Limiting Spectral Distributions of Large Dimensional Random Matrices

Arup Bose[*]
Indian Statistical Institute, Kolkata
Sourav Chatterjee[†]
Stanford University, California
Sreela Gangopadhyay[‡]
Indian Statistical Institute, Kolkata

Abstract

Models where the number of parameters increases with the sample size, are becoming increasingly important in statistics. This necessitates a close look at the statistical properties of eigenvalues of random matrices whose dimension increases indefinitely.

There are several properties of the eigenvalues that one would be interested in and the literature in this area is already huge. In this article we focus on one important aspect: the existence and identification of the limiting spectral distribution (LSD) of the empirical distribution of the eigenvalues.

We describe some of the general tools used in establishing the LSD and how they have been applied successfully to establish results on the LSD for certain types of matrices. Some of the matrices for which the LSD has been established and the nature of the limit laws known are described in detail.

We also discuss a few open problems and partial solutions for some of these. We introduce a few new ideas which seem to hold some promise in this area. We also establish an invariance result for random Toeplitz matrix.

Keywords: Large dimensional random matrix, eigenvalues, limiting spectral distribution, Marčenko-Pastur law, semicircular law, circular law, Wigner matrix, sample variance covariance matrix, F matrix, Toeplitz matrix, moment method, Stieltjes transform, random probability, Brownian motion, normal approximation.

AMS 2000 Subject Classification: 60E07, 60E10, 62E15, 62E20.

[*]Theoretical Statistics and Mathematics Unit, I.S.I., 203 B.T. Road, Kolkata 700108, India
Email: `abose@isical.ac.in`

[†]Department of Statistics, Stanford University, CA 94305, USA
Email: `souravc@stat.stanford.edu`

[‡]Theoretical Statistics and Mathematics Unit, I.S.I., 203 B.T. Road, Kolkata 700108, India
Email: `res9616@isical.ac.in`

1 Introduction

Random matrices occur commonly in statistics; a familiar example is the sample variance covariance matrix. With increasing examples in statistics of models where the number of parameters increases with the sample size and with demonstrated inadequacy of standard statistical procedures in such models (see for example Bai and Saranadasa (1996)), the study of such matrices have become important to statisticians. Physicists have also been interested in certain types of random matrices as their dimension increases to infinity. We call these *large dimensional random matrices* (LDRMs) and focus on the properties of the eigenvalues of such matrices.

Usually, explicit evaluation of the eigenvalues or their distribution is not possible. But several other interesting questions may be asked. What is the probabilistic behaviour of the maximum or the minimum eigenvalue? Can we say more on the spacings between the different eigenvalues? If we consider the empirical distribution of the eigenvalues, what can we say about its limit, the *limiting spectral distribution* (LSD)? If the LSD exists, can we establish the rate at which the convergence takes place?

Some of these questions have been addressed for different types of matrices of interest both in the statistics and the physics community. A nice recent review article by Bai (1999) discusses some of the history, techniques and results in the area of LDRMs. Additional insight in the general area may be gained from the review works of Hwang (1986), and the books by Mehta (1991) and Girko (1988, 1995). Random matrices have drawn the attention of mathematicians for various reasons (in connection to the Riemann hypothesis for example). The books by Deift (1999) and Katz and Sarnak (1999) deal with the mathematical aspects of random matrices.

The literature on LDRMs is huge. In this article we focus on one specific but very interesting aspect, namely that of existence and identification of the LSD. We describe some of the general tools developed in establishing the LSD and how they have been applied successfully in some cases. Some of these results are described in detail. These include, in particular, the *Wigner matrix* which is important in physics and the *sample variance covariance matrix* which is important in statistics.

There are several interesting matrices for which the current techniques seem inadequate or difficult to apply. We introduce a few new ideas which seem to hold some promise and provide some new results based on these. We hope that this brief review will encourage others to work and contribute in this area.

In Section 2, we give the basic definitions we need and also a list of the common matrices (Wigner, sample covariance, F and others) that have been dealt with in the literature.

Section 3 discusses the two main methods (the moment method and the method of Stieltjes transform) that are used to establish the LSD.

In Section 4 we give most of the known results on the LSD along with brief discussions. We also establish a couple of new results. We also discuss a couple of open problems that have been of interest and put forth some ideas.

2 Preliminaries

The purpose of this section is to provide the basic definitions and to list the matrices that have received attention in the literature.

2.1 Basic definitions

Unless otherwise stated, the entries of all matrices are complex in general. I shall always denote an identity matrix whose order will be clear from the context.

Definition 1 *(Empirical Spectral Distribution (ESD)) For any square matrix A, the probability distribution P which puts equal mass on each eigenvalue of A is called the* Empirical Spectral Distribution or measure (ESD) *of A.*

Thus, if λ is an eigenvalue (characteristic root) of an $n \times n$ matrix A_n of multiplicity m, then the ESD puts mass m/n at λ. Note that if the entries of A are random, then P is a *random probability*. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all the eigenvalues, then the *empirical spectral distribution function (ESDF)* of A_n is given by

$$F_n(x,\ y) = n^{-1} \sum_{i=1}^{n} I\{\text{Re}\lambda_i \le x,\quad \text{Im}\lambda_i \le y\}.$$

The *expected spectral distribution function* of A_n is defined as $E(F_n(\cdot))$. This expectation always exists and is a distribution function. The corresponding probablity distribution is often known as the *expected spectral measure*. Note that typically, the order of A_n tends to infinity as $n \longrightarrow \infty$.

Definition 2 *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of square matrices with the corresponding ESD $\{P_n\}_{n=1}^{\infty}$. The Limiting Spectral Distribution (or measure) (LSD) of the sequence is defined as the weak limit of the sequence $\{P_n\}$, if it exists. If $\{A_n\}$ are random, the limit is understood to be in some probabilistic sense, such as "almost surely" or "in probability".*

Some of the main problems that the theory of LDRM seeks to address are:

1. Whether the LSD exists for certain classes of LDRM.

2. Whether the expected spectral measures converge.

3. If the LSD exists, establish the rates of convergence.

4. The behaviour of the extreme eigenvalues (when the matrices are Hermitian).

5. Studying other properties of the ensemble of eigenvalues.

As we have already said, our focus in this article would be the first issue. Often the second problem is easier to settle than the first and is used as an intermediate result to address the first issue. We shall not discuss these points here. The literature on the last two issues is also very rich. In particular, there are some very elegant probabilistic results known for the limiting behaviour of the maximum eigenvalue and for the separation of eigenvalues. There are also innumerable unanswered questions in this area. For more information, we point the reader towards Bai (1999), Bai and Yin (1988), Bai, Yin and Krishnaiah (1986, 1987) and to the published and unpublished works of Jack Silverstein (see *http://www4.ncsu.edu:8030/~jack/*).

2.2 Some LDRMs of interest

We describe some random matrices encountered frequently in the literature on LDRMs. For a complex random variable X, its variance is defined to be $E|X - E(X)|^2$.

Wigner Matrix: A *Wigner matrix* (Wigner (1955, 1958)) of order n and scale parameter σ is a Hermitian matrix of order n, whose entries above the diagonal are independent complex random variables with zero mean and variance σ^2, and whose diagonal elements are i.i.d. real random variables. This matrix is of considerable interest to physicists.

Sample Covariance Type Matrices: Suppose $\{x_{jk}, j, k = 1, 2, \dots\}$ is a double array of i.i.d. complex random variables with mean zero and variance 1. Write $\mathbf{x_k} = (x_{1k}, \dots, x_{pk})'$ and let $X_n = [\mathbf{x_1} \ \mathbf{x_2} \ \cdots \ \mathbf{x_n}]$. In LDRM literature, the matrix

$$S_n = n^{-1} X_n X_n^*$$

is called a *sample covariance matrix* (in short an S matrix). As a concrete example, if $\{x_{ij}\}$ are real normal random variables with mean zero and variance one, then S_n is a Wishart matrix. Note that we do not centre the matrices at the sample means as is conventional in defining the sample covariance matrix in the statistics literature. This however, does not affect the LSD.

Now let $T_n^{1/2}$ be any $p \times p$ Hermitian matrix, independent of X_n. Define

$$B_n = n^{-1} T_n^{1/2} X_n X_n^* T_n^{1/2}.$$

The matrices B_n are called *sample covariance type* matrices. It may be noted that this includes all Wishart matrices. Also observe that the eigenvalues of B_n are the same as those of $n^{-1} X_n X_n^* T_n = S_n T_n$. One example of the latter product form is the *multivariate F matrix* $F_n = S_{1n}S_{2n}^{-1}$ where $S_{in},\ i = 1,\ 2$ are independent Wishart matrices. This has motivated the study of LSD of matrices of the form B_n and S_nT_n.

Toeplitz Matrix: Let $\{x_0, x_1, \dots\}$ be a sequence of i.i.d. real random variables with mean zero and variance σ^2. The $n \times n$ matrix T_n whose $(i,j)th$ entry is $x_{|i-j|}$ is a random *Toeplitz matrix*. Non random Toeplitz matrices have been around in mathematics for a long time and their properties are very well understood. See for example the classic book by Grenander and Szego (1984). Recent information on this matrix may be found in Böttcher and Silbermann (1990, 1999). See also Gray (2002).

Hankel Matrix: Let $\{x_0, x_1, \dots\}$ be a sequence of i.i.d. real random variables with mean zero and variance σ^2. The $(i,j)th$ entry of the $n \times n$ random Hankel matrix is x_{i+j-1}. They are very closely related to the Toeplitz matrices. See the references cited above for the Toeplitz matrices.

