BENIN JOURNAL OF
STATISTICS
ISSN 2682-5767
Vol. 6, pp. 22– 43 (2023)

Modeling Over-dispersion with two-Parameter Discrete Distributions

B. H. Lawal

LOYAB Statistical Consulting & Data Analytics
Adeleke Road, Lamodi, Offa, Kwara State, Nigeria.

(*Received: 24 August 2022; Accepted: 15 March 2023*)

Abstract. In this paper we compare the performances of several two parameter discrete distributions in modeling over-dispersed count data. Particularly, we are interested in the performances of two newly proposed two parameter distributions, namely, the ATPPSD 'another two-parameter Poisson-Sujatha Distribution' ATPPSD (Shanker *et al.*, 2020) and the Bell-Tuchard (BT) distribution with some of other well known two-parameter discrete distributions: the negative binomial (NB); the generalized Poisson (GP), the Poisson-Lindley (GPL), the discrete Weibull (DW) and the Poisson-inverse Gaussian (PIG) distributions. These distributions are applied to a variety of data sets exhibiting over dispersion. The two distributions perform poorly in most of the data set examples. Zero-inflated (ZI) versions of the models are also implemented because the regular models perform poorly with data exhibiting excess zeros.In most cases, the two distributions grossly underestimate the observed variances in the data sets and this subsequently lead to their poor fitting performances. The PIG and DW distributions will be suitable alternative models to the NB and GP models for modeling over dispersed count data.They perform in many cases better than the NB and GP models but the latter two models are very reliable and they both perform very well in most of the examples. For moderately over-dispersed data, the ATPPSD and BT distributions seem to do well.

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

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

Classification codes: 60E07; 60A05; 62F10

1. Introduction

For count data exhibiting over or under dispersion, the most often used discrete distributions are the negative binomial (NB), the generalized Poisson (GP) and other two parameter based distributions such as the Sankaran (1970) Poisson-Lindley (GPL), the Com-Poisson (Shmueli *et al.*, 2005; Sellers *et al.*, 2012)

Corresponding author. Email: blawal66@gmail.com

(CP) , the Bardwell & Crow (1964); Lawal (2017b) Hyper-Poisson (HP), the discrete Weibull (DW), and the Holla (1966) Poisson Inverse Gaussian (PIG) amongst several others.

These distributions all have extra dispersion parameters that transform their variance functions from linear (as in the case of Poisson) to quadratic or cubic functions. For instance, the variance functions of the NB and GP are respectively, $\mu(1+k\mu)$ and $\mu(1+\alpha\mu)^2$. All these alternative distributions to the one-parameter Poisson have received considerable attention in the literature.

Recently, several two-parameter mixture distributions with the Sunjatha distributions are introduced. These include,the Poisson-Sujatha (PSD) (Shanker, 2016), a generalization of the Poisson-Sujatha distribution (AGPSD) (Shanker and Shukla, 2019), the two-parameter Poisson-Sujatha distribution (TPPSD) proposed in Shanker and Shukla (2020) and the discrete Poisson-Akash distribution (Shanker, 2017) to name just a few, Our discussion in this paper however, focuses on the performances of these two-parameter distributions with the well established ones, such as the NB and the GP distributions. We particularly compare the 'another two-parameter Poisson-Sujatha Distribution' which is appropriately called ATPPSD(θ,α) and recently proposed in Shanker *et al.* (2020) and the Bell-Tuchal (Castellares *et al.*, 2019) distributions. The ATPPSD having a random variable Y, is a mixture of the Poisson $P(\lambda)$ and ATPSD(θ,α) distributions. That is,

$$Y|\lambda \sim \text{Poisson}(\lambda), \quad \text{and} \quad \lambda|\theta,\alpha \sim \text{ATPSD}(\theta,\alpha)$$

Thus,

$$\begin{aligned} f(y|\lambda) &= \frac{e^{-\lambda}\lambda^y}{y!}, \quad y=0,1,\ldots,\ \lambda>0;\ \text{and} \\ f(\lambda|\theta,\alpha) &= \frac{\theta^3}{\theta^2+\alpha\theta+2\alpha}(1+\alpha\lambda+\alpha\lambda^2)e^{-\theta\lambda} \end{aligned} \tag{1}$$

where $\lambda>0, \theta>0, \alpha>0$.

The ATPSD is the Mussie and Shanker (2019) 'another two-parameter Sujatha distribution'.

The mixture expression above can be succinctly written as $P(\lambda) \wedge ATPSD(\theta,\alpha) \sim$ ATPPSD. In this case, the mixture model is obtained by evaluating the well-known integral:

$$f(y) = \int_0^\infty f(Y|\lambda) f(\lambda|\theta,\alpha) d\lambda \tag{2}$$

The resulting unconditional pmf from (2) being therefore, 'another two-parameter Poisson-Sujatha Distribution' (ATPPSD) defined as

$$f(y|\theta,\alpha) = \left[\frac{\theta^3}{\theta^2+\alpha\theta+2\alpha}\right]\left[\frac{\alpha(y^2+\theta+3)+\alpha y(\theta+4)+(\theta^2+2\theta+1)}{(\theta+1)^{y+3}}\right],\ y=0,1,\ldots \tag{3}$$

for $\theta > 0$ and $\alpha \geq 0$.
Aderoju (2020) called this distribution the new generalized Poisson-Sujatha distribution (NGPSD). We will however refer to this distribution here as the ATPPSD.
The mean and variance of the distribution is given respectively in Shanker *et al.* (2020) as:

$$\mu = \left[\frac{\theta^2 + 2\alpha(\theta+3)}{\theta[\theta^2 + \alpha(\theta+2)]}\right] \tag{4a}$$

$$\sigma^2 = \left[\frac{\theta^2(\theta+2) + 2\alpha\{12 + 6\theta + \theta^2)\}}{\theta^2\{\theta^2 + \alpha(\theta+2)\}}\right] - \mu^2 \tag{4b}$$

with (4b) reducing to:

$$\sigma^2 = \frac{\theta^4(\theta+1) + \alpha\theta^2(16 + 12\theta + 3\theta^2) + 2\alpha^2(6 + 12\theta + 6\theta^2 + \theta^3)}{\theta^2[\theta^2 + \alpha(2+\theta)]^2} \tag{5}$$

The various distributional properties of this distribution have been fully discussed in Shanker *et al.* (2020). When $\alpha = 1$ the distribution reduces to the Poisson-Sujatha distribution (PSD) with pmf (Shanker, 2016b):

$$P(Y|\theta) = \left[\frac{\theta^3}{\theta^2 + \theta + 2}\right]\left(\frac{y^2 + (\theta+4)y + (\theta^2 + 3\theta + 4)}{(\theta+1)^{y+3}}\right); \quad y = 0, 1, \ldots, \ \theta > 0 \tag{6}$$

The ATPPSD reduces to the geometric when $\alpha = 0$.

1.1 *The Discrete Bell-Touchard Distribution*

The Bell-Touchard discrete distribution (Castellares *et al.*, 2019) has the pmf for a random variable Y having a BT distribution given by:

$$f(y|\alpha, \theta) = \frac{e^{\theta(1-e^{\alpha})}\alpha^y T_y(\theta)}{y!}, \quad y = 0, 1, 2, \ldots \tag{7}$$

where $\alpha > 0, \theta > 0$ and $T_y(\theta)$ are the Touchard polynomials defined as:

$$T_n(\theta) = e^{-\theta}\sum_{k=0}^{\infty}\frac{k^n\theta^k}{k!} \tag{8}$$

such that $T_0(\theta) = 1, T_1(\theta) = \theta$ and so on. When $\theta = 1$, then we have

$$T_n(1) = B_n = \frac{1}{c}\sum_{k=0}^{\infty}\frac{k^n}{k!}$$

where $_n$ are the Bell numbers (Comtet, 2010) with for instance, $B_0 = B_1 = 1; B_2 = 2, \ldots, B_{10} = 115975$ and so on. Its mean and variance (Castellares *et al.*, 2019) are given respectively in (9)

$$\mu = \theta\alpha e^{\alpha} \tag{9a}$$

$$\sigma^2 = \theta(1+\alpha)\alpha e^{\alpha} \tag{9b}$$

The dispersion index (DI) is $1+\alpha$ and since $\alpha > 0$, thus, DI > 1 which implies that the distribution will be most suitable for over-dispersed count data.
In this paper, we will also explore two-parameter distributions NB, GP, and GPL with the TPPSD but with eight different sets of data which include the distribution of epileptic seizure data employed in Aderoju (2020). These distributions are briefly described below:

1.2 *The Negative Binomial-NB*

The Negative binomial distribution (NB) has the probability mass function (pmf):

$$f(y; r, p) = \frac{\Gamma(r+y)}{y!\Gamma(r)} p^y (1-p)^r, \quad y = 0, 1, \ldots \tag{10}$$

where $r \in (0, \infty) > p$ and $p \in (0, 1)$. The mean and variance of the NB model with parameters r and p in (10) are given respectively in (11a) and (11b) respectively.
Hence,

$$\mu = rp/(1-p) \quad \Longrightarrow \; p \; = \frac{\mu}{r+\mu} \tag{11a}$$

$$\sigma^2 = rp/(1-p)^2 \quad \Longrightarrow \sigma^2 = \mu + \frac{\mu^2}{r} \tag{11b}$$

Of course the NB is a mixture of the Poisson-Gamma distributions.

