Journal of Multivariate Analysis 130 (2014) 387–408

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Phi-divergence statistics for the likelihood ratio order: An approach based on log-linear models

Nirian Martin [a], Raquel Mata [b], Leandro Pardo [b,*]

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

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

ARTICLE INFO

Article history:
Received 18 September 2013
Available online 13 June 2014

AMS subject classifications:
62F03
62G10
62E20

Keywords:
Phi-divergence test statistics
Inequality constrains
Likelihood ratio ordering
Log-linear modeling

ABSTRACT

When some treatments are ordered according to the categories of an ordinal categorical variable (e.g., extent of side effects) in a monotone order, one might be interested in knowing whether the treatments are equally effective or not. One way to do that is to test if the likelihood ratio order is strictly verified. A method based on log-linear models is derived to make statistical inference and phi-divergence test-statistics are proposed for the test of interest. Focused on log-linear modeling, the theory associated with the asymptotic distribution of the phi-divergence test-statistics is developed. An illustrative example motivates the procedure and a simulation study for small and moderate sample sizes shows that it is possible to find phi-divergence test-statistic with an exact size closer to nominal size and higher power in comparison with the classical likelihood ratio.

1. Introduction

In this paper we are interested in comparing I treatments when the response variable is ordinal with J categories. We can consider each treatment type to be each of the I ordinal categories of a variable X. We shall denote by Y the response variable and its conditional probabilities by

$$\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T, \quad i = 1, \ldots, I,$$

with

$$\pi_{ij} = \Pr(Y = j | X = i), \quad j = 1, \ldots, J.$$

For the ith treatment and for each individual taken independently from a sample of size n_i, its response is classified to be $\{1, \ldots, J\}$ according to the conditional distribution of $Y|X = i$, $\boldsymbol{\pi}_i$. In this setting the J-dimensional random variable associated with the observed frequencies,

$$\boldsymbol{N}_i = (N_{i1}, \ldots, N_{iJ})^T,$$

is multinomially distributed with parameters n_i and $\boldsymbol{\pi}_i$. Assuming that the different treatments are independent, the probability distribution of the $I \times J$ dimensional random variable $\boldsymbol{N} = (\boldsymbol{N}_1^T, \ldots, \boldsymbol{N}_I^T)^T$ is product-multinomial. We are going to consider a motivation example, taken from Section 5 in [7], in order to clarify the problem considered in this paper. In Table 1, duodenal ulcer patients of a hospital are cross-classified according to an increasing order of $I = 4$ severity degrees of the operation, and the extent of side effects, categorized as None, Slight and Moderate ($J = 3$).

* Corresponding author.
E-mail address: lpardo@mat.ucm.es (L. Pardo).

http://dx.doi.org/10.1016/j.jmva.2014.06.004
0047-259X/

Table 1
Extent of size effect of four treatments.

	None	Slight	Moderate
Treatment 1	61	28	7
Treatment 2	68	23	13
Treatment 3	58	40	12
Treatment 4	53	38	16

We shall consider that Treatment $i+1$ is at least as good as Treatment i, for $i = 1, \ldots, I-1$ simultaneously, if $\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)}$ is non-decreasing for all $j \in \{1, \ldots, J\}$, i.e.,

$$\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} \leq \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)}, \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\}. \tag{1}$$

This is the so called "likelihood ratio ordering", sometimes called also "local ordering" (see [21], Chapter 6). Our interest is focused on testing whether Treatment $i+1$ is "strictly" better than Treatment i, for $i = 1, \ldots, I-1$ simultaneously, or equivalently, if (1) holds with at least one strict inequality, which is the alternative hypothesis of the following hypothesis test

$$H_0 : \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} = \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\}, \tag{2a}$$

$$H_1 : \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} \leq \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\} \tag{2b}$$
$$\text{and } \frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)} < \frac{\Pr(Y=j+1|X=i+1)}{\Pr(Y=j+1|X=i)} \quad \text{for at least one } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\}.$$

