On Some Two-Parameter Exponential Discrete Distributions

Abstract

In this paper we considered discrete two-parameter distribution models with variance functions in the ABM and LM classes. Four example frequency data sets that have received considerable attention in the literature are employed for comparative studies. We also extend our analyses to count data having co-variates (GLM cases) that exhibit over-dispersion, under-dispersion and excess zeros. Models considered here include the NB, the generalized Poisson (GP), the new logarithmic distribution (NLD), the new geometric discrete Pareto distribution (NGDP), the Poisson-Lindley-beta prime distribution (PLB) as well as the Bell-Tuchard (BTD), the discrete Weibull (DWD) and the Poisson-inverse Gaussian (PIG) distributions. All these models belong to the exponential dispersion models for count data. Several measures of comparison (-log-likelihood, Pearson's grouped X^2, Wald's goodness-of-fit statistic X_W^2 as well as the root mean square error-RMSE) were employed. Our findings indicate that the ABM and LM classes, when $r \geq 2$ do not perform any better than either NB or GP with $r = 1, 2$ respectively and these classes are not computational flexible.

For the data sets considered here, the choice of NB, GP, PIG, PCB, NLD, NGDP and DWD, would be more than sufficient in giving good fits to such data sets. The latter three are particularly suited to fitting under-dispersed count data. No single distribution stands out as overall compelling.

Keywords: Exponential, Overdispersion, Poisson-Inverse, discrete Weibull, Tuchard, empirical means and variances,SAS PROC NLMIXED.
MSC 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) (CP) , the Hyper-Poisson-HP (Bardwell & Crow, 1964; Lawal, 2017), the discrete Weibull-DW (Holla,1966), and the Poisson Inverse Gaussian-PIG (Wilmot,1987) 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. These and other models that will be described in this paper belong to the class of exponential dispersion distributions

1.1 Exponential dispersion Models

The mean parameterization of the class of distributions belonging to the exponential dispersion models has the probability mass function (Bar-Lev *et al.*, 2023),

$$f(m,p) \doteq \mu_p(y) e^{y\psi_p(m)-\phi_p(m)}, \quad y = 0,1,2,\ldots \tag{1}$$

where

$$\psi_p(\mu) = \int \frac{1}{V_p(\mu)}\, d\mu; \quad \text{and} \quad \phi_p(\mu) = \int \frac{1}{V_p(\mu)}\, d\mu$$

with $V_p(\mu)$ being the variance function and p the dispersion parameter and μ is the mean parameter of the model. Belonging to this class of discrete exponential dispersion (EDM) distributions are the Hinde-Demétrio distribution $\mathscr{HD}_2(q,\theta)$ defined as

$$f(y;p;\theta,\sigma) = c(y;p;\sigma)\exp\{\theta y - \sigma K_p(\theta)\}, y \in S_p \tag{2}$$

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 $p \geq 1$.

1.2 Variance Functions for EDM Class

The EDM class may have different variance functions but we specify only three types of variance functions:

(a) The EDM class that is characterized by 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 $p \geq 1$. The Hinde-Demétrio family of distributions belongs to this class

(b) The ABM class (Awad, Bar-Lev & Makov, 2016), named after the authors, with $\mu > 0$ and variance function:

$$V_p(\mu) = \mu\left(1 + \frac{\mu}{p}\right)^r, \quad p > 0,\ r = 1,2,\ldots \tag{3}$$

Belonging to this class is the Poisson ($r = 0$), the Negative binomial-NB ($r = 1$) and the generalized Poisson- GP ($r = 2$).

(c) The LM class (Letac & Mora, 1990), named after the authors with $\mu > 0$ and variance function:

$$V_p(\mu) = \frac{\mu}{\left(1 - \frac{\mu}{p}\right)^r}, \quad p > \mu, \ r = 1, 2, \ldots \tag{4}$$

The ABM and LM classes have variance functions satisfying the conditions that:

(i) they are concentrated on the nonnegative integers, leading to a polynomial variance functions (VFs), and

(ii) the variance functions are exponential functions.

Condition (i) is a consequence of the sufficiency and necessary conditions for VF established in Letac & Mora (1990). The Hinde-Demétrio class also satisfies these properties in (i) and (ii). Other VFs satisfying these conditions are the,

- The Poisson-Tweedie mixture (Jørgensen & Kokonendji, 2016; Kokonendji *et al.*, 2004) with variance function $V(\mu) = \mu + p^{1-\gamma}\mu^{\gamma}$
- The Poisson-exponential-Tweedie (Jørgensen & Kokonendji, 2016) with variance function $V(\mu) = \mu + \mu^2 + p^{1-\gamma}\mu^{\gamma}$

However, we will restrict our investigation in this paper to the class of two-parameter models having the ABM variance functions. This class include all the two-parameter distributions being presented in the following sections Our goal is to compare the ABM class of distributions in Bar-Lev *et al.* (2023) with performances with easily applicable flexible models like the NB, GP, DW and the PIG. The class of LM models employed in Bar-Lev *et al.* (2003), while not very flexible to implement does not in our opinion belong to the two-parameter class of exponential dispersion models, but rather to the three-parameter class-simply because r is estimated even though the authors claimed that r is fixed. How do we know the best value of r for some of the specific data. The authors stated

> *We have computed our fitting models for power parameter r in the range of* $r = 1, \ldots, 10$

Thus r is searched until a suitable value is obtained in addition to the parameters m and p for this class of models. Why is $r = 10$ for the LM model for data sets in examples III and IV ib Bar-Lev *et al.* for example? Why not 3 or 4? How did we have r to be 10?

2 Material & Methods

The various two-parameter models considered therefore in this paper are briefly described below:

2.1 The Negative Binomial-NB

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

$$f(y;r,p) = \frac{\Gamma(r+y)}{y!\Gamma(r)} p^y (1-p)^r, \quad y = 0, 1, \ldots \tag{5}$$

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 (5) are given respectively in (6a) and (6b) respectively.

Hence,

$$\mu = rp/(1-p) \quad \Longrightarrow p = \frac{\mu}{r+\mu} \tag{6a}$$

$$\sigma^2 = rp/(1-p)^2 \quad \Longrightarrow \sigma^2 = \mu + \frac{\mu^2}{r} \tag{6b}$$

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

2.2 The Generalized Poisson Distribution-GP:

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

$$\Pr(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left\{-\frac{\mu_i(1+\alpha y_i)}{(1+\alpha\mu_i)}\right\}, \quad y_i = 0, 1, \ldots \tag{7}$$

with mean

$$\mathrm{E}(Y_i) = \mu_i; \quad \text{and} \quad \mathrm{Var}(Y_i) = \mu_i(1+\alpha\mu_i)^2. \tag{8}$$

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 $\alpha = 0$.

2.3 The New Geometric Discrete Pareto Distribution-NGDP

The NGDP proposed in Bhati & Bakouch (2019) has the pmf:

$$f(y|q,\alpha) = \frac{q^y}{(y+1)^\alpha} - \frac{q^{(y+1)}}{(y+2)^\alpha}, \quad y = 0, 1, 2, \ldots, \quad 0 < q < 1, \ \alpha \geq 0. \tag{9}$$

Its mean and variance can be computed from expressions in (10a) and 10b) respectively,

$$\mu_y = q\Phi(q, \alpha, 2) \tag{10a}$$

$$\sigma_y^2 = 2q\Phi(q, \alpha-1, 2) - q\Phi(q, \alpha, 2)[3 + q\Phi(q, \alpha, 2)] \tag{10b}$$

where $\Phi(z, s, a) = \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s}$

2.4 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:

$$f(y|\alpha,\theta) = \frac{e^{\theta(1-e^{\alpha})}\alpha^{y}T_y(\theta)}{y!}, \quad ,y = 0,1,2,\dots \tag{11}$$

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

$$T_n(\theta) = e^{-\theta}\sum_{k=0}^{\infty}\frac{k^n\theta^k}{k!} \tag{12}$$

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 B_n are the Bell numbers (Comtet, 2010) with for instance, $B_0 = B_1 = 1; B_2 = 2, \dots, B_{10} = 115975$ and so on. Its mean and variance (Castellares *et al.*, 2019) are given respectively in (13)

$$\mu = \theta\alpha e^{\alpha} \tag{13a}$$

$$\sigma^2 = \theta(1+\alpha)\alpha e^{\alpha} \tag{13b}$$

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.

