Kybernetika

Nirian Martin; Leandro Pardo

Choosing the best ϕ-divergence goodness-of-fit statistic in multinomial sampling with linear constraints

Kybernetika, Vol. 42 (2006), No. 6, 711--722

Persistent URL: http://dml.cz/dmlcz/135746

KYBERNETIKA — VOLUME 42 (2006), NUMBER 6, PAGES 711 – 722

CHOOSING THE BEST ϕ–DIVERGENCE GOODNESS–OF–FIT STATISTIC IN MULTINOMIAL SAMPLING FOR LOGLINEAR MODELS WITH LINEAR CONSTRAINTS

Nirian Martin and Leandro Pardo

In this paper we present a simulation study to analyze the behavior of the ϕ-divergence test statistics in the problem of goodness-of-fit for loglinear models with linear constraints and multinomial sampling. We pay special attention to the Rényi's and I_r-divergence measures.

Keywords: loglinear models, multinomial sampling, restricted maximum likelihood estimator, goodness-of-fit, I_r-divergence measure, Rényi's divergence measure

AMS Subject Classification: 62H15, 62H17

1. INTRODUCTION

Consider a sample of size $n \in \mathbb{N}$, $Y_1, Y_2, \ldots, Y_n$ with realizations from $\mathcal{Y} = \{1, 2, \ldots, k\}$ and independent and identically distributed according to a probability distribution $\boldsymbol{p}(\boldsymbol{\theta}_0)$. If $k = IJ$ we have a two-way contingency table. This distribution is assumed to be unknown, but belonging to a known family

$$\mathcal{P} = \left\{ \boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots, p_k(\boldsymbol{\theta}))^{\mathrm{T}} : \boldsymbol{\theta} \in \Theta \right\}$$

of distributions on $\mathcal{Y}$ with $\Theta \subset \mathbb{R}^{t+1}$.
The true value $\boldsymbol{\theta}_0$ of parameter $\boldsymbol{\theta} = (u, \theta_1, \ldots, \theta_t)^{\mathrm{T}} \in \Theta$ is assumed to be unknown. Let $\widehat{\boldsymbol{p}} = (\widehat{p}_1, \ldots, \widehat{p}_k)^{\mathrm{T}}$ for

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

The statistic $(N_1, \ldots, N_k)$ is obviously sufficient for the statistical model under consideration and is multinomially distributed with parameters $(n; \boldsymbol{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots$ $\ldots, p_k(\boldsymbol{\theta}))$. We denote

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

and $\boldsymbol{m}(\boldsymbol{\theta}) = (m_1(\boldsymbol{\theta}), \ldots, m_k(\boldsymbol{\theta}))^{\mathrm{T}}$.

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

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

represents the class of the loglinear models associated with $\boldsymbol{X}$. We suppose, in the following that $\boldsymbol{J} = (1, \overset{k}{\ldots}, 1)^{\mathrm{T}} \in \mathcal{C}(\boldsymbol{X})$. Taking into account (2), the parameter space is defined by

$$\Theta' = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{t+1} : \log \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{\theta} \text{ and } \boldsymbol{J}^{\mathrm{T}} \boldsymbol{m}(\boldsymbol{\theta}) = n \right\}.$$

Now in addition to the previous model we shall assume that we have $s-1 < t$ linear constrains defined by

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

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

$$\Theta^* = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{t+1} : \log \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{\theta} \text{ and } \boldsymbol{L}^{\mathrm{T}} \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{d} \right\} \tag{5}$$

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

We have seen in (3) that a loglinear model relates the logarithms of the expected frequencies of cells to a linear model. This model can be seen as a set of linear constraints imposed on the logarithms of the expected cell frequencies. However there are hypotheses that impose linear constraints on the expected cell frequencies and not on their logarithms. This situation was formulated in (5). Some practical situations require loglinear models when expected frequencies are subject to linear constraints. In Haber and Brown [6] can be seen some interesting examples of this model as well as a historical perspective about the development of this model.

The classical goodness-of-fit test statistics for testing if our data are from a considered loglinear model in which the expected frequencies are subject to linear constraints are

$$X^2 = \sum_{j=1}^{k} \frac{(N_j - m_j(\widehat{\boldsymbol{\theta}}))^2}{m_j(\widehat{\boldsymbol{\theta}})} \quad \text{or} \quad G^2 = 2 \sum_{j=1}^{k} N_j \log \frac{N_j}{m_j(\widehat{\boldsymbol{\theta}})},$$

where $\widehat{\boldsymbol{\theta}}$ is the restricted maximum likelihood estimator of $\boldsymbol{\theta} \in \Theta^*$ defined by

$$\widehat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta^*} \boldsymbol{h}^{\mathrm{T}} \boldsymbol{\theta}, \tag{6}$$

where $\boldsymbol{h}^{\mathrm{T}} = (\boldsymbol{n}^*)^{\mathrm{T}} \boldsymbol{X}$, $\boldsymbol{n}^* = (N_1, \ldots, N_k)^{\mathrm{T}}$.

