MATTEX 2018,
CONFERENCE PROCEEDING, v. 1, pp. 113- 120

ON THE COMPOUND BINOMIAL DISTRIBUTION*

KRASIMIRA Y. KOSTADINOVA

ABSTRACT: *In this paper a compound binomial distribution with geometric compounding distribution is analyzed. We derive some of its basic properties, recursion formulas and probability mass function. Then, the method of moments estimation for parameters of a compound binomial distribution is applied.*

KEYWORDS: *Compound distributions, geometric distribution, binomial distribution, moment estimation method.*

1 Introduction

In the present paper we consider a compound probability distribution, which has the form

$$N = X_1 + X_2 + \ldots + X_Z. \tag{1}$$

Here the random variable Z belongs to the class of the Generalized Powers Series Distributions. The binomial, negative binomial, logarithmic series and Poisson distributions belong to this class, see [10]. The random variables X_i are independent, identically distributed as random variable X, which has a geometric distribution with parameter $\gamma \in (0,1)$. In [3] is defined a compound Generalized Powers Series Distributions, where the compounding random variable X has a shifted geometric distribution with parameter $1-\gamma$, denoted by $X \sim Ge_1(1-\gamma),\ \gamma \in [0,1)$.

In the insurance risk theory, some of the Generalized Powers Series Distributions are used for the number of claims during a given time period. The random sum N is interpreted as the aggregate claim amount in the risk model and we say that the random variable N has a compound distribution. When Z has a Poisson distribution, then the random variable N has a Pólya-Aeppli distribution, see [2]. When Z has a negative binomial distribution, then the random variable N has a compound Pólya distribution, see [4].

In the literature there are many generalizations of the classical risk model. One such generalization is by compounding of the counting process, which is the homogeneous Poisson process in the classical risk model. For example in [7] is defined Pólya-Aeppli process. Another such generalization is by mixing, see for example [1], [5] and [6].

The rest of this article is organized as follows. In Section 2, we consider a compound binomial distribution and derive its probability mass function, some properties and recursion formulas. In Section 3, the method of moments estimation for parameters of a compound binomial distribution is given.

2 Compound binomial distribution

In this section we consider the random sum (1), where the random variables X_i are independent, identically distributed as random variable X. In this paper, we suppose that the random variable X has a geometric distribution with parameter $\gamma \in (0,1)$, denoted by $X \sim Ge(\gamma)$. The probability mass function and probability generating function of the random variable X are given by

$$q_i = P(X = i) = \gamma(1-\gamma)^i, \ \ i = 0,1,\ldots \tag{2}$$

*Supported by Scientific Research Grant RD-08-145/2018 of Shumen University.

and

$$\psi_1(s) = \frac{\gamma}{1-(1-\gamma)s}, \; |s| < \frac{1}{1-\gamma}. \tag{3}$$

Let the random variable Z is independent of the random variables $X_i,\ i = 1,2,\ldots$ We suppose that the random variable Z has a binomial distribution with parameters $n \in \mathbf{N}$ and $\theta \in (0,1)$, denoted by $Z \sim Bi(n,\theta)$. Then the random variable N has a compound binomial distribution with compounding random variable X.

The probability mass function and probability generating function of the random variable Z are given by

$$P(Z=i) = \binom{n}{i}\theta^i(1-\theta)^{n-i}, \;\; i = 0,1,\ldots,n \tag{4}$$

and

$$\psi_Z(s) = (1-\theta+\theta s)^n. \tag{5}$$

Then the probability generating function of N is given by

$$\psi(s) = \psi_N(s) = (1-\theta+\theta\psi_1(s))^n, \tag{6}$$

where $\psi_1(s)$ is the probability generating function of the compounding distribution, given by (3).

Definition 2.1. *The probability distribution of the random variable N, defined by the probability generating function (6) and compounding distribution, given by (2) and (3) is called a compound binomial distribution with notation $N \sim compBi(n,\theta,\gamma)$.*

Remark 2.1. *The mean and the variance of the compound binomial distribution are given by*

$$E(N) = \frac{n\theta(1-\gamma)}{\gamma},$$

$$Var(N) = \frac{(1-\gamma)((1-\gamma)(2-\theta)+\gamma)}{\gamma^2}n\theta.$$

For the Fisher index of dispersion we obtain

$$FI(N) = \frac{Var(N)}{E(N)} = 1 + \frac{(1-\gamma)(2-\theta)}{\gamma} > 1,$$

i.e. the compound binomial distribution is over-dispersed.

