

On Zero-Truncated Negative Binomial with Excess Ones

ABSTRACT

In this paper, Zero-truncated negative **binomial distribution** is modified to include excess ones to improve goodness-of-fit. This is necessary when data are dispersed and zero has been eliminated from data **structurally**. However, when the ones are unduly large, the proportion of this excess must be recognized and estimated to improve the fit. This development is applied using **real** data from a national survey.

Keywords: excess ones, negative Binomial, zero-truncated, goodness-of-fit

Mathematics subject classification: 60E05, 62E15.

1 Introduction

In modeling count data of rare events, Poisson distribution is usually referred to as the based model. The probability mass function (pmf) of the Poisson distribution can be expressed as:

$$g(j|\mu) = \begin{cases} \frac{\mu^j e^{-\mu}}{j!}, & \text{for } j = 0, 1, 2, \dots \\ 0, & \text{elsewhere} \end{cases}, \quad (1)$$

where $\mu > 0$.

However, the model has a strict assumption of equi-dispersion. Hence, different generalizations of the model or mixture of the model had been proposed in literature. The Negative Binomial [1], Generalized Poisson, Com-Poisson model [2], the Negative Binomial-Sushila linear model [3], negative binomial-weighted Lindley [4] and a new Generalized Poisson mixed distribution [5] are some of the distributions proposed to account for over- and under-dispersion, which are common scenarios in count data.

When zero is missing structurally in count data, the models need to be adjusted. This leads to the development of zero-truncated models, such as zero-truncated Poisson distribution [6] and zero-truncated Negative Binomial distribution [7]. Most recently introduced are the Zero-truncated negative binomial-Erlang distribution [8], a zero-truncated distribution [9] and Lagrangian Zero-truncated Poisson Distribution [10] among others.

Suppose $g_0(j|\mu)$ is the original distribution for count data including zero counts. Then the zero-truncated version of $g_0(j|\mu)$ can be defined as

$$P(j|\mu) = \frac{g_0(j|\mu)}{1 - g_0(0|\mu)}; \quad j = 1, 2, 3, \dots; \mu > 0$$

For example, the pmf of the zero-truncated Poisson distribution can be expressed as

$$\begin{aligned} g_1(j|\mu) &= \frac{\mu^j e^{-\mu}}{j!} \div (1 - e^{-\mu}) \\ &= \frac{\mu^j}{j! (e^\mu - 1)}, \quad j = 1, 2, 3, \dots \end{aligned} \quad (2)$$

where $\mu > 0$.

Moreover, models to account for excess one in zero-truncated count data are not popularly discussed in literature, hence, the essence of this study is to fill that gap. The aim of this work is to develop zero-truncated Negative Binomial with excess ones and discuss its properties and application to a real-life dataset.

2 Material and Methods

The probability mass function (pmf) of Zero-Truncated Negative Binomial is given as:

$$f(j|k, p) = \begin{cases} \binom{k+j-1}{j} \frac{(1-p)^j p^k}{(1-p^k)}, & \text{for } j = 1, 2, \dots \\ 0, & \text{elsewhere} \end{cases} \quad (3)$$

where $0 < p < 1$ and $k \geq 1$

Definition 1: Suppose J is a random variable of Zero-Truncated Negative Binomial distribution with excess ones (ZTNBw) with parameters k, p and ω , denoted as $J \sim ZTNBw(k, p, \omega)$. The pmf is given by:

$$f(J = j|k, p, \omega) = \begin{cases} \omega + (1 - \omega)f(1|k, p), & \text{if } j = 1 \\ (1 - \omega)f(j|k, p) & \text{if } j = 2, 3, \dots \end{cases}$$

where, ω is the inflation or excess parameter.

Therefore, the pmf of Zero-Truncated Negative Binomial distribution with excess ones is expressed as

$$f(j|k, p, \omega) = \begin{cases} \omega + (1 - \omega) \frac{k(1-p)p^k}{(1-p^k)}, & \text{if } j = 1 \\ (1 - \omega) \binom{k+j-1}{j} \frac{(1-p)^j p^k}{(1-p^k)} & \text{if } j = 2, 3, \dots \end{cases} \quad (4)$$

where $0 < p < 1$, $0 < \omega < 1$ and $k \geq 1$

Corollary 1: If $J \sim ZTNBw(k, p, \omega)$ be a Zero-Truncated Negative Binomial with excess ones, then some characteristics of J are as follows:

$$E(J) = (1 - \omega)\mu$$

$$Var(J) = (1 - \omega)\sigma^2 + \omega(1 - \omega)\mu^2$$

where $\sigma^2 = Var(J|J > 1)$ and $\mu = E(J|J > 1)$

3 Estimation of Parameters of ZTNBw

The maximum likelihood method for parameter estimations of $ZTNBw(k, p, \omega)$ distribution is implemented in this section. Let $J_1, J_2, J_3, \dots, J_n$ be a set of independent and identically distributed random variables with ZTNBw distribution's sample and j_i ($i = 1, 2, 3, \dots, n$) be its

corresponding set of sample values. Let $\theta = (k, p, \omega)^T$ be the vector of the ZTNBw parameters.

