Sankhyā : The Indian Journal of Statistics
2016, Volume 78-B, Part 1, pp. 78-95

Goodness of Fit of Product Multinomial Regression Models to Sparse Data

Dianliang Deng
University of Regina, Regina, Canada
Sudhir R. Paul
University of Windsor, Windsor, Canada

Abstract

Tests of goodness of fit of sparse multinomial models with non-canonical links is proposed by using approximations to the first three moments of the conditional distribution of a modified Pearson Chi-square statistic. The modified Pearson statistic is obtained using a supplementary estimating equation approach. Approximations to the first three conditional moments of the modified Pearson statistic are derived. A simulation study is conducted to compare, in terms of empirical size and power, the usual Pearson Chi-square statistic, the standardized modified Pearson Chi-square statistic using the first two conditional moments, a method using Edgeworth approximation of the p-values based on the first three conditional moments and a score test statistic. There does not seems to be any qualitative difference in size of the four methods. However, the standardized modified Pearson Chi-square statistic and the Edgeworth approximation method of obtaining p-values using the first three conditional moments show power advantages compared to the usual Pearson Chi-square statistic, and the score test statistic. In some situations, for example, for small nominal level, the standardized modified Pearson Chi-square statistic shows some power advantage over the method using Edgeworth approximation of the p-values using the first three conditional moments. Also, the former is easier to use and so is preferable. Two data sets are analyzed and a discussion is given.

AMS (2000) *subject classification.* Primary 62H15, 62J12; Secondary 62F05.
Keywords and phrases. Pearson statistic, sparse data, multinomial model, family of power-divergence statistics, size, power.

1 Introduction

Count data in the form of contingency tables arise in practice in which some of the cell counts are very small. Such a phenomenon of sparseness in contingency tables occurs, for example, for data in the form of proportions in which the binomial indices are small. In product multinomials, such sparseness occur as a result of the multinomial indices being small. In cross sectional studies, it occurs if the overall sample size is small.

The problem of goodness of fit in contingency tables has been well discussed in literature. Goodness of fit in contingency tables are usually tested using either the Pearson χ^2 statistic or the likelihood ratio statistic. However, these test statistics do not work well for sparse data. Many authors consider modifications to the Pearson and likelihood ratio statistics by using higher order moment approximations of the conditional distributions. See, for example, Koehler and Larntz (1980) and Koehler (1986). These authors derive approximations to the first three moments of the conditional distributions of the Pearson χ^2 statistic and the likelihood ratio statistic for data in the form of proportions and count data.

Furthermore, Lewis et al. (1984) derive the explicit expressions of first three moment for the Pearson χ^2 statistic in the two-way contingency table. Stafford (1995) presents the computer algebra procedure for the calculation of exact cumulants for the Pearson χ^2 and Zelterman statistics in the r-way contingency tables. McCullagh (1985) obtains approximations to the first three moments of the unconditional and conditional distributions of the Pearson chi-square statistic for canonical exponential family regression models. McCullagh (1986) argues that it is the conditional distribution of the statistic and not its marginal distribution that is relevant for assessing goodness of fit. He obtains conditional distributions of the Pearson chi-square statistic and the likelihood ratio chi-square statistic for discrete data for the case where the data are extensive but sparse. Farrington (1996) extends the results of McCullagh (1985) to models with non-canonical links using an estimating-equations approach. Paul and Deng (2000) derive approximations to the first three moments of the unconditional and conditional distribution of the modified deviance statistic, again, using an estimating-equations approach. Further, Paul and Deng (2012) extend Farrington's results and obtain approximations to the fourth moments of the unconditional and conditional distribution of the modified Pearson statistic with non-canonical links. They develop a procedure using all four conditional moments to assess goodness of fit properties of the conditional distribution of the Pearson statistic, for non-canonical generalized linear models with data that are extensive but sparse, by Edgeworth approximation of p-values.

For product multinomials also, goodness of fit in contingency tables can be tested by using the Pearson Chi-squared statistic, or, more generally, the family of power-divergence statistic SD_λ proposed by Cressie and Read (1984). But, as in the case of proportional and count data, the approximation of the Pearson statistic to the Chi-square distributions is not good for sparse data. Some authors have suggested that the goodness of fit for sparse product multinomial models could be assessed by using asymptotic normality of the Pearson Chi-square statistic or a family of power-divergence statistics SD_λ (see Dale (1986), Osius and Rojek (1992) and Fagerland et al. (2008)). These authors show that Pearson Chi-square statistic and SD_λ has an asymptotic normal distribution. Furthermore, Kim et al. (2009) considered an estimated based test for large sparse multinomial distributions based on the test statistic proposed by Zelterman (1987). But they did not discuss the asymptotics for the proposed statistic and thus did not give empirical results.

In this paper, we derive approximations to the first three moments of the unconditional and the conditional distributions of the modified Pearson statistic for assessing goodness of fit of multinomial models with non-canonical links. As in McCullagh (1986) and Farrington (1996), the modified Pearson statistic is obtained through a supplementary estimating equation for the dispersion parameter. A simulation study is then conducted to compare, in terms of empirical size and power, the usual Pearson Chi-square statistic, the standardized modified Pearson Chi-square statistic using the first two conditional moments, a method using Edgeworth approximation of the p-values based on the first three conditional moments, and a score test statistic.

The modified Pearson statistic and its unconditional and conditional moments are obtained in Section 2. Methods for assessing goodness of fit are discussed in Section 3. A simulation study comparing the four methods is conducted in Section 4. Two examples are analyzed in Section 5 and a discussion is given in Section 6.

2 Estimating Equations and Goodness of Fit

2.1. The modified pearson statistic. Let $\boldsymbol{y}_i, i = 1, ..., n$, denote independent multinomial random vectors with denominator m_i and p-dimensional probability vector $\boldsymbol{\pi}_i = (\pi_{i1}, \cdots, \pi_{ip})^T$ and covariance matrix $\boldsymbol{\Sigma}_i$ where $\sum_{j=1}^{p} \pi_{ij} = 1$ and $\boldsymbol{\Sigma}_i = m_i \operatorname{diag}(\pi_{ij}) - m_i \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{ip}) = m_i \boldsymbol{\pi}_i$. Consider a multivariate generalized linear regression model with the link function

$$\boldsymbol{\mu}_i = \boldsymbol{h}_i(\boldsymbol{\eta}_i) = (h_{i1}(\boldsymbol{\eta}_{i1}), \cdots, h_{ip}(\boldsymbol{\eta}_{ip}))^T, \tag{2.1}$$

where $\boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{ip})^T$. There are two ways to choose the linear predictor $\boldsymbol{\eta}_i$ for multinomial models. One is that $\boldsymbol{\eta}_i = (\boldsymbol{X}_{i1}^T \beta, \cdots, \boldsymbol{X}_{ip}^T \beta)^T$, where $\boldsymbol{X}_{ij} = (X_{ij1}, \cdots, X_{ijq})^T$ is a q-dimensional vector of covariates, and $\boldsymbol{\beta} =$