It is known that when $FI(N) < 1$, the distribution is called under-dispersed. When $FI(N) = 1$, the distribution is equi-dispersed and when $FI(N) > 1$, the distribution is over-dispersed, see [8] and [12].

From Remark 2.1 follows that the compound binomial distribution is over-dispersed related to the Poisson distribution, which $FI(N) = 1$. This makes the compound binomial distribution suitable for financial data.

2.1 The Probability Mass Function

The probability function of the random variable N is given by expanding the probability generating function $\psi(s)$ in powers of s. Denote by $f(i) = P(N = i),\ i = 0, 1, \ldots,$ the probability mass function of the random variable N. We rewrite the probability generating function of (6) in the form

$$\psi(s) = \sum_{m=0}^{n} \binom{n}{m} \left(\frac{\theta\gamma}{1-(1-\gamma)s} \right)^{m} (1-\theta)^{n-m}. \tag{7}$$

Denote by $\psi^{(i)}(s) = \dfrac{\partial^{(i)}\psi(s)}{\partial s^i}$, for $i = 0, 1, \ldots,$ the derivatives of $\psi(s)$. From (7) we get the following

$$\psi^{(i)}(s) = (1-\gamma)^i \sum_{m=1}^{n} \binom{n}{m} \frac{m(m+1)\ldots(m+i-1)}{(1-(1-\gamma)s)^{m+i}} (\theta\gamma)^m (1-\theta)^{n-m}. \tag{8}$$

From [2], it is known that

$$f(i) = \left. \frac{\psi^{(i)}(s)}{i!} \right|_{s=0}. \tag{9}$$

Theorem 2.1. *The probability mass function of the compound binomial distribution, denoted by $N \sim compBi(n, \theta, \gamma)$, is given by*

$$\begin{aligned} f(0) &= (1-\theta+\theta\gamma)^n, \\ f(i) &= (1-\gamma)^i \textstyle\sum_{m=1}^{n} \binom{n}{m}\binom{m+i-1}{i} (\theta\gamma)^m (1-\theta)^{n-m}, \quad i = 1, 2, \ldots \end{aligned} \tag{10}$$

Proof. The initial value $f(0) = (1-\theta+\theta\gamma)^n$ follows simply from the probability generating function $\psi(0) = f(0)$. Then (10) follows from (8) and (9).

□

On Figure 1 is given a graphic of the probability mass function of the binomial distribution, denoted by $Z \sim Bi(n, \theta)$. On Figure 2 is given a graphic of the probability mass function of the compound binomial distribution, denoted by $N \sim compBi(n, \theta, \gamma)$. A graphic of the probability mass function of $Z \sim Bi(n, \theta)$ and $N \sim compBi(n, \theta, \gamma)$ is given on Figure 3. All these graphics are made for $n = 20,\ \theta = 0.7$ and $\gamma = 0.5$.

From Figure 2 we see that the tail distribution is longer than the tail distribution on Figure 1. From Figure 3, one can see that the tail of the compound binomial distribution becomes heavier than the tail of the binomial distribution.

The following proposition gives an extension of the Panjer recursion formulas (see [9]).

Proposition 2.1. *The probability mass function of the compound binomial distribution satisfies the following recursions*

$$(1-\theta+\theta\gamma)if(i) = (1-\gamma)\left[(1-\theta)(i-1)f(i-1) + n\theta \sum_{j=0}^{i-1} q_j f(i-j-1)\right],\ i = 2, 3, \ldots \tag{11}$$

and $f(1) = \frac{n\theta(1-\gamma)}{1-\theta+\theta\gamma} q_0 f(0)$ *with* $f(0) = (1-\theta+\theta\gamma)^n$.

