Symmetry **2010**, *2*, 1108-1120; doi:10.3390/sym2021108

OPEN ACCESS

symmetry

ISSN 2073-8994

www.mdpi.com/journal/symmetry

Review

Minimum Phi-Divergence Estimators and Phi-Divergence Test Statistics in Contingency Tables with Symmetry Structure: An Overview

Leandro Pardo [1,⋆] **and Nirian Martín** [2]

[1] Department of Statistics and O.R., Complutense University of Madrid, 28040 Madrid, Spain

[2] Department of Statistics, Carlos III University of Madrid, 28903 Getafe (Madrid), Spain

⋆ Author to whom correspondence should be addressed; E-Mail: lpardo@mat.ucm.es.

Received: 12 March 2010; in revised form: 7 May 2010 / Accepted: 10 June 2010 / Published: 11 June 2010

Abstract: In the last years minimum phi-divergence estimators (MϕE) and phi-divergence test statistics (ϕTS) have been introduced as a very good alternative to classical likelihood ratio test and maximum likelihood estimator for different statistical problems. The main purpose of this paper is to present an overview of the main results presented until now in contingency tables with symmetry structure on the basis of (MϕE) and (ϕTS).

Keywords: minimum phi-divergence estimator; phi-divergence statistics; symmetry model

Classification: MSC 62B10, 62H15

---

1. Introduction

An interesting problem in a two-way contingency table is to investigate whether there are symmetric patterns in the data: Cell probabilities on one side of the main diagonal are a mirror image of those on the other side. This problem was first discussed by Bowker [1] who gave the maximum likelihood estimator as well as a large sample chi-square type test for the null hypothesis of symmetry. The minimum discrimination information estimator was proposed in [2] and the minimum chi-squared estimator in [3]. In [4–7] new families of test statistics, based on ϕ-divergence measures, were introduced. These families contain as a particular case the test statistic given by [1] as well as the likelihood ratio test.

Let X and Y denote two ordinal response variables, X and Y having I levels. When we classify subjects on both variables, there are I^2 possible combinations of classifications. The responses (X, Y) of a subject randomly chosen from some population have a probability distribution. Let $p_{ij} = \Pr(X = i, Y = j)$, with $p_{ij} > 0$, $i, j = 1, ..., I$. We display this distribution in a rectangular table having I rows for the categories of X and I columns for the categories of Y. Consider a random sample of size n on (X, Y) and we denote by n_{ij} the observed frequency in the (i, j)th cell for $(i, j) \in I \times I$ with $\sum_{i=1}^{I} \sum_{j=1}^{I} n_{ij} = n$. The classical problem of testing for symmetry is given by

$$H_0 : p_{ij} = p_{ji}, \quad (i, j) \in I \times I \tag{1}$$

versus

$$H_1^* : p_{ij} \neq p_{ji}, \text{ for at least one } (i, j) \text{ pair} \tag{2}$$

This problem was considered for the first time by Bowker [1] using the Pearson test statistic

$$X^2 = \sum_{i=1}^{I} \sum_{\substack{j=1 \\ i<j}}^{I} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \tag{3}$$

for which established that $X^2 \sim \chi_k^2$ for large n, where $k = \frac{1}{2} I(I - 1)$.

In some real problems (*i.e.*, medicine, psychology, sociology, *etc.*) the categorical response variables (X, Y) represent the measure after or before a treatment. In such situations our interest is to determine the treatment effect, *i.e.*, if $X \geq Y$ (we assume that X represents the measure after the treatment and Y before the treatment). In the following we understand that X is preferred or indifferent to Y, according to joint likelihood ratio ordering, if and only if (iff) $p_{ij} \geq p_{ji}$ $\forall i \geq j$. In this situation the alternative hypothesis is

$$H_1 : p_{ij} \geq p_{ji}, \text{ for all } i \geq j \tag{4}$$

This problem was first considered by El Barmi and Kochar [8], who presented the likelihood ratio test for the problem of testing

$$H_0 : p_{ij} = p_{ji} \text{ against } H_1 : p_{ij} \geq p_{ji}, \forall i \geq j$$

and considered the application of it to a real life problem: They tested if the vision of both the eyes, for 7477 women, is the same against the alternative that the right eye has better vision than the left eye. In [5] these results were extended using ϕ-divergence measures.

In this paper we present an overview on contingency tables with symmetry structure on the basis of divergence measures. We pay especial attention to the family of ϕ-divergence test statistics for testing H_0 versus H_1^*, H_0 against H_1 and also for testing H_1 against the alternative H_2 of no restrictions over p_{ij}'s, *i.e.*,

$$H_1 : p_{ij} \geq p_{ji}, \forall i \geq j, \text{ against } H_2 : p_{ij} \ngeq p_{ji}, \forall i \geq j \tag{5}$$

It is interesting to observe that not only we consider ϕ-divergence test statistics but also we consider minimum ϕ-divergence estimators in order to estimate of the parameters of the model.

2. Phi-divergences Measures

We consider the set

$$\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (p_{ij};\ 1 \leq i, j \leq I,\ (i,j) \neq (I,I)) \text{ with } p_{ij} > 0 \text{ and } \textstyle\sum_{i=1}^{I}\sum_{j=1,(i,j)\neq(I,I)}^{I} p_{ij} < 1\} \quad (6)$$

and we denote $\boldsymbol{p}(\boldsymbol{\theta}) = \boldsymbol{p} = (p_{11}, ..., p_{II})^T$, $p_{II} = 1 - \sum_{i=1}^{I}\sum_{j=1,(i,j)\neq(I,I)}^{I} p_{ij}$ or equivalently by the $I \times I$ matrix $\boldsymbol{p}(\boldsymbol{\theta}) = [p_{ij}]$.

