Thailand Statistician
January 2008; 6(1) : 27-46
http://statassoc.or.th
Contributed paper

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 π_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 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)~multinomial(0.10,0.35,0.45,0.10), (X_1, X_2)~ multinomial (0.50,0.30,0.10,0.10), and (X_1, X_2)~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~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 true model intercepts are $\alpha_1 = \log\frac{p_1}{p_2 + p_3}$, $\alpha_2 = \log\frac{p_1 + p_2}{p_3}$, 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~multinomial (0.05,0.20,0.75) with (X_1, X_2)~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~ multinomial (0.25,0.50,0.25) with (X_1, X_2)~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, β_{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.

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 χ^2-statistic or the likelihood ratio χ^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 $K \longrightarrow \infty$ as, the sample size, $n \longrightarrow \infty$ 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, $n \longrightarrow \infty$) 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 $G^2(M_0 \mid M_1)$ statistic for testing the goodness-of-fit of models.

The model comparison statistic $G^2(M_0 \mid M_1)$ 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 $G^2(M_I)$ 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 $G^2(M_0 \mid M_1)$ is approximately chi-squared [4]. The test statistic comparing two models is identical to the difference between $G^2(M_0)$ - $G^2(M_I)$, goodness-of-fit statistics (deviances) for the two models. Then,

$$\begin{aligned} G^2(M_0 \mid M_1) &= -2\left(L_0 - L_1\right) \\ &= -2\left(L_0 - L_s\right) - \left[-2\left(L_1 - L_s\right)\right] \\ &= G^2(M_0) - G^2(M_I) \end{aligned}$$

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 $G^2(M_0) - G^2(M_I)$ 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 $G^2(M_0 \mid M_1)$ 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, $n \to \infty$, 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

$P(Y \le j \mid x) = p_1 + p_2 + \ldots\ldots + p_j$, $j = 1,\ldots, K$. Then,

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

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

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

A model that simultaneously uses all cumulative logit is

$$\text{logit } P(Y \le j \mid x) = \alpha_j + \mathbf{x}'\boldsymbol{\beta}, \quad j = 1,\ldots, K-1.$$

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).

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

It can be expressed in the form

$$\log\left[\frac{p_1 + p_2 + \ldots\ldots + p_j}{p_{j+1} + p_2 + \ldots\ldots + p_K}\right] = \alpha_j + \mathbf{x}'\boldsymbol{\beta}, \quad j = 1,\ldots, K-1.$$

Each cumulative logit has its own intercept. The $\{\alpha_j\}$ are increasing in j, since P($Y \le j \mid$ 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,...,K-1=2, the model (1) consists of two simultaneously cumulative link-functions for solving the model parameters in the following equations:

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

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

Where, α_j are the intercept parameters,

$\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ is a vector of coefficients corresponding to X's, and

P(Y $\le j$ | x) = $p_1 + p_2 + \ldots\ldots + p_j$, and P(Y $> j$ | x) = $p_{j+1} + p_2 + \ldots\ldots + p_K$, j =1, ..., K-1.

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

The proportional odds ratio model (**minimal**):

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

$$= \alpha_j + \beta_1 x_{1i} + \beta_2 x_{2i}, j = 1,2, K = 3, \quad i = 1,2,...n. \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

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

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

$$= \alpha_j + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}, j = 1,2, K = 3, \quad i = 1,2,...n. \ldots.(3)$$

These models for any K ≥ 3 are often called the proportional odds models [11]. It is based on the assumption that the effects of the explanatory variables $X_1, ..., X_p$ 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 π_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)~multinomial(0.10,0.35,0.45,0.10), (X_1, X_2)~ multinomial (0.50,0.30,0.10,0.10), and (X_1, X_2)~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~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 β_{12} from 0-4.5 (increment 0.3), corresponding to the proportion of Y = 1, 2, 3 respectively. Consequently, the categorical response variable $Y_i, i = 1, .., n,$ 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:

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

$R^2_L = \frac{[L_O - L_M]}{L_O} = 1 - \left[\frac{L_M}{L_O}\right]$,

(The log likelihood ratio R^2, [14-16])