2. Materials and Method

2.1 *The generalized Poisson-Lindley (GPL) Distribution*

The GPL having $Y \sim GPL(\alpha, \gamma = 1, \theta)$, proposed in Mahmoudi and Zakerzadeh (2010) has the pmf given by:

$$f(y; \alpha, \gamma = 1, \theta) = \frac{\Gamma(y+\alpha)\theta^{\alpha+1}\left(\alpha + \frac{y+\alpha}{1+\theta}\right)}{y!\Gamma(1+\alpha)(1+\theta)^{y+\alpha+1}} \tag{12}$$

It is a mixture of Poisson distribution and the generalized Lindley (GL) distribution (Zakerzadeh and Dolati, 2009). Its moments are:

$$E(Y) = \frac{\alpha(1+\theta)+1}{\theta(1+\theta)} = \mu \tag{13a}$$

$$E(Y^2) = \mu + \frac{(\alpha+1)[\alpha(1+\theta)+2]}{\theta^2(1+\theta)} - \mu^2 \tag{13b}$$

Hence, variance is:

$$\sigma^2 = \frac{\alpha(\theta+1)^3 + \theta^2 + 3\theta + 1}{\theta^2(\theta+1)^2} \tag{14}$$

and the Dispersion Index is:

$$1 + \frac{\alpha(\theta+1)^2 + 2\theta + 1}{\alpha\theta(\theta+1)^2 + \theta(\theta+1)}$$

indicating over-dispersion for values of α and θ, with equi-dispersion occurring if

$$\frac{\alpha(\theta+1)^2 + 2\theta + 1}{\alpha\theta(\theta+1)^2 + \theta(\theta+1)} = 0$$

2.1.1 The Generalized Poisson Distribution-GP:

The type I generalized Poisson regression (GPI) model has the following pmf:

$$\Pr(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left\{-\frac{\mu_i(1+\alpha y_i)}{(1+\alpha\mu_i)}\right\}, \quad y_i = 0, 1, \ldots \tag{15}$$

with mean

$$\mathrm{E}(Y_i) = \mu_i; \quad \text{and} \quad \mathrm{Var}(Y_i) = \mu_i(1+\alpha\mu_i)^2. \tag{16}$$

Consul and Famoye (1992) have also considered the GPI model for over-dispersed data because like the NB model, the GP also has a dispersion parameter α. The GP reduces to the Poisson when $\alpha = 0$.

2.1.2 The Poisson-inverse Gaussian (PIG) Distribution

The Poisson-inverse Gaussian (PIG) distribution was introduced by Willmot (1987) and has the pmf:

$$f(y|\mu,\beta) = \begin{cases} p_0 & y = 0 \\ \dfrac{p_0\mu^y}{y!} \displaystyle\sum_{k=0}^{y-1} \frac{(y-1+k)!}{(y-1-k)!\,k!} \left(\frac{\beta}{2\mu}\right)^k (1+2\beta)^{-\left(\frac{y+k}{2}\right)}, & y = 1, 2, \ldots \end{cases} \tag{17}$$

where $p_0 = \exp\left\{\frac{\mu}{\beta}[1-(1+2\beta)^{1/2}]\right\}$, with $\mu > 0$ and $\beta > 0$.
Its mean an variance are given respectively as:

$$E(Y) = \mu \quad \text{and} \quad \sigma^2 = \mu(1+\beta) \tag{18}$$

and since $\beta > 0$, hence the dispersion index (DI) > 1. Thus, the PIG would be most suitable for over-dispersed count data.

2.1.3 The Discrete Weibull Distribution

The discrete Weibull distribution was introduced by Nakagawa and Osaki (1975) as a discrete counterpart of the continuous Weibull distribution and is usually referred to as 'type I discrete Weibull distribution', in order to distinguish it from two other models proposed later by Stein and Dattero (1984) (type II discrete Weibull) and Padgett and Spurrier (1985) (type III discrete Weibull). It is derived from the continuous Weibull distribution with probability mass function (pmf) given by

$$f_t(t;\lambda,\beta) = \lambda\beta t^{\beta-1}e^{-\lambda t^{\beta}} \tag{19}$$

for $\lambda, \beta > 0$ and the corresponding cdf is:

$$F_t(t;\lambda,\beta) = 1 - e^{-\lambda t^{\beta}} \tag{20}$$

Following(italia), for a random variable Y = $\lfloor T \rfloor$, where $\lfloor T \rfloor$ is the largest integer equal or smaller than the r.v T in (19), then the pmf defined on the non-negative integers only can be shown to be:

$$\begin{aligned} P(y;q,\beta) = F_t(y+1) - F_t(y) \qquad &= e^{-\lambda y^{\beta}} - e^{-\lambda(y+1)^{\beta}} \\ = q^{y^{\beta}} - q^{(y+1)^{\beta}} \; y = 0,1,\ldots \end{aligned} \tag{21}$$

where $q = e^{-\lambda}$ and $0 < q < 1$. The model in (20) is the type I Discrete Weibull distribution, proposed in Nakagawa and Osaki (1975). It has the cumulative distribution function (cdf) given by:

$$F(y;q,\beta) = \begin{cases} 1 - q^{(y+1)^{\beta}} & \text{for } y = 0,1,2,\ldots \\ 0 & \text{if } y < 0. \end{cases}$$

Some properties of the $DW(q,\beta)$ are,

- $\Pr(0) = 1 - q$. Thus, when q is small, then we would have an excess zero.

- The dispersion index DI $= \frac{\sigma^2}{\mu}$ can be: underdispersed, overdispersed or equi-dispersed for $DI < 1$, $DI > 1$ or $DI = 1$ respectively.

Other properties of the $DW(q, \beta)$ are succinctly described in Kalktawi *et al.* (2015).
The mean and variance of the DW do not have closed form expressions, however, the mean and variance can be computed from the following infinite sums viz:

$$\mathrm{E}(Y) = \sum_{y=1}^{\infty} q^{y^\beta} \tag{22a}$$

$$\mathrm{E}(Y^2) = 2\sum_{y=1}^{\infty} y q^{y^\beta} - \mathrm{E}(Y) \tag{22b}$$

The expression in (22a) for instance leads to a closed expression if and only if $\beta = 1$, in which case $\mathrm{E}(Y) = \frac{q}{1-q}$. From (22), we observe that E(Y), for a fixed q is a decreasing function of β. Khan *et al.*(1989) have shown that

$$\mathrm{E}(Y) < \mathrm{E}(T) = \left(-\frac{1}{\log q}\right)^{\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right) < E(Y) + 1 \tag{23}$$

3. Estimation

For a single observation i, the log-likelihood for the ATPPSD, Bell-Tuchard, GPL, NB,GP, DW and PIG models are presented respectively in LL1 to LL7 in (24).

$$\begin{aligned}
\mathrm{LL1} &= 3\ \log(\theta) + \log\ \left[\alpha(y_i^2 + \theta + 3) + \alpha y_i(\theta + 4) + (\theta^2 + 2\theta + 1)\right] \\
&\quad - \log\ (\theta^2 + \alpha\theta + 2\alpha) - (y_i + 3)\ \log\ (\theta + 1) && \text{(24a)} \\
\mathrm{LL2} &= \theta + [1 - \exp(\alpha)] + y\log(\alpha) + \log[T_y(\theta)] - \log y! && \text{(24b)} \\
\mathrm{LL3} &= \log[\Gamma(y + \alpha)] + (\alpha + 1)\log(\theta) + \log\left[\alpha + \frac{y + \alpha}{1 + \theta}\right] \\
&\quad - \log y! - \log[\Gamma(1 + \alpha)] - (y + \alpha + 1)\log(1 + \theta) && \text{(24c)} \\
\mathrm{LL4} &= \log[\Gamma(ry)] + y\log(p) + r\log(1 - p) - \log y! - \log[\Gamma(r)] && \text{(24d)} \\
\mathrm{LL5} &= y_i\ \log\left(\frac{\mu_i}{1 + \alpha\mu_i}\right) + (y_i - 1)\log(1 + \alpha y_i) - \frac{\mu_i(1 + \alpha y_i)}{1 + \alpha\mu_i} - \log(y_i!) && \text{(24e)} \\
\mathrm{LL6} &= \log\left[q^{y^\beta} - q^{(y+1)^\beta}\right] && \text{(24f)} \\
\mathrm{LL7} &= \begin{cases} \log(p0) & \text{if } y = 0 \\ \log(p0) + y\log(\mu) - \log(y!) + \log(Q) & \text{if } y > 0 \end{cases} && \text{(24g)}
\end{aligned}$$

where,

$$Q = \sum_{k=0}^{y-1} \frac{(y-1+k)!}{(y-1-k)!\,k!} \left(\frac{\beta}{2\mu}\right)^k (1+2\beta)^{-\left(\frac{y+k}{2}\right)}$$

and p0 is as defined earlier.
Maximum-likelihood estimation of (24) is 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 and Bates, 1995) and our choice optimization algorithm here is the Newton-Raphson techniques.
- Can also be implemented in R using package *optim*

