International Journal of Statistics and Probability; Vol. 6, No. 5; September 2017
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Correcting for Non-Sum to 1 Estimated Probabilities in Applications of Discrete Probability Models to Count Data

Bayo H. Lawal[1]

[1] Department of Statistics & Mathematical Sciences, Kwara State University, Malete. Nigeria

Correspondence: Department of Statistics & Mathematical Sciences, Kwara State University, Malete. Nigeria. E-mail: bayo.lawal@kwasu.edu.ng

Received: July 10, 2017 Accepted: July 26, 2017 Online Published: August 24, 2017

doi:10.5539/ijsp.v6n5p119 URL: https://doi.org/10.5539/ijsp.v6n5p119

Abstract

In this paper, we examine some often ignored or assumed problems relating with fitting probability models to count data either exhibiting over, equi, or under dispersion. Of particular concern are last category truncated data, where most often, expected values in this last category are collapsed together so that the sum of the expected values sum to the sample size in the data. That is, so that $\sum_{i=0}^{k} \hat{m}_i = n$, the sample size. We shall for illustrative purposes in this paper, consider the following distributions: the negative binomial (NB), the Inverse trinomial (IT), the hyper-Poisson (HP), the Quasi-negative binomial (QNBD), the extended com-Poisson distribution (ECOMP) as well as the negative binomial-exponential distribution (NBGE).Though, we have restricted our discussion to these six distributions, other distributions may also be employed but the patterns are always the same, that is, the sum of the estimated probabilities does not equal 1.00 and consequently, the sum of the expected values is always less or equal (Poisson case only) the sample size in the observed data. We propose a common procedure to rectify this problem for both right truncated or non-truncated frequency count data exhibiting either excess zeros or regular frequency data.

Keywords: Adaptive Gaussian quadrature, cumulative probabilities, Over-dispersion, right truncated, zero-inflated models

1. Introduction

Count data are often modeled with the Poisson distribution-it being the underlying probability model for count data. However, its use had been restricted because of the absence of dispersion parameter in its function since both mean and variance are equal, thus leading to equi-dispersion: the ratio of the variance to the mean, which in this case equals 1.00. For data exhibiting long tails or over or under dispersion, alternative models such as the negative binomial (NB), the generalized Poisson, Famoye *et al.* (2004), the double Poisson and several other distributions, such as, the NB-generalized exponential, Aryuyuen & Bodhisuwan (2013) have been proposed. Other distributions that have received considerable attention in the literature are the Poisson Inverse Gaussian, the NB-Lindley Zamani & Ismail (2010); Lord *et al.* (2011); Geedipally *et.* (2012) and many others.

For frequency count data, all these distributions have something in common when applied to real life data, namely, the sum of estimated probabilities under these models, does not often add to 1 as would be expected. Consequently, the sum of the expected values under these models therefore do not necessarily add up to the sample size n. Most authors just combine the expected values in the last category with the left over, so that $\sum_{i=0}^{k} \hat{m}_i = n$. We shall illustrate this in the next sections.

For this study, we plan to implement a procedure that corrects this anomaly when these distributions are applied to frequency data. Consequently, we will be employing an array of distributions such as the two-parameter type distributions, the negative binomial (NB), the hyper-Poisson and the Com-Poisson (CMP) distributions. We will also employ the three parameter type distributions (the Quasi-negative binomial, the Inverse Tri-nomial and the generalized negative binomial-exponential distribution), as well as the four parameter type distribution (the extended Com-Poisson distribution). These distributions cover a broad spectrum of possible distributions usually employed for count data. These six to seven models will be applied to two example frequency data. The second data set has excess zeros and we would apply our procedure in the light of these excess zeros to the zero-inflated versions of the six models. We present these distributions in the following sections.

2. Probability Models Considered in this Study

We begin our discussion in this paper with brief introductions to some of these distributions.

2.1 The Poisson Distribution

The Poisson distribution has the form:

$$\Pr(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!};\ y = 0, 1, \ldots, \tag{1}$$

$$E(Y) = \lambda; \quad \mathrm{Var}(Y) = \lambda$$

2.2 The Negative Binomial-NB

The negative-binomial model has the probability distribution:

$$\Pr(Y = y) = \binom{n+y-1}{y} p^n (1-p)^y;\ y = 0, 1, \ldots \tag{2}$$

$$E(Y) = nq/p = \mu; \quad \mathrm{Var}(Y) = nq/p^2$$

Thus,

$$\mu = \frac{n(1-p)}{p} \Longrightarrow p = \frac{n}{n+\mu}$$

$$\mathrm{Var}(Y) = \frac{n(1-p)}{p^2} \Longrightarrow \mathrm{Var}(Y) = \mu + \frac{\mu^2}{n}$$

2.3 The Com-Poisson Distribution

For a random variable Y, Shumueli *et al.* (2005) introduced the Conway-Maxwell Poisson (COM-Poisson) distribution defined by:

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

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 distribution when $\nu = 1$. The means and variance of Y are respectively given as:

$$E(Y) = \frac{1}{Z(\lambda, \nu)} \sum_{j=0}^{\infty} \frac{j\,\lambda^j}{(j!)^\nu} \tag{5}$$

and,

$$\mathrm{Var}(Y) = \frac{1}{Z(\lambda, \nu)} \sum_{j=0}^{\infty} \frac{j^2\,\lambda^j}{(j!)^\nu} - E(Y)^2 \tag{6}$$

2.4 The Hyper-Poisson Distribution

The hyper-Poisson (HP) distribution first proposed by Bardwell & Crow (1964) and Crow & Bardwell (1965) is a two-parameter discrete distribution with probability density function (pdf)

$$P(Y = y|\lambda, \beta) = \frac{\Gamma(\beta)}{\Gamma(\beta + y)} . \frac{\lambda^y}{\phi(1, \beta; \lambda)};\ y = 0, 1, \ldots, ; \quad \beta, \lambda > 0 \tag{7}$$

where

$$\phi(1,\beta;\lambda) = \sum_{k=0}^{\infty} \frac{(1)_k}{(\beta)_k}.\frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{\Gamma(\beta)}{\Gamma(\beta+k)}\lambda^k \tag{8}$$

and

$$(\beta)_k = \beta(\beta+1)(\beta+2)\dots(\beta+k-1) = \frac{\Gamma(\beta+k)}{\Gamma(\beta)};\ k = 1,2,\dots,$$

is the confluent hyper-geometric series in which $(\beta)_0 = 1$. Expressions for the mean and variance of the HP distribution are provided in Kumar & Nair (2014) as:

$$\begin{aligned} \mu &= \frac{\phi(2,\beta+1,\lambda)}{\phi(1,\beta,\lambda)}.\frac{\lambda}{\beta} \\ \sigma^2 &= \frac{1}{\beta}\left[\frac{2}{\beta+1}\frac{\phi(3,\beta+2,\lambda)}{\phi(1,\beta,\lambda)} - \frac{1}{\beta}\frac{[\phi(2,\beta+1,\lambda)]^2}{[\phi(1,\beta,\lambda)]^2}\right]\lambda^2 + \mu \end{aligned} \tag{9}$$

where:

$$\begin{aligned} \phi(1,\beta;\lambda) &= \sum_{k=0}^{\infty} \frac{\Gamma(\beta)}{\Gamma(\beta+k)}\lambda^k \\ \phi(2,\beta+1;\lambda) &= \sum_{k=0}^{\infty} \frac{(k+1)\,\Gamma(\beta+1)}{\Gamma(\beta+k+1)}\lambda^k \\ \phi(3,\beta+2;\lambda) &= \sum_{k=0}^{\infty} \frac{(k+2)(k+1)}{2}.\frac{\Gamma(\beta+2)}{\Gamma(\beta+k+2)}\lambda^k \end{aligned}$$

Alternatively, Lawal (2017) has used the expressions below to obtain the mean and variance of the HPP distribution. His results agree with using expressions in (9).

$$\begin{aligned} E(Y) &= \sum_{j=0}^{\infty} jP(Y=y|\lambda,\beta) \\ \text{Var}(Y) &= \sum_{j=0}^{\infty} j^2P(Y=y|\lambda,\beta) - [E(Y)]^2 \end{aligned}$$

2.5 The Quasi-Negative Binomial-QNBD

The quasi-negative binomial distribution Janardan (1975); Sen & Jain (1996) and very recently by Li *et al.* (2011) has the probability mass function Hassan & Bilal (2008) of the form:

$$P(Y=y) = \binom{a+y-1}{y}\frac{\theta_1(\theta_1+\theta_2 y)^{y-1}}{(1+\theta_1+\theta_2 y)^{a+y}}, \quad y = 0,1,\dots, \tag{10}$$

This is equivalent to QNBD proposed in Li *et al.* (2011) which has the alternative probability mass function:

$$P(Y=y) = \begin{cases} \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\left(\frac{1}{1+cy}\right)\left(\frac{1+cy}{1+b+cy}\right)^y\left(\frac{b}{1+b+cy}\right)^\alpha, & y = 0,1,\dots \\ 0 & \text{for } y > m \text{ if } c < 0 \end{cases} \tag{11}$$

The properties of both alternative distributions have been outlined in the papers referred to above. In this paper, we will employ the QNBD defined in (10).

2.6 The Inverse-Trinomial Distribution-IT

The inverse trinomial distribution Shimizu & Yanagimoto (1991) which is derived from the Lagrangian expression has the probability mass function of the form :

$$P(Y=y) = \frac{\lambda p^\lambda q^y}{y+\lambda}\sum_{t=0}^{\lfloor y/2\rfloor}\frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\frac{pr}{q^2}\right)^t \tag{12}$$

y=0,1,...; $\lambda > 0, p \geq r$ and $p + q + r = 1$. It is so named because its cumulant generating function is the inverse of that for the trinomial distribution. The IT model was employed for overdispersed medical count data by Phang & Ong (2013). It is a special case of the negative binomial distribution (NB)

$$LL = \log(\lambda) + \lambda\log(p) + y\log(q) - \log(y+\lambda) + \log Q(y,\lambda)$$

where

$$Q(y,\lambda) = \sum_{t=0}^{\lfloor y/2\rfloor} \frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\frac{pr}{q^2}\right)^t.$$

The mean and variance of the IT are, for $p > r$ given respectively as:

$$E[X] = \lambda\left\{\frac{1-(p-r)}{p-r}\right\} = \mu \tag{13a}$$

$$\text{Var}[X] = \frac{\lambda}{(p-r)^2}\left\{1-(p-r)+\frac{2r}{p-r}\right\} = \sigma^2 \tag{13b}$$

2.7 The Negative Binomial-Generalized Exponential Distribution NB-GE

The NB-GE distribution with parameters r, α, β is a mixture of the NB and the generalized exponential exponential distributions, viz:

$$Y|\pi \sim NB(r, \pi = \exp(-\lambda)), \quad \text{and} \quad \lambda \sim GE(\alpha,\beta)$$

with the resulting unconditional pmf being given by:

$$f(y;r,\alpha,\beta) = \binom{r+y-1}{y}\sum_{j=0}^{y}(-1)^j\binom{y}{j}\left[\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)}\right] \tag{14}$$

for $y = 0, 1, \ldots,$ and $r, \alpha, \beta > 0$.

The means and variances of the NB-GE distribution in (14) are:

$$E(Y) = r(\delta_1 - 1) \qquad \text{and} \qquad \text{Var}(Y) = r\delta_2(r+1) - r\delta_1(1+r\delta_1) \tag{15}$$

where

$$\delta_1 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{1}{\beta})}{\Gamma(\alpha-\frac{1}{\beta}+1)}, \quad \text{and} \quad \delta_2 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{2}{\beta})}{\Gamma(\alpha-\frac{2}{\beta}+1)}.$$

2.8 The Extended COM-Poisson (ECOMP) Distribution

The pmf of a random variable Y having the extended COM-Poisson distribution with parameters $(\nu, \alpha, , \beta)$ is given in Chakraborty & Imoto (2016) by:

$$f(y;\nu,\alpha,\beta) = \frac{[(\nu)_y]^\beta}{{}_1S^\beta_{\alpha-1}(\nu,1;p)}.\frac{p^y}{(y!)^\alpha} = \frac{[\Gamma(\nu+y)]^\beta}{[\Gamma(\nu)]^\beta\,{}_1S^\beta_{\alpha-1}(\nu,1;p)}.\frac{p^y}{(y!)^\alpha} \tag{16}$$

where

$${}_1S^\beta_{\alpha-1}(\nu,1;p) = \sum_{j=0}^{\infty}\frac{[\Gamma(\nu+j)]^\beta}{[\Gamma(\nu)]^\beta}.\frac{p^j}{(j!)^\alpha}$$

The distribution is defined in the parameter space

$$\Theta_{ECOMP} = \{\nu \geq 0, p > 0, \alpha > \beta\} \cup \{\nu > 0, 0 < p < 1, \alpha = \beta\}$$

To re-emphasize our problem, one common feature of the distributions described above, and indeed for most distributions employed for count regression models is that they all defined to have infinite range. Consequently, for a real life data that takes values $Y = 0, \ldots, k$, it is most common to observe that the expected probabilities under any of the above models are not necessarily summing to 1.00 within the range $0 \leq Y \leq k$ as expected for a probability mass function, and consequently, the expected values will also not sum to n, the sample size. To overcome this, the practice has often been to add this shor fall expected values to the last category expected value, that is, category k in our case.

