

On Some Mixture Models for Over-dispersed Binary Data

Bayo H. Lawal¹

¹ Department of Statistics & Mathematical Sciences, Kwara State University, Malete, Nigeria

Correspondence: Bayo H. Lawal, Department of Statistics & Mathematical Sciences, Kwara State University, Malete, Nigeria. E-mail: bayo.lawal@kwasu.edu.ng

Received: December 19, 2016 Accepted: January 10, 2017 Online Published: February 27, 2017

doi:10.5539/ijsp.v6n2p134 URL: <https://doi.org/10.5539/ijsp.v6n2p134>

Abstract

In this paper, we consider several binomial mixture models for fitting over-dispersed binary data. The models range from the binomial itself, to the beta-binomial (BB), the Kumaraswamy distributions I and II (KPI & KPII) as well as the McDonald generalized beta-binomial mixed model (McGGB). The models are applied to five data sets that have received attention in various literature. Because of convergence issues, several optimization methods ranging from the Newton-Raphson to the quasi-Newton optimization algorithms were employed with SAS PROC NLMIXED using the Adaptive Gaussian Quadrature as the integral approximation method within PROC NLMIXED. Our results differ from those presented in Li, Huang and Zhao (2011) for the example data sets in that paper but agree with those presented in Manoj, Wijekoon and Yapa (2013). We also applied these models to the case where we have a k vector of covariates $(x_1, x_2, \dots, x_k)'$. Our results here suggest that the McGGB performs better than the other models in the GLM framework. All computations in this paper employed PROC NLMIXED in SAS. We present in the appendix a sample of the SAS program employed for implementing the McGGB model for one of the examples.

Keywords: beta-binomial, kumaraswamy, McDonald generalized beta-binomial

1. Introduction

The binomial outcome data are widely encountered in many real world applications. The Binomial distribution often fails to model the binomial outcomes since the variance of the observed binomial outcome data exceeds the nominal Binomial distribution variance, a phenomenon known as over-dispersion. One way of handling overdispersion is modeling 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 (Kumaraswamy, 1980) is another well utilized member of this class. In this paper we focus the emphasis on the McDonald's Generalized Beta distribution (Manoj *et al.*, 2015) of the first kind as the mixing distribution and would employ the newly introduced the McDonald Generalized Beta-Binomial distribution (McGGB) Manoj *et al.* (2013). Some theoretical properties of McGGB are already discussed in Manoj *et al.* (2013) would not again be discussed here. The parameters of the McGGB distribution are estimated via maximum likelihood estimation technique. Real world datasets are modeled by using the new McGGB mixture distribution, and it is shown that this model gives better fit than its nested models, namely, the beta-binomial and the Kumaraswamy type II models. We also compared the distributions for the generalized linear model when we have covariates. The teratology data in Moore & Tsiatis (1991) are employed. Over-dispersion in binomial regression could arise as a result of several reasons or combinations of reasons. In most cases, this could be due to the binomial model grossly underestimating the response variance or it could also be caused by positive correlation between the dependent responses. Hilbe (2011) has presented an elegant overview of this for the Poisson model. When over-dispersion occurs, then standard errors of parameter estimates are often underestimated and this often leads to wrong conclusions with regards to the significance of the predictor variables involved. The McGGB and other mixing models consider here therefore model the parameter π of the binomial with continuous distributions defined in the interval (0,1).

In this paper, we would consider the following binomial mixture distributions in addition to the binomial (used here as a reference model only).

- The beta-binomial
- The Kumaraswamy binomial I & II binomial models
- The Macdonald's generalized beta-binomial distribution

1.1 The Binomial Model

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

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

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

$$E(Y) = n\pi \quad (2a)$$

$$\text{Var}(Y) = n\pi(1 - \pi) \quad (2b)$$

2. Mixture Models for the Binomial

We consider in this section the following mixture distributions for the binomial:

- The beta-binomial
- The Kumaraswamy binomial I & II binomial models
- The Macdonald's generalized beta-binomial distribution

We develop each of these models in turn in what follows:

The mixture model is obtained by evaluating the well-known integral:

$$f(y) = \int_0^1 f_{Y|\pi}(y) f_{\pi}(\pi | \Theta) d\pi \quad (3)$$

for $y = 0, 1, \dots, n$ and Θ is the parameter space of the mixing distribution.

