STATISTICAL ANALYSIS OF INFLATED HIMANSHU DISTRIBUTION AND ITS APPLICATIONS TO LIFETIME DATA

Abhishek Agarwal[1] and Himanshu Pandey[2]

[1]Department of Mathematics and Statistics, D.D.U Gorakhpur University, Gorakhpur, U.P (India)
abhishek.agarwal014@gmail.com
[2]Department of Mathematics and Statistics, D.D.U Gorakhpur University, Gorakhpur, U.P (India)
himanshu_pandey62@yahoo.com

Abstract

This study suggested an inflated version of the Himanshu distribution. Basic statistical characteristics and the equation for its mass function have been given. By using the method of moments, the estimation of the parameters involved in the suggested inflated distribution has been determined. The suggested inflated model's applicability and flexibility for fitting real-world data sets are clearly demonstrated by an application of the model to the lifetime data set.

Key words: Himanshu distribution, Inflated distribution, Method of moments, Zeroth cell frequency method, Lifetime data sets.

1. INTRODUCTION

There is a critical need in lifespan analysis for some new distributions that are more dependable and adaptable to current real-world data that exhibit significant levels of skewness and kurtosis. Researchers currently confront a significant issue with statistical modeling and applications. In the light of this problem several researchers have proposed different form of distribution function such as Pandey and Jai Kishun(2010), Pandey etal.(2014), Shankar(2017), Shukla and Shankar(2017,2019,2020), Agarwal and Pandey(2022) for the analysis of real set count data in the field of biological science, social science and so on.

2. PROPOSED INFLATED DISTRIBUTION

The regular distributions are extended to inflated form since they cannot accommodate a large number of zeros. The inflated distribution used to examine multiple zeros in real-world count data. The most common distributions, which have been employed in numerous literary works, are the inflated Poisson and inflated Negative Binomial distributions. Agarwal and Pandey (2022) has introduced a Himanshu distribution in discrete nature with probability mass function-

$$P(X=x) = p^n(1-p^n)^x \quad ; \quad \begin{matrix} x = 0,1,2,\dots \\ 0 < p < 1 \\ n \in I^+ \end{matrix} \tag{2.1}$$

$$with\ Mean = \frac{1-p^n}{p^n} \qquad Variance = \frac{1-p^n}{p^{2n}}$$

which is suitable for the set of data related to real life problem under the following assumptions-

a) Let α and $(1-\alpha)$be the probability that event occur at a survey point and does not occur at a survey point respectively.

b) The event X follows a Himanshu distribution as given in (2.1). The proposed inflated distribution of (2.1) takes the following form-

$$P(X=x) = \begin{cases} 1-\alpha+\alpha p^n & ; x = 0 \\ \alpha p^n(1-p^n)^x & ; x = 1,2,3,\dots \end{cases}$$

3. STATISTICAL ANALYSIS

3.1 Moments and Related Measures-

The r[th] moment about origin of the distribution can be given as-

$$\mu_r' = \sum_{x \in R(x)} x^r p(x) = \sum_{x=0}^{\infty} x^r p(x) = \sum_{x=1}^{\infty} \alpha x^r p^n (1-p^n)^x$$

By putting $r = 1,2,3,4$ in above equation we get-

$$Mean = \mu_1' = \alpha p^n \sum_{x=1}^{\infty} x(1-p^n)^x = \alpha\left(\frac{1-p^n}{p^n}\right)$$

$$\mu_2' = \alpha p^n \sum_{x=1}^{\infty} x^2(1-p^n)^x = \frac{\alpha(p^{2n} - 3p^n + 2)}{p^{2n}}$$

$$\mu_3' = \alpha p^n \sum_{x=1}^{\infty} x^3(1-p^n)^x = -\frac{\alpha}{p^{3n}}(p^{3n} - 7p^{2n} + 12p^n - 6)$$

$$\mu_4' = \alpha p^n \sum_{x=1}^{\infty} x^4(1-p^n)^x = \frac{\alpha}{p^{4n}}(p^{4n} - 15p^{3n} + 50p^{2n} - 60p^n + 24)$$

