Journal of Probability
and Statistical Science
21(1), 1-7, Feb. 2023

An Inflated Model Based on Size Biased Himanshu Distribution and its Applications

Abhishek Agarwal
Deen Dayal Upadhyay Gorakhpur University

Himanshu Pandey
Deen Dayal Upadhyay Gorakhpur University

ABSTRACT
The purpose of this paper is to develop an inflated probability model through a size biased Himanshu distribution based on certain assumptions. The proposed model involved certain parameters which are estimated by the method of maximum likelihood for the valuable inferences. The suitability of the proposed model is tested by fitting it to an observed set of data.

Keywords: Inflated Model, Himanshu Distribution, Size biased, Method of Maximum likelihood, Goodness of fit.

1. Introduction

The frequency distribution based on real-life data is generally distributed, which allows for frequent zero-valued observations. In this situation, the inflated models of the distributions required to analyze the real-life count data that exhibits many zeros.

The Inflated Poisson and Inflated Negative Binomial are the most popular models that have been used in many literatures under certain conditions.

Size biasing of the distribution is explained in the following way –

Let a random variable X has original probability distribution $P_o(x;\theta)$; $\begin{matrix} x = 0,1,2,\ldots \\ \theta > 0 \end{matrix}$

Suppose the sample units are weighted are selected with probability proportional to some measure. x^α, where α is a positive integer. Then the corresponding size-biased probability distribution of order α can be defined by its probability mass function –

$$P_1(x;\theta) = \frac{x^\alpha . P_0(x;\theta)}{\mu'_\alpha} \tag{1.1}$$

$$\text{where } \mu'_\alpha = E(X^\alpha) = \sum_{x=0}^{\infty} x^\alpha \, P_0(x\,;\theta)$$

If $\alpha = 1$,then the distribution is known as simple size-biased and is applicable to size-biased sampling in sampling theory.

Recently, Pandey and Yadav (2005), Pandey et al. (2015), and Dubey and Pandey (2020) have suggested probability models for vital events.

Received October 2022, revised December 2022, in final form December 2022.

Abhishek Agarwal(corresponding author), Himanshu Pandey is affiliated with the Department of Mathematics and Statistics, D.D.U Gorakhpur University, Gorakhpur, Uttar Pradesh 273009,India, abhishek.agarwal014@gmail.com , himanshu_pandey62@yahoo.com

2. Proposed Inflated Model

Agarwal and Pandey (2022) have introduced a discrete type Himanshu distribution with probability mass function:-

$$P(X = x) = p^n(1-p^n)^{\,x} \quad ; \quad \begin{array}{l} x = 0, 1, 2, \ldots \\ 0 < p < 1 \\ n \in I^+ \end{array} \qquad (2.1)$$

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

$$\mu_3 = \frac{(1-p^n)(2-p^n)}{p^{3n}}, \qquad \mu_4 = \frac{(1-p^n)(9-9p^n+p^{2n})}{p^{4n}}$$

The size biased form of the Himanshu distribution (2.1) is as follows-

$$P_2(x;\theta) = xp^{2n}(1-p^n)^{\,x-1} \quad ; \quad \begin{array}{l} x = 1,2,3, \ldots \\ 0 < p < 1 \\ n \in I^+ \end{array} \qquad (2.2)$$

$$\text{with Mean} = \frac{2}{p^n} - 1, \qquad \text{Variance} = 2\left(\frac{1-p^n}{p^{2n}}\right),$$

$$\mu_3 = \frac{2(1-p^n)(2-p^n)}{p^{3n}}, \qquad \mu_4 = \frac{2(1-p^n)(p^{2n}-12p^n+12)}{p^{4n}}$$

The inflated probability model based on size biased Himanshu distribution is as follows–

$$P(X = x) = \begin{cases} 1-\alpha & ; \quad x = 0 \\ \alpha\, xp^{2n}(1-p^n)^{x-1} & ; \quad x = 1, 2, \ldots \end{cases} \qquad (2.3)$$

where α and p are parameters with

$$\text{Mean} = \frac{\alpha(2-p^n)}{p^n}, \qquad \text{Variance} = \frac{\alpha}{p^{2n}}\{(1-\alpha)p^{2n} - (6-4\alpha)p^n + (6-4\alpha)\}$$