2.5 The New logarithmic Distribution-NLD

Gomez-Deniz, Sarabia and Calderin-Ojeda (2011) proposed the new logarithmic distribution (NLD) whose pmf has the form:

$$f(y|\alpha,\theta) = \frac{\log(1-\alpha\theta^{y}) - \log(1-\alpha\theta^{y+1})}{\log(1-\alpha)}; \quad y = 0,1,\dots, \; 0 < \theta < 1; \; \alpha < 1(\alpha \neq 0) \tag{14}$$

Its mean and variance can be computed from expressions in (15a) and 15b) respectively,

$$\mu_Y = \frac{1}{\log(1-\alpha)}\sum_{y=1}^{\infty}\log(1-\alpha\theta^{y}) \tag{15a}$$

$$\sigma_Y^2 = \frac{1}{\log(1-\alpha)}\sum_{y=1}^{\infty}(2y-1)\log(1-\alpha\theta^{y}) - \mu_Y^2 \tag{15b}$$

2.6 The Hyper-Poisson Distribution

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

$$P(Y=y|\theta,\lambda)=\frac{\Gamma(\lambda)}{\Gamma(\lambda+y)}.\frac{\theta^y}{\phi(1,\lambda;\theta)};\ y=0,1,\ldots,;\quad \lambda>0,\theta>0 \tag{16}$$

where, $\phi(1,\lambda;u)=\sum_{k=0}^{\infty}\frac{(1)_k}{(\lambda)_k}.\frac{u^k}{k!}=\sum_{k=0}^{\infty}\frac{\Gamma(\lambda)}{\Gamma(\lambda+k)}u^k$, and, $(\lambda)_k=\frac{\Gamma(\lambda+k)}{\Gamma(\lambda)}$; $k=1,2,\ldots,$ is the confluent hyper-geometric series in which $(\lambda)_0=1$.

Its mean and variance (Kumar and Nair, 2014) are:

$$\mu=\frac{\phi(2,\lambda+1,\theta)}{\phi(1,\lambda,\theta)}.\frac{\theta}{\lambda},\quad \text{and}\quad \sigma^2=\frac{1}{\lambda}\left[\frac{2}{\lambda+1}\frac{\phi(3,\lambda+2,\theta)}{\phi(1,\lambda,\theta)}-\frac{1}{\lambda}\frac{[\phi(2,\lambda+1,\theta)]^2}{[\phi(1,\lambda,\theta)]^2}\right]\theta^2+\mu$$

where:

$$\phi(1,\lambda;\theta)=\sum_{k=0}^{\infty}\frac{\Gamma(\lambda)}{\Gamma(\lambda+k)}\theta^k,\quad \phi(2,\lambda+1;\theta)=\sum_{k=0}^{\infty}\frac{(k+1)\,\Gamma(\lambda+1)}{\Gamma(\lambda+k+1)}\theta^k \text{ and } \phi(3,\lambda+2;\theta)=\sum_{k=0}^{\infty}\frac{(k+2)(k+1)}{2}.\frac{\Gamma(\lambda+2)}{\Gamma(\lambda+k+2)}\theta^k$$

2.7 The Poisson-inverse Gaussian (PIG) Distribution

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

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

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.

2.8 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 has a probability mass function (pmf),

$$f(y;q,\beta)=q^{y^\beta}-q^{(y+1)^\beta}\quad y=0,1,\ldots \tag{19}$$

where $q=e^{-\lambda}$ and $0<q<1$. The model in (19) is the type I Discrete Weibull distribution, proposed in Nakagawa and Osaki (1975). 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:

$$\mathbb{E}(Y) = \sum_{y=1}^{\infty} q^{y^{\beta}} \tag{20a}$$

$$\mathbb{E}(Y^2) = 2\sum_{y=1}^{\infty} y q^{y^{\beta}} - \mathbb{E}(Y) \tag{20b}$$

The expression in (20a) for instance leads to a closed expression if and only if $\beta = 1$, in which case $\mathbb{E}(Y) = \dfrac{q}{1-q}$. From (20), we observe that $\mathbb{E}$(Y), for a fixed q is a decreasing function of β. Khan Khalique and Abouammoh (1989) have shown that

$$\mathbb{E}(Y) < \mathbb{E}(T) = \left(-\frac{1}{\log q}\right)^{\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right) < E(Y) + 1 \tag{21}$$

2.9 Poisson-Lindley Beta prime Distribution-PLB

Gomez, Hernandez & Fernandez (2016) proposed the two-parameter PLB prime distribution such that,

$$Y|\beta \sim PL(\theta)$$
$$\theta \sim BP(\alpha, \beta)$$

where $PL(\theta)$ is the Poisson-Lindley presented in [36] and $\theta \sim BP(\alpha, \beta) = \dfrac{1}{B(\alpha,\beta)}\dfrac{\theta^{\alpha-1}}{(1+\theta)^{\alpha+\beta}}$ is the beta-prime distribution. Consequently, the pmf of the PLB distribution therefore i

$$f(y|\alpha, \beta) = \frac{\alpha(1+\alpha)\Gamma(\alpha+\beta)\Gamma(\beta+y)}{\Gamma(\beta)\Gamma(\alpha+\beta+y+3)}\left[(\beta+y)(2+y)+\alpha+2\right],\ y = 0, 1, \ldots,\ \alpha, \beta > 0. \tag{22}$$

Its mean and variance are given respectively in (23a) and (23b).

$$\mu = \frac{\beta(2\beta+\alpha+1)}{(\alpha+\beta)(\alpha-1)}, \qquad \alpha > 1 \tag{23a}$$

$$\sigma^2 = \frac{\beta[6+2\beta(5+3\beta)+\alpha(\alpha+4\beta+1)]}{(\alpha+\beta)(\alpha-1)(\alpha-2)} - \mu^2, \quad \alpha > 2 \tag{23b}$$

and because this is a mixture of Poisson distribution, its dispersion index, following [23] is such that DI > 1. Thus the distribution is mainly suitable for fitting over-dispersed count data.

