A New Flexible Discrete Distribution with Applications to Count Data

Anwar Hassan[1,*], Gulzar Ahmad Shalbaf[1], Sheikh Bilal[2], Adil Rashid[2]

[1]*P.G. Department of Statistics, University of Kashmir, Srinagar*
[2]*Goverenment Degree College Ganderbal*

ARTICLE INFO

ABSTRACT

In this paper we have explored a new discrete probability mass function that has been generated through compounding mechanism. This newly proposed probability mass function is essentially a mixture of Poisson and Ailamujia distribution. Furthermore the parameter estimation has also been discussed by using Maximum likelihood estimation (MLE) technique. Moreover, we have also studied some important properties of the proposed model that include factorial moments, raw moments, mean, variance, and coefficient of variation. In the end, the application and potentiality of the proposed model have been tested statistically and it has been shown that the proposed model can be employed to model a real life data set to get an adequate fit that has also corroborated through graphically.

1. INTRODUCTION

In probability distribution theory there are vast numbers of discrete distributions that have specified applications but at times the observable data have distinct features that are not exhibited by these classical discrete distributions. So to overcome these limitations researchers often develop new probability distributions so that these new distributions can be employed in those situations where classical distributions are not providing any adequate fit. There are so many techniques by which we can obtain new distributions such as transmutation, compounding, etc. but compounding of distributions has received maximum attention from the last few decades due to its simplified approach. Some new compound distributions were obtained in the recent past by the numerous researchers for instance, Adil *et al.* [1–4] and Adil and Jan [5–8] obtained a new lifetime distributions to address issues related to count data and lifetime of series system and parallel systems.

2. MATERIALS AND METHODS

A random variable X has a Poisson distribution if for $\lambda > 0$, its probability mass function is

$$p_1(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ where \ x = 0, 1, 2, \ldots \tag{1}$$

The factorial moment of the Poisson distribution and hence mean and variance are given by

$$\mu_{[r]} = \lambda^r \ for \ r = 1, 2, 3, \ldots \tag{2}$$

$$E(X) = \lambda \quad and \quad V(X) = \lambda$$

Also, the probability density function associated with Ailamujia random variate is

$$f_1(X = x, \alpha) = 4x\alpha^2 e^{-2\alpha x}, \ x \geq 0, \alpha > 0. \tag{3}$$

---

*Corresponding author. Email: anwar.hassan2007@gmail.com

The raw moments of Ailamujia distribution (AD) are given by

$$E(X^r) = \int_0^\infty x^r f_1(X = x, \alpha) dx$$

$$\mu'_r = \frac{(r+1)!}{(2\alpha)^r} \tag{4}$$

AD is formulated by Lv *et al.* [9] for several engineering application. This distribution was also studied by Pan *et al.* [10] for interval estimation and hypothesis testing based on sample of small size. The Bayesian estimation of AD was obtained by Long [11] under type ii censoring using the three different priors based on missing data. The minimax estimation of the parameter of Ailamujia model was evaluated by Li [12].

Usually the parameter λ in Poisson distribution is fixed constant but here we will consider a problem in which the parameter λ is itself a random variable following AD with probability density function (3).

3. DEFINITION OF PROPOSED MODEL

If $X|\lambda$ follows $P(\lambda)$ where λ is itself a random variable following AD then determining the distribution that results from marginalizing over λ will be known as a compound of Poisson distribution with that of AD which is called Poisson Ailamujia distribution and is denoted by $PAD(\lambda, \alpha)$. It may be noted here that proposed model will be a discrete since the parent distribution is discrete.

3.1. Theorem

The probability mass function of a compound of $P(\lambda)$ with $AD(\alpha)$

is given by

$$f_{PAD}(X = x, \alpha) = \frac{4\alpha^2(1 + x)}{(1 + 2\alpha)^{x+2}}, \quad x = 0, 1, 2, ..., \alpha > 0.$$

