BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 4, pp. 90– 113 (2021)

On Com-Negative Binomial Distributions

B. H. Lawal*
Department of Statistics & Mathematical Sciences, Kwara State University, Malete, Kwara State

(*Received: 16 August 2020; accepted: 07 January 2021*)

Abstract. In this paper we contrast the two forms of the Com-negative binomial distributions presented in Chakraborty and Ong (2016) and Zhang *et al.* (2018) as applied to over-dispersed count data. These two forms are designated COMNB and CMNB respectively in this paper. We also contrast the performances of the Negative binomial (NB) and the Com-Poisson (CMP) distributions to the former two distributions. These distributions are applied to example data sets that exhibit over-dispersed and utrahigh zero-inflated count data. Because there are no closed form expressions for the means and variances of COMNB, CMNB and the CMP models, the method of moments is proposed to generate these moments and consequently, we are able to generate Wald's test statistics for these distributions rather than the AIC or -2LL as a measure of goodness-of fit. The zero-inflated forms of these distributions are further extended to fit zero-inflated models to some of these data sets. All the models are implemented in SAS PROC NLMIXED. For each distribution considered, MLE estimation based on the log-likelihood functions are obtained using the Adaptive Gaussian Quadrature (usually with 32 q-points) and then optimized by using the Newton-Raphson algorithm. Starting values are obtained from those obtained from employing the Poisson or Negative binomial models. Our results indicate that the COMNB performs much better than the newly proposed CMNB.

Keywords: Overdispersion, empirical means and variances, zero-inflated models, SAS PROC NLMIXED.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

For count data exhibiting over-dispersion or under-dispersion, probability distributions with extra dispersion parameters such as the Negative-binomial (NB),

*Corresponding author. Email: bayo.lawal@kwasu.edu.ng

the generalized Poisson (GP), the Com-Poisson (COMP), and the hyper Poisson distributions amongst many others. These are two type parameter distributions. In an effort to further fit overdispersed data having long tails better, the three parameter distributions such as the quasi-negative binomial (Li *et al.*, 2011; Hassan and Bilal, 2008), the inverse trinomial(Phang *et al.*, 2013, and the negative binomial generalized exponential (NBGE) distribution (Aryuyuen and Bodhisuwan, 2013), the generalized COMP-Poisson (GCOMP) (Imoto, 2014), the Com-Poisson negative binomial (COMNB) in Chakraborty and Ong (2014), as well as the four parameter extended Com-Poisson distribution (ECOMP) presented in Chakraborty and Imoto (2013). We are concerned in this paper in comparing the the performances of the Conway-Maxwell Negative binomial (COMNB) distribution (Chakraborty *et al.*, 2014) with the COM-negative binomial (CMNB) distribution (Zhang *et al.*, 2018). The latter has been applied to ultrahigh zero-inflated count data. We will therefore in this paper compare the performances of these two models using as baseline the negative binomial (NB) and the COM-Poisson (CMP)distributions.

Several frequency count data will be employed in our analysis, three of which were taken from the Zhang *et al. 2018* and the other two are from other sources. For these four models, we shall obtain empirically their means and variances and hence we would be able to compute the Wald's goodness-of fit test statistics. However, we begin our discussion in this paper with brief introductions to some of these distributions.

2. Probability Distributions

We describe in this section, the pmf of the negative binomial (NB), the COM-Poisson (CMP), the COM-POisson negative binomial (COMNB) and the COM-negative binomial (CMNB) distributions.

2.1 *The negative binomial distribution*

We employ here the form of the negative binomial distribution implemented in Zhang *et al.* (2018) which has the pmf:

$$f(y;r,p) = \frac{\Gamma(r+y)}{y!\Gamma(r)} p^y (1-p)^r, \quad y = 0, 1, \ldots, \tag{1}$$

where $r \in (0, \infty)$ and $p \in (0, 1)$. The mean and variance of the NB model with parameters r and p in (1) are given respectively in (2a) and (2b). Hence,

$$\mu = rp/(1-p) \quad \Longrightarrow p = \frac{\mu}{r+\mu} \tag{2a}$$

$$\sigma^2 = rp/(1-p)^2 \quad \Longrightarrow \sigma^2 = \mu + \frac{\mu^2}{r} \tag{2b}$$

2.2 *The Com-Poisson distribution*

The Com-Poisson (here referred to as CMP) distribution, was introduced by Conway and Maxwell (1962) and was later re-introduced in Shmueli *et al.* (2005) and has then received a lot of great attention (Sellers *et al.*, 2011). The distribution is defined for a random variable Y as:

$$f(y;\nu,\lambda) = \frac{\lambda^y}{(y!)^\nu}\frac{1}{Z(\lambda,\nu)}, \quad y = 0,1,2,\cdots, \quad \lambda > 0,\ \nu \geq 0, \tag{3}$$

where

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}. \tag{4}$$

is the the normalizing term and ν is the *dispersion parameter* such that if $(\nu > 1)$ we have under dispersion, and when $(\nu < 1)$, we have overdispersion. The distribution reduces to the Poisson when $\nu = 1 \rightarrow Z(\lambda,\nu) = e^\lambda \rightarrow Y \sim \text{Pois}(\lambda)$. Because the mean and variance of the CMP distribution do not have closed form expressions, Lawal (2017) has suggested that these can be obtained by the method of moments viz:

$$\begin{aligned} E(Y) &= \frac{1}{Z(\lambda,\nu)} \sum_{j=0}^{\infty} \frac{j\,\lambda^j}{(j!)^\nu}, \text{ and} \\ \text{Var}(Y) &= \frac{1}{Z(\lambda,\nu)} \sum_{j=0}^{\infty} \frac{j^2\,\lambda^j}{(j!)^\nu} - E(Y)^2 \end{aligned} \tag{5}$$

Shmueli *et al.* (2005) provided expressions for approximate mean and variances of the distribution as:

$$\begin{aligned} E(Y) &\approx \lambda^{1/\nu} - \frac{\nu-1}{2\nu}, \quad \text{for } \nu \leq 1 \text{ or } \lambda > 10 \\ \text{Var}(Y) &\approx \frac{1}{\nu}\lambda^{1/\nu} \end{aligned} \tag{6}$$

2.3 *COM-Poisson negative binomial distribution-COMNB*

Chakraborty and Iyamote (2016) introduced the COMNB distribution which is one of what they called 'Extended COMway Maxwell-Poisson Distributions'. The properties of the COMNB and its applications to both frequency and right truncated frequency count data were discussed. Ong (2014) has similarly presented some elegant results on the Com-Poisson negative binomial (COMNB). A random variable Y, having the COMNB distribution parameters (ν, p, α) has its pmf in the form given in (7).

$$f(y;\nu,p,\alpha) = \frac{(\nu)_y\, p^y}{(y!)^\alpha\, {}_1H_{\alpha-1}(\nu,1,p)} \quad = \quad \frac{\Gamma(\nu+y)}{\Gamma(\nu)\, {}_1H_{\alpha-1}(\nu,1,p)}.\frac{p^y}{(y!)^\alpha}; \qquad (7)$$

$y = 0, 1, 2, \ldots$, where

$$ {}_1H(\nu,1,p) = \sum_{k=0}^{\infty} \frac{(\nu)_k\, p^k}{(k!)^\alpha} = \sum_{k=0}^{\infty} \frac{\Gamma(k+\nu)\, p^k}{\Gamma(\nu)(k!)^\alpha}$$

and the distribution is defined in the parameter space

$$\Theta_{COM-NB} = \{\nu > 0, p > 0, \alpha > 1\} \cup \{\nu > 0, 0 < p < 1, \alpha = 1\}$$

.

2.4 *The COM-negative binomial distribution-CMNB*

An alternative formulation of the COMNB in (7) is provided in Zhang *et al.* (2018) is the CMNB with parameters (r, ν, p) and has the p.m.f. given by

$$f(y;r,\nu,p) = \left[\frac{\Gamma(r+y)}{y!\Gamma(r)}\right]^\nu p^y(1-p)^r.\frac{1}{C(r,\nu,p)}; \quad y = 0, 1, \ldots \qquad (8)$$

where $r, v \in (0, \infty)$ and $p \in (0, 1)$ with

$$C(r,\nu,p) = \sum_{j=0}^{\infty} \left[\frac{\Gamma(r+j)}{j!\Gamma(r)}\right]^\nu p^j(1-p)^r \qquad (9)$$

being the normalizing constant. The model in (8) is analogous to the negative binomial model in (1) where, $\Gamma(r + y)/y!\Gamma(r)$ in (1) is simply replaced with $[\Gamma(r + y)/y!\Gamma(r)]^\nu$ in (8) with the accompanying normalization constant in (9).

