Phi-Divergence test statistics applied to latent class models for binary data

P. Miranda[1], A. Felipe[2], and N. Martín[3]

Abstract In this paper we present two new families of test statistics for studying the problem of goodness-of-fit of some data to a latent class model for dichotomous questions based on phi-divergence measures. We also treat the problem of selecting the best model out of a sequence of nested latent class models. In both problems, we study the asymptotic distribution of the corresponding test statistics, showing that they share the same behavior as the corresponding maximum likelihood test statistic.

1 Introduction and basic concepts

Latent class modelling is based on the distinction between manifest and latent variables. While manifest variables can be directly observed, like socioeconomic variables, item responses in a questionnaire or some codification of observed behavior, latent variables cannot be observed or measured by means of a yardstick.

In this paper dichotomous observed variables are considered. Consider a set $\mathcal{S}$ of N people: $\mathcal{S} := \{S_1, ..., S_N\}$. Each person S_v is asked to answer to k dichotomous items $I_1, ..., I_k$; let us denote by y_{vi} the answer (right= 1, wrong=0) of person S_v to item I_i and $\boldsymbol{y}_v := (y_{v1}, ..., y_{vk})$ a generic pattern given by person S_v.

A categorical latent variable (categorical unobservable variable) is postulated to exist, whose different levels partition set $\mathcal{S}$ into m mutually exclusive and exhaustive *latent classes* $C_1, ..., C_m$ whose corresponding weights are $w_1, ..., w_m$. Let us denote

$$p_{ji} = Pr(y_{vi} = 1 | S_v \in C_j),\ j = 1, ..., m,\ i = 1, ..., k.$$

Complutense University of Madrid pmiranda@ucm.es · Complutense University of Madrid afelipe@ucm.es · Complutense University of Madrid nimartin@ucm.es

Let $\mathbf{y}_\nu$ be a possible answer vector. We shall assume that in each class the answers for the different questions are stochastically independent.

We will denote by N_ν, $\nu = 1, ..., 2^k$, the number of times that the sequence $\mathbf{y}_\nu$ appears in an N-sample and

$$\widehat{\boldsymbol{p}} := (N_1/N, ..., N_{2^k}/N)$$

the corresponding proportions. The likelihood function L is given by

$$L_{\boldsymbol{y}_1,...,\boldsymbol{y}_{2^k}}(w_1, ..., w_m, p_{11}, ..., p_{mk}) = \frac{N!}{\prod_{\nu=1}^{2^k} n_\nu!} \prod_{\nu=1}^{2^k} \Pr(\boldsymbol{y}_\nu)^{n_\nu}, \tag{1}$$

where n_ν denotes the sample result of N_ν. In this model the unknown parameters are $w_j, j = 1, ..., m$ and $p_{ji}, j = 1, ..., m, i = 1, ..., k$. In order to avoid the problem of obtaining uninterpretable estimations for the item latent probabilities lying outside the interval $[0, 1]$, some authors [?], [?], [?], [?], [?], [?] proposed a linear-logistic parametrization given by

$$p_{ji} = \frac{\exp(x_{ji})}{1 + \exp(x_{ji})},\ j = 1, ..., m,\ i = 1, ..., k,$$

and

$$w_j = \frac{\exp(z_j)}{\sum_{h=1}^{m} \exp(z_h)},\ j = 1, ..., m.$$

Next, restrictions are introduced relating parameters x_{ji}, z_j to some explanatory parameters $\lambda_r, r = 1, ..., t$ and $\eta_s, s = 1, ..., u$, so the final model is given by

$$p_{ji} = \frac{\exp(\sum_{r=1}^{t} q_{jir}\lambda_r + c_{ji})}{1 + \exp(\sum_{r=1}^{t} q_{jir}\lambda_r + c_{ji})},\ j = 1, ..., m,\ i = 1, ..., k, \tag{2}$$

and

$$w_j = \frac{\exp(\sum_{r=1}^{u} v_{jr}\eta_r + d_j)}{\sum_{h=1}^{m} \exp(\sum_{r=1}^{u} v_{hr}\eta_r + d_h)},\ j = 1, ..., m, \tag{3}$$

where $\boldsymbol{Q}_r = (q_{jir}), \boldsymbol{C} = (c_{ji}), \boldsymbol{V} = (v_{jr})$ are fixed. Consequently, in this case the vector of unknown parameters is $\boldsymbol{\theta} := (\boldsymbol{\lambda}, \boldsymbol{\eta})$. It is not difficult to establish [?] that

$$\log L(w_1, ..., w_m, p_{11}, ..., p_{mk}) = -ND_{Kullback}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) + constant.$$

Thus, varying the divergence measure considered, we obtain a family of estimators that includes the MLE.