2.1 The Beta-Binomial (BB)

For the beta-binomial (Skellam, 1946), is 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, $\text{Bin}(n, \pi) \wedge \text{Beta}(\alpha, \beta) \sim \text{BB}$. The resulting unconditional pdf being the beta-binomial defined as:

$$f(y; \alpha, \beta) = \binom{n}{y} \frac{B(\alpha + y, \beta + n - y)}{B(\alpha, \beta)}, \quad y = 0, 1, \dots, n \quad (4)$$

2.2 The Kumaraswamy Distribution

The Kumaraswamy I (Kumaraswamy, 1980) written as KBI or KW(I) model is a mixture of Binomial and p having the Kumaraswamy distribution, that is, is also a mixture of

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

where $\text{KB}(\alpha, \beta)$ has the distribution:

$$f(\pi, \alpha, \beta) = \alpha\beta\pi^{\alpha-1}(1 - \pi^{\alpha})^{\beta-1}, \quad \alpha, \beta > 0 \quad (5)$$

Thus the KW(I) and KW(II) are resulting unconditional probability distributions given in (6) and (8) respectively as:

$$f(y; \beta) = \binom{n}{y} \beta B(y+1, n-y+\beta), \quad \beta > 0 \quad (6)$$

The mean and variance of KW(I) are respectively:

$$E(Y) = n\beta B(2, \beta) \quad (7a)$$

$$\text{Var}(Y) = n^2\beta[B(3, \beta) - \beta B^2(2, \beta)] + n\beta[B(2, \beta) - B(3, \beta)] \quad (7b)$$

The KW(II) on the other hand, has the form:

$$f(y; \alpha, \beta) = \alpha\beta \binom{n}{y} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} B(y+\alpha+\gamma i, n-y+1) \quad (8)$$

where $\alpha, \beta > 0$ and $y = 0, 1, 2, \dots, n$.

Its mean and variance are given respectively as:

$$E(Y) = n\beta B\left[1 + \frac{1}{\alpha}, \beta\right] \quad (9a)$$

$$\text{Var}(Y) = n^2\beta \left[B\left(1 + \frac{2}{\alpha}, \beta\right) - \beta B^2\left(1 + \frac{1}{\alpha}, \beta\right) \right] + n\beta \left[B\left(1 + \frac{1}{\alpha}, \beta\right) - B\left(1 + \frac{2}{\alpha}, \beta\right) \right] \quad (9b)$$

2.3 The McDonald Generalized Beta-Binomial Model (McGGB)

The McGGB distribution with parameters n, α, β, γ is a mixture of binomial $\text{Bin}(n, \pi)$ and the McDonald's generalized beta distribution (McDonald, 1984, 1995) of the first kind $\text{GB1}(\alpha, \beta, \gamma)$ where the latter distribution is defined as:

$$f(\pi; \alpha, \beta, \gamma) = \frac{\gamma}{B(\alpha, \beta)} \pi^{\alpha\gamma-1} (1-\pi)^{\beta-1}; \quad 0 \leq \pi \leq 1 \text{ and } \alpha, \beta, \gamma > 0. \quad (10)$$

Clearly, the GB1 reduces to the beta distribution when $\gamma = 1$ and to the Kumaraswamy distribution when $\alpha = 1$.

Thus the McGGB being a mixture distribution of

$$Y|\pi \sim \text{Bin}(n, \pi), \quad \text{and} \quad \pi \sim \text{GB1}(\alpha, \beta, \gamma)$$

has the resulting unconditional pdf given by:

$$f(y; n, \alpha, \beta, \gamma) = \binom{n}{y} \frac{\gamma}{B(\alpha, \beta)} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} B(y+\alpha\gamma+\gamma i, n-y+1) \quad (11)$$

for $y = 0, 1, \dots, n$ and $\alpha, \beta, \gamma > 0$.

Following Manoj *et al.* (2013), a rearranged pdf of McGGB (n, α, β, γ) is given by:

$$f(y; n, \alpha, \beta, \gamma) = \binom{n}{y} \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + \alpha + \frac{j}{\gamma}, \beta\right) \quad (12)$$

for $y = 0, 1, \dots, n$ and $\alpha, \beta, \gamma > 0$.

It can be easily demonstrated that the McGGB in (12):

- reduces to the Beta-Binomial with parameters n, α, β if $\gamma = 1$.
- reduces to the Kumaraswamy-Binomial model with parameters n, β, γ if $\alpha = 1$.

The mean and variance of the McGBB are given by:

$$E(Y) = n\pi \quad \text{and} \quad \text{Var}(Y) = n\pi(1 - \pi)[1 + (n - 1)\rho] \quad (13)$$

where

$$\pi = \frac{B(\alpha + \beta, \frac{1}{\gamma})}{B(\alpha, \frac{1}{\gamma})}, \quad \text{and} \quad \rho = \frac{\left(\frac{B(\alpha + \beta, 2/\gamma)}{B(\alpha, 2/\gamma)}\right) - \left(\frac{B(\alpha + \beta, 1/\gamma)}{B(\alpha, 1/\gamma)}\right)^2}{\left(\frac{B(\alpha + \beta, 1/\gamma)}{B(\alpha, 1/\gamma)}\right) - \left(\frac{B(\alpha + \beta, 1/\gamma)}{B(\alpha, 1/\gamma)}\right)^2}$$

2.4 Parameter Estimation

For a single observation, the log-likelihood for the binomial, the beta-binomial, KPI, KPIL, and the MCGBB are displayed in expressions (14a) to (14e) respectively.

