Some inference problems for the spiked covariance matrix in the high dimensional context

Addy Bolivar-Cime*
Juarez Autonomous University of Tabasco, Cunduacan, Mexico
addy.bolivar@gmail.com

Victor Perez-Abreu
Center for Research in Mathematics (CIMAT), Guanajuato, Mexico
pabreu@cimat.mx

Doing inference for the covariance matrix of data with high dimension greater than the sample size is very difficult, since in this case the covariance matrix has too many unknown parameters to be estimated with few information. In this talk we consider a kind of covariance matrix called the "spiked covariance matrix", which has a small number of large eigenvalues, say p, well separated from the rest. It is obtained the asymptotic joint distribution of the p largest sample eigenvalues when the dimension d of the data tends to infinity and keeping the sample size n fixed. Using this asymptotic result we study some inference problems for a special case of the spiked covariance matrix in the high dimensional context. More specifically, we develop some results behind hypothesis tests and confidence intervals in two asymptotic settings of the High-Dimension, Low Sample Size (HDLSS) context: when d goes to infinity and n is fixed; and when d, n go to infinity and $d \gg n$. Applying our asymptotic results under a Gaussian assumption, we propose hypothesis tests for our spiked covariance matrix and confidence intervals for the p largest population eigenvalues. It is seen that some classical statistics are also useful in the HDLSS context. We also present a simulation study to assess the behavior of the proposed statistical methodologies.

Key Words: Hypothesis test, Confidence interval, eigen-inference, high dimensional data, HDLSS.