British Journal of Mathematics & Computer Science
21(3): 1-17, 2017; Article no.BJMCS.32184
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

On the Hyper-Poisson Distribution and its Generalization with Applications

Bayo H. Lawal[1*]

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

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/BJMCS/2017/32184
Editor(s):
(1) Prof. Morteza Seddighin, Indiana University East Richmond, USA.
Reviewers:
(1) Antonio Jose Saez-Castillo, University of Jaen, Spain.
(2) Chipo Mufudza, Cukurova University, Turkey.
(3) J. A. Tenreiro Machado, University of La Laguna, Spain.
Complete Peer review History: http://www.sciencedomain.org/review-history/18406

***Received:** 13th February 2017*
***Accepted:** 22nd March 2017*
***Published:** 29th March 2017*

Short Research Article

Abstract

In this paper, we fit the hyper-Poisson, and the Mittag-Leffler function (MLFD) distributions to data exhibiting over and under dispersion. Three frequency data sets were employed with one exhibiting under-dispersion. We also extend these distributions to GLM situations where we have a set of covariates defined in the form $\mathbf{x}'\boldsymbol{\beta}$. In all, we compared the negative-binomial (NB), the generalized Poisson (GP), the Conway-Maxwell Poisson (COMP), the Hyper-Poisson (HP) and the MLFD models to the selected data sets. The generalized linear model (GLM) data employed in this study is the German national health registry data which has 3874 observations with 41.56% being zeros-thus the data is zero-inflated.

Our results contrast the results from these various distributions. Further, theoretical means and variances of each model are computed together with their corresponding empirical means and variances. It was obvious that the two do not match for each of our data sets. The reason being that the models all have infinite range of values than the random variable Y can take, but the data has a finite range of values. It is therefore not unusual for the sum of estimated probabilities being less than 1.00 and consequently, the sum of the expected values are usually less that the sample size n. However, if the range of values of Y are extended beyond the given data value,

*Corresponding author: E-mail: bayo.lawal@kwasu.edu.ng;

both theoretical and empirical moments as expected would be equal. We explore an alternative model for one of the data set. In contrast, most results in the literature sometimes just assume that the last category k has probability $\left(1 - \sum_{k=1}^{y_{(k-1)}} \hat{p}_k\right)$, which does not truly reflect the underlying probability structure from the data.

We have employed SAS PROC NLMIXED in all our computations in this paper with the choice optimization algorithm being the conjugate gradient algorithm. We also computed the Wald test statistic for each data based on both the theoretical and empirical means and variances.
Our results extend previous results on the analyzes of the chosen data in this example. Further, results obtained here indicate that some results in earlier studies on the data employed in this study may be in accurate. In others, our results are consistent with previous analyses on the data sets chosen for this article. While we do not pretend that the results obtained are entirely new, however, the analyses give opportunities to researchers in the field the opportunity of implementing these models in SAS.

Keywords: Hyper-Poisson; negative binomial; overdispersion; underdispersion; GLM.

2010 Mathematics Subject Classification: 53C25, 83C05, 57N16.

1 Introduction

For count data exhibiting over-dispersion or under-dispersion, probability distributions with extra dispersion parameters such as the negative-binomial (NB), the generalized Poisson (GP), the double Poisson and several other distributions (e.g. the Poisson Inverse Gaussian, the NB-Lindley [1], [2], [3], the NB-generalized exponential [4] distribution and many more. Several other distributions have also been employed to model over-dispersed or under-dispersed count data.

While the Poisson distribution is the underlying probability model for count data, its use had been restricted because of the absence of dispersion parameter in its function since both mean and variance are equal, thus leading to equi-dispersion: the ratio of the variance to the mean, which in this case equals 1.00. Consequently, several alternative models utilizing the extended Poisson e.g., the Generalized Poisson [5], the double-Poisson [6], the weighted Poisson, [7]. Most recently however, there has been serious revival of the developments of Poisson based distributions. Notable amongst these are:

1. The Hyper-Poisson Distribution (HP)
2. The Conway-Maxwell: Com-Poisson distribution (COMP) and
3. The Mittag-Leffler function distribution (MLFD)

We will in this paper compare the performances of the above distributions with the negative binomial (NB) and the generalized Poisson (GP) distributions for situations where we have

(a) Frequency count distribution exhibiting over dispersion

(b) A frequency data exhibiting under-dispersion

(c) GLM application of all the models to the German Health data having four covariates

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

2 The Hyper-Poisson Distribution

The hyper-Poisson (HP) distribution first proposed by [8] and [9] is a two-parameter discrete distribution with probability density function (pdf)

$$P(Y = y|\lambda, \beta) = \frac{\Gamma(\beta)}{\Gamma(\beta + y)} \cdot \frac{\lambda^y}{\phi(1, \beta; \lambda)}; \; y = 0, 1, \ldots, ; \quad \beta, \lambda > 0 \tag{2.1}$$

where

$$\phi(1, \beta; \lambda) \quad = \quad \sum_{k=0}^{\infty} \frac{(1)_k}{(\beta)_k} \cdot \frac{\lambda^k}{k!} \quad = \quad \sum_{k=0}^{\infty} \frac{\Gamma(\beta)}{\Gamma(\beta + k)} \lambda^k \tag{2.2}$$

and

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

is the confluent hyper-geometric series in which $(\beta)_0 = 1$.

2.1 Its mean and variance

[10] gave expressions for the mean and variance of HP distribution as:

$$\begin{aligned} \mu &= \frac{\phi(2, \beta + 1, \lambda)}{\phi(1, \beta, \lambda)} \cdot \frac{\lambda}{\beta} \\ \sigma^2 &= \frac{1}{\beta} \left[\frac{2}{\beta + 1} \frac{\phi(3, \beta + 2, \lambda)}{\phi(1, \beta, \lambda)} - \frac{1}{\beta} \frac{[\phi(2, \beta + 1, \lambda)]^2}{[\phi(1, \beta, \lambda)]^2} \right] \lambda^2 + \mu \end{aligned} \tag{2.3}$$

where:

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

The probability generating function of the HP distribution is given by [7] as:

$$g(s) = \frac{\phi(1, \beta + 1, \lambda s)}{\phi(1, \beta, \lambda)} \tag{2.4}$$

Alternatively, we could obtain the mean and variance empirically from expressions:

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

3 The Mittag-Leffler Function Distribution-MLFD (λ, α, β)

This Mittag-Leffler function distribution (MLFD) belongs to the generalized hypergeometric and generalized power series families and also arises as weighted Poison distributions. The MLFD (λ, α, β) has the probability density function (pdf)

$$P(Y = y) = \frac{\lambda^y}{\Gamma(\alpha y + \beta)\ E_{\alpha,\beta}(\lambda)}; \quad y = 0, 1, \dots, \quad \lambda, \alpha, \beta > 0 \tag{3.1}$$

where

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}$$

is the generalized Mittag-Leffler function.