3. Estimation of the parameters

The parameters of the proposed inflated probability model are estimated by the method of maximum likelihood in the following way:

$$L = (1-\alpha)^{f_0}(\alpha p^{2n})^{f_1}[1-\{(1-\alpha)+\alpha p^{2n}\}]^{f-f_0-f_1} \tag{3.1}$$

$$\log L = f_0 \log(1-\alpha) + f_1 \log(\alpha p^{2n}) + (f - f_0 - f_1)\log(\alpha - \alpha p^{2n})$$

Now, by partially differentiating (3.1) with respect to α and equating to zero, we get–

$$\hat{\alpha} = 1 - \frac{f_0}{f}$$

And partially differentiate (3.1) with respect to p and equating to zero, we get–

$$\hat{p} = \left(\frac{f_1}{f - f_0}\right)^{\frac{1}{2n}}$$

where $\hat{\alpha}$ *and* $\hat{p}$ are maximum likelihood estimators of α and p respectively.

Now

$$\frac{E\left(\frac{-\partial^2}{\partial \alpha^2}\log L\right)}{f} = \left[\frac{1}{1-\alpha} + \frac{1}{\alpha}\right]$$

and

$$\frac{E\left(\frac{-\partial^2}{\partial p^2}\log \mathrm{L}\right)}{f} = \frac{4n^2\alpha p^{2n-2}}{1-p^{2n}}$$

Ultimately we get-

$$E(f_0) = f(1-\alpha)$$

$$E(f_1) = f\alpha p^{2n}$$

After estimating all the parameters involved in the proposed model by MLE,

We are interested in calculating the variance of each estimated parameter.

$$FIM = \begin{bmatrix} \frac{1}{1-\alpha} + \frac{1}{\alpha} & 0 \\ 0 & \frac{4n^2\alpha p^{2n-2}}{1-p^{2n}} \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix}$$

$$V(\hat{\alpha}) = \frac{1}{f}\left[\frac{\varphi_{22}}{\varphi_{11}\,\varphi_{22}}\right] = \frac{\alpha(1-\alpha)}{f}$$

$$V(\hat{p}) = \frac{1}{f}\left[\frac{\varphi_{11}}{\varphi_{11}\,\varphi_{22}}\right] = \frac{1}{f}\left(\frac{1-p^{2n}}{4n^2\alpha p^{2n-2}}\right)$$

4. Application

Under the assumption, the proposed inflated model (2.3) at $n = 2$ was applied to data for adult male migrants at household level from two surveys, namely "RDPG-Survey 1978, BHU" and "Sample survey of Rupandehi and Palpa district in Nepal-2011", where *x* denotes the number of male migrants from a households at a survey point and α, (1−α) be the probability that a household is exposed to and does not exposed to risk of migration at a survey point respectively, is properly used to analyze the migration pattern of the human society which is shown in table 1 and 2.

Similarly the proposed inflated model with α, (1−α) being the probability of being exposed to and not exposed to the risk of streptonigrin treatment in the lymphoblast (cells) of the rabbit is used very frequently in Genetics count data sets of catcheside et al. taken from Hassan (2020).

Table 1. Observed and Expected distribution of the number of households according to single Male Migrants from a household in Remote type of villages.

No. of Male migrants aged 15 and over	Remote type of villages	
	Observed No. of households	Expected No. of households
0	872	872
1	176	176
2	59	64.03
3	18	17.48
4	6 } 10	5.49
5	4	
6	0	
7	0	
8	0	
Total	**1135**	**1135**
Mean = 0.3453 Variance = 0.4978 $\hat{\alpha} = 0.2317$ $\hat{p} = 0.9044$ $V(\hat{\alpha}) = 0.00015684$ $V(\hat{p}) = 9.611\ e^{-5}$	$\chi^2 = 4.10$ (after pooling) p-value = 0.1287 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 2. Observed and Expected Frequency of the number of households according to migrants in Nepal.

