.

Linear Mean-Variance Negative Binomial Models Applied to a Set of Orange Tissue-Culture Data

Naratip Jansakul[1], John P. Hinde[2]

[1] Prince of Songkla University, Department of Mathematics (jnaratip@ratree.psu.ac.th)
[2] National University of Ireland, Department of Mathematics (john.hinde@nuigalway.ie)

Abstract

Negative binomial maximum likelihood regression models are commonly used to analyze count data when overdispersion is present. There are various forms of the negative binomial model with different mean-variance relationships, however the most generally used are those with linear and quadratic relationships. We present a Newton-Raphson algorithm for obtaining maximum likelihood estimates of the linear mean-variance negative binomial (NB1) regression model. We also describe the construction of a half-normal plot with a simulated envelope for checking the adequacy of a selected NB1 model. These procedures are illustrated on a set of orange tissue culture data.

Keywords: Count data, Overdispersion, Negative binomial regression.

1. Introduction

Negative binomial (NB) models are very widely used for analyzing overdispersed Poisson counts as all important statistical inferences can be carried out more easily and conveniently than for other types of compound Poisson models (Lawless, 1987). Applications using the NB distribution can be found in many areas, for instance, economics (Hausman et al., 1984), political science (King (1988) and King (1989)), psychology (Gardner et al., 1995) and biostatistics (Alexander et al., 2000). The NB model can be considered as arising from a two-stage model assuming the counts to come from a Poisson distribution with varying mean. Taking the Poisson mean as a gamma distributed random variable leads to the NB model and we can obtain various forms of mean-variance relationship, in particular both linear and quadratic, depending on assumptions about the gamma mixing distribution parameters. The linear mean-variance NB model is obtained by allowing the gamma shape parameter to vary across observations and keeping the scale parameter constant, whereas the quadratic form arises from taking the shape parameter as constant and letting the scale vary. These two variance function models can lead to different models for the mean and also different forms of some associated statistics. Here we will denote the NB model with the linear variance by NB1 and the quadratic variance one by NB2. The NB2 model is a generalized linear model (glm) (Hinde and Demétrio, 1998) when the shape parameter is known. The parameter estimates for the NB2 model can be easily obtained using a full Newton-Raphson method, for example as is in Lawless (1987), or an iterative glm fitting procedure as in Hinde and Demétrio (1998).

This paper concentrates on the maximum likelihood fitting of NB1 models and their application to a real dataset. The paper begins in Section 2 with a short review of Poisson regression. Section 3 describes NB1 models and parameter estimation using a Newton-Raphson procedure. Methods of selecting an appropriate model are described in Section 4. To check the adequacy of a selected model we propose the use of a half-normal plot with a simulated envelope. Details of the construction of this plot are given in Section 5. In Section 6, we consider the application of the NB1 model to a set of orange tissue-culture data. The paper concludes with a brief discussion.

2. Poisson regression and Overdispersion

2.1 Poisson Regression Models

The random variables Y_i, $i = 1,\ldots, n$, represent counts with means m_i, and that $\boldsymbol{x_i} = (x_{i1}, x_{i2},\ldots, x_{ip})^T$ is an associated vector of covariates, with x_{i1} typically equal 1 to include the usual constant term in the model. The standard Poisson regression model assumes that $Y_i \sim \text{Pois}(m_i)$, and is a generalized linear model with variance function

$$\text{Var}(Y_i) = \text{Var}(m_i) = m_i. \qquad (1)$$

The m_i are typically modelled through the canonical log link function by

$$h_i = \log(m_i) = \mathbf{x}_i^T \boldsymbol{ß}$$

.

where β is a p vector of unknown parameters. The maximum likelihood estimate of β is easily obtained using iteratively reweighted least squares (IRLS) and the asymptotic covariance matrix Cov(β) is $(X^T W X)^{-1}$, where W is an $n \times n$ diagonal matrix with i^{th} diagonal element $W_i = \mu_i$, the iterative weight used in the IRLS procedure, see McCullagh and Nelder (1989).

For an appropriate well fitting model, we would expect that the residual deviance and the Pearson chi-square (X^2) would be approximately equal to the degrees of freedom (df). If the residual deviance and X^2 statistic exceed the df, the Poisson regression model may not be adequate, either through some systematic lack of fit, or because the strong assumption from the Poisson model that Var(μ_i) = μ_i is inapropriate; in this case the data are described as overdispersed. If the residual deviance is less than its df, it implies that there is underdispersion in the counts, i.e. the observed variance is less than the nominal Poisson variance. However, in practice, the underdispersion is less common, (McCullagh and Nelder, 1989).

