Monografías del Seminario Matemático García de Galdeano **33**, 301–308 (2006)

PHI-DIVERGENCE TEST STATISTICS IN MULTINOMIAL SAMPLING FOR HIERARCHICAL SEQUENCES OF LOGLINEAR MODELS WITH LINEAR CONSTRAINTS

Nirian Martín and Leandro Pardo

Abstract. We consider nested sequences of hierarchical loglinear models when expected frequencies are subject to linear constraints and we study the problem of finding the model in the the nested sequence that is able to explain more clearly the given data. It will be necessary to give a method to estimate the parameters of the loglinear models and also a procedure to choose the best model among the models considered in the nested sequence under study. These two problems will be solved using the ϕ-divergence measures. We estimate the unknown parameters using the minimum ϕ-divergence estimator (Martín and Pardo [8]) which can be considered as a generalization of the maximum likelihood estimator (Haber and Brown [5]) and we consider a ϕ-divergence test statistic (Martín [7]) that generalize the likelihood ratio test as well as the chi-square test statistic, for testing two nested loglinear models.

Keywords: Loglinear models, multinomial sampling, maximum likelihood estimator, minimum Phi-divergence estimator, Phi-divergence test statistics.

AMS classification: 62H15, 62H17.

§1. Introduction

Loglinear models define a multiplicative structure on the expected cell frequencies of a contingency table. We shall assume that we have k cells (if $k = I \times J$ we get a two-way contingency table) and we denote them by $C_1,\dots,C_k$. Given a random sample $Y_1,Y_2,\dots,Y_n$ with realizations from $\mathscr{Y} = \{C_1,\dots,C_k\}$ we denote by $\widehat{\mathbf{p}} = (\widehat{p}_1,\dots,\widehat{p}_k)^T$ with

$$\widehat{p}_j = \frac{N_j}{n} \quad \text{and} \quad N_j = \sum_{i=1}^{n} I_{\{C_j\}}(Y_i), \quad j = 1,\dots,k. \tag{1}$$

Assuming multinomial sampling and denoting by $p_j(\boldsymbol{\theta}_0) = \Pr(C_j)$, $j = 1,\dots,k$, the statistic $(N_1,\dots,N_k)$ is obviously sufficient for the statistical model under consideration and is multinomially distributed with parameters n and $\mathbf{p}(\boldsymbol{\theta}_0) = (p_1(\boldsymbol{\theta}_0),\dots,p_k(\boldsymbol{\theta}_0))$. We shall denote,

$$m_j(\boldsymbol{\theta}_0) \equiv E(N_j) = np_j(\boldsymbol{\theta}_0), \quad j = 1,\dots,k, \tag{2}$$

and $\mathbf{m}(\boldsymbol{\theta}_0) = (m_1(\boldsymbol{\theta}_0),\dots,m_k(\boldsymbol{\theta}_0))^T$.

Given a $k\times(t+1)$ matrix $\mathbf{X}$ and $\mathrm{rank}(\mathbf{X})=t+1$, the set

$$C(\mathbf{X})=\{\log\mathbf{m}(\boldsymbol{\theta})\in\mathbb{R}^k:\log\mathbf{m}(\boldsymbol{\theta})=\mathbf{X}\boldsymbol{\theta},\ \boldsymbol{\theta}\in\mathbb{R}^{t+1}\}\tag{3}$$

represents the class of the loglinear models associated with $\mathbf{X}$. We suppose, in the following, that $\mathbf{J}=\left(1,\ldots^{(k}\ldots,1\right)^T\in C(\mathbf{X})$. Taking into account (2), the parameter space is defined by $\Theta'=\{\boldsymbol{\theta}\in\mathbb{R}^{t+1}:\log\mathbf{m}(\boldsymbol{\theta})=\mathbf{X}\boldsymbol{\theta}\ \text{ and }\ \mathbf{J}^T\mathbf{m}(\boldsymbol{\theta})=n\}$. Now in addition to the previous model we shall assume that we have $r-1<t$ linear constrains defined by

