

On Modeling the Double and Multiplicative Binomial Models as Log-Linear Models

Bayo H Lawal*

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

Abstract

In this paper we have fitted the double binomial and multiplicative binomial distributions as log-linear models using sufficient statistics. This approach is not new as several authors have employed this approach, most especially in the analysis of the Human sex ratio in [1]. However, obtaining the estimated parameters of the distributions may be problematic, especially for the double binomial where the parameter estimate of π may not be readily available from the Log-Linear (LL) parameter estimates. Other issues associated with the LL approach is its implementation in the generalized linear model with covariates. The LL uses far more parameters than the procedure that employs conditional log-likelihoods functions where the marginal likelihood functions are minimized over the parameter space. This is the procedure employed in SAS PROC NLMIXED. The two procedures are essentially equivalent for frequency data. For models with covariates, the LL uses far more parameters and the marginal likelihood functions approach are employed here on three data set having covariates.

Keywords

Double Binomial; Multiplicative Binomial; Log-Linear; Marginal Likelihood Functions

Introduction

In the formulations of the multiplicative binomial distribution, in Altham and its corresponding double binomial distribution in Efron, both distributions were characterized with intractable normalizing constants $c(n, \psi)$ and $c(n, \pi)$ respectively [2,3]. Consequently, these models were implemented in by utilizing a generalized linear model with a Poisson distribution and log link to the frequency data. This approach has earlier being similarly employed in [1,4]. This approach which employs joint sufficient statistics in both distributions was earlier proposed in Lindley & Mersch [5].

Both distributions are fitted using a Poisson regression model having sufficient statistics from both distributions as explanatory variables with the frequencies being the mean dependent variables. For the Double binomial model (DBM), the sufficient statistics are $y \log(y)$ and $(n - y) \log(n - y)$. Similar joint sufficient statistics for the multiplicative binomial model (MBM) are y and $y(n - y)$ with the offset being

$$Z = \log \binom{n}{y} \text{ for both models. For instance, for the DBM the model would be:}$$

$$\log(n_i / z) = y + \theta \quad (1)$$

$$\text{Where } \theta = \theta_1 + \theta_2$$

$$\theta_1 = \begin{cases} 0 & \text{if } y = 0 \\ y \log(y) & \text{otherwise} \end{cases}, \quad \theta_2 = \begin{cases} 0 & \text{if } y = n \\ n - y \log(n - y) & \text{otherwise} \end{cases}$$

Similarly, for the multiplicative binomial, the model is estimated by the log-linear model (or Poisson Model):

$$\log\left(\frac{n_i}{z}\right) = y + \delta \quad (2)$$

$$\text{Where, } \delta = \delta_1 + \delta_2 \text{ and}$$

$$\delta_1 = \begin{cases} 0 & \text{if } y = 0 \\ y & \text{otherwise} \end{cases}, \quad \delta_2 = \begin{cases} 0 & \text{if } y = n \\ y(n - y) & \text{otherwise} \end{cases}$$

However, in recent times, both models have been fully formulated with the intractable

Article Information

DOI: 10.31021/acs.20181103

Article Type: Research Article

Journal Type: Open Access

Volume: 1 Issue: 1

Manuscript ID: ACS-1-103

Publisher: Boffin Access Limited.

Received Date: 20 October 2017

Accepted Date: 20 November 2017

Published Date: 12 January 2018

*Corresponding author:

Bayo H Lawal

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

Citation: Lawal BH. On Modeling the Double and Multiplicative Binomial Models as Log-Linear Models. Adv Comput Sci.2018 Jan; 1(1):103

Copyright: © 2018 Lawal BH. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 international License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

normalizing constants fully formulated. These distributions are described later in this paper with the normalizing constants fully formulated. In this study, we compare fitting these two probability models to two example frequency data, two data set examples arising from teratology studies, and randomized complete block design example having binary outcome by the method of sufficient statistics described above and by the method of numerically maximizing the marginal likelihood function arising from engaging the problem as a mixed generalized linear model. The SAS PROC NLMIXED which performs the Maximum Likelihood estimation numerically by using the Adaptive Gaussian Quadrature and Newton-Raphson optimization algorithm. We shall designate the sufficient Statistics-Poisson regression approach as LL, while the marginal likelihood function maximization via PROCNLMIXED is designated MgL in this study. The sufficient statistics procedure uses a Poisson regression with an offset and is implemented in SAS PROC GENMOD.

Models Under Consideration

We describe in the following section, the two probability distribution models employed in this paper.

The Multiplicative Binomial Model-MBM

[2,6,7] Lovinson proposed an alternative form of the two-parameter exponential family generalization of the binomial distribution first introduced by [2] which itself was based on the original Cox representation as:

$$f(y) = \frac{\binom{n}{y} \psi^y (1-\psi)^{n-y} \omega^{y(n-y)}}{\sum_{j=0}^n \binom{n}{j} \psi^j (1-\psi)^{n-j} \omega^{j(n-j)}}, y = 0, 1, \dots, n \quad (3)$$

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

The normalizing constant is

$$c(n, \psi) = \sum_{j=0}^n \binom{n}{j} \psi^j (1-\psi)^{n-j} \omega^{j(n-j)}$$

the denominator expression in (3) in this case. [8] presented an elegant characteristics of the multiplicative binomial distribution including its four central moments. His treatment includes generation of random data from the distribution as well as the likelihood profiles and several examples-some of which are similarly employed in this presentation. Following [8] the probability π of success for the Bernoulli trial, that is, $P(Y = 1)$ can be computed from the following expression in (4) as:

$$p_i = \psi_i \frac{K_{n-i}(\psi, \omega)}{K_n(\psi, \omega)}, \text{ for } i = 1 \quad (4)$$

Where:

$$K_{n-a}(\psi, \omega) = \sum_{y=0}^{n-a} \binom{n-a}{y} \psi^y (1-\psi)^{n-a-y} \omega^{(y+a)(n-a-y)}, a = 1, 2, \dots, n \quad (5)$$

with p defined as in (4), Ψ therefore can be defined as the probability of success weighted by the intra-units association measure ω which measures the dependence among the binary responses of the n units. Thus if $\omega = 1$, then $p = \Psi$ and we have independence among the units. However, if $\omega = 1$, then, $p = \Psi$ and the units are not independent.