The ϕ-divergence between two probability distributions $\boldsymbol{p} = [p_{ij}]$, $\boldsymbol{q} = [q_{ij}]$ was introduced independently by [9] and [10]. It is defined as follows:

$$D_\phi(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{I}\sum_{j=1}^{I} q_{ij}\phi\left(\frac{p_{ij}}{q_{ij}}\right), \quad \phi \in \Phi^* \quad (7)$$

where Φ^* is the class of all convex functions $\phi : [0, \infty) \to \mathbb{R} \cup \{\infty\}$, such that $\phi(1) = 0$, $\phi''(1) > 0$; and we define $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u\to\infty} \phi(u)/u$. For every $\phi \in \Phi^*$ that is differentiable at $x = 1$, the function ψ given by

$$\psi(x) = \phi(x) - \phi'(1)(x - 1)$$

also belongs to Φ^*. Then we have $D_\psi(\boldsymbol{p}, \boldsymbol{q}) = D_\phi(\boldsymbol{p}, \boldsymbol{q})$, and ψ has the additional property that $\psi'(1) = 0$. Because the two divergence measures are equivalent, we can consider the set Φ^* to be equivalent to the set

$$\Phi \equiv \Phi^* \cap \{\phi : \phi'(1) = 0\}$$

An important family of ϕ-divergences in statistical problems, is the power divergence family

$$\begin{aligned} \phi_{(\lambda)}(x) &= \frac{x^{\lambda+1} - x + \lambda(1-x)}{\lambda(\lambda+1)};\ \lambda \neq 0, \lambda \neq -1 \quad (8) \\ \phi_{(0)}(x) &= \lim_{\lambda\to 0} \phi_{(\lambda)}(x) = x \ln x - x + 1 \\ \phi_{(-1)}(x) &= \lim_{\lambda\to -1} \phi_{(\lambda)}(x) = -\ln x + x - 1 \end{aligned}$$

which was introduced and studied by [11]. Notice that $\phi_{(\lambda)} \in \Phi$. In the following we shall denote the power-divergence measures by $D_{\phi_{(\lambda)}}(\boldsymbol{p}, \boldsymbol{q})$, $\lambda \in \mathbb{R}$. For more details about ϕ-divergence measures see [12].

3. Hypothesis Testing: H_0 versus H_1^*

We define

$$B = \{(a_{11}, ..., a_{1I},\ a_{22}, ..., a_{2I}, ..., a_{I-1I-1}, a_{I-1I})^T \in \mathbb{R}_+^{\frac{I(I+1)}{2}-1} : \textstyle\sum_{i\leq j} a_{ij} < 1, i, j = 1, .., I\}$$

the hypothesis (1) can be written as

$$H_0 : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\beta}), \quad \boldsymbol{\beta} = (p_{11}, ..., p_{1I}, p_{22}, ..., p_{2I}, ..., p_{I-1I-1}, p_{I-1I})^T \in B \quad (9)$$

where the function $\boldsymbol{\beta}$ is defined by $\boldsymbol{\beta} = (g_{ij}; i, j = 1, ..., I, (i, j) \neq (I, I))$ with

$$g_{ij}(\boldsymbol{\beta}) = \begin{cases} p_{ij}, & i \leq j \\ p_{ji}, & i > j \end{cases}, \qquad i, j = 1, ..., I-1$$

and

$$g_{Ij}(\boldsymbol{\beta}) = p_{jI},\ j = 1, ..., I-1$$
$$g_{iI}(\boldsymbol{\beta}) = p_{iI},\ i = 1, ..., I-1$$

Note that $\boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})) = (g_{ij}(\boldsymbol{\beta}); i, j = 1, ..., I)^T$, where

$$g_{II}(\boldsymbol{\beta}) = 1 - \textstyle\sum_{i,j=1,(i,j)\neq(I,I)}^{I} g_{ij}(\boldsymbol{\beta})$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ can be defined as

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in B} D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))) \text{ a.s.}$$

where $D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})))$ is the Kullback–Leibler divergence measure (see [13,14]) defined by

$$D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))) = \sum_{i=1}^{I}\sum_{j=1}^{I} \widehat{p}_{ij} \log \frac{\widehat{p}_{ij}}{g_{ij}(\boldsymbol{\beta})}$$

We denote by $\widehat{\boldsymbol{\theta}} = \boldsymbol{g}(\widehat{\boldsymbol{\theta}})$ and by $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}) = (p_{11}(\widehat{\boldsymbol{\theta}}), ..., p_{II}(\widehat{\boldsymbol{\theta}}))^T$. It is well known that $p_{ij}(\widehat{\boldsymbol{\theta}}) = \frac{\widehat{p}_{ij}+\widehat{p}_{ji}}{2}$, $i = 1, ..., I, j = 1, ..., I$. Using the ideas developed in [15], we can consider the minimum ϕ_2-divergence estimator ($M\phi_2E$) replacing the Kullback–Leibler divergence by a ϕ_2-divergence measure in the following way

$$\widehat{\boldsymbol{\theta}}^{\phi_2} = \arg\min_{\boldsymbol{\beta}\in B} D_{\phi_2}((\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})))); \quad \phi_2 \in \Phi^* \tag{10}$$

where

$$D_{\phi_2}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))) = \sum_{i=1}^{I}\sum_{j=1}^{I} g_{ij}(\boldsymbol{\beta})\phi_2\left(\frac{\widehat{p}_{ij}}{g_{ij}(\boldsymbol{\beta})}\right)$$

