Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 2, pp. 543 ~ 557

Empirical Comparisons of Disparity Measures for Three Dimensional Log-Linear Models

Y. S. Park[1] · C. S. Hong[2] · D. B. Jeong[3]

Abstract

This paper is concerned with the applicability of the chi-square approximation to the six disparity statistics: the Pearson chi-square, the generalized likelihood ratio, the power divergence, the blended weight chi-square, the blended weight Hellinger distance, and the negative exponential disparity statistic. Three dimensional contingency tables of small and moderate sample sizes are generated to be fitted to all possible hierarchical log-linear models: the completely independent model, the conditionally independent model, the partial association models, and the model with one variable independent of the other two. For models with direct solutions of expected cell counts, point estimates and confidence intervals of the 90 and 95 percentage points of six statistics are explored. For model without direct solutions, the empirical significant levels and the empirical powers of six statistics to test the significance of the three factor interaction are computed and compared.

Keywords : Blended weight chi-squared statistic, Blended weight Hellinger distance statistic, Disparity measure, Log-linear model, Negative exponential disparity measure, Power divergence statistic

1. Introduction

The hypotheses testing for appropriate models fitted to categorical data is one of the most important applications of the chi-square distribution. There are two well-known goodness-of-fit test statistics: the Pearson chi-square X^2 and the

1) Department of Statistics, Sungkyunkwan University, Seoul, 110-745, KOREA.
E-mail : park9457@daum.net

2) Corresponding Author : Professor, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, KOREA. E-mail:cshong@skku.ac.kr

3) Associate Professor, Department of Information Statistics, Kangnung National University, Kangnung, 210-702, KOREA. E-mail: dj@kangnung.ac.kr

generalized likelihood ratio statistic G^2. Many authors have compared these two test statistics as well as some other less used goodness-of-fit test statistics like the Neyman modified chi-square, the Freeman-Tukey and the modified log likelihood ratio statistics. See, for example, Cochran (1952), Hoeffding (1965), West and Kempthorne (1972), Chapman (1976), Lantz (1978), Koehler and Lantz (1980) and Moore and Spruill (1975). Cressie and Read (1984) and Read and Cressie (1988) developed a class of goodness-of-fit test statistics called the family of the power divergence statistics denoted by $\{I(\lambda);\lambda\in R\}$. Cressie and Read (1984) showed that the statistic $I(2/3)$ is an excellent and compromising alternative to any other goodness-of-fit test statistics based on a multinomial distribution.

An even more general class of goodness-of-fit test statistics of which the family of the power divergence statistics is a subclass was studied. These test statistics, which are called the disparity test statistics, have been derived following the minimum disparity estimation approach of Lindsay (1993). Basu and Sarkar (1994) and Lindsay (1993) introduced two other subfamilies of disparity tests, which are named as the blended weight chi-squared family $BWCS(\lambda)$ and the blended weight Hellinger distance family $BWHD(\lambda)$. Basu and Sarkar (1994) derived the asymptotic chi-square distribution of the disparity tests, and showed that the blended weight Hellinger distance family, like the power divergence statistics, is excellent compromise to any other goodness-of-fit statistics when $BWHD(1/9)$ as well as $BWCS(1/3)$ under a multinomial distribution. Another subfamily of disparity tests is investigated by Jeong and Sarkar (2000). This statistic is the negative exponential disparity test $NED(\lambda)$, whose family includes the Pearson's chi-square as a member. When $\lambda=4/3$, the negative exponential disparity test might be preferred to the power divergence statistic $I(2/3)$ based on a multinomial model.

These disparity test statistics including the power divergence statistics $I(2/3)$, the blended weight chi-squared statistics $BWCS(1/3)$, the blended weight Hellinger distance statistics $BWHD(1/9)$ and the negative exponential disparity statistics $NED(4/3)$ are introduced in Section 2 except the well-known X^2 and G^2 statistics.

Several rules have been published to deal with the situation where the chi-square approximation can be applied to the distribution of these statistics for a given categorical data, model and sample size. Most rules are formulated in terms of the minimum cell expectation (MCE). For the Pearson chi-square statistic, Fisher (1941) suggested the rule that the MCE should be at least 5. Cramer (1946) requires MCE=10 and Kendall(1948) needs MCE=20. Cochran (1954) gave the following rule: if the number of those cells where the expected frequency is smaller than 5 does not exceed one fifth of the total number of cells, then the MCE may be as small as 1, i.e., if $t_5 < t/5$, then MCE=1, where t_5 is the number of those cells where the expected frequency is smaller than 5 and t is the number of the cells of the table. The most liberal rules are due to Yarnold (1970) and Fienberg (1979), based on the results of Larntz (1978). Yarnold (1970) suggested that the number of cells is greater than 3, then the MCE might be as small as $5t_5/t$. Fienberg (1979) suggested that the sample size might be as small as 4 or 5 times t.

