RESEARCH ARTICLE

*Corresponding author.

manashstats@gmail.com

One Inflated Binomial Distribution and its Real-Life Applications

Tousifur Rahman[1], Partha Jyoti Hazarika[1], Manash Pratim Barman[1]*

1 Department of Statistics, Dibrugarh University, Dibrugarh, 786004, Assam, India

Abstract

Objective: To introduce a one-inflated Binomial distribution (OIBD) and discuss its applications. **Methods:** Study its distributional properties, reliability characteristics, and estimation of its parameters using the method of moment estimation (MM) and maximum likelihood estimation (MLE). A simulation study has been conducted to see the behaviour of the MLEs. Two real-life examples are used to examine the pertinent of the proposed distribution. **Findings :** The proposed one-inflated binomial distribution (OIBD) provides better fitting in terms of AIC, BIC, and KS test comparison to the other known distributions. **Novelty:** Develop a new statistical distribution to study the count data having inflated frequency at count one, along with the different statistical properties. The practical utility of the distribution is also discussed with real-life examples.

Keywords: One inflated Binomial distribution; MM; MLE; KS; AIC and BIC

1 Introduction

It has been observed that count data with a surplus amount of zero's, one's, two's, three's, etc are common in studies related to health, insurance, agriculture, etc.[1,2]. These inflated frequencies may create trouble in data analysis due to over-dispersion at a particular count. The already existed distributions developed for studying count data may not be suitable in the presence of inflated frequency. Hence, inflated models are generated to overcome such situation and to overcome workable irregularity in parameter estimation[1]. The inflated distribution is a mixture between a point mass at a particular count and any other count distribution supported by non-negative integers[3].

To overcome the situation of surplus zero's Neyman[4] and Feller[5] first introduced the idea of zero inflation. The structural properties and MLE's of inflated discrete distribution inflated at zero was studied by Gupta et al.[6] The results of Gupta et al.[6] were extended by Murat and Szynal[7]. They studied that discrete distributions may inflate at any point say. Lambert[8] proposed a zero-Inflated Poisson regression (ZIP) model with surplus zeroes with an example of a manufacturing defect. Ridout et al.[9] furnishes an aspect of zero-inflated counts on creating contraption of happening as well as acceptable designing structure by citing examples of sexual behaviour and species abundance. By taking on the methodology of Lambert's[8] ZIP regression model, Hall[10] derived a zero-Inflated Binomial Model. Two-inflated binomial distribution was used by Singh et al.[11] to investigate the mechanism of son

preference through the modelling of the pattern of male children in Uttar Pradesh, where family size and sex composition are dominated by strong son preference. The parameters of a zero-inflated Poisson distribution were estimated by Beckett et al.[12] and they modelled some natural calamities data using these parameters. Beckett et al.[12] also juxtapose MLEs and MMEs regarding Standardized bias and standardized mean squared error. Zero-inflated binomial distribution was characterized by Najundan et al.[13]; Zero-inflated negative binomial distribution was characterized by Suresh et al.[14]; Zero-inflated Poisson distribution was characterized by Najundan et al.[15]. Alshkaki[16] extended the zero-inflated Poisson distribution to zero-one-inflated Poisson distribution and also studied its structural properties and estimates its parameters by method of maximum likelihood and method of moments. Mwalili et al.[17] studied a zero-inflated negative binomial model to gratify extravagant zeros, an extension of negative binomial distribution. Alshkaki[18] studied the structural properties and estimated the parameters by the method of moments of zero-one inflated negative binomial distribution. Sakthivel and Rajitha[19] proposed a probability based inflation estimator for zero-inflated Poisson model, which will be helpful for inferences about the inflation parameter. Jornsation and Bodhisuwan[20] proposed a zero-one inflated negative binomial-beta exponential distribution along with its distributional properties and estimate its parameters using method of maximum likelihood.

The Binomial distribution is a well-known non-negative integer-valued discrete distributions. While sampling binomial data, it is often observed that either the number of counts of zero or One or two has a higher frequency than that expected, which can be explained by an appropriate inflated distribution, say, zero-inflated or one-inflated or two-inflated binomial distribution.

The probability distribution of the number of successes, so obtained is called the Binomial Probability distribution. The binomial distribution is a discrete distribution as X can take only the integral values, viz. 0,1,2,3,4…,n

A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is

$$P(x;n,p) = P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}; & x=0,1,2,\ldots,n \\ 0; & otherwise \end{cases} \tag{1}$$

It is denoted by $B(n,p)$

The Zero-Inflated Binomial Distribution (ZIBD) was proposed by Hall[10]. Let $X \sim B(n,p)$ as given in (1), let $\alpha \in (0,1)$ be an extra proportion added to the proportion of zero of the random variable X, then the random variable X defined by

$$P(x;n,p,\alpha) = P(X=x) = \begin{cases} \alpha + (1-\alpha)(1-p)^n & ; x=0 \\ (1-\alpha)\binom{n}{x} p^x (1-p)^{n-x} & ; x=1,2,3,4,\ldots,n \\ 0 & ; otherwise \end{cases} \tag{2}$$

$$where\ 0 < \alpha < 1,\ 0 < p < 1$$

is said to have a zero-inflated binomial distribution and it is denoted by $X \sim ZIBD(n,p,\alpha)$

Note: If $\alpha \to 0$, then the above distribution reduces to standard binomial distribution (1).

The Two-Inflated Binomial Distribution (TIBD) was proposed by Singh et al.[11]. Let $X \sim B(n,p)$ as given in (1), let $\alpha \in (0,1)$ be an extra proportion added to the proportion of two of the random variable X, then the random variable X defined by

$$P(x;n,p,\alpha) = P(X=x) = \begin{cases} \alpha + (1-\alpha)\binom{n}{2} p^2 (1-p)^{n-2} & ; x=2 \\ (1-\alpha)\binom{n}{x} p^x (1-p)^{n-x} & ; x=0,1,3,4,\ldots,n \\ 0 & ; otherwise \end{cases} \tag{3}$$

Where $0 < \alpha < 1,\ 0 < p < 1$, is said to have a two inflated binomial distribution and it is denoted by $X \sim TIBD(n,p,\alpha)$

Note: If $\alpha \to 0$, (3) reduces to standard binomial distribution (1).

Here in this paper, the researchers propose a one-inflated binomial distribution along with its distributional properties, reliability characteristics and consider the method of moment estimation (MM) and maximum likelihood estimation (MLE) to estimate its parameters. A simulation study has been conducted to see the behaviour of the MLEs. Here two real-life data sets are used to examine the pertinent of the proposed distribution.

2 One Inflated Binomial Distribution (OIBD)

Let $X \sim B(n,p)$ as given in (1), let $\alpha \in (0,1)$ be an extra proportion added to the proportion of one of the random variable X, then the random variable X defined by

$$P(x;n,p,\alpha) = P(X=x) = \begin{cases} \alpha + n(1-\alpha)p(1-p)^{n-1} & ; x = 1 \\ (1-\alpha)\binom{n}{x}p^x(1-p)^{n-x} & ; x = 0, 2, 3, 4, \ldots, n \\ 0 & ; \text{otherwise} \end{cases} \quad (4)$$

Where $0 < \alpha < 1$, $0 < p < 1$, is said to have one-inflated binomial distribution and in the rest of the article it will be denoted by $OIBD(n,p,\alpha)$

Some particular cases: When

1. $\alpha \to 0$, $OIBD(n,p,\alpha)$ reduces to $B(n,p)$.
2. $\alpha \to 0$, $n \to \infty$, p is very small and $np = \lambda(finite)$, $OIBD(n,p,\alpha)$ reduces to Poisson distribution.
3. $\alpha \to 0$, $n \to \infty$, and p is very small, $OIBD(n,p,\alpha)$ reduces to Normal distribution.

