Comparing Treatments Based on Bivariate Categorical Responses

Chihiro Hirotsu
Meisei University, Faculty of Science and Technology
2-1-1 Hodokubo, Hino-City
Tokyo 191-8506, Japan
hirotsu@ge.meisei-u.ac.jp

Satoshi Aoki
University of Tokyo, Graduate School of Engineering
7-3-1 Hongo,Bunkyo-Ku
Tokyo 113-8656, Japan
aoki@stat.t.u-tokyo.ac.jp

1. Introduction

There are some but not many articles dealing with the ordered alternative for 3-way interaction effects in a 3-way contingency table. Hirotsu (1982) considered testing the ordered alternative

$$H_1 : \beta_1 \leq \ldots \leq \beta_K$$

for the log odds ratios β_K in the K ordered 2×2 proposed the cumulative chi-squared test based on a complete class lemma. Such a problem occurs when there are K subgroups classified by age, for example and each age group is further classified into 2×2 subgroups by smoker or nonsmoker and getting cancer or not. Then if the odds ratio of getting cancer between the smoker and nonsmoker is denoted by β_K for the k th age group, we are naturally interested in the monotone hypothesis like H_1. Barmi (1997) derived an order restricted likelihood ratio test for the same problem. In the present paper we propose a maximal contrast type test statistic for the problem and derives an exact algorithm for calculating the p value whereas the previous methods are based on the asymptotic theory. The method can be extended to a more general situation comparing treatments based on bivariate ordered categorical responses where the data are obtained in the form of $2 \times J \times K$ contingency table.

2. Mathematical formulation

Let the variable y_{11k}, $k = 1, \ldots, K$, be distributed independently as a generalized hypergeometric distribution $L_k(\beta_k)$ with given margins $y_{i \bullet k}$, $y_{\bullet jk}\,(i, j = 1,2)$ and the log odds ratio parameter β_K,

$$L_k(\beta_k) = C^{-1} \exp(y_{11k}\beta) / (\textstyle\prod_{i,j} y_{ijk}!),$$

so that the likelihood function of data is given by $L = \prod_k L_k(\beta_k)$. Then our test against the ordered alternative H_1 is based on the maximal component T of the cumulative efficient scores t_k, $k = \ldots, u-1$ evaluated and standardized under the null hypothesis

$$H_0 : \beta_1 = \ldots = \beta_K, \quad T = \max t_k, \quad t_k = \{Y_{11k} - E_0(Y_{11k})\} / (V_{11k})^{1/2}, \quad Y_{11k} = y_{111} + \cdots + y_{11k}$$

The calculation of the expectation and variance for standardization of Y_{11k} under H_0 is shown in Hirotsu (1982) and the test based on T is supported by a complete class lemma given there.

3. Exact algorithm for calculating p value

The p value of the test is obtained if the joint distribution function

$$F = \Pr(t_1 \leq t, \ldots, t_{K-1} \leq t \mid H_0\}$$

is obtained, where we consider a conditional distribution given the sufficient statistic $y_{11\bullet} = Y_{11k}$ under H_0, so that the following distribution theory is actually conditional on all the marginal totals $y_{ij\bullet}$, $y_{i\bullet k}$, $y_{\bullet jk}$ which we shall omit from the notation for brevity. Now we have a very efficient recurrence formula for calculating F based on the obvious Markov property of y_{11k}, $k = 1, \ldots, K$.

Define a conditional probability

$$F_k(Y_{11k}) = \Pr(t_1 \le t, \ldots, t_k \le t \mid Y_{11k}, H_0\},$$

and initial function

$$F_1(Y_{111}) = \begin{cases} 1, & if \quad t_1 \le t, \\ 0, & otherwise. \end{cases}$$

Then we have a recurrence formula

$$F_{k+1}(Y_{11k+1}) = \sum_{Y11k} F_{k+1}(Y_{11k+1}) f_0(Y_{11k} \mid Y_{11k+1}), \quad k = 1, \ldots, K-1,$$

where the range of Y_{11k} is restricted by the given marginal totals and also by Y_{11k+1}. Finally the cumulative distribution function F is obtained as Y_{11k}, where $F_K(Y_{11k})$ is the fixed marginal total $y_{11\bullet}$.

4. The conditional null distribution $f_0(Y_{11k} \mid Y_{11k+1})$

It should be noted that the conditional null distribution $f_0(Y_{11k} \mid Y_{11k+1})$ is not a simple null distribution for the accumulated table excepting the first step of $k = 1$ since in the multiplicative model the amalgamation invariance does not hold for the no three-way interaction (Plackett, 1981). However, again a simple recurrence formula is obtained for calculating $f_0(Y_{11k} \mid Y_{11k+1})$.

5. Concluding remarks

A maximal contrast type test is proposed for testing monotone hypothesis in K odds ratios for which a very efficient and exact algorithm is available for calculating the p value. Some power comparisons of the available methods will be given. The statistic is useful also for a more general situation comparing two treatments based on bivariate ordered categorical responses, where the data are obtained in the form of a $2 \times J \times K$ contingency table.

REFERENCES

Barmi, H. (1997). Testing for or against a trend in odds ratios. Commun. Statist.-Theory Meth., 26(8), 1877-1891.

Hirotsu, C. (1982). Use of cumulative efficient scores for testing ordered alternatives in discrete models. Biometrika, 69(3), 567-577.

Plackett, R. L. (1981). The analysis of categorical data, 2nd edition. London: Griffin.

RESUME