$(\beta_1, \cdots, \beta_q)$ is a vector of q regression parameters. Another case is that $\boldsymbol{\eta}_i = (\boldsymbol{Z_i^T}\boldsymbol{\beta}_1, \cdots, Z_i^T\boldsymbol{\beta}_p)^T$, where $\boldsymbol{Z}_i = (Z_{i1}, \cdots, Z_{iq})^T$ is a q-dimensional vector of covariates, and $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{jq})^T$ is a vector of q regression parameters for $j = 1, \cdots, p$. Also, $\boldsymbol{\beta}_r = (\beta_{1r}, \cdots, \beta_{pr})^T$ is a vector of p regression parameters for $r = 1, \cdots, q$. In this case, there are a total of pq regression parameters. Let q' denote the number of the regression parameters. Then, in the former case $q' = q$ and in the later case $q' = qp$. In both cases $\sum_{j=1}^p h_{ij} = m_i$ for $i = 1, 2, \cdots, n$.

Maximum likelihood estimates of the regression parameters $\beta_1, ..., \beta_q$ are obtained as solutions of the q' quasi-likelihood estimating equations $g_r(\hat{\boldsymbol{\beta}}) = 0, r = 1, ..., q$, where

$$g_r(\boldsymbol{\beta}) = \sum_{i=1}^n (\boldsymbol{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \beta_r} = \sum_{i=1}^n \sum_{j=1}^p \frac{y_{ij} - \mu_{ij}}{\mu_{ij}} \frac{\partial \mu_{ij}}{\partial \beta_r}, \tag{2.2}$$

where $\boldsymbol{\Sigma}_i^{-1}$ is the generalized inverse of $\boldsymbol{\Sigma}_i$, that is, $\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i$. Note that, in the second case, equation (2.2) is a p-dimensional vector of equations and thus there are $q' = pq$ likelihood estimating equations.

Goodness of fit is assessed by extending the model to a wider family with covariance matrix $\phi \boldsymbol{\Sigma}_i$ and assessing the departure from the value $\phi = 1$. As in Farrington (1996) and Paul and Deng (2000), the dispersion parameter ϕ is estimated by using a supplementary unbiased estimating equation. The supplementary estimating equation proposed here is $g_d(\hat{\boldsymbol{\beta}}, \hat{\phi}) = 0$, where $d = q' + 1$ and

$$\begin{aligned} g_d(\boldsymbol{\beta}, \phi) &= \sum_{i=1}^n \boldsymbol{a_i^T}(\boldsymbol{y}_i - \boldsymbol{\mu}_i) + \sum_{i=1}^n [(\boldsymbol{y}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) - \phi(p-1)] \\ &= \sum_{i=1}^n \sum_{j=1}^p \left[a_{ij}(y_{ij} - \mu_{ij}) + \frac{(y_{ij} - \mu_{ij})^2}{\mu_{ij}} \right] - n\phi(p-1) \end{aligned} \tag{2.3}$$

which is unbiased. The functions $\boldsymbol{a}_i$ define a family of first-order correction terms to the Pearson statistic, where $\boldsymbol{a}_i = \boldsymbol{a}_i(\boldsymbol{\mu}_i) = (a_{i1}(\mu_{i1}), \cdots, a_{ip}(\mu_{ip}))^T$, which depend on $\boldsymbol{\beta}$ not on ϕ. Obviously, we have

$$\begin{aligned} \hat{\phi} &= \frac{1}{n(p-1)} \sum_{i=1}^n \hat{\boldsymbol{a}}_i^T (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i) + \sum_{i=1}^n [(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)] \\ &= \frac{1}{n(p-1)} \sum_{i=1}^n \sum_{j=1}^p \left(\hat{a}_{ij}(y_{ij} - \hat{\mu}_{ij}) + \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \right), \end{aligned}$$

where $\hat{\boldsymbol{a}}_i = \boldsymbol{a}_i(\hat{\boldsymbol{\mu}}_i), \hat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i(\hat{\boldsymbol{\mu}}_i)$ and $\hat{\boldsymbol{\mu}}_i$ are the maximum likelihood estimates of $\boldsymbol{\mu}_i$. Therefore, $X_*^2 = n(p-1)\hat{\phi}$ defines a modified Pearson Chi-square statistic. In particular, for $\boldsymbol{a}_i = 0$, we have the usual Pearson statistic $X^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(y_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$.

2.2. Moments. For convenience, we introduce some notations. Define $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{ip})^T$ for $i = 1, \cdots, n$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n)^T$. Note that $\boldsymbol{X}_i$ is a $p \times q$ matrix and $\boldsymbol{Z}$ is a $n \times q$ matrix. Define $\boldsymbol{X} = (\boldsymbol{X}_{\boldsymbol{1}}^{\boldsymbol{T}}, \cdots, \boldsymbol{X}_{\boldsymbol{n}}^{\boldsymbol{T}})_{np\times q}^T$ for the first case and $\boldsymbol{X} = (\boldsymbol{Z} \otimes \boldsymbol{I}_p)_{np\times np}$ for the second case, where $\otimes$ means the Kronecker product of matrices and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. Note that $\boldsymbol{X}$ is a $np \times qp$ matrix. Also, define that $h'_{i(j,k)} = \frac{\partial h_{ij}}{\partial \eta_{ik}}, h''_{i(j,k,l)} = \frac{\partial^2 h_{ij}}{\partial \eta_{ik}\partial \eta_{il}}$, $\boldsymbol{h}'_i = \{h'_{i(j,k)}\}_{p\times p}$, $\boldsymbol{h}' = \mathrm{diag}(h'_i)$, $\boldsymbol{\Sigma}^{-1} = \mathrm{diag}(\Sigma_i^{-1})$ $\boldsymbol{W} = \boldsymbol{h}'^T \boldsymbol{\Sigma}^{-1} \boldsymbol{h}'$, and $\boldsymbol{Q} = (Q_{(ij,i'j')}) = \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^T$.

