Contingency-Table Sparseness under Cumulative Logit Models for Ordinal Response Categories and Nominal Explanatory Variables with Two-Factor Interaction

Sujin Sukgumphaphan and Veeranun Pongsapukdee*

Department of Statistics, Faculty of Science, Silpakorn University, Nakhon Pathom 73000, Thailand.

*Author for correspondence; e-mail: veeranun@su.ac.th

Received: 10 July 2007
Accepted: 19 September 2007.

Abstract

In this article the sparseness and the assessing goodness of fit of cumulative models for ordinal response categories and nominal explanatory variables with two-factor interaction are investigated. The sparseness is computed from the number of occurrence of at least one empty cell in each simulation in 1,000 simulations. The magnitude of goodness-of-fit statistics, the coefficients of determination or $R^2$ analogs, the likelihood ratio statistic, $G_M$, AIC (Akaike Information Criterion, [2]), and BIC (Bayesian Information Criterion, Schwarz, 1978) are calculated. The simulations have been conducted for the multinomial logit models with $K=3$ response categories and two random explanatory variables $X_1$ and $X_2$ whose joint distribution of $(X_1, X_2)$ is assumed to be multinomial with probabilities $\pi_1, \pi_2, \pi_3$, and $\pi_4$, corresponding to $(X_1, X_2)$ values of (0, 0), (0, 1), (1, 0), (1, 1), respectively. Three sets of $\left(\pi_1, \pi_2, \pi_3, \pi_4\right)$ are studied to represent different distributional shapes, which were chosen to induce possibly strong effects such that $\beta_1 = \log 2$, $\beta_2 = \log 3$, and $\beta_{12} = 0.0 - 4.5$, namely $(X_1, X_2) \sim$ multinomial(0.10, 0.35, 0.45, 0.10), $(X_1, X_2) \sim$ multinomial (0.50, 0.30, 0.10, 0.10), and $(X_1, X_2) \sim$ multinomial (0.25, 0.25, 0.25, 0.25). Four sets of the
three ordered category distributing corresponding with the \((X_1, X_2)\) were again generated through the models under the proportions of \((p_1, p_2, p_3)\), namely
\[Y \sim \text{multinomial}(p_1, p_2, p_3): (0.05, 0.20, 0.75), (0.25, 0.50, 0.25), (0.5, 0.20, 0.25), \text{ and } (0.33, 0.33, 0.33)\] from which it follows that the true model intercepts are
\[\alpha_1 = \log \frac{p_1}{p_2 + p_3}, \quad \alpha_2 = \log \frac{p_1 + p_2}{p_3}\]
corresponding to the proportions of \(Y = 1, 2, 3\) respectively. Four sample sizes of 600, 800, 1,000, and 1,500 units were performed. Each condition was carried out for 1,000 repeated simulations using the developed macro program run with the Minitab Release 11 [17].

The results indicate that the minimum sparseness of contingency tables and the maximum of goodness-of-fit statistics, \(R^2\) analogs and BIC, occur for the distribution of
\[Y \sim \text{multinomial} (0.05, 0.20, 0.75)\] with \((X_1, X_2) \sim \text{multinomial}(0.25, 0.25, 0.25, 0.25)\) as well as when each distribution of \(Y\) and \((X_1, X_2)\) is equally symmetric proportions. In contrast, the maximum sparse cells occur for the distributions of \(Y\sim \text{multinomial} (0.25, 0.50, 0.25)\) with \((X_1, X_2) \sim \text{multinomial} (0.50, 0.30, 0.10, 0.10)\). In addition, when \((X_1, X_2)\) is \((0.25, 0.25, 0.25, 0.25)\), it always gives less tendency of sparseness than those when \((X_1, X_2)\) are asymmetric, as the sample size become large. Moreover, the number of sparseness tends to increase as the interaction parameter, \(\beta_{12}\) increases; however, it is also relatively decreased when the sample sizes increase. Hence, for the true model with correlated structures are presented, the sparseness of the contingency tables increases as the interaction- parameter increases, and the rate of increasing will decrease as the sample sizes increase. These results indicate and confirm some association patterns in the models and the contingency tables. Therefore, when the distribution of \(Y\) is either equally symmetry or that’s in increasing ordered proportions, corresponding with those of \((X_1, X_2)\) are also symmetric, the moderate to small sample sizes are possible; however, when most distributions are asymmetric we do recommend only the large sample sizes for suitable analysis of the association and sparse contingency tables.

