Type II non-central Pólya-Aeppli distribution

Cite as: AIP Conference Proceedings **2172**, 100009 (2019); https://doi.org/10.1063/1.5133602
Published Online: 13 November 2019

Meglena Lazarova, and Krasimira Kostadinova

Type II Non-central Pólya-Aeppli Distribution

Meglena Lazarova[1,a)] and Krasimira Kostadinova[2,b)]

[1] *Faculty of Applied Mathematics and Informatics,*
Technical University of Sofia, 1756 Sofia, Bulgaria
[2] *Faculty of Mathematics and Informatics,*
Konstantin Preslavsky University of Shumen, 9700 Shumen, Bulgaria

[a)] meglena.laz@tu-sofia.bg
[b)] kostadinova@shu.bg

Abstract. In this paper the Type II Non-central Pólya-Aeppli distribution is introduced. It is a sum of two independent variables, one that is an Inflated-parameter binomial distributed and another one, a Pólya-Aeppli distributed. The probability mass function, moments, recursion formulas and some properties of the defined distribution are derived.

INTRODUCTION

A huge activity of generalizing of the classical discrete distributions have been achieved during the last decade. The necessity of this process arises to give an application of the extended versions for modelling different kinds of dependent count or frequency data structure which can be used in the field of Finance, Biometrics, Insurance, Econometrics, etc., see Bowers et al. [1], Johnson et al., [2] and Rolski et al., [3]. This leads to more flexible distributions which can be used as a base for improved count data models with immediate application in the insurance. Such distributions are used for describing the number of the arrival claims to the insurance company. There are at least four specific distributions in the literature that are applicable in insurance as a claim number distributions: the Poisson, negative binomial, binomial and the logarithmic series distribution, see [2]. Actually the logarithmic series distribution is used rarely and it can be obtained as a limiting distribution of the truncated at zero negative binomial distribution. These distributions are members of the family of Power Series Distributions (PSDs) and this family has a probability generating function (PGF) given by

$$\psi(s) = \frac{g(\theta s)}{g(\theta)}, \tag{1}$$

where $g(\theta)$ is a positive, finite and differentiable series function and $\theta > 0$ is a series parameter. The corresponding random variable is defined on the non-negative integers. If the random variable N is defined on any subset of the non-negative integers and its PGF is given by (1) then N belongs to the family of Generalized Power Series Distributions (GPSDs) or to the classes of their generalizations, see [2], [4], [5]. In [6] an extension of the classical discrete distributions by including an additional parameter $\rho \in [0, 1)$ is introduced. The new probability distributions are useful for modeling heterogeneity and they are known as the Inflated-parameter Power Series distributions (IPSDs) and Inflated-parameter Generalized Power Series distributions (IGPSDs). A random variable N belongs to the family of IPSDs if it is defined on the non-negative integers and its PGF has the form

$$\psi_N(s) = \frac{g(\theta\psi_1(s))}{g(\theta)}, \tag{2}$$

where the series function $g(\theta)$ is a positive, finite and differentiable and

$$\psi_1(s) = \frac{(1-\rho)s}{1-\rho s} \tag{3}$$

Proceedings of the 45th International Conference on Application of Mathematics in Engineering and Economics (AMEE'19)
AIP Conf. Proc. 2172, 100009-1–100009-6; https://doi.org/10.1063/1.5133602
Published by AIP Publishing. 978-0-7354-1919-3/$30.00

is the PGF of the shifted geometric distribution with success probability $1-\rho$, $Ge_1(1-\rho)$. The random variable N belongs to the family of IGPSDs if it is defined on any subset of the non-negative integers and it has a PGF given by (2). The Inflated-parameter binomial distribution, Inflated-parameter Poisson distribution, Inflated-parameter negative binomial distribution and Inflated-parameter logarithmic series distribution belong to the family of IGPSDs, see [7]. The Pólya-Aeppli distribution was derived by Anscombe in 1950, see [8] from a model of randomly distributed colonies. Anscombe states that in 1930 the distribution was given by A. Aeppli in a thesis and in 1930 developed by G. Pólya. For this reason he called it a Pólya-Aeppli distribution. It is a compound Poisson with geometric compounding distribution. The probability mass function (PMF) and the PGF of the compounding distribution are given by

$$P(X=m) = (1-\rho)\rho^{m-1}, \qquad m = 1, 2, \dots$$

and (3).