The pmf plots of $OIBD(n,p,\alpha)$ with different choice of parameters values of n, p and α to study the variety of shapes are provided in Figure 1.
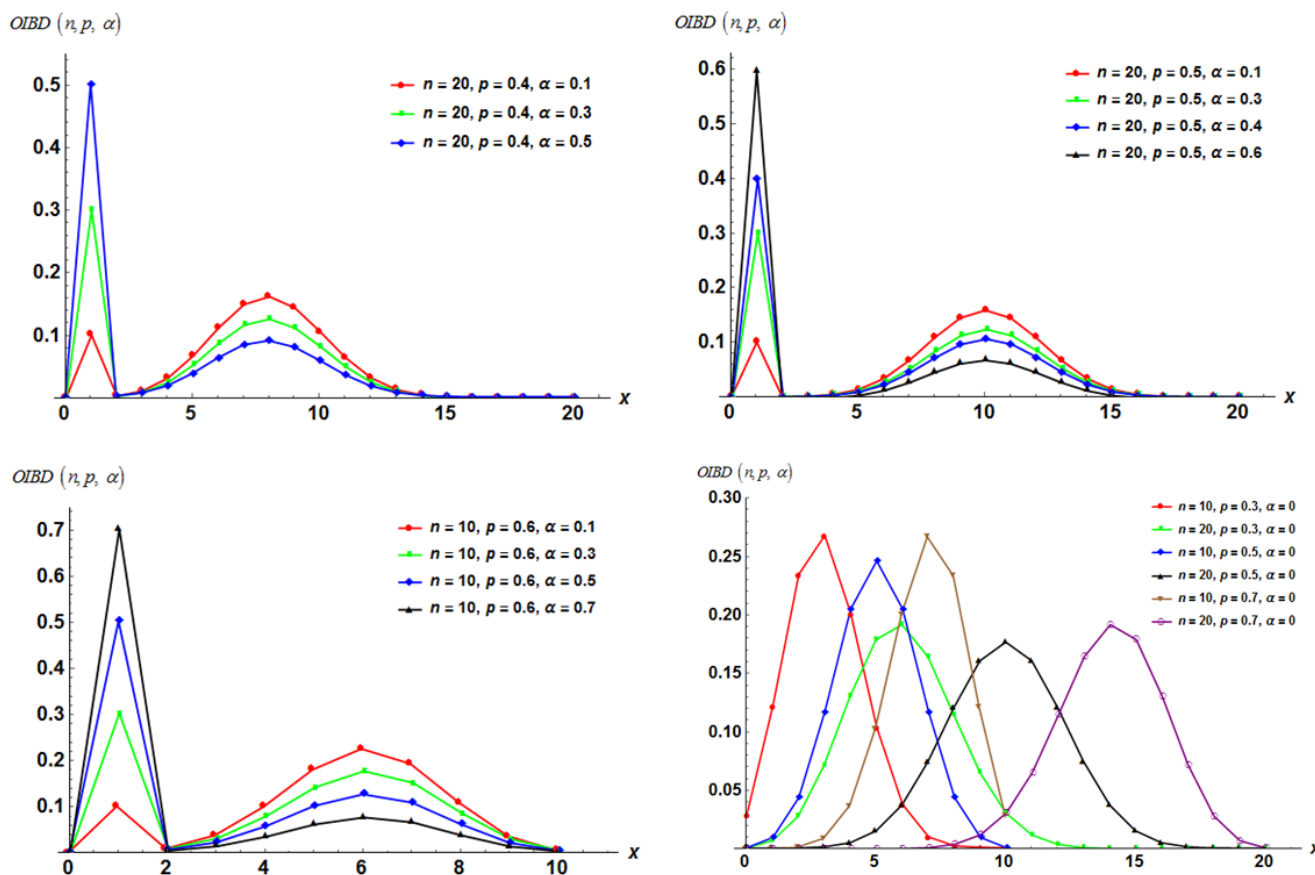

Fig 1. Shape of $OIBD(n,p,\alpha)$ with different choice of parameters

3 Distributional Properties

3.1 Moments

Theorem 1: If $X \sim OIBD(n, p, \alpha)$, then its rth order moments about zero is as follows

$$\mu'_r = E(X^r) = \alpha + (1-\alpha)\sum_{j=0}^{r} \frac{S(r,j)n!p^r}{(n-r)!} \tag{5}$$

Proof: If $X \sim OIBD(n, p, \alpha)$ then the rth order moments about zero is

$$\mu'_r = E(X^r)$$

$$= \sum_{x=0}^{n} x^r p(X = x)$$

$$= \alpha + \sum_{x=0}^{n} x^r (1-\alpha) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \alpha + (1-\alpha) \sum_{x=0}^{n} x^r \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \alpha + (1-\alpha) \sum_{j=0}^{r} \frac{S(r,j)n!p^r}{(n-r)!}$$

$$\sum_{x=0}^{n} x^r \binom{n}{x} p^x (1-p)^{n-x} = \sum_{j=0}^{r} \frac{S(r,j)n!p^r}{(n-r)!}$$

Where $S(r, j)$ is the first kind of Stirling number.

Proved.

In particular, the first four moments of $OIBD(n, p, \alpha)$ can be obtained as

$$\mu'_1 = E(X) = \alpha + (1-\alpha)np \tag{6}$$

$$\mu'_2 = E(X^2) = \alpha + (1-\alpha)np(np+1-p) \tag{7}$$

$$\mu'_3 = E(X^3) = \alpha + (1-\alpha)\left(n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np\right) \tag{8}$$

$$\mu'_4 = E(X^4) = \alpha + (1-\alpha)\left(\begin{array}{c} n(n-1)(n-2)(n-3)p^4 + \\ 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np \end{array} \right) \tag{9}$$

