A New Method for Bounding Rates of Convergence of Empirical Spectral Distributions

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

Abstract

The probabilistic properties of eigenvalues of random matrices whose dimension increases indefinitely has received considerable attention. One important aspect is the existence and identification of the limiting spectral distribution (LSD) of the empirical distribution of the eigenvalues. When the LSD exists, it is useful to know the rate at which the convergence holds. The main method to establish such rates is the use of Stieltjes transform. In this article we introduce a new technique of bounding the rates of convergence to the LSD. We show how our results apply to specific cases such as the Wigner matrix and the Sample Covariance matrix.

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

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

Revision 2
April 27, 2004

[*]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: abose@isical.ac.in

1 Introduction

Random matrices with increasing dimensions are called *large dimensional random matrices* (LDRMs). A nice 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), Bose et. al. (2003) and the books by Mehta (1991) and Girko (1988, 1995). Random matrices have also drawn the attention of mathematicians for various reasons. The books by Deift (1999) and Katz and Sarnak (1999) deal with the mathematical aspects of random matrices.

Suppose A_n is an $n \times n$ Hermitian matrix with eigenvalues (characteristic roots) $\lambda_0, \lambda_1, \ldots, \lambda_{n-1}$. Then the *empirical spectral distribution (ESD)* function of A_n is defined as

$$F_n(x) = n^{-1} \sum_{i=0}^{n-1} I\{\lambda_i \leq x\}.$$

The corresponding probability measure P_n is known as the *empirical spectral measure*. Note that if $\{A_n\}$ are random, then F_n and P_n are random: $F_n(x)$ is a random variable for every x and for every element in the basic probability space $F_n(\cdot)$ is a distribution function. Also, if F and G are two random distribution functions, then their Kolmogorov distance $\|F - G\|_\infty = \sup_{r \in \mathbb{Q}} |F(r) - G(r)|$ is also a random variable. For any distribution function G (which may be random) on $\mathbb{R}$, its (random) characteristic function is defined as $\varphi_G(t) = \int e^{itx} dG(x)$. When talking about the convergence of distribution functions we shall mean weak convergence, and use the notation "$F_n \Longrightarrow F$", as usual. Note that since weak convergence of probability measures on $\mathbb{R}$ is metrizable, the concept of "convergence in probability" is well-defined for distribution functions. Also, it is well known that if F is a continuous distribution function, then $F_n \Longrightarrow F$ if and only if $\|F_n - F\|_\infty \to 0$, and so if F is continuous then $F_n \stackrel{P}{\Longrightarrow} F$ if and only if $\|F_n - F\|_\infty \stackrel{P}{\longrightarrow} 0$.

If $\{A_n\}_{n=1}^\infty$ is a sequence of square matrices with the corresponding ESD $\{P_n\}_{n=1}^\infty$, (typically with the dimension of A_n increasing with n), 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 the matrices are random, the limit is understood to be in a probabilistic sense, either "almost surely" or "in probability".

The *expected spectral distribution function* of A_n is defined as $E(F_n(\cdot))$. This expectation always exists and is a nonrandom distribution function. The corresponding probability measure is called the *expected spectral measure.*

There are essentially two general tools available to establish the LSD: the moment method and the Stieltjes transform method. Often the expected distribution function is easier to deal with. The weak convergence of $E(F_n)$ then serves as an intermediate step in showing the weak convergence of F_n.

When the LSD exists, it is useful to know the rate at which the convergence holds. The main method to establish such rates is the use of Stieltjes transform. In this article we establish some general results useful in establishing the probabilistic weak convergence of F_n from the convergence of $E(F_n)$ and the corresponding rates of convergence. We apply these to establish some new rates of convergence. The rate will be measured in terms of the following two quantities:

$$\Delta^*(F,G) = E\|F-G\|_\infty = E\ \sup_{r\in\mathbb{Q}} |F(r)-G(r)|$$

and when G is non random,

$$\Delta(F,G) = \|E(F)-G\|_\infty = \sup_{r\in\mathbb{Q}} |E(F(r))-G(r)|$$

Given a random Hermitian matrix A of order n, the *empirical characteristic function* of A is the characteristic function of the empirical spectral distribution of A. Let us call it φ. From the spectral decomposition of A, it is easy to see that

$$\varphi(t) = \frac{1}{n}\mathrm{tr}(e^{itA})$$

where n is the order of A and e^M denotes $\sum_{k=0}^{\infty}\frac{1}{k!}M^k$, as usual. We shall hence forth deal with the spectral measure of A through this characteristic function.

Our approach is to obtain bounds for $\mathrm{Var}(\varphi(t))$ and then using Esseen's lemma or otherwise, deduce the concentration of the spectral measure near its mean, and also get the magnitude of concentration using the bounds.

