Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 124 (2014) 18–50

stochastic processes and their applications

www.elsevier.com/locate/spa

Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails

Richard A. Davis[a], Oliver Pfaffel[b,*], Robert Stelzer[c]

[a] *Department of Statistics, Columbia University, NY, USA*
[b] *TUM Institute for Advanced Study & Department of Mathematics, Technische Universität München, Germany*
[c] *Institute of Mathematical Finance, Ulm University, Germany*

Received 29 October 2012; received in revised form 1 May 2013; accepted 13 July 2013
Available online 20 July 2013

Abstract

We study the joint limit distribution of the k largest eigenvalues of a $p \times p$ sample covariance matrix XX^{T} based on a large $p \times n$ matrix X. The rows of X are given by independent copies of a linear process, $X_{it} = \sum_j c_j Z_{i,t-j}$, with regularly varying noise (Z_{it}) with tail index $\alpha \in (0, 4)$. It is shown that a point process based on the eigenvalues of XX^{T} converges, as $n \to \infty$ and $p \to \infty$ at a suitable rate, in distribution to a Poisson point process with an intensity measure depending on α and $\sum c_j^2$. This result is extended to random coefficient models where the coefficients of the linear processes (X_{it}) are given by $c_j(\theta_i)$, for some ergodic sequence (θ_i), and thus vary in each row of X. As a by-product of our techniques we obtain a proof of the corresponding result for matrices with iid entries in cases where p/n goes to zero or infinity and $\alpha \in (0, 2)$.

MSC: primary 60B20; 62G32; secondary 60G55; 62H25

Keywords: Random matrix theory; Heavy-tailed distribution; Random matrix with dependent entries; Largest singular value; Sample covariance matrix; Largest eigenvalue; Linear process; Random coefficient model

1. Introduction

Recently there has been increasing interest in studying *large dimensional data sets* that arise in finance, wireless communications, genetics and other fields. Patterns in these data can often be summarized by the *sample covariance matrix*, as done in multivariate regression and dimension

* Corresponding author. Tel.: +49 15258932561.
E-mail addresses: rdavis@stat.columbia.edu (R.A. Davis), o.pfaffel@gmx.de (O. Pfaffel), robert.stelzer@uni-ulm.de (R. Stelzer).

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.07.005

reduction via factor analysis. Therefore, our objective is to study the asymptotic behavior of the eigenvalues $\lambda_{(1)} \geq \cdots \geq \lambda_{(p)}$ of a $p \times p$ sample covariance matrix $XX^{\top}$, where the *data matrix* X is obtained from n observations of a high-dimensional stochastic process with values in $\mathbb{R}^p$. Classical results in this direction often assume that the entries of X are independent and identically distributed (iid) or satisfy some moment conditions. For example, the Four Moment Theorem of Tao and Vu [39] shows that the asymptotic behavior of the eigenvalues of $XX^{\top}$ is determined by the first four moments of the distribution of the iid matrix entries of X. Our goal is to weaken the moment conditions by allowing for heavy-tails, and the assumption of independent entries by allowing for dependence within the rows and columns. Potential applications arise in portfolio management in finance, where observations typically have heavy-tails and dependence.

Assuming that the data comes from a multivariate normal distribution, one is able to compute the joint distribution of the eigenvalues $(\lambda_{(1)}, \ldots, \lambda_{(p)})$, see [26]. Under the additional assumption that the dimension p is fixed while the sample size n goes to infinity, Anderson [1] obtains a central limit like theorem for the largest eigenvalue. Clearly, it is not possible to derive the joint distribution in a general setting where the distribution of X is not invariant with respect to orthogonal transformations. Furthermore, since in modern applications with large dimensional data sets, p might be of similar or even larger order than n, it might be more suitable to assume that both p and n go to infinity, so Anderson's result may not be a good approximation in this setting. For example, considering a financial index like the S&P 500, the number of stocks is $p = 500$, whereas, if daily returns of the past 5 years are given, n is only around 1300. In genetic studies, the number of investigated genes p might easily exceed the number of participating individuals n by several orders of magnitude. In this *large n*, *large p* framework results differ dramatically from the corresponding *fixed p*, *large n* results—with major consequences for the statistical analysis of large data sets [27].

Spectral properties of large dimensional random matrices is one of many topics that has become known under the banner *Random Matrix Theory (RMT)*. The original motivation for RMT comes from mathematical physics [20,42], where large random matrices serve as a finite-dimensional approximation of infinite-dimensional operators. Its importance for statistics comes from the fact that RMT may be used to correct traditional tests or estimators which fail in the 'large n, large p' setting. For example, Bai et al. [4] give corrections on some likelihood ratio tests that fail even for moderate p (around 20), and El Karoui [21] consistently estimates the spectrum of a large dimensional covariance matrix using RMT. Thus statistical considerations will be our motivation for a random matrix model with heavy-tailed and dependent entries.

Before describing our results, we will give a brief overview of some of the key results from RMT for real-valued sample covariance matrices $XX^{\top}$. A more detailed account on RMT can be found, for instance, in the textbooks [2,5], or [31]. Here X is a real $p \times n$ random matrix, and p and n go to infinity simultaneously. Let us first assume that the entries of X are iid with variance 1. Results on the *global behavior of the eigenvalues* of $XX^{\top}$ mostly concern the *spectral distribution*, that is the random probability measure of its eigenvalues $p^{-1}\sum_{i=1}^{p} \epsilon_{n^{-1}\lambda_{(i)}}$, where ϵ denotes the Dirac measure. The spectral distribution converges, as $n, p \to \infty$ with $p/n \to \gamma \in (0, 1]$, to a deterministic measure with density function

$$\frac{1}{2\pi x\gamma}\sqrt{(x_+ - x)(x - x_-)}\,\mathbf{1}_{(x_-,x_+)}(x), \qquad x_\pm := (1 \pm \sqrt{\gamma})^2,$$

where $\mathbf{1}$ denotes the indicator function. This is the so called *Marčenko–Pastur law* [30,41]. One obtains a different result if $XX^{\top}$ is perturbed via an affine transformation [30,33]. Partially based on these results, [6,7,35,43] treat the case where the rows of X are given by independent copies