Electron. Commun. Probab. **27** (2022), article no. 38, 1–10.
https://doi.org/10.1214/22-ECP481
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Central limit theorem for $C\beta E$ pair dependent statistics in mesoscopic regime*

Ander Aguirre† Alexander Soshnikov‡

Abstract

We extend our results on the fluctuation of the pair counting statistic of the Circular Beta Ensemble $\sum_{i\neq j} f(L_N(\theta_i - \theta_j))$ for arbitrary $\beta > 0$ in the mesoscopic regime $L_N = \mathcal{O}(N^{2/3-\epsilon})$. In addition, we obtain similar results for bipartite statistics.

Keywords: circular beta ensemble; pair counting statistic; central limit theorem.
MSC2020 subject classifications: 60B20; 60F05.
Submitted to ECP on October 21, 2021, final version accepted on July 27, 2022.

1 Introduction

The Circular Beta Ensemble ($C\beta E$) is a random point process of $N \geq 2$ particles on the unit circle; where the joint probability density of the particles $\theta_j \in [0, 2\pi)$, $1 \leq j \leq N$, with respect to the Lebesgue measure is given by:

$$p_{\beta,N}(\theta_1, \cdots, \theta_N) = \frac{1}{Z_{\beta,N}} \prod_{j<k} |e^{i\theta_j} - e^{i\theta_k}|^{\beta}. \tag{1.1}$$

Here $\beta > 0$ and $Z_{\beta,N}$ is the normalization constant:

$$Z_{\beta,N} = (2\pi)^N \frac{\Gamma(1 + \frac{\beta N}{2})}{(\Gamma(1 + \frac{\beta}{2}))^N}.$$

The $C\beta E$ generalizes the classical ensembles of random unitary matrices (COE/CUE/CSE) introduced by Dyson in the 1960s in the context of quantum physics (see e.g. [5]-[8]). The $C\beta E$ can be interpreted as a Coulomb gas, or system of N repelling particles, with β taking the role of the inverse temperature. It can also be viewed as the limiting invariant distribution of a stochastic evolution process on the eigenvalues known as the *circular Dyson Brownian motion* (see [18]). An explicit sparse random matrix model with eigenvalue distribution matching the $C\beta E$ was introduced in [12]. To understand the fluctuation of the eigenvalues one can study *linear statistics* of the form $\sum_{i=1}^{N} f(L_N\theta_i)$, $1 \leq L_N \leq N$. The Central Limit Theorem for linear statistics of eigenvalues of the $C\beta E$ was proved in [9] (see also [3] for the special case $\beta = 2$) and

*Research has been partially supported by the Simons Foundation Collaboration Grant for Mathematicians #312391.
†University of California at Davis, United States of America. E-mail: aaguirre@ucdavis.edu
‡University of California at Davis, United States of America. E-mail: soshniko@math.ucdavis.edu

extended beyond the macroscopic regime to $L_N \gg 1$ in [17] (for $\beta = 2$) and [13] (for arbitrary β). Recently, in [1] and [2] we studied *pair dependent* statistics of the form:

$$S_N(f) = \sum_{i \neq j} f_{L_N}(\theta_i - \theta_j), \tag{1.2}$$

where $f_{L_N}(\theta) = f(L_N\theta)$ for $\theta \in [-\pi, \pi)$ and is extended 2π-periodically to the whole real line. In the global regime we take $L_N = 1$ and f to be a sufficiently smooth function on the unit circle. In the mesoscopic regime ($L_N \to \infty$, $\frac{L_N}{N} \to 0$) and the local regime ($L_N = N$) we consider f to be a smooth compactly supported function on the real line.