To do this, we need to be able to express the random matrix A as a function of independent real random variables $x_1, x_2, \ldots, x_m$, where m is large. Then for each t, $\varphi(t)$ is also a function of $x_1, x_2, \ldots, x_m$. Typically, we show that this function is *slowly varying*, that is, either the partial derivatives are bounded by small numbers in sup norm, or the expected value of the norm-squared of $\nabla\varphi(t)$ is small. This, followed by an application of a Poincaré type inequality (when we have a bound on $\|\nabla\varphi(t)\|_{L^2}$) or an Efron-Stein type inequality (when we have bounds on the partials) will produce a bound on $\mathrm{Var}(\varphi(t))$.

The bounds on the partial derivatives and the gradient of $\varphi(t)$ (as a function of $x_1, x_2, \ldots, x_m$) are obtained by using the identity

$$\frac{\partial\varphi(t)}{\partial x_j} = \frac{1}{n}\mathrm{tr}\left(it\frac{\partial A}{\partial x_j}e^{itA}\right)$$

coupled with the careful use of the fact that e^{itA} is a unitary matrix. It may be mentioned that in none of our examples shall we need to compute e^{itA} explicitly.

We shall employ the above approach to a few examples, including large dimensional Wigner and Sample Covariance matrices, and obtain improved rates of convergence under suitable conditions. Simulation results suggest that our bounds for $\mathrm{Var}(\varphi(t))$ have the correct exponent for n in all cases.

We now introduce some notations. For a complex random variable X, its variance is defined to be $E|X - E(X)|^2$. For $L > 0$, define the probability density

$$h_L(x) = \frac{1 - \cos Lx}{\pi L x^2}.$$

Let H_L be the corresponding distribution function. The characteristic function of H_L is given by $\psi_L(t) = (1 - \frac{|t|}{L}) I(|t| \le L)$. Note that $\int_{-\infty}^{\infty} |\psi_L(t)| dt = L$.

Finally, the *convolution* $F * G$ of F and G is defined in the usual way. That is, $F * G(x) = \int F(x - y) dG(y) = \int G(x - y) dF(y)$.

Now suppose we have a complex matrix A which is a (componentwise) differentiable function of a real or complex scalar variable u. The following two simple Lemmae will be useful. We omit their proofs.

Lemma 1 *If $A(u)$ is an elementwise differentiable map from $\mathbb{R}$ or $\mathbb{C}$ into $\mathbb{C}^{n \times n}$ then*

$$\frac{d}{du} \mathrm{tr}(e^A) = \mathrm{tr} \left(\frac{dA}{du} e^A \right)$$

Lemma 2 *If A is Hermitian and t is real, then e^{itA} is a unitary matrix. In particular, for any vector x, $|e^{itA} x| = |x|$, (where $|\cdot|$ denotes the Euclidean norm) and also all entries of e^{itA} have modulus ≤ 1.*

2 Main Results

We first establish a bound on the expected Kolmogorov distance. This will be eventually used to establish rates of convergence for the ESD.

Theorem 1 *Suppose F is a random distribution function on $\mathbb{R}$ with (random) characteristic function φ. Suppose* $\mathrm{Var}(\varphi(t)) \le Ct^2$ *for each t. If G is a nonrandom distribution function on $\mathbb{R}$, such that* $\sup_{x \in \mathbb{R}} |G'(x)| \le \lambda$ *, then*

$$\Delta^*(F, G) \le 2\Delta(F, G) + \frac{8(3)^{1/2} \lambda^{1/2}}{\pi} C^{1/4}$$

where Δ and Δ^* are as defined in the introduction.

Proof Let $F_0 = E(F)$, and let η be the characteristic function of F_0. Then by assumption,

$$E|\varphi(t) - \eta(t)| \le \sqrt{C} |t|.$$

By Lemma 1 (Esseen's lemma) of Feller (1966, page 510),

$$\|F - G\|_\infty \le 2\|F * H_L - G * H_L\|_\infty + \frac{24\lambda}{\pi L}.$$

Now

$$\begin{aligned} \|F * H_L - G * H_L\|_\infty &\le \|F_0 * H_L - G * H_L\|_\infty + \|F * H_L - F_0 * H_L\|_\infty \\ &\le \|F_0 - G\|_\infty + \|F * H_L - F_0 * H_L\|_\infty. \end{aligned}$$

So by applying the inversion formula (see Feller 1966, page 482-484) and the hypothesis about $\mathrm{Var}(\varphi(t))$,

$$\begin{aligned} E\|F * H_L - F_0 * H_L\|_\infty &\le \frac{1}{\pi}\int_{-\infty}^{\infty} |\psi_L(t)| \frac{E|\varphi(t) - \eta(t)|}{|t|} dt \\ &\le \frac{C^{1/2}L}{\pi}. \end{aligned}$$

Combining all these observations, we have

$$\Delta^*(F, G) \le 2\Delta(F, G) + \frac{2\sqrt{C}L}{\pi} + \frac{24\lambda}{\pi L}.$$

Choosing $L^2 = 12\lambda C^{-1/2}$ gives the desired conclusion. □

Remark 1: The following result linking the convergence of expected Kolmogorov distance with the convergence of the characteristic function may also be proved by a similar convolution argument. We omit the proof.