\[\text{Keyword: contingency table, goodness of fit, interaction effect, multinomial cumulative logit models, sparseness.}\]
1. Introduction

Traditionally, goodness of fit in contingency tables is tested by using either the Pearson $\chi^2$-statistic or the likelihood ratio $\chi^2$-statistic. The asymptotic properties of these statistics are studied on the assumption that the expected cell frequencies become large. Contingency tables with relatively few observations or having small or empty cell counts are referred to as sparse [19]. Sparse tables occur when the sample size $n$ is small. They also occur when $n$ is large but so is the number of cells. These empty cells are of two types: sampling zeros and structural zeros. For sampling zeros, cell counts $n_i$ will be greater than zero with sufficient large $n$ but for structural zeros, observations are impossible. A count of zero value is permissible outcome for a Poisson or multinomial variable [1]. For $(I \times J \times K)$ contingency tables, the nonstandard setting in which $K \rightarrow \infty$ as, the sample size, $n \rightarrow \infty$ is called sparse-data asymptotic. The asymptotic theory for likelihood-ratio and Wald tests require the number of parameters (and hence $K$) to be fixed. Ordinary ML estimation then breaks down because the number of parameters is not fixed, instead having the same order as the sample size. In particular, an approximate chi-squared distribution holds for the likelihood-ratio and Wald statistics for testing conditional independence only when the strata or grouped marginal totals generally exceed about 5 to 10 and $K$ is fixed and small relative to $n$. An alternative approach uses sparse asymptotic approximation that applies when the number of cells, $N$ increases as $n$ increases. For this approach, $\{\mu_i\}$ need not increase, as they must do in the usual (fixed $N$, $n \rightarrow \infty$ ) large-sample theory. Nonetheless, often some associations are not affected by empty cells and give stable results for the various analyses, whereas some others that are affected are unstable. Although empty cells and sparse tables need not affect parameter estimates of interest, they can cause sampling distribution goodness-of-fit statistics to be far from chi-squared [1]. Thus, to handle this problem in this paper we choose the most versatile $G^2(M_0 | M_1)$ statistic for testing the goodness-of-fit of models.

The model comparison statistic $G^2 (M_0 | M_1)$ often has an approximate chi-squared null distribution even when separate $G^2 (M_i)$ do not. For instance, when a predictor is continuous or a contingency table has very small fitted values, the sampling
distribution of $G^2(M_1)$ may be far from chi-squared. However, if the degrees of freedom for the comparison statistic is modest (as in comparing two models that differ by a few parameter), the null distribution of $G^2(M_0 | M_1)$ is approximately chi-squared [4]. The test statistic comparing two models is identical to the difference between $G^2(M_0) - G^2(M_1)$, goodness-of-fit statistics (deviances) for the two models. Then, 

$$G^2(M_0 | M_1) = -2(L_0 - L_1)$$

$$= -2(L_0 - L_s) - [-2(L_1 - L_s)]$$

$$= G^2(M_0) - G^2(M_1)$$

has the form of $-2$ (log likelihood ratio) for testing that $M_0$ holds against the alternative that $M_1$ holds. In addition, theory for likelihood–ratio tests suggests that when the simpler model holds, the asymptotic distribution of $G^2(M_0) - G^2(M_1)$ is chi-squared with the difference of degrees of freedom of the two models.

Moreover, these tests can perform well even for the large sparse tables, as long as the difference of degrees of freedom is small compared to the sample size [7]. The $G^2(M_0) - G^2(M_1)$ converges to its limiting chi-squared distribution more quickly than does $G^2(M_0)$, which depends also on individual cell counts.