After detailed derivations following McCullagh (1987), we obtain the unconditional moments of the modified Pearson χ^2 statistic which are given in what follows (proofs can be obtained from the authors).

$$
\begin{aligned}
&E(X_*^2) = n(p-1)E(\hat{\phi}) = n(p-1) - q' + \sum_{i=1}^n \frac{1}{m_i} \sum_{j,k,l=1}^p \frac{1}{\mu_{ij}} h'_{i(j,k)} h'_{i(j,l)} Q_{(ik,il)} \\
&-\frac{1}{2}\sum_{i=1}^n \sum_{j,k,l=1}^p (a_{ij} + \mu_{ij}^{-1}) h''_{i(j,k,l)} Q_{(ik,il)} \\
&+\frac{1}{2}\sum_{i,i'=1}^n \sum_{j,j',k,k',l',m'=1}^p (a_{ij} + \mu_{ij}^{-1}) \mu_{i'j'}^{-1} h'_{i(j,k)} h'_{i'(j',m')} h''_{i'(j',k',l')} Q_{(ik,i'm')} \\
&\times Q_{(i'k',i'l')} + O(n^{-1}), \\
&\mathrm{var}(X_*^2) = n^2(p-1)^2 \,\mathrm{var}(\hat{\phi}) \\
&= 2(p-1)n - 2(p-1)\sum_{i=1}^n \frac{1}{m_i} + \sum_{i=1}^n \sum_{j=1}^p (a_{ij} + \mu_{ij}^{-1})^2 \mu_{ij} \\
&- \sum_{i=1}^n \frac{1}{m_i} (\sum_{j=1}^p (a_{ij} + \mu_{ij}^{-1}) \mu_{ij})^2 \\
&- \sum_{i,i'=1}^n \sum_{j,j',k,k'=1}^p (a_{ij} + \mu_{ij}^{-1})(a_{i'j'} + \mu_{i'j'}^{-1}) h'_{i(j,k)} h'_{i'(j',k')} Q_{(ik,i'k')} + O(1)
\end{aligned}
$$

and

$$
\kappa_3(X_*^2) = n^3(p-1)^3 \kappa(\hat{\phi})
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{p}\{-[P_{ij}\mu_{ij}^{-1} - (a_{ij} + \mu_{ij}^{-1})]^3\mu_{ij} - 3(\sum_{k=1}^{p}\pi_{ik}a_{ik})[P_{ij}\mu_{ij}^{-1} \\
&-(a_{ij} + \mu_{ij})]^2\mu_{ij} + \frac{3(2m_i - p - 2)}{m_i}[P_{ij}\mu_{ij}^{-1} - (a_{ij} + \mu_{ij}^{-1})]^2\mu_{ij} \\
&-6(a_{ij} + \mu_{ij}^{-1})[P_{ij}\mu_{ij}^{-1} - (a_{ij} + \mu_{ij}^{-1})]\mu_{ij} - 12(1 - \frac{1}{m_i})[P_{ij}\mu_{ij}^{-1} - (a_{ij} + \mu_{ij}^{-1})]\} \\
&+\sum_{i=1}^{n}\{4(1 - \frac{1}{m_i})\sum_{j=1}^{p}\mu_{ij}^{-1} + 2m_i[\sum_{j=1}^{p}\pi_{ij}(a_{ij} + \mu_{ij}^{-1})]^3 \\
&-6(m_i - 1)[\sum_{j=1}^{p}\pi_{ij}(a_{ij} + \mu_{ij}^{-1})]^2 - 12p\frac{m_i - 1}{m_i}[\sum_{j=1}^{p}\pi_{ij}(a_{ij} + \mu_{ij}^{-1})] \\
&-3\sum_{j=1}^{p}(a_{ij} + \mu_{ij}^{-1})^2\mu_{ij} + \frac{m_i - 1}{m_i^2}[4p(2m_i - 7) - 8(m_i - 3)]\} + O(1)
\end{aligned}
$$

where for $i = 1, 2, \cdots, n, j = 1, 2, \cdots, p$,

$$
P_{ij} = \sum_{i'=1}^{n}\sum_{j',k,k'=1}^{p}(a_{i'j'} + \mu_{i'j'}^{-1})h'_{i'(j',k')}h'_{i(j,k)}Q_{(i'k',ik)}.
$$

Now we consider the conditional moments of $\boldsymbol{X}_*^2$. By using the formulae in McCullagh (1987), Section 5.6, it can be shown that

$$
\text{cov}(\hat{\boldsymbol{\beta}}, \hat{\phi}) = O(n^{-2}). \tag{2.4}
$$

Expanding equations $g_r, r = 1, 2, ...q$ (equation (2.2)) and g_d (equation (2.3)) about $\hat{\boldsymbol{\beta}}$ and following steps similar to Farrington (1996), using equation (2.4), the unconditional moments of $\boldsymbol{X}_*^2$, following McCullagh (1987) and after detailed calculations, we obtain the conditional moments of X_*^2 given $\hat{\beta}$ as

$$
\begin{aligned}
&E(X_*^2|\hat{\boldsymbol{\beta}}) = n(p-1) - q' + \sum_{i=1}^{n}\frac{1}{m_i}\sum_{j,k,l=1}^{p}\frac{1}{\hat{\mu}_{ij}}\hat{h}'_{i(j,k)}\hat{h}'_{i(j,l)}\hat{Q}_{(ik,il)} \\
&-\frac{1}{2}\sum_{i=1}^{n}\sum_{j,k,l=1}^{p}(\hat{a}_{ij} + \hat{\mu}_{ij}^{-1})\left(2\hat{h}''_{i(j,k,l)} - \hat{\mu}_{ij}^{-1}\hat{h}'_{i(j,k)}\hat{h}'_{i(j,l)}\right)\hat{Q}_{(ik,il)} \\
&+\frac{1}{2}\sum_{i,i'=1}^{n}\sum_{j,j',k,k',l',m'=1}^{p}(\hat{a}_{ij} + \hat{\mu}_{ij}^{-1})\left(2\hat{\mu}_{i'j'}^{-1}\hat{h}''_{i'(j',k',l')} - \hat{\mu}_{i'j'}^{-2}\hat{h}'_{i'(j',k')}\hat{h}'_{i'(j',l')}\right) \\
&\times\hat{h}'_{i(j,k)}\hat{h}'_{i'(j',m')}\hat{Q}_{(ik,i'm')}\hat{Q}_{(i'k',i'l')} + O(n^{-1}),
\end{aligned}
$$

$$\begin{aligned}\mathrm{var}(X_*^2|\hat{\boldsymbol{\beta}}) &= 2(p-1)n - 2(p-1)\sum_{i=1}^{n}\frac{1}{m_i}\\ &+\sum_{i=1}^{n}\sum_{j=1}^{p}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})^2\hat{\mu}_{ij} - \sum_{i=1}^{n}\frac{1}{m_i}(\sum_{j=1}^{p}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})\hat{\mu}_{ij})^2\\ &-\sum_{i,i'=1}^{n}\sum_{j,j',k,k'=1}^{p}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})(\hat{a}_{i'j'}+\hat{\mu}_{i'j'}^{-1})\hat{h}'_{i(j,k)}\hat{h}'_{i'(j',k')}\hat{Q}_{(ik,i'k')}+O(1)\end{aligned}$$

