Dhaka Univ. J. Sci. 72(1): 56-62, 2024 (January)

ISSN 1022 - 2502, E-ISSN 2408- 8528
DOI: https://doi.org/10.3329/dujs.v72i1.71247

Proposing a New Estimator of Overdispersion for Multinomial Data

Farzana Afroz

Department of Statistics, University of Dhaka, Dhaka - 1000, Bangladesh

(Received : 27 August 2023; Accepted : 13 December 2023)

Abstract

The classical approach of estimating overdispersion parameter, ϕ, by Pearson's goodness of fit statistic is not appropriate when the data are sparse. We have considered several estimators of ϕ, derived from the Pearson's statistic and the deviance statistic for multinomial data. The proposed estimator of ϕ depending on the deviance statistic is shown to perform the best for increasing level of sparsity and overdispersion, regarding the root mean squared error. As a practical example dead recovery data collected on Herring gulls from Kent Island, Canada are considered. A parametric extra variation model finite mixture distribution is used in the simulation study.

Keywords: Multinomial, Overdispersion, Sparse data, finite mixture distribution, Dead recovery model

I. Introduction

Large, sparse multinomial data occur in numerous fields of studies, where many out the set of response categories will not be observed at all. Overdispersion is common while dealing with sparse data. In statistics quasi likelihood approach, proposed by Wedderburn[1] is a way of dealing with overdispersion in data. To be specific for multinomial data, this approach assumes that the variance of the response variable Y is proportional to that specified by the multinomial model. That is $\mathrm{Var}(Y) = \phi\mathrm{V}$, where ϕ, is the measure of the amount of overdispersion and V is the variance function. This approach would be more robust compared to fitting any parametric extra variation model since it does not require any specification of the distribution for the response variable. For sparse data even when the sample size is very large most of the possible responses will be zero. The expected frequencies will be extremely small. The rule of thumb of a minimum expected frequency E ($\boldsymbol{y}$) ≥5 will not be met, so the goodness of fit statistics will not essentially follow chi square distribution. So, it is not reasonable to use Pearson's goodness of fit statistic, P to estimate the dispersion parameter, ϕ, in this situation. Farrington[2] derived a more general goodness of fit statistic, P_F which asymptotically shows smaller variance compared to P, specially for sparse data. Farrington[2] considered an assumption on the 3rd cumulant of the response variable and derived the estimator of ϕ by dividing P_F with the degrees of freedom. Fletcher[3] considered a different assumption on the third cumulant of the response variable which is less restrictive compared to Farrington[2] and proposed a new estimator which has smaller asymptotic variance compared to all the existing estimators. Deng and Paul[4] derived a modified deviance statistic for discrete data and show through a simulation study that the modified deviance statistic has some power advantage over the modified Pearson statistic of Farrington[2]. We extended this modified deviance statistic for multinomial data and derived a new estimator of overdispersion following the procedure adopted by Fletcher[3].

II. Overdispersion Multinomial Model

For modelling overdispersion in multinomial data two parametric extra variation models have been frequently used in the literature. These models are Dirichlet-multinomial distribution, due to Mosimann[5] and a finite mixture distribution proposed by Morel[6]. In this paper we used finite mixture distribution for simulation purpose. Let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{ik_i})'$ denote the observations from a typical cluster of size n_i, where $\sum_{j=1}^{k_i} y_{ij} = n_i$. The component Y_{ij} denotes the count for the jth category, $j = 1, 2, \dots, k_i$, in cluster i, $i = 1, \dots, m$, and $Y_{ik_i} = n_i - (Y_{i1} + Y_{i2} + \dots + Y_{ik_i-1})$. We use the lower case $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{ik_i})'$ to denote the realized values of the random variables $\boldsymbol{Y}_i$. Let $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ik_i})'$ be a probability vector such that $0 < \pi_{i1} + \pi_{i2} + \cdots + \pi_{ik_i-1} < 1$ and $0 < \pi_{ij} < 1$. Let $\pi_{ik_i} = 1 - (\pi_{i1} + \pi_{i2} + \cdots + \pi_{ik_i-1})$.In finite mixture distribution, the overdispersion is believed to arise due to clumped sampling. Let $\boldsymbol{T}_i, \boldsymbol{T}_{i1}^0, \dots, \boldsymbol{T}_{in_i}^0, i = 1, \dots, m$, be iid random variables from multinomial distribution with k_i categories, parameter $\boldsymbol{\pi}_i$ and cluster size 1. That is $\boldsymbol{T}_i, \boldsymbol{T}_{i1}^0, \dots, \boldsymbol{T}_{in_i}^0 \sim M_{k_i}(\boldsymbol{\pi}_i, 1)$. Also let, $U_{i1}, U_{i2}, \dots, U_{il}$ be iid uniform (0, 1) random variables. Considering a predetermined constant $\rho, (0 < \rho < 1)$, define

