Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 2734–2760

stochastic processes and their applications

www.elsevier.com/locate/spa

On the empirical spectral distribution for matrices with long memory and independent rows

F. Merlevède[a,*], M. Peligrad[b]

[a] Université Paris-Est, LAMA (UMR 8050), UPEM, CNRS, UPEC, France

[b] Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, OH 45221-0025, USA

Received 29 June 2015; received in revised form 25 February 2016; accepted 26 February 2016
Available online 7 March 2016

Abstract

In this paper we show that the empirical eigenvalue distribution of any sample covariance matrix generated by independent samples of a stationary regular sequence has a limiting distribution depending only on the spectral density of the sequence. We characterize this limit in terms of Stieltjes transform via a certain simple equation. No rate of convergence to zero of the covariances is imposed, so, the underlying process can exhibit long memory. If the stationary sequence has trivial left sigma field the result holds without any other additional assumptions. This is always true if the entries are functions of i.i.d.

As a method of proof, we study the empirical eigenvalue distribution for a symmetric matrix with independent rows below the diagonal; the entries satisfy a Lindeberg-type condition along with mixingale-type conditions without rates. In this nonstationary setting we point out a property of universality, meaning that, for large matrix size, the empirical eigenvalue distribution depends only on the covariance structure of the sequence and is independent on the distribution leading to it. These results have interest in themselves, allowing to study symmetric random matrices generated by random processes with both short and long memory.

MSC: 60F05; 60F15; 60G42; 60G60

Keywords: Random matrices; Stieltjes transform; Martingale approximation; Lindeberg method; Empirical eigenvalue distribution; Spectral density; Sample covariance matrix

* Corresponding author.
E-mail addresses: florence.merlevede@u-pem.fr (F. Merlevède), peligrm@ucmail.uc.edu (M. Peligrad).

http://dx.doi.org/10.1016/j.spa.2016.02.016

1. Introduction and the main result

Due to the fact that random matrices appear in many applied fields, their empirical spectral distribution is a subject of intense research. Earlier works, pioneered by the celebrated paper by Wigner [44], deal with symmetric matrices having independent entries below the diagonal. Only in the last three decades there has been an intense effort to weaken the hypotheses of independence and various forms of weak dependence have been considered. The progress was in general achieved first for Gaussian random matrices. For this case the joint distribution of eigenvalues is tractable. Among the papers for symmetric Gaussian matrices with correlated entries we mention the works of Khorunzhy and Pastur [27], Boutet de Monvel et al. [12], Boutet de Monvel and Khorunzhy [11], Chakrabarty et al. [16].

Our paper is essentially motivated by the study of large sample covariance matrices, which is a very important topic in multivariate analysis and signal processing.

The spectral analysis of large-dimensional sample covariance matrices has been actively studied starting with the seminal work of Marčenko and Pastur [30] who considered independent random samples from an independent multidimensional vector. Later, also for the independent case, Wachter [43] established the almost sure results and recently Jin et al. [26] generalized it to auto-cross covariance matrix. A big step forward was the study of the dependent case represented in numerous papers. Basically, the entries of the matrix were allowed to be linear combinations of an independent sequence. The first paper where such a model was considered is by Yin and Krishnaiah [48] followed by important contributions by Yin [47], Silverstein [38], Silverstein and Bai [39], Hachem et al. [25], Pfaffel and Schlemm [33], Yao [46], Davis et al. [19], Pan et al. [32], Liu et al. [28], Bhattacharjee and Bose [8] among others.

A departure from linear models was considered by Bai and Zhou [3] who derived the limiting spectral distribution of large sample covariance matrices provided that the true covariance matrix has bounded spectral norm and the entries satisfy a dependence type condition. This dependence condition, sometimes called "good vector condition" is satisfied for Gaussian vectors or for isotropic vectors with log-concave distribution as shown in [31]. Note that the circular law for random matrices with independent isotropic unconditional log-concave rows has been proved by Adamczak [1]. As applications of their main result, Bai and Zhou [3] exhibited the limiting spectral distributions of Spearman's rank correlation matrices, sample correlation matrices and sample covariance matrices from finite populations. When applied to linear models the conditions imposed in the paper by Bai and Zhou can be verified when the innovations are square integrable and the coefficients are absolutely summable as shown in [46,32]. It should be mentioned that the bounded spectral norm condition imposed to the true covariance matrix does not allow to derive the limiting spectral distribution of large sample covariance matrices associated with linear processes exhibiting long range dependence.

Recently, Banna and Merlevède [4] considered samples from a stationary process whose variables are functions of i.i.d. and proved, under a dependence condition implying the absolute summability of the covariances, that the asymptotic behavior of the empirical eigenvalue distribution can be obtained by analyzing a Gaussian matrix with the same covariance structure. In [5], this result has been improved and extended to large covariance matrices associated with square integrable variables that are functions of an i.i.d. random field. In this latter paper, it is proved that no extra assumptions are needed to reduce the study of empirical spectral distribution to the one of a Gaussian matrix with the same covariance structure.

Even if many models encountered in time series analysis can be rewritten as functions of an i.i.d. sequence, this assumption is not completely satisfactory since many stationary processes