Theorem 2 *Let $\{F_n, n \ge 1\}$ (random) and F (nonrandom) be distribution functions on $\mathbb{R}$, with characteristic functions $\{\varphi_n, n \ge 1\}$, and φ. Suppose F is differentiable everywhere with bounded derivative. Then the following are equivalent:*
(a) $\Delta^*(F_n, F) \longrightarrow 0$
(b) $\varphi_n(t) \longrightarrow \varphi(t)$ in probability for each $t \in \mathbb{R}$
(c) $E|\varphi_n(t) - \varphi(t)| \longrightarrow 0$ for each t.

Note the condition on the variance in the statement of Theorem 1. The following result on bound for variances of functions of independent random variables is useful while applying Theorem 1 to ESD. Part (b) follows from part (a). The earliest version of part (a) is credited to Hoeffding (unpublished work) and different versions are due to Efron and Stein (1981), Steele (1986) and Devroye (1991). A proof may be found in Györfi et al. (2002).

Theorem 3 *(Efron-Stein type inequality).*
(a) Suppose $Z_1, \ldots, Z_n, Z_1^*, \ldots, Z_n^*$ are independent m-dimensional random vectors

where Z_i has the same distribution as Z_i^* for all i. Suppose that $f : (\mathbb{R}^m)^n \to \mathbb{C}$ satisfies $E|f(Z_1, \ldots, Z_n)|^2 < \infty$. Then

$$\mathrm{Var}(f(Z_1, \ldots, Z_n)) \leq \frac{1}{2} \sum_{k=1}^{n} E|f(Z_1, \ldots, Z_n) - f(Z_1, \ldots, Z_{k-1}, Z_k^*, Z_{k+1}, \ldots, Z_n)|^2.$$

(b) If $f : \mathbb{R}^n \to \mathbb{C}$ is Lipschitz in each coordinate with Lipschitz constants $M_1, M_2, \ldots, M_n$, then for independent square integrable real random variables $X_1, X_2, \ldots, X_n$,

$$\mathrm{Var}(f(X_1, X_2, \ldots, X_n)) \leq \sum_{j=1}^{n} M_j^2 \mathrm{Var}(X_j).$$

Better results can be obtained if $X_1, X_2, \ldots, X_n$ are i.i.d. from a distribution F which has the following property:

POIN *There exists a constant $K > 0$ such that if $X \sim F$, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a (locally) absolutely continuous map, then* $\mathrm{Var}(g(X)) \leq KE|g'(X)|^2$.

Remark 2: Such inequalities are known as "Poincaré Inequalities" in the literature. It may be noted that (a) if X satisfies **POIN** with constant K, then for any $c \in \mathbb{R}$, cX satisfies **POIN** with constant Kc^2, and (b) for any distribution function satisfying **POIN**, the variance inequality holds for absolutely continuous functions $g : \mathbb{R} \longrightarrow \mathbb{C}$ as well. There is a huge literature on Poincaré and isoperimetric inequalities for probability measures, and we have included some of that in our list of references. The fact that the one dimensional Gaussian distribution satisfies **POIN** has been a part of folklore and has been known since 1930s. See for example Beckner (1989). That the multidimensional Gaussian distribution also satisfies **POIN** has been known since 1950s. See for example Brascamp and Lieb (1976). All distributions with log-concave densities (i.e. densities of the form $e^{U(x)}$ where U is a concave function) satisfy **POIN**. A complete characterization of all absolutely continuous distributions which satisfy **POIN** is available in Muckenhoupt (1972).

The next result, which follows from the Efron-Stein inequality is very well known and is provable under weaker assumptions. See Ledoux (2000).

Theorem 4 *If $X_1, X_2, \ldots, X_n$ are independent and satisfy* **POIN** *with Poincaré constants bounded by K, then for any C^1 map $f : \mathbb{R}^n \longrightarrow \mathbb{C}$,*

$$\mathrm{Var}(f(X_1, \ldots, X_n)) \leq KE|\nabla f(X_1, \ldots, X_n)|^2$$

where $|\cdot|$ denotes the Euclidean norm.