2 Goodness-of-fit tests

LCM for binary data fit is assessed by comparing the observed classification frequencies to the expected frequencies predicted by the LCM for binary data. When dealing with the MLE, the difference is formally assessed with a likelihood ratio test statistic or with a chi-square test statistic:

$$G^2 = 2N \sum_{\nu=1}^{2^k} \hat{p}_\nu \log \frac{\hat{p}_\nu}{p(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})},\ X^2 = \sum_{\nu=1}^{2^k} \frac{\left(n_s - Np(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})\right)^2}{Np(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\eta}})} \tag{4}$$

It is known that the asymptotic distribution of the test statistics G^2 and X^2 is a chi-square distribution with $2^k - (u + t) - 1$ degrees of freedom [**?**]. These statisitcs can be extended in two ways: first, differences betweeen observed and expected values can be measured in terms of a divergence measure. Next, estimation of the parameters to obtain the expected values can be obtained in terms of a divergence measure.

Definition 1. We define the **phi-divergence family** of test statistics for testing goodness-of-fit for latent class models for binary data as

$$T_{\phi_1}^{\phi_2} := \frac{2N}{\phi_1''(1)} D_{\phi_1}\left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\hat{\boldsymbol{\theta}}_{\phi_2})\right), \tag{5}$$

where we are using ϕ_2 for estimation and ϕ_1 for comparing with the observed data.

For this family the following holds.

Theorem 1. *Under the hypothesis that the LCM for binary data with parameters* $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_t)$ *and* $\boldsymbol{\eta} = (\eta_1, ..., \eta_u)$ *holds, the asymptotic distribution of the family of test statistics* $T_{\phi_1}^{\phi_2}$ *given in (5) is a chi-square distribution with* $2^k - (u + t) - 1$ *degrees of freedom.*

It is noteworthy that the asymptotic distribution depends on neither ϕ_1 nor ϕ_2, i.e. it is the same for any functions ϕ_1 and ϕ_2 considered.

3 Nested latent class models

Suppose that our model fits the data, i.e. we conclude that the data can be explained through a LCM with m classes. Then, it could be the case that several sets of parameters could fit the data. If two LCM fit the data but one of them has a reduced number of parameters, then this model should be considered the more appropriate. In this section we deal with the problem of selecting the best model from a nested sequence of LCM.

In general, we shall assume that we have s LCM $\{M_l\}_{l=1,\ldots,s}$ in such a way that the parameter space associated to M_l, $l = 1, ..., s$, is $\boldsymbol{\Theta}_{M_l}$ and

$$\boldsymbol{\Theta}_{M_s} \subset \boldsymbol{\Theta}_{M_{s-1}} \subset .... \subset \boldsymbol{\Theta}_{M_1} \subset \mathbb{R}^r$$

holds. Let us denote $\dim(\boldsymbol{\Theta}_{M_l}) = h_l$; $l = 1, ..., s$, with

$$h_s < h_{s-1} < .... < h_1 \leq r,$$

i.e., the parameters of one LCM are a subset of the parameters of the previous LCM. Our strategy is to test successively

$$H_{l+1} : \boldsymbol{\theta} \in \boldsymbol{\Theta}_{M_{l+1}} \text{ against } H_l : \boldsymbol{\theta} \in \boldsymbol{\Theta}_{M_l},\ l = 1, ..., s-1, \tag{6}$$

and we continue to test as long as the null hypothesis is accepted, and choose the LCM M_l with parameter space $\boldsymbol{\Theta}_{M_l}$ according to the first l satisfying that H_{l+1} is rejected (as null hypothesis) in favor of H_l (as alternative hypothesis).

The classical expressions for solving (**??**) are

$$G^2_{A-B} = 2\sum_{\nu=1}^{2^k} n_\upsilon \log \frac{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^A\right)}{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)}, X^2_{A-B} = N\sum_{\nu=1}^{2^k} \frac{\left(p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^A\right) - p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)\right)^2}{p\left(\boldsymbol{y}_\nu, \widehat{\boldsymbol{\theta}}^B\right)} \tag{7}$$

Hence, proceeding as in the previous section, we can define two new families of test statistics. A generalization of G is given by

$$S^{\phi_1,\phi_2}_{A-B} = \frac{2N}{\phi_1''(1)}\left(D_{\phi_1}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A_{\phi_2})\right) - D_{\phi_1}\left(\hat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B_{\phi_2})\right)\right), \tag{8}$$

and a generalization of X is

$$T^{\phi_1,\phi_2}_{A-B} = \frac{2N}{\phi_1''(1)} D_{\phi_1}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}^A_{\phi_2}), \boldsymbol{p}(\widehat{\boldsymbol{\theta}}^B_{\phi_2})\right). \tag{9}$$

Now, the following can be shown.