In general, when there is overdispersion and we fail to take it into account, it can lead to misinterpretation of the fitted model, (Cox, 1983 and Hinde and Demétrio, 1998) since the overdispersion produces :

(i) smaller standard errors of the parameter estimates than the true values. Therefore we may incorrectly choose explanatory variables for the model that are not required;
(ii) too large a reduction of deviance associated with model selection tests. This again leads to selecting overly complex models.

3. Linear Mean-Variance NB Models

If Y_i, $i = 1, \ldots, n$, are now negative binomial distibuted counts with mean μ_i, and dispersion parameter α with $Y_i \sim$ NB1(μ_i, α) , the probability mass function (p.m.f.) is given by

$$f(y_i;\mu_i,\alpha) = \begin{cases} \dfrac{\Gamma(y_i + \alpha^{-1}\mu_i)}{y_i!\,\Gamma(\alpha^{-1}\mu_i)} \dfrac{\alpha^{y_i}}{(1+\alpha)^{y_i+\alpha^{-1}\mu_i}} & , \; y_i = 0, 1, \ldots, \; \alpha > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

with $\mathbf{E}(Y_i) = \mu_i$ and Var(Y_i) = $\mu_i (1 + \alpha)$. When $\alpha \to 0$, this NB1 model reduces to a Poisson Model.

3.1 Maximum Likelihood Estimation for the NB1 Distribution

For observed values $y_1, \ldots, y_n$, the NB1 log-likelihood, $\ell = \ell(\boldsymbol{\mu}, \alpha)$, is given by

$$\ell = \sum_{i=1}^{n} \left\{ y_i \log \alpha - \left(y_i + \frac{\mu_i}{\alpha} \right) \log(1 + \alpha) + \mathrm{dlg}(y_i, \alpha^{-1}\mu_i) - \log y_i! \right\}, \tag{3}$$

where dlg(y, α) = log $\Gamma(y + \alpha)$ – log $\Gamma(\alpha)$.

The NB1 is not a standard glm-type exponential family distribution, even when the overdispersion parameter α is known, and standard glm fitting methods will not apply. So here we consider a general Newton-Raphson iterative scheme. The first and second derivatives with respect to the underlying parameters are

$$\frac{\ell}{\beta_j} = \sum_i \alpha^{-1} \mathrm{ddg}(y_i, \alpha^{-1}\mu_i) - \frac{\log(1+\alpha)}{\alpha} \; \mu_i x_{ij}, \;\; j = 0, 1, \ldots, p \tag{4}$$

$$\frac{\partial^2 \ell}{\partial \beta_j \, \partial \beta_k} = - \sum_i \alpha^{-1} \mathrm{ddg}(y_i, \alpha^{-1}\mu_i) - \frac{\log(1+\alpha)}{\alpha} - \alpha^{-2} \mu_i \mathrm{dtg}(y_i, \alpha^{-1}\mu_i) \; \mu_i x_{ij} x_{ik}, \;\; j, k = 0, 1, \ldots, p \tag{5}$$

$$\frac{\ell}{\alpha} = -\alpha^{-2} \sum_i \frac{\mu_i - y_i}{1 + \alpha^{-1}} - \mu_i \log(1 + \alpha) + \mu_i \mathrm{ddg}(y_i, \alpha^{-1}\mu_i) \tag{6}$$

.

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_i 2\alpha^{-3}\left[\frac{m_i - y_i}{1+\alpha^{-1}} - m_i \log(1+\alpha) + m_i \mathrm{ddg}(y_i, \alpha^{-1} m_i)\right] + \alpha^{-4}\left[\frac{y_i - m_i}{(1+\alpha^{-1})^2} + \frac{\alpha m_i}{1+\alpha^{-1}} - m_i^2 \mathrm{dtg}(y_i, \alpha^{-1} m_i)\right], \tag{7}$$

and

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \alpha} = \alpha^{-2} \sum_i \left[\log(1+\alpha) - \mathrm{ddg}(y_i, \alpha^{-1} m_i)\right] + \alpha^{-1} m_i \mathrm{dtg}(y_i, \alpha^{-1} m_i) - \frac{\alpha}{1+\alpha} m_i x_{ij}, \tag{8}$$

