ON THE COMPOUND PÓLYA DISTRIBUTION

Krasimira Y. Kostadinova

ABSTRACT: *In this paper we define a compound Pólya distribution with geometric compounding distribution. We derive some of its basic properties, recursion formulas and probability mass function. The method of moments estimation for parameters of a compound Pólya distribution is applied. Finally, a definition of a compound Pólya process and its probability mass function is given.*

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

1 Introduction

The Pólya process is defined and analyzed in [1]. It is a mixed Poisson process with Gamma mixing distribution. In a similar way in [4] the I-Pólya process is defined as a mixed Pólya-Aeppli process with Gamma mixing distribution. The I-Pólya process is defined also as a pure birth process, see also [3]. In this paper we consider the I-Pólya process as a compound Pólya process with geometric compounding distribution.

In the next Section 2 we define the compound Pólya distribution and derive its probability mass function, some properties and recursion formulas. The method of moments estimation for parameters of a compound Pólya distribution is given in Section 3. In Section 4, the compound Pólya process and its probability mass function is given.

2 Compound Pólya distribution

In this section we consider the random sum

(1) $$N = X_1 + X_2 + \ldots + X_Z,$$

where the random variables (r.v.'s) X_i are independent, identically distributed (iid) as X r.v.'s, and the r.v. Z is independent of the r.v.'s X_i, $i = 1, 2, \ldots$. We suppose that the r.v. Z has a negative binomial distribution with parameters $r \in \mathbf{N}$ and $\theta \in (0, 1)$, denoted by $Z \sim NB(r, \theta)$. Then the r.v. N has a compound negative binomial distribution with compounding r.v. X. Let the r.v. X has a geometric distribution with parameter $\gamma \in (0, 1)$, denoted by $X \sim Ge(\gamma)$. The probability mass function (PMF) and probability generating function (PGF) of X are given by

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

and

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

The PMF and PGF of Z are given by

(4) $$P(Z = i) = \binom{r+i-1}{i}(1-\theta)^r\theta^i, \quad i = 0, 1, \ldots$$

and

$$\psi_Z(s) = \left(\frac{1-\theta}{1-\theta s}\right)^r, \ |s| < \frac{1}{\theta}. \tag{5}$$

Then the PGF of N is given by

$$\psi_N(s) = \left(\frac{1-\theta}{1-\theta\psi_1(s)}\right)^r, \tag{6}$$

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

Definition 2.1. *The probability distribution of N, defined by the PGF (6) and compounding distribution, given by (2) and (3) is called a compound Pólya distribution.*

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

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

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

For the Fisher index of dispersion we obtain

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

i.e. the compound Pólya distribution is over-dispersed related to the Poisson distribution.

2.1 The Probability Mass Function

The probability function of $N(t)$ is given by expanding the PGF $\psi(s)$ in powers of s. Denote by $f(i) = P(N(t) = i)$, $i = 0, 1, 2, \dots$, the probability mass function of $N(t)$. We rewrite the PGF of (6) in the form

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

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

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

From [2], it is known that

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

Theorem 2.1. *The probability mass function of $N(t)$ is given by*

$$
(10)\qquad
\begin{aligned}
f(0) &= \left(\tfrac{1-\theta}{1-\theta\gamma}\right)^r, \\
f(i) &= (1-\gamma)^i(1-\theta)^r \sum_{m=1}^{\infty} \tbinom{r+m-1}{m}\tbinom{m+i-1}{i}(\theta\gamma)^m, \quad i = 1, 2, \ldots .
\end{aligned}
$$

Proof. The initial value $f(0) = \left(\frac{1-\theta}{1-\theta\gamma}\right)^r$ follows simply from the PGF $\psi(0) = f(0)$. Then (10) follows from (8) and (9). □

On Figure 1 and Figure 2 are given graphics of the PMF of negative binomial distribution and the compound Pólya distribution correspondingly, and on Figure 3 is given a graphic of the PMF of $Z \sim NB(r, \theta)$ and the compound Pólya distribution for $\gamma = 0.5$, $\theta = 0.7$ and $r = 5$.