It is important to note that $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of the loglinear model $\log \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{\theta}$ with multinomial sampling, under the assumption that relation (4) is satisfied, i. e., $\widehat{\boldsymbol{\theta}}$ is the restricted multinomial maximum likelihood estimator. We can see that $\boldsymbol{\theta}$ varies in Θ^*. If we were interested in the multinomial maximum likelihood estimator of the parameter $\boldsymbol{\theta}$ associated with the loglinear model $\log \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{X\theta}$ the definition given in (6) would be valid but instead of considering that $\boldsymbol{\theta}$ varies in Θ^* one has to assume that $\boldsymbol{\theta}$ varies in Θ'.

Equivalently, the restricted maximum likelihood estimator can be defined as,

$$\widehat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta^*} D_{\text{Kullback}} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}) \right), \tag{7}$$

where $D_{\text{Kullback}} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}) \right)$ is the Kullback–Leibler divergence between the probability vectors $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\boldsymbol{\theta})$(see Kullback [7])

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

We can observe that $G^2 = 2nD_{\text{Kullback}} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right)$. The asymptotic distribution of X^2 and G^2 is a chi-square with $k-t+s-2$ degrees of freedom according to Haber and Brown [6]. It is interesting to observe that X^2 involve two divergence measures, one of them the Kullback–Leibler divergence for estimation and the other one, the Pearson's divergence

$$D_{\text{Pearson}} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right) = \frac{1}{2} \sum_{j=1}^{k} \frac{\left(\widehat{p}_j - p_j(\widehat{\boldsymbol{\theta}}) \right)^2}{p_j(\widehat{\boldsymbol{\theta}})},$$

for testing $X^2 = 2nD_{\text{Pearson}} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right)$. In the case of G^2 we are using Kullback–Leibler divergence for testing and estimation. Kullback–Leibler divergence as well as Pearson's divergence are particular cases of the ϕ-divergence measure defined simultaneously by Csiszár [4] and Ali and Silvey [3]. This family of divergence measures is defined in our model by

$$D_{\phi} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right) = \sum_{j=1}^{k} p_j(\widehat{\boldsymbol{\theta}}) \phi \left(\frac{\widehat{p}_j}{p_j(\widehat{\boldsymbol{\theta}})} \right), \ \phi \in \Phi^*$$

where Φ^* is the class of all convex functions $\phi(x)$, $x \geq 0$, such that $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$ and $0\phi(\kappa/0) = \kappa \lim_{u\to\infty} \phi(u)/u$ for $\kappa \geq 0$.). For more details about ϕ-divergences see Vajda [12]. In Pardo and Menéndez [9] was established, assuming that $\log \boldsymbol{m}(\boldsymbol{\theta}) = \log n\boldsymbol{p}(\boldsymbol{\theta}) \in \mathcal{C}(\boldsymbol{X})$, i. e., $\boldsymbol{\theta} \in \Theta^*$, and $\widehat{\boldsymbol{\theta}}$ satisfies (7), that the family of test statistics

$$T_n^{\phi} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right) = \frac{2n}{\phi''(1)} D_{\phi} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right)$$

converges in law to a chi-square distribution with $k-t+s-2$ degrees of freedom.

Another extension of the Kullback–Leibler divergence was defined initially by Rényi [11] and extended later by Liese and Vajda [8]: Rényi's divergence measure. We shall use the expression given by Liese and Vajda to measure the distance between the nonparametric estimator $\widehat{\boldsymbol{p}}$ and the parametric estimator $\boldsymbol{p}(\widehat{\boldsymbol{\theta}})$,

$$D_{\text{Rényi}}^{r} \left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}) \right) = \frac{1}{r(r-1)} \log \sum_{j=1}^{k} \widehat{p}_j^{\,r} \, p_j(\widehat{\boldsymbol{\theta}})^{1-r}, \ r \neq 0, 1.$$

It is immediate that

$$\lim_{r\to 1} D^r_{\text{Rényi}}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right) = D_{\text{Kullback}}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right)$$

and

$$\lim_{r\to 0} D^r_{\text{Rényi}}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right) = D_{\text{Kullback}}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{p}}\right).$$

Rényi's divergence measure was not previously used in loglinear models. Section 2 is devoted to present some theoretical results for this divergence measure in the context considered previously and in Section 3 a simulation study is carried out to establish that it is possible to get some test statistics based on divergence measures that are good alternatives to the classical likelihood ratio and Pearson test statistic for goodness-of-fit based on multinomial sampling in loglinear models with linear constraints. We consider in our study the family of Rényi's test statistics as well as the I_r -divergence test statistics. The I_r-divergence test statistics are based on the I_r-divergence measures introduced and studied by Liese and Vajda [8]. This is the first known a simulation study carried out in loglinear models with linear constraints using ϕ-divergences because in the cited paper of Pardo and Menéndez [9] only theoretical results were obtained.