where ddg(y,α) and dtg(y,α) denote the differences of the di-gamma and tri-gamma functions. These are defined by

$$\mathrm{ddg}(y,\alpha) = \frac{\partial}{\partial \alpha}(\mathrm{dlg}(y,\alpha)) = \vartheta(y+\alpha) - \vartheta(\alpha)$$

$$= \begin{cases} 0, & y = 0 \\ \sum_{t=0}^{y-1} (\alpha + t)^{-1}, & y > 0, \end{cases}$$

where ϑ is the di-gamma function, and

$$\mathrm{dtg}(y,\alpha) = \frac{\partial^2}{\partial \alpha^2}(\mathrm{dlg}(y,\alpha)) = \zeta(y+\alpha) - \zeta(\alpha)$$

$$= \begin{cases} 0, & y = 0 \\ \sum_{t=0}^{y-1} (\alpha + t)^{-2}, & y > 0, \end{cases}$$

where ζ is the tri-gamma function.

Let $\mathbf{s}(\beta, \alpha)$ be the vector of score functions defined by

$$\mathbf{s}(\beta, \alpha) = \begin{pmatrix} s_\beta(\beta, \alpha) \\ s_\alpha(\beta, \alpha) \end{pmatrix} = \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \alpha} \end{pmatrix}$$

and let $I(\beta, \alpha)$ be the $(p+1) \times (p+1)$ observed information matrix, which we partition as

$$I(\beta, \alpha) = \begin{pmatrix} I_{\beta\beta}(\beta, \alpha) & I_{\beta\alpha}(\beta, \alpha) \\ I_{\alpha\beta}, (\beta, \alpha) & I_{\alpha\alpha}(\beta, \alpha) \end{pmatrix}, \tag{9}$$

where $I_{\beta\beta} = -\frac{\partial^2 \ell}{\partial \beta \partial \beta^T}$ is the $p \times p$ symmetric matrix, $I_{\alpha\alpha} = -\frac{\partial^2 \ell}{\partial \alpha^2}$ is a scalar and $I_{\alpha\beta} = I_{\beta\alpha}^T = -\frac{\partial^2 \ell}{\partial \alpha \partial \beta}$, is a $1 \times (p+1)$ matrix.

Writing $\beta^{(m)}$ and $\alpha^{(m)}$ as the estimates at the m^{th} iteration, the standard Newton-Raphson iterative scheme gives

$$\begin{pmatrix} \beta^{(m+1)} \\ \alpha^{(m+1)} \end{pmatrix} = \begin{pmatrix} \beta^{(m)} \\ \alpha^{(m)} \end{pmatrix} + \left[I^{(m)}\right]^{-1} \mathbf{s}^{(m)}, \tag{10}$$

where $I^{(m)}$ and $\mathbf{s}^{(m)}$ are $I(\beta, \alpha)$ and $\mathbf{s}(\beta, \alpha)$ evaluated at $\beta = \beta^{(m)}$ and $\alpha = \alpha^{(m)}$. The iteration (10) must be carried out until convergence, which can be assessed using a stopping rule such as

$$|\alpha^{(m+1)} - \alpha^{(m)}| < \hat{I} \quad \text{or} \quad |\ell^{(m+1)} - \ell^{(m)}| < \hat{I}.$$

The procedure requires good initial values, which can be obtained as follows:

.

- β ; fit a standard Poisson regression model to obtain $\beta^{(0)}$ and initial estimates of the fitted values $m_i^{(0)}$.
- α ; equate the Pearson X^2 statistic from the Poisson fit to its expected value under the NB1 model, to give

$$\alpha^{(0)} = (n-p)^{-1} \sum_i \frac{\left(y_i - m_i^{(0)}\right)^2}{m_i^{(0)}},$$

this is in fact the quasi-likelihood estimate of the overdispersion parameter from the constant overdispersion Poisson model.

The asymptotic variance of β and α are the diagonal elements of $I^{-1}(\beta, \alpha)$, and are automatically provided at the final iteration. This iterative procedure is simply implemented in any computer software that can handle matrices, such as, SPlus and the free software **R** (The Comprehensive R Archive Network : http://lib.stat.cmu.edu/R/CRAN/).

