Circular Law and Arc Law for Truncation of Random Unitary Matrix

Zhishan Dong[1], Tiefeng Jiang[2] and Danning Li[3]
University of Minnesota

Abstract

Let $\mathbf{V}$ be the $m \times m$ upper-left corner of an $n \times n$ Haar-invariant unitary matrix. Let $\lambda_1, \cdots, \lambda_m$ be the eigenvalues of $\mathbf{V}$. We prove that the empirical distribution of a normalization of $\lambda_1, \cdots, \lambda_m$ goes to the circular law, that is, the uniform distribution on $\{z \in \mathbb{C}; |z| \leq 1\}$ as $m \to \infty$ with $m/n \to 0$. We also prove that the empirical distribution of $\lambda_1, \cdots, \lambda_m$ goes to the arc law, that is, the uniform distribution on $\{z \in \mathbb{C}; |z| = 1\}$ as $m/n \to 1$. These explain two observations by Życzkowski and Sommers (2000).

Keywords: random matrix, circular law, arc law, Haar-invariant unitary matrix, truncation of matrix, empirical distribution.

AMS 2000 Subject Classification: Primary 60B20, 60B15; secondary 60F05, 62H10, 62E20.

[1]School of Statistics, University of Minnesota, 224 Church Street, S.E., MN55455, dongzs@stat.umn.edu.

[2]School of Statistics, University of Minnesota, 224 Church Street, S.E., MN55455, tjiang@stat.umn.edu.
The research of Tiefeng Jiang was supported in part by NSF FRG Grant DMS-0449365.
[3]School of Statistics, University of Minnesota, 224 Church Street, S.E., MN55455, lixx0700@stat.umn.edu.

1 Introduction

Random unitary matrices (Dyson, 1962) are used in many physical applications such as chaotic scattering and conductance in mesoscopic system (Beenakker, 1997) and statistical properties of periodically driven quantum systems (Haake, 2010). In applications, people often consider truncations of large dimensional Haar unitary matrices, which are used to describe quantum systems with absorbing boundaries (Casati, Maspero and Shepelyansky, 1999). Applications of such truncated matrices are found in optical and semiconductor superlattices (Glück, Kolovsky and Korsch, 2002), problems of quantum conductance (Forrester, 2006), distribution of resonances for open quantum maps (Fyodorov and Sommers, 2003; Schomerus and Jacquod, 2005; Nonnenmacher and Zworski, 2007; Pedrosa, Carlo, Wisniacki and Ermann, 2009).

On the other hand, truncations of Haar-invariant unitary and orthogonal matrices have application in statistics. In particular, the singular values of such a truncation have the same distribution as those of a Jacobi matrix (Collins, 2005). In literature, a Jacobi matrix is also called a MANOVA matrix, which has been used extensively in the multivariate analysis from the field of Statistics. Some applications of the truncations of Haar-invariant matrices can be found in, for example, Eaton (1989), Diaconis, Eaton and Lauritzen (1992) and Jiang (2009).

Certain theoretical work on truncations of Haar invariant matrices were carried out by several authors. For example, Życzkowski and Sommers (2000) derived the distributions of the absolute values of the eigenvalues and simulated the empirical distributions of the eigenvalues; Petz and Réffy (2004, 2005) analyzed the entries and proved a large deviation principle for the eigenvalues; Jiang (2006, 2009) used independent normal random variables to approximate the entries; Khoruzhenko, Sommers and Życzkowski (2010) studied the distribution of the eigenvalues of a block of an Haar orthogonal matrices, and Forrester (2010) further studied the same problem and deduced a formula for the zeros of the Kac random polynomial.

In this paper, we shall study the spectra of a truncated block of an Haar unitary matrix. To be precise, for each $n \geq 2$, let $m = m_n < n$ be a positive integer. Let $\{\mathbf{U}_n;\, n \geq 1\}$ be a sequence of Haar-invariant unitary matrices defined on the same probability space, where $\mathbf{U}_n$ is an $n \times n$ matrix for each $n \geq 1$. Denote by $\mathbf{U}_{[n,m]}$ the $m \times m$ upper-left corner of $\mathbf{U}_n$, and by $\lambda_1, \cdots, \lambda_m$ the eigenvalues of $\mathbf{U}_{[n,m]}$. Since the spectral norm of $\mathbf{U}_{[n,m]}$ is bounded by that of $\mathbf{U}_n$, which is one, we know that $\max_{1\leq i\leq m} |\lambda_i| \leq 1$. Define

$$\mu_m := \frac{1}{m}\sum_{i=1}^{m} \delta_{\lambda_i}.$$

Besides their deep theoretical results, Życzkowski and Sommers (2000) made several simulations about μ_m. They found that the behavior of μ_m is decided by the ratio m/n.

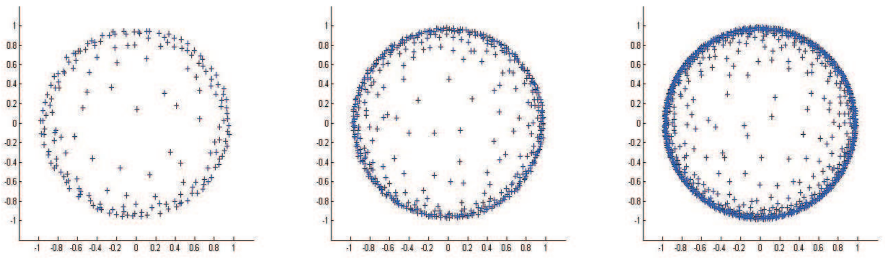

Figure 1: *Arc Law as in Theorem 1 for $U_{[n,m]}$, where the first is $U_{[200,185]}$, the second is $U_{[500,477]}$, the third is $U_{[1000,968]}$.*

In fact, taking $n = 5$, $m = 2$ and $m = 4$, respectively, and independently generating the m-dimensional vector $(\lambda_1, \cdots, \lambda_m)$ many times, they observed that

> *"For $U_{[5,4]}$, there exist several eigenvalues close to the unit circle, while for stronger truncation $U_{[5,2]}$ the eigenvalues are clustered closer to the origin."* (1)

Through the study of the large deviations for the eigenvalues $\lambda_1, \cdots, \lambda_m$, Petz and Réffy (2005) obtained the following nice result: *If $m/n \to \alpha \in (0, 1)$, then μ_m converges weakly to the distribution ν_0 with the probability density function*

$$d\nu_0 = \frac{(\alpha^{-1} - 1)r}{\pi(1 - r^2)^2} dr\, d\varphi, \qquad z = re^{i\varphi} \tag{2}$$

on $\{z \in \mathbb{C};\ |z| \le \sqrt{\alpha}\,\}$. The density, as a function defined on the complex plane, does not depend on the angle φ. Therefore it is rotation-invariant.

In this paper we will study μ_m for the other two cases: $\alpha = 1$ and $\alpha = 0$.

THEOREM 1 (Arc Law) *Assume $\lim_{n\to\infty} m/n = 1$, then, with probability one,*

$$\mu_m := \frac{1}{m}\sum_{i=1}^{m} \delta_{\lambda_i}$$

converges weakly to the uniform distribution on $\{z \in \mathbb{C};\ |z| = 1\}$.

Recall the first part of the observations in (1). Thinking that $m = 4$ is "very close to" $n = 5$, Theorem 1 says that not only "several eigenvalues close to the unit circle", there are $m - o(m)$ of the eigenvalues close to the circle evenly, that is, they are uniformly distributed on the unit circle in the limiting sense. See Figure 1.

