Statistics and Applications {ISSN 2454-7395 (online)}
Volume 19, No. 1, 2021 (New Series), pp 187–198

A Note on the Folklore of Free Independence

Arijit Chakrabarty[1], Sukrit Chakraborty[1] and Rajat Subhra Hazra[1]
[1] *Theoretical Statistics and Mathematics Unit*
Indian Statistical Institute, 203 B.T. Road, Kolkata 700108

Dedicated to the memory of Professor Aloke Dey

Received: 12 November 2020; Revised: 22 January 2021; Accepted: 25 January 2021

Abstract

It is shown that a Wishart matrix of standard complex normal random variables is asymptotically freely independent of an independent random matrix, under minimal conditions, in two different sense of asymptotic free independence.

Key words: Voiculescu's theorem; Random matrix theory; Asymptotic free independence; Wishart matrix.

AMS Subject Classifications: Primary 60B20; Secondary 46L54.

1. Introduction

Since the seminal discovery of [10], there have been several folklores regarding free independence. For example, one such folklore is that any two independent Wigner matrices are asymptotically freely independent, and another is that any Wishart matrix is asymptotically freely independent of a deterministic matrix. While such folklores are true, more often than not, there are a few problems. The first and foremost problem is that the meaning of the phrase "asymptotically freely independent" varies with context. A widely used definition is in terms of the normalized expected trace (or without the expectation). Unfortunately, with this definition, the claim of asymptotic free independence can easily fail, in the absence of any other assumption. The counter example in [7] is noteworthy. This articulates the second problem with the folklore, which is that the required assumptions are usually missing. Nevertheless, in the literature, there are several rigorous proofs of various versions of Voiculescu's theorem; see, for example, the monographs [9], [2] and [8]. The reader will notice that the versions in the above references are not monotonic in strength, that is, one version does not necessarily imply another. In other words, there is no general theorem regarding asymptotic free independence from which most results of interest follow.

This note is a modest attempt at settling some of the issues mentioned above in a specific example. Theorems 1 and 2 claim asymptotic free independence of a Wishart matrix W_N of standard complex normal random variables and an independent matrix Y_N, under two different definitions of asymptotic free independence. The former is the usual definition, in terms of normalized expected trace, while the latter is in terms of the limiting spectral distribution of random matrices, which is weaker than the former. In both the above mentioned

Corresponding Author: Arijit Chakrabarty
Email: arijit.isi@gmail.com

theorems, the limiting spectral distribution of Y_N is assumed to be compactly supported, at the least. This assumption is relaxed in Theorem 3, a consequence of which is that the claim is also significantly weakened. The proofs of Theorems 2 and 3 are based on truncation arguments.

We choose to work with the complex normal distribution because they yield the strongest results in that the assumptions on Y_N become minimal. This is why, for example, Theorem 22.35 of [9] assumes the distribution to be complex normal. It is worth noting that Theorem 2 of [1] and the results in [6] are similar in spirit. Although the results are stated for a Wishart matrix, they hold for a Wigner matrix as well.

2. The Results

Let $(Z_{i,j} : i,j \in \mathbb{N})$ be a family of i.i.d. standard complex Normal random variables. That is, $(\Re(Z_{i,j}) : i,j \geq 1)$ and $(\Im(Z_{i,j}) : i,j \geq 1)$ are independent families of i.i.d. real $N(0,1/2)$ random variables. Suppose that $(M_N : N \geq 1)$ is a sequence of positive integers such that

$$\lim_{N\to\infty} \frac{N}{M_N} = \lambda \in (0,\infty)\,. \tag{1}$$

For each $N \geq 1$, let X_N be the $M_N \times N$ random matrix defined by

$$X_N(i,j) := Z_{i,j},\ 1 \leq i \leq M_N,\ 1 \leq j \leq N\,.$$

For $N \geq 1$, define an $N \times N$ random Hermitian matrix by

$$W_N := \frac{1}{M_N} X_N^* X_N\,.$$

Notice that for $1 \leq i,j \leq N$,

$$W_N(i,j) = \frac{1}{M_N} \sum_{k=1}^{M_N} \overline{Z_{k,i}} Z_{k,j}\,.$$

Hence W_N is a Wishart matrix.