$$LL1 = \log \binom{n}{y} + y \log(\pi) + (n - y) \log(1 - \pi) \quad (14a)$$

$$LL2 = \log \binom{n}{y} + \log[B(\alpha + y, \beta + n - y)] - \log[B(\alpha, \beta)] \quad (14b)$$

$$LL3 = \log \binom{n}{y} + \log(\beta) + \log[B(y + 1, n - y + \beta)] \quad (14c)$$

$$LL4 = \log(\alpha\beta) + \log \binom{n}{y} + \log \left[\sum_{i=0}^{\infty} (-1)^i \binom{\beta - 1}{i} B(y + \alpha + \alpha i, n - y + 1) \right] \quad (14d)$$

$$LL5 = \log \binom{n}{y} + \log \left(\frac{1}{B(\alpha, \beta)} \right) + \log \left[\sum_{j=0}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + \alpha + \frac{j}{\gamma}, \beta\right) \right] \quad (14e)$$

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 (Pinheiro & Bates, 1995) 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 Powell (1977) & Beale (1972). 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. Although the results differ very slightly, on the whole, they all agree very well. Thus we may note here that each of these give slightly different parameter estimates. They all give values that are very close.

$$LL = \log \binom{n}{y} + \log \left(\frac{1}{B(\alpha, \beta)} \right) + \log \left[\sum_{j=0}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + \alpha + \frac{j}{\gamma}, \beta\right) \right] \quad (15)$$

Since the estimation of the Kumaraswamy-binomial often leads to non-convergence, we would employ the log-likelihood of the McGBB since it reduces to the KB for $\alpha = 1$. Thus its likelihood will now be of the form:

$$LL = \log \binom{n}{y} + \log \left(\frac{1}{B(1, \beta)} \right) + \log \left[\sum_{j=0}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + 1 + \frac{j}{\gamma}, \beta\right) \right] \quad (16)$$

3. Application to Data Examples

We have applied the above models to five data sets that have previously being analyzed by various authors. The five data sets are described in Examples I to V.

3.1 Data Example I

This example is taken from Nelder & Mead (1965) and relates to the number of candidates having an “alpha”, i.e. at least 15 scores out of a total 20 points from each of nine questions employed in assessing the final class of candidates in an examination. Here, $Y \sim \text{Bin}(9, \pi)$. There were a total of 209 candidates for the examination and Table 1 gives the distribution of these scores for the 209 candidates as well as the expected values and goodness-of fit-test under the five models. For the binomial case for instance, the expected probabilities are $P(Y = y) = \binom{9}{y} \hat{\pi}^y (1 - \hat{\pi})^{9-y}$ and consequently, the expected values are $\hat{m}_y = 209 \times P(Y = y)$, $y = 0, 1, 2, \dots, 9$. The computed probabilities and expected values, together with the cumulative estimated probabilities and expected values under the McGBB model for this data are presented in Appendix II. We note here that the estimated probabilities sum to 1.0

Table 1. Expected values under the five Models and corresponding GOF test

Y	Freq	Five Different Models				
		BIN	BB	KPI	KPII	McGBB
0	63	37.6895	68.5406	72.2772	69.3715	68.3907
1	67	71.1132	52.9068	50.9881	52.3367	54.0165
2	34	59.6345	36.5973	34.6924	36.1891	36.5044
3	18	29.1717	23.5102	22.5741	23.3822	22.9847
4	11	9.1736	14.0102	13.8807	14.0466	13.6015
5	8	1.9232	7.6362	7.9248	7.7252	7.4840
6	4	0.2688	3.7048	4.0861	3.7833	3.7357
7	3	0.0242	1.5234	1.8140	1.5705	1.6158
8	1	0.0013	0.4810	0.6301	0.5006	0.5514
9	0	0.0000	0.0897	0.1324	0.0942	0.1151
Total	209	209.00	209.00	209.00	209.00	209.00
	-	$\hat{\pi} = 0.1733$	$\hat{\alpha} = 1.1541$	$\hat{\beta} = 4.7578$	$\hat{\beta} = 5.4530$	$\hat{\alpha} = 38.8887$
	-		$\hat{\beta} = 5.4564$	-	$\hat{\gamma} = 1.0837$	$\hat{\beta} = 4.5288$
						$\hat{\gamma} = 0.0522$
X^2		91.7731	7.9710	8.8851	8.6413	7.0160
d.f		8	7	8	7	6
-2LL		791.9	708.1	708.6	708.3	706.8
AIC		793.9	712.1	710.6	712.3	712.8
<hr/>						
6+	8	0.2943				
X^2		253.8570				
d.f		5				
<hr/>						
8+	1		0.5707	0.7625	0.5948	0.6665
X^2			8.1196	8.6094	8.3249	6.7030
d.f.			6	7	6	5

