Benin Journal of
Statistics
ISSN 2682-5767
Vol. 2, pp. 1– 15 (2019)

GLMM applications in some binary count models

B. H. Lawal*

Department of Statistics & Mathematical Sciences, Kwara State University, Malete, Kwara State

Abstract. In this paper, we employed the generalized linear mixed models for the binomial, the beta-binomial, the multiplicative binomial and the Com-Poisson Binomial distributions. These are applied to two examples of over-dispersed binomial data with covariates. The logistic linear model is employed for comparative purposes only. SAS PROC NLMIXED is employed for implementing these models. For the logistic-normal model, we also compare our results from PROC NLMIXED with those from PROC GLIMMIX in SAS, and R packages *glimmer*, and STATA program *melogit*. For this case, our results agree with those obtained from PROC NLMIXED.The conditional log-likelihoods functions are integrated out using the adaptive Gaussian Quadrature (usually with 32 q-points) and the optimized by using either the Newton-Raphson or Nelder-Mead Simplex algorithms. Starting values are obtained by specifying a large range of grid values for each parameter of the models.

Keywords: beta binomial, com-poisson binomial, glmm, multiplicative binomial, adaptive Gaussian quadrature.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

For data having binary outcomes, the underlying distribution is the binomial. For a random variable Y having a 'success' probability π and thus, a 'failure' probability $(1-\pi)$, if the number of trials is n, then $Y \sim \text{Binom}(n, \pi)$. For such data having covariates, $x_1, x_2, \cdots, x_p$, the usual model would be the binomial or logistic model:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots, \beta_p x_{pi} \tag{1}$$

The model in (1) assumes that the responses from one subject to another are independently distributed. However, for data arising from teratology or similar studies, this may not necessarily be the case as there might be intra-correlation, say, for example amongst subjects in the same litter. In such a situation the estimated model variance in (1) will grossly underestimate the true variance and this will lead to the model not fitting the data and consequently leading to overdispersion. As an example, the data in Table 1 which relates to the frequency of males in 6115 families with 12 children in Sax-ony, previously analyzed in Sokal and Rohlf (1969). The data is originally from Geissler (1889) and had similarly been analyzed in Borges *et al.* (2014).

Table 1.: Distribution of males in 6115 families with 12 children

Y	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
count	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115

The observed mean and variance of this data set are respectively, $\mu_y = 6.2306$ and $\sigma_y^2 = 3.4898$, giving a corresponding observed π_i, the proportion of males in a 12-child family to be 0.5192.

*Corresponding author. Email: bayo.lawal@kwasu.edu.ng

When a logistic model of constant probability of success in the form, $\text{logit}(\pi) = \beta_0$ is applied to the above data, the estimated mean and variance are respectively, $\bar{y} = 6.2304$ and $s^2 = 2.9956$ with a dispersion parameter (DP=$\bar{y}/s^2 = 6.2304/2.9956 = 2.0798$) and a corresponding estimated proportion, $\hat{\pi} = 1/(1 + \exp(-0.0769)) = 0.5192$, where, -0.0769 being the parameter estimate of β_0 in the binomial model. We observe immediately that the data is over-dispersed and that the estimated variance under the binomial model grossly underestimates the observed variance in the data viz: of $12\hat{\pi}(1 - \hat{\pi}) = 2.9958$ is less 3.4898.

In this paper, we shall adopt a variance function of the form in (2), viz:

$$\text{Var}(Y) = n_i\pi_i(1 - \pi_i)[1 + (n - 1)\rho]= \phi n_i\pi_i(1 - \pi_i) \tag{2}$$

where $|\rho| < 1$ is the intra-correlation coefficient between subjects. Clearly, when $\rho = 0$ then the variance function becomes that of the binomial. Otherwise, we see that the variance under the binomial will always be less than the observed. In such situations, we can attempt to employ other binary based distributions that would account for this extra-variation, such as the multiplicative binomial,and the Com-Poisson binomial, both of which are considered in this paper. Other models in this category are the double binomial, Feirer *et al.* (2013), and the correlated binomial model, Kupper and Haseman (1978). These distributions having extra dispersion parameters discussed here are the multiplicative and the Com-Poisson which are briefly discussed in the following sections.

1.1 *The Multiplicative Binomial Model (MBM)*

Lovinson (1998) proposed an alternative form of the two-parameter exponential family generalization of the binomial distribution first introduced by Altham (1978) which itself was based on the original Cox's (1972) representation as:

$$f(y) = \frac{\binom{n}{y}\boldsymbol{\psi}^y(1 - \boldsymbol{\psi})^{n-y}\,\boldsymbol{\omega}^{y(n-y)}}{\sum_{j=0}^{n}\binom{n}{j}\boldsymbol{\psi}^j(1 - \boldsymbol{\psi})^{n-j}\,\boldsymbol{\omega}^{j(n-j)}},\ y = 0, 1, \ldots, n. \tag{3}$$

where $0 < \boldsymbol{\psi} < 1$ and $\boldsymbol{\omega} > 0$. When $\boldsymbol{\omega} = 1$ the distribution reduces to the binomial with $\pi = \boldsymbol{\psi}$. If $\boldsymbol{\omega} = 1$, $n \to \infty$, and $\boldsymbol{\psi} \to 0$, then $n\boldsymbol{\psi} \to \mu$ and the MBD reduces to Poisson(μ).

