Journal of Statistical Modelling: Theory and Applications
Vol. 2, No. 2, 2021, pp. 25-39
Yazd University Press 2021

JSMTA

Research Paper

The inverse trinomial distribution and its application in modeling over-dispersed count Data

Bayo H. Lawal*
LOYAB Statistical Consulting & Data Analytics, Adeleke Road, Lamodi, Offa. Kwara State Nigeria.

Received: December 13, 2021/ Revised: February 21, 2022 / Accepted: February 23, 2022

Abstract: In this paper, we apply the inverse trinomial distribution with parameters (p, q, λ) to count data exhibiting over-dispersion. We compare results for the mean parameterized inverse trinomial and the parameter λ as linear predictors in the generalized linear model case. Our results also demonstrate methods for obtaining the means and variances for the zero-inflated and zero-truncated inverse trinomial distributions. Results obtained here indicate that the generalized Poisson type II has a close relationship with the inverse trinomial distributions. Several data examples are employed in this paper for both frequency and data having covariates cases. SAS PROC NLMIXED is employed using adaptive Gaussian quadrature and the Newton-Raphson as the optimizer.

Keywords: Empirical means and variances; Overdispersion; Zero-inflated models.
Mathematics Subject Classification (2010): 60E05, 62E15, 62P10.

1 Introduction

For data exhibiting over-dispersion (that is, data displaying variability bigger than the mean), various distributions have been suggested to analyze such data. These include the Negative binomial and the generalized Poisson, both have additional dispersion parameters k and α. Lawal (2019) has also employed the three-parameter distributions, the quasi-negative binomial-QNBD (Li et al., 2011), the negative binomial generalized exponential (NBGE), and the inverse trinomial (IT) distributions to model overdispersed count data. Over-dispersion can sometimes manifest as a result of clustering or

*Corresponding author: blawal66@gmail.com

heterogeneity in the data, or it could be as a result of excess zeros in the data (Hilbe, 2013). Sometimes, it could be as a result of the data being truncated at zero. In this paper, we shall focus only on IT distribution. We shall apply the IT model to data exhibiting (i) over-dispersion, (ii) excess zeros, and (iii) zero-truncation. We shall extend these analyses to sets of data with: 1. Frequency count: This presents a problem of expected probabilities not summing to 1.00 (Lawal, 2018). In addition, the expressions for the means and variances are not often available or wrongly expressed. We shall compute empirical moments in each case and show that our procedure gives exactly the same results in cases where expressions for the means and variance of zero-inflated IT (ZIT) or zero-truncated IT (ZTIT) exist.
2. GLM models (that is, models having covariates): With computed empirical moments computed, we shall be able to compute the Wald's test statistics for each of the models being considered in this paper. Further, the NB and generalized Poisson have means μ, which are not functions of either k or α, respectively. Hence, both NB and GP are modeled such that the linear predictor function is given by,

$$\mu_i = \exp(\boldsymbol{x\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}); i = 1, 2, \ldots, n.$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ are the p explanatory variables (covariates) and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ are the $(p+1)$ parameters to be estimated. The IT however has a mean that is a function of some of its parameters. We shall re-parameterize this mean in our applications.

1.1 The inverse trinomial distribution

The IT distribution, (Shimizu and Yanagimoto, 1991), which is derived from the Lagrangian expression has the probability mass function (pmf) of the form

$$P(Y = y) = \frac{\lambda p^\lambda q^y}{y + \lambda} \sum_{t=0}^{|y/2|} \frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!} \times \left(\frac{pr}{q^2}\right)^t, \tag{1}$$

$y = 0, 1, \ldots$; $\lambda > 0, p \geq r$ and $p + q + r = 1$. It is so named because its cumulant generating function is the inverse of that for the trinomial distribution, Yanagimoto (1989). The IT model was employed for overdispersed medical count data in Phang and Loh (2014). A zero-truncated application of the model was also proposed in Phang and Ong (2006), while Phang et al. (2013) observed that the IT distribution presents "a stochastic formulation as a classical one dimensional random walk distribution and is another example of a distribution in the Takac family (Letac and Mora, 1990) with a cubic variance function of the mean". Phang et al. (2013) model the ITD as a three-parameter distribution (λ, p, r). We have modeled the ITD here as a three parameter distribution (λ, p, r) since, $p + q + r = 1$. Its probability generating function (pgf) is given by

$$H(s) = \left[\frac{2p}{(1 - qs) + \sqrt{(1 - qs)^2 - 4prs^2}}\right]^\lambda, \tag{2}$$

where $\lambda > 0$, $0 \leq 4pr/(1-q)^2 < 1$ and $p + q + r = 1$. If $r = 0$, $H(S)$ becomes the pgf of the negative binomial (NB) distribution. Following Khang and Ong (2007), if we substitute $-\theta/\log p$ for λ in (2), with $\theta > 0$, then, (2) becomes

$$H(s) = \exp\left[\theta\left\{\frac{1}{\log p}\log\left\{\frac{1-qs}{2} + \sqrt{\left(\frac{1-qs}{2}\right)^2 - prs^2}\right\} - 1\right\}\right]. \tag{3}$$