No. of Migrants	Observed no. of households	Expected no. of households
0	623	623
1	126	126
2	42	45.71
3	13	12.43
4	4 } 7	3.86
5	2	
6	1	
Total	**811**	**811**
Mean = 0.3465 Variance = 0.3452 $\hat{\alpha} = 0.2318$ $\hat{p} = 0.9047$ $V(\hat{\alpha})$ =0.0002195 $V(\hat{p})$ =0.0001339	$\chi^2 = 2.87$ (after pooling) p-value=0.2381 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 3. Mammalian cytogenetic dosimetry lesions in Rabbit lymphoblast induced by streptonigrin (NSC-45383) Exposure-70 µg/Kg

Class/Exposure	Observed Frequency	Expected Frequency
0	200	200
1	57	57
2	30	27.93
3	7	10.26
4	4 } 6	4.80
5	0	
6	2	
Total	**300**	**300**
Mean = 0.5533 Variance = 0.8857 $\hat{\alpha} = 0.3300$ $\hat{p} = 0.8688$ $V(\hat{\alpha}) = 0.000737$ $V(\hat{p}) = 0.0003596$	$\chi^2 = 1.48$ (after pooling) p-value = 0.4771 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

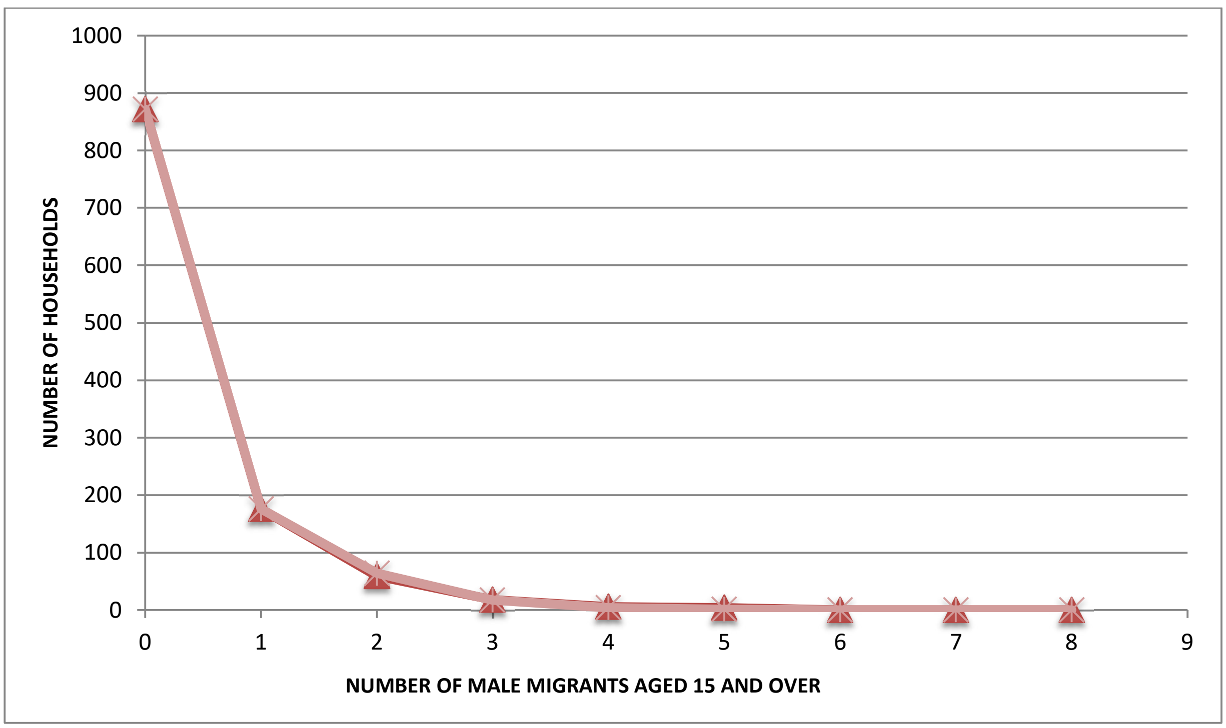

Figure 1.Graphical presentation showing Observed and Expected number of households having adult male migrants aged (15 years and above) in Remote type of villages.