Elamir(2013) presented some elegant characteristics of the multiplicative binomial distribution, including its four central moments. His treatment includes generation of random data from the distribution as well as the likelihood profiles and several examples-some of which are similarly employed in this chapter.

The probability π of success for the Bernoulli trial, that is, $P(Y = 1)$ can be computed from the following expression in (4) as:

$$\pi_i = \boldsymbol{\psi}^i\frac{\kappa_{n-i}(\boldsymbol{\psi}, \boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi}, \boldsymbol{\omega})};\quad i = 1, 2. \tag{4}$$

and with $\kappa(.)$ as defined in (5), where:

$$\kappa_{n-a}(\boldsymbol{\psi}, \boldsymbol{\omega}) = \sum_{y=0}^{n-a}\binom{n-a}{y}\boldsymbol{\psi}^y(1 - \boldsymbol{\psi})^{n-a-y}\,\boldsymbol{\omega}^{(y+a)(n-a-y)}. \tag{5}$$

and from (5), we have:

$$\begin{aligned}\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega}) &= \sum_{y=0}^{n}\binom{n}{y}\boldsymbol{\psi}^y(1-\boldsymbol{\psi})^{n-y}\,\boldsymbol{\omega}^{y(n-y)},\\ \kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega}) &= \sum_{y=0}^{n-1}\binom{n-1}{y}\boldsymbol{\psi}^y(1-\boldsymbol{\psi})^{n-1-y}\,\boldsymbol{\omega}^{(y+1)(n-1-y)},\\ \kappa_{n-2}(\boldsymbol{\psi},\boldsymbol{\omega}) &= \sum_{y=0}^{n-2}\binom{n-2}{y}\boldsymbol{\psi}^y(1-\boldsymbol{\psi})^{n-2-y}\,\boldsymbol{\omega}^{(y+2)(n-2-y)}.\end{aligned} \tag{6}$$

Thus, from (5), $\pi_1 = \boldsymbol{\psi}[\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})/\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})]$, $\boldsymbol{\psi}$ therefore can be defined as the probability of success weighted by the intra-units association measure $\boldsymbol{\omega}$ which measures the dependence among the binary responses of the n units. Thus if $\boldsymbol{\omega} = 1$, then $\pi = \boldsymbol{\psi}$ and we have independence among the units. However, if $\boldsymbol{\omega} \neq 1$, then, $\pi \neq \boldsymbol{\psi}$ and the units are not independent.

The mean and variance of the MBD are given respectively as:

$$E(Y) = n\pi_1, \tag{7a}$$

$$\mathrm{Var}(Y) = n\pi_1 + n(n-1)\pi_2 - (n\pi_1)^2, \tag{7b}$$

The corresponding two-parameter exponential family representation (Feirer *et al.*, 2013) is also given by:

$$f(y|\psi,\omega) = \binom{n}{y}\frac{1}{\sum_{j=0}^{n}\binom{n}{j}\boldsymbol{\psi}^j(1-\boldsymbol{\psi})^j\,\boldsymbol{\omega}^{j(n-j)}} \times \exp\left(y\log\frac{\boldsymbol{\psi}}{1-\boldsymbol{\psi}} + (n-y)y\log\boldsymbol{\omega}\right). \tag{8}$$

1.2 *The Com-Poisson Binomial (CPB) model*

The probability density function for the Com-Poisson Binomial distribution (Borges *et al.*, 2014) is given by:

$$f(y|n,p,\nu) = \frac{\binom{n}{y}^{\nu}\pi^y(1-\pi)^{n-y}}{\sum_{k=0}^{n}\binom{n}{k}^{\nu}\pi^k(1-\pi)^{n-k}}, \quad y = 0,1,\ldots,n, \tag{9}$$

With $\pi \in (0,1)$ and $\nu \in$ R. If $\nu = 1$, the model reduces to the binomial distribution and values of $\nu > 1$ indicate underdispersion, while values of $\nu < 1$ similarly indicate overdispersion with respect to the binomial distribution.

The Com-Poisson distribution (Conway and Maxwell, 1961) is given in (10),

$$f(y) = \frac{\lambda^y}{(y!)^\nu}\frac{1}{Z(\lambda,\nu)}, \quad y = 0,1,2,\cdots, \quad \lambda > 0,\ \nu \geq 0. \tag{10}$$

where the the normalizing term $Z(\lambda, \nu)$ is defined as:

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}. \tag{11}$$

An approximation to the CPB distribution in the limit $n \to \infty$ and with $\lambda = n^\nu p$ is given in Shmueli *et al.* (2005). Following Borges *et al.* (2014), if we let θ be defined as:

$$\theta = \frac{\pi}{1-\pi}. \tag{12}$$

and dividing both the denominator and numerator of the expression in (9) by a factor of $(1-\pi)^m (m!)^\nu$, we thus have:

$$f(y|n, \theta, \nu) = \frac{\theta^y}{(y!)^\nu} \frac{1}{Z(\theta, \nu)}, \quad y = 0, 1, 2, \cdots, \quad \theta > 0, \ \nu \geq 0, \tag{13}$$

where the normalizing term is defined as:

$$Z(\theta, \nu) = \sum_{j=0}^{n} \frac{\theta^j}{[j!(n-j)!]^\nu}. \tag{14}$$

The various properties of the CPB or the Com-Poisson have been presented in various papers Borges *et al.* (2014), Shmueli *et al.* (2005), and Kadane *et al.* (2006) applied the CPB to the number of killings in rural Norway. The means and variance of Y_i are respectively given as:

$$E(Y) = \sum_{j=0}^{n} \frac{j\,\theta^j}{Z(\theta, \nu)[j!(n-j)!]^\nu}, \quad \text{and} \tag{15a}$$

$$\text{Var}(Y) = \sum_{j=0}^{n} \frac{j^2\,\theta^j}{Z(\theta, \nu)[j!(n-j)!]^\nu} - [E(Y)]^2. \tag{15b}$$

