International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:11, No:1, 2017

Moment Estimators of the Parameters of Zero-One Inflated Negative Binomial Distribution

Rafid Saeed Abdulrazak Alshkaki

Abstract**—**In this paper, zero-one inflated negative binomial distribution is considered, along with some of its structural properties, then its parameters were estimated using the method of moments. It is found that the method of moments to estimate the parameters of the zero-one inflated negative binomial models is not a proper method and may give incorrect conclusions.

Keywords**—**Zero one inflated models, negative binomial distribution, moments estimator, non-negative integer sampling.

I. Introduction

THE negative binomial distribution (NBD), which is a well-known discrete distribution, has so many real live applications. It is the distribution in which we are looking for the number of failures observed until a specific number of successes occurs, see Johnson et al. [1] for further details. Recently, models based on zero-inflated negative binomial distribution (ZINBD) were studied by many researchers. In particular, Sharma and Landge [2] used the ZINB regression for modeling heavy vehicle crash rate on Indian rural highway. Saengthong et al. [3] introduced the ZINB – crack distribution consisting of a mixture of Bernoulli distribution and NBD, which is an alternative distribution for the excessive zero counts and overdispersion, and studied some of its properties and parameter estimates. Fang et al. [4] considered a hierarchical regression-based approach using a ZINB mixed model to evaluate associations with disease state while adjusting for potential confounders for two organisms of interest from a study of human microbiota sequence data in oesophagitis. Preisser et al. [5] proposed a marginalized ZINB regression model to the evaluation of a school-based fluoride mouth rinse program on dental caries in 677 children. Nanjundan and Nika [6] discuss the maximum likelihood estimation (MLE) and the method of moments estimation (ME) of the parameters of the zero-inflated power series models, and compared them asymptotically.. For further detailed details of the NBD see Johnson et al. [1], and for its recent real-life applications and inflated form see Banik and Kibria [7].

In this paper, we introduce in Section II, the definition of the NBD and its zero-one inflated form, followed in Section III by some of its structural properties, namely, its mean, variance, and generating functions. Then in Section IV, we consider ME method of the parameters, followed by numerical examples in Section V, representing five different cases of NBDs in order to check the accuracy of this method.

Rafid Saeed Abdulrazak Alshkaki is an associate professor of statistics at Ahmed Bin Mohammed Military College, Doha, Qatar (e-mail; rafid@abmmc.edu.qa).

II. The NBDs and its Zero-One Inflated Form

Let $k > 0$ and $\theta \in (0,1)$, then the discrete random variable (rv) X having a probability mass function (pmf);

$$P(X = x) = \begin{cases} \binom{k+x-1}{x}\theta^x (1-\theta)^k, & x = 0,1,2,3, \\ 0, & \text{otherwise}, \end{cases} \quad (1)$$

is said to have a NBD with parameters k and θ. We will denote that by writing $X \sim \text{NBD}(k,\theta)$. See Johnson et al. [1], for other forms and parameterizations of the NBD.

Let the rv $X \sim \text{NBD}(k,\theta)$, as given in (1), let $\alpha \in (0,1)$ be a proportion added to the proportion of zeros of X, and let $\beta \in (0,1)$ be another proportion added to the proportion of ones of X, such that $0 < \alpha + \beta < 1$, then the rv Z defined by;

$$P(Z = z) = \begin{cases} \alpha + (1-\alpha-\beta)(1-\theta)^k, & z = 0 \\ \beta + k(1-\alpha-\beta)\,\theta\,(1-\theta)^k, & z = 1 \\ (1-\alpha-\beta)\binom{k+z-1}{z}\theta^z (1-\theta)^k, & z = 2,3,4,\dots \\ 0, & \text{otherwise}, \end{cases} \quad (2)$$