4. Applications

The above models are applied to the distribution of epileptic seizures (Chakraborty, 2010). The data has $n = 351$ observations with observed mean $\mu = 1.5442$ and corresponding observed variance being $\sigma^2 = 2.8830$. Consequently, the dispersion index (DI) is $1.8671 > 1$, indicating that the data is over-dispersed. Our results for the implementation of these distributions are presented in Table 1. We summarize the results from the Table as follows:

Table 1: Observed and Expected values under the various Models

Y	count	P	NB	GP	GPL	ATPPSD	BT	PIG	DW
0	126	74.9354	120.2197	118.1122	121.5086	122.2712	125.8531	114.1046	120.1194
1	80	115.7122	92.9960	95.8102	91.4895	89.5487	80.3830	100.7742	92.8754
2	59	89.3391	59.1732	59.8862	58.7114	58.8075	61.1512	61.4528	59.0357
3	42	45.9846	34.9447	34.4855	35.0930	35.7925	38.5677	33.7758	35.1327
4	24	17.7519	19.8372	19.2356	20.0986	20.6229	22.0839	18.1988	20.0714
5	8	5.4823	10.9888	10.5921	11.1793	11.4049	11.7983	9.8892	11.1307
6	5	1.4109	5.9867	5.8060	6.0859	6.1103	5.9538	5.4638	6.0289
7	4	0.3112	3.2224	3.1806	3.2586	3.1922	2.8638	3.0723	3.2024
8	3	0.0601	1.7187	1.7447	1.7218	1.6339	1.3223	1.7558	1.6727
Total	351.00	350.9878	349.0873	348.8532	349.1467	349.3841	349.9772	348.4873	349.2694
		(0.0122)	(1.9127)	(2.1468)	(1.8533)	(1.6159)	(2.3450)	(2.5127)	(1.7306)
$Y \leq y_a$		12	31	32	30	30	29	47	30
		$\hat{\mu}$=1.5442	$\hat{p}$=0.4990	$\hat{\mu}$=1.5442	$\hat{\alpha}$= 1.2920	$\hat{\alpha}$=1.3156	$\hat{\alpha}$=0.8828	$\hat{\mu}$=1.5442	$\hat{q}$=0.6578
			$\hat{r}$=1.5501	$\hat{\tau}$=0.2705	$\hat{\theta}$=1.1390	$\hat{\theta}$=1.3716	$\hat{\theta}$=0.7235	$\hat{\beta}$=1.0285	$\hat{\beta}$=1.1561
μ	1.5442								
σ^2	2.8830								
$\bar{y}$		1.5442	1.5442	1.5442	1.5447	1.5448	1.5442	1.5442	1.5429
s^2		1.5442	3.0825	3.1038	3.0928	3.0433	2.9073	3.1323	3.0374
X_g^2		117.8674	5.6656	7.1279	5.0862	4.2040	2.5588	10.5297	5.4867
d.f		5	6	6	6	6	6	6	6
p-value		0.0000	0.4617	0.3092	0.5328	0.6491	0.8618	0.1040	0.4831
X_W^2		653.4723	327.4076	325.1018	326.2649	331.5707	347.0760	322.1478	332.2126
d.f.		349	348	348	348	348	348	348	348
-2LL		1272.1	1189.9	1191.7	1189.2	1188.1	1185.8	1195.6	1189.5
AIC		1274.1	1193.9	1195.7	1193.2	1192.1	1189.8	1199.6	1193.5
BIC		1278.0	1201.6	1203.4	1200.9	1199.8	1197.5	1207.3	1201.2

From the results in Table 1, we can make comparisons with results presented in Aderoju (2020), viz:

(a) The results for the NB presented in Aderoju (2020) are completely wrong. The parameter estimates under the NB are as presented in Table 1 together with the correct expected frequencies, which lead to grouped Pearson's

$X^2 = 5.6656$ on 6 d.f. (p-value=0.4617), thus a good fit. Aderoju has reported X_g^2 to be 22.53 on 6 d.f.

All the X_g^2 computed here have taken into consideration the problem of small expected frequencies and has accordingly applied the Lawal (1980) rule with the appropriate d.f.

(b) The sum of the expected values under each of the models do not add to $n = 351$. Consequently, we can add the differences (presented in parentheses) to the last category to make the sum in each add to 351. Thus for the ATPPSD for instance, this would be (1.6339+1.6159=3.2498). Lawal (2017a) has provided an alternative way of handling this situation which is peculiar to all count model distributions.

(c) We observe here that all the models produced estimated means that are very close to the observed mean, but the estimated variances (with the exception of the Poisson) are all higher than the observed variance of 2.8830. The BT has a theoretical variance of 2.9073 for this data set and this value is the closest to the observed variance of 2.8830 amongst the distributions, and also produces an estimated mean of 1.5442.

(d) The Wald's test statistics $X_W^2 = \sum_{i=1}^{351} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}_i^2}$ is lowest for the PIG model because it has the largest estimated variance of 3.1323.

(e) In general, all these other models fit the data except the Poisson, however, the Bell-Touchard (BT) distribution is the most parsimonious for this data set based on the grouped X_g^2 GOF of 2.5588 on 6 d.f (pvalue=0.8618). This is closely followed by the ATPPSD.

(f) The parameter q in the discrete Weibull is modeled here in the logit form.

As observed in Lawal (2017a, 2019), one common feature of all discrete distributions for frequency count regression models is that they all have infinite range. Consequently for real life data, like the data in Table 1, where, $Y = 0, \ldots, 8$ for example, it is most common to observe that estimated probabilities under any of the above models are not necessarily summing to 1.00 within the range $0 \leq Y \leq 8$ 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 the shortfall expected values to the last category expected value, that is category 8 in our case.

The ATPPSD and BT as discrete probability distribution are no exceptions to this problem of estimated probabilities not summing to 1.00 within the range of actual data. We present below the following results under the implementation of the ATPPSD model to the data in Table 1.

Here, under each of the estimated models, the likelihoods are obtained and the corresponding expected probabilities computed. With these, the means are obtained as $\sum_{j=0}^{k} j\hat{p}_j$, with, $E(Y^2) = \sum_{j=0}^{k} j^2\hat{p}_j$, and hence the corresponding variance of Y.

We present in Table 2 these computations for the ATPPSD model, where, $\hat{\pi}_j$ is

the estimated probability at $Y = j$, $\sum_{i\leq j}\hat{\pi}_i$ are the cumulative probabilities. Similarly, $\hat{m}_j$ and $\sum_{i\leq j}\hat{m}_i$ are the predicted expected values and the corresponding cumulative expected frequencies respectively. $\hat{\mu}_j$, vv and $\hat{\sigma}_j^2$ are the expressions $\sum_{j=0}^{k} j\hat{\pi}_j$, $E(Y^2) = \sum_{j=0}^{k} j^2\hat{\pi}_j$, and variance of Y respectively.

