric number of summands N, and NMn distributed summands. The case of Poisson distributed N, is investigated in Smith [8].

Here we denote by $MGe(p_1, p_2, \dots, p_k)$ the $NBi(1, p_1, p_2, \dots, p_k)$ distribution, in order to point out that it is a multivariate geometric.

Theorem 3.1. *Let $(Y_{i1}, Y_{i2}, \dots, Y_{ik}) \sim MGe(p_1, p_2, \dots, p_k)$, $i = 1, 2, \dots$ and the r.v. $N - n \sim NBi(s; p_N)$ be independent. Denote by $p_0 = 1 - (p_1 + p_2 + \dots + p_k)$. Assume that $(X_{1N}, X_{2N}, \dots, X_{kN})$ is defined in* (1)*. Then*

$$(5)\quad \begin{aligned}(X_{1N}, X_{2N}, \dots, X_{kN}) &\stackrel{d}{=} \\ &\stackrel{d}{=} NMn\left(s, \frac{p_1}{1-(1-p_N)p_0}, \frac{p_2}{1-(1-p_N)p_0}, \dots, \frac{p_k}{1-(1-p_N)p_0}\right),\end{aligned}$$

i.e. $C\,ShNBi\,MGe(s, p_N; p_1, p_2, \dots, p_k) \equiv$

$$\equiv NMn\left(s, \frac{p_1}{1-(1-p_N)p_0}, \frac{p_2}{1-(1-p_N)p_0}, \dots, \frac{p_k}{1-(1-p_N)p_0}\right).$$

Proof. The p.g.f. of a random sum with equal number of summands, p.g.f. of $N - n$ and p.g.f. of $MGe(p_1, p_2, \dots, p_k)$ imply,

$$E(z_1^{X_{1N}} z_2^{X_{2N}} \cdots z_k^{X_{kN}}) =$$

$$= \left\{\frac{p_N G(z_1^{Y_{11}}, z_2^{Y_{21}}, \dots, z_k^{Y_{k1}})}{1-(1-p_N)G(z_1^{Y_{11}}, z_2^{Y_{21}}, \dots, z_k^{Y_{k1}})}\right\}^s = \left\{\frac{p_N \frac{1-(p_1+p_2+\dots+p_k)}{1-p_1z_1-p_2z_2-\dots-p_kz_k}}{1-(1-p_N)\frac{1-(p_1+p_2+\dots+p_k)}{1-p_1z_1-p_2z_2-\dots-p_kz_k}}\right\}^s =$$

$$= \left\{\frac{\frac{p_Np_0}{1-(1-p_N)p_0}}{1-\frac{p_1}{1-(1-p_N)p_0}z_1-\frac{p_2}{1-(1-p_N)p_0}z_2-\dots-\frac{p_k}{1-(1-p_N)p_0}z_k}\right\}^s.$$

Now the uniqueness of p.g.f. entails (5). □

Theorem 3.2. *Let $(Y_{i1}, Y_{i2}, \dots, Y_{ik}) \sim MGe(p_1, p_2, \dots, p_k)$, $i = 1, 2, \dots$ and the r.v. $N \sim Ge(1 - p_N)$ be independent. If $(X_{1N}, X_{2N}, \dots, X_{kN})$ is defined in* (1) *and $p_0 = 1 - (p_1 + p_2 + \dots + p_k)$, then*

i) $(X_{1N}, X_{2N}, \dots, X_{kN}) \stackrel{d}{=} I_A\xi$, *where the random vector*

$$\xi \sim Ge\left(\frac{p_1}{1-(1-p_N)p_0}, \frac{p_2}{1-(1-p_N)p_0}, \dots, \frac{p_k}{1-(1-p_N)p_0}\right)$$

and the Bernoulli r.v. I_A is independent on the random vector ξ and $P(I_A = 1) = P(A) = 1 - p_N$.

ii) *For all* $r = 1, 2, \ldots, k$ *and any subset* $i_1, i_1, \ldots, i_r$ *of* $\{1, 2, \ldots, n\}$,

$$(X_{i_1N}, X_{i_2N}, \ldots, X_{i_rN}) \sim$$
$$\sim ZI\, MGe\left(1 - p_N; \frac{p_{i_1}}{p_0 + \sum_{s=1}^{r} p_{i_s}}, \frac{p_{i_2}}{p_0 + \sum_{s=1}^{r} p_{i_s}}, \ldots, \frac{p_{i_r}}{p_0 + \sum_{s=1}^{r} p_{i_s}}\right).$$

Proof. The proof of i.) follows analogously to the proof of Th. 3.1 and we omit it. The proof of ii) follows by the properties of p.g.f. □

Note. The result in i) means that

$$C\, Ge\, MGe(p_N; p_1, p_2, \ldots, p_k) \equiv$$
$$\equiv ZI\, MGe\left(\frac{p_1}{1 - (1 - p_N)p_0}, \frac{p_2}{1 - (1 - p_N)p_0}, \ldots, \frac{p_k}{1 - (1 - p_N)p_0}\right).$$

REFERENCES

[1] B. V. Gnedenko, V. Y. Korolev. Random summation: limit theorems and applications. CRC Press, 1996.

[2] J. Grandell. Mixed Poisson processes, Vol. **77**. CRC Press, 1997.

[3] N. L. Johnson, S. Kotz, N. Balakrishnan. Discrete multivariate distributions, vol. **165**, New York, Wiley, 1997.

[4] P. Jordanova. On the total duration of negative surplus of a risk process with two-step premium function, *Applications and Applied Mathematics*, **2**, No 2 (2007), 107–121.

[5] P. Jordanova. On multivariate discrete compound distributions with one and the same number of summands. Proceedings of MATTEX 2016, Shumen, Bulgaria, vol. **1**, 2016, 127–135 (in Bulgarian).

[6] Ch. C. O. Iwunor. Estimation of parameters of the inflated geometric distribution for rural out-migration. *Genus*, **51**, No 3/4 (1995), 253–260.

[7] L. Minkova. Distributions in Insurance Risk Models, Thesis for the degree of Doctor of Science, Faculty of Mathematics and Informatics, Sofia University, 2012.

[8] G. E. J. Smith. Parameter estimation in some multivariate compound distributions, Ph. D. Thesis, Department of Mathematics, The University of British Columbia, Canada, 1965.

[9] B. Sundt, R. Vernic. Recursions for convolutions and compound distributions with insurance applications. Springer Science and Business Media, 2009.

[10] G. E. Willmot. Compound geometric residual lifetime distributions and the deficit at ruin. *Insurance: Mathematics and Economics* **30**, No 3 (2002), 421–438.

[11] G. E. Willmot. On higher-order properties of compound geometric distributions. *Journal of Applied Probability*, **39**, No 2 (2002), 324–340.

Pavlina K. Jordanova
Faculty of Mathematics and Informatics
Shumen University
115, Universitetska St
9712 Shumen, Bulgaria
e-mail: pavlina_kj@abv.bg

СМЕСИ, СЛУЧАЙНИ СУМИ И ИНФЛАЦИИ В НУЛАТА НА ОТРИЦАТЕЛНОТО БИНОМНО РАЗПРЕДЕЛЕНИЕ

Павлина Калчева Йорданова

Статията разглежда връзката между смесеното и съставното многомерно геометрично и отрицателно биномно разпределение, както и техните модификации с допълнително тегло в нулата. Структурата на зависимост на многомерните случайни суми (съставни разпределения), разглеждани в тази статия, се управлява от два фактора: равния брой на събираемите, които участват в координатите, и многомерното разпределение на координатите с еднакви индекси.