is said to have a zero-one inflated negative binomial distribution (ZOINBD) with parameters k, θ , α and β. This will denote that by writing $Z \sim \text{ZOINBD}(k,\theta;\alpha,\beta)$. We note that when α=0, then (2) reduces to the ZINBD. Similarly, when α=0 and β=0, then (2) reduces to the NBD, see Alshkaki [8] for further details. Although, it does not fit in the nature of the supposed model, the inflation parameters α and β may also take negative values providing that $\alpha \in \left(\max\left\{-1, -(1-\beta)\frac{(1-\theta)^k}{1-(1-\theta)^k}\right\}, 0\right)$ and $\beta \in \left(\max\left\{-1, -(1-\alpha)\frac{k\theta\,(1-\theta)^k}{1-k\theta\,(1-\theta)^k}\right\}, 0\right)$ without violating that (2) as being a pmf. This situation represents the excluding proportion of zeros and ones, respectively, from the standard model given by (1).

III. Some Structural Properties

Let $X \sim \text{NBD}(k,\theta)$ then its mean, variance and probability generating function (pgf) are given by, $\frac{k\theta}{1-\theta}$, $\frac{k\theta}{(1-\theta)^2}$, and $\left(\frac{1-\theta}{1-\theta t}\right)^k$, respectively. Let the rv $Z \sim \text{ZOINBD}(k,\theta;\alpha,\beta)$, then it is easy to find that;

$$E(Z) = \beta + (1-\alpha-\beta)\left(\frac{k\theta}{1-\theta}\right) = \beta + (1-\alpha-\beta)E(X),$$

and that

$$Var(Z) = \beta(1-\beta) + (1-2\beta)\left(\frac{k\theta}{1-\theta}\right) + [1-(\alpha+\beta)^2]\left(\frac{k\theta}{1-\theta}\right)^2 Var(Z) = \beta(1-\beta) + (1-2\beta)\left(\frac{k\theta}{1-\theta}\right) + [1-(\alpha+\beta)^2]\left(\frac{k\theta}{1-\theta}\right)^2 = \beta(1-\beta) + (1-2\beta)E(X) + \theta[1-(\alpha+\beta)^2]Var(X) = \beta(1-\beta) + (1-2\beta)E(X) + \theta[1-(\alpha+\beta)^2]Var(X)$$

The pgf $G_Z(s)$ and the moment generating function (mgf) $M_Z(t)$, are respectively:

$$G_Z(t) = \alpha + \beta t + (1-\alpha-\beta)\left(\frac{1-\theta}{1-\theta t}\right)^k = \alpha + \beta t + (1-\alpha-\beta)G_X(t),$$

and

$$M_Z(t) = \alpha + \beta e^t + (1-\alpha-\beta)\left(\frac{1-\theta}{1-\theta e^t}\right)\alpha + \beta e^t + (1-\alpha-\beta)M_X(t)$$

IV. Moment Estimators

Using the pgf, mgf or obtaining them directly, the first four distribution moments about the origin for the ZOINBD can be found to be,

$$\mu_1' = \beta + (1-\alpha-\beta)\left(\frac{k\theta}{1-\theta}\right)$$

$$\mu_2' = \beta + (1-\alpha-\beta)\left[\frac{k\theta(1+k\theta)}{(1-\theta)^2}\right]$$

$$\mu_3' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^3}[1+(1+3k)\theta+k^2\theta^2]$$

$$\mu_4' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^2}[1+6\theta+7k\theta] + (1-\alpha-\beta)k(k+1)(k+2)\frac{\theta^3}{(1-\theta)^4}[6-3\theta+k\theta]\mu_4' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^2}[1+6\theta+7k\theta] + (1-\alpha-\beta)k(k+1)(k+2)\frac{\theta^3}{(1-\theta)^4}[6-3\theta+k\theta]$$

