Dhaka Univ. J. Sci. 69(2): 96-100, 2021 (July)

ISSN 1022 - 2502, E-ISSN 2408- 8528
DOI: https://doi.org/10.3329/dujs.v69i2.56489

Studying the Third Cumulant of the Mixture of Dirichlet-multinomial Distributions

Farzana Afroz

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

(Received : 12 July 2021 ; Accepted : 14 October 2021)

Abstract

Traditionally, the overdispersion parameter φ is estimated by using Pearson's lack of fit statistic X^2or the Deviance statistic D, which do not perform well in the case of sparse data. This paper particularly focuses on an estimator φ_{new} of overdispersion parameter which was proposed for sparse multinomial data. The estimator was derived on the basis of an assumption on the 3rd cumulant of the response variable.When the data comes from the Dirichlet-multinomial distribution φ_{new} is known to have the lowest root mean squared error comparing to the other three estimators. In this paper the 1st to 3rd order raw moments of the finite mixture of Dirichlet-multinomial distributions are derived, which results in complicated mathematical expressions. Furthermore, it is found that the 3rd cumulant of this mixture does not satisfy the assumption which is considered in the derivation of φ_{new}.

Keywords: multinomial distribution, overdispersion, Dirichlet-multinomial,finite-mixture, moments,cumulant.

I. Introduction

The correlation among the observations making up the counts of multinomial data creates overdispersion. Ignoring overdispersion may lead to the serious under-estimation of the standard errors and consequently the model parameters will be incorrectly tested. Overdispersed multinomial data may arise in many areas, such as mark-recovery and mark-recapture modelling, household surveys, DNA sequence analysis, hyperspectral image (HSI) classification. Several approaches can be found in the literature for handling multinomial data that exhibit overdispersion. The most common likelihood approaches for modelling multinomial overdispersed data are Dirichlet-multinomial distribution (Mosimann[1]) and Finite mixture distribution (Morel and Nagaraj[2]). Another approach is 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]. The quasi-likelihood method is simple to apply as only the first two moments of the response variables need to be specified. Also, the maximum likelihood estimate is the same as the maximum quasi likelihood estimate for GLMs. In quasi-likelihood method it is assumed that $\mathrm{Var}(Y) = \varphi V$, where V is the variance function, $\varphi > 1$ indicates overdispersion. The classical estimators of overdispersion are

$$\hat{\varphi}_P = \frac{X^2}{n-p} \text{ or } \hat{\varphi}_D = \frac{D}{n-p}$$

where X^2 is the Pearson's lack of fit statistics and D is the residual deviance, n is the total number of observations and p is the number of parameters estimated. Asymptotically both X^2 and D follow χ^2 distribution with $n-p$ degrees of freedom. However, the asymptotic results of these test statistics depends on the total counts being sufficiently large, which is unlikely to be the case for sparse data containing many small counts.

In 1986 McCullagh[7] argued that for assessing goodness of fit the conditional distribution of the test statistic is more relevant than the marginal distribution. For sparse discrete data McCullagh[7]derived conditional moments of X^2 and D. Farrington[8]used an estimating equations approach to extend the results of McCullagh[7] to models with any type of link function.

Fletcher[9] considered the problem of estimating the overdispersion parameter φ when fitting a generalized linear model to sparse data. He proposed a new estimator of φ that has a smaller variance than Wedderburn[3]'s and Farrington[8]'s estimator, subject to a condition on the third cumulant (κ_{3i}) of the response variable. Under the assumption $\kappa_{3i} = \alpha\kappa_{3i}^0$, where $\alpha > \varphi^2$ and $\kappa_{3i}^0$is the third cumulant of the Poisson distribution, Fletcher[9] showed that that

$$var(\hat{\varphi}_{new}) \le var(\hat{\varphi}_F) < var(\hat{\varphi}_P)$$

where, $\hat{\varphi}_F$ is the Farrington[8]'s estimator. Therefore, the assumption on the 3rd cumulant is needed to justify that the new estimator is more efficient compared to the other estimators. Through simulation study Fletcher[9] showed that the proposed estimator has the lowest level of root mean squared error (RMSE) for the increasing level of overdispersion compared to the Wedderburn[3]'s and Farrington[8]'s estimator when the data comes from Negative binomial distribution and Neyman Type A distribution.