There has been a long debate on the application of the chi-square approximation to statistics based on small sample sizes. Most of the works done in this area have been concerned with the Pearson chi-square and the generalized likelihood ratio statistics. Rudas (1986) and Hosmane (1987) obtained simulation results on the small and moderate sample behaviour of these statistics adding the power divergence statistic $I(2/3)$.

Rudas (1986) considered the completely independent model for two dimensional tables and the conditionally independent model for three dimensional tables. Hosmane (1987) explored the partial association model for three dimensional tables to test the interaction. Whereas the partial association model does not have direct solutions of expected cell counts, the others among three dimensional models have direct solutions. Hence in this paper, all possible three dimensional hierarchical log-linear models are considered: the completely independent model ([A][B][C] model), the conditionally independent models ([AB][AC], [AB][BC], [AC][BC] models), the models with one variable independent the other two ([AB][C], [AC][B], [A][BC] models) and the partial association model ([AB][BC][AC] model). We obtain and compare results of simulation studies on the small and moderate sample behaviour of $BWHD(1/9)$, $BWCS(1/3)$ and $NED(4/3)$ statistics as well as the Pearson chi-square X^2, the generalized likelihood ratio G^2 and the power divergence $I(2/3)$ statistics in order to explore the applicability of the chi-square approximation.

For models with direct solutions of expected cell counts, several contingency tables fitted to each model are generated by using marginal probabilities corresponding to sufficient configurations. Then point estimators and approximate 95% confidence intervals of 90 and 95% points for these six statistics are computed and discussed in Section 3. For the partial association model without direct solutions of expected cell counts, some categorical data are generated by using simulation on each term in the partial association log-linear model. The six statistics are computed for testing the three factor interaction, so that the empirical significant levels and powers are compared in Section 4.

2. Disparity Measure

Suppose the sample space is a countable set, without loss of generality $X = \{0,1,\cdots,K\}$, with K possibly infinite, and that $m_\beta(x)$ is a family of probability densities on X indexed by $\beta \in \Omega$. To avoid technicalities, it will be assumed that $m_\beta(x)>0$ for all $x \in X$. Moreover, suppose that n independent and identically distributed observations $X_1, \cdots, X_n$ are made from $m_\beta(x)$. Let $d(x)$ be the proportion of the n observations which has value x. Define the Pearson residual function $\delta(x)$ as

$$\delta = [d(x) - m_\beta(x)]/m_\beta(x). \tag{2.1}$$

Note that the model-weighted sum of the squared residuals, $\sum m_\beta(x_i)\delta(x_i)^2$, is the Pearson's chi-squared distance. And it is important to note that these residuals are not standardized to have identical variances, so that these residuals have range $[-1, \infty]$. Suppose that $G(\cdot)$ is a real-valued thrice-differentiable function on $[-1, \infty)$, with $G(0)=0$. Lindsay (1994) defined the disparity measure determined by G to be

$$\rho(d, m_\beta) = \sum m_\beta(x_i) G(\delta(x_i)). \tag{2.2}$$

If G is assumed to be strictly convex, then Csiszar (1963) showed that the disparity measure is nonnegative, and is zero only when $d = m_\beta$ by Jensen's inequality. An important class of such measures is the Cressie-Read family of the power divergence measures, defined by

$$\begin{aligned} I(\lambda) &= \sum d(x_i) \frac{\{[d(x_i)/m_\beta(x_i)]^\lambda - 1\}}{\lambda(\lambda+1)} \\ &= \sum m_\beta(x_i) \frac{\{(1+\delta(x_i))^{\lambda+1} - 1\}}{\lambda(\lambda+1)} \end{aligned}.$$