3.1 Properties of MLFD (λ, α, β)

Following [11], the following are some of the characteristics of the MLFD distribution:

(a) When $\alpha = \beta = 1$, then the MLFD (λ, α, β) reduces to the Poisson distribution with parameter λ.

(b) When $\alpha = 0$, and $\beta(\geq 0)$, then MLFD (λ, α, β) becomes the geometric distribution with parameter λ provided $|\lambda| < 1$

(c) When $\alpha = 1$, and $\beta(\geq 0)$ then MLFD (λ, α, β) reduces to the HP (λ, β)

[11] also discussed extensively, the properties of the MLFD including empirical plots of its probability mass functions for various combination values of (λ, α, β). Also discussed are its generating and cumulative distribution functions along with expressions for its moments. It was shown that the MLFD$(\lambda, \alpha, 1)$ has the distribution of a queuing system.
[11] further expressed the probability generating $E[s^x]$ of the MLFD as:

$$P(s) \quad = \quad E_{\alpha,\beta}(\lambda s)/E_{\alpha,\beta}(\lambda) \quad = \quad \frac{\sum_{k=0}^{\infty}(\lambda s)^k/\Gamma(\alpha k + \beta)}{\sum_{k=0}^{\infty}(\lambda)^k/\Gamma(\alpha k + \beta)} \tag{3.2}$$

and it can be shown (see Appendix I) that the mean and variance of the MLFD can be obtained from the following expressions: viz:

$$\mu = \sum_{j=1}^{\infty} \frac{j\lambda^j}{\Gamma(j\alpha + \beta)} \Big/ \sum_{k=0}^{\infty} \frac{\lambda^k}{\Gamma(k\alpha + \beta)} \quad = \quad \mu_1^{'} \tag{3.3}$$

$$\mathrm{E}[X(X-1)] = \sum_{j=2}^{\infty} \frac{j(j-1)\lambda^j}{\Gamma(j\alpha + \beta)} \Big/ \sum_{k=0}^{\infty} \frac{\lambda^k}{\Gamma(k\alpha + \beta)} \quad = \quad \mu_2^{'} \tag{3.4}$$

Consequently,

$$\mathrm{var}(X) = \mu_2^{'} + \mu_1^{'} - \mu_1^{'\,2} \tag{3.5}$$

The mean and variance reduce to that of the hyper-Poisson (HP) when $\alpha = 1$ in (3.3) and (3.5) respectively.

4 The Com-Poisson Distribution

For a random variable Y, [12] introduced the Conway-Maxwell Poisson (COM-Poisson) distribution defined by:

$$f(y|\lambda,\nu) = \frac{\lambda^y}{(y!)^\nu}\frac{1}{Z(\lambda,\nu)}, \quad y_i = 0,1,2,\cdots, \quad \lambda > 0,\ \nu \geq 0. \tag{4.1}$$

Where

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}. \tag{4.2}$$

is the the normalizing term and ν is the *dispersion parameter* such that if $\nu > 1$ we have under dispersion, and when $\nu < 1$, we have over-dispersion. The distribution reduces to the Poisson distribution when $\nu = 1$. The means and variance of Y_i are respectively given as:

$$E(Y) = \frac{1}{Z(\lambda,\nu)}\sum_{j=0}^{\infty}\frac{j\,\lambda^j}{(j!)^\nu} \tag{4.3}$$

and,

$$\text{Var}(Y) = \frac{1}{Z(\lambda,\nu)}\sum_{j=0}^{\infty}\frac{j^2\,\lambda^j}{(j!)^\nu} - E(Y)^2 \tag{4.4}$$

The basic properties of the Com-Poisson model have been presented in [12] and [13] and most recently by [14] who provided an excellent review of this distribution, its properties and applications. [15] has also modeled this distribution for frequency data. The model has been extended in [16] and a Com-Poisson type negative binomial model was proposed by [11]. The distribution has been found to be most useful for under-dispersed count data.

5 Estimation

The log likelihood of a single observation i from HP, MLFD and COMP are given in expressions (5.1a) to (5.1c) respectively:

$$\text{LL1} = y_i\log(\lambda) + \log\Gamma(\beta) - \log\Gamma(y_i+\beta) - \log\left[\sum_{k=0}^{\infty}\frac{\Gamma(\beta)}{\Gamma(\beta+k)}\lambda^k\right] \tag{5.1a}$$

$$\text{LL2} = y_i\log(\lambda) - \log\Gamma(\alpha y_i+\beta) - \log\left[\sum_{k=0}^{\infty}\frac{\lambda^k}{\Gamma(\alpha k+\beta)}\right] \tag{5.1b}$$

$$\text{LL3} = y_i\log\lambda_i - \nu\log y_i! - \log Z(\lambda_i,\nu) \tag{5.1c}$$

Maximum-likelihood estimations of the above models are carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y,\Theta)$ over the parameter space Θ numerically. The integral approximations in PROC NLMIXED is the Adaptive Gaussian Quadrature [17] and the Conjugate Gradient optimization algorithm in PROC NLMIXED (**CONGRA**) of [18] and [19] and the quasi-Newton algorithm (QUANEW) were employed in our computations. Convergence is often a major problem here and the choice of starting values is very crucial. For each of the cases considered here, the above two optimizing algorithms were applied in turn to ascertain accuracy and consistency.

Although the results differ very slightly, on the whole, they both agree very well. Thus we may note here that each of these give slightly different parameter estimates. They all give values that are very close.

Other distributions considered in this study are the negative binomial (NB) and the generalized Poisson distribution (GP). Both have the log-likelihoods functions:

$$\text{LL4} = \log\Gamma(y_i + \frac{1}{k}) - \log\Gamma(y_i + 1) - \log\Gamma(\frac{1}{k}) + y_i\log(k\mu_i) - (y_i + \frac{1}{k})\log(1 + k\mu_i) \quad (5.2a)$$

$$\text{LL5} = y_i\ \log\left(\frac{\mu_i}{1 + \alpha\mu_i}\right) + (y_i - 1)\log(1 + \alpha y_i) - \frac{\mu_i(1 + \alpha y_i)}{1 + \alpha\mu_i} - \log(y_i!) \quad (5.2b)$$

While the dispersion parameter for the NB is designated as k, that for the GP is α and the corresponding means and variances for the two models are given respectively in (5.3).

$$\begin{aligned} E(Y) &= \mu_i; \quad \text{and} \quad \text{Var}(Y_i) = \mu_i + k\,\mu_i^2. \\ E(Y) &= \mu_i; \quad \text{and} \quad \text{Var}(Y_i) = \mu_i(1 + \alpha\mu_i)^2. \end{aligned} \quad (5.3)$$

The GP reduces to the Poisson when $\alpha = 0$ and the dispersion factor $\text{Var}(Y_i)/\text{E}(Y_i) = (1 + \alpha\,\mu_i)^2$. If $\alpha > 0$, then $\text{Var}(Y_i) > \text{E}(Y_i)$ and the GP will model count data with over-dispersion. Similarly, when $\alpha < 0$, then $\text{Var}(Y_i) < \text{E}(Y_i)$ and the GP will in this case be modeling under-dispersed count data.

