Dhaka Univ. J. Sci. 68(1): 45-48, 2020 (January)

Comparison Between Two Multinomial Overdispersion Models Through Simulation

Farzana Afroz[*] and Zillur Rahman Shabuz

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

(Received : 20 May 2019 ; Accepted : 7 January 2020)

Abstract

A key assumption when using the multinomial distribution is that the observations are independent. In many practical situations, the observations could be correlated or clustered and the probabilities within each cluster might vary, which may lead to overdispersion. In this paper we discuss two well-known approaches to model overdispersed multinomial data, the Dirichlet-multinomial model and the finite-mixture model. The difference between these two models has been illustrated via simulation study. The forest pollen data is considered as a practical example of overdisperse multinomial data. The overdispersion parameter, ϕ, has been estimated using two classical estimators.

Keywords: multinomial distribution, overdispersion, Dirichlet-multinomial,finite-mixture, simulation.

I. Introduction

Overdispersion is the presence of excess variability in a data set, relative to the statistical model in use, meaning that the data exhibit more variation than the model predicts. When overdispersion is present in a data set the estimate of regression coefficients are still asymptotically unbiased, although the standard errors may be seriously underestimated, and we may therefore incorrectly assess the significance of individual regression parameters. A variable may wrongly appear to be a significant predictor, and we will tend to select overly complex models. Likewise, confidence intervals will be too narrow, i.e., there will be more uncertainty in the data than we have allowed for. Examples of multinomial data with overdispersion arise in many areas, such as mark-recovery and mark-recapture modelling, household surveys, DNA sequence analysis, hyperspectral image (HSI) classification. Many practical examples of overdispersed multinomial data are discussed in the literature. For example, Mosimann[1] presented a classical environmental example on forest pollen, while Koehler and Wilson[2] considered modelling data on housing satisfaction.

Several approaches can be found in the literature for handling multinomial data that exhibit overdispersion. These include the quasi-likelihood (QL) method introduced by Wedderburn[3]and McCullagh and Nelder[4] , and generalized estimating equations (GEE)by Liang and Zeger[5] and Zeger and Liang[6]. Note that GEE is a generalization of QL method, therefore have the similar robustness properties. For explicit modelling of overdispersed multinomial data the well-known approaches are the Dirichlet-multinomial model considered by Mosimann[1], generalized linear mixed models by Wolfinger and O'connell[7] and finite-mixture model discussed by Morel and Nagaraj[8].The quasi-likelihood estimation technique is appealing, because it only requires specification of the first two moments of the response variable. The quasi-likelihood is a function similar to the likelihood and for GLMs the maximum quasi-likelihood estimate (MQLE) is identical to the maximum likelihood estimate (MLE).

Generally, the actual mechanism in which the overdispersion arise is unknown, therefore it is difficult to choose a single overdispersion model for a given set of data. However it is possible to select a model by checking its performance through simulation. In this paper we focus on comparing two different models, Dirichlet-multinomial model[1] and finite-mixture model[8]using forest pollen data[1]through simulation study as they are mostly used in the literature for modeling overdispersed multinomial data.

II. Overdispersion Multinomial Model

Let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{ik_i-1})'$denote the observations from a typical cluster of size m_i. Here Y_{ij} denotes the count in cluster i and category j ($i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k_{i-1}$) and $Y_{ik_i} = m_i - (Y_{i1} + Y_{i2} + \cdots + Y_{ik_i-1})$. We use the lower case $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{ik_i-1})'$ and $y_{ik_i} = m_i - (y_{i1} + y_{i2} + \cdots + y_{ik_i-1})$ to present the observed values of $\boldsymbol{Y}_i$ and Y_{ik_i} respectively. Suppose $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ik_i-1})'$ be a probability vector with $0 < \pi_{i1} + \pi_{i2} + \cdots + \pi_{ik_i-1} < 1$ and $0 < \pi_{ij} < 1$ such that $\sum_{j=1}^{k_i} \pi_{ij} = 1$.

