

On the principal components of sample covariance matrices

Abstract

I describe recent results on spiked covariance matrices, describing multivariate data with nontrivial correlations. In principal component analysis, one extracts the leading contribution to the covariance by analysing the top eigenvalues and associated eigenvectors of the sample covariance matrix. I give sharp large deviation bounds relating the principal components of the true covariance matrix to those of the sample covariance matrix. In particular, I discuss the behaviour of eigenvalues and eigenvectors at and near the so-called BBP phase transition, at which outlier eigenvalues are created or annihilated.