We denote $\widehat{\boldsymbol{\theta}}^{S,\phi_2} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{\phi_2})$ and we have (see [7,16])

$$\sqrt{n}(\boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{\phi_2}) - \boldsymbol{\beta}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_F^S(\boldsymbol{\beta})^{-1}\right)$$

where

$$\boldsymbol{I}_F^S(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}\boldsymbol{B}(\boldsymbol{\theta})^T(\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^T\boldsymbol{B}(\boldsymbol{\theta}))^{-1}\boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$$

being $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \mathrm{diag}(\boldsymbol{\theta}) - \boldsymbol{\theta}\boldsymbol{\theta}^T$ and $\boldsymbol{B}(\boldsymbol{\theta}) = \left(\frac{\partial h_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{ij}}\right)_{\frac{I(I-1)}{2}\times(I^2-1)}$. The functions h_{ij} are given by

$$h_{ij}(\boldsymbol{\theta}) = p_{ij} - p_{ji}, \qquad i < j, i = 1, ..., I-1, j = 1, ..., I$$

It is not difficult to establish that the matrix $\boldsymbol{I}_F^S(\boldsymbol{\theta})$ can be written as

$$\boldsymbol{I}_F^S(\boldsymbol{\theta}) = \boldsymbol{M}_{\boldsymbol{\beta}}^T\boldsymbol{I}_F(\boldsymbol{\beta})^{-1}\boldsymbol{M}_{\boldsymbol{\beta}}$$

where $\boldsymbol{I}_F(\boldsymbol{\beta})$ is the Fisher information matrix corresponding to $\boldsymbol{\beta} \in B$.

If we consider the family of power divergences we get the *minimum power-divergence estimator,* $\widehat{\boldsymbol{\theta}}^{S,\lambda}$ *of* $\boldsymbol{\theta}$, under the hypothesis of symmetry, whose expression is given by

$$\widehat{\theta}_{ij}^{S,\lambda} = \frac{\left(\dfrac{\widehat{p}_{ij}^{\lambda+1} + \widehat{p}_{ji}^{\lambda+1}}{2}\right)^{\frac{1}{\lambda+1}}}{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{I}\left(\dfrac{\widehat{p}_{ij}^{\lambda+1} + \widehat{p}_{ji}^{\lambda+1}}{2}\right)^{\frac{1}{\lambda+1}}}, \qquad i,j = 1,...,I \tag{11}$$

For $\lambda = 0$ we get

$$\widehat{\theta}_{ij}^{S,0} = \frac{\widehat{p}_{ij} + \widehat{p}_{ji}}{2}, \; i,j = 1,...,I$$

hence, we obtain the maximum likelihood estimator for symmetry introduced by [1]. For $\lambda = -1$, we obtain as a limit case

$$\widehat{\theta}_{ij}^{S,-1} = \frac{(\widehat{p}_{ij}\widehat{p}_{ji})^{\frac{1}{2}}}{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{I}(\widehat{p}_{ij}\widehat{p}_{ji})^{\frac{1}{2}}}, \; i,j = 1,...,I$$

i.e., the minimum discrimination estimator for symmetry introduced and studied in [2]. For $\lambda = 1$ we get the minimum chi-squared estimator for symmetry introduced in [3],

$$\widehat{\theta}_{ij}^{S,1} = \frac{\left(\dfrac{\widehat{p}_{ij}^{2} + \widehat{p}_{ji}^{2}}{2}\right)^{1/2}}{\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{I}\left(\dfrac{\widehat{p}_{ij}^{2} + \widehat{p}_{ji}^{2}}{2}\right)^{1/2}}$$

We denote $\widehat{\boldsymbol{\theta}}^{\phi_2} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}}^{\phi_2})$ and by

$$\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2}) = (p_{11}(\widehat{\boldsymbol{\theta}}^{\phi_2}), ..., p_{II}(\widehat{\boldsymbol{\theta}}^{\phi_2}))^T \tag{12}$$

the $(M\phi_2 E)$ of the probability vector that characterizes the symmetry model. Based on $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ it is possible to define a new family of statistics for testing (1) that contains as a particular case Pearson test statistic as well as likelihood ratio test. This family of statistics is given by

$$T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2}) \equiv \frac{2n}{\phi_1''(1)} D_{\phi_1}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{\phi_2})) = \frac{2n}{\phi_1''(1)} \sum_{i=1}^{I}\sum_{j=1}^{I} p_{ij}(\widehat{\boldsymbol{\theta}}^{\phi_2})\phi_1\left(\frac{\widehat{p}_{ij}}{p_{ij}(\widehat{\boldsymbol{\theta}}^{\phi_2})}\right) \tag{13}$$

We can observe that the family (13) involves two functions ϕ_1 and ϕ_2, both belonging to Φ^*. We use the function ϕ_2 to obtain the $(M\phi_2 E)$ and ϕ_1 to obtain the family of statistics. If we consider $\phi_1(x) = \frac{1}{2}(x-1)^2$ and $\phi_2(x) = x\log x - x + 1$ we get Pearson test statistic whose expression was given in (3) and for $\phi_1(x) = \phi_2(x) = x\log x - x + 1$ we get the likelihood ratio test given by

$$G^2 = 2\sum_{i=1}^{I}\sum_{\substack{j=1\\ i<j}}^{I} n_{ij}\log\frac{2n_{ij}}{n_{ji} + n_{ij}} \tag{14}$$

In the following theorem the asymptotic distribution of $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ is obtained.