Now we demonstrate an application of the above results to find rates of convergence for some random matrices:

Example 1 (Wigner Matrices). 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. Several results on its LSD and rates of convergence of the ESD are known. Wigner (1955) assumed the entries to be i.i.d. real Gaussian and established the convergence of $E(F_n)$ to the semi-crcular law. 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, 1971) obtained almost sure convergence assuming independence of the entries and finiteness of moments. Bai (1999) generalised 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. There is a related result of Trotter (1984) also. Boutet de Monvel, Khorunzhy and Vasilchuk (1996) obtained some other generalizations of Wigner's results with weakly dependent Gaussian sequences.

For our purpose, we shall take the elements to be real. Suppose that W_n is a Wigner matrix with random independent entries $(X_{jk}^{(n)})$ having common variance 1. We shall drop the superscript n for ease of notation. In many situations, the LSD of $n^{-1/2}W_n$ exists and is given by the famous semi-circle law

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{x}(4-y^2)^{1/2}I_{[-2,\ 2]}dy.$$

Consider the following "basic assumptions":

(W1) $E(X_{jk}) = 0,\ E(X_{jk}^2) = 1.$

(W2) $\sup_{i,j,n} EX_{ij}^8 < \infty.$

(W3) $\sum E\left(X_{ij}^4 I_{\{|X_{ij}|\geq \epsilon n^{1/2}\}}\right) = o(n^2)$ for any $\epsilon > 0.$

Let F_n be the ESD of $n^{-1/2}W_n$. Bai (1993a) proved that under the above assumptions, $\Delta(F_n, F) = O(n^{-1/4})$ which was improved by Bai, Miao and Tsay (1997) to $\|F_n - F\|_\infty = O_p(n^{-1/4})$. In Bai, Miao and Tsay (2002), this was further improved to $\|F_n - F\|_\infty = O_p(n^{-2/5})$.

Suppose we strengthen the third assumption to

(W3*) $\sum E\left(X_{ij}^8 I_{\{|X_{ij}|\geq \epsilon n^{1/2}\}}\right) = o(n^2)$ for any $\epsilon > 0.$

Then they also showed that $\Delta(F_n, F) = O(n^{-1/2})$.
Further, suppose that the basic assumptions hold and in addition assume that

(W3**) $\sup_n \sup_{ij} E|X_{ij}|^k < \infty$ for every $k \geq 1$.

Then $\|F_n - F\|_\infty = O(n^{-2/5+\eta})$ almost surely for every $\eta > 0$.

We will show here how our results may be applied under minimal conditions to obtain weaker rate results, and under stronger conditions, new and stronger results.

Fix any $n \geq 1$. Suppose we write the elements of $\mathbb{R}^{n(n+1)/2}$ as tuples of the form (a_{jk}), where j runs from 1 to n, and for each j, k runs from 1 to j. Then, we can have a map $W_n : \mathbb{R}^{n(n+1)/2} \longrightarrow \mathbb{R}^{n\times n}$ which takes a tuple (a_{jk}) to the Wigner matrix whose (j,k)-th entry is $n^{-1/2}a_{jk}$ if $j \geq k$, and $n^{-1/2}a_{kj}$ otherwise. Then $W_n^{jk} = \frac{\partial W_n}{\partial a_{jk}}$ is a constant matrix whose (j,k)-th and (k,j)-th entries are $n^{-1/2}$ and all other entries are zero. Thus, if we fix some $t \in \mathbb{R}$ and define φ_n^t to be the empirical characteristic function of W_n evaluated at t, then it follows from the results of the preceding section that

$$\frac{\partial \varphi_n^t}{\partial a_{jk}} = n^{-1}\mathrm{tr}\left(it\frac{\partial W_n}{\partial a_{jk}} e^{itW_n}\right) = n^{-1}\mathrm{tr}\left(itW_n^{jk}e^{itW_n}\right).$$

Now, if we let $B = e^{itW_n}$, and denote its elements by b_{jk}, it follows that

$$\frac{\partial \varphi_n^t}{\partial a_{jk}} = \frac{it(b_{jk}+b_{kj})}{n\sqrt{n}} = \frac{2itb_{jk}}{n\sqrt{n}}$$

The last equality holds because B is symmetric. Thus,

$$|\nabla \varphi_n^t|^2 = \sum_{j\geq k} |\frac{\partial \varphi_n^t}{\partial a_{jk}}|^2 \leq \sum_{j,k} \frac{4t^2|b_{jk}|^2}{n^3} = \frac{4t^2}{n^2}.$$

