



SDI Review Form 1.6

Journal Name:	British Journal of Mathematics & Computer Science
Manuscript Number:	Ms_BJMCS_33475
Title of the Manuscript:	Employing the Double, Multiplicative and The Com-Poisson Binomial Distributions for modeling Over and Under-dispersed Binary Data
Type of the Article	Original Research Papers

General guideline for Peer Review process:

This journal's peer review policy states that **NO** manuscript should be rejected only on the basis of '**lack of Novelty**', provided the manuscript is scientifically robust and technically sound.

To know the complete guideline for Peer Review process, reviewers are requested to visit this link:

(<http://www.sciencedomain.org/page.php?id=sdi-general-editorial-policy#Peer-Review-Guideline>)



SDI Review Form 1.6

PART 1: Review Comments

	Reviewer's comment	Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
<u>Compulsory</u> REVISION comments		
<u>Minor</u> REVISION comments	General Comments: In this manuscript the authors(s) investigate the performances of the binomial (BN), the beta-binomial (BB), the multiplicative binomial (MBM), the Com-Poisson binomial (CPB) and the double binomial (DBM) model when fitting over-dispersed binary data. Their results show that all these models (excluding the binomial) behave reasonably well in modeling over-dispersed binary data, with or without covariates. However, the multiplicative binomial and the double binomial models slightly behave better. Therefore, it is always advisable for one to explore all possible models in order to come to right choice of model for a given data set. This manuscript may be a fairly simple fitting investigation over certain data, however shows some rather interesting results. <i>Note: The text in red below represents the reviewer's suggested changes/additions. The yellow boxed text indicates a rephrased text (by the reviewer) that author(s) need to check if it is what they were trying to say.</i>	



SDI Review Form 1.6

	Otherwise, i.e. if the author(s) do not agree with the suggested text (in the yellow boxes), they should provide a new rephrased text (as the original one has problematic syntax or it were not explained properly). Moreover, the bold form of [yellow boxed] text indicates a notice to the author(s) from the reviewer. General Comments about the manuscript: (a) Please use [1], [2], etc. and not (1), (2), etc. when you referring to citations (inside your manuscript) of the References section. (b) Please use full stop at the end of your line equations when your paragraph/phrase is ending with them. Use commas at the end of your equation lines when your paragraph/phrase continues. Minor Comments: Page 100: Abstract: Please rewrite as "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) model. Applications of these models to several well-known data sets, exhibiting under-dispersion and over-dispersion, where 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 Generalized 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	Corrected. I disabled the <code>\usepackage{nabbib}</code> in the LATEX pre-amble. Effected. Corrected
--	--	---

	and time). Our results indicate that all the models under consideration (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 behave slightly better for these specific data sets. While this result may not be necessarily generalized to other varieties of over- and under-dispersed data, the investigation of all possible models it is always advisable in order the right choice of model to be achieve the for a given data set.”.	Effected
	Page 101: 1st paragraph: Please rewrite as “... The baseline model for binary data is of course the binomial distribution model. However ,...”. 2nd paragraph: Please rewrite as “To overcome this problems, several distributions have been utilized, especially mixture models such as the beta-binomial in [29], the Kumarasmawany in [17]) and most recently the McDonald's generalized beta-distribution (McGBB) in [22]. Each of these models separately [something (i.e. a verb) is missing!] the probability of success π with beta, Kumaraswamy, 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 over-dispersed binomial data. In other words, no single mixture of 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.”. Please apply also general comment (a) throughout the manuscript. Eq. (2.1): Please rewrite as: $\dots; \quad y = 0, 1, \dots, n, \quad 0 < \pi < 1.$ Please apply general comment (b) throughout the manuscript. Eq. (2.2a)-(2.2b): Please put a comma at the end of (2.2a) and a full-stop at	Done



