Journal of Advances in Mathematics and Computer Science

23(3): 1-17, 2017; Article no.JAMCS.33475

Previously known as British Journal of Mathematics & Computer Science

ISSN: 2231-0851

Employing the Double, Multiplicative and The Com-Poisson Binomial Distributions for modeling Over and Under-dispersed Binary Data

Bayo H. Lawal[1*]

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

Author's contribution

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

Article Information

DOI: 10.9734/JAMCS/2017/33475

Editor(s):

(1) Sun-Yuan Hsieh, Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan.

Reviewers:

(1) Thomas L. Toulias, Technological Educational Institute of Athens, Greece.

(2) N.V.Krishna Prasad, Gandhi Institute of Technology and Management, India.

Complete Peer review History: http://www.sciencedomain.org/review-history/20015

Received: 17th April 2017

Accepted: 11th May 2017

Published: 12th July 2017

Original Research Article

Abstract

In this paper, we compare the performances of several models for fitting over-dispersed binary data. The distribution models considered in this study include the binomial (BN), the beta-binomial (BB), the multiplicative binomial (MBM), the Com-Poisson binomial (CPB) and the double binomial (DBM) models. Applications of these models to several well known data sets exhibiting under-dispersion and over-dispersion were considered in this paper. We applied these models to two frequency data sets and two data sets with covariates that have been variously analysed in the literature. The first relates to the Portuguese version of Duke Religiosity Index in a sample of 273 (202 women, 71 Male) postgraduate students of the faculty of Medicine of University of Sao Paulo. The second set that employs the Generalize Linear Model (GLM) is the correlated binary data which studies the cardiotoxic effects of doxorubicin chemoteraphy on the treatment of acute lymphoblastic leukemia in childhood. In the first data set, we have a single covariate, Sex (0,1) and two covariates in the second data set (dose and time).

Our results indicate that all the models considered here (excluding the binomial) behave reasonably well in modeling over-dispersed binary data with or without covariates, although both the multiplicative binomial and the double binomial models slightly behave better for these

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

specific data sets. While this result may not be necessarily generalized to other variety of over and under-dispersed data, we would however, encourage the investigation of all possible models so that the right applicable model can be employed for a given data set under consideration. All analyses were carried out using PROC NLMIXED in SAS.

Keywords: *Beta-binomial; double binomial; multiplicative binomial; Com-Poisson binomial; over-dispersion; under-dispersion.*

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

1 Introduction

Data with binary outcomes are very common and are widely encountered in many real world applications. The baseline model for binary data is of course the binomial distribution model.However, in many situations the binomial model fails to fit such data, simply because the variance of the observed often exceeds the expected variance of $n_i\hat{\pi}_i(1 - \hat{\pi}_i)$ under the binomial model, which consequently leads to over-dispersion in the binomial data.

To overcome this problems, several distributions have been utilized, especially mixture models such as the beta-binomial in [1], the Kumarasmawany in [2] and most recently the McDonald's generalized beta-distribution (McGBB) in [3]. Each of these distributions, separately models the probability of success π with beta, Kumaraswany, and exponential (continuous type) distributions respectively. The Beta-Binomial (BB) distribution has received considerable attention in the literature. While all these mixture distributions have been studied extensively in the literature, unfortunately, none of them can handle comprehensively all the complexities arising from various overdispersed binomial data.In other words no single mixture distributions fits all possible over-dispersed binary data. Thus, we continue to explore and investigate an alternative or even an already well established distribution in the pursuit of overcoming over-dispersion in binary data.

In this paper, we present both the multiplicative [4] [5], the Com-Poisson binomial [6] and the double binomial [7], [8] and [9] models as alternatives to the binomial model. We also compare our results to those obtained from the beta-binomial.

Our focus in this paper is based on a recent paper by [9] which compares the performances of the multiplicative binomial and the double binomial models to data arising from ink transmissions onto paper [9]. Because the Com-Poisson binomial distribution is also a member of the two-parameter exponential family, this model is therefore considered along with the multiplicative and double binomial models. SAS PROC NLMIXED is employed to estimate the parameters of these models after formulating their log-likelihoods.

We present in the following sections, brief descriptions of the models and their means and variances, together with their log likelihoods. Four example data sets are employed, two dealing with frequency counts and the other two data sets having single covariates. These data sets are appropriately described at the relevant sections of this paper.

2 The Binomial (BIN) Model

The random variable $Y = \sum y_i$, where $y_i \sim \text{Bernoulli}(p)$ has for a fixed n the binomial distribution:

$$f(y, \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}; \quad , y = 0, 1, \ldots, n, \quad 0 < \pi < 1. \tag{2.1}$$

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

$$E(Y) = n\pi, \tag{2.2a}$$
$$\mathrm{Var}(Y) = n\pi(1-\pi). \tag{2.2b}$$

The corresponding one-parameter exponential family representation of the binomial is given by:

$$f(y,\pi) = \binom{n}{y}(1-\pi)^n \times \exp\left[y \log \frac{\pi}{1-\pi}\right].$$

3 The Multiplicative Binomial (MBM) Model

In [5], an alternative form of the two-parameter exponential family generalization of the binomial distribution first introduced in [4] which itself was based on the original representation in [10] is given by,:

$$f(y) = \frac{\binom{n}{y}\boldsymbol{\psi}^y(1-\boldsymbol{\psi})^{n-y}\,\boldsymbol{\omega}^{y(n-y)}}{\sum_{j=0}^{n}\binom{n}{j}\boldsymbol{\psi}^j(1-\boldsymbol{\psi})^{n-j}\,\boldsymbol{\omega}^{j(n-j)}},\ y = 0, 1, \ldots, n. \tag{3.1}$$

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).

In [11], the author presented some elegant characteristics of the multiplicative binomial distribution, including its four central moments. The author's 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 presentation.

