Statistical Inference in Constrained Latent Class Models for Multinomial Data based on ϕ-divergence measures

A. Felipe[1], N. Martín[2], P. Miranda[1]*, and L. Pardo[1]

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

[2]Dept. of Statistics and O.R. II: Decision Methods, Complutense University of Madrid, Madrid, Spain

Abstract

In this paper we explore the possibilities of application of ϕ-divergence measures in inferential problems in the field of latent class models for multinomial items. We start with the problem of estimating the model parameters. As explained below, minimum ϕ-divergence estimators are a natural extension of the maximum likelihood estimator, that is the usual estimator in this problem; we study the asymptotic properties of minimum ϕ-divergence estimators, showing that they share the same asymptotic behavior as the maximum likelihood estimator. To compare the efficiency and robustness of the minimum ϕ-divergence estimators when the sample size is not big enough to apply the asymptotic results, we have carried out an extensive simulation study. Next, we deal with the problem of testing whether a model of latent classes for multinomial data fits a set of data; again, ϕ-divergence measures can be used to generate a family of test statistics generalizing both the likelihood ratio and the chi-squared test statistics. Finally, we treat the problem of choosing the best model out of a sequence of nested latent class models; as before, ϕ-divergence measures can handle the problem and we derive a family of test statistics based on them; we study the asymptotic behavior of these test statistics, showing that it is the same as the classical test statistics. To shed light on differences for small and moderate sample sizes, we have carried out a simulation study.

Keywords: Latent class models, Minimum phi-divergence estimator, Maximum likelihood estimator, Phi-divergence test statistics, Goodness-of-fit, Nested latent class models, Asymptotic distribution.

*Corresponding author: P. Miranda. pmiranda@mat.ucm.es

1. **Introduction**

Latent class modelling deals with situations where manifest and latent variables occur. While manifest variables can be directly observed (e.g. item responses in a questionnaire), this is not the case for latent variables. On the other hand, many theories in Psychology or Sociology rely on the basis of the existence of latent variables (e.g. socioeconomic status) having influence on the manifest variables; thus, these latent variables should be taken into account; however, all that can be assessed are manifest indicators, as income or success or failure in a task. The paradigm of latent class modelling is the principle of local independence, which means that the observed associations of the manifest variables are caused by a latent variable. Otherwise said, if this latent variable remains constant, the relationship between the observed variables disappears. Therefore, latent class modelling is a statistical method aiming to establish the relation between an answer pattern on a set of manifest variables and a latent categorical variable. This probabilistic approach allows the statistical assessment of a model fitting the observed data and permits the unmixing of these observed data into several unobservable data sets coming from different subpopulations.

Latent class models (LCM) were introduced in [22] as a tool for building clustering based on dichotomous observed variables and were later studied in [5], [15], [19] and [20] among many others. From this starting point, latent class models have been successfully applied in different fields (see e.g. [6], [3], [1], [12], [21], [17], [18], [25], [2]). An exhaustive online survey can be found at [27]. One of them is due to Formann [14], [15] and applies to dichotomous response variables; his proposal is characterized by two separated models, one for the class sizes and another one for the latent response probabilities. Following this line, in [13] a family of estimators based on phi-divergence measures were developed for binomial data. A simulation study carried out in that paper suggested that these estimators are competitive with the classical maximum likelihood estimator in this situation.

Formann [16] extended the previous model for dichotomous response variables to the case of polytomous response variables; he used the maximum likelihood estimator (MLE) and the chi-square and likelihood ratio statistics to deal with statistical inferences. As this paper works on the same assumptions, let us explain this model with more detail: Consider a set $\mathcal{P}$ of N people, $\mathcal{P} := \{P_1, ..., P_N\}$; each of the individuals $P_v, v = 1, ..., N$, answers to k polytomous items $I_1, ..., I_k$. Question I_i has g_i possible answers; answers to question i are denoted by $k_{i1}, ..., k_{ig_i}$. Thus, the answer of person P_v to question I_i can be stored in a vector $\mathbf{a_{vi}} := (a_{vi1}, ..., a_{vig_i})$, where

$$a_{vij} := \begin{cases} 1 & \text{if the answer of } P_v \text{ to } I_i \text{ is } k_{ij} \\ 0 & \text{otherwise} \end{cases} .$$

To explain the statistical relationships of the observed variables, a categorical latent variable is postulated to exist, whose different levels define a partition on $\mathcal{P}$ into (fixed) m mutually exclusive and exhaustive latent classes $C_1, ..., C_m$ whose corresponding relative sizes are $w_1, ..., w_m$, (so $\sum_{i=1}^{m} w_i = 1, w_i \geq 0$). We denote

$$p_{jih} := Pr(a_{vih} = 1 | P_v \in C_j),\ j = 1, ..., m,\ i = 1, ..., k,\ h = 1, ..., g_i.$$

Note again that $\sum_{h=1}^{g_i} p_{jih} = 1$. We shall assume that within each class the answers for the different questions are stochastically independent. Thus, given a possible answer vector $\mathbf{a}_\nu := (\mathbf{a}_{\nu\mathbf{1}}, ..., \mathbf{a}_{\nu\mathbf{k}})$, it follows

$$Pr(\mathbf{a}_\nu | P_v \in C_j) = \prod_{i=1}^{k} \prod_{h=1}^{g_i} p_{jih}^{a_{vih}},$$

and consequently,

$$Pr(\mathbf{a}_\nu) = \sum_{j=1}^{m} w_j \prod_{i=1}^{k} \prod_{h=1}^{g_i} p_{jih}^{a_{vih}}. \tag{1}$$

Therefore, the latent class models can be understood as a finite mixture model in which the component distributions are assumed to be multi-way cross-classification tables with all variables mutually independent. There are $g := \prod_{i=1}^{k} g_i$ possible answer vectors $\mathbf{a}_\nu$ whose probability of occurrence are given by Eq. (1). We will denote by N_ν, $\nu = 1, ..., g$, the number of times that the sequence $\mathbf{a}_\nu$ appears in an N-sample and

$$\hat{\mathbf{p}} = (N_1/N, ..., N_g/N). \tag{2}$$

the corresponding sample proportions. The likelihood function L is given by

$$L_{\mathbf{a_1},...,\mathbf{a_g}}(w_1, ..., w_m, p_{111}, ..., p_{mkg_k}) = Pr(N_1 = n_1, ..., N_g = n_g) = \frac{N!}{\prod_{\nu=1}^{g} n_\nu!} \prod_{\nu=1}^{g} Pr(\mathbf{a}_\nu)^{n_\nu}. \tag{3}$$

By n_ν we are denoting a realization of the random variable $N_\nu, \nu = 1, ..., g$. In this model the unknown parameters are w_j and p_{jih}. However, this basic model of unconstrained latent class analysis frequently fails to fit the data. In order to avoid this problem, in [16] it is proposed a linear-logistic parametrization for the probabilities w_j and p_{jih} given by

$$w_j = \frac{exp(z_j)}{\sum_{l=1}^{m} exp(z_l)}, \; j = 1, ..., m, \tag{4}$$

and

$$p_{jih} = \frac{exp(x_{jih})}{\sum_{l=1}^{g_i} exp(x_{jil})}, \; j = 1, ..., m, \; i = 1, ..., k, \; h = 1, ..., g_i. \tag{5}$$

Now, the new parameters of the model are z_j and x_{jih}. Finally, restrictions are introduced relating these parameters to some other explanatory parameters $\lambda_r, r = 1, ..., t$, and $\eta_s, s = 1, ..., u$, through

$$z_j = \mathbf{v}_j^t \boldsymbol{\eta}, \; x_{jih} = \mathbf{q}_{jih}^t \boldsymbol{\lambda},$$

where vectors $\mathbf{v}_j, \mathbf{q}_{jih}$ are known. So we arrive to a model with $t+u$ parameters, $\boldsymbol{\theta} := (\boldsymbol{\lambda}, \boldsymbol{\eta})$, where $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are defined as $\boldsymbol{\lambda} := (\lambda_1, ..., \lambda_t)$, $\boldsymbol{\eta} := (\eta_1, ..., \eta_u)$. Let us denote by $\boldsymbol{\Theta}$ the set in which the parameter $\boldsymbol{\theta}$ varies, i.e. the parametric space. These $t+u$ unknown parameters can be estimated by maximum likelihood using Eq. (3), and from these estimations, the corresponding estimations of p_{jih} and w_j.

Let us now deal with the concept of divergence measure. These measures appear to quantify differences between two probability distributions. There are many divergence measures, but in this paper we will consider the family of **ϕ-divergence measures** introduced in [10]. Given two discrete probability distributions $\mathbf{p} = (p_1, ..., p_r), \mathbf{q} = (q_1, ..., q_r)$ over the same sample space with r elements, we define the ϕ-divergence between $\mathbf{p}$ and $\mathbf{q}$ as

$$D_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^r q_i \phi\left(\frac{p_i}{q_i}\right),$$

where ϕ is a convex function for $x > 0$ satisfying $\phi(1) = 0, 0\phi(0/0) = 0$ and $0\phi(p/0) = p \lim_{x\to\infty} \frac{\phi(x)}{x}$. Let us denote the set of all functions ϕ in these conditions by Φ^*. Now, let $\phi \in \Phi^*$ be differentiable at $x = 1$; then, the function $\psi(x) := \phi(x) - \phi'(1)(x-1)$ also belongs to Φ^* and has the additional property that $\psi'(1) = 0$. This property, together with the convexity and $\psi(1) = 0$, implies that $\psi(x) \geq 0$ for any $x \geq 0$. Moreover, $D_\psi(\mathbf{p}, \mathbf{q}) = D_\phi(\mathbf{p}, \mathbf{q})$. Since the two divergence measures coincide, we can consider the set Φ^* to be equivalent to the set $\Phi := \Phi^* \cap \{\phi : \phi'(1) = 0\}$. More details about ϕ-divergence measures can be seen e.g. in [7] and [26]. In particular, taking $\phi_0(x) = x \log x - x + 1$, we obtain the so-called Kullback-Leibler divergence measure, given by

$$D_{Kullback}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^r p_i \log \frac{p_i}{q_i}.$$

Another important example of divergence measure is the Pearson divergence measure given by

$$D_{Pearson}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^r \frac{1}{2} \frac{(p_i - q_i)^2}{q_i}.$$

In the examples in the paper and in the simulations we shall consider the **power-divergence family** introduced in [9], and given by

$$\phi(x) \equiv \phi_a(x) = \begin{cases} \frac{1}{a(a+1)}(x^{a+1} - x - a(x-1)) & \text{if } a \neq 0, a \neq -1 \\ x \log x - x + 1 & \text{if } a = 0 \\ -\log x + x - 1 & \text{if } a = -1 \end{cases}. \quad (6)$$

Then, $D_a(\mathbf{p}, \mathbf{q})$ is given by

$$D_a(\mathbf{p},\mathbf{q}) = \begin{cases} \frac{1}{a(a+1)}\sum_{i=1}^{g}\left(\frac{p_i^{a+1}}{q_i^a}-1\right) & \text{if } a\neq 0, a\neq -1 \\ D_{Kullback}(\mathbf{p},\mathbf{q}) & \text{if } a=0 \\ D_{Kullback}(\mathbf{q},\mathbf{p}) & \text{if } a=-1 \end{cases}. \tag{7}$$

The plan of the paper is as follows. In Section 2 we introduce minimum phi-divergence estimators for $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$; these estimators can be seen as a natural extension of the maximum likelihood estimator obtained from Eq. (3); besides, the asymptotic distribution of these estimators is obtained. In Section 3 we focus on two problems of testing, namely the problem of goodness-of-fit and the problem of selecting a model out of a sequence of nested models; these problems are classically solved through the Pearson or the likelihood ratio chi-squared test statistics; we shall introduce a new family of test statistics based on minimum phi-divergence that contains as a particular case the classical test statistics; as in Section 2, we derive the asymptotic distribution of the members of these families of test statistics. Section 4 is devoted to a simulation study; in it, we compare ϕ-divergence estimators and ϕ-divergence test statistics with the classical estimators and test statistics for small and moderate sample sizes. We finish with the conclusions.