Considering the same assumption on the 3rd cumulant as Fletcher[9], Afroz[10] proposed an estimator $\hat{\varphi}_{new}$ for multinomial data, which can be defined as follows:

suppose there are n independent multinomial random variables $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{ik})'$having mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ik})'$, then

$$\hat{\varphi}_{new} = \frac{\hat{\varphi}_P}{1+\bar{s}},$$

where $\bar{s} = \frac{\sum_{ij}(y_{ij}-\hat{\mu}_{ij})}{n(k-1)}$ and $\hat{\mu}_{ij}$ is the quasi likelihood estimate of μ_{ij}. Also, Farrington[8]'s estimator for multinomial data can be defined as

$$\hat{\varphi}_F = \hat{\varphi}_P - \frac{n(k-1)}{n(k-1)-\mathrm{p}}\bar{s}$$

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

Afroz[10]showed that $\hat{\varphi}_{New}$ has the lowest RMSE comparing to $\hat{\varphi}_P, \hat{\varphi}_D$ and $\hat{\varphi}_F$, when the data are highly sparse and follow Dirichlet-Multinomial distribution. However, it is shown that $\hat{\varphi}_{New}$ did not performed the best when the data are generated from the finite mixture of Dirichlet-multinomial distributions. Afroz[10]derived the 3rd cumulant of the Dirichlet-multinomial distribution and showed that it satisfies the assumption discussed in Fletcher[9]. The motivation of this paper comes from a question raised in Afroz[10], whether the 3rd cumulant of the mixture of Dirichlet-multinomial distributions satisfy the assumption or not. In this paper, the 1st to 3rd order raw moments of the mixture of Dirichlet-multinomial distribution sare derived, also using the raw moments the 3rdcumulant is derived.

II. Dirichlet-multinomial Distribution

Let $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{i(k-1)})'$denote the observations from a typical cluster of size m. Here Y_{ij} denotes the count in cluster i and category j ($i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k-1$) and $Y_{ik} = m - (Y_{i1} + Y_{i2} + \cdots + Y_{i(k-1)})$. The Dirichlet-multinomial distribution proposed by Mosimann[1] provides a way to model categorical data exhibiting overdispersion. Here the overdispersion is believed to arise from the fact that the cell probabilities $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \dots, \pi_{i(k-1)})'$ vary randomly according to a Dirichlet distribution, with probability density function

$$f(\boldsymbol{\pi}_i) = \frac{\Gamma(c)}{\prod_{j=1}^{k} \Gamma(cp_{ij})} \prod_{j=1}^{k} \pi_{ij}^{cp_{ij-1}}$$

where $c = \frac{1-\tau}{\tau}, 0 < \tau < 1$, and $\Gamma(.)$ denotes the gamma function. Thus it is assumed that $\boldsymbol{Y}_i|\boldsymbol{\pi}_i$is distributed as a k-dimensional multinomial random variable and the marginal distribution of $\boldsymbol{Y}_i$ is then the Dirichlet-multinomial, with probability mass function

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \frac{m!}{y_{i1}! y_{i2}! \dots y_{ik}!} \frac{\Gamma(c)}{\Gamma(m+c)} \frac{\prod_{j=1}^{k} (y_{ij} + cp_{ij})}{\prod_{j=1}^{k} \Gamma(cp_{ij})}.$$

The mean and variance of $\boldsymbol{Y}_i$ are as follows

$$E(\boldsymbol{Y}_i) = m\pi_i$$

and

$$\text{var}(\boldsymbol{Y}_i) = \{1 + \tau(m-1)\}\, m\{\text{diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}_i'\}$$

where $\text{diag}(\boldsymbol{\pi}_i)$ is a diagonal matrix with diagonal elements $\pi_{i1}, \pi_{i2}, \dots, \pi_{i(k-1)}$. The term $\{1 + \tau(m-1)\}$ indicates extra variation comparing to the usual covariance of a multinomial distribution. Note that for $\tau = 0$ the Dirichlet-multinomial distribution and the usual multinomial distribution have common covariance matrix. Newcomer et al.[11] derived the higher order moments of the two commonly used multinomial overdispersion models. Following the results of Newcomer et al.[11]the first through third order moments of Dirichlet-multinomial distribution are