6 Data Examples

For all the examples in this article (both frequency count and GLM), we have written our own programs to implement them in PROC NLMIXED in SAS. While we are aware that SAS PROC COUNTREG and PROC HFMM can fit some of these models, however, these customized programs can not answer all the question we need to ask regarding these models. The implementation of these models is based on the log-likelihood functions presented in LL1 to LL5 in the previous section.
In this study, we considered three frequency data sets that appeared in the literature.

6.1 Example data I

The data in this example is the frequency data set on insurance claims and incapacity caused by sickness or accident which was analyzed in [20] and [21]. The data is presented in 1.

Table 1

Y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Freq	187	185	200	164	107	68	49	39	21	12	11	2	5	2	3	1	1056

For this data, the observed mean and variance are respectively, $\bar{y} = 2.80587$ and $s^2 = 6.41062$. Consequently, the index of dispersion(ID) $s^2/\bar{y} = 2.8255$ which clearly indicates over-dispersion since ID> 1. Thus there is a need to correct for over dispersion in this data. The results of applying the models discussed in this paper to this data are presented in 2.

Results in 2 indicate that the MLFD, the NB, the GP, the COMP, the HP and the Poisson in that order have the lowest -LL (log likelihood) in that order. Thus the MLFD would seem to be the best model for this data but is closely followed by the Negative binomial. because of optimization problems, we will discuss all the models later in this paper. The most obvious observation in relates to the estimated variance under the Poisson. It grossly underestimates the observed variance of 6.41062 in the data. We see clearly that all the other models tried to adjust their variances and are all not too far from the observed variances. We will discuss this in more details later.

Table 2. Parameter estimates for the models

Distribution	MLEs	-LL	AIC	Mean	var
Poisson	$\hat{\lambda}$= 1.0317	2484.96	4971.9	2.8058	2.8058
NB	$(\hat{\lambda}, \hat{k})$=(1.0317, 0.4687)	2265.80	4535.6	2.8058	6.4958
GP	$(\hat{\lambda}, \hat{\tau})$=(1.0317, 0.1907)	2267.05	4538.1	2.8058	6.6125
COMP	$(\hat{\lambda}, \hat{\nu})$=(1.3895, 0.2492)	2666.05	4536.1	2.7970	6.1849
HP	$(\hat{\lambda}, \hat{\beta})$=(13.0675, 13.6849)	2270.15	4544.3	2.8022	6.2272
MLFD	$(\hat{\lambda}, \hat{\alpha}, \hat{\beta})$ =(0.6114, 0.1282, 0.1603)	2265.30	4536.8	2.7923	6.2198

Apart from the Poisson model, all other models considered in this study the (NB, GP, COMP, HP, and the MLFD) produce estimated probabilities over the range of Y, (in example data I, this is $0 \leq Y \leq 15$) that do not add to 1.00 and consequently, the sum of the expected values do not sum to 1506 in this case. This is true of all these probability models since the range of values of Y is usually infinite even though we have real life data with finite range of values of Y, like the [0,15] in this case. To further accentuate this, we present in 3, for instance, the predicted probabilities and expected values, together with their cumulative values under model HP applied to the data in 1.

Under the HP model, the parameter estimates agree with those in [11]. In 3 are the estimated probabilities $\hat{p}_y$, the estimated cumulative probabilities, $P(Y \leq y)$, the expected values $\hat{m}_y$, the cumulative expected values $\sum_{y=0}^{y_i} \hat{m}_y$, and products $\sum_{y=0}^{y_i} y_i.\hat{p}_y$ and $\sum_{y=0}^{y_i} y_i^2.\hat{p}_y$.

Table 3. Estimated probabilities and expected values under the HP model

Y	$\hat{p}_y$	$P(Y \leq y)$	$\hat{m}_y$	$\sum_{y=0}^{y} \hat{m}_y$	$\sum_{y=0}^{y_i} y_i.\hat{p}_y$	$\sum_{y=0}^{y_i} y_i^2.\hat{p}_y$
0	0.19104	0.19104	201.734	201.73	0.00000	0.0000
1	0.18242	0.37345	192.633	394.37	0.18242	0.1824
2	0.16233	0.53578	171.416	565.78	0.50707	0.8317
3	0.13524	0.67102	142.811	708.59	0.91278	2.0489
4	0.10592	0.77693	111.849	820.44	1.33645	3.7435
5	0.07826	0.85520	82.646	903.09	1.72777	5.7001
6	0.05473	0.90993	57.799	960.89	2.05617	7.6705
7	0.03633	0.94627	38.369	999.26	2.31051	9.4509
8	0.02295	0.96922	24.239	1023.50	2.49414	10.9200
9	0.01383	0.98305	14.607	1038.10	2.61863	12.0404
10	0.00797	0.99102	8.414	1046.52	2.69831	12.8372
11	0.00440	0.99542	4.642	1051.16	2.74667	13.3691
12	0.00233	0.99774	2.458	1053.62	2.77460	13.7042
13	0.00118	0.99893	1.250	1054.87	2.78999	13.9043
14	0.00058	0.99951	0.612	1055.48	2.79811	14.0180
15	0.00027	0.99978	0.289	1055.77	2.80221	14.0795
16	0.00012	0.99991	0.132	1055.90	2.80421	14.1115
17	0.00005	0.99996	0.058	1055.96	2.80514	14.1273
18	0.00002	0.99998	0.025	1055.98	2.80556	14.1349
19	0.00001	0.99999	0.010	1055.99	2.80574	14.1384
20	0.00000	1.00000	0.004	1056.00	2.80582	14.1399