The two-parameter exponential family representation of the distribution is presented in (16) (Lawal, 2017)

$$f(y|n, \pi, \nu) = \binom{n}{y}^\nu \frac{1}{\sum_{k=0}^{n} \binom{n}{k}^\nu \left(\frac{\pi}{1-\pi}\right)^k} \times \exp\left(y \log \frac{\pi}{1-\pi}\right). \tag{16}$$

If we let $\mu = \theta^{1/\nu}$, then $\theta = \mu^\nu$. Hence, the pmf in (13) becomes:

$$f(y|n, \mu, \nu) = \left(\frac{\mu^y}{(y!)}\right)^\nu \frac{1}{S(\mu, \nu)}, \quad y = 0, 1, 2, \cdots, n; \quad \mu > 0, \ \nu \geq 0, \tag{17}$$

where, $S(\mu, \nu) = \sum_{j=0}^{n} \left[\frac{\mu^j}{j!(n-j)!}\right]^\nu$. The formulation in (17) is based on Guikema and Coffel (2008). This model will be designated here as CPB_μ.

1.3 *Mixture binomial models*

Mixture binomial models provide an alternative way of handling overdispersion in binary data, which is to model the success probability of the Binomial distribution using a continuous distribution defined on the standard unit interval. The resultant general class of univariate discrete distributions is known as the class of Binomial mixture distributions. The Beta-Binomial (BB) distribution is a prominent member of this class of distributions. The Kumaraswamy-Binomial (KB) distribution (Kumuraswamy, 1980) is another well utilized member of this class. Others in this class is The Two-Sided Power-Binomial (TSP) Distribution (Ali, 2019) and the three parameter McDonald Generalized Beta-Binomial distribution(McGBB), Manjor *al.* (2015). Some theoretical properties of McGBB, KB, BB and the TCP distributions are already discussed in the literature and are therefore not being discussed here. The parameters of all the models will be estimated via maximum likelihood estimation technique. We would apply these models to the example datasets in this chapter. One of the consequences in applying the binomial model to data with inherent over-dispersion is that the model would not fit. The McGBB and other mixing models consider here therefore model the parameter π of the binomial with continuous distributions defined in the interval (0,1). We will however focus in this study on the logistic-Normal and Beta-binomial models in this category of mixed models.

Generally, a mixture model is obtained by evaluating the well-known integral as:

$$f(y) = \int_0^1 f_{Y|\pi}\pi^y f_\pi(\pi|\boldsymbol{\theta})d\pi \tag{18}$$

for $y = 0, 1, \ldots, n$ and $\boldsymbol{\theta}$ is the parameter space of the mixing distribution. We discuss in what follows the logistic-normal and beta-binomial mixture models for mitigating against over dispersion in binary data.

1.4 *The Logistic-Normal Distribution (LND)*

Lindsey and Altham (1982) employed the normal distribution as the mixing distribution for the logit of the Bernoulli probability in a way that is similar to that employed in Hinde (1982) for over-dispersed Poisson data, viz.

$$\int_{-\infty}^{\infty} f(y; p(\lambda))\phi(\lambda; \mu, \psi^2)d\lambda \tag{19}$$

where $f(y; p)$ is the binomial function with p as the corresponding probability, and with $\log(p/(1-p)) = \lambda$, and $\phi(.)$ the normal density with μ the mean logit and ψ the standard deviation. The expression in (19) requires numerical integration and PROC NLMIXED in SAS has several procedures for evaluating this integral. We implement this model only for the toxicological data using the default adaptive Gaussian Quadrature procedure in SAS PROC NLMIXED. The logistic-normal model has the form:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mathbf{X}\boldsymbol{\beta} + u_i \tag{20}$$

where,

$$r_{ij}|u \sim \text{Binomial}(n_{ij}, \pi_{ij}), \quad \text{with } u_i \sim N(0, \sigma_u^2)$$

1.5 *The Beta-Binomial (BB) model*

The beta-binomial in Skellam (1946) is, of course, a mixture of the binomial $\text{Bin}(n, \pi)$ and the beta distribution $\text{Beta}(\alpha, \beta)$, where,

$$Y|p \sim \text{Bin}(n, p), \quad \text{and} \quad p \sim \text{Beta}(\alpha, \beta).$$