3.1.1 Results

In this paper, we have employed the binomial model as a baseline model for comparative purposes. The expected values in column 3 in Table 1 are those under the binomial model designated as ‘BIN’. Under the natural binomial model for this data, $X^2 = 455.6136$ on 208 d.f with estimated dispersion parameter $\hat{\phi} = 2.1904 \gg 1$ indicating strong over-dispersion in the data. The model produces very many small expected frequencies undermining the approximate distribution of Pearson’s X^2 as a χ^2 distribution. Consequently, if we adopt the Lawal’s (1980) rule, which states that expected values can be as small as $r/d^{3/2}$ (where r is the number of expected values less than 3, and d is the degrees of freedom under such model) without violating the χ^2 assumption, we see that the minimum expected values required for the BIN model for example will be $\frac{5}{8^{3/2}} = 0.22$ since only five expected values are less than 3 and the d.f. = $(10 - 2) = 8$. Hence cells 6,7,8,

and 9, that is, 6+ will be collapsed, now giving a new $X^2 = 80.000$ on 5 d.f. This presented in the bottom panel of Table 1. Corresponding expected minimum expectations for the BB, KPI, KPII and McGGB are 0.1620, 0.1326, 0.1620 and 0.2041 respectively, resulting in categories 8 and 9 (i.e. 8+) being collapsed. The expected values, X^2 and the appropriate degree of freedom are presented in the bottom panel of the data.

3.2 Data Example II-Terrorism in the US

The data in this example relate to the incidents of terrorism in the United States and was analyzed in Xiaohu Li *et al.* (2011). The data was originally presented in Jenkins & Johnson (1975) who compiled a chronology of incidents in international terrorism between 01/1968 and 04/1974. The number of incidents and frequency of occurrences are presented in columns 1 and 2 respectively in Table 2. Here Y is the number of incidents recorded.

Table 2. Goodness of fit tests for Terrorism Incidents in the US

Y	Freq	Five Different Models				
		BN	BB	KPI	KPII	McGGB
0	38	33.7216	39.2961	40.2839	39.5565	39.4298
1	26	29.7544	22.0298	20.8953	21.7029	22.0184
2	8	10.5015	9.9312	9.6744	9.8906	9.7633
3	2	1.8532	3.6187	3.7991	3.6764	3.5988
4	1	0.1635	0.9751	1.1444	1.0144	1.0186
≥ 5	1	0.0058	0.1491	0.2029	0.1592	0.1711
Total	76	76.00	76.00	76.00	76.00	76.00
	-	$\hat{\pi} = 0.1500$	$\hat{\alpha} = 1.2126$ $\hat{\beta} = 6.8161$	$\hat{\beta} = 5.6395$ -	$\hat{\alpha} = 1.1012$ $\hat{\beta} = 6.7530$	$\hat{\alpha} = 8.5468$ $\hat{\beta} = 5.4116$ $\hat{\gamma} = 0.2261$
X^2		-	6.7125	5.6673	6.4794	5.8002
d.f		4	3	4	3	2
-2LL	-	188.3	180.4	180.6	180.5	180.1
3+	4	2.0225				
X^2		3.5459				
d.f		2				
4+	2		1.1242	1.3473	1.1736	1.1897
X^2			2.5401	2.8345	2.6199	2.3525
d.f.			2	3	2	1

For this data, we would need to collapse cells 3,4,5 for the binomial and cells 4 and 5, for the other four models. Clearly, the KPI seems to give the most parsimonious model for this data, but clearly the other models also fit the data well. The McGGB even though has the smallest X^2 but also has the smallest degrees of freedom because of the three parameters estimated under the model. In spite of several optimization methods mentioned above in SAS PROC NLMIXED, we could not reproduce the results of expected values presented in Li *et al.* (2011). All the methods lead to similar results presented above. It is our opinion that there are inherent errors in the computational aspects of the results in Li *et al.*. The -2 log likelihood (-2LL) obtained in our analyses are much lower than those presented in Li *et al.*. The generated expected probabilities in our models all sum to 1 and the expected values are computed on this basis. Our results therefore are much different from those of Li *et al.* and would conclude that Li *et al.* perhaps need to revisit their computations and hopefully would observe that the new results will agree with those we presented in Tables 2 and 3 in this paper.

3.3 Data Example III-Terrorism Data in Argentina

This example data is again originally from Jenkins & Johnson (1975) and relate to the number of incidents of terrorism in Argentina between 1968 and 1974.

For this data, cells are collapsed for both the binomial and KPI models. The other three models have minimum expected values that satisfy the Lawal's rule. Here, the Beta-binomial, KPII and the McGGB all fit the data well but the beta-binomial seems to be the most parsimonious for the data. Again here, the expected values obtained are different from those presented in Li *et al.* (2011). The reasons presented earlier for Table 2 results also apply to these discrepancies in our results. We believe that our computations are very accurate.