and

$$\begin{aligned}\kappa_3(X_*^2|\hat{\boldsymbol{\beta}}) &= \sum_{i=1}^{n}\sum_{j=1}^{p}\{-[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]^3\hat{\mu}_{ij}\\ &-3(\sum_{k=1}^{p}\hat{\pi}_{ik}\hat{a}_{ik})[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-(\hat{a}_{ij}+\hat{\mu}_{ij})]^2\hat{\mu}_{ij}+\frac{3(2m_i-p-2)}{m_i}[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]^2\hat{\mu}_{ij}\\ &-6(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]\hat{\mu}_{ij}-12(1-\frac{1}{m_i})[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]\}\\ &+\sum_{i=1}^{n}\{4(1-\frac{1}{m_i})\sum_{j=1}^{p}\hat{\mu}_{ij}^{-1}+2m_i[\sum_{j=1}^{p}\hat{\pi}_{ij}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]^3\\ &-6(m_i-1)[\sum_{j=1}^{p}\hat{\pi}_{ij}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]^2-12p\frac{m_i-1}{m_i}[\sum_{j=1}^{p}\hat{\pi}_{ij}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})]\\ &-3\sum_{j=1}^{p}(\hat{a}_{ij}+\hat{\mu}_{ij}^{-1})^2\hat{\mu}_{ij}+\frac{m_i-1}{m_i^2}[4p(2m_i-7)-8(m_i-3)]\}+O(1),\end{aligned}$$

where, for $i=1,2,\cdots,n, j=1,2,\cdots,p$,

$$\hat{P}_{ij}=\sum_{i'=1}^{n}\sum_{j',k,k'=1}^{p}(\hat{a}_{i'j'}+\hat{\mu}_{i'j'}^{-1})\hat{h}'_{i'(j',k')}\hat{h}'_{i(j,k)}\hat{Q}_{(i'k',ik)}.$$

All these expressions of the unconditional and conditional moments, although look tedious, are necessary for the calculation of the test statistics $\boldsymbol{X}_*^2$, Z_* and Z. Moreover, from the next section, one can see that these expressions are tremendously simplified by choosing the specific values of a_{ij}. For $a_{ij}=0$, the approximations to the first three conditional moments of the usual Pearson statistic X^2 are

$$E(X^2|\hat{\boldsymbol{\beta}})=n(p-1)-q'+\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j,k,l=1}^{p}\frac{1}{\hat{\mu}_{ij}}\hat{h}'_{i(j,k)}\hat{h}'_{i(j,l)}\hat{Q}_{(ik,il)}$$

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j,k,l=1}^{p}\hat{\mu}_{ij}^{-1}\left(2\hat{h}''_{i(j,k,l)}-\hat{\mu}_{ij}^{-1}\hat{h}'_{i(j,k)}\hat{h}'_{i(j,l)}\right)\hat{Q}_{(ik,il)}$$
$$+\frac{1}{2}\sum_{i,i'=1}^{n}\sum_{j,j',k,k',l',m'=1}^{p}\hat{\mu}_{ij}^{-1}\left(2\hat{\mu}_{i'j'}^{-1}\hat{h}''_{i'(j',k',l')}-\hat{\mu}_{i'j'}^{-2}\hat{h}'_{i'(j',k')}\hat{h}'_{i'(j',l')}\right)$$
$$\times\hat{h}'_{i(j,k)}\hat{h}'_{i'(j',m')}\hat{Q}_{(ik,i'm')}\hat{Q}_{(i'k',i'l')}+O(n^{-1}),$$
$$\mathrm{var}(X^2|\hat{\beta})=2(p-1)n-(p^2+2p-2)\sum_{i=1}^{n}\frac{1}{m_i}+\sum_{i=1}^{n}\sum_{j=1}^{p}\hat{\mu}_{ij}^{-1}$$
$$-\sum_{i,i'=1}^{n}\sum_{j,j',k,k'=1}^{p}\hat{\mu}_{ij}^{-1}\hat{\mu}_{i'j'}^{-1}\hat{h}'_{i(j,k)}\hat{h}'_{i'(j',k')}\hat{Q}_{(ik,i'k')}+O(1)$$

and

$$\kappa_3(X^2|\hat{\beta})=\sum_{i=1}^{n}\sum_{j=1}^{p}\{-[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-\hat{\mu}_{ij}^{-1}]^3\hat{\mu}_{ij}+\frac{3(2m_i-p-2)}{m_i}[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-\hat{\mu}_{ij}^{-1}]^2\hat{\mu}_{ij}$$
$$-(18-\frac{12}{m_i})[\hat{P}_{ij}\hat{\mu}_{ij}^{-1}-\hat{\mu}_{ij}^{-1}]+(1-\frac{4}{m_i})\hat{\mu}_{ij}^{-1}\}$$
$$+\sum_{i=1}^{n}\{\frac{2p^3-18(m_i-1)p^2}{m_i^2}+\frac{m_i-1}{m_i^2}[4p(2m_i-7)-8(m_i-3)]\}+O(1),$$

where for $i=1,2,\cdots,n, j=1,2,\cdots,p$,

$$\hat{P}_{ij}=\sum_{i'=1}^{n}\sum_{j',k,k'=1}^{p}\hat{\mu}_{i'j'}^{-1}\hat{h}'_{i'(j',k')}\hat{h}'_{i(j,k)}\hat{Q}_{(i'k',ik)}.$$

3 Assessing Goodness of Fit

Equation (2.4) in Section 2.2 means that the modified Pearson statistic and the regression parameters are asymptotically uncorrelated. Further, it can be established that

$$E(g_r g_d)=\sum_{i=1}^{n}(\boldsymbol{a}_i+\boldsymbol{\mu}_i^{-1})^T\frac{\partial\boldsymbol{\mu}_i}{\partial\beta_r}.$$