Table 2: Moments Computation under ATPPSD Model

j	$\hat{\pi}_j$	$\sum_{i\leq j}\hat{\pi}_i$	$\hat{m}_j$	$\sum_{i\leq j}\hat{m}_i$	$\hat{\mu}_j$	vv	$\hat{\sigma}_j^2$
0	0.34835	0.34835	122.2712	122.2712	0.000000	0.000000	0.000000
1	0.25512	0.60348	89.5487	211.8198	0.255124	0.255124	0.190036
2	0.16754	0.77102	58.8075	270.6273	0.590210	0.925295	0.576948
3	0.10197	0.87299	35.7925	306.4198	0.896129	1.843052	1.040005
4	0.05875	0.93175	20.6229	327.0427	1.131148	2.783128	1.503633
5	0.03249	0.96424	11.4049	338.4477	1.293611	3.595446	1.922016
6	0.01741	0.98165	6.1103	344.5580	1.398061	4.222144	2.267569
7	0.00909	0.99074	3.1922	347.7502	1.461724	4.667781	2.531146
8	0.00466	0.99540	1.6339	349.3841	1.498964	4.965707	2.718813**
9	0.00234	0.99774	0.8223	350.2064	1.520048	5.155465	2.844918
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.00001	0.99999	0.0050	350.9957	1.544543	5.425638	3.040025
17	0.00001	0.99999	0.0023	350.9980	1.544655	5.427546	3.041586
18	0.00000	1.00000	0.0011	350.9991	1.544710	5.428541	3.042411
19	0.00000	1.00000	0.0005	350.9996	1.544737	5.429054	3.042841
20	0.00000	1.00000	0.0002	350.9998	1.544751	5.429317	3.043063
21	0.00000	1.00000	0.0001	350.9999	1.544757	5.429450	3.043176
22	0.00000	1.00000	0.0000	351.0000	1.544760	5.429517	3.043234
23	0.00000	1.00000	0.0000	351.0000	1.544761	5.429550	3.043263
24	0.00000	1.00000	0.0000	351.0000	1.544762	5.429567	3.043277
25	0.00000	1.00000	0.0000	351.0000	1.544762	5.429575	3.043284
26	0.00000	1.00000	0.0000	351.0000	1.544763	5.429579	3.043288
27	0.00000	1.00000	0.0000	351.0000	1.544763	5.429581	3.043289
28	0.00000	1.00000	0.0000	351.0000	1.544763	5.429582	3.043290
29	0.00000	1.00000	0.0000	351.0000	1.544763	5.429582	3.043291
30	0.00000	1.00000	0.0000	351.0000	1.544763	5.429583	3.043291***

In the appendix is a SAS program that generates the results in Table 2. At $Y = 8$,

- $\hat{\pi}_8 = 0.00466$ and $P(Y \leq 8) = 0.99540 < 1.0000$, hence, not yet a probability mass function
- $\hat{m}_8 = 1.8339$-expected value and thus the cumulative sum of expected values=$349.3841 < n = 351$
- The empirical mean and variance at $Y = 8$ are 1.498964 and 2.718813 respectively. These are far from the theoretical means computed from expressions in (4) which are respectively 1.54476 and 3.04329.
- These theoretical means are not achieved until $Y = 30$. At $Y = 30$, the computed mean and variance agree with the theoretical moments computed from expressions (4a) and (4b) respectively.
- In Table 1 are the values of $Y = y_a$ at which these values are obtained for each of the distributions.

5. Models Applied to other Data sets

We present in the following sections applications of the distributions to various data sets exhibiting various characteristics. The data sets employed here

include those with excess zeros, and over-dispersed data sets. The behaviors of the ATPPSD and the BT are compared with these other distributions.

5.1 *Example II: Insurance Claims Data*

This example is 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 3 are therefore the distribution of claims from 4000 vehicle polices.

Table 3: Distribution of Claims from an Insurance Company

Y	Count	P	NB	GP	GPL	ATPPSD	BT	PIG	DW
0	3719	3668.5422	3719.2220	3719.1171	3718.7800	3681.4477	3718.1079	3718.5830	3719.1842
1	232	317.3289	229.9009	231.1393	229.5930	293.1083	227.2675	234.5413	230.9350
2	38	13.7245	39.9106	38.4134	41.3966	23.4067	46.4030	34.8566	38.6770
3	7	0.3957	8.4156	8.4195	8.1604	1.8738	7.1203	8.3175	8.3249
4	3	0.0086	1.9313	2.1076	1.6484	0.1503	0.9608	2.4461	2.0678
5	1	0.0001	0.4648	0.5705	0.3361	0.0121	0.1231	0.8035	0.5663
Total	4000	4000.0000	3999.8453	3999.7680	3999.9135	3999.9989	3999.9826	3999.5479	3999.7882
		(0.4044)	(0.6195)	(0.8026)	(0.4226)	(2.0372)	()		
		$\hat{\mu}$=0.0865	$\hat{p}$=0.2854	$\hat{\mu}$=0.0865	$\hat{\alpha}$=0.1332	$\hat{\alpha}$=0.5213	$\hat{\alpha}$=0.3472	$\hat{\mu}$=0.0865	$\hat{q}$=0.0702
			$\hat{r}$=0.2166	$\hat{\tau}$=2.1741	$\hat{\theta}$=3.9018	$\hat{\theta}$=12.1794	$\hat{\theta}$=0.1760	$\hat{\beta}$=0.4404	$\hat{\beta}$= 0.7229
μ	0.0865								
σ^2	0.1225								
$\bar{y}$		0.0865	0.0865	0.0865	0.0864	0.0866	0.0865	0.0865	0.0865
s^2		0.0865	0.1210	0.1221	0.1192	0.0941	0.1165	0.1246	0.1221
X_g^2		344.1898	1.1738	0.6734	2.3660	61.6539	2.5594	0.6972	0.6918
d.f.		2	3	3	3	1	1	3	3
p-value		0.0000	0.7593	0.8794	0.5000	0.0000	0.1096	0.8739	0.8751
X_W^2		5665.5607	4048.6975	4013.9084	4109.9549	5207.9813	4205.3349	3933.2935	4014.0163
d.f.		3998	3997	3997	3997	3997	3997	3997	3997
-2LL		2492.2	2367.1	2366.8	2367.9	2414.5	2372.4	2367.0	2366.8
AIC		2494.2	2371.1	2370.8	2371.9	2418.5	2376.4	2371.0	2370.8
BIC		2500.4	2383.7	2383.4	2384.4	2431.1	2389.0	2383.0	2383.4

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, GPL GP, PIG and DW models fit this data set well, but the DW is the most parsimonious. The ATPPSD does poorly for this data set.

The percentage of observed zeros is not too far from that expected under the Poisson and the data is not zero-inflated. Here again, the X_g^2 computed have taken into consideration the problem of small expected frequencies and has accordingly applied the Lawal (1980) rule. We also observe that computed means and variances of the NB, GP and GPL are very close to those from the observed data, especially the computed variances. The BT does much better than the ATPPSD.

5.2 *Genetics: Chromatid Aberrations*

The data in Table 4 give the distribution of number of chromatid abberations (0.2 g. chinon I, 24 hours) on chemically induced chromosome aberrations in cultures of human leukocytes. The data has been analyzed by Loeschke and Kohler (1976) and Janadan and Schaeffer (1977).

The most parsimonious model here being the PIG both in terms of AIC, BIC and grouped Pearson's X_g^2, with a p-value of 0.8694 after applying the Lawal's (1980) rule. We observe here that both the ATPPSD and the BT models underes-

timate the variance of the data, hence, they fit poorly. As observed earlier, none of the models have estimated probabilities summing to 1.00 within the range [0,7] of the observed data.

Table 4: Observed and Expected values under the various Models

Y	count	P	NB	GP	GPL	ATPPSD	BT	PIG	DW
0	268	231.3576	270.1749	270.0660	269.2444	259.1486	267.2108	269.2058	270.5137
1	87	126.6683	78.5520	79.9695	78.7115	90.6991	76.7325	83.4962	78.5290
2	26	34.6754	29.8381	28.8742	30.8543	32.1471	35.7758	27.1236	29.6252
3	9	6.3283	12.2203	11.6553	12.5425	11.5044	13.4900	10.5483	12.0673
4	4	0.8662	5.1864	5.0432	5.1261	4.1454	4.6321	4.6944	5.1317
5	2	0.0948	2.2473	2.2876	2.0919	1.5004	1.5021	2.2822	2.2532
6	1	0.0087	0.9872	1.0736	0.8510	0.5445	0.4645	1.1774	1.0130
7+	3	0.0007	0.4375	0.5170	0.3450	0.1978	0.1378	0.6335	0.4642
Total	400.00	400.00	399.6441	399.4864	399.7667	399.8874	399.9456	399.1614	399.5873
$Y \leq y_a$		7	21	28	28	20	19	42	29
		$\hat{\mu}$=0.5475	$\hat{p}$=0.4690	$\hat{\mu}$= 0.5475	$\hat{\alpha}$=0.4740	$\hat{\alpha}$=0.0921	$\hat{\alpha}$= 0.6453	$\hat{\mu}$=0.5475	$\hat{q}$=0.3237
			$\hat{r}$=0.6200	$\hat{\tau}$=0.7194	$\hat{\theta}$=1.5776	$\hat{\theta}$=2.0474	$\hat{\theta}$=0.4450	$\hat{\beta}$=1.0580	$\hat{\beta}$= 0.8694
μ	0.5475								
σ^2	1.1256								
$\bar{y}$		0.5475	0.5475	0.5475	0.5464	0.5480	0.5475	0.5475	0.5469
s^2		0.5475	1.0310	1.0637	0.9881	0.8605	0.9008	1.1268	1.0430
X^2_g		39.1563	5.3321	3.4848	7.5178	13.9158	12.4628	1.2522	4.7900
d.f		2	4	4	4	4	3	4	4
p-value		0.0000	0.2549	0.4802	0.1109	0.0076	0.0060	0.8694	0.3094
X^2_W		820.2694	435.5968	422.2078	454.4883	521.9093	498.5472	398.5717	430.5776
d.f.		398	398	398	398	398	398	398	398
AIC		881.0	803.7	802.4	805.1	807.6	813.4	800.8	803.4
BIC		885.0	811.7	810.3	813.1	815.6	821.4	808.8	811.4

5.3 *Example III: Medical Vaccine Data*

The data in this example was analyzed in Phang and Loh (2014) and relate to vaccine adverse event count, where 4020 observed systemic adverse events for four injections administered to each of the 1005 study participants tabulated by the number of such adverse events occurrences. The data is presented in Table 5.