For a random Hermitian $N \times N$ matrix Z, its "empirical spectral distribution" and "expected empirical spectral distribution", denoted by $\mathrm{ESD}(Z)$ and $\mathrm{EESD}(Z)$, respectively, are probability measures on $\mathbb{R}$, defined as

$$\begin{aligned} \mathrm{ESD}(Z) &= \frac{1}{N}\sum_{i=1}^N \mathbf{1}(\lambda_i \in \cdot)\,, \\ \mathrm{EESD}(Z) &= \frac{1}{N}\sum_{i=1}^N P(\lambda_i \in \cdot)\,, \end{aligned}$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of Z, counted with multiplicity.

It is well known that as $N \to \infty$,

$$\mathrm{ESD}(W_N) \to \nu_\lambda\,,$$

weakly in probability, where ν_λ, with λ as in (1), is the Marčenko-Pastur distribution, defined by

$$\nu_\lambda(dx) = \begin{cases} \left(1-\frac{1}{\lambda}\right)\mathbf{1}(0\in dx) + \frac{1}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{\lambda x}\mathbf{1}_{[\lambda_-,\lambda_+]}(x)\,dx, & \lambda>1\,,\\ \frac{1}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{\lambda x}\mathbf{1}_{[\lambda_-,\lambda_+]}(x)\,dx, & \lambda\le 1\,, \end{cases}$$

with $\lambda_\pm = (1\pm\sqrt{\lambda})^2$.

For each $N\ge 1$, Y_N is an $N\times N$ random complex Hermitian matrix, **independent** of $(Z_{i,j} : i,j\in\mathbb{N})$. The exact assumption on the spectrum of Y_N will vary from result to result, and hence will be mentioned in the statements of the respective results. However, at the very least, there exists a (non-random) probability measure μ on $\mathbb{R}$ such that

$$\mathrm{ESD}(Y_N)\to\mu\,, \tag{2}$$

weakly in probability, as $N\to\infty$.

The statements of the following results are based on the theory of C^*-probability spaces. A reader unacquainted with this may look at [9]. It is known that given probability measures μ_1 and μ_2 which are supported on a compact subset of $\mathbb{R}$, there exist a C^*-probability space $(\mathcal{A},\varphi)$, and two freely independent self-adjoint elements $a_1,a_2\in\mathcal{A}$ such that

$$\varphi\left(a_i^n\right) = \int_{-\infty}^{\infty} x^n\mu_i(dx),\ n\in\mathbb{N},\ i=1,2\,.$$

The probability measures μ_1 and μ_2 are called the distributions of a_1 and a_2, and denoted by $\mathcal{L}(a_1)$ and $\mathcal{L}(a_2)$, respectively.

The first result shows asymptotic free independence between W_N and Y_N in the sense of normalized expected trace.

Theorem 1: Assume that μ is compactly supported, and that for each $n\in\mathbb{N}$,

$$\lim_{N\to\infty}\mathrm{E}\left[\frac{1}{N}\mathrm{Tr}(Y_N^n)\right] = \int_{-\infty}^{\infty}x^n\mu(dx)\,,\ \text{and} \tag{3}$$

$$\lim_{N\to\infty}\mathrm{Var}\left[\frac{1}{N}\mathrm{Tr}(Y_N^n)\right] = 0\,. \tag{4}$$

Then, there exists a C^*-probability space $(\mathcal{A},\varphi)$, in which there are two freely independent self-adjoint elements w and y, having distributions ν_λ and μ, respectively, and satisfying the following: For every polynomial p in two variables having complex coefficients,

$$\lim_{N\to\infty}\frac{1}{N}\mathrm{ETr}\left[p\left(W_N,Y_N\right)\right] = \varphi\Big(p(w,y)\Big)\,. \tag{5}$$

Consequently, if $p\left(W_N,Y_N\right)$ has real eigenvalues, a.s., for all N, then as $N\to\infty$,

$$\mathrm{EESD}\left(p\left(W_N,Y_N\right)\right)\xrightarrow{w}\mathcal{L}\left(p\left(w,y\right)\right)\,. \tag{6}$$

Remark 1: When Y_N is deterministic, the assumptions of Theorem 1 just mean that

$$\lim_{n\to\infty}\frac{1}{N}\mathrm{Tr}(Y_N^n) = \int_{-\infty}^{\infty}x^n\mu(dx)\,,$$

which is stronger than (2). In general, (3) and (4) together imply (2) whenever μ is determined by its moments, which is necessarily the case if μ is compactly supported.