3 Estimation

For a single observation i, the log-likelihood for the NB, GP, NGBD, BTD, NLD, DW, PIG, PLB and HPP models are presented respectively in LL1 to LL9 in (24).

$$\text{LL1} = \log[\Gamma(r+y)] + y\log(p) + r\log(1-p) - \log y! - \log[\Gamma(r)] \tag{24a}$$

$$\text{LL2} = y_i \log\left(\frac{\mu_i}{1+\alpha\mu_i}\right) + (y_i - 1)\log(1+\alpha y_i) - \frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i} - \log(y_i!) \tag{24b}$$

$$\text{LL3} = \log\left[\frac{q^y}{(y+1)^\alpha} - \frac{q^{(y+1)}}{(y+2)^\alpha}\right] \tag{24c}$$

$$\text{LL4} = \theta + [1-\exp(\alpha)] + y\log(\alpha) + \log[T_y(\theta)] - \log y! \tag{24d}$$

$$\text{LL5} = \log\left[\frac{\log(1-\alpha\theta^y) - \log(1-\alpha\theta^{y+1})}{\log(1-\alpha)}\right] \tag{24e}$$

$$\text{LL6} = \log\left[q^{y^\beta} - q^{(y+1)^\beta}\right] \tag{24f}$$

$$\text{LL7} = \begin{cases} \log(p0) & \text{if } y = 0 \\ \log(p0) + y\log(\mu) - \log(y!) + \log(Q) & \text{if } y > 0 \end{cases} \tag{24g}$$

$$\text{LL8} = \log\left[\frac{\alpha(1+\alpha)\Gamma(\alpha+\beta)\Gamma(\beta+y)}{\Gamma(\beta)\Gamma(\alpha+\beta+y+3)}\right] + \log\left[(\beta+y)(2+y)+\alpha+2\right] \tag{24h}$$

$$\text{LL9} = y_i\log(\theta) + \log\Gamma(\lambda) - \log\Gamma(y_i+\lambda) - \log\left[\sum_{k=0}^{\infty}\frac{\Gamma(\lambda)}{\Gamma(\lambda+k)}\theta^k\right] \tag{24i}$$

where for (24g),

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

and p0 is as defined earlier. Maximum-likelihood estimation of all log-likelihoods in (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 [33] and our choice optimization algorithm here is the Newton-Raphson techniques.

4 Applications

The above models are applied to the four example frequency data sets presented in Bar-Lev *et al.* (2023). We also apply the models to the Congo data that is based on 314 observations, and with several covariates. The data displayed high over-dispersion. The following fit criteria are employed to access the performances of the various models

- Pearson's goodness-of-fit test,

$$X^2 = \sum_{i=0}^{K}\frac{(y_i - \hat{m}_i)^2}{\hat{m}_i}$$

- -2LL: Twice the log-likelihood
- AIC: Akaike Information criterion
- BIC: Bayesian Information Criterion

- Root mean squared error (RMSE) defined as:

$$\sqrt{\frac{1}{K+1}\sum_{i=0}^{K}(y_i-\hat{m}_i)^2}, \quad \text{such that}, \quad \sum_{i=0}^{K}\hat{m}_i = n.$$

- Wald's test Statistic,

$$X^2 = \sum_{j=1}^{n}\frac{(y_j-\hat{m}_j)^2}{\hat{\sigma}_j^2}$$

where n is the number of observations in the data.

4.1 Example I: 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 1 are therefore the distribution of claims from 4000 vehicle polices. The data first appeared in Gossiaux and Lemaire (1981), re-analyzed in Kokonendji *et al.* (2004) and recently re-analyzed in Bar-Lev *et al.* (2023).

Y	Count	NB	GP	PLB	NGDP	BTD	PIG	DWD	NLD
0	3719	3719.2220	3719.1171	3719.5369	3718.3785	3718.1079	3718.5830	3719.1842	3719.0596
1	232	229.9009	231.1393	234.2635	234.0088	227.2675	234.5413	230.9350	228.6547
2	38	39.9106	38.4134	35.4804	36.0903	46.4030	34.8566	38.6770	41.8537
3	7	8.4156	8.4195	8.0567	8.1284	7.1203	8.3175	8.3249	8.3239
4	3	1.9313	2.1076	2.3281	2.2620	0.9608	2.4461	2.0678	1.6810
5	1	0.4648	0.5705	1.3344	1.1321	0.1231	0.8035	0.5663	0.4271
Total	4000	3999.8453	3999.7680	3999.4566	3999.5886	3999.9826	3999.5479	3999.7882	3999.9134
		$\hat{p}$=0.2854	$\hat{\mu}$=0.0865	$\hat{\alpha}$=10.1031	$\hat{\alpha}$=-3.0451	$\hat{\alpha}$=0.3472	$\hat{\mu}$=0.0865	$\hat{q}$=0.0702	$\hat{\alpha}$=0.9529
		$\hat{r}$=0.2166	$\hat{\tau}$=2.1741	$\hat{\beta}$=0.6820	$\hat{q}$=0.5811	$\hat{\theta}$=0.1760	$\hat{\beta}$=0.4404	$\hat{\beta}$= 0.7229	$\hat{\theta}$=0.2027
μ	0.0865								
σ^2	0.1225								
$\bar{y}$		0.0865	0.0865	0.0866	0.0865	0.0865	0.0865	0.0865	0.0866
s^2		0.1210	0.1221	0.1257	0.1236	0.1165	0.1246	0.1221	0.1200
X_g^2		1.1738	0.6734	0.6172	0.5313	2.5594	0.6972	0.6918	2.4178
d.f.		3	3	3	3	1	3	3	3
p-value		0.7593	0.8794	0.8925	**0.9120**	0.1096	0.8739	0.8751	0.4903
X_W^2		4048.6975	4013.9084	**3,899.6001**	3,963.5921	4205.3349	3933.2935	4014.0163	4085.2871
d.f.		3997	3997	3997	3997	3997	3997	3997	3997
-LL		1183.55	1183.40	1183.55	1183.45	1186.20	1183.50	**1183.05**	1183.95
RMSE		1.3782	**0.7933**	1.4926	1.2848	4.0562	1.7617	0.8452	2.2311

Table 1: Distribution of claims from an Insurance Company

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 the data can be considered to having slightly excess zeros. We observe the following from our analyses and results in Table 1.

1. As observed in Lawal (2017) all the models (as typical of discrete distributions) produce cumulative sum of expected values not summing to the sample size n which is 4000 here. For example for the PLB and NGDP, $\sum_{i=0}^{5}\hat{m}_i = \{3999.4566, 3999.5886\}$ respectively, with actual estimated expected value $\hat{m}_5 = \{0.7910, 0.7207\}$ respectively (not reported). Thus, for the PLB, the reported $\hat{m}_5 = 1.3344 = 4000 - 3999.4566 + 0.7910$. This is similarly applied to all the models.

2. In computing Pearson's group X^2, we have employed the Lawal's (1980) rule which allows expected values can be as small as $rd^{-3/2}$, where r is the number of expected values less than 3, and d is the corresponding d.f. For the GP for example, $r = 2, d = 3$, hence the minimum expected value can be as small as $2 \times 3^{-3/2} = 0.3849$. Hence, we do not need to collapse categories at

all as 0.5705 satisfy this requirement. The resulting Parson's X_g^2 would therefore be based on $K + 1 - 2 - 1 = 6 - 3 = 3$ d.f. since two parameters are estimated from the model. This is why we have 3 d.f. rather than the 2 d.f. presented in Bar-Lev *et al.* and consequent different X^2.

3. Based on the grouped Pearson's X^2, the most parsimonious model is the NGDP with a p-value of 0.9120 which is better than the LM(r=4) in Bar-Lev *et al.*

4. The Discrete Weibull (DWD) produces the lowest -log-likelihood (-LL) of 1183.05, while the generalized Poisson produces the lowest root-mean square error (rmse) of 0.7933 with the former being performing better than the ABM(r=10) in Bar-Lev *et al.* (2023).

5. All the model produce estimated means 0f 0.0865 which are consistent with the observed sample mean. However, the PLB, NGDP and the PIG over-estimated the observed variance and hence they produce lower Wald's GOF.