The logarithmic function in (3)

$$G(s) = \frac{1}{\log p}\log\left[\frac{1-qs}{2} + \sqrt{\left(\frac{1-qs}{2}\right)^2 - prs^2}\right], \tag{4}$$

and satisfies $G(1) = 1$. Further, expression in (4) can be re-written Khang and Ong (2007), for $r = 0$ as

$$G(s) = \frac{\log(1-qs)}{\log(1-q)}. \tag{5}$$

The expression in (5) is the pgf of the logarithmic distribution (LD) with parameter q, where the pmf of the LD has the form

$$P(Y = y) = \frac{\alpha q^y}{y}, \quad y = 0, 1, 2, \ldots,$$

where $0 < q < 1$ and $\alpha = -1/\log(1-\theta)$. The LD has been widely used in fitting long-tailed data, and various generalizations of the LD have been proposed.

The mean and variance of ITD are

$$\mu = \lambda\left[\frac{1-(p-r)}{p-r}\right], \quad \sigma^2 = \frac{\lambda}{(p-r)^2}\left[1-(p-r) + \frac{2r}{p-r}\right]. \tag{6}$$

Its log-likelihood for a single observation becomes

$$L(y|p, r, \lambda) = \log(\lambda) + \lambda\log(p) + y\log(q) - \log(y+\lambda) + \log Q(y|p, r, \lambda),$$

where

$$Q(y|p, r, \lambda) = \sum_{t=0}^{|y/2|} \frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!} \times \left(\frac{pr}{q^2}\right)^t. \tag{7}$$

1.2 Zero-inflated inverse trinomial distribution

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 $f(y), y = 0, 1, \ldots$ with parameters $\boldsymbol{\theta}$, in the form

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

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. Thus, the probability mass function for the zero-inflated IT (ZIIT) is given by

$$f(y|p, r, \lambda, \phi) = \begin{cases} \phi + (1-\phi)\, p^\lambda, & \text{if } y = 0 \\ (1-\phi)\, f(y), & \text{if } y > 0, \end{cases}$$

where $f(y)$ is the pmf in (1). Its mean and variances are given by Phang et al. (2013).

$$\mu = (1-\phi)\left[\frac{\lambda(1-(p-r))}{(p-r)}\right], \tag{8}$$

$$\sigma^2 = \mu\left(\frac{\lambda-\phi}{p-r}\right) + \frac{\phi\mu^2}{1-\phi} + \frac{2r}{(p-r)^2}(\lambda+\phi). \tag{9}$$

1.3 Zero-truncated models

For a random variable Y with a discrete distribution, where the value of $Y = 0$ can not be observed, the zero-truncated random variable Y_t has the pmf

$$\Pr(Y_t = y) = \frac{\Pr(Y = y)}{\Pr(Y > 0)}, \quad y = 1, 2, 3, \ldots.$$

For the zero-truncated ITD, with parameter $\Pr(Y > 0) = 1 - \Pr(Y = 0) = 1 - p^\lambda$. Hence, the pmf of zero-truncated ITD random variable Y_t becomes

$$f_{zt}(y|p, r, \lambda) = \frac{\lambda p^\lambda q^y}{(y+\lambda)} . \frac{1}{(1-p^\lambda)} \sum_{t=0}^{|y/2|} \frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!} \times \left(\frac{pr}{q^2}\right)^t,$$

for $y = 1, 2, \ldots$. Its corresponding log-likelihood for a single observation becomes

$$L(p, r, \lambda) = \log(\lambda) + \lambda \log(p) + y \log(q) - \log(y+\lambda) + \log Q(y, \lambda) - \log(1-p^\lambda),$$

where $Q(y, \lambda)$ is as defined in (7). Its means and variances will be computationally obtained.

1.4 Parameter estimation

Maximum-likelihood estimations of the above models are carried out with PROC NLMIXED in SAS, which minimizes the function $-L(y, \Theta)$ 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. Convergence is often a major problem here and the choice of starting values is very crucial. By choosing carefully our initial values, convergence is achieved in all cases considered in this paper.

2 Methodology

Because we suspect that the expression for the variance of the zero-inflated IT model as expressed in (9) is suspect, we will demonstrate here alternative ways (based on method of moments) to generate the mean and variance of both the regular IT and its zero-inflated (ZIIT) counterpart. To demonstrate this, we shall fit both models to the Motor vehicle records data set which relates to the number of violation points on the motor vehicle records from Flynn and Francis (2009). This is a skewed distribution with a spike at zero. The data has a mean $\mu = 1.7100$ and a variance $\sigma^2 = 4.6612$ and a dispersion index of 2.7258, thus it is highly over-dispersed considering that the sample size is large $n = 10303$. Further the observed proportion of zeros is 45.2% as against expected percentage of zeros under the Poisson model of $100 * \exp(-1.7100)$=18.09%. Thus this data set is highly zero-inflated.