2. RÉNYI'S TEST STATISTIC FOR LOGLINEAR MODELS

If we consider the functions

$$h_r(x) = \begin{cases} \frac{1}{r(r-1)} \log\left(r\left(r-1\right)x+1\right), & r\neq 0,1 \\ x, & r=0,1 \end{cases} \tag{8}$$

and

$$\phi_r\left(x\right) = \begin{cases} \frac{1}{r(r-1)}\left(x^r - r\left(x-1\right)-1\right), & r\neq 0,1 \\ x\log x - x + 1, & r=1 \\ -\log x + x - 1, & r=0, \end{cases} \tag{9}$$

we find that Rényi's divergence can be given as follows

$$D^r_{\text{Rényi}}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = h_r\left(D_{\phi_r}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))\right).$$

We can observe that ϕ_r is a convex function with $\phi_r\left(1\right) = \phi'_r\left(1\right) = 0$ and $\phi''_r\left(1\right) = 1$, i.e., $D_{\phi_r}\left(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right)$ is a ϕ-divergence between the probability vectors $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\widehat{\boldsymbol{\theta}})$. More precisely, it is the I_r-divergence

$$I_r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \frac{1}{r\left(r-1\right)}\left(\sum_{j=1}^{k}\widehat{p}_j^{\,r}\; p_j(\widehat{\boldsymbol{\theta}})^{1-r} - 1\right), \quad r(1-r)\neq 0.$$

For a complete study of its properties, see Liese and Vajda [8]. For testing

$$H_0 : \boldsymbol{p} = \boldsymbol{p}(\boldsymbol{\theta}),\ \ \boldsymbol{\theta}\in\Theta^*$$

we consider in this paper the I_r-divergence test statistics

$$I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \begin{cases} \frac{2n}{r(r-1)}(\sum_{j=1}^{k} \widehat{p}_j^r \; p_j(\widehat{\boldsymbol{\theta}})^{1-r} - 1), & r \neq 0,1 \\ 2nD_{\text{Kullback}}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})), & r = 1 \\ 2nD_{\text{Kullback}}(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{p}}), & r = 0, \end{cases}$$

as well as the Rényi's family of test statistics given by,

$$T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \begin{cases} \frac{2n}{r(r-1)} \log \sum_{j=1}^{k} \widehat{p}_j^r \; p_j(\widehat{\boldsymbol{\theta}})^{1-r}, & r \neq 0,1 \\ 2nD_{\text{Kullback}}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})), & r = 1 \\ 2nD_{\text{Kullback}}(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{p}}), & r = 0. \end{cases}$$

We can observe that $I^2(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ coincides with the classical Pearson test statistic X^2 and $I^1(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$and T^1 with the likelihood ratio test. In the next theorem we present the asymptotic distribution of the family $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$.

Theorem 1. We consider the class of loglinear models associated with $\boldsymbol{X}$, $C(\boldsymbol{X})$, and we shall assume that we have the $s-1 < t$ linear constraints given in (4). The asymptotic distribution of the family of test statistics $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$, under the hypothesis of $\boldsymbol{\theta} \in \Theta^*$, is a chi-square with $k-t+s-2$ degrees of freedom.

Proof. By the first order Taylor expansion of $h_r(x)$ around $x=0$ we obtain

$$\begin{aligned} T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = & \frac{2n}{\phi_r''(1)h_r'(0)} h_r\left(D_{\phi_r}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))\right) \\ = & \frac{2n}{\phi_r''(1)} D_{\phi_r}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) + 2no\left(D_{\phi}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))\right), \end{aligned}$$

where h and ϕ are defined in (8) and (9), respectively. By Pardo and Menéndez [9] we know that $\frac{2n}{\phi''(1)} D_{\phi}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ converges in law to a chi-square with $k-t+s-2$ degrees of freedom. Therefore $2no\left(D_{\phi_r}(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))\right) = o_P(1)$ and $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ converges in law to a chi-square distribution with $k-t+s-2$ degrees of freedom. □

For testing $H_0 : \boldsymbol{p} = \boldsymbol{p}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta^*$ we can use the families of test statistics $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ or $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$; if it is too large, H_0 is rejected. When $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) > c$ or $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) > c$, we reject H_0, where c is specified so that the size of the test is α :

$$\Pr\left(T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}})) > c \;\middle|\; H_0\right) = \alpha;\;\; \alpha \in (0,1)\,. \tag{10}$$

The same for $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$. If we are able to get the value of c from the equation (10) then we obtain exact tests based on T_n^r and $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ which are obviously equivalent. In general it is not possible to get the exact test and we have the necessity to consider the asymptotic tests. In this case $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ are not equivalent, cf. Remark 1. Based on the previous theorem

$$c = \chi^2_{k-t+s-2,\alpha}, \tag{11}$$

where $\Pr\left(\chi^2_{k-t+s-2} > \chi^2_{k-t+s-2,\alpha}\right) = \alpha$. The choice of (11) in (10) guarantees only an asymptotic size-α test. The same asymptotic critical point is obtained for $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ on the basis of the results in Pardo and Menéndez [9]. In the simulation study of Section 3 we study for what choices of r in $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ is the relation (10) most accurately attained.