Derivative of a Transition Matrix in a Markov Process: Consider a Markov process with n states and transition probabilities $p_{ij}(t)$. Let x_{ij} denote the derivative (with respect to t) of p_{ij} at $t = 0$. Then the $n \times n$ matrix (x_{ij}) is known as the *transition density matrix*. The entries of this matrix satisfy the conditions $\sum_{j=1}^{n} x_{ij} = 0$ for every i, $1 \le i \le n$. Motivated by this, consider the matrix

$$M_n = \begin{bmatrix} -\sum_{i=2}^{n} x_{1i} & x_{12} & x_{13} & \dots & x_{1(n-1)} & x_{1n} \\ x_{21} & -\sum_{i=1\neq 2}^{n} x_{2i} & x_{23} & \dots & x_{2(n-1)} & x_{2n} \\ & & & \vdots & & \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{n(n-1)} & -\sum_{i=1}^{n-1} x_{ni} \end{bmatrix} \qquad (1)$$

where $x_{jk} = x_{kj}$ $j < k$ are iid real random variables. We will refer to it as the *Markov matrix* in the sequel for convenience.

I.I.D entries: The matrix with i.i.d. entries (real or complex) has also received considerable attention in the literature and has given rise to the so called circular law conjecture.

3 Two Methods

We now describe in some detail the two most powerful tools which have been used quite often in establishing LSDs. One is the *moment method* and the other is the *method of Stieltjes Transforms*.

3.1 The Moment Method

Suppose $\{Y_n\}$ is a sequence of real valued random variables. Suppose that there exists some (nonrandom) sequence β_h such that $E(Y_n^h) \to \beta_h$ for every positive integer h where $\{\beta_h\}$ satisfies *Carleman's condition:*

$$\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty. \tag{2}$$

It is well-known that then there exists a distribution function F, such that for all h,

$$\beta_h = \int x^h dF(x) \text{ and } Y_n \text{ converges to } F \text{ in distribution.} \tag{3}$$

For a positive integer h, the h-th moment of the ESD of an $n \times n$ matrix A, with characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ has the following nice form:

$$h\text{-th moment of the ESD of A } = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^h = \frac{1}{n}\operatorname{tr}(A^h) = \beta_h(A) \text{ (say)} \tag{4}$$

Now, suppose $\{A_n\}$ is a sequence of random matrices such that

$$\beta_h(A_n) \longrightarrow \beta_h. \tag{5}$$

Here the convergence takes place either "in probability" or "almost surely" and $\{\beta_h\}$ are nonrandom. Now, if $\{\beta_h\}$ satisfies Carleman's condition then we can say that the LSD of the sequence $\{A_n\}$ is F (in the corresponding "in probability" or "almost sure" sense). We are tacitly assuming that the LSD has all moments finite.

Note that the computation of $\beta_h(A_n)$ involves computing the trace of A_n^h or at least its leading term. This ultimately reduces to counting the number of contributing terms in the following expansion, (a_{ij} denotes the $(i,j)th$ entry of A):

$$\operatorname{tr}(A^h) = \sum_{1 \le i_1, i_2, \ldots, i_h \le n} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{h-1} i_h} a_{i_h i_1} \tag{6}$$

The method, though straightforward, is not practically manageable in a wide variety of cases. The combinatorial arguments involved in the counting become quite unwieldy and even practically impossible as h and n increase. In cases where this method has been successful, the combinatorial arguments are very intricate.

The relation (5) can often be verified by showing that $E(\beta_h(A_n)) \longrightarrow \beta_h$ and $V(\beta_h(A_n)) \longrightarrow 0$. But even if all moments of the LSD exists, there is no guarantee that $E(\beta_h(A_n))$ are finite. Thus, to implement this method, the elements of A_n are first appropriately truncated. Of course then one has to verify that the effect of truncation is negligible.

This method has been successfully applied for the Wigner matrix, the sample covariance matrix and the F matrices and recently for Toeplitz, Hankel and Markov matrices. See Bai (1999) for some of the arguments in connection to Wigner, sample covariance and F matrices. For the arguments concerning Toeplitz, Hankel and Markov matrices see Bryc, Dembo and Jiang (2003).

3.2 Stieltjes Transform Method

Stieltjes transforms play an important role in deriving LSDs. They have also been useful in studying rates of convergence but we shall not discuss the latter here.

Definition 3 *For any function G of bounded variation on the real line, its Stieltjes Transform m_G is defined on $\{z : z = u + iv, v \neq 0\}$ as*

$$m_G(z) = \int_{-\infty}^{\infty} \frac{1}{x - z} G(dx). \tag{7}$$

We shall be concerned with cases where G is the cumulative distribution function of some probability distribution on the real line. *If a sequence of Stieltjes transforms converges, the corresponding distributional convergence holds.*

If A has real eigenvalues $\lambda_i, \quad 1 \leq i \leq n$, then the Stieltjes transform of the ESD of A is

$$m_A(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z} = \frac{1}{n} \mathrm{tr}[(A - zI)^{-1}]. \tag{8}$$

Let $\{A_n\}$ be a sequence of random matrices *with real eigenvalues* and let the corresponding sequence of Stieltjes transforms be $\{m_n\}$. If $m_n \longrightarrow m$ in some suitable manner, where m is a Stieltjes transform, then the LSD of the sequence $\{A_n\}$ is the unique probability on the real line whose Stieltjes transform is the function m. The convergence of the sequence $\{m_n\}$ is often verified by first showing that it satisfies some (approximate) recursion equation. Solving the limiting form of this equation identifies the Stieltjes transform of the LSD.

This method has been successfully applied for the Wigner matrix and the sample covariance type matrices. See Bai (1999) for more details on the use of this transform to derive the convergence of the ESD and on the rate at which the convergence takes place.

4 Limiting Spectral Distributions

4.1 Wigner matrix and the Semi-Circular Law

If the entries of the Wigner matrix are real normal with mean zero and, variances 1 and 1/2 respectively for the entries on and above the diagonal, then the joint distribution of its eigenvalues can be calculated explicitly. If $\lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues, then it is not difficult to prove, (see Mehta (1991)), that the joint density is:

$$f(\lambda_1, \dots \lambda_n) = \frac{\exp(-\sum_{i=1}^{n} \lambda_i^2/2)}{2^{n/2} \prod_{i=1}^{n} \Gamma(\frac{p+1-i}{2})} \prod_{i<j}^{n} (\lambda_i - \lambda_j).$$

However, the distribution of the eigenvalues cannot be found in a closed form if we drop the normality assumption. Nevertheless, even if the normality assumption does not hold and we assume the entries to be real, quick calculations of the first two moments will convince the reader that the correct scaling for convergence is $n^{-1/2}$. That is, one should look at $n^{-1/2}W_n$.

The *semi-circular law* S with scale parameter σ arises as the LSD spectral distribution of $n^{-1/2}W_n$. It has the density function

$$p_\sigma(s) = \begin{cases} \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - s^2} & \text{if } |s| \leq 2\sigma, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This is also known as the *quarter circle law* (see Girko and Repin (1995)). All its odd moments are zero. The even moments are given by

$$\int s^{2k} p_\sigma(s) ds = \frac{(2k)!\sigma^{2k}}{k!(k+1)!}.$$

Wigner (1955) assumed the entries to be i.i.d. real Gaussian and established the convergence of E(ESD) of $n^{-1/2}W_n$ to the semi-circular law (9). Assuming the existence of finite moments of all orders, Grenander (1963, pages 179 and 209) established the convergence of the ESD in probability. Arnold (1967) obtained almost sure convergence under the finiteness of the fourth moment of the entries.

All the earlier proofs use the tedious Moment Method. The Stieltjes transform method can also be used. To give the reader some idea, in Bai (1999) it is shown that using the relation (8) and an appropriate partitioning of the matrix , the Stieltjes transform m_n of $n^{-1/2}W_n$ satisfies the following approximate recursion equation:

$$m_n(z) = -\frac{1}{z + \sigma^2 m_n(z)} + \delta_n, \quad \delta_n \to 0. \tag{10}$$

Using this, the Stieltjes transform of the LSD satsifies $m(z) = -\frac{1}{z+\sigma^2 m(z)}$. This equation has two solutions for each z. From some other considerations, it is shown that the correct solution is

$$m(z) = -\frac{1}{2\sigma^2}[z - \sqrt{z^2 - 4\sigma^2}] \tag{11}$$

which is indeed the Stieltje's transform of the semicircular law (9). We state the result as a Theorem.

Theorem 1 *If $\{W_n\}$ is a sequence of Wigner matrices of order $n \times n$ with scale parameter σ then with probability 1, the ESDF of $n^{-\frac{1}{2}}W_n$ tends to the semicircular law S given in (9) with scale parameter σ.*

Incidentally, Bai (1999) generalises the result of Arnold (1967) by considering Wigner matrices whose entries above the diagonal are not necessarily identically distributed and have no moment restrictions except that they have finite variance.