For $\lambda = -2, -1, -0.5, 0$, and 1, one obtains the well-known measures: Neyman chi-squared measure divided by 2, Kullback-Leibler divergence measure, the twice-squared Hellinger measure, the likelihood disparity, Pearson chi-squared measure divided by 2, respectively. Cressie and Read (1994) suggested that a member of this class, $I(2/3)$, could be used as a good alternative to any other measures. For any fixed number λ in $[0, 1]$ and $\bar{\lambda} = 1 - \lambda$, Lindsay (1993) introduced two modified distance measures. One is the blended weight chi-squared disparity which is defined as

$$BWCS(\lambda) = \sum \frac{[d(x_i) - m_\beta(x_i)]^2}{2[\lambda d(x_i) + \bar{\lambda} m_\beta(x_i)]}$$

and for this statistic $G(\delta) = 2^{-1}\delta^2/(\lambda\delta + 1)$. Here Pearson chi-squared measure corresponds to $\lambda = 0$ and Neyman's corresponds to $\lambda = 1$. Le Cam (1986, pp. 47) considered the case $\lambda=0.5$, showing that it is squared distance satisfying the triangle inequality. The other weighting scheme that generalizes Hellinger distance is the blended weight Hellinger distance measure such as

$$BWHD(\lambda)=\sum\frac{[d(x_i)-m_\beta(x_i)]^2}{2[\lambda\sqrt{d(x_i)}+\bar{\lambda}\sqrt{m_\beta(x_i)}]^2}\ ,$$

which corresponds to $G(\delta)=2^{-1}\{\delta/[\lambda(\delta+1)^{1/2}+(1-\lambda)]\}^2$. This family includes Neyman chi-squared measure, Pearson chi-squared measure, and Hellinger distance for $\lambda=1,\ 0,$ and 0.5, respectively. Jeong and Sarkar (2000) proposed the generalized negative exponential disparity family as the following:

$$NED(\lambda)=\sum\frac{\exp\left[-\lambda\left(\frac{d(x_i)}{m_\beta(x_i)}-1\right)\right]-1+\lambda\left(\frac{d(x_i)}{m_\beta(x_i)}-1\right)}{\lambda^2}m_\beta(x_i)\ ,$$

which corresponds to

$$G(\delta)=\begin{cases}(e^{-\lambda\delta}+\lambda\delta-1)/\lambda^2 & \text{if } \lambda\neq 0,\\ \delta^2/2 & \text{if } \lambda=0.\end{cases}$$

This family includes Pearson chi-squared measure and the negative exponential disparity introduced by Lindsay (1993) for $\lambda=0$ and 1, respectively. Basu and Sarkar (1994) showed that the blended weighted Hellinger distance statistic and the blended weighted chi-square statistic are excellent compromise to any other goodness-of-fit statistics when $BWHD(1/9)$ and $BWCS(1/3)$. Jeong and Sarkar (2000) derived that the negative exponential disparity statistic, $NED(\lambda)$, might be preferred to the power divergence statistics $I(2/3)$ when $\lambda=4/3$.

For a sequence of n observations on a multinomial distribution with probability vector $\boldsymbol{\pi}=(\pi_1,\cdots,\pi_k)$ and $\sum_{i=1}^k\pi_i=1$, let $\rho_G(\mathbf{d},\boldsymbol{\pi})$ be a disparity measure defined in (2.2). Then consider $D_{\rho G}=2n\rho_G(\mathbf{d},\boldsymbol{\pi})$ as a test statistic for the simple null hypothesis H_0: $\boldsymbol{\pi}=\boldsymbol{\pi}_0$, $\pi_{i0}>0$ for all i. Basu and Sarkar (1994) showed that the disparity test statistic $D_{\rho G}$ has an asymptotic chi-squared distribution with $k-1$ degrees of freedom, $\chi^2_{(k-1)}$, under the null hypothesis.

3. Models with direct solutions

This Monte Carlo simulation work extends the research of Rudas (1986) to $BWHD(1/9)$, $BWCS(1/3)$, and $NED(4/3)$ statistics as well as X^2, G^2, and $I(2/3)$ statistics to explore the applicability of the chi-square approximation based on small sample sizes. This small sample behavior of the six disparity statistics is studied for 5 three-dimensional contingency tables of the completely independent model ([A][B][C] model), 6 three-dimensional contingency tables of the conditionally independent model ([AC][BC] model) - these are identical methods of Rudas(1986), and 4 three-dimensional contingency tables of the model with one variable independent of the other two ([AB][C] model), additionally. For small sample sizes which are 2 or 3 times the total number of cells, point estimators and approximate 95% confidence intervals for the 90 and 95% points of these six disparity statistics will be discussed about applicability of the chi-square approximation.

The properties of small sample behavior of not only goodness-of-fit test statistics (X^2, G^2 and $I(2/3)$) but also other statistics ($BWHD(1/9)$, $BWCS(1/3)$ and $NED(4/3)$) are studied by using 15 three-dimensional contingency tables: 5 contingency tables among 15 tables are on the completely independent model ([A][B][C] model), 6 are on the conditionally independent model ([AC][BC] model), and last 4 are on the model with one variable independent of the other two ([AB][C] model).