We notice immediately, that for our data, $P(Y \leq 15) = 0.99978 < 1.0000$ with $\sum \hat{m}_i = 1055.77 < 1056$. Under this circumstance, $(\bar{y}, s^2) = (2.8022, 6.2272)$. However, the theoretical values are respectively 2.80587 and 6.26810, using expressions in (2.3), since $\phi(1, \hat{\beta}, \hat{\lambda}) = 5.2346$; $\phi(2, \hat{\beta}+1, \hat{\lambda}) = 15.3816$ and $\phi(3, \hat{\beta}+2, \hat{\lambda}) = 34.9148$ in this case. Because the HP model has infinite range for its random variable Y (this is true for all models in this study), we see from 3 that it was not until $Y = 20$ that we have the cumulative probability summing to 1 and the sum of expected values being 1056 (the observed sum in the data). Clearly, for the range of our data ($0 \leq Y \leq 15$), the empirical means are:$\bar{y} = 2.802211$ and $s^2 = 14.079537-(2.802211)^2 = 6.2272$ and that the estimated probabilities only sum to one for Y in the range $y = 0, 1, \ldots, 20$. In this case, $\hat{\mu} = 2.805820$ and $\hat{\sigma}^2 = 14.139906 - (2.805782)^2 = 6.2673$. The theoretical values of the mean and variance using expressions in (2.3) are respectively, $\hat{\mu}_T = 2.8059$ and $\hat{\sigma}_T^2 = 6.2681$. The empirical and theoretical Wald test statistics computed with empirical and theoretical values are respectively $X_E^2 = 1086.0767$ and $X_T^2 = 1078.9876$, where,

$$X^2 = \sum_{i=0}^{15} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}_i^2}. \tag{6.1}$$

Perhaps it should be pointed out here that the same problems with regards to cumulative estimated probabilities not summing to one over the range of values Y in the data is encountered for the MLFD, Com-Poisson, NB and GP models because they are all defined over an infinite range of values of Y. Consequently, we would have to contend with these observations and realities when we employ these models to fit our data.

Under the MLFD(λ, α, β), we observe that,

$$\sum_{y=0}^{15} \hat{p}_y = 0.99921; \quad \text{while,} \quad \sum_{y=0}^{24} \hat{p}_y = 1.0000$$

Consequently, the sum of the expected values over these two ranges are respectively 1055.17 and 1056. The latter equaling the sample size in the data. Here again the empirical mean and variance in the range $0 \leq Y \leq 15$ are $\bar{y} = 2.792326$ and $s^2 = 14.016856 - (2.792326)^2 = 6.2198$, and in the range $0 \leq Y \leq 24$, we have $\hat{\mu} = 2.805782$ and $\hat{\sigma}^2 = 14.250377 - (2.805782)^2 = 6.37796$. The theoretical values of the mean and variance using expressions in (3.3) and (3.5) are respectively, $\hat{\mu}_T = 2.80587$ and $\hat{\sigma}_T^2 = 6.37979$. Similarly, the Wald test statistics computed with empirical and theoretical values are respectively $X_E^2 = 1087.3979$ and $X_T^2 = 1060.2793$, where X^2 is as defined in (6.1).

For the Com-Poisson model $\sum_{y=0}^{15} \hat{p}_y = 0.99948$ with $\sum \hat{m}_y = 1055.4519 < n$, while $\sum_{y=0}^{28} \hat{p}_y = 1.0000$ with of course, $\sum \hat{m}_y = 1056 = n$. Consequently,the empirical central moments are $\bar{y} = 2.7900$ and $s^2 = 6.1192$. These contrast with the theoretical values $\mu_T = 2.8059$ and $\sigma_T^2 = 6.2873$ and corresponding values of the GOF being $X_E^2 = 1105.2885$ and $X_T^2 = 1075.6976$. It is again note worthy to observe that in the range $0 \leq Y \leq 28$, the empirical means and variance are 2.8059 and 6.2872, these are as expected, very close to the theoretical values.

While the mean and variance of the observed data are 2.80587 and 6.41062 respectively, we see from tables 2 & 4, that the Poisson model does not fit because its variance does not adjust for over-dispersion in the data. For both the NB and GP, the empirical estimated means and variances are not equal to the theoretical values displayed in (4.4)

Thus, in these cases, the variances equal theoretically, 6.4958 and 6.6125 respectively. Both models have theoretical variances not too far from the observed data variance of 6.4106. However, for the GP for instance, the empirical mean and variance are respectively 2.7767 and 6.2475. The reason being that for the GP for instance:

$$\sum_{y=0}^{15} f(y|\hat{\mu},\hat{\tau}) = 0.99836 < 1.0000 \quad \text{but} \quad \sum_{y=0}^{29} f(y|\hat{\mu},\hat{\tau}) = 1.0000.$$

In both cases, (i) $\sum_{y=0}^{15} y \times \hat{p}(y) = 2.77667$ and $\sum_{y=0}^{15} y^2 \times \hat{p}(y) = 13.9574$ giving us an empirical variance of 6.2475. (ii) In the second case however, $\sum_{y=0}^{29} y \times \hat{p}(y) = 2.805796$ and $\sum_{y=0}^{29} y^2 \times \hat{p}(y) = 14.482996$ giving us an empirical variance of 6.6096.

Table 4. Expected frequencies, empirical and theoretical moments, together with the GOF statistics under the six models

		Six Distributions						R-Truncated	
Y	n_y	P	NB	GP	HP	MLFD	COMP	NB*	COMP*
0	187	63.840	176.120	169.779	201.734	182.096	188.151	176.355	188.422
1	185	179.125	213.463	218.980	192.633	207.143	201.172	213.386	201.131
2	200	251.301	189.990	195.089	171.416	185.573	180.970	189.803	180.799
3	164	235.040	148.705	150.072	142.811	148.036	147.150	148.550	146.980
4	107	164.873	108.411	107.262	111.849	109.917	111.373	108.328	111.261
5	68	92.522	75.545	73.566	82.646	77.611	79.735	75.525	79.690
6	49	43.268	51.021	49.223	57.799	52.755	54.549	51.042	54.555
7	39	17.343	33.676	32.428	38.369	34.789	35.912	33.716	35.947
8	21	6.083	21.840	21.151	24.239	22.374	22.868	21.885	22.915
9	12	1.896	13.969	13.707	14.607	14.087	14.141	14.011	14.187
10	11	0.532	8.834	8.845	8.414	8.708	8.518	8.870	8.557
11	2	0.136	5.535	5.692	4.642	5.296	5.010	5.564	5.041
12	5	0.032	3.441	3.657	2.458	3.174	2.884	3.463	2.906
13	2	0.007	2.125	2.347	1.250	1.878	1.627	2.141	1.642
14	3	0.001	1.305	1.506	0.612	1.098	0.901	1.316	0.911
15	1	0.000	0.797	0.966	0.289	0.635	0.491	2.044	1.055
Total	1056	1056.00	1054.78	1054.27	1055.77	1055.17	1055.452	1056.00	1056.00
$Y \leq a$		14	26	29	20	24	28	15	15
$\bar{y}$		2.8059	2.7856	2.7767	2.8022	2.7923	2.7901	2.8044	2.8058
s^2		2.8059	6.2500	6.2475	6.2272	6.2198	6.1848	6.4298	6.2734
μ_T	**2.8059**	2.8059	2.8059	2.8059	2.8059	2.8059	2.8059	2.8073	2.8069
σ_T^2	**6.4106**	2.8059	6.4956	6.6126	6.2681	6.3798	6.2873	6.5133	6.3033
X_e^2		2410.3767	1082.1838	1082.6958	1086.0769	1087.3979	1093.5626	1051.8532	1078.0763
X_t^2		2410.3767	1041.1925	1022.7800	1078.9878	1060.2793	1075.6976	1038.3637	1072.9549
d.f.		1054	1053	1053	1053	1052	1053	1053	1053