I. $E(Y_{ij}) = mp_{ij}$,

II. $E(Y_{ij}{}^2) = m(m-1)\frac{(cp_{ij}+1)}{(c+1)} + mp_{ij}$,

III. $E(Y_i{}^3) = m(m-1)(m-2)\frac{(cp_{ij}+1)(cp_{ij}+2)}{(c+1)(c+2)} P_i + 3m(m-1)\frac{(cp_{ij}+1)}{(c+1)} + mp_{ij}$.

III. Example

In order to examine the numerical differences between the estimators, mark-recovery data on herring gulls (Larus argentatus) from Kent Island in Canada are used here. From 1934 to 1939, 31,694 fledging gulls were banded, of which 1099 were recoverd after death. The detailed description can be found in Paynter[12]. Suppose, s is the probability that a bird survives a year and r is the probability that a banded dead bird is reported. Simple form of Seber[13] model is used to model cell probabilities of the data. After fitting product multinomial model r and s are estimated as 0.035 and 0.655 respectively. In order to, construct a suitable confidence interval of the parameters, r and s a proper estimator of φ is needed. Now a simulation study is performed to select the best estimator of φ by generating data from Dirichlet-multinomial distribution. The recovery and survival probabilities were set to the estimated value sand φ is varied from 2 to 5 in the simulation study. Now the root mean squared error for all the 4 estimators are calculated for each level of φ. From the results displayed in Table-1 it is apparent that, $\hat{\varphi}_{new}$ outperforms the other estimators for different levels of φ. It is noticeable that RMSE increases for the increasing values of φ, however $\hat{\varphi}_{new}$ has the lowest RMSE among all the estimators in each case.

Table 1. Simulation results for the Dirichlet-multinomial distribution

φ	$\hat{\varphi}_P$	$\hat{\varphi}_F$	$\hat{\varphi}_D$	$\hat{\varphi}_{new}$
2	3.112	2.167	1.203	0.634
3	9.879	7.418	1.967	1.144
4	5.085	3.527	2.759	1.584
5	11.576	9.434	3.581	2.175

Therefore, from the simulation results it is clear that $\hat{\varphi}_{new}$ should be used to construct the confidence interval of the parameters.

IV. Mixture of Dirichlet-multinomial Distributions

Suppose that the multinomial data $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{ik})$ is collected from a population which is a mixture of two groups having different parameters $\boldsymbol{\pi}_1 = (\pi_{11}, \pi_{12}, \dots, \pi_{1k})$ and $\boldsymbol{\pi}_2 = (\pi_{21}, \pi_{22}, \dots, \pi_{2k})$. Each observation comes from group-1 and group-2 with probability w_1 and w_2 respectively, where $w_1 + w_2 = 1$. Let us consider the following setup with subscript i removed for simplicity

$$\boldsymbol{Y} = \boldsymbol{Y}_1 + \boldsymbol{Y}_2$$

$$\boldsymbol{Y}_1 = (Y_{11}, \dots, Y_{1k}), \qquad \boldsymbol{Y}_2 = (Y_{21}, \dots, Y_{2k})$$

where

$$\boldsymbol{Y}_1|m_1, \boldsymbol{\pi}_1 \sim \text{Multinom}(m_1, \boldsymbol{\pi}_1),$$

$$\boldsymbol{Y}_2|m_2,\boldsymbol{\pi}_2 \sim \text{Multinom}(m_2,\boldsymbol{\pi}_2),$$
$$m_1+m_2=m,$$

and

$\boldsymbol{\pi}_1$~Dirichlet$(c_1p_{11},\ldots,c_1p_{1k})$,

$\boldsymbol{\pi}_2$~Dirichlet$(c_2p_{21},\ldots,c_2p_{2k})$.