Three-dimensional tables of the independence model are generated by marginal probability tables. Marginal probability tables are generated and listed in <Table A> of Appendix. Three-dimensional tables of the conditionally independent model are generated by two two-dimensional joint probability tables consisted with the first and third, and the second and third variables. The tables of the model with one variable independent of the other two are generated by a joint probability table of the first and second, and a marginal probability table of the third variable. The degrees of freedom associated with the models and tables are 2, 3, 4, 5, 7, 10, 12, 16 and 20. Their sample sizes and degrees of freedom associated with the corresponding models and tables are shown in <Table B> of Appendix.

For each table and given sample size, 1,000 sample contingency tables are generated as Rudas (1984) did. From each sample the values of the six statistics are calculated. Then $(10\alpha+1)th$ order statistic is a point estimator of the α% point of the true distribution of the each statistic. Methods to obtain approximate confidence intervals for the true percentage points could be understood from David (1970). In order to compare the results of Rudas (1986) with those of disparity measures ($BWHD(1/9)$, $BWCS(1/3)$ and $NED(4/3)$), all these Monte Carlo methods used in this work are basically identical with those of Rudas (1986).

<Table 1> Chi-square approximation ratio

model	Statistic	90 % point	95 % point
Completely independent model ([A][B][C]model)	X^2	17/26	17/26
	$I(2/3)$	18/26	17/26
	G^2	4/26	7/26
	$BWCS(1/3)$	19/26	15/26
	$BWHD(1/9)$	20/26	20/26
	$NED(4/3)$	18/26	15/26
Conditionally independent model ([AC][BC]model)	X^2	19/30	25/30
	$I(2/3)$	17/30	24/30
	G^2	0/30	1/30
	$BWCS(1/3)$	12/30	16/30
	$BWHD(1/9)$	13/30	21/30
	$NED(4/3)$	7/30	14/30
One variable independent model ([AB][C]model)	X^2	18/19	16/19
	$I(2/3)$	17/19	16/19
	G^2	0/19	0/19
	$BWCS(1/3)$	14/19	13/19
	$BWHD(1/9)$	17/19	16/19
	$NED(4/3)$	13/19	13/19

Comparison results among $BWHD(1/9)$, $BWCS(1/3)$ and $NED(4/3)$ are summarized at <Table 1>, which shows the proportions that the percentage point of each limiting distribution exists in 95% confidence intervals of 90 and 95% point estimators. From <Table 1>, we could find that $BWHD(1/9)$ statistic is much more appropriate for applicability of the chi-square approximation with small sample sizes than $BWCS(1/3)$ and $NED(4/3)$ statistics.

4. Models without direct solutions

The concept of three-factor (or second-order) interaction was first introduced by Bartlett (1935) in $2\times2\times2$ contingency table. Goodman (1964) proposed several methods for analyzing the three-factor interaction in a three dimensional $I\times J\times K$ table (also see Bhapker and Koch (1968), Grizzle, Starmer and Koch (1969)). Larntz (1978) and Haber (1984) compared several tests of no three-factor interaction for only $3\times3\times3$ and $2\times2\times2$ contingency tables respectively.

The null hypothesis of no three-factor interaction can be written in terms of the following interaction term u_{ABC} or the partial association log-linear model ([AB][AC][BC] model):

$$H_0: \log m_{ijk} = u + u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{AB(ij)} + u_{AC(ik)} + u_{BC(jk)}\ (u_{ABC(ijk)} = 0), \quad (4.1)$$

where m_{ijk} is the expectation of the (ijk) cell count for $i = 1, \cdots, I$, $j = 1, \cdots, J$ and $k = 1, \cdots, K$. Several chi-square statistics are available in the literature for testing the null hypothesis in $I \times J \times K$ contingency table. Hosmane (1987) performed tests of this null hypotheses by using the Pearson chi-square, the generalized likelihood ratio and the power divergence statistics for several $3 \times 3 \times 3$ categorical data. In this work, three more disparity statistics mentioned in Section 2 are supplemented to test these null hypotheses.

