

E. S. Meckes, M. W. Meckes

**A SHARP RATE OF CONVERGENCE FOR THE
EMPIRICAL SPECTRAL MEASURE OF A RANDOM
UNITARY MATRIX**

ABSTRACT. We consider the convergence of the empirical spectral measures of random $N \times N$ unitary matrices. We give upper and lower bounds showing that the Kolmogorov distance between the spectral measure and the uniform measure on the unit circle is of the order $\log N/N$, both in expectation and almost surely. This implies in particular that the convergence happens more slowly for Kolmogorov distance than for the L_1 -Kantorovich distance. The proof relies on the determinantal structure of the eigenvalue process.

Let $U \in \mathcal{U}(N)$ be a random matrix, distributed according to Haar measure. Denote the eigenvalues of U by $e^{i\theta_1}, \dots, e^{i\theta_N}$, and let μ_N denote the empirical spectral measure of U ; that is,

$$\mu_N := \frac{1}{N} \sum_{j=1}^N \delta_{e^{i\theta_j}}.$$

It is easy to see by symmetry that $\mathbb{E}\mu_N = \nu$ for every N , where ν is the uniform probability measure on the unit circle in the complex plane.

The convergence of the empirical spectral measure of a random matrix to a limiting distribution, as the size of the matrix tends to infinity, has been studied extensively for a variety of random matrix ensembles, most notably for Wigner matrices. In particular, the empirical spectral measure converges to the semicircle law in the Komogorov distance at rate $(\log N)^c/N$ (see [8]).

In the context of random unitary matrices, the convergence of μ_N to the uniform measure on the circle (weakly, in probability) was first proved in [6]. In [9] a large deviations principle was proved which in particular shows that convergence occurs with probability 1. In earlier work (see [12]), we have quantified this convergence, with respect to the L_1 -Kantorovich

Key words and phrases: random matrices; empirical spectral measures; determinantal point processes.

distance W_1 . Specifically, there are absolute constants C_1 and C_2 such that

$$\mathbb{E}W_1(\mu_N, \nu) \leq \frac{C_1 \sqrt{\log(N)}}{N}, \quad (1)$$

and, with probability 1,

$$W_1(\mu_N, \nu) \leq \frac{C_2 \sqrt{\log(N)}}{N} \quad (2)$$

for all sufficiently large N .

In this note, we consider instead the Kolmogorov distance

$$d_K(\mu_N, \nu) = \sup_{0 \leq \theta < 2\pi} \left| \frac{1}{N} \mathcal{N}_\theta - \frac{\theta}{2\pi} \right|,$$

where \mathcal{N}_θ is the number of eigenvalues $e^{i\theta_j}$ of U with $0 \leq \theta_j \leq \theta$. That is, we are interested in upper and lower bounds for the supremum of the stochastic process

$$X_\theta := \left| \frac{1}{N} \mathcal{N}_\theta - \frac{\theta}{2\pi} \right|$$

indexed by $\theta \in [0, 2\pi)$.

Theorem 1. *There are universal constants $c_1, c_2, c_3 > 0$ such that*

$$c_1 \frac{\log(N)}{N} \leq \mathbb{E}d_K(\mu_N, \nu) \leq c_2 \frac{\log(N)}{N}$$

for all N , and with probability 1,

$$d_K(\mu_N, \nu) \leq c_3 \frac{\log(N)}{N}$$

for all sufficiently large N .

After the first version of this paper was written, we were informed by Paul Bourgade of the results of [1], which in particular show convergence in probability of $d_K(\mu_N, \nu)$. Combining the results of [1] with our methods, we prove the following improvement of the first part of Theorem 1.

Theorem 2. *For every $p > 0$,*

$$\frac{N}{\log N} d_K(\mu_N, \nu) \xrightarrow{L_p} \frac{1}{\pi}$$

as $N \rightarrow \infty$.

One interesting consequence of the theorems together with the bounds (1) and (2) proven in [12] is that in this setting, the expected rate of convergence of μ_N to ν in the L_1 -Kantorovich distance is strictly faster than the expected rate of convergence in the Kolmogorov distance. This is in contrast to the setting of more classical limit theorems, for which the rates are often of the same order; e.g., for i.i.d. samples, the rate is $N^{-1/2}$ in both metrics.