Remark 2: The claim (6) is an immediate consequence of (5), whenever p is such that the eigenvalues of $p(W_N, Y_N)$ are a.s. real. For example, if W_N is non-negative definite, then the above holds for

$$p(x, y) = xy\,.$$

In the next result, both the hypotheses and the claim are weakened to (2) and (6), respectively. In other words, this results proves asymptotic free independence in the sense of (6) as opposed to (5).

Theorem 2: If μ, as in (2), is compactly supported, then for every polynomial p in two variables having complex coefficients such that $p\,(W_N, Y_N)$ has real eigenvalues, a.s., for all N, (6) holds.

The last result deals with the case when the support of μ is possibly unbounded. For measures with possibly unbounded support, '$\boxplus$' and '$\boxtimes$' denote their free additive and multiplicative convolutions, respectively. For the latter, at least of one of the two measures has to be supported on the non-negative half line. See [5] for the details.

Theorem 3: If (2) holds for a probability measure μ which is not necessarily compactly supported, then

$$\begin{aligned}\mathrm{EESD}(Y_N + W_N) &\xrightarrow{w} \mu \boxplus \nu_\lambda\,, \text{and}\\ \mathrm{EESD}(Y_N W_N) &\xrightarrow{w} \mu \boxtimes \nu_\lambda\,,\end{aligned}$$

as $N \to \infty$.

Remark 3: Theorems 1 - 3 hold true, if the Wishart matrix is replaced by a Wigner matrix with standard complex normal entries, and the Marčenko-Pastur distribution is replaced by the semicircle law.

3. Some Facts

For the proofs of the results mentioned in Section 2, a few facts will be needed, which are stated here. The proofs are omitted because the results are either elementary or can be found in a cited reference.

The first one is a comparison between ranks of deterministic matrices.

Fact 3.1: Let p be a polynomial in two variables, with complex coefficients. Then, there exists a finite constant C, depending only on the polynomial p, such that

$$\mathrm{Rank}\,(p(A, B) - p(A', B)) \le C\mathrm{Rank}(A - A')\,,$$

for square matrices A, A', B of the same order.

The next result, which is also based on rank, follows from Theorem A.43, page 503, of [4].

Fact 3.2: For probability measures μ_1 and μ_2 on $\mathbb{R}$, let $d(\mu_1, \mu_2)$ denote their Lévy distance, defined by

$$d(\mu_1, \mu_2) := \inf \left\{ \varepsilon > 0 : \mu_1 \left((-\infty, x - \varepsilon] \right) \le \mu_2 \left((-\infty, x] \right) \le \mu_1 \left((-\infty, x + \varepsilon] \right) \right\} .$$

For $N \times N$ random Hermitian matrices A and B, it holds that

$$d \left(\mathrm{EESD}(A), \mathrm{EESD}(B) \right) \le \frac{1}{N} \mathrm{E} \left[\mathrm{Rank}(A - B) \right] .$$

The following fact essentially follows from uniform integrability. Nonetheless, a proof is given.

Fact 3.3: For each $N \ge 1$, suppose that Y_N is an $N \times N$ random Hermitian matrix satisfying (3) and (4). Then it holds that for any $n \ge 1$ and $k_1, \ldots, k_n \ge 0$,

$$\lim_{N \to \infty} N^{-n} \mathrm{E} \left(\prod_{i=1}^n \mathrm{Tr} \left(Y_N^{k_i} \right) \right) = \prod_{i=1}^n \alpha_{k_i} ,$$

where α_n denotes the right hand side of (3).

