Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1407–1423
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Residual analysis and outliers in loglinear models based on phi-divergence statistics

A.K. Gupta, T. Nguyen, L. Pardo*,[1]

Deparment of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA

Received 30 September 2004; received in revised form 4 March 2006; accepted 30 March 2006
Available online 9 June 2006

Abstract

In this paper we consider new families of residuals and influential measures, under the assumption of multinomial sampling, for loglinear models. These new families are based on ϕ-divergence test statistic. The asymptotic normality of the standardized residuals is obtained as well as the relation of the new family of influential measures with the appropriate Cook's distance in this context. The expression of the new family of residuals is obtained in two important problems: independence and symmetry in two-dimensional contingence tables. A numerical example illustrates the results obtained.

MSC: Primary, 62H15;; secondary 62H17

Keywords: Loglinear models; Multinomial sampling; Minimum ϕ-divergence estimator; ϕ-divergence test statistics;Residuals; Outliers; Cook's distance

1. Introduction

Consider a sample $Y_1, Y_2, \ldots, Y_n$ of size $n \in \mathbb{N}$ with realizations from $\mathscr{Y} = \{1, 2, \ldots, M\}$ and independently and identically distributed (i.i.d.) according to the probability distribution $\boldsymbol{p}(\boldsymbol{\theta}_0) = (p_1(\boldsymbol{\theta}_0), \ldots, p_M(\boldsymbol{\theta}_0))^{\mathrm{T}}$. This distribution is assumed to be unknown, but belongs to a known family

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

of distributions on $\mathscr{Y}$ with $\Theta \subset \mathbb{R}^{M_0}$ $(M_0 < M - 1)$. Here and in the sequel, "T" denotes the vector or matrix transpose. In other words, the true value $\boldsymbol{\theta}_0$ of parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{M_0})^{\mathrm{T}} \in \Theta$ is assumed to be unknown. We denote $\widehat{\boldsymbol{p}} = (\widehat{p}_1, \ldots, \widehat{p}_M)^{\mathrm{T}}$ where

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

* Corresponding author.
E-mail address: Leandro_Pardo@mat.ucm.es (L. Pardo).
[1] Work initiated when he was visiting as the Distinguished Lukacs Professor on leave from the Complutense University of Madrid.

0378-3758/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2006.03.005

The statistic $(N_1, \dots, N_M)$ is obviously sufficient for the statistical model under consideration and is multinomially distributed; that is,

$$\Pr(N_1 = n_1, \dots, N_M = n_M) = \frac{n!}{n_1! \cdots n_M!} p_1(\boldsymbol{\theta}_0)^{n_1} \times \cdots \times p_M(\boldsymbol{\theta}_0)^{n_M}, \tag{2}$$

for integers $n_1, \dots, n_M \geqslant 0$ such that $n_1 + \cdots + n_M = n$.

In what follows, we assume that $\boldsymbol{p}(\boldsymbol{\theta})$ belongs to the general class of loglinear models. That is, we assume

$$p_i(\boldsymbol{\theta}) = \exp(\boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{\theta}) \Big/ \sum_{v=1}^{M} \exp(\boldsymbol{w}_v^{\mathrm{T}} \boldsymbol{\theta}), \quad i = 1, \dots, M, \tag{3}$$

where $\boldsymbol{w}_i = (w_{i1}, \dots, w_{iM_0})^{\mathrm{T}}$, $i = 1, \dots, M$. We denote by $\boldsymbol{W} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_M)^{\mathrm{T}}$ the $M \times M_0$ matrix, which is assumed to have full column rank $M_0 < M - 1$ and the columns linearly independent of the $M \times 1$ column vector $\boldsymbol{J}_M = (1, \dots, 1)^{\mathrm{T}}$. We denote by $\boldsymbol{X}$ the matrix defined by

$$\boldsymbol{X} = (\boldsymbol{J}_M, \boldsymbol{W}). \tag{4}$$

Cressie and Pardo (2000), assuming $\phi''(1) > 0$, considered the following minimum ϕ-divergence estimator in loglinear models:

$$\widehat{\boldsymbol{\theta}}_\phi = \arg \min_{\boldsymbol{\theta} \in \Theta} D_\phi(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})), \tag{5}$$

where

$$D_\phi(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = \sum_{j=1}^{M} p_j(\boldsymbol{\theta}) \phi \left(\frac{\widehat{p}_j}{p_j(\boldsymbol{\theta})} \right) \tag{6}$$

is the ϕ-divergence measure defined simultaneously by Ali and Silvey (1966) and Csiszar (1967). We shall assume that $\phi \in \Phi^*$, where Φ^* is the class of all convex functions $\phi(x)$, $x > 0$, such that at $x = 1$, $\phi(1) = 0$, and at $x = 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u \to \infty} \phi(u)/u$. 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 $D_\psi(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta})) = D_\phi(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\theta}))$, 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^*. A complete study about ϕ-divergence measures can be seen in Vajda (1989), Pardo (2006) and some interesting results in loglinear models on the basis of the ϕ-divergence measures can be seen in Cressie and Pardo (2000, 2002b), Cressie et al. (2003) and Pardo and Pardo (2003, 2004).

Cressie and Pardo (2000) established, assuming that $\phi''(1) > 0$, that the minimum ϕ-divergence estimator for loglinear models, defined in (5), has the property that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_0) \xrightarrow[n \to \infty]{L} \mathrm{N}\left(\boldsymbol{0}_{M_0 \times 1}, (\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{W})^{-1}\right), \tag{7}$$

where $\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} \boldsymbol{W}$ is the Fisher information matrix associated with the loglinear model defined in (3) and$\boldsymbol{\Sigma}_{\boldsymbol{p}(\boldsymbol{\theta}_0)} = \mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta}_0)) - \boldsymbol{p}(\boldsymbol{\theta}_0)\boldsymbol{p}(\boldsymbol{\theta}_0)^{\mathrm{T}}$, which $\mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta}_0))$ defined by $\mathrm{diag}(\boldsymbol{p}(\boldsymbol{\theta}_0)) = \mathrm{diag}(p_1(\boldsymbol{\theta}_0), \dots, p_M(\boldsymbol{\theta}_0))$. The asymptotic variance–covariance matrix of $\widehat{\boldsymbol{\theta}}_\phi$ can be estimated by

$$\widehat{\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_\phi)} = \frac{1}{n} \left(\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{p}(\widehat{\boldsymbol{\theta}}_\phi)} \boldsymbol{W} \right)^{-1}. \tag{8}$$

It is well known that the maximum likelihood estimator (MLE) for loglinear models, see Cressie and Pardo (2000), is obtained from (5) if we consider $\phi(x) = x \log x - x + 1$, i.e., the Kullback–Leibler divergence measure. In the following, the MLE will be denoted by $\widehat{\boldsymbol{\theta}}$.