While it is desirable to have results comparable to (1) and (2) for the more familiar and widely used Kolmogorov metric, the interest stems in large part from the connection between Kolmogorov bounds and maximal eigenvalue spacing; a large gap between successive eigenvalues corresponds to a large arc to which the spectral measure assigns no mass. There is great interest in the asymptotics of the maximal eigenvalue spacing for random unitary matrices, in part because of the connection to the Riemann zeta function; the distribution of the maximal eigenvalue spacings for $N \times N$ random unitary matrices are conjectured to predict the statistics of spacings between successive zeroes of the zeta function at height T along the critical line, when $N \approx \log\left(\frac{T}{2\pi}\right)$. A significant recent contribution on the maximal eigenvalue spacing was made in [2], where it was shown that if $\mathcal{T}^{(N)}$ is the maximum eigenvalue gap of a uniform $U \in \mathbb{U}(N)$, then

$$\frac{N}{\sqrt{32 \log(N)}} \mathcal{T}^{(N)} \xrightarrow{L_p} 1$$

for all $p > 0$. This implies in particular that $\mathbb{E}d_K(\mu_N, \nu) \geq \frac{c\sqrt{\log(N)}}{N}$; Theorems 1 and 2 show that the correct rate is in fact $\frac{\log(N)}{N}$.

A crucial property underpinning the proofs of the theorems is that the eigenvalue angles $\theta_1, \dots, \theta_N$ are a determinantal point process on $[0, 2\pi]$, with symmetric kernel

$$K_N(x, y) = \frac{\sin\left(\frac{N(x-y)}{2}\right)}{\sin\left(\frac{x-y}{2}\right)}$$

(see [15, Chapter 11]). In particular, the following properties of the eigenvalue counting function are consequences of the d.p.p. structure.

Proposition 3.

- (1) *Let $A \subseteq [0, 2\pi]$, and let \mathcal{N}_A denote the number of eigenvalue angles of U lying in A . Then there are independent Bernoulli random*

variables ξ_1, \dots, ξ_N such that

$$\mathcal{N}_A \stackrel{d}{=} \sum_{j=1}^N \xi_j.$$

- (2) *The eigenangle process of U is negatively associated: if $A, B \subseteq [0, 2\pi]$ are disjoint, then*

$$\mathbb{P}[\mathcal{N}_A \geq s, \mathcal{N}_B \geq t] \leq \mathbb{P}[\mathcal{N}_A \geq s] \mathbb{P}[\mathcal{N}_B \geq t].$$

The first part of Proposition 3 follows from the corresponding property for a quite general class of determinantal point process, due to Hough–Krishnapur–Peres–Virág [10, Theorem 7]. The second part is again a consequence of a more general statement about determinantal point processes, this time due to Ghosh [7, Theorem 1.4].

The representation of the counting function as a sum of independent Bernoulli random variables is a powerful tool; it opens the doors to countless results of classical probability. (For other uses of this idea in the theory of random unitary matrices, see [12, 14]; see also [4, 5, 13] for related approaches in other random matrix ensembles.) We will be particularly interested in the tail probabilities

$$\mathbb{P}[\mathcal{N}_I - \mathbb{E}\mathcal{N}_I > t],$$

for $t > 0$ and I an interval to be specified; note that by rotation invariance, this is equal to $\mathbb{P}[\mathcal{N}_\theta - \mathbb{E}\mathcal{N}_\theta > t]$, where θ is the length of I . In the classical setting of a sum of independent random variables, an upper bound on such tail probabilities is given by Bernstein’s inequality (see e.g. [16, Lemma 4.3.4]), while a lower bound was proved by Kolmogorov (see [11, Hilfssatz IV]).

Proposition 4. *Let X_1, \dots, X_n be independent random variables, with $|X_j| \leq M$ almost surely, for each j . Let*

$$S_n := \sum_{j=1}^n X_j \quad s_n^2 = \text{Var}(S_n).$$

Then

- (1) *for all $x > 0$,*

$$\mathbb{P}[S_n - \mathbb{E}S_n > xs_n] \leq \exp\left(-\min\left\{\frac{x^2}{4}, \frac{xs_n}{2M}\right\}\right),$$

and

(2) if $x \geq 512$ and $a := \frac{xM}{s_n} \leq \frac{1}{256}$, then for

$$\varepsilon = \max \left\{ 64\sqrt{a}, \frac{32\sqrt{\log(x^2)}}{x} \right\},$$

$$\mathbb{P} [S_n - \mathbb{E}S_n > xs_n] \geq e^{-\frac{x^2}{2}(1+\varepsilon)}.$$

By part 3 of Proposition 3, the conclusions of Proposition 4 apply to the counting functions \mathcal{N}_I with $M = 1$; for them to give usable estimates, formulae (or at least asymptotics) for the means and variances of the counting functions are needed. The mean is trivial to compute by symmetry. Rather precise asymptotics can be determined for the variance, as a further application of the determinantal point process structure of the ensemble of eigenvalues. The estimates in the following lemma were proved in [12, 14].