A related result of Trotter (1984) is worth mentioning, specially because of the distinctly different method employed. He too considered matrices whose entries are real independent Gaussian variables. To quote his result, define the distance d between any two probability measures μ and ν as

$$d(\mu, \nu)^2 = \inf E(X - Y)^2,$$

the infimum being taken over all pairs of random variables X, Y defined on the same probability space having distribution μ and ν respectively. It may be mentioned that this metric was introduced by Mallows (1972) and its properties have been studied in Bickel and Freedman (1981). Trotter's main theorem is:

Theorem 2 *Let $A_n = (2n)^{\frac{1}{2}}(M_n + M_n')$ where M_n is an $n \times n$ standard Gaussian random matrix. Then $\lim_n E\{d(F_{A_n}, S)^2\} = 0$.*

Boutet de Monvel, Khorunzhy and Vasilchuk (1996) obtained some other generalizations of Wigner's results with weakly dependent Gaussian sequences as entries.

4.2 Sample Covariance type matrices and the Marčenko-Pastur Law

Suppose X_n is a $p \times n$ matrix whose entries are i.i.d. complex random variables with mean 0 and variance σ^2. The sample covariance matrix is defined as $S_n = \frac{1}{n}X_nX_n^*$. Note that we have not subtracted the sample mean while defining the sample covariance matrix. This does not affect the treatment of the LSD. Apart from being important to statisticians for a variety of reasons (see below), the spectral theory of large dimensional sample covariance matrices also finds wide application in signal detection and array processing. See Silverstein and Combettes (1992a) and (1992b).

If the entries are i.i.d. normal with mean zero then S_n is a Wishart matrix with population covariance matrix I. In that case much is known about the distribution of eigenvalues of S_n and related matrices. See Anderson (1984).
As a generalization of S_n, the *covariance type matrix* B_n is any matrix of the form $n^{-1}T_n^{1/2}X_nX_n^*T_n^{1/2}$ where T_n is a $p \times p$ Hermitian matrix independent of X_n. Suppose that both n and p are large. There is at least one important reason why one would be interested in such mtarices. Usually p represents the number of explanatory variables and in traditional statistical models, this is held fixed. However, when the number of explanatory variables is very large compared to n, it is natural to

formulate it as a case where both n and p are tending to infinity. See Johnstone (2001) for examples where both n and p are very large and are of the same order. See also Donoho (unpublished work) for more examples.

Note that the eigenvalues of B_n are the same as those of $n^{-1}S_nT_n$. Hence, one would naturally be led to the study of the ESD and LSD of *product* matrices. Such product matrices would cover the sample covariance matrices when the population covariance matrix is not a multiple of the identity matrix. Moreover, the *multivariate F* statistics $F = S_1S_2^{-1}$, where S_1 and S_2 are sample covariance matrices, is also of the product form.

It may be noted that many invariant tests are functions of eigenvalues of matrices of the form F. For example, invariant tests of the general linear hypothesis depend on the sample only through the eigenvalues of product matrices $F = S_1S_2^{-1}$ where S_1, S_2 are Wishart. However the test statistics constructed by classical methods perform inadequately when the dimension of the data is of the same order as the sample size. For instance, consider a two sample problem where we wish to test the hypothesis $H : \mu_1 = \mu_2$ against $K : \mu_1 \neq \mu_2$ where μ_1 and μ_2 are the means of two multivariate populations of dimension p. The classical Hotelling test is not well-defined when the dimension p is large compared to the sample sizes n_1 and n_2. Bai and Saranadasa (1996) proposed an asymptotic normal test. Interestingly, they used the results on LSD of F matrices to show that when both p and n tend to infinity, their test is more powerful than Hotelling's test even when the latter is well-defined.

So as expected, the product matrices and specially the sample covariance type matrices are very well-studied. There is a host of results available for their LSD and asymptotic behaviour of their extreme eigenvalues. Since in this paper we concentrate on LSD only, we present some important results related to LSD.

The so called *Marčenko-Pastur law* with scale index σ^2 is indexed by $0 < y < \infty$. It has a positive mass at 0 if $y > 1$. Elsewhere it has a density:

$$p_y(x) = \begin{cases} \frac{1}{2\pi x y \sigma^2}\sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

and a point mass $1 - \frac{1}{y}$ at the origin if $y > 1$, where $a = a(y) = \sigma^2(1 - \sqrt{y})^2$ and $b = b(y) = \sigma^2(1 + \sqrt{y})^2$.

The LSD of S_n was first established by Marčenko and Pastur (1967). Subsequent work on S_n may be found in Grenander and Silverstein (1977), Wachter (1978), Jonsson (1982), Yin (1986), Yin and Krishnaiah (1985) and Bai and Yin (1988). We state below the result for i.i.d. entries.

Theorem 3 *Suppose that $\{x_{ij}\}$ are i.i.d. complex variables with variance σ^2.*

(i) If $p/n \longrightarrow y \in (0, \infty)$ then the ESD of S_n converges almost surely to the Marčenko-Pastur law (12) with scale index σ^2.

(ii) If $p/n \longrightarrow 0$ then the ESD of $W_n = \sqrt{\frac{n}{p}}(S_n - \sigma^2 I_p)$ converges almost surely to the semicircular law (9) with scale index σ^2.

There are several versions of this result under variations of i.i.d. condition of the entries of X_n. For example Bai (1999) proved the following:

Theorem 4 *Suppose that for each n, the entries of X_n are independent complex variables, with a common mean and variance σ^2. Assume that $p/n \to y \in (0, \infty)$ and that for any $\delta > 0$,*

$$\frac{1}{\delta^2 np} \sum_{j,k} E\left((x_{jk})^2 I_{(|x_{jk}^{(n)}| \geq \delta\sqrt{n})} \right) \to 0. \tag{13}$$

Then ESD of S_n tends almost surely to the Marčenko-Pastur law (12) with ratio index y and scale index σ^2.

When $p/n \to 0$ as $p \to \infty$ and $n \to \infty$ the theorem takes the following form:

Theorem 5 *Suppose that for each n, the entries of X_n are independent and complex random variables with a common mean and variance σ^2. Assume that for each constant $\delta > 0$, as $p \to \infty$ with $p/n \to 0$,*

$$\frac{1}{p\delta^2}\sqrt{np} \sum_{jk} \left((x_{jk}^{(n)})^2 I_{(|x_{jk}^{(n)}| \geq \delta(np)^{1/4})} \right) = o(1)$$

and

$$\frac{1}{np^2} \sum_{jk} \left((x_{jk}^{(n)})^4 I_{(|x_{jk}^{(n)}| \geq \delta(np)^{1/4})} \right) = o(1).$$

Then with probability 1 the ESD of $W_n = \sqrt{\frac{n}{p}}(S_n - \sigma^2 I_p)$ tends to the semicircular law (9) with scale index σ^2.

In case of multivariate $F = S_1 S_2^{-1}$, where S_1 and S_2 are independent Wishart, Wachter (1978, 1980) showed the existence of the LSD and its explicit form may be found in Silverstein (1985). See also Bai, Yin and Krishnaiah (1987), Yin, Bai and Krishnaiah (1983) and Wachter (1980) for related results.

Later it has been shown that the same LSD persists even if we do not start with the Wisharts but just assume that the orginal variables involved in the construction of

S_1 and S_2 are respectively i.i.d with enough moments. See Bai and Yin (1993), Yin (1986), Bai, Yin and Krishnaiah (1985), Yin and Krishnaiah (1983), Wachter (1980) for the details.

For results on general products of the form S_nT_n where S_n is a sample covariance matrix and T_n is independent of S_n, see Yin and Krishnaiah (1983), Bai, Yin and Krishnaiah (1986), Silverstein (1995), Silverstein and Bai (1995) and Yin (1986). We quote one such result from Yin (1986):

Theorem 6 *Suppose x_{ij}, the entries of X_n are iid with finite variance. Suppose T_n is a $p \times p$ non-negative definite random matrix independent of X_n and for each fixed h, $\frac{1}{p}tr(T_n^h) \to \alpha_h$ in probablity/almost surely, where the sequence $\{\alpha_h\}$ satisfies Carleman's condition. Now, if $p/n \to y \in (0, \infty)$, then the LSD of S_nT_n exists in probablity/almost surely.*

Bai (1999) relaxed some of the conditions of the above theorem. He considered X_n with entries as independent complex random variables satisfying (13) and T_n to be $p \times p$ random Hermitian matrices independent of X_n whose LSD exists in probability/almost surely. Then he showed that the LSD of S_nT_n exists in probability/almost surely.

Silverstein and Bai (1995) considered matrices of the type $B_n = A_n + n^{-1}X_n^*T_nX_n$ and proved a general existence theorem. The setup of this theorem reflects the situations encountered in multivariate statistics. Examples of B_n can be found in the analysis of multivariate linear models and error-in-variables models, where the sample covariance matrix of the covariates is ill-conditioned. The between-covariance matrix in MANOVA plays the role of A_n. In this setup A_n is introduced to reduce the instability in the direction of eigenvectors corresponding to small eigenvalues. Note that the result below is powerful enough to guarantee the existence of the LSD for S_n in certain heteroscedastic cases. However, the LSD can be computed only through its Stieltjes transform.

Below, $\to^v$ denotes vague convergence, that is convergence without preservation of the total variation.

Theorem 7 *Suppose that for each n, the entries of $X_n = (\mathbf{x_1}, \ldots, \mathbf{x_n})$, $p \times n$ are i.i.d. complex random variables with $E(|x_{11} - E(x_{11}^2)|) = 1$ and that $T_n = diag(\tau_1^n, \ldots \tau_p^n)$, τ_i^n are real, and the ESDF of T_n converges almost surely to a probability distribution function H as $n \to \infty$. Suppose that $B_n = A_n + \frac{1}{n}X_n^*T_nX_n$ where A_n is Hermitian satisfying $F_{A_n} \to^v F_a$ almost surely. Assume X_n, T_n and A_n are independent.*

If $\frac{p}{n} \to y > 0$, almost surely then the ESDF of B_n converges vaguely to a d.f. F with Stieltjes transform $m(z) = m_a(z - y\int \frac{\tau dH\tau}{1+\tau m(z)})$ where m_a is the Stieltjes transform of F_a and z is a complex variable with imaginary part > 0.