2. Parameter estimation: Minimum Phi-divergence estimators

Let us denote $Pr(\mathbf{a}_\nu)$ by $p(\mathbf{a}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta})$. Based on Eq. (3), the maximum likelihood estimator (MLE) is obtained by maximizing in $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ the log-likelihood function,

$$\sum_{\nu=1}^{g} n_\nu \log p(\mathbf{a}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta}).$$

This expression can be written as

$$\sum_{\nu=1}^{g} n_\nu \log p(\mathbf{a}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta}) = N\sum_{\nu=1}^{g} \frac{n_\nu}{N} \log p(\mathbf{a}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta}) = -ND_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) + constant,$$

i.e. we must select $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ minimizing the Kullback-Leibler divergence measure between the probability vector $\hat{\mathbf{p}}$, defined in (2) and $\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})$ defined by

$$\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}) := \left(p(\mathbf{a_1}, \boldsymbol{\lambda}, \boldsymbol{\eta}), ..., p(\mathbf{a_g}, \boldsymbol{\lambda}, \boldsymbol{\eta})\right).$$

Based on this, we can extend MLE as follows:

Definition 1. *Given a LCM for polytomous response variables with parameters* $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_t)$ *and* $\boldsymbol{\eta} = (\eta_1, ..., \eta_u)$, *the* **minimum ϕ-divergence estimator** *MϕE of* $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$ *is any* $\hat{\boldsymbol{\theta}}_\phi = (\hat{\boldsymbol{\lambda}}_\phi, \hat{\boldsymbol{\eta}}_\phi)$ *satisfying*

$$D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_\phi, \hat{\boldsymbol{\eta}}_\phi)) = \inf_{(\boldsymbol{\lambda}, \boldsymbol{\eta})\in\boldsymbol{\Theta}} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})). \tag{8}$$

To obtain the MϕE we must solve the following system of equations:

$$\frac{\partial D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))}{\partial s_j} = 0, \; j = 1, ..., u+t \tag{9}$$

being

$$s_j := \left\{ \begin{array}{ll} \lambda_j, & j = 1, ..., t \\ \eta_{j-t}, & j = t+1, ..., t+u \end{array} \right. .$$

Let us obtain these expressions:

$$\frac{\partial D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))}{\partial \lambda_{\alpha}} = \sum_{\nu=1}^{g} \left\{ \frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \lambda_{\alpha}} \phi\left(\frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \right) - \frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \phi'\left(\frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \right) \frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \lambda_{\alpha}} \right\},$$

with

$$\frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \lambda_{\alpha}} = \sum_{j=1}^{m} w_j \prod_{i=1}^{k} \prod_{h=1}^{g_i} p_{jih}^{a_{\nu ih}} \left[\sum_{i=1}^{k} \sum_{h=1}^{g_i} a_{vih} \left[q_{jih\alpha} - \sum_{l=1}^{g_i} p_{jil} q_{jil\alpha} \right] \right].$$

Similarly,

$$\frac{\partial D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))}{\partial \eta_{\beta}} = \sum_{\nu=1}^{g} \left\{ \frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \eta_{\beta}} \phi\left(\frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \right) - \frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \phi'\left(\frac{\hat{p}_{\nu}}{p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})} \right) \frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \eta_{\beta}} \right\},$$

with

$$\frac{\partial p(\mathbf{a}_{\nu}, \boldsymbol{\lambda}, \boldsymbol{\eta})}{\partial \eta_{\beta}} = \sum_{j=1}^{m} w_j \prod_{i=1}^{k} \prod_{h=1}^{g_i} p_{jih}^{a_{\nu ih}} \left(v_{j\beta} - \sum_{l=1}^{m} w_l v_{l\beta} \right).$$

The solution of the system of equations in (9) constitutes the necessary conditions for function $D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$ to have an extreme point at $\boldsymbol{\theta}^* = (\lambda_1^*, ..., \lambda_t^*, \eta_1^*, ..., \eta_u^*)$, but in general it is difficult to check whether it is indeed a minimum phi-divergence estimator. Apart from the problem that a solution of the previous system may fail to minimize $D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$, we have to deal with the problem that several local minimums of $D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$ could exist. To obtain a good approximation of a global minimum and avoid a local minimum or a stationary point, we present in Section 4 a multistart optimization algorithm. Remark however that these problems are not due to function ϕ, as they also appear when dealing with MLE (see [16] for more details).

Let us denote by $(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) = (\lambda_1^0, ..., \lambda_t^0, \eta_1^0, ..., \eta_u^0)$ the true value of the parameter $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ and let us assume that it is an interior point of the parametric space $\boldsymbol{\Theta}$. Let us denote by $\boldsymbol{\Delta}_g$ the set

$$\boldsymbol{\Delta}_g := \left\{ \mathbf{p} = (p_1, ..., p_g)^t : p_{\nu} \geq 0, \; \nu = 1, ..., g, \; \sum_{\nu=1}^{g} p_{\nu} = 1 \right\}. \tag{10}$$

In this section we shall assume that Birch's conditions (see [4]) adapted to this problem hold:

i) $p(\mathbf{a}_\nu, \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) > 0$, $\nu = 1, ..., g$. Thus, $\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) = (p(a_1, \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0), ..., p(a_g, \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0))$ is an interior point of $\boldsymbol{\Delta}_g$. In the following, we will denote $p(\mathbf{a}_\nu, \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$ by $p_\nu(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$.

ii) The mapping $\mathbf{p} : \boldsymbol{\Theta} \to \boldsymbol{\Delta}_g$ assigning to any $(\boldsymbol{\lambda}, \boldsymbol{\eta})$ the vector $\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})$ is continuous and totally differentiable at $(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$.

iii) The Jacobian matrix

$$\mathbf{J}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) := \left(\frac{\partial p_\nu(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)}{\partial s_j} \right)_{\substack{\nu=1,...,g \\ j=1,...,t+u}}$$

is of rank $t + u$.

iv) The inverse mapping of $\mathbf{p}$ is continuous at $\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$.

Now, the following can be proved:

Theorem 1. *Under Birch's conditions and assuming $\phi(t)$ is twice continuously differentiable at any $t > 0$, the MϕE $\hat{\boldsymbol{\theta}}_\phi$ defined in Eq. (8), for the LCM for multinomial data satisfies*

$$\hat{\boldsymbol{\theta}}_\phi = (\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T + I_F(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^{-1} \mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)}^{-\frac{1}{2}} (\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)) + o(\|\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)\|),$$

where $I_F(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$ is the Fisher information matrix for the latent class model with multinomial data, defined by $I_F(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) := (\mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T \mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0))^{-1}$ and $\mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) := \mathbf{D}_{\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)}^{-\frac{1}{2}} \mathbf{J}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$ and by $\mathbf{D}_{\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)}$ we are denoting the diagonal matrix whose diagonal is given by $\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)$.

Proof: See Appendix.

We can observe in this theorem that the expansion obtained for the MϕE in LCM for multinomial data does not depend on the function ϕ. Note that this also happened when dealing with binary data, as shown in [13].

Theorem 2. *Under the assumptions of the previous theorem, the MϕE $\hat{\boldsymbol{\theta}}_\phi$ for the LCM for multinomial data satisfies*

(i) $\sqrt{N}(\hat{\boldsymbol{\theta}}_\phi - (\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T) \xrightarrow[N\to\infty]{L} \mathcal{N}(\mathbf{0}, (\mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T \mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0))^{-1})$.

(ii) $\sqrt{N}(\mathbf{p}(\hat{\boldsymbol{\theta}}_\phi) - \mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)) \xrightarrow[N\to\infty]{L} \mathcal{N}(\mathbf{0}, \mathbf{J}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T \left(\mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^t \mathbf{A}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)\right)^{-1} \mathbf{J}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0))$.

Proof: See Appendix.

Following i), MϕE share the same asymptotic properties as MLE, as the asymptotic distribution does not depend on ϕ. Moreover, if we pay attention to the asymptotic variance-covariance matrix, we can see that it is the inverse of the Fisher information matrix of the model; thus, the MϕE are BAN (Best Asymptotically Normal) estimators. Note however that, although all phi-divergence estimators share the same distribution for big sample sizes, differences could appear when the sample size is not big enough to apply the asymptotic distribution.

