CONVERGENCE RATES AND POWERS OF SIX POWER-DIVERGENCE STATISTICS FOR TESTING INDEPENDENCE IN 2 BY 2 CONTINGENCY TABLE

SHI-YUN SHEN

Convergence Rates and Powers of Six Power-Divergence Statistics for Testing Independence in 2 by 2 Contingency Table

by

©Shi-Yun Shen

A thesis submitted to the School of Graduate

Studies in partial fulfillment of the

requirements for the degree of

Master of Science

Department of Mathematics and Statistics

Memorial University of Newfoundland

July 1991

St. John's Newfoundland Canada

ISBN 0-315-68267-1

Abstract

This thesis investigates the convergence rates to the limiting null distribution and the powers of six test statistics of the power-divergence family (Cressie and Read, 1984) for testing independence in the 2 by 2 Contingency Table. This family of statistics can be expressed by

$$2I^{\lambda}(x:m) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2} \sum_{j=1}^{2} x_{i,j}[(\frac{x_{i,j}}{m_{i,j}})^{\lambda} - 1]$$

which is indexed by the parameter λ, $x_{i,j}$ are observed cell frequencies and $m_{i,j}$ are expected cell frequencies. It can easily be seen that the Pearson's X^2 ($\lambda = 1$), the log likelihood ratio statistic G^2 ($\lambda = 0$), the Freeman-Tukey statistic T^2 ($\lambda = -1/2$), the modified log likelihood ratio statistic MG^2 ($\lambda = -1$), the Neyman modified chi-square statistic MX^2 ($\lambda = -2$) and the Cressie-Read statistic ($\lambda = 2/3$), are all special cases.

For calculating the convergence rates and the powers of these six statistics, an iterative procedure for obtaining the minimum power-divergence estimates for the unkown parameters will be presented. It is found that among these six statistics, the convergence rate of Pearson's X^2 ($\lambda = 1$) to the limiting null distribution is the best. For the power of the test, for different alternatives, each of X^2, G^2 and MX^2 is the most powerful. It is also found that the power of the test depends not only on the noncentrality parameter but on the location of alternative hypothesis. The working rules for deciding which statistic is to be used will also be presented for the practitioner.

Key Words: Convergence Rate, Power, Power-Divergent Family, Independence Model, Minimun-Distance Estimator. Asymptotic Distribution, Non-Central Chi-Squared Distribution.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Chu-In Charles Lee, for his guidance and support in conducting this thesis. It was my pleasure to work under his supervision; his valuable advice and suggestions enabled my work to be conducted more efficiently and effectively throughout its duration. Also his hospitality and friendship made my stay in St. John's a truly memorable experience.

I take great pleasure in acknowledging the instructions I received from Professors J. Burry, C. Dalzell, L. Dionne, V. Gadag, R. Jain, A. Luong, and B. Sutradhar for their valuable lectures and strict training in my graduate programme courses.

I am grateful to the School of Graduate Studies and the Department of Mathematics and Statistics for providing me with financial support in the form of a Graduate Student Fellowship and Graduate Assistantship to make my stay at the Memorial Unversity possible.

Also, I would like to thank my wife, Hong, for her encouragement and love that I received throughout my graduate studies for the degree of Master of Science.

Contents

List of Tables

Nomenclature

N	$\cdots$	sample size in the 2 by 2 contingency table
$x_{i,j}$	$\cdots$	observed cell frequencies in the ith row and jth column
x	$\cdots$	vector of $x_{i,j}$
$m_{i,j}$	$\cdots$	expected cell frequencies in the ith row and jth column
m	$\cdots$	vector of $m_{i,j}$
$m_{i,j}^{(\lambda)}$	$\cdots$	estimate of $m_{i,j}$
$m^{(\lambda)}$	$\cdots$	vector of $m_{i,j}^{(\lambda)}$
$p_{i,j}$	$\cdots$	unknown cell probabilities
$p_{i,-}$	$\cdots$	unknown marginal probabilities
$\hat{p}_{i,-}$	$\cdots$	estimate of $p_{i,-}$
$p_{i,-}^{(\lambda)}$	$\cdots$	estimate of $p_{i,-}$ for a given λ
$p_{i,j}^{(\lambda)}$	$\cdots$	estimate of $p_{i,j}$ for a given λ
$p^{(\lambda)}$	$\cdots$	vector of $p_{i,j}^{(\lambda)}$
H_o	$\cdots$	null hypothesis
H_a	$\cdots$	alternative hypothesis
$2NI^{\lambda}(\frac{x}{N} : p)$	$\cdots$	power-divergence family of statistics
χ_1^2	$\cdots$	chi-square distribution with one degree of freedom
$\chi_1^2(N\delta^{(\lambda)})$	$\cdots$	noncentral chi-square distribution with one degree of freedom and noncentrality parameter $N\delta^{(\lambda)}$
X^2	$\cdots$	Pearson's chi-square statistic
G^2	$\cdots$	log likelihood ratio statistic
T^2	$\cdots$	Freeman-Tukey statistic
GM^2	$\cdots$	modified log likelihood ratio statistic
MX^2	$\cdots$	Neyman modified chi-square statistic
r	$\cdots$	odds ratio
α	$\cdots$	size of the test or significance level

Chapter 1

Introduction

1.1 2 by 2 Contingency Tables and its applications

Let A and B denote two categorical variables, both having 2 levels. When we classify subjects on both variables, there are 4 possible combinations of classifications. The responses (A,B) of a subject randomly chosen from the same population have a probability distribution. We display this distribution in a table having two rows for the categories of A, and two columns for those of B . The cells of the table represent the 4 possible outcomes. We denote $p_{i,j}$ as the probability that (A,B) falls in the cell in row i and column j. When the cells contain frequency counts of outcomes, the table is called a 2 by 2 contingency table.

In order to study the associations between the two different classifying variables, we will first consider the following example from Agresti (1989, p. 29-30), to illustrate the basic terminology and concepts of the 2 by 2 contingency table. Comparing the results of car accidents with seat-belt use, based on records of accidents in 1988 compiled by the Department of Highway Safety and Motor Vehicles in the state of Florida, 577,006 car accidents were classified according to variable A: accident type (A_1: belt use, A_2: no use), and variable B: accident severity (B_1: fatal, B_2: non-fatal). The following table, representing the resulting 2 by 2 contingency table of frequencies, analyzes the association between these two variables.

Table 1.1: Observed frequencies of car accidents in the 2 by 2 contingency table

	Fatal	Nonfatal	Total
No use	1601	162,527	164,128
Belt use	510	412,368	412,878
Total	2111	574,895	577,006

The resulting table will have 4 cells, and we define $x_{i,j}$ to be the number of observations classified as in A_i and B_j, which is the cell frequency pertaining to the cell in row i and column j of this table. The marginal freqencies are the row totals and column totals obtained by the summation of the appropriate cell frequencies, and are denoted as $x_{i,+} = \sum_{j=1}^{2} x_{i,j}$, i=1,2, and $x_{+,j} = \sum_{i=1}^{2} x_{i,j}$, j=1,2. The total cell frequency is denoted as $N=x_{+,+} = \sum_{i=1}^{2}\sum_{j=1}^{2} x_{i,j}$.

The above application of the 2 by 2 contingency table, in studying the relationship between the belt use with the severity of accident, can also have applications in many other studies, such as, studying the associations between Lung Cancer with the Smoking Level (Doll and Hill 1952), Income with Job Satisfication (Norusis, 1988), and Gun Registration with Death Penalty (Clogg and Shockey, 1988).

The analysis of relationships or associations between cross-classified variables has developed extensively over the last twenty years. For a comprehensive coverage of the studies of cross-classified categorical data, see, for example, Bishop, Fienberg and

Holland (1975), Haberman (1978, 1979) and Agresti (1989). Lloyd (1988) reviewed the 40-year old controversy over the correct analysis of 2 by 2 tables.

1.2 Testing the Model of Independence

What is the proper probability model of $p_{i,j}$ for the data in the table? The model must reflect the way in which the data was collected. There are several different sampling procedures which could lead to this table. A single random sample size N might be selected, and then categorized in two ways. For example, a random sample of persons can be treated with or without belt use and for accident fatality or nonfatality. We call this model A. Under model A, the cell frequencies $x_{i,j}$ have a single multinomial distribution,

$$f(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) = \frac{N!}{x_{1,1}!x_{1,2}!x_{2,1}!x_{2,2}!} p_{1,1}^{x_{1,1}} p_{1,2}^{x_{1,2}} p_{2,1}^{x_{2,1}} p_{2,2}^{x_{2,2}}.$$

The marginal frequencies are all random, and satisfy $x_{1,+} + x_{2,+} = x_{+,1} + x_{+,2}$=N.

It might also separately select two independent random samples of persons with lung cancer and those without, then categorize them according to heavy or light smoking. Under this model, say model B , the table contains two independent binomial distributions, and $x_{1,+}$ and $x_{2,+}$ are therefore, no longer random.

Hypotheses for these models are stated in terms of the cell probabilities $p_{i,j}$ for the sampled populations. Each model imposes different constraints on the $p_{i,j}$. Model A requires only that:

$$\sum_{i=1}^{2}\sum_{j=1}^{2} p_{i,j} = 1.$$

Model B states that for i=1,2,

$$p_{i,+} = \sum_{j=1}^{2} p_{i,j} = 1.$$

The most common hypothesis in the 2 by 2 contingency table formalize the statement that there is no association between the two categories. For example, there is no association between the seat-belt use and severity of accident, income and job satisfaction, or aspirin use and heart attacks. In model A, this is the hypothesis of independence:

$$H_o : p_{i,j} = p_{i,+} * p_{+,j}, i = 1, 2, \text{and}, j = 1, 2. \tag{1.1}$$

where $p_{i,+}$ and $p_{+,j}$ represent the unknown marginal probabilities.

As the model of independence (1.1) is the most often considered in the 2 by 2 contingency table, this thesis will be concentrated upon in testing this model by using the following power-divergence statistics.