In the present paper we introduce and analyze a Type II Non-central Pólya-Aeppli distribution as a sum of independent Inflated-parameter binomial distribution and a Pólya-Aeppli distribution. This is a counting distribution that can be applied as a corresponding distribution of a counting process in risk theory. The paper is organized as follows. In the next sections we consider the probability generating function, the probability mass function and the moments of the Type II Non-central Pólya-Aeppli distribution. For specific values of the parameters we give some graphics and make some conclusions.

PROBABILITY GENERATING FUNCTION

The Type II Non-central Pólya-Aeppli distribution is a sum of independent Inflated-parameter binomial distribution and a Pólya-Aeppli distribution. Suppose that the first random variable N_1 has an Inflated-parameter binomial distribution, i.e.

$$P(N_1 = m) = \begin{cases} (1-\pi)^n, & m = 0, \\ \displaystyle\sum_{i=1}^{min(m,n)} \binom{n}{i}\binom{m-1}{i-1}[\pi(1-\rho)]^i(1-\pi)^{n-i}\rho^{m-i}, & m = 1, 2, \dots, \end{cases}$$

where $\pi \in (0,1)$, $\rho \in [0,1)$ and $n \in \{1, 2, \dots\}$. We use the notation $N_1 \sim IBi(\pi, \rho, n)$.

Remark 1. In the case of $n = 1$, the $IBi(\pi, \rho, n)$ distribution becomes the Inflated-parameter Bernoulli distribution with parameters π and ρ, see [6].

The PGF of the Inflated-parameter binomial distribution is given by

$$\psi_1(s) = \left(1 - \pi + \pi\frac{(1-\rho)s}{1-\rho s}\right)^n.$$

The mean and the variance of the Inflated-parameter binomial distribution are given by

$$E(N_1) = \frac{n\pi}{1-\rho} \quad \text{and} \quad Var(N_1) = \frac{n\pi(1-\pi+\rho)}{(1-\rho)^2}.$$

The related Fisher index of dispersion is

$$FI(N_1) = \frac{1-\pi+\rho}{1-\rho}.$$

The second random variable N_2 with given parameters $\lambda > 0$ and $\rho \in [0,1)$ has a Pólya-Aeppli distribution with PMF given by

$$P(N_2 = m) = \begin{cases} e^{-\lambda}, & m = 0, \\ e^{-\lambda}\displaystyle\sum_{j=1}^{m}\binom{m-1}{j-1}\frac{[\lambda(1-\rho)]^j}{j!}\rho^{m-j}, & m = 1, 2, \dots \end{cases} \tag{4}$$

We use the notation $N_2 \sim PA(\lambda, \rho)$. The distribution given in formula (4) is also called an Inflated-parameter Poisson distribution, see [2].

Remark 2. In the case of $\rho = 0$ the $PA(\lambda, \rho)$ distribution coincudes with the Poisson distribution, see [2].

The mean and the variance of the Pólya-Aeppli distribution are given by

$$E(N_2) = \frac{\lambda}{1-\rho} \quad \text{and} \quad Var(N_2) = \frac{\lambda(1+\rho)}{(1-\rho)^2}.$$

The related Fisher index of dispersion is

$$FI(N_2) = \frac{1+\rho}{1-\rho}.$$

As the Fisher index for the Poisson distribution is equal to one it is seen that for $\rho \neq 0$, the Pólya-Aeppli distribution is over-dispersed related to the Poisson distribution.