Let $\mathbf{X}_n = (x_{ij})$ be an $n \times n$ Haar-invariant orthogonal or unitary matrix, that is, $\mathbf{X}_n$ generates the Haar probability measure on the classical compact group $O(n)$ or $U(n)$. It is

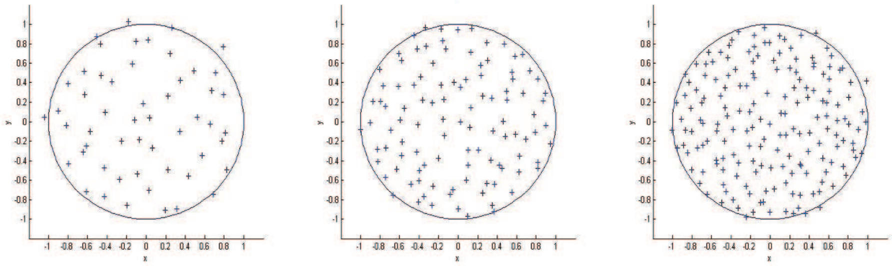

Figure 2: *Circular Law as in Theorem 2 for $U_{[n,m]}$, where the first is $U_{[200,53]}$, the second is $U_{[500,105]}$, the third is $U_{[1000,177]}$.*

known that the empirical distribution of the eigenvalues of $\mathbf{X}_n$ converges to the arc law as $n \to \infty$. See, for example, Diaconis and Shahshahani (1994) and Diaconis and Evans (2001) in this direction. Theorem 1 says that the arc law still holds if the size of a truncated block of an Haar unitary matrix is large enough.

To state the result for $\alpha = 0$, we need to review some terminology. Denote by $\mathcal{M}(\mathbb{C})$ the Borel probability measures defined on the complex plane $\mathbb{C}$. Let μ and ν be two probability measures on $\mathbb{C}$ (or $\mathbb{R}^2$). Define

$$\rho(\mu, \nu) \quad = \quad \sup_{\|f\|_L \leq 1} \left| \int_{\mathbb{C}} f(x)\, \mu(dx) - \int_{\mathbb{C}} f(x)\, \nu(dx) \right|, \tag{3}$$

where f is a bounded Lipschitz function defined on $\mathbb{C}$ with $\|f\| = \sup_{x \in \mathbb{C}} |f(x)|$, and $\|f\|_L = \|f\| + \sup_{x \neq y} |f(x) - f(y)|/|x-y|$. This metric generates the topology of the weak convergence of the probability measures on $\mathbb{C}$ (see, e.g., chapter 11 from Dudley (2002)), that is, μ_n converges to μ weakly as $n \to \infty$ if and only if $\lim_{n\to\infty} \int f(x)\, \mu_n(dx) = \int f(x)\, \mu(dx)$ for every bounded and continuous function $f(x)$ defined on $\mathbb{C}$, and if and only if $\rho(\mu_n, \mu) \to 0$ as $n \to \infty$. We endow $\mathcal{M}(\mathbb{C})$ with the standard weak convergence topology, which is, as explained, the same one generated by $\rho(\mu, \nu)$ in (3).

THEOREM 2 (Circular Law) *Suppose $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Let μ_0 be the uniform probability measure on $\{z \in \mathbb{C};\ |z| \leq 1\}$ and*

$$\mu_m := \frac{1}{m} \sum_{i=1}^{m} \delta_{\sqrt{\frac{n}{m}}\lambda_i}, \;\; n \geq 1.$$

Then,
(i) $\rho(\mu_m, \mu_0) \to 0$ in probability as $n \to \infty$;
(ii) With probability one, μ_m converges weakly to μ_0 as $n \to \infty$ if $\lim_{n\to\infty} m_n/\sqrt{\log n} = \infty$.

In view of the second part of the observation in (1), thinking that $m = 2$ "is of a small portion" of $n = 5$, we indeed see from Theorem 2 that "the eigenvalues are clustered closer to the origin." Further, the theorem tells us that by magnifying the eigenvalues with $\sqrt{\frac{n}{m}}$ times, they are asymptotically distributed on the unit disc uniformly. See Figure 2.

Now we make some comments about Theorem 2.

First, by measuring the variation distance between the entries of $\mathbf{U}_{[n,m]}$ and m^2 independent and identically distributed complex Gaussian random variables, Jiang (2009) showed that (i) of Theorem 2 holds when $m \to \infty$ and $m = o(\sqrt{n})$ as $n \to \infty$. Theorem 2 improves the order to $m = o(n)$. In fact, the order $m = o(n)$ is the best one to make the circular law hold since a different limit law in (2) appears when $m/n \to \alpha \in (0, 1)$.

Second, let $\mathbf{Y}_n = (y_{ij})$ be an $n \times n$ matrix where y_{ij}'s are independent and identically distributed random variables. Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of $\mathbf{Y}_n$. The circular law problem, that is, the empirical distribution of $\lambda_1/\sqrt{n}, \cdots, \lambda_n/\sqrt{n}$ goes to the uniform distribution on $\{z \in \mathbb{C};\ |z| \leq 1\}$ as $n \to \infty$, has been studied by some authors including, for example, Girko (1984a, 1984b), Bai (1997), Pan and Wang (2010), Tao and Vu (2008, 2009, 2010). Theorem 2 presents a circular law when the entries of the matrix, unlike those of $\mathbf{Y}_n$, are dependent random variables.

Finally, based on the ratio $m/n \to \alpha$ for $\alpha = 0$, $\alpha \in (0, 1)$ and $\alpha = 1$, Theorem 1, (2) and Theorem 2 give a complete characterization of the empirical distributions of $\lambda_1, \cdots, \lambda_m$ as $n \to \infty$.

The proofs of Theorem 1 ($\alpha = 1$) and Theorem 2 ($\alpha = 0$) are different from the proof of (2) ($\alpha \in (0, 1)$) by Hiai and Petz (1998). The limiting law in (2) is proved by large deviations. Now we state our methodology.

Theorem 1 is proved by using a polynomial method. Let $g(z) = \det(z\mathbf{I}_m - \mathbf{U}_{[n,m]}) = z^m + b_{m-1}z^{m-1} + \cdots + b_0$ be the characteristic polynomial of $\mathbf{U}_{[n,m]}$. We first study the coefficients b_i's by using the symmetric functions associated with the partitions of integers, an identity from Diaconis and Evans (2001), the Selberg integral and the Jacobi ensembles. Based on the work of Erdös-Turán (1950), Granville (2007) and Hughes and Nikeghbali (2008) give the convergence of the empirical distribution of the roots of a polynomial converging to the uniform distribution on $\{z \in \mathbb{C};\ |z| = 1\}$. The two steps are then combined to complete the proof.

The main focus of the proof of Theorem 2 is to estimate $P(\rho(\mu_m, \mu_0) \geq \epsilon)$. By using an inequality we first transform the original problem to that on the Ginibre ensemble. An important ingredient of the estimate is a large deviation for the eigenvalues of the Ginibre ensemble by Hiai and Petz (1998).

Remark 1.1 *In Theorems 1 and 2, we consider blocks of Haar unitary matrices. Similar results should also be true for Haar orthogonal matrices. In fact, Theorem 5 from Jiang (2009) shows that Theorem 2 still holds if* $\mathbf{U}_n$ *is replaced by an Haar orthogonal matrix and* $m = o(\sqrt{n})$*. Further, Khoruzhenko, Sommer and Życzkowski (2010) and Forrester (2011) studied the density function of the eigenvalues of a truncated block of an Haar orthogonal* *matrix, which is not as explicit or clean as the case for an Haar unitary matrix. We leave this as a future work.*

The rest of the paper is organized as follows: the proof of Theorem 1 is given in Section 2.1; the proof of Theorem 2 is given in Section 2.2.

2 Proofs

From now on, for brevity of notation, we write $m = m_n$ if there is no confusion.

2.1 The Proof of Theorem 1

Let $\mathbf{U}_n$ be an $n \times n$ Haar-invariant unitary matrix, that is, the entries of the unitary matrix $\mathbf{U}_n$ are random variables having the same joint distribution as that of $\mathbf{V}\mathbf{U}_n$ and that of $\mathbf{U}_n\mathbf{V}$ for any $n \times n$ unitary matrix $\mathbf{V}$.