We then have

$$E(\pi_{1j})=\frac{c_1p_{1j}}{\sum_j c_1p_{1j}},\qquad E(\pi_{2j})=\frac{c_2p_{2j}}{\sum_j c_2p_{2j}}$$
$$Var(\pi_{1j})=\tau_1p_{1j}(1-p_{1j}),$$
$$Var(\pi_{2j})=\tau_2p_{2j}(1-p_{2j}).$$

Now following the results of Afroz[1] the expected value and variance of Y_j can be obtained as follows

$$\begin{aligned}
\mu_1' &= E_{m,\pi_j,Y_j}(Y_{1j}+Y_{2j})\\
&= E_m\left(E_{\pi_j,Y_j|m}(Y_{1j}+Y_{2j})\right)\\
&= E_m\left(E_{\pi_j}(E_{Y_j|\pi_j,m}(Y_{1j}+Y_{2j}))\right)\\
&= E_m\left(E_{\pi_j}(m_1\pi_{1j}+m_2\pi_{2j})\right)\\
&= E_m(m_1p_{1j}+m_2p_{2j})\\
&= m(w_1p_{1j}+w_2p_{2j}) \qquad (1)
\end{aligned}$$

Similarly

$$\begin{aligned}
\mu_2 &= Var_{m,\pi_j,Y_j}(Y_{1j}+Y_{2j})\\
&= E_m\left(Var_{\pi_j,Y_j|m}(Y_{1j}+Y_{2j})\right)\\
&+Var_m\left(E_{\pi_j,Y_j|m}(Y_{1j}+Y_{2j})\right)\\
&= E_m\big(m_1(1+((m_1-1)\tau_1)p_{1j}(1-p_{1j})\\
&+m_2(1+((m_2-1)\tau_2)p_{2j}(1-p_{2j})\big)\\
&+Var_m(m_1p_{1j}+m_2p_{2j})\\
&= (1+(m-1)\tau_1w_1)mw_1p_{1j}(1-p_{1j})\\
&+(1+(m-1)\tau_2w_2)mw_2p_{2j}(1-p_{2j})\\
&+p_{1j}^2mw_1(1-w_1)+p_{2j}^2mw_2(1-w_2)\\
&-2mw_1w_2p_{1j}p_{2j} \qquad (2)
\end{aligned}$$

Now the second order raw moment μ_2' of Y_j can be calculated by using Equations (1) and (2). For the simplified form of μ_2', the software Mathematica[14] Version 12.1 is used. The final form of μ_2' is as follows

$$\begin{aligned}
\mu_2' = \mu_2+\mu_1'^2 = {}&-mp_{1j}w_1\big(p_{1j}w_1(1+(m-1)\tau_1)+\\
&\tau_1w_1(1-m)-1\big)-2mp_{1j}p_{2j}w_1w_2+\big(mp_{1j}w_1+\\
&mp_{2j}w_2\big)^2-m^2p_{2j}w_1w_2\big(p_{2j}w_2((m-1)\tau_2+1)+\\
&(1-m)\tau_2w_2-1\big)\left(1+(m-1)p_{1j}w_1\tau_1-p_{1j}^2(w_1-1-\right.\\
&\left.\tau_1w_1+m\tau_1w_1)\right)-2mp_{1j}p_{2j}w_1w_2+\big(mp_{1j}w_1+\\
&mp_{2j}w_2\big)^2+mw_2\left(1+(m-1)\tau_2p_{2j}w_2-p_{2j}^2(w_2(1-\right.\\
&\left.(1-m)\tau_2)-1)\right) \qquad (3)
\end{aligned}$$

In order to derive 3rd order raw moment μ_3' of Y_j, results from Newcomer et al[11] are used. The derivation of μ_3' is as follows