From table 4, the observed mean and variance are presented in boldface, $\bar{y}$ and s^2 are the empirical mean and variance for the data under various distributions in the range $0 \leq Y \leq 15$. On the other hand, μ_T and σ_T^2 are the theoretical empirical mean and variance of the distributions. All the distributions in the first panel, excluding the Poisson have means that are very close to the observed mean of 2.8059 and they all have adjusted variances that are very close to the observed variance of 6.4106. The row $\mathbf{Y} \leq \mathbf{a}$ indicated the value of $Y = a$ required for the estimated probabilities to sum to 1. Clearly, none of them apart from the Poisson, equals 15, the largest values of Y in the example data. The NB model fits very well, and it is quite obvious based on this data that all these models behave equally well but then, the NB is much easier to fit than the other models. However, for under-dispersed data as we will see in Example III, the NB does not often converge well. We consider these distributions in Example II.

6.2 Example II

The data in this example was analyzed in [22] and relate to vaccine adverse event count, where 4020 observed systemic adverse events for four injections administered to each of the 1005 study participants tabulated by the number of such adverse events occurrences. The data is presented in table 5.

Table 5. The actual frequencies correspond to 4020 observed systemic Adverse events for four injection for each of th e1005 stud

		Six Probability Models					
Y	n_y	P	NB	GP	HP	MLFD	COMP
0	1437	890.76	1409.08	1389.97	1442.45	1444.92	1420.98
1	1010	1342.34	1068.65	1098.74	1009.33	1008.58	1039.18
2	660	1011.43	670.65	675.75	661.52	660.18	670.48
3	428	508.06	391.63	384.92	407.74	406.83	402.023
4	236	191.41	220.16	213.07	237.18	236.84	228.85
5	122	57.69	120.88	116.68	130.62	130.66	125.12
6	62	14.49	65.32	63.69	68.30	68.50	66.19
7	34	3.12	34.89	34.79	34.00	34.21	34.05
8	14	0.59	18.47	19.04	16.14	16.31	17.10
9	8	0.10	9.71	10.45	7.33	7.44	8.41
10	4	0.01	5.08	5.75	3.19	3.26	4.06
11	4	0.00	2.64	3.18	1.33	1.37	1.92
12	1	0.00	1.37	1.76	0.53	0.55	0.90
Total	4020	4020	4018.54	4017.78	4019.68	4019.66	4019.25
$Y \leq a$		10	21	23	16	17	24
$\bar{y}$		1.5070	1.5019	1.4991	1.5058	1.5055	1.5044
s^2		1.5070	2.9373	2.9438	2.8804	2.8892	2.8918
μ_T	**1.5070**	1.5070	1.5070	1.5070	1.5069	1.5067	1.5070
σ_T^2	**2.9034**	1.5070	2.9944	3.0343	2.8920	2.9014	2.9199
X_e^2		7743.2231	3972.6936	3963.9860	4051.1400	4038.7458	4035.1448
X_t^2		7743.2487	3896.8990	3845.6237	4034.8471	4021.7738	3996.3349
d.f.		4018	4017	4017	4017	4016	4017

6.3 Results

Our results here are very similar to those in table 4, with the Generalized Poisson doing very well here. Again, in terms of estimated probabilities, the HP and MLFD models will give estimated sum of probabilities for the least values of Y (in this case, 16 & 17) as compared to NB and GP with 21 and 23 respectively. Again, the variances are adjusted for each of these distributions to account for over-dispersion and all very close to the observed mean and variance of the example data.

It is evident from results in tables 4 and 5 that apart from the Poisson model (and to some extent the NB), all models considered here do not have their estimated cumulative probabilities summing to 1 and consequently, the sum of expected values do not necessarily add to the sample size n. It is particularly more serious for the NB and GP models. The HP, Com-Poisson (COMP) and the MLFD give values much closer to the theoretical assumptions. For instance, in table 4, the true sample size is 1056, but we see that the GP gives 1054.27 while the HP and COMP give respectively 1055.73 and 1055.452 (much closer to the true value). Most authors often add this difference to the last category expected value, such that for instance, in table 4, $\hat{m}_{15}$ under the GP model now becomes $(1056 - 1054.72) + 0.966 = 2.696$ instead of the estimated 0.966 under this model. This way, the sum of the estimated frequencies now sum to $n = 1056$. We can overcome this subjective approach by instead fit a model with the log-likelihood of a single observation i of the form:

$$LL = (1-\delta)\log[P(Y_i = y_i)] + \delta \log[P(Y_i \geq C)]; \quad \text{where} \quad \delta = \begin{cases} 0 & \text{for } y_i < C \\ 1 & \text{for } y_i \geq C \end{cases}. \tag{6.2}$$

and C is the last category of the data. In our example I, this would be C=15. To accomplish this, first we compute:

$$W = \sum_{j=0}^{C-1} f(y_j) \tag{6.3}$$

for each of the distributions NB, GP, HP, MLFD and COMP. The expression in (6.2) for the Com-Poisson (COMP) model for instance would be:

$$\text{LL} = (1 - \delta)(LL) + \delta \, \log(1 - W) \tag{6.4}$$

where LL represents the log-likelihood LL3 in (5.1c) for the Com-Poisson model. Thus, for the Com-Poisson model, the above in (6.4) becomes:

$$LL = (1 - \delta)[(y \log(\lambda) - \nu \log(y!) - \log\{Z(\lambda, \nu)\}] + \delta \log(1 - W); \tag{6.5}$$

Here,

$$Z(\mu, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}; \quad \text{and} \quad W = 1 - \frac{1}{Z(\lambda, \nu)} . \sum_{k=0}^{C-1} \frac{\lambda^k}{(k!)^\nu}.$$