Lemma 2.1 *(Theorem 2.1 from Diaconis and Evans (2001)) Consider $a = (a_1, \cdots, a_k)$ and $b = (b_1, \cdots, b_k)$ with $a_j, b_j \in \{0, 1, 2, \cdots\}$. Then for $n \geq \sum_{j=1}^k j a_j \vee \sum_{j=1}^k j b_j$,*

$$\mathbb{E}\Big[\prod_{j=1}^{k} (\mathrm{Tr}(\mathbf{U}_n^j))^{a_j} \overline{(\mathrm{Tr}(\mathbf{U}_n^j))^{b_j}}\Big] = \delta_{ab} \prod_{j=1}^{k} j^{a_j} a_j! \tag{4}$$

where δ_{ab} is Kronecker's delta.

See also Diaconis and Shahshahani (1994) for this. The following well-known formula can be found in many places, e.g., Mehta (1991) and Forrester (2007).

Lemma 2.2 *(Selberg integral) Let $N \geq 2$ be an integer, and α, β and γ be positive numbers. Then*

$$\int_0^1 \int_0^1 \cdots \int_0^1 \prod_{i=1}^{N} x_i^{\alpha-1}(1-x_i)^{\beta-1} \cdot \prod_{1 \leq j < k \leq N} |x_j - x_k|^{2\gamma}\, dx_1\, dx_2 \cdots dx_N$$
$$= \prod_{l=0}^{N-1} \frac{\Gamma(1+\gamma+l\gamma)\Gamma(\alpha+l\gamma)\Gamma(\beta+l\gamma)}{\Gamma(1+\gamma)\Gamma(\alpha+\beta+(N+l-1)\gamma)}.$$

Let $f_d(x) = a_0 + a_1 x + \cdots + a_d x^d$ where $a_0, \cdots, a_d$ are complex numbers with $a_0 a_d \neq 0$. Define

$$L(f) = \frac{1}{\sqrt{|a_0 a_d|}} \sum_{j=0}^{d} |a_j|. \tag{5}$$

The following result comes from Theorem 1.3 from Granville (2007), which is an improvement of Erdös-Turán (1950). See also Theorem 7 from Hughes and Nikeghbali (2008) in the same spirit.

Lemma 2.3 *Suppose $f_1, f_2, \cdots$ is a sequence of polynomials of complex coefficients where f_d has degree d and $f_d(0) \neq 0$. Let $\lambda_1, \cdots, \lambda_d$ be the roots of f_d and*

$$\mu_d = \frac{1}{d} \sum_{j=1}^{d} \delta_{\lambda_j}$$

be the empirical measure for $d \geq 1$. If $\frac{1}{d} \log L(f_d) \to 0$ as $d \to \infty$, then μ_d converges weakly to the uniform distribution on $\{z \in \mathbb{C}; |z| = 1\}$.

From Lemma 2.3, it is trivial to see that the empirical distribution of the roots of a Kac polynomial (the coefficients of the polynomial are independent standard normals) converges weakly to the uniform distribution on $\{z \in \mathbb{C}; |z| = 1\}$ as the degree of the polynomial goes to infinity. The same is also true for a Littlewood polynomial (the coefficients of the polynomial are independent symmetric Bernoulli random variables).

Lemma 2.4 *Let $\Gamma(x)$ be the gamma function and a be a real number. Then*

$$\lim_{x \to +\infty} \frac{\Gamma(x+a)}{x^a \Gamma(x)} = 1.$$

Proof. By the Stirling formula (see, Gamelin (2001) or from Ahlfors (1979)),

$$\log \Gamma(z) = z \log z - z - \frac{1}{2} \log z + \log \sqrt{2\pi} + \frac{1}{12z} + O\left(\frac{1}{x^3}\right)$$

as $x = \mathrm{Re}\,(z) \to +\infty$. Then

$$\log \frac{\Gamma(x+a)}{\Gamma(x)} = (x+a)\log(x+a) - x \log x - a - \frac{1}{2}\left(\log(x+a) - \log x\right) + O\left(\frac{1}{x}\right) \qquad (6)$$

as $x \to \infty$. First, use the fact that $\log(1+t) \sim t + O(t^2)$ as $t \to 0$ to get

$$(x+a)\log(x+a) - x \log x = a \log x + a + O\left(\frac{1}{x}\right)$$

as $x \to \infty$. Evidently, $\log(x+a) - \log x = O(1/x)$ as $x \to +\infty$. Thus, we have from (6)

$$\log \frac{\Gamma(x+a)}{\Gamma(x)} = a \log x + O\left(\frac{1}{x}\right)$$

as $x \to \infty$. The proof is complete. ■

For $n \geq 2$ and variables $x_1, \cdots, x_n$, recall the elementary symmetric function and power symmetric function as follows.

$$e_k = e_k(x_1, \cdots, x_n) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}; \qquad (7)$$

$$p_k = p_k(x_1, \cdots, x_n) = \sum_{i=1}^{n} x_i^k \quad \text{and} \quad p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots \qquad (8)$$

for any partition $\lambda = (\lambda_1, \lambda_2, \cdots)$ with $e_0 = p_0 = 1$.

Lemma 2.5 *Let the characteristic polynomial of an $n \times n$ Haar invariant unitary matrix $\mathbf{U}_n$ be denoted by*

$$f_n(z) = det(z\mathbf{I}_n - \mathbf{U}_n) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$$

for all $z \in \mathbb{C}$. Then $E(|a_k|^2) = 1$ for $0 \le k \le n-1$.

Proof. First, let the eigenvalues of $\mathbf{U}_n$ be $z_1, \cdots, z_n$. Then we know

$$a_{n-k} = (-1)^k e_k(z_1, \cdots, z_n) \tag{9}$$

for $1 \le k \le n$. Second, we know the basic formula between e_k and p_λ:

$$e_k = \sum_{\lambda \vdash k} \epsilon_\lambda z_\lambda^{-1} p_\lambda \tag{10}$$

where the sum runs over all partition $\lambda = (1^{m_1} 2^{m_2} \cdots)$ of k, $\epsilon_\lambda = (-1)^{k-l(\lambda)}$ with $l(\lambda) = m_1 + m_2 + \cdots$ being the length of the partition λ and

$$z_\lambda = \prod_{j \ge 1} j^{m_j} m_j!.$$

See Proposition 7.7.6 from Stanley (2001) or (2.14′) from Macdonald (1998). From (9) and (10) we see that

$$E(|a_{n-k}|^2) = E(|e_k|^2) = \sum_{\lambda \vdash k,\, \rho \vdash k} \epsilon_\lambda \epsilon_\rho z_\lambda^{-1} z_\rho^{-1} E\left(p_\lambda \overline{p}_\rho\right) \tag{11}$$

for all $1 \le k \le n$. Notice $p_k = p_k(z_1, \cdots, z_n) = tr(\mathbf{U}_n^k)$. Thus, for $\lambda = (1^{m_1} 2^{m_2} \cdots)$, we have

$$p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots = \prod_{j \ge 1} \left(\mathrm{tr}(\mathbf{U}_n^j)\right)^{m_j}.$$

It then follows from (4) that

$$E(|p_\lambda|^2) = z_\lambda \quad \text{and} \quad E\left(p_\lambda \overline{p}_\rho\right) = 0 \text{ for } \lambda \ne \rho.$$

By (11) we have

$$E(|a_k|^2) = \sum_{\lambda \vdash k} \frac{1}{z_\lambda} = 1$$

for $k = 0, 1, \cdots, n-1$, where the last identity holds since the conjugacy class associated with λ in the symmetric group S_k has $k!/z_\lambda$ permutations, so that $\sum_{\lambda \vdash k} k!/z_\lambda = k!$. ■

Lemma 2.6 *For $1 \le m \le n$, let $\mathbf{U}_n$ be an $n \times n$ Haar unitary matrix and $\mathbf{U}_{[n,m]}$ be the $m \times m$ upper-left corner of $\mathbf{U}_n$. Set $g(z) = det(z\mathbf{I}_m - \mathbf{U}_{[n,m]}) = z^m + b_{m-1}z^{m-1} + \cdots + b_0$. Then*

$$E(|b_{m-k}|^2) = \frac{\binom{m}{k}}{\binom{n}{k}}$$

for all $k = 1, 2 \cdots, m$.