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

$$E(Y) = np_1 \quad (6a)$$

$$\text{Var}(Y) = np_1 + n(n-1) p_2 - np_1^2 \quad (6b)$$

The Double Binomial (DBM) Model

In Feirer et al. [7], the double binomial distribution was presented, having the pdf form:

$$f(y; \pi, \phi) = \frac{\binom{n}{y} [y^y (n-y)^{n-y}]^{1-\phi} [\pi / (1-\pi)]^{y\phi}}{\sum_{j=0}^n \binom{n}{j} [j^j (n-j)^{n-j}]^{1-\phi} [\pi / (1-\pi)]^{j\phi}}, y = 0, 1, \dots, n \quad (7)$$

Again, the normalizing constant in this case is the denominator expression given by

$$c(n, \psi) = \sum_{j=0}^n \binom{n}{j} [j^j (n-j)^{n-j}]^{1-\phi} [\pi / (1-\pi)]^{j\phi}$$

Applications

We apply the models discussed above to two frequency data and to teratology data sets having four and two treatment groups. We first present the analyses for the two frequency data sets in Tables 1 through 5. The estimation of the parameters under each model for the MgL approach uses SAS PROC NLMIXED, using the following log-likelihoods for the MBM (LL1) and DBM (LL2) respectively. The procedure was discussed earlier in the paper.

$$LL1 = \log \left(\binom{n}{y} \right) + y \log(\psi) + y(n-y) \log \omega - \log \left[\sum_{j=0}^n \binom{n}{j} [\psi^j (1-\psi)^{n-j} \omega^{j(n-j)}] \right]$$

$$LL2 = \log \left(\binom{n}{y} \right) + (1-\phi) [y \log(\psi) + (n-y) \log(n-y)] + y\phi \log \left(\frac{\pi}{1-\pi} \right) - \log \left[\sum_{j=0}^n \binom{n}{j} [j^j (n-j)^{n-j}]^{1-\phi} \left(\frac{\pi}{1-\pi} \right)^{j\phi} \right]$$

Example Data Set I-Geissler Data

The Distribution of males in 6115 families with 12 children in Saxony, previously analyzed in Sokal & Rohlf [10] is presented in Table 1. The data is originally from Geissler [11] and had similarly been analyzed in [12]. Here $Y \sim \text{binomial}(12, \pi)$. The frequencies are presented as counts having a total sum of 6115. The observed mean for the data is $\bar{y} = 6.2306$ and the corresponding variance is $s^2 = 3.4898$. Under the binomial model, the estimated mean is 6.2304 and estimated variance being 2.9956. Hence the estimated dispersion parameter $DP = \bar{y} / s^2 = 2.07$ indicating over-dispersion in the data. The estimated probability of occurrence under the binomial model is $\hat{\pi} = 0.5192$. The binomial does not fit the data ($X^2 = 110.5051$ on 11 d.f., $p\text{-value} = 0.0000$) because the variance of the data is grossly under estimated by the model. The results of the application of the double binomial and the multiplicative models to this data are presented in Table 2. Further, the mixed model approach is based on one more degree of freedom as it estimates one parameter less than the LL model approach. The Mixed model approach gives the parameter estimates of the distribution. We can obtain the equivalent parameters estimates from the Log-linear (LL) approach for the multiplicative models as follows:

$$\hat{\omega} = e \times p(\hat{\delta}) = e \times p(-0.02615) = 0.9742, \hat{\psi} = 1 / [1 + e \times p\{-(\hat{\delta} + \hat{y})\}] = 0.5165 \quad (8)$$

For the DB, $\hat{\phi}$ can equivalently be obtained as $1 - \hat{\theta} = 1 - 0.140205 = 0.8598$, but the estimated probability $\hat{\pi}$ seems intractable in this case and no equivalent solution is available in this case. We may note here that the estimate $\hat{\psi} = 0.5165$ under the multiplicative model is not an estimate of the success probability π . For this data, we must use the expressions in (4) and (5) to obtain this estimate. Here, $k_{(\eta-1)} = 0.42723$ and $k_{(\eta-1)} = 0.42499$.

$$\text{Consequently, } \hat{\pi} = \hat{\psi} \left(\frac{k_{(\eta-1)}}{k_{\eta}} \right) = 0.5165 \left(\frac{0.42723}{0.42499} \right) = 0.5192.$$

The mean and Variance can therefore be computed respectively from (6a) and (6b). Alternatively, the means and variances can be empirically obtained from the fitted models using the elementary principles of

$$E(Y) = \sum_{i=0}^n y_i \hat{p}_i \text{ and}$$

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

Table 1: Distribution of Males in 6115 families with 12 children

Log-Linear			Marginal Likelihood	
	MBM	DBM	MBM	DBM
MLE	Int=0.853840	Int=-3.096918	$\hat{\theta}=0.5165$	$\hat{\pi}=0.5192$
	$\hat{y}=0.092157$	$\hat{y}=0.065977$	$\hat{\omega}=0.9742$	$\hat{\phi}=0.8598$
	$\hat{\delta}=-0.026150$	$\hat{\theta}=0.140205$	$\hat{\pi}=0.5192$	
-2LL	104.5372	103.1298	24985.8	24984.3
AIC	110.5372	109.1297	24990	24988.3
X^2	14.5354	13.0421	14.5354	13.0421
G^2	14.4686	13.0616	14.4686	13.0616
d.f	10	10	11	11