And central moments can be given as-

$$Variance = \mu_2 = \frac{(\alpha-\alpha^2)p^{2n} + (2\alpha^2 - 3\alpha)p^n + (2\alpha - \alpha^2)}{p^{2n}}$$

$$\mu_3 = \frac{\alpha(1-p^n)\{p^{2n}(2\alpha^2-3\alpha+1)+p^n(-4\alpha^2+9\alpha-6)+2(\alpha^2-3\alpha+3)\}}{p^{3n}}$$

$$\mu_4 = \frac{\alpha(1-p^n)\{p^{3n}(3\alpha^3-6\alpha^2+4\alpha-1)+p^{2n}(-9\alpha^3+24\alpha^2-28\alpha+14)+p^n(9\alpha^3-30\alpha^2+48\alpha-36)-3(\alpha^3-4\alpha^2+8\alpha-8)\}}{p^{4n}}$$

Coefficient of variation of the distribution can be given as-

$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{(\alpha-\alpha^2)p^{2n}+(2\alpha^2-3\alpha)p^n+(2\alpha-\alpha^2)}}{(\alpha-\alpha p^n)}$$

Coefficient of skewness can be given as-

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[\alpha(1-p^n)\{p^{2n}(2\alpha^2-3\alpha+1)+p^n(-4\alpha^2+9\alpha-6)+2(\alpha^2-3\alpha+3)\}]^2}{\{(\alpha-\alpha^2)p^{2n}+(2\alpha^2-3\alpha)p^n+(2\alpha-\alpha^2)\}^3}$$

Coefficient of kurtosis of the distribution can be given as-

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\alpha(1-p^n)\{p^{3n}(3\alpha^3-6\alpha^2+4\alpha-1)+p^{2n}(-9\alpha^3+24\alpha^2-28\alpha+14)+p^n(9\alpha^3-30\alpha^2+48\alpha-36)-3(\alpha^3-4\alpha^2+8\alpha-8)\}}{\{(\alpha-\alpha^2)p^{2n}+(2\alpha^2-3\alpha)p^n+(2\alpha-\alpha^2)\}^2}$$

Index of dispersion can be obtained as-

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{(\alpha-\alpha^2)p^{2n}+(2\alpha^2-3\alpha)p^n+(2\alpha-\alpha^2)}{p^n(\alpha-\alpha p^n)}$$

3.2 Estimation of Parameters-The parameters α and p are estimated by equating Zeroth, First cell theoretical frequencies to observed frequencies and theoretical mean to observed mean respectively. Pandey (2003); which is converted into the following equations,

$$\frac{f_0}{f} = 1 - \alpha + \alpha p^n \tag{3.2.1}$$

$$\frac{f_1}{f} = \alpha p^n (1 - p^n) \tag{3.2.2}$$

$$E(x) = \bar{X} = \frac{\alpha}{p^n}(1 - p^n) \tag{3.2.3}$$

4. GOODNESS OF FIT OF PROPOSED DISTRIBUTION

The proposed distribution has been fitted to number of data sets at n=2 to test in goodness of fit. Method of moments has been used to fit the proposed inflated Himanshu distribution. The data set for adult male migrants at the household level was obtained from two surveys, the "RDPG- Survey 1978, BHU" and the "Sample Survey of Rupandehi and Palpa District in Nepal, 2011"; Western Uttar Pradesh data was obtained from Gupta et al. (2016); and the genetic count data set related to the streptonigrin treatment in rabbit lymphoblast was obtained from Hassan et al (2020).

Table 1: Observed and Expected Frequency of the number of households according to the Migrants in Nepal.

Number of Migrants	Observed no. of households	Expected no. of households
0	623	623
1	126	126
2	42	41.64
3	13	13.76
4	4 } 7	6.6
5	2	
6	1	
Total	**811**	**811**
Mean = 0.3464 Variance = 0.5682 $\hat{\alpha} = 0.7016$ $\hat{p} = 0.8182$	$\chi^2 = 0.067$ (after pooling) p-value= 0.9670 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 2: Observed and Expected distribution of the number of households according to single male migrants from a household in Remote type of villages.