Following [11], the probability π of success for the Bernoulli trial, that is, $P(Y = 1)$ can be computed from the following expression in (3.2) as:

$$\pi_1 = \boldsymbol{\psi}\frac{\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})}, \tag{3.2}$$

where:

$$\kappa_{n-a}(\boldsymbol{\psi},\boldsymbol{\omega}) = \sum_{y=0}^{n-a}\binom{n-a}{y}\boldsymbol{\psi}^y(1-\boldsymbol{\psi})^{n-a-y}\,\boldsymbol{\omega}^{(y+a)(n-a-y)}. \tag{3.3}$$

with π defined as in (3.2), $\boldsymbol{\psi}$ 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.

We may note here that the relationship between the probability of success π defined in [9] in the multiplicative binomial is related to $\boldsymbol{\psi}$ with the expressions below. The mean and variance of the MBD are given respectively as:

$$E(Y) = n\pi_1, \tag{3.4a}$$

$$\text{Var}(Y) = n\pi_1 + n(n-1)\pi_2 - (n\pi_1)^2, \tag{3.4b}$$

where,

$$\pi_i = \boldsymbol{\psi}^i \frac{\kappa_{n-i}(\boldsymbol{\psi}, \boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi}, \boldsymbol{\omega})}. \tag{3.5}$$

and with $\kappa(.)$ as defined previously in (3.3). Thus, π_1 and π_2 are computed respectively as:

$$\pi_1 = \boldsymbol{\psi} \left[\frac{\kappa_{n-1}(\boldsymbol{\psi}, \boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi}, \boldsymbol{\omega})} \right] \quad \text{and} \quad \pi_2 = \boldsymbol{\psi}^2 \left[\frac{\kappa_{n-2}(\boldsymbol{\psi}, \boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi}, \boldsymbol{\omega})} \right], \tag{3.6}$$

and from (3.3), we have:

$$\begin{aligned}
\kappa_n(\boldsymbol{\psi}, \boldsymbol{\omega}) &= \sum_{y=0}^{n} \binom{n}{y} \boldsymbol{\psi}^y (1-\boldsymbol{\psi})^{n-y} \, \boldsymbol{\omega}^{y(n-y)}, \\
\kappa_{n-1}(\boldsymbol{\psi}, \boldsymbol{\omega}) &= \sum_{y=0}^{n-1} \binom{n-1}{y} \boldsymbol{\psi}^y (1-\boldsymbol{\psi})^{n-1-y} \, \boldsymbol{\omega}^{(y+1)(n-1-y)}, \\
\kappa_{n-2}(\boldsymbol{\psi}, \boldsymbol{\omega}) &= \sum_{y=0}^{n-2} \binom{n-2}{y} \boldsymbol{\psi}^y (1-\boldsymbol{\psi})^{n-2-y} \, \boldsymbol{\omega}^{(y+2)(n-2-y)}.
\end{aligned} \tag{3.7}$$

The corresponding two-parameter exponential family representation is also given by:

$$f(y|\psi, \omega) = \binom{n}{y} \frac{1}{\sum_{j=0}^{n} \binom{n}{j} \boldsymbol{\psi}^j (1-\boldsymbol{\psi})^j \, \boldsymbol{\omega}^{j(n-j)}} \times \exp\left(y \log \frac{\boldsymbol{\psi}}{1-\boldsymbol{\psi}} + (n-y) y \log \boldsymbol{\omega} \right). \tag{3.8}$$

4 The Com-Poisson Binomial (CPB) Model

The probability density function for the Com-Poisson Binomial distribution is given by:

$$f(y|n, p, \nu) = \frac{\binom{n}{y}^{\nu} \pi^y (1-\pi)^{n-y}}{\sum_{k=0}^{n} \binom{n}{k}^{\nu} \pi^k (1-\pi)^{n-k}}, \quad y = 0, 1, \ldots, n, \tag{4.1}$$

With $\pi \in (0, 1)$ and $\nu \in \mathbb{R}$. 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 [12] is given in (4.2),

$$f(y_i) = \frac{\lambda_i^{y_i}}{(y_i!)^{\nu}} \frac{1}{Z(\lambda_i, \nu)}, \quad y_i = 0, 1, 2, \cdots, \quad \lambda_i > 0, \ \nu \geq 0. \tag{4.2}$$

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

$$Z(\lambda_i, \nu) = \sum_{j=0}^{\infty} \frac{\lambda_i^j}{(j!)^\nu}. \tag{4.3}$$

An approximation to the CPB distribution in the limit $n \to \infty$ and with $\lambda = n^\nu p$ is given in [13]. Following [6], if we let θ be defined as:

$$\theta = \frac{\pi}{1-\pi}. \tag{4.4}$$

and dividing both the denominator and numerator of the expression in (4.1) by a factor of $(1-\pi)^m(m!)^\nu$, we thus have:

$$f(y|n,\theta,\nu) = \frac{\theta^y}{(y!)^\nu}\frac{1}{Z(\theta,\nu)}, \quad y = 0,1,2,\cdots, \quad \theta > 0,\ \nu \geq 0, \tag{4.5}$$

where the the normalizing term is defined as:

$$Z(\theta, \nu) = \sum_{j=0}^{n} \frac{\theta^j}{[j!(n-j)!]^\nu}. \tag{4.6}$$

The various properties of the CPB or the Com-Poisson have been presented in various papers [6], [13], and [14] applied the CPB to the number of killings in rural Norway.

For the analysis of the data sets in our examples in this paper, we shall employ SAS PROC NLMIXED to implement the models discussed in this paper. PROC NLMIXED uses several optimization techniques in its computations. We have adopted the dual Quasi-Newton and conjugate-gradient techniques for our computation. The chosen method of integral approximations of the marginal likelihood is the adaptive Gaussian quadrature as defined in [15].