Theorem 1 *The asymptotic distribution of $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ is chi-squared with $m = I(I-1)/2$ degrees of freedom.*

Proof. See Chapter 8 in [12].

Thus, for a given significance level $\alpha \in (0,1)$, the critical value of $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ may be approximated by $\chi^2_{m,\alpha}$, the upper $100\alpha\%$ of the chi-square distribution with m degrees of freedom, *i.e.*, reject the hypothesis of symmetry iff

$$T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2}) \geq \chi^2_{m,\alpha} \tag{15}$$

Now we are going to analyze the power of the test. Let $\boldsymbol{q} = (q_{11}, ..., q_{II})^T$ be a point at the alternative hypothesis, *i.e.*, there exist at least two indexes i and j for which $q_{ij} \neq q_{ji}$. We denote by $\boldsymbol{\theta}_a^{\phi_2}$ the point on Θ verifying

$$\boldsymbol{\theta}_a^{\phi_2} = \arg\min_{\boldsymbol{\theta}\in\Theta_0} D_{\phi_2}(\boldsymbol{q}, \boldsymbol{p}(\boldsymbol{\theta}))$$

where Θ_0 is given by

$$\Theta_0 = \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\beta}) \text{ for some } \boldsymbol{\beta} \in B\}$$

It is clear that

$$\boldsymbol{\theta}_a^{\phi_2} = (f_{ij}(\boldsymbol{q});\ i,j = 1, ..., I,\ \ (i,j) \neq (I,I))^T$$

and

$$\boldsymbol{p}(\widetilde{\boldsymbol{\theta}}_a^{\phi_2}) = (f_{ij}(\boldsymbol{q});\ i,j = 1, ..., I,\ )^T \equiv \boldsymbol{f}(\boldsymbol{q})$$

with

$$f_{II}(\boldsymbol{q}) = 1 - \sum_{i=1}^{I} \sum_{\substack{j=1 \\ (i,j)\neq(I,I)}}^{I} f_{ij}(\boldsymbol{q})$$

The notation $f_{ij}(\boldsymbol{q})$ indicates that the elements of the vector $\boldsymbol{\theta}_a^{\phi_2}$ depend on $\mathbf{q}$. For instance, for the power-divergence family $\phi_{(\lambda)}(x)$ we have

$$f_{ij}(\boldsymbol{q}) = \frac{\left(q_{ij}^{\lambda+1} + q_{ji}^{\lambda+1}\right)^{\frac{1}{\lambda+1}}}{\sum_{i=1}^{I}\sum_{j=1}^{I}\left(q_{ij}^{\lambda+1} + q_{ji}^{\lambda+1}\right)^{\frac{1}{\lambda+1}}}, \qquad i,j = 1, ..., I$$

We also denote

$$\widehat{\boldsymbol{\theta}}^{S,\phi_2} = (p_{ij}^{S,\phi_2};\ i,j = 1, ..., I,\ \ (i,j) \neq (I,I))^T$$

and then

$$\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{S,\phi_2}) = (p_{ij}^{S,\phi_2};\ i,j = 1, ..., I)^T \equiv \boldsymbol{f}(\widehat{\boldsymbol{p}})$$

where $\boldsymbol{f} = (f_{ij};\ i,j = 1, ..., I)^T$. If the alternative $\boldsymbol{q}$ is true we have that $\widehat{\boldsymbol{p}}$ tends to $\boldsymbol{q}$ and $\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^{S,\phi_2})$ to $\boldsymbol{p}(\boldsymbol{\theta}_a^{\phi_2})$ in probability.

If we define the function

$$\Psi_{\phi_1}(\boldsymbol{q}) = D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{q}))$$

we have

$$\Psi_{\phi_1}(\widehat{\boldsymbol{p}}) = \Psi_{\phi_1}(\boldsymbol{q}) + \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\partial D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{q}))}{\partial q_{ij}} (\widehat{p}_{ij} - q_{ij}) + o\left(\|\widehat{\boldsymbol{p}} - \boldsymbol{q}\|\right)$$

Then the random variables

$$\sqrt{n}\left(D_{\phi_1}(\widehat{\boldsymbol{p}}, \boldsymbol{f}(\widehat{\boldsymbol{p}})) - D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{q}))\right)$$

and

$$\sqrt{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial D_{\phi_1}(\boldsymbol{q},\boldsymbol{g}(\boldsymbol{q}))}{\partial q_{ij}}(\widehat{p}_{ij} - q_{ij})$$

have the same asymptotic distribution. If we define

$$l_{ij} = \frac{\partial D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{q}))}{\partial q_{ij}} \tag{16}$$

and $\boldsymbol{l} = (l_{ij};\ i,j = 1, ..., I)^T$, we have

$$\sqrt{n}(D_{\phi_1}(\widehat{\boldsymbol{p}},\boldsymbol{f}(\widehat{\boldsymbol{p}})) - D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{q}))) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\boldsymbol{0},\boldsymbol{l}^T\boldsymbol{\Sigma}_{\boldsymbol{q}}\boldsymbol{l}) \tag{17}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{q}} = \mathrm{diag}(\boldsymbol{q}) - \boldsymbol{q}\boldsymbol{q}^T$.