$$\boldsymbol{T}_{il} = \begin{cases} \boldsymbol{T}_i \ if \ U_{il} \le \rho \\ T_{il}^0 \ if \ U_{il} > \rho \end{cases}$$

for $l = 1, 2, \dots, n_i$. Now let

*Author for correspondence. e-mail: fafroz87@du.ac.bd

$$\boldsymbol{Y}_i = \sum_{l=1}^{n_i} \boldsymbol{T}_{il}.$$

Here the random vector $\boldsymbol{Y}_i$ has a finite mixture distribution. Using the indicator function notation, we have

$$\boldsymbol{Y}_i = \boldsymbol{T}_i \sum_{l=1}^{n_i} I(U_{il} \le \rho) + \sum_{l=1}^{n_i} \boldsymbol{T}_{il}^0 I(U_{il} > \rho), \qquad (1)$$

where $I(.)$ is the indicator function. Equation (1) leads to the following representation,

$$\boldsymbol{Y}_i = \boldsymbol{T}_i N_i + (\boldsymbol{A}_i | N_i) \qquad (2)$$

where $N_i \sim \text{binomial}(\rho, n_i)$, $\boldsymbol{T}_i \sim M_{k_i}(\boldsymbol{\pi}_i, 1)$, N_i and $\boldsymbol{T}_i$ are independent, and $(\boldsymbol{A}_i|N_i) \sim M_{k_i}(\boldsymbol{\pi}_i, n_i - N_i)$ if $N_i < n_i$. When $N_i = n_i$, $\boldsymbol{A}_i|N_i$ becomes zero. The vector of counts $\boldsymbol{Y}_i$ in (2) has two parts. The first part, given by $\boldsymbol{T}_i N_i$, duplicates the response given by $\boldsymbol{T}_i, N_i$ times. This reflects the fact that in cluster sampling, some of the response within the cluster are similar. The second part, given by $(\boldsymbol{A}_i|N_i)$, is made up of $n_i - N_i$ independent responses. Suppose, $\boldsymbol{y}_i$ is a realization of $\boldsymbol{Y}_i$. Then the probability function of $\boldsymbol{Y}_i$ is given by,

$$P_M(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \sum_{j=1}^{k_i} \pi_{ij} \text{pr}\left(\boldsymbol{B_{ij}} = \boldsymbol{y_i}\right), \qquad (3)$$

where $\boldsymbol{B}_{ij}$ is distributed as $M_{k_i}\left((1-\rho)\boldsymbol{\pi}_i + \rho \boldsymbol{e}_{ij}\right)$, $j = 1, 2, \dots, k_i$, $\boldsymbol{e}_{ij}$ is the jth column of the $k_i \times k_i$ identity matrix, $\boldsymbol{B}_{ik_i}$ is a multinomial random variable with parameter $(1-\rho)\boldsymbol{\pi}_i$ and cluster size n_i. The probability function PM(.) in equation (3) is the probability function of a mixture of k_i multinomial distribution. The mean and covariance matrix of the above mixture are as follows:

$E(\boldsymbol{Y}_i) = n_i \boldsymbol{\pi}_i$ and

$\text{var}(\boldsymbol{Y}_i) = \{1 + \rho^2(n_i - 1)\} n_i \left\{diag(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}_{\boldsymbol{i}}^{\boldsymbol{T}}\right\}.$

Therefore, the amount of overdispersion is

$$\phi = 1 + \rho^2(n_i - 1). \qquad (4)$$

The estimation of this model by maximum likelihood technique can be mathematically expensive, hence the method of quasi -likelihood estimation can be employed.

