Independent Component Analysis via Distance Covariance

David S. Matteson[1] **and Ruey S. Tsay**[2]
[1] Cornell University, U.S.
[2] University of Chicago, U.S.

This paper introduces a novel statistical framework for independent component analysis (ICA) of multivariate data. We propose methodology for estimating and testing the existence of mutually independent components for a given dataset, and a versatile resampling-based procedure for inference. Independent components are estimated by combining a nonparametric probability integral transformation with a generalized nonparametric whitening method that simultaneously minimizes all forms of dependence among the components. U-statistics of certain Euclidean distances between sample elements are combined in succession to construct a statistic for testing the existence of mutually independent components. The proposed measures and tests are based on both necessary and sufficient conditions for mutual independence. When independent components exist, one may apply univariate analysis to study or model each component separately. Univariate models may then be combined to obtain a multivariate model for the original observations. We prove the consistency of our estimator under minimal regularity conditions without assuming the existence of independent components *a priori*, and all assumptions are placed on the observations directly, not on the latent components. We demonstrate the improvements of the proposed method over competing methods in simulation studies. We apply the proposed ICA approach to two real examples and contrast it with principal component analysis.

Keywords: Dimension reduction, Multivariate analysis, Mutual independence test, Nonparametric methods, Principal component analysis, Unsupervised learning.

References:

Matteson, D. S. and Tsay, R. S. (2013). Independent Component Analysis via Distance Covariance *ArXiv e-prints 1306.4911,* `http://arxiv.org/abs/1306.4911`.