Theorem 2. *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 (**??**), it follows*

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

4 An example with real data

In order to shed light about the results established in the previous sections, let us deal with a real data. We consider the interview data collected by [?] and analized later in [?]; this model is explained in [?] and [?]. The experiment consists in evaluating the answers of 3398 schoolboys to two questions about their membership in the "leading crowd" on two occasions t_1 and t_2 (October, 1957 and May, 1958). Thus, in this model we have 4 questions and there are four manifest variables (answers to both questions at both moments); these answers can only be "low" (value 0) and "high" (value 1), so that the manifest variables are dichotomous; thus, there are 16 possible vector of answers. The sample data, i.e., the number of times each possible vector of answers appears, is given in Table 2.

October, 1957/ May, 1958	00	01	10	11
00	554	338	97	85
01	281	531	75	184
10	87	56	182	171
11	49	110	140	458

Table 1 The set of data collected by Coleman.

Next, 4 latent classes are considered, namely
$C_1 \equiv$ low agreement in question 1 and low agreement in question 2.
$C_2 \equiv$ low agreement in question 1 and high agreement in question 2.
$C_3 \equiv$ high agreement in question 1 and low agreement in question 2.
$C_4 \equiv$ high agreement in question 1 and high agreement in question 2.

Consequently, there are 16 probability values p_{ji} to be estimated; let us start with the problem of goodness-of-fit; for this, we consider the first hypothesis appearing in [?], namely "The attitudinal changes between times t_1 and t_2 are dependent on the positions (low, high) of the respective classes on the underlying attitudinal scales at t_1". This implies that the probabilities depend just on the definition of the class. Thus, a model with 8 parameters λ_i is considered; λ_1 means low agreement in the first question at time t_1, λ_2 means high agreement in the first question at time t_1, λ_3 means low agreement in the second question at time t_1, λ_4 means high agreement in the second question at time t_1, and $\lambda_5, \lambda_6, \lambda_7, \lambda_8$ are the same parameters at time t_2; indeed, as we have four possible combinations of answers at each time, we have four parameters for t_1 and four for t_2; moreover, only two parameters at time t_1 appear at matrix $\mathbf{Q}_i$ due to the definition of the classes; and the same applies for t_2. We write the values for matrices $\mathbf{Q}_i$ as they appear in

[?] in Table 3. In our notation, the matrices $\mathbf{Q}_i$ can be derived considering the i-th column in the table and dividing it in four columns of four elements each (each corresponding to a latent class).

Class	Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
1	1	1	0	0	0	0	0	0	0
	2	0	0	1	0	0	0	0	0
	3	0	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	1	0
2	1	1	0	0	0	0	0	0	0
	2	0	0	0	1	0	0	0	0
	3	0	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	0	1
3	1	0	1	0	0	0	0	0	0
	2	0	0	1	0	0	0	0	0
	3	0	0	0	0	0	1	0	0
	4	0	0	0	0	0	0	1	0
4	1	0	1	0	0	0	0	0	0
	2	0	0	0	1	0	0	0	0
	3	0	0	0	0	0	1	0	0
	4	0	0	0	0	0	0	0	1

Table 2 The model design according to Formann

Note that the hypothesis is that the attitudinal changes between times t_1 and t_2 are dependent upon the items as well as on the classes. For this reason, the part corresponding each latent class can be partitioned in four submatrices of size 2×4. The submatrices lying on the main diagonal are the same by the hypothesis defining the model and the two other submatrices are null. The differences among them are due to the differences in the latent classes. Next, $c_{ij} = 0, \ \forall i, j$ (as we have explained when values c_{ij} were introduced in Section 1). Finally, 4 parameters η_j are considered, taking as matrix $\mathbf{V}$ the identity matrix and $d_j = 0, \forall j$. Remark however that one of these parameters can be removed as they sum up to 1. Thus, we have a model with 11 parameters.

It is noteworthy that our model assumes that answers to the questions are conditionally independent given the latent class. In this example, we are dealing with repeated responses to two questions, so this assumption may be unrealistic. However, this assumption is made in the original paper of Goodman [?] and also in [?], and we follow this assumption for the sake of the example.