III. Estimators of Overdispersion Parameter

Suppose, the independent multinomial random vector, $\boldsymbol{Y}_i, i = 1, .., m$, has the covariance matrix, $\boldsymbol{\Sigma_i} = n_i diag(\pi_{ij}) - n_i \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T$ and mean vector, $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ik})' = n_i \boldsymbol{\pi}_i$. Assume a multivariate generalized linear regression model with the link function $\boldsymbol{\mu}_i = h_i(\eta_i) = \left(h_{i1}(\eta_{i1}), \dots, h_{ip}(\eta_{ik_i})\right)^T$,

where $\eta_{ij} = \boldsymbol{X}_{\boldsymbol{ij}}^{\boldsymbol{T}} \boldsymbol{\beta}, \boldsymbol{X}_{ij} = (\boldsymbol{X}_{ij1}, \dots, \boldsymbol{X}_{ijq})^T$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)$ is a vector of q regression parameters. Maximum likelihood estimates of the regression parameters $\beta_1, \dots, \beta_q$ are obtained as solutions of the q' quasi-likelihood estimating equations $g_u(\widehat{\boldsymbol{\beta}}) = 0, u = 1, 2, \dots, q$ where

$$g_u(\boldsymbol{\beta}) = \sum_{i=1}^m (\boldsymbol{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \frac{\delta \boldsymbol{\mu}_i}{\delta \boldsymbol{\beta}_u}.$$

Goodness of fit is assessed by extending the model to a wider family with covariance matrix $\phi \boldsymbol{\Sigma}_i$ and evaluating the departure from the value $\phi = 1$. Deng and Paul[4] used an unbiased supplementary estimating equation $g_{q+1}(\widehat{\boldsymbol{\beta}}, \hat{\phi}) = 0$ for estimating the dispersion parameter ϕ. Where,

$$\begin{aligned} &g_{q+1}(\boldsymbol{\beta}, \phi) \\ &= \sum_{i=1}^m \boldsymbol{a}_i^T (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \\ &+ \sum_{i=1}^m [(\boldsymbol{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) - \phi(k_i - 1)] \\ &= \sum_{i=1}^m \sum_{j=1}^{k_i} a_{ij} \left(y_{ij} - \mu_{ij}\right) \\ &\quad + \sum_{i=1}^m \sum_{j=1}^{k_i} \frac{\left(y_{ij} - \mu_{ij}\right)^2}{\mu_{ij}} - \sum_{i=1}^m (k_i - 1)\phi \qquad (5) \end{aligned}$$

The function a_{ij} in equation (5) define a family of first-order correction terms to the Pearson statistic. Afroz et al.[7] consider the function $a_{ij} = a\mu_{ij}^{-1}$ and for the values $a = 0, -1, -\phi$ derive three estimators of ϕ which are as follows

$$\hat{\phi}_P = \frac{1}{\sum_{i=1}^m (k_i - 1) - q} \left\{ \frac{\sum_{i=1}^m \sum_{j=1}^{k_i} \left(y_{ij} - \hat{\mu}_{ij}\right)^2}{\hat{\mu}_{ij}} \right\},$$

$$\hat{\phi}_{Fa} = \hat{\phi}_P - \frac{\sum_{i=1}^m \sum_{j=1}^{k_i} s_{ij}}{\sum_{i=1}^m (k_i - 1) - q},$$

$$\hat{\phi}_* = \frac{\hat{\phi}_P}{1 + \bar{s}},$$

where $s_{ij} = \frac{(y_{ij} - \hat{\mu}_{ij})}{\hat{\mu}_{ij}}$ and $\bar{s} = \frac{\sum_{i=1}^m \sum_{j=1}^{k_i} s_{ij}}{\sum_{i=1}^m (k_i - 1)}$.

IV. Proposed Estimator Based on Deviance Statistic

The deviance statistic for multinomial data in the previous setup can be written as follows:

$$\begin{aligned} D &= \sum_{i=1}^m \sum_{j=1}^{k_i} d_{ij} \\ &= 2 \sum_{i=1}^m \sum_{j=1}^{k_i} \left\{ y_{ij} \log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) - (y_{ij} - \hat{\mu}_{ij}) \right\} \qquad (6) \end{aligned}$$

As in Paul and Deng[8], the estimator of ϕ can be found from the solution of the supplementary equation $g_d(\widehat{\boldsymbol{\beta}}, \hat{\phi}) = 0$, where, $d = q + 1$ and $g_{q+1}(\boldsymbol{\beta}, \phi)$

$$\begin{aligned} &= \sum_{i=1}^m \boldsymbol{a}_i^T (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \\ &+ \sum_{i=1}^m \left\{ \frac{\boldsymbol{d}_i(\boldsymbol{y}_i, \boldsymbol{\mu}_i)}{\boldsymbol{\kappa}_1^{(i)}(\boldsymbol{\mu}_i)} - \phi(k_i - 1) \right\} \end{aligned}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{k_i} a_{ij}\left(y_{ij}-\mu_{ij}\right)$$

$$+\sum_{i=1}^{m}\sum_{j=1}^{k_i}\frac{d_{ij}(y_i,\mu_i)}{\kappa_1^{(ij)}}-\sum_{i=1}^{m}(k_i-1)\phi \qquad (7)$$