Table 3. Goodness of fit tests on Argentina Terrorism Data

Y	Freq	Five Different Models				
		BIN	BB	KPI	KPII	McGBB
0	46	30.2054	46.3620	36.8823	46.3510	46.3339
1	15	30.1282	12.6172	20.7649	12.8390	12.9117
2	5	12.5213	7.0643	10.7511	6.9975	6.9845
3	3	2.7754	4.4849	4.9675	4.3816	4.3570
4	5	0.3460	2.8903	1.9462	2.8239	2.8065
5	1	0.0230	1.7468	0.5847	1.7367	1.7312
6	1	0.0006	0.8344	0.1034	0.8702	0.8753
		$\hat{\pi} = 0.1425$	$\hat{\alpha} = 0.3108$ $\hat{\beta} = 1.8531$	$\hat{\beta} = 5.6571$	$\hat{\beta} = 1.7195$ $\hat{\gamma} = 0.3719$	$\hat{\alpha} = 10.1455$ $\hat{\beta} = 1.6955$ $\hat{\gamma} = 0.0443$
X^2		1789.16	3.4397	16.4866	3.3810	3.3677
D.f.		5	4	5	4	3
p-value			0.4871		0.4962	0.3383
-2LL		244.00	189.30	200.90	189.20	189.20
4+	7	0.3696				
X^2		139.3368				
d.f.		3				
p-value		0.0000				
5+	2			0.6881		
X^2				15.0032		
d.f.				4		
p-value				0.0047		

Table 4. Goodness of fit tests for the alcohol Drinking Data in Week I Under the Five Models

Y	Freq	Five Different Models				
		BIN	BB	KPI	KPII	McGBB
0	47	1.595	54.6205	38.4654	54.3128	51.3027
1	54	13.407	42.0028	39.9088	41.7253	45.6820
2	43	48.303	38.9019	41.6669	38.8598	43.1804
3	40	96.682	38.5399	43.8892	38.6929	41.6112
4	40	116.110	40.0691	46.8548	40.3775	40.5276
5	41	83.665	43.9956	51.1733	44.4135	40.0029
6	39	33.492	53.0899	58.5900	53.5016	41.7843
7	95	5.746	87.7803	78.4517	87.1169	94.9089
		$\hat{\pi} = 0.5456$	$\hat{\alpha} = 0.7229$ $\hat{\beta} = 0.5809$	$\hat{\beta} = 0.7468$	$\hat{\beta} = 0.5931$ $\hat{\gamma} = 0.7038$	$\hat{\alpha} = 0.0366$ $\hat{\beta} = 0.1937$ $\hat{\gamma} = 24.3004$
X^2		789.0557	9.5143	20.3224	9.9907	2.1559
d.f.		6	5	6	5	4
p-value		0.0000	0.0902	0.0024	0.0755	0.7071
-2LL		2406.30	1626.90	1637.80	1627.40	1619.40

3.4 Data Example IV: Alcohol Drinking Data in Week I

The alcohol drinking data refers to the numbers of alcohol consumption days in two reference weeks that are separately reported by a randomly selected sample of 399 respondents in the Netherlands in 1983 (Alanko & Lemmens, 1996). These are displayed in Tables 4 and 5 for week I and II respectively. The choice of these data sets is motivated by the fact that they have also been analyzed in Rodriguez-Avi *et al.* (2007) and Li *et al.* (2011). Thus we can compare our results with theirs.

For these two data sets, the number of days Y , an individual consumes alcohol out of $n = 7$ days in a particular week is considered distributed binomial, however, the success probability π of consuming alcohol on a particular day can not be assumed constant because of person-to-person variation and thus a mixture binomial distribution could be applied to these data, with π being modeled by either a beta distribution or a Kumaraswamy distribution in the interval $[0,1]$. This is the approach adopted in Manoj *et al.* (2013) and Li *et al.* (2011).

3.5 Data Example V: Alcohol Drinking Data in Week II

The description of this data set is as earlier presented in 3.4 and in Table 5 are presented the expected values as well as the goodness-of fit tests under the five models.

Table 5. Goodness of fit tests for for the alcohol Drinking Data in Week II Under the Five Models

Y	Freq	Seven Different Models				
		BIN	BB	KPI	KPII	McGBB
0	42	1.6574	47.9022	40.4814	47.7196	46.0752
1	47	13.7932	42.9141	41.7310	42.8214	45.1044
2	54	49.1949	41.9502	43.2416	41.9595	44.6300
3	40	97.4774	42.5022	45.1337	42.5927	44.3245
4	49	115.8881	44.3031	47.6296	44.4530	44.1480
5	40	82.6654	47.8099	51.2075	47.9897	44.3265
6	43	32.7596	54.8871	57.2026	55.0292	46.6065
7	84	5.5638	76.7311	72.3726	76.4369	83.7849
		$\hat{\pi} = 0.5456$	$\hat{\alpha} = 0.8576$ $\hat{\beta} = 0.7008$	$\hat{\beta} = 0.7904$	$\hat{\beta} = 0.7068$ $\hat{\gamma} = 0.8528$	$\hat{\alpha} = 0.0406$ $\hat{\beta} = 0.2892$ $\hat{\gamma} = 24.1410$
X^2		662.5698	9.7615	11.8695	9.8789	4.0644
d.f		6	5	6	5	4
p-value		0.0000	0.0823	0.0649	0.0787	0.3974
-2LL		2295.60	1642.80	1645.20	1642.90	1637.0