Choosing $a_{ij}=-\mu_{ij}^{-1}$ yields $E(g_r g_d)=0$, and thus local orthogonality of the estimates of the regression parameters and the estimate of the dispersion parameter is obtained. This means that $\hat{\phi}$ varies only slowly with $\boldsymbol{\beta}$ in the neighborhood of $\hat{\boldsymbol{\beta}}$ (Cox and Reid, 1987). For this choice of the a_{ij}, using expansion of g_d (see Firth, 1987), the modified Pearson statistic

$$X_*^2 = n(p-1)\hat{\phi} \;=\; \sum_{i=1}^{n} -(\hat{\boldsymbol{\mu}}_i^{-1})^T(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i) + \sum_{i=1}^{n}[(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)^T\hat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)]$$

depends only weakly on $\boldsymbol{\beta}$, given $\hat{\boldsymbol{\beta}}$. This choice of $\boldsymbol{a}_i$ also minimizes the variance of the modified Pearson statistic X_*^2. This can be seen by writing $\text{var}(X_*^2)$ as

$$\text{var}(X_*^2) = \boldsymbol{\alpha}^T(\boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T})\boldsymbol{\alpha} + 2(p-1)\sum_{i=1}^{n}(1 - m_i^{-1}) + O(1),$$

where $\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{a}+\boldsymbol{\mu}^{-1})$ with $\boldsymbol{a}+\boldsymbol{\mu}^{-1} = (a_{11}+\mu_{11}^{-1}, \cdots, a_{1p}+\mu_{1p}^{-1}, \cdots, \cdots, a_{n1}+\mu_{n1}^{-1}, \cdots, a_{np}+\mu_{np}^{-1})^T$, and $X^* = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{h}'\boldsymbol{X}$. The first term on the right-hand side of the above equation is the residual sum of squares from the regression of $\boldsymbol{\alpha}$ on X^*. The choice $a_{ij} = -\mu_{ij}^{-1}$ makes this 0. Further, for $a_{ij} = -\mu_{ij}^{-1}$, we have

$$E(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{E}(X_*^2) = n(p-1) - q' + \sum_{i=1}^{n}\sum_{j,k,l=1}^{p}\frac{1}{m_i\hat{\mu}_{ij}}\hat{h}'_{i(j,k)}\hat{h}'_{i(j,l)}\hat{Q}_{(ik,il)} + O(n^{-1}),$$

$$\text{var}(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{\text{var}}(X_*^2) = 2(p-1)n - 2(p-1)\sum_{i=1}^{n}\frac{1}{m_i} + O(1),$$

and

$$\begin{aligned}\kappa_3(X_*^2|\hat{\boldsymbol{\beta}}) &= \hat{\kappa}_3(X_*^2) \\ &= \sum_{i=1}^{n}\{4(1 - \frac{1}{m_i})\sum_{j=1}^{p}\hat{\mu}_{ij}^{-1} + \frac{m_i - 1}{m_i^2}[4p(2m_i - 7) - 8(m_i - 3)]\} + O(1).\end{aligned}$$

In particular, for $p = 2$, we note that $h_{i2} = m_i - h_{i1}$ and $\eta_{i2} = 0$. Therefore, if $a_{ij} = -\mu_{ij}^{-1}$ for $j = 1, 2, a_i = -\frac{1-2\pi_i}{V_i}$. Also, for $i = 1, \cdots, n$, $h'_{i(2,1)} = -h'_{i(1,1)} = h'_i$, $Q_{(i1,i1)} = Q_{ii}$ and $h'_{i(j,l)} = 0, Q_{(ik,il)} = 0$, for $k \neq 1$ or $l \neq 1$, thus,

$$\sum_{i=1}^{n}\sum_{j,k,l=1}^{p}\frac{1}{m_i\mu_{ij}}h'_{i(j,k)}h'_{i(j,l)}Q_{(ik,il)} = \sum_{i=1}^{n}\sum_{j=1}^{2}\frac{h'_ih'_i}{m_i\mu_{ij}}Q_{ii} = \sum_{i=1}^{n}\frac{h_i'^2}{m_i^2\pi_i(1-\pi_i)}Q_{ii}.$$

Hence, after incorporating a degree-of-freedom correction, for the binomial case we have

$$E(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{E}(X_*^2) = n - q' + \sum_{i=1}^{n} \frac{\hat{h}_i'^2}{m_i^2 \hat{\pi}_i(1-\hat{\pi}_i)} \hat{Q}_{ii} + O(n^{-1}),$$

$$\text{var}(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{\text{var}}(X_*^2) = 2(1 - \frac{q'}{n}) \sum_{i=1}^{n} \frac{m_i - 1}{m_i} + O(1),$$

and

$$\kappa_3(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{\kappa}_3(X_*^2)$$
$$= 8(n-q')\left[1 - \frac{1}{n}\sum_{i=1}^{n} \frac{5m_i - 4}{m_i^2} + \frac{1}{2n}\sum_{j=1}^{n} \frac{m_i - 1}{m_i^2}\{\hat{\pi}_i(1-\hat{\pi}_i)\}^{-1}\right] + O(1),$$

which coincide with the expressions obtained by Farrington (1996).

It can be seen that the forms of the approximations to the first three moments of the conditional distribution of the modified Pearson statistic X_*^2 are quite simple to use. In particular, for $m_i = m (i = 1, 2, \cdots, n)$, up to the first order of n,

$$E(X_*^2|\hat{\boldsymbol{\beta}}) \doteq n(p-1) - q'(1 - m^{-1}),$$
$$\text{var}(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{\text{var}}(X_*^2) \doteq 2n(p-1)(1 - m^{-1}),$$

and

$$\kappa_3(X_*^2|\hat{\boldsymbol{\beta}}) = \hat{\kappa}_3(X_*^2) \doteq 4(1 - m^{-1}) \sum_{i=1}^{n}\sum_{j=1}^{p} \hat{\mu}_{ij}^{-1}$$
$$+ nm^{-2}(m-1)[4p(2m-7) - 8(m-3)].$$

Now, the method based on the first two conditional moments is to calculate the p-value $P(Z_* \geq z|\hat{\boldsymbol{\beta}})$, where

$$Z_* = \{X_*^2 - E(X_*^2|\hat{\boldsymbol{\beta}})\}/[\text{var}(X_*^2|\hat{\boldsymbol{\beta}})]^{1/2}$$

which has an approximate standard normal distribution. The method based on the first three conditional moments is to calculate the approximate p-value using the Edgeworth approximation

$$P(Z_* \geq z) \doteq 1 - \Phi(z) + (z^2 - 1)\rho_3 \phi(z)/6,$$

where ρ_3 is the approximate standardized conditional third moment (see McCullagh, 1986).

4 Simulation

In this section a limited simulation study is conducted to compare, in terms of empirical size and power, the usual Pearson Chi-square statistic X^2, the standardized modified Pearson Chi-square statistic X_*^2, namely, $Z_* = (X_*^2 - E(X_*^2|\boldsymbol{\beta})/\sqrt{\text{var}(X_*^2|\boldsymbol{\beta})}$, a method Z using Edgeworth approximation of the p-values using the first three conditional moments as described in Section 3 and a score test statistic S obtained in Paul et al. (1989), where

$$S = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\{(y_{ij}-\mu_{ij})^2-(y_{ij}-\mu_{ij})\}/2\pi_{ij} - \frac{m_i}{2}(p-1)}{\sqrt{\frac{p-1}{2}\sum_{i=1}^{n} m_i(m_i-1)}}.$$

which has a standard normal distribution.