Here, $\kappa_1^{(ij)}=E(d_{ij})$ and a_{ij} is a set of function to be specified. Following Paul and Deng[8], suppose, $a_{ij}=-a\kappa_{11}^{(ij)}/\left(\mu_{ij}\kappa_1^{(ij)}\right)$, where $\kappa_{11}^{(ij)}=E\left\{\left(d_{ij}-\kappa_1^{(ij)}\right)\left(y_{ij}-\mu_{ij}\right)\right\}$. The particular choice of $a=0$and $\kappa_1^{(ij)}=1$ give us usual deviance estimator of ϕ. Now choosing $a=0,-1,-\phi$ and allowing unbiasedness correction the estimators of ϕ based on deviance statistic are as follows:

$$\hat{\phi}_{d_1}=\frac{2}{\sum_{i=1}^{m}(k_i-1)-q}\left\{\frac{\sum_{i=1}^{m}\sum_{j=1}^{k_i} y_{ij}\,log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right)}{\hat{\kappa}_1^{(ij)}}\right\},$$

$$\hat{\phi}_{d_2}=\hat{\phi}_{d_1}-\frac{\sum_{i=1}^{m}\sum_{j=1}^{k_i} s'_{ij}}{\sum_{i=1}^{m}(k_i-1)-q},$$

$\hat{\phi}_{New}=\frac{\hat{\phi}_{d_1}}{1+\bar{s}'}$, where$i=1,2,\dots,m$ and$j=1,2,\dots,k_i$, $s'_{ij}=\frac{\kappa_{11}^{(ij)}(y_{ij}-\hat{\mu}_{ij})}{\hat{\kappa}_1^{(ij)}}$ and $\bar{s}'=\frac{\sum_{i=1}^{m}\sum_{j=1}^{k_i} s'_{ij}}{\sum_{i=1}^{m}(k_i-1)}$.

Here $\hat{\phi}_{New}$ is the proposed estimator. We denoted usual deviance estimator by $\hat{\phi}_{d_0}$ which is defined as,

$$\hat{\phi}_{d_0}=\frac{2}{\sum_{i=1}^{m}(k_i-1)-q}\sum_{i=1}^{m}\sum_{j=1}^{k_i}\left\{y_{ij}\,log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right)\right\}.$$

Closed form expressions for $\kappa_1^{(ij)}$ and $\kappa_{11}^{(ij)}$ are not found. The approximation of d_{ij}utilizing Taylor expansion does not function well. Therefore, in this study we compute $\kappa_1^{(ij)}$ and $\kappa_{11}^{(ij)}$ straight as follows

$$\kappa_1^{(ij)}=E\left(d_{ij}\right)=\sum_{h=0}^{\infty} d_{ij}\,(h)P_{ijh}$$

$$\kappa_{11}^{(ij)}\left(\mu_{ij}\right)=E\left(d_{ij}-\kappa_1^{(ij)}\right)\left(y_{ij}-\mu_{ij}\right)$$

$$=\sum_{h=0}^{\infty}\left(h-\mu_{ij}\right)d_{ij}\,(h)P_{ijh}$$

where for count data, $P_{ijh}=\exp\left(-\lambda_{ij}\right)\lambda_{ij}^{h}/h!$, here λ_{ij} is the expected count in the (i,j)th cell.

V. Practical Example

We illustrate the difference between the estimators of dispersion parameters using Herring gulls (Larus argentatus) data presented by Paynter[9]. From 1934 through 1939 total 31, 694 gulls were banded on Kent Island, Grand Manan, New Brunswick, Canada. A detail description of data can be found in Paynter[9, 10]. These 31, 694 banding yielded 1, 099 recoveries. There are 6 years of ringing and 29 years of follow up. In the literature different methods have been proposed for analysing such kind of recovery data. Recovery data arises in several bird banding and other animal tagging studies. The major objective of these kind of experiments is to estimate parameters related to population survival. A number of animals are captured, banded, and released at the beginning of each time period, for several equal time periods. In bird banding studies, the time period is usually one calendar year. Birds then die through natural mortality, hunting, etc. Records are made of bands returned from dead birds. Suppose for m consecutive years, n_i $(i=1,2,\dots,m)$ is the no of birds banded and released back in to the population at the beginning of ith year, y_{ij} is the number of band recoveries in period j $(j=1,2,\dots,k_i,\ k_i\geq m)$ originally banded in year i. Then we can display the recovery data in symbolic notation as bellow:

Table 1. Recovery data

Year banded	Number banded	Year of recovery					Not recovered
		1	2	3	…	k_i-1	k_i
1	n_1	y_{11}	y_{12}	y_{13}	…	y_{1k_1-1}	n_1-w_1
2	n_2		y_{22}	y_{23}	…	$y_{2k_{2-1}}$	n_2-w_2
.	.						
m	n_m				…	y_{mk_m-1}	n_m-w_m

We have considered Seber[11] parameterization for modelling recovery data. Suppose, s is the probability that the individual survives the year, r is the probability that a dead marked individual is reported during each period between releases. Though, Seber[11] considered r and s as time dependent, for simplicity we considered them to be time independent. For the ith cohort suppose $w_i=\sum_{j=1}^{k_{i-1}} y_{\,ij}$, $i=1,2,\dots,m$ is the aggregate of all recoveries. The probability of non-recovery of a banded bird in year 1 is $\{1-r(1-s^{k_1-1})\}$. The joint probability function of $y_{\,ij}$, considering all the non-recovered birds are at the last category k_i in each year i can be written as follow

$$f(y)=\frac{n_1!}{\prod_{j=1}^{k_1-1}(y_{1j})!(n_1-w_1)!}\{(1-s)r\}^{y_{11}}\{s(1-s)r\}^{y_{12}}\dots\left\{s^{(k_1-1)}(1-s)r\right\}^{y_{1,k_1-1}}\{1-r(1-s^{k_1-1})\}^{n_1-w_1}$$

$$\times\frac{n_2!}{\prod_{j=1}^{k_2-1}(y_{2j})!(n_2-w_2)!}\{(1-s)r\}^{y_{22}}\{s(1-s)r\}^{y_{23}}\dots\left\{s^{(k_2-1)}(1-s)r\right\}^{y_{2,k_2-1}}\{1-r(1-s^{k_2-1})\}^{n_2-w_2}\times\dots$$

$$\times\frac{n_m!}{\prod_{j=1}^{k_m-1}(y_{mj})!(n_m-w_m)!}\{(1-s)r\}^{y_{mm}}\dots\{1-r(1-s^{k_m-1})\}^{n_m-w_m}.$$

We can fit the above product multinomial model to the Herring gulls data. By applying maximum likelihood method, the estimates of the parameters were found to be $\hat{r} = 0.035, \hat{s} = 0.655$, with approximate 95% confidence intervals $(0.0335, 0.037)$ and $(0.638, 0.672)$respectively. We estimate the dispersion parameter of the data by several estimators the results are listed below:

Table 2. Estimates of overdispersion parameter

Pearson based		Deviance based	
$\hat{\phi}_P$	4.409	$\hat{\phi}_{d_0}$	1.619
$\hat{\phi}_{Fa}$	1.906	$\hat{\phi}_{d_1}$	4.057
$\hat{\phi}_*$	1.304	$\hat{\phi}_{d_2}$	2.203
		$\hat{\phi}_{New}$	1.421

Thus $\hat{\phi}_P$ and $\hat{\phi}_{d_1}$ suggest that there is substantial overdispersion, whereas $\hat{\phi}_{Fa}$ and $\hat{\phi}_{d_2}$ suggest moderate overdispersion and finally, $\hat{\phi}_*$, $\hat{\phi}_{d_0}$and $\hat{\phi}_{New}$ suggest that there is relatively little overdispersion.

VI. Simulation

Afroz et al.[7] show that, $\hat{\phi}_*$ is the best estimator compared to the estimators $\hat{\phi}_P$, $\hat{\phi}_{Fa}$ and $\hat{\phi}_{d_0}$ for sparse multinomial data in terms of root mean squared error. Therefore, in this simulation we only compared the proposed estimator $\hat{\phi}_{New}$ with $\hat{\phi}_*$. We now report the results of the simulation studies on the performance of the two estimators of the overdispersion parameter by using finite mixture distribution Morel[6].