The likelihood function of ZTNBw distribution is

$$L(\theta) = \prod_{i=1}^n \left\{ I_{j_i=1} \left(\omega + (1-\omega) \frac{k(1-p)p^k}{(1-p^k)} \right) + I_{j_i>1} \left((1-\omega) \binom{k+j_i-1}{j_i} \frac{(1-p)^{j_i} p^k}{(1-p^k)} \right) \right\} \quad (5)$$

The log-likelihood function of the ZTNBw(k, p, ω) as

$$\mathcal{L}(\theta) = \sum_{i=1}^n \left\{ I_{j_i=1} \log \left(\omega + (1-\omega) \frac{k(1-p)p^k}{(1-p^k)} \right) + I_{j_i>1} \log \left((1-\omega) \binom{k+j_i-1}{j_i} \frac{(1-p)^{j_i} p^k}{(1-p^k)} \right) \right\} \quad (6)$$

Obtaining the first and second partial derivative of equation (6) with respect to each parameter and equate them to zero as:

$$\frac{\partial \mathcal{L}}{\partial \omega} = 0, \quad \frac{\partial^2 \mathcal{L}}{\partial \omega^2} = 0, \quad \frac{\partial \mathcal{L}}{\partial k} = 0, \quad \frac{\partial^2 \mathcal{L}}{\partial k^2} = 0, \quad \frac{\partial \mathcal{L}}{\partial p} = 0 \quad \text{and} \quad \frac{\partial^2 \mathcal{L}}{\partial p^2} = 0$$

The solution of the parameters can be solved simultaneously by using numerical optimization using “*optim*” or “*nlm*” functions in the R software [11].

To compare the ZTNBw distribution with some other distributions, we consider the criteria like Bayesian information criterion (BIC) and Akaike Information Criterion (AIC) as well as the Log-Likelihood (Loglik) values. The best model is the one with minimum values for its model adequacy measures, such as the AIC, BIC and Loglik. Similarly, the best fitted model is the one having a minimum value for the goodness-of-fit statistic (χ^2). The AIC and BIC can be obtained by using the formulas as follows:

$$AIC = 2q - 2\log L,$$

and

$$BIC = q \log n - 2\log L,$$

where q is the number of parameters, n is the sample size and logL is the maximized value of log-likelihood function and are shown in Table 1.

4 Application of the Model

The data used here were reproduced from Nigeria Demographic and Health Survey 2008 [12], Nigeria: representing counts of mothers that experienced at least one dead child within twelve months preceding data collection. The sample's mean and variance are 2.061 and 2.145, respectively.

Note that the proportion of ones Poisson and Zero-truncated Poisson can accommodate are 2564 and 3773 respectively while there are 4770 ones in the data. This shows that there are more ones in the data than Zero-truncated Poisson model can accommodate.

Table 1: The expected and observed number of mothers of the complete fertility that experienced at least one child death

Y	Observed Frequency	ZTP	ZTNB	ZTNBw
1	4770	3772.7	4716.7	4770.0
2	2388	3159.5	2465.5	2362.0
3	1237	1764.0	1270.1	1284.2
4	677	738.6	649.5	671.9
5	350	247.4	330.6	343.3
6	177	69.1	167.8	172.6
7	97	16.5	85.0	85.7
8	37	3.5	43.0	42.2
9	20	0.6	21.7	20.7
10	12	0.1	11.0	10.0
11 - 15	7	0.0	11.1	9.4
Total	9772	9772	9772	9772
MLE		$\hat{\mu} = 1.6749$	$\hat{p} = 0.9988$ $\hat{k} = 1.0912$	$\hat{p} = 0.8578$ $\hat{k} = 1.5324$ $\hat{\omega} = 0.0747$
Loglik		-14623.28	-13939.39	-13935.93
X^2		2445.82	10.99178	5.4286
(P-value)		(< 0.001)	(0.2022)	(0.6078)
d.f		6	8	7
AIC		29248.56	27882.79	27877.87
BIC		29247.25	27883.39	27876.47

The results in Table 1 shows the goodness-of-fit of Zero-truncated Poisson (ZTP) and Zero-truncated Negative Binomial (ZTNB) distributions compared with the newly proposed ZTNBw distribution. Considering the Loglikelihood, X^2 and p-values it is obvious that the ZTNBw

distribution gave better fit than ZTNB model. The ZTP (with p-value < 0.001) fails to fit here at 5% level of significant, ZTNB (with p-value = 0.2022) fits the data while the ZTNBw, which account for the excess ones in the data fits the data far better (p-value = 0.6078). Figure 1 also displayed how close the fitted values of ZTNBw and ZTNB models to the observed data.

