UNIVERSITÉ LIBRE DE BRUXELLES

Rank Tests for PCA under Weak Identifiability

Laura Peralvo Maroto[1]

[1] *ECARES et Département de Mathématique, Université Libre de Bruxelles*

December 12, 2023 - 14:45

Abstract

In this presentation, we consider the problem of testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, in a triangular array framework where n observations are randomly sampled from a p-dimensional elliptical distribution with shape matrix $\mathbf{V}_n$, where $\boldsymbol{\theta}$ is the (fixed) leading unit eigenvector of $\mathbf{V}_n$ and $\boldsymbol{\theta}_0$ is a given unit p-vector. The dependence of the shape matrix on the sample size allows us to consider challenging asymptotic scenarios in which the parameter of interest $\boldsymbol{\theta}$ is unidentified in the limit, because the ratio between both leading eigenvalues of $\mathbf{V}_n$ converges to one. We carefully study the corresponding limiting experiments under such *weak identifiability*, and we show that these may be LAN or non-LAN. While earlier work in the framework was strictly limited to Gaussian distributions, where the study of local log-likelihood ratios could simply rely on explicit expressions, our asymptotic investigation allows for essentially arbitrary elliptical distributions. This requires original results on quadratic mean differentiable families for triangular arrays of observations, that are likely to be of interest in other models, too. Even in non-LAN experiments, our results enable us to investigate, through Le Cam's first and third lemmas, the asymptotic null and non-null properties of multivariate rank tests.

What can be said about the robustness of these tests under weak identifiability? Do they still satisfy the nominal level constraint irrespective of the amount of weak identifiability? What about efficiency? We'll try to answer these questions during the presentation.

Keywords: Elliptical densities, limiting experiments, multivariate signs and ranks, principal component analysis, spiked scatter matrices, triangular arrays of observations, weak identifiability.