Johannes Heiny (Aarhus University)

A comparison of high-dimensional sample covariance and correlation matrices of a heavy-tailed time series

Joint with Thomas Mikosch.

In Principal Component Analysis one studies the sample covariance or sample correlation matrix, both of which often lead to the same result. In this talk, we first analyze the joint distributional convergence of the largest eigenvalues of the sample covariance matrix of a p-dimensional heavy-tailed time series when p converges to infinity together with the sample size n. Assuming a regular variation condition with tail index $\alpha < 4$, we employ a large deviations approach to show that the extreme eigenvalues are essentially determined by the extreme order statistics from an array of iid random variables. The asymptotic behavior of the extreme eigenvalues is then derived routinely from classical extreme value theory. The resulting approximations are strikingly simple considering the high dimension of the problem at hand.

Then we compare the behavior of the eigenvalues of the sample covariance and sample correlation matrices and argue that the latter seems more robust, in particular in the case of infinite fourth moment.

We show that the largest and smallest eigenvalues of a sample correlation matrix stemming from n independent observations of a p-dimensional time series with iid components converge almost surely to $(1+\sqrt{\gamma})^2$ and $(1-\sqrt{\gamma})^2$, respectively, as $n \to \infty$, if $p/n \to \gamma \in (0,1]$ and the truncated variance of the entry distribution is "almost slowly varying", a condition we describe via moment properties of self-normalized sums. Moreover, the empirical spectral distributions of these sample correlation matrices converge weakly, with probability 1, to the Marčenko–Pastur law.

Key Words: PCA, regular variation, sample covariance matrix, sample correlation matrix, extreme eigenvalues, spectral distribution.