Proof. It is known that

$$(-1)^k b_{m-k} = \sum_{1 \le j_1 < \cdots < j_k \le m} \det(\mathbf{U}_{j_1, \cdots, j_k}) \tag{12}$$

for any $1 \le k \le m$, where $\mathbf{U}_{j_1,\cdots,j_k}$ is the $k \times k$ principal minor of $\mathbf{U}_{[n,m]}$, hence the $k \times k$ principal minor of $\mathbf{U}_n$ formed by rows $j_1, \cdots, j_k$ and columns $j_1, \cdots, j_k$ of $\mathbf{U}_n$. By the Haar invariant property, exchanging any two different rows or/and exchanging two different columns do not change the joint distribution of the entries. Thus, $\det(\mathbf{U}_{j_1,\cdots,j_k})$ and $\det(\mathbf{U}_{1,\cdots,k})$ have the same probability distribution. It follows that

$$E\Big(|\det(\mathbf{U}_{j_1,\cdots,j_k})|^2\Big) = E\Big(|\det(\mathbf{U}_{1,\cdots,k})|^2\Big) \tag{13}$$

for any $1 \le j_1 < \cdots < j_k \le m$. On the other hand, any two different $k \times k$ minors $\det(\mathbf{U}_{j_1,\cdots,j_k})$ and $\det(\mathbf{U}_{l_1,\cdots,l_k})$ are uncorrelated, that is,

$$E\Big(\det(\mathbf{U}_{j_1,\cdots,j_k}) \cdot \overline{\det(\mathbf{U}_{l_1,\cdots,l_k})}\Big) = 0 \tag{14}$$

for any $1 \le j_1 < \cdots < j_k \le m$ and $1 \le l_1 < \cdots < l_k \le m$ with $\{j_1, \cdots, j_k\} \ne \{l_1, \cdots, l_k\}$. In fact, suppose $j_s \notin \{l_1, \cdots, l_k\}$ for some $s \in \{1, \cdots, k\}$. For any $\theta \in \mathbb{R}$, set $\mathbf{A} = \mathrm{diag}(1, 1, \cdots, 1, e^{i\theta}, 1, \cdots, 1)$ where $e^{i\theta}$ appears at the s-th position. Then, by the Haar invariance, $\mathbf{A}\mathbf{U}_n$ and $\mathbf{U}_n$ have the same distribution. Notice the transform $\mathbf{A}\mathbf{U}_n$ only changes the j_s-th row of $\mathbf{U}_n$ by multiplying each entry in the row with $e^{i\theta}$, and keep all of the other entries of $\mathbf{U}_n$ the same. Consequently,

$$\begin{aligned} E\Big(\det(\mathbf{U}_{j_1,\cdots,j_k}) \cdot \overline{\det(\mathbf{U}_{l_1,\cdots,l_k})}\Big) &= E\Big(\det((\mathbf{A}\mathbf{U}_n)_{j_1,\cdots,j_k}) \cdot \overline{\det((\mathbf{A}\mathbf{U}_n)_{l_1,\cdots,l_k})}\Big) \\ &= e^{i\theta} \cdot E\Big(\det(\mathbf{U}_{j_1,\cdots,j_k}) \cdot \overline{\det(\mathbf{U}_{l_1,\cdots,l_k})}\Big) \end{aligned}$$

for any $\theta \in \mathbb{R}$, which implies (14). From (12), (13) and (14) we conclude

$$E|b_{m-k}|^2 = \binom{m}{k} \cdot E\Big(|\det(\mathbf{U}_{1,\cdots,k})|^2\Big) \tag{15}$$

for all $1 \le k \le m$. Now define

$$\det(z\mathbf{I}_n - \mathbf{U}_n) = z^n + a_{n-1}z^{n-1} + \cdots + a_0.$$

Taking $m = n$ in (15) and by Lemma 2.5 we have that

$$1 = E|a_{n-k}|^2 = \binom{n}{k} \cdot E\Big(|\det(\mathbf{U}_{1,\cdots,k})|^2\Big)$$

for all $1 \le k \le n$. This and (15) yield the desired conclusion. ■

Lemma 2.7 *Let $m = m_n$ be an integer with $1 \le m < n$ for each $n \ge 2$. Let $\{\mathbf{U}_n;\, n \ge 1\}$ be a sequence of Haar unitary matrices defined on the same probability space, where $\mathbf{U}_n$ is an $n \times n$ matrix for each $n \ge 2$. Denote by $\mathbf{W}_{[n,m]}$ the $m \times m$ lower-right corner of $\mathbf{U}_n$. If $\lim_{n\to\infty} m/n = 0$, then*

$$\lim_{n\to\infty} \frac{1}{n} \log |det(\mathbf{W}_{[n,m]})| = 0 \quad a.s. \tag{16}$$

Proof. For any $\epsilon > 0$, we claim that

$$P\Big(\frac{1}{n}\Big| \log |\det(\mathbf{W}_{[n,m]})|\Big| > \epsilon\Big) \le e^{-n\epsilon} \tag{17}$$

as n is sufficiently large. If this is true, then the sum of the above probabilities over all $n \ge 2$ is finite. Therefore, (16) follows from the Borel-Cantelli lemma.

Let $\mathbf{U}_{[n,m]}$ the $m \times m$ upper-left corner of $\mathbf{U}_n$. By the Haar invariance, $\mathbf{U}_{[n,m]}$ and $\mathbf{W}_{[n,m]}$ have the same distribution. Thus, to prove (17), or to prove the lemma, it suffices to show, for any $\epsilon > 0$,

$$P\Big(\frac{1}{n}\Big| \log |\det(\mathbf{U}_{[n,m]})|\Big| > \epsilon\Big) \le e^{-n\epsilon} \tag{18}$$

as n is sufficiently large. Now let us prove this.

Set $\mathbf{V}_m = \mathbf{U}_{[n,m]}\mathbf{U}^*_{[n,m]}$ for all n. Notice $[\det(\mathbf{V}_m)]^{1/2} = |\det(\mathbf{U}_{[n,m]})| \le 1$ since $\mathbf{U}_{[n,m]}$ is a contraction map, and hence all of its eigenvalues are inside the unit circle on the complex plane. Therefore,

$$\begin{aligned}
P\Big(\frac{1}{n}\Big| \log |\det(\mathbf{U}_{[n,m]})|\Big| > \epsilon\Big) &= P\Big(\frac{1}{[\det(\mathbf{V}_m)]} > e^{2n\epsilon}\Big) \\
&\le e^{-2nt\epsilon} E\frac{1}{[\det(\mathbf{V}_m)]^t} = e^{-2nt\epsilon} E\frac{1}{(\lambda_1 \cdots \lambda_m)^t} \qquad (19)
\end{aligned}$$

for any $t > 0$ by the Markov inequality, where $0 \le \lambda_1 \le \cdots \le \lambda_m$ are the eigenvalues of $\mathbf{V}_m$. By (3.6) and (3.15) from Forrester (2004) or (2.5) from Jiang (2009) we know that $\lambda_1, \cdots, \lambda_m$ actually come from the Jacobi ensemble with probability density function

$$\begin{aligned}
& f(\lambda_1, \cdots, \lambda_m) \\
= & \frac{1}{C(m,n)} \prod_{i=1}^{m} (1-\lambda_i)^{n-2m} \cdot \prod_{1 \le i < j \le m} (\lambda_i - \lambda_j)^2 I(0 \le \lambda_1 \le \cdots \le \lambda_m \le 1), \qquad (20)
\end{aligned}$$

where $C(m,n) > 0$ is a constant depending on m and n. Since $f(\lambda_1, \cdots, \lambda_m)$ is a probability density function,

$$\begin{aligned} C(m,n) &= \int\limits_{0\le\lambda_1\le\cdots\le\lambda_m\le 1}\cdots\int \prod_{i=1}^{m}(1-\lambda_i)^{n-2m}\cdot \prod_{1\le i<j\le m}(\lambda_i-\lambda_j)^2\, d\lambda_1\cdots d\lambda_m \\ &= \frac{1}{m!}\int_0^1\cdots\int_0^1 \prod_{i=1}^{m}(1-\lambda_i)^{n-2m}\cdot \prod_{1\le i<j\le m}(\lambda_i-\lambda_j)^2\, d\lambda_1\cdots d\lambda_m. \end{aligned}$$

