Permutation and rotation tests based on pairwise distance measures in high-dimensional data

Siegfried Kropf[1] and Daniela Kose[1]

[1] Institute of Biometry and Medical Informatics, Otto von Guericke University Magdeburg, Germany

When the number of variables exceeds the sample size in multivariate data, traditional multivariate tests are no longer applicable. One alternative approach is dimension reduction by principal component methods oder related techniques. In this context, Läuter et al. in [1] proposed exact parametric tests based on left-spherically distributed matrices. One of them is the principal component test (PC test).
In the talk we consider a different strategy to cope with the high dimension. We consider tests based on pairwise distance or similarity measures between the sample elements. In a comparison of independent samples, e.g., the distances between two observation vectors from different samples should be the same as those from two vectors from the same sample on the average under the null hypothesis, whereas they would be different under the alternative hypothesis. This basic idea can easily be utilized in exact permutation tests for some simple but important test problems as shown in [2]. Permutation tests, however, are not applicable in very small samples, due to a restricted number of possible permutations, and they are only asymptotical in more complex designs.
Rotation tests (cf. [3]) can be considered as parametric analogue to the permutation tests based on left-spherically distributed sample matrices. Instead of a finite number of permutations, an infinite (in practice very large) number of random rotations of the sample matrices is considered to construct the conditional null distribution of the test statistic enabling rather free choices of the test statistic. This again is utilized in tests based on pairwise distance or similarity measures where the pairwise measures can be adapted to the field of application. Given the necessary normality assumptions, these tests can even be used in very small samples and in more complex test designs, cf. [4].
The efficiency of tests based on pairwise distances is demonstrated and compared to PC tests in examples from microbial fingerprints and from occupational medicine. In these comparisons, the tests based on pairwise measures perform very well.

References

[1] Läuter J, Glimm E, Kropf S (1998) Multivariate tests based on left-spherically distributed linear scores. Linkage disequilibrium mapping identifies a 390 kb region associated with CYP2D6 poor drug metabolising activity. Annals of Statistics 26:1972-1988

[2] Kropf S, Heuer H, Grüning M, Smalla K (2004) Significance test for comparing complex microbial community fingerprints using pairwise similarity measures. Journal of Microbiological Methods 57/2:187-195

[3] Langsrud Ø(2005) Rotation tests. Statistics and Computing 15:53-60

[4] Kropf S, Lux A, Eszlinger M, Heuer H, Smalla K (2007) Comparison of independent samples of high-dimensional data by pairwise distance measures. Biometrical Journal 49:230-241