The research in [1]-[2] was motivated by a classical result of Montgomery on pair correlation of zeros of the Riemann zeta function [15]-[16]. Assuming the Riemann Hypothesis, Montgomery studied the distribution of the "non-trivial" zeros on the critical line $\{1/2 \pm \gamma_n\}$. In particular, for sufficiently large T, fast decaying f with $\operatorname{Supp} \mathcal{F}(f) \subset [-\pi, \pi]$, and rescaling $\tilde{\gamma_n} = \frac{\gamma_n}{2\pi}\log(\gamma_n)$ he considered the statistic:

$$\sum_{0<\tilde{\gamma}_j \neq \tilde{\gamma}_k < T} f(\tilde{\gamma}_j - \tilde{\gamma}_k).$$

The results of [15]-[16] imply that the two-point correlations of the (rescaled) critical zeros coincide in the limit with the local two point correlations of the eigenvalues of a CUE ($\beta = 2$) random matrix.

The asymptotic distribution of the pair counting statistic (1.2) depends on the speed of the growth of L_N, regularity (smoothness) properties of the test function f, and the value of the inverse temperature $\beta > 0$. The results of [1] deal with the limiting behavior of (1.2) in three different regimes, namely macroscopic ($L_N = 1$), mesoscopic ($1 \ll L_N \ll N$) and microscopic ($L_N = N$). In the macroscopic (unscaled) $L_N = 1$ case it was shown that $S_N(g)$ has a non-Gaussian fluctuation in the limit $N \to \infty$ provided g is a sufficiently smooth function on the unit circle. In particular (see Theorem 2.1 in [1]),

$$S_N(g) - \mathbb{E}S_N(g) \xrightarrow{\mathcal{D}} \frac{4}{\beta}\sum_{m=1}^{\infty} \hat{g}(m)m(\varphi_m - 1)$$

where φ_m are i.i.d. exponential random variables with $\mathbb{E}(\varphi_m) = 1$, and

$$\hat{g}(m) = \frac{1}{2\pi}\int_0^{2\pi} g(x)e^{-imx}dx, \quad m \in \mathbb{Z},$$

are the Fourier coefficients of g. The result was proved under the optimal condition $g' \in L^2(\mathbb{T})$ for $\beta = 2$, and under slightly sub-optimal conditions for $\beta \neq 2$.

In the case of a slowly growing variance (i.e. when $\sum_{m=-N}^{N} |\hat{f}(m)|^2 m^2$ is a slowly growing sequence) the asymptotic fluctuation becomes Gaussian (see [2]). The determinantal structure of the correlation functions of the CUE ($\beta = 2$) enabled us to study the pair counting statistic up to the microscopic regime. In particular, a pair counting statistic was shown to have limiting Gaussian fluctuation provided f is sufficiently smooth. However, for arbitrary $\beta \neq 2$ the growth of L_N in [1] was restricted to $L_N = o(N^{\varepsilon})$. In this note, we extend the results of [1] for $S_N(f)$ and arbitrary $\beta > 0$ to $L_N = \mathcal{O}(N^{2/3-\epsilon})$ in the mesoscopic regime. Next, we formulate the main result.

Theorem 1.1. *Let L_N be growing to infinity so that $L_N = \mathcal{O}(N^{2/3-\epsilon})$, where $\epsilon > 0$ is arbitrary small and $f \in C_c^\infty$ be even, smooth and compactly supported. Then as $N \to \infty$*

$$\frac{S_N(f) - \mathbb{E}S_N(f)}{\sqrt{L_N}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{\pi\beta^2}\int_{\mathbb{R}} |\hat{f}(\xi)|^2\xi^2 d\xi\right). \tag{1.3}$$

Here $\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-itx} dx$ denotes the Fourier transform of f, and the notation $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Remark 1.2. We do not expect the exponent $2/3$ to be optimal and note that for $\beta = 2$ the result of Theorem 1.1 holds for all $L_N = o(N)$ (see Theorem 2.4 in [1]). The main difficulty in extending the result of the theorem to all $L_N = o(N)$ lies in controlling the joint distribution of the traces of powers (2.2) for k growing sufficiently fast. Even for one of three special values of β, namely $\beta = 4$, the variance of $|k|^{-1/2} \operatorname{Tr} U_N^k$, somewhat unexpectedly, is unbounded for $|k|$ sufficiently close to N ([11]). For $\beta = 2$ one can avoid these additional difficulties by analyzing explicit formulas for the joint cumulants of $\operatorname{Tr} U_N^k$.