Dirichlet-multinomial distribution

The probability mass function of a Dirichlet-multinomial distribution is

$$P_{DM}(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \frac{m_i!}{y_{i1}!y_{i2}!\dots y_{ik_i}!}\frac{\Gamma(c)}{\Gamma(m_i+c)}\frac{\prod_{j=1}^{k_i}(y_{ij}+c\pi_{ij})}{\prod_{j=1}^{k_i}\Gamma(c\pi_{ij})}, \quad (1)$$

where $c \equiv c_\rho = \rho^{-2}(1-\rho^2)$, $0 < \rho < 1$ and $\Gamma(.)$ is the gamma function. The Dirichlet-multinomial distribution is derived considering two steps. Suppose $\boldsymbol{P}_i = (P_{i1}, P_{i2}, \dots, P_{ik_i-1})'$ denote a probability vector such that $0 < P_{i1} + P_{i2} + \cdots + P_{ik_i-1} < 1$ and $0 < P_{ij} < 1$ for $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k_{i-1}$ and $P_{ik_i} = 1 - (P_{i1} + P_{i2} + \cdots + P_{ik_i-1})$. Suppose the conditional probability of $\boldsymbol{Y}_i$

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

given $\boldsymbol{P}_i$ is multinomial that is $\boldsymbol{Y}_i|\boldsymbol{P}_i \sim M_{k_i}(\boldsymbol{P}_i, m_i)$. If $\boldsymbol{P}_i$ be a Dirichlet random variable with probability mass function

$$\gamma(\boldsymbol{P}_i;\, \boldsymbol{\pi}_i, \rho) = \frac{\Gamma(c)}{\prod_{j=1}^{k_i}\Gamma(c\pi_{ij})}\prod_{j=1}^{k_i} P_i^{c\pi_{ij}-1}, \quad (2)$$

then it can be shown that equation (1) is the probability mass function of $\boldsymbol{Y}_i$. The mean and variance of $\boldsymbol{Y}_i$ are as follows

$$E(\boldsymbol{Y}_i) = m_i\pi_i \quad (3)$$

and

$$\text{var}(\boldsymbol{Y}_i) = \{1 + \rho^2(m_i - 1)\}m_i\{\text{diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i\boldsymbol{\pi}_i'\} \quad (4)$$

where $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ik_i-1})'$and $\text{diag}(\boldsymbol{\pi}_i)$ is a diagonal matrix with diagonal elements $\pi_{i1}, \pi_{i2}, \ldots, \pi_{ik_i-1}$. The term $\{1 + \rho^2(m_i - 1)\}$ indicates extra variation comparing to the usual covariance of a multinomial distribution. Note that for $\rho = 0$ the Dirichlet-multinomial distribution and the usual multinomial distribution have common covariance matrix.

Finite-mixture distribution

In finite-mixture distribution, it is assumed that the overdispersion arises for clumped sampling. Let us consider k_i dimensional independent and identically distributed (iid) multinomial random variables $\boldsymbol{T}_i, \boldsymbol{T}_{i1}^0, \ldots, \boldsymbol{T}_{im_i}^0$. Suppose each variable has parameters $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ik_i-1})'$ with cluster size 1. Let $U_{i1}, U_{i2}, \ldots, U_{ik_i}$ be iid uniform (0, 1) then we define the random variable $\boldsymbol{Y}_i$ for the real number ρ $(0 < \rho < 1)$ as,

$$\boldsymbol{Y}_i = \boldsymbol{T}_i \sum_{l=1}^{m_i} I(U_{il} \le \rho) + \sum_{l=1}^{m_i} \boldsymbol{T}_{il}^0 I(U_{il} > \rho), \quad (5)$$

where$I(.)$ is the indicator function. Equation (5) leads to the following representation,