Therefore,

$$Variance, \ V(X) = (1-\alpha)\left[(np-1)^2\alpha + np(1-p)\right] \tag{10}$$

The plots of mean and variance of the proposed distribution with different choice of parameters to study their variations are shown in the Figures 2 and 3.
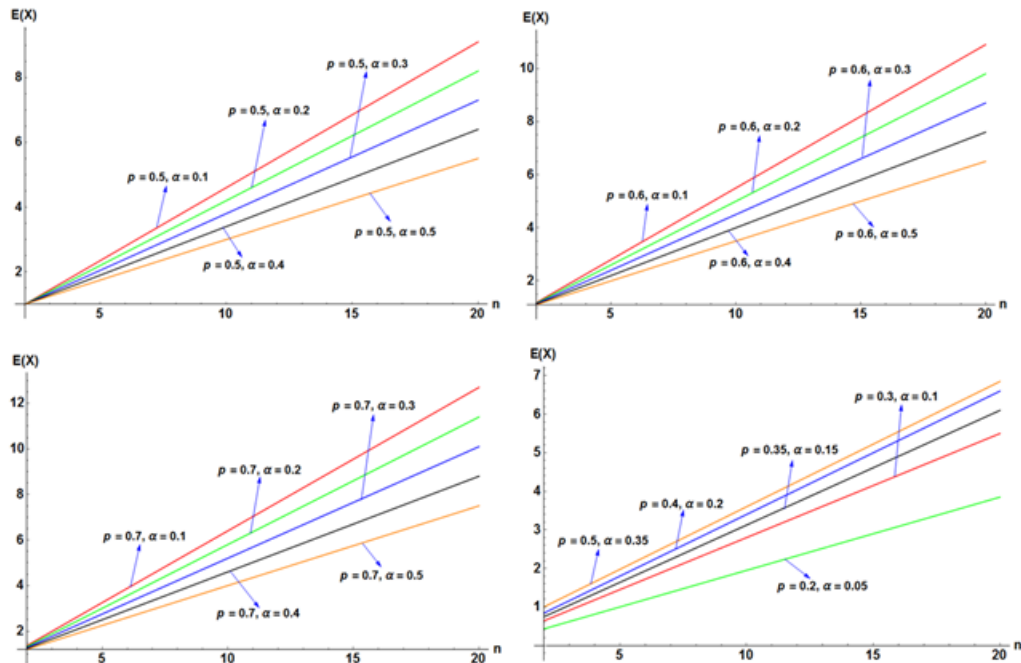
Fig 2. Mean against n for different values of p and α
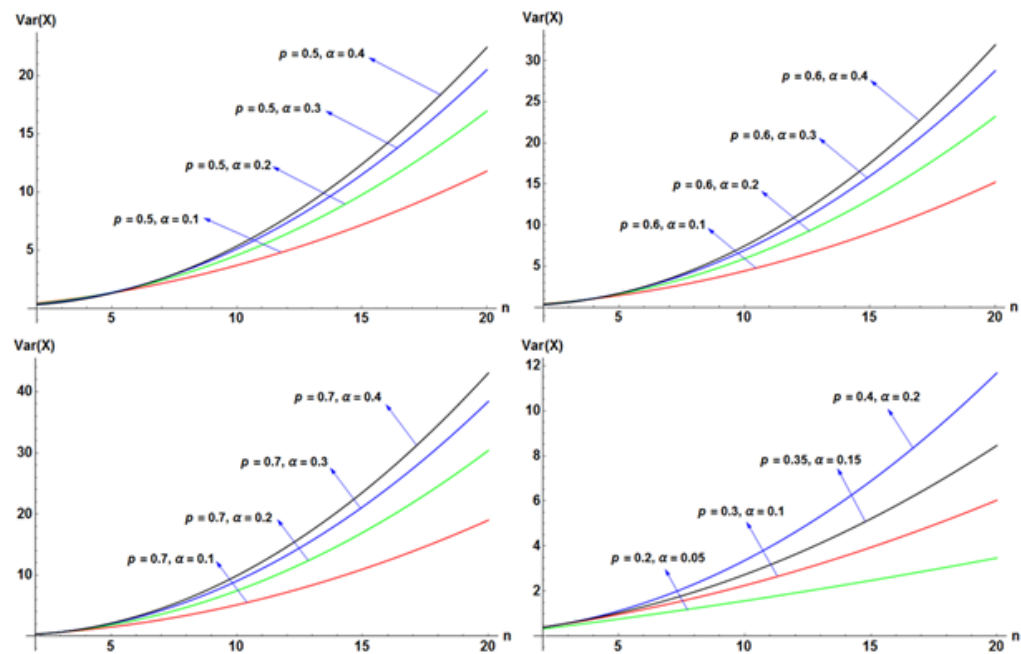

Fig 3. Variance against n for different values of p and α

From the Figure 2 it is clear that as α decreases and p and n increases, the mean of the proposed distribution increases. From the Figure 3 it is clear that as α decreases, the variance of the proposed distribution decreases.

3.1.1 Coefficient of Skewness
If $X \sim OIBD(n, p, \alpha)$ the Pearson's β_1 coefficient is as follows

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\left[\mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3\right]^2}{\left[\mu_2' - \mu_1^2\right]^3}$$

$$= \frac{\left[\begin{array}{l} np^2((n-1)(3+(n-2)p) - 3(\alpha(2\alpha-3)+1)) \\ +np^3(1-\alpha)(2(1-\alpha) - 3n(n-1)) + \alpha\left(1 - 3(n-1)np^2 - 2\alpha\right) \\ +np(1-6\alpha(1-\alpha)) \end{array}\right]^2}{(1-\alpha)[np(1-2\alpha + p(n-2+\alpha)) + \alpha]^3}$$

(11)

The plots of coefficient of skewness of the proposed distribution for different choice of parameters are shown in the Figure 4

Fig 4. β_1 against n for different values of p *and* α

From the Figure 4 it is observed that

1. As $p = 0.5$ and $\alpha \geq 0.5, \beta_1$ decreases
2. As $p \geq 0.5$ and $\alpha < 0.1, \beta_1$ increases
3. As $p < 0.5$ and $\alpha < 0.1, \beta_1$ decreases
4. As $p \neq 0.5$ and $0.1 \leq \alpha \leq 0.4$ & $0.6 \leq \alpha \leq 0.9$, β_1 increases
5. As $\alpha = 0.5, \beta_1$ decreases for $p(0 < p < 1)$

Remark 1

- As $\alpha \to 0$ and $p = 0.5$ and then the coefficient of skewness $\beta_1 \to 0$ i.e. the proposed distribution tends to symmetric distribution.

3.1.2 Coefficient of Kurtosis
If $X \sim OIBD(n, p, \alpha)$ the Pearson's β_2 coefficient is as follows

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\left[\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4\right]}{\left[\mu_2' - \mu_1'^2\right]^2}$$

$$= \frac{\begin{bmatrix} (n-1)np^2(7+(n-2)p(6+(n-3)p)) \\ +\alpha\left(4(n-1)np^2(3+(n-2)p) + 3\alpha\left(1-2(n-1)np^2\right) - 3\alpha^2\right) \\ +np^2(1-\alpha)\begin{pmatrix} 3p^2(1-\alpha)^2 + 4 + 6np^2(1-\alpha) \\ -6n^2p^2(1-\alpha) - 18(1-\alpha)\alpha \end{pmatrix} \\ +4(1-\alpha)n^2p^3\left(n^2p + 3n(1-p-\alpha) - (3-2p-3\alpha)\right) - np(1-2\alpha) \\ \left(1-6(1-\alpha)\alpha + 6p^2(1-\alpha)^2\right) \end{bmatrix}}{(1-\alpha)[np(1-2\alpha + p(n-2+\alpha)) + \alpha]^2} \qquad (12)$$