For the motivation example, the null hypothesis means that all the treatments have equal side effects, while the alternative hypothesis means that as the more severe treatment is, greater the probability of having side effects is. Note that if we multiply on the left and the right hand side of (2a) and (2b) by $\left(\frac{\Pr(Y=j|X=i+1)}{\Pr(Y=j|X=i)}\right)^{-1}$ we obtain

$$H_0 : \vartheta_{ij} = 1, \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\}, \tag{3a}$$

$$H_1 : \vartheta_{ij} \geq 1 \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\} \tag{3b}$$
$$\text{and } \vartheta_{ij} > 1 \quad \text{for at least one } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\},$$

where $\vartheta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}$ represent the "local odds ratios", also called cross-product ratios.

It is very important to clarify that the counts, for analyzing the likelihood ratio ordering, are referred to I independent samples, each composed by a sample of n_i $(i = 1, \ldots, I)$ independent J-dimensional multinomial random variables of size 1, in such a way that $\pi_{ij} = \Pr(Y = j|X = i) = \Pr\left(\mathcal{M}(1, \boldsymbol{\pi}_i) = \boldsymbol{e}_j\right)$, where $\boldsymbol{e}_j$ is the jth unit vector in dimension J. If we denote by $n = \sum_{i=1}^{I} n_i$ the total of the sample sizes, we can consider the joint distribution to be

$$p_{ij} = \Pr(X = i, Y = j) = \Pr(X = i)\Pr(Y = j|X = i) = \frac{n_i}{n}\pi_{ij}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J.$$

Having a sample of n individuals, each of them is first assigned for a treatment group of size n_i, $i = 1, \ldots, I$, and so $\Pr(X = i) = \frac{n_i}{n}$ represents the probability for an individual to be assigned randomly to Treatment i. In a second stage, $\Pr(Y = j|X = i) = \pi_{ij}$ represents the probability for an individual of Treatment group i to be assigned to the jth category of the ordinal categorical variable of interest (e.g., extent of size effect). From a theoretical point of view, the fixed size of each multinomial random variable can be weakened assuming that for each realization $\nu_i = \lim_{n\to\infty} \frac{n_i}{n}$, $i = 1, \ldots, I$, as dealt in Section 4. We display the joint distribution of the whole sample in a rectangular table having I rows for the categories of X and J columns for the categories of Y, and we denote $\boldsymbol{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_I)^T$, with $\mathbf{p}_i = (p_{i1}, \ldots, p_{iJ})^T$, $i = 1, \ldots, I$, the corresponding $I \times J$ matrix and

$$\boldsymbol{p} = \mathrm{vec}(\boldsymbol{P}^T) = (\mathbf{p}_1^T, \ldots, \mathbf{p}_I^T)^T \tag{4}$$

a vector obtained by stacking the columns of $\boldsymbol{P}^T$ (i.e., the rows of matrix $\boldsymbol{P}$). Note that the components of $\boldsymbol{P}$ are ordered in lexicographical order in $\boldsymbol{p}$. The local odds ratios can be expressed only in terms of joint probabilities

$$\vartheta_{ij} = \frac{p_{ij}p_{i+1,j+1}}{p_{i+1,j}p_{i,j+1}} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}, \quad \text{for every } (i,j) \in \{1, \ldots, I-1\} \times \{1, \ldots, J-1\}. \tag{5}$$

The likelihood ratio ordering has been extensively studied in order statistics. In the literature related to order restricted inference for categorical data analysis, the likelihood ratio ordering has received little attention. The definition given in (1) is not specific for multinomial random variables, actually is very similar for any random variable, not necessarily discrete. In [1], it is mentioned that very important families of random variables, such as the one-parameter exponential family of distributions, have the likelihood ratio ordering property with respect to the parameter. For two independent multinomial samples ($I = 2$), Dykstra et al. [9] established the asymptotic distribution of the likelihood ratio test-statistic and Dardanoni