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

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

AIC = G_M - 2 (Δdf), BIC = G_M- (log(n))(Δdf), [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 = sample size,

L_O = the log likelihood function for the reduced model,

L_M = the log likelihood function for the current model containing more parameters,

G_M = -2 (L_O - L_M) = the model chi-square statistic = change in deviances,

Δdf = 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 (X_1, X_2)'s distributions, β_{12} , and each sample size (n). For β_{12} = 0, it corresponds to the cumulative model with main effects or without interaction term, whereas, for $\beta_{12} \neq 0$, 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 β_{12} increase (Table 1-3 in appendix 3). For those statistics, the likelihood ratio statistics and the BIC decrease as the sample sizes and β_{12} are large. Thus, all statistics do confirm some association patterns in the contingency tables and vary dependently upon the distributions of Y and (X_1, X_2), 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 β_{12} among the three distributions of X_1 and X_2 with multinomial $(\pi_1, \pi_2, \pi_3, \pi_4)$: (X_1, X_2)~(0.10,0.35,0.45,0.10), (X_1, X_2)~ (0.50,0.30,0.10,0.10), and (X_1, X_2)~(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 (X_1, X_2) is symmetric (0.25,0.25,0.25,0.25) compared with those when (X_1, X_2) are asymmetric, (0.10,0.35,0.45,0.10) and (0.50,0.30,0.10,0.10), respectively. The latter two (X_1, X_2) proportions always give

Table 1. Means of RN, BIC, and SPARSENESS classified by β_{12}, sample sizes, distributions of Y's and (X1,X2) ~ multinomial (0.10,0.35,0.45,0.10).

Distributions		Y~(0.05,0.20, .75)			Y~(0.25, 0.50, 0.25)			Y~(0.55, 0.20, 0.25)			Y~(0.33, 0.33, 0.33)		
n	β_{12}	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean
Sample size 600	0.0	0.058679	1083.98	48	0.060107	1142.01	68	0.052387	876.30	75	0.063811	1141.69	3
	2.1	0.212123	1051.40	66	0.142682	1086.00	736	0.093214	837.13	858	0.155567	1061.33	555
	4.5	0.322947	992.86	810	0.175407	1067.04	990	0.103920	826.22	999	0.185940	1037.27	988
Sample size 800	0.0	0.057409	1463.52	31	0.058316	1527.46	34	0.052828	1172.94	12	0.058012	1528.82	3
	2.1	0.226641	1412.36	45	0.140179	1455.15	633	0.097928	1115.07	763	0.136609	1438.84	473
	4.5	0.346772	1325.27	694	0.172824	1426.87	987	0.111255	1104.13	997	0.164936	1412.32	983
Sample size 1000	0.0	0.057858	1826.36	11	0.059951	1905.22	3	0.049243	1458.03	3	0.053304	1938.15	3
	2.1	0.230354	1759.94	12	0.159612	1793.69	467	0.093715	1384.13	676	0.111015	1856.88	676
	4.5	0.353092	1646.95	590	0.199785	1748.35	980	0.106758	1368.58	992	0.131566	1834.23	992
Sample size 1500	0.0	0.057337	2753.04	0	0.058683	2869.97	0	0.048043	2191.54	1	0.060091	2880.40	0
	2.1	0.233535	2646.73	0	0.156734	2704.91	357	0.089488	2088.94	545	0.147409	2693.48	158
	4.5	0.356469	2475.62	437	0.194259	2642.49	963	0.103027	2067.81	992	0.178252	2636.77	945

Table 2. Means of RN, BIC, and SPARSENESS classified by β_{12}, sample sizes, distributions of Y's and (X1,X2) ~ multinomial (0.50,0.30,0.10,0.10).