While this practice is most common, there exists however situations, where such a practice may not lead to the right decision being taken. The case in point is if the last category, k has been truncated. That is, we have observations designated in $k+$ categories. We give example below, which is adapted from Hassan & Bilal (2008) and relates to the number of absenteeism among shift workers in steel industry as reported in Arbous & Sichel (1954). The category 25+ actually stands for counts in categories (25-48) and are combined in that category with a frequency count of 16. Thus our last category is $k = 25$ in this data. The distribution of Y and the corresponding frequencies are displayed in columns 1 and 2 respectively in Table 1.

Table 1. Distribution of absenteeism among Shift Workers

		P		NB		CP		HPP		QNBD	
Y	Count	Fit	Cum	Fit	Cum	Fit	Cum	Fit	Cum	Fit	Cum
0	7	0.0344	0.034	10.7965	10.796	14.4774	14.477	17.178	17.178	12.963	12.963
1	16	0.3056	0.340	15.6027	26.399	16.0126	30.490	16.986	34.163	15.875	28.838
2	23	1.3573	1.697	17.8064	44.206	16.7297	47.220	16.674	50.838	16.910	45.748
3	20	4.0190	5.716	18.5175	62.723	16.9060	64.126	16.250	67.088	17.086	62.833
4	23	8.9254	14.642	18.3189	81.042	16.6849	80.811	15.724	82.812	16.775	79.608
5	24	15.8569	30.499	17.5657	98.608	16.1674	96.978	15.107	97.919	16.165	95.774
6	12	23.4763	53.975	16.4875	115.095	15.4329	112.411	14.412	112.331	15.367	111.141
7	13	29.7917	83.767	15.2369	130.332	14.5463	126.957	13.653	125.984	14.454	125.594
8	9	33.0802	116.847	13.9158	144.248	13.5610	140.518	12.843	138.827	13.474	139.068
9	9	32.6504	149.497	12.5917	156.840	12.5206	153.039	11.999	150.826	12.464	151.533
10	8	29.0035	178.501	11.3086	168.148	11.4604	164.499	11.133	161.959	11.450	162.983
11	10	23.4218	201.923	10.0937	178.242	10.4081	174.907	10.259	172.217	10.452	173.435
12	8	17.3381	219.261	8.9629	187.205	9.3850	184.292	9.390	181.607	9.483	182.919
13	7	11.8474	231.108	7.9238	195.129	8.4070	192.699	8.536	190.144	8.555	191.474
14	2	7.5172	238.625	6.9788	202.107	7.4852	200.184	7.709	197.853	7.674	199.148
15	12	4.4517	243.077	6.1262	208.234	6.6268	206.811	6.916	204.768	6.846	205.993
16	3	2.4716	245.549	5.3623	213.596	5.8357	212.647	6.163	210.931	6.073	212.067
17	5	1.2915	246.840	4.6816	218.278	5.1136	217.761	5.456	216.388	5.358	217.425
18	4	0.6373	247.477	4.0781	222.356	4.4598	222.220	4.799	221.187	4.701	222.126
19	2	0.2980	247.775	3.5451	225.901	3.8724	226.093	4.194	225.381	4.102	226.228
20	2	0.1323	247.908	3.0761	228.977	3.3482	229.441	3.642	229.023	3.558	229.786
21	5	0.0560	247.964	2.6647	231.642	2.8833	232.324	3.142	232.166	3.068	232.854
22	5	0.0226	247.986	2.3048	233.946	2.4735	234.798	2.694	234.859	2.630	235.484
23	2	0.0087	247.995	1.9908	235.937	2.1143	236.912	2.295	237.154	2.240	237.724
24	1	0.0032	247.998	1.7173	237.654	1.8009	238.713	1.943	239.097	1.896	239.620
25+	16	0.0011	247.999	1.4797	239.134	1.5288	240.242	1.635	240.732	1.594	241.214
$F(y)$		0.9999		0.9643		0.9687		0.9707		0.9726	
ML Estimates	-	$\hat{\mu} = 8.8831$		$\hat{\mu} = 8.8827$, $\hat{k} = 0.5794$		$\hat{\lambda} = 1.1060$, $\hat{\nu} = 0.0822$		$\hat{\lambda} = 135.36$, $\hat{\beta} = 136.89$		$\hat{\alpha} = 1.3447$, $\hat{\theta}_1 = 7.9775$, $\hat{\theta}_2 = -0.0936$	
X_T^2		1439.5486		234.1975		246.5021		247.1028		na	
X_E^2		1439.6407		365.1146		344.8319		325.6193		339.4635	
d.f.		247		246		246		246		245	
X_G^2		4099.00		31.9082		38.5179		42.5328		38.6978	
d.f.		24		23		23		23		22	

Also presented in Table 1, are the expected values, ML estimated parameters, both true Wald's Goodness-of-Fit test statistic, X_T^2, its corresponding empirical value X_E^2 and the grouped X^2 values with their corresponding degrees of freedom for five of the models, namely, (P, NB, CMP, HPP, and QNBD). We shall discuss the estimation procedures employed for these models but first, we observe the following from our example data in Table 1.

(a) Apart from the Poisson model, all the other models have $\sum_{i=0}^{25} \hat{m}_i < 248$, the sample size $n = 248$. Corresponding fits of IT and ECOMP gives $\sum_{y=0}^{25} \hat{m}_i$ of 237.707 and 239.3451 respectively. In all the cases, $\sum_{y=0}^{25} \hat{m}_i < n$, where $n = 248$ for this data set. The cumulative estimated probabilities under each model is given by the $F(y)$.

For the NB and QNBD for instance, these sums are respectively, 239.134 and 241.214 at $k = 25$. Thus if we add these shortfalls to the last category for instance, we will get the following for {NB, CMP, HPP, QNBD}={10.3457, 9.2868, 8.903, 8.380}. The corresponding contributions to X_G^2 in this case are {3.0903, 4.8528, 5.6574, 6.9289}. Here,

$$X_G^2 = \sum_{k=1}^{25} \frac{(f_k - \hat{m}_k)^2}{\hat{m}_k}.$$

(b) The parameter estimates under each model are provided under the panel 'ML estimates'. Clearly, each $F(y) < 1.00$ in all cases.