4.2 Example II: Frequency of accidents by Machinists

The data set in Table 2 presents the frequency of accidents experienced by machinists (Bliss & Fisher, 1953). It was employed in Bar-Lev *et al.* (2023). The data has 414 observations with mean $\bar{y} = 0.4831$ and $s^2 = 1.0106$ giving a dispersion index $DI = 2.0919 > 1$. Thus the data is over-dispersed.

Y	Count	NB	GP	PLB	NGDP	BTD	PIG	DWD	NLD
0	296	296.7039	296.3418	295.6200	296.6526	294.7762	295.1454	296.6880	296.5044
1	74	71.0079	72.8253	74.8132	72.3213	67.1716	76.8568	71.4953	70.6928
2	26	26.4123	25.4286	24.8824	25.4691	32.8903	23.6382	26.1946	26.6759
3	8	10.9922	10.3772	9.7925	10.4653	12.6533	9.2536	10.7449	11.2256
4	4	4.8177	4.6263	4.3224	4.6795	4.3965	4.2366	4.7038	4.9186
5	4	2.1754	2.1834	2.0762	2.2066	1.4518	2.1353	2.1531	2.1917
6	1	1.0015	1.0724	1.0650	1.0788	0.4608	1.1457	1.0191	0.9838
7	0	0.4674	0.5424	0.5762	0.5415	0.1410	0.6421	0.4953	0.4431
8	1	0.2204	0.2806	0.3259	0.2773	0.0418	0.3715	0.2460	0.3642
Total	414.0000	413.7986	413.6779	413.4737	413.6920	413.9833	413.4252	413.7400	413.8356
		$\hat{p}$=0.5046	$\hat{\mu}$=0.4831	$\hat{\alpha}$=7.6464	$\hat{\alpha}$=1.0527	$\hat{\alpha}$=0.7514	$\hat{\mu}$=0.4831	$\hat{q}$=0.2834	$\hat{\alpha}$=0.7873
		$\hat{r}$=0.4743	$\hat{r}$=0.9209	$\hat{\beta}$=2.4039	$\hat{q}$= 0.5880	$\hat{\theta}$=0.3033	$\hat{\beta}$=1.2208	$\hat{\beta}$= 0.8037	$\hat{\theta}$=0.4516
μ	0.4831								
σ^2	1.0106								
$\bar{y}$		0.4831	0.4831	0.4842	0.4831	0.4831	0.4831	0.4828	0.4845
s^2		0.9752	1.0085	1.0680	1.0084	0.8461	1.0729	0.9892	0.9681
X_g^2		2.6316	2.1962	2.3268	2.2088	11.0806	2.3925	2.4819	2.8093
d.f.		5	5	5	5	4	5	5	5
p-value		0.7566	**0.8214**	0.8023	0.8196	0.0257	0.7926	0.7792	0.7294
X_W^2		428.0135	413.8603	390.8044	413.9061	493.3032	**389.0353**	421.9360	431.1572
d.f.		411	411	411	411	411	411	411	411
-2LL		764.1	**763.6**	763.8	**763.6**	771.3	763.8	763.9	764.3
RSME		1.6028	1.1364	**1.0237**	1.2363	3.7283	1.4911	1.4537	1.7246

Table 2: Distribution of accidents experienced by machinists

Based on results in Table 2, we observe the following:

- In terms of Pearson's group X^2, the generalized Poisson (GP) is the most parsimonious.
- Based on -2LL, AIC and BIC, both the generalized Poisson (GP) and NGDP provide the best fits
- Based on the RMSE, the most parsimonious model is the PLB with rmse=1.0237 and this is slightly followed by the GP. We note here that PLB provides a smaller rmse than the ABM(r=10) proposed in Shaul *et al.* (2023).
- Both the GP and NGDP provide estimated variances that are very close to the observed variance of 1.0106 in the data. In addition, like all the other models, they both provided estimated means of 0.4831 that are close to the observed mean in the data.

- The most parsimonious model in terms of the Wald's test statistic is the PIG with $X_W^2 = 389.0353$ on 411 d.f. (p-value=0.7754). This is closely followed by the PLB with $X_W^2 = 390.8044$ on 411 d.f. with (p-value=0.7558). This is not surprising as both PLB and PIG models produce the largest variances (these in fact over-estimated the observed variance in the data) and Wald's $X^2$the is a function of the reciprocal of the estimated variance. Hence both producing lower GOF values.

- The PLB produces the smallest rmse of 1.0237 and this is followed by the GP.

4.3 Example III: Frequency counts of kidney cysts using steroid

The data in Table 3 gives the counts of cysts of kidneys using steroids analyzed in Bar-Lev *et al.* (2023). It has 110 observations with observed mean $\mu = 1.3909$ and observed variance of 6.1118, giving a dispersion index (DI) of 4.3941. Again, this data set is highly over-dispersed.

Y	0	1	2	3	4	5	6	7	8	9	10	11	Total
Count	65	14	10	6	4	2	2	2	1	1	1	2	110

Table 3: Frequency of cysts of kidneys on those using steroids

$$\mu = 1.3901, \quad \sigma^2 = 6.1118, \quad DP = 4.3941$$

For this data set and that in Table 6, we are presenting only, the maximum-likelihood parameter estimates, and other measures of fit. The expected values are not presented. The results of applying all the models discussed in the previous sections to this data set are presented in Table 4.

	NB	GP	PLB	NGDP	BTD	PIG	DWD	NLD
MLE	$\hat{p}$=0.8122	$\hat{\mu}$=1.3901	$\hat{\alpha}$=2.4080	$\hat{\alpha}$=0.9480	$\hat{\alpha}$=2.1988	$\hat{\mu}$=1.3909	$\hat{q}$=0.4215	$\hat{\alpha}$=0.9645
	$\hat{r}$=0.3216	$\hat{\tau}$=1.0855	$\hat{\beta}$=1.3502	$\hat{q}$= 0.8604	$\hat{\theta}$=0.0702	$\hat{\beta}$=6.7323	$\hat{\beta}$= 0.6291	$\hat{\theta}$=0.7752
$\bar{y}$	1.3901	1.3909	1.5580	1.3821	1.3909	1.3909	1.4158	1.3709
s^2	7.4064	8.7615	25.3507	7.5088	4.4493	10.7548	8.4618	7.2019
X_g^2	1.1268	2.3809	3.1016	3.0852	9.6233	5.4869	1.3452	1.8529
d.f.	7	6	6	6	6	6	6	7
p-value	**0.9925**	0.8816	0.7960	0.7981	0.1414	0.4830	0.9691	0.9675
X_W^2	89.9482	76.0366	26.4001	88.7222	149.7306	**61.9437**	78.7377	92.5087
d.f.	107	107	107	107	107	107	107	107
-2LL	335.1	337.5	339.5	337.1	341.8	341.7	**336.0**	335.0
RSME	**0.9336**	1.6537	1.9854	2.2614	2.1163	2.7955	1.3140	0.9384

Table 4: Distribution of accidents experienced by machinists

Results in Table 4 indicate the followings:

- The NB gives the most parsimonious model when grouped Pearson's X^2 GOF is employed witha p-value of 0.9925 and is the only one based on 7 degrees of freedom. The others are based on 6 d.f. Again these contrast with the 4 d.f. reported for the results for this data set reported in Bar-Lev *et al.* (2023). The NB is closely followed by the DWD and GP. The latter two also fit well the data.

- In terms of Wald's GOF, the PLB gives a Wald's X_W^2 of 26.4001 on 107 d.f. (p-value=1.0000). This very unusual. The PLB grossly over-estimates the variance of the data-being 25.3507, hence the very low value of Wald's X^2. This is attributable to the non-convergence of the distribution to its true mean and variance at $k = 11$. At $k = 11$, the empirical mean and variance of the distribution are respectively 1.0711 and 3.6317 and did not converge to its theoretical mean and variance until $k = 2000$-an unusual case. We would therefore go with the PIG for this criterion.