In order to study if the data are from a LCM for binary data in the conditions explained before, we shall consider the particular family of phi-divergence measures introduced and studied in [?]: The *power divergence family*. This family is obtained from

$$\phi(x) \equiv \phi_a(x) = \begin{cases} \frac{1}{a(a+1)}(x^{a+1} - x - a(x-1)) & a \neq 0, a \neq -1 \\ x \log x - x + 1 & a = 0 \\ -\log x + x - 1 & a = -1 \end{cases} \quad (10)$$

In Felipe et al. (2015) it was established, on the basis of a simulation study, that a competitive alternative to the MLE is the MϕE obtained from Eq. (**??**) with $a = 2/3$, i.e.,

$$\hat{\boldsymbol{\theta}}_{2/3} = arg \min_{(\boldsymbol{\lambda},\boldsymbol{\eta}) \in \boldsymbol{\Theta}} D_{2/3}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})),$$

being

$$D_{2/3}(\hat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) = \frac{9}{10} \left(\sum_{j=1}^{2^k} \frac{\hat{p}_j^{5/3}}{p_j(\boldsymbol{\lambda}, \boldsymbol{\eta})^{2/3}} - 1 \right).$$

Therefore, in order to study the most general situation, we are going to consider in our study the MϕE obtained with $\phi(x)$ defined in Eq. (10) for $a = 2/3$ in order to get an estimation of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$**.** Remark however that in this case the values obtained are very similar to those obtained in [**?**] using MLE. In Table 4 we present the values obtained for these parameters, as well as the estimation of the probabilities and the weights of the latent classes.

Parameter		Parameter	
$\hat{\lambda}_1$	-2.34292610	$\hat{p}_{1,1}$	0.08762969
$\hat{\lambda}_2$	1.72393168	$\hat{p}_{1,2}$	0.30144933
$\hat{\lambda}_3$	-0.84040580	$\hat{p}_{1,3}$	0.11256540
$\hat{\lambda}_4$	1.56524945	$\hat{p}_{1,4}$	0.28671773
$\hat{\lambda}_5$	-2.06480043	$\hat{p}_{2,1}$	0.08762969
$\hat{\lambda}_6$	2.29928080	$\hat{p}_{2,2}$	0.82710532
$\hat{\lambda}_7$	-0.91137901	$\hat{p}_{2,3}$	0.11256540
$\hat{\lambda}_8$	2.01252338	$\hat{p}_{2,4}$	0.88210569
$\hat{\eta}_1$	0.50480183	$\hat{p}_{3,1}$	0.8463457
$\hat{\eta}_2$	0.16964329	$\hat{p}_{3,2}$	0.30144933
$\hat{\eta}_3$	-0.87356633	$\hat{p}_{3,3}$	0.90881746
$\hat{\eta}_4$	-0.00424661	$\hat{p}_{3,4}$	0.28671773
$\hat{w}_1$	0.38936544	$\hat{p}_{4,1}$	0.84863457
$\hat{w}_2$	0.27848377	$\hat{p}_{4,2}$	0.82710532
$\hat{w}_3$	0.09811597	$\hat{p}_{4,3}$	0.90881746
$\hat{w}_4$	0.23403482	$\hat{p}_{4,4}$	0.88210569

Table 3 Estimations of the parameters for Example in Section 4.

Once the estimations of the parameters obtained, we are interested in studying the goodness-of-fit of our data to the model proposed at the beginning of the section. Following Section 2, we shall consider the family of test statistics, $T_{\phi_a}^{\phi_{2/3}}$, obtained from $\phi_a(x)$ with $a = -1, -1/2,\ 0,\ 2/3,\ 1, 1.5, 2, 2.5$

and 3,i.e.,

$$T_{\phi_a}^{\phi_{2/3}} = \begin{cases} \frac{2N}{a(a+1)} \left(\sum_{\nu=1}^{2^k} \frac{\widehat{p}_v^{a+1}}{p\left(\boldsymbol{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3}\right)^a} - 1 \right) & \text{if } a \neq 0, -1 \\ 2 \sum_{\nu=1}^{2^k} n_v \log \frac{\widehat{p}_v}{p\left(\boldsymbol{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3}\right)} & a = 0 \\ 2N \sum_{\nu=1}^{2^k} p\left(\boldsymbol{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3}\right) \log \frac{p\left(\boldsymbol{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3}\right)}{\widehat{p}_v} & a = -1 \end{cases} . \quad (11)$$

The results are presented in Table 5.

a	-1	-1/2	0	2/3	1	3/2	2	5/2	3
$T_{\phi_a}^{\phi_{2/3}}$	1.279	1.278	1.277	1.277	1.277	1.277	1.278	1.279	1.281

Table 4 Statistics for different divergence measures.

On the other hand, the distribution of these statistics under the null hypothesis that the model fits the data is a χ^2 with 16-11-1=4 degrees of freedom; as $\chi^2_{4;0.05} = 9.49$, we conclude that we have no evidence to reject our model.

Notice that the values for all test statistics are very similar; this was expected, as the sample size under consideration is big enough ($N = 3398$) to apply the asymptotic result of Theorem 1. Indeed, a similar result is obtained if the classical test statistics are used (see [**?**]).