That is, the response variable Y has the binomial

$$f(y|p) = \binom{n}{y} p^y (1-p)^{(n-y)}, \quad y = 0, 1, \ldots, n \tag{21}$$

and the probability p has the beta distribution defined as:

$$f(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \ \alpha > 0, \beta > 0. \tag{22}$$

where $B(\alpha, \beta)$ is the beta function with parameters α and β such that

$$B(\alpha, \beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1} dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Thus, $\text{Bin}(n, \pi) \wedge \text{Beta}(\alpha, \beta) \sim BB$, Hence,the compound distribution is given by:

$$\begin{aligned} f(y, \alpha, \beta) &= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \, dp \\ &= \binom{n}{y} \frac{1}{B(\alpha, \beta)} \int_0^1 p^{\beta+y-1}(1-p)^{\beta+n-y-1} \, dp \\ &= \binom{n}{y} \frac{B(\alpha+y, \beta+n-y)}{B(\alpha, \beta)}, \quad y = 0, 1, \ldots, n \end{aligned} \tag{23}$$

with $\text{Var} = np(1-p)[1 + \rho(n-1)]$, where $\rho^2 = \dfrac{1}{\alpha+\beta+1}$ is the dispersion parameter which measures the pairwise correlation between the clusters (Bernoulli trials). Clearly, as $\rho \to 0$, the variance $\text{var}(Y) \to np(1-p)$, thus, reducing to the binomial distribution. Further,$\pi = \dfrac{\alpha}{\alpha+\beta}$ and $0 < \pi < 1$ and is the probability of 'success' and that $\alpha_i + \beta_i = \text{constant}$.

The mean and variance of the beta-binomial are given by:

$$E(Y) = n\pi \qquad \text{and} \qquad \text{Var}(Y) = n\pi(1-\pi)[1 + \rho^2(n-1)]. \tag{24}$$

where

$$\pi = \frac{\alpha}{\alpha+\beta}, \quad \text{and} \quad \rho^2 = \frac{1}{\alpha+\beta+1}.$$

The beta-binomial model in (23) can be parameterized as:

$$\alpha = \theta\tau^{-1}, \quad \text{and} \quad \beta = (1-\theta)\tau^{-1},$$

where $\tau > 0$ and $0 < \theta < 1$. The mean and variance under this formulation are respectively:

$$E(Y|n, |\theta, \tau) = n\theta, \quad \text{var}(Y) = n\theta(1-\theta)\left[1 + (n-1)\frac{\tau}{1+\tau}\right]$$

$$\theta = \hat{\pi}_{ij}\alpha = \theta/\tau; \quad \beta = (1-\theta)/\tau\rho^2 = \frac{1}{\alpha+\beta+1} = \frac{\hat{\tau}}{1+\hat{\tau}}$$

and the log-likelihood becomes:

$$\begin{aligned} LL &= z + \log[\Gamma(r+\alpha)] + \log[\Gamma(n-r+\beta)] - \log[\Gamma(n+\alpha+\beta)] \\ &\quad - \log[\Gamma(\alpha)] - \log[\Gamma(\beta] + \log[\Gamma(\alpha+\beta)] \end{aligned}$$

where $z = \log[\Gamma(n+1)] - \log[\Gamma(r+1)] - \log[\Gamma(n-r+1)]$

2. Materials and methods

To model each of the distributions presented in the preceding sections, our modeling is of the following form:

$$\pi_i = 1/(1+\exp(-lr)) \tag{25a}$$

$$\theta_i = 1/(1+\exp(-lr)) \tag{25b}$$

$$\psi_i = 1/(1+\exp(-lr)) \tag{25c}$$

$$\mu_i = \exp(lr) \tag{25d}$$

where

$$\mathbf{lr} = \mathbf{x}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}})$$

- The logistic-normal is modeled with π defined in (25a)
- The beta-binomial is modeled with θ defined as in (25b)
- The Multiplicative binomial is modeled with ψ defined in (25c).
- The Com-Poisson binomial model is modeled with μ defined in (25d).

For a single observation, the log-likelihoods for the binomial, beta-binomial, the multiplicative binomial, and the Com-Poisson binomial are displayed in expressions (26a) to (26d), respectively.

$$LL1 = z + y\log(\pi) + (n-y)\log(1-\pi) \tag{26a}$$

$$\begin{aligned} LL2 &= z + \log[\Gamma(y+\alpha)] + \log[\Gamma(n-y+\beta)] - \log[\Gamma(n+\alpha+\beta)] \\ &\quad - \log[\Gamma(\alpha)] - \log[\Gamma(\beta] + \log[\Gamma(\alpha+\beta)] \end{aligned} \tag{26b}$$

$$LL3 = z + y\log(\boldsymbol{\psi}) + y(n-y)\log\boldsymbol{\omega} - \log\left[\sum_{j=0}^{n}\binom{n}{j}\boldsymbol{\psi}^j(1-\boldsymbol{\psi})^{n-j}\,\boldsymbol{\omega}^{j(n-j)}\right] \tag{26c}$$

$$LL4 = y\log\theta - \nu\log(y!) - \nu\log(n-y_))! - \log Z(\theta,\nu) \tag{26d}$$

where $z = \log\binom{n}{y}$ and $Z(\theta, \nu)$ is as defined in (14).

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 approximation employed is the Adaptive Gaussian Quadrature (Pinheiro & Bates, 1995) and the optimization algorithms utilized are either the Nelder-Mead Simplex method(NMSIMP), or the Newton-Raphson method with line search (NEWRAP). Choice of starting values are a problem but this is accomplished by using appropriate values over grids of values for our parameters.