Number of Male Migrants (Aged 15 and over)	Observed no. of households	Expected no. of households
0	872	872
1	176	176
2	59	58.04
3	18	19.15
4	6	9.81
5	4	
6	0	
7	0	
8	0	
	$\left.\right\}$ 10	
Total	**1135**	**1135**
Mean = 0.3453 Variance = 0.5672 $\widehat{\alpha} = 0.7011$ $\hat{p} = 0.8184$	$\chi^2 = 0.087$ (after pooling) p-value= 0.9574 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 3: Observed and Expected Frequency of the total number of households according to the Migrants in Western Uttar Pradesh.

Number of Migrants	Observed no. of households	Expected no. of households
0	2679	2679
1	445	445
2	78	82.35
3	19	17.65
4	3	
	$\left.\right\}$ 22	
Total	**3224**	**3224**
Mean = 0.2078 Variance = 0.2592 $\widehat{\alpha} = 0.9150$ $\hat{p} = 0.9027$	$\chi^2 = 1.30$ (after pooling) p-value= 0.2542 $\chi^2_{(1)} = 3.84$ at 5% level of significance	

Table 4: Mammalian cytogenetic dosimetry lesions in Rabbit Lymphoblast induced by streptonigrin (NSC-45383) Exposure-70$\mu g/kg$

Class/Exposure $\mu g/kg$	**Observed Frequency**	**Expected Frequency**
0	200	200
1	57	57
2	30	23.60
3	7	9.77
4	4	
5	0 } 6	9.63
6	2	
Total	**300**	**300**
Mean = 0.5533 Variance = 1.030 $\hat{\alpha} = 0.7831$ $\hat{p} = 0.7654$	$\chi^2 = 3.87$(after pooling) p-value= 0.1444 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

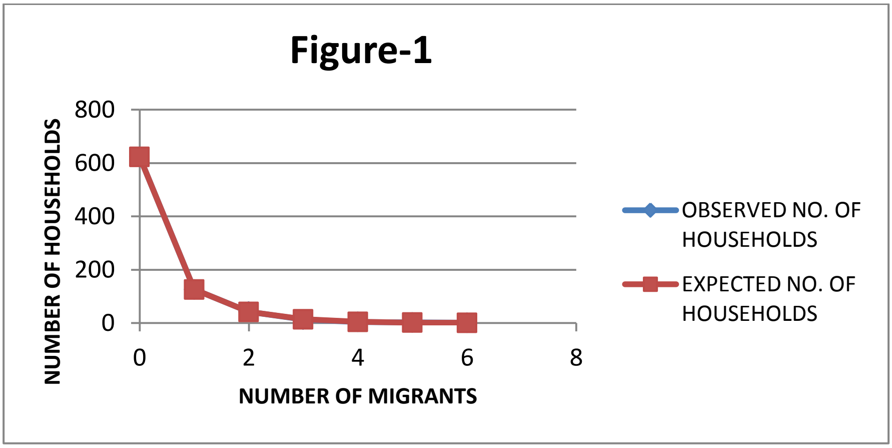

Figure 1: Graphical Presentation Showing Observed and Expected Frequency of the Number of Households according to the Migrants in Nepal.

