

The discrete poisson-garima distribution

Abstract

In this paper, a discrete Poisson-Garima distribution has been obtained by compounding Poisson distribution with Garima distribution introduced by Shanker.¹ The general expression for the n th factorial moment has been derived and hence moments about origin and central moments have been obtained. The expression for coefficient of Variation, skewness, kurtosis and index of dispersion has been given. Maximum likelihood estimation and the method of moments have been discussed for estimating the parameter of the distribution. Two examples of real data set have been given to test the goodness of fit of the discrete Poisson-Garima distribution and the fit has been compared with Poisson and Poisson-Lindley distributions.

Keywords: garima distribution, poisson-lindley distribution, compounding, moments, skewness, kurtosis, index of dispersion, estimation of parameter, goodness of fit

Volume 5 Issue 2 - 2017

Rama Shanker

Department of Statistics, Eritrea Institute of Technology, Eritrea

Correspondence: Rama Shanker, Department of Statistics, Eritrea Institute of Technology, Eritrea, Email shankerrama2009@gmail.com

Received: December 13, 2017 | **Published:** February 13, 2017

Introduction

Shanker¹ introduced a lifetime distribution named Garima distribution having probability density function

$$f(x; \theta) = \frac{\theta}{\theta+2} (1+\theta+x) e^{-\theta x} ; x > 0, \theta > 0 \quad (1.1)$$

to model behavioral science data. It has been shown by Shanker¹ that Garima distribution gives much better fit than exponential and Lindley² distributions and the new lifetime distributions introduced by Shanker³⁻⁶ namely Shanker, Akash, Aradhana and Sujatha distributions. The first four moments about origin of Garima distribution obtained by Shanker¹ are given by

$$\begin{aligned} \mu_1' &= \frac{\theta+3}{\theta(\theta+2)}, \quad \mu_2' = \frac{2(\theta+4)}{\theta^2(\theta+2)} \\ \mu_3' &= \frac{6(\theta+5)}{\theta^3(\theta+2)}, \quad \mu_4' = \frac{24(\theta+6)}{\theta^4(\theta+2)} \end{aligned}$$

The central moments of Garima distribution obtained by Shanker¹ are as follows

$$\begin{aligned} \mu_2 &= \frac{\theta^2+6\theta+7}{\theta^2(\theta+2)^2} \\ \mu_3 &= \frac{2(\theta^3+9\theta^2+21\theta+15)}{\theta^3(\theta+2)^3} \\ \mu_4 &= \frac{3(3\theta^4+36\theta^3+134\theta^2+204\theta+111)}{\theta^4(\theta+2)^4} \end{aligned}$$

Shanker¹ has studied various properties of Garima distribution including its shape, moments, skewness, kurtosis, hazard rate function, mean residual life function, stochastic ordering, mean deviations, order statistics, Bonferroni and Lorenz curves, entropy measure and stress-strength reliability, amongst others. The estimation of the parameter of Garima distribution has been discussed by Shanker¹ using both maximum likelihood estimation and the method of moments.

In this paper, a Poisson mixture of Garima distribution named, Poisson-Garima distribution (PGD) has been proposed and its various statistical and mathematical properties have been investigated.

The estimation of its parameter has been studied using maximum likelihood estimation and method of moments. Since Poisson-Lindley distribution (PLD), a Poisson mixture of Lindley² distribution and introduced by Sankaran,⁷ gives better fit than Poisson distribution, the Poisson-Garima distribution is expected to give better fit than both Poisson and Poisson-Lindley distribution due to the fact that Garima distribution gives better fit than Lindley distribution. The goodness of fit of the Poisson-Garima distribution has been discussed and it has been compared with that of Poisson and Poisson-Lindley distributions.