The means and variance of Y_i are respectively given as:

$$E(Y) = \sum_{j=0}^{n} \frac{j\,\theta^j}{Z(\theta,\nu)[j!(n-j)!]^\nu}, \quad \text{and} \tag{4.7a}$$

$$\text{Var}(Y) = \sum_{j=0}^{n} \frac{j^2\,\theta^j}{Z(\theta,\nu)[j!(n-j)!]^\nu} - [E(Y)]^2. \tag{4.7b}$$

The two-parameter exponential family representation of the distribution is presented in (4.8).

$$f(y|n,\pi,\nu) = \binom{n}{y}^\nu \frac{1}{\sum_{k=0}^{n}\binom{n}{k}^\nu\left(\frac{\pi}{1-\pi}\right)^k} \times \exp\left(y \log\frac{\pi}{1-\pi}\right). \tag{4.8}$$

5 The Double Binomial (DBM) Model

In [9], the double binomial distribution was presented, having the pdf form:

$$f(y;\pi,\phi)=\frac{\binom{n}{y}\left[y^y(n-y)^{n-y}\right]^{1-\phi}\left[\pi/(1-\pi)\right]^{y\phi}}{\sum_{j=0}^{n}\binom{n}{j}\left[j^j(n-j)^{n-j}\right]^{1-\phi}\left[\pi/(1-\pi)\right]^{j\phi}},\ y=0,1,\ldots,n. \tag{5.1}$$

and following [9], the double binomial can be written as a two parameter exponential family distribution in the form:

$$f(y;\pi,\phi)=\binom{n}{y}y^y(n-y)^{n-y}\frac{1}{\sum_{j=0}^{n}\binom{n}{j}(j^j(n-j)^{n-j})^{1-\phi}(\pi/(1-\pi))^{j\phi}}$$
$$\times\exp\left(-[y\log(y+(n-y)\log(n-y)]\phi+y\phi\log\frac{\pi}{1-\pi}\right).$$

We see that the expression above factorizes appropriately.

6 The Beta-Binomial (BB) Model

The beta-binomial in [1] is, of course, a mixture of the binomial $\text{Bin}(n,\pi)$ and the beta distribution $\text{Beta}(\alpha,\beta)$, where,

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

That is, $\text{Bin}(n,\pi)\wedge\text{Beta}(\alpha,\beta)\sim BB$, with resulting unconditional pdf presented in (6.1).

$$f(y;\alpha,\beta)=\binom{n}{y}\frac{B(\alpha+y,\beta+n-y)}{B(\alpha,\beta)},\quad y=0,1,\ldots,n. \tag{6.1}$$

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

$$E(Y)=n\pi\qquad\text{and}\qquad\text{Var}(Y)=n\pi(1-\pi)[1+\rho^2(n-1)]. \tag{6.2}$$

where

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

7 Estimation

For a single observation, the log-likelihoods for the binomial, the multiplicative binomial, the Com-Poisson binomial, the double binomial and the beta-binomial are displayed in expressions (7.1a) to

(7.1e) respectively.

$$LL1 = \log \binom{n}{y} + y\log(\pi) + (n-y)\log(1-\pi) \tag{7.1a}$$

$$LL2 = \log \binom{n}{y} + y\log(\boldsymbol{\psi}) + y(n-y)\log \boldsymbol{\omega} - \log\left[\sum_{j=0}^{n}\binom{n}{j}\boldsymbol{\psi}^j(1-\boldsymbol{\psi})^{n-j}\,\boldsymbol{\omega}^{j(n-j)}\right] \tag{7.1b}$$

$$LL3 = y_i\log\theta_i - \nu\log[y_i!(m-y_i)!] - \log Z(\theta_i,\nu) \tag{7.1c}$$

$$\begin{aligned} LL4 = \log \binom{n}{y} &+ (1-\phi)[y\log y + (n-y)\log(n-y)] + y\phi\log\left(\frac{\pi}{1-\pi}\right) \\ &- \log\left[\sum_{j=0}^{n}\binom{n}{j}\left(j^j(n-j)^{n-j}\right)^{1-\phi}\left(\frac{\pi}{1-\pi}\right)^{j\phi}\right] \end{aligned} \tag{7.1d}$$

$$LL5 = \log \binom{n}{y} + \log[B(\alpha+y,\beta+n-y)] - \log[B(\alpha,\beta)] \tag{7.1e}$$

Maximum-likelihood estimations of the above models are carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y,\Theta)$ over the parameter space Θ numerically. The integral approximations in PROC NLMIXED is the Adaptive Gaussian Quadrature [15] and several optimization algorithms: namely:the quasi-Newton algorithm (QUANEW), the Nelder-Mead Simplex method(NMSIMP), the Newton-Raphson method with line search (NEWRAP) and the Conjugate Gradient method (CONGRA) of [16] [17]. Convergence is often a major problem here and the choice of starting values is very crucial. For each of the cases considered here, the above four optimizing algorithms were applied in turn to ascertain accuracy and consistency.

For the double binomial however, while the parameters are estimated via PROC NLMIXED in SAS with the above optimization and integration techniques, the corresponding estimated probabilities are estimated using the function **ddoublebinom(0, 36, 0.9968, 0.0187)** in package **rmutil** [18] in R.

8 Applications

We apply the models discussed above to the frequency of males in 6115 families with 12 children in Sax-ony, previously analyzed in [19]. The data is originally from Geissler [20] and had similarly been analyzed in [6]. The data is presented in Table 1.

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
count	3	24	104	286	670	1033	1343	1112	829	478	181	45	7

Here $Y \sim \text{binomial}(12,\pi)$. The frequencies are presented as counts having a total sum of 6115.