Distributions		Y~(0.05,0.20, .75)			Y~(0.25, 0.5, 0.25)			Y~(0.55, 0.20, 0.25)			Y~(0.33, 0.33, 0.33)		
n	β_{12}	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean
Sample size 600	0.0	0.092791	1002.20	0	0.096472	1164.43	105	0.081008	971.28	152	0.092754	1199.72	50
	2.1	0.234714	972.45	87	0.172185	1117.85	778	0.116830	939.52	927	0.147693	1158.46	789
	4.5	0.330918	924.25	862	0.199730	1101.15	998	0.124169	933.49	999	0.164782	1146.38	988
Sample size 800	0.0	0.095893	1346.15	0	0.098667	1553.21	23	0.094352	1296.73	18	0.110945	1590.57	2
	2.1	0.265161	1295.56	11	0.193330	1474.51	640	0.140867	1242.37	758	0.186667	1511.54	566
	4.5	0.376119	1215.86	695	0.226971	1443.93	985	0.154078	1227.56	996	0.211336	1487.02	989
Sample size 1000	0.0	0.117160	1672.38	0	0.119330	1938.14	1	0.092974	1644.05	8	0.113630	1984.51	1
	2.1	0.311074	1597.65	2	0.232552	1818.12	457	0.138355	1576.73	682	0.196920	1874.05	351
	4.5	0.437604	1467.59	547	0.273434	1772.13	977	0.151304	1561.95	995	0.224653	1839.20	982
Sample size 1500	0.0	0.094653	2517.92	0	0.103104	2933.07	0	0.078734	2452.18	3	0.095011	3011.83	0
	2.1	0.257192	2434.43	2	0.202582	2776.60	373	0.115188	2368.67	652	0.159471	2888.93	327
	4.5	0.366935	2286.15	523	0.238673	2716.21	956	0.125701	2349.83	993	0.180292	2851.71	979

Table3. Means of RN, BIC, and SPARSENESS classified by β_{12}, sample sizes, distributions of Y's and (X1,X2) ~ multinomial (0.25,0.25,0.25,0.25).

Distributions		Y~(0.05,0.20, .75)			Y~(0.25, 0.5, 0.25)			Y~(0.55, 0.20, 0.25)			Y~(0.33, 0.33, 0.33)		
n	β_{12}	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean	RN Mean	BIC Mean	SPARS Mean
Sample size 600	0.0	0.116262	1073.23	1	0.119802	1115.17	1	0.097560	871.77	1	0.119847	1118.68	0
	2.1	0.408356	978.37	0	0.311012	963.17	374	0.189795	776.29	609	0.282095	961.67	247
	4.5	0.584123	817.59	460	0.380440	906.53	958	0.216386	755.09	990	0.335963	914.03	955
Sample size 800	0.0	0.124660	1430.32	0	0.117021	1495.08	0	0.099318	1167.42	0	0.120732	1504.52	0
	2.1	0.429295	1292.72	0	0.311231	1287.55	268	0.193366	1038.69	423	0.276914	1307.51	134
	4.5	0.608017	1059.36	318	0.382902	1209.89	937	0.220855	1011.17	985	0.329991	1243.65	941
Sample size 1000	0.0	0.113219	1810.38	0	0.118671	1872.84	0	0.102035	1438.73	0	0.122004	1875.77	0
	2.1	0.414791	1642.97	0	0.305482	1626.84	186	0.204218	1260.43	222	0.288977	1603.70	83
	4.5	0.594478	1359.05	249	0.375111	1532.73	903	0.237176	1216.43	965	0.345035	1520.09	877
Sample size 1500	0.0	0.118020	2703.23	0	0.122105	2814.05	0	0.096457	2187.08	0	0.115583	2823.25	0
	2.1	0.420350	2452.16	0	0.321496	2415.86	78	0.194758	1933.23	117	0.274387	2445.30	24
	4.5	0.599675	2024.14	121	0.391742	2267.54	830	0.223310	1873.44	955	0.328945	2327.39	824

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, β_{12}, increases, however, it decreases when the sample sizes increase (Figure 2-4).

For Figure 1-4, use symbols below.