The PGF of the Pólya-Aeppli distribution with parameters $\lambda > 0$ and $\rho \in [0, 1)$ is given by

$$\psi_{N_2}(s \mid \lambda) = e^{-\lambda\left[1-\psi_1(s)\right]},$$

where $\psi_1(s)$ is the PGF of the shifted geometric distribution, given in (3).

Considering that $N = N_1 + N_2$, where N_1 and N_2 are independent random variables we obtain that the PGF of the Type II Non-central Pólya-Aeppli distributed random variable N is given by

$$\psi_N(s) = \left(1 - \pi + \frac{\pi(1-\rho)s}{1-\rho s}\right)^n e^{-\lambda(1-\psi_1(s))}, \tag{5}$$

where $\psi_1(s)$ is given in (3). This leads to the next definition.

Definition 1. *The random variable N with PGF given in (5) is referred to a Type II Non-central Pólya-Aeppli distribution. We use the notation* $N \sim NPAD_{TypeII}(\pi, n, \lambda, \rho)$.

PROBABILITY MASS FUNCTION

Lemma 1. The probability mass function of the Type II Non-central Pólya-Aeppli distribution is given by

$$p_m = \begin{cases} e^{-\lambda} \sum_{l=0}^{m} \sum_{j=0}^{l} \frac{\lambda^j (1-\rho)^j \rho^{l-j}}{j!} \binom{n}{n-m+l} \binom{n+l-1}{l-j} (1-\pi)^{n-m+l} (\pi-\rho)^{m-l}, & m = 0, 1, \ldots, n, \\ e^{-\lambda} \sum_{l=m-n}^{m} \sum_{j=0}^{l} \frac{\lambda^j (1-\rho)^j \rho^{l-j}}{j!} \binom{n}{n-m+l} \binom{n+l-1}{l-j} (1-\pi)^{n-m+l} (\pi-\rho)^{m-l}, & m = n+1, n+2, \ldots \end{cases} \tag{6}$$

Proof. Representing the PGF (5) in the form

$$\psi_N(s) = \frac{[1 - \pi + (\pi - \rho)s]^n}{(1-\rho s)^n} e^{-\lambda} e^{\frac{\lambda(1-\rho)s}{1-\rho s}}$$

and after some transformations we obtain the following form of the PGF

$$\psi_N(s) = \sum_{i=0}^{n} \binom{n}{i} (1-\pi)^i (\pi-\rho)^{n-i} s^{n-i} e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j (1-\rho)^j s^j}{j!} \sum_{k=0}^{\infty} \binom{n+j+k-1}{k} (\rho s)^k,$$

where $\psi_N(s) = \sum_{m=0}^{\infty} p_m s^m$. The PMF given in formula (6) is obtained after several substitutions applied over the equation above and equating the coefficients in front of s^m on the both sides for fixed values $m = 0, 1, \ldots, n$ and $m = n+1, n+2, \ldots$

For specific values of the distributions' parameters $\pi,\ n,\ \lambda$ and ρ we construct some useful graphics for the PMF of the Type II Non-central Pólya-Aeppli distribution. Initially in the first three graphics, given in Figure 1 we change the values of the parameter ρ and fix the values of the other three parameters. We are interested in the sensitivity of the distribution influenced by the change of the parameter ρ. It is seen that by changing the parameter $\rho = 0.25,\ \rho = 0.5$ and $\rho = 0.75$ the distribution forms a longer right tail.