Poisson-garima distribution

Assuming that the parameter λ of Poisson distribution follows Garima distribution (1.1), the Poisson mixture of Garima distribution can be obtained as

$$P(X=x) = \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{\theta}{\theta+2} (1+\theta+\lambda) e^{-\theta \lambda} d\lambda \quad (2.1)$$

$$\begin{aligned} &= \frac{\theta}{(\theta+2)x!} \int_0^\infty e^{-(\theta+1)\lambda} \left[(1+\theta)\lambda^x + \theta\lambda^{x+1} \right] d\lambda \\ &= \frac{\theta}{\theta+2} \cdot \frac{\theta x + (\theta^2 + 3\theta + 1)}{(\theta+1)^{x+2}} ; x=0,1,2,3,\dots, \theta>0 \end{aligned} \quad (2.2)$$

This probability mass function (p.m.f.) has been named as "Poisson-Garima distribution (PGD)".

It should be noted that Sankaran⁷ obtained Poisson-Lindley distribution (PLD) having probability mass function (p.m.f)

$$P(X=x) = \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}} ; x=0,1,2,\dots, \theta>0 \quad (2.3)$$

by compounding Poisson distribution with Lindley distribution, introduced by Lindley² having probability density function (p.d.f)

$$f(x, \theta) = \frac{\theta^2}{\theta+1} (1+x) e^{\theta x} ; x > 0, \theta > 0 \quad (2.4)$$

The graphs of the pmf of Poisson-Garima distribution (PGD) and Poisson-Lindley distribution (PLD) for varying values of the parameter are shown in the figure 1