If we consider the maximum likelihood estimator instead of minimum ϕ-divergence estimator, we get

$$\boldsymbol{l}^T\boldsymbol{\Sigma}_{\boldsymbol{q}}\boldsymbol{l} = \sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} q_{ij}(m_{ij}^{\phi_1})^2 - \left(\sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} q_{ij}m_{ij}^{\phi_1}\right)^2$$

where

$$m_{ij}^{\phi_1} = \frac{1}{2}\phi_1\left(\frac{2q_{ij}}{q_{ij}+q_{ji}}\right) + \frac{1}{2}\phi_1\left(\frac{2q_{ji}}{q_{ij}+q_{ji}}\right) + \frac{q_{ij}}{q_{ij}+q_{ji}}\left(\phi_1'\left(\frac{2q_{ij}}{q_{ij}+q_{ji}}\right) - \phi_1'\left(\frac{2q_{ji}}{q_{ij}+q_{ji}}\right)\right)$$

It is also interesting to observe, if we consider the power divergence measure, that

$$\begin{aligned} m_{ij}^{\lambda} = {} & \frac{1}{2\lambda(\lambda+1)}\left(\left(\frac{2p_{ij}}{p_{ij}+p_{ji}}\right)^{\lambda+1} + \left(\frac{2p_{ji}}{p_{ij}+p_{ji}}\right)^{\lambda+1} - 2\right) \\ & + \frac{1}{\lambda}\frac{p_{ji}}{p_{ij}+p_{ji}}\left(\left(\frac{2p_{ij}}{p_{ij}+p_{ji}}\right)^{\lambda} - \left(\frac{2p_{ji}}{p_{ij}+p_{ji}}\right)^{\lambda}\right) \end{aligned}$$

For $\lambda \to 0$ and $\lambda = 1$ we get

$$m_{ij}^0 = \log\frac{2p_{ij}}{p_{ij}+p_{ji}} \qquad \text{and} \qquad m_{ij}^1 = \frac{p_{ij}^2 - 3p_{ji}^2 + 2p_{ij}p_{ji}}{2(p_{ij}+p_{ji})^2}$$

respectively. Therefore, the corresponding asymptotic variances are given by

$$\sigma_{(0)}^2 = \sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} p_{ij}\left(\log\frac{2p_{ij}}{p_{ij}+p_{ji}}\right)^2 - \left(\sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} p_{ij}\log\frac{2p_{ij}}{p_{ij}+p_{ji}}\right)^2$$

and

$$\sigma_{(1)}^2 = \sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} p_{ij}\left(\frac{p_{ij}^2 - 3p_{ji}^2 + 2p_{ij}p_{ji}}{2(p_{ij}+p_{ji})^2}\right)^2 - \left(\sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I} p_{ij}\frac{p_{ij}^2 - 3p_{ji}^2 + 2p_{ij}p_{ji}}{2(p_{ij}+p_{ji})^2}\right)^2$$

Based on the previous result we can formulate the following theorem.

Theorem 2 *The asymptotic power for the test given in (15), at the alternative* $\boldsymbol{q}$*, is given by*

$$\beta_{n,\phi_1,\phi_2}(\boldsymbol{q}) = 1 - \Phi_n\left(\frac{1}{\boldsymbol{l}^T\boldsymbol{\Sigma_q}\boldsymbol{l}}\left(\frac{\phi_1''(1)}{2\sqrt{n}}\chi^2_{m,\alpha} - \sqrt{n}D_{\phi_1}(\boldsymbol{q}, \boldsymbol{f}(\boldsymbol{\theta}_a^{\phi_2}))\right)\right)$$

where $\Phi_n(x)$ *is a sequence of distributions functions tending uniformly to the standard normal distribution function* $\Phi(x)$*.*

We consider a contiguous sequence of alternative hypotheses that approaches the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta}))$, for some unknown $\boldsymbol{\beta} \in B$, at the rate $O\left(n^{-1/2}\right)$. Consider the multinomial probability vector

$$p_{n,ij} = p_{ij}(\boldsymbol{g}(\boldsymbol{\beta})) + d_{ij}n^{-1/2},\ i = 1, ..., I,\ j = 1, ..., I$$

where $\boldsymbol{d} = (d_{11}, ..., d_{II})^T$ is a fixed $I^2 \times 1$ vector such that $\sum_{i=1}^{I}\sum_{j=1}^{I} d_{ij} = 0$, recall that n is the total count parameter of the multinomial distribution and $\boldsymbol{\beta} \in B$. As $n \to \infty$, the sequence of multinomial probabilities $\{\boldsymbol{p}_n\}_{n\in\mathbb{N}}$ with $\boldsymbol{p}_n = (p_{n,ij},\ i = 1, ..., I,\ j = 1, ..., I)^T$, converges to a multinomial probability in H_0 at the rate of $O\left(n^{-1/2}\right)$. Let

$$H_{1,n} : \boldsymbol{p}_n = \boldsymbol{p}(\boldsymbol{g}(\boldsymbol{\beta})) + \boldsymbol{d}n^{-1/2}, \qquad \boldsymbol{\beta} \in B \tag{18}$$

In the next theorem we present the asymptotic distribution of the family of test statistics $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ defined in (13), under the contiguous alternative hypotheses given in (18).