4. Selecting an Appropriate Model

Testing the Poisson assumption against the NB1 alternative corresponds to testing $H_0 : \alpha = 0$ against $H_1 : \alpha > 0$. The commonly used test statistics, the likelihood ratio test (LRT) defined by $-2\{\ell(\boldsymbol{\mu}) - \ell(\boldsymbol{\mu}, \alpha)\}$, where $\ell(\boldsymbol{\mu})$ and $\ell(\boldsymbol{\mu}, \alpha)$ are maximized log-likelihood estimates under the Poisson and NB1 model, respectively, and the Wald test specified by $\frac{\alpha^2}{\mathrm{Var}(\alpha)}$, are both applicable here. Although somecare is required as the null hypothesis is on the boundary of the parameter space (e.g. the null distribution of the LRT is not the usual χ_1^2 distribution), and also the alternative hypothesis is one-sided as we are only testing for overdespersion.)

Selecting an appropriate model among all possible NB1 regression is straightforward using the standard likelihood criteria, for example, Akaike information criterion (AIC) (Akaike, 1973) or Baysian information criterion (BIC) given in Schwarz (1978). These criteria simply require the maximized log-likelihood value from the NB1 distribution fit and are defined as:

$$\begin{aligned} \mathrm{AIC} &= -2\ell + 2(\text{number of fitted parameters}) \\ \mathrm{BIC} &= -2\ell + \log(n) \times (\text{number of fitted parameters}). \end{aligned}$$

5. Model Checking

A model diagnostic technique that has been found to be useful for checking the adequacy of fitted models is the use of half-normal plots with a simulated envelope. This technique was first proposed by (Atkinson, 1985). He applied the plot to check model adequacy using Pearson residuals or Cook's statistics in normal regression. The technique was further developed for glms using (standardized) Pearson residuals and (standardized) deviance residuals by Williams (1987). Williams claimed that the plot can detect both outliers and overdispersion in both Poisson and binomial regression models.

Even though the NB1 regression model is not a glm, we can define its complete p.m.f. and hence the log-likelihood function. The associated (standardized) Pearson residual, or the standardized studentized residual, for the NB1 model can be obtained by using the general definition, $\frac{(y-m)}{\sqrt{\mathrm{Var}(Y)}}$ (Lawless, 1987). Denoting the standardized Pearson residual for an NB1 fit by $\dot{r}_{Pi}$, the i^{th} component is

$$\dot{r}_{Pi} = \frac{y_i - m_i}{\sqrt{m_i(1+\alpha)}}.$$

NB1 deviance residuals cannot be obtained simply based on the usual deviance expression for glms: $-2\{\ell(m,\alpha;y) - \ell(y,\alpha;y)\}$, as some of the individual components can be negative. Nelder (1991) pointed out that the log-likelihood (3) does not have the

.

property that its mode occurs at m = y unless $y = 0$. He used $y_i + \frac{1}{2}$ as the approximate mode of ℓ and then approximated the deviance component for y_i by

$$\begin{cases} \dfrac{2m_i \log(1+\alpha)}{\alpha}, & y_i = 0 \\ -2 \left\{ y_i + \dfrac{1}{2} - m_i \right\} \dfrac{\log(1+\alpha)}{\alpha} + d\lg(y_i, \alpha^{-1} m_i) - d\lg(y_i, \alpha^{-1}(y_i + \dfrac{1}{2}) \right\}, & y_i > 0 \end{cases}$$

Jansakul (2001) explored this approximation and found that $y + \frac{1}{2}$ is not an adequate approximation of the mode. Her investigations indicated that there is no simple form for the mode of ℓ, but values such as $y + k$, where $\frac{\alpha}{2+1/y} < k < \frac{\alpha}{2}$ are likely to be close to giving the mode, and for large y, $k \approx \frac{\alpha}{2}$ works well. Using this simple form gives the deviance residuals $r_{D,i}$ = sgn($y_i - m_i$) $\sqrt{D_i}$ for the NB1 model, where

$$D_i = \begin{cases} \dfrac{2m_i \log(1+\alpha)}{\alpha}, & y_i = 0 \\ -2 \left\{ y_i + \dfrac{\alpha}{2} - m_i \right\} \dfrac{\log(1+\alpha)}{\alpha} + d\lg(y_i, \alpha^{-1} m_i) - d\lg(y_i, \alpha^{-1} \left\{ y_i + \dfrac{\alpha}{2} \right\}, & y_i > 0 \end{cases}$$