Proof: Using the definition (3), the probability mass function of a compound of $P(\lambda)$ with $AD(\alpha)$ can be obtained as

$$f_{PAD}(X = \lambda, \alpha) = \int_0^\infty p_1(x|\lambda) f_1(x, \lambda) d\lambda$$

$$f_{PAD}(X = \lambda, \alpha) = \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} 4\lambda \alpha^2 e^{-2\alpha\lambda} d\lambda$$

$$f_{PAD}(X = \lambda, \alpha) = \frac{4\alpha^2}{x!} \int_0^\infty e^{-(1+2\alpha)\lambda} \lambda^{(x+2)-1} d\lambda$$

$$f_{PAD}(X = \lambda, \alpha) = \frac{4\alpha^2(1 + x)}{(1 + 2\alpha)^{x+2}}, \quad x = 0, 1, 2, ..., \alpha > 0.$$

$$f_{PAD}(X = x, \alpha) = \frac{4\alpha^2(1 + x)}{(1 + 2\alpha)^{x+2}}, \quad x = 0, 1, 2, ..., \alpha > 0. \tag{5}$$
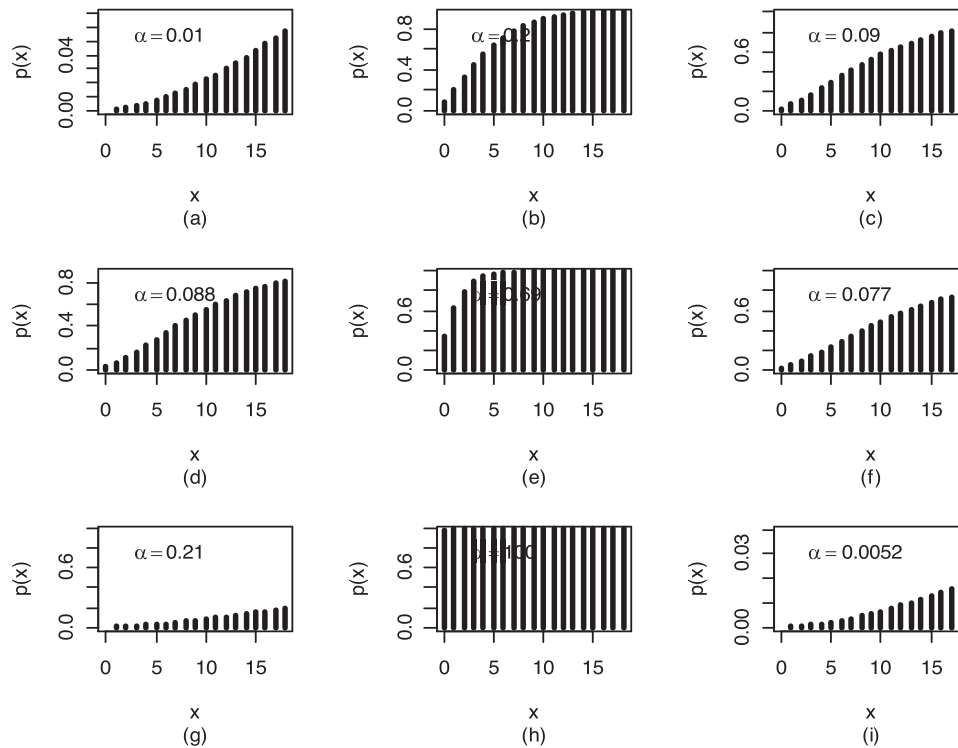
3.2. Cumulative Distribution Function of the Proposed Model

$$F_X(x) = \sum_{i = x_i \le x} P(X = x_i)$$

$$= \sum_{x=0}^{x} \frac{4\alpha^2(x+1)}{(1 + 2\alpha)^{x+2}}$$

$$= \frac{4\alpha^2}{(1 + 2\alpha)^2} \sum_{x=0}^{x} \frac{(x+1)}{(1 + 2\alpha)^x}$$

$$F_X(x) = 1 - \frac{(4\alpha + 2\alpha x + 1)}{(1 + 2\alpha)^{x+2}}$$

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

4. FACTORIAL MOMENTS OF PROPOSED MODEL

We shall obtain factorial moments of the proposed model which is very useful to study some of the most important properties such as mean, variance, standard deviation, etc.