4.3 Toeplitz and related matrices

Consider the space l_2 of all sequences which are square summable. Suppose that $\{x_i\}$ belongs to l_2. Then the Toeplitz matrices with these entries are looked upon as transformations from l_2 to l_2. This is the basic starting point of the theory of *non-random* Toeplitz matrices. See for example Böttcher and Silbermann (1990, 1999) and the references there. Much is known about the existence of the ESD in such non random cases.

Random Toeplitz matrix plays a significant role in statistical analysis, particularly in time-series analysis. In time-series analysis the covariance matrix is a Toeplitz matrix. It comes up in several spectral estimation methods. For instance given a real signal with autocorrelation coefficient $r(i)$ the coefficients of the auto-regressive model can be estimated by solving the corresponding Yule-Walker equations:

$$A \times Y = C$$

where Y is an unknown vector, C is a vector of known coefficients and A is a Toeplitz matrix with elements $r(i)$.

However, no sequence of (non zero) i.i.d. $\{x_i\}$ can be in l_2 in any sense and it appears that the strong machinery available for nonrandom Toeplitz matrices is not of much use. At the time of submission of the first version of this article, the existence of LSD of the Toeplitz matrix remained an open problem. From a private communication from one of the authors, we learnt that Bryc, Dembo and Jiang (2003) have settled the question of the existence of the LSD of the random Toeplitz matrix. They used complicated combinatorial arguments to prove its existence, though its closed form is not yet known. They also established the existence of the LSD for the Hankel matrices and identified the LSD for Markov matrices (as the free convolution of the standard Gaussian and semicirle laws). Following is one of the main theorems proved by Bryc, Dembo and Jiang (2003).

Theorem 8 *Let the entries of the Toeplitz matrix T_n be i.i.d. real-valued random variables with mean zero and variance one. Then with probability one, the ESD of $\frac{1}{\sqrt{n}}T_n$ converges weakly as $n \to \infty$ to a non-random symmetric probability measure which does not depend on the distribution of the entries of T_n and has unbounded support.*

It is not hard to compute the first four moments of the ESD $n^{-1/2}T_n$. The limits of first four moments turn out to be 0, 1, 0 and 8/3. Recently Hammond and Miller (2003) obtained some useful results for the moments of ESD of a random Toeplitz matrix normalized by $\sqrt{n}$. They show that while the odd moments tend to 0, the even moments obey the following bounds:

$$\beta_{2k}(\frac{T_n}{\sqrt{n}}) \leq \frac{(2k)!}{2^k k!} + O_k(\frac{1}{n}).$$

Note that the first term above is the Gaussian $2k$-th moment.

We simulated 50 Gaussian Toeplitz matrices of order 200. In Figure 1, we overlap the kernel-smoothed density estimates for the ESDs of each of these 50 matrices. This graph gives an idea of the concentration of the random distribution of the eigenvalues. The expected distribution is estimated by simluating 500 Gaussian Toeplitz matrices of order 200. The kernel estimate based on all the $500 \times 200 = 10^5$ eigenvaluese is given in Figure 2. Inspite of the look, this curve is not Gaussian (recall that the limiting fourth moment is not 3).

4.3.1 An Invariance Result

In an attempt to explore the structure of the Toeplitz matrix we discovered an invariance result. To develop this invariance result, we shall look at matrices as operators on $C[0,1]$. Take any square matrix A of order n. Define the bounded linear operator $C_A : C[0,1] \longrightarrow C[0,1]$ by the following.

Take any $f \in C[0,1]$. Let $x = (f(\frac{0}{n}), f(\frac{1}{n}), \ldots, f(\frac{n-1}{n}))^T$. Let $y = Ax$. Define $C_A(f)$ to be the function on $[0,1]$ whose graph is the polygonal line joining $(0, y_0)$, $(\frac{1}{n}, y_1)$, $(\frac{2}{n}, y_2)$, $\ldots$, $(\frac{n-1}{n}, y_{n-1})$, $(1, y_{n-1})$.

Recall that in a Banach algebra $\mathcal{B}$ with identity e, the spectrum of an element $x \in \mathcal{B}$, denoted by $\sigma(x)$, is defined by

$$\sigma(x) = \{\lambda : \lambda e - x \text{ does not have an inverse in } \mathcal{B}\} \tag{14}$$

The following are easy to verify (with obvious meanings for the notations):

$$\left.\begin{array}{rcl} C_{A+B} & = & C_A + C_B \\ C_{\lambda A} & = & \lambda C_A \\ C_{AB} & = & C_A C_B \end{array}\right\} \tag{15}$$

However, it is not true that $C_I = \widetilde{1}$ where $\widetilde{1}$ is the identity operator on $C[0,1]$. Nevertheless, we have the following:

Theorem 9 *Under the notation introduced in this section,* $\sigma(A) \cup \{0\} = \sigma(C_A) \cup \{0\}$.

To prove this, take any λ, such that $\lambda \neq 0$ and $\lambda \notin \sigma(A)$. We shall show that $\lambda\widetilde{1} - C_A$ is invertible in $C[0,1]$ by verifying that $\lambda^{-1}(\widetilde{1} - C_{A(A-\lambda I)^{-1}}) = S$ (say) is the inverse of $\lambda\widetilde{1} - C_A$.

To show this, just use the properties listed in (15) as follows:

$$\begin{aligned} S(\lambda\widetilde{1} - C_A) &= \lambda^{-1}(\widetilde{1} - C_{A(A-\lambda I)^{-1}})(\lambda\widetilde{1} - C_A) \\ &= \lambda^{-1}(\lambda\widetilde{1} - C_A - \lambda C_{A(A-\lambda I)^{-1}} + C_{A(A-\lambda I)^{-1}} C_A) \\ &= \lambda^{-1}(\lambda\widetilde{1} - C_A - C_{A(A-\lambda I)^{-1}\lambda I} + C_{A(A-\lambda I)^{-1}A}) \\ &= \lambda^{-1}(\lambda\widetilde{1} - C_A + C_{A(A-\lambda I)^{-1}(A-\lambda I)}) \\ &= \widetilde{1} \end{aligned}$$

So S is a left-inverse of $(\lambda\widetilde{1} - C_A)$. Similarly, it can be checked that S is a right-inverse of $\lambda\widetilde{1} - C_A$. Thus, we have shown that $\sigma(C_A) \subseteq \sigma(A) \cup \{0\}$. Conversely, if $\lambda \in \sigma(A)$ then $\exists x \neq 0$ such that $Ax = \lambda x$. Let $\widetilde{x}$ be the element of $C[0,1]$ whose graph is the polygonal line joining $(0, x_0)$, $(\frac{1}{n}, x_1)$, $(\frac{2}{n}, x_2)$, $\dots$, $(\frac{n-1}{n}, x_{n-1})$, $(1, x_n)$. Then $\widetilde{x} \neq 0$ and $C_A(\widetilde{x}) = \lambda\widetilde{x}$ by linearity. This proves the reverse inclusion. □

Now consider the sequence of normalised Toeplitz matrices $n^{-1/2}T_n$ where T_n is the Toeplitz matrix of order n formed by the $\{x_i\}$. For simplicity, we will denote the corresponding sequence of operators $\{C_{n^{-1/2}T_n}\}$, as simply $\{C_n\}$.

Before we deal with random elements of $\mathcal{B}$ and convergence, we must clarify that the topology on $\mathcal{B}$ under consideration is the usual operator norm topology. Measurability is in terms of the Borel sigma algebra generated by this topology.

It can be easily checked that the map $A \mapsto C_A$ is continuous from $\mathbb{R}^{n\times n} \to \mathcal{B}$, and hence measurable. So C_n is a legitimate random variable on $\mathcal{B}$.

The asymptotic behaviour of the law of C_n is made somewhat explicit by the following Theorem:

Theorem 10 *(i) If $f_1, f_2, \dots, f_m$ are twice continuously differentiable functions, then $\{(C_n f_1, \dots, C_n f_m)\}_{n=1}^{\infty}$ converges in law, to a limiting distribution which is not dependent on the distribution of the $\{x_i\}$'s.*

(ii) If the sequence $\{C_n\}$ is tight, then it converges in law, and the limiting distribution does not depend on the distribution of the $\{x_i\}$'s.

Proof Assuming (i), proof of (ii) is easy. If P is a limit point of the sequence $\{P_n\}$, where P_n stands for the law of C_n, and $C \sim P$ then for any $f_1, f_2, \dots, f_m \in C^2[0,1]$, by (i), the law of $(Cf_1, \dots, Cf_m)$ is not dependent on the distribution of the $\{x_i\}$'s. Since the C^2 functions are norm dense in $C[0,1]$, the above assertion holds even when f_i's are just continuous. Then, by the usual technique, it can be proved that P is uniquely determined, irrespective of the distribution of the x_i's.