Example 1. *We are going to find the MϕE for the set of data of Example 3 in [16]. In this example, father's status and son's status of 3497 British families are compared. Eight different status are considered and the data are given in the following table:*

Father / Son	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1	*50*	*19*	*26*	*8*	*7*	*11*	*6*	*2*
2	*16*	*40*	*34*	*18*	*11*	*20*	*8*	*3*
3	*12*	*35*	*65*	*66*	*35*	*88*	*23*	*21*
4	*11*	*20*	*58*	*110*	*40*	*183*	*64*	*32*
5	*2*	*8*	*12*	*23*	*25*	*46*	*28*	*12*
6	*12*	*28*	*102*	*162*	*90*	*553*	*230*	*177*
7	*0*	*6*	*19*	*40*	*21*	*158*	*143*	*71*
8	*0*	*3*	*14*	*32*	*15*	*126*	*91*	*106*

Next, two latent classes are considered; each class is characterized by the weights of their marginal distributions and a mobility parameter. Consequently,

$$p_{ijl} = \frac{exp(\alpha_{ji} + \beta_{jl} + |i-l|\gamma_j)}{\sum_{i,l}[exp(\alpha_{ji} + \beta_{jl} + |i-l|\gamma_j)]},\ i,l = 1,...,8,\ j = 1,2.$$

Here, α_{ji} denotes the weight of father's status i in class j, β_{jl} denotes the weight of son's status l in class j, and γ_j is the mobility parameter for class j. Consequently, we have the constraints

$$\alpha_{j8} = -(\alpha_{j1} + ... + \alpha_{j7}),\ \beta_{j8} = -(\beta_{j1} + ... + \beta_{j7}).$$

We will consider model M_2'' in [16], where it is imposed that $\gamma_1 = 0$.

From the definition of the model, it follows that just one question is considered, referring to "family status" with 64 possible answers (i.e. 64 combinations (i,j) in the notation explained before). Then, we have a model with 64-5 categories and 29 parameters: 1 for class sizes, 14 for weights in class C_1, 13 for weights in class C_2 and 1 for mobility parameter γ_2.

For comparing different divergence measures, we shall consider the family of power phi-divergence measures introduced in Eq. (7)*. We have obtained the estimations for several values of a for the data in Table 1; the results can be seen at Table 1.*

It should be noted the differences between these estimations and the estimations obtained by Formann in [16]. The reason is that he considered additional constraints in his model. More concretely, he assumed that $p_{171} = p_{181} = p_{271} = p_{281} = 0$, as sample values are taken as structural; it is also assumed that $p_{2i6} = 0, i = 1,...,8$, i.e. class 2 has no individuals in the sixth category for son's status.

3. Goodness-of-fit and model selection

In this section we study the potential of phi-divergence measures to face the problems of goodness-of-fit and the problem of selecting the best model out of a sequence of nested models.

3.1 Goodness-of-fit

The typical approach to model fit in latent class models for multinomial response variables is to compare the response pattern frequencies observed in the data to the expected values according to the model. The predicted response pattern is computed through the parametric estimations obtained in the latent class model. The two most common measures of goodness of fit are (see [16]) the Pearson chi-squared statistic X^2, and the likelihood ratio statistic G^2, whose expressions are:

$$X^2 = \sum_{\nu=1}^{g} \frac{\left(n_s - Np(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})\right)^2}{Np(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})}, \tag{11}$$

and

$$G^2 = 2N \sum_{\nu=1}^{g} \hat{p}_\nu \log \frac{\hat{p}_\nu}{p(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})}. \tag{12}$$

Both statistics are based on MLE and they measure the degree of concordance between the predicted frequencies and the observed ones. The asymptotic distribution of these statistics is chi-square with $\prod_{i=1}^{k} g_i - (u+t) - 1$ degrees of freedom.

These test statistics can be extended in two possible ways: first, we can use MϕE instead of MLE. Second, both X^2 and G^2 measure the differences between the distribution of expected values according to the model and the distribution of observed values; thus, we can consider other measures of divergence to achieve this task.

Both possibilities are combined in the following family of test statistics.

Definition 2. *We define the family of* **phi-divergence test statistics** *to fit the latent class model for multinomial variables by*

$$T_{\phi_1}^{\phi_2} := \frac{2N}{\phi_1''(1)} D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})) = \frac{2N}{\phi_1''(1)} \sum_{\nu=1}^{g} p(y_\nu, \hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2}) \phi_1 \left(\frac{\hat{p}_\nu}{p(y_\nu, \hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})} \right). \tag{13}$$

where ϕ_1 and ϕ_2 are phi-divergence measures.

Associated to ϕ_2 we have the divergence measure that obtains the estimation $\hat{\boldsymbol{\theta}}_{\phi_2} := (\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})$ for $\boldsymbol{\theta} := (\boldsymbol{\lambda}, \eta)$. Associated to ϕ_1 we obtain the measure of fit. This allows us to consider different measures for the problems of estimating and testing. When $\phi_2(x) = x \log x - x + 1$ we get the MLE; moreover, combining $\phi_2(x) = x \log x - x + 1$ with $\phi_1(x) = \frac{1}{2}(x-1)^2$ we get the Pearson chi-square statistic X^2, and combining $\phi_2(x) = x \log x - x + 1$ with $\phi_1(x) = x \log x - x + 1$ we recover the likelihood ratio statistic G^2. Therefore, the family presented in (13) is a natural extension of X^2 and G^2. Below we obtain the asymptotic distribution of $T_{\phi_1}^{\phi_2}$.

Theorem 3. *Under the hypothesis that the LCM for multinomial data with parameters $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_t)$ and $\boldsymbol{\eta} = (\eta_1, ..., \eta_u)$ holds, the asymptotic distribution of the family of the ϕ-divergence test statistics $T_{\phi_1}^{\phi_2}$ given in Eq. (13) is a chi-square distribution with $g - (u+t) - 1$ degrees of freedom.*

Proof. See Appendix. ■

As for estimators, it happens that the asymptotic distribution does not depend on function ϕ. Possible differences among test statistics may arise for samples with small or moderate size. This point is treated in Section 4.

Example 2. *Let us consider again the situation of Example 1; we will study the goodness-of-fit of the model presented in Example 1 for different values of a for estimation and different values for measuring differences. From Theorem 3, we know that these test statistics follow asymptotically a distribution χ^2_{33}. Thus, for a significance of 0.05, we accept the goodness-of-fit of the model if the corresponding test statistic is less than 47.4. The values of the corresponding test statistics appear in Table* **??**.

3.2 Model selection in nested models

Let us turn to the problem of selecting the best model out of a sequence of nested models. A pair of nested latent models consists of a simple model and a more complex depending on more parameters. The more complex model can be considered an extension of the simpler one, in the sense that it contains some parameters that need to be estimated, while in the simpler they are considered fixed and known. We shall denote by $\boldsymbol{\theta}^A = \left(\boldsymbol{\theta}^{A,1}, \boldsymbol{\theta}^{A,2}, \boldsymbol{\theta}^{A,3}, \boldsymbol{\theta}^{A,4}\right)$ with $\boldsymbol{\theta}^{A,1} = (\lambda_1, ..., \lambda_{t^*})$, $\boldsymbol{\theta}^{A,2} = (\lambda_{t^*+1}, ..., \lambda_t)$, $\boldsymbol{\theta}^{A,3} = (\eta_1, ..., \eta_{u^*})$ and $\boldsymbol{\theta}^{A,4} = (\eta_{u^*+1}, ..., \eta_u)$ the parameters associated to the LCM A and by $\boldsymbol{\theta}^B = \left(\boldsymbol{\theta}^{A,1}, \mathbf{0}, \boldsymbol{\theta}^{A,3}, \mathbf{0}\right)$ the parameters associated to the LCM B. We shall assume that $t + u = h_1$ and $t^* + u^* = h_2$. It is clear that the LCM B is nested in LCM A. Two nested latent class models for multinomial response variables can be compared statistically (see [16]) by considering the difference of their corresponding G^2 values (or their X^2 values). The expression of the classical likelihood ratio test for solving the test

$$H_0 : LCMB \text{ against } H_1 : LCMA \tag{14}$$

is

$$G^2_{A-B} = 2\sum_{\nu=1}^{g} n_\nu \log \frac{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^A\right)}{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)}, \tag{15}$$

and the chi-square test statistic is given by

$$X^2_{A-B} = N\sum_{\nu=1}^{g} \frac{\left(p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^A\right) - p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)\right)^2}{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)}. \tag{16}$$

It is known that these test statistics follow asymptotically a chi-square distribution whose degrees of freedom is the difference between the degrees of freedom of the models [16]. If the value of these statistics is nonsignificant, it is concluded that the simpler model fits as well as the more complex one, and thus, there seems to be no benefit considering the complex model.

On the other hand, if the value of the statistics is significant, this means that, according to data, the additional complexity is needed to achieve an adequate fit.

We can observe that

$$G^2_{A-B} = 2N\left(D_{Kullback}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A)\right) - D_{Kullback}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B)\right)\right) \tag{17}$$

and

$$X^2_{A-B} = \frac{2N}{\phi''(1)} D_{Pearson}(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A), \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B)), \tag{18}$$

Remark that in these test statistics, MLE is applied. As before, we can generalize them considering MϕE instead of MLE and other phi-divergence measures instead of $D_{Kullback}$ or $D_{Pearson}$. Then, a generalization of (17) is obtained by

$$S^{\phi_1,\phi_2}_{A-B} = \frac{2N}{\phi_1''(1)}\left(D_{\phi_1}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A_{\phi_2})\right) - D_{\phi_1}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B_{\phi_2})\right)\right), \tag{19}$$

and a generalization of (18) by

$$T^{\phi_1,\phi_2}_{A-B} = \frac{2N}{\phi_1''(1)} D_{\phi_1}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A_{\phi_2}), \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B_{\phi_2})\right). \tag{20}$$

These extensions have been considered in many statistical applications, see for example [8], [26] and references therein. In the following theorem we shall obtain the asymptotic distribution of the family of test statistics given in (19) and (20).

Theorem 4. *Given the LCM for binary data A, B with parameters $\boldsymbol{\theta}^A = \left(\boldsymbol{\theta}^{A,1}, \boldsymbol{\theta}^{A,2}, \boldsymbol{\theta}^{A,3}, \boldsymbol{\theta}^{A,4}\right)$ and $\boldsymbol{\theta}^B = \left(\boldsymbol{\theta}^{A,1}, \mathbf{0}, \boldsymbol{\theta}^{A,3}, \mathbf{0}\right)$, respectively, and under the null hypothesis given in (14), it follows*

$$S^{\phi_1,\phi_2}_{A-B} \underset{N\longrightarrow\infty}{\overset{\mathcal{L}}{\rightarrow}} \chi^2_{h_1-h_2}, \quad T^{\phi_1,\phi_2}_{A-B} \underset{N\longrightarrow\infty}{\overset{\mathcal{L}}{\rightarrow}} \chi^2_{h_1-h_2}.$$

Proof. See Appendix. ∎

Example 3. *Let us consider again the situation of Example 1. In [16], the model considered in that example is just one of a nested sequence of possible models. Thus, several other situations can be considered:*

- *A model with two classes and two mobility parameters γ_1, γ_2. This is model M_2'.*
- *A model with two classes and one mobility parameter γ_2. This is model M_2'' (the model considered in Example 1).*
- *A model with two classes and no mobility parameters. This is model M_2.*
- *A model with one class and a mobility parameter. This is model M_1'.*
- *A model with one class and no mobility parameter. This is model M_1*

Consequently, we have two sequences of nested models given by

$$M_1->M_1'->M_2''->M_2', \qquad M_1->M_2->M_2''->M_2'.$$

Let us apply the results of this section in order to select the best model. For this, we have considered several different values of parameter a for estimating and several different values for testing. The results obtained are given in next table.