We will now apply these four distributions to example data sets in the rest of this paper.

3. Parameter Estimation

The log likelihood of a single observation i from **NB,CMP, CMNB, COMNB** are given in expressions (10a) to (10d) respectively:

$$\text{LL1} = y\log(p) + r\log(1-p) + \log\Gamma(y+r) - \log(y!) - \log\Gamma(r) \tag{10a}$$

$$\text{LL2} = y\log(\lambda) - \nu\log(y!) - \log Z(\lambda,\nu) \tag{10b}$$

$$\text{LL3} = y\log(p) + r\log(1-p) + \nu\log\Gamma(y+r) - \nu\log(y!) - \nu\log\Gamma(r) - \log C(r,\nu,p) \tag{10c}$$

$$\text{LL4} = y\log(p) + \log\Gamma(y+\nu) - \alpha\log(y!) - \log\Gamma(\nu) - \log H(\nu,1,p) \tag{10d}$$

Maximum-likelihood estimations (we have restricted ourselves to the single component, y_i of the log-likelihood function) in (10) is carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y,\Theta)$ over the parameter space Θ numerically. The integral approximations employed in PROC NLMIXED is the Adaptive Gaussian Quadrature and the choice optimization algorithm here is the Newton-Raphson technique (NEWRAP). Convergence is often a major problem but a good choice of starting values result in rapid convergence.

4. Applications to Example Data Sets

Example I

Our first example data is presented in Table 1 and gives the number of death notices of women 80 years of age and older, has it appeared in the London Times on each day for three consecutive years. The data was analyzed in Hasselblad (1969) and recently re-analyzed in Gupta *et al.* (2014). This example data was selected to illustrate how we can obtain empirical means and variances of all the distributions discussed in the previous sections, since three of them do not have close form expressions for their means and variances and we know that the CMP fits the data very well. Thus, the results of applying the distributions are also provided in the Table. We summarize these results as follows:

(a) The observed mean μ and variance (σ^2) of the data are respectively 2.1569 and 2.6073, giving a dispersion index of 1.21. The sample size of the data is $n = 1096$.

(b) The sum of the expected values under each of the models do not add to $n = 1096$. Consequently, we can add the difference to the last category to make the sum in each add to 1096 giving the last category expected values in parentheses. Lawal (2017) has provided an alternative way of handling this situation which is peculiar to all count model distributions.

(c) The most parsimonious model based on the grouped X_g^2 GOF is the CMP with $X_g^2 = 1.5910$ on 7 d.f.

(d) We observe here that all the models produced estimated means that are very close to the observed mean, but the estimated variances are slightly differ-

Table 1: Models for the frequency counts of death notices

Y	Count	NB	CMNB	COMNB	CMP
0	162	155.6940	160.3459	162.3702	161.6239
1	267	275.7962	270.3133	267.0340	268.3344
2	271	268.9208	265.8266	264.6942	264.9258
3	185	190.8337	192.2163	193.5118	192.9890
4	111	110.0932	112.3289	113.7353	113.3072
5	61	54.7464	55.8618	56.3857	56.2750
6	27	24.3175	24.4216	24.3135	24.3782
7	8	9.8794	9.5976	9.3148	9.4078
8	3	3.7327	3.4461	3.2206	3.2847
9	1	1.3277	1.1446	1.0171	1.0499
		(1.9861)	(1.6420)	(1.4199)	(1.4741)
Total	1096	1095.3416	1095.5026	1095.5972	1095.5758
		(1096)	(1096)	(1096)	(1096)
		$\hat{p}$=0.1787	$\hat{r}$=39.908	$\hat{\nu}$=0.9390	$\hat{\lambda}$=1.6602
		$\hat{r}$=9.9104	$\hat{\nu}$=0.8064	$\hat{p}$=1.7515	$\hat{\nu}$=0.7498
			$\hat{p}$=0.0863	$\hat{\alpha}$=1.7766	
μ	2.1569				
σ^2	2.6073				
$\bar{y}$		2.1569	2.1569	2.1569	2.1569
s^2		2.6264	2.6233	2.6107	2.6143
X_g^2		2.7390	1.7645	1.5906	1.5910
d.f.		7	6	6	7
p-value		0.9081	0.9400	0.9533	0.9790
X_W^2		1087.051	1088.3390	1093.5614	1092.0650
d.f.		1093	1092	1092	1093

ent with the COMNB and CMP providing estimated means and variances that are much closer to the observed variance of 2.6073 in the data.

(e) Consequently, the Wald's test statistics $X_W^2 = \sum_{i=1}^{1096} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}_i^2}$ is lowest for the NB model because it has the largest estimated variance.

5. Computations of Moments

There are no closed form expressions for the means and variances of CMNB, COMNB and CMP distributions. Consequently, if we want to to conduct Wald's test for each of these distributions, we would have to develop a procedure for obtaining these-which are described as empirical means and variances. Thus these empirical means and variances of the distributions in each of the table above are computed via the expressions in (11).

$$E(Y) = \sum_{j=1}^{\infty} j \; f(y|) \tag{11a}$$

$$\text{Var}(Y) = \sum_{j=1}^{\infty} j^2 f(y|)) - [E(Y)]^2 \tag{11b}$$

where $f(y|\hat{\ })$ is the pmf of the distribution under consideration and $\hat{\ }$ is a vector of estimated parameters of the model. For the negative binomial (NB) the estimated theoretical mean and variance are given respectively in (2). Let us illustrate the procedure in (11) with the data set in Example I, that is, the data set in Table 1. For this data set, the estimated theoretical mean and variance are respectively, 2.15693 and 2.62638. However, when the NB model is applied to the data in the range $0 \leq Y \leq 9$, the corresponding estimated mean and variance are 2.15065 and 2.58737, respectively (see Table 2) which are not equal to the theoretical values obtained at convergent Y_m value of 21. Table 2 gives the estimated probabilities under the NB model (using the log-likelihood), the cumulative probabilities, fitted or expected values (fit), cumulative sums of expected values (ss), the mean and the variance (var). At $Y = 9$, the sum of expected values do not add up to the sample size $n = 1096$. It does not get to $n = 1096$ until $Y = 21$, which is well outside the range of Y for the observed data. At that value of Y, the mean and variance agree with the theoretical values respectively. Clearly, this procedure (method of moments) gives the exact theoretical values of the mean and variance at $Y = 21$.

Table 2: The Negative Binomial (NB) model applied to the data in Table 1

Y	prob	cum	fit	ss	mean	v1	var
0	0.14206	0.14206	155.693999	155.693999	0.00000	0.00000	0.00000
1	0.25164	0.39370	275.796224	431.490223	0.25164	0.25164	0.18832
2	0.24537	0.63906	268.920812	700.411034	0.74237	1.23310	0.68199
3	0.17412	0.81318	190.833695	891.244729	1.26473	2.80017	1.20064
4	0.10045	0.91363	110.093184	1001.337912	1.66653	4.40737	1.63006
5	0.04995	0.96358	54.746430	1056.084343	1.91628	5.65614	1.98401
6	0.02219	0.98577	24.317543	1080.401886	2.04941	6.45490	2.25483
7	0.00901	0.99478	9.879357	1090.281243	2.11250	6.89658	2.43391
8	0.00341	0.99819	3.732661	1094.013904	2.13975	7.11455	2.53602
9	0.00121	0.99940	1.327723	1095.341626	2.15065	7.21267	2.58737*
10	0.00041	0.99981	0.448781	1095.790407	2.15475	7.25362	2.61068
11	0.00013	0.99994	0.145194	1095.935601	2.15620	7.26965	2.62043
12	0.00004	0.99998	0.045223	1095.980823	2.15670	7.27559	2.62424
13	0.00001	0.99999	0.013623	1095.994447	2.15686	7.27769	2.62564
14	0.00000	1.00000	0.003985	1095.998431	2.15691	7.27841	2.62613
15	0.00000	1.00000	0.001135	1095.999567	2.15693	7.27864	2.62630
16	0.00000	1.00000	0.000316	1095.999883	2.15693	7.27871	2.62635
17	0.00000	1.00000	0.000086	1095.999969	2.15693	7.27873	2.62637
18	0.00000	1.00000	0.000023	1095.999992	2.15693	7.27874	2.62638
19	0.00000	1.00000	0.000006	1095.999998	2.15693	7.27874	2.62638
20	0.00000	1.00000	0.000002	1095.999999	2.15693	7.27874	2.62638
21	0.00000	1.00000	0.000000	1096.000000	2.15693	7.27874	2.62638**