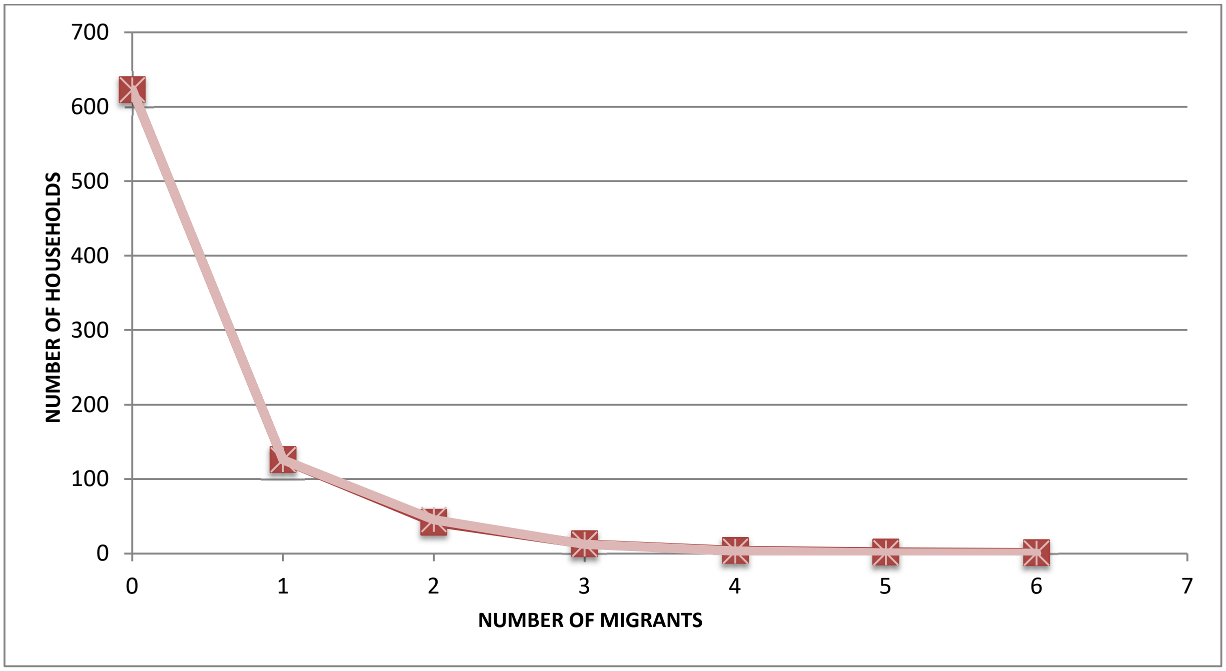

Figure 2.Graphical presentation showing Observed and Expected number of households having adult male migrants aged (15 years and above) in Nepal.

