

New Techniques in High-Dimensional Statistical Analysis: SSE vs. NSSE and Data Transformation

Makoto Aoshima

University of Tsukuba, Japan

Abstract

Any high-dimensional data is classified into two disjoint models: the strongly spiked eigenvalue (SSE) model and the non-SSE (NSSE) model. In actual high-dimensional data, a non-sparse and low-rank structure which contains strongly spiked eigenvalues is often found; a structure which fits the SSE model. Under the SSE model, it may be noted that the asymptotic normality of high-dimensional statistics is not valid because it is heavily influenced by strongly spiked eigenvalues. To enable a unified treatment of both the SSE models and non-SSE models, data transformation techniques that transform the SSE models to the non-SSE models were developed by Aoshima and Yata (*Statist. Sinica*, 2018; *Ann. Inst. Statist. Math.*, 2018). Following these novel methodologies, strongly spiked eigenvalues are accurately detected by using new PCA-type techniques. With the transformed data, one can create a new statistic which can ensure high accuracy of inferences by using asymptotic normality even under the SSE models. I will demonstrate the new techniques to handle ultra-high-dimensional genomic data to solve two-sample problems and classification problems in Aoshima and Yata (*Methodol. Comput. Appl.*, 2018).

[The talk is based on joint work with Kazuyoshi Yata (University of Tsukuba).]