In Table 1 are the expected values, parameter estimates and goodness-of-fit test statistics for both the regular and adjusted IT and ZIIT models. We observe that for the regaular models, the sum of the expected values are less than the sample size $n = 10,300$. The adjusted models in the last two columns correct this anomaly based on the methodology presented in Lawal (2017, 2018). The regular models do not sum to $n = 10300$ or expected cumulative probabilities being 1 until $Y = 59$ for the IT model and $Y = 30$ for the regular ZIIT model. We present the case of the IT in Table 2. As observed in Lawal (2018), this non-summing to 1 of expected probabilities (in frequency count data) is peculiar to all count distributions, the exception being the Poisson.

Table 1: Frequency counts of motor vehicle violations

		Regular models		Adjusted models	
Y	count	IT	ZIT	IT	ZIT
0	4659	4425.6603	4659.0000	4426.3462	4659.0009
1	1467	1644.3544	1348.2128	1644.2361	1349.0425
2	1199	1660.1942	1318.3952	1659.3165	1318.1189
3	966	841.9919	1046.1703	841.7786	1045.6487
4	727	636.5118	743.3909	636.3016	742.9382
5	528	370.1444	486.4691	370.1717	486.2607
6	341	255.8208	300.6443	255.8758	300.6296
7	213	158.7669	177.7541	158.8704	177.8486
8	114	106.5054	101.5783	106.6042	101.7081
9	53	68.3432	56.4836	68.4372	56.6064
10	20	45.4820	30.7224	45.5609	30.8208
11	13	29.7385	16.4088	29.8036	16.4802
12	1	19.7967	8.6318	19.8482	8.6802
13	2	13.0968	4.4830	39.8489	9.2162
Total	10303.000	10276.4074	10298.3445	10303.0000	10303.0000
		$\hat{p} = 0.7672$	$\hat{p} = 0.8445$	$\hat{p} = 0.7669$	$\hat{p} = 0.8436$
		$\hat{r} = 0.1163$	$\hat{r} = 0.0376$	$\hat{r} = 0.1164$	$\hat{r} = 0.0377$
		$\hat{\lambda} = 3.1888$	$\hat{\lambda} = 11.0237$	$\hat{\lambda} = 3.1837$	$\hat{\lambda} = 10.9504$
			$\hat{\phi} = 0.3515$		$\hat{\phi} = 0.3514$
AIC		36321	35958	36311	35953
BIC		36343	35985	36338	35982
X_W^2		9033.7789	10182.9430	9414.9837	10203.0125
d.f		10299	10298		
X_g^2		359.8043	58.0420	386.0020	62.1342
d.f		10	10	9	9
p-value		0.0000	0.0007	0.2395	0.0000

2.1 Computations of means and variances

The theoretical mean and variance of the IT model are given in expressions (6). Those of ZIIT are given respectively in (8) and (9). However, we suspect that the expression for the variance of the ZIIT in (9) is not accurate.

The method moments employed here, generate the means and variances from the following expressions for IT model:

$$E(Y) = \sum_{j=1}^{\infty} j\ f(y|\hat{p}, \hat{r}, \hat{\lambda}), \quad \mathrm{Var}(Y) = \sum_{j=1}^{\infty} j^2 f(y|\hat{p}, \hat{r}, \hat{\lambda})) - [E(Y)]^2,$$

where $f(y|\hat{p}, \hat{r}, \hat{\lambda})$ is the pmf of the IT distribution under estimated parameters $\hat{p}$, $\hat{r}$ and $\hat{\lambda}$. These values are compared with the theoretical values computed from expressions in (6). A similar approach has been used in Lawal (2018).

For the zero-inflated model, the means and variances are estimated from the following expressions in (10)

$$\widehat{E}(Y) = \sum_{j=1}^{\infty}(1-\hat{\phi})\, j\ f(y|\hat{p}, \hat{r}, \hat{\lambda}), \quad \widehat{\mathrm{Var}}(Y) = \sum_{j=1}^{\infty} (1-\hat{\phi})\, j^2 f(y|\hat{p}, \hat{r}, \hat{\lambda})) - [E(Y)]^2. \quad (10)$$

2.2 Probability based values

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 IT 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. $\hat{\mu}_j$, 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.

We observe immediately the following:

• At $Y = 13$, that is, $j = 13$, the sum of expected values and its corresponding mean and variances are (double asterisk) 10276.4074, 1.668512 and 4.772391, which indicate that at this point it is not yet a pdf because $\sum \hat{\pi}_k = 0.997419 < 1.00$.

• The cumulative probability at $Y = 13$ is $0.997419 < 1$. Hence this is not yet a probability distribution within the range of our data $0 \leq Y \leq 13$.

• Although the cumulative probabilities sum to 1.000000 at $j = 50$, but we see that the predicted values keep increasing, and the estimated variance is not yet stable. Had we presented the results to 8 or ten decimal places, we would have observed that the estimated probabilities are not yet zero. Further, had we presented the results to three decimal places, we would not have observed that the probabilities really do not sum to 1.00 until $Y = 57$.