Table 2: Parameter estimates under the five Models

Y	Count	Double Binomial					Multiplicative Binomial				
		$\hat{p}_i$	$\sum \hat{p}_i$	$\sum y_i \hat{p}_i$	V_1	V_1	$\hat{p}_i$	$\sum \hat{p}_i$	$\sum y_i^2 \hat{p}_i$	$\sum y_i^2 \hat{p}_i$	V_2
0	3	0.0005	0.0005	0.0000	0.0000	0.0000	0.0004	0.0004	0.0000	0.0000	0.0000
1	24	0.0038	0.0043	0.0038	0.0038	0.0038	0.0037	0.0041	0.0037	0.0037	0.0037
2	104	0.0171	0.0214	0.0379	0.0721	0.0706	0.0171	0.0212	0.0380	0.0723	0.0708
3	286	0.0503	0.0717	0.1889	0.5250	0.4893	0.0508	0.0721	0.1905	0.5298	0.4936
4	670	0.1068	0.1785	0.6160	2.2333	1.8538	0.1072	0.1793	0.6194	2.2454	1.8617
5	1033	0.1698	0.3483	1.4652	6.4792	4.3325	0.1694	0.3487	1.4665	6.4812	4.3304
6	1343	0.2067	0.5550	2.7056	13.9220	6.6015	0.2057	0.5544	2.7008	13.8867	6.5924
7	1112	0.1938	0.7488	4.0622	23.4178	6.9164	0.1933	0.7478	4.0542	23.3605	6.9239
8	829	0.1390	0.8878	5.1743	32.3146	5.5414	0.1396	0.8874	5.1712	32.2961	5.5553
9	478	0.0748	0.9626	5.8472	38.3710	4.1810	0.0755	0.9629	5.8511	38.4154	4.1803
10	181	0.0289	0.9915	6.1364	41.2628	3.6074	0.0291	0.9920	6.1418	41.3227	3.6009
11	45	0.0074	0.9989	6.2178	42.1579	3.4972	0.0071	0.9992	6.2204	42.1873	3.4939
12	7	0.0011	1.0000	6.2306	42.3116	3.4915	0.0008	1.0000	6.2306	42.3094	3.4893

Table 3: Empirical Means and Variances for both the DBM and MBM

Var of Y being $E(Y^2) - [E(Y)]^2$. These distributions are displayed in Table 3.

In the above Table, Some of the columns are self explanatory. The columns labeled V_1 and V_2 are cumulative values of

$$\sum y_i^2 \hat{p}_i - \left[\sum y_i \hat{p}_i \right]^2$$

for both models respectively. Thus, the mean is the value at $y = 12$. The variance for the DBM for instance, is computed as $42.3116 - (6.2306)^2 = 3.4915$.

In Table 4 are presented the expected values under both models for the two approaches (LL & MgL), both approaches give exact results as expected. The Table also displays the mean of the distributions under both approaches as well as the empirical variances, designated here as var. We recall that for the observed data in Table 1, $\bar{y} = 6.2306$ and $s^2 = 3.4898$. We see from Table 4, that while the two models estimate the mean of the data well, the estimated variance under the binomial model of $12(0.5192)(1-0.5192) = 2.9956$ underestimates the observed variance of the data, and this explains the poor fit to the

data by the binomial model. On the other hand, for the two models, the variance of the observed data are reasonably well estimated, because of the extra parameter in the model (dispersion parameter) of ϕ and ω for the DBM and MBM respectively.

Table 4 also displays the corresponding Pearson's X^2 and the corresponding degrees of freedom (d.f.). Clearly, for this data set, both the double binomial and the multiplicative models fit the data well with the Double binomial being slightly providing a better fit. Although the expected values generated are the same for both fitting approaches, we see that the marginal likelihood (MgL) approach gives a more parsimonious model because it is based on one more degree of freedom.

Data Example II

This example is taken from Nelder & Mead [13] 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. There were a total of 209 candidates for the exam and Table 5 gives the distribution of these scores for the 209 candidates.

Y	Count	DBM		MBM	
		MgL	LL	MgL	LL
0	3	2.956	2.956	2.3486	2.3487
1	24	23.3861	23.386	22.5809	22.581
2	104	104.3138	104.3137	104.8482	104.8484
3	286	307.7531	307.7531	310.8921	310.8923
4	670	652.8854	652.8858	655.6551	655.6551
5	1033	1038.546	1038.546	1036.077	1036.077
6	1343	1264.242	1264.242	1257.907	1257.907
7	1112	1185.039	1185.04	1182.293	1182.293
8	829	850.0634	850.0639	853.7711	853.7724
9	478	457.2186	457.2188	461.9646	461.9659
10	181	176.8358	176.8359	177.7841	177.785
11	45	45.2369	45.2369	43.6925	43.6928
12	7	6.5246	6.5246	5.1858	5.1858
G^2		13.0421	13.0421	14.4686	14.4686
X^2		13.0612	13.0612	14.5354	14.5354
d.f.		10	9	10	9
P-value		0.2213	0.1607	0.1527	0.1066
$\bar{y}$	6.2306	6.2306	6.2306	6.2306	6.2306
var	3.4898	3.4893	3.4915	3.4915	3.4915

Table 4: Expected values under the two models and Approaches with corresponding Pearson's X^2 Statistic values

Results

The results of applying both models DBM and MBM to the data using both approaches (LL and MM) are presented in Table 5. Again, both approaches lead to the same results in terms of expected values. However, the MgL models have one more degree of freedom under both models than the LL approach. Again, to get equivalent parameter estimates from the LL model, we have, for the multiplicative model, $\hat{\omega} = \exp(\hat{\delta}) = \exp(-0.2168) = 0.8051$, $\hat{\psi} = 1/[1+\exp\{-(\hat{\delta} + \hat{\gamma})\}] = 0.3630$. For the double binomial, an equivalent estimate for $\hat{\phi}$ is $\hat{\phi} = 1 - \hat{\theta} = 1 - 0.6072 = 0.3928$. As discussed earlier, the corresponding estimate for π is not readily available. For this model, the multiplicative model is the most parsimonious and fits the data very well.