The plots of coefficient of Kurtosis of the proposed distribution for different choice of parameters are shown in the Figure 5
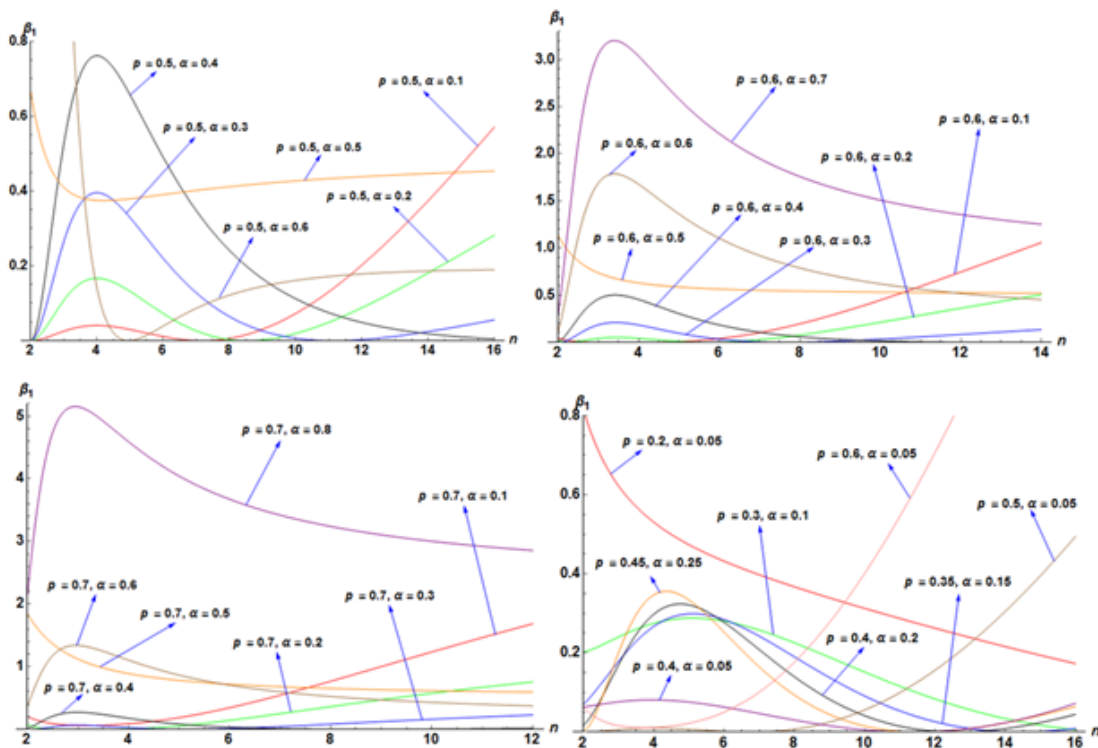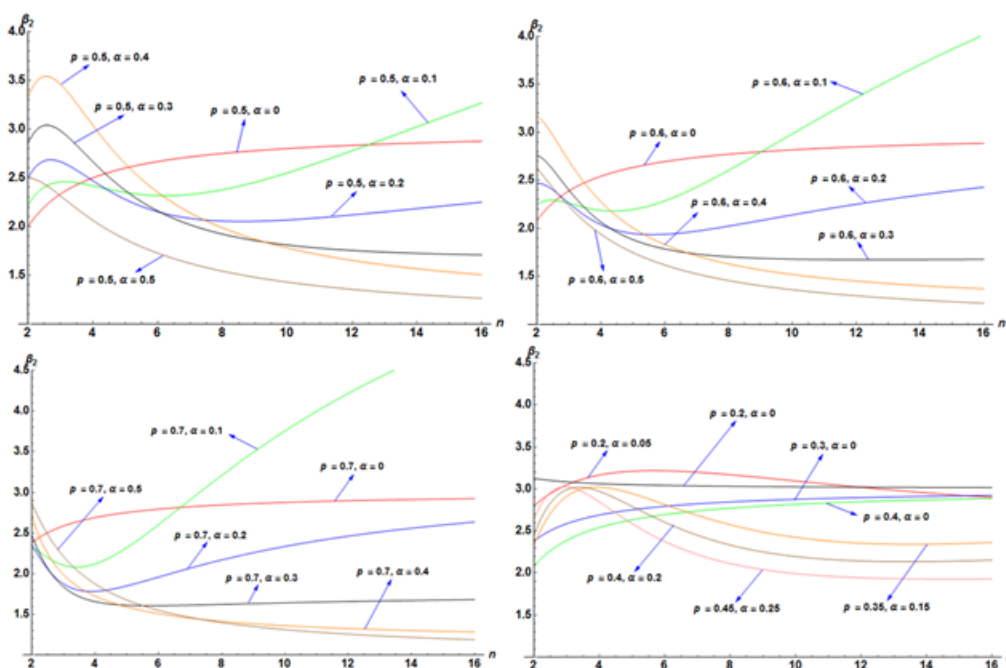

Fig 5. β_2 against n for different values of *p and* α

From the Figure 5 it is observed that
1. As $p \geq 0.5$ and $\alpha = 0.1, \beta_2 > 3$ for large n
2. As $\alpha > 0.1, \beta_2 < 3$ for large n
Remark 2

1. As $\alpha \to 0, 0 < p < 1$, then the coefficient of kurtosis $\beta_2 \to 3$ for large n i.e. the proposed distribution tends to normal.

3.2 Probability Generating Function

Theorem 2 : If $X \sim O\,IBD(n, p, \alpha)$, then its Probability Generating Function(p.g.f), $P_x(S)$ is as follows

$$P_x(S) = \alpha s + (1-\alpha)(1-p+ps)^n \qquad (13)$$

Proof: If $X \sim O\,IBD(n,p,\alpha)$ then, the probability generating function $P_x(S)$

$$P_x(S) = E\left(S^x\right)$$
$$= \sum_{x=0}^{n} P(X = x)s^x$$
$$= \alpha s + (1 - \alpha) \sum_{x=0}^{n} \binom{n}{x} (ps)^x (1 - p)^{n-x}$$
$$= \alpha s + (1 - \alpha)(1 - p + ps)^n$$

Proved.

Remark 3

Putting $S = e^t$ in equation (13), the Moment Generating Function (m.g.f), $M_x(t)$ of OIBD(n,p,α) is as follows

$$M_x(t) = \alpha e^t + (1 - \alpha)\left(1 - p + pe^t\right)^n$$

Remark 4

Putting $S = e^t$ in equation (13), the Characteristic Function, $\varphi_x(t)$ of $OIBD(n,p,\alpha)$ is as follows

$$\varphi_x(t) = \alpha e^{it} + (1 - \alpha)\left(1 - p + pe^{it}\right)^n$$

3.3 Cumulative Distribution Function (CDF)

Theorem 3: If $X \sim O\,IBD(n,p,\alpha)$, *then its CDF of X is as follows*

$$F(x) = P(X \leq x) = \alpha + (1 - \alpha)I_q(n - x, 1 + x)$$

Proof: If $X \sim O\,IBD(n,p,\alpha)$, then its CDF is as follows

$$F(x) = P(X \leq x)$$
$$= \sum_{t=0}^{x} P(X = t)$$
$$= \alpha + \sum_{t=0}^{x} (1 - \alpha) \binom{n}{t} p^t q^{n-t}$$
$$= \alpha + (1 - \alpha) \sum_{t=0}^{x} \binom{n}{t} p^t q^{n-t}$$
$$= \alpha + (1 - \alpha)I_q(n - x, 1 + x)$$

Where $q = 1 - p$ and $\sum_{t=0}^{x} \binom{n}{t} p^t q^{n-t} = I_q(n - x, 1 + x)$

Where $I_q(n - x, 1 + x)$ is an incomplete beta function.

Proved.

The plots of CDF of $OI\,BD(n,p,\alpha)$ with different choice of parameters n, p and α are provided in Figure 6.
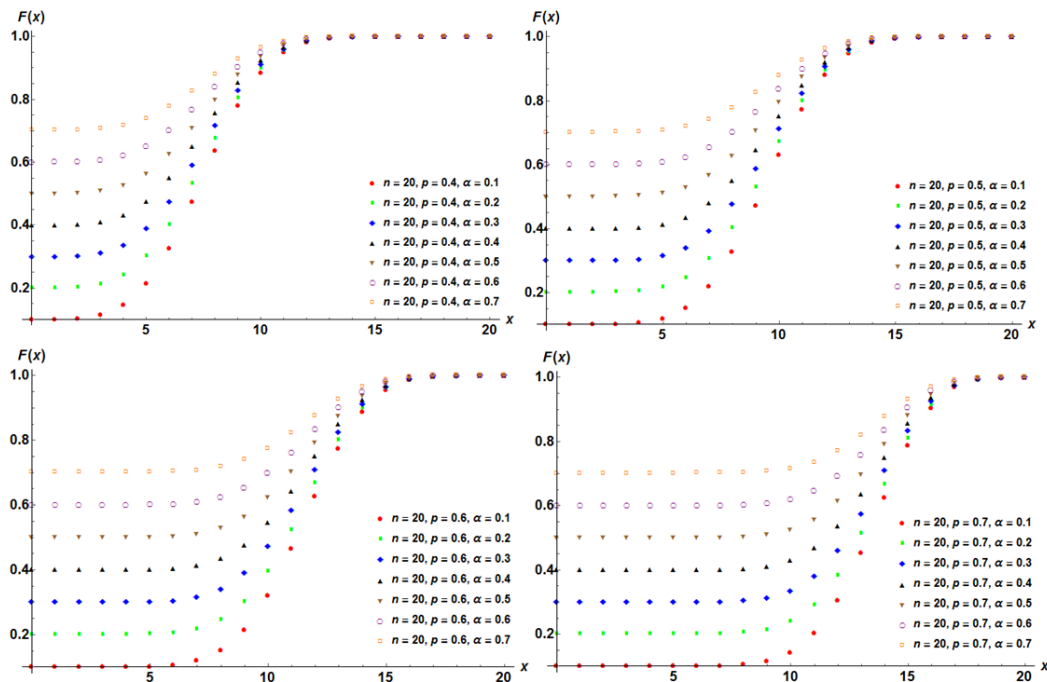
Fig 6. *Shape of CDF with different choice of parameters n, p and α*