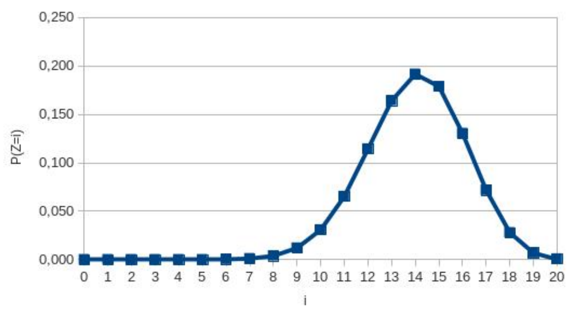

Figure 1. The PMF of Z~Bi(20, 0.7)

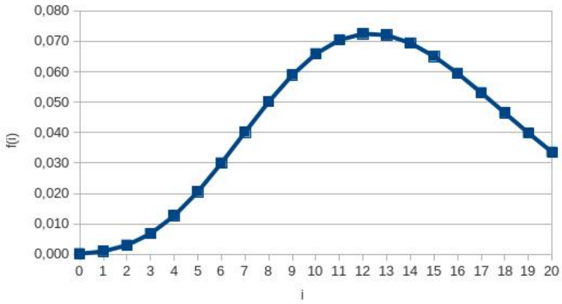

Figure 2. The PMF of N~compBi(20, 0.7, 0.5)

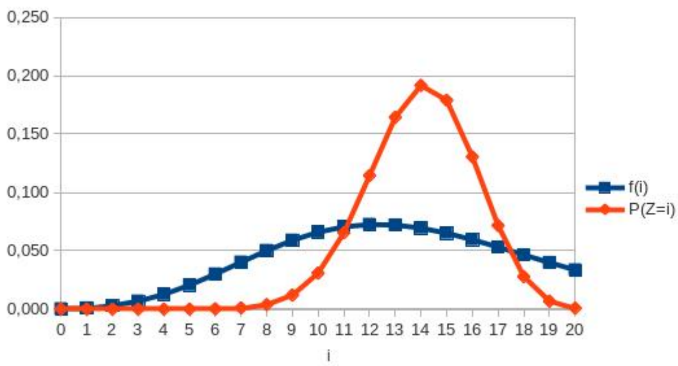

Figure 3. The PMF of Z~Bi(20, 0.7) and N~compBi(20, 0.7, 0.5)

Proof. Differentiation in (6) leads to

$$\frac{\partial \psi(s)}{\partial s} = \frac{n\theta(1-\gamma)}{(1-\theta)(1-(1-\gamma)s)+\theta\gamma}\psi_1(s)\psi(s), \tag{12}$$

where $\psi(s) = \sum_{i=0}^{n} f(i)s^i$, $\frac{\partial \psi(s)}{\partial s} = \sum_{i=0}^{n}(i+1)f(i+1)s^i$, and $\psi_1(s) = \sum_{j=0}^{\infty} q_j s^j$. The equation (12) has the form

$$[(1-\theta)(1-(1-\gamma)s)+\theta\gamma]\sum_{i=0}^{n}(i+1)f(i+1)s^i \;=n\theta(1-\gamma)\sum_{i=0}^{n}f(i)s^i\sum_{j=0}^{\infty}q_js^j.$$

Changing the variable from $i+j=l \Rightarrow i=l-j$, yields

$$[(1-\theta)(1-(1-\gamma)s)+\theta\gamma]\sum_{i=0}^{n}(i+1)f(i+1)s^i \;=n\theta(1-\gamma)\sum_{j=0}^{\infty}q_j\sum_{l=j}^{n+j}f(l-j)s^l.$$

Interchanging the order of summing in the double sum and equivalent transformations results in

$$\begin{aligned}[1-\theta+\theta\gamma]\sum_{i=0}^{n}(i+1)f(i+1)s^i \;&=(1-\theta)(1-\gamma)\sum_{i=1}^{n+1}if(i)s^i\\ &+n\theta(1-\gamma)\left[\sum_{i=0}^{n}\sum_{j=0}^{i}q_jf(i-j)+\sum_{i=n}^{\infty}\sum_{j=i-n}^{i}q_jf(i-j)\right]s^i.\end{aligned}$$

The recursions (11) are obtained by equating the coefficients of s^i on both sides for fixed $i=0,1,2,\dots$ For $i=0$ follows that

$$f(1)=\frac{n\theta(1-\gamma)}{1-\theta+\theta\gamma}q_0f(0).$$

For $i=1,\dots,n$

$$(1-\theta+\theta\gamma)(i+1)f(i+1)=(1-\gamma)\left[(1-\theta)if(i)+n\theta\sum_{j=0}^{i}q_jf(i-j)\right],$$

and hence (11).