Simulation results for varying degrees of overdispersion:

At first, we simulated 10^3 data from finite mixture distribution. The Herring gulls data is used for simulation purpose, where, the number of ringing years is 6 and the number of recovery years is 29. All the birds those are not recovered are considered at the last category. Here r and s were set to 0.035 and 0.655 respectively, also total number of birds banded each year (n_i) are kept fixed to10^3. The overdispersion parameter ϕ is varied from 1 to 5. For each value of ϕ, ρ is calculated using Equation (4). For each simulation 10^3 observations from uniform (0,1) distribution are generated and the number of observations (N_i) less or equal to ρ is calculated. Using the value of r and s the multinomial probability vector $(\boldsymbol{\pi}_i)$ for each year i is computed. Then a random vector $\boldsymbol{T}_i$ from $M_{k_i}(\boldsymbol{\pi}_i, 1)$ distribution is generated, where $k_i = 30$ for the first ringing year$(i = 1)$ and k_i decreases by 1 in each consecutive ringing year. After that $\boldsymbol{T}_i$ is multiplied by N_i and another multinomial vector $(\boldsymbol{A}_i|N_i)$ is generated from $M_{k_i}(\boldsymbol{\pi}_i, n_i - N_i)$ distribution. Hence our final multinomial vector for year i is $\boldsymbol{Y}_i = \boldsymbol{T}_i N_i + (\boldsymbol{A}_i|N_i)$. In the simulated vector, response in a specific category is duplicating N_i times, which essentially produce overdispersion in the dataset. When, $\phi = 1$, ρ becomes zero and the data is simulated from the classical multinomial distribution. Finally, bias, standard error and square root of the mean squared error have been calculated for the estimators, for different levels of ϕ. The resulting figure of simulation is given in figure 1. From figure 1 it can be said that, the estimators are negatively biased.

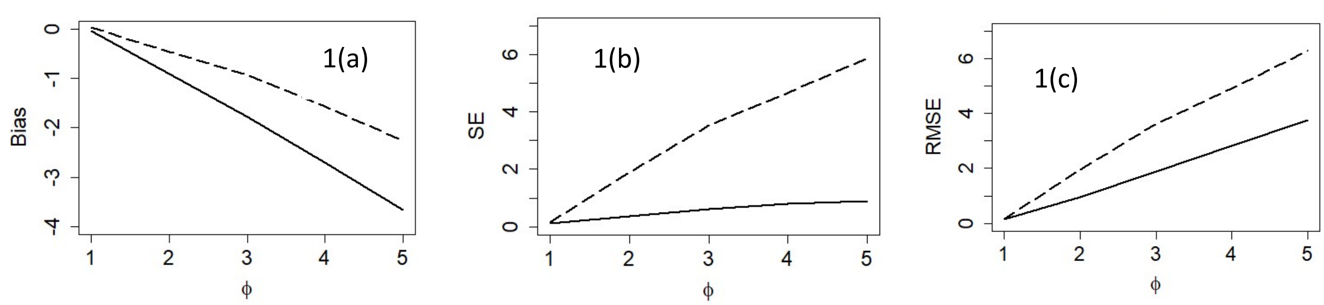

Fig 1. The (a) Bias, (b) SE, and (c) RMSE for $\hat{\phi}_*$(dotted line) and $\hat{\phi}_{New}$ (solid line) for different levels of verdispersion.

Between the two estimators $\hat{\phi}_{New}$ has larger bias and the bias increases with the level of overdispersion ϕ. The proposed estimator $\hat{\phi}_{New}$ has smaller standard error, and consequently it has smaller root mean squared error (rmse) compared to $\hat{\phi}_*$. However, when there is no overdispersion that is $\phi \approx 1$ the estimators considered here shows the same performance, but when overdispersion is present the proposed estimator $\hat{\phi}_{New}$ essentially performs the best.

Simulation results for varying degrees of sparsity:

Next, we varied $r$and s from 0.1 to 0.9, and simulated data using a finite mixture distribution by setting $\phi = 2$ following same procedure mentioned earlier. As a measure of sparsity, we considered the proportion of observations in expected recovery matrix those are less than or equal to 1 and denoted it by E_1. E_1 ranges from 0 to 0.9, for 81 different combinations of r and s. Table 3 displays the different values of r and s and the corresponding values of E_1.

Table 3. The level of sparsity(E_1) produced for different combination of r and s.

Sl	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
s	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
r	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
E_1	0.89	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.81	0.81	0.81	0.81	0.81	0.81	0.78

Sl	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
s	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4
r	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7
E_1	0.78	0.81	0.78	0.78	0.78	0.78	0.74	0.74	0.74	0.74	0.78	0.74	0.74	0.74	0.70	0.70	0.70

Sl	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
s	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6
r	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6
E_1	0.70	0.70	0.74	0.70	0.67	0.67	0.67	0.63	0.63	0.63	0.63	0.67	0.63	0.60	0.60	0.56	0.56

Sl	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
s	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8
r	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5
E_1	0.52	0.52	0.52	0.60	0.52	0.49	0.45	0.41	0.41	0.41	0.38	0.38	0.45	0.34	0.27	0.23	0.20

Sl	69	70	71	72	73	74	75	76	77	78	79	80	81
s	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
r	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
E_1	0.16	0.12	0.12	0.09	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

For each value of E_1 we simulated 10^3 dataset. For each estimator, bias, standard error and root mean squared error (rmse) have been calculated for all the levels of E_1. Figure 2 gives some insight into the behavior of the estimators for different levels of E_1.