In this research we present the analysis of data using the $G^2(M_0 | M_1)$ statistic to study the sparseness obtained from the number of occurrence of at least one empty cell in each simulation in 1,000 simulations and also to investigate the goodness-of-fit statistics for the contingency tables having some sparse cells under situation where sampling zeros are as a part of data set. The primary emphasis is on the statistical models of the multinomial cumulative logit models for the ordinal response categories and nominal explanatory variables including two–factor interaction term. As the associations between the variables in contingency table occur, some patterns of the cell counts are usually presented and are also probably leading to some sparseness of data, especially when the effect of X’s tend to be strong. The purpose is then to analyze the
performance of the above models for fixed \( N, n \rightarrow \infty \), and varied interaction parameter, from 0-4.5, increment 0.3 in terms of goodness-of-fit statistics and the occurrence of sparseness in 1,000 simulations. We aim to study how and when the sparseness occur; meanwhile, the parameter estimation and the goodness-of-fit of the considered models are expected to be working well under the chosen appropriately statistics.

2. The Cumulative Logit Models

The cumulative logit model was originally proposed by Walker and Duncan [22] and later called the proportional odds model by McCullagh [11]. The cumulative logits are defined [1] as

\[
P(Y \leq j \mid x) = \frac{p_1 + p_2 + \cdots + p_j}{1 - p_j}, \quad j = 1, \ldots, K.
\]

Then,

\[
\text{logit } [P(Y \leq j \mid x)] = \log \left( \frac{P(Y \leq j \mid x)}{1 - P(Y \leq j \mid x)} \right)
\]

\[
= \log \left( \frac{P(Y \leq j \mid x)}{P(Y > j \mid x)} \right)
\]

\[
= \log \left( \frac{p_1 + p_2 + \cdots + p_j}{p_{j+1} + p_2 + \cdots + p_K} \right), \quad j = 1, 2, \ldots, K-1.
\]

A model that simultaneously uses all cumulative logit is

\[
\text{logit } P(Y \leq j \mid x) = \alpha_j + x'\beta, \quad j = 1, \ldots, K-1.
\]

This model, which extends the logistic model for binary responses to allow for several ordinal responses, has often involved modeling cumulative logits, generalized cumulative logit models [5] and also those models often used in repeated measurement modeling [12,13]. Consider a multinomial response variable \( Y \) with categorical outcomes, denoted by 1, ..., \( K \) and let \( \mathbf{X}_i \) denote a \( p \)-dimensional vectors of explanatory variables or covariates. The dependence of the cumulative probabilities of \( Y \) on \( X \) for the proportional odds model is often of the form in (1).

\[
\log \left( \frac{P(Y \leq j \mid x)}{P(Y > j \mid x)} \right) = \alpha_j + x'\beta, \quad j = 1, \ldots, K-1.
\] .... (1)

It can be expressed in the form

\[
\log \left( \frac{p_1 + p_2 + \cdots + p_j}{p_{j+1} + p_2 + \cdots + p_K} \right) = \alpha_j + x'\beta, \quad j = 1, \ldots, K-1.
\]
Each cumulative logit has its own intercept. The \{ \alpha_j \} are increasing in \( j \), since \( P( Y \leq j \mid x ) \) increases in \( j \) for fixed \( x \), and the logit is an increasing function of this probability and each cumulative logit uses all \( K \) response categories.

Hence, for \( K=3 \), and \( j = 1, \ldots, K-1=2 \), the model (1) consists of two simultaneously cumulative link-functions for solving the model parameters in the following equations:

\[
\log \left[ \frac{p_{i1}}{p_{i2} + p_{i3}} \right] = \alpha_j + x'\beta, \quad \text{for } j = 1, \ldots, 2 \quad \text{(1.1)}
\]

\[
\log \left[ \frac{p_{i1} + p_{i2}}{p_{i3}} \right] = \alpha_j + x'\beta, \quad \text{for } j = 2 \quad \text{(1.2)}
\]

Where, \( \alpha_j \) are the intercept parameters,

\[
\beta = (\beta_1, \beta_2, \ldots, \beta_p)' \text{ is a vector of coefficients corresponding to } X's, \text{ and}
\]

\[
P( Y \leq j \mid x ) = p_{i1} + p_{i2} + \ldots \ldots + p_j, \text{ and } P( Y > j \mid x ) = p_{i1} + p_{i2} + \ldots \ldots + p_{iK}, \quad j = 1, \ldots, K-1.
\]

Similarly to (1), we have (2) and (3).