Next we consider bipartite statistics:

$$B_N(f) = \sum_{i,j} f_{L_N}(\tau_i - \theta_j), \tag{1.4}$$

where $\{\tau_i\}_{i=1}^N$ and $\{\theta_j\}_{j=1}^N$ come from different ensembles on the unit circle, for example two independent $C\beta E$ ensembles. In particular, the following result holds.

Theorem 1.3. *Let $\{\tau_i\}_{i=1}^N$ and $\{\theta_j\}_{j=1}^N$ be point configurations from two independent $C\beta E$ ensembles, $L_N = \mathcal{O}(N^{2/3-\epsilon})$, with $\epsilon > 0$ is arbitrary small and $f \in C_c^\infty$ be even, smooth and compactly supported. Then $\mathbb{E} B_N(f) = \frac{N^2}{2\pi L_N} \int_{\mathbb{R}} f(x) dx$ and*

$$\frac{B_N(f) - \mathbb{E} B_N(f)}{\sqrt{L_N}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{2}{\pi\beta^2} \int_{\mathbb{R}} |\hat{f}(t)|^2 t^2 dt\right). \tag{1.5}$$

Remark 1.4. The result of Theorem 1.3 could be extended to study bipartite statistics involving different sizes $N_1,\ N_2$ and different values $\beta_1,\ \beta_2 > 0$.

Remark 1.5. The asymptotic behavior of bipartite statistics $B_N(g)$ in the global regime $L_N = 1$ can be treated similarly to Theorem 2.1 in [1]. In particular, the mean of a bipartite statistic is given by $\mathbb{E} \sum_{i,j=1}^N g(\tau_i - \theta_j) = \hat{g}(0) N^2$, and, for a sufficiently smooth test function g,

$$B_N(g) - \mathbb{E} B_N(g) \xrightarrow{\mathcal{D}} \frac{2}{\beta} \sum_{m=1}^{\infty} \hat{g}(m) m \phi_m,$$

where ϕ_m are i.i.d. centered double exponential (Laplace) random variables with $\operatorname{Var}(\phi_m) = 2$.

We will denote a $N \times N$ random $C\beta E$ matrix by U_N. The notation $a_N = o(b_N)$ means that $a_N / b_N \to 0$ as $N \to \infty$. The notation $a_N = \mathcal{O}(b_N)$ means that the ratio of a_N and b_N is bounded in N. In Section 2, we recall some preliminary material. Theorems 1.1 and 1.3 are proved in Section 3. Local bipartite statistics are studied in Section 4.

2 Preliminary material

Consider an even real-valued function g on the unit circle that can be represented by the Fourier series:

$$g(x) = \sum_{m=-\infty}^{\infty} \hat{g}(m) e^{imx}, \quad \hat{g}(m) = \frac{1}{2\pi} \int_0^{2\pi} g(x) e^{-imx} dx.$$

Fourier expanding the pair dependent statistic we obtain:

$$\sum_{1 \le i \ne j \le N} g(\theta_i - \theta_j) = 2 \sum_{m=1}^{\infty} \hat{g}(m) \left| \sum_{j=1}^{N} \exp(im\theta_j) \right|^2 + \hat{g}(0) N^2 - N g(0).$$

