Modelling Negative Binomial as a substitute model to Poisson for raters agreement on ordinal scales with sparse data

A. O. Adejumo, C. Heumann and H. Toutenburg
Department of Statistics
Ludwig-Maximilians-Universität, München
Ludwigstr. 33, D-80539 Munich,
Germany

26th May 2004

Abstract

The Poisson distribution has been widely used for modelling rater agreement using loglinear models. Mostly in all life or social science researches, subjects are being classified into categories by rater, interviewers or observers and most of these tables indicate that the cell counts are mixtures of either too big values and two small values or zeroes which are sparse data. We refer to sparse as a situation when a large number of cell frequencies are very small. For these kinds of tables, there are tendencies for overdispersion in which the variance of the outcome or response exceeds the nominal variance, that is, when the response is greater than it should be under the given model or the true variance is bigger than the mean. In these types of situations assuming Poisson models means we are imposing the mean-variance equality restriction on the estimation. This implies that we will effectively be requiring the variance to be less than it really is, and also, as a result, we will underestimate the true variability in the data. Lastly, this will lead us to underestimating the standard errors, and so to overestimating the degree of precision in the coefficients. The Negative Binomial, which has a variance function, would be better for modelling rater agreement with sparse data in the table in order to allow the spread of the observations or counts. We observed that assuming Negative Binomial as the underline sampling plan is better for modelling rater agreement when there are sparse data in a limited number of example.

keywords: Negative Binomial, sparse data, overdispersion, chance and beyond chance agreements, rater.

1 Introduction

In most analyzes on categorical data, like any cross-classified table between two raters, the researchers frequently prefer assuming Poisson distribution as the underlying sampling plan rather than using binomial, multinomial, or hypergeometric or any other distribution because of its added advantage of having the same mean and variance. The two major assumptions on Poisson distribution for count data are assumptions of independence and of constant rates (Breslow, 1984). In this paper we propose Negative Binomial as a better underlying assumption for modelling the rater agreement most especially when the resulted cross-classified table of ratings is sparse. A square table can be used to display joint ratings of two rater or observers. If the reliability test between the two raters has been satisfied, it is highly expected that some of the cells counts in the diagonal will be very big while those off the diagonal will be very small or zero which leads to sparse data. And in a situation like this, there is a very high tendency of observing overdispersion which is one of the main causes of lack of fit. Overdispersion refers to the situation that the variance of the outcome or response exceeds the nominal variance under the given model. McCullagh and Nelder (1989) said that overdispersion is the rule rather than the exception. Overdispersion has two major adverse effects. Firstly, the summary statistics have a larger variance than anticipated under the simple model, and secondly, there is a possible loss of efficiency in using statistics appropriate for the assumed distribution. Deviance and Pearson Chi-square divided by the degrees of freedom are used to detect overdispersion or underdispersion in the Poisson regression. Values greater than 1 indicate overdispersion, that is, the true variance is bigger than the mean, values smaller than 1 indicate underdispersion, the true variance is smaller than the mean. Evidence of underdispersion or overdispersion indicates are an inadequate fit of the Poisson model. We can test for overdispersion with a likelihood ratio test based on Poisson and this Negative Binomial distribution. This tests equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean. See also Cameron and Trivedi (1990) for more on overdispersion.

Before the introduction of modelling in rater agreement many authors have described various testing procedures and various modifications of kappa liked statistic for raters agreement measures Barnhart and Williamson (2002); Shoukri (2004) and so on. Many categorical scales are quite subjective and reliability assessment which has to be conducted at the initial stage depends on evaluation of agreement between the raters, interviewers or observers involved in the measurements or tests, so as to assess the level of variability in the measurement procedure to be used in the final tests. Due to the fact that most of these authors have described raters agreement using generalization of Cohen's kappa, some authors like Tanner and Young (1985a); Maclure and Willett (1987) showed difficulties with summarizing agreement between two raters by a single measure. These difficulties include: loss of information from summarizing the table by a single number; sensitivity of value to the form of the marginal distributions; and subsequence dangers in comparing values of kappa between two tables. These authors then proposed modelling the structure of agreement between raters. Modelling that involves loglinear and latent class models for raters agreement measures. In this research work, modelling approach to rater agreement will be considered with attention on ordinal scales data.

To these effects, in place of Poisson, we propose a natural variance function distribution called Negative Binomial as an underlying sampling plan for modelling rater agreement when there are sparse data. Negative Binomial model was proposed because it is a parametric model and more dispersed than Poisson, and it allows the spread of the counts in the table (Cameron and Trivedi 1998).

We used generalized linear model (GLM) approach to fit both Poisson and Negative Binomial models, and the algorithms for these models can be executed with any of Splus or R-program. In sections 2 and 3 we present models as well as their properties and estimation of parameters as well as modelling and in section 4 we have various empirical examples on ordinal categorial scale data followed by the summary of results and conclusion in section 5.