3.5.1 Results

Our results in Tables 4 and 5 agree with those presented in Manoj *et al.* (2013), For both data, the McGBB model is most preferable to the other models. Of course the binomial model does not fit the data at all as expected. At the 5% nominal level, the beta-binomial, KPI, KPII and the McGBB fit the data however, the McGBB is the most parsimonious with p-values of 0.7071 and 0.3974 respectively for data in Tables 4 and 5. We note here that the KPI does not fit the week 1 data. The poor fit of the binomial model for instance can be attributed to the expected means and variances under the binomial compared to the observed values of these parameters. For Table 4 for instance, these are displayed in the following:

Model	\bar{y}	$\hat{\sigma}^2$
BIN	3.8192	1.7362
McGBB	3.8451	6.2607
Observed Data	3.8195	6.2538

While the observed means and variance of the data are $\mu = 3.8195$ and $\sigma^2 = 6.2538$, we see that the binomial model grossly underestimates the variance of the data, while the McGBB estimates the variance well. We thus see that the McGBB and the other mixture models model the variance of the observed much better and hence fits the data much better than the binomial.

4. Extension to GLM

In this section we extend the applications of the models proposed above to the case where we have covariates. Suppose we have a vector of covariates, say, $(x_1, x_2, \dots, x_t)'$. We can incorporate this into our models—we present an example data for illustration:

4.1 Example: Teratology Data

Teratology is the study of abnormalities of physiological development. The data below 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 injection), group 2 (injections on days 7 and 10), group 3 (days 0 and 7), 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. Due to non measured covariates and genetic variability the probability of death may vary from litter to litter within a particular treatment group. There are 31, 12, 5 and 10 rats respectively in groups 1, 2, 3 and 4, and presented below are y/n for the data, where for instance, 1/10 means that $y = 1$ and $n = 10$.

group1	31	1/10,	4/11,	9/12,	4/4,	10/10,	9/11,	9/9,	11/11,	10/10,	7/10,	
		12/12,	9/10,	8/8,	9/11,	4/6,	7/9,	14/14,	7/12,	9/11,	8/13,	
		5/14,	10/10	10/12	8/13	10/10	3/14,	13/13,	3/4,	8/8,	5/13,	12/12
group2	12	1/10,	1/3,	1/13,	0/12,	4/14,	2/9,	2/13,	1/16,	0/11,	0/4,	0/12
group3	5	0/8,	1/11,	0/14,	1/14,	0/11						
group4	10	0/3,	0/13,	2/9,	2/17,	0/15,	0/2,	1/14,	0/8,	0/6,	0/17	

If we let π_{ij} denote the probability of death for fetus j in litter i , then,

- For the binomial, the set up is:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = a_0 + a_2x_{2i} + a_3x_{3i} + a_4x_{4i}$$

where

$$x_{gi} = \begin{cases} 1 & \text{if litter } i \text{ is in group } g \\ 0 & \text{otherwise} \end{cases}$$

- For the beta-binomial, Kumaraswamy II (fitted as nested within the McGBB) and McGBB models, the set up is as follows in (17a) to (17c) respectively:

$$\alpha = \exp(a_0 + a_2x_2 + a_3x_3 + a_4x_4), \quad \gamma = 1, \beta > 0 \quad (17a)$$

$$\beta = \exp(a_0 + a_2x_2 + a_3x_3 + a_4x_4), \quad \alpha = 1, \gamma > 0 \quad (17b)$$

$$\alpha = \exp(a_0 + a_2x_2 + a_3x_3 + a_4x_4), \quad \beta, \gamma > 0 \quad (17c)$$

where a_0, a_2, a_3, a_4 are to be estimated from the models in addition to the the other parameters of the models.

The use of $\exp(\mathbf{x}'\mathbf{a})$ is some times preferable to the use of the logit - $1/[1 + \exp(-\mathbf{x}'\mathbf{a})]$ (see Martinez, 2015). For a general set of covariates, the above, for instance, α becomes $\alpha = a_0 + a_1x_1 + a_2x_2 + \dots, +a_tx_t$.

5.2 Computing the Pearson's GOF X^2

We generate the Pearson's goodness-of fit test statistic X^2 with the the basic GOF defined as:

$$X^2 = \sum_{i=1}^n \frac{[y_i - E(Y)]^2}{\text{Var}(Y)} \quad (18)$$

where $n = 58$ the total number of observations in our data. Expressions for $E(Y)$ and $\text{Var}(Y)$ have been presented for each of the models considered in this study. While those of the binomial are presented in expressions (2), those for the McGBB and its nested models BB and KPII can be generated from the expressions in (13) appropriately.

5.2.2 Results

We present in Table 6 the results of applying these models to the teratology data.