We now prove (i). For simplicity asume $m = 1$. Generalization to the case $m > 1$ is easy. So fix $f \in C^2[0,1]$. Let

$$S_n = \sum_{i=0}^{n-1} x_i \tag{16}$$

with $S_0 = 0$. Let B_n denote the function whose graph is the polygonal line connecting $(0, \frac{S_0}{\sqrt{n}})$, $(\frac{1}{n}, \frac{S_1}{\sqrt{n}})$, $(\frac{2}{n}, \frac{S_2}{\sqrt{n}})$, $\dots$, $(1, \frac{S_n}{\sqrt{n}})$. It is well known that $\{B_n\}$ converges in law to the Brownian Motion on $[0,1]$, which we shall generically denote by B. Let $g_n = C_n f$. Let $\mathcal{B}$ denotes the Banach algebra of all bounded linear operators from $C[0,1]$ into itself. We shall show that there exists an element $\varphi \in \mathcal{B}$, depending only on f, and not on the distribution of the x_i's, such that

$$\|g_n - \varphi(B_n)\| \xrightarrow{P} 0. \tag{17}$$

(We are using the sup norm.) Now since $B_n \Longrightarrow B$, therefore $\varphi(B_n) \Longrightarrow \varphi(B)$ and so, by (17), $g_n \Longrightarrow \varphi(B)$. This will establish the claim.

Given $u \in C[0,1]$, φ is defined by:

$$\varphi(u)(x) = f(0)u(x) + f(1)u(1-x) - \int_0^x f'(x-y)u(y)dy - \int_0^{1-x} f'(x+y)u(y)dy \tag{18}$$

It is easy to check that $\varphi \in \mathcal{B}$. Let us first compute $|g_n(x) - \varphi(B_n)(x)|$ at each x. We shall first consider $x = \frac{k}{n}$ for $k = 0, 1, \ldots, n-1$. Then,

$$g_n(\frac{k}{n}) = \sum_{j=1}^{k} \frac{x_j}{\sqrt{n}} f(\frac{k-j}{n}) + \sum_{j=0}^{n-k-1} \frac{x_j}{\sqrt{n}} f(\frac{k+j}{n}) \tag{19}$$

Now the first term in the above expansion can be written as

$$\begin{aligned}
& \sum_{j=1}^{k} \frac{x_j}{\sqrt{n}} [\sum_{i=1}^{j} \{f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})\}] + f(\frac{k}{n}) \sum_{j=1}^{k} \frac{x_j}{\sqrt{n}} \\
= \; & \sum_{i=1}^{k} [\{f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})\} \sum_{j=i}^{k} \frac{x_j}{\sqrt{n}}] + f(\frac{k}{n}) \sum_{j=1}^{k} \frac{x_j}{\sqrt{n}} \\
= \; & (f(0) - f(k/n)) \sum_{j=0}^{k} \frac{x_j}{\sqrt{n}} - \\
& \sum_{i=1}^{k} [\{f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})\} \sum_{j=0}^{i-1} \frac{x_j}{\sqrt{n}}] + f(\frac{k}{n}) \sum_{j=1}^{k} \frac{x_j}{\sqrt{n}} \\
= \; & \sum_{i=1}^{k} \{f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})\} \frac{S_i}{\sqrt{n}} + f(0) \frac{S_{k+1}}{\sqrt{n}} - f(\frac{k}{n}) \frac{x_0}{\sqrt{n}}
\end{aligned}$$

Similarly, the second term can be written as

$$\begin{aligned}
& \sum_{j=0}^{n-k-1} \frac{x_j}{\sqrt{n}} f(\frac{j+k}{n}) \\
= \; & \sum_{j=0}^{n-k-1} \frac{x_j}{\sqrt{n}} [\sum_{i=j}^{n-k-1} \{f(\frac{i+k}{n}) - f(\frac{i+k+1}{n})\}] + f(1) \sum_{j=0}^{n-k-1} \frac{x_j}{\sqrt{n}} \\
= \; & \sum_{i=0}^{n-k-1} \{f(\frac{i+k}{n}) - f(\frac{i+k+1}{n})\} \frac{S_i}{\sqrt{n}} + f(1) \frac{S_{n-k}}{\sqrt{n}}
\end{aligned}$$

Let $M = \sup_{0 \le x \le 1} f(x)$, $M_1 = \sup_{0 \le x \le 1} f'(x)$, $M_2 = \sup_{0 \le x \le 1} f''(x)$, $D_n = \sup_{0 \le x \le 1} |B_n(x)|$, and $\delta_n = n^{-1/2} \max_{0 \le i \le n-1} |x_i|$.

Fix k, $0 \le k \le n-1$. Define h on $[0, k/n)$ by

$$h(y) = \frac{S_i}{\sqrt{n}} n[f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})] \tag{20}$$

if $\frac{i-1}{n} \le y < \frac{i}{n}$ for some i, $1 \le i \le k$.
Take any $y \in [0, k/n)$. Suppose $\frac{i-1}{n} \le y < \frac{i}{n}$ where $1 \le i \le k$. Then, by the Mean Value Theorem,

$$h(y) = -\frac{S_i}{\sqrt{n}} f'(z) \;\text{ for some } z \in (\frac{k-i}{n}, \frac{k-i+1}{n}).$$

Let $y_1 = k/n - y$. Then $|z - y_1| < 1/n$

$$\begin{aligned}
& |h(y) + f'(k/n - y)B_n(y)| \\
= \;& |-\frac{S_i}{\sqrt{n}} f'(z) + \frac{S_i}{\sqrt{n}} f'(y_1) - \frac{S_i}{\sqrt{n}} f'(y_1) + B_n(y) f'(y_1)| \\
\le \;& |\frac{S_i}{\sqrt{n}}| \frac{1}{n} M_2 + |\frac{S_{i-1}}{\sqrt{n}} - \frac{S_i}{\sqrt{n}}||f'(y_1)| \\
\le \;& \frac{1}{n} D_n M_2 + M_1 \delta_n
\end{aligned}$$

Thus,

$$\begin{aligned}
& |\sum_{i=1}^{k} \frac{S_i}{\sqrt{n}} \{f(\frac{k-i}{n}) - f(\frac{k-i+1}{n})\} + \int_0^{k/n} f'(k/n - y) B_n(y) dy| \\
= \;& |\int_0^{k/n} h(y) dy + \int_0^{k/n} f'(k/n - y) B_n(y) dy| \\
\le \;& \int_0^{k/n} |h(y) + f'(k/n - y) B_n(y)| dy \\
\le \;& \frac{1}{n} D_n M_2 + M_1 \delta_n
\end{aligned} \tag{21}$$

It is well known that $D_n = O_P(1)$, and $\delta_n = o_P(1)$. Hence left side of (21) converges to 0 in probabilty. Now note that the right side is not dependent on k. Hence, the maximum of the left side over k, $0 \le k \le n-1$, also tends to zero in probability. Similar arguments on other expressions finally show that

$$\max_{0 \le k \le n} |g_n(k/n) - \varphi(B_n)(k/n)| \xrightarrow{P} 0. \tag{22}$$

Now, $\{\varphi(B_n)\}$ is a tight family. So, by standard arguments, $V_{1/n}(\varphi(B_n)) \xrightarrow{P} 0$ as $n \longrightarrow \infty$, where $V_\delta(u) := \sup_{|t-s| \le \delta} |u(t) - u(s)|$. Also, if $k/n \le x < (k+1)/n$, then $|g_n(x) - g_n(k/n)| \le |g_n((k+1)/n) - g_n(k/n)| \le 2\delta_n M + \delta_n M_1$, as can be easily checked from (19). Now using all the preceding observations, it is easy to show (17), completing the proof of the theorem. □

4.3.2 A close relative to Toeplitz matrix

In a Toeplitz matrix, each diagonal has equal entries. Consider a matrix where each *anti-diagonal* has equal entries in a symmetric fashion. Thus the $(i,j)th$ entry of such a matrix of order n equals x_{i+j-2}. Call this matrix A_n. As in the Toeplitz case, the natural normalisation turns out to be $n^{-1/2}$ and we consider $X_n = n^{-1/2}A_n$. With some work, its eigenvalues can be explicitly computed (below $\lfloor x \rfloor$ denotes the largest integer less than or equal to x) as:

$$\begin{cases} \lambda_{0,n} & = n^{-1/2}\sum_{t=0}^{n-1} x_t \\ \lambda_{n/2,n} & = n^{-1/2}\sum_{t=0}^{n-1}(-1)^t x_t, \text{ if } n \text{ is even} \\ \lambda_{k,n} = -\lambda_{n-k,n} & = \sqrt{a_{k,n}^2 + b_{k,n}^2}, \ 1 \le k \le \lfloor \frac{n-1}{2} \rfloor. \end{cases}$$

where,

$$a_{k,n} = \frac{1}{\sqrt{n}}\sum_{l=0}^{n-1} x_l \ \cos(2\pi lk/n) \ \text{ and, } \ b_{k,n} = \frac{1}{\sqrt{n}}\sum_{l=0}^{n-1} x_l \ \sin(2\pi lk/n).$$

Note that if the entries are i.i.d. $N(0,\ 1)$, then $\{a_{k,n}, b_{k,n}\}$ are i.i.d normal variables. Hence in this case the ESD is essentially a symmetrised version of the usual *empirical distribution* of the square root of a chi-squared distribution with two degrees of freedom. The latter has the density:

$$f(x) = |x| \exp(-x^2), \quad -\infty < x < \infty.$$

Since the empirical distribution converges to the true distribution almost surely, ESD converges to the above law in this special case. When the entries are not necessarily Gaussian, Bose and Mitra (2002) proved the following Theorem. The main idea in the proof is to use *normal approximation* for sums of independent variables to show that for each x, $E(F_{X_n}(x)) \to F(x)$ and $V(F_{X_n}(x)) \to 0$.