Following the general procedure for constructing the half-normal plots with a simulated envelope given in Vieira et al. (2000), a plot for checking a selected NB1 model using (standardized) deviance residuals can be constructed as follows:

- Fit a NB1 model to obtain $\hat{\mu}, \hat{\alpha}$ and calculate the ordered absolute values of deviance residuals $r_{D,i}$;
- Simulate nineteen samples for the response variable under the fitted model, by first generating $e_{0j,i}$ where $e_{0j,i} \sim \Gamma(\alpha^{-1} m_i, 1)$, $j = 1,\ldots, 19$, $i = 1,\ldots, n$, calculating $e_{ji} = e_{0j,i} \times \alpha m_i^{-1}$ then simulating $Y_{ji} \sim \text{Pois}(m_i e_{ji})$ to give $Y_{ji} \sim \text{NB1}(m_i, \alpha)$, i.e. 19 datasets based on the fitted model.
- Refit the model, using the same explanatory variables, to each sample and calculate the ordered absolute values of the deviance residuals, $r^*_{j(D,i)}$, $j = 1, \ldots, 19$, $i = 1,\ldots, n$;
- For each i calculate the minimum, maximum and the mean of the $r^*_{j(D,i)}$;
- Plot these values and the observed $r_{D,i}$ against the half-normal scores (expected order statistics); $\Phi^{-1}\{(i + n - \frac{1}{8}) / (2n + \frac{1}{2})\}$, where Φ is the normal cumulative density function (Demétrio and Hinde, 1997).

If the selected model is adequate, the observed $r_{D,i}$ should lie within the simulated envelope.

Demétrio and Hinde (1997) gave a GLIM macro to construct such plots with special emphasis on overdispersed models (i.e. constant overdispersion Poisson and NB2 models for extra Poisson variation). These are easily adapted for the NB1 deviance residuals as all that is required are two macros, one to calculate the NB1 deviance residuals and the other to simulate from an NB1 distribution.

6. Application: An Orange Tissue-culture Experiment

The orange variety *Valencia* was used in a tissue-culture experiment conducted in Brazil to study the effect of six carbohydrate sources (maltose, glucose, galactose, lactose, sucrose and glycerol) on the stimulation of somatic embryos

.

from callus cultures. The response variable is the number of embryos observed after approximately four weeks. The experiment was a completely randomized block design with the above six sugars at dose levels of 18, 37, 75, 110 and 150 mM for the first five and 6, 12, 24, 36, and 50 mM for the glycerol, and 5 replicates of each treatment, see Tomaz et al. (2001), for further details of the experiment and histological analyzes. The main interest was in the dose-response relationship for the sugars (maltose, galactose and lactose) that produced large numbers of embryos. The number of embryos produced is highly variable, see Figure 1, with marked differences between the three sugars. Table 1 presents the mean and variance of the number of embryos, classified by sugars and dose levels (excluding 1 missing value). Most of the sample overdispersion index values (relative to a baseline Poisson distribution) exceed 3, and give strong evidence of overdispersion. In their analysis Tomaz et al. (2001) used a quadratic response function over the dose levels and a simple constant overdispersion Poisson model fitted by quasi-likelihood to take account of overdispersion. Here we use this dataset to illustrate the use of maximum likelihood estimation for the NB1 distribution.

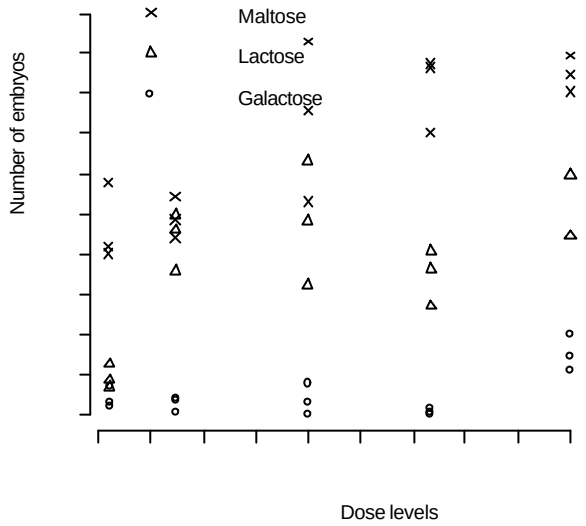

Figure 1: Orange (*Valencia*) tissue culture data

Table 1:Orange (*Valencia*) tissue culture data : Mean and variance (Var) of number of embryos classified by sugars and dose levels