Let now $f \in C_c^\infty(\mathbb{R})$ be even, smooth and compactly supported, and $L_N \to \infty$ as $N \to \infty$. Then for sufficiently large L_N we can view $f(L_N x)$ as a smooth compactly supported function on the unit circle and the pair counting statistic (1.2) can be written as:

$$S_N(f) = \sum_{1\le j\ne k\le N} f_{L_N}(\theta_j - \theta_k) = \frac{\hat{f}(0)N^2}{\sqrt{2\pi}L_N} + \sum_{k=1}^{\infty} \frac{2}{\sqrt{2\pi}L_N}\hat{f}\left(\frac{k}{L_N}\right)\left(|\operatorname{Tr} U_N^k|^2 - N\right), \quad (2.1)$$

where $\hat{f}$ denotes the Fourier transform of f. We will use the following notation for traces of powers of a random unitary matrix:

$$T_N^{(k)} := \sum_{j=1}^{N} e^{ik\theta_j} = \operatorname{Tr} U_N^k, \quad k = 0, \pm 1, \pm 2, \dots. \quad (2.2)$$

Our proof relies on the results of Johansson and Lambert [10], who estimated the Wasserstein-2 distance between a random vector of traces of powers of a $C\beta E$ matrix U_N and a random vector of independent Gaussians of matching variance. We refer to Theorem 1.5 in [10]. In the Appendix, we justify the claim in Remark 1.1 of [10] that enables us to extend results to arbitrary $\beta > 0$ using Proposition 2.3 below.

Let $T_d = \left(T_N^{(k)}\right)_{k=1}^d$ be the vector of the traces of the first d powers of a random $C\beta E$ matrix U_N and $G_d = \left(\sqrt{\frac{2}{\beta}k}Z_k\right)_{k=1}^d$, where Z_k are i.i.d. complex $\mathcal{N}(0,1)$. Reformulated in terms of the pair (T_d, G_d), their result states:

Proposition 2.1 (Johansson and Lambert [10])**.** *Let $2d \le N$. and $\{e^{i\theta_j}\}_{j=1}^N$ be drawn from the $C\beta E$ with $\beta > 0$. Then as $N \to \infty$ we have the following bound:*

$$\mathcal{W}_2(T_d, G_d) = \mathcal{O}\left(\frac{d^2}{N}\right). \quad (2.3)$$

We recall that the Wasserstein-p distance between two probability measures on a normed space is defined as:

$$\mathcal{W}_p(\mu,\nu) := \left(\inf\{\mathbb{E}||X-Y||^p : (X,Y) \text{ is a r.v. such that } X\sim\mu,\ Y\sim\nu\}\right)^{1/p},$$

where $p \ge 1$ and the notation $X \sim \mu$ means that a random variable X has probability distribution μ. The Wasserstein distance takes values in $[0,\infty]$. For $p = 1$, the Wasserstein-1 distance $\mathcal{W}_1(\mu,\nu)$ is also known as the Kantorovich-Monge-Rubinstein metric and can be equivalently written as:

$$\mathcal{W}_1(\mu,\nu) := \sup\left\{\left|\int f d\mu - \int f d\nu\right| : f \text{ is 1-Lipschitz}\right\}.$$

In other words, the supremum is taken over all real-valued functions f that satisfy $|f(x) - f(y)| \le d(x,y)$, where d is the metric on the underlying metric space.

Remark 2.2. Important earlier results of Döbler and Stolz [4] and Webb [18] bounded from above the Wasserstein-1 distance $\mathcal{W}_1(T_d, G_d)$. In particular, it was shown in [4] that for $\beta = 2$ one has $\mathcal{W}_1(T_d, G_d) = \mathcal{O}(d^{5/2}/N)$. Webb proved in [18] the bound $\mathcal{W}_1(T_d, G_d) = \mathcal{O}(d^{7/2}/N)$ for arbitrary β. These results are strengthened by (2.3) since

$$\mathcal{W}_1(\mu,\nu) \le d^{1/2}\mathcal{W}_2(\mu,\nu).$$

Potential further improvements of these bounds should lead to improvement of the results of Theorem 1.1 and Theorem 1.3.

Finally, we will require the following bound on the moments of $T_N^{(k)}$.