The proportional odds ratio model (minimal):

\[
\log \left[ \frac{P( Y \leq j \mid x )}{P( Y > j \mid x )} \right] = \alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2}, \quad j = 1, 2, K = 3, \quad i = 1, 2, \ldots, n. \quad \text{(2)}
\]

The proportional odds ratio with two-factor-interaction model (Interaction):

\[
\log \left[ \frac{P( Y \leq j \mid x )}{P( Y > j \mid x )} \right] = \alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2}, \quad j = 1, 2, K = 3, \quad i = 1, 2, \ldots, n. \quad \text{(3)}
\]

These models for any \( K \geq 3 \) are often called the proportional odds models [11]. It is based on the assumption that the effects of the explanatory variables \( X_1, \ldots, X_p \) are the same for all categories, on the logarithmic scale. It probably also represents the most widely used ordinal categorical model at the present time. The models (2) and (3) are extended to several \( X \)'s and corresponded to the main effect and interaction effect models, respectively.
3. Simulation and Statistical Analyses

From the models (2) and (3) in section 2, the simulations have been conducted for the multinomial logit models with K=3 response categories and two random explanatory variables $X_1$ and $X_2$ whose joint distribution of $(X_1, X_2)$ is assumed to be multinomial with probabilities $\pi_1, \pi_2, \pi_3$, and $\pi_4$, corresponding to $(X_1, X_2)$ values of (0, 0), (0, 1), (1, 0), (1, 1), respectively. Three sets of $(\pi_1, \pi_2, \pi_3, \pi_4)$ are studies to represent different distributional shapes, which were chosen to induce possibly strong effects such that $\beta_1 = \log 2$, $\beta_2 = \log 3$, and $\beta_{12} = 0.0 - 4.5$, namely $(X_1, X_2) \sim \text{multinomial}(0.10, 0.35, 0.45, 0.10)$, $(X_1, X_2) \sim \text{multinomial}(0.50, 0.30, 0.10, 0.10)$, and $(X_1, X_2) \sim \text{multinomial}(0.25, 0.25, 0.25, 0.25)$. Four sets of the three ordered category distributing corresponding with the $(X_1, X_2)$ were again generated through the models studies in the forms of (1.1)-(1.2), (2)-(3) under the proportions of $(p_1, p_2, p_3)$, namely $Y \sim \text{multinomial}(p_1, p_2, p_3)$: (0.05, 0.20, 0.75), (0.25, 0.50, 0.25), (0.5, 0.20, 0.25), and (0.33, 0.33, 0.33) from which it follows that the model parameters to be used in each condition are $\alpha_1 = \log \frac{p_1}{p_2 + p_3}$, $\alpha_2 = \log \frac{p_1 + p_2}{p_3}$, $\beta_1 = \log 2$, and $\beta_2 = \log 3$ for varied $\beta_{12}$ from 0-4.5 (increment 0.3), corresponding to the proportion of $Y = 1, 2, 3$ respectively. Consequently, the categorical response variable $Y_i$, $i = 1, \ldots, n$, of which the data are corresponded with $X$’s under the true models, will be random at each setting of fixed values of the explanatory variables $(X_1, X_2)$ through the cut points and the specified proportions. Four sample sizes were specified to vary from $n = 600, 800, 1,000$ and $1,500$ units. All results were performed for 816 (=4 x 3 x 4 x 17) conditions. Each of which for each model was carried out 1,000 replicates of data sets.

Statistical analyses in assessing goodness of fit of models consist of several statistics which were computed for each combination of the model conditions: the likelihood ratio statistics, the generalized coefficients of determination or $R^2$ analogs, AIC (Akaike Information Criterion, [2]), BIC (Bayesian Information Criterion, [21]) and the number of occurrence of sparseness in 1,000 sets are evaluated.
All the statistics were computed using the following formulae:

\[ G_M = -2 (L_O - L_M) \], the model chi-square statistic or the likelihood ratio statistic.