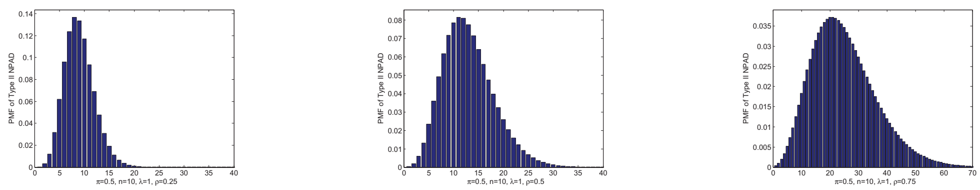

FIGURE 1. Probability mass function of Type II NPAD with fixed parameters $\pi = 0.5,\ n = 10,\ \lambda = 1$

In Figure 2 we introduce another three graphics for the PMF of Type II NPAD. Here we fix the parameters $n,\ \lambda,\ \rho$ and change the values of the parameter π. Thus we are interested in the distribution's sensitivity provoked by the changing of the parameter $\pi = 0.25,\ \pi = 0.6.$ and $\pi = 0.75$. It is seen that the changing of the parameter π leads to a symmetrical behaviour of the PMF.

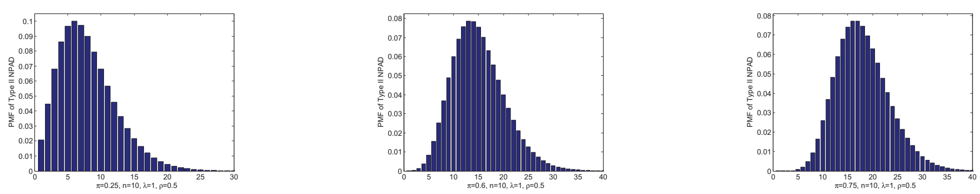

FIGURE 2. Probability mass function of Type II NPAD with fixed parameters $n = 10,\ \lambda = 1,\ \rho = 0.5$

In Figure 3 we fix the parameters $\pi,\ n,\ \lambda$ and change the values of the parameter ρ. On the left graphic of Figure 3 we take $\rho = 0$ and thus we obtain the PMF of the sum of two independent Poisson and Inflated-parameter binomial distributions. On the right graphic we take the value $\rho = 0.6$. For a larger value of the parameter ρ the distribution becomes a right tailed distribution.

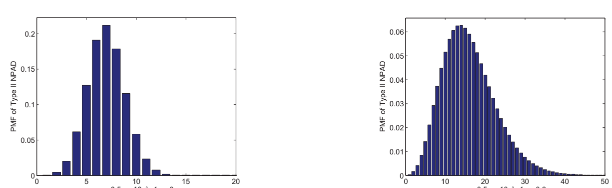

FIGURE 3. Probability mass function of Type II NPAD with fixed parameters $\pi = 0.5,\ n = 10,\ \lambda = 1$

Proposition 1 The probability mass function of the Type II Non-central Pólya-Aeppli distribution satisfies the following recursions

$$\begin{aligned}(1-\pi)mp_m &= [(3\rho-\pi-2\rho\pi)(m-1)+(1-\rho)(\lambda(1-\pi)+n\pi)]p_{m-1}\\ &+[(1-\rho)(\lambda(\pi-\rho)-n\pi\rho)-(3\rho^2-2\rho\pi-\rho^2\pi)(m-2)]p_{m-2}\\ &+\rho^2(\rho-\pi)(m-3)p_{m-3},\quad m=3,4,\dots,\end{aligned}\tag{7}$$

where $p_1=[\lambda(1-\rho)+\dfrac{n\pi(1-\rho)}{1-\pi}]p_0$ and $p_0=(1-\pi)^n e^{-\lambda}$.

Proof. Upon substituting $s=0$ in the PGF given in (5) we obtain the initial value p_0. The differentiation in (5) leads to

$$\psi_N'(s)=\psi_N(s)\frac{(1-\rho)[\lambda(1-\pi)+n\pi+(\lambda(\pi-\rho)-n\pi\rho)s]}{(1-\rho s)^2(1-\pi-(\rho-\pi)s)},\tag{8}$$

where $\psi_N(s)=\sum_{i=0}^{\infty}p_i s^i,\ \psi_N'(s)=\sum_{i=0}^{\infty}(i+1)p_{i+1}s^i$.