8.1 Results

We present in Table 2, the parameter estimates under the binomial (BN), the beta-binomial (BB), the multiplicative binomial (MBM), the Com-Poisson binomial (CPB) and the double binomial (DBM) models.

Table 2. Parameter estimates under the five Models

	Models				
	BIN	BB	MBM	CPB	DBM
	$\hat{\pi} = 0.5192$	$\hat{\pi} = 0.5192$	$\hat{\psi} = 0.5165$	$\hat{\theta} = 1.0682$	$\hat{\pi} = 0.5191$
		$\hat{\rho} = 0.0150$	$\hat{\omega} = 0.9742$	$\hat{\nu} = 0.8434$	$\hat{\phi} = 0.8602$
-2LL	25068.34	24986	24986	24985	24824
AIC	25070.34	24990	24990	24989	24828
Mean	6.2304	6.2306	6.2306	6.2306	6.2297
Var	2.9956	3.4897	3.4893	3.4918	3.4878
X_W^2	7122.8133	6114.19	6115.0059	6110.4869	6117.5773
d.f	6114	6113	6113	6113	6113

The BB parameters are estimated from the log-likelihood formulation in (7.1e) and the estimated probability here is given by, $\hat{\alpha}/(\hat{\alpha} + \hat{\beta}) = 0.5192$ and the intra-class correlation ρ^2 is estimated to be $0.1225 = 1/(\hat{\alpha} + \hat{\beta} + 1)$. For the MBM the estimate probabilities are $\hat{\pi}_1 = 0.5192$ and $\hat{\pi}_2 = 0.2733$. The X_W^2 is the Wald test statistic:

$$X^2 = \sum_{i=0}^{N} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}_i^2}; \quad , N = 6115. \tag{8.1}$$

The $\hat{m}_i$ in Table 2, as above is given by the mean=$n\hat{\pi}$. The estimated variances are also presented as Var. For the observed data in Table 1, $\bar{y} = 6.2306$ and $s^2 = 3.4898$. We see from Table 2, that while the five models estimate the mean of the data well, the estimated variance under the binomial model of 2.9956 underestimates the observed variance of the data, and this explains the poor fit to the data as exhibited by the binomial model. On the other hand, for the other four models, the variance of the observed data are reasonably well estimated. The Wald's Goodness-Of-Fit (GOF) test statistic suggests that the Com-Poisson binomial model fits data best, but again, they all fit the data well.

In Table 3 are the expected values under each of the five models, the corresponding Pearson‘s X^2 and the corresponding degrees of freedom (d.f.). Clearly, for this data set, apart from the binomial model, all the other four models fit the data, but the double binomial is the most parsimonious in this case, doing slightly better than the Com-Poisson binomial model.

Table 3. Expected Values under the five models and corresponding Pearson's X^2 Statistic Values

Y	count	BN	BB	MPD	COMP	DB
0	3	0.9328	2.3487	2.3486	2.6797	2.9390
1	24	12.0888	22.5746	22.5809	23.2758	23.3191
2	104	71.8032	104.8238	104.8482	104.7036	104.1809
3	286	258.4751	310.8757	310.8921	308.7432	307.6766
4	670	628.0550	655.7208	655.6551	653.5383	653.1269
5	1033	1085.2107	1036.2201	1036.0769	1037.6988	1039.2163
6	1343	1367.2794	1257.9632	1257.9074	1262.3570	1265.0121
7	1112	1265.6303	1182.1426	1182.2927	1184.0487	1185.3592
8	829	854.2466	853.5574	853.7711	850.8785	849.7391
9	478	410.0126	461.9057	461.9646	458.6614	456.5902
10	181	132.8357	177.8755	177.7841	177.4821	176.3454
11	45	26.0825	43.7809	43.6925	45.0189	45.0228
12	7	2.3473	5.2109	5.1858	5.9140	6.4723
Total	6115	6115	6115	6115	6115	6115
X^2		110.5051	14.4692	14.5354	13.3597	13.0457
d.f		11	10	10	10	10

9 Example II

The data in Table 4 is from [9] and relate to the counts of blocks with Y successfully printed pixels from sample 201 (B).

Table 4. Counts of blocks with y successfully printed pixels from sample 201 (B)-[9]

y	count	y	count	y	count	y	count
0	204	10	265	20	686	30	2257
1	121	11	296	21	728	31	2713
2	132	12	345	22	865	32	3239
3	155	13	355	23	880	33	4022
4	144	14	382	24	1064	34	5551
5	186	15	455	25	1267	35	5999
6	216	16	492	26	1242	36	219,358
7	169	17	502	27	1459		
8	254	18	586	28	1753		
9	240	19	592	29	1947		

It is assumed here that $n = 36$, and $y = 0, 1, 2, \ldots, n$ with probability π. Here, the total sample size $N = 261, 121$ and are as distributed in Table 2 We also assume here that Y has an underlying binomial distribution with parameters $n = 36$ and success probability π, that is, $y \sim \text{binomial}(36, \pi)$.

In Table 5, we present the parameter estimates under the five models, together with their corresponding estimated means and variances and Wald's GOF test statistic.