□

In the next proposition we give an alternative recursion formulas.

Proposition 2.2. *The probability mass function of the compound binomial distribution satisfies the recursions*

(13)
$$\begin{aligned}(1-\theta+\theta\gamma)if(i) \;&=(1-\gamma)\left[(i-1)(2-2\theta+\theta\gamma)+n\theta\gamma\right]f(i-1)\\ &-(1-\gamma)^2(1-\theta)(i-2)f(i-2),\ i=2,3,\dots\end{aligned}$$

and $f(1)=\frac{n\theta\gamma(1-\gamma)}{1-\theta+\theta\gamma}f(0)$ *with* $f(0)=(1-\theta+\theta\gamma)^n$.

Proof. Differentiation in (6) leads to

(14)
$$\psi'(s)=\frac{n\theta}{1-\theta+\theta\psi_1(s)}\psi_1'(s)\psi(s),$$

where $\psi(s)=\sum_{i=0}^{n}f(i)s^i$, $\frac{\partial\psi(s)}{\partial s}=\sum_{i=0}^{n}(i+1)f(i+1)s^i$, and

$$\psi_1'(s)=\frac{(1-\gamma)\gamma}{(1-(1-\gamma)s)^2}$$

is the derivative of (3). So, the equation (14) has the form

$$[(1-\theta)(1-(1-\gamma)s)+\theta\gamma](1-(1-\gamma)s)\sum_{i=0}^{n}(i+1)f(i+1)s^i = n\theta\gamma(1-\gamma)\sum_{i=0}^{n}f(i)s^i,$$

or equivalently

$$(1-\theta+\theta\gamma)\sum_{i=0}^{n}(i+1)f(i+1)s^i = (2-2\theta+\theta\gamma)(1-\gamma)\sum_{i=1}^{n+1}if(i)s^i$$
$$-(1-\gamma)^2(1-\theta)\sum_{i=2}^{n+2}(i-1)f(i-1)s^i+n\theta\gamma(1-\gamma)\sum_{i=0}^{n}f(i)s^i.$$

The recursions are obtained by equating the coefficients of s^i on both sides for fixed $i=0,1,2,\dots$ □

3 Method of moments

3.1 The Procedure

Let $X_1, X_2, \dots,$ are independent, identically distributed random variables, which have some distribution. Then the kth moment of the distribution is defined by

$$\mu_k = E(X^k).$$

For example $\mu_1 = E(X)$ and $\mu_2 = Var(X)+(E(X))^2$.

The procedure follows these four steps (see for example [11]):

1) If the model has m parameters, we compute m moments, $\mu_1, \mu_2, \dots, \mu_m$ and obtain m equations with m unknowns.

2) Then we solve so that these m parameters as a function of the moments, i.e. every parameter we express by $\mu_i,\ i=1,2,\dots,m$.

3) After that, based on the data $X=(X_1,X_2,\dots,X_n)$, we compute the first m sample moments, $\overline{X^m} = \frac{1}{n}\sum_{i=1}^{n} X_i^m$.

4) We replace the distributional moments μ_m by the sample moments $\overline{X^m}$.

3.2 Example - compound binomial distribution

Let us denote by $\mu_k = E(N^k)$, the kth moment of the random variable N. Using that

$$\mu_1 = E(N) = \frac{n\theta(1-\gamma)}{\gamma},$$

$$\mu_2 = E(N^2) = \mu_1\left[1+\mu_1+\frac{(1-\gamma)(2-\theta)}{\gamma}\right],$$

$$\mu_3 = E(N^3) = \frac{\mu_1^3}{(n\theta)^2}\left[(n-1)(n-2)\theta^2+6(n-1)\theta+6\right]+3\mu_2-2\mu_1,$$