For empirical size, simulations have been conducted for the multinomial model with dimension $p = 4$, and continuous covariates have been chosen to induce very strong regression effects, for sample sizes varying from $n = 10$ to $n = 100$ and multinomial denominators $m = 10$ and 20. For generating multinomial data, we used the link functions $\eta_{ij} = \beta_0 + X_{ij}\beta_1, \pi_{ij} = \exp(\eta_{ij})/\{1 + \sum_{k=1}^{p-1}(\exp(\eta_{ik})\}$ and $\pi_{ip} = 1/\{1 + \sum_{k=1}^{p-1}(\exp(\eta_{ik})\}$ for $i = 1, \cdots, n$ and $j = 1, \cdots, p-1$.

We then extended the simulation study to compare power of the four methods. For power comparison, we simulated data from the Dirichlet-multinomial distribution with mean vector $m\pi$ and covariance matric $\{1 + \rho^2(m-1)\}\Sigma_i$ by using the algorithm given by Morel (1992) where π_{ij} are the same as above and ρ is the over-dispersion parameter. Simulations have been conducted with $m = 10, 20$,sample sizes $n = 10, 20$, and over-dispersion parameter ρ=0.15 for α =0.01, 0.05, 0.10. Behavior of the three statistics in terms of power is similar for other values of ρ which are not given here. Results of the empirical sizes and powers are shown in Figures 1, 2 and 3.

The simulation results in Figure 1 show that for $\alpha = 0.01$, there is no qualitative difference in size of the four methods for different sample sizes. However, the modified statistic Z^* shows better power performance than the other three statistics Z, X^2 and S. Further, power performance of the statistic Z, based on Edgeworth-approximation statistic, is better than that of the usual Pearson statistic X^2 and the score test statistic S.

From Figure 2 we can see that for $\alpha = 0.05$, the four test statistics are somewhat conservative for small sample sizes ($n = 10, 20, 40$) but maintain nominal level well for large sample sizes. In terms of power, the method Z

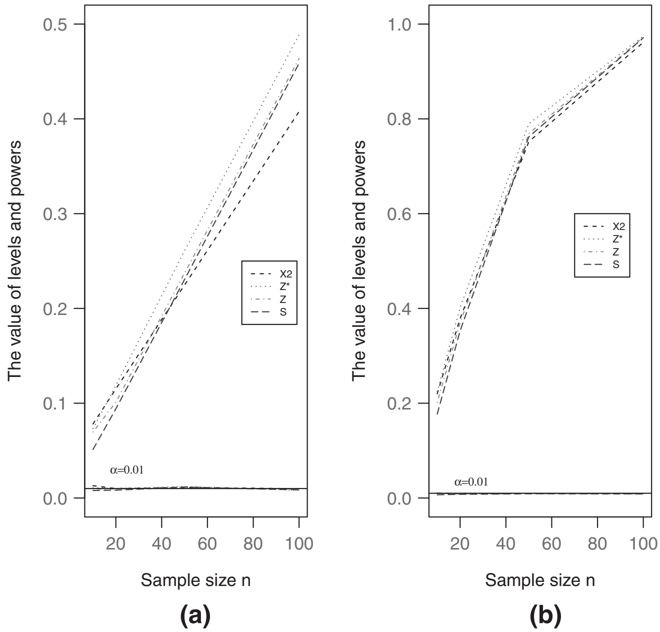

Figure 1: A comparison of the empirical powers and levels of test statistics X^2, Z_*, Z and S: **a** $\alpha = 0.01, \phi = 0.15, m = 10$; **b** $\alpha = 0.01, \phi = 0.15, m = 20$

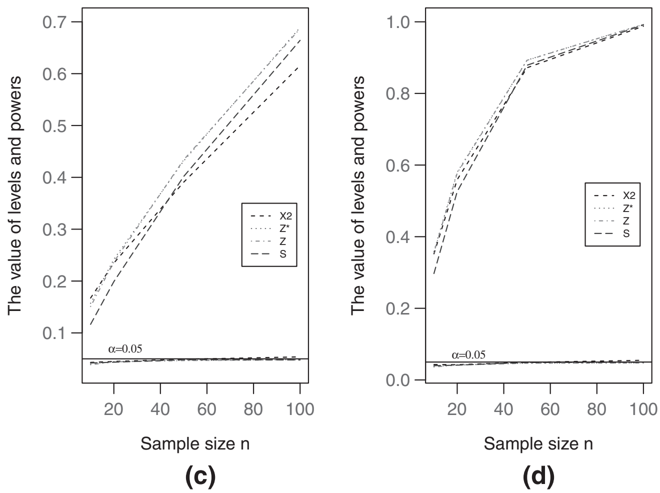

Figure 2: A comparison of the empirical powers and levels of test statistics X^2, Z_*, Z and S: **c** $\alpha = 0.05, \phi = 0.15, m = 10$; **d** $\alpha = 0.05, \phi = 0.15, m = 20$

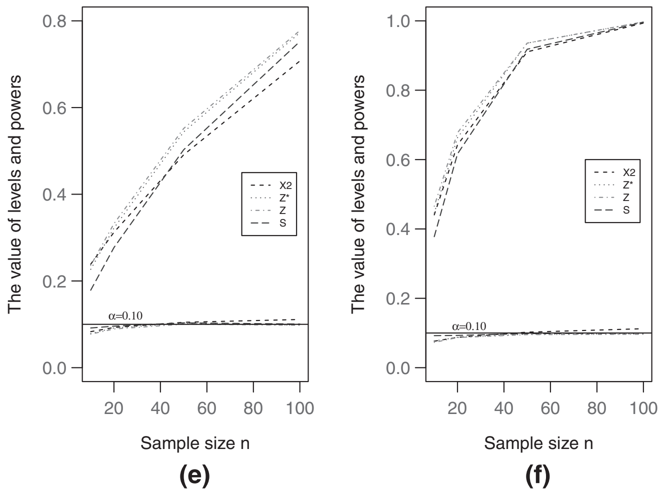

Figure 3: A comparison of the empirical powers and levels of test statistics X^2, Z_*, Z and S: **e** $\alpha = 0.10, \phi = 0.15, m = 10$; **f** $\alpha = 0.10, \phi = 0.15, m = 20$

using the Edgeworth approximation of the p-values using the first three conditional moments and the standardized modified statistic Z^* shows similar advantage compared to the usual chi-squared statistic X^2, and the score test statistic S.