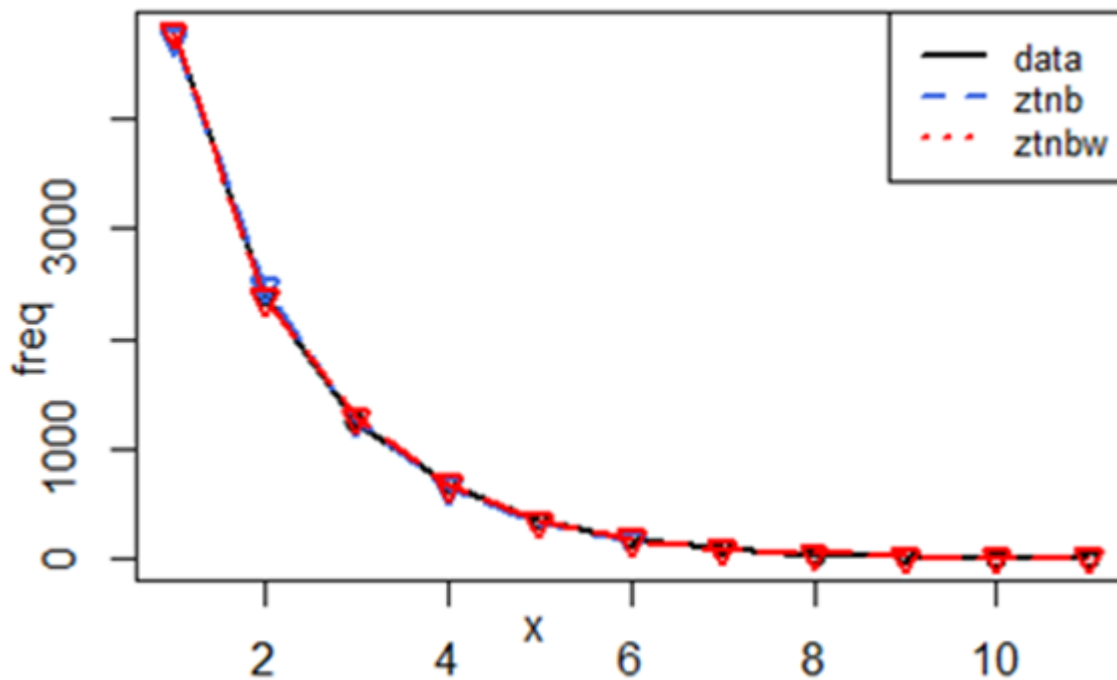


Figure 1: Fitness of ZTNB and ZTNBw to the data.

5 Conclusion

In this study Zero-Truncated Negative Binomial (ZTNB) distribution was modified to account for excess ones similar to application of zero-inflated model when there are excess zeros in count data. The excess ones make data more skewed and there is need to account for that to get more accurate fit. The parameter estimation through maximum likelihood method was presented. The proposed ZTNBw **distribution's** application was compared with ZTP and ZTNB models by using real data. The results show that ZTNBw model provides a better fit compared to the ZTNB distribution.

Competing Interests

The authors have declared that no competing interests exist.

Reference

1. Fisher, R. A. (1941). The negative binomial distribution. *Ann. Eugen., Lond.*, 11, 182.
2. Shmueli, G, Minka, T.P, Kadane, J.B, Borle, S and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Appl. Stat.* 54, 127–142.
3. Yamrubboon, D., Thongteeraparp, A., Bodhisuwan, W., Jampachaisri, K., and Volodin, A., (2019). Bayesian Inference for the Negative Binomial-Sushila Linear Model. *Lobachevskii Journal of Mathematics*, Vol. 40, No. 1, pp. 42–54.
4. Denthet, S. and Promin, P. (2019). The negative binomial-weighted Lindley distribution. *Decision Science Letters*. 8 ; 317–322.
5. Aderoju, S.A. (2020). A New Generalized Poisson Mixed Distribution and its application. *Applied Mathematical Sciences*, Vol. 14, no. 5, 229–234; <https://doi.org/10.12988/ams.2020.914185>
6. David, F. N. and Johnson, N. L. (1952). The truncated Poisson. *Biometrics*, 8, 275.
7. Sampford, M. R. (1955). The Truncated Negative Binomial Distribution. *Biometrika*, Vol. 42, No. 1/2, pp. 58-69; <https://www.jstor.org/stable/2333422>.
8. Winai B., Chookait P., Rujira B., and Luckhana S. (2017). Zero-truncated negative binomial - Erlang distribution. *AIP Conference Proceedings* 1905, 050011; doi:10.1063/1.5012230
9. Aderoju, S.A. and Jolayemi, E.T. (2018). On modeling Zero-truncated count data. *Journal of Global Research in Mathematical Archives (JGRMA)*, Volume 5, No.6, 7-13.
10. Irshad, M.R.; Chesneau, C.; Shibu, D.S.; Monisha, M.; Maya, R. (2022). Lagrangian Zero-truncated Poisson Distribution: Properties Regression Model and Applications. *Symmetry*, 14, 1775. <https://doi.org/10.3390/sym14091775>
11. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
12. National Population Commission (NPC) [Nigeria] and ICF International. (2008). Nigeria Demographic and Health Survey 2008. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International.