(c) The mean and variance of the observed data are respectively, $\mu = 8.8831$ and $\sigma^2 = 51.7717$, giving a dispersion parameter DP=5.8281. Thus the data is highly over-dispersed.

(d) In Table 2 are presented the following:

Table 2. Statistics Computed from implementation of the designated five Models

	Models				
Param	P	NB	COMP	HPP	QNBD
F(25)	0.9999	0.9643	0.9687	0.9707	0.9726
F(y*)	F(25)	F(98)	F(78)	F(58)	F(51)
μ	8.8831	8.8831	8.8831	8.8831	8.9061
σ^2	8.8831	54.6018	51.8763	51.7502	48.7628
$\bar{y}$	8.8830	7.7448	7.9222	8.0064	8.00967
$\hat{\sigma}^2$	8.8825	35.9036	37.7476	39.8570	38.2273

1. $F(25) < 1$ for all models
2. $F(y*)$ gives the value of Y required to attain an F(y) to be 1. Thus, k has to be 25, 98, 78, 58 and 51 respectively for these estimated probabilities to sum to 1 and of course in this case, the expected values also sum to $n = 248$. When we realize that the range of k here is $0 \leq k \leq 25$ we can see that we need to make the necessary adjustments to realize our expectations.
3. The estimates μ and σ^2 are the theoretical means and variances under these models for the data. These are used to compute Wald's X_T^2, where $X_T^2 = \sum_{i=1}^{248} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$. It is based on $248 - p - 1$ d.f.
4. The estimated $\bar{y}$ and $\hat{\sigma}^2$ are the estimated means and variances at $k = 25$, representing reality from the data and are thus classified as empirical estimates on which the Wald's test statistic X_E^2 are computed, where $X_E^2 = \sum_{i=1}^{248} \frac{(y_i - \bar{y})^2}{\hat{\sigma}_i^2}$. It is based on $n - p - 1$ degrees of freedom. Here p is the number of parameters being estimated.
5. For the grouped data, we employed Pearson's $X_G^2 = \sum_{k=0}^{25} \frac{(f_k - \hat{m}_k)^2}{\hat{m}_k}$ statistic. However, we ascertain that the expected values satisfy the Lawal (1980) rule for the χ^2 approximation to be valid. The model is based on $k - p$ degrees of freedom.
6. We see that why the theoretical means and variances (apart from the Poisson, which has a much lower variance) of the models are much closer to the observed mean and variance of the data, however, when the models are implemented, the empirical means and variances of all the distributions do not match up with the observed values. The estimated variances grossly underestimates the true variance of the data. These we plan to correct with the new procedure described at a later section of this paper.

3. Estimation

The log-likelihood of a single observation i from P, NB, HPP, COMP, IT, ECP, QNBD, NBGE distributions are given in expressions (17a) to (17h) respectively:

$$\text{LL1} = -\lambda + y\log(\lambda) - \log(y!) \tag{17a}$$

$$\text{LL2} = \log\Gamma(y_i + \frac{1}{k}) - \log\Gamma(y_i + 1) - \log\Gamma(\frac{1}{k}) + y_i\log(k\mu_i) - (y_i + \frac{1}{k})\log(1 + k\mu_i) \tag{17b}$$

$$\text{LL3} = y_i\log(\lambda) + \log\Gamma(\beta) - \log\Gamma(y_i + \beta) - \log\left[\sum_{k=0}^{\infty}\frac{\Gamma(\beta)}{\Gamma(\beta+k)}\lambda^k\right] \tag{17c}$$

$$\text{LL4} = y_i\log\lambda_i - \nu\log y_i! - \log\left[\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}\right]. \tag{17d}$$

$$\text{LL5} = \log(\lambda) + \lambda\log(p) + y\log(q) - \log(y+\lambda) + \log\left[\sum_{t=0}^{[y/2]}\frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\frac{pr}{q^2}\right)^t\right] \tag{17e}$$

$$\text{LL6} = y_i\log(y_i) + \beta\log\Gamma(\nu + y_i) - \beta\log\Gamma(\nu) - \alpha\log(y!) - \log\left[\sum_{j=0}^{\infty}\frac{[\Gamma(\nu+j)]^\beta}{[\Gamma(\nu)]^\beta}.\frac{p^j}{(j!)^\alpha}\right] \tag{17f}$$

$$\begin{aligned}\text{LL7} = {} & \log(a+y-1)! - \log(y!) - \log(a-1)! + \log(\theta_1) + (y-1)\log(\theta_1 + \theta_2 y) \\ & - (a+y)\log(1 + \theta_1 + \theta_2 y)\end{aligned} \tag{17g}$$

$$\text{LL8} = \sum_{i=1}^{n}\log[\Gamma(r+y_i) - \Gamma(r) - \Gamma(y_i+1)] + \sum_{i=1}^{n}\log\left(\sum_{j=0}^{y}(-1)^j\binom{y}{j}\left[\frac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)}\right]\right) \tag{17h}$$

Maximum-likelihood estimations of the above models are carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y,\Theta)$ over the parameter space Θ numerically. The integral approximations in PROC NLMIXED is the Adaptive Gaussian Quadrature Pinheiro & Bates (1995) and the Newton-Rapson Conjugate Gradient optimization algorithm in PROC NLMIXED (**NEWRAP**). To obtain a quicker convergence, the Conjugate Gradient or quasi-Newton optimization algorithms of Powell (1977) and Beal (1972) are initially employed in our computations (the latter is the default in PROC NLMIXED). Convergence is often a major problem here and the choice of starting values is very crucial. For each of the cases considered here, the above two initial optimizing algorithms were applied in turn to ascertain accuracy and consistency. Although the results differ very slightly, on the whole, they all agree very well. Thus, we may note here that each of these give slightly different parameter estimates. They all give values that are very close.