CONTINUACION DEL EJEMPLO ANTERIOR

4. **Simulation study**

In this section we present a simulation study to see the behavior of the new minimum phi-divergence estimators and the families of phi-divergence test statistics for latent class models for multinomial data for different sample sizes. Instead of consider a theoretical latent class model, we have considered 100 different models.

We start explaining the way we have defined the 100 models. In each model, we have considered $k = 3$ questions. The values of g_i (the number of possible answer to each question) are: $g_1 = 2, g_2 = 3, g_3 = 4$. Next, we have considered $m = 3$ latent classes, and the number of parameters is given by $t = 3$ parameters $\lambda_1, \lambda_2, \lambda_3$ and $u = 3$ parameters η_1, η_2, η_3. The theoretical values of these parameters are: $\lambda_1^0 = 0.1, \lambda_2^0 = 0.4, \lambda_3^0 = 0.7, \eta_1^0 = 0.1, \eta_2^0 = 0.2, \eta_3^0 = 0.3$. For each model, we have taken matrices $Q_i, i = 1, 2, 3, V$ such that all coordinates vary randomly on $(0, 3)$. And for each model, we have generated 1000 different samples for each of the 8 different sample sizes N. The values of N are 200, 300, 400, 500, 600, 800, 1000, 3000.

In this simulation study we have considered the divergence measures defined in Eq. (6). In this sense, the minimum ϕ-divergence estimator is given by $\hat{\boldsymbol{\lambda}}_\phi, \hat{\boldsymbol{\eta}}_\phi$ satisfying

$$D_a(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_\phi, \hat{\boldsymbol{\eta}}_\phi)) = \inf_{(\boldsymbol{\lambda},\boldsymbol{\eta})\in\boldsymbol{\Theta}} D_a(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})), \tag{21}$$

where

$$D_a(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) := \frac{1}{a(a+1)} \sum_{i=1}^{g} \left(\frac{\hat{p}_i^{a-1}}{\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})_i^a} - 1 \right).$$

We have studied the following values of a: -1, -1/2, -1/4, -1/8, 0, 2/3, 1, 3/2, 3.

Let us start with the problem of estimating the parameters. First, we have to solve Eq. (21); specially, we have to minimize the probability of obtaining a local minimum instead of the global minimum for the function; on the other hand, we do not want the complexity of the algorithm increasing too much; for this, we proceed as follows for each simulation $l, l = 1, ..., 1000$. First, we generate N_{in} initial points (we use $N_{in} = 100$); next, we improve each point in its neighborhood using a low computational cost procedure (more concretely, a variant of the Hooke and Jeeves algorithm); if the improvement is satisfactory, we apply a good optimization algorithm (a quasi-Newton method) from this improved point to obtain a local optimum, and next we try to solve $\nabla D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) = 0$ from this local optimum maintaining the reduction of $D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$ function.

For fixed a and given model s, $s = 1, ..., 100$, for simulation l we get the values

$$\hat{\lambda}^j_{s,l},\ j = 1, ..., t, \quad \hat{\eta}^k_{s,l},\ k = 1, ..., u,$$

i.e. we obtain two vectors $\hat{\boldsymbol{\lambda}}^{(l)}_s = (\hat{\lambda}^1_{s,l}, ..., \hat{\lambda}^t_{s,l})$ and $\hat{\boldsymbol{\eta}}^{(l)}_s = (\hat{\eta}^1_{s,l}, ..., \hat{\eta}^u_{s,l})$ being $\hat{\boldsymbol{\theta}}^{(l)}_s = (\hat{\boldsymbol{\lambda}}^{(l)}_s, \hat{\boldsymbol{\eta}}^{(l)}_s)$ the minimum power-divergence estimator obtained for the l-th simulation using $D_a(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$. Next, $\hat{\boldsymbol{\theta}}_s = (\hat{\boldsymbol{\lambda}}_s, \hat{\boldsymbol{\eta}}_s)$ is defined as

$$\hat{\lambda}^j_s = \frac{1}{n} \sum_{l=1}^{n} \hat{\lambda}^j_{s,l}, \quad \hat{\eta}^k_s = \frac{1}{n} \sum_{l=1}^{n} \hat{\eta}^k_{s,l}.$$

For each a we compute the mean squared error for each λ_j and η_k

$$mse(\lambda_j) = \frac{1}{100} \sum_{s=1}^{100} (\hat{\lambda}^j_s - \lambda^0_j)^2, \quad mse(\eta_k) = \frac{1}{100} \sum_{s=1}^{100} (\hat{\eta}^k_s - \eta^0_k)^2$$

and also the mean squared error for $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$

$$mse_{\boldsymbol{\lambda}} = \frac{1}{t} \sum_{j=1}^{t} mse(\lambda_j), \quad mse_{\boldsymbol{\eta}} = \frac{1}{u} \sum_{k=1}^{u} mse(\eta_k).$$

We have also computed

$$mse_{\boldsymbol{\lambda},\boldsymbol{\eta}} = \frac{1}{t+u} (t\ mse_{\boldsymbol{\lambda}} + u\ mse_{\boldsymbol{\eta}})$$

for each combination (N, a). Similarly, we present the values of $mse_{\mathbf{p}}, mse_{\mathbf{w}}$ and

$$mse_{\mathbf{p},\mathbf{w}} = \frac{1}{j(g+1)} (gj\ mse_{\mathbf{p}} + j\ mse_{\mathbf{w}}).$$

In Tables 2 and 3, we present the simulated averages for the 100 models. From these tables, several conclusions may arise: first, it can be seen that the different errors diminish for every value a when the sample size N grows; this is in consonance with the asymptotic results developed in Section 2; moreover, it can be seen that the differences among the values of a are less significant when N increases; this was expected from the asymptotic results, too. Next, the possible differences among the estimators could only appear for small and moderate sample sizes; in this sense, it can be seen that $a = 2/3$ provides the best estimations when dealing with $\mathbf{p}, \mathbf{w}$; for the parameters $\boldsymbol{\lambda}, \boldsymbol{\eta}$, it seems that $a = 0$ i.e. the MLE and $a = -1/8$ are the best values. Thus, it may be concluded that there are several estimators that could be competitive with the classical MLE.

Next, we are going to deal with the size and power of phi-divergence test statistics. For studying the power, we have considered several alternative hypothesis; for the k-th alternative hypothesis, we have considered modified versions of Q_{jk} given by $Q_{jk} := Q_j + f_k Q'$, where Q' is such that all coordinates are generated randomly on $(0, 1)$ and the different f_k are 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2 for $k = 1, ..., 10$ respectively. We have considered two nominal sizes, namely 0,05 and 0,1. The simulated size is given by

$$\hat{\alpha}_n^a := \frac{\sharp T_{\phi_a}^{\phi_{2/3}} > \chi^2_{g.l.;0.05}}{n}, \quad \left(\text{or } \hat{\alpha}_n^a := \frac{\sharp T_{\phi_a}^{\phi_{2/3}} > \chi^2_{g.l.;0.1}}{n} \right).$$

Remember that for testing the goodness of fit, the family of test statistics that we have developed in Section 3 can apply a divergence measure for estimating and another one (possibly different) for testing. In this study, we have considered $a = 2/3$ for estimating, as we have seen in Tables 2 and 3 that this seems to be a value that obtains good estimations for $\mathbf{p}$ and $\mathbf{w}$. Besides, when several possibilities are compared in terms of power, it is necessary to fix the phi-divergence test statistics that are going to be selected in the sense that only test statistics with good behavior in terms of size are considered. For this, it is usual to fix upper and lower bounds for the size simulated levels; a popular option for these bounds have been proposed by Dale in [11] (see e.g. [8], [24]). Based on this criterion, we only consider the phi-divergence test statistics whose simulated exact size $\hat{\alpha}_n^a$ satisfies $|logit(1 - \hat{\alpha}_n^a) - logit(1 - \alpha)| \leq d$ where $logit(p) = log(\frac{p}{1-p})$ and $d \in (0.035, 0.07)$. We have chosen $d = 0.4$, so that we only take under consideration the phi-divergence test statistics such that the corresponding size is

$$\hat{\alpha}_n^a \in (0.0327, 0.07625) \tag{22}$$

for the nominal size $\alpha = 0.05$, and

$$\hat{\alpha}_n^a \in (0.06616, 0.1484) \tag{23}$$

for $\alpha = 0.1$. In Table 4, we present the results of sizes for several values of a. It can be seen that the phi-divergence test statistics satisfying Eqs. (22) and (23) increases with the sample size; again, this was expected from the asymptotic results presented in Section 3. On the other hand, it can be observed that this convergence is very slow for negative values of a.

From Table 4, we have decided to consider the values for a given by $-1/8, 0, 2/3, 1, 3/2$. Now, in Tables 5 and 6, it can be seen that the test statistic for $a = -1/8$ seems to show a better behavior than the corresponding test statistics for $a = 0$ (the likelihood ratio test statistic) and $a = 1$ (the chi-squared test statistic).

5. **Conclusions**

In this paper we have introduced some new inferential tools based on divergence measures in the framework of LCM for multinomial data. First, we have considered the problem of estimating the parameters that define the model; we have defined a family of estimators that contain the MLE as a special case, and we have proved that the asymptotic distribution of any member of this family coincides with the asymptotic distribution of the MLE. Besides, we have introduced new families of test statistics, the phi-divergence test statistics, for dealing with the problems of goodness-of-fit and selecting a model out of a nested sequence for latent class models with multinomial data. For this, we have used the fact that the classical likelihood ratio test and chi-squared test statistic are special cases of phi-divergence test statistics. Thus, our family seems a natural extension of these test statistics. Besides, any test statistic in these new families shares the same asymptotic properties as the classical test statistics.

The possible differences between the test statistics in these families should appear on the performance for reduced sample sizes. To study these differences, we have conducted a simulation study. From this simulation study, it seems that there are estimators in the family defined in the paper that are competitive with the classical MLE. Similarly, when dealing with goodness of fit, we have obtained a phi-divergence test statistic that seems to behave better than the classical test statistics based on the Pearson chi-square test statistic and the maximum likelihood test statistic.

