Multivariate Nonparametric Methods Based on Spatial Signs and Ranks

Hannu Oja

University of Turku, Finland

Abstract

Classical multivariate statistical inference methods (Hotelling's T^2, multivariate analysis of variance, multivariate regression, etc.) are based on the use of L_2 criterion function: If $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_n)^T$ is the matrix of p-variate residuals, then in the estimation one minimizes the criterion function

$$\frac{1}{n}\sum_{i=1}^{n} ||\mathbf{z}_i||^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{z}_i \quad (\text{or} \quad \sum_{i=1}^{n} ||\mathbf{z}_i||^2_{\mathbf{W}} = \sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{W}^{-1} \mathbf{z}_i).$$

In this talk we consider alternative L_1 criterion functions with related tests and estimates.

The univariate concepts of sign and rank are based on the ordering of the data. Unfortunately, in the multivariate case there are no natural orderings of the data points. An approach utilizing objective or criterion functions is therefore often used to extend the concepts of sign and rank to the multivariate case. The multivariate spatial sign $\mathbf{u}_i$, multivariate spatial (centered) rank $\mathbf{r}_i$, and multivariate spatial signed-rank $\mathbf{q}_i$, $i = 1, ..., n$ may be implicitly defined using the three L_1 criterion functions

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} ||\mathbf{z}_i|| &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_i^T \mathbf{z}_i, \\
\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n} ||\mathbf{z}_i - \mathbf{z}_j|| &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_i^T \mathbf{z}_i, \text{and} \\
\frac{1}{4n}\sum_{i=1}^{n}\sum_{j=1}^{n} (||\mathbf{z}_i - \mathbf{z}_j|| + ||\mathbf{z}_i + \mathbf{z}_j||) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i.
\end{aligned}$$

Note also that the sign, centered rank, and signed-rank may be seen as (location) *scores* $\mathbf{T}_i = \mathbf{T}(\mathbf{z}_i)$ corresponding to the three objective functions. ($\mathbf{T}(\mathbf{z}) = \mathbf{z}$ is the score function for the regular L_2 criterion.) Note that the first objective function, the *mean deviation* of the residuals, is the basis for the so called least absolute deviation (LAD) methods; it yields different median-type estimates and sign tests in the one-sample, two-sample, c-sample and

finally general linear model settings. The second objective function is the *mean difference* of the residuals. The second and third objective functions generate Hodges-Lehmann type estimates and rank and signed-rank tests for different location problems.

In the talk, we review the theory of the multivariate spatial sign and rank methods, tests and estimates, in the one sample, several samples and, finally, multivariate linear regression cases. See Möttönen and Oja (1995), Oja and Randles (2004), Oja (2010) and Nordhausen and Oja (2011). Transformation-retransformation technique is used to obtain affine invariant tests and equivariant estimates. Limiting distribution of the tests and estimates are given and their asymptotical efficiencies are discussed and compared. The theory is illustrated with examples.

Acknowledgement This talk is based on the joint work with several people, Jyrki Möttönen, Jaakko Nevalainen, Klaus Nordhausen, Ron Randles, Seija Sirkiä and Sara Taskinen among others.

Keywords

Multivariate analysis of variance, Multivariate Hodges-Lehmann estimator, Multivariate L_1 regression, Multivariate median,

References

Möttönen, J. and H. Oja (1995). Multivariate spatial sign and rank methods. *J. Nonp. Statist.*, *5*, 201-213.

Nordhausen, K. and Oja, H. (2011). Multivariate L1 methods: The package MNM. *Journal of Statistical Software*, 43(5), 1-28.

Oja, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks.* Springer, New York

Oja, H. and R. Randles (2004). Multivariate nonparametric tests. *Statistical Science*, *19*, 598-605.