• From $Y = 57$, the means and variances are stable and thus, $Y = 57$ is the values of Y at which the means and variance can be obtained.

The results of these computations are presented in Table 3.

Results here, indicate that for the ZIT, all the three methods give the same values for its mean and variance. However, for the IT, the theoretical variance is highly inflated, 25.511905, which indicates that the expressions for this variance as presented

Table 2: Moments computation under IT model

j	$\hat{\pi}_j$	$\sum_{i\leq j}\hat{\pi}_i$	$\hat{m}_j$	$\sum_{i\leq j}\hat{m}_i$	$\hat{\mu}_j$	w	$\hat{\sigma}_j^2$
0	0.429551	0.429551	4425.66028	4425.66028	0.000000	0.000000	0.000000
1	0.159600	0.589150	1644.35442	6070.01470	0.159600	0.159600	0.134128
2	0.161137	0.750287	1660.19417	7730.20887	0.481874	0.804147	0.571945
3	0.081723	0.832010	841.991853	8572.20072	0.727042	1.539654	1.011064
4	0.061779	0.893789	636.511783	9208.71250	0.974160	2.528123	1.579136
5	0.035926	0.929715	370.144444	9578.85695	1.153789	3.426270	2.095041
6	0.024830	0.954545	255.820811	9834.67776	1.302767	4.320140	2.622938
7	0.015410	0.969955	158.766916	9993.44468	1.410636	5.075219	3.085326
8	0.010337	0.980292	106.505442	10099.9501	1.493334	5.736808	3.506760
9	0.006633	0.986925	68.343233	10168.2934	1.553034	6.274108	3.862192
10	0.004414	0.991340	45.482032	10213.7754	1.597179	6.715553	4.164572
11	0.002886	0.994226	29.738526	10243.5139	1.628929	7.064806	4.411396
12	0.001921	0.996148	19.796750	10263.3107	1.651987	7.341496	4.612436
13	0.001271	0.997419	13.096792	10276.4074	1.668512	7.556322	4.772391**
14	0.000849	0.998268	8.747090	10285.1545	1.680398	7.722723	4.898987
15	0.000566	0.998834	5.833824	10290.9884	1.688891	7.850124	4.997772
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	0.000000	1.000000	0.000011	10303.0000	1.709987	8.239593	5.315537
51	0.000000	1.000000	0.000008	10303.0000	1.709987	8.239595	5.315539
52	0.000000	1.000000	0.000005	10303.0000	1.709987	8.239596	5.315540
53	0.000000	1.000000	0.000004	10303.0000	1.709987	8.239597	5.315541
54	0.000000	1.000000	0.000003	10303.0000	1.709987	8.239598	5.315541
55	0.000000	1.000000	0.000002	10303.0000	1.709987	8.239599	5.315542
56	0.000000	1.000000	0.000001	10303.0000	1.709987	8.239599	5.315542
57	0.000000	1.000000	0.000001	10303.0000	1.709987	8.239599	5.315543***
58	0.000000	1.000000	0.000001	10303.0000	1.709987	8.239600	5.315543
59	0.000000	1.000000	0.000000	10303.0000	1.709987	8.239600	5.315543
60	0.000000	1.000000	0.000000	10303.0000	1.709987	8.239600	5.315543
61	0.000000	1.000000	0.000000	10303.0000	1.709987	8.239600	5.315543

Table 3: Results of estimation of moments

	Regular			
	IT		ZIT	
Moments	μ	σ^2	μ	σ^2
Theoretical	1.709987	5.315543	1.709987	25.511905
Probaility Based	1.709987	5.315543	1.709987	4.726031
Moments	1.709987	5.315543	1.709987	4.726031

	Adjusted			
	IT		ZIT	
	μ	σ^2	μ	σ^2
Theoretical	-	-	-	-
Probability Based	1.702422	5.100384	1.709268	4.706399
Moments	-	-	-	-
Observed	1.7100	4.6612		

in Phang and Ong (2006) and expressed in (9) is not correct. Further, we see that while both the IT and ZIT models produce estimated means that are very close to the observed mean of 1.7100 in the data, the estimated variance produced by the IT model grossly overestimates the observed variance of 4.6612. However, the zero-inflated model corrects this somehow but is still not very close to the observed value. From the above therefore, we are assured that the method of moments presented in the earlier sections would be suitable to produce the means and variances when our models are applied to data having covariates. These empirical means and variances are needed to compute the Wald's Goodness-of-fit test statistic: $X_W^2 = \sum_{i=1}^{N} \frac{y_i - \hat{\mu})^2}{\hat{\sigma}^2}$.

2.3 Zero-Truncation example

For this example, we shall employ the Chey (2002) moth data from lightly logged rain forest site in Sabah, Malaysia. The variable Y in Table 4 is the number of species represented by n individuals.