Regression Model Formulations

When there are covariates in our data, the sufficient statistic approach here into referred to as Log-linear (LL) does not lend itself to easier formulation and implementation. Lindsey and Altham [13] employed this approach to fitting amongst others, the two models considered in this study to the distribution of males in families in Saxony during 1885-1976 (the human sex ratio data). This approach employs far too many parameters, 13 to be precise when the same group of models can be implemented with only four parameters with the same results. Further, the implementations under this approach are not readily available. Thus in this study, we will employ the alternative MgL procedure that utilizes PROC NLMIXED in SAS. One advantage of this is that it will be based on more degrees of freedom than the log-linear model. We see for the frequency data in Tables 1 for instance, that the LL approach is based on 1 d.f. more than the MgL model based.

Example I: Teratology-Ossification on the Phalanges

Teratology is the study of abnormalities of physiological development. The offspring of animals that were exposed to a toxin during pregnancy are studied for malformation. The number of malformed offspring in a litter of size n is not typically distributed binomial because the responses of the offspring from the same litter are not independent, hence their sum does not constitute a binomial r.v. Thus, data in teratological studies exhibit over-dispersion because

of the correlation among responses from off springs in the same litter.

[14] report data from a completely randomized design that studies the teratogenicity of phenytoin in 81 pregnant mice. The treatment structure of the experiment is an augmented factorial. In addition to an untreated control, mice received 60 mg/kg of phenytoin (PHT), 100 mg/kg of trichloropropene oxide (TCPO), and their combination. The design was augmented with a control group that was treated with water. As in [15], the two control groups are combined here into a single group. The presence or absence of ossification in the phalanges on both the right and left forepaws on each of the fetuses is considered a measure of the teratogenic effect. The data is presented below. For the control for instance, there are 35 pair of observations designated as (n, r) . Thus, the numbers of rats in each group are respectively $\{35, 19, 16, 11\}$.

```

control 35 8 8 9 9 7 9 0 5 3 3 5 8 9 10 5 8 5 8 1 6 0 5
        8 8 9 10 5 5 4 7 9 10 6 6 3 5 8 9 7 10 10 10
        1 6 6 6 1 9 8 9 6 7 5 5 7 9 2 5 5 6 2 8 1 8
        0 2 7 8 5 7
PHT     19 1 9 4 9 3 7 4 7 0 7 0 4 1 8 1 7 2 7 2 8 1 7
        0 2 3 10 3 7 2 7 0 8 0 8 1 10 1 1
TCPO    16 0 5 7 10 4 4 8 11 6 10 6 9 3 4 2 8 0 6 0 9
        3 6 2 9 7 9 1 10 8 8 6 9
PHT2    11 2 2 0 7 1 8 7 8 0 10 0 4 0 6 0 7 6 6 1 6 1 7

```

Suppose Y_{ij} denote the number of deaths in litter i . Further, let p_{ij} be the probability of a fetus in litter i dying. Y_{ij} has the overdispersed binomial distribution with mean $n_i p_{ij}$ and variance $n_i p_{ij} (1 - p_{ij}) \phi$, with ϕ characterizing the correlation between any two fetuses within the same litter.

The probability of fetal death is modeled with the logit link viz:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_2 z_{2i} + \beta_3 z_{3i} + \beta_4 z_{4i} \quad (9)$$

We have assumed here that β_0 is similar across litters, and that,

Y	Count	Double Binomial		Multiplicative Binomial	
		MgL	LL	MgL	LL
0	3	2.956	2.956	2.3486	2.3487
1	24	23.3861	23.386	22.5809	22.581
2	104	104.3138	104.3137	104.8482	104.8484
3	286	307.7531	307.7531	310.8921	310.8923
4	670	652.8854	652.8858	655.6551	655.6551
5	1033	1038.546	1038.546	1036.077	1036.077
6	1343	1264.242	1264.242	1257.907	1257.907
7	1112	1185.039	1185.04	1182.293	1182.293
8	829	850.0634	850.0639	853.7711	853.7724
9	478	457.2186	457.2188	461.9646	461.9659
$\bar{y}$	1.5598	1.5598	-1.5598	1.5598	1.5598
S^2	2.8245	2.6428	2.6428	20.811	20.811
		$\hat{\pi}=0.1537$	Int= -7.7413	$\hat{\psi}=0.3630$	Int= 4.1887
		0.3928	-0.6701	0.8051	-0.3457
			0.6072	$\hat{\pi}=0.1733$	-0.2168
-2LL		713	51.4985	703.1	41.6374
AIC		717	57.4985	707.1	47.6374
X^2		13.9366	13.9366	2.6948	2.6948
G^2		12.9164	12.9164	3.0554	3.0554
d.f.		7	6	7	6

Table 5: Expected Values under the two models and approaches with corresponding Pearson's X^2 Statistic Values

$$z_2 = \begin{cases} 1 & \text{if PHT} \\ 0 & \text{otherwise} \end{cases}, z_3 = \begin{cases} 1 & \text{if TCPO} \\ 0 & \text{otherwise} \end{cases}, z_4 = \begin{cases} 1 & \text{if PHT}_2 \\ 0 & \text{otherwise} \end{cases}$$

Thus, the 'control' treatment is the reference category in this set up. Our analysis begins by fitting three different models to the data. These models are described briefly below:

- The model that assumes $p_0 = p_1 = p_2 = p_3$ with a common dispersion parameter ϕ or ω for the double binomial and the multiplicative models respectively, and, where p_0 and p_1, p_2, p_3 refer respectively to the corresponding probabilities in the control and other treatment groups.

$$\text{Here, } p_{ij} = \frac{1}{[1 + e^{x p(-\beta_0)}]}, \phi = e^{a_0} \text{ and}$$

$\omega = \exp(c_0)$ with $a_0 \neq c_0$. These ensure that the dispersion parameters are positive.