1.3 Pearson's X^2, Loglikelihood Ratio G^2 and the Power-Divergence Family of Statistics

As a test criterion for the null hypothesis, Karl Pearson (1900) proposed the test :

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(x_{i,j} - m_{i,j})^2}{m_{i,j}} \tag{1.2}$$

where $x_{i,j}$ and $m_{i,j}$ are observed and expected cell frequencies for the ith and jth cell respectively, and suggested that the asymptotic distribution of X^2 is a χ^2 distribution with 3 degrees of freedom. The latter caused some confusion and controversy in practical applications and was not settled until 20 years later. Fisher (1922,1924) investigated the 2 by 2 contingency tables , and pointed out that the limiting distribution of X^2 depends on the method of estimation for the unknown parameters of $p_{i,\cdot}$ and $p_{\cdot,j}$. With a different method of estimation, X^2 may have a large test value, and the null hypothesis would be rejected for the large value; or the limiting distribution of X^2 will be other than the chi-squared distribution. It is therefore necessary to state that the method of estimation is to minimize the objective function. Among others, Fisher (1924), Moore (1978), and Berkson (1980) have also considered this class of Minimun Distance Estimation.

Cochran (1952) gave a review of the early development of the Pearson's chi-squared test X^2, and discussed a variety of competing tests, including the log likelihood ratio test G^2:

$$G^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} x_{i,j} ln(\frac{x_{i,j}}{m_{i,j}}) \tag{1.3}$$

The question of which one is the best has long provided interest, speculation and controversy in literature.

Cochran (1936) felt that the G^2 distribution might be better represented by a continuous curve than the X^2 distribution in a contingency table, and he (1952) concluded that there is little to distinguish X^2 from G^2. Fisher (1950) found the X^2 test to be governed by and more sensitive to high observations in several cells than the G^2 test. Chapman (1976) concluded that the difference between the exact X^2 test probabilties and the chi-squared probabilities is usually smaller than that between the exact G^2 test probabilities and the chi-squared probabilities. West and

Kempthorne (1971) plotted the sensitivity of X^2 and G^2 for a few cases of two, three, and four cells, and concluded that there existed different regions for which one is better than other. In a Monte Carlo study, Kallenberg, Oosterhoff, and Schriever (1985) showed that X^2 and G^2 have similar powers for testing the equiprobable null hypothesis for the cells with small frequencies. Lee (1987) confirmed that the powers of the X^2 and G^2 tests depend on the noncentrality parameters in testing for ordered restriction on multivariate parameters. Three of the potential models are discussed by Kroll (1989), they are the hypergeometric independence trial, the double-binomial comparative trial, and the multinomial double dichotomy trial for testing independence in 2 by 2 contingency table.

Cressie and Read (1984) introduced a power-divergence family of goodness-of-fit statistics, I^λ, defined as :

$$2 \quad (x : m) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2} \sum_{j=1}^{2} x_{i,j}[(\frac{x_{i,j}}{m_{i,j}})^\lambda - 1]. \tag{1.4}$$

Statistics (1.4) are obtained by the continuity for the cases of $\lambda = -1$ and $\lambda = 0$. Using the fact that $ln(t) = lim_{h \to 0}(t^h - 1)/h$, we obtain:

$$lim_{\lambda \to 0} 2I^\lambda(x : m) = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} x_{i,j} ln(\frac{x_{i,j}}{m_{i,j}})$$

and

$$lim_{\lambda \to -1} 2I^\lambda(x : m) = 2 \sum_{i=1}^{2} \sum_{j=1}^{2} m_{i,j} ln(\frac{m_{i,j}}{x_{i,j}}).$$

It is noted that when $\lambda = 1$, the statistic (1.4) is the Pearson's X^2; $\lambda = 0$, the log likelihood ratio statistic G^2; $\lambda = -1/2$, the Freeman-Tukey (1950) statistic T^2;

$$T^2 = 4\sum_{i=1}^{2}\sum_{j=1}^{2}(\sqrt{x_{i,j}} - \sqrt{m_{i,j}})^2 \tag{1.5}$$

$\lambda = -1$, the modified log likelihood ratio statistic or minimum discrimination information statistic GM^2 (see Kullback, 1959,1985);

$$GM^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} m_{i,j} ln(\frac{m_{i,j}}{x_{i,j}}); \tag{1.6}$$

$\lambda = -2$; the Neyman (1949) modified chi-square statistic MX^2;

$$MX^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(x_{i,j} - m_{i,j})^2}{x_{i,j}}. \tag{1.7}$$

The theory underlying these statistics, as of most omnibus tests of fit, is a large-sample theory. Under the null hypothesis H_o, this theory is well understood. When N tends to infinity in the 2 by 2 contingency table, all the members of I^λ have the same limiting null distribution, the chi-square distribution with one degree of freedom χ_1^2, as they are all equivalent, due to Cressie and Read (1984). For Pearson's X^2, this large sample approximation is surprisingly accurate for moderate and small sample sizes N, especially when the cells are equiprobable (see Yarnold 1972); however, the approximation is markedly less accurate for λ far from 1 (Larntz 1987 and Read 1984). Cressie and Read (1984) found the chi-squared approximation adequate for λ between 1/3 and 3/2.

Cressie and Read (1984) proposed a test which is between X^2 ($\lambda = 1$) and G^2 ($\lambda = 0$), viz. $\lambda = 2/3$, to take advantage of the desirable properties of both. The test:

$$2I^{2/3}(x:m) = \frac{9}{5}\sum_{i=1}^{2}\sum_{j=1}^{2} x_{i,j}[(\frac{x_{i,j}}{m_{i,j}})^{2/3} - 1] \quad (1.8)$$

has been called the Cressie-Read Statistic by Rudas (1986), and earlier, Moore (1984) named the family (1.4) the Cressie-Read statistics. As an omnibus test of goodness-of-fit, Cressie and Read (1984) showed it to be the most competitive in this family of (1.4). More details can be found in Read (1984) and Read and Cressie (1988).

We will present iterative procedures in Chapter 2 to obtain the minimum power-divergence estimators for the unknown parameters $p_{i,j}$, which are the BAN estimators defined by Neyman (1949); and which are also in the class of the minimum distance estimators. In Chapter 3, we will investigate the convergence rates to the null limiting chi-squared distribution of the aforementioned six different statistics for testing the model of independence in the 2 by 2 contingency tables, by using the minimun power-divergence estimators for the unknown parameters. In Chapter 4, comparisons between the powers of these statistics for testing independence will be investigated. Discussions can be found in the last Chapter.

Chapter 2

Minimum Power-Divergence Estimation for Testing Independence in the 2 by 2 Contingency Table

2.1 Asymptotic Equivalence of the Power-Divergence statistics

We consider testing the hypothesis of independence in the 2 by 2 contingency table

$$H_o : p_{i,j} = p_{i,+} * p_{+,j}, i = 1,2, \text{and}, j = 1,2. \quad (2.1)$$

Let the sample size be N and let $x_{i,j}$ be the observed cell frequencies, i=1,2 and j=1,2. The random vector x=$(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2})$ has a multinomial distribution with sample size N and cell probabilities p=$(p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2})$. The parameters $p_{1,+}, p_{2,+}, p_{+,1}$ and $p_{+,2}$ will be estimated under certain principles and their estimates are denoted by $\hat{p}_{1,+}$ $\hat{p}_{2,+}$ $\hat{p}_{+,1}$ and $\hat{p}_{+,2}$. The expected frequencies under (2.1) are

$$m_{i,j} = N\hat{p}_{i,+} * \hat{p}_{+,j}, i = 1,2, \text{and}. j = 1,2.$$

Cressie and Read (1984) introduced a family of power-divergence statistics,

$$2I^{\lambda}(x:m) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2}\sum_{j=1}^{2} x_{i,j}[(\frac{x_{i,j}}{m_{i,j}})^{\lambda} - 1] \qquad (2.2)$$

for testing the null hypothesis (2.1). The null distributions of the statistics (2.2) are asymptotically equivalent based on Theorem 2.2 of Cressie and Read (1984).

Theorem 2.1: Under the null hypothesis, the family of power-divergence statistics are asymptotically equivalent, in the sense that

$$2I^{\lambda}(x:m) = 2I^{1}(x:m) + o_p(1) \qquad (2.3)$$

where $o_p(1)$ represents a term that converges to 0 as $N \to \infty$.

Proof: The statistics (2.2) can be written as

$$2I^{\lambda}(x:m) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2}\sum_{j=1}^{2} m_{i,j}[(1+\frac{x_{i,j}-m_{i,j}}{m_{i,j}})^{\lambda-1} - 1]$$

provided that $\lambda \neq 0$ nor -1. Now let $v_{i,j} = (x_{i,j} - m_{i,j})/m_{i,j}$, and expand in a Taylor series for each λ, giving

$$2I^{\lambda}(x:m) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2}\sum_{j=1}^{2} m_{i,j}[(\lambda+1)v_{i,j} + \frac{\lambda(\lambda+1)}{2}v_{i,j}^2 + o_p(1)]$$

$$= [\sum_{i=1}^{2}\sum_{j=1}^{2} m_{i,j}v_{i,j}^2 + o_p(1)] = 2I^{1}(x:m) + o_p(1)$$

An identical result holds for the special cases of λ=0 and -1, also by a Taylor series expansion. This completes the proof of the asymptotic equivalence of the power-divergence statistics. □

This result is sufficient for us to conclude that each member of the power-divergence family has the same asymptotic null distribution.

2.2 BAN estimator for the unknown parameters under Birch's Regularity Conditions

In order to test the null hypothesis (2.1), we need to estimate the unknown marginal probability $p_{i,+}$, i=1,2 and $p_{+,j}$, j=1,2. Moore (1978) recommended that the MLE, which is equivalent to minimizing (2.2) with respect to $m_{i,j}$ when $\lambda=0$, is always used in the contingency table, namely $x_{i,+}/N$ and $x_{+,j}/N$ respectively for $p_{i,+}$ and $p_{+,j}$. This gives the estimated expected frequencies:

$$m_{i,j} = N * (\frac{x_{i,+}}{N}) * (\frac{x_{+,j}}{N}) = \frac{x_{i,+}x_{+,j}}{N}$$

A natural procedure is to obtain the estimate which minimize (2.2) according to the specific value of λ. This leads to the following definition of the minimum power-divergence estimate $m_{i,j}^{(\lambda)}$ of $m_{i,j}$:

Definition: $m_{i,j}^{(\lambda)}$ satisfying the following equation is called the minimum power-divergence estimate of $m_{i,j}$ for a specific value of λ.

$$I^{\lambda}(x : m^{(\lambda)}) = \inf_{m} I^{\lambda}(x : m). \qquad (2.4)$$

Since $(x^{-\lambda} - 1)/\lambda(\lambda + 1)$ is strictly convex for $x > 0$, including the limiting forms for $\lambda = 0$ and -1. The strict convexity of $I^{\lambda}(x : m)$ ensures uniqueness of the estimate $m^{(\lambda)}$ (see Read and Cressie 1988, Appendix 2. page 159). Various authors (e.g. Fisher (1924), Moore (1978), Berkson (1980)) have considered the minimun power-divergence estimator, and Parr (1981) has provided an extensive bibliography for this estimation.