2 Models and their properties

The Poisson model can be used to model the number of occurrences of an event of interest or the rate of occurrence of an event of interest, as a function of one or more independent variables or factors. The Poisson distribution can be expressed in terms of a linear regression model with dependent variable Y, cell frequency or number of occurrences of an event, and independent variables or factors X_i, $i = 1, \ldots, p$, Cameron and Trivedi (1998). So Poisson distribution for a random variable y given X_i, $i = 1, \ldots, p$ is given as

$$P(Y = y \mid x_1, \ldots, x_p) = \frac{\lambda^y e^{-\lambda}}{y!} \quad for\ y = 0, 1, 2, \ldots, \tag{2.1}$$

The mean and the variance of Poisson distribution are equal, that is, $E(Y) = V(Y) = \lambda$. The Poisson model for y_i assumes that the y_i are independent.

Suppose that $y \sim Poisson(\lambda)$, but λ itself is a random variable with a gamma distribution. That is suppose

$$y/\lambda \sim Poisson(\lambda),$$

$$\lambda \sim Gamma(\alpha, \beta),$$

where $Gamma(\alpha, \beta)$ is the gamma distribution with mean $\alpha\beta$ and variance $\alpha\beta^2$, whose density is

$$P(\lambda) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} \exp(-(\lambda/\beta)) & \text{for} \quad \lambda, \alpha, \beta > 0, \\ \\ 0 & otherwise. \end{cases} \tag{2.2}$$

Then it is easy to show that the unconditional distribution of y is the Negative Binomial,

$$f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!} \left(\frac{\beta}{1+\beta}\right)^y \left(\frac{1}{1+\beta}\right)^\alpha, \quad for\ y = 0, 1, 2, \ldots, \tag{2.3}$$

This distribution has mean

$$E(Y) = E[E(Y \mid \lambda)] = E(\lambda) = \alpha\beta$$

and variance

$$\begin{aligned} Var(Y) &= E[var(Y \mid \lambda)] + var[E(Y \mid \lambda)] \\ &= var(\lambda) + E(\lambda) \\ &= \alpha\beta + \alpha\beta^2 \end{aligned} \tag{2.4}$$

For building a regression model, it is natural to express the Negative Binomial distribution in terms of the parameters $\mu = \alpha\beta$ and $\alpha = \frac{1}{k}$, so that

$$\begin{aligned} E(Y) &= \mu \ \ and \\ Var(Y) &= \mu + k\mu^2 \end{aligned} \tag{2.5}$$

Notice that the variance function is quadratic. Therefore the distribution of y is then

$$f(y) = \frac{\Gamma(k^{-1} + y)}{\Gamma(k^{-1}) y!} \left(\frac{k\mu}{1+k\mu}\right)^y \left(\frac{1}{1+k\mu}\right)^{\frac{1}{k}}, \quad for\ y = 0, 1, 2, \ldots, \tag{2.6}$$

which approaches $Poisson(\mu)$ as $k \to 0$ Adejumo (2002). k is not a dispersion parameter, large value of k correspond to greater amounts of overdispersion . Anscombe (1949) had worked on estimation of parameter k for single samples and for several samples. See also Santos-Silva (1997), Cameron and Trivedi (1998) and others for more on Negative Binomial models.

3 Modelling and estimation of parameters

Loglinear models have been used to model agreement in terms of components, such as chance agreement and beyond chance agreement by displaying patterns of agreement among raters, Tanner and Young (1985). They expressed statistical independence of the ratings or chance agreement and agreement beyond chance respectively when there are rr raters through the following loglinear models,

$$\log(v_{ij\ldots l}) = u + u_i^{R_1} + u_j^{R_2} + \ldots + u_l^{R_{rr}}, \quad i, j, \ldots l = 1, \ldots m \tag{3.1}$$

and

$$\log(v_{ij\ldots l}) = u + u_i^{R_1} + u_j^{R_2} + \ldots + u_l^{R_{rr}} + \delta_{ij\ldots l}, \quad i, j, \ldots l = 1, \ldots m \tag{3.2}$$

where $v_{ij\ldots l}$ is the expected cell counts in the $ij\ldots lth$ cell of the joint rr-dimensional cross classification of the ratings, u is the overall effect, u^{R_i} is the effect due to categorization by the rth raters in the cth category $(r = 1, \ldots, rr; c = 1, \ldots, m)$, $\sum_{i=1}^{m} u_i^{R_1} = \ldots = \sum_{l=1}^{m} u_l^{R_{rr}} = 0$ and parameter $\delta_{ij\ldots l}$ represents agreement beyond chance for the $ij\ldots lth$ cell.