Table 4: Parameter estimates under ITD

Y	Count	ZTNB	ZTGP	ZTECOMP	ZTITD
1	140	128.0879	136.1538	139.0684	136.4323
2	36	45.6199	41.3043	38.2503	41.1600
3	17	21.6591	18.7331	17.3800	18.6260
4	13	11.5677	10.0612	9.8539	10.0108
5	6	6.5897	5.9346	6.2337	5.9145
6	2	3.9103	3.7159	4.1695	3.7103
7	4	2.3866	2.4249	2.8644	2.4262
8	2	1.4869	1.6313	1.9873	1.6355
9	2	0.9411	1.1232	1.3783	1.1285
10	0	0.6031	0.7877	0.9494	0.7931
11	1	0.3904	0.5607	0.6468	0.5657
12	0	0.2548	0.4041	0.4346	0.4085
13	0	0.1675	0.2942	0.2876	0.2981
14	0	0.1107	0.2161	0.1871	0.2194
15	1	0.2243	0.6548	0.3087	0.6710
Total	224				
		$\hat{\mu} = 0.00112$	$\hat{\mu} = 0.00982$	$\hat{\nu} = 1.5853$	$\hat{p} = 0.4746$
μ	1.98661	$\hat{k} = 2211.91$	$\hat{\tau} = 99.9539$	$\hat{p} = 5.0161$	$\hat{r} = 0.0715$
σ^2	3.65453			$\hat{\alpha} = -9.0000$	$\hat{\lambda} \approx 0.00$
				$\hat{\beta} = -9.6245$	
-2L		593.1	589.7	588.7	589.7
AIC		597.1	593.7	596.7	595.7
BIC		603.9	600.5	598.3	605.9
$\hat{\mu}_{zT}$		1.98661	1.98661	1.98632	1.98661
$\hat{\sigma}^2_{zt}$		2.95256	3.87452	3.63118	3.90882
Wald's X^2		275.9250	210.3385	226.4631	208.4926
X_g^2		12.5312	6.6208	6.6194	6.5534
d.f		11	12	10	11
pvalue		0.3251	0.8816	0.7608	0.8340

Results of the zero-truncated models on Chey's (Chey, 2002) moth data are presented in Table 4. We have also fitted the zero-truncated negative binomial (ZTNB), the zero-truncated generalized Poisson (ZTGP) and the zero-truncated extended Com-Poisson distribution-ZTEC

OMP, Chakraborty and Imoto (2016). The probability distributions for these three ZT distributions are presented in (11) to (13) respectively:

$$f_{zt}(y_t|\mu, k) = \frac{\dfrac{\Gamma(y_i + \frac{1}{k})}{\Gamma(y_i + 1)\,\Gamma(\frac{1}{k})} \left(\dfrac{1}{1 + k\mu_i}\right)^{1/k} \left(\dfrac{k\mu_i}{1 + k\mu_i}\right)^{y_i}}{1 - (1 + k\mu_i)^{-1/k}}, \tag{11}$$

$$f_{zt}(y_t|\mu, \alpha) = \frac{\left(\dfrac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \dfrac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp\left(-\dfrac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)}\right)}{1 - \exp\left[-\dfrac{\mu_i}{(1 + \alpha\mu_i)}\right]}, \tag{12}$$

$$f_{zt}(y_t|\nu, p, \alpha, \beta) = \frac{[\Gamma(\nu + y)]^\beta}{[\Gamma(\nu)]^\beta\, [{}_1S^\beta_{\alpha-1}(\nu, 1; p) - 1]} . \frac{p^y}{(y!)^\alpha}, \tag{13}$$

where ${}_1S^\beta_{\alpha-1}(\nu, 1; p)$ in (13) is defined as:

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

and the distribution is defined in the parameter space

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

The observed mean and variance for this data set are respectively, $\mu = 1.98661$ and $\sigma^2 = 3.65452$, giving a dispersion index of $1.8395 > 1$. Thus, this data set is overdispersed. The following are our observations when these models are applied to the data in Table 4.

• As expected in the interval $1 \leq Y \leq 15$, none of the models have cumulative probabilities summing to 1. The cumulative probabilities under each model are respectively, $\{0.99933, 0.99779, 0.99916, 0.99773\}$.

• The corresponding sums of expected frequencies under each model respectively are $\{223.84925, 223.50512, 223.81099, 223.49178\}$.

• The cumulative probabilities sum to 1.00 at Y equals $\{56, 86, 34, 80\}$ respectively. The theoretical means and variances are thus computed at these points.

• Thus at $Y = 15$ non-of these distributions have probabilities summing to 1, which in turn affects the sum of the expected values not being $n = 224$, the observed sample size.

• The theoretical means of ZTNB, ZTGP and ZTIT are all equal to the observed mean in the data.

• The theoretical variances are however underestimated in ZTNB, and overestimated in all others.

• Clearly the ZTGP and the ZTIT are more parsimonious than the other models. The difference in the Wald's X^2 between the ZTGP and the ZTIT being as a result of slightly inflated theoretical variance of the latter, which lowers the Wald's test statistic X^2.