Why then should we consider alternative methods to the MLE ? Rao (1961,1962) defined a second-order efficiency criterion, for which he showed that the method of Maximum Likelihood Estimation provides the unique optimum estimate. However, the sovereignty of MLE has been called into question by a number of authors (e.g. Berkson 1980, Parr 1981, Harris and Kanji 1983). Therefore, it seems quite reasonable to recommend the minimum power-divergence estimate. For example, if we decide to use Pearson's X^2 (λ=1) for the test, then we might estimate m by $m^{(1)}$.

In the case of testing independence in the 2 by 2 contingency table, the way to estimate $p_{i,j}$ is to choose the $p_{i,j}^{(\lambda)}$ that is closest to $\frac{x}{N}$, which satisfies:

$$I^{\lambda}(\frac{x}{N} : p^{(\lambda)}) = \inf_{p} I^{\lambda}(\frac{x}{N} : p) \qquad (2.5)$$

where p is a vector of $p_{i,j}$ and $p^{(\lambda)}$ a vector of $p_{i,j}^{(\lambda)}$.

Ensuring that the minimum power-divergence estimate $p_{i,j}^{(\lambda)}$, a function $p_{i,j}^{(\lambda)}$ = F($p_{i,+}^{(\lambda)}$,$p_{+,j}^{(\lambda)}$) under H_o, exists, Birch (1964) defined a set of six regularity conditions sufficient to ensure that the null model H_o really has 2 parameters, and that F satisfies various requirements. The conditions are:

(1): there is a 2 dimensional open neighbourhood of ($p_{1,+}^{(\lambda)}$, $p_{+,1}^{(\lambda)}$), which is completely contained in its domain P_o,

(2): F($p_{1,+}^{(\lambda)}$,$p_{+,1}^{(\lambda)}$) > 0, for each of the two coordinates.

(3): F is totally differentiable at ($p_{1,+}^{(\lambda)}$,$p_{+,1}^{(\lambda)}$), with partial derivatives.

(4): the Jacobian of F of $p_{1,+}$ and $p_{+,1}$ is of full rank 2, (5): the inverse mapping F^{-1} is continuous at $F(p_{1,\cdot}^{(\lambda)}, p_{\cdot,1}^{(\lambda)})$.

(6): the mapping F is continuous at every point in its domain P_o.

Birch (1964) showed that any estimate satisfying the regularity conditions is best asymptotically normal (BAN) defined by Neyman (1949), which have three important properties:

(A): They are consistent, i.e., the estimate converges to the true value of the estimated parameter as $n \to \infty$.

(B): They are asymptotically normally distributed. The asymptotical chi-squared distribution for $2NI^{\lambda}(\frac{x}{N} : p^{(\lambda)})$ is based on this result.

(C): They are asymptotically efficient, no other estimator can have a smaller limiting variance, as $N \to \infty$.

By a straightforward generalization of the argument provided by Birch (1964) we obtain the following theorem:

Theorem 2.2: Under the above regularity conditions, any minimum power-divergence estimator under H_o in the 2 by 2 contingency table is **BAN**.

Birch (1964) showed that the MLE (λ=0) is BAN. In another case, the minimum chi-squared estimator (λ=1) was proved to be BAN by Holland (1967), and the general case for any λ was proved by Read and Cressie (1988, Appendix 5. pp. 163-166). It follows that $p^{(\lambda)}$ is a BAN estimate under H_o, and $\sqrt{N}(\frac{x}{N} - p^{(\lambda)})$ converges in distribution, as $N \to \infty$, to a multivariate normal random vector with variance-covariance matrix $A - pp'$, here A is a 4×4 diagonal matrix with diagonal entries $p_{1,1}$ $p_{1,2}$ $p_{2,1}$ and $p_{2,2}$, and p is a column vector of $p_{i,j}$. It is well known that

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(x_{i,j} - Np_{i,j}^{(1)})^2}{Np_{i,j}^{(1)}}$$

can be written as a quadratic form of $\sqrt{N}(\frac{x}{N} - p^{(1)})$, and X^2 converges in distribution to a chi-squared random variable with one degree of freedom under H_o. Therefore, by the asymptotic equivalence of the power-divergence family (2.3), it indicates that for testing independence H_o in the 2 by 2 contingency table the power-divergence statistics are asymptotically chi-squared distributed with one degree of freedom under H_o.

2.3 Iterative Procedure of Minimum Power-Divergence estimate in the cases of λ >0

Under the independence model H_o in the 2 by 2 tables, there are no closed-form solutions of the unknown parameters $p_{i,+}$ and $p_{+,j}$ for members of the power-divergence statistics, except for the maximum likelihood estimate when λ=0. This is one of the reasons why Moore (1978) recommended that the MLE is always used for the contingency table. Though the minimum power-divergence estimate attracts many authors' interest, very few use it in the contingency table to test the hypothesis of independence.

The problem, for $\lambda = 1$, has been studied by Cohen (1982), and recently, a general method to compute the minimum power-divergence estimate $p_{i,+}^{(\lambda)}$ and $p_{+,j}^{(\lambda)}$ was discussed by Bohning and Holling (1986).

The power-divergence statistics (2.2) under H_o can be expressed as a function of $p_{i,+}$ and $p_{+,j}$ through $f_{i,j} = C_{i,j}p_{i,+}^{-\lambda}p_{+,j}^{-\lambda}$:

$$2NI^{\lambda}(\frac{x}{N}:p)=\sum_{i=1}^{2}\sum_{j=1}^{2}f_{i,j}-C$$

where $C_{i,j}=2x_{i,j}^{\lambda-1}/\lambda(\lambda-1)N^{\lambda}$ for i=1,2 and j=1,2 and C=$2N/\lambda(\lambda-1)$ are constants for a given λ. If $f_{i,j}$ is strictly convex for all i and j, then so is $\sum_{i=1}^{2}\sum_{j=1}^{2}f_{i,j}$.

Theorem 2.3: For any $\lambda>0$, the power-divergence statistics for testing the hypothesis of (2.1) is a convex function of $p_{1,+}$ and $p_{+,1}$.

Proof: Suppose that F is a twice continuously differentiable real-valued function on an open convex set C in R^n. Then F is convex on C if and only if its Hessian matrix $Q_x=(q_{ij}(x))$,

$$q_{ij}(x)=\frac{d^2F}{dx_idx_j}(x_1,\ldots,x_n)$$

is positive semi-definite for every $x\in$C. (see, Rockafellar, 1972, pp.27)

Therefore, the convexity of $f_{i,j}$ over $p_{1,+}>0$, $p_{+,1}>0$ is equivalent to the following three inequalities:

$$\lambda(\lambda+1)p_{1,+}^{-(\lambda+2)}p_{+,1}^{-\lambda}\geq 0,$$

$$\lambda(\lambda+1)p_{1,+}^{-\lambda}p_{+,1}^{-(\lambda+2)}\geq 0,$$

and

$$\lambda^2(2\lambda+1)p_{1,+}^{-2(\lambda+1)}p_{+,1}^{-2(\lambda+1)}\geq 0.$$

The sufficient condition of convexity of the power-divergence statistics is $\lambda>0$. This completes the proof. □

Remark 1: Bohning and Holling (1986) stated, incorrectly, that the sufficient condition of convexity of the power-divergence statistics is λ >-1/2. There is a defect in their proof, in which they required a less restrictive condition that the determinant of its Hessian Matrix is positive. Consider the function

$$f(x, y) = -(2x - 3y)^2 - x^2.$$

The determinant of its Hessian Matrix is 36. However, it is not convex. Therefore, their conclusion should be amended as Theorem 2.3 above.

For $\lambda = 0$, the estimates of $p_{1,+}$ and $p_{+,1}$ are the MLE, $x_{1,+}/N$ and $x_{+,1}/N$, denoted by $p_{1,+}^{(0)}$ and $p_{+,1}^{(0)}$, respectively. For $\lambda > 0$, the convexity of the statistics ensures that the minimum power-divergence estimates exist and are unique. Although there are no closed form expressions for these estimates, we shall present the following well-known iterative procedure, which is commonly used in the analysis of contingency tables and linear models to obtain the estimates for the unknown parameters. Since $p_{2,+}$ and $p_{+,2}$ are redundant under H_o, we denote the power-divergence statistics $I^{(\lambda)}(p_{1,+}, p_{+,1})$ as a function of $p_{1,+}$ and $p_{+,1}$

Iterative Procedure 1:

Step 0. (initial value).

Choose any initial value $p_{1,+}^{(0)}$, $0 < p_{1,+}^{(0)} < 1$, and let n=0.