$$\mathbf{C}^T\mathbf{m}(\boldsymbol{\theta})=\mathbf{d}^*,\tag{4}$$

where $\mathbf{C}$ and $\mathbf{d}^*$ are $k\times(r-1)$ and $(r-1)\times 1$ matrices, respectively. If we consider the linear constraint $\mathbf{J}^T\mathbf{m}(\boldsymbol{\theta})=n$, associated to the multinomial sampling, we can write the parameter space for this new model by

$$\Theta^*=\{\boldsymbol{\theta}\in\mathbb{R}^{t+1}:\log\mathbf{m}(\boldsymbol{\theta})=\mathbf{X}\boldsymbol{\theta}\ \text{ and }\ \mathbf{L}^T\mathbf{m}(\boldsymbol{\theta})=\mathbf{d}\},$$

where $\mathbf{L}=(\mathbf{J},\mathbf{C})$, $\mathbf{d}=\left(n,(\mathbf{d}^*)^T\right)^T$ and $\mathrm{rank}(\mathbf{L})=\mathrm{rank}(\mathbf{L}^T,\mathbf{d})=r$.

The problem that has motivated our research involves a nested sequence of hypotheses

$$H_l:\mathbf{p}=\mathbf{p}(\boldsymbol{\theta}),\quad \boldsymbol{\theta}\in\Theta_0^{(l)},\quad l=1,\ldots,m,\quad m\le t<k-1,\tag{5}$$

where $\Theta_0^{(1)}\supset\Theta_0^{(2)}\supset\cdots\supset\Theta_0^{(m)}$ with $\dim(\Theta_0^{(l)})=t_l+1$, $\mathrm{rank}(\mathbf{L}_l)=r_l$, $l=1,\ldots,m$, such that

$$t_{l+1}\le t_l\ \text{ and }\ r_{l+1}\ge r_l,\quad l=1,\ldots,m-1,\tag{6}$$

where at least one of both inequalities is a strict inequality. In this framework, there is an integer m^* $(1\le m^*\le m)$ for which H_{m^*} is true but H_{m^*+1} is not true. A common strategy for making inference on m^* (e.g., Cressie and Read [2, p. 42]) is to test successively,

$$H_{Null}:H_{l+1}\ \text{ against }\ H_{Alt}:H_l,\quad l=1,\ldots,m-1,\tag{7}$$

where we continue to test as long as the null hypothesis is accepted, and we infer m^* to be the first l for which H_{l+1} is rejected as a null hypothesis. The full operating characteristics of this sequence of tests of nested hypotheses are not known. Our goal in this paper is to present ϕ-divergence test statistics for testing a sequence of nested hypotheses as given in (5).

§2. Minimum ϕ-divergence estimator

Since the parameter values in $\{\Theta_0^{(l)}:l=1,\ldots,m\}$ are generally unknown, most tests require their estimation. In this context the maximum likelihood estimator, under the linear constrains given in (4) is defined by

$$\widehat{\boldsymbol{\theta}}^{(r)}=\arg\max_{\boldsymbol{\theta}\in\Theta^*}\mathbf{h}^T\boldsymbol{\theta},$$

where $\mathbf{h}^T = (\mathbf{n}^*)^T \mathbf{X}$ and $\mathbf{n}^*$ is an observation from $(N_1, \ldots, N_k)$. It is a simple exercise to establish that equivalently $\widehat{\boldsymbol{\theta}}^{(r)}$ can be defined by

$$\widehat{\boldsymbol{\theta}}^{(r)} = \arg\min_{\boldsymbol{\theta} \in \Theta^*} D_{Kullback}\left(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})\right), \tag{8}$$

where $D_{Kullback}\left(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})\right)$ is the Kullback-Leibler (see Kullback [6]) divergence between the probability vectors $\widehat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\theta})$,

$$D_{Kullback}\left(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})\right) = \sum_{j=1}^{k} \widehat{p}_j \log \frac{\widehat{p}_j}{p_j(\boldsymbol{\theta})}.$$

The definition (8) hints at a much more general inference framework based on divergence measures, which was investigated by Martín and Pardo [8]. In the next several paragraphs, we give the essential details of the framework for estimation and hypothesis testing there.