• Both ZTGP and ZTIT behave alike, this is not surprising because as noted in Khang and Ong (2007), the generalized Poisson distribution is a special case of the ITD distribution.
• ZTIT is a special case of the generalized logarithmic distribution (GLD) in this application with $\lambda \approx 0$ in the ZTIT.
• The grouped $X_g^2 = \sum_{k=1}^{15} \frac{(f_k - \hat{m}_k)^2}{\hat{m}_k}$ in table 4 are generated using the Lawal (1980) rule that the minimum expected values can be as low as $r/d^{3/2}$ where r is the number of expected values less than 3, and r is the degree of freedom for the χ^2 approximation to be valid.

3 GLM applications

We employ two data sets that have been previously analyzed to illustrate fitting the distributions discussed above. The first data set is the NMES (The US National Medical Expenditure Survey 1987 and 1988). The data has previously been analyzed in Deb and Trivedi (1997). The data has been used to model medical care. However, we will model the response variable, HOSP, the number days stayed in hospitals. The other covariates modeled are : number of chronic conditions (NUMCHRON), age (AGE), sex (MALE), private insurance indicator (PRIVINS), Medicaid indicator (MEDICAID) and self-perceived health status (EXCLHLTH and POORHLTH). Thus,

$$\boldsymbol{x\beta} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i},$$

where, x_1 = NUMCHRON, x_2 = age, x_3 = sex, x_4 = PRIVINS, x_5 = MEDICAID, x_6 = EXCLHLTH and x_7 = POORHLTH.

However, before we consider the GLM (i.e models with covariates) applications of these distributions to our data, we want to first focus on the distribution of the response variable Y-the number of hospital stays. The data has $n = 4406$ observations and the range of Y is [0,8] with 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 zero responses. The IT and Zero-Inflated models applied to the response variable Y indicate that $\hat{\phi} = 0$ indicating that, the IT model is suitable for the frequency count Y giving parameter estimates $\hat{p} = 0.6568$, $\hat{q} = 0.3231$ and $\hat{\lambda} = 0.5187$ with the grouped $X^2 = 7.3592$ with 6 d.f. and Wald's $X^2 = 4679.16$ with 4402 d.f.

3.1 GLM formulation

For data having covariates $x_1, x_2, \ldots, x_p$, the linear predictor function is defined as $\log(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ and consequently, $E(Y_i) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$.

GLM models for the Poisson, NB and generalized Poisson (Type I) have all been modeled as a function of the mean μ_i for these distributions. That is, $\mu_i = \exp(\boldsymbol{X\beta})$. Thus, to capture this formulation for the IT, we recall that for the IT model, $\mu = \lambda \left[\frac{1-(p-r)}{p-r}\right]$. Hence, $\lambda = \left[\frac{p-r}{1-(p-r)}\right].e^{\boldsymbol{x}_i'\boldsymbol{\beta}}$.

4 Results

The results of applying all the above models to the NMES data set are presented in Table 5. Only the ZIT model results are presented and our analysis indicate that only three of the covariates are actually needed to model the zero-component of the ZIIT model. The parameter estimates from using λ or mean parameterized IT are exactly the same. The only difference between only the intercept terms. The IT$_\mu$ intercept parameter is however similar to only those of NB and GP1. Consequently, there is no reason to prefer the IT$_\mu$ over the IT$_\lambda$. Consequently, we will adopt IT$_\lambda$ in this study.

Table 5: Parameter Estimates and GOF values under the Models

	P	NB	GP	IT$_\mu$	IT$_\lambda$	ZIT$_\lambda$
Covariate						
Intercept	-3.405**	-3.846**	-3.392**	-3.392**	-2.6057	-0.6723
NUMCHRON	0.264**	0.287**	0.256**	0.256**	0.256**	0.149**
AGE	0.184**	0.237**	-0.188**	0.188**	0.188**	0.02290
MALE	0.1215*	0.1699*	0.1490*	0.1490*	0.1490*	-0.0563
PRIVINS	0.1953*	0.16610	0.14450	0.14450	0.14450	0.14350
MEDICAID	0.2332*	0.20710	0.22780	0.22780	0.22780	0.21480
EXCLHLTH	-0.719**	-0.697**	-0.571**	-0.571**	-0.571**	-0.535**
POORHLTH	0.614**	0.605**	0.589**	0.589**	0.589**	0.584**
Zero-Model						
Intercept						7.790**
NUMCHRON						-0.835**
AGE						-1.005**
MALE						-1.191**
Parms.		$\hat{k}$=1.7563	$\hat{\delta}$=0.2256	$\hat{p}$=0.7026	$\hat{p}$=0.7026	$\hat{p}$=0.7284
estimates				$\hat{r}$=0.0157	$\hat{r}$=0.0157	$\hat{r}$=0.0132
-2LL	6086.5	5710.9	5696.7	5696.6	5696.6	5673.5
AIC	6102.5	5728.9	5714.7	5716.6	5716.6	5701.5
BIC	6153.7	5786.4	5772.2	5780.5	5780.5	5791.0
X^2	7065.54	4746.04	4203.71	4201.94	4201.94	4183.31
d.f.	4398	4397	4397	4396	4396	4391

* sig at 5% ; ** sig at 0.01%

The parameter estimates for the GP, IT$_\mu$ and IT$_\lambda$ are the same, further supporting the observations in Khang and Ong (2007) of the GP being a special form of the IT. Any of the GP, IT and ZIIT would be preferable. However, the GP2 model is the most parsimonious for this data set.