Based on Tanner and Young (1985), for a two dimensional cross classification of ratings, the independence model with agreement beyond chance will be

$$\log(v_{ij}) = u + u_i^{R_1} + u_j^{R_2} + \delta_{ij}, \quad i, j = 1, \ldots m, \tag{3.3}$$

where δ_{ij} can be homogenous or nonhomogeneous defined respectively as

$$\delta_{ij} = \begin{cases} \delta & \text{If} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{3.4}$$

or

$$\delta_{ij} = \begin{cases} \delta_1 & \text{If} \quad i = j = 1, \\ \delta_2 & \text{If} \quad i = j = 2, \\ \vdots & \qquad \vdots \\ \delta_m & \text{If} \quad i = j = m, \\ 0 & \text{if} \quad i \neq j. \end{cases} \tag{3.5}$$

This illustrates the fact that model (3.3) is sensitive only to discrepancies that are present on the main diagonal.

We also consider the uniform association model proposed by Goodman (1979) under the assumption of Poisson model also, given as

$$\log(v_{ij}) = u + u_i^{R_1} + u_j^{R_2} + \beta u_i v_j, \quad i, j = 1, \ldots m, \tag{3.6}$$

where $u_1 < u_2 < \ldots < u_m$ or $v_1 < v_2 < \ldots < v_m$ are fixed scores assigned to the response categories so as to reflect the category ordering.

In addition to these, Agresti (1988) also proposed a model of agreement plus linear-by-linear association, which is the combination of the model of Tanner and Young (1985) and the uniform association model of Goodman (1979) for bivariate cross-classifications of ordinal variables. The model is

$$\log(v_{ij}) = u + u_i^{R_1} + u_j^{R_2} + \beta u_i v_j + \delta_{ij}, \quad i, j = 1, ...m \tag{3.7}$$

where

$$\delta_{ij} = \begin{cases} \delta & \text{If} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{3.8}$$

$u_1 < u_2 < \ldots < u_m$ or $v_1 < v_2 < \ldots < v_m$ are fixed scores assigned to the response categories, and u, $\lambda_i^{R_1}$, $\lambda_j^{R_2}$ and v_{ij} are as defined in equation (3.1).

In the same way, instead of assuming Poisson as we have in the above models of Tanner and Young (1985) and Agresti (1988), let the distribution of y_{ij}, the count frequency, follow a Negative Binomial in equation (2.6), with parameters μ_{ij} and k. In particular, that means, relaxing the assumption of Poisson on equality of mean and variance, since the mean and variance of Negative Binomial are not the same whenever $k > 0$ in equation (2.5). That is,

$$y_{ij} \sim Negbin(\mu_{ij}, k).$$

Based on the generalized linear model (GLM) assumptions approach, let us assume that

$$g(\mu_{ij}) = \eta_{ij} = x'_{ij}\beta, \tag{3.9}$$

where g is a link function, x_{ij} contains n×p covariates, where n is the number of cells and p is the number of parameters involve in the model. The most common choice for g is the log link, see McCullagh and Nelder (1989) and Fahrmeir and Tutz (2001).

The log likelihood function is

$$l(\theta, \phi) = \sum_{i=1}^{n} \left(\frac{(y_i\theta_i - b(\theta_i))}{\phi} + c(y_i, \phi) \right) \tag{3.10}$$

where θ subsumes all of the θ_i. It could also be written as a function of β and ϕ because (given the x_i), β determines all the θ_i. The main way of maximizing β is by maximizing (3.10).

Early approaches to rater agreement measurement before the introduction of modelling, have focused on the observed proportion of agreement, suggesting that chance agreement can be ignored. To these effect, just like Tanner and Young (1985), the log link for g will be structured in such a way that this chance agreement and beyond shall be accommodated but under the Negative Binomial sampling plan.

So the log link for g will be an independence model which can also incorporate agreement beyond chance given as

$$\log(\mu_{ij}) = \lambda_0 + \lambda_i^{Rx_1} + \lambda_j^{Rx_2} + \alpha_{ij} \tag{3.11}$$

where μ_{ij} is the expected cell count in cell i, j; λ_0 is the intercept; $\lambda_i^{Rx_1}$ is the effect of raters 1 in the ith category; $\lambda_j^{Rx_2}$ is the effect of raters 2 in the jth category and α_{ij} can be homogeneous or non-homogeneous agreement effect between raters 1 and raters 2 in i, j categories which we define just like the δ_{ij} in (3.4) respectively as

$$\alpha_{ij} = \begin{cases} \alpha & \text{If} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{3.12}$$

and

$$\alpha_{ij} = \alpha_i I(i = j) = \begin{cases} \alpha_1 & \text{If} \quad i = j = 1, \\ \alpha_2 & \text{If} \quad i = j = 2, \\ \vdots & \qquad \vdots \\ \alpha_I & \text{If} \quad i = j = I, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{3.13}$$