- The model here has $p_i = p_j$ $i \neq j$ and the dispersion parameters are functions of the covariates. That is, $\phi = \exp(a_0 + a_2 z_2 + a_3 z_3 + a_4 z_4)$ and $\omega = \exp(c_0 + c_2 z_2 + c_3 z_3 + c_4 z_4)$.
- The model here has $p_i \neq p_j$ with the p s modeled as in (9) and the dispersion parameters are modeled as functions of the covariates as in the preceding case.

Results

From the results in Table 6, the two cases (II & III) with variable dispersion parameters fit better than the model in case I, where the dispersion is uniform across the four groups. Of the models in Cases II and III, the models in case III fits much better than those in case II. Case II models assume that the four groups have a common estimated probability π , which are estimated respectively as 0.2125 and 0.2158 in both the DBM and MBM. However, the models in III which assume heterogeneous success probabilities across the four groups and variable dispersion parameters (that are functions of the covariates) fit better than those in case II. The DBM here is based

on $X^2 = 115.1333$ on 72 d.f. The estimated π s under the MBM are functions of n , hence these values are different for different n in the final model (Case III). We may note here that the Ψ s should not be mistaken for the success probabilities.

Data Example II-Trout Egg Data

The data in Table 7 from Manly [16] relate to the number of surviving eggs from boxes of eggs that were buried at five different locations in a stream and at four different times a box from a location was sampled. The data is presented as y/n where y is the number surviving and n is the number of eggs in the box.

The model of interest here is:

$$\log it(p_{ij}) = \beta_0 + \sum_{k=1}^4 \beta_k z_k + \sum_{l=1}^3 \beta_l x_l \quad (10)$$

where z_k are four dummy variables for location effects, and x_l are three dummy variables representing the Time effects. The structure here is that of a randomized block design having locations as blocks and Survival times as treatments. Thus, the structure of the Pearson's X^2 would be for Location (L) and Survival time (T):

Source	d.f.
L T	4
T L	3
Residual*	12

The degree of freedom of 12 refers only to the binomial model. For all other distributions, the d.f. must account for the additional dispersion parameter estimates. Under the Binomial model $X^2 = 63.9639$ on 12 d.f, giving an estimated dispersion parameter of $5.3303 > 1$, indicating that the data is highly overdispersed.

Because of the overdispersion in the data, we now apply our models, DBM and the MBM to the data, giving the results in Table 8.

Parameters	Case I		Case II		Case III	
	DBM	MBM	DBM	MBM	DBM	MBM
	$\pi = 0.4808$	$\psi = 0.4977$	$\pi = 0.2125$	$\psi = 0.4977$	$\psi_0 = 0.7956$	$\psi_0 = 0.5566$
		$\pi = 0.4912$	-	$\pi = 0.2158$	$\pi_1 = 0.2158$	$\psi_2 = 0.2814$
			-	-	$\psi_2 = 0.4949$	$\psi_2 = 0.4988$
			-	-	$\pi_3 = 0.0000$	$\psi_3 = 0.4484$
	$\hat{\phi} = 0.1224$	$\omega = 0.7463$	$\phi_0 = 0.0000$	$\phi_0 = 0.7342$	$\phi_0 = 0.1732$	$\omega_0 = 0.7506$
			$\phi_1 = 0.6830$	$\omega_1 = 0.7658$	$\phi_1 = 0.6856$	$\omega_1 = 0.9120$
			$\omega_2 = 0.0574$	$\omega_2 = 0.7981$	$\phi_2 = 0.2219$	$\omega_2 = 0.7981$
			$\phi_3 = 0.1054$	$\phi_3 = 0.6028$	$\phi_3 = 0.0004$	$\omega_3 = 0.6145$
-2LL	329.3874	340.1538	309.7278	328.3978	291.8264	295.9996
AIC	333.3874	344.1538	319.7278	338.3978	307.8264	311.9996
X ²	161.85	158.9242	144.0727	159.0039	115.1333	118.7587
d.f.	78	78	75	75	72	72

Table 6: Parameter estimates for the Models in all the Cases

Location in stream	Survival Period (weeks)			
	4	7	8	11
1	89/94	94/98	77/86	141/155
2	106/108	91/106	87/96	104/122
3	119/123	100/130	88/119	91/125
4	104/104	80/97	67/99	111/132
5	49/93	11/113	18/88	0/138

Table 7: Number of Surviving eggs against number of eggs in a box

Models in (A) fit both the double binomial and the multiplicative binomial with constant dispersion parameter. For this group of models, the multiplicative binomial performs much better with constant dispersion parameter of 0.9884, very close to 1, indicating there is partial independence in the data ignoring the effects of locations. Models in B, have variable dispersion parameters that are functions of the covariates (Time), that is, $\phi = \exp(a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3)$ and $\omega_i = \exp(c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3)$. Under this formulation, the double binomial computation does not converge, but that of the multiplicative binomial converged. This model gives a Pearson X² of 5.0120 on d.f. The results of this final model are presented in Table 9. Note that for the multiplicative, the estimated probabilities of success π which are not the same as the ϕ in the model formulation in (3) are computed using expressions in (4) and (5). Note that $\psi \neq \pi$. The column labeled $\sum X^2$ gives the cumulative contributions of observation towards X². The value 5.0120 is the sum of all 20 contributions towards X². Under the final multiplicative model, the estimated average probabilities of surviving in the first 4, 7, 8 and 11 weeks are respectively {0.8854, 0.6999, 0.6831, 0.6656}.

