Robust monitoring of a multivariate data stream

Daniel Kosiorowski and Małgorzata Snarska

Cracow University of Economics, Poland

Abstract

A data stream could be defined as continuous sequence of ordered observations of indeterminate length. Because the data are arriving continuously and there is no known end to it, the classical approach of reading in all data and then processing them is not feasible. Data stream carry signals that appear randomly, are irregularly spaced and the time duration between successive signals is not deterministic but random. Additionally data streams generally are generated by multivariate non-stationary models of unknown form.

In this paper we present two approaches to robust analysis of multivariate economic data streams. We use the information-theoretic approach proposed by [1] based on the relative Kullback-Leibler entropy and bootstrapping to extract possible changes and Kulldorff's spatial scan statistics to identify regions where large changes have occurred. The second proposal appeals to the data depth concept. We propose a variant of a multivariate Wilcoxon statistic (a scale change), depth vs. depth plot based statistics (a location shift) and robust semi-parametric regression to monitor a linear relationship between multivariate data stream components.

Keywords

Data stream, Robust spectral analysis, Depth function, Relative entropy.

References

[1] Dasu, T., S. Krishnan, S. Venkatasubramanian, and K. Yi (2006). An information-theoretic approach to detecting changes in multi-dimensional data streams. *Proceedings of the 38th Symposium on the Interface of Statistics, Computing Science, and Applications (Interface '06)*, Pasadena, CA.

[2] Jureckova, J. and J. Kalina (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli 18(1)*, 229-251.

[3] Wellman, R. and Ch.H. Müller (2010). Depth notions for orthogonal regression. *J. Multivariate Anal. 101*, 2358–2371.