Proof: Fix $n \ge 1$ and $k_1, \ldots, k_n \ge 0$. A consequence of (3) and (4) is that for all fixed $k \ge 1$,

$$\frac{1}{N} \mathrm{Tr} \left(Y_N^k \right) \xrightarrow{P} \alpha_k \, , N \to \infty \, .$$

Therefore,

$$N^{-n} \prod_{i=1}^n \mathrm{Tr} \left(Y_N^{k_i} \right) \xrightarrow{P} \prod_{i=1}^n \alpha_{k_i} \, , N \to \infty \, . \tag{7}$$

Let

$$k = \sum_{i=1}^n k_i \, ,$$

which we assume without loss of generality to be at least 1, and observe that

$$\begin{aligned} N^{-n} \left| \prod_{i=1}^n \mathrm{Tr} \left(Y_N^{k_i} \right) \right| &= \prod_{i=1}^n \left| \int_{-\infty}^{\infty} x^{k_i} \left(\mathrm{ESD}(Y_N) \right) (dx) \right| \\ &\le \prod_{i=1}^n \int_{-\infty}^{\infty} |x|^{k_i} \left(\mathrm{ESD}(Y_N) \right) (dx) \\ &\le \prod_{i=1}^n \left(\int_{-\infty}^{\infty} x^{2k} \left(\mathrm{ESD}(Y_N) \right) (dx) \right)^{k_i/2k} \\ &= \left(\frac{1}{N} \mathrm{Tr} \left(Y_N^{2k} \right) \right)^{1/2} , \end{aligned}$$

the penultimate line following from the Lyapunov inequality. Thus,

$$\limsup_{N \to \infty} \mathrm{E} \left[\left(N^{-n} \prod_{i=1}^n \mathrm{Tr} \left(Y_N^{k_i} \right) \right)^2 \right] \le \lim_{N \to \infty} \mathrm{E} \left(\frac{1}{N} \mathrm{Tr} \left(Y_N^{2k} \right) \right) = \alpha_{2k} < \infty \, ,$$

the equality being implied by (3). Hence,

$$\left(N^{-n}\prod_{i=1}^{n}\mathrm{Tr}\left(Y_N^{k_i}\right):N\geq 1\right)$$

is an uniformly integrable family, which in conjunction with (7) completes the proof. □

The next fact is elementary.

Fact 3.4: Let $Z_1,\ldots,Z_N$ be i.i.d. standard complex normal, that is for each $i=1,\ldots,N$, the real and imaginary parts of Z_i are independent $N(0,1/\sqrt{2})$. If Z denotes the column vector whose i-th component is Z_i, and U is an $N\times N$ deterministic unitary matrix, then the components of UZ are also i.i.d. standard complex normal.

The next fact has essentially been proved in page 386 of [9]. As mentioned therein, an $N\times N$ Haar unitary matrix is a random matrix distributed according to the Haar measure on the group of $N\times N$ unitary matrices. Before stating the fact, we need to introduce a few notations. Let S_n denote the group of permutations on $\{1,\ldots,n\}$ for $n\geq 1$. A permutation is identified with the partition of $\{1,\ldots,n\}$, induced by the cyclic decomposition. For $\alpha\in S_n$, $\#\alpha$ denotes the number of blocks in α, that is the number of cycles. For any block $\theta\in\alpha$, $\#\theta$ denotes the length of the cycle θ. For example, for

$$\alpha\in S_4,$$

defined by

$$\alpha(1)=2,\ \alpha(2)=4,\ \alpha(3)=3,\ \alpha(4)=1\,,$$

we write

$$\alpha=\{(1,2,4),(3)\}\,,$$

and hence $\#\alpha=2$. If the elements of α, as listed above, are labelled as θ_1 and θ_2, respectively, then

$$\#\theta_1=3,\ \#\theta_2=1\,.$$

Fact 3.5: For a fixed N, let A and B be deterministic $N\times N$ Hermitian matrices. If U is an $N\times N$ Haar unitary matrix, then for any $1\leq n\leq N$ and $k_1,\ldots,k_n\geq 0$,

$$\begin{aligned}&\mathrm{ETr}\left[\prod_{i=1}^{n}\left(UA^{k_i}U^*B\right)\right]\\&=\sum_{\alpha,\beta\in S_n}\mathrm{Wg}(N,\alpha^{-1}\beta)\left(\prod_{\theta\in\alpha}\mathrm{Tr}\left(A^{\sum_{i\in\theta}k_i}\right)\right)\left(\prod_{\theta\in\beta^{-1}\gamma}\mathrm{Tr}\left(B^{\#\theta}\right)\right)\,,\end{aligned}$$

where Wg is the Weingarten function defined by

$$\mathrm{Wg}(N,\alpha)=\mathrm{E}\left[U(1,1)\ldots U(n,n)\overline{U(1,\alpha(1))}\ldots\overline{U(n,\alpha(n))}\right]\,,$$

for $\alpha\in S_n$, $N\geq n$ and

$$\gamma=\{(1,\ldots,n)\}\in S_n\,.$$

The following has essentially been proved in the course of the proof of Theorem 23.14 of [9].