Many authors including Cochran (1952), Hoeffding (1965), West and Kempthorne (1972), Chapman (1976), Lantz (1978), Koehler and Lantz (1980), and Moore and Spruill(1975) worked to compare the behaviors of a lot of goodness-of-fit test statistics for multinomial models. In this paper $3 \times 3 \times 3$ contingency tables are considered. These categorical data are generated by using the log-linear model. Values of the eighteen u terms $\{u_{A(i)}, u_{B(j)}, u_{C(k)}, u_{AB(ij)}, u_{AC(ik)}, u_{BC(jk)}\ ;\ i, j, k = 1, 2\}$ in the partial association model are simulated from the uniform distribution having appropriate intervals. The rest u terms are obtained satisfying usual restriction on u terms such as $u_{A(3)} = -(u_{A(1)} + u_{A(2)})$, $u_{B(3)} = -(u_{B(1)} + u_{B(2)})$, $\cdots$, $u_{BC(13)} = -(u_{BC(11)} + u_{BC(12)})$, $\cdots$. For each (i, j, k), cell counts for $i, j, k\ (=1, 2, 3)$, $x_{ijk}\ (\geq 1)$, satisfying H_0 in (4.1) are calculated by the following equation

$$x_{ijk} = N \frac{\exp(u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{AB(ij)} + u_{AC(ik)} + u_{BC(jk)})}{\sum_{i,j,k} \exp(u_{A(i)} + u_{B(j)} + u_{C(k)} + u_{AB(ij)} + u_{AC(ik)} + u_{BC(jk)})},$$

where N is a sample size. Fienberg (1979) suggested that the sample size might be as small as 4 or 5 times the total number of cells of the table, so that the cell counts x_{ijk} are obtained with sample sizes 50, 70, 100 and 200. Then estimates of the expected counts $\{\widehat{m}_{ijk}\}$ are obtained using the iterative proportional fitting method (see Bishop, Fienberg, and Holland (1975), for more detail). Note that $d(x_{ijk}) = x_{ijk}/N$ and $m_\beta(x_{ijk}) = \widehat{m}_{ijk}/N$ in (2.1).

Independent random samples are obtained 1,000 times for the partial association log-linear model. We explore behaviors of the six disparity test statistics discussed in Section 2. For each simulation, the disparity statistics $I(2/3)$, X^2, G^2, $BWCS(1/3)$, $BWHD(1/9)$ and $NED(4/3)$ are computed for testing H_0. The empirical levels of significance attained, viz. $\widehat{\alpha}$, are calculated as the proportion of times the values of test statistics exceed the asymptotic critical value $\chi^2_{v,\alpha}$ for the nominal values $\alpha = 0.05,\ 0.01$ with degrees of freedom $v = (I-1)(J-1)(K-1) = 8$.

The values of $\hat{\alpha}$ are given in <Table 2>.

In this work, all theses statistics perform more or less the same for large samples. For moderate samples, the statistics $I(2/3)$ and $BWHD(1/9)$ attain quite close levels to the nominal values $\alpha=0.05$ and 0.01, whereas X^2, G^2, $BWCS(1/3)$ and $NED(4/3)$ statistics do not perform well. Among them, the statistic X^2 rejects much more often than expected. The statistics G^2, $BWCS(1/3)$ and $NED(4/3)$ seem to work very poorly compared to any other statistics considered in our study.

We also study the power of the six disparity test statistics. Since the partial association log-linear model nests the conditionally independent model ([AB][AC] model), which could nest the model with one variable independent the other two ([AB][C] model), we might regard both [AB][AC] and [AB][C] models as the null hypotheses :

$$[AB][AC] : \log m_{ijk}=u+u_{A(i)}+u_{B(j)}+u_{C(k)}+u_{AB(ij)}+u_{AC(ik)}$$
$$[AB][C] : \log m_{ijk}=u+u_{A(i)}+u_{B(j)}+u_{C(k)}+u_{AB(ij)} .$$

The empirical powers of the disparity statistics which simulate for samples of size 50, 70, 100 and 200 to test the [AB][AC] model and the [AB][C] model against the [AB][AC][BC] model are listed in <Table 3>, respectively. These tables indicate that the empirical powers exceed the asymptotic critical value $\chi^2_{v,\alpha}$ for the nominal values $\alpha=0.05$ and 0.01 with degrees of freedom $I(J-1)(K-1)=12$ and $(IJ-1)(K-1)=16$, respectively.