Step 1. Compute

$$p_{+,1}^{(n+1)} = \frac{(x_{1,1}^{\lambda+1}(1 - p_{1,+}^{(n)})^{\lambda} + x_{2,1}^{\lambda+1}(p_{1,+}^{(n)})^{\lambda})^{\frac{1}{\lambda+1}}}{(x_{1,1}^{\lambda+1}(1 - p_{1,+}^{(n)})^{\lambda} + x_{2,1}^{\lambda+1}(p_{1,+}^{(n)})^{\lambda})^{\frac{1}{\lambda+1}} + (x_{1,2}^{\lambda+1}(1 - p_{1,+}^{(n)})^{\lambda} + x_{2,2}^{\lambda+1}(p_{1,+}^{(n)})^{\lambda})^{\frac{1}{\lambda+1}}} \qquad (2.6)$$

Step 2. Compute

$$p_{1,+}^{(n+1)} = \frac{(x_{1,1}^{\lambda+1}(1-p_{+,1}^{(n+1)})^{\lambda} + x_{1,2}^{\lambda+1}(p_{+,1}^{(n+1)})^{\lambda})^{\frac{1}{\lambda+1}}}{(x_{1,1}^{\lambda+1}(1-p_{+,1}^{(n+1)})^{\lambda} + x_{1,2}^{\lambda+1}(p_{+,1}^{(n+1)})^{\lambda})^{\frac{1}{\lambda+1}} + (x_{2,1}^{\lambda+1}(1-p_{+,1}^{(n+1)})^{\lambda} + x_{2,2}^{\lambda+1}(p_{+,1}^{(n+1)})^{\lambda})^{\frac{1}{\lambda+1}}} \tag{2.7}$$

Replace n by n+1 and go to step 1. The sequence $(p_{1,+}^{(n)}, p_{+,1}^{(n)})$ converges to the desired solution.

Theorem 2.4: Let $\lambda > 0$ and $(p_{1,+}^{(n)}, p_{+,1}^{(n)})$ be a sequence obtained by the Iterative Procedure 1 for any initial value between 0 and 1. Then the sequence converges to the same limit $(p_{1,+}^{(\lambda)}, p_{+,1}^{(\lambda)})$ and this limit minimizes the power-divergence statistics, i.e.,

$$I^{(\lambda)}(\frac{x}{N} : p^{(\lambda)}) \leq I^{(\lambda)}(\frac{x}{N} : p)$$

for any p=$(p_{i,j})$, with $p_{i,j} = p_{i,+} * p_{+,1}$ where $p_{i,j}^{(\lambda)} = p_{i,+}^{(\lambda)} * p_{+,j}^{(\lambda)}$.

Proof: As in (2.2), I^{λ} is a function of the two variables $p_{1,+}$ and $p_{+,1}$, and $I^{\lambda}(p,q) > 0$ if $p \neq$q. We differentiate I^{λ} with respect to $p_{+,1}$ and $p_{1,+}$, and set them equal to zero, thus obtaining the equations (2.8) and (2.9):

$$p_{+,1} = \frac{(x_{1,1}^{\lambda+1}(1-p_{1,+})^{\lambda} + x_{2,1}^{\lambda+1}p_{1,+}^{\lambda})^{\frac{1}{\lambda+1}}}{(x_{1,1}^{\lambda+1}(1-p_{1,+})^{\lambda} + x_{2,1}^{\lambda+1}p_{1,+}^{\lambda})^{\frac{1}{\lambda+1}} + (x_{1,2}^{\lambda+1}(1-p_{1,+})^{\lambda} + x_{2,2}^{\lambda+1}p_{1,+}^{\lambda})^{\frac{1}{\lambda+1}}} \tag{2.8}$$

$$p_{1,+} = \frac{(x_{1,1}^{\lambda+1}(1-p_{+,1})^{\lambda} + x_{1,2}^{\lambda+1}p_{+,1}^{\lambda})^{\frac{1}{\lambda+1}}}{(x_{1,1}^{\lambda+1}(1-p_{+,1})^{\lambda} + x_{1,2}^{\lambda+1}p_{+,1}^{\lambda})^{\frac{1}{\lambda+1}} + (x_{2,1}^{\lambda+1}(1-p_{+,1})^{\lambda} + x_{2,2}^{\lambda+1}p_{+,1}^{\lambda})^{\frac{1}{\lambda+1}}} \tag{2.9}$$

Let $\theta = p_{1,+}/(1-p_{1,+})$ and $\phi = p_{+,1}/(1-p_{+,1})$, one-to-one transformations, we obtain:

$$\theta^{\lambda-1} = \frac{x_{1,1}^{\lambda-1} + x_{1,2}^{\lambda-1}\phi^{\lambda}}{x_{2,1}^{\lambda+1} + x_{2,2}^{\lambda-1}\phi^{\lambda}} \tag{2.10}$$

$$\phi^{\lambda-1} = \frac{x_{1,1}^{\lambda+1} + x_{2,1}^{\lambda-1}\theta^{\lambda}}{x_{1,2}^{\lambda+1} + x_{2,2}^{\lambda-1}\theta^{\lambda}} \tag{2.11}$$

The two parameters θ and ϕ are bounded, $\phi \in [x_{2,1}/x_{2,2},\ x_{1,1}/x_{1,2}]$ and $\theta \in [x_{1,2}/x_{2,2},\ x_{1,1}/x_{2,1}]$, if the odds ratio $r = x_{1,1}x_{2,2}/x_{1,2}x_{2,1} > 1$, and $\phi \in [x_{1,1}/x_{1,2}, x_{2,1}/x_{2,2}]$, $\theta \in [x_{1,1}/x_{2,1}, x_{1,2}/x_{2,2}]$ if $r < 1$. Since

$$\frac{d\theta^{\lambda+1}}{d\phi} \times \frac{d\phi^{\lambda+1}}{d\theta} > 0,$$

where

$$\frac{d\theta^{\lambda+1}}{d\phi} = \frac{\lambda\phi^{2\lambda-1}(x_{1,2}^{\lambda+1}x_{2,1}^{\lambda+1} - x_{2,2}^{\lambda-1}x_{1,1}^{\lambda+1})}{(x_{2,1}^{\lambda+1} + x_{2,2}^{\lambda+1}\phi^{\lambda})^2}$$

$$\frac{d\phi^{\lambda+1}}{d\theta} = \frac{\lambda\theta^{2\lambda-1}(x_{1,2}^{\lambda+1}x_{2,1}^{\lambda+1} - x_{2,2}^{\lambda+1}x_{1,1}^{\lambda+1})}{(x_{1,2}^{\lambda+1} + x_{2,2}^{\lambda+1}\theta^{\lambda})^2}$$

the functions of (2.10) and (2.11) are either both monotonically increasing, or both monotonically decreasing, depending upon whether the odds ratio $r > 1$ or $r < 1$. Since $p_{1,+}$ and $p_{+,1}$ are monotone functions of θ and ϕ respectively and the functions (2.8) and (2.9) are bounded over [0,1], we have that $(p_{1,+}^{(n)}, p_{+,1}^{(n)})$ converges monotonically to a limit, say $(p_{1,+}^{(\lambda)}, p_{+,1}^{(\lambda)})$. By the continuity of equations (2.8) and (2.9), the limit $(p_{1,+}^{(\lambda)}, p_{+,1}^{(\lambda)})$ satisfies these two equations, and since $I^{\lambda}(x/N : p)$, under the null hypothesis H_o: $p_{i,j} = p_{i,+} * p_{+,j}$, is a convex function of $p_{1,+}$ and $p_{+,1}$, $p^{(\lambda)}$ is the desired solution. □

Remark 1: When the odds ratio r=1,

$$Nx_{1,1} = (x_{1,1} + x_{1,2} + x_{2,1} + x_{2,2})x_{1,1} =$$

$$x_{1,1}(x_{1,1} + x_{1,2}) + x_{1,1}x_{2,1} + x_{1,2}x_{2,1} = x_{1,+} * x_{+,1}$$

the estimates $p^{(\lambda)}_{1,+}$ and $p^{(\lambda)}_{+,1}$ are $\frac{x_{1,1}}{x_{1,+}}$ and $\frac{x_{1,1}}{x_{+,1}}$, respectively, and the values of every power-divergence statistic $I^{(\lambda)}(x/N : p^{(\lambda)})$ will be zero.

So far, for $\lambda > 0$ by using the iterative procedure presented above, we can obtain the minimun power-divergence estimates.

2.4 Iterative Procedure of Minimun Power-Divergence Estimates for $\lambda < 0$

When $\lambda < 0$, $I^{\lambda}(x/N : p)$ may not necessarily be a convex function of $p_{1,+}$ and $p_{+,1}$. We shall present the conditions in Theorem 2.5 to ensure that the Iterative Procedure 1 is available to obtain the minimun power-divergence estimate for λ=-2, and λ=-1/2. This approach may be applied to all other cases when $\lambda < 0$.

Theorem 2.5: For any $\lambda < 0$ and $\lambda \neq -1$, let $(p^{(n)}_{1,+}, p^{(n)}_{+,1})$ be a sequence obtained by the Iterative Procedure 1. If $r > 1$ and

$$(\frac{\lambda}{\lambda+1})^2 \frac{x^2_{2,2}x^{4\lambda+1}_{1,2}x^{4\lambda+1}_{2,1}(1-r^{\lambda+1})^2}{(x^{\lambda+1}_{2,1}x^{\lambda}_{1,2} + x^{\lambda-1}_{2,2}x^{\lambda}_{1,1})^2(x^{\lambda+1}_{1,2}x^{\lambda}_{2,1} + x^{\lambda+1}_{2,2}x^{\lambda}_{1,1})^2} < \frac{1}{2} \qquad (2.12)$$

or if $r < 1$ and

$$(\frac{\lambda}{\lambda+1})^2 \frac{x^3_{1,2}x^3_{2,1}(1-r^{\lambda+1})^2}{x^2_{1,1}(x_{1,2}+x_{2,2})^2(x_{2,1}+x_{2,2})^2} < \frac{1}{2} \qquad (2.13)$$

then the sequence converges to a unique limit $(p_{i,+}^{(\lambda)}, p_{+,j}^{(\lambda)})$, and this limit minimizes the power-divergence statistics, i.e.,

$$I^{\lambda}(x/N : p^{(\lambda)}) \leq I^{\lambda}(x/N : p)$$

for any p=$(p_{i,j})$ with $p_{i,j} = p_{i,+} * p_{+,j}$.