Table 5. Parameter estimates under the five Models

	Models				
	BIN	BB	MBM	CPB	DBM
	$\hat{\pi} = 0.9639$	$\hat{\pi} = 0.9648$	$\hat{\psi} = 0.5303$	$\hat{\theta} = 1.1237$	$\hat{\pi} = 0.9968$
	-	$\hat{\rho} = 0.6562$	$\hat{\omega} = 0.8949$	$\hat{\nu} = -0.2917$	$\hat{\phi} = 0.0187$
-2LL	1,840,559	483,556	911,827	632,168	248,647
AIC	1,840,561	483,560	911,831	632172	248,651
mean	34.6987	34.7328	34.6987	34.6987	30.2627
Var	1.2543	19.6483	18.7012	22.6995	59.0810
X^2_W	3,893,291	248,458	261,121.48	215,126.94	169, 625.31
d.f.	212,121	212,120	212,120	212,120	212,120

The estimated π's under the binomial and the double binomial for example are respectively, 0.9639 and 0.9968. For the binomial for instance, the mean=$36 \times 0.9639 = 34.6987$ and the variance is estimated as $36 \times 0.9639 \times (1 - 0.9639) = 1.2543$. However, for the observed data in Table 4, $\bar{y} = 34.6987$ and $s^2 = 18.7013$. We see again from Table 5, that while the five models estimate the mean of the data well, the estimated variance under the binomial model of 1.2543 grossly underestimates the observed variance of the data, and this again explains the very poor fit to the data as exhibited by the binomial model. On the other hand, the beta-binomial and the multiplicative binomial reasonably estimated the observed data variance of this data set well. The Com-Poisson estimate is also not too far from the 18.7013 but the double binomial over-estimates the variance of the observed data. The Wald's GOF test statistic seems to fit best in the double binomial model. The Akaike Information Criterion (AIC) and -2LL values also support the double binomial asthe model providing the best fit for this data set. The estimated probabilities under the multiplicative model are respectively, $\hat{\pi}_1 = 0.9638$ and $\hat{\pi}_2 = 0.9429$.

In Table 6 are presented the expected values under each of the five models. The beta-binomial fits this data best when the data is aggregated over the various values of y, that is, grouped.

Table 6. Expected Values under the five models and corresponding X^2 Statistic Values

Y	count	BN	BB	MPD	CPB	DBM
0	204	0.0000	114.5162	1595.5381	2490.1177	1689.5201
1	121	0.0000	150.0848	1329.0081	983.8881	763.9615
2	132	0.0000	175.5709	671.9961	479.7792	673.1709
3	155	0.0000	197.2481	274.7949	265.5642	660.7565
4	144	0.0000	217.1259	102.1481	161.2576	678.8770
5	186	0.0000	236.1422	36.7842	105.4470	714.9485
6	216	0.0000	254.8381	13.3538	73.3948	764.4623
7	169	0.0000	273.5765	5.0216	53.9480	825.7038
8	254	0.0000	292.6319	1.9946	41.6385	898.1945
9	240	0.0000	312.2343	0.8491	33.6031	982.1153
10	265	0.0000	332.5931	0.3917	28.2629	1078.0635
11	296	0.0000	353.9129	0.1976	24.7119	1186.9435
12	345	0.0000	376.4046	0.1097	22.4175	1309.9198
13	355	0.0000	400.2950	0.0674	21.0658	1448.4024
14	382	0.0000	425.8357	0.0460	20.4808	1604.0517
15	455	0.0000	453.3135	0.0350	20.5819	1778.7984
16	492	0.0000	483.0617	0.0298	21.3645	1974.8754
17	502	0.0000	515.4747	0.0283	22.8959	2194.8642
18	586	0.0000	551.0262	0.0302	25.3258	2441.7548
19	592	0.0000	590.2943	0.0361	28.9111	2719.0277
20	686	0.0000	633.9940	0.0484	34.0646	3030.7609
21	728	0.0000	683.0242	0.0726	41.4384	3381.7747
22	865	0.0000	738.5321	0.1216	52.0678	3777.8291
23	880	0.0000	802.0095	0.2271	67.6248	4225.8995
24	1064	0.0007	875.4339	0.4714	90.8702	4734.5709
25	1267	0.0086	961.4874	1.0828	126.4873	5314.6165
26	1242	0.0971	1063.9042	2.7375	182.6678	5979.8765
27	1459	0.9585	1188.0473	7.5662	274.2401	6748.6415
28	1753	8.2153	1341.9109	22.6634	429.0942	7645.9291
29	1947	60.4299	1537.9704	72.7552	702.0030	8707.4329
30	2257	375.9784	1796.8532	246.7011	1205.9636	9986.8351
31	2713	1940.3788	2155.3386	866.5108	2187.8086	11570.5117
32	3239	80,84.2671	2686.1279	3,068.2483	4,224.7488	13,610.5244
33	4022	26,128.8818	3,556.2415	10,524.8583	8785.2897	16,410.8568
34	5551	61,474.8218	52,55.2783	32,818.7076	20,041.6599	20,711.9560
35	5999	93,668.4528	10,122.0242	82,761.7158	51,897.1772	29,118.7833
36	219,358	69,378.5091	219,016.6420	126,694.051	165,853.1380	79,775.7891
X^2		$> 1.2E50$	2,870.6458	82,519,387.7	237,058.033	351,757.251

The corresponding degrees of freedom are 35 d.f. for the binomial and 34 d.f. for the other four models. Clearly, none of the models fit this data.

10 Generalized Linear Model Application

In this section, we would employ the BN (logistic model), MBM, CPB and the DBM models to data having covariates. For data having covariates $(x_1, x_2, \ldots, x_p)'$, the probability of success π can be modeled as:

$$\pi_{ij} = \frac{1}{1 + \exp(-\mathbf{x}'\mathbf{b})},$$

where $(b_0, b_1, b_2, \ldots, b_p)'$ are parameter estimates to be estimated. We apply the five models to two sets of binary response data having covariates. These are presented in the next section.

10.1 Example data I: The intrinsic religiosity index

The data in Table 7 is reproduced from [21] and relates to Portuguese version of Duke Religion Index in a sample of 273 (202 women, 71 Male) postgraduate students of the faculty of Medicine of University of Sao Paulo. The index is a five-item measure of religious involvement; for details, see [22]. The maximum number of points on the scale is $n = 18$ and the counts in the data are the number of points scored by Women and Men. The means and variances of men and women are respectively, Men=$\{\bar{y}_m = 13.5352,\ s_m^2 = 13.1095\}$; women=$\{\bar{y}_w = 15.4158,\ s_w^2 = 6.3635\}$.

