A topic for a Master degree project in mathematical statistics

by Johannes Heiny (johannes.heiny@math.su.se), January 2023

Topic: High-dimensional sample correlation matrices

Measuring dependence and testing for independence are fundamental problems in statistics. In this context sample correlations are an indispensable tool in modern statistics. Consider a p-dimensional population $\mathbf{x} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p$ where the coordinates X_i are i.i.d. non-degenerated random variables with mean zero. For a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ from the population we construct the sample covariance matrix $\mathbf{S}_n = n^{-1} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t^\top$ and the sample correlation matrix $\mathbf{R}$,

$$\mathbf{R}_n = \{\mathrm{diag}(\mathbf{S}_n)\}^{-1/2} \, \mathbf{S}_n \{\mathrm{diag}(\mathbf{S}_n)\}^{-1/2} \,,$$

where $\mathrm{diag}(\mathbf{S}_n)$ denotes the diagonal matrix with the same diagonal elements as $\mathbf{S}_n$.

The aim of this project is to study limit theorems for the trace $\mathrm{tr}(\mathbf{R}_n^2)$ in a high-dimensional setting where the sample size n converges to infinity together with the dimension p. We apply the method of moments and calculate the expectations

$$\mathbb{E}\big[\big(\,\mathrm{tr}(\mathbf{R}_n^2)\big)^m\big] \,, \qquad m \in \mathbb{N} \,, \tag{0.1}$$

and their limits using some combinatorics. Some special cases are well-understood. For example, in the case $p = n$ and $\mathbb{E}[X_1^4] < \infty$, the properly standardized $\mathrm{tr}(\mathbf{R}_n^2)$ satisfies a central limit theorem and the limits of the moments in (0.1) are the moments of a Gaussian distribution.

The main goals are as follows:

(i) find conditions (on X_1 and the relationship of p and n) under which $\mathrm{tr}(\mathbf{R}_n^2)$ is asymptotically Gaussian;

(ii) find conditions under which $\mathrm{tr}(\mathbf{R}_n^2)$ is asymptotically non-Gaussian.

To get started, some simulation experiments might be useful.