Table 5: Parameter Estimates for the Injection Study Data

		Eight Probability Models							
Y	count	P	NB	GP	GPL	ATPPSD	BT	PIG	DW
0	1437	890.757	1409.083	1389.968	1418.119	1427.031	1444.0324	1350.4344	1410.5098
1	1010	1342.340	1068.653	1098.740	1054.743	1032.794	948.7219	1154.3773	1065.5600
2	660	1011.430	670.653	675.746	668.470	669.739	705.4562	689.1260	667.8474
3	428	508.063	391.633	384.9‘9	394.624	402.444	435.9538	374.2808	393.1594
4	236	191.408	220.157	213.068	223.233	228.924	244.1122	200.6188	222.6488
5	122	57.689	120.881	116.677	122.648	124.989	127.2175	108.9210	122.5773
6	62	14.489	65.318	63.695	65.953	66.113	62.5125	60.2918	65.9953
7	34	3.119	34.887	34.786	34.883	34.102	29.2413	34.0243	34.8803
8	14	0.588	18.471	19.038	18.208	17.234	13.1161	19.5363	18.1452
9	8	0.098	9.712	10.449	9.402	8.564	5.6727	11.3871	9.3090
10	4	0.015	5.078	5.753	4.811	4.195	2.3758	6.7233	4.7169
11	4	0.002	2.643	3.178	2.443	2.030	0.9668	4.0138	2.3634
12	1	0.000	1.371	1.761	1.233	0.972	0.3834	2.4192	1.1721
$Y \leq y_a$		12	34	32	33	28	27	45	29
Parameter		$\hat{\mu}$=1.5070	$\hat{p}$=0.4967	$\hat{\mu}$=1.5070	$\hat{\alpha}$=1.2946	$\hat{\alpha}$=1.3501	$\hat{\alpha}$=0.8302	$\hat{\mu}$=1.5070	$\hat{q}$= 0.6491
Estimates			$\hat{r}$=1.5268	$\hat{\tau}$=0.2780	$\hat{\theta}$=1.1654	$\hat{\theta}$=1.4015	$\hat{\theta}$=0.7914	$\hat{\beta}$=1.0539	$\hat{\beta}$=1.1470
μ	1.5070								
σ^2	2.9034								
$\bar{y}$		1.5070	1.5070	1.5070	1.5072	1.5071	1.5070	1.5070	1.5065
s^2		1.5070	2.9944	3.0343	2.9835	2.9382	2.7580	3.0952	2.9661
X^2_g		> 1500.0	11.069	19.704	8.427	3.909	17.7870	48.5101	10.7392
d.f.		6	10	10	10	9	9	10	10
p-value		0.0000	0.3522	0.0322	0.5872	0.9173	0.0377	0.0000	0.3782
X^2_W		7743.22	3896.91	3845.62	5600.52	3971.39	4230.87	3769.99	3934.05
AIC		14464	13485	13496	13482	13478	13488	13,525	13,483
BIC		14471	13498	‘13508	13495	13491	13500	13,537	13,496

The observed mean and variance for this data set are 1.5070 and 2.9034 respec-

tively. In this example, we see that the ATPPSD model is the most parsimonious. The reason is that both its means and variances are very close to the observed values for the data. The ATPPSD is closely followed by the GPL model. Again (apart from the Poisson) none of the models have their cumulative probabilities and expected values summing to 1.00 and $n = 4020$ respectively within the data range $0 \leq Y \leq 12$. The values of Y, where these sums add to the appropriate values are presented as y_a, where for instance, for the ATPPSD, this would be $y_a = 28$, which is well outside the range of Y. Consequently again, the X_g^2 are computed with the last category being adjusted as is often the case. All the expected values here satisfy the Lawal (1980) rule for the application of the χ^2 distribution.
In terms of the Wald's test statistic, the most parsimonious model would be the PIG model. This is not unexpected as it has the largest estimated variance of 3.0952.

5.4 *Example IV: Accident Data*

This example is presented in Greenwood and Yule (1920). The data in Table 6 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 has excess zeros. In Table 6 are the results of applications of these distributions to this data set.

Table 6: Distribution of Number of accidents among machine operators

Y	Count	P	NB	GP	GPL	ATPP	BT	PIG	DW
0	447	406.3125	445.8864	445.1728	446.3985	442.1930	446.7131	443.3309	445.5437
1	132	189.0263	134.8957	136.7721	133.7239	139.3642	129.3145	141.1393	135.3634
2	42	43.9698	43.9920	43.0768	44.4548	44.4708	49.3925	41.2279	43.9955
3	21	6.8186	14.6924	14.2714	14.9354	14.2603	15.5372	13.3448	14.6194
4	3	0.7930	4.9647	4.9260	5.0017	4.5722	4.4489	4.7829	4.9247
5	2	0.0738	1.6893	1.7548	1.6652	1.4611	1.1940	1.8478	1.6753
Total	647	646.9939	646.1205	645.9739	646.1795	646.3216	646.6003	645.6737	646.1221
		$\hat{\mu}$=0.4652	$\hat{p}$=0.3497	$\hat{\mu}$=0.4652	$\hat{\alpha}$=0.7364	$\hat{\alpha}$=0.3004	$\hat{\alpha}$=0.4744	$\hat{\mu}$=0.4652	$\hat{q}$=0.3114
			$\hat{r}$=0.8651	$\hat{\tau}$=0.5251	$\hat{\theta}$=2.2446	$\hat{\theta}$=2.6571	$\hat{\theta}$=0.6102	$\hat{\beta}$=0.5677	$\hat{\beta}$=0.9673
μ	0.4652								
σ^2	0.6919								
$\bar{y}$		0.4652	0.4652	0.4652	0.4654	0.4653	0.4652	0.4652	0.4653
s^2		0.4652	0.7154	0.7203	0.7150	0.6861	0.6859	0.7293	0.7136
X_g^2		70.3711	3.9091	4.3456	3.5172	4.3136	3.6585	6.1269	3.8357
d.f.		3	3	3	3	3	3	3	3
p-value		0.0000	0.2714	0.2265	0.3185	0.2295	0.3008	0.1056	0.2798
X_W^2		960.8	624.8	620.52	625.11	651.43	651.61	612.84	626.34
d.f.		645	644	644	644	644	644	644	644
AIC		1236.4	1188.5	1189.2	1188.3	1188.8	1188.0	1191.1	1188.6
BIC		1240.8	1197.5	1198.1	1197.2	1197.7	1196.0	1200.0	1197.5

For this data set, the GPL model is the most parsimonious model based on the grouped Pearson's X_g^2 of 3.5172 on 3 d.f. (pvalue=0.3185). However, based on AIC and BIC selection criteria, the chosen model would be the Bell-Touchard distribution. On the other hand, is parsimony is based on Wald's Test statistics, the PIG model (with over estmated variance of the data) will be chosen. In such competing selection situations, selection will be based on the fitted values of the model which are reflected in the Pearson's X^2 goodness-of-fit statistic. 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:

$$f(y;p;\theta,\sigma) = c(y;p;\sigma)\exp\{\theta y - \sigma K_p(\theta)\}, y \in S_p \tag{25}$$

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. Results in Table 6 indicate that with the exception of the Poisson and the PIG, all the other models perform better than the Kokonendji and Malouche $HD_2(q,\theta)$ model.

5.5 *Example V: Death Notices*

This example data is presented in Table 7 and gives the number of death notices of women 80 years of age and older, as 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) and Lawal (2021). The data has an observed mean and variance $\mu = 2.1569$ and $\sigma^2 = 2.6073$ respectively with a dispersion index of 1.2088-indicating moderate over-dispersion in the data.