However, the log link for g for ordinary statistical independence or chance agreement between the ratings of the two raters can also be expressed as

$$\log(\mu_{ij}) = \lambda_0 + \lambda_i^{Rx_1} + \lambda_j^{Rx_2} \tag{3.14}$$

Also for agreement plus linear-by-linear association model, as we have stated above, the corresponding model under this underlying assumption will be

$$log(m_{ij}) = \lambda_0 + \lambda_i^{Rx_1} + \lambda_j^{Rx_2} + \beta u_i v_j + \alpha_{ij}, \quad i, j = 1, 2, \ldots, m \tag{3.15}$$

where

$$\alpha_{ij} = \begin{cases} \alpha & \text{If} \quad i = j, \\ 0 & \text{if} \quad i \neq j, \end{cases} \tag{3.16}$$

where $u_1 < u_2 < \ldots < u_m$ or $v_1 < v_2 < \ldots < v_m$ are fixed scores assigned to the response categories, and λ_0, $\lambda_i^{Rx_1}$, $\lambda_i^{Rx_2}$ and m_ij are as defined in equation (3.11).

The fact that $G(\mu) = x_{ij}\beta$ suggests a crude approximation estimate: regress $G(y)$ on x_{ij}, perhaps modifying y_{ij} in order to avoid violating range restrictions (such as taking log(0)), and accounting for the differing variances of the observations.

We shall consider these models under the families of Poisson and Negative Binomial in the algorithms for the following maximization of parameter estimates procedures.

Fisher scoring iteration is the widely used technique for maximizing the GLM likelihood over β.

3.1 Negative Binomial model

The loglikelihood contribution of a single observation is

$$l = log\Gamma(k^{-1} + y) - log\Gamma(k^{-1}) + ylog\left(\frac{k\mu}{1 + k\mu}\right) + k^{-1}log\left(\frac{1}{1 + k\mu}\right) - log(y!), \tag{3.17}$$

For a fixed value of k, this has the form of a GLM. Therefore it can be fitted by Fishers scoring, iteration re-weight least square (IRWLS), method. One iteration is

$$\beta^{(t+1)} = (X'WX)^{-1}X'Wu,$$

where the weights are

$$W = Diag\left[var(y_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)^2\right]^{-1}$$

and

$$u = \eta + \left(\frac{\partial\eta}{\partial\mu}\right)(y - \mu).$$

is the working variate. Recall from equation (2.5) that, $Var(y_i) = \mu_i + k\mu_i^2$, so under log *link*

$$\frac{\partial\eta}{\partial\mu} = \frac{1}{\mu},$$

the weights become

$$w_i = \frac{\mu_i^2}{\mu_i + k\mu_i^2},$$

and the working variate is

$$u_i = X_i{}'\beta + \frac{(y_i - \mu_i)}{\mu_i}.$$

As we usually do under the Poisson model, we regress u on X using the u_i's as weights to get the new estimate of β and continue these processes until the value of β converges. That is, the iteration is stopped when the consecutive iteration values are close or the loglikelihood values are maximized. In most applications k will be unknown. We can find the ML estimate for k by grid search. For each value of k along a grid, we can find the maximized value of the loglikelihood. Plotting this profile loglikelihood versus k will allow us to identify the ML estimate $\widehat{k}$. The estimated $\widehat{\beta}$ at $k = \widehat{k}$ is the ML estimate for β.

The ML estimates for k and β are asymptotically uncorrelated. This means that uncertainty in the estimate of k does little to affect the precision of $\widehat{\beta}$. If we were to use expectation rather than observed information, it would make no difference whether we regard k as fixed at $\widehat{k}$.

3.2 Cases outlines

The following cases shall be considered with the algorithm written based on the assumptions and properties of both Poisson and Negative Binomial models discussed in the previous sub sections.

- Case 1:- When $\beta = 0$ and $\alpha_{ij} = 0$, this results to equation 3.14
- Case 2:- When $\beta = 0$ and $\alpha_{ij} \neq 0$, this results to equation 3.11
- Case 3:- When $\beta \neq 0$ and $\alpha_{ij} \neq 0$, this results to equation 3.15

4 Empirical examples

4.1 Example 1.

Consider the data on cross-classification of two independent interpretations of Sputum cytology slides for lung by two observers taken from Becker (1990) based on the following categorical scales: 1=Negative, 2=Ambigious cells, 3=suspect, 4=positive, and 5=technically unsatisfactory.