$$\boldsymbol{Y}_i = \boldsymbol{T}_i N_i + (\boldsymbol{X}_i|N_i) \quad (6)$$

where $N_i \sim binomial(\rho, m_i)$, $\boldsymbol{T}_i \sim M_{k_i}(\boldsymbol{\pi}_i, 1)$, N_i and $\boldsymbol{T}_i$ are independent, and $(\boldsymbol{X}_i|T_i) \sim M_{k_i}(\boldsymbol{\pi}_i, m_i - N_i)$ if $N_i < m_i$. It can be shown that $\boldsymbol{Y}_i$ is infact a mixture of k_i dimensional multinomial random variables and can be written as

$$\boldsymbol{Y}_i = \sum_{j=1}^{k_i} \pi_{ij}\boldsymbol{X}_{ij}, \quad (7)$$

where $\boldsymbol{X}_{ij}$ is distributed as $M_{k_i}((1-\rho)\boldsymbol{\pi}_i + \boldsymbol{e}_{ij}, m_i)$ for $j = 1, 2, \ldots, k_{i-1}$ where $\boldsymbol{e}_{ij}$ is the jth column of the $(k_i - 1) \times (k_i - 1)$ identity matrix and m_i is the cluster size, $\boldsymbol{X}_{ik_i}$ is multinomial random variable with parameter $(1-\rho)\boldsymbol{\pi}_i$ and having the same cluster size. Note that the Dirichlet-multinomial distribution and finite-mixture distribution have common mean vector and covariance matrix.

However the maximum likelihood estimation of the parameters of these models could be computationally intensive. Therefore the quasi-likelihood estimation technique, that is specifying only the relationship between the mean and variance of the response variable, can be applied. In particular, we assume that the variance is ϕ times the variance specified by the generalized linear model. The term, ϕ provides the measure of the amount of overdispersion.

III. Estimators of Overdispersion Parameters

Suppose $\boldsymbol{Y}_i$ (for $i = 1, 2, \ldots, n$), be independent multinomial random vectors with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ik_i})' = m_i\boldsymbol{\pi}_i$ and covariance matrix $\phi\boldsymbol{\Sigma}_i$. Where $\boldsymbol{\Sigma}_i = m_i\text{diag}(\boldsymbol{\pi}_i) - m_i\boldsymbol{\pi}_i\boldsymbol{\pi}_i'$ is the covariance matrix of $\boldsymbol{Y}_i$ under a multivariate generalized linear regression model. Suppose that $\boldsymbol{\mu}_i = \boldsymbol{h}_i(\boldsymbol{\eta}_i) = (h_{i1}(\eta_{i1}), \ldots, h_{ik_i}(\eta_{ik_i}))'$ where $\boldsymbol{h}_i(.)$ is the inverse link function with $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{ik_i})'$. Now $\boldsymbol{\eta}_i = (\boldsymbol{X}_{i1}'\boldsymbol{\beta}, \ldots, \boldsymbol{X}_{ik_i}'\boldsymbol{\beta})'$ where $\boldsymbol{X}_{ij} = (X_{ij1}, \ldots, X_{ijq})'$is the vector of covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$ is the vector of regression parameters.

Wedderburn[3] suggested estimating ϕ by

$$\hat{\phi}_P = \frac{X^2}{\text{df}} = \sum_{ij}\frac{(y_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} / \{\sum_{i=1}^{n}(k_i - 1) - q\}, \quad (8)$$

where X^2 is Pearson's goodness of fit statistic and df is the residual degrees of freedom. We consider another estimator of ϕ based on the usual deviance statistic, defined as

$$\hat{\phi}_D = \frac{D}{\text{df}} = 2\sum_{ij} y_{ij}\log(y_{ij}/\hat{\mu}_{ij}) \;/\; \{\sum_{i=1}^{n}(k_i - 1) - q\} \quad (9)$$

where D is the residual deviance and df is the residual degrees of freedom. Asymptotically both D and X^2will have a χ^2distribution with $\{\sum_{i=1}^{n}(k_i - 1) - q\}$ degrees of freedom. If the model is adequate, we expect $\hat{\phi}$ to be close to 1; as the mean of a randomvariable with a χ^2_{df} distribution is equal to its degrees of freedom (df).