4. New Procedure

We can overcome the above subjective approach by instead fit a model with the log-likelihood of a single observation i of the form:

$$LL = (1-\delta)\log[P(Y_i = y_i)] + \delta\log[P(Y_i \geq k)], \tag{18}$$

where

$$\delta = \begin{cases} 0 & \text{for } y_i < k \\ 1 & \text{for } y_i \geq k \end{cases}.$$

and k is the last category of the data. In our example in Table 1, this would be $k = 25$. To accomplish this, first we compute:

$$\omega = \sum_{j=0}^{k-1} f(y_j) \tag{19}$$

for each of the distributions NB, CP, HPP, ECP, QNBD, IT and NBGE, The expression in (18) for the HPP model for example would be:

$$\text{LL} = (1-\delta)(LL) + \delta\ \log(1-\omega) \tag{20}$$

where LL represents the log-likelihood LL3 in (17c) for the Hyper-Poisson model. Thus, for the Hyper-Poisson model, the above in (20) becomes:

$$\begin{aligned} LL = &(1-\delta)[y_i \log(\lambda) + \log\Gamma(\beta) - \log\Gamma(y_i+\beta) - \log(\phi(\beta,\lambda)] \\ &+ \delta \log(1-\omega) \end{aligned} \tag{21}$$

Here,

$$\phi(\beta,\lambda) = \sum_{k=0}^{\infty} \frac{\Gamma(\beta)}{\Gamma(\beta+k)}\lambda^k \quad \text{and} \quad \omega = 1 - \frac{1}{\phi(\beta,\lambda)}.\sum_{j=0}^{k-1}\frac{\Gamma(\beta)}{\Gamma(\beta+j)}\lambda^j$$

Once the parameters λ and β are successfully estimated, the estimated probabilities and expected values are computed viz:

$$\hat{p}_i = \begin{cases} \exp\{i\log(\hat{\lambda}) + \log\Gamma(\hat{\beta}) - \log\Gamma(i+\hat{\beta}) - \log\hat{\phi}\} & \text{for } 0 \le i \le (k-1) \\ \exp\{\log(1-\hat{\omega})\} & \text{for } i = k \end{cases} \tag{22}$$

Here again,

$$\hat{\phi}(\hat{\beta},\hat{\lambda}) = \sum_{k=0}^{\infty} \frac{\Gamma(\hat{\beta})}{\Gamma(\hat{\beta}+k)}\hat{\lambda}^k \quad \text{and} \quad \hat{\omega} = 1 - \frac{1}{\hat{\phi}(\hat{\beta},\hat{\lambda})}.\sum_{j=0}^{k-1}\frac{\Gamma(\hat{\beta})}{\Gamma(\hat{\beta}+j)}\hat{\lambda}^j$$

The results of implementing this approach (sometimes referred to as *right truncated*) is presented in Table 3 for the seven models. Clearly now, the estimated probabilities sum to 1.00 in the range $0 \le Y \le 25$ and consequently, the expected values also sum to n. Chakraborty & Imoto (2016) have used a similar procedure.

Table 3. Results from the Implementation of the Procedure

		Models						
Y	Count	NB	CP	ECP	QNB	HPP	IT	NBGE
0	7	12.355	16.684	6.907	9.305	18.970	9.769	8.685
1	16	16.257	17.226	18.486	15.874	18.137	16.405	15.784
2	23	17.677	17.160	20.867	19.203	17.272	19.364	19.596
3	20	17.874	16.740	20.620	20.187	16.385	19.927	20.764
4	23	17.392	16.088	19.355	19.718	15.484	19.177	20.255
5	24	16.526	15.285	17.710	18.448	14.575	17.796	18.837
6	12	15.452	14.386	15.974	16.803	13.667	16.172	17.019
7	13	14.279	13.432	14.287	15.044	12.766	14.514	15.103
8	9	13.078	12.455	12.711	13.319	11.879	12.930	13.256
9	9	11.896	11.479	11.272	11.707	11.012	11.467	11.559
10	8	10.760	10.522	9.977	10.245	10.169	10.145	10.044
11	10	9.687	9.598	8.821	8.943	9.356	8.963	8.715
12	8	8.689	8.715	7.796	7.798	8.575	7.915	7.562
13	7	7.767	7.881	6.890	6.798	7.830	6.990	6.567
14	2	6.924	7.100	6.091	5.930	7.122	6.176	5.713
15	12	6.157	6.373	5.387	5.178	6.455	5.460	4.981
16	3	5.464	5.701	4.767	4.528	5.829	4.831	4.353
17	5	4.839	5.085	4.223	3.967	5.244	4.278	3.814
18	4	4.279	4.521	3.743	3.483	4.700	3.792	3.352
19	2	3.778	4.009	3.321	3.064	4.197	3.365	2.953
20	2	3.330	3.546	2.950	2.701	3.734	2.989	2.610
21	5	2.933	3.128	2.623	2.387	3.311	2.657	2.313
22	5	2.580	2.752	2.334	2.114	2.924	2.365	2.056
23	2	2.267	2.417	2.080	1.877	2.574	2.107	1.832
24	1	1.990	2.117	1.855	1.670	2.257	1.879	1.638
25+	16	13.771	13.600	16.953	17.710	13.578	16.568	18.638
Total	248	248.00	248.00	248.00	248.00	248.00	248.00	248.00
ML Estimates	-	$\hat{\mu} = 9.3189$ $\hat{k} = 0.6527$	$\hat{\lambda} = 1.0325$ $\hat{\nu} = 0.0517$	$\hat{\nu} = 0.1684$ $\hat{p} = 0.7388$ $\hat{\beta} = -0.7225$ $\hat{\alpha} = -0.7738$	$\hat{\alpha} = 2.5405$ $\hat{\theta}_1 = 2.6409$ $\hat{\theta}_2 = 0.0802$	$\hat{\lambda} = 243.71$ $\hat{\beta} = 254.90$	$\hat{p} = 0.3394$ $\hat{q} = 0.5610$ $\hat{\lambda} = 2.9931$	$\hat{\alpha} = 18.7176$ $\hat{\beta} = 2.6710$ $\hat{r} = 3.0601$
$\bar{y}$	8.8831	8.9438	8.9411	8.8435	8.8101	8.9358	8.8617	8.7979
s^2	51.7717	48.7326	50.5937	50.3520	49.8801	52.1445	50.1559	50.4271
AIC		1520.5	1527.0	1518.0	1518.4	1530.7	1517.2	1518.3
X_E^2		262.4224	252.7676	253.9202	256.3934	245.2475	254.9596	256.9035
d.f.		246	246	244	245	246	245	245
X_G^2	-	29.2406	34.6457	24.8582	28.0935	37.6379	25.7397	28.7451
d.f.		24	24	22	23	24	23	23

We see here from Table 3 that the sums of expected values all sum to 248, the sample size as expected. Further, the empirical means and variances are much closer to the true values of 8.8831 and 51.7717 than in the earlier models. Consequently, the empirical Wald's GOF X_E^2 are much smaller than those observed in Table 1. Also, the grouped Pearson's X_G^2 fit much better. The ECP gives the lowest values but it is based on 22 d.f., while the Inverse tri-nomial (IT) gives a X_G^2 of 25.7397 on 23 d.f. and more parsimonious.