For all distributions, since usually, this value of Y_m is often unknown, we propose that we can estimate the empirical (or numerical) mean and variance from the following expressions in (11), where for the NB distribution for instance:

$$E(Y) = \sum_{j=1}^{\infty} j\ f(y|\hat{p}, \hat{r}) \tag{12a}$$

$$\text{Var}(Y) = \sum_{j=1}^{\infty}\ j^2 f(y|\hat{p}, \hat{r})) - [E(Y)]^2 \tag{12b}$$

Here, $f(y|\hat{p}, \hat{r})$ is the pmf of the NB distribution under the estimated parameters $\hat{p}$ and $\hat{r}$. We allow the maximum value of j be set to 200, although 21 would still have been alright. The results obtained using the expressions in (12) for the NB distribution are presented below and they all agree with those obtained using the maximum values of $Y_m = 21$.

Moments	Theoretical	Empirical
μ	2.15693	2.15693
$\hat{\sigma}^2$	2.62638	2.62638

Since the theoretical values of the NB are correctly reproduced by this procedure, this therefore allows us to use the same to generate empirical means and variances of the CMNB, COMNB and CMP. We shall illustrate this for the CMP distribution, and again let us use the data in Example I, since the CMP fits the data very well from results in Table 1, we can compare our results with expressions provided in Gupta *et al.* (2014) for the mean and variance of the CMP. Using the recurrence relations established in Gupta (1974) and subsequently in Gupta *et al.* (2014), the mean and variance of the r.v having the CMP distribution can be obtained via:

$$\begin{aligned} E(Y) &= \lambda.\frac{\partial}{\partial \lambda}[\log Z(\lambda, \nu)] \\ Var(Y) &= \lambda.\frac{\partial^2}{\partial \lambda^2}[\log Z(\lambda, \nu)] \end{aligned} \tag{13}$$

where

$$\begin{aligned} \frac{\partial \log Z(\lambda, \nu)}{\partial \lambda} &= \frac{\displaystyle\sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j!)^{\nu}}}{\displaystyle\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu}}} \\ \frac{\partial^2 \log Z(\lambda, \nu)}{\partial \lambda^2} &= \frac{\displaystyle\sum_{j=2}^{\infty} \frac{j(j-1)\lambda^{j-2}}{(j!)^{\nu}}}{\displaystyle\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^{\nu}}} \end{aligned} \tag{14}$$

However, when we employ the procedure in (11), we have for the CMP distribution

$$E(Y) = \sum_{j=1}^{\infty} j\ f(y|\hat{\lambda}, \hat{\nu}) \tag{15a}$$

$$\mathrm{Var}(Y) = \sum_{j=1}^{\infty} j^2 f(y|\hat{\lambda}, \hat{\nu})) - [E(Y)]^2 \tag{15b}$$

where $f(y|\hat{\lambda}, \hat{\nu})$ is the pmf of the distribution under estimated parameters $\hat{\lambda}$, and $\hat{\nu}$. In practice, a very large value, say 200, is used instead of infinity.

For this example, we have the following:

$$\begin{aligned}
\sum_{j=0}^{200} \frac{\lambda^j}{(j!)^\nu} &= 6.78117 \\
\sum_{j=1}^{200} j\ f(y|\hat{\lambda}, \hat{\nu}) &= 14.62655 \\
\sum_{j=1}^{200} j^2 f(y|\hat{\lambda}, \hat{\nu})) &= 49.27666 \\
\sum_{j=1}^{200} \frac{j\lambda^{j-1}}{(j!)^\nu} &= 8.80990 \\
\sum_{j=2}^{200} \frac{j(j-1)\lambda^{j-2}}{(j!)^\nu} &= 12.57081
\end{aligned} \tag{16}$$

Thus, the mean and variance are computed from expressions in (15) as:

$$\bar{y} = \frac{14.62655}{6.78117} = 2.15693; \quad s^2 = \frac{49.62655}{6.78117} - \bar{y}^2 = 2.61432$$

Similarly, using the Gupta *et al.* (2014) expressions in (13), the mean and variance are computed, here, $\hat{\lambda} = 1.66024$, viz:

$$\bar{y} = 1.66024 \times \frac{8.80990}{6.78117} = 2.15694; \quad s^2 = 1.66024 \times \frac{12.57081}{6.78117} = 3.07772$$

These results are summarized below:

Moments	Using expressions in (11)	Using Gupta *et al.* (2014)	At $Y_m = 18$
μ	2.1569	2.1569	2.1569
s^2	2.6143	3.0777	2.6143

While the Gupta *et al.* (2014) produces an equivalent mean, it does not produce an equivalent variance. Since the CMP distribution fits this data set very well, it must be the case that its mean and variance are very close to the observed mean and variance of the data. The mean and variance obtained from expressions in (15) are very close to the observed mean and variance and thus the model fits the data well. Consequently, the variance expressions from Gupta *et al.* grossly overestimate the variance and is therefore not reliable. Consequently, the Wald's goodness-of-fit test statistics X^2 defined above are already presented as part of Table 1. Presented below are the estimated probabilities, cumulative probabilities, expected values, the cumulative expected values, mean, and variance (generated via log-likelihood).

Table 3: The Com-Poisson (CPM) applied to the data in Table 1

Y	prob	cum	fit	ss	mean	v1	var
0	0.14747	0.14747	161.623946	161.623946	0.00000	0.00000	0.00000
1	0.24483	0.39230	268.334441	429.958387	0.24483	0.24483	0.18489
2	0.24172	0.63402	264.925766	694.884153	0.72827	1.21171	0.68133
3	0.17608	0.81010	192.988988	887.873141	1.25653	2.79648	1.21762
4	0.10338	0.91349	113.307167	1001.18031	1.67006	4.45060	1.66151
5	0.05135	0.96483	56.275022	1057.45533	1.92679	5.73424	2.02174
6	0.02224	0.98707	24.378169	1081.83350	2.06024	6.53498	2.29039
7	0.00858	0.99566	9.407786	1091.24128	2.12033	6.95559	2.45979
8	0.00300	0.99866	3.284652	1094.52594	2.14430	7.14739	2.54935
9	0.00096	0.99961	1.049869	1095.57581	2.15293	7.22498	2.58990
10	0.00028	0.99990	0.310078	1095.88588	2.15575	7.25327	2.60600
11	0.00008	0.99997	0.085265	1095.97115	2.15661	7.26269	2.61172
12	0.00002	0.99999	0.021965	1095.99311	2.15685	7.26557	2.61357
13	0.00000	1.00000	0.005329	1095.99844	2.15691	7.26640	2.61412
14	0.00000	1.00000	0.001223	1095.99966	2.15693	7.26661	2.61427
15	0.00000	1.00000	0.000266	1095.99993	2.15693	7.26667	2.61431
16	0.00000	1.00000	0.000055	1095.99999	2.15693	7.26668	2.61432
17	0.00000	1.00000	0.000011	1096.00000	2.15693	7.26668	2.61432
18	0.00000	1.00000	0.000002	1096.00000	2.15693	7.26669	2.61432

We observe that in the range of our data $0 \leq Y \leq 9$, sum of the estimated cumulative probabilities is 0.99961 with cumulative fitted values of 1095.57581 which does not sum to the observed sample size of $n = 1096$. At $Y = 9$, the estimated mean and variance are respectively 2.15293 and 2.58990 which are not equal to the expected mean and variance. However, the probabilities add to 1 at $Y = 18$ and at this point the means and variances are equivalent to those obtained from expressions in (15).

Consequently therefore, we are going to employ the approach outlined in (11) to compute the means and variances of the CMNB and the COMNB. Results based on this are presented in Table 1.