4.1. Theorem

The factorial moments of order "r" of the proposed model is given by

$$\mu_{[r]}(X) = \frac{(r+1)!}{(2\alpha)^r}, \ \ X = 0, 1, 2, ..., \alpha > 0.$$

Proof: The factorial moment of order r of Poisson distribution is

$$\mu_{[r]} = \lambda^r \ \ for \ \ r = 1, 2, ...$$

Since λ itself is a random variable following AD, therefore one can obtain factorial moment of the proposed model by using the definition

$$\mu_{[r]}(x) = E_\lambda \left[m_r \left(X|\lambda \right) \right]$$

Where $m_r \left(X|\lambda \right)$ are factorial moments of order "r" of a Poisson distribution.

We know that

$$m_r \left(X|\lambda \right) = \lambda^r$$

Therefore

$$\mu_{[r]}(x) = E[\lambda^r]$$

$$= \int_0^\infty \lambda^r \, 4\lambda\alpha^2 e^{-2\alpha\lambda} d\lambda$$

$$= 4\alpha^2 \int_0^\infty \lambda^{(r+2)-1} e^{-2\alpha\lambda} d\lambda$$

$$\mu_{[r]}(X) = \frac{(r+1)!}{(2\alpha)^r}, \quad X = 0, 1, 2, ..., \alpha > 0.$$

For r = 1, we get the mean of $PAD(\alpha)$

$$\mu_{[1]}(x) = \frac{4\alpha^2(1+1)!}{(2\alpha)^{1+2}}$$

$$\mu_{[1]}(x) = \frac{4\alpha^2(2)!}{(2\alpha)^3}$$

$$mean = \mu_{[1]}(x) = \frac{1}{\alpha}$$

$$\mu_{[2]}(x) = \frac{4\alpha^2(2+1)!}{(2\alpha)^{2+2}}$$

$$\mu_{[2]}(x) = \frac{3}{2\alpha^2}$$

$$\mu_{[2]}(x) = \mu_2' - \mu_1'$$

$\mu_2' = \dfrac{3 + 2\alpha}{2\alpha^2}$ And hence Variance $= \dfrac{2\alpha + 1}{2\alpha^2}$ SD $= \dfrac{\sqrt{2\alpha + 1}}{\alpha\sqrt{2}} = \dfrac{1}{\alpha}\sqrt{(2\alpha + 1)/2}$

Moreover the CV $= \sqrt{\dfrac{2\alpha + 1}{2}} \times 100$

5. PARAMETER ESTIMATION

The estimation of a parameter of $PAD(\alpha)$ model will be discussed here in this section through maximum likelihood estimation.

5.1. Maximum Likelihood Estimation

The estimation of a parameter of $PAD(\alpha)$ model via maximum likelihood estimation method requires the log likelihood function of $PAD(\alpha)$ as

Let $X_1, X_2, ..., X_n$ are i.i.d $PAD(\alpha)$. Then

$$L = \prod_{i=1}^{n} P(X = x, \alpha)$$

$$L = \prod_{i=1}^{n} \frac{4\alpha^2(x_i + 1)}{(1 + 2\alpha)^{x_i + 2}}$$

$$L = \frac{\left(4\alpha^2\right)^n \prod_{i=1}^{n}(x_i + 1)}{(1 + 2\alpha)^{\sum x_i + 2n}}$$

$$\log L = n \log 4 + 2n \log \alpha + \sum_{i=1}^{n} \log(x_i + 1) - \left(\sum_{i=1}^{n} x_i + 2n\right) \log(1 + 2\alpha).$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{2n}{\alpha} - \frac{2\left(\sum_{i=1}^{n} x_i + 2n\right)}{(1 + 2\alpha)}$$

For estimating the unknown parameter we set the first order partial derivative equal to zero

$$\frac{\partial \log L}{\partial \alpha} = \frac{2n}{\alpha} - \frac{2\left(\sum_{i=1}^{n} x_i + 2n\right)}{(1 + 2\alpha)} = 0$$

Hence we get that $\alpha = \dfrac{1}{\bar{x}}$

6. APPLICATION IN BIOLOGICAL SCIENCE

Genetics is the biological science which deals with the mechanisms responsible for similarities and differences among closely related species. The term "genetic" is derived from the Greek word "genesis" meaning grow into or to become. So, genetic is the study of heredity and hereditary variations. It is the study of transmission of body features that is similarities and difference, from parents to offspring's and the laws related to this transmission, any difference between individual organisms or groups of organisms of any species, caused either by genetic difference or by the effect of environmental factors, is called variation. Variation can be shown in physical appearance, metabolism, behavior, learning and mental ability, and other obvious characters. In this section the potentiality of proposed model will be justified by fitting it to the reported genetics count data sets of Catcheside *et al.* (Tables 1 and 2).