4.1 Data Set II

As a second example, the data set in Table 4, is taken from Aryyuen & Bodhisuwan (2013) and gives the number of hospital stays by United States residents aged 66 and above. The data was originally presented in Flynn (2009) but recently re-analyzed in Aryyuen & Bodhisuwan (2013) with zero inflated models. Here, Y_i is the number of hospital stays and n_i is the frequency in each category. The sample size here is n = 4406.

Table 4. Parameter Estimates and Expected values under the various Models

		Expected Values Under the Six Models					
Y	n_i	NB	COMP	HPP	QNBD	IT	NBGE
0	3541	3544.41	3399.79	3400.15	3540.940	3542.80	3540.87
1	599	583.49	776.42	776.28	601.642	593.76	601.74
2	176	177.49	177.31	177.19	165.996	169.93	166.44
3	48	62.25	40.49	40.433	56.836	59.25	56.58
4	20	23.28	9.25	9.22	22.185	23.01	21.96
5	12	9.03	2.11	2.10	9.495	9.56	9.39
6	5	3.59	0.48	0.48	4.361	4.16	4.34
7	1	1.45	0.11	0.11	2.120	1.87	2.13
8	4	0.59	0.03	0.03	1.080	0.86	1.10
Total	4406	4405.0000	4405.9926	4405.993	4404.655	4405.220	4404.547
ML Estimates		$\hat{r}$ = 2.6957 $\hat{\lambda}$ = 0.2960	$\hat{\lambda}$ = − $\hat{\nu}$ = 0.000	$\hat{\lambda}$ = 929.97 $\hat{\beta}$ = 4073.34	$\hat{a}$ = 14.1597 $\hat{\theta}_1$ = 0.0156 $\hat{\theta}_2$ = 0.0165	$\hat{p}$ = 0.6568 $\hat{q}$ = 0.3231 $\hat{\lambda}$ = 0.5187	$\hat{\alpha}$ = 0.5625 $\hat{\beta}$ = 8.6148 $\hat{r}$ = 3.4023
-2LL		6019.2	6136	6136.2	6015.0	6015.7	6015.0
AIC		6023.2	6140	6140.2	6021.0	6021.7	6021.0
μ	0.2960	0.2960	0.2960	0.2958	0.2960	0.2959	0.2960
σ^2	0.5571	0.5321	0.3834	0.3832	0.5601	0.5454	0.5629
$\bar{y}$		0.2950	0.2960	0.2958	0.2928	0.2942	0.2925
s^2		0.5235	0.3834	0.3831	0.5270	0.5287	0.5259
W_E		4687.75	6400.47	6406.15	4657.11	4642.10	4402.31
W_T		4612.19	6398.20	6403.82	4381.33	4499.86	4360.07
d.f.		4403	44403	4403	4402	4402	4402

As in the previous example, under the models considered for this data, again, the sums of the expected frequencies in each model do not add to n = 4406. Both the estimated variances of the Com-Poisson and Hyper-Poisson models grossly underestimate the variance of the observed data. However, the NG, IT, QNB and the NBGE models give estimated variances that are closer to the observed variance of 0.5571 in the data, than the HPP, and CP models. The estimated means of all the models are very close to the observed mean of 0.2960. It is noted here that, both the QNB and the NBGE fit the data best. The dispersion parameter is 1.89 for this data, which by the size of the data gives significant indication of over-dispersion in the data. This is not necessarily unexpected because of the excess zeros in the data. We examine the effect of this in the next section.

When the procedure outlines earlier is employed on these data, the corresponding results for the right truncated models are displayed in Table 5.

Table 5. Results from the Implementation of the Procedure

Y	Count	Models						
		NB	CP	ECP	QNB	HPP	IT	NBGE
0	3541	3544.833	3471.876	3541.399	3540.051	3399.002	3543.400	3540.399
1	599	581.911	696.730	599.950	606.002	776.922	591.788	604.724
2	176	177.693	168.623	167.748	162.758	177.526	170.126	163.459
3	48	62.624	45.537	56.936	55.230	40.551	59.667	55.546
4	20	23.544	13.291	21.594	21.837	9.260	23.324	21.940
5	12	9.184	4.121	8.830	9.651	2.114	9.757	9.670
6	5	3.669	1.342	3.818	4.651	0.482	4.274	4.638
7	1	1.490	0.456	1.725	2.403	0.110	1.936	2.380
8	4	1.052	4.025	4.000	3.415	0.033	1.728	3.246
Total	4406	4406.00	4406.00	4406.00	4406.00	4406.00	4406.00	4406.00
ML Estimates	-	$\hat{\mu} = 0.2966$ $\hat{k} = 2.7203$	$\hat{\lambda} = 0.2007$ $\hat{\nu} = -0.2703$	$\hat{\nu} = 0.9222$ $\hat{p} = 0.3574$ $\hat{\beta} = 9.2240$ $\hat{\alpha} = 9.0506$	$\hat{\alpha} = 6.2721$ $\hat{\theta}_1 = 0.0355$ $\hat{\theta}_2 = 0.0330$	$\hat{\lambda} = 699.490$ $\hat{\beta} = 3060.250$	$\hat{p} = 0.6527$ $\hat{q} = 0.3270$ $\hat{\lambda} = 0.5107$	$\hat{\alpha} = 0.6943$ $\hat{\beta} = 6.7101$ $\hat{r} = 2.1974$
$\bar{y}$		0.2964	0.2923	0.2959	0.2962	0.2962	0.2964	0.2962
s^2		0.5329	0.4650	0.5542	0.5588	0.3838	0.5444	0.5575
-2LL		6014.9	6039.3	6007.7	6007.9	6134.2	6010.5	6007.9
AIC		6018.9	6043.3	6015.7	6013.9	6138.2	6016.5	6013.9
X_E^2		4605.3459	5278.2617	4428.1272	4391.6755	6393.4352	4508.0241	4402.3110
d.f.		4403	4403	4401	4402	4403	4402	4402
X_G^2		14.2428	44.6147	3.7369	3.7768	640.527	7.1245	3.7780
d.f.		6	6	4	5	6	5	5

Results from Table 5 again indicate that the right-truncated models behave much better than those in Table 4. The estimated probabilities sum to 1.00, as well as the estimated frequencies summing to 4402. For both data considered here, the QNB, IT, the NB-GE, and ECP all give means and variances that are very close to the means and variances of the observed data. Also for both data sets, the HPP and Com-Poisson give estimated variances that grossly underestimate the true variances and both therefore are not very good for modeling these data sets. The ECP, QNB and NBGE all fit the data very well, but the QNB model is the most parsimonious for this data set. It also gives the lowest empirical Wald's test statistic of 4391.6755 on 4402 d.f.