Fact 3.6: For a fixed $n \geq 1$ and $\alpha \in S_n$,

$$\phi(\alpha) := \lim_{N\to\infty} N^{2n-\#\alpha}\mathrm{Wg}(N,\alpha) \text{ exists and is real.}$$

Furthermore, if $(\mathcal{A},\varphi)$, w and y are as in the statement of Theorem 1, then for $n \geq 1$ and $k_1,\ldots,k_n \geq 0$,

$$\begin{aligned}&\varphi\left(w^{k_1}y\ldots w^{k_n}y\right)\\ &= \sum_{\substack{\alpha,\beta\in S_n:\\ \#(\alpha^{-1}\beta)+\#\alpha+\#(\beta^{-1}\gamma)=2n+1}} \left[\phi(\alpha^{-1}\beta)\left(\prod_{\theta\in\alpha}\varphi\left(w^{\sum_{i\in\theta}k_i}\right)\right)\left(\prod_{\theta\in\beta^{-1}\gamma}\varphi\left(y^{\#\theta}\right)\right)\right].\end{aligned}$$

The following result is Corollary 2 of [3].

Fact 3.7: For a fixed $N \in \mathbb{N}$, there exists a measurable map

$$\psi : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N},$$

where $\mathbb{C}^{N\times N}$ is the space of all $N \times N$ matrices with complex entries, such that $\psi(M)$ is an unitary matrix for every $M \in \mathbb{C}^{N\times N}$, and

$$\psi(M)^*M\psi(M)$$

is upper triangular for every M.

4. Proofs

Proof of Theorem 1: Let $(\mathcal{A},\varphi)$, w and y be as in the statement. In order to prove the claim, all that needs to be shown is that

$$\lim_{N\to\infty} N^{-1}\mathrm{E}\left[\mathrm{Tr}\left(W_N^{k_1}Y_N\ldots W_N^{k_n}Y_N\right)\right] = \varphi\left(w^{k_1}y\ldots w^{k_n}y\right), \tag{8}$$

for fixed $n \geq 1$ and $k_1,\ldots,k_n \geq 0$.

The foremost task is to show that the expectation on the left hand side of (8) exists. To that end, it suffices to show that there exists N_0 such that

$$\mathrm{E}\left[|Y_N(i,j)|^n\right] < \infty \text{ for all } N \geq N_0,\ 1 \leq i,j \leq N. \tag{9}$$

Fix $N \geq 1$ and enumerate the eigenvalues of Y_N in ascending order by $\lambda_1,\ldots,\lambda_N$. Notice that

$$\begin{aligned}\sum_{i,j=1}^N |Y_N(i,j)|^{2n} &\leq \left[\sum_{i,j=1}^N |Y_N(i,j)|^2\right]^n\\ &= \left(\sum_{i=1}^N \lambda_i^2\right)^n\\ &\leq N^{n-1}\sum_{i=1}^N \lambda_i^{2n}\\ &= N^{n-1}\mathrm{Tr}(Y_N^{2n}).\end{aligned}$$

Since (3) implies that the expectation of the right hand side is finite for N large, an N_0 satisfying (9) exists.