Lemma 5.

(1) For $\theta \in [0, 2\pi]$,

$$\mathbb{E}\mathcal{N}_\theta = \frac{N\theta}{2\pi}.$$

(2) For $\theta \in [0, 2\pi]$,

$$\text{Var } \mathcal{N}_\theta \leq \log(eN).$$

(3) If $\frac{3\pi}{2N} \leq \theta \leq \frac{\pi}{2}$,

$$\frac{1}{3\pi^2} \log \left(\frac{2N\theta}{3\pi} \right) \leq \text{Var } \mathcal{N}_\theta \leq \frac{1}{2} \log(e^{3/2}N\theta).$$

With these ingredients in place, we now turn to upper and lower bounds on $\mathbb{E}d_K(\mu_n, \nu)$.

Proof of Theorem 1. We consider the upper bounds first. If $\frac{2\pi k}{N} \leq \theta < \frac{2\pi(k+1)}{N}$, then

$$\mathcal{N}_\theta - \frac{N\theta}{2\pi} \leq \mathcal{N}_{\frac{2\pi(k+1)}{N}} - (k+1) + 1$$

and

$$\mathcal{N}_\theta - \frac{N\theta}{2\pi} \geq \mathcal{N}_{\frac{2\pi k}{N}} - k - 1,$$

so that

$$d_K(\mu_N, \nu) \leq \frac{1}{N} \sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| + \frac{1}{N}. \quad (3)$$

As discussed above, Proposition 4 can be applied to the counting function $\mathcal{N}_{\frac{2\pi k}{N}}$. Part 1 of Proposition 4 and part 2 of Lemma 5 imply that

$$\begin{aligned} \mathbb{P} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| > x \right] &\leq \sum_{k=1}^N \mathbb{P} \left[\left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| > x \right] \\ &\leq 2 \sum_{k=1}^N \exp \left(- \min \left\{ \frac{x^2}{4 \operatorname{Var}(\mathcal{N}_{\frac{2\pi k}{N}})}, \frac{x}{2} \right\} \right) \quad (4) \\ &\leq 2N \exp \left(- \min \left\{ \frac{x^2}{4 \log(eN)}, \frac{x}{2} \right\} \right). \end{aligned}$$

Now, since $\left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| \leq N$ for all k , it follows from the estimate above that for any $x > 0$,

$$\begin{aligned} \mathbb{E} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| \right] &\leq x + N \mathbb{P} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| > x \right] \\ &\leq x + 2N^2 \exp \left(- \min \left\{ \frac{x^2}{4 \log(eN)}, \frac{x}{2} \right\} \right). \end{aligned}$$

Setting $x = 4 \log(eN)$, this implies

$$\mathbb{E} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| \right] \leq 4 \log(eN) + \frac{2}{e^2}.$$

The claimed upper bound on $\mathbb{E}d_K(\mu_N, \nu)$ now follows from (3).

Setting $x = 6 \log(eN)$ in (4) yields

$$\mathbb{P} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| > 6 \log(eN) \right] \leq \frac{2}{e^3 N^2}.$$

The almost sure rate of convergence now follows from (3) and the Borel–Cantelli lemma.