The last equality follows from the fact that B is unitary.
(It is worth mentioning that it is a well-known result that for any Lipschitz function $f : \mathbb{R} \to \mathbb{C}$, if we define $T_f(W_n) = n^{-1}\sum_{j=1}^n f(\lambda_j)$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of W_n, then $|\nabla T_f(W_n)|^2 \leq 2n^{-1}\|f\|_{\mathrm{Lip}}^2$. See, for example, Horn and Johnson (1985) or Simon (1979). For the case of complex entries, a similar result holds, too. See Guionnet and Zeitouni (2000)). Applying these observations to the scenario where a_{jk} are random, and noting that $\sup_{-2\leq x\leq 2} F'(x) = \pi^{-1}$, we have:

Theorem 5 *If W_n is a random real Wigner matrix whose entries on and above the diagonal are independent and satisfy* **POIN** *with Ponicaré constants uniformly bounded by K, then $\mathrm{Var}(\varphi_n(t)) \leq 4Kt^2/n^2$. Consequently, by Theorem 1, if F denotes the semicircular law, then*

$$\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + \frac{8(6)^{1/2}K^{1/4}}{\pi^{3/2}}n^{-1/2}$$

where F_n denotes the empirical c.d.f. of $n^{-1/2}W_n$.

In this context, it should be mentioned that general bounds for $P(|\mathrm{tr}(f(W_n)) - E(\mathrm{tr}(f(W_n)))| > t)$ where f is a Lipschitz function, may be obtained by using the results of Guionnet and Zeitouni (2000, Theorem 1.1). However, if f is not convex (as is the case here), then the stronger assumption that the distribution of the entries satisfy a logarithmic Sobolev inequality instead of **POIN** is required for those bounds to hold. Those bounds would imply the variance bound on the empirical characteristic function that we need. However, since f in this problem is not convex, and since we are only interested in variance bounds for applying Theorem 1, the stronger assumption seems to be unnecessary.

Now note that by Lemma 2, the elements of e^{itW_n} are bounded in modulus by 1, and this implies

$$\left\| \frac{\partial \varphi_n^t}{\partial a_{jk}} \right\|_\infty \leq 2|t| n^{-3/2}.$$

So, if we don't assume **POIN**, we can still have the following result under remarkably weak conditions, by invoking Theorems 3 and 1:

Theorem 6 *If W_n is a random real Wigner matrix, whose entries on and above the diagonal are independent with variance uniformly bounded by* 1*, then*

$$\mathrm{Var}(\varphi_n(t)) \leq \frac{4t^2}{n^3} \sum_{j \geq k} \mathrm{Var}(x_{jk}) \leq \frac{4t^2}{n}.$$

Hence if F_n denotes the empirical c.d.f. of $n^{-1/2}W_n$ and F denotes the semicircular law, then

$$\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + \frac{8(6)^{1/2}}{\pi^{3/2}} n^{-1/4}.$$

Remark 3 A recent result of Götze and Tikhomirov (2003) which appeared after this article was submitted supercedes Theorem 5. There it is shown that if $M_4 = \sup_{j,k} EX_{jk}^4$, then $\Delta(F_n, F) \leq CM_4^{1/2} n^{-1/2}$. If further the observations are Gaussian, then Götze and Tikhomirov (2002) show that $\Delta(F_n, F) = O(n^{-2/3})$. Theorem 6, however, seems to be new.

Example 2 (Sample covariance matrices). Suppose X is a real $p \times n$ matrix with entries x_{jk}, which are i.i.d. real random variables with mean zero and unit variance. Let $S = \frac{1}{n} XX^T$. In case, the entries are i.i.d. normal, much is known about the distribution of eigenvalues of S and related matrices. See Anderson (1984). The LSD of S was first established by Marčenko and Pastur (1967). Subsequent work on S may be found in Grenander and Silverstein (1977), Wachter (1978), Jonsson (1982), Yin (1986), Yin and Krishnaiah (1985) and Bai and Yin (1988a). If $y_n = p/n \longrightarrow y \in (0, 1]$ then the ESD of S_n converges almost surely to the law $F_y(\cdot)$ with the Marčenko-Pastur density

$$f_y(x) = \begin{cases} \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $a = a(y) = (1-\sqrt{y})^2$ and $b = b(y) = (1+\sqrt{y})^2$. It can be easily shown that the density is bounded by $\lambda = \left[\pi\sqrt{y}(1-y)\right]^{-1}$.

In cases where $y > 1$, the LSD exists but has a point mass at the origin. If $y = 0$, then a scaling and a centering are required for the LSD of S_n to exist. See Bai (1999) or Bose et al. (2003) for the precise results. We do not consider these cases. For versions of this result under variations of the above conditions, see the above references.