Theorem 3 *Under* $H_{1,n}$*, given in (18), the family of test statistics* $T_n^{\phi_1}(\widehat{\boldsymbol{\theta}}^{\phi_2})$ *is asymptotically noncentrality chi-squared distributed with* $I(I-1)/2$ *degrees of freedom and noncentrality parameter*

$$\delta = \frac{1}{2}\sum_{i=1}^{I}\sum_{\substack{j=1\\ i\neq j}}^{I}\frac{d_{ij}^2}{p_{ij}} - \sum_{i=1}^{I}\sum_{\substack{j=1\\ i<j}}^{I}\frac{d_{ij}d_{ji}}{p_{ij}}$$

Proof. See Chapter 8 in [12].

An interesting a simulation study can be seen in [7]. In that study some interesting alternative test statistics appear to the classical Pearson test statistics and likelihood ratio test.

4. Hypothesis Testing: H_0 versus H_1 and H_1 versus H_2

In this section we consider the three hypotheses, H_0, H_1, H_2 given in (1), (4), (5) respectively and some test statistics based on ϕ-divergence

$$D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(0)}) \text{ and } D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)}) \tag{19}$$

for testing H_0 against H_1 and H_1 against H_2.

In the expression (19), $\widehat{\boldsymbol{p}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{p}$ given by $\widehat{\boldsymbol{p}} = [\widehat{p}_{ij}]$, where $\widehat{\boldsymbol{p}}_{ij} = n_{ij}/n$; and $\widehat{\boldsymbol{p}}^{(0)}$ and $\widehat{\boldsymbol{p}}^{(1)}$ denote the MLEs of $\boldsymbol{p}$ under H_0 and H_1 respectively. These MLEs were obtained by [8]. Let

$$\theta_{ij} = \frac{p_{ij}}{p_{ij} + p_{ji}}, \text{ for } i > j$$

then $H_0 : \theta_{ij} = 1/2$ (for $i > j$) and $H_1 : \theta_{ij} \geq 1/2$ (for $i > j$), and

$$\widehat{\theta}_{ij} = \frac{n_{ij}}{n_{ij} + n_{ji}}, \ \widehat{\theta}_{ij}^{(0)} = 1/2 \ \text{ and } \ \widehat{\theta}_{ij}^{(1)} = \max\left(\frac{n_{ij}}{n_{ij} + n_{ji}}, \frac{1}{2}\right) \ \text{ for } i > j$$

It follows that $\widehat{\boldsymbol{p}}^{(0)}$ and $\widehat{\boldsymbol{p}}^{(1)}$are given by

$$\widehat{p}_{ij}^{(0)} = \frac{n_{ij} + n_{ji}}{2n} \text{ and } \widehat{p}_{ij}^{(1)} = \frac{n_{ij} + n_{ji}}{n} \max\left(\frac{n_{ij}}{n_{ij} + n_{ji}}, \frac{1}{2}\right)$$

Then we have

$$D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(0)}) = \sum_{i=1}^{I} \sum_{\substack{j=1 \\ i>j}}^{I} \frac{n_{ij} + n_{ji}}{2n} \left(\phi(2\widehat{\theta}_{ij}) + \phi(2(1 - \widehat{\theta}_{ij}))\right)$$

and

$$D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)}) = \sum_{i=1}^{I} \sum_{\substack{j=1 \\ i>j}}^{I} \frac{n_{ij} + n_{ji}}{n} \left(\widehat{\theta}_{ij}^{(1)} \phi\left(\frac{\widehat{\theta}_{ij}}{\widehat{\theta}_{ij}^{(1)}}\right) + (1 - \widehat{\theta}_{ij}^{(1)}) \phi\left(\frac{(1 - \widehat{\theta}_{ij})}{(1 - \widehat{\theta}_{ij}^{(1)})}\right)\right)$$

To solve the problem of testing H_1 against H_2, [8] consider the likelihood ratio test statistic

$$T_{12} = 2 \sum_{i=1}^{I} \sum_{\substack{j=1 \\ i>j}}^{I} \left(n_{ij} \ln \widehat{\theta}_{ij} + n_{ji} \ln(1 - \widehat{\theta}_{ij}) - n_{ij} \ln(\widehat{\theta}_{ij}^{(1)}) - n_{ji} \ln(1 - \widehat{\theta}_{ij}^{(1)})\right)$$

This statistic is such that

$$T_{12} = 2n D_{KL}(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)}) \tag{20}$$

where $D_{KL}(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)})$ is the Kullback–Leibler divergence given by (20) with $\phi(x) = \phi_{(0)}(x)$ defined above. Then the likelihood ratio test statistic is based on the closeness, in terms of the Kullback–Leibler divergence measure, between the probability distributions $\widehat{\boldsymbol{p}}$ and $\widehat{\boldsymbol{p}}^{(1)}$. Thus, one could measure the closeness between the two probability distributions using a more general divergence measure if we are able to obtain its asymptotic distribution. One appropriate family of divergence measures for that purpose is the ϕ-divergence measure.