Theorem 11 *Let $\{x_i\}$ be i.i.d. with mean zero and variance 1 and $E|x_1|^3 < \infty$. Then at each argument, the ESD of $X_n = n^{-1/2}A_n$ converges in L_2 to the LSD with density f given above. Hence the ESD converges to this distribution in probability.*

We provide two new variations of the above theorem using different approaches. Suppose, $\{x_l\}$ is a *weakly* stationary sequence of random variables with mean zero and autocovariance function $c(u) = E(x_l x_{l+u})$. If $\sum_{u=0}^{\infty} |c(u)| < \infty$, then this autocovariance function has an associated density, commonly known as the *spectral density* (see Brockwell and Davis (1991) for example) which is given by

$$f(\lambda) = \frac{1}{2\pi}\sum_{u=0}^{\infty} c(u)\exp(-i\lambda u), \quad -\pi \le \lambda \le \pi.$$

The *periodogram* at frequency λ of $\{x_l\}$ is defined to be

$$I(\lambda) = \frac{1}{2\pi n} d_x^n(\lambda) d_x^n(-\lambda).$$

where $d_x^n(\lambda) = \sum_{l=0}^{n-1} x_l \exp(-i\lambda l)$. Therefore, interestingly,

$$I(\frac{2\pi k}{n}) = \frac{1}{2\pi}\lambda_{k,n}^2.$$

For any random variables $Y_1, \dots Y_k$, its *cumulant*, $cum(Y_1, \dots Y_k)$ is defined to be the coefficient of $\prod_{j=1}^k (it_j)$ in the expansion of $\log[E\exp\{i\sum_{j=1}^k t_j Y_j\}]$. Now suppose $\{x_l\}$ satisfies the following condition: Let

$$c_h(u_1, u_2, \dots u_{h-1}) = cum(x_{u_1}, x_{u_2}, \dots x_{u_{h-1}})$$

Then

$$\sum_{u_1, u_2, \dots u_{h-1}}^{\infty} (1 + |u_j|) c_h(u_1, u_2, \dots u_{h-1}) < \infty, \text{ for } 1 \leq j \leq h-1, \text{ and } h \geq 2 \quad (23)$$

By Theorem 2 of Chiu (1988),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{|j| \leq n} I^k(2\pi j/n) = \frac{k!}{2\pi} \int_{-\pi}^{\pi} f^k(\lambda) d\lambda \quad \text{almost surely.}$$

The left side equals $\frac{1}{n}\sum_j \frac{1}{(2\pi)^k}(a_j^2 + b_j^2)^k = \frac{1}{(2\pi)^k}\int x^{2k} dF_n$ where F_n is the ESD of X_n. Hence $\int x^{2k} dF_n \to (2\pi)^{k-1} k! \int_{-\pi}^{\pi} f^k(\lambda) d\lambda = \beta_{2k}$ (say) almost surely.

By Stirling's formula for large K,

$$\sum_{k \geq K} \beta_{2k}^{-\frac{1}{2k}} \sim \sum_{k \geq K} (2\pi)^{-\frac{1}{2} + \frac{1}{4k}} e^{\frac{1}{2}} k^{-\frac{1}{2} - \frac{1}{4k}} (\int_{-\pi}^{\pi} f^k(\lambda) d\lambda)^{-\frac{1}{2k}}.$$

But $|f(\lambda)| \leq \sum |c(u)| < M < \infty$. So the right side above is divergent. Hence $\{\beta_{2k}\}$ satisfies Carleman's condition. Thus we have proved the following theorem:

Theorem 12 *Suppose that $\{x_i\}$ is weakly stationary with mean zero and auto-covariance function satisfying (23). Then ESD of $X_n = n^{-1/2} A_n$ converges weakly almost surely. The odd moments of the LSD are zero and the moments of order $2k$ are given by $(2\pi)^{k-1} k! \int_{-\pi}^{\pi} f^k(\lambda) d\lambda$ where $f(\lambda) = \frac{1}{2\pi}\sum c(u) e^{-i\lambda u}$.*

We now deal with the case where the second moment of the entries need not be finite.

Suppose X is a metric space. Let $\mathbf{P}(\mathbf{X})$ denote the space of all probability measures on $\mathcal{B}_{\mathcal{X}}$, the Borel σ-field of X. Let $(\Omega, \mathcal{A}, Q)$ be a probability space. A *random measure* λ is defined to be a measurable map $\lambda : \Omega \to \mathbf{P}(\mathbf{X})$. Then $Q\lambda^{-1} \in \mathbf{P}(\mathbf{P}(\mathbf{X}))$ where $Q\lambda^{-1}$ is defined by $Q\lambda^{-1}(M) = Q\{\omega \in \Omega : \lambda(\omega) \in M\}$ where M is a Borel set of $\mathbf{P}(\mathbf{X})$. We say that $Q\lambda^{-1}$ is the distribution of λ.

Now let λ_n be a sequence of random measures with distributions $Q\lambda_n^{-1}$. We say that λ_n converges to a random measure λ *in distribution* if $Q\lambda_n^{-1} \to Q\lambda^{-1}$ weakly, that is, if $\int_{\mathbf{P(X)}} f(x) dQ\lambda_n^{-1}(x) \to \int_{\mathbf{P(X)}} f(x) Q\lambda^{-1}(x)$ for all continuous bounded function f on $\mathbf{P(X)}$.

For any probability measure P on $\mathbf{R^2}$, let $\hat{P}(t_1, t_2) = \int_{\mathbf{R^2}} \exp\{it_1 x + it_2 y\} dP(x, y)$ denote its characteristic function at (t_1, t_2). Let μ be an element of $\mathbf{P}(\mathbf{P}(\mathbf{R^2}))$ which satisfies for each $k \geq 1$,

$$\int \prod_{i=1}^{k} \hat{P}(t_{i1}, t_{i2}) d\mu(P) = \theta_k(t_{11}, t_{12}, \ldots t_{k1}, t_{k2}) \tag{24}$$

where

$$\theta_k(t_{11}, t_{12}, \ldots t_{k1}, t_{k2}) = \exp[-\int_{I^k} |\sum_{i=1}^{k} \{t_{i1} \cos(2\pi x_i) + t_{i2} \sin(2\pi x_i)\}| d\mathbf{x}]$$

and I_k is the k-dimensional unit cube. Such a μ exists by Lemma (34) of Freedman and Lane (1981).

Let $g : \mathbf{P}(\mathbf{R^2}) \to \mathbf{P}(\mathbf{R})$ be defined by $g(m) = l$ where $l(A) = m(f^{-1}(A))$ for any Borel set A in $\mathbf{R}$ and $f(x, y) = \sqrt{(}x^2 + y^2)$. Then we prove the following:

Theorem 13 *If $\{x_i\}$ are i.i.d. such that $n^{-\frac{1}{\alpha}} \sum_{j=1}^{n} x_j$ converges in distribution to a symmetric stable law of index $\alpha < 2$ then the ESD of the $X_n = n^{-1/\alpha} A_n$ which is a random measure, converges in distribution to a random measure whose distribution is given by $\nu = \mu g^{-1}$ with μ as defined above in (24).*

Proof: Define

$$Y_{ns} = n^{-\frac{1}{\alpha}} \sum_{j=1}^{n} x_j \cos(2\pi js/n) \text{ and } Z_{ns} = n^{-\frac{1}{\alpha}} \sum_{j=1}^{n} x_j \sin(2\pi js/n).$$

From the arguments in Freedman and Lane (1980) it follows very easily that the joint empirical distribution G_n of $\{(Y_{ns}, Z_{ns})\}_{1 \leq s \leq n}$ converges in distribution to a random measure whose distribution is μ.

On the other hand, the sth eigenvalue of $X_n = n^{-1/\alpha} A_n$ is $\lambda_s = \sqrt{Y_{ns}^2 + Z_{ns}^2}$. Convergence of $\{G_n\}$ in distribution ensures that the ESD F_n converges and to the limit $\nu := \mu g^{-1}$. This proves the Theorem. □

4.3.3 Hankel and Markov matrices

In addition to the Toeplitz matrix Bryc, Dembo and Jiang (2003) solved the problem of existence of LSD of two other related matrices, the random Hankel and the Markov matrices. The main results they obtained for Hankel and Markov matrices are the following:

Theorem 14 *Let the entries of the Hankel matrix H_n be i.i.d. real-valued random variables with mean zero and variance one. With probability one, the ESD of $\frac{1}{\sqrt{n}}H_n$ converges weakly, as $n \to \infty$, to a non-random symmetric probability measure which does not depend on the distribution of the entries of H_n, has unbounded support and is not unimodal.*

To state the theorem on the Markov matrices, define the *free convolution* of two probability measures μ and ν as the probability measure whose nth cumulant is the sum of the nth cumulants of μ and ν). The proof of the following result involves intricate combinatorial argument. For details, see Bryc, Dembo and Jiang (2003).