Using the method of moments introduced in (11), the empirical mean and variance are computed with the results for the first five observations displayed in Table 4.

Here, **suma** is the normalizing constant. **sumb** is the expression in (12a), while, **sumc** is the first part of the expression in (12b). Similarly, **sumd** and **sume** are from the expressions in (16). We observe that the ensuing computed empirical mean and variance agree with those generated from the probability distribution. Since there are no covariates in the data, the above display will be

Table 4: Computation of the moments of the first five Observations

Obs	suma	sumb	sumc	sumd	sume	mean	var	m1	v1
1	6.78117	14.6265	49.2767	8.80990	12.5708	2.15693	2.61432	2.15693	3.07772
2	6.78117	14.6265	49.2767	8.80990	12.5708	2.15693	2.61432	2.15693	3.07772
3	6.78117	14.6265	49.2767	8.80990	12.5708	2.15693	2.61432	2.15693	3.07772
4	6.78117	14.6265	49.2767	8.80990	12.5708	2.15693	2.61432	2.15693	3.07772
5	6.78117	14.6265	49.2767	8.80990	12.5708	2.15693	2.61432	2.15693	3.07772

replicated for all the $n = 1096$ observations.

5.1 *Examples*

Example II

This is the car insurance claim data from a Chinese insurance company (Wang and Lei, 2000) and recently re-analyzed in Zhang *et al.* (2018). The data exhibits ultra-high zero inflation.

Table 5: Distribution of insurance claims in a Chinese Insurance Company

Y	Count	NB	CMNB	COMNB	CMP
0	27141	27165.8231	27170.3756	27153.2342	26617.3256
1	5789	5664.0585	5667.0255	5725.9548	6416.5947
2	1443	1563.3450	1552.4894	1516.8727	1546.7894
3	457	466.6827	465.6390	451.7546	372.8635
4	155	144.5631	146.1237	145.5712	89.8800
5	56	45.7569	47.1167	49.7770	21.6656
6	27	14.6889	15.4697	17.8508	5.2225
7	2	4.7626	5.1451	6.6611	1.2589
8	1	1.5557	1.7279	2.5719	0.3034
9	1	0.5111	0.5847	1.0231	0.0731
Total	35072	35071.7476	35071.6972	35071.2715	35071.9768
		$\hat{p}$=0.3435	$\hat{r}$=0.9906	$\hat{\nu}$=0.8870	$\hat{\lambda}$=0.2411
		$\hat{r}$=0.6069	$\hat{\nu}$=57.674	$\hat{p}$=0.2377	$\hat{\nu}$=0.0001
			$\hat{p}$=0.3594	$\hat{\alpha}$=0.7600	
μ	0.3176				
σ^2	0.4913				
$\bar{y}$		0.3176	0.3176	0.3176	0.3176
s^2		0.4838	0.4860	0.4904	0.4185
X^2		27.4828	23.5908	14.9810	300.5555
d.f.		7	6	6	7
p-value		0.0003	0.0006	0.0204	0.0000
X_W^2		35616.34	35071.69	35137.05	41173.7429
d.f.		35069	35068	35068	35069

Again none of the cumulative sums of expected values add up to exactly $n = 35,072$. We have therefore added the difference to the last category such that for the NB model the expected value in the last category, viz $\hat{m}_9 = 0.7635$, giving a Pearson's GOF for the model of 27.4828 on 7 d.f. We have done similarly for the other models, giving respectively, estimated values for {CMNB, COMNB, CMP}={0.8875, 1.7516, 0.0963} although these are not displayed but the GOF statistics are computed with these revised estimates. For this data set, the COMNB is the most parsimonious and performs much better than the Zhang *et al.* (2018) CMNB. The estimated mean and variance of the COMNB distribution are very close to the corresponding observed values from the data.

For this data for instance, the theoretical mean and variance for the NB model is obtainable at $Y = 22$ (see results in Table 6), because that is when the cumulative sum of expected values add to the sample size $n = 35,072$.

Table 6: The Negative Binomial (NB) model applied to the data in Table 5

Y	prob	cum	fit	ss	mean	v1	var
0	0.77457	0.77457	27165.823070	27165.823070	0.00000	0.00000	0.00000
1	0.16150	0.93607	5664.058515	32829.881585	0.16150	0.16150	0.13542
2	0.04458	0.98065	1563.345043	34393.226628	0.25065	0.33980	0.27697
3	0.01331	0.99395	466.682676	34859.909304	0.29057	0.45956	0.37513
4	0.00412	0.99807	144.563113	35004.472417	0.30706	0.52551	0.43122
5	0.00130	0.99938	45.756925	35050.229342	0.31358	0.55812	0.45979
6	0.00042	0.99980	14.688864	35064.918206	0.31609	0.57320	0.47329
7	0.00014	0.99993	4.762634	35069.680840	0.31704	0.57986	0.47934
8	0.00004	0.99998	1.555693	35071.236533	0.31740	0.58269	0.48195
9	0.00001	0.99999	0.511078	35071.747611	0.31753	0.58387	0.48305
10	0.00000	1.00000	0.168666	35071.916277	0.31758	0.58436	0.48350
11	0.00000	1.00000	0.055870	35071.972147	0.31759	0.58455	0.48368
12	0.00000	1.00000	0.018564	35071.990712	0.31760	0.58462	0.48375
13	0.00000	1.00000	0.006184	35071.996896	0.31760	0.58465	0.48378
14	0.00000	1.00000	0.002065	35071.998961	0.31760	0.58467	0.48379
15	0.00000	1.00000	0.000691	35071.999652	0.31760	0.58467	0.48380
16	0.00000	1.00000	0.000231	35071.999883	0.31760	0.58467	0.48380
17	0.00000	1.00000	0.000078	35071.999961	0.31760	0.58467	0.48380
18	0.00000	1.00000	0.000026	35071.999987	0.31760	0.58467	0.48380
19	0.00000	1.00000	0.000009	35071.999996	0.31760	0.58467	0.48380
20	0.00000	1.00000	0.000003	35071.999998	0.31760	0.58467	0.48380
21	0.00000	1.00000	0.000001	35071.999999	0.31760	0.58467	0.48380
22	0.00000	1.00000	0.000000	35072.000000	0.31760	0.58467	0.48380

Example III

The data in this example is presented in Zhang *et al.* (2018) as Table 2. It is a 10,000 synthetic (or simulated) observations from the COMNB distribution with parameters $r = 0.005$, $\nu = 0.1$ and $p = 0.5$, that is a, CMNB (0.005,0.1,0.5). The observed mean of the data is 0.6821. Thus, under the Poisson model, we would expect $\exp(-0.6821)$%=50.56% of the data to be zeros. This data has 64.42% zeros, thus it is highly zero-inflated. The results of applying the four models to the data are presented in Table 7.

The parameter estimates under the COMNB model have 95% confidence intervals $r = [-0.0349, 0.0737]$. It is clear that this parameter is not significant but the interval includes the true value of r. However, the corresponding 95% confidence intervals for $p = [0.4817, 0.5273]$ and that for $\nu = [0.0335, 0.2436]$. Both are clearly significant and these intervals include the true values of the parameters. Results from Table 7 indicate that both the NB (pvalue=0.0716) and the CMP (pvalue=0) do not fit the data, particularly the CMP. This is because while their estimated means are equal to the observed mean of 0.6821, the NB overestimates the observed variance of the data, while the CMP grossly underestimates the observed variance. Consequently, the two models do not fit the data. On the other hand both the CMNB (pvalue=0.5801) and the COMNB (pvalue=0.5186) fit the data well but the CMNB fits better in this case. This is not surprising since the data was simulated from a CMNB distribution. We observe that the estimated means and variances from both distributions are very close to the observed values, particularly the COMNB model.