Table 7: Models for the frequency counts of Death Notices

Y	Count	P	NB	GP	GPL	ATPPSD	BT	PIG	DW
0	162	126.7844	155.6940	155.0261	155.7159	236.2784	158.1816	153.3423	156.5517
1	267	273.4657	275.7962	276.2127	275.7881	264.3793	273.6696	277.1488	278.4826
2	271	294.9237	268.9208	269.5270	268.9012	213.7483	266.9110	271.0949	263.1223
3	185	212.0437	190.8337	190.9293	190.8257	149.6387	190.9472	191.2643	189.3131
4	111	114.3411	110.0932	109.8851	110.0958	96.3236	111.0463	109.4029	112.2854
5	61	49.3252	54.7464	54.5505	54.7517	58.6418	55.4136	54.0498	56.8184
6	27	17.7319	24.3175	24.2301	24.3212	34.3005	24.5234	23.9868	25.0037
7	8	5.4638	9.8794	9.8671	9.8811	19.4650	9.8396	9.8174	9.6881
8	3	1.4731	3.7327	3.7477	3.7332	10.7870	3.6360	3.7751	3.3342
9	1	0.3530	1.3277	1.3445	1.3278	5.8645	1.2520	1.3824	1.0259
Total	1096	1095.9057	1095.3416	1095.3198	1095.3418	1089.4271	1095.4203	1095.2647	1095.6254
		(1096)	1096	(1096)	(1096)	(1096)	22	20	18
		$\hat{\mu}$=2.1569	$\hat{p}$=0.1787	$\hat{\mu}$=2.1569	$\hat{\alpha}$=9.8725	$\hat{\alpha}$=2000.00	$\hat{\alpha}$=0.2205	$\hat{\mu}$=2.1569	$\hat{q}$=0.8572
			$\hat{r}$=9.9104	$\hat{\tau}$=0.0477	$\hat{\theta}$=4.6590	$\hat{\theta}$=1.2111	$\hat{\theta}$=7.8459	$\hat{\beta}$=0.2121	$\hat{\beta}$=1.7142
μ	2.1569								
σ^2	2.6073								
$\bar{y}$		2.1569	2.1569	2.1569	2.1569	2.1653	2.1569	2.1569	2.1565
s^2		2.1569	2.6264	2.6233	2.6266	4.1138	2.6326	2.6144	2.6152
X^2_g		26.9746	2.7390	2.9247	2.7350	73.8447	2.1487	3.4467	1.9212
d.f.		8	7	7	7	7	7	7	7
p-value		0.0007	0.9081	0.8919	0.9084	0.0000	0.9512	0.8408	0.9641
X^2_W		1323.641	1087.051	1088.325	1086.959	694.031	1084.499	1092.023	1091.707
d.f.		1094	1093	1093	1093	1093	1093	1093	1093
AIC							3985.0	3986.4	3984.6
BIC							3995.0	3996.4	3994.6

The ATPPSD fits poorly here but the BT fits very well. However, the most parsimonious model here being the DW. The GPL,GPL. NB and GP are also suitable candidates. Notice the ratio of the parameters for the ATPPSD model

here: $2000/1.2111 = 1615.39$. The ATPPSD grossly over estimates the variance of the observed data.

6. Zero-Inflated ATPPSD, GPL and BT

The ATPPSD and BT tend to perform poorly for data sets having more than 80% of their observations being zeros. We therefore, consider zero-inflated GPL, ATPPSD and BT models in this section. Zero-inflated models for the NB, GP and P have been exhaustively considered in various literature. The 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 for a pmf $f(y), y = 0, 1, \ldots$ in the form:

$$f(y|\theta,\phi) = \begin{cases} \phi + (1-\phi)\, f(0), & \text{for } y = 0 \\ (1-\phi)\, f(y), & \text{for } y = 1, 2, \ldots \end{cases} \tag{26}$$

where ϕ is the extra proportion of zeros, such that $0 \leq \phi < 1$ and Y is the count random variable with specified parameters. ϕ is modeled here in the logit form. Specifically, the ZI-NGPSD probability mass function becomes,

$$f(y|\alpha,\theta,\phi) = \begin{cases} \phi + (1-\phi)\, f(0) & \text{if } y = 0 \\ (1-\phi)\; f(y) & \text{if } y > 0 \end{cases} \tag{27}$$

where $f(y)$ on the RHS of (27) is the probability mass functions in (3) and $f(0) = P(Y = 0)$ such that,

$$P(Y=0) = \left[\frac{\theta^3}{\theta^2 + \alpha\theta + 2\alpha}\right]\left[\frac{\alpha(\theta+3) + (\theta^2+2\theta+1)}{(\theta+1)^3}\right]$$

Similarly, for the GPL, $f(0)$ is given by: $f(0) = \dfrac{\alpha(2+\theta)\theta^{\alpha+1}}{\Gamma(1+\alpha)(1+\theta)^{\alpha+2}}$, while for the BT distribution, $f(0) = e^{\theta(1-e^{\alpha})}$. These are the expressions for the zero-part of the likelihood in (26).

Implementing the ZI-GPL, ZI-ATPPSD, ZI-BT and the ZIPIG models for the data set in Table 8 and the results therein. The data set in Table 8, is the distribution of one of the response variables HOSP=the number of days stayed in hospital from the NMES (The US National Medical Expenditure Survey 1987 and 1988). The data has previously been analyzed in Deb and Trivedi (200/) and recently by Lie *et al.* (2011) and Wogrin and Bodhisuwan (2017). Here, Y_i is the number of hospital stays and count is the frequency in each category. The sample size here is $n = 4406$. The data have a sample mean $\bar{y} = 0.2960$ and sample variance $s^2 = 0.5571$ and consequently a dispersion parameter of 1.88 which clearly indicates over-dispersion. Also the data has excess zeros with 80.4% of the data having zeros.

Table 8: Parameter Estimates and Expected values under the various Models

		Regular Models					Zero-Inflated Models			
Y	count	GPL	ATPPSD	BT	PIG	DW	ZIGPL	ZIATPPSD	ZIBT	ZIPIG
0	3541	3540.8016	3401.1094	3533.7148	3536.4538	3544.7244	3541.0087	3541.0000	3541.0000	3541.0000
1	599	579.8887	773.6037	562.9048	618.4163	587.1748	579.6051	574.6805	556.5355	601.4059
2	176	186.8588	177.6970	219.1836	154.7011	174.5332	186.8543	192.1207	213.9888	165.8911
3	48	64.1171	41.1165	66.1548	5.4985	60.4913	64.1502	64.7890	67.8342	56.6547
4	20	22.3272	9.5622	17.9181	22.1577	22.8960	22.3520	21.9958	19.5687	22.4181
5	12	7.8065	2.2309	4.6175	10.1774	9.2037	7.8195	7.5041	5.2879	9.7342
6	5	2.7316	0.5213	1.1478	4.8908	3.8699	2.7376	2.5686	1.3536	4.4938
7	1	0.9555	0.1219	0.2757	2.5598	1.6860	0.9580	0.8809	0.3309	2.1655
8	4	0.3340	0.0285	0.0642	1.3565	0.7562	0.3350	0.3024	0.0778	1.0769
ML		$\hat{\alpha}$=0.2210	$\hat{\alpha}$=0.1940	$\hat{\alpha}$=0.6195	$\hat{\mu}$=0.2960	$\hat{q}$=0.1955	$\hat{\alpha}$=0.2321	$\hat{\alpha}$=0.0784	$\hat{\alpha}$=0.4733	$\hat{\mu}$=0.3952
Estimates		$\hat{\theta}$=1.9109	$\hat{\theta}$=3.7126	$\hat{\theta}$=0.2572	$\hat{\beta}$=0.9322	$\hat{\beta}$=0.7667	$\hat{\theta}$=1.9124	$\hat{\theta}$=2.1851	$\hat{\theta}$=0.6247	$\hat{\beta}$=0.7797
							$\hat{\phi}$=0.0181	$\hat{\phi}$=0.4093	$\hat{\phi}$=0.3765	$\hat{\phi}$=0.2511
AIC		6030.2	6134.7	6066.6	6020.5	6020.9	6032.2	6035.0	6063.4	6020.5
BIC		6043.0	6147.5	6079.4	6033.3	6033.7	6051.4	6054.2	6082.6	6039.7
μ	0.2960									
σ^2	0.5571									
$\bar{y}$		0.2955	0.2964	0.2960	0.2960	0.2957	0.2955	0.2960	0.2960	0.2960
s^2		0.5118	0.3866	0.4793	0.5719	0.5397	0.5121	0.5105	0.4889	0.5561
X^2_w		4794.70	4612.12	5120.15	4291.41	4546.80	4792.23	4807.47	3509.15	4413.00
d.f.		4403	4403	4403	4403	4403	4402	4402	4402	4402
X^2_g		18.1839	>100.00	75.7546	5.8997	9.2790	18.1470	19.0920	62.1234	4.8101
d.f.		5	4	3	6	6	4	3	2	5
p-value		0.0027	0.0000	0.0000	0.4345	0.1585	0.0012	0.0003	0.0000	0.4395

Results from the left panel in Table 8 indicate that the PIG and DW models fit the data well, with the PIG being the most parsimonious model. The other regular models fit poorly, although the GPL performs slightly better than the BT and ATPPSD. The reason for this is that both ATPPSD and BT grossly underestimate the observed variance of the data. These estimates being 0.3866 and 0.4793 respectively. However, the GPL gives a slightly higher estimate of the variance - being 0.5118, which is closer to the true variance than the other two, while PIG and DW provide estimated variances that are even much closer to the observed variance in the data.