5. Zero-Inflated Models

We present in this section the effect of applying the procedure employed in the last section to data exhibiting excess zeros (like the data in Table 4) where 80.4% of the data are zeros. Lawal (2017) has fitted the zero-inflated negative binomial-generalized exponential distribution (ZINBGE) to the accident data in Table 4. We present here the results of fitting the ZINB, the ZIIT, the ZIQNB and the ZINBGE to the data in Table 4. To accomplish these, we recall that a zero-inflated (ZI) model 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:

$$\Pr(Y = y|\phi) = \begin{cases} \phi + (1-\phi)\ \Pr(Y=0) & \text{if } y_i = 0 \\ (1-\phi)\ \Pr(Y = y_i)) & \text{if } y_i = 1, 2, \cdots \end{cases} \tag{23}$$

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, ZIIT, ZIQNB and ZINBGE models are given respectively in expressions (24) to (27).

$$\Pr(Y_i = y_i) = \begin{cases} \phi + (1-\phi)(1+k\mu_i)^{-k^{-1}}, & y_i = 0 \\ (1-\phi)\dfrac{\Gamma(y_i + k^{-1})}{y_i!\ \Gamma(k^{-1})}\dfrac{(k\,\mu_i)^{y_i}}{(1+k\mu_i)^{y_i+k^{-1}}} & y_i > 0 \end{cases} \tag{24}$$

$$\Pr(Y = y) = \begin{cases} \phi + (1-\phi)\, p^{\lambda} & \text{if } y = 0 \\ (1-\phi)\dfrac{\lambda p^{\lambda} q^{y}}{y+\lambda}\displaystyle\sum_{t=0}^{\lfloor y/2 \rfloor}\dfrac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\dfrac{pr}{q^2}\right)^{t} & \text{if } y > 0 \end{cases} \tag{25}$$

$$\Pr(Y=y)=\begin{cases}\phi+(1-\phi)\left[\dfrac{1}{(1+\theta_1)^a}\right] & \text{if } y=0\\[2ex] (1-\phi)\dbinom{a+y-1}{y}\dfrac{\theta_1(\theta_1+\theta_2 y)^{y-1}}{(1+\theta_1+\theta_2 y)^{a+y}}, & \text{if } y>1\end{cases} \tag{26}$$

$$\Pr(Y_i=y)=\begin{cases}\phi+(1-\phi)\left(\dfrac{\Gamma(\alpha+1)\Gamma(1+\frac{r}{\beta})}{\Gamma(\alpha+\frac{r}{\beta}+1)}\right), & \text{if } y=0\\[2ex] (1-\phi)\dbinom{r+y-1}{y}\displaystyle\sum_{j=0}^{y}(-1)^j\binom{y}{j}\left(\dfrac{\Gamma(\alpha+1)\Gamma(1+\frac{r+j}{\beta})}{\Gamma(\alpha+\frac{r+j}{\beta}+1)}\right) & \text{if } y>1\end{cases} \tag{27}$$

From the above, it is not too difficult to formulate the corresponding log-likelihoods. The results of implementing these models are given in the first panel in Table 6.

Table 6. MLEs, expected values and empirical means and variances for the four models

		Models				Right Truncated ZI Models			
Y	n_i	ZINB	ZIIT	ZIQNB	ZINBGE	ZINB	ZIIT	ZIQNB	ZINBGE
0	3541	3544.410	3542.802	3541.076	3541.017	3544.834	3543.403	3541.007	3541.024
1	599	583.484	593.762	600.810	600.846	581.910	591.786	601.901	602.106
2	176	177.494	169.931	167.372	167.802	177.692	170.125	166.721	165.793
3	48	62.251	59.254	56.702	56.362	62.624	59.667	55.428	55.792
4	20	23.281	23.014	21.841	21.669	23.544	23.324	21.388	21.750
5	12	9.032	9.562	9.273	9.243	9.184	9.757	9.298	9.487
6	5	3.588	4.159	4.260	4.282	3.669	4.274	4.452	4.522
7	1	1.449	1.870	2.090	2.122	1.490	1.936	2.307	2.315
8	4	0.592	0.862	1.085	1.113	1.052	1.728	3.499	3.210
Total		4405.580	4405.216	4404.510	4404.456	4406.00	4406.000	4406.000	4406.00
MLE		$\hat{\phi}=0.0000$	$\hat{\phi}=0.0000$	$\hat{\phi}=0.0060$	$\hat{\phi}=0.1647$	$\hat{\phi}=0.0000$	$\hat{\phi}=0.0000$	$\hat{\phi}=0.2180$	$\hat{\phi}=0.1445$
		$\hat{\mu}=0.2960$	$\hat{p}=0.6568$	$\hat{\alpha}=0.5252$	$\hat{\alpha}=1.0345$	$\hat{\mu}=0.2966$	$\hat{p}=0.6527$	$\hat{\alpha}=1.8224$	$\hat{\alpha}=2.0183$
		$\hat{k}=2.6957$	$\hat{q}=0.3231$	$\hat{\theta}_1=0.5203$	$\hat{\beta}=7.3548$	$\hat{k}=2.7203$	$\hat{p}=0.3270$	$\hat{\theta}_1=0.1719$	$\hat{\beta}=5.9202$
			$\hat{\lambda}=0.5187$	$\hat{\theta}_2=0.0509$	$\hat{r}=2.2017$		$\hat{\lambda}=0.5107$	$\hat{\theta}_2=0.0580$	$\hat{r}=1.1005$
μ	0.2960								
σ^2	0.5571								
$\bar{y}$		0.2950	0.2942	0.2924	0.2923	0.2964	0.2965	0.2961	0.2961
s^2		0.5235	0.5287	0.5244	0.5243	0.5329	0.5444	0.5568	0.5556
-2LL		6019.2	6015.7	6015.1	6015.0	6014.9	6010.5	6007.4	6007.7
AIC		6025.2	6023.7	6023.1	6023.0	6020.9	6018.5	6015.4	6015.7
W_E		4687.7627	4612.1862	4340.5217	4680.9653	4605.3546	4508.0324	4407.4183	4416.9216
d.f.		4403	4402	4402	4402	4402	4402	4402	4402
X^2		25.4209	15.4115	11.2699	10.8060	14.2428	7.1246	3.2806	3.5316
d.f.		6	5	5	5	5	5	5	5