Consider the $\phi-$divergence defined by Csiszár [3] and Ali and Silvey [1]

$$D_{\phi}(\mathbf{p}, \mathbf{q}) \equiv \sum_{j=1}^{k} q_j \phi\left(\frac{p_j}{q_j}\right), \qquad \phi \in \Phi^*, \tag{9}$$

where Φ^* is the class of all convex functions $\phi(x)$, $x > 0$, such that at $x = 1$, $\phi(1) = 0$, $\phi''(1) > 0$, and at $x = 0$, $0\,\phi(0/0) = 0$ and $0\,\phi(p/0) = p \lim_{u\to\infty} \phi(u)/u$. For every $\phi \in \Phi^*$ that is differentiable at $x = 1$, the function $\psi(x) \equiv \phi(x) - \phi'(1)(x-1)$ also belongs to Φ^*. Then we have $D_{\psi}(\mathbf{p}, \mathbf{q}) = D_{\phi}(\mathbf{p}, \mathbf{q})$, and ψ has the additional property that $\psi'(1) = 0$. Because the two divergence measures are equivalent, we can consider the set Φ^* to be equivalent to the set $\Phi \equiv \Phi^* \cap \{\phi : \phi'(1) = 0\}$. In what follows, we give our theoretical results for $\phi \in \Phi$, but we often apply them to choices of functions in Φ^*. For more details about ϕ-divergences see Pardo[10].

Based in (8) and (9), in the cited paper of Martín and Pardo, was considered the minimum ϕ-divergence estimator in loglinear models when we have some linear constraints and multinomial sampling, $\widehat{\boldsymbol{\theta}}_{\phi}^{(r)}$, is given by

$$\widehat{\boldsymbol{\theta}}_{\phi}^{(r)} = \arg\min_{\boldsymbol{\theta} \in \Theta^*} D_{\phi}(\widehat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})). \tag{10}$$

In the next section we shall use this estimator to define a family of test statistics for testing the nested hypotheses $\{H_l : l = 1, \ldots, m\}$ given in (5).

§3. Phi-divergence test statistics

In this section for testing nested hypotheses $\{H_l : l = 1, \ldots, m\}$ given in (5), we test

$$H_{Null} : H_{l+1} \text{ against } H_{Alt} : H_l, \quad l = 1, \ldots, m-1,$$

using the family of test statistics

$$T_{\phi_1,\phi_2}^{(l)} = \frac{2n}{\phi_1''(1)} D_{\phi_1}\left(\mathbf{p}\left(\widehat{\boldsymbol{\theta}}_l^{(r),\phi_2}\right), \mathbf{p}\left(\widehat{\boldsymbol{\theta}}_{l+1}^{(r),\phi_2}\right)\right), \tag{11}$$

where $\widehat{\boldsymbol{\theta}}_l^{(r),\phi_2}$ and $\widehat{\boldsymbol{\theta}}_{l+1}^{(r),\phi_2}$ are defined by (10). When $T_{\phi_1,\phi_2}^{(l)} > c$, we reject H_{Null} in (7), where c is specified so that the size of the test is α:

$$\Pr\left(T_{\phi_1,\phi_2}^{(l)} > c \,\middle|\, H_{l+1}\right) = \alpha, \quad \alpha \in (0,1). \tag{12}$$

In the next theorem we establish that, under $H_{Null} : H_{l+1}$, the test statistic $T_{\phi_1,\phi_2}^{(l)}$ converges in distribution to a chi-squared distribution with $t_l - t_{l+1} - r_l + r_{l+1}$ degrees of freedom ($\chi^2_{t_l - t_{l+1} - r_l + r_{l+1}}$), $l = 1, \ldots, m-1$. Thus, c could be chosen as the $(1-\alpha)$-th quantile of a $\chi^2_{t_l - t_{l+1} - r_l + r_{l+1}}$ distribution,

$$c = \chi^2_{t_l - t_{l+1} - r_l + r_{l+1}}(1-\alpha), \tag{13}$$

where $\Pr(\chi^2_f \le \chi^2_f(p)) = p$. Notice that, when $\phi_1(x) = \phi_2(x) = x\log x - x + 1$, we obtain the usual likelihood-ratio test, and that, when $\phi_1(x) = \frac{1}{2}(x-1)^2$ and $\phi_2 = x\log x - x + 1$, we obtain the Pearson test statistic (e.g. Haber and Brown [5]).