From <Tables 3>, the powers of all these statistics perform more or less the same for large samples. For moderate samples, the statistics $I(2/3)$ and $BWHD(1/9)$ attain similar powers. We might say that whereas the statistic X^2 is over-powered, the statistics G^2, $BWCS(1/3)$ and $NED(4/3)$ are under-powered than the statistics $I(2/3)$ and $BWHD(1/9)$. This investigation is consistent with the conclusions induced from <Tables 2>. One also find that the powers of [AB][C] model are more likely than those of [AB][AC] model, since the null model ([AB][C] model) is nested by the null model ([AB][AC] model). Since the completely independent model ([A][B][C] model) is nested by [AB][C] model, the empirical powers of the disparity statistics to test [A][B][C] model against [AB][AC][BC] model have much larger than those of [AB][C] model. Therefore the empirical powers of the disparity statistics to test [A][B][C] model are not described in this paper.

<Table 2> Empirical levels at $\alpha=0.05$, and 0.01 for testing [AB][AC][BC]

Statistics	Sample size (N)			
	50	70	100	200
X^2	49 (22)	101 (44)	117 (30)	67 (13)
$I(2/3)$	22 (6)	52 (8)	56 (10)	48 (10)
G^2	3 (0)	9 (1)	20 (3)	38 (11)
$BWCS(1/3)$	2 (0)	2 (1)	13 (1)	32 (8)
$BWHD(1/9)$	17 (2)	47 (8)	49 (9)	46 (10)
$NED(4/3)$	0 (0)	1 (1)	11 (1)	27 (8)

<Table 3> Empirical powers at $\alpha=0.05$, and 0.01 for testing [AB][AC] and [AB][C] against [AB][AC][BC]

Statistics	[AB][AC] model				[AB][C] model			
	50	70	100	200	50	70	100	200
X^2	363 (171)	530 (363)	652 (456)	809 (668)	716 (527)	856 (763)	939 (864)	988 (975)
$I(2/3)$	233 (103)	403 (254)	559 (377)	792 (659)	626 (443)	811 (708)	924 (835)	987 (974)
G^2	143 (87)	273 (170)	478 (317)	780 (655)	575 (412)	761 (647)	902 (807)	986 (972)
$NED(4/3)$	109 (60)	168 (114)	390 (250)	746 (606)	439 (312)	648 (497)	834 (719)	980 (967)
$BWHD(1/9)$	227 (100)	393 (246)	552 (375)	792 (659)	625 (440)	808 (701)	921 (834)	987 (974)
$BWCS(1/3)$	121 (67)	203 (130)	428 (269)	756 (618)	475 (328)	690 (548)	867 (752)	984 (968)

We conclude in this work that the Cressie-Read's power divergence statistic $I(2/3)$ and the blended weight Hellinger distance statistic $BWHD(1/9)$ attain almost close levels that are very close to the nominal values and have similar powers. It is found that the statistic X^2 gets larger values than those empirical levels of $I(2/3)$ and $BWHD(1/9)$ and its powers are larger than those of $I(2/3)$ and $BWHD(1/9)$, whereas the statistics G^2, $BWCS(1/3)$ and $NED(4/3)$ obtain less levels and their powers are smaller than those of $I(2/3)$ and $BWHD(1/9)$. Based on these facts the Cressie-Read's power divergence statistic $I(2/3)$ and the blended weight Hellinger distance statistic $BWHD(1/9)$ are the best tests with respect to size and power for testing the partial association model.

5. Conclusions

In order to explore the applications of the chi-square approximation we consider the six disparity statistics: the Pearson chi-square X^2, the generalized likelihood ratio G^2, the power divergence $I(2/3)$, the blended weight chi-square $BWCS(1/3)$, the blended weight Hellinger distance $BWHD(1/9)$ and the negative exponential disparity $NED(4/3)$ statistics.

Three dimensional categorical data which are designed in this paper are fitted to all possible hiararchical log-linear models: the completely independent model ([A][B][C] model), the conditionally independent models ([AB]][AC], [AB][BC], [AC][BC] models), the models with one variable independent the other two ([AB][C], [AC][B], [A][BC] models) and the partial association model ([AB][BC][AC] model). These models are divided into two groups. The first group has direct solutions of expected cell counts, and the second does not have direct solutions. The first contains the completely independent model, the conditionally independent models and the models with one variable independent the other two, and the second contains the partial association model.

Several contingency tables fitted to each model which has direct solutions of expected cell counts are generated by using marginal probabilities corresponding to sufficient configurations. Then point estimators and approximate 95% confidence intervals of 90 and 95% points for these six statistics are obtained. For the partial association model which does not have direct solutions of expected cell counts, some categorical data are generated by using simulation on each term in the partial association log-linear model. The six statistics are computed for testing the three factor interaction, so that the empirical significant levels and powers are obtained.