**Table 1** | Distribution of number of Chromatid aberrations (0.2 g chinon 1, 24 hours).

		Fitted Distribution	
Number of Aberrations	Observed Frequency	Poisson	PAD
0	268	231.3	246.1
1	87	126.1	106.1
2	26	34.7	34.3
3	9	6.3	9.8
4	4	0.8	2.6
5	2	0.1	0.6
6	1	0.1	0.1
7+	3	0.1	0.04
Total	400	400	400
Chi-square estimate		40.8	10.05
Parameter estimation		$\hat{\theta} = 0.55$	$\hat{\alpha} = 1.82$
p-Value		<.0001	.018

PAD, Poisson Ailamujia Distribution.

Table 2 | Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by Streptonigrin (NSC-45383) Exposure -70 µg/kg.

		Fitted Distribution	
Class/Exposure (µg/kg)	Observed Frequency	Poisson	PAD
0	200	172.5	184.6
1	57	95.4	79.5
2	30	26.4	25.7
3	7	7.3	4.9
4	4	1.9	0.7
5	0	.5	0.1
6	2	0.1	0.0
Total	300	300	300
Parameter estimation		$\hat{\theta} = 0.55$	$\hat{\alpha} = 1.82$
Chi-square estimate		29.6	9.4
p-Value		<.0001	.09

PAD, Poisson Ailamujia Distribution.

Fitting of ALPD model ,Table 1
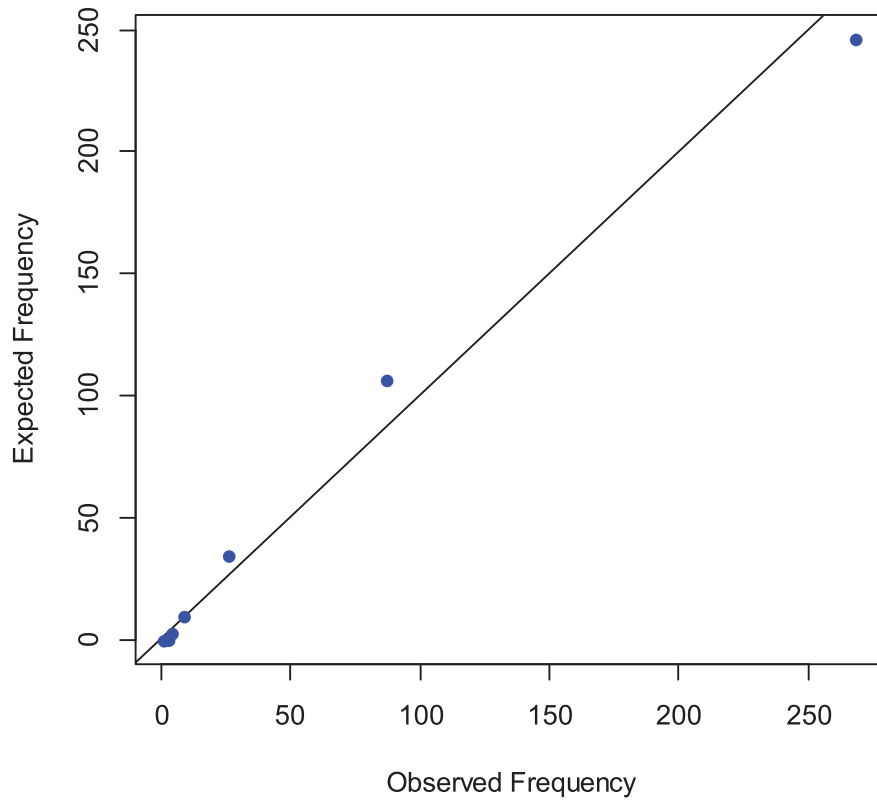
Fitting of ALPD model ,Table 2
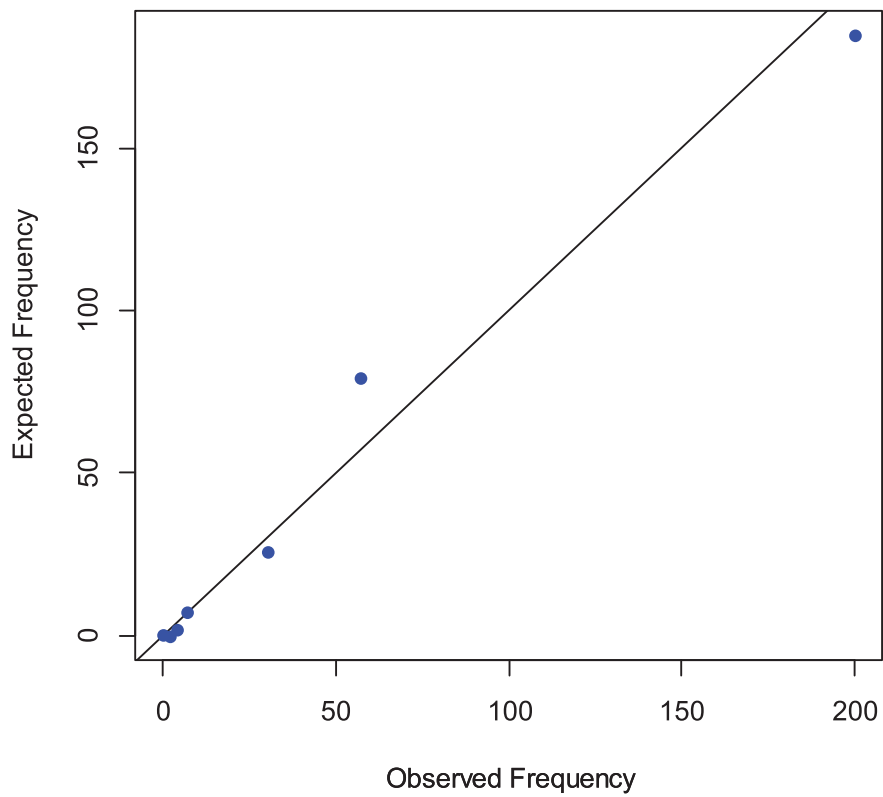
7. CONCLUSIONS