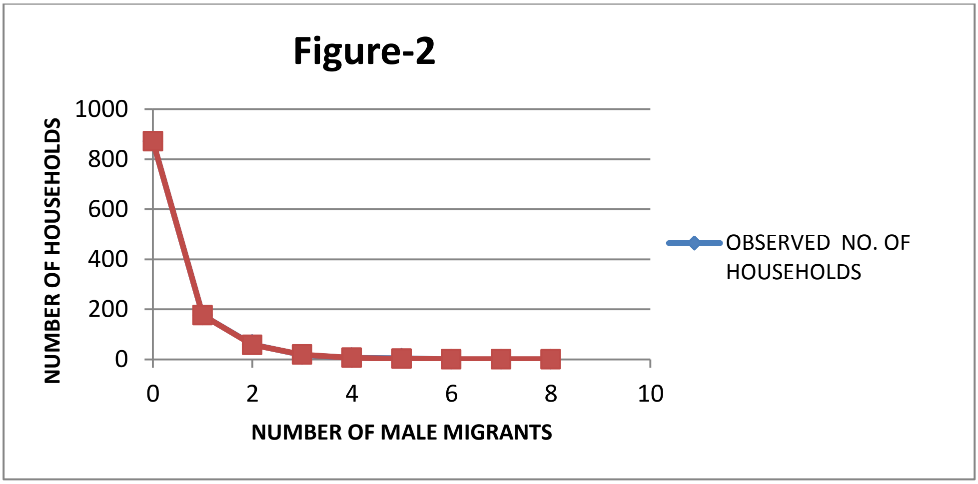

Figure 2: Graphical Presentation Showing observed and Expected distribution of the number of households according to single male migrants from a household in Remote type of villages.

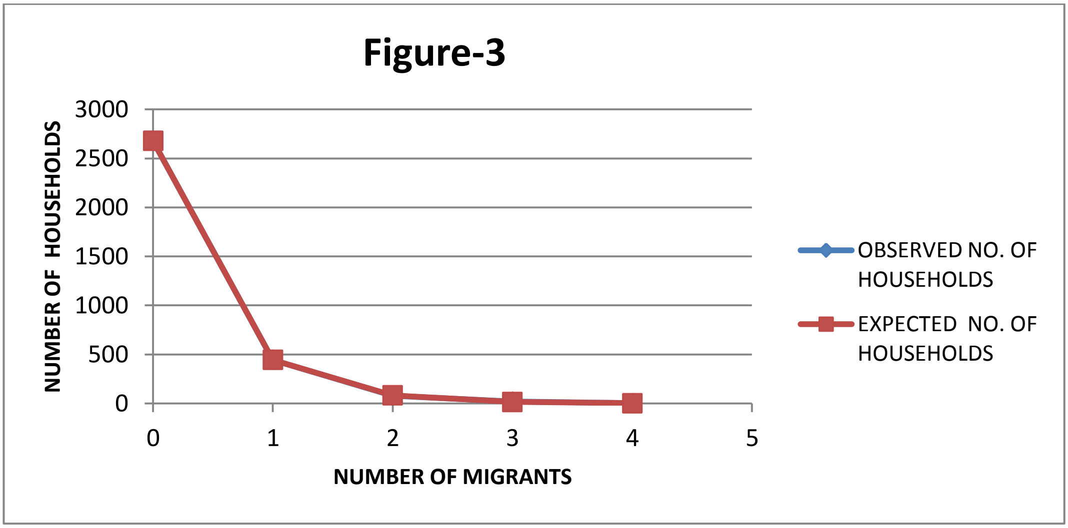

Figure 3: Graphical Presentation Showing Observed and Expected Frequency of the total number of households according to the Migrants in Western Uttar Pradesh.