Table 7: Fitting results to simulated data from a CMNB(0.005,0.1,0.5)

Y	Count	NB	CMNB	COMNB	CMP
0	6442	6420.5633	6444.7274	6440.7105	5944.9498
1	1866	1950.8620	1882.7480	1877.1421	2410.7070
2	874	837.2996	865.1369	862.8398	977.5538
3	435	394.3493	413.1548	420.8716	396.4030
4	188	193.9682	200.4634	206.6216	160.7434
5	101	97.8383	98.1169	100.6485	65.1822
6	55	50.1676	48.2893	48.4198	26.4317
7	19	26.0234	23.8569	22.9719	10.7182
8	7	13.6156	11.8193	10.7459	4.3463
9	8	7.1712	5.8681	4.9579	1.7624
10	2	3.7970	2.9183	2.2573	0.7147
11	2	2.0150	1.4533	1.0148	0.2898
12	1	1.0775	0.7246	0.4507	0.1175
T0tal	10000	9998.7521	9999.2772	9999.6522	9999.9198
		(2.3254)	(1.4474)	(0.7985)	(0.4875*)
		$\hat{p}$=0.5545	$\hat{p}$=0.5045	$\hat{p}$=0.7614	$\hat{\lambda}$=0.4055
		$\hat{r}$=0.5479	$\hat{r}$=0.0194	$\hat{\alpha}$=1.1956	
			$\hat{\nu}$=0.1386	$\hat{\nu}$=0.3828	$\hat{\nu}$=0.0000
μ	0.6821				
σ^2	1.4752				
$\bar{y}$		0.6821	0.6813	0.6821	0.6821
s^2		1.5312	1.4935	1.4724	1.1474
X_G^2		17.1251	7.5495	8.1546	279.9009
d.f.		10	9	9	9
p-value		0.0716	0.5801	0.5186	0.0000
X_W^2		9632.9878	9876.6645	10018.199	12855.496
d.f.		9997	9996	9996	9997

Example IV

This example is again taken from Zhang *et al.* (2018) and relate to claim counts of third party liability vehicle insurance in a Zaire insurance company (Willmot, 1987). The data in Table 8 are therefore the distribution of claims from 4000 vehicle polices.

The observed data has a mean of 0.0865 and thus under the Poisson model the percentage of expected zeros would be $\exp(-0.0865)$=91.72%. However the observed data has about 93.98% zeros. Clearly with this percentage, the NB model fits this data set well because the percentage of observed zeros is not too far from that expected under the Poisson and this is the main reason the NB model fits well this data. The data is not zero-inflated. Both the COMNB and Zhang *et al.* (2018) CMNB fit the data as well but the former is better than the Zhang *et al.* CMNB. Observe that the X_g^2 computed are different from those presented in Zhang*et al.* because we have taken into consideration the problem of small expected frequencies and has accordingly applied the Lawal (1980) rule. The CMP does not fit well at all. We also observe that computed means and variances of the other three models are very close to those from the observed data.

Table 8: Distribution of claims from an Insurance Company

Y	Count	NB	CMNB	COMNB	CMP
0	3719	3719.2220	3719.0766	3719.0327	3681.5462
1	232	229.9009	231.7116	231.7447	293.1006
2	38	39.9106	37.7353	37.8660	23.3347
3	7	8.4156	8.3970	8.2928	1.8578
4	3	1.9313	2.1810	2.1431	0.1479
5	1	0.4648	0.6213	0.6204	0.0118
Total	4000	3999.8453	3999.7229	3999.6997	3999.9990
		(0.6195)	(0.8984)	(0.9207)	(2.0185)
		$\hat{p}$=0.2854	$\hat{r}$=0.9267	$\hat{\nu}$=0.3144	$\hat{\lambda}$= 0.0796
		$\hat{r}$=0.2166	$\hat{\nu}$=24.7880	$\hat{p}$=0.1982	$\hat{\nu} \approx 0.000$
			$\hat{p}$=0.4107	$\hat{\alpha}$=0.6729	
μ	0.0865				
σ^2	0.1225				
$\bar{y}$		0.0865	0.0865	0.0865	0.0865
s^2		0.1210	0.1226	0.1227	0.0940
X_g^2		1.1738	0.5537	0.5517	62.2989
d.f.		3	2	2	1
p-value		0.7593	0.7582	0.7589	0.0000
X_W^2		4048.6975	3995.8750	3992.9867	5214.6308
d.f.		3997	3996	3996	3997

5.2 *Data Set V*

The data in Table 9 is the single-vehicle roadway departure fatal crashes that occurred on 32,672 rural two-lane horizontal curves between 2003 and 2008, Lord and Geedipally (2011). About 90% of the data experienced no crashes giving rise to excess zeros for the data.

Table 9: Observed and expected values under the various models

Y	Count	NB	CMNB	COMNB	CMP	NB-L
0	29087	29101.5496	29114.8553	29098.6795	28719.8554	29099.7347
1	2952	2859.1680	2851.6209	2891.8779	3474.0763	2904.4275
2	464	549.3741	535.4603	517.3517	420.2391	499.7932
3	108	122.7542	124.8763	118.3130	50.8339	116.5471
4	40	29.3497	32.4543	31.5448	6.1491	33.2889
5	9	7.2929	9.0008	9.3992	0.7438	11.0241
6	5	1.8578	2.6068	3.0551	0.0900	4.0911
7	2	0.4816	0.7787	1.0665	0.0109	1.6629
8	3	0.1264	0.2381	0.3955	0.0013	0.7283
9	1	0.0335	0.0741	0.1546	0.0002	0.3396
10	1	0.0090	0.0234	0.0633	0.0000	0.1670
Total	32672.0	32671.9967	32671.9898	32671.9011	32672.0000	32671.8044
		(2.5516)	(1.1253)	(1.7787)	(0.8462)	(1.4305)
		$\hat{p}$=0.2860	$\hat{p}$=0.3596	$\hat{p}$=0.1732	$\hat{\lambda}$=0.1210	$\hat{\theta}$=10.0549
		$\hat{r}$=0.3435	$\hat{r}$=0.9967	$\hat{\alpha}$=0.6076		$\hat{r}$=1.1311
			$\hat{\nu}$=399.33	$\hat{\nu}$=0.5738	$\hat{\nu} \approx 0.000$	
μ	0.1376					
σ^2	0.2044					
$\bar{y}$		0.1376	0.1376	0.1376	0.1376	0.1375
s^2		0.1927	0.1959	0.2534	0.1566	0.1988
X_g^2		58.1720	49.9984	26.5038	485.3629	14.8767
d.f		5 (7+)	5 (8+)	5 (8+)	3 (5+)	6 (8+)
p-value		0.0000	0.0000	0.0001	0.0000	0.0212
X_W^2		34643.6	34085.3	26346.9	42639.24	33594.1
d.f.		32,668	32,667	32,667	32,668	32,668

For this data set, the observed mean and variance are respectively 0.1376 and

0.2044. None of the models NB, CMP, CMNB and COMNB fits this data set. Again as expected, the sums of expected values in all the models do not sum to the sample size of $n = 32672$. The percentage of observed zeros is 89.03%, while under the Poisson this would be $100\exp(-0.1376)$=87.14%. This is not too far from the observed percentage of zeros. Thus this data set does not exhibit zero inflation with regards to the Poisson distribution. The group's Pearson X_g^2 are computed based on the Lawal's (1980) rule on minimum expected values required for the χ^2 approximation to be valid. Thus, under the COMNB model the categories (8+, that is, 8, 9 and 10) are collapsed to give the expected count of 1.7787 (this includes the shortfall from the sample size, viz: 32672-32671.9011=0.0989). The COMNB is the most parsimonious among the four distributions and it does perform much better than the Zhang *et al.* (2018) CMNB distribution. In the last column are the results of applying the negative binomial-Lindley model to the data. Lord and Geedipally (2011) applied the NB-L to the data in Table 9.

The Negative binomial Lindley pmf (Zamani and Ismail, 2010) is given by:

$$f(y;\theta,r) = \frac{\Gamma(y+r)}{y!\,\Gamma(r)}.\left(\frac{\theta^2}{\theta+1}\right)\sum_{j=0}^{y}\frac{\Gamma(y+1)}{\Gamma(j+1)\Gamma(y-j+1)}.(-1)^j\left[\frac{\theta+r+j-1}{(\theta+r+j)^2}\right] \tag{17}$$

We note here that the expected values for $Y = 8$ and $Y = 9$ are not zeros as presented in Lord and Geedipally (2011). Its mean and variance are given respectively in expressions (18a) and (18b).:

$$E(Y) = \left[\frac{\theta^3}{(\theta+1)(\theta-1)^2} - 1\right] r \tag{18a}$$

$$E(Y^2) = (r+r^2)\left[\frac{\theta^3(\theta-1)}{(\theta+1)(\theta-2)^2}\right] - (r+2r^2)\left[\frac{\theta^3}{(\theta+1)(\theta-1)^2}\right] + r^2 - [E(Y)]^2 \tag{18b}$$

The theoretical mean and variance of the distribution for this data set computed from the expressions in (18) are μ=0.1375 and σ^2=0.1988. These values are not realized for this data set until Y=25, well outside the range of $0 \leq Y \leq 10$ of the observed data. At $Y = 10$, the empirical computed mean and variance are 0.1374 and 0.1978, which are not yet equal to the theoretical values. The NB-L model fits better than the other models for this data set with a p-value of 0.0212 (not yet a good fit). However, if we combine categories 7+, we would have computed $X_g^2 = 10.8350$ on 5 d.f. and a pvalue of 0.0548, a slightly good fit for the NB-L.

