Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 123 (2013) 2899–2920

stochastic processes and their applications

www.elsevier.com/locate/spa

Asymptotic theory for maximum deviations of sample covariance matrix estimates

Han Xiao[a,*], Wei Biao Wu[b]

[a] *Department of Statistics and Biostatistics, Rutgers University, 501 Hill Center, 110 Frelinghuysen Road, Piscataway, NJ 08854, United States*

[b] *Department of Statistics, The University of Chicago, 5734 S. University Ave, Chicago, IL 60637, United States*

Available online 28 March 2013

Abstract

We consider asymptotic distributions of maximum deviations of sample covariance matrices, a fundamental problem in high-dimensional inference of covariances. Under mild dependence conditions on the entries of the data matrices, we establish the Gumbel convergence of the maximum deviations. Our result substantially generalizes earlier ones where the entries are assumed to be independent and identically distributed, and it provides a theoretical foundation for high-dimensional simultaneous inference of covariances.

MSC: primary 62H15; 62H10; secondary 62E20

Keywords: Covariance matrix; High dimensional analysis; Maximal deviation; Tapering; Test for bandedness; Test for covariance structure; Test for stationarity

1. Introduction

Let $\boldsymbol{X}_n = \left(X_{ij}\right)_{1\le i\le n, 1\le j\le m}$ be a data matrix whose n rows are independent and identically distributed (i.i.d.) as some population distribution with mean vector $\boldsymbol{\mu}_n$ and covariance matrix Σ_n. High dimensional data increasingly occur in modern statistical applications in biology, finance and wireless communication, where the dimension m may be comparable to the number of

* Corresponding author.
E-mail addresses: hxiao@stat.rutgers.edu (H. Xiao), wbwu@galton.uchicago.edu (W.B. Wu).

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.03.012

observations n, or even much larger than n. Therefore, it is necessary to study the asymptotic behavior of statistics of $\boldsymbol{X}_n$ under the setting that $m = m_n$ grows to infinity as n goes to infinity.

In many empirical examples, it is often assumed that $\Sigma_n = I_m$, where I_m is the $m \times m$ identity matrix, so it is important to perform the test

$$H_0 : \Sigma_n = I_m \tag{1}$$

before carrying out further estimation or inference procedures. Due to high dimensionality, conventional tests often do not work well or cannot be implemented. For example, when $m > n$, the likelihood ratio test (LRT) cannot be used because the sample covariance matrix is singular; and even when $m < n$, the LRT is drifted to infinity and leads to many false rejections if m is also large [1]. Ledoit and Wolf [16] found that the empirical distance test [21] is not consistent when both m and n are large. The problem has been studied by several authors under the "large n, large m" paradigm. Bai et al. [1] and Ledoit and Wolf [16] proposed corrections to the LRT and the empirical distance test respectively. Assuming that the population distribution is Gaussian with $\boldsymbol{\mu}_n = 0$, [14] used the largest eigenvalue of the sample covariance matrix $\boldsymbol{X}_n^\top \boldsymbol{X}_n$ as the test statistic, and proved that its limiting distribution follows the Tracy–Widom law [27]. Here we use the superscript $^\top$ to denote the transpose of a matrix or a vector. His work was extended to the non-Gaussian case by Soshnikov [24] and Péché [22], where they assumed the entries of $\boldsymbol{X}_n$ are i.i.d. with sub-Gaussian tails.

Let $x_1, x_2, \ldots, x_m$ be the m columns of $\boldsymbol{X}_n$. In practice, the entries of the mean vector $\boldsymbol{\mu}_n$ are often unknown, and are estimated by $\bar{x}_i = (1/n) \sum_{k=1}^n X_{ki}$. Write $x_i - \bar{x}_i$ for the vector $x_i - \bar{x}_i \mathbf{1}_n$, where $\mathbf{1}_n$ is the n-dimensional vector with all entries being one. Let $\sigma_{ij} = \mathrm{Cov}(X_{1i}, X_{1j})$, $1 \le i, j \le m$, be the covariance function, namely, the (i, j)th entry of Σ_n. The sample covariance between columns x_i and x_j is defined as

$$\hat{\sigma}_{ij} = \frac{1}{n}(x_i - \bar{x}_i)^\top (x_j - \bar{x}_j).$$

In high-dimensional covariance inference, a fundamental problem is to establish an asymptotic distributional theory for the maximum deviation

$$M_n = \max_{1 \le i < j \le m} |\hat{\sigma}_{ij} - \sigma_{ij}|.$$

With such a distributional theory, one can perform statistical inference for structures of covariance matrices. For example, one can use M_n to test the null hypothesis $H_0 : \Sigma_n = \Sigma^{(0)}$, where $\Sigma^{(0)}$ is a pre-specified matrix. Here the null hypothesis can be that the population distribution is a stationary process so that Σ_n is Toeplitz, or that Σ_n has a banded structure.

It is very challenging to derive an asymptotic theory for M_n if we allow dependence among $X_{11}, \ldots, X_{1m}$. Many of the earlier results assume that the entries of the data matrix $\boldsymbol{X}_n$ are i.i.d.. In this case $\sigma_{ij} = 0$ if $i \ne j$. The quantity

$$L_n = \max_{1 \le i < j \le m} |\hat{\sigma}_{ij}|$$

is referred to as the *mutual coherence* of the matrix $\mathbf{X}_n$, and is related to compressed sensing (see for example [9]). Jiang [13] derived the asymptotic distribution of L_n.

Theorem 1 (*[13]*). *Suppose $X_{i,j}, i, j = 1, 2, \ldots$ are independent and identically distributed as ξ which has variance one. Suppose $\mathbb{E}|\xi|^{30+\epsilon} < \infty$ for some $\epsilon > 0$. If $n/m \to c \in (0, \infty)$, then*