Proof: We shall prove the case of $r > 1$, and the case of $r < 1$ follows similarly. Let D=$\{(x_1, x_2)$: $a_i \leq x_i \leq b_i$, $i = 1, 2\}$ with given a_1, a_2, b_1 and b_2. Let g(x)=$(g_1(x), g_2(x))$ be a continuous function which maps from D into D itself. Ortega (1972, p.153) showed that g has at least one fixed point in D, i.e., there exists $x_o \in$D such that g(x_o)=x_o, and the fixed point is unique if g has continuous partial derivatives and if there exists a constant k<1 such that

$$|det(\frac{\partial g_i}{\partial x_j})| \leq \frac{k}{2}. \tag{2.14}$$

It suffices to consider the mapping

$$g_1(\theta, \phi) = (\frac{x_{1,1}^{\lambda+1} + x_{1,2}^{\lambda+1}\phi^{\lambda}}{x_{2,1}^{\lambda+1} + x_{2,2}^{\lambda+1}\phi^{\lambda}})^{\frac{1}{\lambda+1}} \tag{2.15}$$

$$g_2(\theta, \phi) = (\frac{x_{1,1}^{\lambda+1} + x_{2,1}^{\lambda+1}\theta^{\lambda}}{x_{1,2}^{\lambda+1} + x_{2,2}^{\lambda+1}\theta^{\lambda}})^{\frac{1}{\lambda+1}} \tag{2.16}$$

If $r > 1$, then $g_1 \in [\frac{x_{1,2}}{x_{2,2}}, \frac{x_{1,1}}{x_{2,1}}]$, $g_2 \in [\frac{x_{2,1}}{x_{2,2}}, \frac{x_{1,1}}{x_{1,2}}]$, and because of $\lambda < 0$, $\theta^{\lambda} \in [(\frac{x_{1,1}}{x_{2,1}})^{\lambda}, (\frac{x_{1,2}}{x_{2,2}})^{\lambda}]$, $\phi^{\lambda} \in [(\frac{x_{1,1}}{x_{1,2}})^{\lambda}, (\frac{x_{2,1}}{x_{2,2}})^{\lambda}]$. By (2.14)

$$|\frac{\partial g_1}{\partial \phi} \times \frac{\partial g_2}{\partial \theta}| =$$

$$(\frac{\lambda}{\lambda+1})^2 \frac{\theta^{\lambda-1}\phi^{\lambda-1}x_{1,2}^{2(\lambda+1)}x_{2,1}^{2(\lambda-1)}(1-r^{\lambda-1})^2}{g_1^{\lambda}g_2^{\lambda}(x_{2,1}^{\lambda+1}+x_{2,2}^{(\lambda+1)}\phi^{\lambda})^2(x_{1,2}^{\lambda+1}+x_{2,2}^{(\lambda-1)}\theta^{\lambda})^2} \leq$$

$$(\frac{\lambda}{\lambda+1})^2 \frac{x_{2,2}^2 x_{1,2}^{4\lambda-1}x_{2,1}^{4\lambda+1}(1-r^{\lambda+1})^2}{(x_{1,2}^{\lambda+1}x_{2,1}^{\lambda}+x_{2,2}^{\lambda+1}x_{1,1}^{\lambda})^2(x_{2,1}^{\lambda+1}x_{1,2}^{\lambda}+x_{2,2}^{\lambda+1}x_{1,1}^{\lambda})^2} < \frac{1}{2}$$

then the condition (2.12) will hold for $r > 1$. So under the condition (2.12) or (2.13), $(p_{1,+}^{(n)}, p_{+,1}^{(n)})$ will be convergent to the unique point $(p_{1,+}^{(\lambda)}, p_{+,1}^{(\lambda)})$ for any given initial value, and will satisfy (2.15) and (2.16). Therefore, by θ_λ and ϕ_λ, we can obatin $p_{1,+}^{(\lambda)}$ and $p_{+,1}^{(\lambda)}$, which are the Minimum Power-Divergence Estimates. □

If $x = (x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2})$ does not satisfy the conditions (2.12) and (2.13) for a given λ, then the linear equations $\theta = g_1(\theta,\phi)$ and $\phi = g_2(\theta,\phi)$ in (2.15) and (2.16) may have more than one solution. Using the following Iterative Procedure 2, we can obtain finitely many limits.

Iterative Procedure 2:

Step 0: Choose a small quantity ε. Choose a small initial value, say $\theta_1^{(0)}$=.0101 (i.e., $p_{1,+}$ = .01) and a large initial value, say $\theta_1^{(0)}$=99.0 ($p_{1,+}$ = .99), using the Iterative Procedure 1 to obtain limits a and b, respectively. If a> $b-\varepsilon$, stop this procedure. Otherwise let L=$\{a,b\}$ and I=$\{(a,b)\}$.

Step 1: Choose (a,b)∈I, and let I=I-$\{(a,b)\}$. Use the Iterative Procedure 1 with the initial value $\theta_3^{(0)}$=m=(a+b)/2. The procedure is terminated if either a limit c is reached or a value c lying outside the interval (a,b) is observed.

If c ≤a then I=I∪$\{(m,b)\}$, if m< $b-\varepsilon$.

If c≥b then I=I∪$\{(a,m)\}$, if a< $m-\varepsilon$.

If $a < c < m$ then L=L∪$\{c\}$, I=I∪$\{(a,c)\}$ if a< $c-\varepsilon$ and I=I∪$\{(m,b)\}$ if m< $b-\varepsilon$.

If $m < c < b$ then L=L∪$\{c\}$, I=I∪$\{(a,m)\}$ if a< $m-\varepsilon$, I=I∪$\{(c,b)\}$ if c< $b-\varepsilon$.

If m=c then L=L∪$\{m\}$, I=I∪$\{(a,m)\}$ if a< $m-\varepsilon$, I=I∪$\{(m,b)\}$ if m< $b-\varepsilon$.

Repeat step 1 until I is empty.

Step 2: The minimum power-divergence estimate is the limit in L which has the smallest value of the given statistic.

In the case of λ=-1, the power-divergence statistic is defined by the continuity of (2.4) as $\lambda \to -1$, and $I^{(-1)}$ has the following expression:

$$GM^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} Np_{i,+}p_{+,j}ln(\frac{Np_{i,+}p_{+,j}}{x_{i,j}}) \tag{2.17}$$

Differentiate GM^2 with respect to $p_{1,+}$ and $p_{+,1}$, and then set them equal to zero; thus we obtain:

$$p_{1,+} = \frac{\frac{x_{1,2}}{x_{2,2}}}{r_1^{p_{+,1}} + \frac{x_{1,2}}{x_{2,2}}} \tag{2.18}$$

$$p_{+,1} = \frac{\frac{x_{2,1}}{x_{2,2}}}{r_1^{p_{1,+}} + \frac{x_{2,1}}{x_{2,2}}} \tag{2.19}$$

here $r_1 = 1/r$, r is the odds ratio $\frac{x_{1,1}x_{2,2}}{x_{1,2}x_{2,1}}$.

By using the Iterative Procedure 1 through (2.18) and (2.19) for any initial value $p_{1,+}^{(0)}$, there is the series $(p_{1,+}^{(n)}, p_{+,1}^{(n)})$. By Theorem 2.5, the sufficient condition for the unique limit $(p_{1,+}^{(-1)}, p_{+,1}^{(-1)})$ is

$$\frac{\partial p_{1,+}}{\partial p_{+,1}} \times \frac{\partial p_{+,1}}{\partial p_{1,+}} = \frac{(lnr_1)^2 r_1^{(p_{1,+}+p_{+,1})} x_{2,1}x_{1,2}}{(r_1^{p_{+,1}} + \frac{x_{1,2}}{x_{2,2}})^2 (r_1^{p_{1,+}} + \frac{x_{2,1}}{x_{2,2}})^2 x_{2,2}^2}$$

$$\leq \frac{(lnr_1)^2}{4} < \frac{1}{2}$$

$$(lnr_1)^2 < 2 \tag{2.20}$$

Therefore, if $x=(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2})$ satisfies the condition (2.20), the minimum power-divergence estimate can be obtained by the Iterative Procedure 1.

If the condition (2.20) fails to satisfy, we can obtain all possible limits by using the Iterative Procedure 2.

Remark 2: For λ=-2 and -1/2, from (2.10) and (2.11), we can obtain a polynomial of five degrees and two degrees, respectively in θ, such as follows:

$$f(\theta) = t(\theta^2 + u)^2(\theta - v) + (\theta^2 + v)(\theta - u)$$

$u = x_{1,2}/x_{2,2}$, $v = x_{1,1}/x_{2,1}$ and $t=x_{1,1}/x_{1,2}$ for $\lambda = -2$ and

$$g(\theta) = a\theta^2 - (2a + b)\theta + a$$

$a=(\sqrt{x_{1,1}x_{2,1}}+\sqrt{x_{1,2}x_{2,2}})^2$ and $b=(x_{1,2}-x_{2,2}-x_{2,1}+x_{1,1})^2$ for λ =-1/2. The Iterative Procedure 2 can produce, at most, five limits for $\lambda = -2$, and two for λ=-1/2. In the simulated study, we found that there are, at most, three limits for λ=-2. In which case, one of the limits is a local maximun, while the other two limits are local minimums.

Example: We shall demonstrate the use of the above Iterative Procedure to the car accidents data in Chapter 1. It does not satisfy the conditions for the unique limit of the Iterative Procedure in Theoren 2.5, we shall use two endpoints as initial values. However, they converge to the single limit for each λ. Thererfore, the limits are the minimum power-divergence estimates. The convergent sequences $(p^n_{1,+}, p^n_{+,1})$ are provided in the Table 2.1.

Table 2.1: Example of the iterative procedure for the car accident data

$\lambda < 0$	λ	$p^0_{1,+}$	$p^1_{+,1}$	$p^1_{1,+}$	$p^2_{+,1}$	$p^2_{1,+}$	$p^3_{+,1}$	$p^3_{1,+}$
	-2	.99	.0097	.2962	.0017	.2831	.0016	.2831
		.01	.0013	.2829	.0016	.2831	.0016	.2831
	-1	.99	.0096	.2867	.0022	.2836	.0022	.2836
		.01	.0013	.2832	.0022	.2836	.0022	.2836
	-1/2	.99	.0037	.2845	.0037	.2844	.0037	.2844
		.01	.0037	.2844	.0037	.2844	.0037	.2844
$\lambda > 0$	2/3	.99	.0037	.2844	.0037	.2844	.0037	.2844
	1	.99	.0013	.2899	.0052	.2845	.0053	.2845

When $\lambda = 0$, the $p^{(\lambda)}_{1,+}$ and $p^{(\lambda)}_{+,1}$ of the MLE are $\frac{x_{1,+}}{N}$=.02844 and $\frac{x_{+,1}}{N}$=.0037, respectively. The Iterative Procedure 1 is efficient; normally, six to seven steps are sufficient to obtain the limit.

