Statistical Methodology 20 (2014) 11–26

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Consistency of large dimensional sample covariance matrix under weak dependence

Monika Bhattacharjee, Arup Bose *

Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form
23 June 2013
Accepted 9 August 2013

Dedicated to the memory of Kesar Singh

Keywords:
High-dimensional data
Covariance matrices
Cross covariances
Regularization
Banding
Tapering
Convergence rate
Operator norm

ABSTRACT

Convergence rates for banded and tapered estimates of large dimensional covariance matrices are known when the vector observations are independent and identically distributed. We investigate the case where the independence does not hold. Our models can accommodate suitable patterned cross covariance matrices. These estimators remain consistent in the operator norm with appropriate rates of convergence under suitable class of models.

1. Introduction

New technologies and methods in medical sciences, image processing and the internet, and many other fields of science generate data where the dimension is high and the sample size is small relative to the dimension. For example, microarray data [7] contains the gene expression for tens of thousands of genes (rows) on a few observations (columns). Another example is fMRI data, which measures the hemodynamic response in hundreds of thousands of voxels (rows) for only a few subjects or replicates (columns). Similarly, the Netflix movie rating data [2] contains the rating information for approximately 480,000 customers (columns) on 18,000 movies (rows). Let $X_{p\times n}$ denote the

* Corresponding author. Tel.: +91 9830496250.
E-mail addresses: monaiidexp.gamma@gmail.com (M. Bhattacharjee), bosearu@gmail.com, abose@isical.ac.in (A. Bose).

1572-3127/$ – see front matter
http://dx.doi.org/10.1016/j.stamet.2013.08.005

corresponding data matrix:

$$X_{p\times n} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & x_{p3} & \dots & x_{pn} \end{bmatrix}$$

where the dimension $p = p(n)$ is assumed to be increasing with the sample size n. This type of data matrix has been modeled by identical Gaussian distribution of the columns

$$C_{ip} = (x_{1i}, x_{2i}, \dots, x_{pi})' \sim \mathcal{N}(\mu, \Sigma_p) \quad \forall i = 1, 2, \dots, n,$$

with mean vector $\mu \in \mathbb{R}^p$ and variance–covariance matrix $\Sigma_p = ((\sigma_{ij}))_{p\times p}$. The estimation of the large covariance matrix Σ_p is crucial for statistical inference procedures.

Usually one assumes further that the columns are independent/exchangeable. Thus, genes in microarrays, pixels in images, voxels in fMRIs and movies in Netflix movie-rating data are considered as dependent features whereas respectively the samples, repeated images, images with respect to different subjects or replications and customers are modeled to be independent.

However, this assumption has been questioned. Many have suggested that, in microarrays, the arrays are not independent (e.g., [13,8,11,12]). Latent variables such as age, gender, family history, underlying health status, measurement process, laboratory conditions may be responsible for the dependency between two patients. For the Netflix movie-rating data, a particular type of movies are likely to have similar ratings from customers having similar tastes. The latent variable time may be responsible for the dependency between two replications in fMRI data sets. Specific examples of this lack of independence can be found in [1].

Hence, there is need for models which allow for dependence between columns. [8] proposed the matrix-variate normal as a model for microarrays. Mean-restricted matrix-variate normal was considered by [1]. This distribution, denoted by $X_{p\times n} \sim \mathcal{N}_{p,n}(\nu, \mu, \Sigma_p, \Delta)$, has separate mean and covariance parameters for the rows, $\nu \in \mathcal{R}^p$, $\Sigma_p = ((\sigma_{ij}))_{p\times p}$, and the columns, $\mu \in \mathcal{R}^n$, $\Delta = ((\delta_{ij}))_{n\times n}$. If the matrix is transformed into a vector of length np, we have that $vec(X) \sim \mathcal{N}(vec(M), \Omega)$, where $M = ((\nu_i + \mu_j))_{p\times n}$, $\Omega = \Delta \otimes \Sigma_p$ and $\otimes$ is the Kronecker product between two matrices. In this model the correlation between columns is controlled without considering the effect of the components (rows); that is,

$$\frac{corr(x_{ki}, x_{lj})}{corr(x_{mi}, x_{mj})} = \frac{\delta_{kl}}{\sqrt{\delta_{kk}\delta_{ll}}} \quad \forall i, j = 1, 2, \dots p \text{ and } m = 1, 2, \dots, n.$$

We will assume that $C_{ip} \sim \mathcal{N}_p(0, \Sigma_p)$ $\forall i = 1, 2, \dots, n$ are identically distributed and the distribution is Gaussian with zero mean. However, we will allow dependence of appropriate nature between the columns. We call this dependence the *cross covariance structure*. In this paper we work under three different restrictions on cross covariance structures. In one case, the restriction is on the growth of the powers of the trace of certain matrices derived from the cross covariance structure. In the second case, the dependence among any two columns weakens as the lag between them increases and in the third case we assume weak dependence among the last few columns. See Section 2 for details.

The existing methods to estimate Σ_p (under column independence) involves banding or tapering of the sample variance–covariance matrix. [4] proved that suitably banded and tapered estimators are both consistent in the operator norm for the sample variance–covariance matrix as long as $n^{-1}\log p \to 0$ uniformly over some fairly natural well-conditioned families of covariance matrices. They also obtained some explicit rates.

In the first case, we show that the convergence rate of the banded estimator is the same as in the *i.i.d.* case of [4] (see Theorem 3.1 of Section 3) under a trace condition. We also provide some sufficient conditions that imply this trace condition. The other two cases do not fall under the purview of Theorem 3.1. Under appropriate conditions we obtain explicit rates of convergence for the banded estimators (see Theorems 3.2 and 3.3). In particular, for all three cases, the estimators continue to remain consistent in the operator norm.