Table 7. Counts of Points in IR sub-scale for women and Men individuals

	Number of Points (r)												
Sex	6	7	8	9	10	11	12	13	14	15	16	17	18
Women	3	2	3	1	3	2	5	12	21	24	55	31	40
Men	4	3	4	2	4	1	4	5	8	8	12	10	6
Total	7	5	7	3	7	3	9	17	29	32	67	41	46

The five models discussed in the previous sections are now applied to this data set, with *sex* (1 for women, 0 for men) as the single covariate.

10.2 Results

For the multiplicative model we employ the log-link for parameter ψ such that $0 < \psi < 1$ and the association parameter is modeled with a log-link such that $\log(\omega) = a_0$. For the Com-Poisson binomial, we model the dispersion parameter ν similarly with a log-link. For the double-binomial model, both the probability of success π and the dispersion parameter ϕ are modeled respectively with the logit and the log links. The model of interest is then given by:

$$\begin{aligned} \pi_{ij} &= \Pr[\mathbf{Y}_{ij} = 1|\text{Sex}_j], \\ \log\left(\frac{r_{ij}}{n - r_{ij}}\right) &= \frac{e^{\beta_0+\beta_1 \text{Sex}_j}}{1 + e^{\beta_0+\beta_1 \text{Sex}_j}}; i = 6, 7, \ldots, 18; j = 0, 1. \end{aligned} \tag{10.1}$$

where, $n = 18$ and,

$$\text{Sex} = \begin{cases} 1 & \text{if Women} \\ 0 & \text{if Men} \end{cases}$$

The results of our analysis are presented in Table 8. Based on Wald's test statistic, we observe that the multiplicative binomial (MBM) model gives the best fit with $X_W^2 = 226.3633$ on 273-3=270 degrees of freedom as compared with 851.268 on 271 d.f. from the baseline binomial model.

Table 8. Parameter estimates and GOF X^2 under various Models

Parameters	BN	BB	MBM	CPB	DBM
$\hat{\beta}_0$	1.1091	1.1503	0.5043	1.2136	1.6962
$\hat{\beta}_1$	0.6769	0.5942	0.0497	0.2933	2.2501
$\hat{a}_0$	na	-0.6757	0.8681	0.0852	-2.4040
$\hat{\pi}_M$	0.7520	0.7596	0.7346	0.7709	0.8450
$\hat{\pi}_W$	0.8564	0.8513	0.8547	0.8185	0.9810
Mean-M	13.536	13.6721	13.2223	13.535	13.1231
Var - M	3.357	9.6424	17.3143	10.9841	10.3415
Mean-W	15.415	15.3227	15.3595	15.4158	14.7728
Var - W	2.2136	6.6851	7.3916	6.4278	6.0105
Disp.	na	$\hat{\rho} = 0.1137$	$\hat{\omega} = 0.8681$	$\hat{\nu} = 0.8681$	$\hat{\phi} = 0.0899$
X_w^2	851.268	286.901	226.3633	282.534	316.392
d.f	271	270	270	270	270
-2LL	1465.40	1209.1	1178.8	1198.4	954.9
AIC	1469.4	1215.1	1184.8	1204.4	960.9

In terms of the -2LL and AIC as measures of fit, the double-binomial seems to be the best and is followed by the multiplicative binomial. The Com-Poisson binomial performs better than the beta-binomial in this example with GOF X^2_W of 282.534 and 286.901 respectively. Thus based on this data, we would probably recommend the multiplicative binomial model. Based on this model, we must note that the parameter $\boldsymbol{\psi}$ in the multiplicative binomial probability distribution in (3.8) is not the success probability. For our data, under the MBM, $\hat{\boldsymbol{\psi}}_M = 0.5043$ and $\hat{\boldsymbol{\psi}}_W = 0.5541$, with $\hat{\boldsymbol{\omega}} = 0.8681$. Consequently, using expressions in (3.5) and (3.6), we have $\hat{\pi}_M =$

$$\pi_1 = \boldsymbol{\psi}\left[\frac{\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})}\right].$$

For this data, $\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})$ and $\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})$ are computed as 0.000206331 and 0.000133972 respectively for Women and 0.000048785 and 0.000071058 for men. Thus, from (3.5), we have,

$$\hat{\pi}_1^M = \boldsymbol{\psi}_M \frac{\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})} = \frac{0.50432 \times 0.000048785}{0.000071058} = 0.73457,$$
$$\hat{\pi}_1^W = \boldsymbol{\psi}_W \frac{\kappa_{n-1}(\boldsymbol{\psi},\boldsymbol{\omega})}{\kappa_n(\boldsymbol{\psi},\boldsymbol{\omega})} = \frac{0.55406 \times 0.000206331}{0.000133972} = 0.85331.$$

Similarly, with $\kappa_{n-2}(\boldsymbol{\psi},\boldsymbol{\omega}) = \{0.000325101, 0.000112154\}$ for women and men respectively. It is not too difficult therefore to compute $\hat{\pi}_2$ from (3.6) as: $\hat{\pi}_2^M = 0.58471$ and $\hat{\pi}_2^W = 0.74492$. Consequently, the means and variances are computed using expressions in (3.4) and these values are displayed in Table 8.

10.3 Example data II

This example is from [23] and was originally published 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 9.