Proposition 2.3 (Jiang and Matsumoto [11])**.** *Let $\{e^{i\theta_j}\}_{j=1}^N$ be drawn from the $C\beta E$ and T_N^k defined as in (2.2). For $0 \le k \le N$ we have:*

$$\mathbb{E}|T_N^{(k)}|^{2m} \le \left(1 + \frac{\left|\frac{2}{\beta} - 1\right|}{N - K + \frac{2}{\beta}} \mathbf{1}(\beta > 2)\right)^K \times \left(\frac{2}{\beta}\right)^m \times k^m \times m! \qquad (2.4)$$

where $K = km \le N$.

3 Mesoscopic case

Proof of Theorem 1.1. Let $L_N = \mathcal{O}\left(N^{2/3-\epsilon}\right)$ be going to infinity with N, $\epsilon > 0$ be arbitrary small, and $d = \lfloor L_N N^{\epsilon/2} \rfloor$. Going forward we can ignore the constant term

$$\frac{1}{\sqrt{2\pi} L_N} \hat{f}(0) N^2 - N f(0)$$

appearing in (2.1) since it disappears upon centralization. Taking into account the smoothness of f we will approximate the pair counting statistic $S_N(f)$ by a truncated version

$$S_{N,d}(f) = \sum_{k=1}^{d} \frac{2}{\sqrt{2\pi} L_N} \hat{f}\left(\frac{k}{L_N}\right) |T_N^{(k)}|^2, \qquad (3.1)$$

and compare its distribution with the distribution of

$$S_d = \sum_{k=1}^{d} \frac{2}{\sqrt{2\pi} L_N} \hat{f}\left(\frac{k}{L_N}\right) \frac{2k}{\beta} |Z_k|^2. \qquad (3.2)$$

In Lemma 3.1 we show that the error of the approximation (3.1), namely

$$V_{N,d}(f) = \sum_{k>d} \frac{2}{\sqrt{2\pi} L_N} \hat{f}\left(\frac{k}{L_N}\right) |T_N^{(k)}|^2$$

is negligible in the limit of large N. In Lemma 3.2., we show that (3.2) converges in distribution to a centered Gaussian with variance $\frac{4}{\pi\beta^2} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \xi^2 d\xi$. Finally, the main result follows from the Wasserstein distance bound in Lemma 3.3.

Lemma 3.1. *Let $f \in C_c^\infty(\mathbb{R})$ be even, smooth and compactly supported and further assume that $L_N = o(N^{2/3-\epsilon})$ and $d = \lfloor L_N N^{\epsilon/2} \rfloor$. Then*

$$\frac{V_{N,d}(f) - \mathbb{E}\left(V_{N,d}(f)\right)}{\sqrt{L_N}} \xrightarrow{L^1} 0. \qquad (3.3)$$

Proof. From the triangle inequality:

$$\mathbb{E}|\mathrm{LHS}(3.3)| \le \frac{1}{\sqrt{L_N}} \sum_{k \ge d} \frac{2}{\sqrt{2\pi} L_N} \left|\hat{f}\left(\frac{k}{L_N}\right)\right| \mathrm{Var}(T_N^{(k)}), \qquad (3.4)$$

where we recall that $T_N^{(k)}$ denotes $\mathrm{Tr}(U_N^k)$, the trace of the k-th power of a random unitary matrix U_N. The bound of Proposition 2.3 gives us that $\mathrm{Var}(T_N^{(k)}) \le Ck$ for $k \le N/3$. For $k \ge N/3$ we may bound it trivially by N^2. Next, we will use the fact that the Fourier transform of a function $f \in C_c^\infty(\mathbb{R})$ is in Schwartz space and thus decays faster than any power.

$$\text{RHS}(3.4) \leq \frac{1}{\sqrt{L_N}} \sum_{k \geq d}^{N/3} \frac{2C'k}{\sqrt{2\pi} L_N} \left(\frac{k}{L_N}\right)^{-\gamma} + \frac{1}{\sqrt{L_N}} \sum_{k > N/3} \frac{2C'' N^2}{\sqrt{2\pi} L_N} \left(\frac{k}{L_N}\right)^{-\gamma}$$

$$= \mathcal{O}\left(\frac{L_N^{\gamma-3/2}}{d^{\gamma-2}} + N^{3-\gamma} L_N^{\gamma-3/2}\right). \tag{3.5}$$