In this paper we have proposed a new model by compounding Poison distribution with Ailamujia Distribution (AD) and it has been shown that proposed model can be nested to different compound distributions. Furthermore, we have derived several properties of proposed model such as factorial moments, mean, variance. In addition to this parameter estimation of the proposed model has been discussed by means of method of moments and MLE. Finally, the application of the proposed model have been explored in genetics, based on the goodness of fit test and it has been shown that our model offers a better fit as compared to classical Poisson distribution. We hope that the proposed model will serve as an alternative to various models available in literature.

REFERENCES

1. R. Adil, A. Zahoor, T.R. Jan, J. Mod. Appl. Stat. Methods. 7 (2018), 1–25.
2. R. Adil, A. Zahoor, T.R. Jan, Pak. J. Stat. Oper. Res. 10 (2018), 139–155.
3. R. Adil, A. Zahoor, T.R. Jan, J. Appl. Math. Inf. Sci. Lett. Stat. Anal. 6 (2018), 113–121.
4. R. Adil, A. Zahoor, T.R. Jan, J. Stat. Appl. Probab. 3 (2014), 451.
5. R. Adil, T.R. Jan, J. Mod. Appl. Stat. Methods. 9 (2014), 17–24.
6. R. Adil, T.R. Jan, J. Mod. Appl. Stat. Methods. 13 (2014), 18.
7. R. Adil, T.R. Jan, Int. J. Math. Stat. 17 (2016), 23–38.
8. R. Adil, T.R. Jan, J. Reliab. Stat. Stud. 6 (2013), 11–19.
9. H.Q. Lv, L.H. Gao, C.L. Chen, J. Acad. Armed Force Eng. 16 (2002), 48–52.
10. G.T. Pan, B.H. Wang, C.L. Chen, Y.B. Hang, M.T. Dang, Appl. Stat. Manag. 28 (2009), 468–472.
11. B. Long, Math. Pract. Theor. (2015), 186–192.
12. L.P. Li, Sci. J. Appl. Math. Stat. 4 (2016), 229–233.