MedCrave
Step into the World of Research
Biometrics & Biostatistics International Journal

Research Article
Open Access
CrossMark

The Poisson-Adya distribution

Abstract

In this paper a Poisson mixture of Adya distribution called Poisson-Adya distribution has been suggested. The expressions of statistical constants including coefficients of variation, skewness, kurtosis and index of dispersion have been obtained and their behavior for varying values of parameter has been studied. It is observed that the obtained distribution is unimodal, has increasing hazard rate and over-dispersed. Maximum likelihood estimation and method of moment have been discussed for estimating parameter. Finally, the goodness of fit of the proposed distribution and its comparison with Poisson and Poisson-Lindley distributions has been given.

Keywords: Adya distribution, compounding, unimodality, over-dispersion, estimation, goodness of fit

Volume 11 Issue 3 - 2022

Rama Shanker,[1] Kamlesh Kumar Shukla[2]
[1]Department of Statistics, Assam University, Silchar, Assam, India
[2]Department of Mathematics, Noida International University, Gautam Buddh Nagar, India

Correspondence: Rama Shanker, Department of Statistics, Assam University, Silchar, Assam, India, Tel 291 7412260; Email shankerrama2009@gmail.com

Received: July 20, 2022 | **Published:** August 17, 2022

Introduction

The Poisson distribution is a suitable distribution for data having equi-dispersion (mean equal to variance). But in real life situation, it has been observed that most of the datasets being stochastic in nature are either over-dispersed (variance greater than mean) or under-dispersed (variance less than mean). During recent decades an attempt has been made by different researchers to derive over-dispersed one parameter discrete distribution by compounding Poisson distribution with one parameter continuous lifetime distributions. A popular one parameter discrete distribution for over-dispersed (variance greater than the mean) is the Poisson-Lindley distribution (PLD) proposed by Sankaran[1]. PLD is a Poisson mixture of Lindley distribution introduced by Lindley[2]. Further, it has been observed that these one parameter discrete distributions are not suitable for some over-dispersed datasets from biological sciences due to their levels of over-dispersion. Shanker & Hagos[3] have detailed discussion on applications of PLD for data arising from biological sciences, as the data from biological sciences are, in general, over-dispersed. It has been observed by Shanker & Hagos[3] that in some biological science data PLD does not give better fit and hence there is a need for another over-dispersed discrete distribution is required.

Shanker, et al[4] proposed a one parameter continuous lifetime distribution named Adya distribution, defined by its probability density function (pdf) and cumulative density function (cdf) given by

$$f(x;\theta)=\frac{\theta^3}{\theta^4+2\theta^2+2}(\theta+x)^2 e^{-\theta x} \quad ;x>0,\ \theta>0 \tag{1.1}$$

$$F(x,\theta)=1-\left[1+\frac{\theta x\left(\theta x+2\theta^2+2\right)}{\theta^4+2\theta^2+2}\right]e^{-\theta x};x>0,\theta>0 \tag{1.2}$$

Shanker et al[4] derived Adya distribution as a convex combination of exponential $(2,\theta)$, gamma $(2,\theta)$ and gamma $(3,\theta)$ distributions with respective proportions $\frac{\theta^4}{\theta^4+2\theta^2+2}$, $\frac{2\theta^4}{\theta^4+2\theta^2+2}$ and $\frac{2\theta^4}{\theta^4+2\theta^2+2}$ respectively. Its various statistical properties including moments and moments-based measures, hazard rate function, mean residual life function, stochastic ordering, deviations from the mean and the median, Bonferroni and Lorenz curves, and stress-strength reliability, estimation of parameter and applications are available in Shanker et al[4].

In the present paper a Poisson mixture of Adya distribution has been derived and its statistical constants including coefficients of variation, skewess, kurtosis and index of dispersion have been studied. The Unimodality, increasing hazard rate and over-dispersion of the distribution have been explained. Estimation of parameter using method of moment and maximum likelihood has been discussed. Applications, goodness of fit and its comparison with other one parameter discrete distributions are presented.