Now, taking $\alpha = 1$, $\beta = n-2m+1$, $\gamma = 1$ and $N = m$ in Lemma 2.2, we obtain

$$C(m,n) = \frac{1}{m!}\prod_{l=0}^{m-1}\frac{\Gamma(l+2)\Gamma(l+1)\Gamma(n-2m+l+1)}{\Gamma(2)\Gamma(n-m+l+1)}. \tag{21}$$

On the other hand, for fixed $t \in (0,1)$, by (20),

$$\begin{aligned} & E\frac{1}{(\lambda_1\cdots\lambda_m)^t} \\ = {} & \frac{1}{C(m,n)}\int\limits_{0\le\lambda_1\le\cdots\le\lambda_m\le 1}\cdots\int \prod_{i=1}^{m}\lambda_i^{-t}(1-\lambda_i)^{n-2m}\cdot \prod_{1\le i<j\le m}(\lambda_i-\lambda_j)^2\, d\lambda_1\cdots d\lambda_m \\ = {} & \frac{1}{m!C(m,n)}\int_0^1\cdots\int_0^1 \prod_{i=1}^{m}\lambda_i^{-t}(1-\lambda_i)^{n-2m}\cdot \prod_{1\le i<j\le m}(\lambda_i-\lambda_j)^2\, d\lambda_1\cdots d\lambda_m. \end{aligned}$$

Take $\alpha = 1-t$, $\beta = n-2m+1$, $\gamma = 1$ and $N = m$ in Lemma 2.2 to get

$$\begin{aligned} E\frac{1}{(\lambda_1\cdots\lambda_m)^t} &= \frac{1}{m!C(m,n)}\prod_{l=0}^{m-1}\frac{\Gamma(l+2)\Gamma(l-t+1)\Gamma(n-2m+l+1)}{\Gamma(2)\Gamma(n-m-t+l+1)} \\ &= \prod_{l=0}^{m-1}\frac{\Gamma(l-t+1)}{\Gamma(l+1)}\cdot\frac{\Gamma(n-m+l+1)}{\Gamma(n-m+l-t+1)}. \end{aligned}$$

Rearranging the index l above, and joint with (19) we conclude

$$\begin{aligned} & P\Big(\frac{1}{n}\Big|\log|\det(\mathbf{U}_{[n,m]})|\Big| > \epsilon\Big) \\ \le {} & e^{-2nt\epsilon}\prod_{l=1}^{m}\frac{\Gamma(l-t)}{\Gamma(l)}\cdot\prod_{l=1}^{m}\frac{\Gamma(n-m+l)}{\Gamma(n-m+l-t)}. \end{aligned} \tag{22}$$

Now by Lemma 2.4, there exist integer $m_0 \ge 2$ such that

$$\frac{1}{2}x^{-t} \le \frac{\Gamma(x-t)}{\Gamma(x)} \le 2x^{-t}$$

for all $x \ge m_0$. Thus,

$$\prod_{l=1}^{m}\frac{\Gamma(l-t)}{\Gamma(l)} \le C(m_0,t)\cdot\frac{1}{((m_0-1)!)^t}\prod_{l=m_0}^{m}\frac{\Gamma(l-t)}{\Gamma(l)} \le C(m_0,t)\frac{2^m}{(m!)^t} \tag{23}$$

for $m \geq m_0$, where

$$C(m_0, t) = ((m_0 - 1)!)^t \prod_{l=1}^{m_0-1} \Big(1 + \frac{\Gamma(l-t)}{\Gamma(l)}\Big)$$

is a constant depending on m_0 and t only. Therefore, from (23),

$$\prod_{l=1}^{m} \frac{\Gamma(l-t)}{\Gamma(l)} \leq C(m_0, t) \frac{2^m}{(m!)^t} \tag{24}$$

for all $m \geq m_0$. It is easy to check that (24) also holds when $m \leq m_0 - 1$. Thus, (24) is true for all m.

On the other hand, from the assumption, we see that $n - m \to \infty$ as $n \to \infty$. By the same argument as in the above,

$$\prod_{l=1}^{m} \frac{\Gamma(n-m+l)}{\Gamma(n-m+l-t)} \leq 2^m \prod_{l=1}^{m} (n-m+l)^t = 2^m \prod_{j=n-m+1}^{n} j^t = 2^m \big(\frac{n!}{(n-m)!}\big)^t \tag{25}$$

as n is sufficiently large. Combing (22), (24) and (25), we arrive at

$$P\Big(\frac{1}{n}\Big| \log |\det (\mathbf{U}_{[n,m]})|\Big| > \epsilon\Big) \leq C(m_0, t) e^{-2nt\epsilon} 4^m \binom{n}{m}^t$$

as n is sufficiently large. By the Stirling formula $m! = \sqrt{2\pi m}\, m^m e^{-m+\frac{\theta_m}{12m}}$ with $\theta_m \in (0, 1)$, we have that $\binom{n}{m} \leq \big(\frac{ne}{m}\big)^m$ for all $1 \leq m \leq n$. Consequently,

$$\begin{aligned} & P\Big(\frac{1}{n}\Big| \log |\det (\mathbf{U}_{[n,m]})|\Big| > \epsilon\Big) \\ \leq\ & C(m_0, t) \cdot \exp\Big\{ - \Big[(2\epsilon)\frac{n}{m} - \log \frac{n}{m} - \log(4^{1/t} e)\Big] mt\Big\} \leq C(m_0, t) \cdot e^{-3nt\epsilon/2} \end{aligned}$$

as n is sufficiently large, where we use the inequality $(2\epsilon)\frac{n}{m} - \log \frac{n}{m} - \log(4^{1/t}e) \geq (1.5\epsilon)\frac{n}{m}$ in the last step since $\lim_{n\to\infty} \frac{n}{m} = \infty$. Taking $t = 3/4$, we get (18). The proof is completed. ■

Lemma 2.8 *Let $n \geq 2$ and $\mathbf{A}$ be an $n \times n$ unitary matrix. Write*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{pmatrix}$$

where $\mathbf{A}_1$ and $\mathbf{A}_4$ are square matrices. Then $|det(\mathbf{A}_1)| = |det(\mathbf{A}_4)|$.