2.1 *Applications*

Generalized linear mixed models are applied to two well analyzed two data sets in this study. The data sets employed are the teratology data and the cardio-vascular data both of which are fully discussed in the following sections:

2.2 *Example I: Teratology Data*

Teratology is the study of abnormalities of physiological development. The data in Table **??** gives the results of studies on the effects of dietary regiments or chemical agents on fetal developments in rats. (Moore & Tsiatis, 1991). Female rats on iron-deficient diets were assigned to four groups. Group 1 (placebo), group 2 (injections on days 7 and 10), group 3 (days 7 and 10), and group 4 (Injections weekly). 58 rats were made pregnant, sacrificed after three weeks, and the total number of dead fetuses was counted in each litter, as well as the mother's hemoglobin level denoted in Table **??** as h. Due to non measured covariates and genetic variability the probability of death may vary from litter to litter within a particular treatment group.(n, y). Recent analysis of the data was presented in Agresti (2015) employing the beta-binomial.

Table 2.: Results of Teratology Studies on Female Rats

Groups	Size	h	(n,y)	h	(n,y)	h	(n,y)	h	(n,y)	h	(n,y)
GP1	31	4.1	10/1	3.2	11/4	4.7	12/9	3.5	4/4	3.2	10/10
		5.9	11/9	4.7	9/9	4.7	11/11	3.5	10/10	4.8	10/7
		4.3	12/12	4.1	10/9	3.2	8/8	6.3	11/9	4.3	6/4
		3.1	9/7	3.6	14/14	4.1	12/7	4.8	11/9	4.7	13/8
		4.8	14/5	6.7	10/10	5.2	12/10	4.3	13/8	3.9	10/10
		6.3	14/3	4.4	13/13	5.2	4/3	3.9	8/8	7.7	13/5
		5.0	12/12								
GP2	12	8.6	10/1	11.1	3/1	7.2	13/1	8.8	12/0	9.3	14/4
		9.3	9/2	8.5	13/2	9.4	16/1	6.9	11/0	8.9	4/0
		11.1	1/0	9.0	12/0						
GP3	5	11.2	8/0	11.5	11/1	12.6	14/0	9.5	14/1	9.8	11/0
GP4	10	16.6	3/0	14.5	13/0	15.4	9/2	14.5	17/2	14.6	15/0
		16.5	2/0	14.8	14/1	13.6	8/0	14.5	6/0	12.4	17/0

Thus for litter 1, the mother's hemoglobin level was 4.1 and one dead fetus out of 10 offsprings. Similarly, for litter 58, the hemoglobin level was 12.4 with zero fetal death in 17 offsprings.

To implement the GLMM models for the distributions considered above, we let π_{ij} denote the probability of death for fetus j in litter i. Then,

- For the binomial GLMM (logistic-binomial), the set up is:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 z_{2i} + \beta_2 z_{3i} + \beta_3 z_{4i} + \beta_4 h_{ij} + u_{ij} \quad = \mathbf{z}'\boldsymbol{\beta} \tag{27}$$

where

$$z_2 = \begin{cases} 1 \text{ if GP2} \\ 0 \text{ otherwise} \end{cases}, \quad z_3 = \begin{cases} 1 \text{ if GP3} \\ 0 \text{ otherwise} \end{cases}, \quad z_4 = \begin{cases} 1 \text{ if GP4} \\ 0 \text{ otherwise} \end{cases}$$

Thus, GP1 is the reference category in this set up. Thus, from (27), $\pi_{ij} = 1/[1 + \exp(-\mathbf{z}'\boldsymbol{\beta})]$ and, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ are to be estimated from the models in addition to the other parameters of the models.

The results of applying the above models to the data in Table 2 are displayed in Table 3.

Table 3.: Parameter and GOF Statistics under the Five Models

Parameter	BIN	BIN^a	BB	MBM	CPB
Intercept	2.1795	2.6250	3.3850	5.1124	1.6829
	(0.5056)	(1.2066)	(1.2475)	(2.0689)	(1.3969)
gp2	-2.4748*	-3.7614	-3.0029*	-0.1679	-2.2034
	(0.5045)	(1.2447)	(1.2046)	(0.1142)	(2.0654)
gp3	-3.1527*	-4.7565	-3.0523	-0.9458	-2.8776
	(0.9433)	(1.8938)	(1.8317)	(1.2370)	(2.6102)
gp4	-2.0520	-3.8229	-1.4166	2.0501	-2.1266
	(1.0629)	(2.5681)	(2.4868)	(1.4775)	(2.6012)
hemo	-0.2190*	-0.1773	-0.3662	-0.6954*	-0.1348
	(0.1020)	(0.2448)	(0.2465)	(0.2841)	(0.1649)
			$\hat{\tau} = 0.0112$	$\hat{\omega}* = 0.9552$	$\hat{\nu} = 0.5115$
			(0.0578)	(0.0792)	(0.6094)
ρ			0.0111		
-2LL	240.40	183.29	184.4	188.9	183.0
σ^2	na	2.3089	2.0867	1.8124	0.7256
	0.4122	(0.8748)	(1.0731)	(1.7897)	(1.4290)
X_W^2	155.8163	7.1119*	16.5452	17.8363	22.5917
d.f	53	52	51	51	51