Remark 1. We are going to analyze the relation existing between the powers of $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ as well as between the size using the asymptotic critical point given in (11). To avoid the problems with empty cells we are going to assume that $r > 0$. We shall denote by $\alpha_r^{\text{Rényi}}$, $\beta_r^{\text{Rényi}}$, α_r and β_r, size and power for $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and size and power for $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$, respectively. It is obvious that

$$T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})) \begin{cases} \underset{\text{st}}{<} I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})), & \text{if} \quad r > 1 \\ \underset{\text{st}}{>} I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})), & \text{if} \quad 0 < r < 1 \\ \underset{\text{a.s.}}{=} I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})), & \text{if} \quad r = 1 \end{cases}$$

because $h_r(x) < x$ if $r > 1$, $h_r(x) > x$ if $0 < r < 1$ and $h_r(x) = x$ if $r = 1$. We denote by $X_1 <_{st} X_2$ that $\Pr(X_1 \geq x) < \Pr(X_2 \geq x)$ for every $x \in \mathbb{R}^+$. Taking into account that our procedure of testing uses the asymptotic critical value $c = \chi^2_{k-t+s-2,\alpha}$ we have

$$\alpha_r^{\text{Rényi}} \begin{cases} < \alpha_r, & \text{if} \quad r > 1 \\ > \alpha_r, & \text{if} \quad 0 < r < 1 \end{cases} \quad \text{and} \quad \beta_r^{\text{Rényi}} \begin{cases} < \beta_r, & \text{if} \quad r > 1 \\ > \beta_r, & \text{if} \quad 0 < r < 1. \end{cases}$$

Remark 2. In the same way as we have used the family $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ for testing when the data are from $\boldsymbol{p}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta^*$, we can also use the family $S_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = T_n^r(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{p}})$, i.e., we can change the position of the arguments in the divergence measure. We are going to establish the asymptotic distribution of this family of test statistics. We consider the function $\varphi_r(x) = \frac{1}{r(r-1)}\left(x^{-r+1} - r(1-x) - x\right)$, which is convex for $x > 0$ and satisfying $\varphi_r(1) = \varphi_r'(1) = 0$ and $\varphi_r''(1) = 1$, i.e., $\varphi_r \in \Phi^*$. It is also easy see that

$$D_{\varphi_r}\left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right) = D_{\phi_r}\left(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{p}}\right).$$

Now by applying the result of Pardo and Menéndez [9] we obtain that

$$\widetilde{I}_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})) = \frac{2n}{\varphi_r''(1)} D_{\varphi_r}\left(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}})\right) = I_n^r(\boldsymbol{p}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{p}})$$

converges in law to the chi-square distribution with $k-t+s-2$ degrees of freedom and using a similar argument as in the previous theorem we get that $S_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ converges in law to the chi-square distribution with $k-t+s-2$ degrees of freedom.

The asymptotic chi-squared approximation, $c = \chi^2_{k-t+s-2,\alpha}$, is checked for a loglinear model in the simulation study given in Section 3. We give a small illustration of those results now. Figures 1 and 2 show departures of the exact size from the nominal size of $\alpha = 0.05$ for the loglinear model with constaints considered in (12)–(13) for the null hypothesis and for various choices of r and for small to large sample sizes. In Figure 1 we used the test statistics $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and in Figure 2 the test statistic $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$.

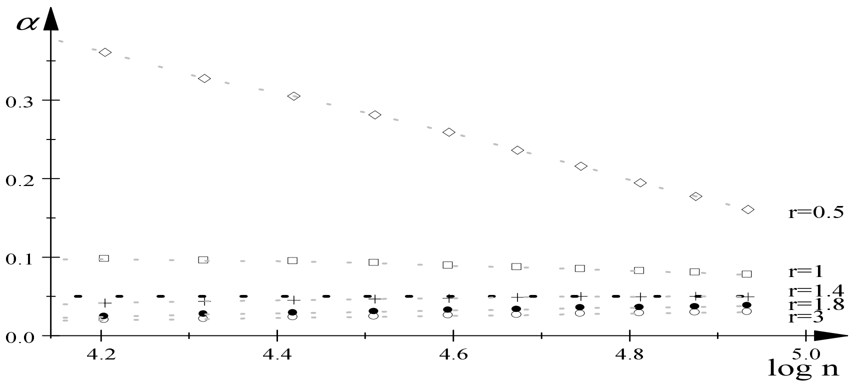

Fig. 1. Exact size as a function of $x = \log n$ for $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$.