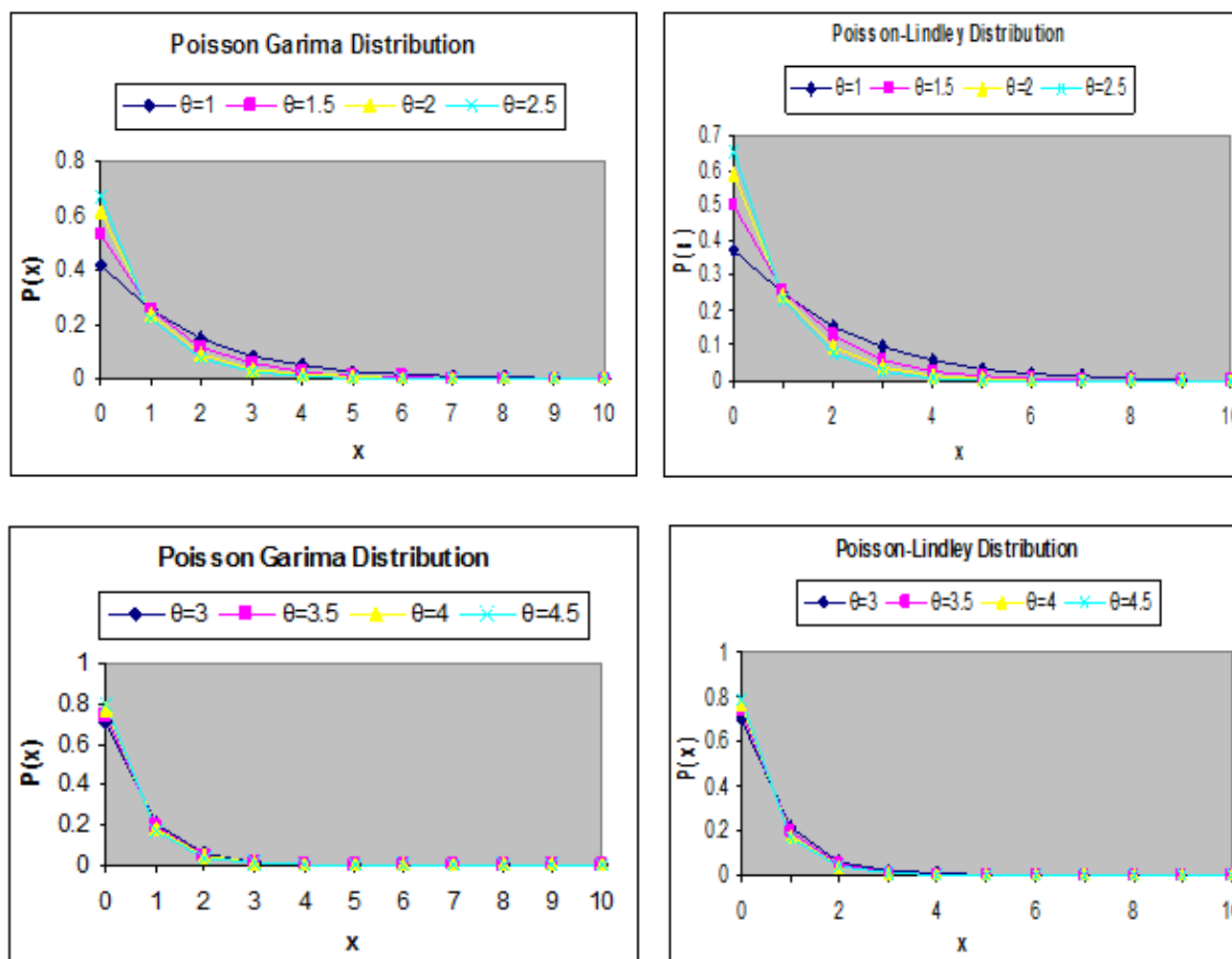


Figure 1 Graphs of the pmf of PGD and PLD For Varying Values of the Parameter.

Moments and related measures of PGD

The r th factorial moment about origin of Poisson-Garima distribution (2.2) can be obtained as

$$\mu_{(r)}' = E\left[E\left(X^{(r)}|\lambda\right)\right], \text{ where}$$

$$\begin{aligned} X^{(r)} &= X(X-1)(X-2)\dots(X-r+1) \\ &= \frac{\theta}{\theta+2} \int_0^\infty \sum_{x=0}^\infty x^{(r)} \frac{e^{-\lambda} \lambda^x}{x!} (1+\theta+\theta\lambda) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta}{\theta+2} \int_0^\infty \lambda^r \sum_{x=r}^\infty \frac{e^{-\lambda} \lambda^{x-r}}{(x-r)!} (1+\theta+\theta\lambda) e^{-\theta\lambda} d\lambda \end{aligned}$$

Taking $x+r$ in place of x within bracket, we get

$$\mu_{(r)}' = \frac{\theta}{\theta+2} \int_0^\infty \lambda^r \left[\sum_{x=0}^\infty \frac{e^{-\lambda} \lambda^x}{x!} \right] (1+\theta+\theta\lambda) e^{-\theta\lambda} d\lambda$$

The expression within the bracket is clearly unity and hence we have

$$\mu_{(r)}' = \frac{\theta}{\theta+2} \int_0^\infty \lambda^r (1+\theta+\theta\lambda) e^{-\theta\lambda} d\lambda$$

Using gamma integral and some algebraic simplification, we get finally a general expression for the r th factorial moment of PGD (2.2) as

$$\mu_{(r)}' = \frac{r!(\theta+r+2)}{\theta^r(\theta+2)}; r=1,2,3,\dots \quad (3.1)$$

Substituting $r=1,2,3$, and 4 in (3.1), the first four factorial moments can be obtained and using the relationship between factorial moments and moments about origin, the first four moments about origin of the PGD (2.2) are obtained as

$$\mu_1' = \frac{\theta+3}{\theta(\theta+2)}$$

$$\mu_2' = \frac{\theta^2+5\theta+8}{\theta^2(\theta+2)}$$

$$\mu_3' = \frac{\theta^3 + 9\theta^2 + 30\theta + 30}{\theta^3(\theta+2)}$$

$$\mu_4' = \frac{\theta^4 + 17\theta^3 + 92\theta^2 + 204\theta + 144}{\theta^4(\theta+2)}$$

