

Nonparametric test for conditional independence in two-way contingency tables

Gery Geenens, geenens@stat.ucl.ac.be

Institut de Statistique, Université catholique de Louvain, voie du Roman Pays, 20,
1348 Louvain-la-Neuve, Belgium

Keywords: contingency tables; conditional independence; chi-square test; nonparametric regression; long-range dependent process.

AMS: 62H17, 62E20, 62G08, 62G20, 62M07.

Abstract

Consider a two-way contingency table, built on two categorical variables R and S . A fundamental question in this context is whether R and S are independent or not. This over-studied testing problem mostly relies on the chi-square or the likelihood ratio test statistics. The principal drawback of this classical way of doing is that probabilities for an individual to fall in a cell of the table are assumed to be equal from one individual to another, so that we are not treating each individual as such but rather a group of supposed homogeneous units. This is often highly unrealistic, since in most of the practical situations some possibly known characteristics of each individual ought to influence or be associated with R , S , or both, and therefore influence the whole dependence structure of the table. In this case, a more judicious idea seems to analyze the conditional joint distribution of R and S given the vector of covariates, say X , and then to test for the conditional independence between R and S given X . In the literature, a striking fact is that the estimation of conditional probabilities associated with categorical responses, given a vector of covariates, is almost always treated via logistic regression methods, most of the time with very few validation of this parametric assumption. In this talk, we first present a nonparametric estimation procedure for the conditional probabilities to fall in each cell of the table. These estimates can be used as such, or be employed to validate a parametric assumption, like the logistic one. Secondly, we propose a generalization of the chi-square and the likelihood ratio tests to the case of conditional independence test, based on the above-mentioned nonparametric estimates of the conditional probabilities. The asymptotic law of the proposed test statistics is derived, and a simulation study illustrates the performance of the procedure in finite-sample situations.

Acknowledgements: Financial support from the "Interuniversity Attraction Pole", Phase VI from the Belgian Science Policy is gratefully acknowledged.