For the binomial-logistic model, the residual marginal variance is $Var(u_{ij}) = \sigma^2 + \frac{\pi^2}{3}$ and consequently, the intra-class correlation coefficient is $\dfrac{\sigma^2}{\sigma^2 + \frac{\pi^2}{3}}$ under this model. For our data , this would be, $\dfrac{2.3089}{2.3089 + 3.2925} = 0.4122$. Under the binomial model for instance, the Wald's GOF is 155.8163 on 53 d.f, giving an estimate of the dispersion parameter to be $2.9400 > 1$, thus there is clear over-dispersion in the data as a result of the intra-class correlation among the offsprings within a litter. The results in Table 3 indicate that the logistic-binomial model BIN^a does well when compared with all other distributions. The estimated variance function for the logistic-binomial can take one of two forms presented in (28a) and (28b) respectively. The former is the Williams (1982) type III variance function, which approximates the variance and the latter is the estimated variance presented in Hinde and Dimétrio (2007).

$$\text{Var}(Y) \approx n_i\pi_i(1 - \pi_i)[1 + \sigma^2(n_i - 1)\pi_i(1 - \pi_i)] \tag{28a}$$

$$\text{Var}(Y) = n_i^2\sigma^2\pi_i^2(1 - \pi_i)^2 \tag{28b}$$

With E(Y) still being $n_i\pi_i$, the Wald test statistic is computed as:

$$X_W^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{m}_i)^2}{\text{Var}}, \quad i = 1, 2, \ldots, N(= 58).$$

Thus, the Wald's GOF under logistic-binomial model of 7.1119 is based on the Williams (1982) type III variance function. However, if we use the expression for the variance Hinde & Dimetrio (2007), then X_w^2 would be 15.8857 on 52 d.f. which is still a very good fit. The multiplicative, the beta-binomial and the Com-Poisson behave very well. However, the BB fits much better with an a variance estimated component of $\hat{\sigma}^2 = 2.0867$ and $\hat{\rho}^2 = 0.0111$. Of course, the GLM versions of these distributions also provide several possibilities for modeling the data in Table 2. Table 4 gives the estimated probabilities under these models for the data in Table 2.

Table 4.: Estimated probabilities for each group under the various Models

Group	BIN	BIN^a	BB	MBM	CPB
GP1	0.7614	0.7795	0.7775	0.7915	0.7805
GP2	0.0961	0.0897	0.0822	0.1228	0.0926
GP3	0.0343	0.0258	0.0326	0.0295	0.0264
GP4	0.0444	0.0367	0.0423	0.0408	0.0402

The estimated probabilities under the GLMM models are very similar (safe for GP2 under MBM). However, the beta-binomial and the logistic-binomial are the easiest to implement in terms of early convergence over initial grid values for the parameters.

2.3 *Example Data II: Cardiotoxic Effects of Chemotheraphy*

This example is from (Nelson *et al.*, 2006) and was originally published at http://www.stat.sc.edu/ kerrie/cardiodata.html. The data is a correlated binary data which studies the cardiotoxic effects of doxorubicin chemoteraphy on the treatment of acute lymphoblastic leukemia in childhood. The data set is presented in Table 5.

In this study, 24 patients previously cured of leukemia had a long-term followup visit to determine how their hearts were functioning. For each subject on a visit, six similar tests of heart function were performed, with the result of each test being coded as normal/abnormal. Thus, we have $N = 24$ clusters, each patient serving as a cluster, and $n_i = 5$ or $n_i = 6$ observations per cluster (some patients have only 5, and not 6 tests performed). In Table 5, ID=Patient number, r is the number of abnormal heart tests, n is the number of tests, time=time since chemotherapy (in years), and dose=1 if High and 0 if low dosage.

Let the response variable be $\mathbf{Y}_{ij}$ from patient i having a j^{th} heart test such that:

$$\mathbf{Y}_{ij} = \begin{cases} 1 & \text{if abnormal} \\ 0 & \text{if normal} \end{cases}$$

Suppose the probability of an abnormal result is π_i, then we have:

$$\begin{aligned} \pi_i &= \Pr[\mathbf{Y}_{ij} = 1 | \text{Dose}_i, \text{Time}_i] \\ &= \frac{e^{\beta_0 + \beta_1 \text{Dose}_i + \beta_2 \text{Time}_i}}{1 + e^{\beta_0 + \beta_1 \text{Dose}_i + \beta_2 \text{Time}_i}}, \end{aligned} \tag{29}$$

where *Dose* is 1 if high and 0 if low, and Time_i is the time in years since the last chemotherapy.