4 Reliability Characteristics

4.1 Survival Function (SF)

Theorem 4 : If $X \sim OIBD(n, p, \alpha)$, then its Survival Function (SF) of X is as follows

$$S(x) = P(X \geq x) = \alpha + (1 - \alpha)I_p(x, n - x + 1) \tag{14}$$

Proof: If $X \sim O\,IBD(n, p, \alpha)$, then its Survival Function *(SF)* is as follows

$$S(x) = P(X \geq x)$$

$$= \sum_{i=x}^{n} P(X = t)$$

$$= \alpha + \sum_{t=x}^{n}(1 - \alpha)\left(\begin{array}{c} n \\ t \end{array}\right) p^t q^{n-t}$$

$$= \alpha + (1 - \alpha)\sum_{i=x}^{n}\left(\begin{array}{c} n \\ t \end{array}\right) p^t q^{n-t}$$

$$= \alpha + (1 - \alpha)I_p(x, n - x + 1)$$

Where $q = 1 - p$ and $\sum_{t=x}^{n}\left(\begin{array}{c} n \\ t \end{array}\right) p^t q^{n-t} = I_p(x, n - x + 1)$

Where $I_p(x, n - x + 1)$ is an incomplete beta function.

Proved.

The plots of Survival Function (SF) of $OIBD(n, p, \alpha)$ with different choice of parameters n, p and α are provided in Figure 7.

4.2 Failure Rate (FR)

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from one-inflated binomial distribution as given by (4)
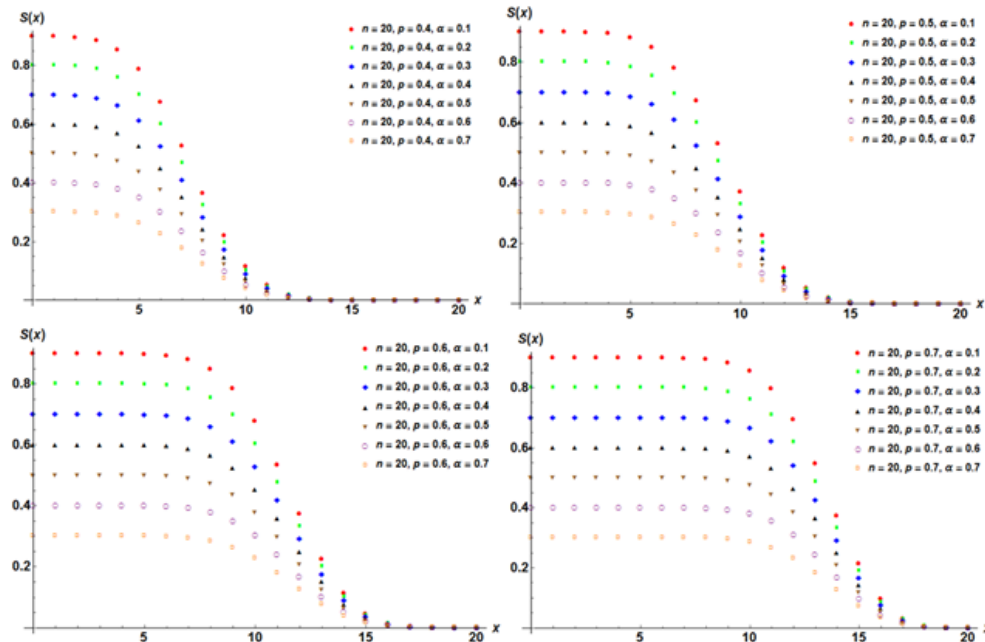
Fig 7. *Shape of SF with different choice of parameters n, p and α*

Define Y be the number of $X_i s$ taking the value 1. Then equation (4) can be inscribe as follows

$$P(X = x_i) = \left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right]^Y \left[(1-\alpha)\begin{pmatrix} n \\ x_i \end{pmatrix} p^{x_i}(1-p)^{n-x_i}\right]^{1-Y}$$

and using s(x) from equation (14)

The failure rate (FR) of $OIBD(n, p, \alpha)$ is given by

$$R(x) = \frac{p(x)}{S(x)} = \frac{\left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right]^Y \left[(1-\alpha)\begin{pmatrix} n \\ x_i \end{pmatrix} p^{x_i}(1-p)^{n-x_i}\right]^{1-Y}}{\alpha + (1-\alpha)I_p(x, n-x+1)}$$

The plots of Failure Rate (FR) of $OIBD(n, p, \alpha)$ with different choice of parameters n, p and α are provided in Figure 8.

5 Parameter Estimation

5.1 Method of Moment Estimation (MM)

The parameters p and α of (4) can be obtained using the method of moments as follows:

Considering the first two moments from Equations (6) and (7)

From Equations (6)

$$\alpha = \frac{\mu'_1 - np}{1 - np} \tag{15}$$

From Equation (7)

$$\mu'_2 = \alpha + (1-\alpha)n^2 p^2 + (1-\alpha)np - (1-\alpha)np^2$$

Using Equation (6)

$$\mu'_2 = \mu'_1 + (1-\alpha)\left(n^2 p^2 - np^2\right) \tag{16}$$
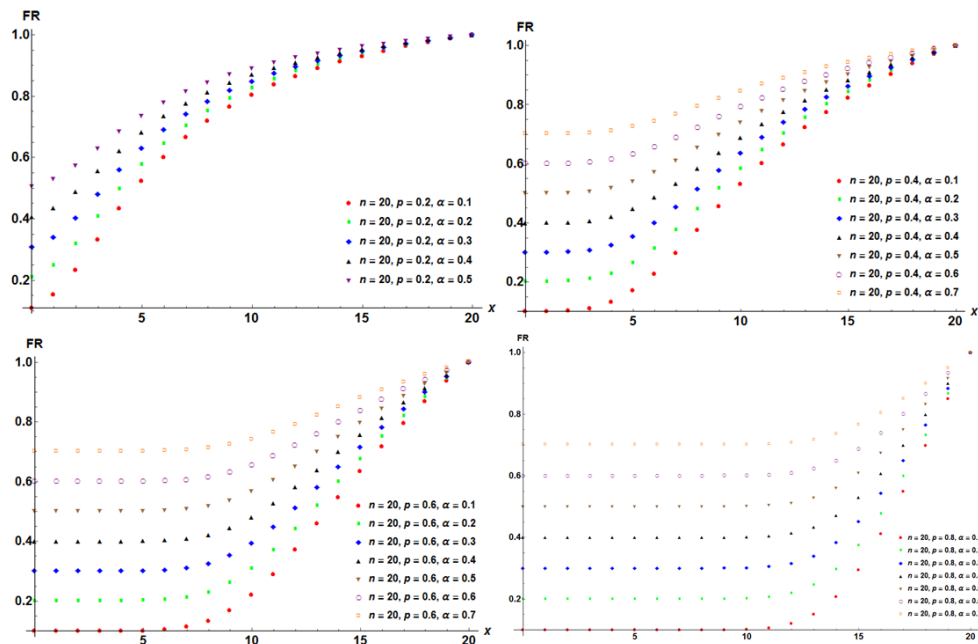
Fig 8. Shape of FR with different choice of parameters n, p and α

Putting the value of Equation (15) to Equation (16)

$$\mu_2' - \mu_1' = (n^2 p^2 - np^2)\left(1 - \frac{\mu_1' - np}{1 - np}\right)$$

$$\frac{\mu_2' - \mu_1'}{1 - \mu_1'} = \frac{n^2 p^2 - np^2}{1 - np} = M \text{ (say)}$$