5 ZIT GLM formulations

In this section, we shall apply the zero-truncated models discussed above to the National Health Insurance Scheme (NHIS) data presented in Adesina et al. (2021) and which is fully described in Mendeley Data web site, *https:/data.mendeley/z7wznk53cf/8*. The data, obtained from health facilities in Ota, Ogun State, Nigeria has 1647 patients. The response variable of interest here is Y-the number of encounter visits to the doctors the patient had in a specified period. The predictors in the data set are: covariates eclass-class of admission (in patient=1, outpatient=0), follow-up (follow-up=1, no follow-up=0), sex(male=1, female=0) and age of patient.

The variable Y has $\bar{y} = 3.3892$; $s^2 = 11.5987$, giving a dispersion index of $3.4223 > 1$, thus indicating strong over-dispersion. Further, Y has the range $[1, 27]$, thus it is truncated at $Y = 0$.

We have thus applied the ZT models discussed in the previous sections to the full data with four covariates: sex, age, fup and ecs. Here, the ZTP and ZTNB are modeled with $\mu_i = \exp(a_0 + a_1\text{sex} + a_2\text{age} + a_3\text{fup} + a_4\text{eclass})$. The zero-truncated generalized Poisson employed here is the GP type II, proposed in Consul (1989), Consul and Famoye (1992) for implementing GLM generalized Poisson model and has the zero-truncated pmf given by:

$$f_{ZT}(y_i; \theta_i, \delta) = \frac{\theta_i(\theta_i + \delta y_i)^{y_i-1} e^{-\theta_i-\delta y_i}}{y_i!(1-e^{-\theta_1})},$$

with $y_i = 1, \ldots,$ $\theta_i > 0$, $0 \leq \delta < 1$. Expressions for the mean and variance of the un-truncated version are provided in Joe and Zhu (2005). Thus the distribution can be modeled in the form: $\log\left(\frac{\theta_i}{1-\delta}\right) = \boldsymbol{x}'\boldsymbol{\beta}$.

The log-likelihood for a single observation would therefore be given by:

$$\text{L} = \log(\theta_i) + (y_i - 1)\log(\theta_i + \delta y_i) - (\theta_i + \delta y_i) - \log(y_i!) - \log(1 - e^{-\theta_i}),$$

where, $\theta_i = \exp(\boldsymbol{x}'\boldsymbol{\beta} + \text{offset})$ and $\text{offset} = \log(1-\delta)$.

This version of the generalized Poisson or its zero-truncated counterpart are the ones implemented in SAS PROC HPFMM, STATA and R package *glmmTMB*. We have modeled the ZTIT in terms of its parameter λ, while ZTCOMNB and ZTECOMP are modeled in terms of their parameter p,

$$\begin{aligned}\lambda &= \exp(b_0 + b_1\text{sex} + b_2\text{age} + b_3\text{fup} + b_4\text{eclass}),\\ p &= \exp(a_0 + a_1\text{sex} + a_2\text{age} + a_3\text{fup} + a_4\text{eclass}).\end{aligned}$$

Results in Table 6 are those from implementing the GLM (with covariates) versions of all the zero-truncated models with outcome variable Y. From Table 6 it appears that models ZTGP2, ZTGP1 and ZTIT are suitable candidates for parsimony, with the ZTGP1 and ZTGP2 having one less parameter and much easier to model. However, based on the Wald's goodness-of-fit test statistic, the most parsimonious model is the ZTIT, with a p-value of 0.7343 and closely followed by ZTGP1 and ZTGP2 respectively. We observe here again that the parameter estimates under the ZTGP2 and ZTIT are very similar-indication our earlier observation on th relationship between the generalized Poisson and the Inverse trinomial distribution.