IV. Practical Example

We considered forest pollen data[1] as a practical example. Four different types of pollen were counted in order to understand the past vegetation character of an area of Mexico. The types of pollen were *Pinus*, *Abies*, *Quercus* and *Alnus.* Mosimann[1]presented $n = 73$ sets of data each with $m = 100$ counts. Table1 provides the estimates ofϕ for pollen data after fitting the usual multinomial model.

Table 1. Estimates of ϕ for pollen data

Pearson based		Deviance based	
$\hat{\phi}_P$	2.599	$\hat{\phi}_D$	2.941

Thus both $\hat{\phi}_P$ and $\hat{\phi}_D$ suggest a fair amount of overdispersion. Note that the **VGAM** package in **R** is used to fit the usual multinomial model.

V. Simulation Study

We generated a random sample of size 1000 from the Dirichlet-multinomial distribution using the estimated probabilities of the usual multinomial model. Considering the estimates of ϕ for pollen data (Table 1) we fix ϕ at 3. We also generated another random sample of same size from the finite-mixture distribution using the same simulation setting to make comparison between the two models.

We summarized the empirical bias (BIAS), standard error (SE) and root mean square error (RMSE) for each estimator of ϕ (Pearson and deviance estimator) using the simulated data. The results for the Dirichlet-multinomial distribution and the finite-mixture distribution are shown in Table 2.

Table 2. Simulation results for the Dirichlet-multinomial and finite-mixture distribution

		Dirichlet-multinomial	Finite-mixture
$\hat{\phi}_P$	BIAS	-0.026	-0.117
	SE	0.385	0.762
	RMSE	0.386	0.771
$\hat{\phi}_D$	BIAS	-0.179	-0.806
	SE	0.275	0.390
	RMSE	0.328	0.896

From the above results we can see that both estimators, $\hat{\phi}_P$ and $\hat{\phi}_D$, have larger bias, standard error and root mean square error when the data generated from the finite-mixture distribution. Hence we can conclude that the Dirichlet-multinomial model performs better than the finite-mixture model to pollen data.

Next we examine the performance of the two estimators by constructing boxplots. The figures are shown below

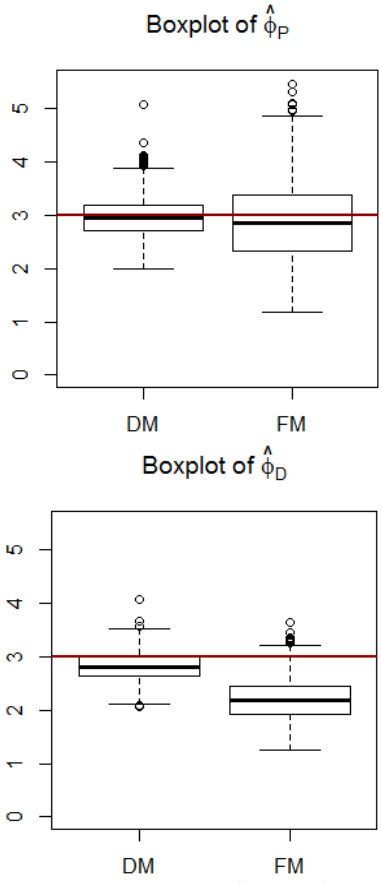

Fig. 1. Boxplots for the estimators $\hat{\phi}_P$ and $\hat{\phi}_D$ for two different distributions: Dirichlet-multinomial (DM) and finite-mixture (FM). The line through the point 3 indicates the true value of ϕ.

The boxplots agree with the results displayed in Table 2. That is the estimators are negatively biased for both the distributions. Finite-mixture distribution has the largest bias for $\hat{\phi}_D$. On the other hand $\hat{\phi}_P$ has the largest variance for the same distribution. Overall both estimator sperform better in case of Dirichlet-multinomial distribution compared to finite-mixture distribution.

Further we examined more closely which distribution better fits the generated data. We compared the quantile-quantile (q-q) plots for the standardized residuals of the simulated data. The figures are as follows.

