Available online at www.sciencedirect.com

ScienceDirect

Information Sciences 177 (2007) 5024–5032

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

# Simulation study of the tests of uniform association based on the power-divergence ☆

M.M. Mayoral [a], J.A. Pardo [b,*]

[a] Department of Methodology of the Investigation and T.C., Complutense University of Madrid, Spain
[b] Department of Statistics and O.R. I, Complutense University of Madrid, Spain

Received 21 March 2005; received in revised form 30 March 2007; accepted 16 April 2007

**Abstract**

In this paper, a simulation study is presented to analyze the behavior of the family of test statistics proposed by Conde and Salicrú [J. Conde, M. Salicrú, Uniform association in contingency tables associated to Csiszár divergence, Statistics and Probability Letters, 37 (1998) 149–154] using the $\phi$-divergence measures, that include as special case the power-divergence [N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, Journal of the Royal Statistic Society, Series B, 46 (1984) 440–464] for the analysis of uniform association between two classification processes, based on the local odd ratios. For the above test statistics the significance level and its power are evaluated for different sample sizes when we consider a $3 \times 2$ contingency table.


*Keywords:* Power-divergence; Uniform association; Local odd ratio; Contingency table

## 1. Introduction

Let $X$ and $Y$ denote two categorical response variables, $X$ and $Y$ having $I$ and $J$ levels, respectively. When we classify subjects on both variables, there are $I \times J$ possible combinations of classifications. The responses $(X, Y)$ of a subject randomly chosen from some population have a probability distribution. Let $p_{ij} = P(X = i, Y = j)$, with $p_{ij} > 0, i = 1, \ldots, I, j = 1, \ldots, J$ and we denote by $p = (p_{11}, \ldots, p_{IJ})^{\mathrm{T}}$ the joint distribution of $X$ and $Y$. We display this distribution in a contingency table having $I$ rows for the categories of $X$ and $J$ columns for the categories of $Y$. We consider the $I \times J$ contingency table with ordered rows and adjacent columns. If the responses are independent we have $p_{ij} = p_{i.} \times p_{.j}$ for all $i$ and $j$, where $p_{i.} = \sum_{j=1}^{J} p_{ij}$ and $p_{.j} = \sum_{i=1}^{I} p_{ij}$. When the responses are not independent, there is an association between them, and the question arises of how the association can be measured. It is possible to consider single summary numbers that describe relationships between the categorical response variables, $X$ and $Y$. Examples include Yule's coefficient,

☆ This work was supported by Grant MTM2006-06872 and UCM2006-910707
* Corresponding author.
*E-mail address:* jpardo@mat.ucm.es (J.A. Pardo).

0020-0255/$ - see front matter © 2007 Published by Elsevier Inc.
doi:10.1016/j.ins.2007.04.009

Pearson's coefficients, Tschuprow's coefficient and Cramer's coefficient among others. These numbers, are generally called "measures of association" , and summarize the deviations of the $p_{ij}$ from the hypotheses of independence.

Bartolucci and Forcina [1], Menéndez et al. [11], Tomizawa [17] and references therein used the odd ratios to establish any measure of association of rows and columns for an $I \times J$ contingency table. In this contingency table the association may be viewed in terms of the association in each of the $2 \times 2$ subtables formed from the full $I \times J$ table. It is well-known that the $(I-1)(J-1)$ local odd ratios formed from adjacent rows $i$ and $i+1$, and adjacent columns $j$ and $j+1$

$$\theta_{ij} = p_{ij}p_{i+1,j+1}/p_{i,j+1}p_{i+1,j}, \quad i = 1,\ldots,I-1, \; j = 1,\ldots,J-1$$

determine all odd ratios that can be formed from the full $I \times J$ contingence table. The value $\theta_{ij}$ is a measure of the row–column interaction in the $2 \times 2$ subtable formed from adjacent rows $i$ and $i+1$ and adjacent columns $j$ and $j+1$. Interesting topics in categorical data analysis can be seen, for example, in [2,20,19,21] and references therein.