Proof. Let $\mathbf{A}_1$ be $p \times p$ and $\mathbf{A}_4$ be $q \times q$ with $p \geq 1$, $q \geq 1$ and $p + q = n$. Without loss of generality, assume $p \geq q$. Looking at the $(1, 1)$-block of $\mathbf{A}^*\mathbf{A} = I_n$, we get $\mathbf{A}_1^*\mathbf{A}_1 + \mathbf{A}_3^*\mathbf{A}_3 = I_p$. It follows that

$$|\det(\mathbf{A}_1)|^2 = \det(\mathbf{A}_1^*\mathbf{A}_1) = \det(I_p - \mathbf{A}_3^*\mathbf{A}_3) = \det(I_q - \mathbf{A}_3\mathbf{A}_3^*) \tag{26}$$

since the eigenvalues of $\mathbf{A}_3^*\mathbf{A}_3$ are the same as those of $\mathbf{A}_3\mathbf{A}_3^*$ plus 0 with $p-q$ fold. Looking at the $(2,2)$-block of $\mathbf{A}\mathbf{A}^* = I_n$, we have $\mathbf{A}_3\mathbf{A}_3^* + \mathbf{A}_4\mathbf{A}_4^* = I_q$. Thus

$$|\det(\mathbf{A}_4)|^2 = \det(I_q - \mathbf{A}_3\mathbf{A}_3^*),$$

which together with (26) yields the conclusion. ■

Remark 2.1 *Let $n \geq 2$ and $\mathbf{U}_n = (u_{ij})$ be an $n \times n$ Haar unitary matrix and $\mathbf{U}_{[n,n-1]}$ be the $(n-1)\times(n-1)$ upper-left corner of $\mathbf{U}_n$. According to Lemma 2.8 and the distribution of the entries of $\mathbf{U}_n$ (see, for example, Jiang 2010), we have*

$$|det(\mathbf{U}_{[n,n-1]})|^2 = |u_{nn}|^2 \stackrel{d}{=} \frac{\xi_1^2 + \xi_2^2}{\sum_{i=1}^{2n} \xi_i^2} \tag{27}$$

where $\{\xi_1, \cdots, \xi_{2n}\}$ are independent and identically distributed random variables with distribution $N(0,1)$, and "$\stackrel{d}{=}$" means the last two random variables in (27) have the same probability distribution. Thus, $|det(\mathbf{U}_{[n,n-1]})|^2$ is asymptotically $\chi^2(2)/(2n)$ as n is large. Heuristically, $|det(\mathbf{U}_{[n,n-1]})|$ is close to $|det(\mathbf{U}_n)| = 1$. However, (27) shows that $det(\mathbf{U}_{[n,n-1]})$ is in fact very small. This seems counterintuitive.

Proof of Theorem 1. Set $g(z) = \det(z\mathbf{I}_m - \mathbf{U}_{[n,m]}) = z^m + b_{m-1}z^{m-1} + \cdots + b_0$. By Lemma 2.3, to prove the theorem, it is enough to show that

$$\frac{1}{m}\log\Big(\frac{1}{\sqrt{|b_0|}}\sum_{j=0}^{m}|b_j|\Big) \to 0 \ \ a.s. \tag{28}$$

as $n \to \infty$, where $b_m := 1$.

First, for any $\epsilon > 0$,

$$P\Big(\frac{1}{m^2}\sum_{j=0}^{m}|b_j| \geq \epsilon\Big) \leq \frac{1}{m^2\epsilon^2}\cdot E\Big(\frac{1}{m}\sum_{j=0}^{m}|b_j|\Big)^2 \leq \frac{1}{m^2\epsilon^2}\cdot\frac{1}{m}\sum_{j=0}^{m}E(|b_j|^2)$$

by a convex inequality. From Lemma 2.6 we see that

$$E(|b_{m-j}|^2) = \frac{\binom{m}{j}}{\binom{n}{j}} \leq 1$$

for all $1 \leq j \leq m$. Recalling $b_m := 1$, the above two assertions give that

$$\sum_{n=1}^{\infty}P\Big(\frac{1}{m_n^2}\sum_{j=0}^{m_n}|b_j| \geq \epsilon\Big) \leq \sum_{n=1}^{\infty}\frac{1}{m_n^2\epsilon^2} < \infty$$

for any $\epsilon > 0$ since $m_n/n \to 1$ as $n \to \infty$. By the Borel-Cantelli lemma, we have that

$$\frac{1}{m^2}\sum_{j=0}^{m}|b_j| = \frac{1}{m_n^2}\sum_{j=0}^{m_n}|b_j| \to 0 \ \ a.s.$$

as $n \to \infty$. Since $b_m = 1$, it is not hard to see

$$\frac{1}{m} \log \sum_{j=0}^{m} |b_j| \to 0 \quad a.s. \tag{29}$$

as $n \to \infty$.

Second, notice $|b_0| = |\det(\mathbf{U}_{[n,m]})|$. Write

$$\mathbf{U}_n = \begin{pmatrix} \mathbf{U}_{[n,m]} & * \\ * & \mathbf{W}_{[n,m'_n]} \end{pmatrix}$$

where $\mathbf{W}_{[n,m'_n]}$ is the $m'_n \times m'_n$ lower-right corner of $\mathbf{U}_n$ with $m'_n := n - m$. By Lemma 2.8, $|b_0| = |\det(\mathbf{W}_{[n,m'_n]})|$. From Lemma 2.7,

$$\frac{1}{n} \log |b_0| = \frac{1}{n} \log |\det(\mathbf{W}_{[n,m'_n]})| \to 0 \ a.s. \tag{30}$$

as $n \to \infty$ since $m'_n/n = 1 - (m/n) \to 0$ by the assumption. Then (29) and (30) yield (28). The proof is completed. ■

2.2 The Proof of Theorem 2

Suppose $\mathbf{X}_n = (X_{ij})$ is an $n \times n$ matrix whose n^2 entries are independent and identically distributed complex normal random variables with $E(X_{11}) = 0$ and $E(|X_{11}|^2) = 1/n$. In other words, X_{11} and $(\xi + i\eta)/\sqrt{2n}$ have the same distribution where ξ and η are i.i.d. $N(0,1)$-distributed real random variables. In literature, $\mathbf{X}_n$ is called a complex Ginibre ensemble, see Ginibre (1965). Let $\lambda_1, \cdots, \lambda_n$ be the (complex) eigenvalues of $\mathbf{X}_n$. Their joint probability density function is given by

$$f(\lambda_1, \cdots, \lambda_n) = C_n \cdot e^{-n\sum_{i=1}^{n} |\lambda_i|^2} \cdot \prod_{1 \le i < j \le n} |\lambda_i - \lambda_j|^2 d\lambda_1 \cdots d\lambda_n \tag{31}$$

for all $\lambda_1, \cdots, \lambda_n \in \mathbb{C}$, where $d\lambda_1 \cdots d\lambda_n := \prod_{i=1}^{n} d\,\mathrm{Re}(\lambda_i) \prod_{i=1}^{n} d\,\mathrm{Im}(\lambda_i)$ is the Lebesgue integral in $\mathbb{R}^{2n}$ and

$$C_n = \frac{n^{n(n+1)/2}}{\pi^n \prod_{k=1}^{n} k!}. \tag{32}$$

See, for example, (1.35) and (1.36) from Ginibre (1965) (note the notation $dz\,dz^* = 2\,dx\,dy$ for $z = x + iy$ in the paper).

Let μ_0 be the uniform probability distribution on the complex unit disc $\{z \in \mathbb{C};\ |z| \le 1\}$. Given $\lambda_1, \cdots, \lambda_n$, define

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i} \tag{33}$$

as the empirical distribution of $\lambda_1, \cdots, \lambda_n$. Recall $\mathcal{M}(\mathbb{C})$ above (3) and the associated weak topology. We need the following large deviation result.

Lemma 2.9 *(Theorem 9 from Hiai and Petz (1998)). Suppose $\lambda_1, \cdots, \lambda_n$ have the joint density function $f(\lambda_1, \cdots, \lambda_n)$ as in (31). Then $\{P_n;\, n \geq 1\}$ satisfies the large deviation principle with speed $\{n^2;\, n \geq 1\}$ and a good rate function $I(\mu)$, which takes the unique minimum at μ_0 and $I(\mu_0) = 0$. In particular,*

$$\limsup_{n\to\infty} \frac{1}{n^2} \log P\left(P_n \in F\right) \leq -\inf_{\mu \in F} I(\mu) < 0$$

for any closed set $F \subset \mathcal{M}(\mathbb{C})$ with $\mu_0 \notin F$.