SDI Review Form 1.6

	the end of (2.2b). Please apply general comment (b) throughout the manuscript. Page 102: Section title: Please rewrite as “The Multiplicative Binomial Model (MBM)”. 1st paragraph: Please rewrite as “In [21] an alternative form of the two-parameter exponential family generalization of the binomial distribution have been proposed, firstly introduced in [1] which itself was based on the original representation in [6], as: [something (e.g. an equation) is missing!]”. 2nd paragraph: Please rewrite as “In [9], the author presented some elegant characteristics of the multiplicative binomial distribution, including its four central moments. The author’s treatment includes...”. Please apply general comment (a) throughout the manuscript. Line before eq. (3.5): Please rewrite as “with $\kappa(\cdot)$ defined previously...”. Eq. (3.5): Delete “,” and put a comma at the end of eq. (3.5). Please apply general comment (b) throughout the manuscript. Page 103: Section title: Please rewrite as “The Com-Poisson Binomial (CPB) Model”. Eq. (4.1.): Please replace symbol $\cdot^{\cdot}$ with $\dots$ (use \dots in latex source) and put a comma at the end. Please apply general comment (b) throughout the manuscript. Line after eq. (4.1): Please rewrite as “with $\pi \in (0,1)$ and $v \in \mathbb{R}$” (in latex, use \mathbb{R} for the symbol of the real numbers’ set). Line after eq. (4.3): Please rewrite as “An approximation to the CPB distribution in the limit $n \rightarrow \infty$ and with $\lambda = n^v p$ is given in [28]”.	Done Effected Yes, equation missing. I have inserted it. Corrected
--	---	--



SDI Review Form 1.6

	Page 104: Section title: Please rewrite as “The Double Binomial Model (DBM)”. 1st paragraph: Please rewrite as “In [10], the double binomial distribution were presented, having pdf form:”. Line before eq. (5.2): Please rewrite as “and factorizes appropriately as”. Line after eq. (5.2): Please rewrite as “The beta-binomial in [29] is, of course, a mixture of the binomial $\text{Bin}(n, \pi)$ and the beta distribution $\text{Beta}(\alpha, \beta)$, where...”. Page 105: 1st paragraph: Please rewrite as “We present in Table 2, the...”. Line after eq. (7.1): Please rewrite as “The $\hat{m}_i$ in Table 2, as above, is given by...”. Paragraph after eq. (7.1): Please rewrite as “...The Wald’s Goodness-Of-Fit (GOF) test statistic seems that the Com-Poisson binomial model fits data best, but again, they all fit the data well.”. Page 107: Last paragraph: Please rewrite as “In Table 5 we present the parameter estimates under...”. Page 108: End of 1st paragraph: Please rewrite as “The Akaike Information Criterion (AIC) and the -2LL values also support the double binomial as the model providing the best fit for this data set.”. Page 109:	Done Done Done Done Done Done
--	---	---



SDI Review Form 1.6

	Eq.: Please rewrite as “$\pi_{ij} = \frac{1}{1 + \exp(-\mathbf{x}'\mathbf{b})}$”. Latex source: $\pi_{ij} = \frac{1}{1 + \exp(\mathbf{x}'\mathbf{b})}$. Page 110: 1st paragraph: Please rewrite as “...The index is a five item measure of religious involvement; for details, see [23].”. Line before eq. (9.1): Please rewrite as “...The model of interest is then given by”. Page 112: 1st paragraph: Please rewrite as “This example is from [25] and was originally published in http://www.stat.sc.edu/kerrie/cardiodata.html”. 2nd paragraph: Please rewrite as “Tests of heart functions were conducted.”. Eq. (9.2): Please use for Y the latex script $\mathbf{Y}$. See also suggestion for the eq. in pg. 109. Page 113: Table 9: Please remove the right vertical line or add a left vertical line to the table. 1st paragraph: Please delete the following part (duplicated with the one in previous page): “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. 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	Corrected Corrected Corrected Effectuated Done
--	--	---



SDI Review Form 1.6

	dose=1 if High and 0 if low dosage.” Eqs: In eq. (9.2) and the one above that, please use for Y the latex source $\{\bf Y\}$. See also suggestions in pg. 112. Page 114: Eq. (9.3): Please rewrite as “$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}'\boldsymbol{\beta}$, that is,”. Latex source script: $\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}'\boldsymbol{\beta}$, that is,”. Use the same technique for the eq. in the line before Table 10. Last paragraph: Please rewrite as “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, indicates a strong over-dispersion 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, the $\hat{\phi}$ 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:”. Page 115: First paragraph: Please rewrite as “The Wald GOF is very susceptible in this case when $r = 0$. For instance, note that for cluster 2 and $r = 0$, the contribution...”. Line after Table 11: Please rewrite as “The values of G^2 and X^2 columns in Table 11 are...”.	Done Effectuated Done Done Done
--	---	--

	Page 116: First paragraph: Please rewrite as “Clearly, each of the models under investigation in this paper (excluding the binomial)...”.	Done Done Done Done
<u>Optional/General</u> comments		