Proceeding towards (8), fix $N \geq N_0$, and let

$$\mathcal{F} := \sigma\left(X_N, Y_N\right) ,$$

that is $\mathcal{F}$ is the smallest σ-field with respect to which the entries of X_N and Y_N are measurable. Let U_N be a Haar unitary matrix independent of $\mathcal{F}$. Fact 3.4 implies that conditioned on U_N, the entries of $U_N X_N$ are i.i.d. standard complex Normal. That is, the conditional joint distribution of the entries of $U_N X_N$, given U_N, is the same as that of X_N. Therefore

$$(U_N W_N U_N^*, Y_N) \stackrel{d}{=} (W_N, Y_N) \ .$$

As a result,

$$\begin{aligned} C_N &:= \mathrm{E}\left[\mathrm{Tr}\left(W_N^{k_1} Y_N \dots W_N^{k_n} Y_N\right)\right] \\ &= \mathrm{E}\left[\mathrm{Tr}\left((U_N W_N U_N^*)^{k_1} Y_N \dots (U_N W_N U_N^*)^{k_n} Y_N\right)\right] \\ &= \mathrm{E}\left[\mathrm{Tr}\left(U_N W_N^{k_1} U_N^* Y_N \dots U_N W_N^{k_n} U_N^* Y_N\right)\right] \\ &= \mathrm{E}\,\mathrm{E}_{\mathcal{F}}\left[\mathrm{Tr}\left(U_N W_N^{k_1} U_N^* Y_N \dots U_N W_N^{k_n} U_N^* Y_N\right)\right] , \end{aligned}$$

where $\mathrm{E}_{\mathcal{F}}$ is the conditional expectation given $\mathcal{F}$. By an appeal to Fact 3.5,

$$\begin{aligned} &\mathrm{E}_{\mathcal{F}}\left[\mathrm{Tr}\left(U_N W_N^{k_1} U_N^* Y_N \dots U_N W_N^{k_n} U_N^* Y_N\right)\right] \\ = &\sum_{\alpha,\beta\in S_n} \mathrm{Wg}(N,\alpha^{-1}\beta)\left(\prod_{\theta\in\alpha}\mathrm{Tr}\left(W_N^{\sum_{i\in\theta}k_i}\right)\right)\left(\prod_{\theta\in\beta^{-1}\gamma}\mathrm{Tr}\left(Y_N^{\#\theta}\right)\right) . \end{aligned}$$

Taking the unconditional expectation of both sides, and using the independence of W_N and Y_N, we get that

$$C_N = \sum_{\alpha,\beta\in S_n} \mathrm{Wg}(N,\alpha^{-1}\beta)\mathrm{E}\left(\prod_{\theta\in\alpha}\mathrm{Tr}\left(W_N^{\sum_{i\in\theta}k_i}\right)\right)\mathrm{E}\left(\prod_{\theta\in\beta^{-1}\gamma}\mathrm{Tr}\left(Y_N^{\#\theta}\right)\right) . \tag{10}$$

It is well known that for all $k \in \mathbb{N}$,

$$\begin{aligned} \lim_{N\to\infty}\mathrm{E}\left(N^{-1}\mathrm{Tr}(W_N^k)\right) &= \varphi(w^k) , \\ \lim_{N\to\infty}\mathrm{Var}\left(N^{-1}\mathrm{Tr}(W_N^k)\right) &= 0 . \end{aligned}$$

Combining the above with Fact 3.3 yields that

$$\lim_{N\to\infty}\mathrm{E}\left(\prod_{\theta\in\alpha}N^{-1}\mathrm{Tr}\left(W_N^{\sum_{i\in\theta}k_i}\right)\right) = \prod_{\theta\in\alpha}\varphi\left(w^{\sum_{i\in\theta}k_i}\right) . \tag{11}$$

Similarly, (3), (4) and Fact 3.3 together imply that

$$\lim_{N\to\infty} \mathrm{E}\left(\prod_{\theta\in\beta^{-1}\gamma} N^{-1}\mathrm{Tr}\left(Y_N^{\#\theta}\right)\right) = \prod_{\theta\in\beta^{-1}\gamma} \varphi\left(y^{\#\theta}\right) . \tag{12}$$

Rewrite (10) as

$$N^{-1}C_N = \sum_{\alpha,\beta\in S_n} N^{\#\alpha+\#(\beta^{-1}\gamma)-1}\mathrm{Wg}(N,\alpha^{-1}\beta) \mathrm{E}\left(\prod_{\theta\in\alpha} N^{-1}\mathrm{Tr}\left(W_N^{\sum_{i\in\theta}k_i}\right)\right) \mathrm{E}\left(\prod_{\theta\in\beta^{-1}\gamma} N^{-1}\mathrm{Tr}\left(Y_N^{\#\theta}\right)\right) .$$

