Robust Semi-parametric PCA

Hung Hung[1], Su-Yun Huang[2], Shinto Eguchi[3]

[1]*Institute of Epidemiology and Preventive Medicine and Department of Public Health, National Taiwan University, Taiwan*
[2] *Institute of Statistical Science, Acacemia Sinica, Taiwan*
[3]*Institute of Statistical Mathematics, Tokyo, Japan*

Abstract

Principal component analysis (PCA) is one of the most popular dimension reduction methods. The usual PCA is known to be sensitive to the presence of outliers, and thus many robust PCA methods have been developed. Among them, the Tyler's M-estimator is shown to be the most robust scatter estimator under the elliptical distribution. However, when the underlying distribution is contaminated and deviates from ellipticity, Tyler's M-estimator might not work well. In this article, we apply the semiparametric theory to propose a robust semiparametric PCA, which is shown to be a re-weighted Tyler's M-estimator. The merits of our proposal are twofold. First, it is robust to both heavy-tailed elliptical outliers and nonelliptical outliers. Second, it pairs well with a data-driven tuning procedure, which is based on active ratio and can adapt to different degrees of data outlyingness. Simulation studies and a data analysis demonstrate the superiority of our method.

Keywords: active ratio, elliptical distributions, influence function, PCA, robustness, semiparametric theory, Tyler's M estimator.