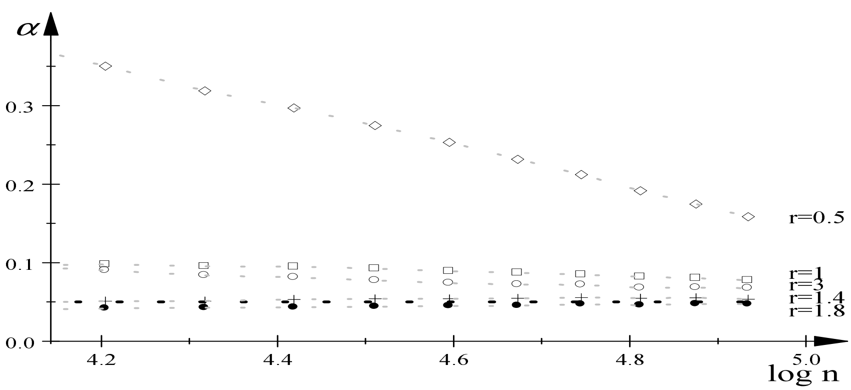

Fig. 2. Exact size as a function of $x = \log n$ for $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$.

Previous pictures show behavior of the exact size for nominal size of $\alpha = 0.05$ for different values of r in $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ including the behavior for the likelihood ratio test ($r = 1$).

3. SIMULATION STUDY

In this section we present a simulation study to see the behavior of the Rényi's test statistics as well as the I_r -divergence test statistics in the model of quasi-independence with marginal homogeneity. This model in a 4×4 contingency table is given by

$$\log m_{ij}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \delta_i I\left(i = j\right), \; i, j = 1, 2, 3, 4, \tag{12}$$

where $\sum_{i=1}^{4} \theta_{1(i)} = \sum_{j=1}^{4} \theta_{2(j)} = 0$, and the linear constraints

$$\boldsymbol{L}^{\mathrm{T}} \boldsymbol{m}(\boldsymbol{\theta}) = \boldsymbol{d}, \tag{13}$$

where

$$\boldsymbol{L}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & -1 & 0 \end{pmatrix},$$

and $\boldsymbol{d} = (n, 0, 0, 0)^{\mathrm{T}}$.

There are many practical situations in two-way contingency tables with I and J levels for the two nominal response variables X and Y in which there is a correspondence between row and column variables but diagonal cells tend to be large. These large diagonal cells often contribute significantly to the poor fit of the independence model. One substantively interesting hypothesis is whether the rest of the table satisfies the independence hypothesis net of the diagonal cell. This leads to the quasi-independence model. For more details about the quasi-independence model see Agresti [1], Powers and Xie [10], Andersen [2] and references therein.

However the study of some real situations requires to include linear constraints on the expected cell frequencies associated with the loglinear model of quasi-independence. A nice real example of this situation can be seen in Section 4.1 of Haber and Brown [6]. They considered a loglinear model of quasi-independence with marginal homogeneity to model the frequency of ewes according to the number of lambs born in two consecutive years.

The theoretical model considered by us is defined by the parameters

$$\begin{aligned} &\exp(\theta_{1(1)}) = \exp(\theta_{2(1)}) = 0.8835, \quad \exp(\theta_{1(2)}) = \exp(\theta_{2(2)}) = 0.9639, \\ &\exp(\theta_{1(3)}) = \exp(\theta_{2(3)}) = 1.0448, \quad \exp(\delta_1) = 5.5455, \quad \exp(\delta_2) = 5.1557, \\ &\exp(\delta_3) = \exp(\delta_4) = 4.5714, \end{aligned} \tag{14}$$

and we understand that $(\theta_{1(1)}, \theta_{1(2)}, \theta_{1(3)}, \theta_{2(1)}, \theta_{2(2)}, \theta_{2(3)}, \delta_1, \delta_2, \delta_3, \delta_4)^{\mathrm{T}}$ is $(\theta_1, \ldots, \theta_t)^{\mathrm{T}}$ according to the notation used in Section 1. These values give the following probability vector

$p_{ij}(\boldsymbol{\theta})$	1	2	3	4	$p_{i*}(\boldsymbol{\theta})$
1	0.1355	0.0267	0.0289	0.0311	0.2222
2	0.0267	0.1502	0.0315	0.0339	0.2422
3	0.0289	0.0315	0.1561	0.0367	0.2531
4	0.0311	0.0339	0.0367	0.1807	0.2822
$p_{*j}(\boldsymbol{\theta})$	0.2222	0.2422	0.2531	0.2822	1.0000