		Dose levels (mM)				
Sugars		18	37	75	110	150
	Mean	233.00	245.33	369.67	407.00	424.33
Maltose	Var	2368.00	654.33	9952.33	2356.00	506.33
	o.i.	10.16	2.67	26.92	5.79	1.19
	Mean	47.33	219.33	239.33	174.33	260.50
Lactose	Var	224.33	1310.33	5854.33	1234.33	2964.50
	o.i.	4.74	5.97	24.46	7.08	11.38
	Mean	21.67	14.00	18.33	4.00	75.67
Galactose	Var	185.33	76.00	408.33	13.00	508.30
	o.i.	8.55	5.43	22.28	3.25	6.72

o.i. denotes overdispersion index $= \dfrac{\text{Var}}{\text{Mean}} - 1.$

Writing $\boldsymbol{\mu}$ for the vector of the mean numbers of embryos and taking sugar (S) and dose as factors, fitting the full interaction Poisson regression model ($\log(\boldsymbol{\mu}) = \text{S} * \text{DOSE}$), the residual deviance is 298.04 on 29 df, which as expected

shows strong evidence of overdispersion. The half-normal plot, Figure 2 (a), also indicates greater variation than in the Poisson model as all the Poisson deviance residuals lie above the upper envelope.

Fitting the corresponding NB1 model with the full interaction between dose and sugar gives a likelihood ratio test statistic for overdispersion of 166.59 on 1 df. The model certainly fits the data much better than the Poisson model. This model is equivalent to fitting a model with an interaction between sugar and a quartic polynomial over the actual dose levels ($\log(\boldsymbol{\mu}) = S * (D + D^2 + D^3 + D^4)$). This suggests that we might consider simplifying the model by fitting lower order polynomials over the dose levels. The model that seems best in terms of both AIC and BIC is $\log(\boldsymbol{\mu}) = S * (D + D^2 + D^3)$, see Table 2. This model gives a separate cubic dose relationship for each of the sugar types. The corresponding half-normal plot; Figure 2(b) indicates that the model is consistent with the data.

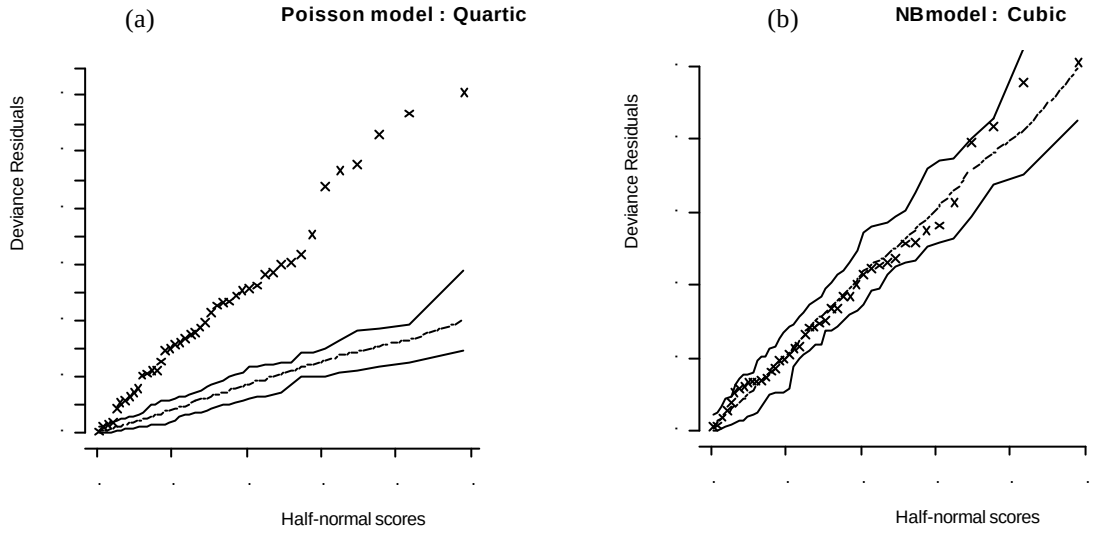

Figure 2: Orange (*Valencia*) tissue culture data : Half-normal Plots based on Poisson and NB1 model

Table 2: Orange (*Valencia*) tissue-culture data: Statistics for Poisson and NB1 models