(X_1, X_2)~multiomial(0.10, 0.35, 0.45, 0.10)

(X_1, X_2)~multiomial(0.50, 0.30, 0.10, 0.10)

(X_1, X_2)~multiomial(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)

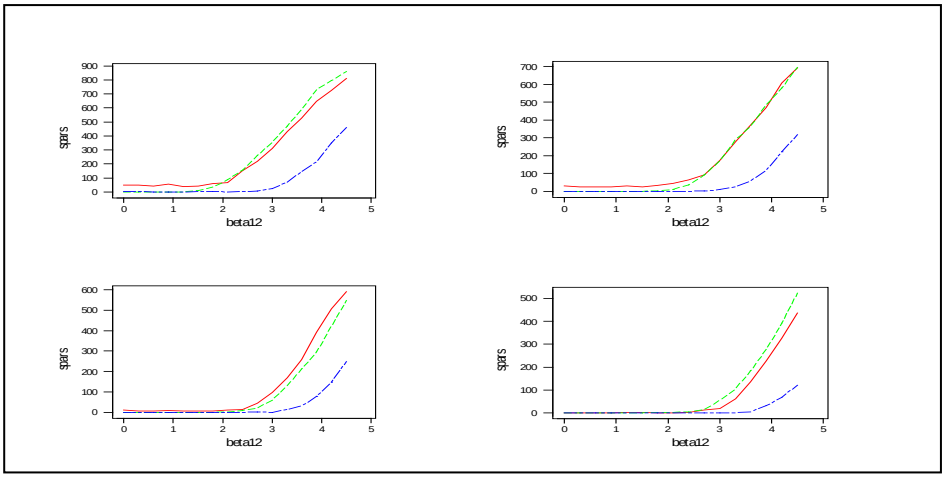

Figure 1. Sparseness plots versus 3 distributions of (X_1, X_2) under Y~ multinomial (0.05,0.20,0.75) and n = 600, 800, 1000 and 1500.

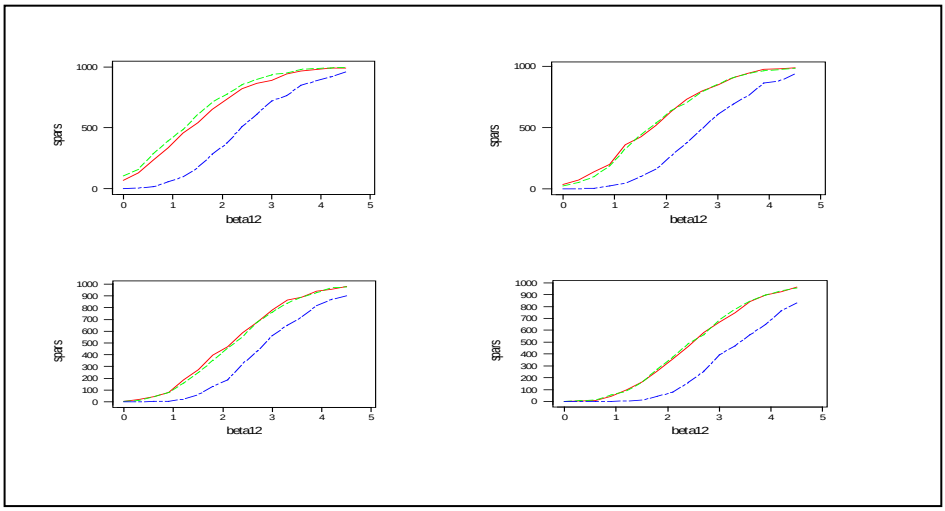

Figure 2. Sparseness plots versus 3 distributions of (X_1, X_2) under Y ~ multinomial (0.25,0.50,0.25) and n = 600, 800, 1000 and 1500.