The Coefficients of determination, \( R^2 \) analogs:

\[ R^2_C = \frac{G_M}{(G_M + n)}, \quad \text{(The contingency coefficient } R^2, [3]) \]

\[ R^2_L = \frac{[L_O - L_M]}{L_O} = 1 - \left[ \frac{L_M}{L_O} \right], \quad \text{(The log likelihood ratio } R^2, [14-16]) \]

\[ R^2_M = 1 - \left[ \frac{L_O}{L_M} \right]^2, \quad \text{(The geometric mean squared improvement per observation } R^2, [6,10,20]) \]

\[ R^2_N = \frac{\left[ 1 - \left( \frac{L_O}{L_M} \right)^2 \right]}{\left[ 1 - \left( \frac{L_O}{L_M} \right)^2 \right]}, \quad \text{(The adjusted geometric mean squared improvement } R^2, [18,20]) \]

\[ \text{AIC} = G_M - 2 (\Delta df), \quad \text{BIC} = G_M - (\log(n))(\Delta df), [9]. \]

The sparseness is computed from the number of occurrence of at least one empty cell in each simulation in 1,000 simulations. Whereas,

\[ n = \text{sample size}, \]

\[ L_O = \text{the log likelihood function for the reduced model}, \]

\[ L_M = \text{the log likelihood function for the current model containing more parameters}, \]

\[ G_M = -2 (L_O - L_M) = \text{the model chi-square statistic = change in deviances}, \]

\[ \Delta df = \text{change in degrees of freedom between those of the null and alternative models}. \]

The computer simulation programs were developed using the MINITAB macro language and run by MINITAB release 11 on Pentiums IV [17].
4. Research Results

Several models for analyzing data with ordinal responses have been fitted and also are examined their goodness-of-fits and sparseness. The mean of each goodness-of-fit statistic (RN, BIC) and the sparseness statistic (spars) based on 1,000 simulations are summarized in Table 1 – Table 3. All statistics are classified by Y’s and (X₁, X₂)’s distributions, β₁₂, and each sample size (n). For β₁₂ = 0, it corresponds to the cumulative model with main effects or without interaction term, whereas, for β₁₂ ≠ 0, it does correspond to the model with two-factor interaction effect.

The results show that the magnitude of goodness-of-fit statistics, the coefficients of determination or $R^2$ analogs and the sparseness tend to increase as the sample sizes and the parameter $β₁₂$ increase (Table 1-3 in appendix 3). For those statistics, the likelihood ratio statistics and the BIC decrease as the sample sizes and $β₁₂$ are large. Thus, all statistics do confirm some association patterns in the contingency tables and vary dependently upon the distributions of Y and (X₁, X₂), which will be further study in more details. For $R^2$ analogs, results are all quite similar. We then report only the RN or the Nagelkerke’s $R^2$ analog and BIC statistics.

The number of sparseness in 1,000 simulations for varied $β₁₂$ among the three distributions of X₁ and X₂ with multinomial $(\pi₁, \pi₂, \pi₃, \pi₄)$: (X₁, X₂)~(0.10,0.35,0.45,0.10), (X₁, X₂)~(0.50,0.30,0.10,0.10), and (X₁, X₂)~(0.25,0.25,0.25,0.25) are compared for each sample size and for each distribution of Y~ multinomial (p₁, p₂, p₃). It is found that the sparseness of the contingency tables reach its minimum when (X₁, X₂) is symmetric (0.25,0.25,0.25,0.25) compared with those when (X₁, X₂) are asymmetric, (0.10,0.35,0.45,0.10) and (0.50,0.30,0.10,0.10), respectively. The latter two (X₁, X₂) proportions always give
### Table 1. Means of RN, BIC, and SPARSENESS classified by $\beta_{12}$, sample sizes, distributions of Y’s and $(X1,X2) \sim \text{multinomial}(0.10,0.35,0.45,0.10)$.