Using the relationship between moments about mean and the moments about origin, the moments about mean of the PGD (2.2) are obtained as

$$\mu_2 = \sigma^2 = \frac{\theta^3 + 6\theta^2 + 12\theta + 7}{\theta^2(\theta+2)^2}$$

$$\mu_3 = \frac{\theta^5 + 10\theta^4 + 42\theta^3 + 87\theta^2 + 84\theta + 30}{\theta^3(\theta+2)^3}$$

$$\mu_4 = \frac{(\theta^7 + 19\theta^6 + 148\theta^5 + 607\theta^4 + 1402\theta^3 + 1816\theta^2 + 1224\theta + 333)}{\theta^4(\theta+2)^4}$$

The coefficient of variation (C.V), coefficient of Skewness ($\sqrt{\beta_1}$), coefficient of Kurtosis (β_2) and index of dispersion (γ) of the PGD (2.2) are thus obtained as

$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{\theta^3 + 6\theta^2 + 12\theta + 7}}{\theta + 3}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\theta^5 + 10\theta^4 + 42\theta^3 + 87\theta^2 + 84\theta + 30}{(\theta^3 + 6\theta^2 + 12\theta + 7)^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{(\theta^7 + 19\theta^6 + 148\theta^5 + 607\theta^4 + 1402\theta^3 + 1816\theta^2 + 1224\theta + 333)}{(\theta^3 + 6\theta^2 + 12\theta + 7)^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{\theta^3 + 6\theta^2 + 12\theta + 7}{\theta(\theta+2)(\theta+3)}$$

To study the nature and behavior of μ_1' , μ_2 , C.V, $\sqrt{\beta_1}$, β_2 and γ of PGD and PLD, values of these characteristics for varying values of parameter θ have been computed and presented in table 1

The graph of the coefficient of variation (C.V), coefficient of skewness ($\sqrt{\beta_1}$), coefficient of kurtosis (β_2), and index of dispersion (γ) of PGD and PLD are presented in figure 2.

Table 1 Values of μ_1' , μ_2 , C.V, $\sqrt{\beta_1}$, β_2 and γ of PGD and PLD for varying values of the parameter θ

Values of θ for Poisson-Garima Distribution						
	1	2	3	4	5	6
μ_1'	1.333333	0.625	0.4	0.291667	0.228571	0.1875
μ_2	2.888889	0.984375	0.551111	0.373264	0.279184	0.221788
CV	1.274755	1.587451	1.855921	2.094697	2.311655	2.511701
$\sqrt{\beta_1}$	1.915904	2.147798	2.355147	2.54717	2.727407	2.897852
β_2	8.210059	9.335601	10.36498	11.36106	12.34549	13.32641
γ	2.166667	1.575	1.377778	1.279762	1.221429	1.18287

Values of θ for Poisson-Lindley Distribution						
	1	2	3	4	5	6
μ_1'	1.5	0.666667	0.416667	0.3	0.233333	0.190476
μ_2	3.25	1.055556	0.576389	0.385	0.285556	0.225624
CV	1.20185	1.541104	1.822087	2.068279	2.290174	2.493742
$\sqrt{\beta_1}$	1.792108	2.083265	2.314307	2.517935	2.704839	2.87957
β_2	7.532544	8.941828	10.10611	11.17187	12.19654	13.203
γ	2.166667	1.583333	1.383333	1.283333	1.22381	1.184524

Table 2 Distribution of mistakes in copying groups of random digits

No. of errors per group	Observed Frequency	Expected Frequency		
		PD	PLD	PGD
0	35	27.4	33.0	33.3
1	11	21.5	15.3	15.1
2	8	8.4	6.8	6.6
3	4	2.2	2.9	2.9
4	2	0.5	2.0	2.1
Total	60	60.0	60.0	60.0
ML estimate		$\hat{\theta} = 0.7833$	$\hat{\theta} = 1.7434$	$\hat{\theta} = 1.628413$
χ^2		7.98	2.20	1.71
d.f.		1	1	2
p-value		0.0047	0.1380	0.4253

