Detecting Rare and Weak Spikes in Large Covariance Matrices

Tracy Ke

Harvard University
E-mail: zke@fas.harvard.edu

Abstract: Given p-dimensional Gaussian vectors X1, ..., Xn, where $p \geq n$, we are interested in testing a null hypothesis where $\Sigma = Ip$ against an alternative hypothesis where all eigenvalues of Σ are 1, except for r of them are larger than 1 (i.e., spiked eigenvalues). We consider a Rare/Weak setting where the spikes are sparse (i.e., 1<<r<<p) and individually weak (i.e., each spiked eigenvalue is only slightly larger than 1), and discover a phase transition: the two-dimensional phase space that calibrates the spike sparsity and strengths partitions into the Region of Impossibility and the Region of Possibility. In Region of Impossibility, all tests are (asymptotically) powerless in separating the alternative from the null. In Region of Possibility, there are tests that have (asymptotically) full power. We consider a CuSum test, a trace-based test, an eigenvalue-based Higher Criticism test, and a Tracy-Widom test, and show that the first two tests have asymptotically full power in Region of Possibility. To use our results from a different angle, we derive new bounds for (a) empirical eigenvalues, and (b) cumulative sums of the empirical eigenvalues, both under the alternative hypothesis. Part (a) is related to those in the literature, but both the settings and results are different. The study requires careful analysis of the L1-distance of our testing problem and delicate Radom Matrix Theory. Our technical devises include (a) a Gaussian proxy model, (b) Le Cam's comparison of experiments, and (c) large deviation bounds on empirical eigenvalues.