Theorem 1. *Suppose that data* $(N_1, \ldots, N_k)$ *are multinomially distributed according to the loglinear model (3). Consider the nested sequence of hypotheses given in (5) and (6). Choose functions* ϕ_1 *and* $\phi_2 \in \Phi$*. Then for testing* $H_{Null} : H_{l+1}$ *against* $H_{Alt} : H_l$*, the asymptotic null distribution of the* ϕ*-divergence test statistic* $T_{\phi_1,\phi_2}^{(l)}$ *is a chi-squared distribution with* $t_l - t_{l+1} - r_l + r_{l+1}$ *degrees of freedom.*

Proof. Based on Theorem 2 in Martín and Pardo [8], it is not difficult to establish that

$$\sqrt{n}\left(\mathbf{p}\left(\widehat{\boldsymbol{\theta}}_i^{(r),\phi_2}\right) - \mathbf{p}(\boldsymbol{\theta}_0)\right) = \mathbf{R}_i\sqrt{n}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o_P(1),$$

with $\mathbf{R}_i = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i\mathbf{H}_i(\boldsymbol{\theta}_0)\mathbf{X}_i^T$, $i = l, l+1$, where

$$\begin{aligned}\mathbf{H}_i(\boldsymbol{\theta}_0) = (\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i)^{-1} - (\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{L}_i \\ \times\left(\mathbf{L}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i)^{-1}\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{L}_i\right)^{-1} \\ \times\mathbf{L}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i)^{-1}, \quad i = l, l+1.\end{aligned}$$

Therefore, $T_{\phi_1,\phi_2}^{(l)} = \mathbf{Z}_l^T\mathbf{Z}_l + o_P(1)$, being

$$\begin{aligned}\mathbf{Z}_l &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}\sqrt{n}\left(\mathbf{p}\left(\widehat{\boldsymbol{\theta}}_l^{(r),\phi_2}\right) - \mathbf{p}\left(\widehat{\boldsymbol{\theta}}_{l+1}^{(r),\phi_2}\right)\right) \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}\sqrt{n}\left(\mathbf{p}\left(\widehat{\boldsymbol{\theta}}_l^{(r),\phi_2}\right) - \mathbf{p}(\boldsymbol{\theta}_0)\right) - \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}\sqrt{n}\left(\mathbf{p}\left(\widehat{\boldsymbol{\theta}}_{l+1}^{(r),\phi_2}\right) - \mathbf{p}(\boldsymbol{\theta}_0)\right) \\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}(\mathbf{R}_l - \mathbf{R}_{l+1})\sqrt{n}(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o_P(1).\end{aligned}$$

The asymptotic distribution of the ϕ-divergence test statistic (11) will be a chi-squared iff the matrix $\boldsymbol{\Sigma}_{Z_l} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}(\mathbf{R}_l - \mathbf{R}_{l+1})\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}(\mathbf{R}_l - \mathbf{R}_{l+1})^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}$, where the matrix $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{p}(\boldsymbol{\theta}_0)\mathbf{p}(\boldsymbol{\theta}_0)^T$ is idempotent and symmetric.

It is clear that

$$\begin{aligned}\mathbf{R}_i\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i\mathbf{H}_i(\boldsymbol{\theta}_0)\mathbf{X}_i^T\left(\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)} - \mathbf{p}(\boldsymbol{\theta}_0)\mathbf{p}(\boldsymbol{\theta}_0)^T\right)\\ &= \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i\mathbf{H}_i(\boldsymbol{\theta}_0)\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}, \quad i = l, l+1,\end{aligned}$$

and $\mathbf{K}_i = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2}\mathbf{R}_i\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{1/2} = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{1/2}\mathbf{X}_i\mathbf{H}_i(\boldsymbol{\theta}_0)\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{1/2}$ is a symmetric matrix. Therefore to establish that $\boldsymbol{\Sigma}_{Z_l} = (\mathbf{K}_l - \mathbf{K}_{l+1})(\mathbf{K}_l - \mathbf{K}_{l+1})$ is an idempotent matrix will be enough to see that $\mathbf{K}_l - \mathbf{K}_{l+1}$ is an idempotent matrix ($\boldsymbol{\Sigma}_{Z_l} = \mathbf{K}_l - \mathbf{K}_{l+1}$). We establish that