Table 3 Distribution of *Pyrausta nubilalis*

No. of insects	Observed Frequency	Expected Frequency		
		PD	PLD	PGD
0	33	26.4		
1	12	19.8		
2	6	7.4	31.5	31.7
3	3	1.8	14.2	14.0
4	1	0.3	6.1	6.0
5	1	0.3	2.5	2.5
			1.0	1.0
			0.7	0.8
Total	56	56.0	56.0	56.0
ML estimate		$\hat{\theta} = 0.7500$	$\hat{\theta} = 1.8081$	$\hat{\theta} = 1.695033$
χ^2		4.87	0.53	0.38
d.f.		1	1	1
p-value		0.0273	0.4660	0.5376

Statistical properties of PGD

The PGD (1.3) is always over dispersed ($\sigma^2 > \mu$).

We have

$$\begin{aligned}\sigma^2 &= \frac{\theta^3 + 6\theta^2 + 12\theta + 7}{\theta^2(\theta+2)^2} \\ &= \frac{\theta+3}{\theta(\theta+2)} \left[\frac{\theta^3 + 6\theta^2 + 12\theta + 7}{\theta(\theta+2)(\theta+3)} \right] \\ &= \frac{\theta+3}{\theta(\theta+2)} \left[1 + \frac{\theta^2 + 6\theta + 7}{\theta(\theta+2)(\theta+3)} \right]\end{aligned}$$

$$= \mu \left[1 + \frac{\theta^2 + 6\theta + 7}{\theta(\theta+2)(\theta+3)} \right] > \mu$$

This shows that PGD (2.2) is always over dispersed.

Unimodality and increasing hazard rate

Since

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

is decreasing function in x , $P(x;\theta)$ is log-concave. Therefore, the PGD has an increasing hazard rate and thus unimodal. Detailed

discussion about relationship between log-concavity, unimodality and increasing hazard rate of discrete distribution can be seen in Grandell.⁸