In this situation we have $k = 16$, $t = 10$ and $s = 4$ and therefore the asymptotic critical point for $\alpha = 0.05$ is $c = 15.507$. The simulated exact sizes, at a nominal size α for a sample size n, $\widehat{\alpha}_{n,r}^{\text{Rényi}}$ and $\widehat{\alpha}_n^r$ are given by

$$\widehat{\alpha}_{r,n}^{\text{Rényi}} = \frac{\text{Number of } T_{n,j}^r > 15.507}{N} \quad \text{and} \quad \widehat{\alpha}_n^r = \frac{\text{Number of } I_{n,j}^r > 15.507}{N},$$

respectively. By $T_{n,j}^r$ and $I_{n,j}^r$ we are denoting the value of $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$, in the jth simulation $(j = 1, \dots, N)$ when the sample size is n respectively. We shall assume in our study $N = 100\,000$ and we consider $n = 65$ and 100. We are going to consider $r = 0.5$, 1, 1.4, 1.8, 2.2, 2.6, 3, 3.4 and 3.8.

In order to study the powers of the test statistics based on $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ we are going to define some alternative hypotheses. We consider the alternative hypotheses a_ϵ by defining the probability distribution

$$p_{ij}^{\epsilon}(\boldsymbol{\theta}) = \begin{cases} (1-\epsilon)\, p_{ij}(\boldsymbol{\theta}), & (i,j) \neq (4,3) \\ (1-\epsilon)\, p_{ij}(\boldsymbol{\theta}) + \epsilon, & (i,j) = (4,3)\,. \end{cases} \tag{15}$$

We shall assume $\epsilon = 0.03$, 0.07 , 0.11, 0.15 and 0.19. The way to obtain the simulated powers $\widehat{\beta}_{r,n}^{\text{Rényi}}$ and $\widehat{\beta}_n^r$ is the same as the way used for getting the simulated exact size but now the simulations are obtained from the probability distribution given in (15).

In Tables 1 and 2 we present the simulated exact size (column labeled with “size”) and the power for the considered alternatives a_ϵ for $n = 65$ and 100, respectively. The row LRT corresponds to the likelihood ratio test $T^1(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I^1(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$.

The trade-off between size behavior and power behavior is a classical problem in hypothesis testing as one of the referees pointed out. Therefore we have evaluated the size-corrected relative local efficiencies

$$\rho_r^{\text{Rényi}}(a_\epsilon) = \frac{(\widehat{\beta}_{n,r}^{\text{Rényi}}(a_\epsilon) - \widehat{\alpha}_{n,r}^{\text{Rényi}}) - (\widehat{\beta}_{n,1}^{\text{Rényi}}(a_\epsilon) - \widehat{\alpha}_{n,1}^{\text{Rényi}})}{\widehat{\beta}_{n,1}^{\text{Rényi}}(a_\epsilon) - \widehat{\alpha}_{n,1}^{\text{Rényi}}}$$

of $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ with respect to the classical likelihood ratio test $T_n^1(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$. In a similar way we define the local efficiencies, $\rho_r(a_\epsilon)$, of $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ with respect to the classical likelihood ratio test $T_n^1(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$. We have only included in the study the test statistics with a simulated exact size less than or equal to 0.1, i. e., the test statistics with simulated exact size less than or equal to the double of the nominal size $\alpha = 0.05$. In Tables 3 and 4 we present the relative efficiencies of $T_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ with respect the likelihood ratio test statistic.

Table 1. Exact size and powers of $T_n^r\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ for $n=65$.

		size	a_ϵ				
	r		$\epsilon=0.03$	$\epsilon=0.07$	$\epsilon=0.11$	$\epsilon=0.15$	$\epsilon=0.19$
	0.5	0.371 32	0.412 14	0.546 56	0.693 18	0.816 82	0.904 38
	1.4	0.039 71	0.066 21	0.180 21	0.366 19	0.570 14	0.746 37
	1.8	0.023 16	0.042 61	0.135 52	0.304 83	0.505 44	0.691 22
T_n^r	2.2	0.018 38	0.034 37	0.115 96	0.268 85	0.459 11	0.645 32
	2.6	0.017 92	0.032 89	0.109 08	0.251 64	0.429 44	0.607 21
	3	0.019 70	0.034 17	0.107 79	0.243 67	0.410 20	0.577 26
	3.4	0.021 77	0.036 29	0.108 99	0.238 62	0.396 21	0.554 42
	3.8	0.024 08	0.038 73	0.109 63	0.234 69	0.384 62	0.535 35
LRT	1	0.097 67	0.134 03	0.276 97	0.470 29	0.660 17	0.810 95
	0.5	0.360 30	0.400 66	0.534 72	0.683 24	0.809 38	0.898 87
	1.4	0.049 05	0.078 94	0.203 26	0.396 88	0.601 09	0.771 22
	1.8	0.040 32	0.067 19	0.188 13	0.379 58	0.586 00	0.759 79
I_n^r	2.2	0.046 71	0.075 52	0.202 73	0.398 34	0.603 29	0.770 09
	2.6	0.063 25	0.097 42	0.236 55	0.438 58	0.637 68	0.793 01
	3	0.091 65	0.131 12	0.283 79	0.490 70	0.680 32	0.822 29
	3.4	0.128 80	0.174 48	0.340 60	0.547 63	0.725 43	0.853 22
	3.8	0.175 33	0.226 07	0.401 08	0.605 74	0.768 10	0.880 72