For the data in this example, its mean and variance are respectively, 0.2960 and 0.5571, with a dispersion parameter DP=1.88 which indicates over dispersion. This is not surprising because of the excess zeros in the data. We see again here for the first part of the results that none of the zero-inflated models have expected values summing to $n = 4406$. Clearly, for this data set, both ZINB and ZIIT are clearly not improving our fits since the estimates of ϕ in both cases are zero. We might as well do with NB and IT models. The ZIQNB and the ZINBGE both fit the data better with grouped X^2 being respectively, 11.2699 and 10.8060 on 5 d.f. The equivalent Wald's test statistic X_W^2 are 4340.5217 and 4680.9653, indicating that the ZIQNB now fits better, each is of course on 4402 degrees of freedom. Because the sum of the $\hat{m}_i < n = 4406$ in all the four models, it is often the case that the last categories usually carries the difference. Thus for instance, most analysis in the literature would assign an expected value of $0.592 + (4406 - 4405.216) = 0.592 + 0.784 = 1.376$ under the ZINB model. We can do the same for the other three models to ensure that the sums all add to $n = 4406$ in this case.

We also observe that the empirical means and variances for the four models fall short of the true values for the data, hence the models did not fit very well. We have employed here the Lawal (1980) rule of the χ^2 approximation to X^2. These are all satisfied in our data. As discussed in the previous section, the results in the second half of the table designated **right truncated** are the same models with truncation at $Y = 7$. The results are much better for these models based on the groups' X^2. For instance, for the ZIQNB, this is 3.2806 on 5 d.f. In particular, both ZIQNB and ZINBGE fit very well. The reasons for this can be traced to their empirical means and variances which are very close to that observed for the data. Further, both the Akaike Information Criterion (AIC) and -2log-likelihood (-2LL) are lowest for both models. Consequently the Wald's test statistics are much lower for the two models. Under the ZIQNB model, the corresponding parameter estimates with the Li *et al.* model would be $\hat{c} = 5.8174$ and $\hat{b} = 0.3374$.

6. Conclusions

The study here indicates that for all models usually considered for modeling count data (the exception being the Poisson in a few situations), the estimated probabilities often do not sum to 1 and consequently, the sum of the expected values does not necessarily sum to the sample size n in the observed data. We have demonstrated here that a right-truncated model approach works better that the usual addition to last category expected values that often is the norm in the literature. This procedure will particularly be better for right truncated long right tail data, where it is sometimes necessary to truncate at some values of $Y = y_t$.

A major problem with estimating the MLE for these distributions is setting the initial values to be used in the optimization algorithm in SAS PROC NLMIXED to achieve convergence. However, by first using the conjugate gradient optimization algorithm **congra**(it converges faster than the Newton-Rapson), one is able to get good initial value estimates.

The SAS programs for implementing these models are available from the author.

References

Arbous, A. G., & Sichel, H. S. (1954). New techniques for the analysis of absenteeism data. *Biometrika*, 41, 77–90.

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

Bardwell, G. E., & Crow, E. L. (1964). A two parameter Family of Hyper-Poisson Distributions. *Journal of the American Statist. Assoc.*, *59*, 133–141. https://doi.org/10.1080/01621459.1964.10480706.

Beale, E. M. L. (1972). *A Derivation of Conjugate Gradients. In: F. A. Lootsma, ed., Numerical Methods for Nonlinear Optimization*, London: Academic Press.

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

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

Crow, E. L., & Bardwell, G. E. (1965). Estimation of the parameters of the Hyper-Poisson Distributions. *In: Patil, G. P. (ed) Classical and Contagious Discrete Distributions*, 27–140. Pergamon, Oxford.

Famoye, F., Wulu, J. J., & Singh, K. P. (2004). *On the generalized Poisson regression model with an application to accident data*. *Journal of Data Science*, *2*, 287–295.

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

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

Guikema, S. D., & Coffelt, J. P. (2008). *A flexible count data regression model for risk analysis*. *Risk Analysis*, *28*(1), 213–223.

Hassan, A., & Bilal, S. (2008). On some properties of Quasi-Binomial Distribution and Its Applications. *Journal of Modern Applied Statistical Methods*, *7*(2), 616–631.

Janardan, K. G. (1975). *Markov-Polya urn-model with predetermined strategies*. *Gujarat Statist. Rev.*, *2*(1), 17–32.

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

Khazraee, S. H., Castillo-Sáez, A. J., Geedipally, S. R., & Lord, D. (2014). Application of the hyper-Poisson generalized linear model for analyzing motor vehicle crashes.

Kumar C. S., & Nair, B. U. (2014). A three parameter hyper-Poisson distribution and some of its properties. *Statistica*, 183–198.

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.

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

Lord, D., & 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., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–312.

Phang, Y. N., & Loh, E. F. (2014). Statistical Analysis for overdispersed medical Count Data. *International Scholarly and Scientific Research*, *8*(2), 292–294.

Phang, Y. N., Sim, S. Z., & Ong, S. H. (2013). Statistical Analysis for the inverse Trinomial Distribution. *Communications in Statistics-Simulation and Computation*, *42*(9), 2073–2085.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.

Powell, M. J. D. (1977). *Restart Procedures for the Conjugate Gradient Method*, *Mathematical Programming*, *12*, 241–254.

Sáez-Castilo, A. J., & Conde-Sánchez, A. (2015). *A hyper-Poisson regression model for Overdispersed and Underdispersed Count Data*. *Computational Statistics and Data Analysis*. https://doi.org/10.1016/j.csda.2012.12.009.

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

Sellers, K. F., Borle, S., & Shmueli, G. (2011). The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 104–116. https://doi.org/10.1002/asmb.918.

Sen, K., & Jain, R. (1996). Generalized Markov-Polya urn-model with pre-determined strategies. it Jour. Statist. Plann. Infer., *54*, 119–133.

Shimizu, K., & Yanagimoto, T. (1991). *The inverse trinomial Distribution*. *Japan Journal of Applied Statistics*, *20*, 89–96.

Shmueli, G., Minka, T., Borle, J., & 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.

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

Copyrights