As in the case of Wigner matrices, the Stieltjes transform method was used to derive rates of convergence results. Bai (1993b) proved that $\Delta(F_n, F_y) = O(n^{-1/4})$ or $O(n^{-5/48})$ depending on how close y_n is to 1. The same rates were obtained for convergence in probability of F_n to F_y in Bai, Miao and Tsay (1997). The most recent results are by Bai, Miao and Yao (2003) who prove several results under the conditions given in Example 1. In particular it follows from their results that if y_n remains bounded away from 1 and suitable combinations of the above conditions hold then $\Delta^*(F_n, F_y) = O(n^{-1/2})$, $||F_n - F_y||_\infty = O_P(n^{-2/5})$ and $||F_n - F_y||_\infty = O_{a.s.}(n^{-2/5+\eta})$.

Now consider S as a function of the entries of X. Clearly,

$$S_{jk} := \frac{\partial S}{\partial x_{jk}} = \frac{1}{n}(YX^T + XY^T)$$

where $Y = \partial X / \partial x_{jk}$. Now the matrix Y has 1 at the (j,k)th position and 0 elsewhere, i.e. $Y = e_{j,p} e_{k,n}^T$ where $e_{m,r}$ is the r-vector with 1 at the mth position and 0 elsewhere. Thus, if $x_{\cdot k}$ denotes the kth column of X and φ_n^t denotes the empirical characteristic function evaluated at t, then

$$\begin{aligned}
\frac{\partial \varphi_n^t}{\partial x_{jk}} &= p^{-1}\mathrm{tr}(itS_{jk}e^{itS}) \\
&= it(np)^{-1}\mathrm{tr}(YX^Te^{itS} + XY^Te^{itS}) \\
&= it(np)^{-1}\mathrm{tr}(e_{j,p}x_{\cdot k}^T e^{itS} + x_{\cdot k}e_{j,p}^T e^{itS}) \\
&= it(np)^{-1}(x_{\cdot k}^T e^{itS} e_{j,p} + e_{j,p}^T e^{itS} x_{\cdot k}) \\
&= \frac{2itz_{kj}}{np}
\end{aligned}$$

where we have written z_{kj} for the jth component of the vector $z_k := e^{itS}x_{\cdot k}$. Note that since e^{itS} is unitary, $\|z_k\| = \|x_{\cdot k}\|$.
Now suppose x_{jk} are random variables. Then using the preceding observations, we have

$$\sum_{j,k} \left|\frac{\partial \varphi_n^t}{\partial x_{jk}}\right|^2 \le \frac{4t^2}{n^2p^2}\sum_{k=1}^n \|z_k\|^2 = \frac{4t^2}{n^2p^2}\sum_{k=1}^n \|x_{\cdot k}\|^2 = \frac{4t^2}{n^2p^2}\sum_{j,k}|x_{jk}|^2 \quad \text{a.s.}$$

and so, under the assumption $\forall j,k,\ E|x_{jk}|^2 \le M^2$, it follows that

$$E|\nabla \varphi_n^t|^2 \le \frac{4M^2t^2}{np}.$$

Applying Theorems 4 and 1, we immediately have the following result:

Theorem 7 *If $\{x_{jk}\}$ are independent and satisfy* **POIN** *with Poincaré constants bounded by K and second moments bounded by M, then $\mathrm{Var}(\varphi_n(t)) \le \frac{4KM^2t^2}{np}$. Consequently, if $y = p/n \in (0,1)$ and F_y denotes the Marčenko-Pastur distribution with parameter y, then*

$$\Delta^*(F_{n,p}, F_y) \le 2\Delta(F_{n,p}, F_y) + \frac{8(6)^{1/2}K^{1/4}M^{1/2}}{\pi^{3/2}[y(1-y)]^{1/2}} n^{-1/2}$$

where $F_{n,p}$ denotes the ESD of S.

(Note that if X is a mean zero random variables satisfying **POIN** with constant K, then automatically $EX^2 = \mathrm{Var}(X) \le K$. So we can put $M = K$ if the entries have zero mean.)

If we don't assume **POIN** but instead impose $\forall j,k,\ |x_{jk}| \le M$ a.s., then

$$|\frac{\partial \varphi_n^t}{\partial x_{jk}}| \le 2|t|(np)^{-1}\|z_k\| = 2|t|(np)^{-1}\|x_{\cdot k}\| \le \frac{2M|t|}{n\sqrt{p}} \quad \text{a.s.}$$

Thus, if the variance of x_{jk} is bounded by 1, then

$$\mathrm{Var}(\varphi_n(t)) \le \frac{4M^2t^2}{n^2p} np = \frac{4M^2t^2}{n}.$$

Hence we get

Theorem 8 *Suppose $y = p/n \in (0,1)$ and x_{jk} are independent with mean zero and variance bounded by 1. Suppose M is such that $P(|x_{jk}| \le M) = 1$. Then*

$$\Delta^*(F_{n,p}, F_y) \le 2\Delta(F_{n,p}, F_y) + \frac{8(6)^{1/2}M^{1/2}}{\pi^{3/2}y^{1/4}(1-y)^{1/2}} n^{-1/4}$$

where $F_{n,p}$ is the EDF of S, as before.