Theorem 15 *Let the entries of a symmetric Markov matrix M_n be i.i.d. random variables with mean zero and variance one. With probability one, the ESD of $\frac{1}{\sqrt{n}}M_n$ converges weakly as $n \to \infty$ to the free convolution of the semicircle and standard normal measures. This measure is a non-random symmetric probability measure with smooth bounded density, does not depend on the distribution of the underlying random variables and has unbounded support.*

4.4 I.I.D. entries and the Circular Law

Since a sequence of i.i.d. random variables exhibits 'invariance behaviour", (for example CLT), it is very natural to ask if such invariance holds for the LSD of matrices with i.i.d. entries. To state such a result, define the *circular law* as simply the uniform distribution on the unit disc of the complex plane. That is, its density is

$$c(x+iy) = \pi^{-1} \text{ if } 0 \leq x^2 + y^2 \leq 1. \tag{25}$$

Let $X_n = ((x_{ij}))_{i,j=1,2,\ldots,n}$. Then the *circular law conjecture* states that

Conjecture. *If $\{x_{ij}\}_{i,j=1,2,\ldots,}$ are i.i.d. complex random variables, with mean 0, variance 1 then the LSD of $\{n^{-1/2}X_n\}$ is the circular law given in (25).*

This conjecture is widely believed to be true. It has been established in certain special cases. However, a resolution of the conjecture in the completely general case has so far not been achieved.

Assuming that the entries are complex normal, (the real and complex parts being i.i.d. real normal with mean zero and variance 1/2 each), Ginibre (1965) showed that the joint density of the (complex) eigenvalues of X_n is given by:

$$c \prod_{j \neq k} |\lambda_j - \lambda_k|^2 \exp\{-\frac{1}{2}\sum_{k=1}^{n} |\lambda_k|^2\}.$$

This was used by Mehta (1991) to verify that the conjecture is true in this case. See also Hwang (1986) who credits it to unpublished work of Jack Silverstein.

Edelman (1997) derived the exact distribution of the (complex) eigenvalues when the entries are real normal, given that there are exactly k real eigenvalues. He used it to show that the expected ESD converges to the circular law.

Girko (1984a, b) gave a proof of the validity of conjecure under some restriction on the densities of the entries. He used the technique of *V-transformation* by which the characteristic function of a non-selfadjoint matrix is expressed in terms of ESD of Hermitian matrices. But researchers have found these proofs extremely difficult to understand.

Bai (1997) proved the circular law under reasonably general conditions by using some of the ideas of Girko (1984a, b). His result may be stated as follows:

Theorem 16 *Suppose the entries have finite $(4+\epsilon)$th moment and either the joint distribution of the real and complex parts has a bounded density or the conditional distribution of the real part given the imaginary part has a bounded density. Then the circular law conjecture is true.*

4.4.1 An Idea

The conjecture has eluded a proof in the general case so far. Note that the eigenvalues of the matrix need not be real. Hence, the moment method or the Stieltjes transform method apparently become inoperative. In this section we suggest a different approach. Though we do not have any rigorous results, we hope that the ideas given here would eventually turn out to be useful.

Let X be a complex valued random variable with law P. Let $B = B(a, r) = \{z : |z - a| < r\}$ denote the open disc of radius r with center a. Let Γ denote the circumference of B, parametrized in the usual way by the path γ. Also, suppose $f(z) := E[(z - X)^{-1}]$ exists $\forall z \in \Gamma$.

Note that if X is a real valued random variable, then $-f$ is the Stieltjes transform of X. The eigenvalues of Hermitian matrices are real and Bai (1999) has mentioned that the Stieltjes transform technique works only for such matrices. However, if the conditions for applying Fubini's theorem hold, we have

$$P(X \in B) = E\left(\frac{1}{2\pi i}\int_\Gamma \frac{d\gamma}{\gamma - X}\right) = \frac{1}{2\pi i}\int_\Gamma f(\gamma)d\gamma \tag{26}$$

This suggests that f may be used for non-Hermitian matrices in a manner similar to m for Hermitian matrices.

Note that if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, then the eigenvalues of $(zI - A)^{-1}$ are precisely $(z-\lambda_1)^{-1}, (z-\lambda_2)^{-1}, \ldots, (z-\lambda_n)^{-1}$. Hence, if z is *not* a characteristic root of the $n \times n$ matrix A, and P is the ESD of A, then $E_P[(z-X)^{-1}] = n^{-1}\text{tr}[(zI-A)^{-1}]$.

For the matrix X_n with iid entries, let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the eigenvalues of $n^{-1/2}X_n$. Define f_n on Ω_n = the complement of the set of all eigenvalues of $n^{-1/2}X_n$ as

$$f_n(z) = \frac{1}{n}\text{tr}\left[(zI - n^{-1/2}X_n)^{-1}\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{z-\lambda_i}. \tag{27}$$

Then $\forall z \in \Omega_n$

$$f_n(z) = \frac{1}{n}\frac{g_n'(z)}{g_n(z)} \tag{28}$$

where g_n is defined on the plane by

$$g_n(z) = \det(zI - n^{-1/2}X_n) = \prod_{i=1}^{n}(z-\lambda_i) \tag{29}$$

Clearly, g_n is a polynomial in z of degree n. Let us write

$$g_n(z) = \sum_{k=0}^{n} c_{nk}z^k \tag{30}$$

where the coefficients c_{nk} are random. Now, for $z \in \Omega_n \cup \{0\}$,

$$f_n(z) = \frac{1}{n}\frac{g_n'(z)}{g_n(z)} = \frac{1}{z}\frac{\sum_{k=0}^{n}\frac{k}{n}c_{nk}z^k}{\sum_{k=0}^{n}c_{nk}z^k} = \frac{1}{z}\frac{\sum_{k=0}^{n}\frac{k}{n}t_{nk}}{\sum_{k=0}^{n}t_{nk}} \tag{31}$$

where $t_{nk} = c_{nk}z^k$. Note the last expression in (31) is z^{-1} times an affine combination of k/n, $k = 0, 1, \ldots, n$. If it so happens that the values t_{nk} are "insignificant" when k/n lies outside a small neighbourhood of some number $\alpha(z)$, then $f_n(z)$ will be approximately equal to $\alpha(z)/z$. So let us inspect the nature of the random variables t_{nk}. Direct evaluation of the determinant $\det(zI - n^{-1/2}X_n)$ shows that

$$c_{nk} = n^{-(n-k)/2}\sum s(i_1, \ldots, i_{n-k}, j_1, \ldots, j_{n-k}) \cdot x_{i_1j_1}x_{i_2j_2}\cdots x_{i_{n-k}j_{n-k}} \tag{32}$$

where the summation is taken over all $i_1, i_2, \ldots, i_{n-k}, j_1, j_2, \ldots, j_{n-k}$ such that $1 \leq i_1 < i_2 < \cdots < i_{n-k} \leq n$ and $(j_1, j_2, \ldots, j_{n-k})$ is a permutation of $(i_1, i_2, \ldots, i_{n-k})$, and $s(i_1, \ldots, i_{n-k}, j_1, \ldots, j_{n-k})$ is either 1 or -1.

Since $\{x_{ij}\}$ are independent with mean 0 and variance 1, any pair of terms in (32) are uncorrelated. To see that, just note that in each product term, all the elements are forced to be distinct since $i_1 < i_2 < \cdots < i_{n-k}$. Also, the number of terms is $\frac{n!}{k!}$, since we can choose $(i_1, \ldots, i_{n-k})$ in ${}^nC_k$ ways, and for each such choice, we can choose $(j_1, \ldots, j_{n-k})$ in $(n-k)!$ ways, and there are no overlaps. Thus,

$$E\left(|t_{nk}|^2\right) = E\left(|c_{nk}z^k|^2\right) = \frac{n!|z|^{2k}}{k!n^{n-k}} = \tau_{nk} \text{ (say)} \tag{33}$$

Now note that

$$\frac{\tau_{n,k+1}}{\tau_{n,k}} = \frac{n|z|^2}{k+1} \tag{34}$$

Hence, τ_{nk} is maximized when k is such that $\frac{k}{n} \leq |z|^2 < \frac{k+1}{n}$ if $|z| < 1$. If $|z| \geq 1$ then the maximum is attained at $k = n$.

From now on assume that $|z| < 1$, $z \neq 0$.

Using (34), it is easy to prove that τ_{nk} are negligible when k/n falls outside $\left(|z|^2(1-n^{-1/3}), |z|^2(1+n^{-1/3})\right) = (\alpha_n, \beta_n)$ (say). More precisely, we can show that

$$\lim_{n\longrightarrow\infty} \frac{\sum_{\alpha_n < \frac{k}{n} < \beta_n} \tau_{nk}}{\sum_{k=0}^{n} \tau_{nk}} = \lim_{n\longrightarrow\infty} \frac{\sum_{\alpha_n < \frac{k}{n} < \beta_n} E\left(|t_{nk}|^2\right)}{\sum_{k=0}^{n} E\left(|t_{nk}|^2\right)} = 1 \tag{35}$$

Now, it is again easy to see, using (32), that if $k \neq j$ then t_{nk} and t_{nj} are uncorrelated. Using this, we get,

$$\begin{aligned}
&\lim_{n\longrightarrow\infty} \frac{E\left|\sum_{k=0}^{n}\left(\frac{k}{n}-|z|^2\right)t_{nk}\right|^2}{E\left|\sum_{k=0}^{n} t_{nk}\right|^2} = \lim_{n\longrightarrow\infty} \frac{\sum E\left|\left(\frac{k}{n}-|z|^2\right)t_{nk}\right|^2}{\sum E|t_{nk}|^2} \\
&= \lim_{n\longrightarrow\infty} \left\{ \frac{\sum_{\frac{k}{n}\notin(\alpha_n,\beta_n)} E\left|\left(\frac{k}{n}-|z|^2\right)t_{nk}\right|^2}{\sum E|t_{nk}|^2} + \frac{\sum_{\frac{k}{n}\in(\alpha_n,\beta_n)} E\left|\left(\frac{k}{n}-|z|^2\right)t_{nk}\right|^2}{\sum E|t_{nk}|^2} \right\} \\
&\leq \lim_{n\longrightarrow\infty} \left\{ \frac{\sum_{\frac{k}{n}\notin(\alpha_n,\beta_n)} E|t_{nk}|^2}{\sum E|t_{nk}|^2} + \frac{n^{-1/3}|z|^2 \sum_{\frac{k}{n}\in(\alpha_n,\beta_n)} E|t_{nk}|^2}{\sum E|t_{nk}|^2} \right\} \\
&= 0
\end{aligned} \tag{36}$$

