

Testing and segmentation of joint and individual components in integrative multi-source factor models

Kyoowon Kim¹, Sungkyu Jung²

¹Department of Statistics, Seoul National University, kwkim1224@snu.ac.kr

²Department of Statistics and Institute for Data Innovation in Science, Seoul National University, sungkyu@snu.ac.kr

Abstract

Disentangling shared (joint) structures from source-specific (individual) variations is a fundamental task in multi-source data integration. Existing joint-individual models often rely on computationally intensive optimization or loose spectral bounds, leading to suboptimal separation accuracy and poor scalability. In this paper, we propose the Multi-Source Sequential Alignment Test (MSSAT). MSSAT leverages the geometric observation that true joint components manifest as closely aligned score subspaces across different data sources. By deriving the asymptotic null distribution of our alignment statistic, we develop a rigorous, resampling-free sequential testing procedure to accurately estimate the joint rank. Extensive simulations and real data applications, including a TCGA multi-omics dataset, demonstrate that MSSAT achieves superior separation accuracy and substantially faster computation compared to competing methods.