As a conclusion, we could say that the LCM proposed (that we will call M_1) fits our data; however, a question arises: Is it possible to find a latent model with a reduced number of parameters that also fits the data? In [**?**], the following reduced models are studied:

M_2 : Attitudinal changes between the two moments are dependent on the latent classes but are independent on the items.

M_3: Attitudinal changes between the two moments are independent both on the items and on the latent classes.

M_4 : There are no attitudinal changes.

These different models imply different number of parameters λ_i. More concretely, model M_2 needs six parameters λ_i, model M_3 needs five parameters and finally model M_4 needs four parameters. The corresponding matrices Q_i for these models can be found in [**?**].

As for M_1, $c_{ij} = 0, \ \forall i, j$ and 4 parameters η_j are considered, taking matrix V as the identity matrix and $d_j = 0, \forall j$.

We can observe that

$$\boldsymbol{\Theta}_{M_1} \supset \boldsymbol{\Theta}_{M_2} \supset \boldsymbol{\Theta}_{M_3} \supset \boldsymbol{\Theta}_{M_4},$$

being $\boldsymbol{\Theta}_{M_i}$ the parameter space associated to the LCM M_i. Therefore, we have a nested sequence of LCM. We consider again $\hat{\boldsymbol{\theta}}_{2/3}$ in order to estimate the parameters of the different models. For testing, we consider the family of phi-divergences test statistics $S_{A-B}^{\phi_a,\phi_{2/3}}$ and $T_{A-B}^{\phi_a,\phi_{2/3}}$ given in (8) and (9), with $\phi_{2/3}(x)$ defined in (10). For ϕ_a we shall take $a = -1, -1/2, 0, 2/3, 3/2, 2, 5/2$ and 3. In Table 6 we present the results obtained.

a/Model	$M_1 - M_2$	$M_2 - M_3$	$M_3 - M_4$	$M_1 - M_2$	$M_2 - M_3$	$M_3 - M_4$
-1	3.761	4.610	31.465	3.431	4.613	31.005
-1/2	3.757	4.593	30.977	3.417	4.604	30.845
0	3.755	4.584	30.769	3.403	4.595	30.722
2/3	3.754	4.578	30.626	3.386	4.585	30.616
1	3.754	4.580	30.659	3.378	4.580	30.587
3/2	3.756	4.586	30.820	3.366	4.574	30.574
2	3.759	4.599	30.991	3.355	4.570	30.597
5/2	3.763	4.617	31.347	3.344	4.566	30.655
3	3.769	4.641	31.765	3.334	4.563	30.749
$\chi^2_{i;0.05}$	5.99	3.84	3.84	5.99	3.84	3.84

Table 5 Results for Example in Section 4 for statistics S (left) and T (right).

As a conclusion, we can adopt LCM M_2 as the best model in all cases. As before, the values obtained are very similar, due to the asymptotic results developed in Section 3.

5 Simulation

Sections 2 and 3 present theoretical results for testing hypothesis in latent models with binary data. These results give the asymptotic distribution theory for the phi-divergence test statistics given in (5), (8) and (9) under the null hypothesis. In this section we present a simulation study to analyze the behavior of these statistics in small samples. We shall analyze the test statistics given in (5).

As before, we consider the minimum power divergence with $a = 2/3$ for estimations. We compare the different test statistics of the family $T_{\phi_a}^{\phi_{2/3}}$ defined in (11). The theoretical LCM with binary data that we shall consider in our simulation study is given by a theoretical model with 5 dichotomous questions and 10 latent classes; next, 7 parameters λ_j and 6 parameters η_k are considered; the corresponding matrices of the model are

$$\mathbf{Q}_1 = \begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{pmatrix},\ \mathbf{Q}_2 = \begin{pmatrix} 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0 \end{pmatrix},\ \mathbf{Q}_3 = \begin{pmatrix} 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0 \end{pmatrix},$$

$$\mathbf{Q}_4 = \begin{pmatrix} 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0 \end{pmatrix},\ \mathbf{Q}_5 = \begin{pmatrix} 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \end{pmatrix},\ \mathbf{Q}_6 = \begin{pmatrix} 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \end{pmatrix},$$

$$\mathbf{Q}_7 = \begin{pmatrix} 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \end{pmatrix}.$$

Matrix $\mathbf{C}$ is the null matrix. Matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \begin{pmatrix} 1\,0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,1 \\ 0\,1\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,1\,1 \\ 0\,0\,0\,0\,1\,0 \end{pmatrix},$$

while $\mathbf{d} = \mathbf{0}$. The theoretical values for vector $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are

$$\boldsymbol{\lambda}_0 = (\lambda_1^0, ..., \lambda_7^0) = (-3, -2, -1, 0, 1, 2, 3), \ \boldsymbol{\eta}_0 = (\eta_1^0, ..., \eta_6^0) = (0.5, 1, 1.5, 2, 2.5, 3).$$

We shall also consider different values of a; more concretely, we consider the values $a = -0.5, 0, 2/3, 1$.