As a generalization of the test statistic given in (20) for testing H_1 against H_2 we introduce the family of test statistics

$$T_{12}^{\phi} = \frac{2n}{\phi''(1)} D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)}) \tag{21}$$

To test H_0 against H_1, El Barmi and Kochar [8] consider the likelihood ratio test statistic

$$T_{01} = 2 \sum_{i=1}^{I} \sum_{\substack{j=1 \\ i>j}}^{I} \left(n_{ij} \ln \widehat{\theta}_{ij}^{(1)} + n_{ji} \ln(1 - \widehat{\theta}_{ij}^{(1)}) - n_{ij} \ln(\tfrac{1}{2}) - n_{ji} \ln(\tfrac{1}{2})\right)$$

It is clear that

$$T_{01} = 2n \left(D_{KL}(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(0)}) - D_{KL}(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)})\right)$$

As a generalization of this test statistic we consider in this paper the family of test statistics

$$T_{01}^{\phi} = \frac{2n}{\phi''(1)} \left(D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(0)}) - D_\phi(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{p}}^{(1)})\right) \tag{22}$$

If $\phi = \phi_{(0)}$ then $T_{12}^{\phi} = T_{12}$ and $T_{01}^{\phi} = T_{01}$, and hence the families of test statistics T_{12}^{ϕ} and T_{01}^{ϕ} can be considered as generalizations of the test statistics T_{12} and T_{01} respectively.

In order to get the asymptotic distribution of the test statistics given in (21) and (22), we first define the so-called chi-bar squared distribution with n degrees of freedom, denoted by $\bar{\chi}_n^2$.

Definition 4 *Let $U = \max(0, Z)$ where $Z \sim \mathcal{N}(0,1)$, so that the c.d.f. of U is given by*

$$F_U(u) = \begin{cases} \Phi(u), & u > 0 \\ 0, & u < 0 \end{cases}$$

where Φ denotes the standard normal cumulative distribution function. Let $V = \sum_{k=1}^{n} U_k^2$, where $U_1, ..., U_n$ are independent and distributed like U, then $V \sim \bar{\chi}_n^2$.

It is readily shown that

$$\mathrm{E}(V) = \frac{1}{2}n \text{ and } \mathrm{Var}(V) = \frac{5}{4}n$$

This distribution is related to χ^2 distribution. It can be readily shown that

$$\Pr(V > v) = \sum_{l=0}^{n} \binom{n}{l} \frac{1}{2^n} \Pr(\chi_l^2 > v)$$

by conditioning on L, the number of non-zero U_is.

Furthermore, like the χ_n^2 distribution, the $\bar{\chi}_n^2$ distribution is stochastically increasing with n. If $V \sim \bar{\chi}_n^2$ and $V \sim \bar{\chi}_{n'}^2$, where $n < n'$, then V is stochastically smaller than $V' : \Pr(V > t) \leq \Pr(V' > t)$. This follows since $V' \sim V + W$ where $W \sim \bar{\chi}_{n-n'}^2$ with V and W independent. For more details about the chi-bar squared distribution see [17].

The following theorem present the asymptotic distribution of T_{01}^{ϕ}.

Theorem 5 *Under H_0, $T_{01}^{\phi} \xrightarrow{\mathcal{L}} \bar{\chi}_K^2$ as $n \to \infty$ where $K = I(I-1)/2$.*

Proof. See [7].

If we consider the family of power divergences given in (8), we have the power divergence family of test statistics defined as

$$T_{01}^{\lambda} = T_{01}^{\phi_{(\lambda)}}$$

which can be used for testing H_0 against H_1. Therefore some important statistics can now be expressed as members of the power divergence family of test statistics T_{01}^{λ}, that is, T_{01}^{1} is the Pearson test statistic (X_{01}^2), $T_{01}^{-1/2}$ is the Freeman–Tukey test statistic (F_{01}^2), T_{01}^{-2} is the Neyman-modified test statistic (NM_{01}^2), T_{01}^{-1} is the modified loglikelihood ratio test statistic (NG_{01}^2), T_{01}^{0} is the loglikelihood ratio test statistic (G_{01}^2) introduced by [8] and $T_{01}^{2/3}$ is Cressie–Read test statistic (see [11]).

Theorem 6 *Under H_1, $T_{12}^{\phi} \xrightarrow{\mathcal{L}} \bar{\chi}_M^2$ as $n \to \infty$, where M is the number of elements in the set $\{(i,j) : i > j, p_{ij} = p_{ji}\} \leq K = \frac{1}{2}I(I-1)$ and*

$$\lim_{n\to\infty} \Pr(T_{12}^{\phi} \geq t) \leq \sum_{l=0}^{K} \binom{K}{l} \frac{1}{2^K} \Pr(\chi_l^2 \geq t)$$

If we consider the family of power divergences given in (8), we have the power divergence family of test statistics defined as

$$T_{12}^{\lambda} = T_{12}^{\phi_{(\lambda)}}$$

which we can use for testing H_1 against H_2.

Remark 7 *In the same way as previously we can obtain the test statistics* $T_{12}^{0} = G_{12}^{2}$, $T_{12}^{-1} = NG_{12}^{2}$, $T_{12}^{1} = X_{12}^{2}$, $T_{12}^{-1/2} = F_{12}^{2}$, $T_{12}^{-2} = NM_{12}^{2}$ *and* $T_{12}^{2/3}$.

We will refer here to the example of [18, Section 9.5], where the test proposed by Bowker [1] is applied. The proposed tests in this paper may be used in the situation such that it is hoped that a new formulation of a drug will reduce some side-effects.