The GLMM model therefore becomes:

$$\begin{aligned} \ln\left(\frac{\pi_i}{1-\pi_i}\right) &= \mathbf{x}' \,\boldsymbol{\beta} + \mathbf{u}; \quad \text{that is,} \\ \text{logit}_{ij} &= \beta_0 + \beta_1 \,\text{Dose}_i + \beta_2 \,\text{Time}_i + u_i \end{aligned} \tag{30}$$

Table 5.: Cardiotoxicity study data

ID	r	n	dose	time
1	4	6	1	13.7
2	0	5	1	15.6
3	3	5	1	4.6
4	4	5	1	13.0
5	0	5	0	6.2
6	1	6	1	15.4
7	2	5	0	6.5
8	0	5	0	4.4
9	1	5	0	9.6
10	3	5	1	11.2
11	3	5	0	8.1
12	3	5	1	13.1
13	1	5	0	10.1
14	4	6	0	8.4
15	1	5	0	4.2
16	1	5	1	13.5
17	1	5	1	17.9
18	1	5	0	8.8
19	2	6	0	5.9
20	3	5	1	13.2
21	4	5	1	14.5
22	4	6	0	8.1
23	0	5	0	8.2
24	4	6	0	8.1

and $\mathbf{u} \sim N(0, \sigma^2)$. Our formulation of the above model is based on the fact that there is no significant interaction between dose and time (Nelson *et al.*, 2006). Since the binary observations are assumed correlated, suppose we let ρ be the correlation (or overdispersion parameter) between two heart measurements on the same subject.

The parameters θ, $\boldsymbol{\psi}$ and θ in the Beta binomial, the multiplicative, and Com-Poisson binomial are modeled as discussed earlier in this paper. The results of implementing these models are presented in Table 6.

Table 6.: Paramter Estimates and Wald's GOF Under the five Models

Parameter	BN	BN^a	BB	MBM	CPB
Int.	-0.2202	-0.2970	-0.2965	-0.6673	-6.1728
	(0.6023)	(0.8355)	(0.7116)	(1.9173)	(10.8569)
Dose	0.9631	1.1045	0.9626	2.1152	17.3792
	(0.5729)	(0.7809)	(0.7194)	(3.2562)	(15.0007)
Time	-0.0638	-0.0702	-0.0591	-0.1295	-1.0349
	(0.0739)	(0.1016)	(0.0886)	(0.2579)	(1.4851)
	na	na	$\hat{\tau} = 0.1274$	$\hat{\omega} = 1.5211$	$\hat{\nu} = 17.8409$
			(0.2816)	(1.8420)	(1.2207)
σ^2	na	0.6146	0.0043	3.8910	340.02
		(0.5339)	(1.0054)	(13.5081)	(471.40)
-2LL	81.8958	79.0	78.9	78.7	78.1
X^2_{w}	35.4218	20.0621	23.7920	7.2461	0.8812
d.f.	21	20	19	19	19

When the binomial model was applied to the data, the Wald's $X^2_W = 35.4218$ on 21 d.f, giving an estimated dispersion parameter (DP) of 1.8668 indicating a strong overdispersion of the data. Clearly, the Com-Poisson binomial fits the data best with estimated variance component being $\hat{\sigma}^2 = 340.02$ which may be higher when compared to the other models. However, we must realize that the Com-Poisson is not structured like the other models because θ is not parametrized in terms of its mean μ. Further, when the variance structure in (28a) is employed to compute the Wald's test statistic for the logistic-binomial, the computed value is $X_w^2 = 8.4396$.

Clearly, based on the -2LL statistic, each of the four GLMM models are very close, but the Wald's GOF for the multiplicative GLMM fits the data best. We observe here that the Wald GOF may be very susceptible to cases when $r = 0$.

3. Conclusions

Clearly, the results will vary with each data set, but based on the results in Examples I and II, each of these models, with the exception of the binomial, will behave very well in modeling over-dispersed data. The procedures introduced in this paper further add to the suit of alternatives for modeling over-dispersed binary data. It should be noted here that both R and STATA do not yet have packages for implementing all the above models-the exception being the logistic-normal. The results of applying both STATA 15 and R packages to model the logistic-normal model for the data in Example I are presented in the appendix. The results agree. Notice that we have employed the Adaptive Gaussian Quadrature with 100 q-points in R. The default being the Laplacian integration which is not as accurate as the AGQ.

Acknowledgments

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

References

Agresti, A. (2015). Foundations of Linear & Generalized Models. Wiley, NY.

Ali, I.Genc, (2019). The two-sided power-Binomial Distribution for modeling count data. Communications in Statistics-Theory and Methods, **48**(10): 2548–2561.

Altham, P.M. (1978). Two generalizations of the binomial distribution. Appl. Statist., **27**: 162-167.

Borges, P., Rodrigues, T. & Balakrishnan, N. (2014). Com-Poisson type generalization of the Binomial distribution and its properties and applications. Statistics & Probability Letters, **87**: 158-166.

Conway, R.W. & Maxwell, W.L. (1961). A queuing model with state dependent service rates. Journal of Industrial Engineering, **12**: 132-136.

Cox, D.R. (1972). The analysis of multivariate binary data. Applied Statistics, **21**: 113-120.

Elamir, E.A.H. (2013). Multiplicative-binomial distribution: Some results and characterization, inference and random data generation. Journal of Statistical Theory and Applications, **12**(1): 92-105.

Feirer V., Friedl, H. and Hirn, U. (2013). Modeling over-and underdispersed frequencies of successful ink transmissions onto paper. Journal of Applied Statistics, **40**(3): 626-643. http://dx.doi.org/10.1080/02664763.2012.750284.

Geissler, A. (1889). Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. Z. *Köngl. Sächs. Statist. Bur.,* **35**: 1-24.

Guikema S.D., & Coffelt J. P. (2008). A flexible count data regression model for risk analysis. Risk Analysis, **28**(1): 213–223.

Hinde, J. (1982). Compound Poisson regression models. Lecture Notes in Statistics, **14**: pp. 109-121.