- In terms of -log-likelihood, the best fitting model is the DWD with -LL=168.0

- The model with the smallest rmse is the NB with **rmse=0.9336**.

4.4 Example IV: Number of claims of automobile liabilities

The data in Table 5 gives the distribution of the number of claims of automobile liability policies (Klugman, Panjer, and Willmot, 1998) and presented as data set 4 in Bar-Lev *et al.* (2023). The data has a dispersion index of 2.1478, thus it is moderately over-dispersed.

Y	0	1	2	3	4	5	6	7	8	9	10	11	Total
Count	99	65	57	35	20	10	4	0	3	4	0	1	298

Table 5: Frequency of claims of automobile liabilities

$\mu = 1.7801, \quad \sigma^2 = 3.6687, \quad DP = 2.1478$

	NB	GP	PLB	NGDP	BTD	PIG	DWD	NLD
MLE	$\hat{p}$=0.5368	$\hat{\mu}$=1.7081	$\hat{\alpha}$=1100.90	$\hat{\alpha}$=0.0000	$\hat{\alpha}$=0.9795	$\hat{\mu}$=1.7081	$\hat{q}$=0.6765	$\hat{\alpha}$=-2.2035
	$\hat{r}$=1.4736	$\hat{\tau}$=0.2821	$\hat{\beta}$=1231.42	$\hat{q}$= 0.6307	$\hat{\theta}$=0.6548	$\hat{\beta}$=1.2472	$\hat{\beta}$= 1.1368	$\hat{\theta}$=0.5431
$\bar{y}$	1.7081	1.7081	1.7112	1.7081	1.7081	1.7081	1.7076	1.7058
s^2	3.6878	3.7505	3.9503	4.6255	3.3810	3.8383	3.6746	3.5744
X_g^2	10.1262	10.0792	**9.9117**	13.4778	14.9950	11.2334	10.3291	10.2108
d.f.	7	7	7	7	7	7	7	7
p-value	0.1815	0.1841	0.1936	0.0613	0.0372	0.1288	0.1707	0.1767
X_W^2	295.4603	290.5213	275.8267	**235.5638**	322.2679	283.8774	296.5260	304.8356
d.f.	295	295	295	295	295	295	295	295
-2LL	1057.5	1058.2	1058.0	1062.9	1057.6	1060.2	1057.7	**1056.8**
AIC	1156.0	1061.5	1062.2	1062.0	1066.9	1061.6		
BIC	1159.7	1068.9	1069.6	1069.4	1074.3	1069.0		
RSME	4.2441	4.8775	4.0491	5.8721	**2.3792**	6.2153	4.2001	3.5230

Table 6: Summary Results of Implementation of Models on Data set 4

$\mu = 1.7081, \quad \sigma^2 = 3.6687, \quad DP = 2.1478$

Table 6 gives the results of applying all the models discussed in the previous sections to the automobile liability claims data in Table 5. As before, we observe the followings from Table 6

- The PLB is the most parsimonious model based on Pearson's grouped X^2 wit a p-value of 0.1936 on 7 d.f. We note that the results in Bar-Lev *et al.* is based on 4 d.f. and the most parsimonious model being the BTD. Our results certainly provide more information as is based on a higher d.f.. The PLB here is slightly followed by the GP with a corresponding p-value of 0.1841.
- Although the NGPD will be considered the most parsimonious model in terms of Wald's GOF, however, it produces and estimate of α that is approximately zero, and thus a higher estimated variance.
- Both the NB and DWD produce estimated variances that are the most close to the observed data variance of 3.6687 and the two models produce Wald's GOF statistics are very similar.
- The NLD model produces the lowest -2 log-likelihood.
- The BTD has the lowest root mean square error (rmse) among all the models. This result is consistent with that reported for the same data in Bar-Lev *et al.* (2023).

5 GLM Applications- The Congo Data Example

In this section, we apply these models to the class attendance dataset which examines the relationship between the number of days absent from high-school and the covariates gender, mathematics score (on a standardized score out of 100) and academic program ('General=1', Academic=2' and 'Vocational=3') of 314 students samples from two urban High schools (see reference in Table 7). The first and last five

id	sex	math	prog	Y
1001	male	63	2	4
1002	male	27	2	4
1003	female	20	2	2
1004	female	16	2	3
1005	female	2	2	3
⋮	⋮	⋮	⋮	⋮
2148	female	8	2	5
2150	male	63	3	3
2151	female	59	3	7
2152	female	46	3	1
2153	male	26	2	1

Table 7: Congo: High school Absent Behaviour Data

Source:*(http://www.ats.ucla.edu/stat/stata/dae/nb-data.dta)*

observations of this dataset are displayed below in Table 7. We shall discuss this dataset more fully in a later section.

To fit our models, first we create the following indicator variables from the above data:

$$\text{gender} = \begin{cases} 1 & \text{if male} \\ 0 & \text{otherwise} \end{cases}; \quad z_1 = \begin{cases} 1 & \text{if prog=2} \\ 0 & \text{otherwise} \end{cases}; \quad \text{and} \quad z_2 = \begin{cases} 1 & \text{if prog=3} \\ 0 & \text{otherwise} \end{cases}.$$

For the linear predictor,

$$\mathbf{x}'\boldsymbol{\beta} = \beta_{0i} + \beta_{1i}\,\text{gender} + \beta_{2i}\,z_1 + \beta_{3i}\,z_2 + \beta_{4i}\,\text{math}$$

the parameters μ_i in NB, GP and PIG; the parameters α_i in BTD and PLB models and the parameters q_i in models , NLD, NGDP and DW are modeled respectively as follows:

$$\mu_i = \exp(\mathbf{x}'\boldsymbol{\beta}); \quad \alpha_i = \exp(\mathbf{x}'\boldsymbol{\beta}); \quad q_i = 1/[1+\exp(-\mathbf{x}'\boldsymbol{\beta})]; \quad i = 1, 2, \ldots, 314.$$

Here, the q_i, are modeled in the logit form. We may first note here that the data is grossly over-dispersed with a dispersion parameter DI=6.5178 under the Poisson model. The results therefore in Table 8 gives the performances of the seven two-parameter distributions considered as alternatives to the Poisson. We therefore have the following summary results, viz:

- Based on the -2 log-likelihood, the PIG has the smallest -2LL and would be considered the most parsimonious on this criterion.
- Based on the Wald's test statistic, X_W^2, the PIG also gives the lowest value on 308 d.f. and with a p-value of 0.9992. All the models fit the data somehow except the Bell-Tuchard (BTD) distribution. The BTD grossly underestimates the variances of the observed data.
- Based on the root mean square error (rmse) defined here as:

  $$\text{RMSE} = \sqrt{\sum_{i=1}^{314} \frac{(y_i - \hat{m}_i)^2}{314}}$$

  The NLD gives the smallest root mean square error (RMSE) and this is followed by the NB, the discrete Weibull (DWD) and the new geometric discrete Pareto distribution (NGDP).
- The estimate of the parameter α for the NGDP, which is modeled as $\mathbf{x}'\boldsymbol{\beta}$ is approximately zero, which would be equivalent to a DWD with parameter $\beta = 1$. For this data example, the estimate of the parameter β in the DWD is $0.9999 \approx 1$, hence the seemingly similar results for both the NGDP and the DWD models in Table 8.