<table>
<thead>
<tr>
<th>Distributions</th>
<th>$Y \sim (0.05,0.20,\ldots,0.75)$</th>
<th>$Y \sim (0.25,0.50,0.25)$</th>
<th>$Y \sim (0.55,0.20,0.25)$</th>
<th>$Y \sim (0.33,0.33,0.33)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RN Mean</td>
<td>BIC Mean</td>
<td>SPARS Mean</td>
<td>RN Mean</td>
</tr>
<tr>
<td>Sample size 600</td>
<td>0.0</td>
<td>0.058679</td>
<td>1083.98</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.212123</td>
<td>1051.40</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>0.322947</td>
<td>992.86</td>
<td>810</td>
</tr>
<tr>
<td>Sample size 800</td>
<td>0.0</td>
<td>0.057409</td>
<td>1463.52</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.226641</td>
<td>1412.36</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>0.346772</td>
<td>1325.27</td>
<td>694</td>
</tr>
<tr>
<td>Sample size 1000</td>
<td>0.0</td>
<td>0.057858</td>
<td>1826.36</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.230354</td>
<td>1759.94</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>0.353092</td>
<td>1646.95</td>
<td>590</td>
</tr>
<tr>
<td>Sample size 1500</td>
<td>0.0</td>
<td>0.057337</td>
<td>2753.04</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.233535</td>
<td>2646.73</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>0.356469</td>
<td>2475.62</td>
<td>437</td>
</tr>
</tbody>
</table>
Table 2. Means of RN, BIC, and SPARSENESS classified by $\beta_{12}$, sample sizes, distributions of Y’s and (X1,X2) ~ multinomial (0.50,0.30,0.10,0.10).

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Y~(0.05,0.20, .75)</th>
<th>Y~(0.25, 0.5, 0.25)</th>
<th>Y~(0.55, 0.20, 0.25)</th>
<th>Y~(0.33, 0.33, 0.33)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>RN Mean</td>
<td>BIC Mean</td>
<td>SPARS Mean</td>
<td>RN Mean</td>
</tr>
<tr>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.092791</td>
<td>1002.20</td>
<td>0</td>
<td>0.096472</td>
</tr>
<tr>
<td>2.1</td>
<td>0.234714</td>
<td>972.45</td>
<td>87</td>
<td>0.172185</td>
</tr>
<tr>
<td>4.5</td>
<td>0.330918</td>
<td>924.25</td>
<td>862</td>
<td>0.199730</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.095893</td>
<td>1346.15</td>
<td>0</td>
<td>0.098667</td>
</tr>
<tr>
<td>2.1</td>
<td>0.265161</td>
<td>1295.56</td>
<td>11</td>
<td>0.193330</td>
</tr>
<tr>
<td>4.5</td>
<td>0.376119</td>
<td>1215.86</td>
<td>695</td>
<td>0.226971</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.117160</td>
<td>1672.38</td>
<td>0</td>
<td>0.119330</td>
</tr>
<tr>
<td>2.1</td>
<td>0.311074</td>
<td>1597.65</td>
<td>2</td>
<td>0.232552</td>
</tr>
<tr>
<td>4.5</td>
<td>0.437604</td>
<td>1467.59</td>
<td>547</td>
<td>0.273434</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.094653</td>
<td>2517.92</td>
<td>0</td>
<td>0.103104</td>
</tr>
<tr>
<td>2.1</td>
<td>0.257192</td>
<td>2434.43</td>
<td>2</td>
<td>0.202582</td>
</tr>
<tr>
<td>4.5</td>
<td>0.366935</td>
<td>2286.15</td>
<td>523</td>
<td>0.238673</td>
</tr>
</tbody>
</table>
Table 3. Means of RN, BIC, and SPARSENESS classified by $\beta_{12}$, sample sizes, distributions of Y’s and (X1,X2) ~ multinomial (0.25,0.25,0.25,0.25).