For each value of a we consider R =10 000 simulations and we reproduce the study for different sample sizes: 200, 300, 400, 500 and 1000. Note that as we have to estimate 13 parameters and there are 32 possible answer vectors, sample sizes under 200 provide frequencies very low, so the results are very different for different samples, and cannot be used to obtain reliable estimations. We must not forget that for $a = 0$ and $a = 1$ we have the likelihood ratio test and the chi-square ratio test statistics, respectively, but the unknown parameters are estimated using the minimum power divergence estimator with $a = 2/3$ instead of the maximum likelihood estimator.

We consider as nominal size $\alpha = 0.05$ and compute the *simulated exact size*

$$\hat{\alpha}_n^a := \frac{\sharp T_{\phi_a}^{\phi_{2/3}} > \chi^2_{g.l.;0.05}}{R}.$$

As explained in [**?**], we only consider the test statistics whose simulated exact size $\hat{\alpha}_n^a$ satisfies

$$|logit(1 - \hat{\alpha}_n^a) - logit(1 - \alpha)| \leq 0.35 \tag{12}$$

where $logit(p) = log(\frac{p}{1-p})$. As a consequence, we only take under consideration the test statistics such that

$$\hat{\alpha}_n^a \in (0.0357, 0.0695). \tag{13}$$

At the same time we obtain the *simulated exact power* for different alternative hypothesis. More concretely, we shall consider a model with a new parameter λ_8 whose corresponding matrix $\mathbf{Q}_8$ is given by

$$\mathbf{Q}_8 = \begin{pmatrix} 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 1\,1\,1\,1\,1 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \end{pmatrix},$$

and where this new parameter takes different values, namely -3, -2, -1.5, -1, -0.8, 0, 0.7, 0.9, 1, 1.3, 1.5, 2. Each of these values is related to an alternative hypothesis, except when considering value 0, that corresponds to the null hypothesis.

Simulating observations from each alternative hypothesis we get the simulated exact power for such alternatives

$$\hat{\beta}^a := \frac{\sharp T_{\phi_a}^{\phi_{2/3}} > \chi^2_{g.l.;0.05}}{R}.$$

In Table 7 we present the simulated exact size as well as the simulated exact power for different values of a and in Figures 1 to 5 we present the pictures for each sample size of the different alternative hypothesis for the test statistic $a = -1/2, 0, 2/3, 1$.

N	a	-3	-2	-1.5	-1	-0.8	0	0.7	0.9	1	1.3	1.5	2
200	-.5	0.776	0.524	0.494	0.478	0.451	0.404	0.544	0.635	0.680	0.822	0.893	0.987
	0	0.483	0.166	0.159	0.150	0.145	0.109	0.230	0.323	0.383	0.608	0.749	0.959
	2/3	0.318	0.061	0.056	0.047	0.046	0.035	0.104	0.176	0.237	0.466	0.633	0.928
	1	0.291	0.049	0.042	0.033	0.035	0.026	0.088	0.152	0.212	0.436	0.607	0.919
300	-.5	0.800	0.447	0.409	0.372	0.337	0.268	0.456	0.583	0.660	0.851	0.938	0.998
	0	0.620	0.198	0.178	0.164	0.141	0.094	0.251	0.394	0.489	0.756	0.891	0.995
	2/3	0.509	0.104	0.087	0.077	0.062	0.040	0.159	0.291	0.381	0.686	0.853	0.993
	1	0.488	0.087	0.071	0.062	0.051	0.033	0.141	0.268	0.356	0.671	0.843	0.993
400	-.5	0.831	0.408	0.382	0.319	0.282	0.193	0.420	0.592	0.681	0.910	0.972	0.999
	0	0.733	0.224	0.202	0.171	0.149	0.081	0.287	0.476	0.587	0.873	0.960	0.999
	2/3	0.667	0.142	0.122	0.100	0.086	0.042	0.212	0.402	0.521	0.844	0.950	0.999
	1	0.653	0.127	0.109	0.087	0.073	0.036	0.199	0.385	0.505	0.836	0.948	0.999
500	-.5	0.879	0.412	0.362	0.306	0.257	0.144	0.421	0.630	0.736	0.946	0.989	1.000
	0	0.820	0.262	0.229	0.201	0.162	0.074	0.335	0.559	0.678	0.932	0.986	1.000
	2/3	0.782	0.190	0.160	0.134	0.107	0.044	0.275	0.506	0.638	0.920	0.983	1.000
	1	0.774	0.176	0.145	0.119	0.094	0.039	0.262	0.496	0.628	0.917	0.983	1.000
1000	-.5	0.992	0.491	0.522	0.463	0.319	0.081	0.608	0.882	0.953	0.999	1.000	1.000
	0	0.991	0.321	0.349	0.336	0.278	0.065	0.585	0.875	0.950	0.999	1.000	1.000
	2/3	0.991	0.235	0.263	0.252	0.239	0.051	0.568	0.872	0.948	0.999	1.000	1.000
	1	0.991	0.217	0.243	0.230	0.228	0.047	0.565	0.871	0.948	0.999	1.000	1.000

