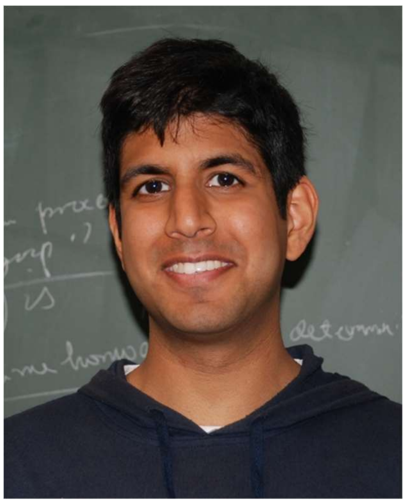
RUTGERS
School of Arts and Sciences

Department of Statistics

DEPARTMENT OF STATISTICS



Rajen Shah
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

RSVP-graphs: Fast High-dimensional Covariance Matrix Estimation Under Latex Confounding

April 15, 2020
11:45am – 12:45pm

Zoom Meeting: Meeting ID: 992 4099 7838
Password:003078

Abstract: We consider the problem of estimating a high-dimensional p × p covariance matrix Σ, given n observations of confounded data with covariance Σ + ΓΓ^T , where Γ is an unknown p × q matrix of latent factor loadings. We propose a simple and scalable estimator based on the projection on to the right singular vectors of the observed data matrix, which we call RSVP. Our theoretical analysis of this method reveals that in contrast to approaches based on removal of principal components, RSVP is able to cope well with settings where the smallest eigenvalue of Γ^T Γ is relatively close to the largest eigenvalue of Σ, as well as when eigenvalues of Γ^T Γ are diverging fast. RSVP does not require knowledge or estimation of the number of latent factors q, but only recovers Σ up to an unknown positive scale factor.

We argue this suffices in many applications, for example if an estimate of the correlation matrix is desired. We also show that by using subsampling, we can further improve the performance of the method. We demonstrate the favourable performance of RSVP through simulation experiments and an analysis of gene expression datasets collated by the GTEX consortium.

Bio: Rajen Shah is a Reader in Statistics at the Statistical Laboratory, which is part of the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge. He is also a fellow of the Alan Turing Institute and an associate editor of Annals of Statistics, JRSSB, and Biometrika. His research interests include high-dimensional statistics and large-scale data analysis.