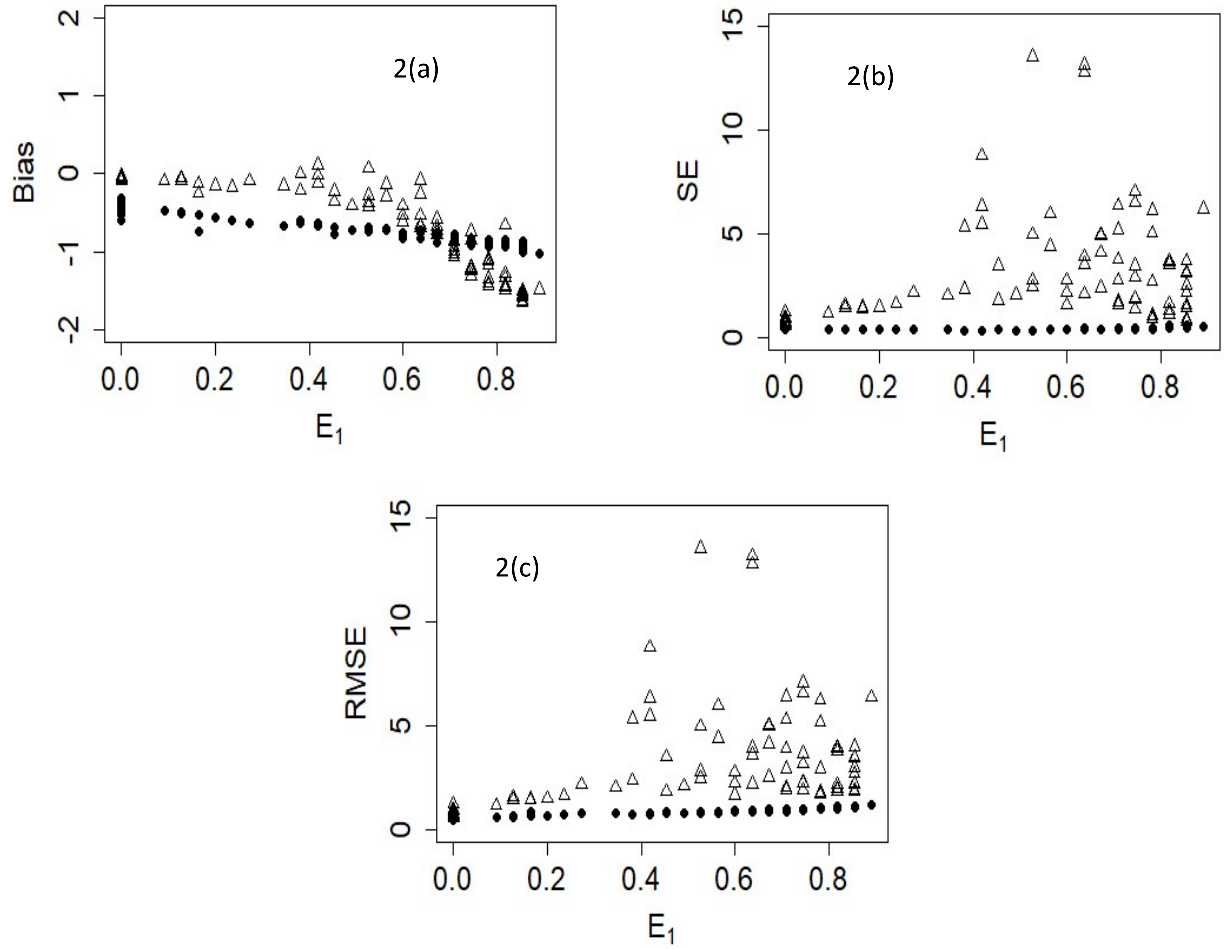

Fig. 2. The (a) Bias, (b)SE, and (c) RMSE for $\hat{\phi}_*$(triangle) and $\hat{\phi}_{New}$ (solid circle) for different levels of sparsity.