In Tables 7 are the averages of the computationally generated means and variances from the 1647 observations in the data for all the zero-truncated models considered in this study. We recollect that the observed mean and variance of the response variable Y are 3.3892 and 11.5987 respectively. All the models generated an average means that are very close to 3.3892. In particular, ZTP, the ZTGP2 and the ZTIT produce average means that are almost identical with the observed mean of the response variable. Clearly, none of the models produce an estimated mean less than 3.0630 and more than 4.6622. For the corresponding averages of the estimated variances and their corresponding ranges, clearly, the ZTP grossly underestimates the true variance of Y. The

Table 6: ZT models on Y with covariates

Parm.	ZTP	ZTNB	ZTGP2	ZTGPI	ZTIT	ZTCOMNB	ZTECOMP
Int	1.0766*	0.3267	0.5377*	0.6086	-0.2076	-0.4143	0.4900
sex	-0.0126	-0.0144	0.0204	-0.0123	0.0213	-0.0034	-0.0034
age	0.0032*	0.0043*	0.0058	0.0038	0.0058	0.0008	0.0008
fup	0.0726	0.1046	0.0470	0.0945	0.0460	0.0171	0.0176
ecs	0.2146	0.3323	0.2312	0.3063	0.2300	0.0474	0.0499
		$\hat{k} = 2.5302$	$\hat{\delta} = 0.5416$	$\hat{\tau} = 0.5625$	$\hat{p} = 0.3812$	$\hat{\alpha} = 0.9406$	$\hat{\nu} = 16.3999$
				-	$\hat{r} = 0.0591$	$\hat{\nu} = 0.6471$	$\hat{\beta} = -0.4368$
							$\hat{\alpha} = -0.2690$
-2L	8550.7	6709.3	6707.7	6705.8	6707.5	6707.6	6705.9
AIC	8560.7	6721.3	6719.7	6717.8	6721.5	6721.6	6721.5
BIC	8587.8	6753.7	6752.2	6750.3	6759.4	6759.4	6765.1
X^2	6403.81	1699.84	1615.35	1612.49	1603.73	1622.38	1643.72
d.f.	1642	1641	1641	1641	1640	1640	1639
pvalue	0.0000	0.1523	0.6695	0.6875	0.7345	0.6168	0.4626

other models give estimated average values of the variances very close to 11.5987 (the observed variance of Y). Some of the models produce estimated observation variances that really high (29.11431 for instance for the ZTGP1). However, both the ZTGP2 and ZTIT produce average variance that are not too far from the true values and this accounts for their outperforming other models.

Table 7: Average and range of estimated means and variances

	Est. means		Est. variances	
Model	Average	Range	Average	Range
ZTP	3.3892	[3.0670, 4.4496]	2.9459	[2.5484, 4.2053]
ZTNB	3.3896	[3.0637, 4.6161]	11.1787	[8.4581, 24.3863]
ZTGP1	3.3900	[3.0630, 4.6622]	11.8300	[8.6915, 29.1143]
ZTGP2	3.3892	[3.1485, 4.0558]	11.7215	[10.4632, 15.2521]
ZIT	3.3892	[3.1481, 4.0527]	11.8072	[10.5404, 15.3389]
ZTCOMNB	3.3890	[3.0810, 4.6485]	11.7672	[8.7838, 30.3872]
ZTECOMP	3.3893	[3.0791, 4.6038]	11.5716	[8.8096, 25.2648]

6 Conclusion

We have demonstrated in this study that the IT model is quite suitable for modeling over-dispersed count data and that like all other count distributions, when applied to frequency data, the estimated probabilities will not sum to one within the range of the data. We also demonstrate how we could obtain the means and variances when the model is applied to data having covariates, and thus Wald's GOF can be easily computed. This procedure is also applied to the Com-Poisson type distributions whose means do not have close form solutions. The type II generalized Poisson distribution is a suitable candidate for the IT model. Because it is easy in implementation and quicker convergence. The SAS programs for implementing these models are available from the author.

References

Adesina, O.S., Agunbiade, D.A. and Oguntunde, P.E. (2021). Flexible Bayesian Dirichlet mixtures of generalized linear mixed models for count data. *Scientific African*, **13**:1–11.

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.

Chey, V.K. (2002). Comparison of moth diversity between lightly and heavily logged sites in a tropical rain forest. *Malayan Nature J.* **56**(1):23–41.

Consul, P.C. (1989). *Generalized Poisson Distributions; Properties and Applications.* New York: Marcel Dekker.

Consul, P.C. and Famoye, F. (1992). Generalized Poisson Regression Model. *Communications in Statistics-Theory and 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.

Flynn, M. and Francis, L.A. (2009). More flexible GLMs zero-inflated models and hybrid models. *Casualty Actuarial Soc*, **2009**:148–224.

Hilbe, J.M. (2013). *Negative Binomial Regression.* Cambridge University Press.

Joe, H. and Zhu, R. (2005). Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **47**(2):219–229.

Khang, T.F. and Ong, S.H. (2007). A new generalization of the logarithmic distribution arising from the inverse trinomial distribution. *Communications in Statistics-Theory and Methods*, **36**:3–21.

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

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

Lawal, H.B. (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.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Annals of Statistics*, **18**:1–37.

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.

Phang, Y.N. and Ong, S.H. (2006). Zero-Inflated inverse trinomial distribution for modeling count Data. *Proceedings, 2nd IMT-GT regional Conference*.

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.

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

Yanagimoto, T. (1989). The inverse binomial distribution as a statistical model. *Communications in Statistics: Theory and Methods*, **18**(10):3629–3633.