Table 2. Exact size and powers of $T_n^r\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and $I_n^r\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ for $n=100$.

		size	a_ϵ				
	r		$\epsilon=0.03$	$\epsilon=0.07$	$\epsilon=0.11$	$\epsilon=0.15$	$\epsilon=0.19$
	0.5	0.254 42	0.312 71	0.494 94	0.711 29	0.873 71	0.957 57
	1.4	0.047 32	0.093 52	0.288 88	0.567 75	0.797 67	0.928 01
	1.8	0.031 34	0.069 28	0.252 56	0.528 94	0.771 02	0.914 00
T_n^r	2.2	0.025 60	0.059 87	0.230 91	0.501 06	0.747 41	0.900 05
	2.6	0.024 57	0.056 61	0.220 47	0.480 60	0.726 87	0.886 15
	3	0.025 21	0.056 27	0.215 65	0.466 27	0.709 41	0.872 29
	3.4	0.026 99	0.057 24	0.213 51	0.455 64	0.693 27	0.857 08
	3.8	0.028 86	0.059 61	0.212 45	0.446 76	0.678 65	0.842 46
LRT	1	0.087 24	0.144 64	0.352 50	0.621 01	0.829 07	0.941 55
	0.5	0.248 69	0.307 18	0.488 06	0.705 60	0.870 15	0.955 83
	1.4	0.053 13	0.103 16	0.306 56	0.587 57	0.811 53	0.933 85
	1.8	0.044 17	0.091 57	0.295 16	0.579 30	0.807 11	0.932 32
I_n^r	2.2	0.046 26	0.096 23	0.306 46	0.590 84	0.814 09	0.934 80
	2.6	0.055 92	0.111 50	0.333 43	0.615 79	0.828 88	0.941 05
	3	0.072 62	0.134 96	0.370 67	0.649 16	0.848 06	0.949 42
	3.4	0.095 77	0.165 77	0.415 18	0.686 98	0.869 48	0.957 37
	3.8	0.126 03	0.203 40	0.462 50	0.724 46	0.889 97	0.965 07

If we observe Table 1 the simulated sizes corresponding to $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ are less than or equal to 0.1 for all r except for $r=0.5$ and for $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ all the values of the interval $[1,3]$ satisfies the condition. For $n=100$ the values of r that satisfies the condition are the same as for $n=65$ in the case of $T_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ and for $I_n^r(\widehat{\boldsymbol{p}},\boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$

we have the new value $r = 3.4$. It is also interesting to observe in Tables 1 and 2 that the values obtained in the simulation study are in accordance with the theoretical results presented in Remark 1.

Table 3 indicates that the size-corrected relative local efficiency for $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ with $r = 3$ is the best and of course better than the likelihood ratio test and chi-square test statistic obtained for $r = 2$ in $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$. Table 4 indicates that the size-corrected relative efficiency for $I_n^r(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ with $r = 3$ and 3.4 are the best. Therefore we can conclude that independently considered of the sample size the test statistic $I_n^3(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\widehat{\boldsymbol{\theta}}))$ is a good alternative to the classical likelihood ratio test and chi-square test statistic for the problem goodness-of-fit in multinomial sampling for loglinear models with linear constraints.

Table 3. Size-corrected relative local efficiencies $\rho_r^{\text{Rényi}}(a_\epsilon)$ and $\rho_r(a_\epsilon)$ for $n = 65$.