Example III-Teratology

The data below is from an unpublished toxicological studies on pregnant mice, Kupper & Haseman (1978). The study is concerned with the effect of compounds on fetal death or the occurrence of some abnormalities in physiological development. Ten pregnant female mice in each of two groups (one group is the control and the other is the treated group) are employed in the study. The data is presented

below for (y/n). y is the number of off springs dead in n litters.

control 10 0/5, 2/6, 0/7, 0/7, 0/8,

0/8, 0/8, 1/9, 2/9, 1/10

TRT 10 0/5, 2/5, 1/7, 0/8, 2/8

3/8, 0/9, 4/9, 1/10, 6/10.

If we let π_{ij} denote the probability of death for fetus j in litter i. Then, we would model this probability for both models with the logit link, viz:

$$\log it(\pi_{ij}) = \beta_0 + \beta_1 trt \quad (11)$$

where (trt=1 if treatment group and 0, otherwise). Again here, we fit three competing models (17) viz:

1. The model that assumes $\pi_1 = \pi_1$ with a common dispersion parameter ϕ or ω for the double binomial and the multiplicative models respectively, and, where π_0 and π_1 refer respectively to the corresponding probabilities in the control and treatment groups. Here, $\log it(\pi_{ij}) = \beta_0$, $\phi = a_0$ and $\omega = c_0$ with $a_0 \neq c_0$
2. The model here has $\pi_0 = \pi_2$ and the dispersion parameters are functions of the covariate. That is, $\phi = a_0 + a_1 trt$ and $\omega = c_0 + c_1 trt$
3. The model here has $\pi_0 \neq \pi_1$ with the π S modeled as in (11) and the dispersion parameters are modeled as functions of the covariates as in preceding case.

Source	Models A		Models B	
	BIN	DBM	MBM	MBM
		$\hat{\phi} = 0.3116$	$\hat{\omega} = 0.9884$	-
Residual (X^2)	63.9639	27.2582	18.5493	5.0120
d.f.	12	11	11	8
-2LL	141.0292	120.4564	125.7706	112.7608
AIC	157.0292	138.4564	143.7706	136.7608

Table 8: Results of Analysis of Data in Table 7.

#	n	y	ψ	π_1	π_2	m_i	ω_i	s^2	$\sum X^2$
1	94	89	0.9895	0.9863	0.9727	92.7093	1.003	1.2635	0.1484
2	98	94	0.9042	0.9059	0.8207	88.7817	0.9997	8.3863	0.4551
3	86	77	0.9226	0.8714	0.7592	74.9446	1.009	8.2376	0.5115
4	155	141	0.7915	0.9327	0.87	144.5644	0.9903	12.016	0.5994
5	108	106	0.9868	0.9821	0.9645	106.0685	1.003	1.8759	0.5994
6	106	91	0.8823	0.8844	0.7822	93.7495	0.9997	10.894	0.6801
7	96	87	0.9044	0.8413	0.7076	80.7662	1.009	10.4533	1.1612
8	122	104	0.7509	0.8805	0.7755	107.4158	0.9903	17.1125	1.2698
9	123	119	0.9736	0.9633	0.928	118.4912	1.003	4.2347	1.272
10	130	100	0.787	0.7902	0.6244	102.7271	0.9997	21.7874	1.3444
11	119	88	0.8235	0.7383	0.5447	87.8609	1.009	16.3371	1.3446
12	125	91	0.5977	0.7114	0.5077	88.9207	0.9903	50.801	1.3933
13	104	104	0.9782	0.9711	0.943	100.9938	1.003	2.8698	1.4827
14	97	80	0.8183	0.8206	0.6734	79.5995	0.9997	14.3821	1.4848
15	99	67	0.8504	0.7776	0.6042	76.9785	1.009	13.1452	2.7782
16	132	111	0.6442	0.7911	0.6268	104.4282	0.9903	37.8028	3.1918
17	93	49	0.5274	0.5241	0.2743	48.7372	1.003	20.3948	3.1932
18	113	11	0.1006	0.0986	0.0097	11.1422	0.9997	10.0944	3.195
19	88	18	0.1238	0.1869	0.0346	16.4498	1.009	10.8223	3.3411
20	138	0	0.0431	0.0121	0.0001	1.6708	0.9903	1.7055	5.012

Table 9: Parameter estimates under the multiplicative model with variable dispersion parameter.

The results of these models are presented in Table 10.