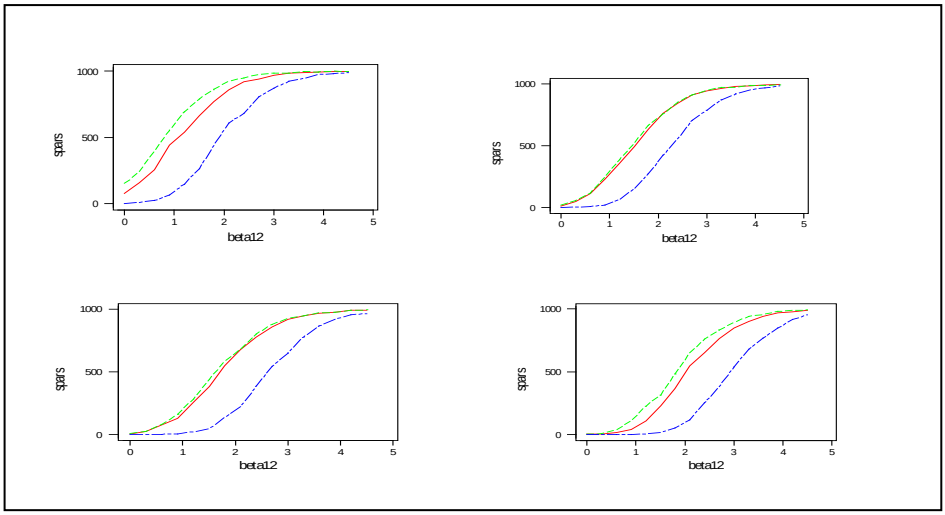

Figure 3. Sparseness plots versus 3 distributions of (X_1, X_2) under Y~ multinomial (0.55,0.20,0.25) and n = 600, 800, 1000 and 1500.

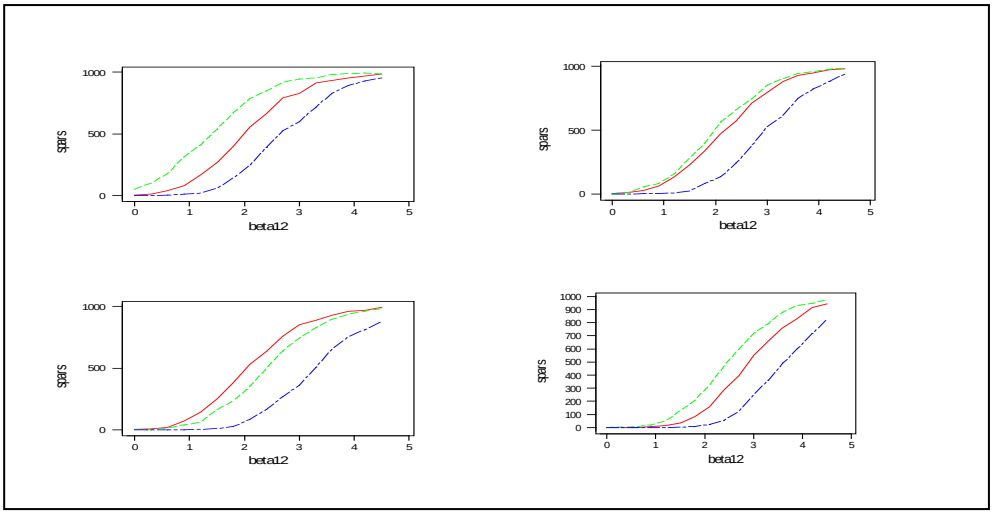

Figure 4. Sparseness plots versus 3 distributions of (X_1, X_2) under Y~ multinomial (0.33,0.33,0.33) and n = 600, 800, 1000 and 1500.