For the lower bound, note first that given probability measures μ and ν on $[0, 2\pi)$,

$$d_K(\mu, \nu) \leq \sup_{0 \leq a \leq b < 2\pi} \left| \mu((a, b]) - \nu((a, b]) \right| \leq 2d_K(\mu, \nu).$$

Let \mathcal{I}_N be a collection of T disjoint subintervals of $[0, 2\pi)$, each of length $N^{-1/2}$; in particular, $T \leq 2\pi\sqrt{N}$. Then by the Bonferroni inequalities,

$$\begin{aligned} \mathbb{P} [d_K(\mu_N, \nu) > x] &\geq \mathbb{P} \left[\sup_{I \in \mathcal{I}_N} (\mu_N(I) - \nu(I)) > 2x \right] \\ &\geq \sum_{I \in \mathcal{I}_N} \mathbb{P} [(\mu_N(I) - \nu(I)) > 2x] \\ &\quad - \frac{1}{2} \sum_{\substack{I, J \in \mathcal{I}_N \\ I \neq J}} \mathbb{P} [(\mu_N(I) - \nu(I)), (\mu_N(J) - \nu(J)) > 2x] \\ &\geq \sum_{I \in \mathcal{I}_N} \mathbb{P} [(\mu_N(I) - \nu(I)) > 2x] \\ &\quad - \frac{1}{2} \sum_{\substack{I, J \in \mathcal{I}_N \\ I \neq J}} \mathbb{P} [(\mu_N(I) - \nu(I)) > 2x] \mathbb{P} [(\mu_N(J) - \nu(J)) > 2x], \end{aligned}$$

where the last estimate follows from the negative association property of part 3 of Proposition 3. Since all of the intervals I have the same length, it follows from the rotation-invariance of both measures that this last expression is exactly

$$TP - \frac{T(T-1)}{2} P^2 \geq \frac{1}{2} TP(2 - TP),$$

where P is the common value of $\mathbb{P} [(\mu_N(I) - \nu(I)) > 2x]$ for $I \in \mathcal{I}_N$. It follows that if x and T can be chosen such that $TP \in [\frac{1}{2}, \frac{3}{2}]$, then

$$\mathbb{P} [d_K(\mu_N, \nu) > x] \geq \frac{3}{8},$$

and therefore

$$\mathbb{E} d_K(\mu_N, \nu) \geq \frac{3}{8} x. \tag{5}$$

Since each I has length $N^{-1/2}$, we have that for sufficiently large N ,

$$\begin{aligned} P &= \mathbb{P} \left[\mu_N(I) - \nu(I) > \frac{\text{Var}(\mathcal{N}_I)}{256N} \right] = \mathbb{P} \left[\mathcal{N}_I - \frac{|I|}{2\pi} > \frac{\text{Var}(\mathcal{N}_I)}{256} \right] \\ &\geq \exp \left(-\frac{5}{2^{17}} \text{Var}(\mathcal{N}_I) \right) \\ &\geq \exp \left(-\frac{5}{2^{19}} \log(e^3 N) \right); \end{aligned}$$

the first estimate follows from part 2 of Proposition 4 and the second follows from part 3 of Lemma 5 with $\theta = N^{-1/2}$. It follows that for all sufficiently large N , $N^{1/2}P \geq 2$, and therefore $TP \in [\frac{1}{2}, \frac{3}{2}]$ for some integer $1 \leq T \leq N^{1/2}$. Then by (5) and Lemma 5,

$$\mathbb{E}d_K(\mu_N, \nu) \geq \frac{3 \operatorname{Var}(\mathcal{N}_{N^{-1/2}})}{2^{12}N} \geq c \frac{\log(N)}{N}$$

for all N large enough. □

Proof of Theorem 2. In [1], the authors state that

$$\frac{1}{\log N} \sup_{0 \leq \pi < 2\pi} \left(\mathcal{N}_\theta - \frac{N\theta}{2\pi} \right) \rightarrow \frac{1}{\pi}$$

in probability. It can similarly be shown [3] that the corresponding infimum converges in probability to $-\frac{1}{\pi}$, from which it follows that

$$\frac{N}{\log N} d_K(\mu_N, \nu) \rightarrow \frac{1}{\pi}$$

in probability.

For a fixed $\varepsilon > 0$,

$$\begin{aligned} & \mathbb{E} \left| \frac{N}{\log N} d_K(\mu_N, \nu) - \frac{1}{\pi} \right|^p \\ & \leq \varepsilon^p + \mathbb{E} \left| \frac{N}{\log N} d_K(\mu_N, \nu) - \frac{1}{\pi} \right|^p \mathbb{1}_{\left| \frac{N}{\log N} d_K(\mu_N, \nu) - \frac{1}{\pi} \right| > \varepsilon} \\ & \leq \varepsilon^p + \sqrt{\mathbb{E} \left| \frac{N}{\log N} d_K(\mu_N, \nu) - \frac{1}{\pi} \right|^{2p}} \sqrt{\mathbb{P} \left[\left| \frac{N}{\log N} d_K(\mu_N, \nu) - \frac{1}{\pi} \right| > \varepsilon \right]}. \end{aligned}$$

The theorem thus follows from the convergence in probability of $\frac{N}{\log N} d_K(\mu_N, \nu)$, if we can show that the sequence of random variables $\frac{N}{\log N} d_K(\mu_N, \nu)$ is bounded in L_{2p} .