Results:

Based on results from the five example data sets above, it is clear that:

(a) None of the models have their expected values sum to the observed sample size withing the range of the data. This is true of all count models (see Lawal, 2017).
(b) The method of moments adequately computes the means and variances of these distributions, especially for those distributions that do not have close form expressions.
(c) The Zhang*et al.* (2018(CMNB does not perform better than the COMNB proposed in Chakrabort (2014).
(d) Computations of the empirical means and variances give us the opportunity to know how the distributions closely match those of the observed moments in the data. It also allows us to compute the Wald's test statistic, although, models with very high estimated variances will provide lower X_W^2, even the estimated variance deviates grossly from the observed variance.

6. Zero-Inflated Models

We present in this section the corresponding zero-inflated models for the four distributions considered in the previous sections.In many situations of insurance data, the zero outcome is clearly different from the non-zero ones. Ignoring the zero outcome in modeling such data usually lead to biasness etc. (Hilbe, 2014). The probability mass function of a zero-inflated distribution (ZI) is a two-part process manifested by the structural zeros part and the process that generates random counts and can be written in the form:

$$\Pr(Y = y|\phi) = \begin{cases} \phi + (1-\phi)\ \Pr(Y=0) & \text{if } y_i = 0 \\ (1-\phi)\ \Pr(Y = y_i)) & \text{if } y_i = 1, 2, \cdots \end{cases} \tag{19}$$

where ϕ is the extra proportion of zeros and Y is the count random variable with specified parameters. ϕ is modeled here in the logit form. Thus, the probability mass function for the ZINB, ZICMP, ZICMNB and ZICOMNB models are given in expressions (20), (21), (24) and (25), respectively.

$$\Pr(Y_i = y_i) = \begin{cases} \phi + (1-\phi)(1-p)^r, & y = 0 \\ (1-\phi)\dfrac{\Gamma(r+y)}{y!\Gamma(r)}p^y(1-p)^r, & y > 1 \end{cases} \tag{20}$$

6.1 *ZICMP*

For the zero-inflated Com-Poisson, we have the probability density function:

$$P(\lambda, \beta; y_i) = \begin{cases} \phi + (1-\phi)\dfrac{1}{Z(\lambda_i, \nu_i)} & \text{if } y_i = 0 \\ (1-\phi)\dfrac{1}{Z(\lambda_i, \nu_i)}\dfrac{\lambda_i^{y_i}}{(y!)^{\nu_i}} & \text{if } y_i = 1, 2, \ldots \end{cases} \tag{21}$$

where

$$Z(\lambda_i, \nu_i) = \sum_{j=0}^{\infty} \frac{\lambda_i^k}{(k!)^{\nu_i}}$$

and the mean and variance of Y_i are respectively given as:

$$E(Y_i) = (1-\phi)\frac{1}{Z(\lambda_i, \nu_i)} \sum_{j=0}^{\infty} \frac{j\,\lambda^j}{(j!)^{\nu}} \tag{22}$$

$$\text{Var}(Y_i) = (1-\phi)\frac{1}{Z(\lambda_i, \nu_i)} \sum_{j=0}^{\infty} \frac{j^2\,\lambda^j}{(j!)^{\nu}} - E(Y)^2 \tag{23}$$

6.2 *ZICMNB*

The zero-inflated p.m.f for the Zhang *et al.* Com negative binomial distribution is given in (24).

$$\Pr(Y_i = y) = \begin{cases} \phi + (1-\phi)(1-p)^r C^{-1} & \text{if } y = 0 \\ (1-\phi) \quad \left[\dfrac{\Gamma(r+y)}{y!\Gamma(r)}\right]^{\nu} p^y(1-p)^r.\dfrac{1}{C(r,\nu,p)}; & \text{if } y > 1 \end{cases} \tag{24}$$

where

$$C(r,\nu,p)=\sum_{j=0}^{\infty}\left[\frac{\Gamma(r+j)}{j!\Gamma(r)}\right]^{\nu}p^{j}(1-p)^{r}$$

6.3 *ZICOMNB*

The p.m.f of the zero-inflated Yasimoto *et al* Com-Poisson negative binomial model is given as:

$$\Pr(Y_i=y)=\begin{cases}\phi+(1-\phi)H^{-1} & \text{if } y=0\\ (1-\phi)\quad \dfrac{\Gamma(\nu+y)}{\Gamma(\nu)\,{}_1H_{\alpha-1}(\nu,1,p)}.\dfrac{p^{y}}{(y!)^{\alpha}}; & \text{if } y>1\end{cases} \tag{25}$$

where

$${}_1H(\nu,1,p)=\sum_{k=0}^{\infty}\frac{(\nu)_k\,p^{k}}{(k!)^{\alpha}}=\sum_{k=0}^{\infty}\frac{\Gamma(k+\nu)\,p^{k}}{\Gamma(\nu)(k!)^{\alpha}}. \tag{26}$$

The parameter estimates of these models are again obtained using SAS PROC NLMIXED, with the choice of adaptive quadrature as the integration method and the Newton-Raphson being the optimization algorithm. These models are applied to three example data sets having different percentages of zeros.

Example VI

This example is presented in Greenwood and Yule (1920). The data in Table 10 provides the frequency distribution of number of accidents among 647 machine operators in a fixed period. The percentage of zeros in the observed data is 69.1$\%$ while the corresponding percentage under the Poisson model is 62.8$\%$. Thus the data is slightly zero-inflated. In the table below are the results of application of the four distributions with their corresponding zero-inflated models. We have presented the results to two decimal places so that we can accommodate all the distributions in the same table.

For this data set, the NB model is the most parsimonious among the left panel distributions. Kokonendji and Malouche (2008) has employed the Hinde-Demétrio distribution $HD_2(q,\theta)$ to the data. This distribution belongs to the class of *discrete exponential dispersion* model (EDM) and is defined as:

$$f(y;p;\theta,\sigma)=c(y;p;\sigma)\exp\{\theta y-\sigma K_p(\theta)\},y\in S_p \tag{27}$$

Table 10: Distribution of number of accidents among machine operators

		Regular Distributions				Zero-Inflated Models			
Y	Count	NB	CMNB	COMNB	CMP	ZINB	ZICMNB	ZICOMNB	ZICMP
0	447	445.89	445.76	447.16	441.57	447.00	447.03	447.16	447.00
1	132	134.90	135.07	129.76	140.21	130.19	130.19	129.76	130.05
2	42	43.99	43.98	47.71	44.52	47.78	47.73	47.71	47.83
3	21	14.69	14.67	15.91	14.13	15.48	15.48	15.91	15.63
4	3	4.96	4.95	4.76	4.49	4.68	4.69	4.76	4.70
5	2	1.69	1.68	1.29	1.42	1.35	1.36	1.29	1.32
Total	647	646.12	646.12	646.58	646.34	646.48	646.48	646.58	646.53
		$\hat{p}$=0.35	$\hat{r}$=1.00	$\hat{\nu}$=0.35	$\hat{\lambda}$=0.32	$\hat{p}$=0.24	$\hat{p}$=0.23	$\hat{p}$=0.83	$\hat{\lambda}$=0.45
		$\hat{r}$=0.87	$\hat{\nu}$=187.0	$\hat{p}$=0.83	$\hat{\nu} \approx 0.00$	$\hat{r}$=2.09	$\hat{r}$=2.66	$\hat{\alpha}$=1.61	$\hat{\nu}$=0.29
			$\hat{p}$=0.35	$\hat{\alpha}$=1.61			$\hat{\nu}$=0.75	$\hat{\nu}$=0.35	
$\hat{\phi}$						0.286	0.276	0.000	0.24
μ	0.4652								
σ^2	0.6919								
$\bar{y}$		0.4652	0.4652	0.4652	0.4652	0.4652	0.4652	0.4652	0.4652
s^2		0.7154	0.7146	0.6936	0.6817	0.6968	0.6967	0.6935	0.6946
X_g^2		3.9091	3.7853	3.0533	4.5213	3.3066	3.2921	3.0533	3.2205
d.f.		3	2	2	3	2	1	1	2
p-value		0.2714	0.1507	0.2173	0.2104	0.1914	0.0696	0.0806	0.1998
X_W^2		624.8	625.5	644.4	655.7	641.49	641.52	644.47	643.47
d.f.		644	643	643	644	643	642	642	643