In this study, 24 patients previously cured of leukemia had a long-term followup visit to determine how the heart was functioning. Tests of heart functions were conducted. For each subject on a visit, there are six similar tests of heart function 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). Here, 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:

$$\begin{aligned} \pi_i &= \Pr[\mathbf{Y_{ij}} = \mathbf{1}|\mathrm{DOSE_i}, \mathrm{TIME_i}], \\ &= \frac{e^{\beta_0+\beta_1 \mathrm{DOSE}_i+\beta_2 \mathrm{TIME}_i}}{1+e^{\beta_0+\beta_1 \mathrm{DOSE}_i+\beta_2 \mathrm{TIME}_i}}, \end{aligned} \tag{10.2}$$

Table 9. 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

where DOSE is 1 if High and 0 if low, and TIME_i is the time in years since the last chemotherapy.

Model (10.1) therefore becomes:

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}'\boldsymbol{\beta}, \quad \text{that is,} \qquad \text{logit}_i = \beta_0 + \beta_1 \,\text{DOSE}_i + \beta_2 \,\text{TIME}_i. \tag{10.3}$$

Our formulation of the above model is based on the fact that there is no significant interaction between dose and time [23].

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 π, $\boldsymbol{\psi}$ and θ and π in the Beta binomial, the multiplicative, Com-Poisson binomial and the double binomial are modeled with the logit-link function. For the CPB for instance $\theta = \log\left(\frac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})}\right)$, where $\mathbf{x}'\boldsymbol{\beta}$ is as defined in (10.3).

Table 10. Results of analyses for the models

Parameters	BN	BB	MBM	CPB	DBM
β_0	-0.2202	-0.2971	-0.1460	0.8749	1.2071
β_1	0.9631	0.9613	0.6689	0.4521	2.5007
β_2	-0.0638	-0.0589	-0.0447	-0.0310	-0.2626
	na	$\hat{\rho} = 0.1139$	$\hat{\omega} = 0.8698$	$\hat{\nu} = 0.5184$	$\hat{\phi} = 0.3482$
-2LL	81.8958	78.9477	79.6	79.7	55.0
AIC	87.8957	86.9473	87.6	87.7	63.8
X_W^2	35.4218	23.9024	25.3575	24.3268	39.0086
d.f.	21	20	20	20	20

The results of applying these models to the cardiotoxicity Study data in Table 9 are presented in Table 10. When the binomial model was applied to the data, $X_W^2 = 35.4218$ on 21 d.f. giving an estimated dispersion parameter (DP) of 1.8668 indicating a strong overdispersion of the data. Under the MBM, CPB and the DBM models, the estimated dispersion parameters are $\hat{\omega} = 0.8698$, $\hat{\nu} = 0.5184$ and $\hat{\phi} = 0.3482$ indicating an over-dispersion in the data. Clearly, based on the -2LL and AIC statistics, the double binomial fits best with respective values 55.0 and 63.8. However, the Wald test statistic under the DB model gives a value of 39.0086. In Table 11 we present the estimated probabilities, $\hat{\phi}$ values, the expected values of each observation as well as the corresponding variances under the double binomial model. In the last column, the cumulative values of the Wald test statistic are given. Recall that the Wald test statistic is defined as:

$$X_W^2 = \sum_{i=1}^{N} \frac{(r_i - \hat{m}_i)^2}{\text{var}}, \quad i = 1, 2, \ldots, N(= 24).$$

The Wald GOF is very susceptible in this case when $r = 0$. For instance, note that for cluster 2 and $r = 0$, the contribution towards Wald's GOF here is 5.1626-0.4747=4.6879. Similarly, for clusters 5,8 and 23, the spikes in the GOF are respectively, 4.6468, 5.3668, 4.1147. The four observations alone contribute a total of 18.8162 towards X_W^2 compared to say contributions of 8.7596 for the four cells under the MBM. The double binomial tends to underestimate the variances for each cluster and more especially for the cases when $r = 0$. However, the Pearson's X^2 and the likelihood-ratio test G^2 are 20.2674 and 11.5774 under the double binomial model. These values fit much better than those of the other models. 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.

Table 11. Computations of some values under the DBM

Cluster	$\hat{\pi}$	$\hat{\phi}$	r	n	dose	time	$\hat{m}$	s^2	G^2	X^2	X_W^2
1	0.52750	0.34823	4	6	1	13.7	3.06993	1.82222	2.1171	0.2818	0.4747
2	0.40399	0.34823	0	5	1	15.6	2.34031	1.16834	2.1171	2.6221	5.1626
3	0.92413	0.34823	3	5	1	4.6	3.37324	0.71392	1.4135	2.6634	5.3578
4	0.57295	0.34823	4	5	1	13.0	2.62094	1.17534	4.7956	3.3890	6.9758
5	0.39625	0.34823	0	5	0	6.2	2.32721	1.16554	4.7956	5.7162	11.6226
6	0.41670	0.34823	1	6	1	15.4	2.78729	1.79826	2.7455	6.8623	13.3990
7	0.37757	0.34823	2	5	0	6.5	2.29534	1.15780	2.1945	6.9003	13.4743
8	0.51290	0.34823	0	5	0	4.4	2.52128	1.18448	2.1945	9.4216	18.8411
9	0.21183	0.34823	1	5	0	9.6	1.98487	1.01588	0.8234	9.9103	19.7959
10	0.68279	0.34823	3	5	1	11.2	2.81067	1.12279	1.2146	9.9230	19.8278
11	0.28495	0.34823	3	5	0	8.1	2.13000	1.09704	3.2695	10.2784	20.5178
12	0.56652	0.34823	3	5	1	13.1	2.61018	1.17694	4.1047	10.3366	20.6469
13	0.19073	0.34823	1	5	0	10.1	1.93912	0.98509	2.7802	10.7914	21.5422
14	0.26917	0.34823	4	6	0	8.4	2.39138	1.60785	6.8956	11.8735	23.1516
15	0.52600	0.34823	1	5	0	4.2	2.54294	1.18358	5.0289	12.8097	25.1630
16	0.54056	0.34823	1	5	1	13.5	2.56703	1.18187	3.1434	13.7662	27.2407
17	0.27035	0.34823	1	5	1	17.9	2.10236	1.08356	1.6573	14.3443	28.3622
18	0.24902	0.34823	1	5	0	8.8	2.06089	1.06157	0.2110	14.8904	29.4224
19	0.41525	0.34823	2	6	0	5.9	2.78355	1.79731	-1.1113	15.1109	29.7640
20	0.56006	0.34823	3	5	1	13.2	2.59941	1.17840	-0.2513	15.1727	29.9002
21	0.47502	0.34823	4	5	1	14.5	2.45876	1.18368	3.6418	16.1388	31.9070
22	0.28495	0.34823	4	6	0	8.1	2.43591	1.63810	7.6096	17.1431	33.4004
23	0.27963	0.34823	0	5	0	8.2	2.11999	1.09227	7.6096	19.2631	37.5151
24	0.28495	0.34823	4	6	0	8.1	2.43591	1.63810	11.5774	20.2674	39.0086