Poisson-Adya distribution

Let X follows Poisson distribution with parameter $\lambda>0$ having pmf

$$P(X\mid\lambda)=\frac{e^{-\lambda}\lambda^x}{x!};x=0,1,2,...$$

Now suppose the parameter λ follows Adya distribution with parameter θ having pdf

$$f(\lambda\mid\theta)=\frac{\theta^3}{\theta^4+2\theta^2+2}(\theta+\lambda)^2 e^{-\theta\lambda};\lambda>0,\theta>0$$

Thus, the marginal pmf of X can be obtained as

$$P(X=x)=\int_0^\infty P(X\mid\lambda)f(\lambda\mid\theta)d\lambda=\int_0^\infty\frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^3}{\theta^4+2\theta^2+2}(\theta+\lambda)^2 e^{-\theta\lambda}d\lambda \tag{2.1}$$

$$=\frac{\theta^3}{\left(\theta^4+2\theta^2+2\right)x!}\int_0^\infty e^{-(\theta+1)\lambda}\lambda^x\left(\theta^2+2\theta\lambda+\lambda^2\right)d\lambda$$

$$=\frac{\theta^3}{\left(\theta^4+2\theta^2+2\right)}\frac{x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)}{(\theta+1)^{x+3}};x=0,1,2,...,\theta>0 \tag{2.2}$$

We name this distribution as Poisson-Adya distribution. In the subsequent sections it has been shown that the pmf of Poisson-Adya distribution (PAD) is unimodal, has increasing hazard rate and over-dispersed. The nature of the pmf of PAD for varying values of parameter has been shown in the following figure1. As the value of parameter increases, the distribution becomes positively skewed and also it is becoming more over-dispersed (Figure 1).

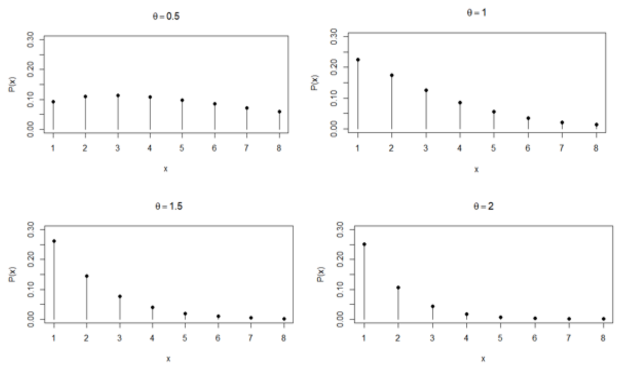

Figure 1 pmf of PAD for varying values of parameter

Statistical constants

Using (2.1), the r th factorial moment about origin, $\mu_{(r)}'$, of PAD can be obtained as

$$\mu_{(r)}' = E\left[E\left(X^{(r)}\,|\,\lambda\right)\right] = \frac{\theta^3}{\theta^4+2\theta^2+2}\int_0^\infty \left[\sum_{x=0}^\infty x^{(r)}\frac{e^{-\lambda}\lambda^x}{x!}\right](\theta+\lambda)^2 e^{-\theta\lambda}\,d\lambda$$

$$=\frac{\theta^3}{\theta^4+2\theta^2+2}\int_0^\infty \lambda^r\left[\sum_{x=r}^\infty \frac{e^{-\lambda}\lambda^{x-r}}{(x-r)!}\right]\left(\theta^2+2\theta\lambda+\lambda^2\right)e^{-\theta\lambda}\,d\lambda$$

$$=\frac{\theta^3}{\theta^4+2\theta^2+2}\int_0^\infty \lambda^r\left(\theta^2+2\theta\lambda+\lambda^2\right)e^{-\theta\lambda}\,d\lambda$$

$$=\frac{r!\left\{\theta^4+2(r+1)\theta+(r+1)(r+2)\right\}}{\theta^r\left(\theta^4+2\theta^2+2\right)};r=1,2,3,...$$

Substituting $r=1,2,3,\&4$ the first four factorial moment about origin of PAD can be obtained as

$$\mu_{(1)}' = \frac{\theta^4+4\theta^2+6}{\theta\left(\theta^4+2\theta^2+2\right)},\quad \mu_{(2)}' = \frac{2\left(\theta^4+6\theta^2+12\right)}{\theta^2\left(\theta^4+2\theta^2+2\right)}$$

$$\mu_{(3)}' = \frac{6\left(\theta^4+8\theta^2+20\right)}{\theta^3\left(\theta^4+2\theta^2+2\right)},\quad \mu_{(4)}' = \frac{24\left(\theta^4+10\theta^2+30\right)}{\theta^4\left(\theta^4+2\theta^2+2\right)}.$$