With some results induced in Sections 3 and 4, we might conclude that both the power divergence $I(2/3)$ and the blended weight Hellinger distance $BWHD(1/9)$ statistics are good compromising alternatives to the Pearson chi-square statistic X^2 for three-dimensional categorical data fitted to all possible hierarchical log-linear models. And X^2, $I(2/3)$ and $BWHD(1/9)$ are more appropriate to the chi-square approximation than others (the generalized likelihood ratio G^2, the blended weight chi-square $BWCS(1/3)$ and the negative exponential disparity $NED(4/3)$ statistics).

Appendix

<Table A> Marginal probability tables

Model	Table	Marginals										
[A][B][C] model		p_{i++}		p_{+j+}			p_{++k}					
	2×2×2	0.6	0.4	0.6	0.4		0.4	0.6				
	2×2×3	0.4	0.6	0.56	0.44		0.2	0.3	0.5			
	2×2×4	0.45	0.55	0.39	0.61		0.2	0.3	0.2	0.3		
	2×3×3	0.4	0.6	0.25	0.5	0.25	0.4	0.4	0.2			
	2×2×6	0.45	0.55	0.39	0.61		0.1	0.1	0.19	0.11	0.2	0.3
[AC][BC] model		p_{i+k}					p_{+jk}					
	2×2×2	0.16	0.45				0.15	0.45				
		0.24	0.15				0.25	0.15				
	2×2×3	0.1	0.1	0.2			0.1	0.21	0.25			
		0.1	0.2	0.3			0.1	0.09	0.25			
	2×3×2	0.2	0.2				0.15	0.1				
		0.2	0.4				0.3	0.2				
							0.15	0.1				
	2×2×5	0.05	0.02	0.18	0.1	0.1	0.02	0.05	0.1	0.1	0.12	
		0.05	0.08	0.12	0.1	0.2	0.08	0.05	0.2	0.1	0.18	
	3×2×5	0.025	0.04	0.1	0.025	0.04	0.04	0.05	0.15	0.06	0.15	
		0.025	0.04	0.1	0.05	0.2	0.06	0.15	0.15	0.04	0.15	
		0.05	0.12	0.1	0.025	0.06						
	3×3×5	0.02	0.1	0.05	0.1	0.05	0.05	0.02	0.14	0.09	0.04	
		0.02	0.05	0.05	0.1	0.1	0.025	0.04	0.03	0.12	0.14	
		0.06	0.05	0.1	0.1	0.05	0.025	0.14	0.03	0.09	0.02	
[AB][C] model		p_{ij+}					p_{++k}					
	2×2×2	0.16	0.45				0.4	0.6				
		0.24	0.15									
	2×3×2	0.1	0.1	0.2			0.4	0.6				
		0.1	0.2	0.3								
	2×3×3	0.1	0.1	0.2			0.2	0.4	0.4			
		0.1	0.2	0.3								
	3×2×5	0.15	0.1				0.1	0.2	0.2	0.2	0.3	
		0.3	0.2									
		0.15	0.1									

<Table B> Degree of freedom and sample sizes

Model	Table	*d.f.*	Sample sizes					
[A][B][C] model	2×2×2	4	15	25	35	45	55	
	2×2×3	7	15	25	35	45	55	
	2×2×4	10	35	45	55	65	75	
	2×3×3	12	45	55	65	75	85	95
	2×2×6	16	50	75	100	125	150	
[AC][BC] model	2×2×2	2	15	25	35	45	55	
	2×2×3	3	15	25	35	45	55	
	2×3×2	4	15	25	35	45	55	
	2×2×5	5	45	55	65	75	85	95
	3×2×5	10	50	75	100	125		
	3×3×5	20	50	75	100	125	150	
[AB][C] model	2×2×2	3	15	25	35	45	55	
	2×3×2	5	15	25	35	45	55	
	2×3×3	10	25	35	45	55	65	
	3×2×5	20	50	75	100	125		

References

1. Bartlett, M. S. (1935). Contingency Table Interaction. *Journal of the Royal Statistical Society,* Suppl. 2, 248-252.
2. Basu, A. and Sarkar, S. (1994). On disparity based goodness-of-fit tests for multinomial models. *Statistics and probability Letters,* 19, 307-312.
3. Bhapker, V. P. and Koch, G. G. (1968). Hypotheses of No Interaction in Multidimensional Contingency Tables. *Technometrics,* 10, 107-123.
4. Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis.* The M.I.T. Press, Cambridge, Massachusetts.
5. Chapman, J. W. (1976). A Comparison of the X^2, $-2\log R$ and the Multinomial Probability Criteria for Significance Testing when Expected Frequencies are Small. *Journal of the American Statistical Association,* 71, 854-863.
6. Cochran, W. G. (1952). The χ^2 Test of Goodness of Fit. *Annals of Mathematical Statistics,* 23, 315-345.
7. Cochran, W. G. (1954). Some methods for strengthening the common