We denote $\pi = (\pi_{11},\ldots,\pi_{(I-1)(J-1)})^{\mathrm{T}}$ and $\pi^u = (\pi^u_{11},\ldots,\pi^u_{(I-1)(J-1)})^{\mathrm{T}}$, where $\pi_{ij} = \theta_{ij}/v$, $v = \sum_{i=1}^{I-1}\sum_{j=1}^{J-1}\theta_{ij}$ and $\pi^u_{ij} = \frac{1}{(I-1)(J-1)}$. Goodman [7] considered the uniform association model, defined by

$$\pi_{ij} = \pi^u_{ij}, \quad i = 1,\ldots,I-1, \; j = 1,\ldots,J-1.$$

Tomizawa [17] proposed a measure to represent the degree of departure from uniform association based on the information in the vector $\pi = (\pi_{11},\ldots,\pi_{(I-1)(J-1)})^{\mathrm{T}}$ and using the Kullback–Leibler divergence. This measure of divergence is a particular case of the power-divergence measure introduced by Cressie and Read [4]. Tomizawa and Hashimoto [18] use the power-divergence to represent the degree of departure from uniform association, which includes Tomizawa's [17] as a special case, but they do not present any test for uniformity association. This family of measures of divergence is defined for two different contingence tables, characterized by the probability distributions $\pi^u$ and $\pi$, by

$$I^{\lambda}(\pi^u,\pi) = \frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{I-1}\sum_{j=1}^{J-1}\pi^u_{ij}\left(\left(\frac{\pi^u_{ij}}{\pi_{ij}}\right)^{\lambda} - 1\right), \quad \lambda \neq 0,-1,$$

$I^0(\pi^u,\pi) = \lim_{\lambda\to 0} I^{\lambda}(\pi^u,\pi)$ and $I^{-1}(\pi^u,\pi) = \lim_{\lambda\to -1} I^{\lambda}(\pi^u,\pi)$. For more details of the power-divergence, see [4,15]. Note that $I^0(\pi^u,\pi)$ is the Kullback–Leibler divergence measure between the probability distributions $\pi^u$ and $\pi$. We can see that $I^{\lambda}(\pi^u,\pi) = 0$ if and only if the uniform association model holds.

In Section 2, we consider a measure based on the power-divergence which represents the degree of departure from uniform association in the $I \times J$ contingency table. Finally, in Section 3 a simulation study is carried out obtaining new test statistics that are good alternatives to the classical test for this problem.

## 2. Test statistic based on power-divergence

In [3] a class of measures was introduced to represent the degree of departure from uniform association for a contingency table based on Csiszár divergence measure. This measure of divergence, between the probability vectors $\pi$ and $\pi^u$ is defined by

$$D_{\phi}(\pi^u,\pi) = \sum_{i=1}^{I-1}\sum_{j=1}^{J-1}\pi_{ij}\phi\left(\frac{\pi^u_{ij}}{\pi_{ij}}\right), \tag{1}$$

where $\phi$ is a convex function for $x > 0$ such that $\phi(1) = 0, \phi''(1) > 0, 0\phi(0/0) = 0$ and $0\phi(p/0) = p\lim_{u\to\infty}\phi(u)/u$. For more details about $\phi$-divergences see [13].

The power-divergence family of Cressie and Read is obtained if we consider

$$\phi(x) \equiv \phi_{(\lambda)}(x) = (\lambda(\lambda+1))^{-1}(x^{\lambda+1} - x + \lambda(1-x)); \quad \lambda \neq 0, \; \lambda \neq -1,$$
$$\phi_{(0)}(x) = \lim_{\lambda\to 0}\phi_{(\lambda)}(x) \quad \text{and} \quad \phi_{(-1)}(x) = \lim_{\lambda\to -1}\phi_{(\lambda)}(x). \tag{2}$$