Let $z_1, z_2, \dots, z_n$ be a random sample from ZOINBD as given by (2), and let,

$$m_k' = \frac{\sum_{i=1}^n z_i^k}{n}, \quad k = 1,2,3,4.$$

be their sample moments about the origin, then the ME of the parameters k, θ, α and β are obtained by solving the following simultaneous equations:

$$m_1' = \beta + (1-\alpha-\beta)\left(\frac{k\theta}{1-\theta}\right)$$

$$m_2' = \beta + (1-\alpha-\beta)\left[\frac{k\theta(1+k\theta)}{(1-\theta)^2}\right]$$

$$m_3' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^3}[1+(1+3k)\theta+k^2\theta^2]$$

$$m_4' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^2}[1+6\theta+7k\theta] + (1-\alpha-\beta)k(k+1)(k+2)\frac{\theta^3}{(1-\theta)^4}[6-3\theta+k\theta]m_4' = \beta + (1-\alpha-\beta)\frac{k\theta}{(1-\theta)^2}[1+6\theta+7k\theta] + (1-\alpha-\beta)k(k+1)(k+2)\frac{\theta^3}{(1-\theta)^4}[6-3\theta+k\theta]$$

Let $\mu_{[r]}$ be the factorial moment of the rv Z defined by

$$\mu_{[r]} = E[Z(Z-1)(Z-2)\dots(Z-r+1)], \; r = 1,2,3,\dots,$$

then, it is easy to find that;

$$\mu_{[1]} = \beta + (1-\alpha-\beta)k\left(\frac{\theta}{1-\theta}\right) \tag{3}$$

and that for $r = 2,3,\dots,$

$$\mu_{[r]} = (1-\alpha-\beta)k(k+1)(k+2)(\dots)(k+r-1)\left(\frac{\theta}{1-\theta}\right)^r \tag{4}$$

For $r = 1,2,3,\dots,$ let $m_{[r]}'$ be the r^{th} sample factorial moments of the rv Z defined by;

$$m_{[r]}' = \frac{\sum_{i=1}^n z_i(z_i-1)(z_i-2)\dots(z_i-r+1)}{n} \tag{5}$$

Now let us, for r=1, 2, 3 and 4, equating the distributional factorial moments $\mu_{[r]}$ given by (3) and (4) with their respective sample factorial moments $m_{[r]}'$ given by (5), in order to solve these equations for the parameters k, θ, α and β. Firstly, from (4) and (5), we have that;

$$\frac{m_{[3]}'}{m_{[2]}'} = (k+2)\left(\frac{\theta}{1-\theta}\right)$$

or equivalently,

$$\hat{\theta} = \frac{m_{[3]}'}{m_{[3]}' + (\hat{k}+2)m_{[2]}'}$$

From (3), (4) and (5) we have that,

$$\frac{m_{[2]}'}{m_{[1]}' - \beta} = (k+1)\left(\frac{\theta}{1-\theta}\right)$$

from which we get that;

$$\hat{\beta} = m_{[1]}' - \frac{m_{[2]}'(1-\hat{\theta})}{(\hat{k}+1)\hat{\theta}}$$

Similarly, from (4), we have that;

$$m_{[2]}' = (1-\alpha-\beta)k(k+1)\left(\frac{\theta}{1-\theta}\right)^2$$

from which we have that;

$$\hat{\alpha} = 1 - \hat{\beta} - \frac{m'_{[2]}}{\hat{k}(\hat{k}+1)}\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right)^2$$

Finally, from (3)-(5) we have that;

$$\frac{m'_{[3]}}{m'_{[2]}} = (k+2)\left(\frac{\theta}{1-\theta}\right)$$

and that;

$$\frac{m'_{[4]}}{m'_{[3]}} = (k+3)\left(\frac{\theta}{1-\theta}\right),$$

hence,

$$\frac{m'_{[4]}\, m'_{[2]}}{\left(m'_{[3]}\right)^2} = \frac{k+3}{k+2}$$

from which we have that;

$$\hat{k} = \frac{3-2v}{v-1}$$

where;