For the zero-inflated versions of the models on the right panel, results indicate that the estimated inflation parameter for the ZI-GPL is not significant and there is therefore no improvement on the ZI-GPL relative to its regular version. On the other hand, the ZI-ATPPSD and ZI-BT have significant estimated parameter ϕ and consequently gives improved performances over the regular models but these improvements are not significant enough for the models to provide adequate fits to the data. The variance estimates are higher but not sufficiently higher enough to provide parsimonious models. The most parsimonious model here is the ZIPIG. On both panels, all the models correctly estimated the observed mean of the data.

The estimated zero-inflated means and variances are computed from the expressions in the table below:

$\hat{\mu}_{ZI}$	$(1-\hat{\phi})\mu$
$\hat{\sigma}^2_{ZI}$	$(1-\hat{\phi})(\sigma^2+\hat{\phi}\mu^2)$

where the μ for ATPPSD, BT and GPL are given respectively in (4a), (9a) and (13a). Similarly, the σ^2 for the ATPPSD, BT and GPL are given in (5), (9b) and (14) respectively.

7. GLM Applications of the Models

In this section, we consider the applications of the above models to data having covariates or explanatory variables. Two data sets that have received considerable attention in the literature are considered: The doctor visits data in the United States and the doctor's visits from the German health Registry. The data and results are presented in the following sections.

7.1 *Doctor visits from United States*

These data consist of 485 observations with the response variable being the number of doctor visits and is from the United States in the year 1986. The explanatory variables are:the number of children in the household, a measure of access to healthcare and a measure of health status (larger positive numbers are associated with poorer health). The response variable, has approximately 50% of zeros, and thus it can be considered as zero excessive data. The response variable has a sample mean of 1.6103 and sample variance of 11.2011, giving us a dispersion index (DI) of $6.9559 > 1$. Thus the data is strongly over-dispersed. These data are available from the Ecdat R package, under the name Doctor. The results of implementing some of the models discussed in the previous sections and their zero-inflated versions are presented in Table 9.

Table 9: MLE and GOF Statistics for the Various Models

Parameters	Regular				Zero-Inflated		
	GPL	BT	PIG	DW	ZIGPL	ZIBT	ZIPIG
intercept	-1.9322**	0.4687**	0.4091*	0.1125	-1.3906*	0.4499**	0.4694*
Children	-0.3676	-0.0649**	-0.1019	-0.1385**	-0.2971	-0.0662**	-0.1062*
access	2.3888*	0.3463**	0.5230	0.4624	1.9364*	0.3514**	0.5499
health	0.5099**	0.0974**	0.2656**	0.2769**	0.4512**	0.0980**	0.2699**
$\hat{\theta}$	0.5705**	0.1808**			0.5831**	0.1963**	
$\hat{\beta}$			4.4284**	0.7824**			4.0689**
$\hat{\phi}$					0.1993*	0.6106**	0.6152**
AIC	1609.5	1647.8	1572.5	1576.7	1609.1	1649.6	1573.7
BIC	1630.4	1668.7	1593.4	1597.7	1634.2	1674.8	1598.8
X^2_W	716.7748	871.2928	438.0371	588.2812	686.4500	853.5290	456.0429
d.f.	480	480	480	480	479	479	479

* Significant at 0.05, ** Significant at 0.01.

Among the regular models the most parsimonious model is the PIG. It has a Wald's GOF of 438.0371 on 480 d.f. This is closely followed by the DW with a Wald's X^2 of 588.2812 on 480 d.f with an ID of 1.2256. This ID should not be confused with that presented in Kalktawi *et al.* (2018) which is 4.9397 (this is based on the ratio of estimated variance and mean for the response variable under the DW model, viz: $7.8735/1.5933$). Actually, the ID should be 1.4256: see - Table 11. The GPL, BT underestimate the variances of some of the 485 observations. Ditto for the DW but to a lesser extent. Also presented are the corresponding zero-inflated models for the GPL, BT and PIG. The means and variances of the zero inflated models are obtained as:

$$\mu_{zit} = (1-\hat{\phi})\mu, \qquad \text{and} \qquad \sigma^2_{zit} = (1-\hat{\phi})[\sigma^2 + \hat{\phi}\mu^2]$$

where for the BT, GPL and PIG, the σ^2 are given in (9b), (14) and (18) respectively. Results from the zero-inflated panel also indicate that while there are small improvements with the ZIGPL and ZIBT over their regular counterparts, the ZIPIG does not give any improvement over its regular counterpart. The little gains from the AIC values are negated with their corresponding BIC values as a result of the estimation of the additional parameter ϕ for the ZI models. The zero inflated ZIDW model always returns a $\hat{\phi} \approx 0.0000$, indicating that the ZI for the DW is not effective as its parameter q is already directly linked to the zeros in the data set.

7.2 *Doctor visits from Germany prior to health reform data*

Our second example of data having covariates is the data set from the German health registry(GHR) for the years 1984-1988. It provides information for the years prior to health reform. The data has 27,326 observations and the four variables: number of visits to doctor during a year (which ranges from 0 to 121), age (which ranges from 25 to 64), years of formal education (spanning from 7 to 18) and household yearly income (in DM/1000). The data has a sample mean of 31,835 and a corresponding variance o 32.3726, thus a DI of $10.1689 >> 1$. The data is therefore over-dispersed. The respond variable has 37.1% of its observation zeros. Under the Poisson, the expected number of zeros, would be $100\exp(-3.1835)$=4.14%. Thus this data display excess zeros and we would explore zero-inflated models for this data set. This dataset is available in the R package COUNT under the name *rwm*.

Table 10: MLE and GOF or various Model-German (GHR) Data

Parameters	Regular					Zero-Inflated		
	GPL	BT	PIG	DW	DW_{II}	ZIGPL	ZIBT	ZIPIG
Intercept	-8.2824**	0.644**	0.7199**	0.2973**	-0.5861**	-6.0479**	0.9019**	0.9972**
age	0.1400**	0.0064**	0.0189**	0.0178**	-0.0141**	0.1080**	0.0057**	0.0183**
educ	-0.054**	-0.017**	-0.0298**	-0.0363**	0.0288**	-0.0796**	-0.0119**	-0.0302**
income	-0.1536**	-0.0140**	-0.0217**	-0.0355**	0.0276**	-0.1134**	-0.0149**	-0.0262**
$\hat{\theta}$	0.2767**	0.0681**				0.2715**	0.0935**	
$\hat{\beta}$			9.7732**	0.7360**	0.7359**			6.6074**
$\hat{\phi}$						0.0026*	0.7113**	0.7305**
AIC	121,905	128,669	121,121	120,327	120,335	121,725	128,545	119,987
BIC	121,946	128,710	121,162	120,368	120,376	121,774	128,595	120,037
X^2_W	50,373	67,214	23,960	33,805	33,759	46,921	63,793	30,481
d.f.	27,321	27,321	27,321	27,321	27,321	27,320	27,320	27,320

* Significant at 0.05, ** Significant at 0.01.

The results of these implementations are presented in Table 10. Results here indicate that for 'regular' models, the PIG is the most parsimonious model. Both DW and DW_{II}, with parameter q formulated in the logit form in the former and in the log-log form in the latter. The parameter estimates under the latter agree with those presented in Kalktawi *et al.* (2018). Both give about the same AIC, BIC and Wald's GOF. The zero-truncated models indicate slight improvements over their 'regula' counterparts. The ZIPIG is the most parsimonious here and its variances and means are estimated lower (and closer) as illustrated in Table 11.

In Table 11 are estimated means, variances, Wald's GOF and the corresponding estimated dispersion indices for the models listed for the response variables (only) for the USA (doctors visits) and the GHR (docvis).