$$\begin{aligned}
\mu_3' &= E_{m,\pi_j,Y_j}(Y_{1j}+Y_{2j})^3\\
&= E_{m,\pi_j,Y_j}(Y_{1j}^3+3Y_{1j}^2Y_{2j}+3Y_{1j}3Y_{2j}^2+3Y_{2j}^3)\\
&= E_m\left(E_{\pi_j}\left(E_{Y_j|\pi_j,m}(Y_{1j}^3+3Y_{1j}^2Y_{2j}+3Y_{1j}3Y_{2j}^2+3Y_{2j}^3)\right)\right)\\
&= E_m\Big(E_{\pi_j}\Big(m_1\pi_{1j}(1-3\pi_{1j}+3m_1\pi_{1j}+2\pi_{1j}^2-3m_1\pi_{1j}^2\\
&+m_1^2\pi_{1j}^2)+3\left(m_1\pi_{1j}(1-\pi_{1j}+m_1\pi_{1j})\right)m_2\pi_{2j}\\
&+3\left(m_2\pi_{2j}(1-\pi_{2j}+m_2\pi_{2j})\right)m_1\pi_{1j}\\
&+m_2\pi_{2j}(1-3\pi_{2j}+3m_2\pi_{2j}+2\pi_{2j}^2-3m_2\pi_{2j}^2\\
&+m_2^2\pi_{2j}^2)\Big)\Big)\\
&= E_m\Bigg(m_1p_{1j}+3m_1(m_1-1)\frac{(C_1p_{1j}+1)p_{1j}}{(C_1+1)}\\
&+m_1(m_1-1)(m_1-2)\frac{(C_1p_{1j}+1)(C_1p_{1j}+2)p_{1j}}{(C_1+1)(C_1+2)}\\
&+3\left(m_1p_{1j}+m_1(m_1-1)\frac{(C_1p_{1j}+1)p_{1j}}{(C_1+1)}\right)m_2p_{2j}\\
&+3\left(m_2p_{2j}+m_2(m_2-1)\frac{(C_2p_{2j}+1)p_{2j}}{(C_2+1)}\right)m_1p_{1j}\\
&+m_2p_{2j}+3m_2(m_2-1)\frac{(C_2p_{2j}+1)p_{2j}}{(C_2+1)}\\
&+m_2(m_2-1)(m_2-2)\frac{(C_2p_{2j}+1)(C_2p_{2j}+2)p_{2j}}{(C_2+1)(C_2+2)}\Bigg)\\
&= mw_1p_{1j}+3m(m-1)w_1^2\frac{(C_1p_{1j}+1)p_{1j}}{(C_1+1)}\\
&+m(m-1)(m-2)w_1^3\frac{(C_1p_{1j}+1)(C_1p_{1j}+2)p_{1j}}{(C_1+1)(C_1+2)}\\
&+3\left(mw_1p_{1j}+m_1(m_1-1)w_1^2\frac{(C_1p_{1j}+1)p_{1j}}{(C_1+1)}\right)mw_2p_{2j}\\
&+3\left(mw_2p_{2j}+m(m-1)w_2^2\frac{(C_2p_{2j}+1)p_{2j}}{(C_2+1)}\right)mw_1p_{1j}\\
&+mw_2p_{2j}+3m(m-1)w_2^2\frac{(C_2p_{2j}+1)p_{2j}}{(C_2+1)}\\
&+m(m-1)(m-2)w_2^3\frac{(C_2p_{2j}+1)(C_2p_{2j}+2)p_{2j}}{(C_2+1)(C_2+2)} \qquad (4)
\end{aligned}$$

Note that in Equation (4), μ_3' is expressed in terms of C_1 and C_2. In order to express μ_3' in terms of τ_1 and τ_1, the software Mathematica[14] is used. The final form of μ_3' is as follows

$$\begin{aligned}\mu_3' &= m\big((p_{1j}w_1 - 3(m-1)p_{1j}\big(p_{1j}(\tau_1-1)-\tau_1\big)w_1^2 \\ &+ \frac{(m-2)(m-1)p_{1j}\big(p_{1j}(\tau_1-1)-2\tau_1\big)\big(p_{1j}(\tau_1-1)-\tau_1\big)w_1^3\big)}{1+\tau_1} \\ &+ p_{2j}w_2 \\ &- 3mp_{1j}p_{2j}w_1\Big((m-1)p_{1j}(\tau_1-1)w_1+\tau_1(w_1-mw_1)-1\Big)w_2 \\ &- 3(m-1)p_{2j}\big(p_{2j}(\tau_2-1)-\tau_2\big)w_2^2 \\ &+ \frac{(m-2)(m-1)p_{2j}\big(p_{2j}(\tau_2-1)-2\tau_2\big)\big(p_{2j}(\tau_2-1)-\tau_2\big)w_2^3}{1+\tau_2} \\ &-3mp_{1j}p_{2j}w_1w_2\Big((m-1)(\tau_2-1)p_{2j}w_2+\tau_2(w_2- \\ &mw_2)-1\Big)\Big). \end{aligned} \tag{5}$$