$$v = \frac{m'_{[4]}\, m'_{[2]}}{\left(m'_{[3]}\right)^2}$$

Since;

$$m'_{[1]} = m'_1 \tag{6}$$

$$m'_{[2]} = m'_2 - m'_1 \tag{7}$$

$$m'_{[3]} = m'_3 - 3m'_2 + 2m'_1 \tag{8}$$

$$m'_{[4]} = m'_4 - 6m'_3 + 11m'_2 - 6m'_1 \tag{9}$$

Therefore, replacing the sample factorial moments in term of their respective sample moments as given in (6)-(9), give the ME of the parameters as follows:

$$\hat{k} = \frac{3\left(m'_3-3m'_2+2m'_1\right)^2-2\left(m'_4-6m'_3+11m'_2-6m'_1\right)\left(m'_2-m'_1\right)}{\left(m'_4-6m'_3+11m'_2-6m'_1\right)\left(m'_2-m'_1\right)-\left(m'_3-3m'_2+2m'_1\right)^2} \tag{10}$$

$$\hat{\theta} = \frac{m'_3-3m'_2+2m'_1}{m'_3-3m'_2+2m'_1+(\hat{k}+2)(m'_2-m'_1)} \tag{11}$$

$$\hat{\beta} = m'_1 - \frac{(m'_2-m'_1)(1-\hat{\theta})}{(\hat{k}+1)\hat{\theta}}, \tag{12}$$

$$\hat{\alpha} = 1 - \hat{\beta} - \frac{(m'_2-m'_1)}{\hat{k}(\hat{k}+1)}\left(\frac{1-\hat{\theta}}{\hat{\theta}}\right)^2 \tag{13}$$

V. Numerical Examples

In this section, five different assumed NBD cases will be used to estimate the ME parameters of the ZOINBD in order to check the accuracy of the resulting estimated frequencies using the estimates obtained by this method. Table I represents these assumed NBD cases. Firstly, for each NBD case represented in Table I, we computed theoretical exact frequencies of that case, then using these frequencies, we computed the sample moments needed to compute the ME of the parameters for each case.

TABLE I
Assumed NBD Cases

Case	Parameters			
	k	θ	α	β
1	2	0.55	0.3	0.1
2	3	0.4	0.5	0.2
3	4	0.5	0.2	0.3
4	5	0.35	0.2	0.1
5	9	0.2	0.5	0.3

Secondly, we used the computed sample moments to estimate the parameter k for each case. Table II represents the assumed and estimated parameter k, from which we found that the estimates are far away from the actual values.

Thirdly, we ignore the estimated value of k and used its assumed values, as given in Table I, and then used the computed sample moments to estimate the parameters θ, α and β, as given in (11)-(13), for each of the assumed distribution case. Table III represents the theoretical and estimated frequencies along with the model theoretical and estimated parameters, as well as testing the accuracy of these estimates using chi-squares goodness of fit test.

From Table III, by comparing between the assumed parameter values with their respective estimated values, as well as, the theoretical frequencies with its respective estimates, we can see that some of them are very close in values, and that is supported by the p-values of the chi-squares goodness of fit test that are closed to one, namely cases 2 and 5 and, while it is statistically significant in case 4, and not significant, on the other hand, in cases 1 and 3. This is also may indicate, generally, a poor method of estimation and it may lead to misleading estimates.

VI. Conclusion

Parameters estimation of the ZOINBD distributions by the method of moment estimators was considered. It is found numerically that this method is generally not an accurate method to estimate the parameters of the ZOINB models and may lead to misleading predication. Thus, further studies are needed to find an accurate method.