Parameter	NB	GP	PLB	NGDP	BTD	PIG	DWD	NLD
	Distributions							
intercept	2.7075	2.6731	1.3055	2.7070	1.3012	2.8408	2.7040	2.7449
	(0.2031)	(0.2384)	(0.3898)	(0.2073)	(0.0567)	(0.1418)	(0.2496)	(0.2462)
gender(Male)	-0.2111	-0.1943	0.1766	-0.2108	-0.0611	-0.2095	-0.2107	-0.2109
	(0.1224)	(0.1290)	(0.0993)	(0.1247)	(0.0236)	(0.0986)	(0.1248)	(0.1262)
Program-Z1	-0.4245	-0.4207	0.3871	-0.4245	-0.1013	-0.6120	-0.4244	-0.4310
	(0.1818)	(0.2160)	(0.1480)	(0.1856)	(0.0278)	(0.1202)	(0.1857)	(0.1896)
Program-Z2	-1.2526	-1.2478	1.0797	-1.2525	-0.3346	-1.3068	-1.2518	-1.2580
	(0.2014)	(0.2276)	(0.1648)	(0.2054)	(0.0432)	(0.1496)	(0.2075)	(0.2089)
math	-0.0062	-0.0058	0.0051	-0.0062	-0.0017	-0.0064	-0.0062	-0.0063
	(0.0025)	(0.0027)	(0.0020)	(0.0025)	(0.0005)	(0.0020)	(0.0025)	(0.0026)
ML	$\hat{r}$=1.0473	$\hat{\alpha}$=0.2917	$\hat{\beta}$=24.0434	$\hat{\alpha}$=≈ 0.000	$\hat{\theta}$=0.1137	$\hat{\beta}$=7.6632	$\hat{\beta}$=0.9990	$\hat{\alpha}$=0.1076
	(0.1082)	(0.0253)	(10.0059)	-	(0.0255)	(1.4242)	(0.0473)	(0.3399)
-2LL	1728.3	1741.7	1728.6	1728.5	1779.8	**1718.0**	1728.5	1728.4
AIC	1740.3	1753.7	1740.6	1740.5	1791.8	1730.0	1740.5	1740.4
X_W^2	345.7758	333.8455	337.9984	333.2553	527.0813	**235.2896**	332.6307	324.3440
d.f.	308	308	308	308	308	308	308	308
p-value	0.0680	0.1491	0.1154	0.1544	0.0000	**0.9992**	0.1602	0.2502
RMSE	6.3431	6.3484	6.3441	6.3432	6.3369	6.3714	6.3432	**6.3427**

Table 8: Parameter estimates & GOF statistics for the Distributions

Z1-Academic, Z2-Vocational; Parameter standard errors are in parenthesis

6 Computation of Model Means and Variances

The Wald's GOF is a function (reciprocal) of estimated variance(s), hence there is need to compute these sample moments. The NB, GP, PIG and the Discrete Bell-Touchard (BTD) all have close expressions for their means and variances in terms of their estimated parameters. However, the DWD, NLD and NGDP and PLB have summation expressions for their means and variances as expressed in the preceeding sections. Three approaches are employed for the computations of these moments, viz:

1. Based on estimated probabilities from $(0, \infty)$. thta is,

$$\begin{aligned} E(Y) &= \sum_{j=0}^{\infty} j\hat{\pi}_j \\ E(Y^2) &= \sum_{j=0}^{\infty} j^2\hat{\pi}_j \end{aligned} \tag{25}$$

   with $\hat{\pi}_j = f(j|\hat{\alpha}, \hat{\theta}),\ j = 0, 2, \ldots$, for the NLD for example and hence $\text{Var}(Y) = E(Y^2) - [E(Y)]^2$

2. Based on similar expressions in (25) and are defined as:

$$\begin{aligned} E(Y) &= \sum_{j=1}^{\infty} jf(j|\hat{\alpha}, \hat{\theta}) \\ E(Y^2) &= \sum_{j=1}^{\infty} j^2 f(j|\hat{\alpha}, \hat{\theta}) \end{aligned} \tag{26}$$

3. Based on the summation expressions, which for the NLD for example, are in (15).

Procedure in (26) is particularly suitable for GLM models where the means and variances vary for each observation. For frequency data, the three approaches lead to the same result. To illustrate this, consider the data in Example I. When the NLD model is implemented, the estimated mean and variance are as presented in Table 9, where the upper limit of the summation is $[\infty = 13]$. At $Y = 13$, sum of expected values add to $n = 4000$ as well as the estimated probabilities $\hat{\pi}_i$ summing to 1.00 (this is necessary for Method 1). At the bottom of panel are the estimated means and variances for the first five observations in the Congo data, when the NLD model is applied. Notice that method 1 is not applicable in this case. The upper summation here is 180. Notice that the estimated values for method 2 are slightly different from those of method 3. Method 2 suffers from floating point errors if the upper summation limit is excessively high.

	Method 1		Method 2		Method 3		
obs #	Mean	Variance	Mean	Variance	Mean	Variance	∞
	0.086575	0.119957	0.086575	0.119957	0.086575	0.119957	13
Congo Data							
1	na	na	5.356592	35.032377	5.356592	35.032377	180
2	na	na	6.711183	53.248609	6.711183	53.248609	180
3	na	na	8.661351	86.106180	8.661352	86.106304	180
4	na	na	8.880930	90.296001	8.880932	90.296196	180
5	na	na	9.694167	106.677444	9.694172	106.678307	180

Table 9: Estimated moments for Data Set I under NLD Model

6.1 Case of Under-Dispersed Data with Covariates

Here, we are going to employ the asthma inhaler data presented in Grunwald *et al.* (2011) and Canale and Dunson (2012), which comprises of 5209 daily count observations from 48 children suffering from asthma during the school day, for a certain period in Denver, Colorado. The students are aged 6 to 13 years. The covariates here are: (i) the percentage of humidity (humidity), (ii) the barometric pressure (in mmHG/1000-pressure), (iii) the average daily temperature (in Fahrenheit degree/100- temperature), and (iv) the morning levels of PM25, which are small air particles less than 25mm in diameter (particles). The linear predictor here is

$$\mathbf{x}'\boldsymbol{\beta} = \beta_0 + \beta_1\text{humidity} + \beta_2\text{pressure} + \beta_3\text{temperature} + \beta_4\text{particle}.$$

The response variable (Y), which is the inhaler use count, has a sample mean of 1.2705, sample variance of 0.9183, and therefore, a dispersion index DI $= 0.6637 < 1$. Thus, the data is under-dispersed. In Table **??** are the results of the applications of the three models with the Poisson as a baseline model. some of the models discussed above to the inhaler data.