Now using Equations (1), (3) and (5) the 3rd cumulant of the mixture of Dirichlet-multinomial distribution can be calculated as

$$\begin{aligned}\kappa_3 &= \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 \\ &= 2\big(m\,p_{1j}w_1 + m\,p_{2j}w_2\big)^3 \\ &\quad + m\Big(\big(p_{1j}w_1 \\ &\quad - 3(m-1)p_{1j}\big(p_{1j}(\tau_1-1)-\tau_1\big)w_1^2\big) \\ &+ \frac{(m+2)(m-1)p_{1j}\big(p_{1j}(\tau_1-1)-2\tau_1\big)\big(p_{1j}(\tau_1-1)-\tau_1\big)w_1^3}{1+\tau_1} \\ &\quad + p_{2j}w_2 \\ &-3m\,p_{1j}p_{2j}w_1w_2\Big((m-1)(\tau_1-1)p_{1j}w_1+\tau_1(w_1-mw_1) \\ &\quad -1\Big) \\ &\quad - 3(m-1)p_{2j}\big(p_{2j}(\tau_2-1)-\tau_2\big)w_2^2 \\ &+ \frac{(m-2)(m-1)p_{2j}\big(p_{2j}(\tau_2-1)-2\tau_2\big)\big(p_{2j}(\tau_2-1)-\tau_2\big)w_2^3}{1+\tau_2} \\ &-3mp_{1j}p_{2j}w_1w_2\Big((m-1)(\tau_2-1)p_{2j}w_2+\tau_2(w_2-mw_2) \\ &\quad -1\Big)\Big) - 3m\big(p_{1j}w_1+p_{2j}w_2\big) \\ &\Big(-mp_{1j}w_1\Big(-1+p_{1j}(1+(m-1)\tau_1)w_1+\tau_1(w_1-mw_1)\Big) \\ &\quad - 2mp_{1j}p_{2j}w_1w_2 \\ &\quad + \big(m\,p_{1j}w_1+m\,p_{2j}w_2\big)^2 \\ &\quad - mp_{2j}w_2\Big(-1+p_{2j}(1+(m-1)\tau_2)w_2 \\ &\quad + \tau_2(w_2 \\ &\quad - mw_2)\Big)\end{aligned} \tag{6}$$

Now, the 3rd cumulant of multinomial distribution is

$$\kappa_3{}^0 = m\big(p_{1j}w_1+p_{2j}w_2\big)\big(1-p_{1j}w_1-p_{2j}w_2\big)\big(1- 2p_{1j}w_1-2p_{2j}w_2\big). \tag{7}$$

From Afroz[10] the third cumulant should satisfy the assumption $\kappa_3 = \alpha\kappa_3{}^0$, where $\alpha \geq \varphi^2$, i.e. α must be positive. Here using Equations (6) and (7) one can calculate the value of $\alpha = \frac{\kappa_3}{\kappa_3{}^0}$.Since, the form of α would be very complex, α is plotted for different values of the parameters $(m, p_1, p_2, w_1, w_2, \tau_1, \tau_2)$ using Mathematica[14].