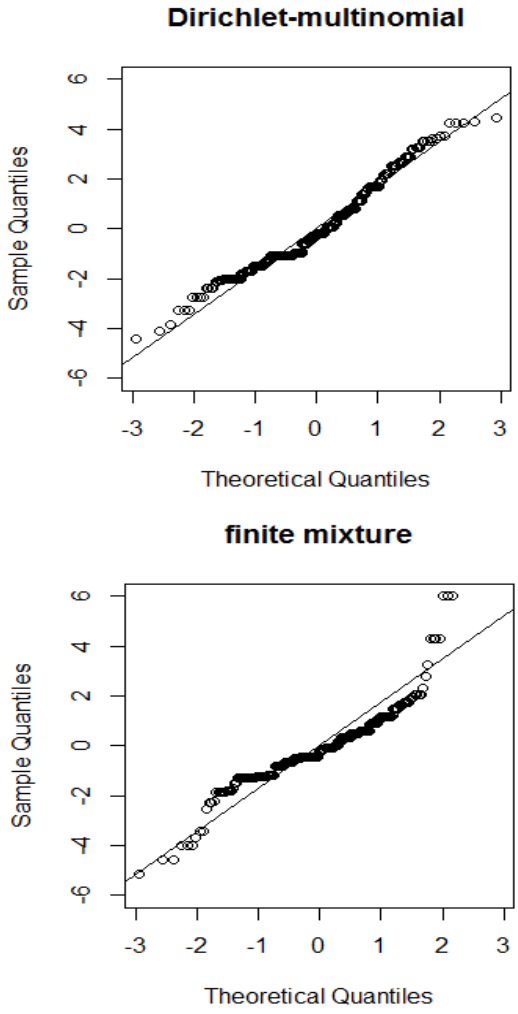

Fig. 2. q-q plot of the standardized residuals

As mentioned earlier, we considered ϕ equal 3 for simulation from both the Dirichlet-multinomial distribution and the finite-mixture distribution. Therefore we expect that the standardized residuals to follow normal distribution with mean0 and standard deviation $\sqrt{3}$. These q-q plots compare the theoretical quantiles to the actual quantiles of the standardized residuals. If the points fall on the straight line with slope $\sqrt{3}$, then the theoretical and realised quantiles are very similar, and the assumption is met. Clearly the standardized residuals of Dirichlet-multinomial

distribution better approximate the normal distribution than the standardized residuals of finite-mixture distribution.

VI. Conclusion

In this paper we have described and used two separate distributions for modeling overdispersion in multinomial data. Though both distributions have same first and second order moments the higher order moments are different[9]. In these models the extra variation arise in separate mechanism, therefore they cannot be meaningfully compared. However for the forest pollen data set, the simulation study shows better approximation for Dirichlet-multinomial model compared to finite-mixture model in our study. It would be interesting to check the performance of one of these two distributions when the overdispersion arises by the other.

References

1. Mosimann, J. E., 1962. On the compound multinomial distribution, the multivariate β- distribution, and correlations among proportions. *Biometrika*, **49(1/2)**,65–82.
2. Koehler, K. J. and J. R. Wilson, 1986. Chi-square tests for comparing vectors of proportions for several cluster samples. *Communications in Statistics - Theory and Methods*, **15(10)**, 2977-2990.
3. Wedderburn, R. W. M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61(3)**, 439–447.
4. McCullagh, P. and J. A. Nelder, 1989. Generalized linear models.Vol. 37, CRC press.
5. Liang, K. Y. and S. L. Zeger, 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, **73(1)**, 13–22.
6. Zeger, S. L. and K. Y. Liang, 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42(1)**, 21–130.
7. Wolfinger, R. and M. O'connell, 1993. Generalized linear mixed models a pseudo-likelihoodapproach. *Journal of Statistical Computation and Simulation*, **48(3-4)**, 233–243.
8. Morel, J. G. and N. K. Nagaraj, 1993. A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, **80(2)**, 363–371.
9. Newcomer, J. T., Neerchal, N. K. 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.*