The first term inside the limit vanishes due to (35), and the second term is dominated by $n^{-1/3}$. Now, *if we could only show that*

$$\lim_{n\longrightarrow\infty} E\left|\frac{\sum_{k=0}^{n}\left(\frac{k}{n}-|z|^2\right)t_{nk}}{\sum_{k=0}^{n} t_{nk}}\right|^2 = 0, \tag{37}$$

then it would follow that

$$f_n(z) = \frac{1}{z}\frac{\sum_{k=0}^{n}\frac{k}{n}c_{nk}z^k}{\sum_{k=0}^{n}c_{nk}z^k} = \frac{1}{z}\frac{\sum_{k=0}^{n}\frac{k}{n}t_{nk}}{\sum_{k=0}^{n}t_{nk}} \xrightarrow{L^2} \frac{|z|^2}{z} = \bar{z} \tag{38}$$

So, if $|z| < 1$ and $z \neq 0$ then $f_n(z) \xrightarrow{L^2} \bar{z}$. For $z = 0$, this can be proved independently. If this holds, then by Jensen's inequality and Fubini's theorem, it is easy to verify that for any ball B with boundary Γ, which is contained in the unit ball,

$$P_n(B) = \frac{1}{2\pi i}\int_\Gamma f_n(z)dz \xrightarrow{L^2} \frac{1}{2\pi i}\int_\Gamma \bar{z}dz = \pi^{-1}\text{area}(B) \tag{39}$$

where P_n denotes the ESD of $n^{-1/2}X_n$. This implies that the LSD of $\{n^{-1/2}X_n\}$ is indeed the circular law.

Acknowledgement. We are grateful to the referee for his careful reading of the article and for his extremely constructive suggestions. He has also suggested important references that we had missed.

References

[1] Anderson, T, W. (1984) *An Introduction to Multivariate Statistical Analysis.* Second edition. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[2] Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20, 262–268.

[3] Arnold, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahr.und Verw. Gebiete*, 19, 191–198.

[4] Bai, Z. D. (1997). Circular law. *Ann. Probab.*, 25, no. 1, 494–529.

[5] Bai, Z. D. (1999) Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9, 611-677 (with discussions).

[6] Bai, Zhidong and Saranadasa, Hewa (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica*, 6 , no. 2, 311–329.

[7] Bai, Z. D. and Yin, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.*, 16, no. 2, 863–875.

[8] Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab*, 21, 1275-1294.

[9] Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1986). On limiting spectral distribution of product of two random matrices when the underlying distribution is isotropic. *J. Multivariate Anal.*, 19, no. 1, 189–200.

[10] Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1987). On limiting empirical distribution function of the eigenvalues of a multivariate F matrix. *Teor. Veroyatnost. i Primenen.*, 32, no. 3, 537–548.

[11] Bickel, Peter J. and Freedman, David A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, no. 6, 1196–1217.

[12] Bose, Arup and Mitra, J. (2002). Limiting spectral distribution of a special circulant. *Stat. Probab. Letters*, 60, 1, 111-120.

[13] Böttcher, A. and Silbermann, B. (1990). *Analysis of Toeplitz operators*. Springer-Verlag, Berlin.

[14] Böttcher, A. and Silbermann, B. (1999). *Introduction to Large Truncated Toeplitz Matrices*, Springer, New York.

[15] Boutet de Monvel, A.; Khorunzhy, A. and Vasilchuk, V. (1996). Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Processes and Related Fields.* 2, no.4, 607–636

[16] Brockwell, P.J. and Davis, R.A (1991). *Time Series: Theory and Methods.* Second edition. Springer, New York.

[17] Bryc, W., Dembo, A. and Jiang, T.(2003). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Preprint*. Also available at *http://arxiv.org/abs/math.PR/0307330*

[18] Chiu, Shean-Tsong (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann. Statist.* 16, no. 3, 1315–1326.

[19] Deift, P. A. (1999). *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert approach.* Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; Amer. Math. Soc., Providence, RI.

[20] Donoho, P. High-dimensional data analysis: The curses and blessings of dimesionality. Technical Report, Department of Statistics, Stanford University.

[21] Edelman, A. (1997). The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60, no. 2, 203–232

[22] Freedman, D. A. and Lane, David (1980). The empirical distribution of Fourier coefficients. *Ann. Statist.* 8, no. 6, 1244–1251

[23] Ginibre, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6, 440–449.

[24] Girko, V. L. (1984a). Circular law (Russian) *Teor. Veroyatnost. i Primenen.* 29, no. 4, 669–679

[25] Girko, V. L. (1984b). On the circle law. *Theory Probab. Math. Statist.*, 28, 15-23.

[26] Girko, V. L. (1988). *Spectral Theory of Random Matrices* (Russian). Probability Theory and Mathematical Statistics, Nauka, Moscow.

[27] Girko, Vyacheslav L. (1995) *Statistical Analysis of Observations of Increasing Dimension.* Translated from the Russian. Theory and Decision Library. Series B: Mathematical and Statistical Methods, 28. Kluwer Academic Publishers, Dordrecht.

[28] Girko, V. L. and Repin, K. Yu. (1995). The quarter-of-circumference law. *Theory Probab. Math. Statist.* No. 50, 67–70; translated from *Teor. Ĭmovīr. Mat. Stat.* No. 50 (1994), 66–69.

[29] Gray, Robert M. (2002). Toeplitz and Circulant Matrices: A review. Available at *http://www-ee.stanford.edu/ gray/toeplitz.html*.

[30] Grenander, U. (1963). *Probabilities on Algebraic Structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Gteborg-Uppsala.

[31] Grenander, U. and Silverstein, J. W. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* 32, no. 2, 499–519.

[32] Grenander, U. and Szegö, G. (1984). *Toeplitz Forms and Their Applications,* Second edition. Chelsea Publishing Co., New York.

[33] Hammond, C. and Miller, S. J. (2003) Eigenvalue spacing distribution for the ensemble of real symmetric Toeplitz matrices. *Prepint.*

[34] Hwang, C. R. (1986). A brief survey on the spectral radius and the spectral distribution of large random matrices with i.i.d. entries. *Random matrices and their applications* (Brunswick, Maine, 1984), 145–152, *Contemp. Math.*, 50, Amer. Math. Soc., Providence, RI.

[35] Johnstone, Iain M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* 29, no. 2, 295–327.

[36] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* 12, no. 1, 1–38.

[37] Katz, Nicholas M. and Sarnak, Peter (1999). *Random Matrices, Frobenius Eigenvalues, and Monodromy.* Amer. Math. Soc. Colloq. Publ., 45. Amer. Math. Soc., Providence, RI.

[38] Mallows, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Statist.* 43, 508–515.

[39] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices, (Russian) *Mat. Sb. (N.S.)* 72 (114), 507–536.

[40] Mehta, M. L. (1991). *Random Matrices*, Academic Press, New York.

[41] Silverstein, J. W. (1985). The limiting eigenvalue distribution of a multivariate F matrix. *SIAM J. Math. Anal.* 16, no. 3, 641–646.

[42] Silverstein, Jack W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* 55, no. 2, 331–339.

[43] Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues for a class of large-dimensional random matrices, *J. Multivariate Anal.*, 54, no. 2, 175–192.

[44] Silverstein, Jack W. and Choi, Sang-Il (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* 54, no. 2, 295–309

[45] Silverstein, J. W. and Combettes, P. L. (1992a). The relevance of spectral theory of large dimensional sample covariance matrix in source detection and estimation. *Proceedings of the sixth SSAP Workshop on Statistical Signal and array processing.* Victoria, B. C. Oct.7-9, 1992, 276-279.

[46] Silverstein, J. W. and Combettes, P. L. (1992b). Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. on signal processing.* Vol. 40, no. 8, 2100-2105.

[47] Trotter, H.F. (1984). Eigenvalue distributions of large Hermitian matrices; Wigner's semicircle law and a theorem of Kac, Murdock and Szegö. *Advances in Math.* 54, 67–82

[48] Wachter, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* 6, 1–18

[49] Wachter, K.W. (1980). The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.* 8, 937-957.

[50] Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, (2), 62, 548–564

[51] Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, (2), 67, 325–327

[52] Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20, no. 1, 50–68

[53] Yin, Y. Q.; Bai, Z. D.; Krishnaiah, P. R.(1983). Limiting behavior of the eigenvalues of a multivariate F matrix. *J. Multivariate Anal.*, 13 no. 4, 508–516.

[54] Yin, Y. Q. and Krishnaiah, P. R. (1985). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.*, 30, 861-867.

Figure 1. Kernel density estimates for the ESD for 50 simulated Toeplitz matrices of order 200 with $N(0,1)$ entries.

Figure 2. Average kernel density estimate of the ESD from 500 simulated Toeplitz matrices of order 200 with $N(0,1)$ entries.