The relationship between moments about origin and factorial moments about origin gives the following four moments about origin

$$\mu_1' = \frac{\theta^4+4\theta^2+6}{\theta\left(\theta^4+2\theta^2+2\right)}$$

$$\mu_2' = \frac{\theta^5+2\theta^4+4\theta^3+12\theta^2+6\theta+24}{\theta^2\left(\theta^4+2\theta^2+2\right)}$$

$$\mu_3' = \frac{\theta^6+6\theta^5+10\theta^4+36\theta^3+54\theta^2+72\theta+120}{\theta^3\left(\theta^4+2\theta^2+2\right)}$$

$$\mu_4' = \frac{\theta^7+14\theta^6+40\theta^5+108\theta^4+294\theta^3+408\theta^2+720\theta+720}{\theta^4\left(\theta^4+2\theta^2+2\right)}.$$

Using the relationship between moments about mean and the moments about origin, moments about the mean are obtained as

$$\mu_2 = \frac{\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12}{\theta^2\left(\theta^4+2\theta^2+2\right)^2}$$

$$\mu_3 = \frac{\left(\begin{array}{l}\theta^{14}+3\theta^{13}+10\theta^{12}+30\theta^{11}+54\theta^{10}+126\theta^9+172\theta^8+264\theta^7+284\theta^6\\+324\theta^5+280\theta^4+216\theta^3+168\theta^2+72\theta+48\end{array}\right)}{\theta^3\left(\theta^4+2\theta^2+2\right)^3}$$

$$\mu_4 = \frac{\left(\begin{array}{l}\theta^{19}+10\theta^{18}+28\theta^{17}+129\theta^{16}+300\theta^{15}+796\theta^{14}+1628\theta^{13}+2952\theta^{12}+4952\theta^{11}\\+6968\theta^{10}+9624\theta^9+11048\theta^8+12368\theta^7+11952\theta^6+10544\theta^5+8544\theta^4\\+5520\theta^3+3648\theta^2+1440\theta+720\end{array}\right)}{\theta^4\left(\theta^4+2\theta^2+2\right)^4}$$

Now, the descriptive measures of PAD including coefficient of variation (C.V), skewness, kurtosis and index of dispersion are obtained as

$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12}}{\theta^4+4\theta^2+6}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\left(\begin{array}{l}\theta^{14}+3\theta^{13}+10\theta^{12}+30\theta^{11}+54\theta^{10}+126\theta^9+172\theta^8+264\theta^7+284\theta^6\\+324\theta^5+280\theta^4+216\theta^3+168\theta^2+72\theta+48\end{array}\right)}{\left(\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12\right)^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^{\,2}} = \frac{\left(\begin{array}{l}\theta^{19}+10\theta^{18}+28\theta^{17}+129\theta^{16}+300\theta^{15}+796\theta^{14}+1628\theta^{13}+2952\theta^{12}+4952\theta^{11}\\+6968\theta^{10}+9624\theta^9+11048\theta^8+12368\theta^7+11952\theta^6+10544\theta^5+8544\theta^4\\+5520\theta^3+3648\theta^2+1440\theta+720\end{array}\right)}{\left(\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12\right)^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12}{\theta\left(\theta^4+2\theta^2+2\right)\left(\theta^4+4\theta^2+6\right)}$$

The nature of coefficients of variation, skewness, kurtosis and index of dispersion of PAD for varying values of parameter are shown in the following Figure 2. It is obvious that the coefficient of variation, skewness, kurtosis and index of dispersion are all increasing for increasing values of parameter (Figure 2).

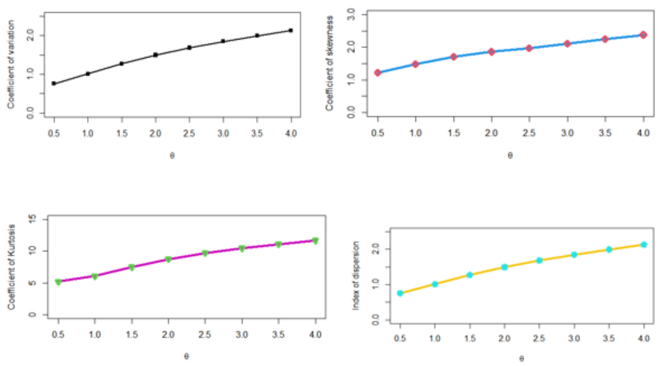

Figure 2 Coefficients of variation, skewness, kurtosis and index of dispersion for varying values of parameter