		a_ϵ					
	r	$\epsilon = 0.03$	$\epsilon = 0.07$	$\epsilon = 0.11$	$\epsilon = 0.15$	$\epsilon = 0.19$	Total
T_n^r	1.4	–0.271 18	–0.216 38	–0.123 82	–0.057 00	–0.009 28	–0.677 66
	1.8	–0.465 07	–0.373 34	–0.244 07	–0.142 60	–0.063 40	–1.288 48
	2.2	–0.560 23	–0.455 78	–0.327 81	–0.216 46	–0.121 04	–1.681 32
	2.6	–0.588 28	–0.491 59	–0.372 77	–0.268 40	–0.173 82	–1.894 86
	3	–0.602 04	–0.508 71	–0.398 91	–0.305 78	–0.218 32	–2.033 76
	3.4	–0.600 66	–0.513 56	–0.418 02	–0.334 32	–0.253 24	–2.119 80
	3.8	–0.597 08	–0.522 88	–0.434 77	–0.359 03	–0.283 22	–2.196 98
LRT	1	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00
I_n^r	1.4	–0.177 94	–0.139 96	–0.066 52	–0.018 59	0.012 46	–0.390 55
	1.8	–0.261 00	–0.175 66	–0.089 52	–0.029 89	0.008 68	–0.547 39
	2.2	–0.207 65	–0.129 86	–0.056 32	–0.010 52	0.014 16	–0.390 19
	2.6	–0.060 23	–0.033 48	0.007 29	0.021 22	0.023 11	–0.042 09
	3	0.085 53	0.071 62	0.070 94	0.046 53	0.024 34	0.298 96

Table 4. Size-corrected relative local efficiencies $\rho_r^{\text{Rényi}}(a_\epsilon)$ and $\rho_r(a_\epsilon)$ for $n = 100$.

		a_ϵ					
	r	$\epsilon = 0.03$	$\epsilon = 0.07$	$\epsilon = 0.11$	$\epsilon = 0.15$	$\epsilon = 0.19$	Total
T_n^r	1.4	–0.195 12	–0.089 35	–0.024 99	0.011 48	0.030 88	–0.267 10
	1.8	–0.339 02	–0.166 03	–0.067 76	–0.002 90	0.033 18	–0.542 53
	2.2	–0.402 96	–0.226 00	–0.109 24	–0.026 99	0.023 57	–0.741 62
	2.6	–0.441 81	–0.261 48	–0.145 64	–0.053 29	0.008 51	–0.893 71
	3	–0.458 89	–0.282 06	–0.173 69	–0.077 69	–0.008 46	–1.000 79
	3.4	–0.473 00	–0.296 84	–0.196 94	–0.101 84	–0.028 35	–1.096 97
	3.8	–0.464 29	–0.307 89	–0.217 08	–0.124 07	–0.047 65	–1.160 98
LRT	1	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00	0.000 00
I_n^r	1.4	–0.128 40	–0.044 60	0.001 26	0.022 34	0.030 91	–0.118 49
	1.8	–0.174 22	–0.053 80	0.002 55	0.028 46	0.039 61	–0.157 40
	2.2	–0.129 44	–0.019 08	0.020 25	0.035 05	0.040 07	–0.053 15
	2.6	–0.031 71	0.046 18	0.048 90	0.041 96	0.036 08	0.141 41
	3	0.086 06	0.123 61	0.080 13	0.045 31	0.026 33	0.361 44
	3.4	0.219 51	0.204 14	0.107 61	0.042 98	0.008 53	0.582 77

ACKNOWLEDGEMENT

This work was supported by Grants MTM 2006-06872 and UCM 2005-910707.

(Received September 1, 2005.)

REFERENCES

[1] A. Agresti: Categorical Data Analysis. Wiley, New York 2002.
[2] E. B. Andersen: The Statistical Analysis of Categorical Data. Springer, New York 1990.
[3] S. M. Ali and S. D. Silvey: A general class of coefficient of divergence of one distribution from another. J. Roy. Statist. Soc. *28* (1966), 131–142.
[4] I. Csiszár: Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizität on Markhoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. *8* (1963), 84–108.
[5] J. R. Dale: Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. J. Roy. Statist. Soc. Ser. B *41* (1986), 48–59.
[6] M. Haber and M. B. Brown: Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. J. Amer. Statist. Assoc. *81* (1986), 477–482.
[7] S. Kullback: Kullback information. In: Encyclopedia of Statistical Sciences (S. Kotz and N. L. Johnson, eds.), Wiley, New York 1985, Volume 4, pp. 421–425.
[8] F. Liese and I. Vajda: Convex Statistical Distances. Teubner, Leipzig 1987.
[9] L. Pardo and M. L. Menéndez: Analysis of divergence in loglinear models when expected frequencies are subject to linear constraints. Metrika *64* (2006), 63–76.
[10] D. A. Powers and Y. Xie: Statistical Methods for Categorical Data Analysis. Academic Press, San Diego 2000.
[11] A. Rényi: On measures of entropy and information. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability *1* (1961), pp. 547–561.
[12] I. Vajda: Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht 1989.

Nirian Martin, Escuela Universitaria de Estadistica, Universidad Complutense de Madrid, Avenida Puerta de Hierro s/n – 28040 Madrid. Spain.
e-mail: nirian@estad.ucm.es

Leandro Pardo, Facultad de Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias 3 – 28040 Madrid. Spain.
e-mail: lpardo@mat.ucm.es