by the method of moments for the parameters n, θ and γ of the compound binomial distribution, we obtain

$$\widehat{n} = \frac{(\overline{X})^2[2\overline{X}\overline{X^2}+2\overline{X^2}(\overline{X})^2-2(\overline{X})^3-(\overline{X^2})^2-(\overline{X})^2-(\overline{X})^4 \mp |u|v]}{u(v^2 \pm |u|v)},$$

$$\widehat{\theta}=\frac{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4\pm|u|v}{\overline{X}[\overline{X^2}+\overline{X^3}+\overline{X}\overline{X^2}-(\overline{X})^2]-2(\overline{X^2})^2},$$

$$\widehat{\gamma}=\frac{\overline{X}[\overline{X}\overline{X^2}+3\overline{X^2}(\overline{X})^2-(\overline{X})^3-(\overline{X^2})^2-(\overline{X})^4+w\mp|u|v]}{\overline{X}[2(\overline{X^2})^2+\overline{X^2}.\overline{X^3}+3\overline{X}(\overline{X^2})^2-\overline{X}\overline{X^3}-\overline{X^3}(\overline{X})^2-\overline{X^2}(\overline{X})^3+w\mp|u|v]-2(\overline{X^2})^3},$$

where $u=\overline{X^2}-\overline{X}-(\overline{X})^2$, $v=\sqrt{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4}$, $w=\overline{X}\overline{X^2}-(\overline{X})^2-(\overline{X})^3-\overline{X^2}(\overline{X})^2$ and $\overline{X^k}=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,2,3.$

Table 1.

i	*Observed*
0	370410
1	43000
2	3930
3	300
4	27
5	3
≥ 6	0

Taking the data from the Table 1 we obtain the following moment estimates for the parameters n,θ and $\gamma:\widehat{n}=2,\ \widehat{\theta}=0.972$ and $\widehat{\gamma}=0.937$.

Concluding remarks

In this paper we have introduced a compound binomial distribution as a compound binomial distribution with geometric compounding distribution. Also, we find the moments, the recursion formulas and probability mass function and then the method of moments estimation for its parameters is applied.

REFERENCES:

[1] Grandell, J., Mixed Poisson Process. Chapman & Hall, London (1997).

[2] Johnson, N.L., Kemp, A.W., Kotz, S., Univariate Discrete Distributions. Third edition, John Wiley & Sons, Hoboken, New Jersey (2005).

[3] Kolev, N., Minkova, L., Neytchev P., Inflated-Parameter Family of Generalized Power Series Distributions and Their Application in Analysis of Overdispersed Insurance Data. ARCH Research Clearing House, **2** (2000), 295–320.

[4] Kostadinova, K.Y., On the compound Pólya distribution. MATTEX 2016, vol. **1** (2016), 51–58.

[5] Lazarova, M.D., Minkova, L.D., I-Delaporte process and applications. Mathematics and Computers in Simulation, vol. **133** (2017), 135–141.

[6] Minkova, L.D., I-Pólya Process and Applications. Commun. Statist.-Theory and Methods, vol. **40** (2011), 2847–2855.

[7] Minkova, L.D., The Pólya-Aeppli process and ruin problems. J. Appl. Math. Stoch. Analysis, **3** (2004), 221–234.

[8] Minkova, L.D., Balakrishnan, N., Compound weighted Poisson distributions. Metrika, **76(4)** (2013), 543–558.

[9] Panjer, H., Recursive evaluation of a family of compound discrete distributions. ASTIN Bull., vol. **12** (1) (1981), 22–26.

[10] Patil, G.P., Certain properties of the generalized power distributions. Ann. Inst. Stat. Math., vol. **14** (1962), 179–182.

[11] Rice, J.A., Mathematical Statistics and Data Analysis. Duxbury Press, Belmont, CA, third edition (2007).

[12] Xekalaki, E., Under- and overdispersion. John Wiley & Sons, Hoboken, New Jersey. In: Teugels, J.L., Sundt, B. (eds) Encyclopedia of actuarial science, vol. **3** (2006), 1700–1705.

Krasimira Yankova Kostadinova,
Faculty of Mathematics and Informatics, Shumen University
E-mail: kostadinova@shu.bg