Then,

$$M = \frac{n^2 p^2 - np^2}{1 - np}$$

$$p^2\left(n^2 - n\right) + npM - M = 0 \tag{17}$$

Solving the quadratic equation (17), we can estimate the value of p. The value of p have been used in equation (15) to estimate the value of α

5.2 Maximum Likelihood Estimation (MLE)

The parameters p and α of (4) can be obtained using the method of maximum likelihood as follows:

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from one-inflated binomial distribution as given by (4) and let for $i = 1, 2, 3, \ldots, n$

$$a_i = \begin{cases} 1; & \text{if } x_i = 1 \\ 0; & \text{otherwise} \end{cases}$$

then for $i = 1, 2, 3, \ldots, n$, (4) can be inscribed as follows

$$P\left(X = x_i\right) = \left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right]^{a_i}\left[(1-\alpha)\begin{pmatrix} n \\ x_i \end{pmatrix}p^{x_i}(1-p)^{n-x_i}\right]^{1-a_i}$$

Hence the likelihood function; $L = L(p, \alpha; x_1, x_2, x_3, \ldots\ldots, x_n)$ will be

$$L = \prod_{i=1}^{n} \left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right]^{a_i} \left[(1-\alpha)\binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}\right]^{1-a_i}$$

$$= \left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right]^{n_0} \prod_{i=1}^{n} \left[(1-\alpha)\binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}\right]^{b_i}$$

Where $b_i = 1 - a_i$ and $n_0 = \sum_{i=1}^{n} a_i$. Note that n_0 represents, respectively the number of one's in the sample.

Therefore

$$\log L = n_0 \log \left[\alpha + n(1-\alpha)p(1-p)^{n-1}\right] + (n-n_0)\log(1-\alpha) + \sum_{i=1}^{n} b_i \log \binom{n}{x_i}$$

$$+ \sum_{i=1}^{n} b_i x_i \log p + \sum_{i=1}^{n} b_i (n-x_i) \log(1-p)$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{n_0 \left(1 - np(1-p)^{n-1}\right)}{\alpha + n(1-\alpha)p(1-p)^{n-1}} - (n-n_0)\left(\frac{1}{1-\alpha}\right) \tag{18}$$

$$\frac{\partial \log L}{\partial p} = n_0 \left[\frac{n(1-\alpha)(1-p)^{n-2}(1-np)}{\alpha + n(1-\alpha)p(1-p)^{n-1}}\right] + \frac{\sum_{i=1}^{n} b_i x_i}{p} - \frac{\sum_{i=1}^{n} b_i (n-x_i)}{1-p} \tag{19}$$

Now, letting $\frac{\partial}{\partial \alpha} \log L = 0$, from Equation (18) that

$$1 - \alpha = \frac{n - n_0}{\dfrac{n_0 \left(1 - np(1-p)^{n-1}\right)}{\alpha + n(1-\alpha)p(1-p)^{n-1}}} \tag{20}$$

$$Let, \ p_1 = \alpha + n(1-\alpha)p(1-p)^{n-1} \tag{21}$$

Setting $\frac{\partial \log L}{\partial p} = 0$, from Equation (19) and using Equation (21)

$$n_0 \left[\frac{n(1-\alpha)(1-p)^{n-2}(1-np)}{p_1}\right] + \frac{\sum_{i=1}^{n} b_i x_i}{p} - \frac{\sum_{i=1}^{n} b_i (n-x_i)}{1-p} = 0 \tag{22}$$

Now, if we replace p_1 by their sample relative frequencies, which is by their sample estimate, the proportion of one's in the sample, i.e. $\hat{p}_1 = n_0/n$, respectively then Equations (20) and (22) reduce to

$$1 - \alpha = \frac{n - n_0}{n\left(1 - np(1-p)^{n-1}\right)} \tag{23}$$

and

$$n^2(1-\alpha)(1-p)^{n-2}(1-np) + \frac{\sum_{i=1}^{n} b_i x_i}{p} - \frac{\sum_{i=1}^{n} b_i (n-x_i)}{1-p} = 0 \tag{24}$$

Using Equation (23), Equation (24) reduces to(25)

$$\left(\frac{\sum_{i=1}^{n} b_i x_i}{p(1-p)} - \frac{n\sum_{i=1}^{n} b_i}{1-p}\right)\left(1 - np(1-p)^{n-1}\right) + n(n-n_0)(1-p)^{n-2}(1-np) = 0 \tag{25}$$

$$A(p) = 0$$

Where

$$A(p) = \left(\frac{\sum_{i=1}^{n} b_i x_i}{p(1-p)} - \frac{n\sum_{i=1}^{n} b_i}{1-p}\right)\left(1 - np(1-p)^{n-1}\right) + n(n-n_0)(1-p)^{n-2}(1-np)$$

Hence Equation (25) can be solved by any numerical procedure say, Newton Rapson, to obtain $\hat{p}$ numerically,

i: e. $A(\hat{p}) = 0$.

Similarly, using Equation (23), α can be estimated

$$\hat{\alpha} = 1 - \frac{n - n_0}{n\left(1 - np(1-p)^{n-1}\right)}$$

Therefore, the maximum likelihood estimates (MLE) of the parameters p and α are given by solving Equation (25) numerically to find $\hat{p}$, with $\widehat{\alpha}$ given by Equation (26) respectively.

When likelihood functions have convoluted frames, Expectation Maximization (EM) algorithm can be used also as a simplest alternative method for finding the MLEs of the parameters.

For the reckoning of the asymptotic variance covariance matrix of the estimates the second order differentiations of the log-likelihood function are furnish here

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -n_0 \left[\frac{\left(1 - np(1-p)^{n-1}\right)^2}{\left(\alpha + n(1-\alpha)p(1-p)^{n-1}\right)^2} \right] - (n - n_0)\left\{ \frac{1}{(1-\alpha)^2} \right\}$$

$$\frac{\partial^2 \log L}{\partial p^2} = n_0 \frac{\left[\begin{array}{c} \left(\alpha + n(1-\alpha)p(1-p)^{n-1}\right)\left[n(n-1)(np-2)(1-\alpha)(1-p)^{n-3}\right] \\ - \left(n(1-\alpha)(1-p)^{n-2}(1-np)\right)^2 \end{array} \right]}{\left(\alpha + n(1-\alpha)p(1-p)^{n-1}\right)^2}$$

$$- \frac{\sum_{i=1}^{n} b_i x_i}{p^2} - \frac{\sum_{i=1}^{n} b_i(n-x_i)}{(1-p)^2}$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial p} = n_0 \left[\frac{n(1-p)^{n-2}(np-1)}{\left(\alpha + n(1-\alpha)p(1-p)^{n-1}\right)^2} \right]$$

The asymptotic variance covariance matrix of the maximum likelihood estimators of p and α for $OIBD(n, p, \alpha)$, can be acquired by inverting the Fisher information matrix (I), given by

$$I = \left[\begin{array}{cc} E\left(-\frac{\partial^2 \log L}{\partial p^2}\right) & E\left(-\frac{\partial^2 \log L}{\partial p \partial \alpha}\right) \\ E\left(-\frac{\partial^2 \log L}{\partial \alpha \partial p}\right) & E\left(-\frac{\partial^2 \log L}{\partial \alpha^2}\right) \end{array} \right]$$

The ingredient of the above Fisher information matrix can be acquired as

$$E\left(-\frac{\partial^2 \log L}{\partial p^2}\right) = \left(-\frac{\partial^2 \log L}{\partial p^2}\right)\bigg|_{p=\hat{p}, \alpha=\hat{\alpha}, n=n}$$

The asymptotic distribution of the maximum likelihood estimator $(\hat{p}, \hat{\alpha})$ is given by

$$\sqrt{n}\left(\begin{array}{c} \widehat{p} \\ \widehat{\alpha} \end{array} \right)_{MLE} \xrightarrow{L} AN\left(\left(\begin{array}{c} \widehat{p} \\ \widehat{\alpha} \end{array} \right), I^{-1} \right), \text{ as } n \to \infty$$