Acknowledgements

This paper was supported by the Spanish Grant MTM-2012-33740.

Appendix

Proof of Theorem 1.

We denote by l^g the interior of the g-dimensional unit cube, where $g := \prod_{i=1}^{k} g_i$. The interior of Δ_g defined in (10) is contained in l^g. Let $W_{(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}$ be a neighborhood of $(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)$, the true value of the unknown parameter $(\boldsymbol{\lambda},\boldsymbol{\eta})$, on which

$$\begin{array}{lll} \mathbf{p}: & \Theta & \to \Delta_g \\ & (\boldsymbol{\lambda},\boldsymbol{\eta}) & \mapsto \mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta}) := (p_1(\boldsymbol{\lambda},\boldsymbol{\eta}),...,p_g(\boldsymbol{\lambda},\boldsymbol{\eta})) \end{array}$$

has continuous second partial derivatives. Consider the application

$$\mathbf{F} := (F_1,...,F_{t+u}) : l^g \times W_{(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} \to \mathbb{R}^{t+u}$$

whose components F_j, $j = 1,...,t+u$ are defined by

$$F_j(\tilde{\mathbf{p}};(\boldsymbol{\lambda},\boldsymbol{\eta})) := \frac{\partial D_\phi(\tilde{\mathbf{p}},\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta}))}{\partial s_j},\ j = 1,...,t+u,$$

where s_j is either λ_j if $j \leq t$ or η_{j-t} if $j > t$ and $\tilde{\mathbf{p}}$ is a g-dimensional probability vector.

Then F_j, $j = 1,...,t+u$ vanishes at $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0);(\boldsymbol{\lambda}_0;\boldsymbol{\eta}_0))$. Since

$$\frac{\partial}{\partial s_r}\left(\frac{\partial D_\phi(\tilde{\mathbf{p}},\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta}))}{\partial s_j}\right) = \sum_{\nu=1}^{g} \phi''\left(\frac{\tilde{p}_\nu}{p_\nu(\boldsymbol{\lambda},\boldsymbol{\eta})}\right)\frac{\tilde{p}_\nu}{p_\nu(\boldsymbol{\lambda},\boldsymbol{\eta})^2}\frac{\partial p_\nu(\boldsymbol{\lambda},\boldsymbol{\eta})}{\partial s_r}\frac{\partial p_\nu(\boldsymbol{\lambda},\boldsymbol{\eta})}{\partial s_j}\frac{\tilde{p}_\nu}{p_\nu(\boldsymbol{\lambda},\boldsymbol{\eta})}$$

The $(t+u)\times(t+u)$ matrix $\mathbf{J}_{\mathbf{F}}(\boldsymbol{\theta}_0)$ associated with function $\mathbf{F}$ at point $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$ is given by

$$\begin{aligned}\frac{\partial \mathbf{F}}{\partial(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} &= \left(\frac{\partial \mathbf{F}}{\partial(\boldsymbol{\lambda},\boldsymbol{\eta})}\right)_{(\tilde{\mathbf{p}},(\boldsymbol{\lambda},\boldsymbol{\eta}))=(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0);\boldsymbol{\lambda}_0;\boldsymbol{\eta}_0)} \\ &= \phi''(1)\left(\sum_{l=1}^{g}\frac{1}{p_l(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}\frac{\partial p_l(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}{\partial s_r}\frac{\partial p_l(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}{\partial s_j}\right)_{\substack{j=1,\dots,t+u\\ r=1,\dots,t+u}}\end{aligned}$$

Next, it is a simple algebra exercise to prove that $J_F(\boldsymbol{\theta}_0)$ is nonsingular. As $\mathbf{J_F}(\boldsymbol{\theta}_0) = \mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)\phi''(1)$, we conclude that this matrix is nonsingular at point $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$.

Applying the Implicit Function Theorem, there exists a neighborhood U of $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$ such that the matrix $\mathbf{J_F}$ is nonsingular (in our case $\mathbf{J_F}$ at $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$ is positive definite and then it is continuously differentiable). Also, there exists a continuously differentiable function

$$\tilde{\boldsymbol{\theta}} : A \subset l^g \to \mathbb{R}^{t+u}$$

such that $\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0) \in A$ and

$$\{(\tilde{\mathbf{p}},(\boldsymbol{\lambda},\boldsymbol{\eta})) \in U : \mathbf{F}(\tilde{\mathbf{p}},(\boldsymbol{\lambda},\boldsymbol{\eta})) = 0\} = \left\{(\tilde{\mathbf{p}},\tilde{\boldsymbol{\theta}}(\tilde{\mathbf{p}})) : \tilde{\mathbf{p}} \in A\right\}. \tag{24}$$

Let us define

$$\psi(\boldsymbol{\lambda},\boldsymbol{\eta}) := D_\phi(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta})).$$

As $\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0) \in A$, we conclude that

$$\mathbf{F}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))) = \frac{\partial D_\phi(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\mathbf{p}(\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))))}{\partial(\boldsymbol{\lambda},\boldsymbol{\eta})} = \mathbf{0}.$$

Briefly speaking, $\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$ is the minimum of function ψ. On the other hand, applying (24),

$$(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))) \in U,$$

and then $\mathbf{J_F}$ is positive definite at $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)))$. Therefore,

$$D_\phi(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\mathbf{p}(\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)))) = \inf_{(\boldsymbol{\lambda},\boldsymbol{\eta})\in\Theta} D_\phi(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta})),$$

and by the ϕ-divergence properties $\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)) = (\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T$, and

$$\frac{\partial \mathbf{F}}{\partial \mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} - \frac{\partial \mathbf{F}}{\partial(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}\frac{\partial(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}{\partial \mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} = \mathbf{0}.$$

Further, we know that

$$\frac{\partial \mathbf{F}}{\partial(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} = \phi''(1)\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0).$$

The (i,j)-th element of the $(t+u)\times g$ matrix $\frac{\partial F_j}{\partial p_i}$ is given by:

$$\frac{\partial}{\partial p_i}\left(\frac{\partial D_\phi(\tilde{\mathbf{p}},\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta}))}{\partial s_j}\right)=\frac{1}{p_i(\boldsymbol{\lambda},\boldsymbol{\eta})}\left(-\frac{p_i}{p_i(\boldsymbol{\lambda},\boldsymbol{\eta})}\phi''\left(\frac{p_i}{p_i(\boldsymbol{\lambda},\boldsymbol{\eta})}\right)\right)\frac{\partial p_i(\boldsymbol{\lambda},\boldsymbol{\eta})}{\partial s_j}$$

and for $(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0),(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))$ we have

$$\frac{\partial}{\partial p_i}\left(\frac{\partial D_\phi(\tilde{\mathbf{p}},\mathbf{p}(\boldsymbol{\lambda},\boldsymbol{\eta}))}{\partial s_j}\right)=\frac{1}{p_i(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}\phi''(1)\frac{\partial p_i(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}{\partial s_j}.$$

Since $\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)=\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}\mathbf{J}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)$, then

$$\frac{\partial\mathbf{F}}{\partial\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}=-\phi''(1)\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}. \tag{25}$$

Then we obtain

$$\frac{\partial(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}{\partial\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}=(\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))^{-1}\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}.$$

A first order Taylor expansion of the function $\tilde{\boldsymbol{\theta}}$ around $\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)$ yields

$$\tilde{\boldsymbol{\theta}}(\tilde{\mathbf{p}})=\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))+\left(\frac{\partial\tilde{\boldsymbol{\theta}}(\tilde{\mathbf{p}})}{\tilde{\mathbf{p}}}\right)_{\tilde{\mathbf{p}}=\boldsymbol{\pi}}(\tilde{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))+o(\|\tilde{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)\|).$$

But $\tilde{\boldsymbol{\theta}}(\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))=(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t$, whence

$$\tilde{\boldsymbol{\theta}}(\tilde{\mathbf{p}})=(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t+(\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))^{-1}\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0}(\tilde{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))+o(\|\tilde{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)\|).$$

It is well-known that the nonparametric estimation $\hat{\mathbf{p}}$ converges almost sure to the probability vector $\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)$. Therefore $\hat{\mathbf{p}}\in A$ and $\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})$ is the unique solution of the system of equations

$$\frac{\partial D_\phi(\hat{\mathbf{p}},\mathbf{p}(\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})))}{s_j}=0,\ j=1,...,t+u,$$

and also $(\hat{\mathbf{p}},\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}}))\in U$. Therefore, $\tilde{\boldsymbol{\theta}}(\hat{\mathbf{p}})$ is the minimum ϕ-divergence estimator, $\hat{\boldsymbol{\theta}}_\phi$, satisfying the relation

$$\hat{\boldsymbol{\theta}}_\phi=(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t+(\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))^{-1}\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^t\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}(\hat{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))+O_p(N^{-1/2}).$$

This finishes the proof. ■

Proof of Theorem 2.
Based on the BAN decomposition of the previous theorem it holds

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_\phi-(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T)=\left(\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)^T\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)\right)^{-1}\mathbf{A}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)\mathbf{D}^{-\frac{1}{2}}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}\sqrt{N}(\hat{\mathbf{p}}-\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0))+O_p(1).$$

By the Central Limit Theorem we conclude $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)) \xrightarrow{L} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)})$, where $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)}$ is given by $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\lambda}_0,\boldsymbol{\eta}_0)} - \mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)\mathbf{p}(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0)^T$.

Now the result holds after some algebra. ■

Proof of Theorem 4.

In ARTICULO CL2 we proved the asymptotic distribution of the first test statistic for LCM for binary data. Let us then prove in this case the asymptotic distribution of the second one, the other being similar.