Let $1 \leq m < n$. Recall $\mathbf{U}_{[n,m]}$ is the $m \times m$ upper-left corner of an $n \times n$ Haar unitary matrix $\mathbf{U}_n$. It is known from Życzkowski and Sommers (2000) and Petz and Réffy (2005) that the joint probability density function of $\lambda_1, \cdots, \lambda_m$, the eigenvalues of $\mathbf{U}_{[n,m]}$, is given by

$$g(\lambda_1, \cdots, \lambda_m) = C_{[n,m]} \prod_{1 \leq i < j \leq m} |\lambda_i - \lambda_j|^2 \cdot \prod_{i=1}^{m} (1 - |\lambda_i|^2)^{n-m-1} d\lambda_1 \cdots d\lambda_m \tag{34}$$

with

$$\begin{aligned} \frac{1}{C_{[n,m]}} &= \pi^m m! \prod_{j=0}^{m-1} \binom{n-m+j-1}{j}^{-1} \frac{1}{n-m+j} \\ &= \pi^m \frac{\prod_{j=1}^{m} j!}{\prod_{j=n-m}^{n-1} j!} \Big[(n-m-1)!\Big]^m. \end{aligned} \tag{35}$$

Lemma 2.10 *Let $\{a_{n,i};\, 1 \leq i \leq n,\, n \geq 1\}$ and $\{t_n;\, n \geq 1\}$ be complex numbers such that $\lim_{n\to\infty} t_n = 1$. Set*

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{a_{n,i}} \quad \textit{and} \quad \mu_n' = \frac{1}{n}\sum_{i=1}^{n} \delta_{(t_n a_{n,i})}$$

for $n \geq 1$. If μ_n converges weakly to a Borel probability measure ν on $\mathbb{C}$, then μ_n' also converges weakly to ν.

Proof. Let $f(x)$ be a real-valued function defined on $\mathbb{C}$ with

$$\sup_{x \in \mathbb{C}} |f(x)| = 1 \quad \text{and} \quad |f(x) - f(y)| \leq \alpha |x - y| \tag{36}$$

for all $x, y \in \mathbb{C}$, where $\alpha \in (0, \infty)$ is a constant. We need to show

$$\frac{1}{n}\sum_{i=1}^{n} f(t_n a_{n,i}) \to \int_{\mathbb{C}} f(x)\, \nu(dx)$$

as $n \to \infty$. By the given condition, the above is true with $t_n \equiv 1$. So it suffices to show

$$H_n := \frac{1}{n}\sum_{i=1}^{n} |f(t_n a_{n,i}) - f(a_{n,i})| \to 0 \tag{37}$$

as $n \to \infty$. For any $r > 0$, set $F_r = \{x \in \mathbb{C};\ |x| \geq r\}$. Using (36) we have

$$\frac{1}{n}\sum_{i=1}^{n}|f(t_n a_{n,i}) - f(a_{n,i})| \ \leq\ \frac{2}{n}\sum_{i=1}^{n} I\{|a_{n,i}| \geq r\} + \alpha r|t_n - 1| = 2\mu_n(F_r) + \alpha r|t_n - 1|.$$

Then $\limsup_{n\to\infty} H_n \leq 2\nu(F_r)$. We get (37) by letting $r \to +\infty$. $\blacksquare$

Proof of Theorem 2. Set $t_n = \sqrt{(n-m-1)/n}$ and

$$\mu'_m = \frac{1}{m}\sum_{i=1}^{m}\delta_{(t_n\sqrt{n/m}\,\lambda_i)} \tag{38}$$

for $n \geq 1$. We will first show that, for any $\epsilon > 0$, there exists a constant $\tau = \tau(\epsilon) > 0$ such that

$$P(\rho(\mu'_m, \mu_0) \geq \epsilon) \leq e^{-\tau m_n^2} \tag{39}$$

as n is sufficiently large, where ρ is as in (3).

(i) If (39) is true, then $\rho(\mu'_m, \mu_0) \to 0$ in probability as $n \to \infty$. Recall that random variable $T_n \to T$ in probability as $n \to \infty$ if and only if for any subsequence $\{n_k\}$ of $\{n\}$, there is a further subsequence $\{n_{k_j}\}$ such that $T_{n_{k_j}} \to T$ almost surely as $j \to \infty$. Notice $\lim_{n\to\infty} t_n = 1$ since $m = m_n = o(n)$ as $n \to \infty$. By Lemma 2.10, we have that $\rho(\mu_m, \mu_0) \to 0$ in probability as $n \to \infty$.

(ii) If (39) is true, by the assumption $\lim_{n\to\infty} m_n/\sqrt{\log n} = \infty$, we see that

$$\sum_{n\geq m_1} P(\rho(\mu'_m, \mu_0) \geq \epsilon) < \infty$$

for any $\epsilon > 0$. It follows from the Borel-Cantelli lemma that, with probability one, $\rho(\mu'_m, \mu_0) \to 0$ as $n \to \infty$, equivalently, μ'_m converges weakly to μ_0 as $n \to \infty$. By Lemma 2.10, with probability one, μ_m converges weakly to μ_0 as $n \to \infty$.

Now we prove (39). From (34), we have

$$\begin{aligned}
& P(\rho(\mu'_m, \mu_0) \geq \epsilon) \\
= \ & C_{[n,m]}\int_{\rho(\mu'_m,\mu_0)\geq\epsilon}\prod_{1\leq i<j\leq m}|\lambda_i - \lambda_j|^2 \cdot \prod_{i=1}^{m}(1 - |\lambda_i|^2)^{n-m-1}\prod_{i=1}^{m} d\,\lambda_i \\
\leq \ & C_{[n,m]}\int_{\rho(\mu'_m,\mu_0)\geq\epsilon}\prod_{1\leq i<j\leq m}|\lambda_i - \lambda_j|^2 \cdot e^{-(n-m-1)\sum_{i=1}^{m}|\lambda_i|^2}\prod_{i=1}^{m} d\,\lambda_i
\end{aligned}$$

where $\prod_{i=1}^{m} d\,\lambda_i = \prod_{i=1}^{m} d\,\mathrm{Re}(\lambda_i)\prod_{i=1}^{m} d\,\mathrm{Im}(\lambda_i)$, and $C_{[n,m]}$ is as in (35) and the inequality $1 - x \leq e^{-x}$ for any $x \in \mathbb{R}$ is used in the last step. Set $y_i = s_n\lambda_i$ for $i = 1, 2, \cdots, m$, where $s_n = \sqrt{(n-m-1)/m}$. Then

$$\begin{aligned}
& P(\rho(\mu'_m, \mu_0) \geq \epsilon) \\
\leq \ & \frac{(s_n)^{-m(m+1)}C_{[n,m]}}{C_m}\int_{\rho(P_m,\,\mu_0)\geq\epsilon} C_m \cdot \prod_{1\leq i<j\leq m}|y_i - y_j|^2 \cdot e^{-m\sum_{i=1}^{m}|y_i|^2}\prod_{i=1}^{m} d\,y_i
\end{aligned} \tag{40}$$

where C_m is as in (32) and

$$P_m := \frac{1}{m}\sum_{i=1}^{m}\delta_{y_i}$$

is the same as the notation in (33). Observe that the integral in (40) is equal to $P(P_m \in F)$ where the underlying probability measure P has density function $f(\lambda_1, \cdots, \lambda_m)$ as in (31) and (32), and where $F := \{\mu \in \mathcal{M}(\mathbb{C});\ \rho(\mu, \mu_0) \geq \epsilon\}$ is a closed set in the weak convergence topology and $\mu_0 \notin F$. By Lemma 2.9, there exists a constant $\tau = \tau(\epsilon) > 0$ such that

$$P(P_m \in F) \leq e^{-2m^2\tau}$$

as n is sufficiently large, and hence $m = m_n$ is sufficiently large. This combining with (40) gives

$$P(\rho(\mu'_m, \mu_0) \geq \epsilon) \leq \frac{(s_n)^{-m(m+1)}C_{[n,m]}}{C_m}e^{-2m^2\tau} \tag{41}$$

as n is sufficiently large. Note that

$$\frac{(s_n)^{-m(m+1)}C_{[n,m]}}{C_m} = (s_n)^{-m(m+1)} \cdot \frac{\prod_{j=n-m}^{n-1} j!}{\Big[(n-m-1)!\Big]^m} \cdot \frac{1}{m^{m(m+1)/2}}. \tag{42}$$