Statistical properties

Over-dispersion

We have

$$\mu_2 = \frac{\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12}{\theta^2\left(\theta^4+2\theta^2+2\right)^2}$$

$$=\frac{\theta^4+4\theta^2+6}{\theta\left(\theta^4+2\theta^2+2\right)}\left[\frac{\theta^9+\theta^8+6\theta^7+8\theta^6+16\theta^5+24\theta^4+20\theta^3+24\theta^2+12\theta+12}{\theta\left(\theta^4+2\theta^2+2\right)\left(\theta^4+4\theta^2+6\right)}\right]$$

$$=\frac{\theta^4+4\theta^2+6}{\theta\left(\theta^4+2\theta^2+2\right)}\left[1+\frac{\theta^8+8\theta^6+24\theta^4+24\theta^2+12}{\theta\left(\theta^4+2\theta^2+2\right)\left(\theta^4+4\theta^2+6\right)}\right]$$

$$=\mu_1'\left[1+\frac{\theta^8+8\theta^6+24\theta^4+24\theta^2+12}{\theta\left(\theta^4+2\theta^2+2\right)\left(\theta^4+4\theta^2+6\right)}\right]$$

This shows that $\mu_2 > \mu_1'$ and thus PAD is always over-dispersed distribution. Therefore, PAD can be used for discrete data sets which are over-dispersed in nature.

Increasing Hazard Rate and Unimodality

It can be easily shown that PAD has increasing hazard rate (IHR) and is unimodal. Since

$$\frac{P(x+1,\theta)}{P(x,\theta)}=\frac{1}{\theta+1}\left[1+\frac{2\left\{x+\left(\theta^2+\theta+2\right)\right\}}{x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)}\right]$$

is a decreasing function of x for a given θ, $P(x,\theta)$ is log-concave. This implies that PAD has an increasing hazard rate and is unimodal. Grandell[5] has detailed discussion about relationship between log-concavity, IHR and Unimodality of discrete distributions.

Parameter estimation

Method of moment estimate

Let $x_1,x_2,...,x_n$ be a random sample of size n from PAD. Equating the first moment about origin to the corresponding sample moment, the MOME $\tilde{\theta}$ of θ is the solution of the following fifth degree polynomial equation

$\bar{x}\theta^5-\theta^4+2\bar{x}\theta^3-4\theta^2+2\bar{x}\theta-6=0$, where $\bar{x}$ is the sample mean.

This equation can be solved using Newton-Raphson method to get the estimate of the parameter.

Maximum Likelihood Estimate

Let be a random sample of size n from PAD and let f_x be the observed frequency in the sample corresponding to $X=x$ $(x=1,2,3,...,k)$ such that $\sum_{x=1}^{k} f_x=n$, where L is the largest observed value having non-zero frequency. The likelihood function L of PAD is given by

$$L=\left(\frac{\theta^4}{\theta^4+2\theta^2+2}\right)^n\frac{1}{(\theta+1)^{\sum_{x=1}^{k} f_x(x+3)}}\prod_{x=1}^{k}\left[x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)\right]^{f_x}$$

The log likelihood function is obtained as

$$\log L=n\log\left(\frac{\theta^4}{\theta^4+2\theta^2+2}\right)-\sum_{x=1}^{k} f_x(x+3)\log(\theta+1)$$
$$+\sum_{x=1}^{k} f_x\log\left[x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)\right]$$

The first derivative of the log likelihood function is given by

$$\frac{d\log L}{d\theta}=\frac{12n}{\theta}-\frac{4n\left(\theta^3+\theta\right)}{\theta^4+2\theta^2+2}-\frac{n(\bar{x}+3)}{\theta+1}+\sum_{x=1}^{k}\frac{\left[(4\theta+2)x+\left(4\theta^3+6\theta^2+6\theta+2\right)\right]f_x}{x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)},$$

where $\bar{x}$ is the sample mean.

The maximum likelihood estimate (MLE), $\tilde{\theta}$ of θ is the solution of the equation $\frac{d\log L}{d\theta}=0$ and is given by the solution of the non-linear equation

$$\frac{12n}{\theta}-\frac{4n\left(\theta^3+\theta\right)}{\theta^4+2\theta^2+2}-\frac{n(\bar{x}+3)}{\theta+1}+\sum_{x=1}^{k}\frac{\left[(4\theta+2)x+\left(4\theta^3+6\theta^2+6\theta+2\right)\right]f_x}{x^2+\left(2\theta^2+2\theta+3\right)x+\left(\theta^4+2\theta^3+3\theta^2+2\theta+2\right)}=0$$

Since this log-likelihood equation cannot be expressed in closed form, it may be difficult to solve it by direct method. Therefore, the MLE of the parameter θ can be computed iteratively by solving log-likelihood equation using Newton-Raphson iteration available in R-software, until sufficiently close values of the parameter θ is obtained. The initial value of the parameter θ can be taken as the value given by method of moment estimate.