Table 4.1: Cross-classification of two independent interpretations of Sputum cytology slides for lung of 100 patients

Category	Second observer					
First observer	1	2	3	4	5	Total
1	26	19	1	0	7	53
2	2	11	5	3	4	25
3	0	1	6	6	0	13
4	0	0	0	4	1	5
5	1	1	0	0	2	4
Total	29	32	12	13	14	100

Table 4.2: Parameter estimates under Poisson model on Sputum cytology data for Case 1: ($G^2 = 89.3338$,$X^2 = 79.2964$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	2.7885975	0.16805904	16.59	< 0.0001
raters1	-0.7093479	0.09356623	-7.58	< 0.0001
raters2	-0.2516554	0.07361791	-3.42	0.001

Table 4.3: Parameter estimates under Poisson model on Sputum cytology data for Case 2: ($G^2 = 59.3106$,$X^2 = 55.6819$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.98958675	0.23177627	8.58	< 0.0001
raters1	-0.69309524	0.09786396	-7.08	< 0.0001
raters2	-0.04692543	0.08655485	-0.54	0.588
alpha	1.27889242	0.23387508	5.47	< 0.0001

4.2 Example 2.

Consider the following data arising from the diagnosis of multiple sclerosis reported by two neurologists, on 69 patients of New Orleans , using the following categorical scales for diagnostic classes: 1=certain multiple sclerosis, 2=probable multiple sclerosis, 3=possible multiple sclerosis, 4=doubtful, unlikely, or definitely not multiple sclerosis (Basu et al. 1999).

Table 4.4: Parameter estimates under Poisson model for Sputum cytology data for Case 3: ($G^2 = 54.9287$,$X^2 = 49.3268$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	2.3594704	0.28683612	8.23	< 0.0001
raters1	-0.2120408	0.12119259	-1.75	0.080
raters2	-1.0128531	0.19182118	-5.28	< 0.0001
alpha	1.0042797	0.27111594	3.70	< 0.0001
beta	0.1552665	0.07411052	2.10	0.036

Table 4.5: Parameter estimates under Negative Binomial model on Sputum cytology data for Case 1: ($G^2 = 49.8358$,$X^2 = 38.3347$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	2.5074186	0.3113741	8.05	< 0.0001
raters1	-0.6709538	0.1253411	-5.35	< 0.0001
raters2	-0.1189377	0.1112238	-1.07	0.285

Table 4.6: Parameter estimates under Negative Binomial model on Sputum cytology data for Case 2: ($G^2 = 34.0581$,$X^2 = 30.2818$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.86113289	0.3493162	5.33	< 0.0001
raters1	-0.72004195	0.1350833	-5.33	< 0.0001
raters2	0.01503443	0.1223015	0.12	0.902
alpha	1.41712289	0.3671640	3.86	< 0.0001

Table 4.7: Parameter estimates under Negative Binomial model on Sputum cytology data for Case 3: ($G^2 = 32.7313$,$X^2 = 28.6627$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	2.1944387	0.4890735	4.49	< 0.0001
raters1	-0.1390016	0.1921464	-0.72	0.469
raters2	-0.9684229	0.2779526	-3.48	< 0.0001
alpha	1.2132863	0.4225651	2.87	0.004
beta	0.1179295	0.1067453	1.10	0.269

5 Results and Conclusion

5.1 Results

We present the summary of the working examples' results obtained using Poisson as well as the Negative Binomial as the underlying sampling plan in the previous

Table 4.8: Cross-classification table of multiple sclerosis assessment for 69 patients.

Category	Neurologist 2				
Neurologist 1	1	2	3	4	Total
1	5	3	0	0	8
2	3	11	4	0	18
3	2	13	3	4	22
4	1	2	4	14	21
Total	11	29	11	18	69

Table 4.9: Parameter estimates under Poisson model on diagnosis of multiple sclerosis data for Case 1: ($G^2 = 60.8985$,$X^2 = 60.8701$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.01464926	0.2855009	3.55	< 0.0001
raters1	0.25379598	0.1106240	2.29	0.022
raters2	0.01739279	0.1076902	0.16	0.872

Table 4.10: Parameter estimates under Poisson model on diagnosis of multiple sclerosis data for Case 2: ($G^2 = 43.9031$,$X^2 = 47.2775$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	0.73003098	0.2774976	2.63	0.009
raters1	0.27470099	0.1167700	2.35	0.019
raters2	-0.06691349	0.1155491	-0.58	0.563
alpha	1.04167944	0.2476778	4.21	< 0.0001

Table 4.11: Parameter estimates under Poisson model on diagnosis of multiple sclerosis data for Case 3: ($G^2 = 36.8981$,$X^2 = 44.4871$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.6201426	0.4127650	3.93	< 0.0001
raters1	-0.6465883	0.2581471	-2.50	0.012
raters2	-0.1952360	0.2180004	-0.90	0.370
alpha	0.5028639	0.3261764	1.54	0.123
beta	0.3353320	0.1308588	2.56	0.010

Table 4.12: Parameter estimates under Negative binomial model on diagnosis of multiple sclerosis data for Case 1: ($G^2 = 34.3501$,$X^2 = 31.7901$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.07108202	0.3745630	2.86	0.004
raters1	0.29094076	0.1512600	1.92	0.054
raters2	-0.06075422	0.1491649	-0.41	0.684

Table 4.13: Parameter estimates under Negative binomial model on diagnosis of multiple sclerosis data for Case 2: ($G^2 = 25.5948$,$X^2 = 26.3745$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	0.7432522	0.3860204	1.93	0.054
raters1	0.3157199	0.1586633	1.99	0.047
raters2	-0.1242281	0.1571962	-0.79	0.429
alpha	1.0511986	0.3620830	2.90	0.004

