An empirical estimator for the sparsity of a large covariance matrix under multivariate normal assumptions

Binyan Jiang

Abstract Large covariance or correlation matrix is frequently assumed to be sparse in that a number of the off-diagonal elements of the matrix are zero. This paper focuses on estimating the sparsity of a large population covariance matrix using a sample correlation matrix under multivariate normal assumptions. We show that sparsity of a population covariance matrix can be well estimated by thresholding the sample correlation matrix. We then propose an empirical estimator for the sparsity and show that it is closely related to the thresholding methods. Upper bounds for the estimation error of the empirical estimator are given under mild conditions. Simulation shows that the empirical estimator can have smaller mean absolute errors than its main competitors. Furthermore, when the dimension of the covariance matrix is very large, we propose a generalized empirical estimator using simple random sampling. It is shown that the generalized empirical estimator can still estimate the sparsity well while the computation complexity can be greatly reduced.