Plot of α for $m = 100$

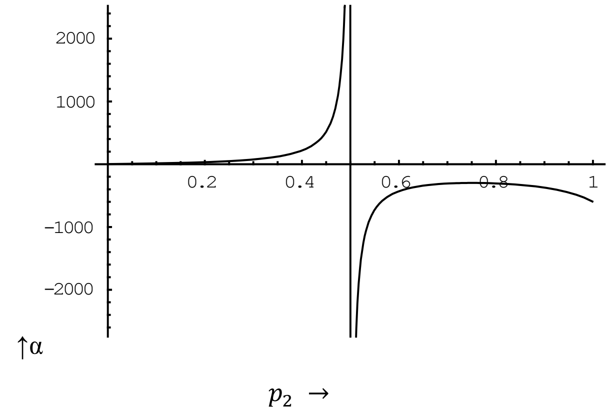

Fig. 1. $m = 100, p_1 = 0.5, w_1 = 0.2, w_2 = 0.8, \tau_1 = \frac{1}{99}, \tau_2 = \frac{1}{99}, 0.001 < p_2 < 0.999$

Plot of α for $m = 10$

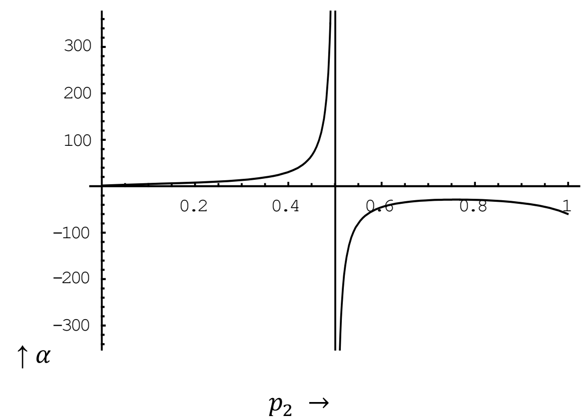

Fig. 2. $m = 10, p_1 = 0.5, w_1 = 0.2, w_2 = 0.8, \tau_1 = \frac{1}{9}, \tau_2 = \frac{1}{9}, 0.001 < p_2 < 0.999$

In figure 1 and 2, α is plotted in the y-axis while p_2 varies from 0.001 to 0.999 in the x-axis and the other parameters kept fixed to certain values. In both the figures α becomes negative when $0.5 < p_2 < 0.999$. Therefore, it is apparent that the 3rd order cumulant of the mixture of Dirichlet-multinomial distribution does not satisfy the assumption of $\alpha \geq \varphi^2$.

V. Discussion and Conclusion

In this paper, the 1st to 3rd order raw moments and the 3rd cumulant (κ_3) of the mixture of Dirichlet-multinomial distributionsare derived. Higher order moments are useful to develop goodness of fit statistics, also use of additional information about the moments in estimation procedure can provide more efficient estimators. Here it is found that, κ_3 does not satisfy the assumption discussed in Fletcher[9]and Afroz[10]. Unlike the present set up one can consider other type of mixtures of Dirichlet-multinomial distributions, where one subpopulation is selected randomly from the two sub-populations and the whole sample comes from the selected sub-population. Also, there are other kind of distributions to model overdispersed multinomial data such as finite-mixture distribution (Morel and Nagaraj[2]). The future research may involve derivation of the moments of such distributions and also checking the assumption on the 3rd cumulant. Furthermore, a new estimator of overdispersion which is more relaxed to the assumption on the third cumulant can be developed.

References

1. Mosimann, J. E. 1962. On the compound multinomial distribution, the multivariate distribution and correlations among proportions. *Biometrika,* **49(1/2),** 65–82.

2. Morel, J. G. & Nagaraj, N. K. 1993. A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, **80(2),** 363–371.

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. & Nelder, J.A. (1989), Generalized linear models. *CRC press* **(37)**.

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. McCullagh,P. 1986. The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association*, **81(393),** 104–107.

8. 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.

9. Fletcher, D. J. 2012. Estimating overdispersion when fitting a generalized linear model to sparse data.*Biometrika* **99(1),** 230–237.

10. Afroz, F. (2018). Estimating Overdispersion in Sparse Multinomial Data. *PhD thesis, University of Otago*, https://ourarchive.otago.ac.nz/handle/10523/8595.

11. 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.*

12. Paynter, R.A. (1947), 'The fate of banded kent island herring gulls', *Bird-banding* **18(4),** 156-170.

13. Seber, G. (1970), 'Estimating time-specific survival and reporting rates for adult birds from band returns', *Biometrika* **57(2),** 313-318.

14. Wolfram Research, Inc., Mathematica, Version 12.1, Champaign, IL (2020).