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

$$\hat{p}_i = \begin{cases} \exp\{i \log(\hat{\lambda}) - \hat{\nu} \log(i!) - \log[Z(\hat{\lambda}, \hat{\nu})]\} & \text{for } 0 \leq i \leq (C - 1) \\ \exp\{\log(1 - \hat{W})\} \end{cases} \tag{6.6}$$

Here again,

$$Z(\hat{\lambda}, \hat{\nu}) = \sum_{j=0}^{\infty} \frac{\hat{\lambda}^j}{(j!)^{\hat{\nu}}}; \quad \text{and} \quad \hat{W} = 1 - \frac{1}{Z(\hat{\lambda}, \hat{\nu})} . \sum_{k=0}^{C-1} \frac{\hat{\lambda}^k}{(k!)^{\hat{\nu}}}$$

The results of implementing this approach (sometimes referred to as *right truncated*) is presented in the last two columns of table 4 for just the NB and COMP distributions only. Clearly now, the estimated probabilities sum to 1.00 in the range $0 \leq Y \leq 15$ and consequently, the expected values also sum to n. However, this does not have much effect. A case where this approach makes a considerable impact is presented in [[16]].

6.4 Example III: insurance data

The third set of data is from [23]. Because the data is under-dispersed, both the NB and GP models do not lend themselves adaptable to fitting under-dispersed count data. In fact for this data set, the two models when applied, give estimates corresponding to the Poisson, with dispersion parameters‘ estimates being in the order of $10^{-6} = 0$. The results of applying the other models considered in this paper to this data set is presented in table 6, where the Y's are the number of insurance claims and the counts are the frequencies of each claim for a given number of claims.

6.5 Results

While both the NB and GP models are not suitable for this data set, the Com-Poisson, HP and MLFD models cope well with this under-dispersed data set. Thus for under-dispersed data, either the HP, MLFD or Com-Poisson distributions will be suitable. The choice of models of course depends on the optimization technique employed as well as the choice of initial parameter estimates. The Com-Poisson is much easier to fit than either the HP or MLFD models.

Table 6. Parameter estimates and expected frequencies under the models applied to this data set

Y	n_y	P	COMP	HP	MLFD
0	121	126.907	123.517	121.692	121.091
1	85	73.796	79.028	82.227	84.247
2	19	21.456	20.811	19.835	18.045
3	1	4.159	3.260	2.912	3.068
4	0	0.605	0.353	0.307	0.471
5	0	0.070	0.029	0.025	0.068
6	1	0.007	0.002	0.002	0.009
Total	227	226.999	227.000	227.000	226.999
		$\hat{\lambda} = 0.5815$	$\hat{\lambda} = 0.6398$	$\hat{\lambda} = 0.3752$	$\hat{\lambda} = 0.1166$
		(0.0506)	(-)	(0.1178)	
			$\hat{\nu} = 1.2807$	$\hat{\beta} = 0.5552$	$\hat{\beta} = 0.0011$
					$\hat{\alpha} = 0.0055$
			(0.342)	(0.2262)	
-2LL		450.2	449.4	448.4	445.6
AIC		452.2	453.4	452.4	451.6
$\bar{y}$		0.5815	0.5815	0.5815	0.5807
s^2		0.5161	0.5344	0.5138	0.5157
μ	0.5815	0.5815	0.5815	0.5815	0.5808
σ^2	0.5719	0.5815	0.5344	0.5138	0.5160
X_e^2		222.3006	241.8664	251.5459	250.6165
X_T^2		222.2576	241.8578	251.5368	250.4801
d.f.		226	225	225	224

7 GLM Applications

In this section, we would employ the NB, GP, COMP, HPP and MLFD models to data having covariates. For data having covariates $(x_1, x_2, \ldots, x_p)'$, the mean and dispersion parameter will be modeled as:

$$\lambda_i = \exp(\mathbf{x}'\mathbf{b})$$
$$\mathbf{disp}. = \exp(\mathbf{x}'\mathbf{a})$$

where $(b_0, b_1, b_2, \ldots, b_p)'$ and $(a_0, a_1, a_2, \ldots, a_p)'$ are parameter estimates to be estimated. The above assumes that we have a variable dispersion parameter ($\boldsymbol{\beta}$ for instance for the HP model) that varies and is dependent on the covariates, otherwise, we could also model the dispersion parameter as a constant by assuming that $\beta_i = \exp(a_0)$, where a_0 is a constant. We explore these two possibilities, by applying these models to the German national health registry data described in the following section.

7.1 The German health data

We will employed the very well analyzed German national health registry data [24] set which comprises of 3874 respondents. The response variable is the number of visits made by a patient to a physician (y) during the year, the age of the individual (age), Outwork (1=if patient is not working; 0=if patient is working), gender (1=female, 0 if male) and marital status (1 if married; 0 if not married). Thus the covariates here are **outwork, age, gender, married**. The response variable Y is the count variable with minimum 0 and maximum 121, mean 3.162881 and variance 39.387611. Thus the index of dispersion(ID) here is 12.4531, indicating a very strong over-dispersion. In addition, 41.58% of the data have zero counts. The Com-Poisson employed here is the one based on [2] formulation and all models will be fitted using PROC NLMIXED in SAS.

7.2 Results

The results of applying these models to the case where we have a constant dispersion parameter (though not necessarily constant for each of the 3874 observations but independent of the covariates). These results are presented in table 7.

Table 7. Application of the models to the german data

	Models				
Parameters	NB	GP	COMP	HP	MLFD
intercept	-0.1068	-0.1339	-0.5610	23.0248	-0.5666
	(0.1132)	(0.1214)	(0.0220)	(0.0688)	(0.0410)
outwork	0.2847	0.2961	0.0583	0.0577	0.0577
	(0.0594)	(0.0662)	(0.0106)	(0.0104)	(0.0105)
age	0.0239	0.0247	0.0053	0.0051	0.0051
	(0.0024)	(0.0027)	(0.0004)	(0.0004)	(0.0004)
gender	0.3211	0.3502	0.0671	0.0663	0.0662
	(0.0564)	(0.0613)	(0.0105)	(0.0104)	(0.0104)
married	-0.1650	-0.1951	-0.0202	-0.0195	-0.0188
	(0.0649)	(0.0746)	(0.0099)	(0.0097)	(0.0097)
dispersion	2.2612	0.6896	0.0067	$> 10^8$	7.3907
	(0.0701)	(0.0186)	(0.0013)	(5047*)	
α					0.0000
					(0.0116)
X^2	5662.2748	4778.8622	11417.4911	11201.4174	11203.5191
-2LL	16625	16668	17420	17365	17365
AIC	16637	16680	17432	17377	17379