Applications

In this section, the applications of PAD have been discussed for three count datasets which are over-dispersed. The goodness of fit of PAD has been compared with Poisson and PLD. The pmf of PLD is given by

$$P(x,\theta)=\frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}};x=0,1,2,...,\theta>0$$

The expected values given by Poisson, PLD and PAD are given in the table for ready comparison. It is very clear from the goodness of fit presented in tables 1, 2, and 3 that PAD provides a better fit over Poisson and PLD (Tables 1-3).

Table 1 Distribution of mistakes in copying groups of random digits, available in Kemp and Kemp[6]

No. of errors per group	Observed frequency	Expected frequency		
		PD	PLD	PAD
0	35	27.4	33	33.1
1	11	21.5	15.3	15.2
2	8	8.4	6.8	6.7
3	4	2.2	2.9	2.8
4	2	0.5	2.0	2.9
Total	60	60	60	60
ML estimate		$\hat{\theta}=0.7833$	$\hat{\theta}=1.7434$	$\hat{\theta}=1.9141$
χ^2		7.98	2.2	1.72
d.f.		1	1	2
p-value		0.0047	0.138	0.4232

Table 2 Distribution of number of chromatid aberrations (0.2 g chinon 1, 24 hours), available in Loeschke & Kohler[7] and Janardan & Schaeffer[8]

No. of chromatid aberrations	**Observed frequency**	**Expected frequency**		
		PD	**PLD**	**PAD**
0	268	231.3	257	258.1
1	87	126.7	93.4	92.5
2	26	34.7	32.8	32.4
3	9	6.3	11.2	11.2
4	4	0.8	3.8	3.8
5	2	0.1	1.2	1.3
6	1	0.1	0.4	0.4
7+	3	0.1	0.2	0.4
Total	400	400	400	400
ML estimate		$\hat{\theta} = 2.380442$	$\hat{\theta} = 2.380442$	[illegible]
χ^2		38.21	6.21	5.21
d.f.		2	3	3
p-value		0.0000	0.1018	0.1577

Table 3 Accidents to 647 women working on high explosive shells in 5 weeks, available in Sankaran[1]

No. of accidents	**Observed frequency**	**Expected frequency**		
		PD	**PLD**	**PAD**
0	447	406	439.5	440.5
1	132	189	142.8	141.5
2	42	45	45	44.7
3	21	7	13.9	14
4	3	1	4.2	4.3
≥ 5	2	0.1	1.2	2.0
Total	647	647	647	647
ML estimate		$\hat{\theta} = 0.465$	$\hat{\theta} = 2.7182$	$\hat{\theta} = 2.7182$
χ^2		61.08	4.82	4.66
d.f.		1	3	2
p-value		0.0273	0.1855	0.1985

Concluding remarks

In this paper a Poisson mixture of Adya distribution called Poisson-Adya distribution (PAD) has been suggested. The expressions of statistical constants including coefficients of variation, skewness, kurtosis and index of dispersion have been obtained and their behavior for varying values of parameter has been studied. It is observed that the obtained distribution is unimodal, has increasing hazard rate and over-dispersed. Maximum likelihood estimation and method of moment have been discussed for estimating parameter. Finally, the goodness of fit of the proposed distribution and its comparison with other one parameter discrete distributions including Poisson and PLD on three datasets from biological science has been presented.

Acknowledgments

None.

Conflicts of interests

The authors declare no conflicts of interest.

References

1. Sankaran M. The discrete Poisson-Lindley distribution, *Biometrics*. 1970;26(1):145–149.
2. Lindley DV. Fiducial distributions and Bayes theorem, *Journal of the Royal Statistical Society*. 1958;20(1):102–107.
3. Shanker R, Hagos F. On Poisson-Lindley distribution and its Applications to Biological Sciences. *Biometrics & Biostatistics International Journal*. 2015;2(4):103–107.
4. Shanker R, Shukla KK, Ranjan A, et al. Adya distribution with properties and application. *Biometrics & Biostatistics International Journal*. 2021;10(3):81–88.
5. Grandell J. *Mixed Poisson Processes*. Chapman &Hall: London; 1997.
6. Kemp CD, Kemp AW. Some properties of the Hermite distribution. *Biometrika*. 1965;52(3):381–394.
7. Loeschke V, Kohler W. Deterministic and Stochastic models of the negative binomial distribution and the analysis of chromosomal aberrations in human leukocytes. *Biometrische Zeitschrift*. 1976;18(6):427–451.
8. Janardan KG, Schaeffer DJ. Models for the analysis of chromosomal aberrations in human leukocytes. *Biometrical Journal*. 1977;19(8):599–612.