Chapter 3

Convergence Rates to the Null Limiting Distribution of the Power-Divergence Family of Statistics

3.1 The Criterions of Cochran and Yarnold

It has been established that the power-divergence family of statistics has limiting null chi-squared distribution with one degree of freedom for testing the hypothesis of independence in 2 by 2 contingency tables. Which chi-squared test in the family is the best has attracted a great deal of interest in literature. There is no uniformly preferable test.

The most important characteristic is the accuracy of the probability of the event $2NI^{\lambda}(\frac{x}{N} : p^{(\lambda)}) > \chi^2_{\alpha}(1)$ as compared to the significance level α. The accuracy depends on the sample size N. A small sample size usually means a less accurate test. However, a larger sample size means higher costs for experiments. Determination of an adequate minimum sample size has attracted many studies, and most have been devoted to the comparisons between X^2 and G^2.

It has long been known that the approximation to the chi-squared distribution for Pearson's X^2 statistic relies on the expected frequencies in each cell being large. Cochran (1952, 1954) provided a complete bibliography of the early discussions regarding this point, and stated (1952) that the approximation is acceptable if the exact power falls within the range of .04 to .06 for the .05 tabular value, and within the range of .007 to .015 for the .01 tabular value. He (1954) recommended the use of much smaller expectations by saying that goodness-of-fit tests of unimodal distributions (such as the normal or poisson): here the expectations will be small only at one or both cells. Group so that the minimum expectation at each cell is at least one.

Tate and Hyer (1973) stated, in a study of the accuracy of the chi-squared approximation for Pearson's X^2 test, that the chi-square probabilities of X^2 may differ, markedly, from the exact cumulative multinomial probabilities. In 1970, Yarnold stated that the X^2 approximation was originally derived under the assumption that all expectations are large. For this reason, many authors recommend that all expectations be at least five, and that neighboring classes be combined if this rule is not satisfied. Other authors recommend a minimum expectation of 10 or 20. Like the criterion of Cochran (1952), Yarnold (1970) presented a new rule with a wider range for acceptable approximations; the range being .0375-.06 for the .05 tabular value, and .006-.0162 for the .01 tabular value. He concluded that if the number of cells, k, is three or more, and if r denotes the number of expectations less than five, then the minimum expectation may be as small as 5r/k. So in the 2 by 2 contingency table, it becomes 5r/4.

3.2 Calculation of the Convergence Rates of the Six Statistics

Using power-divergence statistics tor testing independence in the 2 by 2 contingency table, it is important to study the convergence rate to the null distribution and the corresponding minimum sample size. Simulation studies are conducted for this investigation.

Due to the symmetry of the 2 by 2 contingency table, the values of the power-divergence statistics remain the same after changing the columns, the rows, or transposing the data matrix. It suffices to consider the case:

$$x_{1,2} \leq x_{2,1} \leq x_{2,2}, \text{and}, x_{1,1} \leq x_{2,2}.$$

Let $F_{E_\lambda}(t)$ be the exact distribution function of power-divergence statistics in (1.1) for a fixed λ and for a given $p = (p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}) \in H_o$, satisfying the hypothesis of independence. Let $F_{\chi^2(1)}(t)$ be the χ^2 distribution function with one degree of freedom. It follows, from the asymtotic eqivalence of the power-divergence statistics that

$$F_{E_\lambda}(t) = F_{\chi(1)^2}(t) + o(1) \tag{3.1}$$

holds for all t.

To calculate the F_{E_λ} for any given N ,λ, and $p \in H_o$, the following three steps are performed:

1. For every possible outcome $x = (x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2})$, calculate the exact probability Pr(X=x) as

$$Pr(X = x) = \frac{N!}{x_{1,1}!x_{1,2}!x_{2,1}!x_{2,2}!} p_{1,1}^{x_{1,1}} p_{1,2}^{x_{1,2}} p_{2,1}^{x_{2,1}} p_{2,2}^{x_{2,2}},$$

2. Compute the minimum power-divergence estimate $p^{(\lambda)}$ for every outcome x and the value of the statistic $2NI^{\lambda}(x/N : p^{(\lambda)})$.

3. The cumulative probability of the event $2NI^{\lambda} \leq t$ is the exact distribution $F_{E_\lambda}(t)$.

The accuracy of the approximation can be measured by the difference between 1-$F_{E_\lambda}(t_\alpha)$ and α, the size of the test, where t_α satisfies:

$$1 - F_{\chi^2(1)}(t_\alpha) = \alpha \tag{3.2}$$

The values of $1 - F_{E_\lambda}(t_\alpha)$ are provided in Table A.1 and Table A.2 for α=.05 and α=.01 respectively for the four different parameters p=$(p_{i,j})$, p$\in H_o$.

We shall introduce the notation to represent all possible configuration of the hypothesis of independence by $R_1 = p_{1,2}/p_{1,1}$ and $R_2 = p_{2,1}/p_{1,1}$. For example, the equiprobable hypothesis is represented by (R_1, R_2)=(1,1). Simulation study is conducted for different values of (R_1, R_2) to investigate the behaviours of the six statistics.

3.3 Discussion of the Convergence Rates and The Minimum Sample Sizes

Most often, the equiprobable hypothesis:

$$H_o : p_{i,j} = 1/4, i = 1, 2 \text{and}, j = 1, 2. \tag{3.3}$$

will be assumed for small-sample studies of goodness-of-fit. Read (1984) stated that there are three reasons: First, there have been various studies published indicating that equiprobable class intervals produce the most sensitive test (e.g. see Cohen and Sackrowitz 1975, Spruill 1977). Second, by applying the probability integral transformation, many goodness-of-fit problems are reduced to testing the fit of a uniform distribution on $[0,1]$. Finally, the power-divergence family of statistics is invariant to permutations in the observed frequencies x, when (3.3) is assumed. This greatly reduces the computations for calculating F_{E_λ}.

It is found from Table A.1 and Table A.2 that the commonly used minimum size N=20, for testing the hypothesis of equiprobabilities, is acceptable for $\lambda=1$, 2/3, and 0 at both test levels $\alpha=.05$ and .01 under the criterion of Cochran. The Pearson's X^2 ($\lambda=1$) seems to be the best among all six statistics. When λ decreases, the convergence rates are slow. Therefore, N=40 will become acceptable for λ =-1/2 and -1 under Cochran's criterion; the Neyman-modified MX^2 (λ=-2) requires that N=60.

The cases of (R_1, R_2)=(1,2), (1,4), and (1.5,1.5) have been also investigated, and the powers at the levels of .05 and .01 can also be found in Table A.1 and Table A.2.

As R_2 increases from 1 to 2, the minimum sample size N remains the same at 20, for λ = 1 and 2/3, under Cochran's criterion at both levels α=.05 and .01. For λ=0, -1/2, -1 and -2, the required minimum sample size N will be increased to 30, 40, 50 and 60 respectively.

Simulation studies have shown that with R_1 and/or R_2 increasing, the minimum sample size N will increase too; however, the convergence rates of λ=-1 and -2 are quite slow as compared to that of λ=1, 2/3 and 0.

Upon the above comparisons of the convergence rates and the minimum sample sizes, we found that the X^2 ($\lambda = 1$) and the Cressie-Read statistic ($\lambda = 2/3$) are the best for the four locations of the null hypotheses. It is interesting to note that the log-likelihood ratio statistic G^2 ($\lambda = 0$) does not perform as well as the Pearson's X^2 statistic ($\lambda = 1$). This result is supported by the small-sample studies of Larntz (1978) and Chapman (1976). The statistics MG^2 (λ=-1) and MX^2 (λ=-2) do not perform satisfactorily. This can be attributed to the fact that for $\lambda \leq -1$, the corresponding exact critical region will contain all possible x, with one or two random zeros in the opposite corner; since, in these cases, $2NI^{(\lambda)}(\frac{x}{N} : p^{(\lambda)})$ are ∞.

Chapter 4

The Powers of Tests under the Different Alternative Hypothesis

4.1 Calculation of the Powers of the Six Statistics

The advantage of considering various kinds of statistics in testing independence is that one may choose the most powerful test among them for a specific alternative. Therefore the next step is to compare the performances of these statistics. By the Neyman-Pearson theory of hypothesis testing, tests of the same sample size N are compared by determining their powers against relevant alternatives. If one of the six statistics has the greatest power against all alternatives, it is called the uniformly most powerful test; but no such test exists for testing independence in the 2 by 2 contingency table. Previously, many studies have been conducted in investigating the behaviour of the statistics under the different alternative hypothesis.

It is also known from Drost, Kallenberg, Moore and Oosterhoff (1989) that the powers of the statistics are approximated by a non-central chi-squared distribution with one degree of freedom for testing the independence in the 2 by 2 contingency table. The noncentrality parameter is $N\delta^{(\lambda)}$, where

$$\delta^{(\lambda)} = 2I^{(\lambda)}(q : q^{(\lambda)}) = inf_{q^*} 2I^{(\lambda)}(q : q^*) \tag{4.1}$$

The power of the statistics depends not only on their corresponding noncentrality parameter $\delta^{(\lambda)}$ but also on the location of q. This confirms Lee's argument (1987).

Testing the hypothesis of independence against the alternative q in the 2 by 2 table, the limiting distribution of any member of the power-divergence family of statistics under q is the noncentral chi-square with one degree of freedom, $\chi_1^2(N\delta^{(\lambda)})$. For example, the limiting distribution of Pearson's X^2 under the alternative q is $\chi_1^2(N\delta^{(1)})$ where

$$\delta^{(1)} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(q_{i,j} - q_{i,j}^{(1)})^2}{q_{i,j}^{(1)}}.$$

Simulative study has been conducted for the powers of the tests. Let $G_{E_\lambda}(t)$ be the exact distribution function of the power-divergence statistics for a fixed λ and a given alternative q. Three steps to calculate G_{E_λ}, as in Section 3.2, have been conducted with $p_{i,j}$, $p_{i,j}^{(\lambda)}$ and F_{E_λ} replaced by $q_{i,j}$, $q_{i,j}^{(\lambda)}$ and G_{E_λ} respectively.