Remark 4 Again, it is proved in Götze and Tikhomirov (2003) that under finite twelfth moment, $\Delta(F_{n,p}, F_y) = O(n^{-1/2})$. However, Theorem 8 still appears to be a new result.

Example 3 (Anti-Toeplitz matrix). Suppose $\{x_0, x_1, x_2, \ldots\}$ is a sequence of numbers. The *anti-Toeplitz matrix* of order n defined by this sequence is $A_n = ((x_{(i+j-2) \bmod n}))$. Visually,

$$A_n = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_{n-1} & x_0 \\ x_2 & \cdots & x_{n-1} & x_0 & x_1 \\ & & \vdots & & \\ x_{n-1} & x_0 & x_1 & \cdots & x_{n-2} \end{bmatrix}$$

From the results of Bose and Mitra (2002) and Bose, Chatterjee and Gangopadhyay (2003), it follows that if $\{x_i\}$ are i.i.d. with mean zero and variance 1 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 by

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

Hence the ESD converges to this distribution in probability.
Let $B = e^{itA_n}$. Denote elements of B by b_{ij}, and the empirical characteristic function of A_n evaluated at t by φ_n^t, as usual. Then it can be checked by our usual technique that

$$\frac{\partial \varphi_n^t}{\partial x_k} = \frac{it}{n\sqrt{n}} \sum_{i+j-2=k \bmod n} b_{ji}$$

for $k = 0, 1, \ldots, 2n-2$. Thus,

$$\sum_k \left|\frac{\partial \varphi_n^t}{\partial x_k}\right|^2 \leq \frac{t^2}{n^3} \sum_k \left[2n \sum_{i+j-2=k \bmod n} |b_{ji}|^2 \right] = \frac{2t^2}{n}$$

We used the Cauchy-Schwarz inequality, noting that for each k, there are at most $2n$ pairs of (i,j) such that $i+j-2 = k \bmod n$. The last equality holds due to the fact that $\sum\sum |b_{ij}|^2 = n$, as we observed before. Now we can show, as before, that if x_k are i.i.d. from a density satisfying **POIN**, then $\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + O(n^{-1/4})$. Note that in this case, the eigenvalues can be explicitly obtained and using their form, under suitable conditions, $\Delta(F_n, F)$ is of a much smaller order than $n^{-1/4}$.

In the next two examples on Hankel and Toeplitz matrices, the existence of the LSD were open problems, very recently settled by Bryc, Dembo and Jiang (private communication). However, neither the closed form expressions of the LSD nor the convergence rates are known. Our method, however, applies very easily to give bounds like $\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + O(n^{-1/4})$ (assuming that the limiting distribution has a bounded density). We hope this will considerably ease the task of finding the convergence rates after the limiting distributions are identified.

Example 4 (Hankel matrix). A matrix of the form $H_n = ((x_{i+j-2}))$ (under the same notation as in the previous example) is called a Hankel matrix. Note that the matrix is symmetric. The objective is to investigate the limiting behaviour of the spectral distribution of $n^{-1/2}H_n$. As we said before, the existence of the LSD has been settled. The computations for our method are very similar to the previous example. In fact, here

$$\frac{\partial \varphi_n^t}{\partial x_k} = \frac{it}{n\sqrt{n}} \sum_{i+j-2=k} b_{ji}$$

and so, exactly similar computations as before show that under **POIN** we can again get $\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + O(n^{-1/4})$ if the x_j are i.i.d. with mean zero and varaince 1 and F has a bounded density.

Example 5 (Toeplitz matrix). Under the same notation as before, the $n \times n$ matrix $T_n = ((x_{|i-j|}))$ is called a *Toeplitz matrix* of order n. Some theoretical results and simulations of Bose, Chatterjee and Gangopadhyay (2003) showed that it is plausible that the LSD of $n^{-1/2}T_n$ exists when the variables form one i.i.d. sequence. Recently Bryc, Dembo and Jiang (private communication) has shown that the LSD exists, is unimodal and nonnormal. Exactly the same kind of computations as in the preceding examples show that in this case, too, under **POIN**, $\Delta^*(F_n, F) \leq 2\Delta(F_n, F) + O(n^{-1/4})$, again if the limiting distribution has a bounded density.

Acknowledgement. We are specially thankful to the referee for the very detailed report and suggestions. In particular he corrected us on several issues and provided us with references that we were not aware of. We also thank W. Bryc, Amir Dembo and T. Jiang for providing us with the preprints of their work. Finally, we thank Anirban DasGupta and J.K. Ghosh of Purdue University for helpful comments. The research of Sourav Chatterjee has been supported in part by NSF grant DMS-0072331.