Table 4.14: Parameter estimates under Negative binomial model on diagnosis of multiple sclerosis data for Case 3: ($G^2 = 20.7266$,$X^2 = 26.0728$)

Coefficients	Value	Std. Error	t-value	p-value
intercept	1.8495632	0.5778368	3.20	0.001
raters1	-0.8442744	0.3550601	-2.38	0.017
raters2	-0.2930215	0.2996202	-0.98	0.328
alpha	0.4586100	0.4609252	0.99	0.320
beta	0.4212744	0.1849612	2.28	0.023

section.

Table 5.1: Summary table of results for case 1 (When $\beta = 0$ and $\alpha_{ij} = 0$) using Poisson as underlying distribution

			Poisson model		
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df
1	4	39.9167	79.2964	89.3338	22
2	4.3125	19.8292	60.8701	60.8985	13

Table 5.2: Summary table of results for case 1 (When $\beta = 0$ and $\alpha_{ij} = 0$) using Negative Binomial as underlying distribution

			Negative Binomial model			
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df	$\widehat{k}$
1	4	39.9167	38.3347	49.8358	22	0.25
2	4.3125	19.8292	31.7901	34.3501	13	0.203

Table 5.3: Summary table of results for Case 2 (When $\beta = 0$ and $\alpha_{ij} \neq 0$) using Poisson as underlying distribution

			Poisson model		
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df
1	4	39.9167	55.6819	59.3106	21
2	4.3125	19.8292	47.2775	43.9031	12

Table 5.4: Summary table of results for case 2 (When $\beta = 0$ and $\alpha_{ij} \neq 0$) using Negative Binomial as underlying distribution

			Negative Binomial model			
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df	$\widehat{k}*$
1	4	39.9167	30.2818	34.0581	21	0.25
2	4.3125	19.8292	26.3745	25.5948	12	0.203

Table 5.5: Summary table of results for Case 3 (When $\beta \neq 0$ and $\alpha_{ij} \neq 0$) using Poisson as underlying distribution

			Poisson model		
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df
1	4	39.9167	49.3268	54.9287	20
2	4.3125	19.8292	44.4871	36.8981	11

Table 5.6: Summary table of results for case 3 (When $\beta \neq 0$ and $\alpha_{ij} \neq 0$) using Negative Binomial as underlying distribution

			Negative Binomial model			
Example	Mean	Variance	Pearson's (χ^2)	Like. ratio (G^2)	df	$\widehat{k}$
1	4	39.9167	28.6627	32.7313	20	0.25
2	4.3125	19.8292	26.0728	20.7266	11	0.203

Table 5.7: Summary table of results for α_{ij} and β under Poisson as underlying distribution for Cases 2 and 3

	Case 2		Case 3			
Example	α_{ij}	p-value	α_{ij}	p-value	β	p-value
1	1.2789	< 0.0001	1.0043	< 0.0001	0.1553	0.0741
2	1.0417	< 0.0001	0.5029	0.123	0.3353	0.010

Table 5.8: Summary table of results for α_{ij} and β under Negative Binomial as underlying distribution for Cases 2 and 3

	Case 2		Case 3			
Example	α_{ij}	p-value	α_{ij}	p-value	β	p-value
1	1.4171	< 0.0001	1.2133	0.004	0.1179	0.269
2	1.0512	0.004	0.4586	0.320	0.4213	0.023

By considering the descriptive statistics of each of the data in the examples 1 and 2, it is clearly shown that there are sparse due to their respective large percentages of very small values in the cell counts. Also, their respective variance is greater than their mean, which indicates the existence of overdispersion. In Tables 5.1 to 5.6, the estimates of their respective Pearson's χ^2 and likelihood ratio G^2 under Poisson model assumptions for all the cases are greater than their respective estimates under Negative Binomial model assumptions with the same number of degrees of freedom. In addition to that, by observing each of Figures A.1 to A.4 which are the plots of fitted values against residuals under Poisson model or Negative Binomial model for cases 2 and 3, shows that Negative Binomial has better fit than its Poisson counterpart. Which hence means that Negative Binomial is a good substitute to Poisson for rater agreement data of this form which have sparse data. Tables 5.7 and 5.8 give the estimates for parameters α_{ij} and β under Cases 2 and 3 which can be used to test the null hypothesis $H_o : \alpha_{ij} = 0$ against the alternative hypothesis $H_1 : \alpha_{ij} \neq 0$ for case 2 and the null hypotheses $H_o : \alpha_{ij} = 0$ and $H_o : \beta = 0$ against their respective alternative hypotheses $H_1 : \alpha_{ij} \neq 0$ and $H_1 : \beta \neq 0$, under each of the Poisson and Negative Binomial models as underlying distribution. Based on the p-values from Tables 5.7 and 5.8, we observed that for Case 2 parameter α_{ij} is significant in all the examples under the two distributions. However, for Case 3, α_{ij} is only significant in examples 1 for both distributions, and parameter β is only significant in example 2. This implies that, under case 3 whenever β is significant α_{ij} will not be significant and vice versa. In addition to these results, we used two different data arising from the study reported in Holmquist et al. (1967) that investigated the variability in the classification of carcinoma in situ of the uterine cervix in which seven pathologists were requested to separately evaluate and classify 118 slides, we used (path A and path C) as well as (path E and path F) and we used also one other one on dysplasia assessments data by two different pathologists as presented by Barnhart and Williamson (2002). From all these we have the same conclusion that Negative binomial model fits better for rater agreement data that are sparse than Poisson model.