We choose $t_{.05}^{(\lambda,N,q^{(\lambda)})}$ and $t_{.01}^{(\lambda,N,q^{(\lambda)})}$, which depend on λ, the sample size N and the alternative q (through $q^{(\lambda)}$) so that;

$$1 - F_{E_\lambda}(t_{.05}^{(\lambda,N,q^{(\lambda)})}) = .05$$

and

$$1 - F_{E_\lambda}(t_{.01}^{(\lambda,N,q^{(\lambda)})}) = .01.$$

where $F_{E_\lambda}(t)$ is the exact distribution of power-divergence statistics defined in Chapter 3. Then calculate the powers $P(t_{.05}^{(\lambda,N,q^{(\lambda)})})$ and $P(t_{.01}^{(\lambda,N,q^{(\lambda)})})$ such as:

$$P(t_{.05}^{(\lambda,N,q^{(\lambda)})}) = 1 - G_{E_\lambda}(t_{.05}^{(\lambda,N,q^{(\lambda)})})$$

and

$$P(t_{.01}^{(\lambda,N,q^{(\lambda)})}) = 1 - G_{E_\lambda}(t_{.01}^{(\lambda,N,q^{(\lambda)})})$$

for the six statistics with the sample size N=20, 40, 60, 80, until the power reaches approximately 90 percent.

4.2 Discussion of the Most Powerful Test

We compare the powers of the test statistics for testing $p_{i,j} = p_{i,+} * p_{+,j}$ at the following four alternative locations:

$$q_1 = (q_{11}, q_{12}, q_{21}, q_{22}) = (1/25, 8/25, 8/25, 8/25)$$

$$q_2 = (q_{11}, q_{12}, q_{21}, q_{22}) = (1/15, 1/15, 1/15, 12/15)$$

$$q_3 = (q_{11}, q_{12}, q_{21}, q_{22}) = (3/20, 1/20, 4/20, 12/20)$$

$$q_4 = (q_{11}, q_{12}, q_{21}, q_{22}) = (5/14, 2/14, 2/14, 5/14)$$

Therefore, for q_1, we investigate the powers of statistics when one of the alternative cell probabilities is very small, as compared to the other three equal cell probabilities; q_2 is just the opposite of q_1, one large cell probability and three small equal cell probabilities. For q_3, we investigate the case when q_{11} is greater than q_{12}, and q_{22} greater than q_{21}. For q_4, we investigate the case of two small equal cell probabiblities at the two diagonal corners of the table, and two large probabilities at the other diagonal corners.

The simulated powers for the above alternatives have been presented in Table B.1 and Table B.2 for the significance levels of α=.05 and .01. The $\delta^{(\lambda)}$ for different alternative hypotheses and different λ have also been provided in these tables. The $\delta^{(\lambda)}$ is calculated through (4.1) using the minimum power-divergence estimate $q^{(\lambda)}$.

From Table B.1 and Table B.2, it is found that against q_1, the Neyman Modified MX^2 ($\lambda = -2$) is the most powerful, and the power will increase when λ decreases, as long as N is sufficiently large, such as 60. With the alternative q_3, the log likelihood ratio test G^2 ($\lambda = 0$) is the most powerful; for the alternative q_2, the Pearson's chi-squared statistic X^2 ($\lambda = 1$) would be recommended, as it has the greatest power, and the power will decrease when λ decreases to -2. With the alternative q_4, there are no difference in the powers of these six statistics. For these alternatives and others not listed in the tables, it has been found that the larger the noncentrality parameters, the higher the simulated powers in most cases of the simulations. They confirm the conjecture that the powers will depend on their noncentrality parameters and the alternative q. It is found that there is no uniformly most powerful test.

Chapter 5

Discussion and Conclusion

Another method in calculating the values of the modified loglikelihood statistic MG^2 ($\lambda = -1$) and the Neyman-modified MX^2 ($\lambda = -2$), when there is a cell with zero frequency, is to replace zero by .5 and N by N+.5 to obtain the values of MG^2 and MX^2. Simulative studies have been conducted to investigate the performances of these two methods. The result of the studies indicates that there is no difference between these two methods, because when there is a cell with small frequencies in the table, the values of the above two statistics will always be very large.

In the simulation studies of the Chapter 3, it is found that the chi-squared approximations for the Pearson statistic X^2 (λ=1), the Cressie-Read statistic (λ=2/3), and the log likelihood ratio statistic (λ=0) are adequate for the sample size N=20, at the both .05 and .01 significance level, according to Cochran's (1952) and Yarnold's (1970) criterions for testing equiprobability of the null hypothesis. The Freeman-Tukey statistic T^2 can not be accepted at the .01 level, although it satisfies the criterions at the .05 level for N=20. The sample size N=16 has also been studied for this hypothesis, but none of the statistics could be acceptable for either test level and criterion.

For the unequal probabilities of the null hypothesis, the minimum sample size N will increase as the difference between R_1 and R_2 increases. From Table A.1 and Table A.2, it can be concluded that the difference between R_1 and R_2 will affect the minimum sample size much more for the loglikelihood ratio statistic G^2 ($\lambda=0$), the Freeman-Tukey statistic T^2 and the modified likelihhood statistic MG^2 ($\lambda = -1$) than the Pearson statistic X^2 ($\lambda = 1$) and the Cressie-Read statistic when $\lambda = 2/3$. The convergence rate for the Neyman-modified statistic MX^2 ($\lambda = -2$) is very slow, and it requires a sample size of at least N=60, even for the equal probabilities of the null hypothesis.

Therefore, the Pearson statistic X^2 is the best one for the convergence rate to the null limiting distribution of chi-square with one degree of freedom.

The powers of these test statistics, hen the null hypotheses are false, are simulated and presented in Table B.1 and Table B.2. We can see that, for different alternatives, each of the Pearson's X^2, the loglikelihood ratio statistic G^2 and the Neyman-modified MX^2 would be the most powerful, as we have discussed in Chapter 4. The powers depend on their noncentrality parameters and the alternative q. The Freeman-Tukey statistic T^2 and the modified likelihood statistic MG^2 are rarely used, as their performances are not as good as X^2 and G^2.

Because the Cressie-Read statistic, when $\lambda = 2/3$, appears to do well in most cases of the preceding discussion, it gives new competition for all of the other well known statistics. It should recieve more study and consideration in the future.

References

- Agresti, A. (1989). Categorical Data Analysis. John Wiley and Sons.
- Lerkson, J. (1980). Minimum chi-square ; not maximum likelihood! *Annals of Statistics* 8, *pp*.457 − 487.
- Birch, W. C. (1964). A new proof of the Pearson-Fisher theorem. *Annals of Mathematical Statistics* 35, *pp*.817 − 824.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA, The MIT press.
- Bohning, D. and Holling, H. (1986). On minimizing chi-square distances under the hypothesis of Homogeneity or independence for a two-way contingency table. *Statistics and Probability Letters* 4. *pp*.253 − 258.
- Chapman, J. A. W. (1976) A comparison of the X^2, -2LogR, and Multinomial probability criteria for significance tests when expected frequencies are small. *Journal of American Statistical Association* 71, *pp*.854 − 863.
- Clogg, C. C. and Shockey, J. W. (1988). Multivariate analysis of discrete data. *Handbook of Multivariate Experimental Psychology*. ed. by J.R.Nesselroade and R.B.Cattell. New York: Plenum Press.
- Cochran, W. G. (1936). The χ^2 distribution for the binomial and poisson series, with small expectations. *Annals of Eugenics*, Vol. 7, *pp*.207 − 217
- Cochran, W. G. (1952). The chi^2 test of goodness-of-fit. *Annals of Mathematical Statistics*, 23, *PP*.315 − 345.
- Cochran. W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, *pp*.417 − 451
- Cohen, A. and Sackrowitz, H. B. (1975). Unbiasedness of the chi-square, likelihood ratio, and other goodness of fit tests for the equal cell case. *Annals of Statistics* 3, *pp*.959 − 964.
- Cohen, V. (1982) A chi-square minimum method applied to contingency tables. *Proceedings of the fifth Compstat-Congress, part 2. pp*.55 − 56.
- Cressie, N. and Read, T.C. (1984). Multinomial goodness-of-fit tests. *J. Royal Stat. Soc. Ser. B pp*.440 − 464.
- Doll, R. and Hill, A. B. (1952) A study of the aetiology of carcinoma of the lung. *British. Med. J. 2 pp*.1271 − 1286.
- Drost, F. C., Kallenberg, W. C. M., Moore, D. S. and Oosterhoff, J. (1989). Power approximations to multinomial tests of fit. *Journal of the American Statistical Association*. 84, *pp*.130 − 141.

- Fisher, R. A. (1922). On the interpretation of chi square from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society* 87, *pp*.442 – 450.

- Fisher, R. A. (1924). The conditions under which chi-square measures the discrepancy between observation and hypothesis. *Journal of Royal Statistical Society*, 87. *pp*.442 – 450.

- Fisher, R. A. (1950). The significance of deviations from expectation in a poisson series. *Biometrics*, 6. *pp*.17 – 24.

- Freeman, M.F., and Tukey, J.W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics* 21, *pp*.607 – 611.

- Harris, R. R. and Kanji, G. K. (1983). On the use of minimum chi-square estimation. *The Statistician*, 32. *pp*.379 – 394.

- Haberman, S. J. (1978,1979). Analysis of Qualitative Data. *vols. 1 and 2.* New York: Academic Press.

- Holland, P. W. (1967). A variation on the minimum chi-square test. *J. Math. Psychol. 4*. *pp*.377 – 413.

- Kallenberg, W. C. M., Oosterhoff, J., and Schriever, B. F. (1985). The number of classes in chi-squared goodness-of-fit tests. *Journal of the American Statistical Association* 80. *pp*.959 – 968.

- Kroll, N.E.A. (1989) Testing independence in 2 by 2 contingency tables. *Journal of Educational Statistics* 1989. *pp*.35 – 46.