Now, for $x > 0$, it follows from (4) that

$$\mathbb{E} \left[\sup_{1 \leq k \leq N} \left| \mathcal{N}_{\frac{2\pi k}{N}} - k \right| \right]^{2p} \leq x^{2p} + 2N^{2p+1} \exp \left(- \min \left\{ \frac{x^2}{4 \log(eN)}, \frac{x}{2} \right\} \right).$$

Choosing x to be a sufficiently large multiple of $\log N$ we obtain

$$\mathbb{E} \left| \sup_{1 \leq k \leq N} \mathcal{N}_{\frac{2\pi k}{N}} - k \right|^{2p} \leq C_p (\log N)^{2p}$$

for some constant $C_p > 0$ depending only on p ; together with (3) this implies that

$$\mathbb{E} \left[\frac{N}{\log N} d_K(\mu_N, \nu) \right]^{2p} \leq C'_p. \quad \square$$

REFERENCES

1. L. P. Arguin, D. Belius, P. Bourgade, *Maximum of the characteristic polynomial of random unitary matrices*, — Comm. Math. Phys, **349** (2017), 703–751.
2. G. Ben Arous, P. Bourgade, *Extreme gaps between eigenvalues of random matrices*, — Ann. Probab, **41**(4) (2013), 2648–2681.
3. P. Bourgade, *Personal communication*.
4. S. Dallaporta, *Eigenvalue variance bounds for Wigner and covariance random matrices*, — Random Matrices Theory Appl., **1**(3), 28, 1250007 (2012).
5. S. Dallaporta, *Eigenvalue variance bounds for covariance matrices*, — Markov Process. Related Fields, **21**(1) (2015), 145–175.
6. P. Diaconis, M. Shahshahani, *On the eigenvalues of random matrices*, — J. Appl. Probab., **31A** Studies in Applied Probability (1994), 49–62.
7. S. Ghosh, *Determinantal processes and completeness of random exponentials: the critical case*, — Probab. Theory Related Fields, **163**(3-4) (2015), 643–665.
8. F. Goötze, A. A. Naumov, A. N. Tikhomirov, *Local semicircle law under moment conditions: Stieltjes transform, rigidity and delocalization*, — Teor. Veroyat. Primenen., **62**(1), (2017) 72–103.
9. F. Hiai, D. Petz, *A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices*, — Ann. Inst. H. Poincaré Probab. Statist., **36**(1) (2000), 71–85.
10. J. B. Hough, M. Krishnapur, Y. Peres, B. Virág, *Determinantal processes and independence*, — Probab. Surv., **3** (2006), 206–229.
11. A. Kolmogoroff, *Über das Gesetz des iterierten Logarithmus*, — Math. Ann., **101**(1) (1929), 126–135.
12. E. S. Meckes, M. W. Meckes, *Spectral measures of powers of random matrices*, Electron. Commun. Probab., **18**, No. 78 (2013), 13.
13. E. S. Meckes, M. W. Meckes, *A rate of convergence for the circular law for the complex Ginibre ensemble*, — Ann. Fac. Sci. Toulouse Math. (6), **24**(1) (2015), 93–117.
14. E. S. Meckes, M. W. Meckes, *Self-similarity in the circular unitary ensemble*, — Discrete Anal., Paper No. 9 (2016), 15 pages.
15. M. L. Mehta, Random Matrices, volume 142 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, third edition, (2004).

16. M. Talagrand, *Upper and Lower Bounds for Stochastic Processes. Modern Methods and Classical Problems*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Heidelberg, (2014).

Department of Mathematics,
Applied Mathematics, and Statistics,
Case Western Reserve University,
10900 Euclid Ave., Cleveland, Ohio 44106, U.S.A.
E-mail: `elizabeth.meckes@case.edu`
E-mail: `mark.meckes@case.edu`

Поступило 4 августа, 2017 г.