Setting $\gamma = \frac{2}{\epsilon}$ with $\epsilon > 0$ sufficiently small we obtain that the r.h.s. of (3.5) goes to zero as $N \to \infty$. □

Lemma 3.2. *We have the following convergence in distribution:*

$$\frac{S_d - \mathbb{E} S_d}{\sqrt{L_N}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{\pi \beta^2} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \xi^2 d\xi\right). \tag{3.6}$$

Proof. For $k \geq 1$ let us denote

$$a_k = \frac{2}{\sqrt{2\pi} L_N} \hat{f}\left(\frac{k}{L_N}\right) \frac{2k}{\beta}, \qquad S_d = \sum_{k=1}^{d} a_k |Z_k|^2. \tag{3.7}$$

Then we have

$$\frac{\text{Var}(S_d)}{L_N} = \frac{1}{L_N} \times \sum_{k=1}^{d} \frac{8}{\pi \beta^2} \left|\hat{f}\left(\frac{k}{L_N}\right)\right|^2 \left|\frac{k}{L_N}\right|^2 \longrightarrow \frac{4}{\pi \beta^2} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 \xi^2 d\xi.$$

Since $\max\{a_k^2\}_{k=1}^d = o(\sum_1^d a_k^2)$, the result follows from the Lindeberg-Feller theorem. □

Lemma 3.3. *Let $d \leq N/2$. Then we have the following Wasserstein-1 bound:*

$$\mathcal{W}_1\left(\frac{S_{N,d}}{L_N^{1/2}}, \frac{S_d}{L_N^{1/2}}\right) = \mathcal{O}\left(\frac{d^3}{N L_N^{3/2}}\right), \tag{3.8}$$

Therefore L.H.S(3.8)$\to 0$ if $L_N = \mathcal{O}\left(N^{2/3-\epsilon}\right)$ and $d = \lfloor L_N N^\epsilon \rfloor$.

Proof. Let $(\cdot,\cdot)$ denote standard Euclidean inner product in $\mathbb{C}^d$ and A be a diagonal $d \times d$ matrix with $A_{k,k} = a_k,\ 1 \leq k \leq d$, with a_k as in (3.7). With T_d and G_d as defined in Proposition 2.1 we have:

$$\begin{aligned} L_N^{-1/2}|S_{N,d} - S_d| &= L_N^{-1/2}|(AT_d, T_d) - (AG_d, G_d)| \\ &\leq L_N^{-1/2}|(AT_d, T_d - G_d) + (A(T_d - G_d), G_d)| \\ &\leq L_N^{-1/2}||A|| \times \|T_d - G_d\|_2^2 + 2L_N^{-1/2}||A|| \times \|T_d - G_d\|_2 \times \|G_d\|_2 . \end{aligned}$$

Here $||A||$ denotes the operator norm of A and $\|\cdot\|_2$ is the vector l^2 norm. Since A is a diagonal matrix, one has $||A|| = \max\{|a_k|,\ 1 \leq k \leq d\} = \mathcal{O}\left(L_N^{-1}\right)$. Proposition 2.2 allows us to choose the components of the vector G_d so that

$$\mathbb{E}(\|T_d - G_d\|_2^2) = \mathcal{O}\left(\frac{d^4}{N^2}\right). \tag{3.9}$$

Using $\mathbb{E}(\|G_d\|_2^2) = \mathcal{O}\left(d^2\right)$ and the Cauchy-Schwartz inequality we arrive at:

$$\begin{aligned} L_N^{-1/2}\mathbb{E}|S_{N,d} - S_d| &\le L_N^{-3/2}\mathbb{E}(\|T_d - G_d\|_2^2) + 2L_N^{-3/2}\mathbb{E}(\|T_d - G_d\|_2 \times \|G_d\|_2) \\ &\le \mathcal{O}\left(\frac{d^4}{N^2 L_N^{3/2}}\right) + \mathcal{O}\left(\frac{d^3}{N L_N^{3/2}}\right), \end{aligned}$$ □

Theorem 1.1 follows directly from combining Lemmas 3.1, 3.2 and 3.3. □

The proof of Theorem 1.3 is quite similar with minor changes (such as replacing the approximation of the quadratic form (AT_d, T_d) by the approximation of the bilinear form $(AT_d, \mathcal{T}_d)$, where $\mathcal{T}_d$ is an independent copy of T_d. The details are left to the reader.