where $\theta \in \Theta_p \subseteq \Re$ is the canonical parameter, $\sigma > 0$ is the scale parameter and $c(y; p; \theta)$ is the normalizing constant and $K_p(\theta)$ is the cumulant function. The EDM is characterized by the unit variance function:

$$\mathrm{V}_p(\mu) = \mu + \mu^p, \quad p \in \{0\} \cup [1, \infty)$$

where $\mu > -1$ for $p = 0$ and $\mu > 0$ for $p \geq 1$. When the $HD_2(q, \theta)$ was applied to the above data, the model gives a $X^2 = 4.318$ on 2 d.f. and was considered then, the most parsimonious of the Hinde-Demétrio family of distributions. We observe that all the models considered on the left panel in Table 10 perform better than this. For the zero-inflated right panel models, the ZICMP (zero-inflated Com Poisson) is the most parsimonious. The estimated inflation parameter for the COMNB model is approximately zero and thus the regular model is most appropriate for this data set, and is therefore no surprise that the expected values on both sides of the panel for the COMNB are the same – this yielding the same group X_g^2 but different d.f. since ϕ was estimated from the right panel distribution. However, the COMNB performs better than the Zhang *et al.* CMNB.

Example VII

The data in this example is the Motor vehicle records data set which relates to the number of violation points on the motor vehicle records from Flynn and Francis (2009). This is a skewed distribution with a spike at zero. The zero-inflated Quasi-negative binomial distribution has been shown to fit well the data, especially at the tail end of the data, (Li *et al.*, (2011). We will apply the zero-inflated models discussed in the preceding section to this data and the results are presented in Table 11. The data has a mean $\mu = 1.7100$ and a variance $\sigma^2 = 4.6612$ and thus dispersion index of 2.7258, thus it is highly over-dispersed considering that the sample size is large $n = 10303$. Further the observed proportion of zeros is 45.2% as against expected percentage of zeros under the Poisson model of $100 * \exp(-1.7100)$=18.09%. Thus this data set is highly zero-inflated.

Results here suggest that the ZICOMNB distribution fits best the data. Under this distribution, the estimated inflation factor is quite small, being, $\hat{\phi} = 0.0000113$ and it seems probable that COMNB will fit the data without employing its ZI version. This model fits well because its estimated variance is very close to the observed variance of the data. While all the models produce estimated means close to the observed mean of 1.7100, their estimated variances are not very close to that observed in the data. We also observe that the Wald's statistics for the other three models are mlower than that from the ZICOMNB because their variances are much higher. Clearly, the ZICOMNB performs much better here than the Zhang *et al.* (2018) CMNB distribution.

Table 11: Frequency counts of motor vehicle violations

Y	Count	ZINB	ZICMP	ZICOMNB	ZICMNB
0	4659	4659.000	4659.000	4661.269	4658.991
1	1467	1363.191	1386.909	1435.508	1386.523
2	1199	1314.794	1279.728	1214.254	1280.425
3	966	1041.870	1024.156	995.890	1024.549
4	727	735.959	740.886	749.208	740.785
5	528	481.876	495.580	516.042	495.247
6	341	298.924	310.941	327.547	310.626
7	213	178.091	184.816	193.013	184.629
8	114	102.816	104.819	106.312	104.761
9	53	57.884	57.040	55.059	57.065
10	20	31.924	29.913	26.949	29.973
11	13	17.307	15.171	12.521	15.235
12	1	9.247	7.463	5.543	7.516
13	2	4.880	3.569	2.346	3.608
Total	10303.000	10297.769	10299.991	10301.460	10299.931
μ	1.7100				
σ^2	4.6612				
$\bar{y}$		1.7100	1.7100	1.7057	1.7100
s^2		4.7645	4.7157	4.6662	4.7172
		$\hat{p}$=0.4483	$\hat{\lambda}$=1.1769	$\hat{p}$=2.3226	$\hat{p}$=0.1547
		$\hat{r}$=3.3032	$\hat{\nu}$=0.3511	$\hat{\alpha}$=1.6369	$\hat{r}$=299.74
				$\hat{\nu}$=0.1326	$\hat{\nu}$=0.3564
		$\hat{\phi}$=0.3628	$\hat{\phi}$= 0.3378	$\hat{\phi}$= 0.000	$\hat{\phi}$=0.3382
X^2_W		10078.63	10182.9430	10250.574	10179.68
d.f		10298	10300	10299	10299
X^2		60.4262	33.577	11.556	36.622
d.f		10	10	9	9
p-value		0.0000	0.0007	0.2395	0.0000

Example VIII

The data in Table 12 gives the observed frequency distributions of injury counts for cleaner pre-WRATS* (workplace risk assessment team's) visit. The data was originally analyzed in Carrivick *et al.* (2003). The observed counts in this case are the number of occupational related injuries by cleaners pre-(WRAT) intervention in a 600-bed hospital. Again in this data, we see a preponderance of zero injury-free individuals which is very typical of population-based occupational health data. The observed percentage of zeros is 78.1%, while the observed mean of the data is $\mu = 0.3275$ giving a % of zeros under the Poisson model of 72.1%, thus the data is lightly zero-inflated. The results of applying the NB, CMP, CMNB and COMNB models to the data are presented in the left panel of Table 12. For ZINB, ZICMNB and ZICOMNB, the estimated zero-inflated parameter ϕ are respectively, 0.0000, 0.000178 and 0.000028. Thus, the zero-inflated models are not effective and are therefore not presented in Table 12. However, for the ZICMP model $\hat{\phi} = 0.3362$ and there is considerable improvement in the fit of the ZICMP over the CMP model.

Table 12: Distribution of injury counts and parameter estimates

Y	Count	NB	CMP	CMNB	COMNB	ZICMP
0	267	257.4306	257.6300	267.6528	267.1771	267.0000
1	52	50.6459	63.5563	50.6968	51.9331	50.2232
2	16	15.4747	15.6791	15.1275	14.7217	16.5916
3	4	5.3275	3.8680	5.2518	4.9470	5.4812
4	2	1.9372	0.9542	1.9666	1.8533	1.8107
5	0	0.7269	0.2354	0.7706	0.7518	0.5982
6	0	0.2784	0.0581	0.3113	0.3247	0.1976
7	0	0.1082	0.0143	0.1285	0.1477	0.0653
8	1	0.0425	0.0035	0.0539	0.0702	0.0216
Total	342	341.9720	341.9988	341.9597	341.9266	341.9894
		(0.4571)	(1.2667)	(0.5339)	(0.6159)	(0.2951)
μ	0.3275					
σ^2	0.6197					
$\bar{y}$		0.3275	0.3275	0.3275	0.3266	0.3275
s^2		0.5663	0.4347	0.5771	0.5807	0.5434
		$\hat{p}$=0.4217	$\hat{\lambda}$=0.2467	$\hat{p}$=0.4699	$\hat{p}$=0.2693	$\hat{\lambda}$=0.3304
		$\hat{r}$=0.4491	$\hat{\nu}$=0.000	$\hat{r}$=0.9983	$\hat{\alpha}$=0.7099	
				$\hat{\nu}$=541.88	$\hat{\nu}$=0.7218	$\hat{\nu} \approx$=000
						$\hat{\phi}$=0.3362
X_g^2		1.7593	4.8249	1.5619	1.2954	2.7860
d.f.		5	2	5	5	3
p-value		0.8813	0.0896	0.9058	0.9353	0.4258
X_w^2		373.1587	486.0964	366.1949	363.8851	388.9182
d.f.		339	339	338	338	338

Thus in this case, the, ZINB, ZICMNB and ZICOMNB, are not necessary to explain the variation in this data.