From Fig. 2 it is apparent that, estimator $\hat{\phi}_*$ is less robust compared with the estimator $\hat{\phi}_{New}$. Both the estimators are mostly negatively biased and the bias increases with the level of sparsity. Finally, for finite mixture distribution $\hat{\phi}_*$ shows higher standard error and consequently higher root mean squared error (rmse) compared to the proposed estimator $\hat{\phi}_{New}$.

VII. Conclusion

Afroz et al.[7] proposed an estimator of overdispersion, that was shown to have smaller variance comparing to the existing estimators for sparse multinomial data for the increasing level of overdispersion and sparsity. The proposed estimator by Afroz[7] was derived from Pearson's goodness of fit statistic following the procedure of Fletcher[3] and the simulation was done considering the Dirichlet multinomial distribution[5]. In this paper we derived a new estimator of overdispersion for multinomial data from deviance statistic following the procedure by Fletcher[3] and Deng and Paul[4]. Instead of using Dirichlet multinomial distribution[5] we considered finite mixture distribution[6] for simulation purpose and compared the new estimator of overdispersion with the estimator proposed by Afroz et al.[7] for sparse multinomial data. We found that our proposed estimator is more robust compared to that proposed by Afroz et al.[7] for overdispersed multinomial data for the growing level of overdispersion and sparsity. Afroz et al.[12] made a comparison among Dirichlet multinomial distribution[5] and finite mixture distribution[6] through simulation and mentioned that in these two models the overdispersion arises in different mechanism, therefore they cannot be meaningfully compared. Newcomer et al.[13] showed that the higher order moments from the finite-mixture and Dirichlet-multinomial distributions are different, though the first two moments of these distributions are the same. This could be a reason behind $\hat{\phi}_*$ performing worse for finite mixture distribution. Afroz[14] found that the third cumulant of the mixture of Dirichlet multinomial distribution does not satisfy the assumption considered by Fletcher[3] while deriving the estimator of ϕ for count data. Therefore, checking that assumption on the third cumulant of the finite mixture[6] distribution would be helpful for understanding the results found in this research. Though our proposed estimator performs better it has large rmse, as for example it has rmse approximately 4 while the true value of ϕ is 5. Therefore, it would be worthwhile to find out a reasonable confidence interval for the overdispersion parameter ϕ.

References

1. Wedderburn, R. W. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika*, **61(3)**, 439-447.
2. Farrington, C. P. 1996. On Assessing Goodness of Fit of Generalized Linear Models to Sparse Data. *Journal of the Royal Statistical Society*, Series B (Methodological), **58(2),** 349–360.
3. Fletcher, D. J. 2012. Estimating overdispersion when fitting a generalized linear model to sparse data. *Biometrika,* **99(1),** 230–237.
4. Deng, D., & S.R. Paul, 2016. Goodness of fit of product multinomial regression models to sparse data. *Sankhya B*, **78**, 78-95.
5. Mosimann, J. E. 1962. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika*, **49(1/2)**, 65-82.
6. Morel, J. G. 1992. A simple algorithm for generating multinomial random vectors with extravariation. *Communications in Statistics-Simulation and Computation*, **21(4)**, 1255-1268.
7. Afroz, F., M. Parry, & D. Fletcher, 2020. Estimating overdispersion in sparse multinomial data. *Biometrics*, **76(3)**, 834-842.
8. Paul, S. R., & D. Deng, 2000. Goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **62(2)**, 323-333.
9. Paynter, R. A. 1966. *A new attempt to construct life tables for Kent Island Herring Gulls*. The Museum.
10. Paynter, R.A. 1947, 'The fate of banded kent island herring gulls', *Bird-banding* **18(4),** 156-170.
11. Seber, G. A. F. 1970. Estimating time-specific survival and reporting rates for adult birds from band returns. *Biometrika*, **57(2)**, 313-318.
12. Afroz, F., & Z. R. Shabuz, 2020. Comparison Between Two Multinomial Overdispersion Models Through Simulation. *Dhaka University Journal of Science*, **68(1)**, 45-48.
13. Newcomer, J. T., N.K. Neerchal, and J. G. Morel, 2008. Computation of higher order moments from two multinomial overdispersion likelihood models. *Department of Mathematics and Statistics, University of Maryland, Baltimore, USA.*
14. Afroz, F. 2021. Studying the Third Cumulant of the Mixture of Dirichlet-multinomial Distributions. *The Dhaka University Journal of Science*, **69(2)**, 96-100.