Available online at www.sciencedirect.com

ScienceDirect

Mathematics and Computers in Simulation 81 (2011) 1978–1990

MATHEMATICS AND COMPUTERS IN SIMULATION

www.elsevier.com/locate/matcom

Original article

Estimating Rao's statistic distribution for testing uniform association in cross-classifications

V. Alba-Fernández [a,*], M.D. Jiménez-Gamero [b]

[a] Dpto. de Estadística e I.O., Universidad de Jaén, Paraje Las Lagunillas s.n., 23071 Jaén, Spain
[b] Dpto. de Estadística e I.O., Universidad de Sevilla, 41012 Sevilla, Spain

Received 29 October 2009; received in revised form 25 November 2010; accepted 3 January 2011
Available online 26 February 2011

Abstract

We consider the problem of testing uniform association in cross-classifications with ordered categories taking as test statistic a R_ϕ divergence. The asymptotic null distribution of any test statistic in this class is not free because it depends on the unknown true vector of probabilities, so in practice one has to approximate it in order to get an estimate of the null distribution. As an alternative approach we propose to approximate the null distribution of the test statistic by bootstrapping. We show that the bootstrap yields a consistent null distribution estimator. The finite sample performance of the bootstrap estimator is studied by simulation. We also compare it empirically with the asymptotic null approximation. From the simulations it can be concluded that it is worth calculating the bootstrap estimator, because it is more accurate than the approximation yielded by the asymptotic null distribution which, furthermore, cannot always be exactly computed. Finally, the results are applied to some real data sets.

Keywords: Uniform association; Burbea and Rao's divergence measure; Local odds ratio; Bootstrap; Consistency

1. Introduction

Let X and Y denote two categorical random variables, having I and J levels, respectively. Let $p_{i,j} = P(X = i, Y = j) > 0$, $1 \leq i \leq I$, $1 \leq j \leq J$, and denote $p = (p_{1,1},\ldots,p_{I,J})^t \in \mathbb{R}^{IJ}$. This joint probability distribution is usually displayed in a contingency table having I rows and J columns. We assume that the table has ordered row categories and ordered column categories. If $p_{i,j} = p_{i,.} \times p_{.,j}$, $1 \leq i \leq I$, $1 \leq j \leq J$, where $p_{i,.} = \sum_{j=1}^{J} p_{i,j}$ and $p_{.,j} = \sum_{i=1}^{I} p_{i,j}$, then the responses are independent. When the responses are not independent, then there is an association between them. In this case, a wide range of models can be considered in order to try to give a parsimonious description of the association.

Goodman [6] has proposed a class of models for the analysis of association in a contingency table with ordered rows and ordered columns which is based on the $(I-1)(J-1)$ local odds-ratios of the tables formed from adjacent rows (i.e. rows i and $i+1$) and adjacent columns (i.e. columns j and $j+1$). Let

$$\theta_{i,j} = \frac{p_{i,j}\,p_{i+1,j+1}}{p_{i+1,j}\,p_{i,j+1}}, \qquad 1 \leq i \leq I-1, \;\; 1 \leq j \leq J-1,$$

* Corresponding author. Tel.: +34 953212142; fax: +34 953212034.
E-mail address: mvalba@ujaen.es (V. Alba-Fernández).

0378-4754/$36.00 © 2011 IMACS. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.matcom.2011.01.014