A second-order Taylor expansion of $D_{\phi_1}(\boldsymbol{p}, \boldsymbol{q})$ around $(\boldsymbol{p}(\boldsymbol{\theta}_0), \boldsymbol{p}(\boldsymbol{\theta}_0))$ at $\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right), \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right)$ is given by (see the proof of Theorem 1)

$$\begin{aligned} D_{\phi_1}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right), \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right) &= \frac{\phi_1''(1)}{2}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right)^T \boldsymbol{D}^{-1}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right) \\ &\quad + o(||\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}(\boldsymbol{\theta}_0)||^2) + o(||\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right) - \boldsymbol{p}(\boldsymbol{\theta}_0)||^2), \end{aligned}$$

Therefore,

$$T^{\phi_1,\phi_2}_{A-B} = \sqrt{N}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right)^T \boldsymbol{D}^{-1}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right) + o_p(||1||).$$

Thus, the asymptotic distribution of T^{ϕ_1,ϕ_2}_{A-B} is the same as the one of X^tX with

$$X := \sqrt{N}\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\left(\boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^A}_{\phi_2}\right) - \boldsymbol{p}\left(\widehat{\boldsymbol{\theta}^B}_{\phi_2}\right)\right).$$

On the other hand, we already know that

$$\widehat{\boldsymbol{\theta}^A}_{\phi_2} - \boldsymbol{\theta}_0 = \boldsymbol{J}(\boldsymbol{\theta}_0)\left(\boldsymbol{A}_A^T\boldsymbol{A}_A\right)^{-1}\boldsymbol{A}_A^T\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}(\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}_0)) + o_p(N^{-1/2}),$$

$$\widehat{\boldsymbol{\theta}^B}_{\phi_2} - \boldsymbol{\theta}_0 = \boldsymbol{J}(\boldsymbol{\theta}_0)\left(\boldsymbol{A}_B^T\boldsymbol{A}_B\right)^{-1}\boldsymbol{A}_B^T\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}(\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}_0)) + o_p(N^{-1/2}),$$

Let us define

$$\boldsymbol{W}_A := \boldsymbol{A}_A\left(\boldsymbol{A}_A^T\boldsymbol{A}_A\right)^{-1}\boldsymbol{A}_A^T, \boldsymbol{W}_B := \boldsymbol{A}_B\left(\boldsymbol{A}_B^T\boldsymbol{A}_B\right)^{-1}\boldsymbol{A}_B^T.$$

Consequently, the asymptotic distribution of X coincides with the asymptotic distribution of

$$\sqrt{N}(\boldsymbol{W}_A - \boldsymbol{W}_B)\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}(\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}_0)).$$

Now,

$$\sqrt{N}(\hat{\boldsymbol{p}} - \boldsymbol{p}(\boldsymbol{\theta}_0)) \underset{N\longrightarrow\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^*)$$

being

$$\boldsymbol{\Sigma}^* = (\boldsymbol{W}_A - \boldsymbol{W}_B)\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\boldsymbol{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}(\boldsymbol{W}_A - \boldsymbol{W}_B).$$

Consequently, it suffices to show that $\mathbf{\Sigma}^*$ is a symmetric and idempotent matrix. Symmetry is trivial, whence it suffices to show idempotency. Notice that

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2}\mathbf{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} &= \boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2}\left[\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} - \boldsymbol{p}(\boldsymbol{\theta}_0)\boldsymbol{p}(\boldsymbol{\theta}_0)^T\right]\boldsymbol{D}_{\boldsymbol{p}(\boldsymbol{\theta}_0)}^{-1/2} \\ &= Id - \sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)}\sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)}^T \end{aligned}$$

From the proof of Theorem 1, we know that $\boldsymbol{W}_A\sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)} = \boldsymbol{W}_A\sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)} = \mathbf{0}$. Finally,

$$\mathbf{\Sigma}^* = (\boldsymbol{W}_A - \boldsymbol{W}_B)\left(Id - \sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)}\sqrt{\boldsymbol{p}(\boldsymbol{\theta}_0)}^T\right)(\boldsymbol{W}_A - \boldsymbol{W}_B) = (\boldsymbol{W}_A - \boldsymbol{W}_B)(\boldsymbol{W}_A - \boldsymbol{W}_B).$$

On the other hand,

$$\boldsymbol{W}_A\boldsymbol{W}_B = \boldsymbol{W}_B,\ \boldsymbol{W}_B\boldsymbol{W}_A = \boldsymbol{W}_B,\ \boldsymbol{W}_A\boldsymbol{W}_A = \boldsymbol{W}_A,\ \boldsymbol{W}_B\boldsymbol{W}_B = \boldsymbol{W}_B,$$

whence we conclude that

$$\mathbf{\Sigma}^* = (\boldsymbol{W}_A - \boldsymbol{W}_B)$$

and it is an idempotent matrix. We conclude that

$$T_{A-B}^{\phi_1,\phi_2} \underset{N\longrightarrow\infty}{\overset{\mathcal{L}}{\to}} \chi^2_{tr(\mathbf{\Sigma}^*)}.$$

To obtain the degrees of freedom we compute

$$\begin{aligned} tr(\mathbf{\Sigma}^*) &= tr(\boldsymbol{W}_A - \boldsymbol{W}_B) \\ &= tr(\boldsymbol{W}_A) - tr(\boldsymbol{W}_B) \\ &= tr\left(\boldsymbol{A}_A\left(\boldsymbol{A}_A^T\boldsymbol{A}_A\right)^{-1}\boldsymbol{A}_A^T\right) - tr\left(\boldsymbol{A}_B\left(\boldsymbol{A}_B^T\boldsymbol{A}_B\right)^{-1}\boldsymbol{A}_B^T\right) \\ &= tr\left(\boldsymbol{A}_A^T\boldsymbol{A}_A\left(\boldsymbol{A}_A^T\boldsymbol{A}_A\right)^{-1}\right) - tr\left(\boldsymbol{A}_B^T\boldsymbol{A}_B\left(\boldsymbol{A}_B^T\boldsymbol{A}_B\right)^{-1}\right) \\ &= h_1 - h_2 \end{aligned}$$

This finishes the proof. ■

References

[1] B. Abar and E. Loken. Self-regulated learning and self-directed study in a pre-college sample. *Learning and Individual Differences*, 20:25–29, 2010.

[2] S. Bacci, S. Pandolfi, and F. Pennoni. A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification*, 8(2):125–145, 2014.

[3] P. Biemer. *Latent Class Analysis and Survey Error*. John Wiley and Sons, New York (USA), 2011.

[4] M.W. Birch. A new proof of the pearson-fisher theorem. *Annals of Mathematical Statistics*, 35:817–824, 1964.

[5] C. Clogg. Latent class models for measuring. In *Latent trait and class models*, pages 173–205. Plenum, New York (USA), 1988.

[6] L. Collins and S. Lanza. *Latent class and latent transition analysis for the social, behavioral, and health sciences*. Wiley, New York (USA), 2010.

[7] N. Cressie and L. Pardo. Phi-divergence statistics. In A.H. Elshaarawi and W.W. Plegorich, editors, *Encyclopedia of environmetrics*, volume 13, pages 1551–1555. John Wiley and Sons, New York (USA), 2002.

[8] N. Cressie, L. Pardo, and M.C. Pardo. Size and power considerations for testing loglinear models using ϕ-divergence test statistics. *Statistica Sinica*, 13(2):555–570, 2003.

[9] N. Cressie and T.R.C. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B*, 8:440–464, 1984.

[10] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

[11] J.R. Dale. Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *Journal of the Royal Statistical Society Series B*, 41:48–59, 1986.

[12] B. Feldman, K. Masyn, and R. Conger. New approaches to studying behaviors: A comparison of methods for modelling longitudinal, categorical and adolescent drinking data. *Development Psycology*, 45(3):652–676, 2009.

[13] A. Felipe, P. Miranda, and L. Pardo. Minimum ϕ-divergence estimation in constrained latent class models for binary data. *Psychometrika*, 80(4):1020–1042, 2015.

[14] A. Formann. Linear logistic latent class analysis. *Biometrical Journal*, 24:171–190, 1982.

[15] A. Formann. Constrained latent class models: Theory and applications. *British Journal of Mathematics and Statistical Psicology*, 38:87–111, 1985.

[16] A. Formann. Linear logistic latent class analysis for polytomous data. *Journal of the Amearican Statistical Association*, 87:476–486, 1992.

[17] E. Genge. A latent class analysis of the public attitude towards the euro adoption in Poland. *Advances in Data Analysis and Classification*, 8(4):427–442, 2014.

[18] M. Gnaldi, S. Bacci, and F. Bartolucci. A multilevel finite mixture item response model to cluster examinees and schools. *Advances in Data Analysis and Classification*, In Press.

[19] L.A. Goodman. Exploratory latent structure analysis using Goth identifiable and unidentifiable models. *Biometrika*, 61:215–231, 1974.

[20] J. A. Hagenaars and A.L.M. Cutcheon. *Applied Latent Class Analysis*. Cambridge University Press, Cambridge (UK), 2002.

[21] M. Laska, K. Pash, K. Lust, M. Story, and E.L.A. Ehlinger. Latent class analysis of lifestyle characteristics and health risk behaviors among college youth. *Prevention Sciences*, 10:376–386, 2009.

[22] P. Lazarsfeld. The logical and mathematical foundation of latent structure analysis. In *Studies in Social Psycology in World War II*, volume IV, pages 362–412. Princeton University Press, 1950.

[23] D.A. Linzar and J.B. Lewis. polca: An r package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(10):1–29, 2011.

[24] N. Martin, R. Mata, and L. Pardo. Comparing two treatments in terms of the likelihood ratio order. *Journal of Statistical Computation and Simulation*, 85(17):3512–3534, 2015.

[25] D. L. Oberski. Beyond the number of classes: separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification*, In Press.

[26] L. Pardo. *Statistical inference based on divergence measures*. Chapman and Hall CRC, Boca Raton (USA), 2006.

[27] J. Uebersax. *Latent Structure Analysis*. Web document: http://www.john-uebersax.com/stat/index.htm, 2009.