6 Simulation study

In this section a simulation study has been conducted to see the performance of the estimated parameters. Here, to generate random numbers X from $OIBD(n, p, \alpha) = OIBD(20, p, \alpha)$ we have applied acceptance rejection sampling[21]. By applying this method random samples are generated of size n=100 and 200 with different combination of true values of parameters p and α and finally, MLEs are computed using EM algorithm of R software. Bias and MSE of the parameters given in the Table 1 are

calculated using the following formulae.

$$\text{Bias}(\widehat{\boldsymbol{\theta}}) = E(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$$

$$\text{MSE}(\widehat{\boldsymbol{\theta}}) = E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2$$

$$\widehat{\boldsymbol{\theta}} = \text{ estimated parameter and } \boldsymbol{\theta} = \text{ true parameter}$$

$$\boldsymbol{\theta} = (p, \alpha), \quad \widehat{\boldsymbol{\theta}} = (\widehat{p}, \widehat{\alpha})$$

Here, $r(= \text{ mumber of replication }) = 1000$

From the values of the MSE and biases of the simulation study given in Table 1, it is observed that as the sample size increase the estimated bias an MSE are also gradually decreases which is as expected.

Table 1. Results of Simulation

$\alpha\downarrow$	$p \rightarrow$ $n \rightarrow$	0.5		0.6		0.7		0.8	
		100	200	100	200	100	200	100	200
0.1	Bias($\hat{p}$)	0.0055	0.0052	0.0058	0.0058	0.0058	0.0059	0.0057	0.0058
	MSE($\hat{p}$)	0.0002	0.0001	0.0002	0.0001	0.0002	0.0001	0.0001	0.0001
	Bias($\widehat{\alpha}$)	0.0005	-0.0003	0.0005	-0.0003	0.0005	-0.0003	0.0005	-0.0003
	MSE($\widehat{\alpha}$)	0.0009	0.0004	0.0009	0.0004	0.0009	0.0004	0.0009	0.0004
0.2	Bias($\hat{p}$)	0.0123	0.0123	0.0129	0.0127	0.0129	.0127	0.0129	0.0127
	MSE($\hat{p}$)	0.0003	0.0002	0.0003	0.0002	0.0003	0.0002	0.0003	0.0002
	Bias($\widehat{\alpha}$)	-0.0001	0.0001	-0.0002	0.0001	-0.0002	0.0001	-0.0002	0.0001
	MSE($\widehat{\alpha}$)	0.0015	0.0007	0.0015	0.0008	0.0015	0.0008	0.0015	0.0008
0.3	Bias($\hat{p}$)	0.0215	0.0213	0.0219	0.0218	0.0219	0.0219	0.0219	0.0218
	MSE($\hat{p}$)	0.0007	0.0005	0.0007	0.0005	0.0006	0.0005	0.0006	0.0005
	Bias($\widehat{\alpha}$)	0.0002	-0.0001	0.0002	-0.0001	0.0002	-0.0001	0.0002	-0.0001
	MSE($\widehat{\alpha}$)	0.0020	0.0009	0.0020	0.0009	0.0020	0.0009	0.0020	0.0009
0.4	Bias($\hat{p}$)	0.0340	0.0331	0.0339	0.0340	0.0339	0.0340	0.0339	0.0339
	MSE($\hat{p}$)	0.0014	0.0012	0.0014	0.0012	0.0014	0.0012	0.0013	0.0012
	Bias($\widehat{\alpha}$)	0.0004	-0.0002	0.0004	-0.0002	0.0004	-0.0002	0.0004	-0.0002
	MSE($\widehat{\alpha}$)	0.0024	0.0011	0.0024	0.0011	0.0024	0.0011	0.0024	0.0011
0.5	Bias($\hat{p}$)	0.0507	0.0501	0.0512	0.0510	0.0513	0.0510	0.0512	0.0509
	MSE($\hat{p}$)	0.0029	0.0027	0.0030	0.0028	0.0029	0.0027	0.0029	0.0027
	Bias($\widehat{\alpha}$)	0.0001	0.0006	0.0001	0.0006	0.0001	0.0006	0.0001	0.0006
	MSE($\widehat{\alpha}$)	0.0024	0.0012	0.0024	0.0011	0.0024	0.0012	0.0024	0.0012

7 Real-life examples

The researchers illustrate the application of OIBD with real data set with the inflated count of one and compare it with binomial distribution and ZIBD. The data set consists of the sex composition of child to mothers with parity two in Assam and Meghalaya of India. The data used to fit the model are taken from NFHS-IV conducted in all the states and Union territories of India during 2015-2016 (dhsprogram.com). The prime reason for selecting the two states is the variation in the prevailing social system. Assam being a patriarchal society, a strong son preference exists[22]; whereas, the social system of Meghalaya is based on matriarchal where preference for daughters is observed[23]. Past research showed that sex preference has a positive effect on fertility and contraceptive practices[24–26]. Studies also showed that it is imperative to have one son to continue family in India with a patriarchal family system[27]. Considering these facts, for Assam and Meghalaya, the birth of male and female child are studied with mothers of parity 2. Singh et al.[11] show that sex preference does not exist among mothers with higher parity, that

is why, in this study, mothers with lower parity is considered to study gender preference. The study considers the women who had ever been married in the age group 15-49 and also the birth interval of the considered women is greater than five years (i.e. who completed their family size). Tables 2 and 3 shows the number of son and daughters born to mothers in the states of Assam and Meghalaya. The count one is inflated in both states, so OIBD is fitted and compared with Binomial distribution and ZIBD.

Table 2. Distribution of number of male children out of observed number women in Assam

Number of male births	Observed number of women	$B(n,p)$	$ZIBD(n,p,\alpha)$	$OIBD(n,p,\alpha)$
0	525	660.751877	660.751877	524.980182
1	2016	1744.630246	1744.630246	2015.952960
2	1016	1151.617877	1151.617877	1016.066858
Total	3557	3557	3557	3557
Estimated parameters		$\hat{p} = 0.5690$	$\hat{p} = 0.5690$	$\hat{p} = 0.5818$
			$\hat{\alpha} = 1.0659 \times 10^{-32}$	$\hat{\alpha} = 0.1561$
	$\log L$	-3465.67	-3465.67	-3422.25
	KS test	0.038165	0.038165	0.000019
	p-value	0.006	0.006	0.999
	AIC	6933.34	6935.34	6848.50
	BIC	6932.4386	6933.5372	6846.6972

Table 3. Distribution of number of male children out of observed number women in Meghalaya

Number of female births	Observed number of women	$B(n,p)$	$ZIBD(n,p,\alpha)$	$OIBD(n,p,\alpha)$
0	129	155.234464	154.987176	129.024056
1	360	307.995072	307.996848	359.999856
2	127	152.770464	153.015976	126.976088
Total	616	616	616	616
Estimated parameters		$\hat{p} = 0.4984$	$\hat{p} = 0.4984$	$\hat{p} = 0.4980$
			$\hat{\alpha} = 7.0743 \times 10^{-9}$	$\hat{\alpha} = 0.168844$
	Log L	-604.418	-604.418	-595.595
	KS test	0.042588	0.042234	0.000039
	p-value	0.327	0.333	0.999
	AIC	1210.836	1212.836	1195.19
	BIC	1209.935	1211.0332	1193.3872

Here we fitted our proposed distribution, i.e. one-inflated binomial distribution (OIBD) along with Binomial distribution (BD) and Zero-inflated binomial distribution (ZIBD)[10]. The values of log-likelihood, Akaike information criterion (AIC)[28], Bayesian information criterion (BIC)[28] and the Kolmogorov–Smirnov test (KS test) with p-values are summarized in Tables 2 and 3 for the two states. AIC and BIC are the model selection criteria and KS test for goodness of fit.