Figure 3 demonstrates similar behaviors of the four methods to the case $\alpha = 0.05$. The four methods are somewhat conservative for small sample sizes $(n = 10, 20)$ and hold level reasonably well even for moderate sample sizes except for the usual Pearson test statistic. For example, for $\alpha = 0.10$, $m = 20$ and $n = 10, 20$, the score test statistic holds nominal level reasonably well, whereas the other three statistics show some conservative behavior. Further, in terms of empirical power, the Edgeworth-approximation based statistic Z performs somewhat better than the standardized modified statistic X^2 and both Z and Z^* perform better than the usual Pearson statistic X^2 and the score test statistic S. What is interesting is that even though for $\alpha = 0.10$, $m = 20$ and $n = 10, 20$, the score test statistic holds nominal level better than the other two statistics. However, its power is much smaller than that of the other two statistics.

In summary, for small nominal level α, our recommendation is to use the standardized modified Pearson test statistic Z^*. However, for large values of α, the Edgeworth-approximation based statistic Z can be used for assessing goodness of fit in multinomial regression models.

5 Examples

Example 1. *Fahrmeir and Tutz (1994) provide data on job expectations of students of psychology in Regensburg. In a study on the perspectives of students, psychology students have been asked if they expect to find an adequate job after their graduation. The responses are classified into three categories: "don't expect adequate employment" (category 1), "not sure" (category 2) and "immediately after the degree" (category 3). The data are sparse with 8 out of 13 annual age groups contributing denominators of 3 or less. Using the multinomial model with three categories and the link function* $\pi_{ij} = \exp\{\beta_{j1} + \log(age(i))\beta_{j2}\}/(1 + \sum_{k=1}^{p-1} \exp\{\beta_{k1} + \log(age(i))\beta_{k2}\}$ *for* $j = 1, 2$ *and* $\pi_{i3} = 1/(1 + \sum_{k=1}^{2} \exp\{\beta_{k1} + \log(age(i))\beta_{k2}\}$ *we obtain* $X^2 = 35.18$ *with 22 degrees of freedom.,* $X_*^2 = 22.80$ *with conditional expected value 23.06, conditional variance 28.65, the conditional standardized third moment 0.8256 and the value of the score test statistic is* $S = -1.1286$. *Note, we are dealing with two tailed tests. The p-values of* X^2*, the method based on standardized* X_*^2*, the method based on Edgeworth approximation, and the score statistic* S *are respectively 0.0371, .9619, .8325 and 0.2590*

Note that there are significant differences between the p-values of the procedures based on the modified Pearson Statistic and the p-value of the usual Pearson statistic. The tests based on the modified Pearson statistic and the score test show that the fit of the model is good whereas the test based on the usual Pearson statistic indicates evidence against the model. This is possibly because sparseness of the data results in the inflation of usual Pearson statistic.

Now, for the purpose of comparison the observations with denominator one are first eliminated, and then it can be obtained that $X^2 = 8.03$ with 14 degrees of freedom, $X_*^2 = 11.35$ with conditional expected value 14.53, conditional variance 28.65, the conditional standardized third moment 2.193 and the value of the score test statistic is $S = -1.6658$. The p-values of X^2, the method based on standardized X_*^2, the method based on Edgeworth approximation, and the score statistic S are respectively 0.8878, .5524, .3942 and .0958.

In this case, the p-value of the usual Pearson statistic X^2 is much closer to the p-values of the two procedures based on the standardized modified Pearson statistic Z^*, and the conclusions are also the same. Comparing with the original sparse data, the new dataset does not display the sparseness, in which, the usual Pearson statistic X^2 and the modified Pearson statistic X_*^2 have the similar behaviors. Since the correction term $\sum_{i=1}^{n} \sum_{j=1}^{p} (y_{ij} - \mu_{ij})/\mu_{ij}$ is added to the modified Pearson statistic, which not only reduces the effect of sparseness in the data but also minimizes the variance of the Pearson statistic. This example has indeed shown that the Pearson statistic can be severely affected by the presence of sparseness in the data. Our overall conclusion is that the fit of the model is good.

Example 2. *Agresti (2002) presents a data set on a study of factors influencing the primary food choice of alligators. It used 219 alligators captured in four Florida lakes. The nominal response variable is the primary food type, in volume, found in an alligator's stomach. This had five categories: fish, invertebrate, reptile, bird, and other. More details and the data can be found in Agresti (page 268-269, 2002). The data display sparseness with 40 of 80 cells showing counts less than or equal to one. Now we fit the data by using the following multinomial regression model with five (5) response categories and the link* $\pi_{ij} = \exp\{\beta_{j0} + x_{1i}\beta_{j1} + x_{2i}\beta_{j2} + x_{3i}\beta_{j3} + x_{4i}\beta_{j4} + x_{5i}\beta_{j5}\}/(1 + \sum_{k=1}^{4} \exp\{\beta_{k0} + x_{1i}\beta_{k1} + x_{2i}\beta_{k2} + x_{3i}\beta_{k3} + x_{4i}\beta_{k4} + x_{5i}\beta_{k5}\}$ *for* $j = 1, 4$ *and* $\pi_{i5} = 1/(1 + \sum_{k=1}^{4} \exp\{\beta_{k0} + x_{1i}\beta_{k1} + x_{2i}\beta_{k2} + x_{3i}\beta_{k2} + x_{4i}\beta_{k4} + x_{5i}\beta_{k5}\}$, *where* $x_1 = 1$ *for female, 0 for male;* $x_2 = 1$ *for size* < 2.3, *0 for size* ≥ 2.3*;* $x_3 = 1$ *for lake Hancock, 0 for other;* $x_4 = 1$ *for lake Oklawaha, 0 for other;* $x_5 = 1$ *for lake Trafford, 0 for other. Therefore, we have* $\boldsymbol{\eta} = (\boldsymbol{Z}^T\boldsymbol{\beta}_1, \cdots, \boldsymbol{Z}^T\boldsymbol{\beta}_4)^T$ *where* $\boldsymbol{Z} = (1, x_1, \cdots, x_5)^T$ *is a 6-dimensional vector of covariates, and* $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \cdots, \beta_{j5})^T$ *is a vector of 6 regression parameters for* $j = 1, \cdots, 4$. *There are 24 parameters. The parameter* $\boldsymbol{\beta}_j$ *represents the effect of the covariate to the* j*th component of the multinomial variable and thus* $\beta_{j3}, \beta_{j4}, \beta_{j5}$ *represent the effects of Hancock, Oklawaha, and Trafford to the* j*th component, respectively;* $\beta_{j0} - \beta_{j3} - \beta_{j4} - \beta_{j5}$ *is the effect of Lake George to the* j*th components,* $j = 1, \cdots, 4$.

In order to test the goodness of fit of this model to the data, we obtain $X^2 = 52.57$ on 40 degrees of freedom and $X_*^2 = 44.83$ with conditional expected value 42.08, conditional variance 106.65, the conditional standardized third moment 0.6334, and the value of the score statistic is $S = -1.6707$. The p-values of X^2, the method based on standardized X_*^2, the method based on Edgeworth approximation, and the score statistic S are respectively 0.088, 0.790, 0.7143 and .0948.

