On the Asymptotic Properties of Product-PCA under the High-Dimensional Setting

Hung Hung

Institute of Health Data Analytics and Statistics, National Taiwan University, Taiwan

ABSTRACT

Principal component analysis (PCA) is a widely used dimension reduction method, but is known to be non-robust in the presence of outliers. Recently, product-PCA (PPCA) has been shown to possess the efficiency-loss free ordering-robustness property: (i) in the absence of outliers, PPCA and PCA share the same asymptotic distributions; (ii) in the presence of outliers, PPCA is more ordering-robust than PCA in estimating the leading eigen-subspace. This property makes PPCA different from the branch of the conventional robust PCA methods (which usually suffer the problem of efficiency loss in an exchange of robustness gain), and may deserve further investigations. In this article, we study the high-dimensional asymptotic properties of the PPCA eigenvalues via the techniques of random matrix theory. In particular, we derive the critical value for being distant spiked eigenvalues, the limiting values of the sample spiked eigenvalues, and the limiting spectral distribution of PPCA. These results enable us to more clearly observe the superiorities of PPCA in comparison with PCA. Similar to the case of PCA, the explicit forms of the asymptotic properties of PPCA become available under the special case of the simple spiked model. Numerical studies are conducted to verify our results.

Keywords: efficiency-loss; PCA; random matrix theory; robustness

Automatic Sparse Estimation of High-Dimensional Covariance Matrices

Tetsuya Umino[1], **Kazuyoshi Yata[2]**, Makoto Aoshima[2]

[1] *Degree Programs in Pure and Applied Sciences, Graduate School of Science and Technology, University of Tsukuba*

[2] *Institute of Mathematics, University of Tsukuba*

ABSTRACT

Sparse principal component analysis (SPCA) has been widely investigated as a framework for estimating sparse principal component directions in high-dimensional data. A recent study by Yata and Aoshima (Statistica Sinica, 35) proposed the automatic SPCA (A-SPCA), which determines its threshold adaptively and has been shown to achieve consistency under mild conditions. Motivated by this development, we extend the A-SPCA methodology to the estimation of high-dimensional covariance matrices as well as mean vectors, and we introduce novel automatic sparse estimators within this framework. The proposed estimators inherit the adaptability of A-SPCA and are theoretically proven to be consistent under mild assumptions in high-dimensional settings, including challenging high-dimension, low-sample-size contexts. Their performance is carefully examined through numerical simulations, which demonstrate both accuracy and stability across a range of scenarios. In addition, we extend the approach to high-dimensional regression by employing the proposed estimators. Applications to gene expression data confirm their effectiveness in practice.

Keywords: Cross-covariance matrix; Extended cross-data-matrix methodology; Large p small n; High-dimensional regression

Collaborative and Federated Black-Box Optimization: A Bayesian Optimization Perspective

Raed Al Kontar*

Industrial & Operations Engineering, University of Michigan

ABSTRACT

We focus on collaborative and federated black-box optimization (BBOpt), where agents optimize their heterogeneous black-box functions through collaborative sequential experimentation. From a Bayesian optimization perspective, we address the fundamental challenges of distributed experimentation, heterogeneity, and privacy within BBOpt, and propose three unifying frameworks to tackle these issues: (i) a global framework where experiments are centrally coordinated, (ii) a local framework that allows agents to make decisions conditioned on shared information, and (iii) a predictive framework that enhances local surrogates through collaboration to improve decision-making. We categorize existing methods within these frameworks and highlight key open questions to unlock the full potential of federated BBOpt. Our overarching goal is to shift federated learning from its predominantly descriptive/predictive paradigm to a prescriptive one, particularly in the context of BBOpt — an inherently sequential decision-making problem.

Keywords: Federated Leaning, Personalization, Bayesian Optimization, Experimental Design, Distributed Learning

Effective Permutation Tests for Differences Across Multiple High-Dimensional Correlation Matrices

José A. Sánchez Gómez[1], **Elio Zhang**[2], Yufeng Liu*

[1] *Department of Statistics, University of California at Riverside*

[2] *Department of Mathematics, Seattle University*

* *Department of Statistics, University of Michigan*

ABSTRACT

Testing the equality of two or multiple correlation or covariance matrices is an important problem in biology, finance and many other areas. High dimensionality, where the number of features is much larger than the sample size, causes conventional procedures to perform poorly, as they are often based on limiting distributions of test statistics in the classical large sample size setting. Moreover, their performance is often contingent on whether the matrix of differences is sparse or dense, while such information is rarely available. In this article, we develop a new family of permutation testing procedures to tackle these challenges. The introduced tests are demonstrated to outperform many other competing procedures in terms of size control and power under various settings. In particular, using our variance-stabilizing transformation, the proposed methods provide the best performance for testing correlation or covariance matrix differences in both sparse and dense settings. We establish non-asymptotic guarantees on the power of our test, which ensure its reliability for sparse and dense differential correlation matrices. Through the analysis of gene-expression and brain imaging data, we showcase the high power and accurate size control of our test in high-dimensional statistical applications.

Keywords: Brain Activation; Concentration Inequalities; Covariance Testing; Gene Expression; High-Dimensional Statistics