From the tables it is seen that the values of KS test, AIC and BIC of OIB distribution is smaller than BD and ZIB distribution and the expected frequencies of OIBD is closed to the observed frequencies. From Table 2, it can be observed that the proposed OIBD provides better fitting to the son preference data of Assam with inflated count of one (p-value > 0.05) whereas the Binomial distribution and ZIBD does not fits to the data. In case of Meghalaya, all the three distributions fits well with the daughter preference data with inflated count of one (Table 3), but OIBD has the highest p-value of KS test, indicating that it fits better than the other distribution.

8 Conclusion

In this study, the one-inflated binomial distribution (OIBD) is introduced and studied its distributional properties and reliability characteristics. The parameters are estimated using the method of maximum likelihood estimators. A simulation study has been conducted to see the behaviour of the MLEs. The appropriateness of the fitting distribution is carried out based on the goodness of fit test and some information criteria. To the real-life data set having a higher frequency of count one, it can be observed that the proposed one-inflated binomial distribution (OIBD) provides better fitting than other competitor distributions.

A future prospect of this study may be to proposed an OIBD regression model. This study is centered on only the mixture of over dispersed count data for one as a point mass and binomial distribution i.e. one-inflated binomial distribution (OIBD). There are many other count distributions suitable for one-inflated version, which can be done for future research.

References

1) Mallick A, Joshi R. Parameter Estimation and Application of Generalized Inflated Geometric Distribution. *Journal of Statistical Theory and Applications*. 2018;17(3):491–519. Available from: https://dx.doi.org/10.2991/jsta.2018.17.3.7.

2) Danillo X, Manoel SN, Marcelo B, Vera T. Bourguignon Marcelo and Tomazella Vera. Zero-Modified Poisson-Lindley distribution with applications in zero-inflated and zero-deflated count data. *ArXiv*. 2018. Available from: https://arxiv.org/pdf/1712.04088.pdf.

3) Winai B, Siriporn S, Prasit P. The zero-inflated negative binomial-Erlang distribution: An application to highly pathogenic avian influenza H5N1 in Thailand. *Songklanakarin Journal of Science and Technology*. 2018;40(6):1428–1436. Available from: https://rdo.psu.ac.th/sjstweb/journal/40-6/24.pdf.

4) Neyman J. On a New Class of "Contagious" Distributions, Applicable in Entomology and Bacteriology. *The Annals of Mathematical Statistics*. 1939;10(1):35–57. Available from: https://dx.doi.org/10.1214/aoms/1177732245.

5) Feller W. On a General Class of "Contagious" Distributions. *The Annals of Mathematical Statistics*. 1943;14(4):389–400. Available from: https://dx.doi.org/10.1214/aoms/1177731359.

6) Gupta PL, Gupta RC, Tripathi RC. Inflated modified power series distributions with applications. *Communications in Statistics - Theory and Methods*. 1995;24(9):2355–2374. Available from: https://dx.doi.org/10.1080/03610929508831621.

7) Murat M, Szynal D. Non–zero inflated modified power series distributions. *Communications in Statistics - Theory and Methods*. 1998;27(12):3047–3064. Available from: https://dx.doi.org/10.1080/03610929808832272.

8) Lambert D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*. 1992;34(1):1–14. Available from: https://dx.doi.org/10.2307/1269547.

9) Martin R, John H, Clarice GBD. A score test for testing a zero-inated poisson regression model against zero-inated negative binomial alternatives. *Biometrics*. 2001;57:219–223. Available from: https://doi.org/10.1111/j.0006-341X.2001.00219.x.

10) Hall DB. Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*. 2000;56(4):1030–1039. Available from: https://dx.doi.org/10.1111/j.0006-341x.2000.01030.x.

11) Brijesh PS, S M, Punee KG, K. A probability model for sex composition of children in the presence of son preference. *Demography India*. 2015;44(1&2):50–57. Available from: https://demographyindia.in/article_document/5/2443550-57.pdf.

12) Beckett S, Jee J, Ncube T, Pompilus S, Washington Q, Singh A, et al. Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve, a Journal of Mathematics*. 2014;7:751–767. Available from: https://dx.doi.org/10.2140/involve.2014.7.751.

13) Nanjundan G, Sadik P. A characterization of Zero-inflated binomial model. *International journal of mathematics and computer research*. 2015;3(10):1188–1190. Available from: https://ijmcr.in/index.php/ijmcr/article/view/142.

14) Suresh R, Nanjundan G, Nagesh S, Pasha S. On a Characterization of Zero-Inflated Negative Binomial Distribution. *Open Journal of Statistics*. 2015;05(06):511–513. Available from: https://dx.doi.org/10.4236/ojs.2015.56053.

15) Nanjundan G, Pasha S. A Note on the Characterization of Zero-Inflated Poisson Model. *Open Journal of Statistics*. 2015;05(02):140–142. Available from: https://dx.doi.org/10.4236/ojs.2015.52017.

16) Abdulrazak ARS. On the Zero-One Inflated Poisson distribution. *International journal of Statistical Distributions and Applications*. 2016;2(4):42–48. Available from: https://doi.org/10.11648/j.ijsd.20160204.11.

17) Mwalili SM, Lesaffre E, Declerck D. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical Methods in Medical Research*. 2008;17(2):123–139. Available from: https://dx.doi.org/10.1177/0962280206071840.

18) Abdulrazak ARS. Moment Estimators of the parameters of Zero-One Inflated Negative Binomial Distribution. International journal of mathematical, computational, physical, electrical and computer engineering. *World Academy of Science, Engineering and Technology*. 2017;11(1):38–41. Available from: https://doi.org/10.5281/zenodo.1129117.

19) Sakthivel KM, Rajitha C. Estimation of Zero-Inflation Parameter in Zero-Inflated Poisson Model. *International Journal of Mathematics Trends and Technology*. 2018;56(2):135–140. Available from: https://doi.org/10.14445/22315373/IJMTT-V56P519.

20) Jornsatian C, Bodhisuwan W. Zero-one inflated negative binomial - beta exponential distribution for count data with many zeros and ones. *Communications in Statistics - Theory and Methods*. 2021;p. 1–15. Available from: https://dx.doi.org/10.1080/03610926.2021.1898642.

21) Sheldon R. Simulation. Elsevier Academic Press. 2012. Available from: https://www.elsevier.com/books/simulation/ross/978-0-12-415825-2.

22) National Family Health Survey (NFHS-4), India, 2015-16: Assam. Mumbai: IIPS. . Available from: http://rchiips.org/nfhs/NFHS-4Reports/Assam.pdf.

23) National Family Health Survey (NFHS-4), India, 2015-16: Meghalaya. Mumbai: IIPS. . Available from: http://rchiips.org/nfhs/NFHS-4Reports/Meghalaya.pdf.

24) Das N. Sex preference and fertility behavior: A study of recent Indian data. *Demography*. 1987;24(4):517–530. Available from: https://dx.doi.org/10.2307/2061389.

25) Malhotra A, Vanneman R, Kishor S. Fertility, Dimensions of Patriarchy, and Development in India. *Population and Development Review*. 1995;21(2):281–305. Available from: https://dx.doi.org/10.2307/2137495.

26) Kulkarni PM. Gender preference contraceptive prevalence: Evidence of regional variations. *Economic and Political Weekly*. 1999;34(42-43):3058–3062. Available from: https://www.jstor.org/stable/4408537.

27) Arvind KS, Yadava R, Tiwari AK. Measuring Son Preference Through Number of Children Born. *Demography India*. 2018;47(2):67–82. Available from: http://demographyindia.in/article_document/73/817706.pdf.

28) Kenneth PB, David RA. Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods & Research*. 2004;33(2):261–304. Available from: https://doi.org/10.1177/0049124104268644.