Table 6 Power for different values of N and a and for differenet values of λ_8 (-3, -2, -1.5, -1, -0.8, 0, 0.7, 0.9, 1, 1.3, 1.5, 2). Exact level corresponds to $\lambda_8 = 0$.

As it can be observed in Table 7 (see the column corresponding to 0) and Figures 1 to 5, the simulated level is outside the interval given in (13) for $a = 0, -0.5$ for all sample sizes under consideration; besides, for sample sizes $N = 400, 500, 1000$, the test statistic corresponding to $a = 1$ lays inside this interval. Notice that the test statistic for $a = 2/3$ is the only one laying in this interval for any sample size. As a straightforward conclusion, the test statistic for $a = 2/3$ seems to be the best one for sample sizes $N = 200, 300$, and we just need to choose between $a = 1$ and $a = 2/3$ for $N = 400, 500, 1000$. For making this decision, we focus on the simulated power values, noting that they are higher for $a = 2/3$ than for $a = 1$; we then conclude that $a = 2/3$ seems to show a better behavior than the likelihood ratio test statistic and

Pearson test statistic (with estimations obtained through $a = 2/3$ instead of maximum likelihood) when dealing with LCM for binary data.

6 Conclusions

In this paper we have introduced phi-divergence test statistics in the context of LCM for binary data. In a previous paper, it has been shown that phi-divergence estimators can be a useful tool in this framework; now, we have treated two new problems: the problem of goodness-of-fit and the problem of selecting the best model throughout a nested sequence of models. Classically, as it can be seen for instance in [**?**], these problems have been solved on the basis of the likelihood-ratio-test and the chi-square test statistic. In this paper, we have derived two families of test statistics based on phi-divergence measures that generalize the likelihood-ratio-test and the chi-square test statistic; for these new statistics, we have obtained their asymptotic distribution under the null hypothesis of that LCM fits the data, showing that it coincides with the that of the likelihood-ratio-test and the chi-square test statistic; thus, they show the same behavior as the classical statistics for big sample sizes. To see the applicability of this theory, we have considered a real data situation studied by Goodman [**?**] and Formann [**?**].

At this point, an interesting problem arises: are there differences for small or moderate sample sizes? To deal with this problem, we have carried out a simulation study; from this study, it seems that the phi-divergence test statistic for $a = 2/3$ shows a behavior that can compete with the likelihood-ratio-test and the chi-square test statistic.

7 Section Heading

Use the template `author_TMIDS.tex` together with the Springer document class `svmult.cls` to style the various elements of your chapter content in the Springer layout.

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations. And please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

8 Section Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Use the standard `equation` environment to typeset your equations, e.g.

$$a \times b = c \; , \tag{14}$$

however, for multiline equations we recommend to use the `eqnarray` environment[1].

$$\begin{aligned} a \times b &= c \\ \mathbf{a} \cdot \mathbf{b} &= \mathbf{c} \end{aligned} \tag{15}$$

8.1 Subsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

> Please do not use quotation marks when quoting texts! Simply use the `quotation` environment – it will automatically render Springer's preferred layout.

8.1.1 Subsubsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.1, see also Fig. 1[2]

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

---

[1] In physics texts please activate the class option `vecphys` to depict your vectors in ***boldface-italic*** type - as is customary for a wide range of physical subjects

[2] If you copy text passages, figures, or tables from other works, you must obtain *permission* from the copyright holder (usually the original publisher). Please enclose the signed permission with the manucript. The sources must be acknowledged either in the captions, as footnotes or in a separate section of the book.

Paragraph Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

For typesetting numbered lists we recommend to use the `enumerate` environment – it will automatically render Springer's preferred layout.

1. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
   a. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
   b. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
2. Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.

Subparagraph Heading

In order to avoid simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8, see also Fig. 2.

For unnumbered list we recommend to use the `itemize` environment – it will automatically render Springer's preferred layout.

- Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development, cf. Table 7.