χ^2 test. *Biometrics,* 10, 417-451.
8. Cram´ȩr, H. (1946). *Mathematical Methods of Statistics.* Princeton, NJ : Princeton University Press.
9. Cressie, N. and Read, T. R. C. (1984). Multinomial Goodness-of-fit Tests. *Journal of the Royal Statistical Society,* B, 46, 440-464.
10. Csiszar, I. (1963). Eine informations theoretische Ungleichung ihre Anwendung auf den Beweis der Ergodizit¨ạt von Markoffschen Ketten. *Publ. math. Inst. Hungar. Acad. Sci.,* 3 85-107.
11. David, H. A. (1970). *Order Statistics.* Wiley.
12. Fienberg, S. E. (1979). The use of chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society,* B, 41, 54-64.
13 Fisher, R. A. (1941). *Statistical Methods for Research Workers.* 8th ed., Oliver and Boyd.
14. Goodman, L. A. (1964). Simple Methods for Analyzing Three-factor Interaction in Contingency Tables. *Journal of the American Statistical Association,* 59, 319-352.
15. Grizzle, J. E., Starmer, C. F. and Koch, G. G. (1969). Analysis of Categorical Data by Linear Models. *Biometrics,* 25, 489-504.
16. Haber, M. (1984). A Comparison of Tests for the Hypothesis of No Three-factor Interaction in $2 \times 2 \times 2$ Contingency Tables. *Journal of Statistical Computation and Simulation,* 20, 205-215.
17. Hoeffding, W. (1965). Asymptotically Optimal Test for Multinomial Distribution. *Annals of Mathematical Statistics,* 36, 369-408.
18. Hosmane, B. (1987). An Empirical Investigation of Chi-Square Tests for the Hypothesis of No Three-Factor Interaction in $I \times J \times K$ Contingency Tables. *Journal of Statistical Computation and Simulation,* 28, 167-178.
19. Jeong, D. B. and Sarkar, S. (2000). Negative Exponential Disparity Based Family of Goodness-of-fit Tests for Multinomial Models. *Journal of the Statistical Computation and Simulation,* 65, 43-61.
20. Juritz, J. M., Juritz, W. F. and Stephens M. A. (1983). On the accuracy of simulated percentage points. *Journal of the American Statistical Association,* 78, 441-444.
21. Kendall, M. G. (1948). *The Advanced Theory of Statistics.* vol.1, 5th ed., Griffin.
22. Koehler, K. J. and Larntz, K. (1980). An Empirical Investigation of Goodness-of-fit Statistics for Sparse Multinomials. *Journal of the American Statistical Association,* 75, 336-342.
23. Lantz, K. (1978). Small-sample comparison of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association,* 73, 253-263.
24. Le Cam, L. (1986). *Asymptotic Method in Statistical Decision Theory.* Springer, New York.

25. Lindsay, B. G. (1993). Efficiency versus Robustness: The Case for Minimum Hellinger Distance and Relasted Methods. *The Annals of Statistics,* 22, 1081-1114.
26. Moore, D. S. and Spruill, M. C. (1975). Unified Large-sample Theory of General Chi-squared Statistics Fit Tests of Fit. *Annals of Statistics,* 3, 599-616.
27. Read, T. R. C. and Cressie, N. A. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer, New York.
28. Rudas, T. (1984). Testing goodness-of-fit of log-lnear models based on small samples : a Monte Carlo study. *Colloquia Mathematica Societas János Bolyai,* 45, Goodness-of-fit, North-Holland.
29. Rudas, T. (1986). A Monte Carlo comparison of the small sample behaviour of the Pearson, the Likelihood Ratio and the Cressie-Read Statistic. *Journal of Statistical Computation and Simulation,* 24, 107-120.
30. West, E. N. and Kempthorne, O. (1972). A Comparison of the and Likelihood Ratio Tests for Composite Alternatives. *Journal of Statistical Computation and Simulation,* 1, 1-33.
31. Yarnold, J. K. (1970). The minimum expectation in χ^2 goodness of fit tests and the accuracy of approximation for the null distribution. *Journal of the American Statistical Association,* 65, 864-886.

[received date : Mar. 2006, accepted date : May. 2006]