5.2 Conclusion

We observed that when there are sparse data, relaxing the assumption of Poisson model and assuming a Negative Binomial model that allows the spreads of the observations of the ratings between the two raters is better. In other word, modelling Negative Binomial model instead of Poisson model for rater agreement data that exhibit a large number of small values or zero cells is better in order not to underestimate the true variability in the data and the standard errors and also not to overestimating the degree of precision in the coefficients as well as in the estimates of the underlying statistics. Lastly, Negative Binomial model is a better substitute to Poisson for modelling the agreement of the ratings between two raters with sparse data in the arising table.

6 References

[1] Adejumo, A. O. (2002). On the Estimation of the Relative incident parameter of Negative Multinomial Distribution, *Journal of The Nigerian Statistical Association*; 15, 36-43.

[2] Agresti, A. (1988). A model for agreement between ratings on an ordinal scales, *Biometrics*, 44, 539- 548.

[3] Anscombe, F. J. (1949). The statistical analysis of insect counts based on the Negative Binomial distribution. *Biometrics*, 15, 165-173.

[4] Barnhart, H. X, and Williamson, J. M. (2002). Weighted least-squares approach for comparing correlated kappa. *Biometrics*, 58,1012-1019.

[5] Basu, S., Basu, A., and Raychaudhuri, A. (1999). Measuring agreement between two raters for ordinal response: a model based approach. *The Statistician*, 48(3), 339-348.

[6] Becker, M. P. (1990). Quasisymmetric models for the analysis of square contingency tables. *J. Roy. Statist. Soc. Series B*, 52, 369-378.

[7] Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Appli. Statist.*, 33,38-44.

[8] Cameron, A. C and Trivedi, P. K. (1990). Regression based tests for overdispersion in the Poisson model *J. Econometrics*, 46,347-364.

[9] Cameron, A. C and Trivedi, P. K. (1998). *Regression analysis of count Data.* Cambridge Unviversity press.

[10] Fahrmeir, L., and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models.* Springer.

[11] Goodman L. A. (1979). Simple models for the analysis of association in cross classifications having ordered categories. *J. Amer. Statist. Assoc.*, 74, 537-552.

[12] Holmquist, N. S., McMahon, C. A. and Williams, O. D. (1967). Variability in classification of carcinoma in situ of the uterine cervix. *Archives of Pathology*, 84, 334-345.

[13] Maclure, M., and Willett, W. C. (1987). Misinterpretation and misuse of the kappa statistic. *Amer. J. Epidemiol.*, 126, 161- 169.

[14] McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*, 2nd Edit. Chapman and Hall.

[15] Santos Silva J. M. C. (1997). Generalized Poisson regression for positive count data. *Comm. in Statist.*, Simulation Methods, 26, 1089-1102.

[16] Shoukri, M. M. (2004). *Measures of interobserver agreement.* Chapman and Hall.

[17] Tanner, M. A., and Young, M. A. (1985). Modelling agreement among raters. *J. Amer. Statist. Assoc.*, 80, 175-180.

A Appendix

A.1 Figures A.1 and A.2 are for Case 2 when $\beta = 0$ and $\alpha_{ij} \neq 0$ while Figures A.3 and A.4 are for Case 3 when $\beta \neq 0$ and $\alpha_{ij} \neq 0$

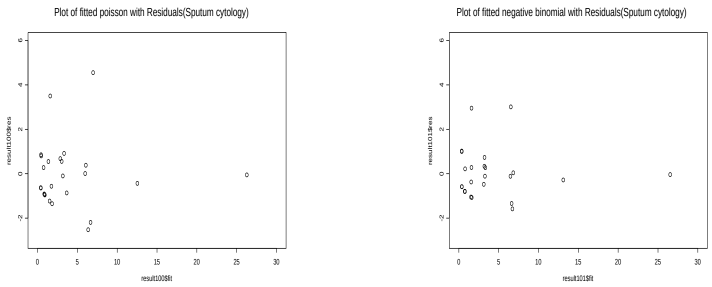

Figure A.1: Left is Fitted Poisson model and right is fitted Negative Binomial model against their respective residuals for Sputum cytology slides