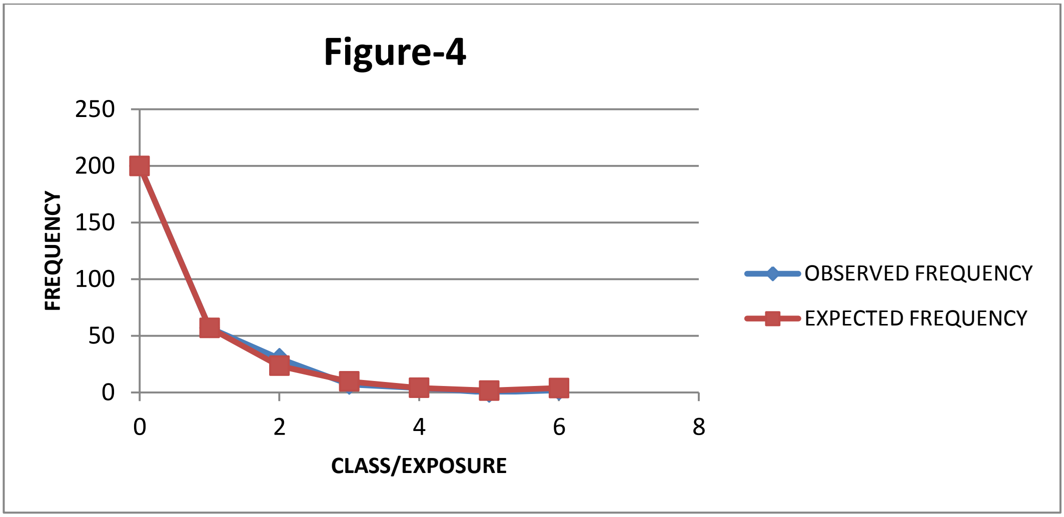

Figure 4: Graphical presentation showing observed and expected frequency on Mammalian cytogenetic dosimetry lesions in rabbit lymphoblast induced by streptonigrin (NSC-45383) Exposure -70$\mu g/kg$.

5. CONCLUSION

The sample data's conformity to the population is determined using a chi-square goodness of fit test. An elevated degree of correlation between your two sets of data is indicated by a low chi-square value. As tables (1–4) clearly indicate that calculated χ^2 is less than the critical χ^2 value at 5% level of significance, we draw the conclusion that there is no significant difference between the observed and predicted value of the given data set. A Chi-square test with a p-value less than or equal to your significance level indicates that there is sufficient evidence to conclude that the observed distribution differs from the expected distribution. According to the value of χ^2 and p-value from (Tables 1-4) and graphical representation between O_i and E_i, the nature and behavior of proposed model found suitable to explain the pattern of migration and genetics count data. Inflated model are better than simple model. Table 1,2,3,4 shows the estimated values of parameters, mean, variance, observed and expected frequencies for different sets of data. The value of χ^2 with degree of freedom and p-value are also given in the tables.

In the light of work done by Rao and Pandey (2017, 2021); it is also possible for the Bayesian analysis of the proposed model.

REFERENCES

[1.] Pandey H., and Jai Kishun.(2010).A probability model for the child mortality in a family. *Revista Colombiana de Estadistica*, 33(1), 1-11.

[2.] Pandey H., Pandey A., and Shukla V. (2014). A displaced probability model for the child mortality in a family. *International Journal of Statistics and Applications,* 4(2), 89-95.

[3.] Shankar R. (2017). Ishita distribution and its application to model lifetime data. *Biometrics and Biostatistics International Journal,* 5(2), 1-9.

[4.] Shukla KK., and Shankar R. (2017). A zero truncated two parameters Poisson Lindley distribution with an application to biological sciences. *Turkiye Klinikleri Journal of Biostatistics*, 9(2), 85-95.

[5.] Shukla KK., and Shankar R.(2019).A discrete Poisson Ishita distribution and its applications. *International Journal of Statistics and Economics,* 20(2), 109-122.

[6.] Shukla KK., Shankar R., and Tiwari M.(2020). Zero truncated Poisson Ishita distribution and its application. *Journal of Scientific Research*, 64(2), 287-294.

[7.] Agarwal A., and Pandey H.(2022).An inflated probability model for the adult male migrants. International *Journal of Statistics and Applied Mathematics*, 7(6), 48-52.

[8.] Agarwal A., and Pandey H.(2022). Himanshu Distribution and its Applications. *Bulletin of Mathematics and Statistics Research,* 10(4), 24-35.

[9.] Pandey H. (2003). Probability model for the number of dependent migrants. *Journal of Stochastic Modeling and Application*, 6(1), 38-43.

[10.] Gupta C.B, Kumar S., and Singh Brijesh P. (2016).Modeling of Rural-Urban migration: A statistical investigation of western Uttar Pradesh (India).*International Journal of Statistics and Applications*, 6(5), 293-299.

[11.] Hassan A., Bilal S., Rashid A., and Ahmad G. (2020). A new flexible discrete distribution with applications to count data. *Journal of Statistical theory and Applications* 19(1):102-108.

[12.] Rao A.K, Pandey H., and Singh K.L.(2017). Bayesian estimation of the parameter of the p-dimensional size biased Rayleigh distribution. *Lithuanian Journal of Statistics* 56(1):88-91.

[13.] Rao A.K., and Pandey H. (2021).Reliability estimation of Weibull Exponential distribution via Bayesian Approach. *Ratio Mathematica*, 40(1), 123-137.