As we can see, there is significant difference between the p-value of the usual Pearson statistic and those of the two procedures based on the modified Pearson statistic.

All four methods show evidence of good fit of the model to the data. However, the test based on the standardized modified Pearson statistic and that based on the Edgeworth approximation show strong evidence in favor of the conclusion that the model fits the data well, where as, the evidence of fit of the model based on the usual Pearson statistic and Score statistic is only marginal (p-value=0.088 for X^2 and that for S is .0948).

6 Discussions

We have derived two procedures for testing goodness of fit of the Product Multinomial Regression Models to Sparse Data. These methods are: a standardized modified Pearson Chi-square statistic using the first two conditional moments, and a method using Edgeworth approximation of the p-values based on the first three conditional moments. These methods are then compared in terms of level and power with the usual Pearson Chi-square statistic and a score test statistic. Simulations show no qualitative difference in size of the four procedures. All four methods hold levels reasonably well. However, the standardized modified Pearson Chi-square statistic X_*^2 and the method Z using Edgeworth approximation of the p-values using the first three conditional moments show power advantages compared to the other two procedures. Data analysis in Section 5 show similar p-values for these two methods (p-values for example 1 are .9619 and .8325 and those for the modified example 1 and example 2 are, respectively, .5524 and .3942, and .790 and .7143). However, the simulation study in Section 4 show that the statistic X_*^2 shows slightly higher power than the procedure Z for small α ($\alpha = 0.01$). The former also takes a less computational effort and so it is recommended for use to test goodness of fit of product multinomial regression models to sparse data.

Cressie and Read (1984) proposed the family of power-divergence statistics SD_λ depending on a real parameter to assess the goodness of fit in multinomial models. Each member of SD_λ is a sum over all cells of "deviation" between observed and expected counts:

$$SD_\lambda = \frac{2}{\lambda+1} \sum_{ij} \left\{ \frac{y_{ij}}{\lambda} \left[\left(\frac{y_{ij}}{\mu_{ij}} \right)^\lambda - 1 \right] - (y_{ij} - \mu_{ij}) \right\}.$$

As we know, the Pearson Statistic can be obtained from SD_λ by taking $\lambda = 1$. Therefore we have obtained approximations to the first three moments to the family of power-divergence statistics SD_λ for a specific value of λ. It may be of interest to derive approximations to the three moments of the conditional distributions of the family of power-divergence statistics SD_λ for any value of λ. We plan to carry this out in a future study along with a comparison, by simulations, of the performance of the procedures based on the modified Pearson statistics and those from the family of power-divergence statistic. The challenge, however, lies in obtaining closed forms of the moments for the family of power-divergence statistics SD_λ. Further extensions would be to develop a procedure based on the first four moments of the modified Pearson statistic and develop procedures based on the first four moments of the modified deviance statistics (to be derived) as in Paul and Deng (2000) and Paul and Deng (2012).

Acknowledgments. This research was partially supported by the Natural Science and Engineering Research Council of Canada.

References

AGRESTI, A. (2002). *Categorical data analysis*, 2nd. John Wiley & Sons, New York.

COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B **49**, 1–39.

CRESSIE, N. and READ, I. (1984). Multinomial goodness-of-fit Tests. *J. R. Statist. Soc.* B **46**, 440–464.

DALE, J. (1986). Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *J. R. Statist. Soc.* B **48**, 48–59.

FAGERLAND, M. W., HOSMER, D. W. and BOFIN, A. M. (2008). Multinomial goodness-of-fit tests for logistic regression models. *Statist. Med.* **27**, 4238–4253.

FAHRMEIR, L. and TUTZ, G. (1994). *Multivariate statistical modelling based on generalized linear models.* Springer-Verlag, New York.

FARRINGTON, C. P. (1996). On assessing goodness of fit of generalized linear models to sparse data. *J. R. Statist. Soc.* B **58**, 349–360.

FIRTH, D. (1987). Discussion of 'parameter orthogonality and approximate conditional inference' by D.R. Cox and N. Reid. *J. R. Statist. Soc.* B **49**, 22–23.

KIM, S., CHOI, H. and LEE, S. (2009). Estimate-based goodness-of-fit test for large sparse multinomial distributions. *Computational Statistics and Data Analysis* **53**, 1122–1131.

KOEHLER, K. J. (1986). Goodness-of-fit tests for log-linear models in sparse contingency tables. *J. Am. Statist. Ass.* **81**, 483–493.

KOEHLER, K. J. and LARNTZ, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *J. Am. Statist. Ass.* **75**, 336-344.

LEWIS, T., SAUNDERS, I. W. and WESTCOTT, M. (1984). The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables. *Biometrika* **71**, 515–522.

MCCULLAGH, P. (1985). On the asymptotic distribution of Pearson's statistic in linear exponential family models. *Int. Statist. Rev.* **53**, 61–67.

MCCULLAGH, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *J. Am. Statist. Ass.* **81**, 104–107.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, London.

MOREL, J. G. (1992). A simple algorithm for generating multinomial random vectors with extra variation. *Comm. Statist.* **21**, 1255–1268.

OSIUS, S. and ROJEK, D. (1992). Normal goodness-of-fit tests for parametric multinomial models with large degree of freedom. *J. Am. Statist. Ass.* **87**, 1145–1152.

PAUL, S. R. and DENG, D. (2000). Goodness of fit of generalized linear models to sparse data. *J. R. Statist. Soc.* B. **62**, 323–333.

PAUL, S. R. and DENG, D. (2012). Assessing goodness of fit of generalized linear models to sparse data using higher order moment corrections. *Sankhya: The Indian Journal of Statistics* (B) **74**, 195–210.

PAUL, S. R., LIANG, K. Y. and SELF, S. G. (1989). On Testing Departure from the Binomial and Multinomial Assumptions. *Biometrics* **45**, 231–236.

STAFFORD, J. E. (1995). Exact cumulant calculations for Pearson χ^2 and Zelterman statistics for r-way contingency tables. *Journal of Computational and Graphical Statistics* **4**, 199–212.

ZELTERMAN, D. (1987). Goodness-of-fit tests for large sparse multinomial distributions. *J. Amer. Statist. Assoc.* **82**, 624–629.

Dianliang Deng
Department of Mathematics and
Statistics,
University of Regina,
Regina,
SK S4S 0A2 Canada
E-mail: deng@uregina.ca

Sudhir R. Paul
Department of Mathematics and
Statistics,
University of Windsor,
Windsor,
ON N9B 3P4 Canada
E-mail: smjp@uwindsor.ca

Paper received: 24 January 2013.