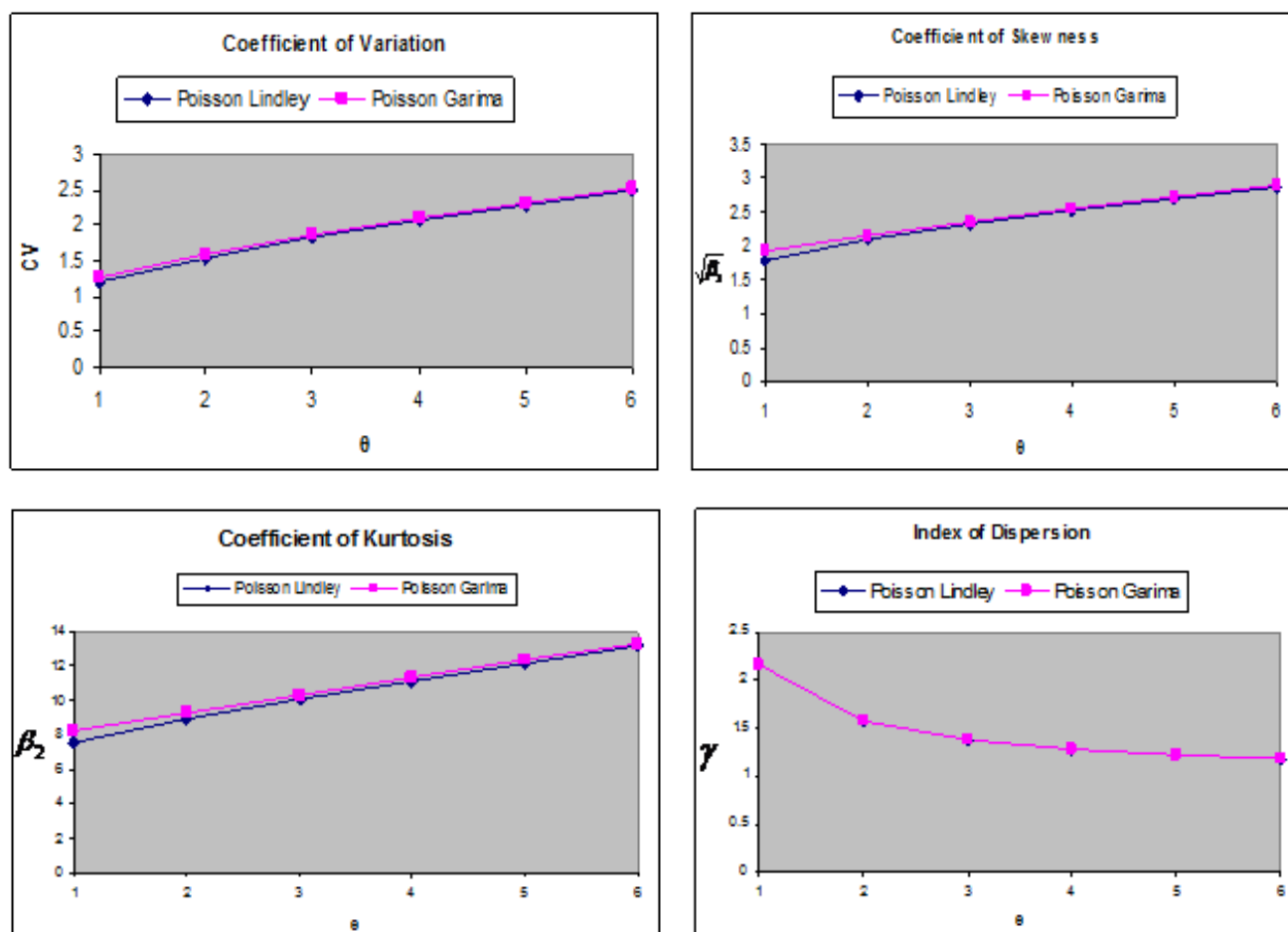


Figure 2 Graphs of $(C.V)$, $(\sqrt{\beta_1})$, (β_2) , and (γ) of PGD and PLD for Varying Values of the Parameter θ .

Generating Functions

Probability generating function: The probability generating function of the PGD (2.2) can be obtained as

$$\begin{aligned}
 P_X(t) &= E(t^X) = \frac{\theta}{(\theta+2)(\theta+1)^2} \left[\theta \sum_{x=0}^{\infty} x \left(\frac{t}{\theta+1} \right)^x + (\theta^2 + 3\theta + 1) \sum_{x=0}^{\infty} \left(\frac{t}{\theta+1} \right)^x \right] \\
 &= \frac{\theta}{(\theta+2)(\theta+1)^2} \left[\frac{\theta(\theta+1)t}{(\theta+1-t)^2} + \frac{(\theta^2 + 3\theta + 1)(\theta+1)}{\theta+1-t} \right] \\
 &= \frac{\theta}{(\theta+2)(\theta+1)} \left[\frac{\theta t}{(\theta+1-t)^2} + \frac{\theta^2 + 3\theta + 1}{\theta+1-t} \right] \\
 &= \frac{\theta^3 + (4-t)\theta^2 + 2(2-t)\theta + (1-t)}{(\theta+1)(\theta+2)(\theta+1-t)^2}
 \end{aligned}$$

Moment Generating Function: The moment generating function of the PGD (2.2) is thus given by

$$M_X(t) = \frac{\theta^3 + (4-e^t)\theta^2 + 2(2-e^t)\theta + (1-e^t)}{(\theta+1)(\theta+2)(\theta+1-e^t)^2}$$

Estimation of parameter

Maximum likelihood estimate (MLE): Let x_1, x_2, \dots, x_n be a random sample of size n from the PGD (2.2) and let f_x be the observed frequency in the sample corresponding to $X=x$ ($x=1, 2, 3, \dots, k$) such that $\sum f_x = n$, where k is the largest observed value having non-zero frequency. The likelihood function L of the PGD (2.2) is given by

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

The log likelihood function is obtained as

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

The first derivative of the log likelihood function is given by

$$\frac{d \log L}{d \theta} = \frac{2n}{\theta(\theta+2)} - \frac{n(\bar{x}+2)}{\theta+1} + \sum_{x=1}^k \frac{(x+2\theta+3)f_x}{\theta x + (\theta^2 + 3\theta + 1)}$$

where \bar{x} is the sample mean.