Table 6. Parameter estimates and GOFs under the four models for the teratology Data

Parameters	BIN	BB	KPII	McBB
	-	$\hat{\beta} = 0.8510$	$\hat{\gamma} = 1.7481$	$\hat{\beta} = 0.1451$ $\hat{\gamma} = 45.8639$
\hat{a}_0	1.1440	0.9924	-0.8121	-2.9930
\hat{a}_2	-3.3225	-2.3814	4.5903	-2.2076
\hat{a}_3	-4.4762	-3.0343	6.5215	-2.8533
\hat{a}_4	-4.1297	-3.1799	5.9134	-2.9923
-2LL	248.3499	198.70	176.80	194.00
X^2	154.7070	44.8798	53.5515	37.5571
d.f.	54	52	52	51
p-value	0.000	0.7475	0.4145	0.9196

Under the Binomial model, we see that the data is clearly over-dispersed with the dispersion parameter estimated as $\hat{\phi} = 154.7070/54 = 2.8649 \gg 1$. Of course the model does not fit the data at all. Of the four models, the McGBB fits best with $X^2 = 37.5571$ on 52 d.f and a corresponding p-value of 0.9196. Interestingly, both the BB and the KPII also fit the data, however, the BB fits better than the KPII. We present in Table 7, the estimated probabilities $\{0.7714, 0.2300, 0.1337, 0.1182\}$ for the four groups as well as the corresponding means and variances for the first five observations in groups 1, 2 and 3 and the last five observations in group 4 together with their corresponding contributions to X^2 .

Table 7. Estimated means and Variances under the McGBB Model

Obs	GP	n	y	$\hat{\pi}$	mean	var	X^2
1	1	10	1	0.7714	7.7139	6.5919	6.8382
2	1	11	4	0.7714	8.4853	7.8413	2.5657
3	1	12	9	0.7714	9.2567	9.1979	0.0072
4	1	4	4	0.7714	3.0856	1.3492	0.6198
5	1	10	10	0.7714	7.7139	6.5919	0.7928
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
32	2	10	1	0.2300	2.2999	9.9196	0.1704
33	2	3	1	0.2300	0.6900	1.0745	0.0894
34	2	13	1	0.2300	2.9899	16.4265	0.2411
35	2	12	0	0.2300	2.7599	14.0764	0.5411
36	2	14	4	0.2300	3.2199	18.9576	0.0321
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
44	3	8	0	0.1337	1.0694	4.3919	0.2604
45	3	11	1	0.1337	1.4705	8.0811	0.0274
46	3	14	0	0.1337	1.8715	12.8841	0.2719
47	3	14	1	0.1337	1.8715	12.8841	0.0590
48	3	11	0	0.1337	1.4705	8.0811	0.2676
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
54	4	2	0	0.1182	0.2363	0.3205	0.1743
55	4	14	1	0.1182	1.6544	11.6595	0.0367
56	4	8	0	0.1182	0.9454	3.9723	0.2250
57	4	6	0	0.1182	0.7090	2.3067	0.2179
58	4	17	0	0.1182	2.0089	17.0164	0.2372

6. Conclusions

In all the data examples considered in this study, the McGBB performs best. We employ the finite series representation of the McDonald's Generalized Beta-Binomial distribution as presented in Manoj *et al.* (2013). The McGBB has considerable improvement over the BB, KPI and KPII models. However, convergence problem for the McGBB makes it daunting for easy use but with the array of optimizations algorithms available in SAS PROC NLMIXED and R studio, the distribution is not more difficult to fit than any of the other models.

We also observe here that the results presented in Liet *al.* (2011) are suspect and our results are confirmed by the use of several SAS procedures as mentioned in the body of this paper. Thus if convergence can be achieved, the three parameter McDonald's generalized beta-binomial model offers a much better alternative to some of the well known mixture models for over-dispersed binary data, at least for the example cases considered in this paper.

Acknowledgements

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

References

- Alanko, T., & Lemmens, P. H. (1996). Response effects in consumption surveys: an application of the beta-binomial model to self-reported drinking frequencies. *Journal of Official Statistics*, 12(3), 253-273.
- Beale, E. M. L. (1972). A Derivation of Conjugate Gradients. in F. A. Lootsma, ed., *Numerical Methods for Nonlinear Optimization*, London: Academic Press.
- 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.
- Haseman, J. K., & Kuper, J. K. (1978). The use of a correlated binomial model for the analysis of a certain toxicological experiments. *Biometrics*, 34, 69-76.
- Hilbe, J. (2011). *Negative Binomial Regression*, Second edition. Cambridge . Cambridge University Press.
- Jenkins, & Johnson (1975). International terrorism: a chronology, 1968-1974, Report R-1597-DOS/ARPA, The Rand Corporation.
- Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1), 79-88. [https://doi.org/10.1016/0022-1694\(80\)90036-0](https://doi.org/10.1016/0022-1694(80)90036-0)
- Lawal, H. B. (1980). Tables of percentage points of Pearson's goodness-of-fit statistic for use with small expectations. *Applied Statistics*, 29, 292-298.
- Li, X. H., Huang Y. Y., & Zhao, X. Y. (2011). The Kumaraswamy Binomial Distribution. *Chinese Journal of Applied Probability and Statistics*, 27(5), 511-521.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica: Journal of the Econometric Society*, 52(3), 647-663. <https://doi.org/10.2307/1913469>
- McDonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66(1-2), 133-152. [https://doi.org/10.1016/0304-4076\(94\)01612-4](https://doi.org/10.1016/0304-4076(94)01612-4)
- Manoj, 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. <https://doi.org/10.5539/ijsp.v2n2p24>
- 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., & Tsatis, A. (1991). Robust estimation of the variance in moment methods for extra-binomial and extra-Poisson variation. *Biometrics*, 47, 383-401.
- Morel, J. G., & Neerchal, N. K. (1997). Clustered binary logistic regression in teratology data using a finite mixture distribution. *Stat. Medicine*, 16, 2843-2853.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308-312. <https://doi.org/10.1093/comjnl/7.4.308>
- 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.