- Kullback, S. (1959). Information Theory and Statistics. New York, John Wiley.

- Kullback, S. (1985). Minimum discrimination information (MDI) estimation. In *Encyclopedia of Statistical Sciences, Volume 5 pp*.527 – 529. New York. Jphn Wiley.

- Larntz, K. (1978). Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of American Statitical Association*, 73. *pp*.253 – 263.

- Lee, C. C. (1987). Chi-squared tests for and against an order restriction on multinomial parameters. *Journal of American Statistical Association*, 82. *pp*.611 – 618.

- Lloyd, C. E. (1988) Some issues arising from the analysis of 2 by 2 contingency tables. *The Australian Journal of Statistics*, *pp*.35 – 46.

- Mitra, S. K. (1958). On the limiting power function of the frequency chi-square test. *Annals of Mathematical Statistics* 29, *pp*.1221 – 1233.

- Moore, D. S. (1978). Chi-square tests. in studies in Statistics, ed. R. V. Hogg, Washington, DC: Mathematical Association of America, *pp*.66 – 106.

- Moore, D. S. (1984). Measures of Lack of Fit from Tests of Chi-Squared Type. *Journal of Statistical Planning and Inference 10. pp*.151 – 166.

- Neyman, J. (1949). Contribution to the theory of the chi-square. *Proceedings of the First Berkeley Symposium. pp*.239 – 273.

- Norusis, M.J. (1988). SPSS-X Advanced Statistics Guide, 2nd edn. New York: McGraw-Hill.

- Ortega, J. M., (1972). *Numerical Analysis: A second course*. Academic Press, New York.

- Parr, W. C. (1981). Minimum distance estimation: a bibliography. *Communications in Statitics - theory and methods*, 10. *pp*.1205 – 1224.

- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine*, 50. *pp*.157 – 172.

- Rao, C. R. (1961). Asymptotic efficiency and limiting information. *Proceedings of the fourth Berkely Symposium on Mathematical Statistics and Probability* 1. *pp*.537 – 548.

- Rao, C. R. (1962). Sfficient estimates and optimum inference procedures in large samples. *Journal of the Royal Statistical Society series B*, 24. *pp*.46 – 72.

- Read, T. R. C. (1984). Small-Sample Comparisons for the power divergence goodness-of-fit statistics. *Journal of American Statistical Association*, 79. *pp*.929 – 935.

- Read, T. R. C. and Cressie, N. A. C. (1988). Goodness-of-fit statistics for discrete multivariate data. Springer-verlag.

- Rockafellar, R. T. (1972). Convex Analysis. Princeton University Press.

- Rudas, T. (1986). A Monte Carlo Comparison of the Small Sample Behaviour of the Pearson, the Likelihood Ratio and the Cressie-read statistics. *J. Stat. Comput. Simul. Vol.24. pp*.107 – 120.

- Spruill,C. (1977). Equally likely intervals in the chi-square test. *Sankhya Series A* 39. *pp*.299 – 302.

- Tate, M. W. and Hyer, L. A. (1973). Inaccuracy of the X^2 test of goodness-of-fit when expected frequencies are small. *Journal of the American Statistical Association* 68. *pp*.836 – 841.

- West, E. N. and Kempthorne, O. (1971). A comparison of the X^2 and G^2 for composite alternative. *Journal of Statistical Computation and Simulation*, 1. *pp.*1 – 33.

- Yarnold, J. K. (1970). The minimum expectation in χ^2 goodness-of-fit tests and the accuracy of approximations for the null distribution. *Journal of American Statistical association*, 65. *pp.*864 – 886.

- Yarnold, J. K. (1972). Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Annals of Mathematical Statistics*, 43. *pp.*1566 – 1580.

Appendices

Appendix A

Tables of upper tail probabilities of power-divergence statistics under H_o: $p_{i,j} = p_{i,+} * p_{+,j}$

Table A.1: Table of upper tail probabilities for α=.05

(R_1, R_2)	λ	N					
		20	30	40	50	60	70
(1, 1)	1	.051*	.050	.051	.057	.052	.051
	2/3	.052*	.050	.051	.057	.052	.051
	0	.059*	.055	.051	.059	.053	.052
	-1/2	.058*	.055	.054	.059	.054	.053
	-1	.061	.060*	.056	.059	.056	.054
	-2	.080	.073	.065	.064	.057*	.056
(1, 2)	1	.049*	.050	.052	.053	.051	.050
	2/3	.052*	.053	.054	.053	.051	.051
	0	.075	.060*	.058	.057	.054	.053
	-1/2	.081	.060*	.059	.058	.056	.054
	-1	.055	.065	.061	.059*	.058	.056
	-2	.065	.071	.066	.062	.060*	.058
(1, 4)	1	.036	.040*	.046	.047	.046	.048
	2/3	.041*	.051	.051	.052	.052	.052
	0	.089	.078	.068	.062	.058*	.057
	-1/2	.131	.101	.079	.065	.058*	.057
	-1	.032	.058	.065	.065	.064	.060*
	-2	.027	.047	.059	.060	.061	.058*
(1.5, 1.5)	1	.049*	.050	.052	.052	.050	.051
	2/3	.052*	.051	.054	.052	.052	.051
	0	.069	.058*	.056	.056	.052	.051
	-1/2	.073	.058*	.056	.056	.052	.051
	-1	.056	.063	.060*	.058	.058	.056
	-2	.069	.071	.063	.061	.059*	.057

* Minimum sample size under Cochran's criterion

Table A.2: Table of upper tail probabilities for α=.01

(R_1, R_2)	λ	N					
		20	30	40	50	60	70
(1, 1)	1	.011*	.010	009	.089	.010	.010
	2/3	.012*	.011	.010	.010	.011	.010
	0	.015*	.014	.012	.011	.012	.011
	-1/2	.024	.016	.015*	.013	.013	.012
	-1	.020	.017	.016	.014*	.013	.012
	-2	.015	.018	.018	.017	.015*	.014
(1, 2)	1	.008*	.008	.009	.009	.009	.010
	2/3	.010*	.010	.010	.010	.010	.010
	0	.017	.015*	.012	.012	.011	.011
	-1/2	.038	.019	.015*	.013	.013	.012
	-1	.015	.016	.016	.014*	.014	.013
	-2	.010	.016	.017	.016	.014*	.014
(1, 4)	1	.003	.004	.006*	.007	.007	.007
	2/3	.005	.005	.008*	.008	.009	.009
	0	.014	.018	.017	.015*	.013	.012
	-1/2	.050	.040	.028	.020	.016	.015*
	-1	.006	.009	.014	.015	.016	.015*
	-2	.004	.008	.014	.016	.016	.015*
(1.5, 1.5)	1	.009*	.009	.009	.009	.010	.010
	2/3	.010*	.010	.010	.010	.010	.010
	0	.015*	.014	.011	.012	.011	.011
	-1/2	.032	.018	.015*	.013	.012	.011
	-1	.017	.016	.016	.014*	.014	.013
	-2	.011	.017	.017	.016	.015*	.014

* Minimum sample size under Cochran's criterion

Appendix B

Tables of the simulated powers of the power-divergence statistics for four different alternatives

Table B.1: The simulated powers for N=20,40,60 and 80 under $\alpha=.05$

$H_a:(q_{11},q_{12},q_{21},q_{22})$	λ	N				$\delta^{(\lambda)}$
		20	40	60	80	
(1/25,8/25,8/25,8/25)	1	.401	.720	.884	.943	.146
	2/3	.415	.732	.897	.958	.153
	0	.441	.753	.907	.964	.168
	-1/2	.458	.756	.910	.968	.181
	-1	.402	.762	.911	.973	.195
	-2	.428	.770	.928	.985	.225
(1/15,1/15,1/15,12/15)	1	.433	.675	.859	.952	.147
	2/3	.431	.663	.853	.948	.143
	0	.428	.661	.839	.932	.130
	-1/2	.408	.517	.646	.801	.111
	-1	.422	.638	.834	.909	.090
	-2	.420	.603	.763	.892	.058
(3/20,1/20,4/20,12/20)	1	.440	.701	.854	.932	.163
	2/3	.436	.700	.859	.938	.166
	0	.483	.742	.886	.956	.170
	-1/2	.447	.712	.849	.910	.168
	-1	.342	.686	.786	.874	.159
	-2	.241	.563	.701	.811	.126
(5/14,2/14,2/14,5/14)	1	.453	.619	.881	.986	.184
	2/3	.451	.620	.882	.986	.185
	0	.452	.620	.882	.987	.190
	-1/2	.431	.617	.880	.984	.195
	-1	.439	.617	.880	.979	.203
	-2	.434	.617	.878	.975	.225

Table B.2: The simulated powers for N=20,40,60 and 80 under α=.01

$H_a : (q_{11}, q_{12}, q_{21}, q_{22})$	λ	N				$\delta^{(\lambda)}$
		20	40	60	80	
(1/25,8/25,8/25,8/25)	1	.150	.447	.712	.830	.146
	2/3	.150	.466	.724	.837	.153
	0	.200	.501	.742	.846	.168
	-1/2	.202	.497	.750	.852	.181
	-1	.168	.524	.757	.856	.195
	-2	.193	.538	.775	.880	.225
(1/15,1/15,1/15,12/15)	1	.242	.504	.718	.834	.147
	2/3	.241	.490	.716	.826	.143
	0	.250	.472	.672	.782	.130
	-1/2	.211	.367	.450	.624	.111
	-1	.223	.447	.652	.814	.090
	-2	.224	.409	.520	.735	.058
(3/20,1/20,4/20,12/20)	1	.224	.483	.621	.792	.163
	2/3	.228	.494	.632	.797	.166
	0	.256	.515	.658	.845	.170
	-1/2	.267	.463	.601	.766	.167
	-1	.152	.416	.582	.723	.159
	-2	.101	.201	.484	.668	.126
(5/14,2/14,2/14,5/14)	1	.267	.423	.704	.862	.184
	2/3	.237	.423	.706	.861	.185
	0	.231	.422	.706	.863	.190
	-1/2	.203	.421	.705	.859	.195
	-1	.250	.417	.701	.855	.203
	-2	.235	.420	.698	.849	.225