Table 11: Summary Statistics for all Models Applied to both Covariate Data sets

	USA Data				German Data			
Model	$\bar{y}$	s^2	X^2	DI	$\bar{y}$	s^2	X^2	DI
P	1.6103	1.6103	3366.660	6.9559	3.1835	3.1835	277,861.9936	10.1688
NB	1.6103	7.5048	722.3878	1.4925	3.1835	23.9805	36,887.5333	1.3500
GP	1.6103	8.6984	623.2592	1.2904	3.1835	29.7315	29,752.2959	1.0889
GPL	1.6087	6.1179	886.1516	1.8347	3.1783	18.3592	48,181.9055	1.7634
BT	1.6103	4.6015	1178.1794	2.4393	3.1835	12.3291	71,747.2896	2.6258
PIG	1.6103	10.4527	518.6565	1.0738	3.1835	36.5062	24,230.9389	0.8868
DW	1.5933	7.8735	688.5702	1.4256	3.1630	25.9332	34,110.4199	1.2484
ZIPIG	1.6103	9.4153	575.8034	1.1921	3.1835	27.6963	31,938.5401	1.1689
Data	1.6103	11.2011			3.1835	32.3726		

Clearly here:

(1) For the USA Data:
- Most of the models estimate the mean of the response variable about right (exceptions being the GPL and DW)
- The variance of the response variable is grossly underestimated, however,the PIG with estimated s^2=10.4527 being about the closest to the observed variance of 11.2011 in the data.
- The PIG is clearly the most parsimonious here with ID of 1.0738.
- The ID of 4.9377 reported in Kalktawi *et al.* (2018) is the ratio 7.8735/1.5933 under the DW model.

(2) For the German Health Data:
- Here too, the observed mean of 3.1835 is well estimated by all models except GPL and DW.
- The generalized Poisson model gives the most parsimonious model, although this model is not considered in our results in Tables **??** and **??**.
- The PIG gives a dispersion index of $0.8868 < 1$ for this data set. This is because the variance is overestimated and thus reduces the Wald's X^2. Consequently, because the data has excess zeros, its zero-inflated version-ZIPIG produces a slightly modified estimate of variance and a dispersion index that is much closer to 1.0. Thus the zero-inflated model here will be better than the regular model.
- The ID of 8.1987 in Kalktawi *et al.* (2018) is the ratio of $s^2/\bar{y}$ under the DW model and far from the true index of dispersion under the model of 1.2484, which is much comparable to GP, ZIPIG and NB models.

8. Conclusion

Based on our results in this paper, the PIG and DW distributions will be suitable models for modeling over-dispersed data. In some cases,these distributions would be much preferred than either the negative binomial (NB) or the generalized Poisson. Not considered here are other two-parameter distributions, such as the hyper-Poisson and the Com-Poisson which are also suitable for modeling under-dispersed count data.
The ATPPSD in particular does not do well in most cases. It performs poorly

if the ratio of its estimated parameter is $>= 10$. For example, in Shanker *et al.* (2020), five data sets were studied. The ratios of these estimated parameters for Tables 1 to 5 in their paper are 1439, 3.328, 8.845. 4.5021 and 3.059 respectively. The ATPPSD fits all except the data in Table 1 of their paper, which has a ratio of $1439 > 10$. It also does poorly for data exhibiting excess zeros.
The ATPPSD however, provides another alternative in the suite of models for fitting over and under-dispersed data sets. While it might not be suitable for all data sets, it might perform better than existing models in data sets such as the one in data sets in which the ratio of its estimated parameters is ≤ 10 such as the data in Table 1.

Acknowledgments

The author would like to express my deepest gratitude to the anonymous reviewer for the insightful comments and suggestions that have greatly improved this paper.

References

Aderoju, S. A. (2020). A New Generalized Poisson Mixed Distribution. Applied Mathematical Sciences, 14(5), 229-234.

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

Barbiero, A. (2015). Discrete Weibull Distributions (Type I and 3). R package version 1.0.1

Barbiero, A. (2017). Discrete Weibull regression for modeling football outcomes. Int. Jour. Business Intelligence & Data Mininig.

Castellares, F.,Ferrari, S.L.P. and Lemonte, A.J. (2018). On the bell distribution and its associated regression model for count data. Applied Mathematical Modelling, 6, 172-185. doi:10.1016/j.apm.2017.12.014.

Chakraborty S. (2010). On Some Distributional Properties of the Family of Weighted Generalized Poisson distribution. Communications in Statistics-Theory and Methods, 39(15), 2767-2788.

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

Comtet, I. (2010). Advanced combinatorics: the art of finite and infinite expansions, Holland:Reidel Publishing Company.

Consul, P.C and Famoye, F. (1992). Generalized Poisson Regression Model. Communication in Statistics: Theory Methods, 21(1), 89-109.

Deb, P. and Trivedi, P. K. (1997). Demand for Medical Care by the Elderly: A Finite Mixture Approach. Journal of Applied Econometrics, 12, 313-336.

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

Greenwood, M., and Yule, G.U. (1920). An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, Journal of Royal Statistics Society, 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.

Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. Journal of American Statistical Association, 64, 1459-1471.

Holla, M. S. (1966). On a Poisson-Inverse Gaussian Distribution, Metrika, 15, 377-384.

Imoto, T. (2014). A generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution. Appl. Math. Comput, 247, 824834.

Janardan, K.G. and Schaeffer, D.J. (1977). Models for the analysis of chromosomal aberrations in human leukocytes. Biometrical Journal, 8, 599612.

Kalktawi, H.S., Vinciotti, V. and Yu, K. (2015). A simple and Adaptive Dispersion regression Model for Count Data. arXiv:1511.00634.

Khan, M.S.A., Khalique, A., and Abouammoh, A. M. (1989) On estimating parameters in a discrete Weibull distribution, IEEE Transactions on Reliability, 38(3), 348-350.

Kokonendji,C.C. and Malouche,D. (2008). A property of Count Distributions in the Hinde-Demetrio Family. Communication in Statistics: 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. (2017a). On the negative binomial-generalized exponential distribution and its applications. Applied Mathematical Sciences, 11, 345-360. doi.org/10.12988/ams.2017.612288

Lawal, H.B. (2017b). On the Hyper-Poisson distribution and its Generalizations with Applications. British Journal of Mathematics and Computer Science, 21(3), 1-17. doi:10.9734/BJMCS/2017/32184.

Lawal, H.B. (2018). Correcting for Non-Sum to 1 Estimated Probabilities in Applications of Discrete Probability Models to Count Data. International Journal of Statistics and Probability, 6, 119-131.

Lawal, H.B. (2019). Quasi-Negative Binomial, Inverse Trinomial and Negative Binomial Generalized Exponential Distributions and their Applications. Journal of Probability and Statistical Science, 17(2), 205-221.

Lawal, H. B. (2021). On Com-Negative Binomial Distributions. Benin Journal of Statistics, Vol. 4, 90-113.

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.

Loeschke, V. and Kohler, W (1976). Deterministic and Stochastic models of the negative binomial distribution and the analysis of chromosomal aberrations in human leukocytes. Biometrische Zeitschrift 18, 427-51.

Mahmoudi, E. and Zakerzadeh,H. (2010). Generalized Poisson-Lindley Distribution. Communication Statistics: Theory and Methods, 39(10), 1785-1798.

Mussie, T. and Shanker R. (2019). Another Two-Parameter Sujatha Distribution with Properties and Applications. Journal of Mathematical Sciences and Modelling, 2(1), 1-13.

Nagakawa, T. and Osaki, S. (1975). The Discrete Weibull distribution. IEEE transactions on Reliability, 24(5), 300-301.

Padgett, W.J., and Spurrier, J.D. (1985). Discrete failure models. IEEE Transactions on Reliability, 34(3), 253-256.

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

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.

Sankaran, M. (1970). The discrete Poisson-Lindley distribution. Biometrics, 26, 145-149.

Shanker, R. (2016). Sujatha distribution and its applications. Statistics in Transition new Series, 17(3), 1-20.

Shanker, R. (2017). The Discrete Poisson-Akash Distribution. International Journal of Probability and Statistics, 6(1), 1-10.

Shanker, R. and Shukla, K.K. (2019). A generalization of Poisson-Sujatha Distribution and its Applications in Ecology. International Journal of Biomathematics, 12(2), 1-11.

Shanker, R., Shukla, K. K. and Leonida, T. A. (2020). Another Two-Parameter Poisson-Sujatha Distributions. International Journal of Statistics and Applications, 10(2), 43-53.

Sellers, K.F., Borle, S. and Shmueli, G. (2012). The COM-Poisson model for count data: a survey of methods and applications. Applied Stochastic Models in Business and Industry, 104-116. 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. Journal of Royal Statistics Society, Series C, 54, 127-142.

Stein, W.E., and Dattero, R. (1984) A new discrete Weibull distribution. IEEE Transactions on Reliability, 33, 196-197.

Willmot, G.E. 1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. Scan Actura J., (3-4), 113-127.

Wongrin, W. and Bodhisuwan, W. (2017). Generalized Poisson-Lindley linear model for count data. Journal of Applied Statistics, 44(15), 2659-2671.

Zakerzadeh, H. and Dolati, A. (2009). Generalized Lindley Distribution. Journal of Mathematical Extension, 3, 13-25.

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