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Sample size 600</th>
<th>Sample size 800</th>
<th>Sample size 1000</th>
<th>Sample size 1500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>Y~(0.05,0.20, .75)</td>
<td>Y~(0.25, 0.5, 0.25)</td>
<td>Y~(0.55, 0.20, 0.25)</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>RN Mean</td>
<td>BIC Mean</td>
<td>SPARS Mean</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.116262</td>
<td>1073.23</td>
<td>1</td>
</tr>
<tr>
<td>2.1</td>
<td>0.0</td>
<td>0.124660</td>
<td>1430.32</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.1</td>
<td>0.408356</td>
<td>978.37</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>0.584123</td>
<td>817.59</td>
<td>460</td>
</tr>
<tr>
<td>0.0</td>
<td>2.1</td>
<td>0.118671</td>
<td>1872.84</td>
<td>0</td>
</tr>
<tr>
<td>4.5</td>
<td>0.305482</td>
<td>0</td>
<td>1626.84</td>
<td>0</td>
</tr>
<tr>
<td>0.0</td>
<td>2.1</td>
<td>0.375111</td>
<td>1532.73</td>
<td>903</td>
</tr>
<tr>
<td>4.5</td>
<td>0.599675</td>
<td>0</td>
<td>2024.14</td>
<td>121</td>
</tr>
<tr>
<td>2.1</td>
<td>0.321496</td>
<td>0</td>
<td>2415.86</td>
<td>0</td>
</tr>
<tr>
<td>4.5</td>
<td>0.391742</td>
<td>0</td>
<td>2267.54</td>
<td>0</td>
</tr>
<tr>
<td>0.0</td>
<td>2.1</td>
<td>0.122105</td>
<td>2814.05</td>
<td>0</td>
</tr>
<tr>
<td>4.5</td>
<td>0.321496</td>
<td>0</td>
<td>2415.86</td>
<td>0</td>
</tr>
<tr>
<td>0.0</td>
<td>2.1</td>
<td>0.391742</td>
<td>2267.54</td>
<td>0</td>
</tr>
<tr>
<td>4.5</td>
<td>0.328945</td>
<td>0</td>
<td>2327.39</td>
<td>824</td>
</tr>
</tbody>
</table>
closely results which are higher than that the former do for $\beta_{12} > 2$ (Figure 1). These results occur similarly for every sample size and every distribution of $Y$’s proportion of response categories. Beside this, the number of sparseness tends to increases as the interaction parameter, $\beta_{12}$, increases, however, it decreases when the sample sizes increase (Figure 2-4).

For Figure 1-4, use symbols below.

- (X₁, X₂)~multinomial(0.10, 0.35, 0.45, 0.10)
- (X₁, X₂)~multinomial(0.50, 0.30, 0.10, 0.10)
- (X₁, X₂)~multinomial(0.25, 0.25, 0.25, 0.25)

For Figure 5-7, use symbols below.

- Y~multinomial(0.05, 0.20, 0.75)
- Y~multinomial(0.25, 0.50, 0.25)
- Y~multinomial(0.55, 0.20, 0.25)
- Y~multinomial(0.33, 0.33, 0.33)

Figure 1. Sparseness plots versus 3 distributions of (X₁, X₂) under Y~ multinomial (0.05,0.20,0.75) and n = 600, 800, 1000 and 1500.
Figure 2. Sparseness plots versus 3 distributions of \((X_1, X_2)\) under \(Y \sim \text{multinomial (0.25,0.50,0.25)}\) and \(n = 600, 800, 1000\) and 1500.

Figure 3. Sparseness plots versus 3 distributions of \((X_1, X_2)\) under \(Y \sim \text{multinomial (0.55,0.20,0.25)}\) and \(n = 600, 800, 1000\) and 1500.
Figure 4. Sparseness plots versus 3 distributions of \((X_1, X_2)\) under \(Y \sim\) multinomial \((0.33,0.33,0.33)\) and \(n = 600, 800, 1000\) and 1500.

Figure 5. Sparseness for 4 distributions of \(Y\) under \((X_1, X_2) \sim\) multinomial \((0.10,0.35,0.45,0.10)\) and \(n = 600, 800, 1000\) and 1500.
Figure 6. Sparseness for 4 distributions of Y under (X₁,X₂) ~ multinomial (0.50,0.30,0.10,0.10) and n = 600, 800, 1000 and 1500.

Figure 7. Sparseness for 4 distributions of Y under (X₁,X₂) ~ multinomial (0.25,0.25,0.25,0.25) and n = 600, 800, 1000 and 1500.
The comparison of the sparseness in contingency tables among the four types of the proportion of response categories of $Y\sim \text{multinomial} \ (p_1, p_2, p_3)$: $Y\sim (0.05,0.20,0.75)$, $Y\sim (0.55, 0.20, 0.25)$, $Y\sim (0.25,0.50,0.25)$, and $Y\sim (0.33,0.33,0.33)$ are also compared for each sample size and for each distribution of $(X_1,X_2) \sim \text{multinomial} \ (\pi_1,\pi_2,\pi_3,\pi_4)$.

