On the permutation test for canonical correlation coefficients.

Yamada Tomoya
Department of Economics, Sapporo Gakuin University
11 Bunkyo-dai
Ebetsu, Hokkaido, JAPAN
tomoya@sgu.ac.jp

Sugiyama Sanae
Department of Food and Nutrition, Seitoku Junior College of Nutrition
1-4-6 Nishi-shinkoiwa
Katsushika, Tokyo, JAPAN

1. Introduction

The permutation test with respect to canonical correlation analysis is considered in this paper. The basic idea of permutation test is attractively simple and free of mathematical assumptions. This test has complexity of calculation by computer. It is serious especially in multivariate analysis. Modern computational power, however, makes permutation test practical even in the multivariate problems. We apply the permutation test to canonical correlation analysis, one of the multivariate analyses. Detailed account for the permutation test is written in Efron, B and Tibshirani, R. J. (1993), and Pesarin, F. (1999).

2. Permutation test for the canonical correlation

In this section, we discuss the algorithm of the permutation test for the canonical correlation, the test statistics and its asymptotic bias. Let $X(p \times 1)$, $Y(q \times 1)$ be two sets of variates having covariance matrix Σ. Without loss of generality, we assume that $p \leq q$. Let ρ_α be the αth canonical correlation and $U_\alpha^{(1)}$ and $U_\alpha^{(2)}$ be the linear combinations of X and Y. Also, let us r_α be the sample canonical correlation and $U_\alpha^{*(1)}$, $U_\alpha^{*(2)}$ be the sample canonical variates of X and Y. We want to test whether the canonical coefficient is some given number $\rho_{\alpha 0}$;

$$H_0; \rho_\alpha = \rho_{\alpha 0}.$$

In particular, we are interested in the first canonical coefficient, because it has <u>most</u> information of the discrepancy between two variates. To do permutation test for the canonical correlation, we consider the permutation of the linear combinations. The advantage that we consider the permutation of the linear combinations is that $U_\alpha^{(i)}$ and $U_\beta^{(j)}$ $(\beta \neq \alpha,\ j = 1, 2)$ are uncorrelated except for $U_\alpha^{(j)}$ $(j \neq i)$ by orthogonality. With respect to ρ_α, we only consider $U_\alpha^{(1)}$ and $U_\alpha^{(2)}$, or $U_\alpha^{*(1)}$ and $U_\alpha^{*(2)}$. So, we do the permutation between $U_\alpha^* = (U_{\alpha 1}^{*(1)}, \cdots, U_{\alpha n}^{*(1)}, U_{\alpha 1}^{*(2)}, \cdots, U_{\alpha n}^{*(2)})$. Note that $U_{\alpha 1}^{(1)}$, $\cdots$, $U_{\alpha n}^{(1)}$, $U_{\alpha 1}^{(2)}$, $\cdots$, $U_{\alpha n}^{(2)}$ are exactly exchangeable if x_1, $\cdots$, x_n, y_1, $\cdots$, y_n are exchangeable, though $U_{\alpha 1}^{*(1)}$, $\cdots$, $U_{\alpha n}^{*(1)}$, $U_{\alpha 1}^{*(2)}$, $\cdots$, $U_{\alpha n}^{*(2)}$ are asymptotically exchangeable. Then, we consider the following transformation:

$$V_\alpha^* = \begin{pmatrix} V_\alpha^{*(1)} \\ V_\alpha^{*(2)} \end{pmatrix} = A^{-1} \begin{pmatrix} U_\alpha^{*(1)} \\ U_\alpha^{*(2)} \end{pmatrix} \qquad where \qquad AA' = \begin{pmatrix} 1 & \rho_{\alpha 0} \\ \rho_{\alpha 0} & 1 \end{pmatrix}.$$

By the transformation, the null hypothesis becomes $H_0^* : \Sigma^* = I$. Transformed data is uncorrelated under the null hypothesis, so it is useful for permutation test. We use the Fisher's z to the transformed data as a test statistic. It is serious for the permutation test that the expectation of the test statistic has bias. In canonical correlation analysis, higher dimension produces more bias. To solve this problem, we consider the modified correlation $\rho_{\alpha 0}^* = \sqrt{\rho_{\alpha 0}^2 + n^{-1/2}(1-\rho_{\alpha 0}^2)(p+q-1-2\rho_{\alpha 0}^2)}$ instead of $\rho_{\alpha 0}$. This modification becomes less bias for the test statistic.

3. Simulation study

We examine the reliability of our method by simulation study under multivariate normal, multivariate contaminated normal and multivariate t case. The results of the simulation study will be given in the presentation.

REFERENCES

Efron, B. Tibshirani, R.J. (1993). *An Introduction to the bootstrap. Chapman & Hall.*

Muirhead, R. J. and Waternaux, C.(1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika.* **67**, 31-43.

Pesarin, F. (1999). *Permutation testing of multidimensional hypotheses by nonparametric combination of dependent tests.*, Padova:Cleup Editrice.

Seo, T, Kanda, T and Fujikoshi, Y.(1994). The effects on the distributions of sample canonical correlations under nonnormality. *Commun. Statist. Theory Met.*, **23**, 2615-2628.

RESUME

In canonical correlation analysis, we are interested in the test of the αth canonical correlation coefficient, especially first coefficient. In this paper, we try permutation test for the canonical correlation analysis and discuss the test statistics and algorithm of the permutation test.