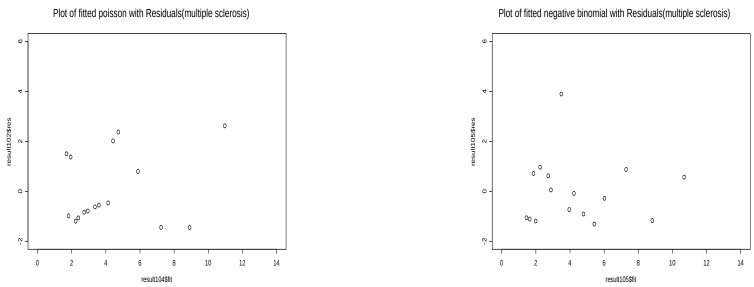

Figure A.2: Left is Fitted Poisson model and right is fitted Negative Binomial model against their respective residuals for diagnosis multiple sclerosis

A.2 Matrix of fitted Poisson for case 2 on Sputum cytology data

$$\begin{pmatrix} 26.2714 & 6.9773 & 6.6574 & 6.3523 & 6.0611 \\ 3.6564 & 12.5342 & 3.3289 & 3.1763 & 3.0306884 \\ 1.8283 & 1.7445 & 5.9801 & 1.5882 & 1.5154229 \\ 0.9142 & 0.8723 & 0.8323 & 2.8531 & 0.7578 \\ 0.4571 & 0.4362 & 0.4162 & 0.3971 & 1.3612 \end{pmatrix}$$

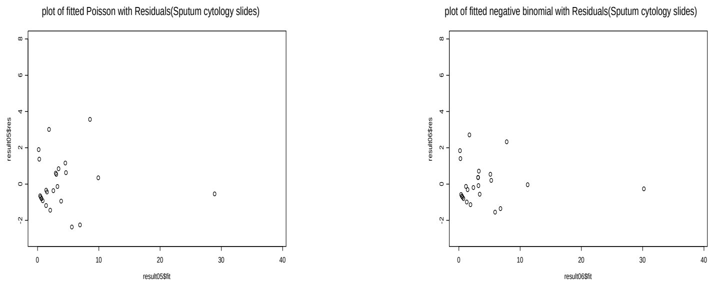

Figure A.3: Left is Fitted Poisson model and right is fitted Negative Binomial model against their respective residuals for Sputum cytology slides

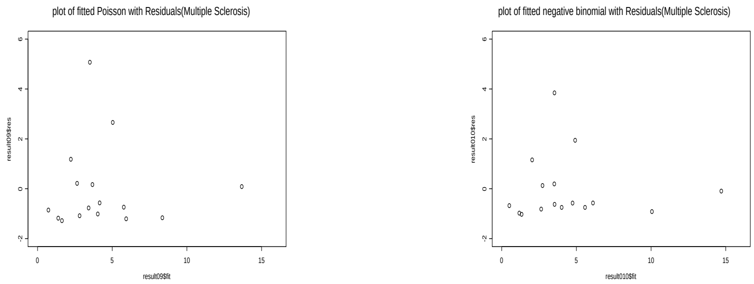

Figure A.4: Left is Fitted Poisson model and right is fitted Negative Binomial model against their respective residuals for diagnosis multiple sclerosis

A.3 Matrix of fitted Negative binomial for Case 2 on Sputum cytology data

$$\begin{pmatrix} 26.5294591 & 6.5284354 & 6.6273283 & 6.7277191 & 6.8296307 \\ 3.1301814 & 13.1083346 & 3.2257317 & 3.2745951 & 3.3241987 \\ 1.5235589 & 1.5466378 & 6.4768918 & 1.5938497 & 1.6179933 \\ 0.7415646 & 0.7527979 & 0.7642013 & 3.2002637 & 0.7875289 \\ 0.3609431 & 0.3664107 & 0.3719611 & 0.3775956 & 1.5812658 \end{pmatrix}$$

A.4 Matrix of fitted Poisson for Case 2 on diagnosis of multiple sclerosis data

$$
\begin{pmatrix}
5.880904 & 1.940833 & 1.815215 & 1.697727 \\
2.731175 & 7.239108 & 2.389072 & 2.234442 \\
3.594601 & 3.361944 & 8.910991 & 2.940832 \\
4.730988 & 4.424780 & 4.138391 & 10.968998
\end{pmatrix}
$$

A.5 Matrix of fitted Negative binomial for case 2 on diagnosis of multiple sclerosis data

$$
\begin{pmatrix}
6.016170 & 1.857115 & 1.640164 & 1.448557 \\
2.883406 & 7.285912 & 2.249068 & 1.986328 \\
3.953859 & 3.491963 & 8.823640 & 2.723745 \\
5.421714 & 4.788341 & 4.228959 & 10.685914
\end{pmatrix}
$$