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 Pólya distribution becomes heavier than the tail of the negative binomial distribution.

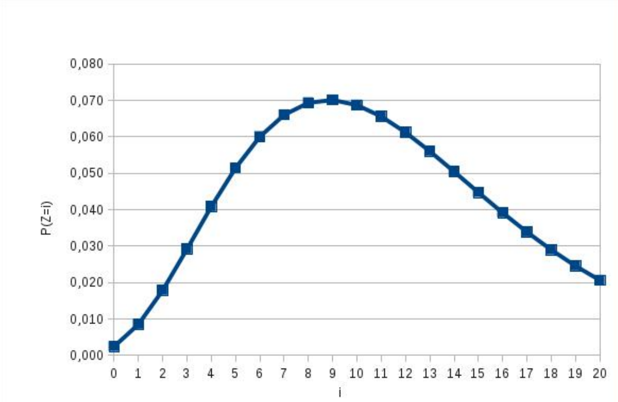

Figure 1. The PMF of Z~NB(5, 0,7)

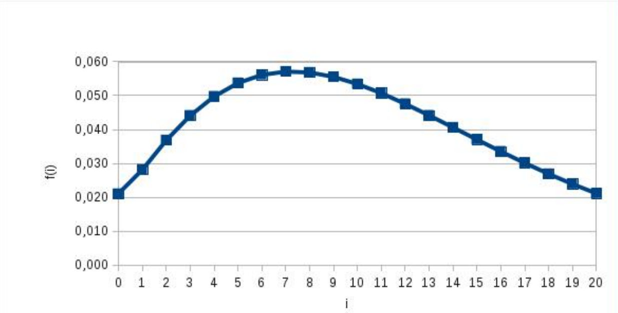

Figure 2. The PMF of N(t)

The following proposition gives an extension of the Panjer recursion formulas (see [5]).

Proposition 2.1. *The PMF of the compound Pólya distribution satisfies the following recursions*

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

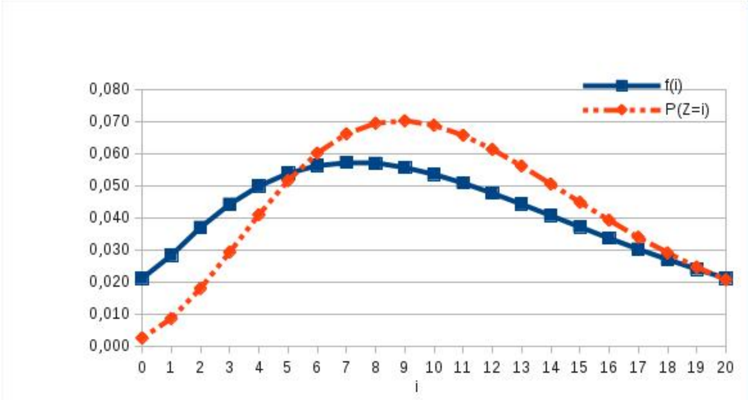

Figure 3. The PMF of Z~NB(5, 0,7) and N(t)

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

Proof. Differentiation in (6) leads to

$$\frac{\partial \psi_{N(t)}(s)}{\partial s} = \frac{(1-\gamma)\theta r}{1-(1-\gamma)s-\theta\gamma}\psi_1(s)\psi_{N(t)}(s), \tag{12}$$

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

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

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

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

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

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

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

$$f(1) = \frac{1-\gamma}{1-\theta\gamma}\theta r q_0 f(0).$$

For $i = 1, 2, \ldots$

$$(1-\theta\gamma)(i+1)f(i+1) = (1-\gamma)if(i) + (1-\gamma)\theta r\sum_{j=0}^{i} q_j f(i-j),$$

and hence (11).

□

In the next proposition we give an alternative recursion formula.