Observe that $j! = (n-m-1)!(n-m)\cdots j \leq (n-m-1)!n^{j-n+m+1}$ for all $n-m \leq j \leq n-1$. Recalling $s_n = \sqrt{(n-m-1)/m}$, we get from (42) that

$$\begin{aligned}\frac{(s_n)^{-m(m+1)}C_{[n,m]}}{C_m} &\leq (s_n)^{-m(m+1)} \cdot \frac{n^{m(m+1)/2}}{m^{m(m+1)/2}}\\ &= \Big(1 - \frac{m+1}{n}\Big)^{-m(m+1)/2}\end{aligned}$$

which joint with (41) yields

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{m^2}\log P(\rho(\mu'_m, \mu_0) \geq \epsilon)\\ \leq\ &\limsup_{n\to\infty}\Big(-2\tau - \frac{m+1}{2m}\log\Big(1 - \frac{m+1}{n}\Big)\Big) = -2\tau\end{aligned}$$

since $m = m_n = o(n)$. Consequently,

$$P(\rho(\mu'_m, \mu_0) \geq \epsilon) \leq e^{-\tau m_n^2}$$

as n is sufficiently large. Thus (39) is obtained. The proof is completed. ■

Acknowledgement We thank Professor Sho Matsumoto for a very useful comment on the proof of Theorem 1.

References

[1] Ahlfors, L. (1979). Complex Analysis, 3rd ed., McGraw-Hill, Inc.

[2] Bai, Z. D. (1997). Circular Law. Ann. Probab. 25, 494-529.

[3] Beenakker, C. (1997). Random-matrix theory of quantum transport, Reviews of Modern Physics 69, 731-808.

[4] Casati, G., Maspero, G. and Shepelyansky, D. (1999). Quantum Poncaré recurrences, Phys. Rev. Lett. 82, 524-527.

[5] Collins, B. (2005). Product of random projections, Jacobi ensembles and universality problems arising from free probability. Probability theory and related fields 133, 315-344.

[6] Diaconis, P. and Evans, S. (2001). Linear functionals of eigenvalues of random matrices. Transactions Amer. Math. Soc. 353, 2615-2633.

[7] Diaconis, P. and Shahshahani, M. (1994). On the eigenvalues of random matrices, J. Appl. Probab. 31A, 49-62.

[8] Diaconis, P., Eaton, M. and Lauritzen, L. (1992). Finite deFinetti theorem in linear models and multivariate analysis. Scand. J. Statist. 19(4), 289-315.

[9] Dudley, R. (2002). Real Analysis and Probability, 2nd ed., Cambridge University Press.

[10] Eaton, M. (1989). Group-Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics, Vol.1. IMS, Hayward, California.

[11] Erdös, P. and Turán, P. (1950). On the distributions of roots of polynomials. Ann. Math. 5(1), 105-119.

[12] Forrester, P. (2010). The limiting Kac random polynomial and truncated random orthogonal matrices, *http://arxiv.org/abs/1009.3066.*

[13] Forrester, P. (2007). Log-Gases and Random Matrices (LMS-34) (London Mathematical Society Monographs). Princeton University Press.

[14] Forrester, P. (2006). Quantum conductance problems and the Jacobi ensemble. J. Phys. A: Math. Gen. 39, 6861-6870.

[15] Fyodorov, Y. and Sommers, H. (2003). Random matrices close to Hermitian or unitary: overview of methods and results. J. Phys. A, 36(12), 3303-3347.

[16] Gamelin, T. W. (2001). Complex Analysis, 1st ed., Springer.

[17] Ginibre, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. Journal of Mathematical Physics 6, 440-449.

[18] Girko, V. L. (1984a). Circle law. Theory Probab. Appl. 4, 694-706.

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

[20] Glück, A., Kolovsky, A. and Korsch, H. (2002). Wannier-Stark resonances in optical and semiconductor superlattices. Phys. Rep. 366, 103-182.

[21] Granville, A. (2007). The distribution of roots of a polynomial in equidistribution in number theory, An introduction NATO Sci. Ser. II Math. Phys. Chem. vol. 237, 93-102.

[22] Hakke, F. (2010). Quantum Signature of Chaos, 3rd ed. Springer and New York.

[23] Hiai, F. and Petz, D. (1998). Logarithmic energy as entropy functional. In Advances in Differential Equations and Mathematical Physics (E. Carlen, E. M. Harrell and M. Loss, eds.) 205-221. Amer. Math. Soc., Providence, RI.

[24] Hughes, C. and Nikeghbali, A. (2008). The zeros of random polynomials cluster uniformly near the unit circle. Compositio Mathematica, 144(3), 734-746.

[25] Jiang, T. (2010). The entries of Haar-invariant matrices from the classical compact groups. Journal of Theoretical Probability 23(4), 1227-1243.

[26] Jiang, T. (2009). Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles. Probability Theory and Related Fields 144(1), 221-246.

[27] Jiang, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? Ann. Probab. 34(4), 1497-1529.

[28] Khoruzhenko, B., Sommers, H. and Życzkowski, K. (2010). Truncation of random orthogonal matrices. Physical Review E 82, 040106(R).

[29] Macdonald, I. (1998). Symmertic Functions and Hall Polynomials, Oxford Mathematical Monographs, 2nd ed., Oxford University Press.

[30] Mehta, M. (1991). Random Matrices, 2nd ed., Academic Press, Boston.

[31] Nonnenmacher, S. and Zworski, M. (2007). Distribution of resonances for open quantum maps, Commun. Math. Phys. 269, 311-365.

[32] Pan, G. and Zhou, W. (2010). Circular law, extreme singular values and potential theory. Journal of Multivariate Analysis 101(3), 645-656.

[33] Pedrosa, J., Carlo, G., Wisniacki, D. and Ermann, L. (2009). Distribution of resonances in the Quantum Open Baker Map. Phys. Rev. E 79, 016215.

[34] Petz, D. and Réffy, J. (2005). Large deviation for the empirical eigenvalue density of truncated Haar unitary matrices. Probability Theory and Related Fields 133(2), 175-189.

[35] Petz, D. and Réffy, J. (2004). On asymptotics of large Haar distributed unitary matrices. Periodica Mathematica Hungarica. 49(1), 103-117.

[36] Schomerus, H. and Jacquod, P. (2005). Quantum-to-classical correspondence in open chaotic systems. J. Phys. A 38, 10663-10682.

[37] Stanley, R. (2001). Enumerative Combinatorics, vol. 2. Cambridge University Press.

[38] Tao, T. and Vu, V. (2010). Random matrices: universality of ESDs and the circular law. Ann. Probab. 38(5), 2023-2065.

[39] Tao, T. and Vu, V. (2009). From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices. Bull. Amer. Math. Soc. 46, 377-396.

[40] Tao, T. and Vu, V. (2008). Random matrices: The Circular Law, Communication in Contemporary Mathematics 10, 261-307.

[41] Życzkowski, K. and Sommers, H. (2000). Truncation of random unitary matrices. J. Phys. A: Math. Gen. 33, 2045-2057.