Parameter / a	-1	-1/2	-1/4	-1/8	0	2/3	1	3/2	3
$\hat{\alpha}_{11}$	-1.959	-2.338	-2.375	-2.397	-2.470	-2.389	-2.374	-2.367	-2.451
$\hat{\alpha}_{12}$	-0.881	-1.478	-1.489	-1.484	-1.468	-1.515	-1.490	-1.510	-1.528
$\hat{\alpha}_{13}$	0.598	0.306	0.316	0.319	0.331	0.312	0.307	0.305	0.310
$\hat{\alpha}_{14}$	1.238	0.905	0.921	0.926	0.937	0.929	0.915	0.917	0.945
$\hat{\alpha}_{15}$	-0.018	-0.718	-0.729	-0.727	-0.729	-0.716	-0.727	-0.722	-0.718
$\hat{\alpha}_{16}$	1.985	1.974	1.984	1.987	2.000	1.994	1.994	1.997	2.016
$\hat{\alpha}_{17}$	-1.035	0.734	0.742	0.744	0.753	0.746	0.740	0.741	0.759
$\hat{\alpha}_{21}$	0.831	0.104	0.077	0.069	0.070	0.062	0.099	0.093	0.058
$\hat{\alpha}_{22}$	0.624	-0.083	-0.107	-0.114	-0.117	-0.114	-0.091	-0.096	-0.124
$\hat{\alpha}_{23}$	0.594	0.237	0.219	0.214	0.213	0.213	0.234	0.233	0.223
$\hat{\alpha}_{24}$	0.512	0.368	0.352	0.348	0.345	0.343	0.357	0.356	0.341
$\hat{\alpha}_{25}$	-0.710	-0.721	-0.721	-0.723	-0.721	-0.727	-0.725	-0.725	-0.730
$\hat{\alpha}_{26}$	0.612	0.881	0.893	0.898	0.896	0.890	0.865	0.864	0.873
$\hat{\alpha}_{27}$	-0.668	-0.254	-0.225	-0.215	-0.212	-0.206	-0.233	-0.227	-0.194
$\hat{\beta}_{11}$	-4.752	-4.920	-4.899	-4.806	-4.061	-3.541	-2.528	-2.499	-3.041
$\hat{\beta}_{12}$	-0.054	-0.403	-0.228	-0.252	-0.376	-0.461	-0.591	-0.614	-0.554
$\hat{\beta}_{13}$	1.830	1.045	1.241	1.229	1.103	1.020	0.858	0.860	0.987
$\hat{\beta}_{14}$	2.351	1.678	1.882	1.872	1.746	1.657	1.476	1.475	1.591
$\hat{\beta}_{15}$	-0.809	0.369	0.526	0.488	0.365	0.283	0.116	0.109	0.106
$\hat{\beta}_{16}$	3.144	2.888	3.101	3.094	2.969	2.890	2.697	2.700	2.838
$\hat{\beta}_{17}$	-1.118	-1.381	-2.573	-2.569	-2.568	-2.580	-2.580	-2.580	-2.580
$\hat{\beta}_{21}$	-1.574	-0.499	-0.479	-0.472	-0.471	-0.459	-0.489	-0.483	-0.442
$\hat{\beta}_{22}$	-0.942	-0.573	-0.554	-0.548	-0.546	-0.545	-0.562	-0.553	-0.527
$\hat{\beta}_{23}$	-1.140	-0.325	-0.304	-0.298	-0.296	-0.290	-0.314	-0.310	-0.294
$\hat{\beta}_{24}$	-0.726	-0.364	-0.347	-0.343	-0.340	-0.330	-0.354	-0.352	-0.332
$\hat{\beta}_{25}$	-0.412	-0.424	-0.412	-0.404	-0.404	-0.405	-0.399	-0.397	-0.371
$\hat{\beta}_{26}$	2.852	1.437	1.390	1.373	1.369	1.348	1.403	1.388	1.303
$\hat{\gamma}_2$	-0.901	-0.561	-0.547	-0.541	-0.539	-0.530	-0.540	-0.533	-0.505

Table 1: Power-divergence estimations of the parameters α_{ji}, β_{ji} and γ_2 for different values of a.

N	a	$mse_{\boldsymbol{\lambda}}$	$mse_{\boldsymbol{\eta}}$	$mse_{\mathbf{p}}$	$mse_{\mathbf{w}}$	$mse_{\boldsymbol{\lambda},\boldsymbol{\eta}}$	$mse_{\mathbf{p},\mathbf{w}}$
200	-1	1.7338	3.0843	0.3379	0.0576	2.4090	0.3267
	-0.5	0.0887	1.8628	0.0689	0.0402	0.9757	0.0678
	-0.25	0.0818	1.8230	0.0651	0.0396	0.9524	0.0641
	-0.125	0.0794	1.8081	0.0638	0.0394	0.9437	0.0629
	0	0.0777	1.8047	0.0628	0.0393	0.9412	0.0618
	0.667	0.0739	1.8205	0.0616	0.0399	0.9472	0.0608
	1	0.0736	1.8573	0.0619	0.0405	0.9655	0.0610
	1.5	0.0736	1.8997	0.0628	0.0412	0.9867	0.0619
	3	0.0795	2.0635	0.0694	0.0448	1.0715	0.0684
300	-1	0.5133	1.7204	0.1265	0.0338	1.1169	0.1228
	-0.5	0.0537	1.2411	0.0427	0.0273	0.6474	0.0420
	-0.25	0.0511	1.2171	0.0411	0.0270	0.6341	0.0406
	-0.125	0.0503	1.2101	0.0408	0.0269	0.6302	0.0402
	0	0.0498	1.2084	0.0403	0.0269	0.6291	0.0398
	0.667	0.0481	1.2197	0.0398	0.0272	0.6339	0.0393
	1	0.0477	1.2296	0.0399	0.0274	0.6387	0.0394
	1.5	0.0482	1.2557	0.0405	0.0279	0.6519	0.0400
	3	0.0517	1.3614	0.0444	0.0301	0.7065	0.0438
400	-1	0.1945	1.1522	0.0626	0.0237	0.6734	0.0610
	-0.5	0.0384	0.9291	0.0309	0.0208	0.4837	0.0305
	-0.25	0.0372	0.9236	0.0303	0.0206	0.4804	0.0299
	-0.125	0.0368	0.9180	0.0301	0.0206	0.4774	0.0297
	0	0.0365	0.9174	0.0299	0.0206	0.4769	0.0295
	0.667	0.0356	0.9242	0.0296	0.0207	0.4799	0.0292
	1	0.0356	0.9269	0.0297	0.0209	0.4812	0.0293
	1.5	0.0358	0.9464	0.0301	0.0211	0.4911	0.0298
	3	0.0380	1.0176	0.0326	0.0226	0.5278	0.0322
500	-1	0.0953	0.8881	0.0392	0.0186	0.4917	0.0383
	-0.5	0.0301	0.7543	0.0244	0.0169	0.3922	0.0241
	-0.25	0.0295	0.7503	0.0240	0.0168	0.3899	0.0237
	-0.125	0.0291	0.7452	0.0237	0.0167	0.3872	0.0234
	0	0.0288	0.7480	0.0236	0.0167	0.3884	0.0233
	0.667	0.0282	0.7505	0.0233	0.0168	0.3894	0.0231
	1	0.0283	0.7511	0.0236	0.0169	0.3897	0.0233
	1.5	0.0285	0.7646	0.0238	0.0171	0.3965	0.0235
	3	0.0303	0.8198	0.0258	0.0181	0.4250	0.0254

Table 2: mse of the simulation study for several sample sizes

N	a	$mse_{\boldsymbol{\lambda}}$	$mse_{\boldsymbol{\eta}}$	$mse_{\mathbf{p}}$	$mse_{\mathbf{w}}$	$mse_{\boldsymbol{\lambda},\boldsymbol{\eta}}$	$mse_{\mathbf{p},\mathbf{w}}$
600	-1	0.0543	0.6847	0.0268	0.0149	0.3695	0.0264
	-0.5	0.0240	0.6127	0.0194	0.0139	0.3184	0.0192
	-0.25	0.0236	0.6059	0.0191	0.0138	0.3147	0.0189
	-0.125	0.0234	0.6064	0.0190	0.0138	0.3149	0.0188
	0	0.0233	0.6063	0.0190	0.0138	0.3148	0.0188
	0.667	0.0228	0.6058	0.0188	0.0138	0.3143	0.0186
	1	0.0228	0.6073	0.0189	0.0139	0.3151	0.0187
	1.5	0.0230	0.6174	0.0192	0.0141	0.3202	0.0190
	3	0.0244	0.6581	0.0205	0.0149	0.3412	0.0203
700	-1	0.0353	0.5757	0.0206	0.0126	0.3055	0.0202
	-0.5	0.0203	0.5329	0.0166	0.0121	0.2766	0.0164
	-0.25	0.0200	0.5299	0.0164	0.0120	0.2749	0.0162
	-0.125	0.0198	0.5281	0.0163	0.0120	0.2739	0.0161
	0	0.0198	0.5306	0.0163	0.0120	0.2752	0.0161
	0.667	0.0194	0.5298	0.0161	0.0120	0.2746	0.0160
	1	0.0194	0.5342	0.0162	0.0120	0.2768	0.0160
	1.5	0.0196	0.5396	0.0164	0.0122	0.2796	0.0162
	3	0.0206	0.5703	0.0175	0.0127	0.2955	0.0173
800	-1	0.0263	0.5015	0.0169	0.0110	0.2639	0.0167
	-0.500	0.0178	0.4669	0.0144	0.0105	0.2423	0.0143
	-0.25	0.0176	0.4678	0.0143	0.0105	0.2427	0.0142
	-0.125	0.0175	0.4687	0.0143	0.0105	0.2431	0.0141
	0	0.0174	0.4699	0.0142	0.0105	0.2437	0.0141
	0.667	0.0172	0.4703	0.0141	0.0105	0.2437	0.0140
	1	0.0172	0.4673	0.0142	0.0105	0.2422	0.0140
	1.5	0.0173	0.4753	0.0143	0.0107	0.2463	0.0142
	3	0.0183	0.4996	0.0153	0.0112	0.2589	0.0151
1000	-1	0.0174	0.3938	0.0125	0.0088	0.2056	0.0124
	-0.5	0.0141	0.3801	0.0115	0.0086	0.1971	0.0114
	-0.25	0.0140	0.3786	0.0114	0.0086	0.1963	0.0113
	-0.125	0.0139	0.3785	0.0114	0.0085	0.1962	0.0112
	0	0.0139	0.3777	0.0113	0.0086	0.1958	0.0112
	0.667	0.0137	0.3803	0.0113	0.0086	0.1970	0.0112
	1	0.0137	0.3813	0.0113	0.0086	0.1975	0.0112
	1.5	0.0138	0.3855	0.0114	0.0087	0.1997	0.0113
	3	0.0143	0.4019	0.0120	0.0090	0.2081	0.0118
3000	-1	0.0045	0.1426	0.0037	0.0030	0.0736	0.0037
	-0.5	0.0045	0.1430	0.0037	0.0030	0.0737	0.0036
	-0.25	0.0045	0.1419	0.0036	0.0030	0.0732	0.0036
	-0.125	0.0045	0.1424	0.0036	0.0030	0.0734	0.0036
	0	0.0045	0.1428	0.0036	0.0030	0.0737	0.0036
	0.667	0.0045	0.1424	0.0036	0.0030	0.0734	0.0036
	1	0.0045	0.1421	0.0036	0.0030	0.0733	0.0036
	1.5	0.0045	0.1428	0.0037	0.0030	0.0736	0.0036
	3	0.0046	0.1446	0.0037	0.0031	0.0746	0.0037