For the models considered in this study, the NB is not suitable for fitting under-dispersed count data because of convergence issues. The Poisson inverse Gaussian and the discrete Bell-Touchard Distribution are mainly suitable for overdispersed data as a result of the structure of their variances. Consequently, we only explore the NLD, NLPG, the GP and the DW models in this case. The results of applying these models are presented in Table 10

	Distributions			
Parameter	NGDP	NLD	DW	HPP
Intercept	-4.9006	-3.7526	-1.6193	-4.0770
humidity	-0.0064	-0.3741**	-0.2477*	-0.1688
pressure	4.9708**	2.7998	5.1593	6.1452
temperature	-0.0471	-0.5706**	-0.4150*	-0.3055
particles	0.0217**	0.0346	0.0317	0.0338*
ML	$\hat{\alpha}$=-2.5050	$\hat{\alpha}$= -1790.77	$\hat{\beta}$=2.1282	$\hat{\lambda}$=0.2656**
-2LL	13448	14134	13,472	13,417
AIC	13460	14,146	13,484	13,429
BIC	13500	14,186	13,524	13,468
X^2_W	4,898.2025	4,332.7955	5,162.9150	5,176.0143
d.f.	5203	5203	5203	5203
p-value	-	-	-	-

Table 10: Parameter estimates & GOF statistics for the Distributions

* significant at the 5% point ** significant at the 10% point

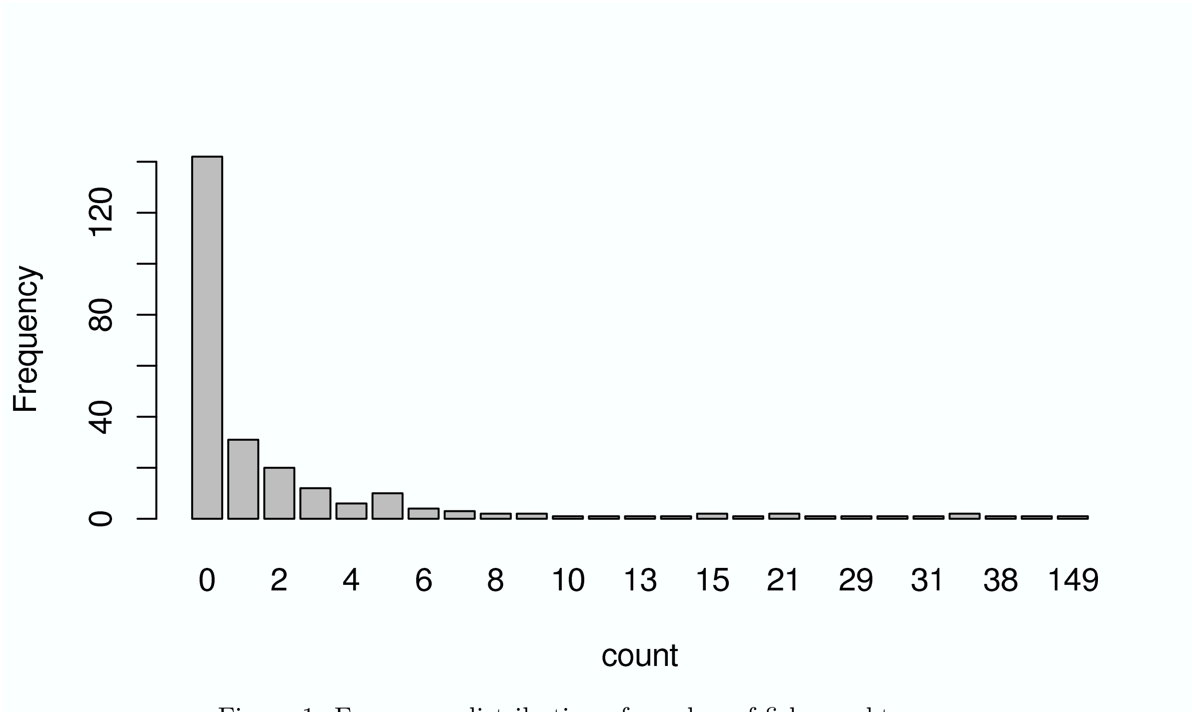

Figure 1: Frequency distribution of number of fish caught

Results:

Results from Table 10 indicate that for under-dispersed data, these distributions sometimes fail to converge. While the NLD here gives a Wald's pvalue of () and being the most parsimonious based on this criterion, however, the NGDP and DW have lower AIC and BIC, with the former being the most parsimonious. More often, the parameter estimates are sometimes suspect with under-dispersed data. Consequently here, the Discrete Weibull (DW) would be the choice model for this data set.

6.2 Over-dispersed Data with Excess zeros

The fish data set presented here comes from wildlife biologists interested in number of fish caught in a certain state park. The response variable Y and the predictor variables are:

- Y-number of fish caught
- child-number of children in the group
- persons-number of people in the group
- camper-[0,1]. Is 1 if a camper is brought to the park and zero otherwise.

There were 250 groups that went to the pack with 142 of them having zero count because they did not catch any fish. Thus 56.8 % of the data are zeros. The histogram distribution of the response variable is presented in figure 1.

6.2.1 Results:

The results of implementing the distributions discussed in the previous sections to the fish data are presented in Table 11. The positive part of the ZI model is modeled as in (27a), while the zero-component is modeled in the logit form (27b).

$$\mu_i = \exp(b_0 + b_1\text{camper} + b_2\text{persons} + b_3\text{children}) \tag{27a}$$

$$\log\left(\frac{\phi_i}{1-\phi_i}\right) = a_0 + a_1\text{children} \tag{27b}$$

Parm.	ZINB	ZIGP	ZIPLB	ZINGDP	ZIBTD	ZIPIG	ZIDW	ZINLD
Int.	-1.6600	-0.8318	3.3079	-1.4996	0.2949	-0.8861	-1.8761	-0..9984
Camper	0.5834*	0.7096**	0.4916**	0.5480*	0.2088**	0.7146**	0.6659**	0.6083**
Persons	1.0516**	0.7670**	0.7166**	1.1070**	0.2583**	0.7838**	0.8412**	1.0363**
Children	-1.2056**	-1.233**	-0.8458**	-1.1200**	-0.3569**	-1.2195**	-1.4922**	-1.2332**
zero-Model								
Int.	-4.4302**	-29.4971	-4.0638**	-3.3912**	-9.2452	-8.3472	-16.7405	-4.7258**
Children	2.9263**	14.7752	2.7526**	2.5778**	4.9485	4.3064	-2.0406	2.9790**
MLE	$\hat{r}$=0.5586	$\hat{\delta}$=0.7374	$\hat{\beta}$=4.8949	$\hat{\alpha}$=0.1529	$\hat{\theta}$=0.0599	$\hat{\beta}$=17.6771	$\hat{\beta}$=0.6925	$\hat{\alpha}$=0.8851
-2LL	799.8	792.0	791.8	807.7	892.7	785.9	796.6	793.5
AIC	813.8	806.0	805.8	821.7	906.7	799.9	810.6	807.5
X^2	417.3731	170.7851	480.3507	448.7525	673.3547	131.7135	671.7992	423.4727
d.f.	243	243	243	243	243	243	243	243
p-value	0.0000	na	0.0000	0.0000	na	na	na	0.0000
$\hat{p}_{\text{zero}}$	0.5685	0.5789	0.5558	0.5516	0.5811	0.5657	0.5795	0.5712

Table 11: Estimated Parameters for the Zero-Inflated Models

Results in Table 11 indicate that:

- The ZIGP, ZIBTD, ZIPIG and ZIDW did not fit well the zero-component of the ZI model. They all return very large standard errors of the zero-component parameters and are therefore not well suited for the ZI model for this data set. These models give either over-estimated or under-estimated variances of the observations, leading to very low or very high Wald's X^2, Consequently, we have labeled the p-values 'na' for these cases.

- Of the remaining models, the ZIPLB is the most parsimonious in terms of the AIC, while the ZINB is the most parsimonious in terms of the Wald's GOF.

- The number of children in the group contributed significantly to the zero observations in the data, that is, the logit link parameters.

- In most of the models, the Poisson part indicates that all the three explanatory variables *camper*, *no of persons* and *no of children* are highly significant. Again, the parameter for children here is negative indicating that the number of children in the group has a negative effect on the number of fish caught-all other variables being the same. For instance, for the ZIPLB, this parameter estimate is -0.8458 indication that for a unit increase in the number of children (keeping the the other predictors constant), the number of fish caught reduces by $\exp(-0.8458)$=42.9%.