The first claim of Fact 3.6 implies that for fixed $\alpha,\beta\in S_n$,

$$\begin{aligned} N^{\#\alpha+\#(\beta^{-1}\gamma)-1}\mathrm{Wg}(N,\alpha^{-1}\beta) &= O\left(N^{\#(\alpha^{-1}\beta)+\#\alpha+\#(\beta^{-1}\gamma)-2n-1}\right) \\ &= O(1)\,, \end{aligned}$$

because

$$\#\alpha+\#(\alpha^{-1}\beta)+\#(\beta^{-1}\gamma)\le 2n+1\,,$$

as shown in (23.4) and the following display on page 387 in [9]. Therefore, letting $N\to\infty$ in (13) and using the first claim of Fact 3.6 along with (11) and (12), we get that

$$\begin{aligned} &\lim_{N\to\infty} N^{-1}C_N \\ = &\sum_{\substack{\alpha,\beta\in S_n: \\ \#(\alpha^{-1}\beta)+\#\alpha+\#(\beta^{-1}\gamma)=2n+1}} \left[\phi(\alpha^{-1}\beta)\left(\prod_{\theta\in\alpha}\varphi\left(w^{\sum_{i\in\theta}k_i}\right)\right)\left(\prod_{\theta\in\beta^{-1}\gamma}\varphi\left(y^{\#\theta}\right)\right)\right]. \end{aligned}$$

The second claim of Fact 3.6 shows that the right hand side of the above equation is the same as that of (8). Thus the latter follows, which completes the proof. □

Proof of Theorem 2: Since μ is compactly supported, let $M>1$ be such that

$$\mu\left([-(M-1),M-1]\right)=1\,.$$

Letting ψ be as in Fact 3.7, define

$$P_N=\psi(Y_N)\,,$$

and

$$T_N:=P_N^*Y_NP_N\,,$$

which is an upper triangular matrix. Define an $N\times N$ matrix T_N' by

$$T_N'(i,j):=\begin{cases}T_N(i,j), & i\neq j\,,\\ T_N(i,i)\mathbf{1}(|T_N(i,i)|\le M), & i=j\,,\end{cases}$$

and let

$$Y_N' := P_N T_N' P_N^* \,. \tag{13}$$

In order to complete the proof, it suffices to show that for a fixed polynomial p satisfying the hypothesis,

$$\mathrm{EESD}\left(p(W_N, Y_N')\right) \xrightarrow{w} \mathcal{L}\left(p(w,y)\right) \,, \tag{14}$$

and

$$\lim_{N\to\infty} d\left(\mathrm{EESD}\left(p(W_N, Y_N)\right), \mathrm{EESD}\left(p(W_N, Y_N')\right)\right) = 0 \,, \tag{15}$$

where d is the Lévy metric, convergence in which is equivalent to weak convergence.

We start by showing (15). To that end, note that

$$\begin{aligned} N^{-1}\mathrm{Rank}(Y_N - Y_N') &= N^{-1}\mathrm{Rank}(T_N - T_N') \\ &= N^{-1}\#\{1 \le i \le N : |T_N(i,i)| > M\} \\ &= \left(\mathrm{ESD}(Y_N)\right)\left([-M,M]^c\right) \,, \end{aligned}$$

the inequality in the second line being based on the fact that $T_N - T_N'$ is a diagonal matrix, and hence

$$N^{-1}\mathrm{Rank}(Y_N - Y_N') \xrightarrow{P} 0 \tag{16}$$

as $N \to \infty$. Fact 3.1 and the bounded convergence theorem show that

$$\lim_{N\to\infty} \mathrm{E}\left[\frac{1}{N}\mathrm{Rank}(p(W_N, Y_N) - p(W_N, Y_N'))\right] = 0 \,.$$

An appeal to Fact 3.2 establishes (15).