Table 3: mse of the simulation study for several sample sizes

N	Size	-1	-0.5	-0.25	-0.125	0	0.6667	1	1.5	3
200	0.10	0.3938	0.2530	0.1791	0.1546	0.1357	0.0854	0.0791	0.0803	0.1489
	0.05	0.3412	0.1736	0.1072	0.0870	0.0724	0.0395	0.0364	0.0384	0.0916
300	0.10	0.2537	0.1831	0.1473	0.1341	0.1234	0.0915	0.0859	0.0864	0.1350
	0.05	0.1900	0.1169	0.0848	0.0735	0.0648	0.0436	0.0411	0.0423	0.0802
400	0.10	0.1971	0.1509	0.1297	0.1210	0.1133	0.0911	0.0875	0.0876	0.1261
	0.05	0.1318	0.0892	0.0700	0.0631	0.0579	0.0431	0.0410	0.0420	0.0718
500	0.10	0.1803	0.1441	0.1283	0.1223	0.1169	0.0991	0.0955	0.0955	0.1283
	0.05	0.1135	0.0819	0.0685	0.0637	0.0600	0.0482	0.0466	0.0476	0.0734
600	0.10	0.1540	0.1265	0.1147	0.1097	0.1057	0.0916	0.0888	0.0895	0.1166
	0.05	0.0920	0.0686	0.0597	0.0564	0.0535	0.0441	0.0427	0.0433	0.0641
700	0.10	0.1478	0.1258	0.1163	0.1119	0.1083	0.0963	0.0939	0.0940	0.1171
	0.05	0.0871	0.0680	0.0609	0.0580	0.0555	0.0476	0.0463	0.0469	0.0646
800	0.10	0.1394	0.1204	0.1125	0.1089	0.1058	0.0957	0.0938	0.0939	0.1147
	0.05	0.0802	0.0649	0.0588	0.0564	0.0544	0.0477	0.0466	0.0469	0.0635
900	0.10	0.1330	0.1170	0.1104	0.1074	0.1048	0.0957	0.0937	0.0938	0.1124
	0.05	0.0754	0.0624	0.0570	0.0550	0.0532	0.0468	0.0457	0.0462	0.0611
1000	0.10	0.1349	0.1201	0.1139	0.1115	0.1088	0.1006	0.0990	0.0991	0.1160
	0.05	0.0756	0.0638	0.0592	0.0572	0.0553	0.0493	0.0483	0.0488	0.0631
3000	0.10	0.1120	0.1082	0.1062	0.1057	0.1050	0.1018	0.1012	0.1011	0.1069
	0.05	0.0593	0.0558	0.0544	0.0538	0.0534	0.0513	0.0510	0.0509	0.0561
9000	0.10	0.1037	0.1023	0.1017	0.1013	0.1010	0.1002	0.0998	0.0995	0.1014
	0.05	0.0522	0.0512	0.0508	0.0506	0.0505	0.0499	0.0497	0.0497	0.0512

Table 4: Sizes of the test for different sample sizes and different values of a. Theoretical values are 0.10 and 0.05.

N	a	0	0.01	0.05	0.2	0.4	0.6	0.8	1	2
200	-1/8	0.1546	0.1522	0.1546	0.1565	0.1971	0.2622	0.3977	0.5991	0.9965
	0	0.1357	0.1335	0.1350	0.1378	0.1757	0.2391	0.3744	0.5800	0.9959
	2/3	0.0854	0.0854	0.0859	0.0899	0.1185	0.1772	0.3053	0.5154	0.9928
	1	0.0791	0.0787	0.0793	0.0832	0.1108	0.1679	0.2937	0.5037	0.9917
	3/2	0.0803	0.0803	0.0819	0.0847	0.1128	0.1705	0.2956	0.5056	0.9912
400	-1/8	0.1210	0.1270	0.1229	0.1356	0.1963	0.3462	0.5804	0.8024	1.0000
	0	0.1133	0.1189	0.1157	0.1284	0.1880	0.3373	0.5730	0.7976	1.0000
	2/3	0.0911	0.0957	0.0933	0.1045	0.1612	0.3053	0.5449	0.7801	1.0000
	1	0.0875	0.0915	0.0897	0.1002	0.1564	0.2993	0.5389	0.7762	0.9999
	3/2	0.0876	0.0926	0.0903	0.1009	0.1579	0.3004	0.5401	0.7755	0.9999
600	-1/8	0.1097	0.1134	0.1134	0.1404	0.2307	0.4461	0.7224	0.8993	1.000
	0	0.1057	0.1092	0.1090	0.1355	0.2253	0.4410	0.7188	0.8980	1.0000
	2/3	0.0916	0.0951	0.0942	0.1206	0.2070	0.4213	0.7049	0.8913	1.0000
	1	0.0888	0.0925	0.0915	0.1174	0.2032	0.4176	0.7021	0.8896	1.0000
	3/2	0.0894	0.0925	0.0914	0.1178	0.2038	0.4178	0.7019	0.8895	1.0000
800	-1/8	0.1089	0.1103	0.1156	0.1359	0.2668	0.5416	0.8149	0.9437	1.0000
	0	0.1058	0.1075	0.1127	0.1323	0.2629	0.5383	0.8131	0.9428	1.0000
	2/3	0.0957	0.0974	0.1021	0.1211	0.2484	0.5255	0.8053	0.9399	1.0000
	1	0.0938	0.0953	0.0996	0.1187	0.2460	0.5230	0.8040	0.9389	1.0000
	3/2	0.0939	0.0954	0.1000	0.1184	0.2460	0.5233	0.8034	0.9390	1.0000
1000	-1/8	0.1115	0.1085	0.1154	0.1436	0.3074	0.6169	0.8710	0.9669	1.0000
	0	0.1088	0.1062	0.1129	0.1407	0.3044	0.6145	0.8698	0.9664	1.0000
	2/3	0.1006	0.0974	0.1044	0.1314	0.2935	0.6063	0.8652	0.9648	1.0000
	1	0.0990	0.0958	0.1028	0.1295	0.2919	0.6044	0.8643	0.9643	1.0000
	3/2	0.0987	0.0964	0.1024	0.1291	0.2923	0.6045	0.8641	0.9642	1.0000
9000	-1/8	0.1013	0.1019	0.1214	0.5032	0.9427	0.9969	0.9998	1.0000	1.0000
	0	0.1010	0.1019	0.1212	0.5031	0.9426	0.9968	0.9998	1.0000	1.0000
	2/3	0.1002	0.1009	0.1205	0.5022	0.9427	0.9967	0.9998	1.0000	1.0000
	1	0.0998	0.1008	0.1203	0.5021	0.9426	0.9967	0.9998	1.0000	1.0000
	3/2	0.0995	0.1007	0.1203	0.5023	0.9428	0.9967	0.9998	1.0000	1.0000

Table 5: Power corresponding to level 0.10 for different values of N and a and for different values of the parameter (0, 0.05, 0.2, 0.4, 0.6, 0.8, 1, 2). Exact level corresponds to value 0

N	a	0	0.01	0.05	0.2	0.4	0.6	0.8	1	2
200	-1/8	0.0870	0.0863	0.0861	0.0889	0.1170	0.1677	0.2853	0.4883	0.9934
	0	0.0724	0.0715	0.0720	0.0745	0.1002	0.1487	0.2638	0.4672	0.9922
	2/3	0.0395	0.0400	0.0409	0.0428	0.0612	0.1008	0.2017	0.3990	0.9869
	1	0.0364	0.0366	0.0373	0.0395	0.0567	0.0952	0.1935	0.3883	0.9853
	3/2	0.0384	0.0387	0.0395	0.0416	0.0591	0.0991	0.1988	0.3942	0.9845
400	-1/8	0.0631	0.0678	0.0654	0.0726	0.1169	0.2391	0.4696	0.7265	0.9999
	0	0.0579	0.0622	0.0604	0.0666	0.1101	0.2308	0.4611	0.7204	0.9999
	2/3	0.0431	0.0467	0.0457	0.0508	0.0904	0.2047	0.4311	0.6978	0.9998
	1	0.0410	0.0446	0.0438	0.0482	0.0873	0.2002	0.4267	0.6934	0.9998
	3/2	0.0420	0.0457	0.0447	0.0495	0.0888	0.2026	0.4294	0.6946	0.9997
600	-1/8	0.0564	0.0585	0.0589	0.0761	0.1422	0.3313	0.6299	0.8538	1.0000
	0	0.0535	0.0553	0.0561	0.0729	0.1377	0.3260	0.6256	0.8516	1.0000
	2/3	0.0441	0.0457	0.0465	0.0622	0.1224	0.3073	0.6099	0.8432	1.0000
	1	0.0427	0.0441	0.0452	0.0603	0.1199	0.3045	0.6070	0.8413	1.0000
	3/2	0.0433	0.0450	0.0459	0.0610	0.1212	0.3067	0.6078	0.8410	1.0000
800	-1/8	0.0564	0.0560	0.0606	0.0735	0.1720	0.4279	0.7420	0.9155	1.0000
	0	0.0544	0.0539	0.0580	0.0713	0.1687	0.4241	0.7397	0.9145	1.0000
	2/3	0.0477	0.0472	0.0505	0.0630	0.1571	0.4118	0.7300	0.9101	1.0000
	1	0.0466	0.0458	0.0492	0.0616	0.1547	0.4103	0.7281	0.9091	1.0000
	3/2	0.0469	0.0463	0.0492	0.0626	0.1557	0.4116	0.7282	0.9084	1.0000
1000	-1/8	0.0572	0.0543	0.0603	0.0784	0.2058	0.5094	0.8155	0.9483	1.0000
	0	0.0553	0.0528	0.0587	0.0763	0.2029	0.5069	0.8139	0.9479	1.0000
	2/3	0.0493	0.0469	0.0518	0.0700	0.1931	0.4973	0.8084	0.9456	1.0000
	1	0.0483	0.0462	0.0508	0.0691	0.1915	0.4949	0.8074	0.9449	1.0000
	3/2	0.0488	0.0466	0.0513	0.0697	0.1926	0.4958	0.8078	0.9450	1.0000
9000	-1/8	0.5006	0.0512	0.0635	0.3873	0.9144	0.9951	0.9996	1.0000	1.0000
	0	0.0505	0.0510	0.0633	0.3871	0.9143	0.9951	0.9996	1.0000	1.0000
	2/3	0.0499	0.0503	0.0624	0.3862	0.9142	0.9951	0.9995	1.0000	1.0000
	1	0.0497	0.0500	0.0622	0.3861	0.9142	0.9951	0.9996	1.0000	1.0000
	3/2	0.0497	0.0500	0.0621	0.3859	0.9143	0.9951	0.9996	1.0000	1.0000

Table 6: Power corresponding to level 0.05 for different values of N and a and for different values of the parameter (0, 0.05, 0.2, 0.4, 0.6, 0.8, 1, 2). Exact level corresponds to value 0