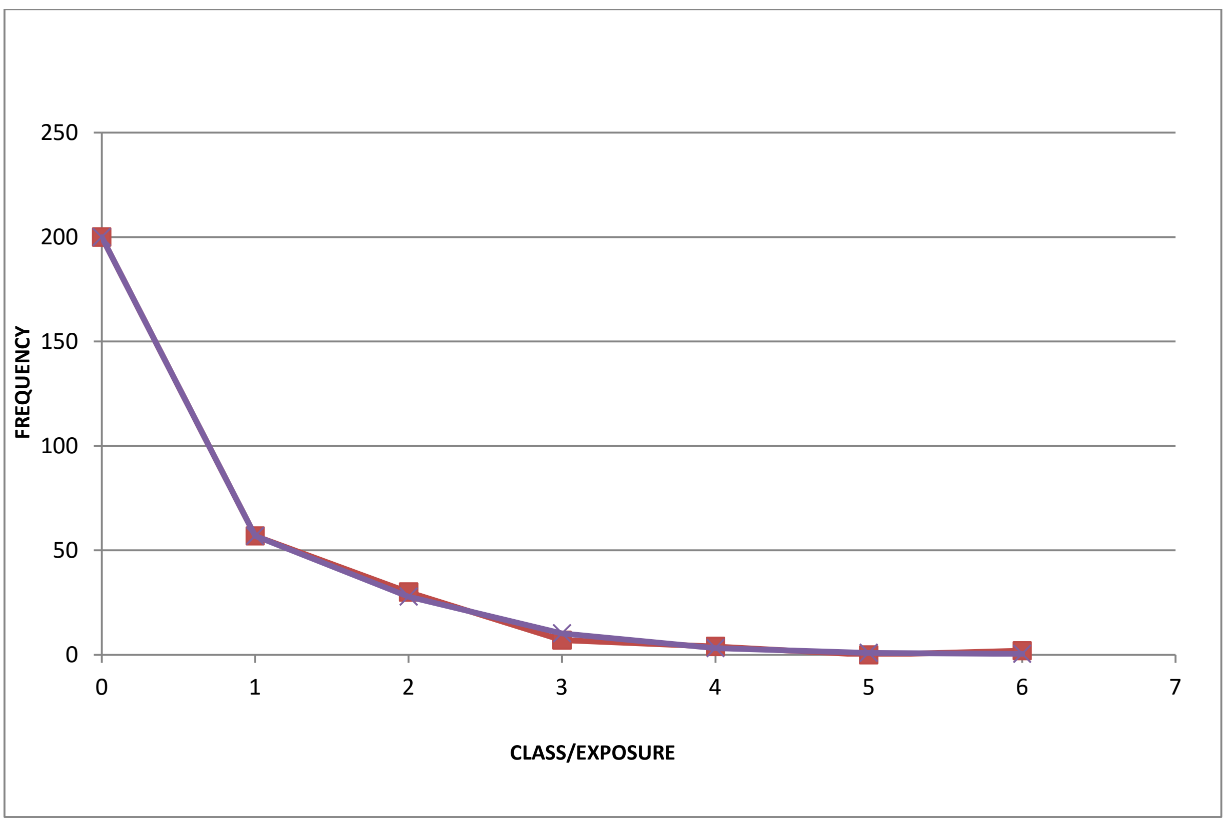

Figure 3.Graphical presentation showing Observed and Expected frequency of Mammalian cytogenetic dosimetry lesions in Rabbit lymphoblast induced by streptonigrin (NSC-45383) Exposure-70 **μ**g/Kg.

5. Conclusion

The study with respect to above tables and graphs shows that the proposed inflated model is a good approximation for the migration data and genetic count data, which implies the proposed model could be very useful for the analysis and interpretation of the count data related to real-life data. The estimated value of the parameters, the value of χ^2 with degree of freedom and $p-value$ are also given in the table, which justifies the suitability of the proposed inflated model. In the light of the work done by Rao and Pandey (2020, 2022), it is also possible to perform the Bayesian analysis of the proposed model.

References

[1]. Agarwal A., Pandey H., (2022). Himanshu distribution and its applications, Bulletin of Mathematics and Statistics Research. 10(4):24−35.

[2]. Agarwal A., Pandey H. (2022). Probability model based on Zero truncation of Himanshu distribution and their applications, Asian Journal of probability and Statistics. 20(4):57-67.

[3]. Agarwal A., Pandey H. (2022). An Inflated Probability Model for the Adult Male Migrants, International Journal of Statistics and applied Mathematics. 7(6):48-52.

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

[5]. Pandey H., Dubey A. (2020). An Inflated probability model for commutation Pattern, Bulletin of Mathematics and Statistics Research. 8(4):101−108.

[6]. Pandey H. et al. (2015). An Inflated Probability Model on Rural out Migration, Journal of Computer and Mathematical Sciences. 6(12):702−711.

[7]. Rao. A.K. Pandey H. (2020). Parameter estimation Dagum distribution, International Journal of Mathematical trends and technology. 66(1):9−15.

[8]. Rao. A.K., Pandey H. (2022). Parameter estimation Generalized Inverse generalized Weibull distribution via Bayesian Approach, European Scholar Journal. 3(3):23−28.

[9]. Yadav A.K, Pandey H. (2005). A probability Model for distribution of level of fecund ability of Migrated couples, Journal of International Academy of Physical Sciences. 9(1):91−98.