After several transformations over the equation (8) we obtain the following equation

$$\begin{aligned}(1-\pi)\sum_{i=0}^{\infty}(i+1)p_{i+1}s^i &= (3\rho-\pi-2\rho\pi)\sum_{i=1}^{\infty}ip_i s^i-(3\rho^2-2\rho\pi-\rho^2\pi)\sum_{i=2}^{\infty}(i-1)p_{i-1}s^i\\ &+\rho^2(\rho-\pi)\sum_{i=3}^{\infty}(i-2)p_{i-2}s^i+(1-\rho)(\lambda(1-\pi)+n\pi)\sum_{i=0}^{\infty}p_i s^i\\ &+(1-\rho)(\lambda(\pi-\rho)-n\pi\rho)\sum_{i=1}^{\infty}p_{i-1}s^i.\end{aligned}\tag{9}$$

The recursions are obtained by equating the coefficients in front of s^i on the both sides of the equation (9) for fixed values of $i=1,2,\dots$ and after substituting $m=i+1$.

MOMENTS

The mean and the variance of the Type II Non-central Pólya-Aeppli distribution are given by

$$E(N)=\frac{\lambda+n\pi}{1-\rho},$$

$$Var(N)=\frac{(\lambda+n\pi)(1+\rho)-n\pi^2}{(1-\rho)^2}.$$

For the Fisher index of dispersion we obtain

$$FI=\frac{Var(N)}{E(N)}=\frac{1+\rho}{1-\rho}-\frac{n\pi^2}{(1-\rho)(\lambda+n\pi)}<\frac{1+\rho}{1-\rho},$$

i.e. the Type II Non-central Pólya-Aeppli distribution is under-dispersed related to the Pólya-Aeppli distribution.

CONCLUDING REMARKS

In the present paper we have analysed the Type II Non-central Pólya-Aeppli distribution with its PGF, PMF, recursion formulas and moments. We give a connection to some well known distributions like Poisson distribution, Pólya-Aeppli distribution and Inflated-parameter binomial distribution. Some useful graphics about the distribution and its parameters are presented and analyzed.

ACKNOWLEDGMENTS

The research of the first author is supported by a project ”Stochastic and simulation models in the field of medicine, social sciences and dynamic systems” funded by the National Science Fund of Ministry of Education and Science of Bulgaria (Contract No. DN 12/11/20 Dec. 2017). The second author is supported by the Bulgarian Ministry of Education and Science under the National Program for Research ”Young Scientists and Postdoctoral Students”. A special gratitude to Leda Minkova for helpful suggestions, stimulating discussions and comments. The authors are also grateful to the anonymous reviewer for his careful reading of the manuscript, critical remarks and suggested improvements.

REFERENCES

[1] N.L.Bowers, H.U.Gerber, J.C.Hickman, D.A.Jones, and C.J.Nesbitt, *Actuarial Mathematics, 2nd edition* (Society of Actuaries, Shaumburg, 1997).
[2] N.L.Johnson, A.W.Kemp, and S.Kotz, *Univariate Discrete Distributions, 3rd edition* (John Wiley & Sons, Hoboken, New Jersey, 2005).
[3] T.Rolski, H.Shmidli, V.Schmidt, and J.Teugels, *Stochastic Processes for Insurance and Finance* (John Wiley & Sons, Chichester, 1999).
[4] R.C.Gupta, Sankhyã: Series B **35**, 288–298 (1974).
[5] P. Consul, Communications in Statistics: Theory and methods **19**, 4653–4666 (1990).
[6] L. Minkova, Communications in Statistics - Theory and Methods **31**, 871–888 (2002).
[7] N.Kolev, L. Minkova, and P.Neytchev, ARCH Research Clearing House **2**, 295–320 (2000).
[8] F.Anscombe, Biometrica **37**, 358–382 (1950).