i) $\mathbf{K}_i\mathbf{K}_i = \mathbf{K}_i, \quad i = l, l+1,$

ii) $\mathbf{K}_l\mathbf{K}_{l+1} = \mathbf{K}_{l+1}.$

Part i) follows because $\mathbf{H}_i(\boldsymbol{\theta}_0)\mathbf{X}_i^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_i\mathbf{H}_i(\boldsymbol{\theta}_0) = \mathbf{H}_i(\boldsymbol{\theta}_0),\ i = l, l+1.$ Part ii) follows taking into account that $\mathbf{X}_{l+1}$ is a submatrix of $\mathbf{X}_l$. There exists a matrix $\mathbf{B}$ such that $\mathbf{X}_{l+1} = \mathbf{X}_l\mathbf{B}$ and

$$\begin{aligned}\mathbf{X}_l\mathbf{H}_l(\boldsymbol{\theta}_0)\mathbf{X}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_{l+1} = \mathbf{X}_l\mathbf{B} - \mathbf{X}_l&\left(\mathbf{X}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_l\right)^{-1}\mathbf{X}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{L}_l\\ &\times\left(\mathbf{L}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_l\left(\mathbf{X}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_l\right)^{-1}\mathbf{X}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{L}_l\right)^{-1}\\ &\times\mathbf{L}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_l\mathbf{B}.\end{aligned}$$

Multiplying on the right side by $\mathbf{H}_{l+1}(\boldsymbol{\theta}_0)$, the last term is zero because $\mathbf{L}_l$ is a submatrix of $\mathbf{L}_{l+1}$ and $\mathbf{H}_{l+1}(\boldsymbol{\theta}_0)\mathbf{B}_{l+1}(\boldsymbol{\theta}_0) = \mathbf{0}_{(t_{l+1}+1)\times r_{l+1}}$, therefore, $\mathbf{L}_l^T\mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{X}_l\mathbf{B}\mathbf{H}_{l+1}(\boldsymbol{\theta}_0) = \mathbf{0}_{r_l\times(t_l+1)}$.

The degrees of freedom of $T_{\phi_1,\phi_2}^{(l)}$ coincides with the trace of the matrix $\boldsymbol{\Sigma}_{Z_l}$. It is not difficult to establish that

$$\operatorname{tr}(\mathbf{K}_i) = t_i + 1 - r_i, \quad i = l, l+1,$$

therefore

$$\operatorname{tr}(\mathbf{K}_l - \mathbf{K}_{l+1}) = t_l - t_{l+1} - r_l + r_{l+1}. \quad \square$$

To test the nested sequence of hypotheses $\{H_l : l = 1, \ldots, m\}$ referred previously, we need an asymptotic independence result for the sequence of test statistics $T_{\phi_1,\phi_2}^{(1)}, T_{\phi_1,\phi_2}^{(2)}, \ldots, T_{\phi_1,\phi_2}^{(m^*)}$, where m^* is the integer $1 \leq m^* \leq m$ for which H_{m^*} is true but H_{m^*+1} is not true. This result is given in the theorem below.

Theorem 2. *Suppose that data $(N_1, \ldots, N_k)$ are multinomially distributed according to the loglinear model (3). We first test $H_{Null} : H_l$ against $H_{Alt} : H_{l-1}$, followed by $H_{Null} : H_{l+1}$ against $H_{Alt} : H_l$. Then, under the hypothesis H_{l+1}, the statistics $T_{\phi_1,\phi_2}^{(l-1)}$ and $T_{\phi_1,\phi_2}^{(l)}$ are asymptotically independent.*