Powell, M. J. D. (1977). Restart Procedures for the Conjugate Gradient Method, *Mathematical Programming*, 12, 241-254.

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. J., & Olmo-Jiménez, M. J. (2007). A generalization of the beta-binomial distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 56(1), 51-61. <https://doi.org/10.1111/j.1467-9876.2007.00564.x>

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. Soc., B10*, 257-261. <https://doi.org/10.2307/2983779>

Appendix

Appendix I SAS PROC NLMIXED Program to Fit the McGBB Model

```
options nodate nonumber ls=85 ps=66;
data arg;
INPUT y count @@;
n=6;
datalines;
0 46 1 15 2 5 3 3 4 5 5 1 6 1
;
run;
data new1;
set arg;
do i=1 to count;
output;
end;
keep y n;
run;

title 'Terrorism Data Argentina';
proc nlmixed data=new1 tech=nmsimp qppts=21;
parms aa=1.5 beta=2 gamma=1;
bounds aa >0, beta >0, gamma >0;
b=lgamma(aa)+lgamma(beta)-lgamma(aa+beta);
z=lgamma(n+1)-lgamma(y+1)-lgamma(n-y+1);
sum=0.0;
t=n-y;
do j=0 to t;
u1=(-1)**j;
u2=(y/gamma)+aa+(j/gamma);
z1=lgamma(t+1)-lgamma(j+1)-lgamma(t-j+1);
z2=lgamma(u2)+lgamma(beta)-lgamma(u2+beta);
b1=z1+z2;
sum=sum+(u1*exp(b1));
keep sum;
end;
LL=z-b+log(sum);
model y ~general(LL);
predict aa out=aa1;
predict beta out=bb;
predict gamma out=cc;
run;

data q1;
set aa1;
aa=pred;
```

```
run;
data q2;
set bb;
beta=pred;
run;
data q3;
set cc;
gamma=pred;
run;
data qq3;
merge q1 q2 q3;
do s=0 to n;
suma=0.0;
k=n-s;
do i=0 to k;
uu1=(-1)**i;
uu2=(s/gamma)+aa+(i/gamma);
zz1=lgamma(k+1)-lgamma(i+1)-lgamma(k-i+1);
zz2=lgamma(uu2)+lgamma(beta)-lgamma(uu2+beta);
bb1=zz1+zz2;
suma=suma+(uu1*exp(bb1));
keep suma;
end;
/* Compute estimated probabilities, cumulative as well as the expected */;
/* frequencies */;

b1=lgamma(aa)+lgamma(beta)-lgamma(aa+beta);
zz=lgamma(n+1)-lgamma(s+1)-lgamma(n-s+1);
LL1=zz-b1+log(suma);
prob=exp(LL1);
fit1=DF*prob;
ss1+fit1;
cum+prob;
output;
end;
stop;
keep s prob cum fit1 ss1 aa beta gamma;
run;

/* Print the estimated probabilities and expected values*/;
proc print data=qq3;
var s prob cum fit1 ss1 ;
run;

title1 'Compute Pearson's GOF for the Data';
data gof;
merge arg qq3;
xx=((count-fit1)**2)/fit1;
sum+xx;
run;
proc print data=gof;
var sum;
format sum 10.4;
run;
```

Appendix II

Estimated probabilities and expected values from the SAS program. ss1 is the cumulative expected values.

y	$\hat{\pi}_i$	$P(Y \leq y)$	\hat{m}_i	ss1
0	0.32723	0.32723	68.3907	68.391
1	0.25845	0.58568	54.0165	122.407
2	0.17466	0.76034	36.5044	158.912
3	0.10997	0.87032	22.9847	181.896
4	0.06508	0.93540	13.6015	195.498
5	0.03581	0.97120	7.4840	202.982
6	0.01787	0.98908	3.7357	206.718
7	0.00773	0.99681	1.6158	208.333
8	0.00264	0.99945	0.5514	208.885
9	0.00055	1.00000	0.1151	209.000

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).