S is a three-level factor for sugar
DOSE is a five-level factor for the dose levels
D is a variate for the dose level

	Models					
Description	$\log(\boldsymbol{\mu})$	a	-2ℓ	df_1	AIC	BIC
Poisson	$S * (D + D^2 + D^3 + D^4)^\dagger$	0	573.143	29	**603.143**	**629.906**
	$S * (D + D^2 + D^3)$	0	648.715	32	672.715	694.125
	$S * (D + D^2)$	0	959.656	35	977.656	993.414
	$S * D$	0	1182.815	38	1194.815	1205.520
NB1	$S * (D + D^2 + D^3 + D^4)$	6.331	406.552	28	438.552	467.099
	$S * (D + D^2 + D^3)$	7.772	413.000	31	**439.000**	**462.194**
	$S * (D + D^2)$	15.360	438.385	34	458.385	476.227
	$S * D$	24.410	457.234	37	471.234	483.724

$S * (D + D^2 + D^3 + D^4)^\dagger$ is equivalent to S*DOSE

Figure 3 shows the plot of the predicted mean number of embryos for cubic model. This suggests that the dose response relationship for maltose and galactose may be approximately linear or quadratic. In order to investigate this, we

.

fitted NB1 regression models with cubic, quadratic and linear functions over the dose levels, see Table 3. The best model suggested by AIC and BIC for each sugar is different. That is the NB1 model with a log-linear model in the dose levels for maltose, a log-quadratic model for galactose and a log-cubic regression model for lactose. The parameter estimates and their standard errors (given in the parentheses) are given as follows:

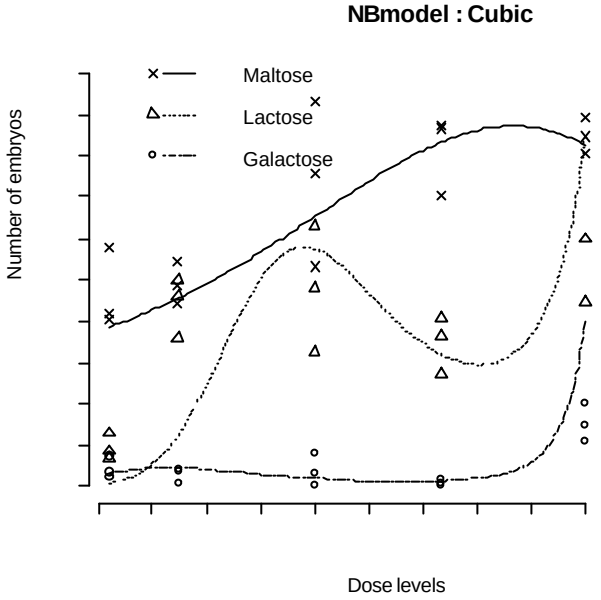

Figure 3: Orange (*Valencia*) tissue culture data : Observed (*symbols*) and estimated (*lines*) values of embryogenic responses

Table 3: Orange (*Valencia*) tissue-culture data: Statistics for NB1 models classified by sugar

	Models					
Sugars	$\log(\boldsymbol{\mu})$	a	-2ℓ	df	AIC	BIC
Maltose						
	$\tilde{D}+\tilde{D}^2+\tilde{D}^3$	5.78	157.84	10	167.84	171.38
	$\tilde{D}+\tilde{D}^2$	5.86	158.03	11	166.03	168.86
	$\tilde{D}$	7.84	161.74	12	**167.84**	**169.82**
Lactose						
	$\tilde{D}+\tilde{D}^2+\tilde{D}^3$	9.01	142.17	9	**152.17**	**155.37**
	$\tilde{D}+\tilde{D}^2$	30.02	157.63	10	165.63	168.18
	$\tilde{D}$	36.52	160.30	11	166.30	168.22
Galactose						
	$\tilde{D}+\tilde{D}^2+\tilde{D}^3$	8.84	112.32	10	122.32	125.86
	$\tilde{D}+\tilde{D}^2$	10.37	114.26	11	**122.26**	**125.08**
	$\tilde{D}$	31.06	127.72	12	133.72	135.84

$\tilde{D}$ denotes a vector of standardized $\tilde{D}_i$; $\tilde{D}_i = \dfrac{D_i - \overline{D}}{\sqrt{\mathrm{Var}(D)}}$, $i = 1,\ldots, n$, where $\overline{D} = n^{-1}\sum_i D_i$.