Proposition 2.2. *The PMF of the compound Pólya distribution satisfies the recursions*

(13) $(1-\theta\gamma)if(i)=(1-\gamma)\left[(i-1)(2-\theta\gamma)+r\theta\gamma\right]f(i-1)-(1-\gamma)^2(i-2)f(i-2),\ i=2,3,\ldots$

and $f(1)=\frac{1-\gamma}{1-\theta\gamma}\theta r\gamma f(0)$ *with* $f(0)=\left(\frac{1-\theta}{1-\theta\gamma}\right)^r$.

Proof. Differentiation in (6) leads to

$$\psi'_{N(t)}(s)=\frac{\theta r}{1-\theta\psi_1(s)}\psi'_1(s)\psi_{N(t)}(s), \tag{14}$$

where $\psi_{N(t)}(s)=\sum_{i=0}^{\infty}f(i)s^i$, $\frac{\partial\psi_{N(t)}(s)}{\partial s}=\sum_{i=0}^{\infty}(i+1)f(i+1)s^i$, 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

$$(1-(1-\gamma)s-\theta\gamma)(1-(1-\gamma)s)\sum_{i=0}^{\infty}(i+1)f(i+1)s^i \quad = r\theta\gamma(1-\gamma)\sum_{i=0}^{\infty}f(i)s^i,$$

or equivalently

$$\begin{aligned}(1-\theta\gamma)\sum_{i=0}^{\infty}(i+1)f(i+1)s^i \quad &= (2-\theta\gamma)(1-\gamma)\sum_{i=1}^{\infty}if(i)s^i\\ &-(1-\gamma)^2\sum_{i=2}^{\infty}(i-1)f(i-1)s^i+r\theta\gamma(1-\gamma)\sum_{i=0}^{\infty}f(i)s^i.\end{aligned}$$

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

3 Method of moments

3.1 The Procedure

Let $X_1,X_2,\ldots,$ are iid r.v.'s, 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 [6]):

1) If the model has m parameters, we compute m moments, $\mu_1,\mu_2,\ldots,\mu_m$ 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 $\mu_i,\ i=1,2,\ldots,m$.

3) After that, based on the data $X=(X_1,X_2,\ldots,X_n)$, 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 Pólya distribution

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

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

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

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

by the method of moments for the parameters r, θ and γ of the compound Pólya distribution, we obtain

$$\widehat{r} = \frac{(\overline{X})^2}{\sqrt{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4}},$$

$$\widehat{\theta} = \frac{2\sqrt{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4}}{\overline{X^2}-(\overline{X})^2-\overline{X}+\sqrt{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4}},$$

$$\widehat{\gamma} = \frac{2\overline{X}}{\overline{X^2}-(\overline{X})^2+\overline{X}-\sqrt{2\overline{X}\overline{X^3}-3(\overline{X^2})^2+(\overline{X})^2+(\overline{X})^4}},$$

where $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1, 2, 3.$

4 Compound Pólya process

We consider the stochastic process $N(t),\ t > 0$ defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ and given by

$$N(t) = \begin{cases} X_1 + X_2 + \ldots + X_{N_1(t)}, & N_1(t) > 0, \\ 0, & N_1(t) = 0, \end{cases} \tag{15}$$

where $X_i,\ i = 1, 2, \ldots$ are iid as X r.v.'s, independent of $N_1(t)$. We suppose that the counting process $N_1(t)$ is a negative binomial process with parameters $r \in \mathbf{N}$ and $1-\pi \in (0, 1)$, where $\pi = \frac{\delta}{\delta+t},\ \delta > 0$. Denote by $N_1(t) \sim NB(r,\ \frac{t}{\delta+t})$. In this case $N(t)$ is a compound negative binomial process. The PMF and PGF of $N_1(t)$ are given by

$$P(N_1(t) = i) = \binom{r+i-1}{i}\left(\frac{\delta}{\delta+t}\right)^r\left(\frac{t}{\delta+t}\right)^i,\quad i = 0, 1, \ldots \tag{16}$$

and

$$\psi_{N_1(t)}(s) = \left(\frac{\delta}{\delta+t-ts}\right)^r,\ |s| < \frac{\delta+t}{t}. \tag{17}$$