Hinde, J. & Demétrio,C.G.B. (2007). Overdispersion: Models and Estimation.

Kadane, J., Shmueli, G., Minka, G., Borle, T. & Boatwright, P. (2006). Conjugate analysis of the Conway Maxwell Poisson distribution. Bayesian Analysis, **1**: 363-374.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. Journal of Hydrology, **46**(1): 79-88.

Kupper L.L. & Haseman, J.K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. Biometrics, **34**: 69-76.

Lawal, H.B. (2017). Employing the double, multiplicative and the Com-Poisson binomial distributions for modeling over- and under-dispersed binary data. Journal of Advances in Mathematics & Computer Science, **23**(3): 1-17. doi:10.9734/BJMCS/2017/33475.

Lindsey, J.K. & Altham, P. K. (1982) Varieties of over and underdispersion models for binary data.

Lovison, G. (1998). An alternative representation of Altham's multiplicative-binomial distribution. Statistics & Probability letters, **36**: 415-420.

Manjor, C., Wijekoom, P. & Yapa, R.D. (2013). The McDonald generalized beta-binomial distribution: A new binomial mixture distribution and simulation based comparison with its nested distributions in handling overdispersion. International Journal of Statistics & Probability, **2**(2): 24-41.

Martinez, E.Z., Achcar, J.A., Aragon, D.C. (2015). Parameter estimation of the beta-binomial distribution: an application using the SAS software. Ciência e Natura, Santa Maria, **37**(4): 12-19.

Moore, D. F. & Tsiatis, A. (1991). Robust estimation of the variance in moment methods for extra-binomial and extra-Poisson variation. Biometrics, 47: 383-401.

Nelson, K.P., Lipsitz, S.R., Fitzmaurice, G.M., Ibrahim, J., Parzen, M. & Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. Journal of Computational and Graphical Statistics, **15**(1): 39–57. DOI:10.1198/106186006X96854.

Pinheiro, J. C. & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics, **4**: 12–35.

Shmueli, G., Minka, T., Borle, J., & Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. J. R. Stat. Soc., Series C, **54**: 127-142.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between sets of trials. J. R. Statist. Sco., B **10**: 257-261.

Sokal, R.R. and Rohlf, F. J. (1969). Biometry: the Principles and Practice of Statistics in Biological Research. San Francisco: Freeman.

Williams, D. (1982). Extra-binomial variation in logistic linear model. Applied Statistics, **31**: 144–148.

Appendix

```
STATA (version 15)
==================

. melogit y z2 z3 z4 h || litter:, binomial(n) intpoints(32)

Mixed-effects logistic regression               Number of obs     =         58
Binomial variable:              n
Group variable:            litter               Number of groups  =         58

                                                Obs per group:
                                                              min =          1
                                                              avg =        1.0
                                                              max =          1

Integration method: mvaghermite                 Integration pts.  =         32

                                                Wald chi2(4)      =      67.05
Log likelihood = -91.644752                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z2 |  -3.761355    1.24467    -3.02   0.003    -6.200862   -1.321847
          z3 |  -4.756476   1.893747    -2.51   0.012    -8.468153   -1.044799
          z4 |  -3.822939   2.568047    -1.49   0.137    -8.856219    1.210341
           h |  -.1772736   .2447665    -0.72   0.469    -.6570072      .30246
       _cons |   2.625032   1.206566     2.18   0.030     .2602057    4.989858
-------------+----------------------------------------------------------------
litter       |
   var(_cons)|    2.30886   .8747942                      1.098726    4.851832
------------------------------------------------------------------------------
```

```
R APPLICATION
=============
rats <-read.table("C:/PAPER2019/rats.txt", header=T)
rats
attach(rats)
gg=as.factor(gp)
z2=ifelse(gg==2,1,0)
z3=ifelse(gg==3,1,0)
z4=ifelse(gg==4,1,0)
Z=data.frame(litter, h, n, r, z2,z3,z4)
u=n-r
library(lme4)
library(nlme)
fit1 <- glmer(cbind(r,u)~z2+z3+z4+h+(1|litter), family=binomial, data=Z,nAGQ=100)
```

```
summary(fit1)

> fit1 <- glmer(cbind(r,u)~z2+z3+z4+h+(1|litter), family=binomial, data=Z,nAGQ=100)
> summary(fit1)
Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite
  Quadrature, nAGQ = 100) [glmerMod]
 Family: binomial  ( logit )
Formula: cbind(r, u) ~ z2 + z3 + z4 + h + (1 | litter)
   Data: Z

     AIC      BIC   logLik deviance df.resid
   123.8    136.2    -55.9    111.8       52

Scaled residuals:
     Min       1Q   Median       3Q      Max
-0.99883 -0.36009 -0.04923  0.61632  0.93921

Random effects:
 Groups Name        Variance Std.Dev.
 litter (Intercept) 2.309    1.519
Number of obs: 58, groups:  litter, 58

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.6250     1.2066   2.176  0.02958 *
z2           -3.7614     1.2447  -3.022  0.00251 **
z3           -4.7565     1.8937  -2.512  0.01202 *
z4           -3.8230     2.5680  -1.489  0.13656
h            -0.1773     0.2448  -0.724  0.46892
---
Signif. codes:  0 **0.001 *0.01 0.05 0.1 1
```