.

Maltose : $\log(\mu) = 5.783 + 0.226\,\tilde{D}$
(0.043) (0.040)
a = 7.836(3.219)

Lactose : $\log(\mu) = 5.564 - 0.604\,\tilde{D} - 0.636\,\tilde{D}^2 + 0.675\,\tilde{D}^3$
(0.091) (0.209) (0.113) (0.135)
a = 9.005(3.824)

Galactose : $\log(\mu) = 1.887 + 0.045\,\tilde{D} + 0.997\,\tilde{D}^2$
(0.420) (0.162) (0.244)
a = 10.374(4.666)

7 Discussion

Fitting NB1 models using a Newton-Raphson iterative procedure is conveniently performed in any computer software that can deal with matrices, in particular, R or SPlus, as the commands for calculating di-gamma and tri-gamma functions are also available. Moreover, the correct asymptotic covariance matrix of the parameter estimates Cov(β, a) is automatically provided at the final iteration.

The half-normal plot with a simulated envelope using the approximated deviance residuals seems to be very useful check on the adequacy of the linear-mean variance negative binomial model. The plot can also be used to check the correct form of the variance function and seems to give some suggestion whether the NB1 or the NB2 variance function is appropriate. An investigation of this will be reported elsewhere.

Acknowledgements: The authors are grateful to M.L. Tomaz and B. M. J. Mendes for providing orange *Valencia* tissue culture data.

References

[1] Akaike, H. (1973). Information theory and extension of the maximum likelihood principle. In B. N. Petrov and F. Csáki (Eds.), *Proceedings 2nd International Symposium on Inference Theory*, Budapest: Akadmiai Kiadó, pp267-281.

[2] Alexander, N., R. Moyeed, and J. Stander (2000). Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics*, Vol. 1, pp453-463.

[3] Atkinson, A. (1985). *Plots, Transformations and Regression. An introduction to graphical methods of diagnostic regression analysis*. Oxford: Clarendon Press.

[4] Demétrio, C. G. B. and J. P. Hinde (1997). Half-normal plots and overdispersion. *GLIM Newsletter*, Vol. 27, pp19-26.

[5] Gardner, W., E. P. Mulvey, and E. C. Shaw (1995). Regression analyses of counts and rates: Poisson, Overdispersed Poisson, and Negative binomial models. *Psychological Bulletin*, Vol. 11, pp392-404.

[6] Hausman, J., B. H. Hall, and Z. Griliches (1984). Econometric models for count data with an application to the patents-R & D relationship. *Econometrica*, Vol. *52*, pp909-938.

[7] Hinde, J. P. and C. G. B. Demétrio (1998). Overdispersion: Models and Estimation. In *The 13th Brazilian Symposium of Probability and Statistics (13o SINAPE)*, Caxambu, Minas Gerais, Brazil. 73p.

[8] Jansakul, N. (2001). Some aspacts of modelling overdispersed and zeroinflated count data. *School of Mathematical Sciences, Exeter University*. PhD Thesis, 286 pages, Unpublished.

[9] King, G. (1988). Statistical models for Political Science event counts: Bias in conventional procedures and evidence for the experimental Poisson regression model. *American Journal of Political Science*, Vol. *32*, pp838-863.

[10] King, G. (1989). Variance specification in event count models: From restrictive assumptions to a generalized estimator. *American Journal of Political Science*, Vol. 33, pp762-784.

[11] Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, Vol. 15, pp209-225.

[12] McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. (second ed.). London: Chapman and Hall.

.

[13] Nelder, J. A. (1991). Generalized linear models with negative binomial or beta-binomial errors. *Department of Mathematics, Imperial College of Science, Technology and Medicine*. 10 pages, Unpublished.

[14] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, Vol. 6, pp461-464.

[15] Tomaz, M. L., B. M. J. Mendes, F. A. Mourão Filho, C. G. B. Demétrio, N. Jansakul, and A. P. M. Rodriguez (2001). Somatic embryogenesis in *citrus spp*: Carbohydrate stimulation and histodifferentiation. *In Vitro Cellular & Developmental Biology – Plant*, Vol. 37, pp446-452.

[16] Vieira, A. M. C., J. P. Hinde, and C. G. B. Demétrio (2000). Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics*, Vol.27, pp373-389.

[17] Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, Vol. 36, pp181-191.