We suppose that the compounding r.v. $X \sim Ge(\gamma)$ is given by PMF (2) and PGF (3). For the PGF of the process $N(t)$, given in (15) we get

$$\psi(s) = \psi_{N(t)}(s) = \left(\frac{\delta}{\delta + t - t\psi_1(s)}\right)^r, \tag{18}$$

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

Definition 4.1. *The stochastic process, defined by the PGF (18) and compounding distribution, given by (2) and (3) is called a compound Pólya process.*

4.1 The Probability Mass Function

The probability function of $N(t)$ is given by expanding the PGF $\psi(s)$ in powers of s. Denote by $P_i(t) = P(N(t) = i)$, $i = 0, 1, 2, \ldots$, the probability mass function of $N(t)$. We rewrite the PGF of (18) in the form

$$\begin{aligned} \psi(s) &= \left(\frac{\delta}{\delta + t}\right)^r \sum_{m=0}^{\infty} \binom{r+m-1}{m} \left(\frac{t}{\delta + t}\psi_1(s)\right)^m \\ &= \left(\frac{\delta}{\delta + t}\right)^r \sum_{m=0}^{\infty} \binom{r+m-1}{m} \left(\frac{\gamma t}{(\delta + t)(1-(1-\gamma)s)}\right)^m. \end{aligned} \tag{19}$$

Denote by $\psi^{(i)}(s) = \dfrac{\partial^{(i)}\psi(s)}{\partial s^i}$, for $i = 0, 1, \ldots$, the derivatives of $\psi(s)$. From (19) we get the following

$$\psi^{(i)}(s) = (1-\gamma)^i \left(\frac{\delta}{\delta + t}\right)^r \sum_{m=1}^{\infty} \binom{r+m-1}{m} \left(\frac{\gamma t}{\delta + t}\right)^m \frac{m(m+1)\ldots(m+i-1)}{(1-(1-\gamma)s)^{m+i}}. \tag{20}$$

From [2], it is known that

$$P_i(t) = \frac{\psi^{(i)}(s)}{i!}\Big|_{s=0}. \tag{21}$$

The result is given in the next theorem.

Theorem 4.1. *The probability mass function of $N(t) \sim$compound Pólya process is given by*

$$\begin{aligned} P_0(t) &= \left(\frac{\delta}{\delta + t - t\gamma}\right)^r, \\ P_i(t) &= (1-\gamma)^i \left(\frac{\delta}{\delta + t}\right)^r \sum_{m=1}^{\infty} \binom{r+m-1}{m} \binom{m+i-1}{i} \left(\frac{\gamma t}{\delta + t}\right)^m, \quad i = 0, 1, \ldots \end{aligned} \tag{22}$$

Proof. The initial value $P_0(t) = \left(\frac{\delta}{\delta+t-t\gamma}\right)^r$ follows from the PGF in (18), $\psi(0) = P_0(t)$. Then (22) follows from (20) and (21). □

Concluding remarks

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

Acknowledgements

This work was supported by grant RD-08-69/2016 of Shumen University, Bulgaria.

References

1. **Grandell, J.** Mixed Poisson Process. Chapman & Hall, London, 1997.
2. **Johnson, N.L., A. W. Kemp, S. Kotz.** Univariate Discrete Distributions. Third edition, John Wiley & Sons, Hoboken, New Jersey, 2005.
3. **Minkova, L.D.** Inflated-parameter modification of the pure birth process. Compt. Randue Bulg. Acad. Sci., vol. 54(11), 17–22, 2001.
4. **Minkova, L.D.** I-Pólya Process and Applications. Commun. Statist.-Theory and Methods, vol. 40, 2847–2855, 2011.
5. **Panjer, H.** Recursive evaluation of a family of compound discrete distributions. ASTIN Bull., vol. 12(1), 22–26, 1981.
6. **Rice, J.A.** Mathematical Statistics and Data Analysis. Duxbury Press, Belmont, CA, third edition, 2007.