Proof. The statistic $T_{\phi_1,\phi_2}^{(l)}$ can be written in the way,

$$T_{\phi_1,\phi_2}^{(l)} = \sqrt{n}\left(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\right)^T\mathbf{M}_l^T\mathbf{M}_l\sqrt{n}\left(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)\right) + o_P(1), \tag{14}$$

where

$$\mathbf{M}_l = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{R}_l - \mathbf{R}_{l+1}) \quad \text{and} \quad \mathbf{R}_i = \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{X}_i \mathbf{H}_i(\boldsymbol{\theta}_0) \mathbf{X}_i^T, \quad i = l, l+1.$$

Similarly,

$$T_{\phi_1,\phi_2}^{(l-1)} = \sqrt{n}\,(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0))^T \mathbf{M}_{l-1}^T \mathbf{M}_{l-1} \sqrt{n}\,(\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) + o_P(1).$$

By Theorem 4 in Searle [11], the quadratic forms $T_{\phi_1,\phi_2}^{(l)}$ and $T_{\phi_1,\phi_2}^{(l-1)}$ are asymptotically independent if $\mathbf{M}_l^T \mathbf{M}_l \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{M}_{l-1}^T \mathbf{M}_{l-1} = \mathbf{0}_{k\times k}$. We have

$$\mathbf{M}_l^T \mathbf{M}_l \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{M}_{l-1}^T \mathbf{M}_{l-1} = \mathbf{M}_l^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} (\mathbf{R}_l - \mathbf{R}_{l+1}) \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{R}_{l-1} - \mathbf{R}_l) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0)}^{-1/2} \mathbf{M}_{l-1}^T,$$

and since $\mathbf{H}_l(\boldsymbol{\theta}_0)\mathbf{R}_l(\boldsymbol{\theta}_0) = \mathbf{0}_{(t_l+1)\times r_l}$, we have $\mathbf{M}_l^T \mathbf{M}_l \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{M}_{l-1}^T \mathbf{M}_{l-1} = \mathbf{0}_{k\times k}$, because,

$$\mathbf{M}_l \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{M}_{l-1}^T = \mathbf{K}_l - \mathbf{K}_{l+1}\mathbf{K}_{l-1} - \mathbf{K}_l + \mathbf{K}_{l+1}\mathbf{K}_l = \mathbf{0}_{k\times k}. \quad \square$$

In general, theoretical results for the test statistic $T_{\phi_1,\phi_2}^{(l)}$ under alternative hypotheses are not easy to obtain. An exception to this is when there is a contiguous sequence of alternatives that approach the null hypothesis H_{l+1} at the rate of $O\left(n^{-1/2}\right)$.

Consider the multinomial probability vector

$$\mathbf{p}_n \equiv \mathbf{p}(\boldsymbol{\theta}_0) + \frac{\mathbf{s}}{\sqrt{n}}, \quad \boldsymbol{\theta}_0 \in \Theta_{l+1} \text{ and } \boldsymbol{\theta}_0 \text{ unknown}, \tag{15}$$

where $\mathbf{s} \equiv (s_1,\ldots,s_k)^T$ is a fixed $k\times 1$ vector such that $\sum_{j=1}^k s_j = 0$, and n is the total-count parameter of the multinomial distribution. As $n \to \infty$, the sequence of multinomial probabilities $\{\mathbf{p}_n\}_{n\in\mathbb{N}}$ converges to a multinomial probability in H_{l+1} at the rate of $O\left(n^{-1/2}\right)$. We call

$$H_{l+1,n} : \mathbf{p}_n = \mathbf{p}(\boldsymbol{\theta}_0) + \frac{\mathbf{s}}{\sqrt{n}}, \quad \boldsymbol{\theta}_0 \in \Theta_{l+1} \text{ and } \boldsymbol{\theta}_0 \text{ unknown}, \tag{16}$$

a sequence of *contiguous alternative hypotheses*, here contiguous to the null hypothesis H_{l+1}.