References

[1] Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis.* Second edition. Wiley Series in 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. (1993a). Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.*, 21, no. 2, 625–648.

[5] Bai, Z. D. (1993b). Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21, no. 2, 649–672.

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

[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., Miao, Baiqi and Tsay, Jhishen (1997). A note on the convergence rate of the spectral distributions of large random matrices. *Statist. Probab. Lett.* 34, no. 1, 95–101.

[9] Bai, Z. D., Miao, Baiqi and Tsay, Jhishen (2002). Convergence rates of the spectral distributions of large Wigner matrices. *Int. Math. J.* 1, no. 1, 65–90.

[10] Bai, Z. D., Miao, Baiqi and Yao, Jian-Feng (2003). Convergence rates of the spectral distributions of large sample covariance matrices. *Siam J. Matrix Anal. Appl.* To appear.

[11] Beckner, W. (1989). A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.*, 105, 2, 397-400.

[12] Bobkov, S. G. (1999). Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.* 27, 4, 1903–1921

[13] Bobkov, S. G. and Götze, F. (1999). Discrete isoperimetric and Poincaré-type inequalities. *Probab. Theory Related Fields*, 114, 2, 245–277.

[14] Bobkov, Sergey G., Götze, Friedrich and Houdré, C. (2001). On Gaussian and Bernoulli covariance representations. *Bernoulli* 7 , no. 3, 439–451

[15] Bobkov, S. G. and Houdré, C. (1996). Variance of Lipschitz functions and an isoperimetric problem for a class of product measures. *Bernoulli*, 2, no. 3, 249–255.

[16] Bobkov, S. G. and Houdré, C. (1997). Isoperimetric Inequalities for Product Probability Measures, *Ann. of Probab.*, 25, 1, 184-205.

[17] Bobkov, S. and Ledoux, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* 107, no. 3, 383–400.

[18] Borovkov, A. A. and Utev, S. A. (1983). On an inequality and a related characterization of the normal distribution. *Theory Probab. Appl.* 28 219-228.

[19] Bose, Arup, Chatterjee, Sourav and Gangopadhyay, Sreela (2003). Limiting Spectral Distributions of Large Dimensional Random Matrices. *Tech. Rep.* No. 01/2003, Stat-Math Unit, Indian Statistical Institute, Kolkata.

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

[21] 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

[22] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log-concave functions, and with an application to the diffusion equation. *J. Funct. Anal.*, 22, 366-389.

[23] Bryc, W., Dembo, A. and Jiang, T. (2004). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Available at the URL* `http://arxiv.org/abs/math.PR/0307330`

[24] 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.

[25] Devroye, Luc (1991). Exponential inequalities in nonparametric estimation. In *Nonparametric functional estimation and related topics* (Spetses, 1990), 31–44, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht.

[26] Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Ann. Statist.* 9, no. 3, 586–596.

[27] Feller, William (1966). *An introduction to probability theory and its applications.* Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney.

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

[29] 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.

[30] Götze, F. and Tikhomirov, A. N. (2002). Rate of convergence to the semicircular law for the Gaussian unitary ensemble. *Teor. Veroyatnost. i Primenen* 47, no. 2, 381–387.

[31] Götze, F. and Tikhomirov, A. N. (2003). Rate of convergence to the semicircular law. *Probab. Theory Rel. Fields*, 127, 228-276.

[32] Götze, F. and Tikhomirov, A. N. (2004) Rate of convergence in probability to the Marčenko-Pastur law. To appear in *Bernoulli.*

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

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

[35] Guionnet, A. and Zeitouni, O. (2000). Concentration of spectral measure for large matrices. *Electron. Comm. Probab.*, 5, 119-136.

[36] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution Free Theory of Nonparametric Regression.* Springer Series in Statistics. Springer, New York.

[37] Horn, R. A. and Johnson, C. (1985). *Matrix Analysis*, Cambridge Univ. Press.

[38] 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.

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

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

[41] Ledoux, M. (2000). Concentration of measure and logarithmic Soboloev inequalities. In *Séminaire de Probabilités XXXV, Lecture Notes in Math.*, 1755, 120-216, Springer.

[42] 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.

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

[44] Muckenhoupt, B. (1972). Hardy's inequality with weights. *Studia Math.* 44, 31-38.

[45] Simon, B. (1979). *Trace ideals and their applications*, Cambridge Univ. Press.

[46] Steele, J. Michael (1986). An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.* 14, no. 2, 753–758.

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

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

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

[52] Yin, Y. Q. and Krishnaiah, P. R. (1985). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Teor. Veroyatnost. i Primenen.*, 30, no. 4, 810–816