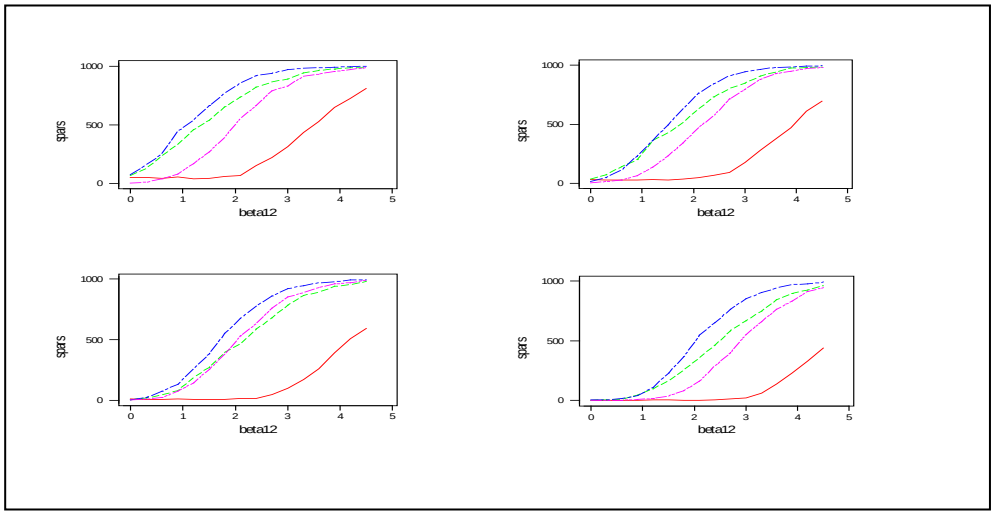

Figure 5. Sparseness for 4 distributions of Y under (X_1,X_2) ~ multinomial (0.10,0.35,0.45,0.10) and n = 600, 800, 1000 and 1500.

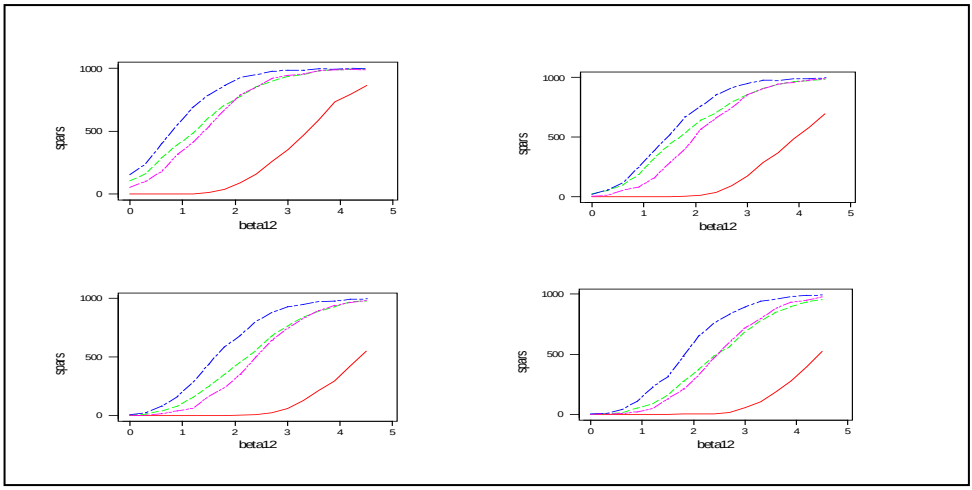

Figure 6. Sparseness for 4 distributions of Y under (X_1,X_2) ~ multinomial (0.50,0.30,0.10,0.10) and n = 600, 800, 1000 and 1500.

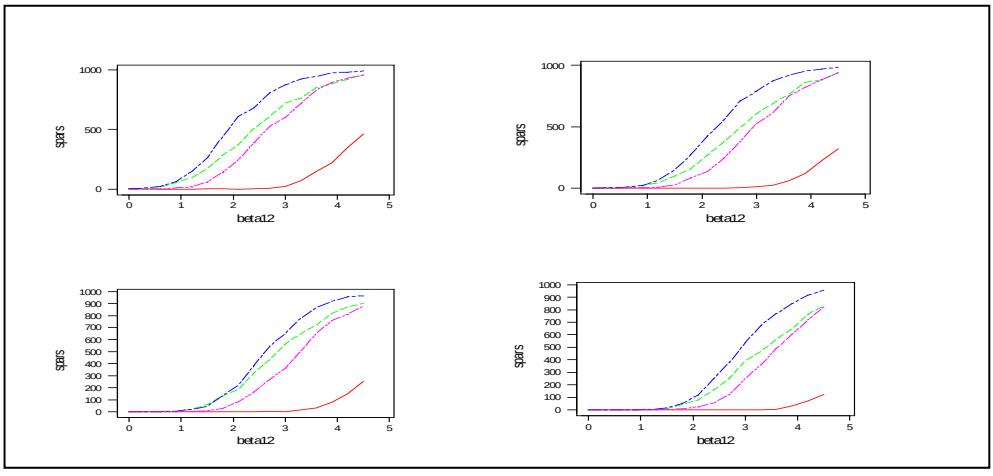

Figure 7. Sparseness for 4 distributions of Y under (X_1,X_2) ~ multinomial (0.25,0.25,0.25,0.25) and n = 600, 800, 1000 and 1500.