Now consider testing $H_{Null} : H_{l+1}$ against $H_{Alt} : H_{l+1,n}$, using the test statistic $T_{\phi_1,\phi_2}^{(l)}$ given by (11). The power of this test is $\pi_n^{(l)} \equiv \Pr\left(T_{\phi_1,\phi_2}^{(l)} > c \,\middle|\, H_{l+1,n}\right)$. In what is to follow, we show that under the alternative $H_{l+1,n}$, and as $n \to \infty$, $T_{\phi_1,\phi_2}^{(l)}$ converges in distribution to a non-central chi-squared random variable with non-centrality parameter μ, where μ is given in Theorem3, and $t_l - t_{l+1} - r_l + r_{l+1}$ degrees of freedom $\left(\chi^2_{t_l - t_{l+1} - r_l + r_{l+1},\mu}\right)$. Consequently, as $n \to \infty$,

$$\pi_n^{(l)} \to \Pr\left(\chi^2_{t_l - t_{l+1} - r_l + r_{l+1},\mu} > c\right). \tag{17}$$

Theorem 3. *Suppose that $(N_1,\ldots,N_k)$ is multinomially distributed according to the loglinear model (3). The asymptotic distribution of the statistic $T_{\phi_1,\phi_2}^{(l)}$, under the contiguous alternative hypotheses (16), is chi-squared with $t_l - t_{l+1} - r_l + r_{l+1}$ degrees of freedom and non-centrality parameter*

$$\mu = \mathbf{s}^T \left(\mathbf{X}_l \mathbf{H}_l(\boldsymbol{\theta}_0) \mathbf{X}_l^T - \mathbf{X}_{l+1} \mathbf{H}_{l+1}(\boldsymbol{\theta}_0) \mathbf{X}_{l+1}^T\right) \mathbf{s}.$$

Proof. By (14), we have $T^{(l)}_{\phi_1,\phi_2} = \mathbf{Z}_l^T \mathbf{Z}_l + o_P(1)$, where

$$\mathbf{Z}_l = \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{R}_l - \mathbf{R}_{l+1}) \sqrt{n} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0)) = (\mathbf{K}_l - \mathbf{K}_{l+1}) \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} \sqrt{n} (\widehat{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta}_0))$$

and $\mathbf{Z}_l \xrightarrow[n\to\infty]{L} \mathcal{N}\left(\boldsymbol{\mu}_s^{(l)}, \boldsymbol{\Sigma}_s^{(l)}\right)$, being

$$\boldsymbol{\mu}_s^{(l)} = \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{R}_l - \mathbf{R}_{l+1}) \mathbf{s} = (\mathbf{K}_l - \mathbf{K}_{l+1}) \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{s}$$

and

$$\boldsymbol{\Sigma}_s^{(l)} = \mathbf{D}^{-1/2}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} (\mathbf{R}_l - \mathbf{R}_{l+1}) \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} (\mathbf{R}_l - \mathbf{R}_{l+1}) \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} = \mathbf{K}_l - \mathbf{K}_{l+1}.$$

The matrix $\boldsymbol{\Sigma}_s^{(l)}$ (see i) and ii) in the previous theorem) is idempotent and symmetric and its trace is $t_l - t_{l+1} - r_l + r_{l+1}$.

We apply a lemma by Ferguson (cf. [4, p. 63]): "Suppose that $\mathbf{Z}_l$ is $\mathcal{N}(\boldsymbol{\mu}_s^{(l)}, \boldsymbol{\Sigma}_s^{(l)})$. If $\boldsymbol{\Sigma}_s^{(l)}$ is idempotent and $\boldsymbol{\Sigma}_s^{(l)} \boldsymbol{\mu}_s^{(l)} = \boldsymbol{\mu}_s^{(l)}$, the distribution of $\mathbf{Z}_l^T \mathbf{Z}_l$ is noncentral chi-square with degrees of freedom equal to rank of the matrix $\boldsymbol{\Sigma}_s^{(l)}$ and noncentrality parameter $\mu = (\boldsymbol{\mu}_s^{(l)})^T \boldsymbol{\mu}_s^{(l)}$". Therefore, the result follows if we establish that $\boldsymbol{\Sigma}_s^{(l)} \boldsymbol{\mu}_s^{(l)} = \boldsymbol{\mu}_s^{(l)}$. Applying that $\mathbf{K}_l - \mathbf{K}_{l+1}$ is an idempotent matrix, we have

$$\boldsymbol{\Sigma}_s^{(l)} \boldsymbol{\mu}_s^{(l)} = (\mathbf{K}_l - \mathbf{K}_{l+1}) (\mathbf{K}_l - \mathbf{K}_{l+1}) \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{s} = (\mathbf{K}_l - \mathbf{K}_{l+1}) \mathbf{D}^{-1/2}_{\mathbf{p}(\boldsymbol{\theta}_0)} \mathbf{s}.$$