The values of G^2 and X^2 columns in Table 11 are respectively, the likelihood ratio test statistic and the Pearson's test statistic defined as:

$$G^2 = 2\sum_{i=1}^{N} r_i \log\left(\frac{r_i}{\hat{m}_i}\right), \quad X^2 = \sum_{i=1}^{N} \frac{(r_i - \hat{m}_i)^2}{\hat{m}_i}.$$

11 Conclusions

Clearly, each of the models under investigation in this paper (excluding the binomial) behave reasonably well in modeling over-dispersed binary data with or without covariates. While the MBM is most parsimonious for the religiosity index data, the double binomial fits best the cardiotoxicity data. It is therefore obvious that no single model fits all possible data and it is always advisable to explore all possible models in other to come to an informed conclusion.

It should be noted here that there are other approaches for overcoming overdispersed or under-dispersed binary data. Some of these are (i)the quasi-likelihood approach resulting in either scaling (via X^2 or deviance-QL(1)) or employing William's (QL2) approach. Alternatively, we could employ the GLMM method with the normal-binomial model and since the data are in clusters, we could also employ the generalized estimating equations (GEE) in fitting especially the two data sets having co-variates. We present the results of applying these approaches, for instance to the data in Table 4. These results are presented in the appendix.

Acknowledgement

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

Competing Interests

Author has declared that no competing interests exist.

References

[1] Skellam JG. 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. 1948;B10:257-261.

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

[3] Manjor C, Wijekoom P, Yapa RD. 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. 2013;2(2):24-41.

[4] Altham PM. Two generalizations of the binomial distribution. Appl. Statist. 1978;27:162-167.

[5] Lovison G. An alternative representation of Altham's multiplicative-binomial distribution. Statistics & Probability letters. 1998;36:415-420.

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

[7] Efron B. Double exponential families and their use in generalized linear regression. J. Amer. Statist. Assoc. 1986;81:709-721.

[8] Lindsey JK, Altham PK. Varieties of over- and underdispersion models for binary data; 1982.

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

[10] Cox DR. The analysis of multivariate binary data. Applied Statistics. 1972;21: 113-120.

[11] Elamir EAH. Multiplicative-binomial distribution: Some results and characterization, inference and random data generation. Journal of Statistical Theory and Applications. 2013;12(1):92-105.

[12] Conway RW, Maxwell WL. A queuing model with state dependent service rates. Journal of Industrial Engineering. 1961;12:132-136, 363.

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

[14] Kadane J, Shmueli G, Minka G, Borle T, Boatwright P. Conjugate analysis of the conway maxwell poisson distribution. Bayesian Analysis. 2006;1:363-374.

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

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

[17] Powell MJD. Restart Procedures for the Conjugate Gradient Method. Mathematical Programming. 1977;12:241-254.

[18] Lindsey J. R codes; 2015.

[19] Sokal RR, Rohlf FJ. Biometry: The principles and practice of statistics in biological research. San Francisco: Freeman; 1969.

[20] Geissler A. Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. Z. Köngl. Sächs. Statist. Bur. 1889;35:1-24.

[21] Martinez EZ, Achcar JA, Aragon DC. Parameter estimation of the beta-binomial distribution: An application using the SAS software. Ciência e Natura, Santa Maria. 2015;37(4):12-19.

[22] Martinez EZ, Santos-Almeida RG, Carvalho ACD. Propriedades da escala de religiosidade de Duke em uma amostra de pós-graduandos. Revista de Psiquiatria Clinica. 2012;39(5):180.

[23] Nelson KP, Lipsitz SR, Fitzmaurice GM, Ibrahim J, Parzen M, Strawderman R. Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. Journal of Computational and Graphical Statistics. 2006;15(1):39-57.
DOI: 10.1198/106186006X96854

Appendix

Results of applying these procedures to a typical data presented in Table 5.

Parameters	BIN	Q(1)	Q(2)	GLMM	GEE
β_0	1.1091	1.1091	1.1091	1.0473	0.7016
	(0.0648)	(0.1148)	(0.1421)	(0.4598)	(0.2617)
β_1	0.6769	0.6769	0.6769	0.1234	0.0089
	(0.0802)	(0.1421)	(0.1421)	(0.6480)	(0.3700)
disp	na	1.7723	$\hat{\rho} = 0.1259$	$\hat{\sigma}^2 = 2.6116$	$\hat{\rho} = 0.3570$
π_M	0.7520	0.7520	0.7520	0.7403	0.6685
π_W	0.8564	0.8564	0.8564	0.7633	0.6705
-2LL	1465.403	1465.403	1465.403	739.2	na

Standard errors are in parentheses.

Clearly here the Generalized linear mixed model utilizing the normal-binomial model fits best based on the -2LL.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/20015