Appendix A. SAS Program for the Example

```
options nodate nonumber ls=85 ps=66;
```

```
data ex1;
```

```
do L=1 to 5;
```

```
do T=1 to 4;
```

```
input y n @@;
```

```
output;
```

```
end; end;
```

```
datalines;
```

```
89 94 94 98 77 86 141 155
```

```
106 108 91 106 87 96 104 122
```

```
119 123 100 130 88 119 91 125
```

```
104 104 80 97 67 99 111 132
```

```
49 93 11 113 18 88 0 138
```

```
;
```

```
run;
```

```
proc print;
```

```
run;
```

```
/*generate indicator variables for Location*/;
```

```
data w1;
```

```
set ex1;
```

```
array x(5) z1-z5;
```

```
do j=1 to 5;
```

```
if j=L then x(j)=1;
```

```
else x(j)=0;
```

```
end;
```

```
drop j;
```

```
run;
```

```
/*generate indicator variables for Time*/;
```

```
data w2;
```

```
set ex1;
```

```
array d(4) x1-x4;
```

```
do k=1 to 4;
```

```
if k=T then d(k)=1;
```

```
else d(k)=0;
```

```
end;
```

```
drop k;
```

```
run;
```

```
data new;
```

```
merge w1 w2;
```

```

run;

proc sort data=new;
by T;
run;

proc nlmixed data=new tech=newwrap maxit=2000;
parms b0=-0.1 b1=1.1 b2=0.4 b3=.1 b4=0.2 s1-s3=0.0 a0=0 a1=0
a2=0 a3=0;
lp=b0+b1*z1+b2*z2+b3*z3+b4*z4+s1*x1+s2*x2+s3*x3;
lr=a0+a1*x1+a2*x2+a3*x3;
omega=exp(lr);
p=1/(1+exp(-lp));
sum=0.0;
do j=0 to n;
z1=lgamma(n+1)-lgamma(j+1)-lgamma(n-j+1);
u=z1+ j*log(p) + (n-j)*log(1-p) + j*(n-j)*log(omega);
sum=sum+exp(u);
end;
keep sum;
z2=lgamma(n+1)-lgamma(y+1)-lgamma(n-y+1);
LL=z2+ y*log(p) + (n-y)*log(1-p) + y*(n-y)*log(omega)-log(sum);
model y~general(LL);
predict p out=aa;
predict omega out=bb;
run;
Ods rtf close;
data q1;
set aa;
psi=pred;
run;
data q2;
set bb;
omega=pred;
run;
data qq4;
merge q1 q2;
suma=0;
do j=0 to n;
zz1=lgamma(n+1)-lgamma(j+1)-lgamma(n-j+1);
u1=zz1+ j*log(psi) + (n-j)*log(1-psi) + j*(n-j)*log(omega);
suma=suma+exp(u1);
end;
sumb=0;
do k=0 to n-1;
zz2=lgamma(n)-lgamma(k+1)-lgamma(n-k);
u2=zz2+ k*log(psi) + (n-k-1)*log(1-psi) + (k+1)*(n-k-1)*log(omega);
sumb=sumb+exp(u2);
end;
sumc=0;
do t=0 to n-2;
zz3=lgamma(n-1)-lgamma(t+1)-lgamma(n-1-t);
u3=zz3+t*log(psi)+(n-t-2)*log(1-psi)+(t+2)*(n-t-2)*log(omega);
sumc=sumc+exp(u3);

```

```

end;
/* Generate p1,p2, expected values and variances */;
p1=psi*(sumb/suma);
p2=(psi*psi)*(sumc/suma);
exp=n*p1;
var=n*p1+(n*(n-1)*p2)-(n*n*p1*p1);
/* Generate Wald, LRT and Pearson's GOFs */;
wald+((y-exp)**2)/var;
if y=0 then lrt=0;
else lrt+2*y*log(y/exp);
XX+((y-exp)**2)/exp;
run;
proc print data=qq4;
var n y psi p1 p2 exp omega var XX LRT Wald;
format psi p1 p2 exp omega var xx LRT Wald 10.4;
run;

```

Results

Results from Table 10 show that for cases I to III, the model for case II is the most parsimonious. The difference between the Likelihood-test statistic, G^2 between models II and III being 0.2307 on 1 d.f (p-value=0.6310), which is not significant. We have used the G^2 rather than the Wald or Pearson's X^2 because only the G^2 statistic has the partitioning property, (see, [18]). However, while this model seems the best, it does not tell us much about the probability of success (π_i , $i = 0, 1$) for each group. The model assumes a common success probability for both groups. Our results further indicate that we probably do not need variable dispersion parameters for both probability models, that is, a common dispersion parameter would be adequate since neither the ϕ or ω associated with the treatment groups are significant in model III. Thus, a reduced model of case III which models the probability of success separately for the treatments but assumes a common dispersion parameter. The models are based on 16 d.f. Here, under the double binomial model, the estimated probabilities of fetal deaths for the control and experimental groups are respectively 0.0552 and 0.2332 and these estimated probabilities are constant across the treatment levels. The corresponding goodness-of-fit values are $G^2 = 9.1437$ and $X^2 = 22.9671$ with common dispersion parameter estimate being $\hat{\phi} = 0.4900$. We notice a considerable discrepancy between the values of G^2 and X^2 for this data here. This is because, of the twenty observations in the data, nine of them have zeros for the values of Y. Consequently, these observations do not contribute to the overall G^2 and this accounts for the lower values of G^2 compared to their corresponding X^2 .

For the MBM, while the estimated ψ are specific to each treatment and constant across each treatment, the estimated probabilities π_1 of successes vary by the number of litters n as outlined in expression (4). Thus for $n = 8$, π_1 equals 0.0769 and 0.2469 respectively for the control and treatment groups. We present in Table 11 the estimated probabilities and other variables under the multiplicative model for this case.

We may note here that for this data, we have also computed the Wald's test Statistic and it seems to give the lowest value of 23.2406. The GOF values are cumulated so that the last values give the sums over all observations.