- The observed proportion of zeros in the data is $142/250 = 0.5680$. All the good models here, ZINB, ZIPLB, ZINGDP and ZINLD give estimated proportion of zeros that are very close to the observed value here.

7 Conclusions

Based on our results in this paper, it's obvious that different distributions fit the example data sets considered in this paper depending on the performance measure in use. This paper believes that the class of models having variance functions ABM and LM with $r \geq 3$ are not really flexible in implementation and some of the well applicable models NB, GP, DWD and PIG are more than adequate to fit these data sets if consideration is solely based on two-parameter discrete models. The NGDP and NLD and their zero-inflated versions are also suitable candidate as well as the Poisson-Lindley Beta prime (PLB) distribution. This gives at least seven variety of two-parameter based distributions. The DW, NLD and NGDP can also be applied to under-dispersed count data. We may note here that fitted ABM(r=1) for example data sets in III and IV in Bale-Lev *et. al.* (2023) is equivalent to the Negative binomial model which is more flexible to implement. The models fail to fit the fish data because of the very strong right-skewness of this data set, and a right truncated model may be more appropriate Saffar & Adnan (2012) and Saffar *et al.* (2019).

The NB, GP, PIG and DWD, PCB,NLD and NGDP 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 the Com-Poisson, the two-parameter Poisson-mixture with the Lindley and the class of Poisson-Sujatha two-parameter distributions.

References

[1] Awad, Y.S., Bar-Lev,S.K. & Makov, U. (2016). A new class counting distributions embedded in the Lee-Carter model for mortality projections. *A Bayesian approach. Technical report*, No 146. Actuarial Research Center, University of Haifa. Israel.

[2] Bardwell, G. E. & 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.

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

[4] Bar-Lev, S.K. & Ridder, A. (2023). Exponential dispersion models for overdispersed zero-inflated count data. *Communications in Statistics-Simulation and Computation*, 52 (7): 3286–3304.

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

[6] Bhati,D. and H.S. Bakouch (2019). A new infinitelydivisible discrete distribution with applications to count data modelling. *Communications in Statistics-Theory and Methods*, 48 (6): 140–16.

[7] Bliss, C.I. & Fisher, R.A. (1953). Fitting the Negative Binomial Distribution to Biological Data. *Biometrics*, 9(2), 176–200.

[8] Canale, A. & Dunson, D.B. (2012). A Bayesian nonparametric model for count functional data. In *46th Scientific meeting of the Italian Statistical Society.*

[9] Castellares,F., Lemonte, A.J., & Moreni-Arenas, G. (2020). On the two-parameter Bell-Tuchard discrete distribution. *Communications in Statistics-Theory and Methods*, 49 (19): 4834–52. doi:10.1080/03610926.2019.1609515.

[10] Consul, P.C, Famoye, F. (1992). Generalized Poisson Regression Model. *Comm. Statist. Theory Methods*, 21(1), 89–109.

[11] El-Morshedy,M, Eliwa,M.S., and Nagy, H. (2020). A new two-parameter exponentiated discrete Lindley distribution: Properties, estimation and applications. *Journal of Applied Statistics*, 47 (2): 354–75.

[12] Gómez-Déniz E, Vásquez-Polo FJ, & Garcia, V. (2017). The modified Borel-Tunner (MBT) regression Model. *REVSTAT Statistical Journal.*

[13] Gómez-Déniz, Sarabia, J.M, and Calderin-Ojeda,E. (2011). A new discrete distribution with actuarial applications. *Insurance: Mathematics and Economics*, 48 (3),406–12.

[14] Gómez-Déniz, E., Hernández-Bastida, A., and Fernández-Sánchez, M.P.A. (2016). A Suitable Discrete Distribution for Modelling Automobile Claim Frequencies. *Bulletin of the Malaysian Mathematical Sciences Society*, 39, 633–647.

[15] Gossiaux, A. & Lamaire, J. (1981). Methodes dájustenant de distributions de sinistres. *Bulletin of the Association of Swiss Actuaries*, 81, 87–95.

[16] Grunwald, G.K., Bruce,S.L., Jiang,L., Strand,G.K. & Rabinovitch, N.A. (2011). Statistical model for under-or-overdispersed clustered and longitudinal count data. *Biometric Journal*, 53(4), 578–594.

[17] Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of American Statistical Association*, 64, 1459–1471.

[18] Hinde, J. and Demétrio, C.G.B. (1998). Overdispersion: Models and Estimation. *Computational Statistics and Data Analysis*, 27, 151–170.

[19] Holla, M. S. (1966). On a Poisson-Inverse Gaussian Distribution, *Metrika*, 15, 377–384.

[20] Jørgensen, B. (1987). Exponential dispersion models (with discussion), *Journal of the Royal Statistical Society*, Ser. B 49(2), 127–162.

[21] Jørgensen, B. & Kokonendji, C. C. (2016). Discrete dispersion models and their Tweedie asymptotics. *AstA Advances in Statistical Analysis*, 100(1), 43–78. doi:10.1007/s10182-015-0250-z.

[22] Kalktawi, H.S., Vinciotti, V. and Yu, K. (2016). A simple and Adaptive Dispersion regression Model for Count Data. *arXiv:1511.00634*.

[23] Karlis, D., Xekalaki, E. (2005). Mixed Poisson distributions. *Int. Stat. Rev.*, 73, 35-58.

[24] Khan, M.S.A., Khalique, A., & Abouammoh, A.M. (1989) On estimating parameters in a discrete Weibull distribution, *IEEE Transactions on Reliability*, 38(3), 348–350.

[25] Klugman,S.A, Panjer, H.H. & Wilmot, G. E. (1998). *Loss Models: From Data to Decisions*. Wiley, New York.

[26] Kokonendji,C.C. & Malouche,D. (2008). A property of Count Distributions in the Hinde-Demetrio Family. *Comm. Statist.-Theory and Methods*, 37: 1823–1834.

[27] Kokonendji, C.C., Dossou-Gbété,S. & Demétrio, C.G.B.(2004). Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *Statistics and Operations Research Transactions*, 28(2),201–14.

[28] 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.

[29] 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.

[30] Letac, G. & Mora, M. (1990). Natural real exponential families with cubic variance functions. *The Annals of Statistics*, 18 (1), 1–37. doi:10.1214/aos/1176347491.

[31] 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.

[32] Nagakawa, T. and Osaki, S. (1975). The Discrete Weibull distribution. *IEEE transactions on Reliability*, 24(5), 300–301.

[33] Pinheiro, J.C. & 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.

[34] Saffari,S.E. & Adnan, R. (2012). Parameter estimation on Zero-inflated Negative Binomial Regression with Right Truncated Data. *Saina Malaysiana*, 41(11), 1483-1487.

[35] Saffari,S.E., Allen J.C., Adnan, R., Ong, S.H., Sim S. Z., & Green, W. (2019). Frequency of visiting Doctor: A right truncated count regression model with excess zeros. *Biostatistics and Biometrics Open Access Journal*, 9(5):555773.

[36] Sankaran, M. (1970). The discrete Poisson-Lindley distribution, *Biometrics*, 26: 145–149.

[37] Shanker, R. (2016). Sujatha distribution and its applications. *Statistics in Transition new Series*, 17(3): 1–20.

[38] Shanker, R. (2017). The Discrete Poisson-Akash Distribution. *International Journal of Probability and Statistics*, 6(1): 1–10.

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

[40] Stein, W.E., & Dattero, R. (1984) A new discrete Weibull distribution. *IEEE Transactions on Reliability*, 33, 196–197.

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

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