Example We consider 158 patients who have been treated with the old formulation and records are available of any side-effects. We might now treat each patient with the new formulation and note incidence of side-effects. Table 1 shows a possible outcome for such an experiment. Do the data in Table 1 provide any evidence regarding a less severity of side-effects with the new formulation of the drug? BA The two test statistics given in (21) and (22) are appropriate for this problem. For

Table 1. Side-effect levels for old and new formulation.

		Side-effect levels-new formulation			
		None	Slight	Severe	Total
Side-effect levels-old formulation	None	83	4	3	90
	Slight	17	22	5	44
	Severe	4	9	11	24
	Total	104	35	19	158

the test statistic T_{01}^{ϕ} given in (22) the null hypothesis is that for all off-diagonal counts in the table the associated probabilities are such that all $p_{ij} = p_{ji}$. The alternative is that $p_{ij} \geq p_{ji}$ for all $i \geq j$. We have computed the members of the family $\{T_{01}^{\lambda}\}$ given in Remark 7 and the corresponding asymptotic p-values $P_{01}^{\lambda} = \Pr(\bar{\chi}_3^2 > T_{01}^{\lambda})$ which are given in the following table:

Table 2. Asymptotic p-values for T_{01}^{λ}.

λ	-2	-1	$-1/2$	0	$2/3$	1
T_{01}^{λ}	14.43	11.48	10.58	9.96	9.46	9.33
P_{01}^{λ}	0.001	0.003	0.004	0.006	0.007	0.008

On the other hand, if we consider the usual Pearson test statistic X^2, we have that the value of this statistic is 9.33. In this case using the chi-squared distribution with 3 degrees of freedom, the corresponding asymptotic distribution found by Bowker [1], $\Pr(\chi_3^2 > X^2) = 0.025$. Then for all

the considered statistics there is evidence of a differing incidence rate for side-effects under the two formulations, moreover this difference is towards less severe side effects under the new formulation. Therefore, the two considered tests lead to the same conclusion: There is a strong evidence of a bigger incidence rate for side-effects under the old formulation. The conclusion obtained in [18] is in accordance with our conclusion.

Acknowledgements

This work was supported by Grants MTM 2009-10072 and BSCH-UCM 2008-910707.

References

1. Bowker, A. A test for symmetry in contingency tables. *J. Am. Statist. Assoc.* **1948**, *43*, 572–574.
2. Ireland, C.T.; Ku, H.H.; Koch, G.G. Symmetry and marginal homogeneity of an $r \times r$ contingency table. *J. Am. Statist. Assoc.* **1969**, *64*, 1323–1341.
3. Quade, D.; Salama, I.A. A note on minimum chi-square statistics in contingency tables. *Biometrics* **1975**, *31*, 953–956.
4. Menéndez, M.L.; Pardo, J.A.; Pardo, L. Tests based on ϕ-divergences for bivariate symmetry. *Metrika* **2001**, *53*, 15–29.
5. Menéndez, M.L.; Pardo, J.A.; Pardo, L. Tests for bivariate symmetry against ordered alternatives in square contingency tables. *Aust. N. Z. J. Stat.* **2003**, *45*, 115–124.
6. Menéndez, M.L.; Pardo, J.A.; Pardo, L. Tests of symmetry in three-dimensional contingency tables based on phi-divergence statistics. *J. Appl. Stat.* **2004**, *31*, 1095–1114.
7. Menéndez, M.L.; Pardo, J.A.; Pardo, L.; Zografos, K. On tests of symmetry, marginal homogeneity and quasi-symmetry in two contingency tables based on minimum ϕ-divergence estimator with constraints. *J. Stat. Comput. Sim.* **2005**, *75*, 555–580.
8. El Barmi, H.; Kochar, S.C. Likelihood ratio tests for symmetry against ordered alternatives in a square contingency table. *Stat. Probab. Lett.* **1995**, *2*, 167–173.
9. Csiszàr, I. Eine Informationstheorestiche Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. The Mathematical Institute of Hungarian Academy of Sciences: Budapest, Hungary, 1963; Volume 8, pp. 84–108.
10. Ali, S.M.; Silvey, S.D. A general class of coefficients of divergence of one distribution from another. *J. Roy. Stat. Soc. Ser. B Stat. Met.* **1966**, *28*, 131–142.
11. Cressie, N.; Read, T.R.C. Multinomial goodness-of-fit tests. *J. Roy. Stat. Soc. Ser. B Stat. Met.* **1984**, *46*, 440–464.
12. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall/CRC: New York, NY, USA, 2006.
13. Kullback, S. *Information Theory and Statistics*; John Wiley: New York, NY, USA, 1959.
14. Kullback, S. Marginal homogeneity of multidimensional contingency tables. *Ann. Math. Stat.* **1971**, *42*, 594–606.
15. Morales, D.; Pardo, L.; Vajda, I. Asymptotic divergences of estimates of discrete distributions. *J. Stat. Plan. Infer.* **1995**, *48*, 347–369.

16. Pardo, J.A.; Pardo, L.; Zografos, K. Minimum ϕ-divergence estimators with constraints in multinomial populations. *J. Stat. Plan. Infer.* **2002**, *104*, 221–237.
17. Robertson, T.; Wrigt, F.T.; Dykstra, R.L. *Order Restricted Statistical Inference*; Wiley: New York, NY, USA, 1988.
18. Sprent, P. *Applied Nonparametric Statistical Methods*; Chapman & Hall: London, UK, 1993.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license http://creativecommons.org/licenses/by/3.0/.