For constant dispersion parameters, the Com-Poisson, HP and MLFD did not perform as well as the NB and the GP models having co-variates. The GP performs much better in this case. In fact for this data, the MLFD has $\hat{\alpha}$ almost zero, which reduces it to the geometric distribution. The Wald test statistics for COMP, HP and MLFD models are unnecessarily high compared to their NB and GP counterparts. The -2LL and AIC values also indicate that both NB and GP are better alternatives to the HP-type models. A further complication relates to convergence issues with COMP, HP and MLFD models. The choice of initial values are very crucial, just as well as the technique for optimization (the conjugate gradient is most adaptable while the Newton-Rapson or Nelder-Mead or quasi-Newton optimization techniques usually create convergence problem. The method of integration in all cases being adaptive quadrature. We consider in the next section, dispersion parameters for each of these models that are functions of all or some of the covariates. Here however, we employ all the covariates.

7.3 Models with varying dispersion parameters

In this section, we represent the dispersion parameters as a function of the covariates. For instance, for the NB and HP models, these are equivalent to:

$$\mathbf{k} = a_0 + a_1 \,\text{outwork} + a_2 \,\text{age} + a_3 \,\text{female} + a_4 \,\text{married}$$
$$\boldsymbol{\beta} = a_0 + a_1 \,\text{outwork} + a_2 \,\text{age} + a_3 \,\text{female} + a_4 \,\text{married}$$

The $a_i, i = 0, 1, 2, 3, 4$ are not the same for both models. Thus for each of the models, the dispersion parameters are modeled as functions of the covariates. The results of this analysis is presented in table 8.

Table 8. German data-results for variable dispersion parameters

	Models				
Parameters	NB	GP	COMP	HP	MLFD
Intercept	-0.0966	-0.1009	-0.56667	10.1143	-0.6414
	(0.1207)	(0.1335)	(0.0217)	(0.1114)	(0.0043)
outwork	0.2497	0.2466	0.0577	0.1744	0.0453
	(0.0591)	(0.0650)	(0.0105)	(0.0642)	(0.0097)
age	0.0236	0.0237	0.0051	0.1907	0.0064
	(0.0024)	(0.0027)	(0.0004)	(0.0023)	(0.0002)
gender	0.3121	0.3150	0.0662	1.0456	0.0720
	(0.0581)	(0.0640)	(0.0104)	(0.1397)	(0.0107)
married	-0.1357	-0.1387	-0.0188	0.2050	-0.0064
	(0.0661)	(0.0759)	(0.0097)	(0.1278)	(0.0092)
dispersion					
intercept	1.6802	0.6358	-8.4057	10.6809	0.0538
	(0.1400)	(0.1214)	(1089.11)*	(0.1001)	(0.3550)*
outwork	-0.1008	-0.1545	-0.6620	0.1167	-0.1798
	(0.0711)	(0.0611)	(449.01)*	-	(-)*
age	-0.0125	-0.0157	-0.2722	0.1856	0.0795
	(0.0029)	(0.0025)	(38.7020)*	(0.0023)	(0.3637)*
gender	-0.3432	-0.3435	-0.7430	0.9794	-0.6399
	(0.0691)	(0.0597)	(315.93)*	(0.1399)	(16.6104)*
married	-0.0871	-0.0396	0.4269	0.2239	-0.0000
	(0.0750)	(0.0597)	(337.61)*	(0.1268)*	(13.9625)*
α					0.0002
					(0.0004)
X^2	5630.1392	4253.2110	11203.5841	11203.5247	12083.6182
d.f.	3864	3864	3864	3864	3863
-2LL	16555	16545	17365	17365	17365
AIC	16575	16565	17385	17385	17387

Results from this analysis indicate that for the COMP, HP and MLFD models, the concept of variable dispersion parameter can be very difficult to implement and all the parameter estimates seem to be unimportant in the model. HP model shows that, perhaps only age can probably be included as a covariate in the dispersion model. It should be obvious that we do not need all the covariates in the dispersion parameter formulation. Usually only a subset of the covariates may prove useful.

On the other hand, both the NB and GP benefited from the use of the covariates. For the GP for instance, the change in -2LL is (-16545+16668)=123 on 5 d.f. which is highly significant, indicating that the GP with variable dispersion provides a better fit than the usual GP model.

8 Conclusions

Based on the results in this study, we have the following observations:

(a) For most practical purposes, both the negative binomial and generalized Poisson models behave very well except for under-dispersed data.

(b) For under-dispersed data, the Com-Poisson, HP and MLFD models provide better fit.

(c) The HP and MLFD converges much faster than the Com-Poisson, especially in the infinite summation of the normalizing constants

(d) The implementation of a variable dispersion parameter incorporating some or all of the covariates creates convergence problems in the HP, COMP and MLFD models, especially with regards to providing initial parameter estimates for the optimization algorithm. Both the NB and GP models handle this fairly well and could well be employed instead of the constant parameter model.

(e) All the models have estimated empirical probabilities that do not necessarily sum to 1.00. Consequently, models that employ right truncation would provide better alternatives to modeling with HP, COMP and MLFD models as well as the NB and the GP models, the latter two providing cumulative estimated probabilities that are even much < 1.00.

Acknowledgement

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

Competing Interests

Author has declared that no competing interests exist.

References

[1] Geedipally SR, Lord D, Dhavala SS. The Negative binomial-Lindley generalized linear model: Characteristics and application using crash data. Accident Analysis & Prevention. 2012;45(2): 258-265.

[2] Lord D, Geedipally SR. The Negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. Accident Analysis and Prevention. 2011;43(5):1738-1742.

[3] Zamani H, Ismail N. Negative binomial-Lindley distribution and its application. J. Math. Stat. 2010;6:4-9.

[4] Aryuyuen S, Bodhisuwan W. Negative binomial-generalized exponential (NB-GE) distribution. Applied Mathematical Sciences. 2013;7(22):1093-1105.

[5] Consul P, Famoye F. Generalized poisson regression-model. Communications in Statistics-Theory and Methods. 1992;21(1):89-109.

[6] Efron B. Double Exponential-families and their use in generalized linear regression. Journal of the American Statist. Assoc. 1986;81:709-721.

[7] Castilo JD, Pérez-Casany. Over-dispersed and under-dispersed poisson generalizations. Journal of Statist. Planning and Inference. 2005;134:486-500.