Example IX

Our last example here is the data in Table 13. The data was originally presented in Klugmann *et al.* (2008) and recently analyzed using the negative binomial-Lindley distribution by Zamani and Ismail (2010). The variable Y is the number of accident, while count represents the number of corresponding claims. There is a total of 94,935 claims.

The results of applying the distributions are also provided in the Table. We summarize these results as follows:

Table 13: Models for the frequency counts of insurance claims

Y	Count	Fitting Distributions NB	CMNB	COMNB	CMP
0	81,714	81718.4465	81717.2915	81715.1172	81620.0976
1	11,306	11283.9588	11285.0903	11299.3140	11447.4912
2	1,618	1645.3989	1646.2832	1628.9700	1605.5362
3	250	244.1699	243.6463	245.6414	225.1807
4	40	36.5485	36.3011	38.4814	31.5822
5	7	5.4990	5.4295	6.2262	4.4295
		(6.4773)	(6.3876)	(7.4760)	(5.1521)
Total	94,935	94,934.0217	94934.0419	94933.7502	94934.2774
		(94935)	(94935)	(94935)	(94935)
		$\hat{p}$=0.1536	$\hat{p}$=0.1518	$\hat{p}$=0.1147	$\hat{\lambda}$=0.1403
		$\hat{r}$=0.8993	$\hat{r}$=0.5149	$\hat{\alpha}$=0.8113	
			$\hat{\nu}$=0.1420	$\hat{\nu}$=1.2056	$\hat{\nu} \approx 0.000$
μ	0.1631				
σ^2	0.1929				
$\bar{y}$		0.1631	0.1631	0.1631	0.1631
s^2		0.1927	0.1927	0.1930	0.1897
X_g^2		1.0069	1.1261	0.2452	7.5947
d.f.		3	2	2	3
p-value		0.7996	0.5695	0.8846	0.0552
X_W^2		95039.37	95073.1091	94917.23	96532.62
d.f.		94932	94931	94931	94932
-LL		44764.5	44764.5	44764.0	44767.5

7. Conclusion

Results from our analyses in this paper indicate that the COMNB model proposed in Chakraborty & Imoto (2016) performs in most all cases better than the recently formulated CMNB in Zhang *et al.* (2018) or the CMP model. Our results also confirm that in all cases, the sum of expected values do not necessarily add up to the sample size n and consequently, the estimated means and variances within the observed range of the random variable Y are never equal to the theoretical values. The method of moments employed here, adequately computes the means and variances of these distributions, especially for those distributions that do not have close form expressions. As mentioned earlier, computations of the empirical means and variances give us the opportunity to know how the distributions closely match those of the observed moments in the data. It also allows us to compute the Wald's test statistic, although, models with very high estimated variances will provide lower X_W^2, even the estimated variance deviates grossly from the observed variance. The SAS programs used in all the analyses here are available from the author.

Acknowledgment

The author is grateful to the editor and anonymous reviewers for their helpful comments and suggestion towards the revision of this paper.

References

Aryuyuen. S., Bodhisuwan, W. (2013). The Negative binomial-generalized exponential (NB-GE) distribution. Applied Mathematical Sciences, **7**(22): 1093–1105

Aryuyuen, S., W. Bodhisuwan, and T. Supapakorn (2014). Zero inflated negative binomial-generalized exponential distribution and its applications. Songklanakarin J. Sci. Technol., **36**(4): 483–491.

Beale, E. M. L. (1972). A derivation of conjugate gradients. In: F. A. Lootsma (editor), Numerical Methods for Nonlinear Optimization, London: Academic Press.

Carrivick, P.J.W, Lee, A.H. and Yau, K. K. W. (2003). Zero-inflated Poisson modeling to evaluate occupational safety interventions. Safety Science, **41**: 53–63.

Chakraborty, S. and Imoto, T. (2016). Extended Conway-maxwell-Poisson distribution and its properties and applications. Journal of Statistical Distributions and Applications, **3**(5):1–199. DOI:10.1186/s40488-016-0044-1

Chakraborty, S. and Ong, S.H. (2016). A COM-type generalization of the negative binomial distribution. Communications in Statistics-Theory and Mathods, **45**: 4117-4135.

Conway, R.W. and Maxwell, W.L. (1961). A queuing model with state dependent service rates. Journal of Industrial Engineering, **12**:132–136.

Flynn, M. (2009). More flexible GLMs zero-inflated models and hybrid models. Casualty Acturial Society E-Forum, USA, pp. 148–224.

Geedipally, S.R., Lord, D. and Dhavala, S.S. (2012). The negative binomial-Lindley generalized linear model: characteristics and application using crash data. Accident Analysis & Prevention, **45**(2): 258–265.

Greenwood, M., and Yule, U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Roy. Statist. Soc. Ser. A, **83**:255–279.

Gupta, R.C., Sim, S.Z. and Ong S.H. (2014). Analysis of discrete data by Conway-Maxell Poisson distribution. AstA Adv. Stat. Anal., **98**: 327–343.

Hassan, A. and Bilal, S. (2008). On some properties of quasi-binomial distribution and its applications. Journal of Modern Applied Statistical Methods, **7**(2): 616–631.

Imoto, T. (2014). A generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution. Appl. Math. Comput., **247**: 825–834.

Kadane, J., Shmueli,G., Minka, G., Borle, T. and Boatwright, P. (2006). Conjugate analysis of the Conway Maxwell Poisson distribution. Bayesian Analysis, 1:363–374.

Klugman,S.A.,Panjer, H.H. and Wilmot, G.E. (2008). Loss models: from data to decision (3rd ed.). John Wiley & Sons, USA.

Kokonendji, C.C. and Malouche, D. (2008). A property of count distributions in the Hinde-Demetrio family. Comm. Statist.-Theory and Methods, **37**: 1823–1834.

Lawal, H.B. (1980). Tables of percentage points of Pearson's goodness-of-fit statistic for use with small expectations. Applied Statistics, **29**: 292-298.

Lawal, H.B. (2017). On the negative binomial-generalized exponential distribution and its applications. Applied Mathematical Sciences, **11**: 345–360.

Lawal, H. B. (2017). Correcting for non-sum to 1 estimated probabilities in applications of discrete probability distributions to count data. International Journal of Statistics & Probability, **6**(5): 119–131.

Li, S., Yang,F., Famoye,F., Lee, C., and Black, D. (2011). Quasi-negative binomial distribution: properties and applications. Computational Statistics and Data Analysis, **55**(20): 2363–2371.

Lord, D., and Geedipally, S.R., (2011). The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. Accident Analysis and Prevention, **43**(5): 1738–1742.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. Computer Journal, **7**: 308–312.

Phang, Y.N. and Loh, E.F. (2014). Statistical analysis for overdispersed medical count data. International Scholarly and Scientific Research, **8**(2):292–294.

Pinheiro, J.C. and Bates, D.M.(1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graomegacal Statistics, **4**: 12–35.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. Mathematical Programming, **12**:241–254.

Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. Annals of Applied Statistics, **4**: 943–961.

Sellers, K.F., Borle, S. and Shmueli, G. (2011). The COM-Poisson model for count data: a survey of methods and applications. Applied Stochastic Models in Business and Industry. DOI: 10.1002/asmb.918.

Shmueli, G., Minka, T., Borle, J., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. J. R. Stat. Soc. Series C, **54**: 127–142.

Wang L.M. and Lei, Y.L. (2000). Simulation and EM algorithm for the distribution of number of claim in the heterogenous portfolio. Comm. Appl. Math. Comput. Sci., **14**(2): 71–78.

Yip, K.C.H and Yau, K.K..W.(2005). On modeling claim frequency data in general insurance with extra zeros. Insurance: Mathematics and Economics, **36**: 153–163.

Zamani, H. and Ismail, N. (2010). Negative binomial-Lindley distribution and its application. J. Math. Stat., **6**: 4–9.

Zhang, H., Tan, K. and Li, B. (2018). COM-negative binomial distribution:modeling overdispersion and utrahigh zero-inflated count data. Front. Math. China, **13**(4): 967–998.