Conclusion

Results presented in the preceding sections showed that while it is relatively easier to fit both the double binomial and the multiplicative binomial with joint sufficient statistics employing Poisson regression for frequency data, this approach cannot easily be implemented with data having co-variables. Further, the sufficient statistics approach is based on more degrees of freedom than the MgL method, which makes the MgL method more parsimonious in all cases. We would encourage the use of the MgL methods in applications of these models

	Case I		Case II		Case III	
Parameters	DBM	MBM	DBM	MBM	DBM	MBM
MLE Est.	$\hat{\pi} = 0.1269$	$\hat{\psi} = 0.3033$	$\hat{\pi}_0 = 0.0853$	$\hat{\pi}_0 = 0.3019$	$\hat{\pi}_0 = 0.0703$	$\hat{\psi}_0 = 0.0624$
					$\hat{\psi}_1 = 0.2293$	$\hat{\psi}_1 = 0.3566$
	$\hat{\phi} = 0.3648$	$\hat{\omega} = 0.8314$	$\hat{\phi}_0 = 0.8035$	$\hat{\omega}_0 = 0.7590$	$\hat{\omega}_0 = 0.7180$	$\hat{\omega}_0 = 1.0412$
			$\hat{\phi}_1 = 0.1772$	$\hat{\omega}_1 = 0.8964$	$\hat{\omega}_1 = 0.4445$	$\hat{\omega}_1 = 0.8514$
-2LL	60.3121	63.5982	57.7621	59.4377	55.6644	57.1084
AIC	64.3121	67.5982	63.7621	65.4377	63.6644	65.1084
X ²	28.6119	41.6813	22.8846	27.5236	22.366	28.159
G ²	11.2724	39.9362	9.6845	32.2834	9.4538	30.5746
d.f.	17	17	16	16	15	15

Table 10: Parameter estimates under the three cases for the two probability models.

#	TRT	n	y	$\hat{\psi}$	$\hat{\pi}_1$	S ²	$\sum G^2$	$\sum X^2$	$\sum Wald$
1	0	5	0	0.1565	0.1086	0.5467	0	0.543	0.5392
2	0	6	2	0.1565	0.0974	0.6076	4.9215	3.9724	3.8374
3	0	7	0	0.1565	0.0868	0.6485	4.9215	4.5799	4.4064
4	0	7	0	0.1565	0.0868	0.6485	4.9215	5.1873	4.9754
5	0	8	0	0.1565	0.0769	0.6702	4.9215	5.8023	5.5396
6	0	8	0	0.1565	0.0769	0.6702	4.9215	6.4172	6.1039
7	0	8	0	0.1565	0.0769	0.6702	4.9215	7.0321	6.6681
8	0	9	1	0.1565	0.0677	0.6743	5.9118	7.2823	6.8942
9	0	9	2	0.1565	0.0677	0.6743	10.6648	10.4546	9.7616
10	0	10	1	0.1565	0.0594	0.6637	11.7067	10.7322	10.0101
11	1	5	0	0.3409	0.2949	1.3354	11.7067	12.2067	11.6382
12	1	5	2	0.3409	0.2949	1.3354	12.926	12.394	11.845
13	1	7	1	0.3409	0.2643	1.9864	11.6956	12.7845	12.2088
14	1	8	0	0.3409	0.2469	2.3069	11.6956	14.7597	13.8999
15	1	8	2	0.3409	0.2469	2.3069	11.7456	14.76	13.9002
16	1	8	3	0.3409	0.2469	2.3069	14.2534	15.2918	14.3554
17	1	9	0	0.3409	0.2282	2.6004	14.2534	17.3455	15.9774
18	1	9	4	0.3409	0.2282	2.6004	19.5864	19.1899	17.4341
19	1	10	1	0.3409	0.2084	2.8434	18.1179	19.7537	17.8473
20	1	10	6	0.3409	0.2084	2.8434	30.8076	27.1123	23.2406

Table 11: Parameter estimates under the multiplicative model with constant dispersion parameter.

to binary count data. Of the two binomial models, the multiplicative binomial seems more consistent and fits much better than the double binomial. Further, it does not have much convergence problems than the DBM.

The SAS programs for implementing all the models discussed in this paper are readily available from the author. Meanwhile, we have attached a typical program in the appendix for implementing the MGM for the Manly data discussed in section 5.3.

References

- Lindsey JK, Altham PK. Analysis of the human sex ratio by using over dispersion models. Appl Statist. 1998; 47: 149-157.
- PM Altham. Two generalizations of the binomial distribution. Appl Statist. 1978; 27: 162-167.
- B Efron. Double exponential families and their use in generalized linear regression. J Amer Statist Assoc. 1986; 81: 709-721.
- Lindsey JK. Modelling Frequency and Count Data. Oxford University Press; 1995.
- Lindsey JK, Mersch G. Fitting and comparing probability distributions with log linear models. Comput Stat Data Anal. 1992; 13: 373-384.
- Lovinson G. An alternative representation of Altham's multiplicative-binomial distribution. Stat Proba lett. 1998; 36: 415-420.
- DR Cox. The analysis of multivariate binary data. Appli Stat. 1972; 21: 113-120.
- EAH Elamir. Multiplicative-binomial distribution: Some results and characterization, inference and random data generation. J Stat Theory Appl. 2013; 12: 92-105.
- Feirer V, Friedl H, Hirn U. Modeling over-and under dispersed frequencies of successful ink transmissions onto paper. J Appl Stat. 2013; 40: 626-643.

10. Sokal RR, Rohlf FJ. Biometry: the Principles and Practice of Statistics in Biological Research. San Francisco: Freeman; 1969.
11. Geissler A. Contribution to the question of the sexual relationship of the born. Sachs Super numerary Bur. 1889; 35: 1-24.
12. Borges P, Rodrigues J, Balakrishnan N. A Com-Poisson type generalization of the Binomial distribution and its properties and applications. J Stat Prob Letters. 2014; 87: 158-166.
13. Nelder JA, Mead R. A simplex method for function minimization. Comput J. 1965; 7: 308-312.
14. Morel JG, Neerchal NK. Clustered binary logistic regression in teratology data using a finite mixture distribution. Stat Med. 1997; 16: 2843-2853.
15. Morel JG, Neerchal NK. Overdispersion Models in SAS. SAS publication. 2012.
16. Manly B. Regression models for proportions with extraneous variance, Biometrie-Praximetrie. 1978; 18: 1-18.
17. Lawal HB. Categorical Data Analysis with SAS and SPSS Applications. Taylor & Francis. 2003; New York, USA.