Proceeding towards (14), in view of Theorem 1 and Remark 2, it suffices to show that (3) and (4) hold with Y_N replaced by Y_N'. Equation (16) and the hypotheses imply that

$$\mathrm{ESD}(Y_N') \to \mu \,,$$

weakly in probability, as $N \to \infty$. Since

$$\left(\mathrm{ESD}(Y_N')\right)\left([-M,M]^c\right) = \left(\mathrm{ESD}(T_N')\right)\left([-M,M]^c\right) = 0, \; N \ge 1 \,,$$

and

$$\mu\left([-M+1, M-1]^c\right) = 0 \,,$$

it follows that for a fixed $n \ge 1$, as $N \to \infty$,

$$\int_{-\infty}^{\infty} x^n \left(\mathrm{ESD}(Y_N')\right)(dx) \xrightarrow{P} \int_{-\infty}^{\infty} x^n \mu(dx) \,.$$

The observations that

$$\frac{1}{N}\mathrm{Tr}\left[(Y_N')^n\right] = \int_{-\infty}^{\infty} x^n \left(\mathrm{ESD}(Y_N')\right)(dx) \,,$$

and that the modulus of the above quantity is bounded by M^n, show, by bounded convergence theorem, that (3) and (4) hold, with Y_N replaced by Y_N'. Theorem 1 now shows (14), which, in turn, completes the proof. □

Proof of Theorem 3: As in the preceding proof, let

$$P_N = \psi(Y_N),\ N \geq 1\,.$$

Fix $M > 0$ and let Y'_N be as in (13), M being suppressed in the notation. Theorem 2 implies that

$$\mathrm{EESD}(Y'_N + W_N) \xrightarrow{w} \mu_M \boxplus \nu_\lambda\,,$$

and

$$\mathrm{EESD}(Y'_N W_N) \xrightarrow{w} \mu_M \boxtimes \nu_\lambda$$

as $N \to \infty$, where

$$\mu_M(B) = \mu(B \cap [-M, M]) + \mu\left([-M, M]^c\right) \mathbf{1}(0 \in B)$$

for every Borel set $B \subset \mathbb{R}$. Proposition 4.13 and Corollary 6.7 of [5] imply, respectively, that as $M \to \infty$,

$$\begin{aligned} \mu_M \boxplus \nu_\lambda &\xrightarrow{w} \mu \boxplus \nu_\lambda\,, \text{ and} \\ \mu_M \boxtimes \nu_\lambda &\xrightarrow{w} \mu \boxtimes \nu_\lambda\,. \end{aligned}$$

In view of Facts 3.1 and 3.2, and recalling that convergence in the Lévy metric defined in the latter is equivalent to weak convergence, it suffices to show that

$$\lim_{M\to\infty} \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\mathrm{Rank}(Y_N - Y'_N)\right] = 0\,.$$

However, arguments as in the proof of Theorem 2 show that for M such that

$$\mu(\{-M, M\}) = 0\,,$$

it holds that

$$\limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\mathrm{Rank}(Y_N - Y'_N)\right] = \mu\left([-M, M]^c\right)\,.$$

Hence the proof follows. $\square$

Acknowledgement

The authors are grateful to an anonymous referee for a careful reading of the manuscript and suggesting changes that helped in improving it.

References

Adhikari, K. and Bose, A. (2019). Brown measure and asymptotic freeness of elliptic and related matrices. *Random Matrices: Theory and Applications*, **8(2)**, 1950007.

Anderson, G. W., Guionnet, A. and Zeitouni. O. (2010). *An Introduction to Random Matrices*. Cambridge University Press.

Azoff, E. A. (1974). Borel measurability in linear algebra. *Proceedings of the American Mathematical Society*, **42(2)**, 346 – 350.

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices.* Springer Series in Statistics, New York, second edition.

Bercovici, H. and Voiculescu, D. (1993). Free convolution of measures with unbounded support. *Indiana University Mathematics Journal*, **42**, 733–773.

Hiai, F. and Petz, D. (2000). Asymptotic freeness almost everywhere for random matrices. *Acta Scientiarum Mathematicarum (Szeged)*, **66**, 809–834.

Male, C. (2017). The limiting distributions of large heavy Wigner and arbitrary random matrices. *Journal of Functional Analysis*, **272**, 1 – 46.

Mingo, J. A. and Speicher, R. (2017). *Free Probability and Random Matrices.* Springer.

Nica, A. and Speicher, R. (2006). *Lectures on the Combinatorics of Free Probability.* Cambridge University Press, New York.

Voiculescu, D. (1991). Limit laws for random matrices and free products. *Inventiones Mathematicae*, **104(1)**, 201–220.