[8] Bardwell GE, Crow EL. A two parameter family of hyper-poisson distributions. Journal of the American Statist. Assoc. 1964;59:133-141.
DOI: 10.1080/01621459.1964.10480706

[9] Crow EL, Bardwell GE. Estimation of the parameters of the Hyper-Poisson distributions. In: Patil, G.P. (ed) classical and contagious discrete distributions. Pergamon, Oxford. 1965;27-140.

[10] Kumar CS, Nair BU. A three parameter hyper-Poisson distribution and some of its properties. Statistica. 2014;183-198.

[11] Chakraborty S, Ong SH. Mittag-Leffler function distribution-A new generalization of hyper-Poisson distribution; 2014.

[12] Shmueli G, Minka T, Borle J, Boatwright P. A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. J. R. Stat. Soc., Series C. 2005;54:127-142.

[13] Kadane J, Shmueli G, Minka G, Borle T, Boatwright P. Conjugate analysis of the Conway Maxwell Poisson distribution. Bayesian Analysis. 2006;1:363-374.

[14] Sellers KF, Borle S, Shmueli G. The COM-Poisson model for count data: A survey of methods and applications. Applied Stochastic Models in Business and Industry. 2011;104-116.
DOI: 10.1002/asmb.918

[15] Lawal HB. Fitting the Conway-poisson model to count data with SAS. ICASTOR Journal of Mathematical Sciences. 2011;5:229-238.

[16] Chakraborty S, Imoto T. Extended conway-maxwell-poisson distribution and its properties and applications. Journal of Statistical Distributions and Applications. 2016;3(5):1-199.
DOI: 10.1186/s40488-016-0044-1

[17] Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics. 1995;4:12-35.

[18] Powell MJD. Restart procedures for the conjugate gradient method. Mathematical Programming. 1977;12:241-254.

[19] Beale EML. A derivation of conjugate gradients. In: F. A. Lootsma, ed., Numerical Methods for Nonlinear Optimization. London: Academic Press; 1972.

[20] Lundberg O. On random processes and their application to sickness and accident statistics. Uppsala: Almquist and Wiksells; 1940.

[21] Phang YN, Sim SZ, Ong SH. Statistical analysis for the inverse trinomial distribution. Communications in Statistics-Simulation and Computation. 2013;42(9):2073-2085.

[22] Phang YN, LohEF. Statistical analysis for overdispersed medical count data. International Scholarly and Scientific Research. 2014;8(2):292-294.

[23] Famoye F, Wulu JJ, Singh KP. On the generalized Poisson regression model with an application to accident data. Journal of Data Science. 2004;2:287-295.

[24] Hilbe JM. German health data for 1984. Negative Binomial Regression. Cambridge University Press; 2007.

Appendix

The probability generating $E[s^x]$ of the MLFD [[11]] can be expressed as:

$$P(s) \quad = \quad E_{\alpha,\beta}(\lambda s)/E_{\alpha,\beta}(\lambda) \quad = \quad \frac{\sum_{k=0}^{\infty}(\lambda s)^k/\Gamma(\alpha k+\beta)}{\sum_{k=0}^{\infty}(\lambda)^k/\Gamma(\alpha k+\beta)} \tag{.1}$$

The numerator can be written for $k = 0, 1, 2, 3, 4, \ldots$ as:

$$\frac{1}{\Gamma(\beta)} + \frac{\lambda s}{\Gamma(\alpha+\beta)} + \frac{\lambda^2 s^2}{\Gamma(2\alpha+\beta)} + \frac{\lambda^3 s^3}{\Gamma(3\alpha+\beta)} + \frac{\lambda^4 s^4}{\Gamma(4\alpha+\beta)} + \frac{\lambda^5 s^5}{\Gamma(5\alpha+\beta)} + \ldots$$

Hence,

$$P^{'}(s) = \frac{\lambda}{\Gamma(\alpha+\beta)} + \frac{2\lambda^2 s}{\Gamma(2\alpha+\beta)} + \frac{3\lambda^3 s^2}{\Gamma(3\alpha+\beta)} + \frac{4\lambda^4 s^3}{\Gamma(4\alpha+\beta)} + \frac{5\lambda^5 s^4}{\Gamma(5\alpha+\beta)} + \ldots, \quad \text{for} \quad k = 1, 2, 3, \ldots \tag{.2}$$

Thus,

$$P^{'}(1) = \frac{\lambda}{\Gamma(\alpha+\beta)} + \frac{2\lambda^2}{\Gamma(2\alpha+\beta)} + \frac{3\lambda^3}{\Gamma(3\alpha+\beta)} + \frac{4\lambda^4}{\Gamma(4\alpha+\beta)} + \frac{5\lambda^5}{\Gamma(5\alpha+\beta)} + \cdots = \sum_{j=2}^{\infty}\frac{j\lambda^j}{\Gamma(j\alpha+\beta)}, \quad \text{for} \quad j = 1, 2, 3, \ldots$$

Hence, $\mu = \mathrm{E}(X)$ is given by

$$\mu = \sum_{j=1}^{\infty}\frac{j\lambda^j}{\Gamma(j\alpha+\beta)}/\sum_{k=0}^{\infty}\frac{\lambda^k}{\Gamma(k\alpha+\beta)} \quad = \quad \mu_1^{'} \tag{.3}$$

Similarly, from (.2), we have:

$$P^{''}(s) = \frac{2\lambda^2}{\Gamma(2\alpha+\beta)} + \frac{6\lambda^3 s}{\Gamma(3\alpha+\beta)} + \frac{12\lambda^4 s^2}{\Gamma(4\alpha+\beta)} + \frac{20\lambda^5 s^3}{\Gamma(5\alpha+\beta)} + \ldots, \quad \text{for} \quad k = 2, 3, \ldots$$

Thus $P^{''}(1)$ is given by:

$$P^{''}(s=1) = \frac{2\lambda^2}{\Gamma(2\alpha+\beta)} + \frac{6\lambda^3}{\Gamma(3\alpha+\beta)} + \frac{12\lambda^4}{\Gamma(4\alpha+\beta)} + \frac{20\lambda^5}{\Gamma(5\alpha+\beta)} + \ldots, = \sum_{j=2}^{\infty}\frac{j(j-1)\lambda^j}{\Gamma(j\alpha+\beta)}.$$

Therefore,

$$\mathrm{E}[X(X-1)] = \sum_{j=2}^{\infty}\frac{j(j-1)\lambda^j}{\Gamma(j\alpha+\beta)}/\sum_{k=0}^{\infty}\frac{\lambda^k}{\Gamma(k\alpha+\beta)} \quad = \quad \mu_2^{'} \tag{.4}$$

Consequently,

$$\mathrm{var}(X) = \mu_2^{'} + \mu_1^{'} - [\mu_1^{'}]^2$$

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/18406