The results show that the number of sparseness from $Y\sim (0.05,0.20,0.75)$ always give the minimum sparseness compared with those when $Y\sim (0.33,0.33,0.33)$ and $Y\sim (0.25,0.50,0.25)$ for each sample size, respectively (Figure 5). These results are also found similarly for other sample sizes and distributions of $(X_1,X_2)$ (Figure 6-7).

5. Conclusion and Discussions

In conclusion, all results indicate that for the models with ordinal response categories and their corresponding nominal explanatory variables with two-factor interaction, the minimum sparseness of contingency tables occurs under the two distributions of $Y\sim \text{multinomial} \ (0.05,0.20,0.75)$ and $(X_1,X_2) \sim \text{multinomial} \ (0.25,0.25,0.25,0.25)$ as well as when each distribution of $Y$ and $(X_1,X_2)$ is equally symmetric proportions. In contrast, the maximum sparse cells occurs for $Y\sim \text{multinomial} \ (0.25,0.50,0.25)$ and $(X_1,X_2) \sim \text{multinomial} \ (0.50,0.30,0.10,0.10)$. In addition, when $(X_1,X_2)$ is equally symmetric $(0.25,0.25,0.25,0.25)$, it always gives less tendency of sparseness than those when $(X_1,X_2)$’s are asymmetric. Moreover, the number of sparseness does increase as the interaction parameter, $\beta_{12}$ increases; however, it is relatively decreased when the sample sizes increase. All goodness-of-fit statistics and sparseness when the sample sizes are large are also consistent. For the true model with correlated structures among the explanatory variables are presented, the sparseness of the contingency tables increase as the interaction parameter increase and the rate of increasing will decrease as the sample size increase. Thus, these results will confirm the correlated structures and indicate some association patterns in the contingency tables among variables. Therefore, in practice, we probably either still be able to use large sample sizes or try to develop more appropriated goodness-of-fit statistics for assessing the model fit to sparse contingency tables. Since many maximum likelihood analyses are unharmed by empty cells. Even when a parameter estimate is infinite, this is not fatal to data analysis and the likelihood ratio confident interval for the true log odds ratio can has
one endpoint that is infinite, such as \((-\infty, U]\) and \((L, \infty]\) for some finite upper and lower bound, respectively.

For the usual sort of contingency table models, a danger with sparse data is that one might not realize that a true estimated effect is infinite and, as a consequence, report estimated effects and results of inferences that are invalid and highly unstable [1]. Also when the pattern of empty cells forces certain fitted values for a model to equal 0, this affect the degrees of freedom for testing model fit [8]. Nonetheless, most existence of estimates in loglinear and logit models is identical for multinomial and independence Poisson sampling in contingency tables. For unsaturated models, suppose that at least one cell is zero but sufficient marginal counts are all positive, the estimates still exist, except when any count is zero in the set of sufficient marginal tables. The empty cells and sparse tables can also cause problems. However, they need not always be problematic.

For the interaction model in this article, we used the likelihood ratio statistic for assessing the model fit and the likelihood can still be maximized. We illustrate that the more highly correlated structure the model is, the sparseness is also possibly increase as well as probably is indicating a pattern in contingency table. A point estimate of \(\infty\) for an effect still usually has a finite lower bound for a likelihood-based confidence interval, and one can use even some small-sample inferential methods rather than asymptotic procedures. In addition, one way to obtain finite estimates of all effects and ensure-convergence of fitting algorithms is usually to adding a small constant to empty cell counts.

Hence, the summary results from our research work with the related points of view it probably is concluded obviously that when the distribution of \(Y\) is not only equally symmetric proportions with multinomial \((p_1, p_2, p_3)\) but also that is in increasing ordered proportions, corresponding with \(X_1, X_2\) are also symmetric distributing such that with multinomial \((\pi_1, \pi_2, \pi_3, \pi_4)\), then the moderate to small sample sizes are possible; however, when those distributions are all in asymmetric patterns we do confirm and recommend only the large sample sizes for the suitable analysis of the association and sparse contingency tables.
5. Acknowledgements

This research was partially supported by the Institute of Research and Development of Silpakorn University. We would also like to thank the Editor-in-Chief, the Executive Editor and all the Referees for their helpful comments.

References