The maximum likelihood estimate (MLE), $\hat{\theta}$ 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

$$\frac{2n}{\theta(\theta+2)} - \frac{n(\bar{x}+2)}{\theta+1} + \sum_{x=1}^k \frac{(x+2\theta+3)f_x}{\theta x + (\theta^2 + 3\theta + 1)} = 0$$

This non-linear equation can be solved by any numerical iteration methods such as Newton- Raphson, Bisection method, Regula –Falsi method etc

Method of Moment Estimate (MOME): Let x_1, x_2, \dots, x_n be a random sample of size n from the PGD (2.2). Equating the first population moment about origin to the corresponding sample moment, the MOME $\hat{\theta}$ of θ is given by

$$\hat{\theta} = \frac{(1-\bar{x}) + \sqrt{4\bar{x}^2 + 8\bar{x} + 1}}{2\bar{x}}; \bar{x} > 0$$

where \bar{x} is the sample mean.

Applications

The PGD has been fitted to a number of data - sets to test its goodness of fit over Poisson distribution (PD) and Poisson-Lindley distribution (PLD). The parameter has been estimated using maximum likelihood estimation. Two examples of observed data-sets, for which the PD, PLD and PGD has been fitted, are presented. The first data-set is due to Kemp and kemp⁹ regarding the distribution of mistakes in copying groups of random digits and the second data-set is due to Beall¹⁰ regarding the distribution of Pyrausta nubilalis.

Conclusions

A discrete Poisson-Garima distribution has been proposed by compounding Poisson distribution with Garima distribution

introduced by Shanker.¹ Expression for r th factorial moment about origin has been derived and hence moments about origin and central moments have been given. The nature and behavior of coefficient of Variation, skewness, kurtosis and index of dispersion of the proposed distribution have been studied for varying values of the parameter. The estimation of parameter has been discussed using both maximum likelihood estimation and method of moments. The goodness of fit of the proposed distribution has been discussed with two examples of real data set and fit has been compared with Poisson and Poisson-Lindley distributions. The goodness of fit of the Poisson – Garima distribution shows that it gives better fit than both Poisson and Poisson-Lindley distribution and hence it can be considered as an important distribution to model discrete data over these two discrete distributions.

Acknowledgement

None.

Conflict of interest

None.

References

- Shanker R. Garima distribution and its application to model behavioral science data. *Biometrics & Biostatistics International Journal*. 2016;4(7):1–9
- Lindley DV. Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series B*. 1958;20(1):102–107.
- Shanker R. a Shanker Distribution and Its Applications. *International Journal of Statistics and Applications*. 2015;5(6):338–348
- Shanker R. Akash Distribution and Its Applications. *International Journal of Probability and Statistics*. 2015;4(3):65–75.
- Shanker R. a Aradhana Distribution and Its Applications. *International Journal of Statistics and Applications*. 2016;6(1):23–34.
- Shanker R. Sujatha Distribution and Its Applications. *Statistics in Transition–New series*. 2016;17(3):1–20.
- Sankaran M. The discrete Poisson–Lindley distribution. *Biometrics*. 1970;26(1):145–149.
- Grandell J. Mixed Poisson Processes. London: Chapman & Hall. 1997.
- Kemp CD, Kemp AW. Some properties of the Hermite distribution. *Biometrika*. 1965;52(3):381–394.
- Beall G. The fit and significance of contagious distributions when applied to observations on larval insects. *Ecology*. 1940;21:460–474.