Fig. 1 If the width of the figure is less than 7.8 cm use the `sidecapion` command to flush the caption on the left side of the page. If the figure is positioned at the top of the page, align the sidecaption with the top of the figure – to achieve this you simply need to use the optional argument `[t]` with the `sidecaption` command

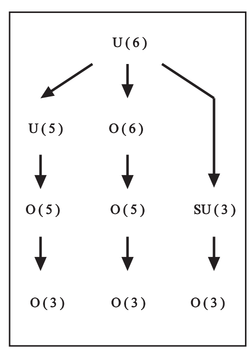
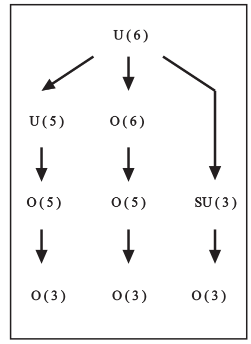


Fig. 2 If the width of the figure is less than 7.8 cm use the `sidecapion` command to flush the caption on the left side of the page. If the figure is positioned at the top of the page, align the sidecaption with the top of the figure – to achieve this you simply need to use the optional argument `[t]` with the `sidecaption` command



  - Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.
  - Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.

- Livelihood and survival mobility are oftentimes coutcomes of uneven socioeconomic development.

Run-in Heading Boldface Version Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Run-in Heading Italic Version Use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Table 7 Please write your table caption here

Classes	Subclass	Length	Action Mechanism
Translation	mRNA[a]	22 (19–25)	Translation repression, mRNA cleavage
Translation	mRNA cleavage	21	mRNA cleavage
Translation	mRNA	21–22	mRNA cleavage
Translation	mRNA	24–26	Histone and DNA Modification

[a] Table foot note (with superscript)

9 Section Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

If you want to list definitions or the like we recommend to use the Springer-enhanced `description` environment – it will automatically render Springer's preferred layout.

Type 1 That addresses central themes pertainng to migration, health, and disease. In Sect. 7, Wilson discusses the role of human migration in infectious disease distributions and patterns.

Type 2 That addresses central themes pertainng to migration, health, and disease. In Sect. 8.1, Wilson discusses the role of human migration in infectious disease distributions and patterns.

9.1 Subsection Heading

In order to avoid simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Use the LaTeX automatism for all your cross-references and citations citations as has already been described in Sect. 8.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

If you want to emphasize complete paragraphs of texts we recommend to use the newly defined Springer class option `graybox` and the newly defined environment `svgraybox`. This will produce a 15 percent screened box 'behind' your text.

If you want to emphasize complete paragraphs of texts we recommend to use the newly defined Springer class option and environment `svgraybox`. This will produce a 15 percent screened box 'behind' your text.

9.1.1 Subsubsection Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Please note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Theorem 3. *Theorem text goes here.*

Definition 2. Definition text goes here.

Proof. Proof text goes here. □

Paragraph Heading

Instead of simply listing headings of different levels we recommend to let every heading be followed by at least a short passage of text. Further on please use the LaTeX automatism for all your cross-references and citations as has already been described in Sect. 8.

Note that the first line of text that follows a heading is not indented, whereas the first lines of all subsequent paragraphs are.

Theorem 4. *Theorem text goes here.*

Definition 3. Definition text goes here.

Proof. Proof text goes here. □

Acknowledgements If you want to include acknowledgments of assistance and the like at the end of an individual chapter please use the `acknowledgement` environment – it will automatically render Springer's preferred layout.

Appendix

When placed at the end of a chapter or contribution (as opposed to at the end of the book), the numbering of tables, figures, and equations in the appendix section continues on from that in the main text. Hence please *do not* use the `appendix` command when writing an appendix at the end of your chapter or contribution. If there is only one the appendix is designated "Appendix", or "Appendix 1", or "Appendix 2", etc. if there is more than one.

$$a \times b = c \tag{16}$$

References may be *cited* in the text by number.[3] The reference list should ideally be *sorted* in alphabetical order. If there are several works by the same author, the following order should be used:

1. all works by the author alone, ordered chronologically by year of publication

[3] Make sure that all references from the list are cited in the text. Those not cited should be moved to a separate *Further Reading* section or chapter.

2. all works by the author with a coauthor, ordered alphabetically by coauthor
3. all works by the author with several coauthors, ordered chronologically by year of publication.

Examples with `spbasic.bst` for a paper in a Contribution [1], a Journal article [2], a Monograph [3] and a Thesis [4] are:

References

1. Brandt J, Hein W (2001) Polymer materials in joint surgery. In: Grellmann W, Seidler S (eds) Deformation and Fracture Behavior of Polymers. Engineering materials. Springer, Berlin Heidelberg New York
2. Che M, Grellmann W, Seidler S (1997) Appl Polym Sci 64:1079–1090
3. Kipphan H (2000) Handbook of printmedia. Springer, Berlin Heidelberg New York
4. Ross DW (1977) Lysosomes and storage diseases. MA Thesis, Columbia University, New York