TABLE II
Assumed and Estimated k

Case	k	$\hat{k}$
1	2	10.76
2	3	6.02
3	4	166.75
4	5	11.82
5	9	14.91

TABLE III
ASSUMED NBD CASES WITH ITS ME PARAMETERS AND FREQUENCIES ESTIMATES AND TESTING

z_i		Case 1		Case 2		Case 3		Case 4		Case 5	
		Theo. Freq.	Est. Freq.	Theo. Freq.	Est. Freq.	Theo. Freq.	Est. Freq.	Theo. Freq.	Est. Freq.	Theo. Freq.	Est. Freq.
0		421.50	464.60	564.80	569.21	231.25	317.35	281.22	299.25	526.84	528.23
1		233.65	145.90	277.76	268.13	362.50	199.20	242.14	204.42	348.32	345.18
2		110.26	138.54	62.21	66.07	78.13	115.48	149.24	161.97	48.32	49.80
3		80.86	95.55	41.47	43.11	78.13	105.92	121.88	128.93	35.43	36.03
4		55.59	61.78	24.88	25.32	68.36	85.00	85.32	87.97	21.26	21.33
5		36.69	38.35	13.93	13.88	54.69	62.37	53.75	54.02	11.06	10.94
6		23.54	23.14	7.43	7.25	41.02	42.91	31.35	30.72	5.16	5.04
7		14.80	13.68	3.82	3.65	29.30	28.11	17.24	16.47	2.21	2.13
8		9.16	7.96	1.91	1.78	20.14	17.73	9.05	8.43	0.88	0.84
9		5.60	4.58	0.93	0.85	13.43	10.84	4.58	4.15	0.33	0.31
10+		8.36	5.99	0.84	0.75	23.05	15.09	4.23	3.67	0.19	0.17
Total		1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Parameters	k	2		3		4		5		9	
	θ	0.55	0.5173	0.4	0.392	0.5	0.459	0.35	0.341	0.2	0.197
	α	0.3	0.292	0.5	0.497	0.2	0.262	0.2	0.207	0.5	0.499
	β	0.1	-0.033	0.2	0.184	0.3	0.099	0.1	0.046	0.3	0.295
Testing	χ^2	66.937		1.166		186.095		9.743		0.098	
	df	7		7		7		7		7	
	p-val.	0.00001		0.991691		0.00001		0.203623		0.999998	

REFERENCES

[1] Johnson, N. L., Kemp, A. W. and Kotz, S. (2005) *Univariate Discrete Distributions*, Third Edition, John Wiley and Sons; New Jersey.

[2] Sharma, A. K., and Landge, V. S. (2013). Zero Inflated Negative Binomial For Modeling Heavy Vehicle Crash Rate On Indian Rural Highway. *International Journal of Advances in Engineering & Technology*, 5(2), 292-301.

[3] Saengthong, P., Bodhisuwan, W., and Thongteeraparp, A. (2015). The Zero Inflated Negative Binomial – Crack Distribution: Some Properties And Parameter Estimation. *Songklanakarin J. Sci. Technol*, 37 (6), 701-711.

[4] Fang, R., Wagner, B. D., Harris, J. K. and Fillon, S. A. (2016). Zero-inflated negative binomial mixed model: an application to two microbial organisms important in oesophagitis. *Epidemiology and Infection*, 144(11), 2447-2455.

[5] Preisser, J. S., Das, K., Long, D. L. and Divaris, K. (2016). Marginalized zero-inflated negative binomial regression with application to dental caries. *Stat Med.* 35(10), 1722-35.

[6] Nanjundan, G. and Nika, T. R. (2013). Estimation of the Parameters of the Zero-One Inflated Power Series Distributions, *International Journal of Statistika and Mathematika*, Vol. 6(3), 109-114.

[7] Banik, S. and Kibria, B. M. G. (2009). On Some Discrete Distributions and their Applications with Real Life Data. *Journal of Modern Applied Statistical Methods*, 8(2), 423-447.

[8] Alshkaki, R. S. A. (2016). On a Characterization of the Zero-One Inflated Negative Binomial *Distribution. Research Journal of Mathematical and Statistical Sciences*, Vol. 4(9), 2016, 1-3.