Now we are going to get the noncentrality parameter,

$$(\boldsymbol{\mu}_s^{(l)})^T \boldsymbol{\mu}_s^{(l)} = \mathbf{s}^T \left(\mathbf{X}_l \mathbf{H}_l(\boldsymbol{\theta}_0) \mathbf{X}_l^T - \mathbf{X}_{l+1} \mathbf{H}_{l+1}(\boldsymbol{\theta}_0) \mathbf{X}_{l+1}^T\right) \mathbf{s}.$$

Now the result follows. □

Remark 1. Theorem 3 can be used to obtain an approximation to the power function of (7), as follows. Write

$$\mathbf{p}\left(\widehat{\theta}_l^{(r),\phi_2}\right) = \mathbf{p}\left(\widehat{\theta}_{l+1}^{(r),\phi_2}\right) + \frac{1}{\sqrt{n}} \left(\sqrt{n}\left(\mathbf{p}\left(\widehat{\theta}_l^{(r),\phi_2}\right) - \mathbf{p}\left(\widehat{\theta}_{l+1}^{(r),\phi_2}\right)\right)\right)$$

and define $\mathbf{p}_n \equiv \mathbf{p}\left(\widehat{\theta}_{l+1}^{(r),\phi_2}\right) + \frac{1}{\sqrt{n}} \mathbf{s}$, where

$$\mathbf{s} = \sqrt{n}\left(\mathbf{p}\left(\widehat{\theta}_l^{(r),\phi_2}\right) - \mathbf{p}\left(\widehat{\theta}_{l+1}^{(r),\phi_2}\right)\right).$$

Then substitute $\mathbf{s}$ into the definition of μ and, finally, μ into the right side of (17).

Acknowledgements

This work was parcially supported by Grants DGES PB2003-892 and UCM2005-910707.

References

[1] ALI, S. M., AND SILVEY, S. D. A general class of coefficient of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B, 28* (1966), 131–142.

[2] CRESSIE, N., AND READ, T. R. C. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B, 46* (1984), 440–464.

[3] CSISZÁR, I. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizität on Markhoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences 8* (1963), 84–108.

[4] FERGUSON, T. S. *A Course in Large Sample Theory*, Texts in Statistical Science. Chapman & Hall, London, 1996.

[5] HABER, M., AND BROWN, M. B. Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. *Journal of the American Statistical Association 81* (1986), 477–482.

[6] KULLBACK, S. Kullback information. In *Encyclopedia of Statistical Sciences*, vol. 4, S. Kotz and N. L. Johnson, Eds. John Wiley & Sons, New York, 1985.

[7] MARTÍN, N. Phi-divergencias en modelos loglineales con restricciones lineales. Trabajo de Investigación. Departamento de Estadística. Facultad de Matemáticas. Universidad Complutense de Madrid, Madrid, 2005.

[8] MARTÍN, N., AND PARDO, L. Minimum phi-divergence estimators for loglinear models with linear constraints and multinomial sampling. *Statistical Papers*, to appear (2006).

[9] PARDO, L., AND MENÉNDEZ, M. L. Analysis of divergence in loglinear models when expected frequencies are subject to linear constraints. *Metrika 64* (2006), 63–76.

[10] PARDO, L. *Statistical Inference based on Divergence Measures*. Statistics: Textbooks and Monographs. Chapman & Hall, New York, 2006.

[11] SEARLE, S. R. *Linear Models*. John Wiley & Sons, New York, 1971.

Nirian Martín
Departamento de Estadística e I. O. III
Universidad Complutense de Madrid
Puerta de Hierro s/n
28040 Madrid, Spain
nirian@estad.ucm.es

Leandro Pardo
Departamento de Estadística e I. O. I
Universidad Complutense de Madrid
Plaza de Ciencias 3
28040 Madrid, Spain
lpardo@mat.ucm.es