The comparison of the sparseness in contingency tables among the four types of the proportion of response categories of Y~multinomial (p_1, p_2, p_3): Y~(0.05,0.20,0.75), Y~(0.55, 0.20, 0.25), Y~(0.25,0.50,0.25), and Y~(0.33,0.33,0.33) are also compared for each sample size and for each distribution of (X_1, X_2) ~ multinomial $(\pi_1, \pi_2, \pi_3, \pi_4)$. The results show that the number of sparseness from Y~(0.05,0.20,0.75) always give the minimum sparseness compared with those when Y~(0.33,0.33,0.33) and Y~(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~ multinomial (0.05,0.20,0.75) and (X1,X2) ~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~ multinomial (0.25,0.50,0.25) and (X_1, X_2) ~ 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, β_{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, ∞) 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 ∞ 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

[1] Agresti, A. Categorical Data Analysis second edition, Wiley, New York, 2002.

[2] Akaike, H. Information Theory and an extension of the maximum Likelihood principle, in 2nd Intenational Symposium on Information Theory (eds.B.N. Petrov and F. Czake). Akademiai, Kiado, Budapest,1973:267-81.

[3] Aldrich, J.H. and F.D. Nelson. Linear Probability Logit and Probit Models**.** Sage, Beverly Hills, 1984.

[4] Aitkin, M.D. Anderson, B. Francis, and J. Hinde. Statistical Modelling in GLIM. Oxford Science Publications, Oxford, 1989.

[5] Cole S.R., P.D. Allison, and C.V. Ananth . Estimation of Cumulative Odds Ratios. Copyright Elsevier Inc. 2003. AEP Vol. 14 No.3 March 2004: 172-178.

[6] Cox, D.R. and E.G., Snell. The Analysis of Binary Data, second edition, London. Chapman and Hall, 1989.

[7] Haberman, S.J. Log-linear Models and Frequency Tables with Small Expected Cell Counts. Ann. Statists. 5, 1977a.:1148-1169.

[8] Haslett, S. Degrees of Freedom and Parameter Estimability in Hierarchically Models for Sparse Complete Contingency Tables. Computat. Statist. Data Anal. 9, 1990: 179-195.

[9] Lawal, H.B. Categorical Data Analysis with SAS and SPSS Applications. Lawrence Erlbaum Associates. Inc., London, 2003.

[10] Maddala, G.S. Limited-Dependent and Qualitative Variables in Econometrics: Cambridge Uni.Press, 1983.

[11] McCullagh, D. Regression Models for Ordinal Data. Journal Royal Statist. Soc. Ser. B42, 1980:109-142.

[12] McCulloch,C. Generalized Linear Models. J. of the American Statistical Association 95 No. 452: 2000: 1320-1324.

[13] McCulloch, C. and S.R. Searle. Generalized, Linear, and Mixed Model. Wiley, New York, 2001.

[14] McFadden, D. The Measurement of Urban Travel Demand. Journal of Public Economics 3, 1974: 303-328.

[15] Menard, S. Applied Logistic Regression Analysis. A Sage University Papers series, CA: Sage, 1995.

[16] Menard, S. Coefficients of Determination for Multiple Logistic Regression Analysis. The American Statistician 54, 2000:17-24.

[17] Minitab Reference Manual and User's Guide Release 11 for Windows ™ Windows 3.1 or 3.11, Window NT, and Windows 95, Windows95, Minitab Inc. June, 1996.

[18] Nagelkerke, N.J.D. A Note on a General Definition of the Coefficient of Determination. Biometrika78,1991:691-692.

[19] Paul, S.R. and D. Deng. Goodness of Fit of Generalized Linear Models to Sparse Data. Royal Statistics Soc. B 62, 2000: 323-333.

[20] Ryan, T.P. Modern Regression Methods. New York : Wiley, 1997.

[21] Schwarz, G. Estimating the dimensions of a model. Annals of statistics, 6: 461-464, 1978.

[22] Walker, S.H. and D.B. Duncan. Estimation of Probability of an event as a function of several independent variables. Biometrika 54, 1967: 167-179.