4 Local bipartite statistics

In this section we study bipartite statistics in the local regime $L_N = N$ for $\beta = 2$. We recall that in Theorem 2.5 in [1] we considered $S_N(f) = \sum_{1\le i\ne j\le N} f_N(\theta_i - \theta_j)$, $f_N(\theta) = f(N\theta)$ for $\theta \in [-\pi, \pi)$, $f_N(\theta + 2\pi) = f_N(\theta)$, and proved the following proposition.

Theorem 4.1. *Let $\beta = 2$ and $f \in C_c^\infty(\mathbb{R})$ be an even, smooth, compactly supported function on the real line. Then $(S_N(f) - \mathbb{E}S_N(f))N^{-1/2}$ converges in distribution to centered real Gaussian random variable with the variance*

$$\begin{aligned} \frac{1}{\pi}\int_{\mathbb{R}} |\hat{f}(t)|^2 \min(|t|, 1)^2 dt - \frac{1}{\pi}\int_{|s-t|\le 1, |s|\vee|t|\ge 1} \hat{f}(t)\hat{f}(s)(1 - |s-t|)dsdt \\ -\frac{1}{\pi}\int_{0\le s,t\le 1, s+t>1} \hat{f}(s)\hat{f}(t)(s+t-1)dsdt. \end{aligned} \tag{4.1}$$

Consider

$$B_N(f) = \sum_{1\le i,j\le N} f(N(\tau_i - \theta_j)), \tag{4.2}$$

where the point configuration $\{\tau_i\}_{i=1}^N$ comes from one of the following three ensembles: (i) an independent copy of $C\beta E$, (ii) a collection of i.i.d. uniformly distributed points on the unit circle, (iii) evenly spaced deterministic sequence. Then the following result holds.

Theorem 4.2. *Let $f \in C_c^\infty(\mathbb{R})$ be an even, smooth, compactly supported function on the real line. Consider $B_N(f)$ defined in (4.2) where $\{\theta_j\}_{j=1}^N$ be a CUE configuration and $\{\tau_i\}_{i=1}^N$ comes from one of the following three ensembles:*

(i) an independent copy of a CUE;
(ii) a sequence of i.i.d. uniformly distributed points on the unit circle;
(iii) an evenly spaced deterministic sequence $\tau_i = \frac{2\pi i}{N},\ i = 1, \dots, N$.

Then $\mathbb{E}B_N(f) = \frac{N}{2\pi}\int_{\mathbb{R}} f(x)dx$, and $(B_N(f) - \mathbb{E}B_N(f))N^{-1/2}$ converges in distribution to centered real Gaussian random variable with variance $\sigma^2(f)$, where

$$\sigma^2(f) = \begin{cases} \frac{1}{2\pi}\int_{\mathbb{R}} |\hat{f}(t)|^2 \min(|t|, 1)^2 dt & \text{in the case } (i), \\ \frac{1}{2\pi}\int_{\mathbb{R}} |\hat{f}(t)|^2 \min(|t|, 1) dt & \text{in the case } (ii), \\ \frac{1}{2\pi}\sum_{l\ne 0} |\hat{f}(l)|^2 & \text{in the case } (iii). \end{cases} \tag{4.3}$$

Proof of Theorem 4.2. As always $T_N^{(k)} := \sum_{j=1}^N e^{ik\theta_j}$ and denote $\mathcal{T}_N^{(k)} = \sum_{j=1}^N e^{ik\tau_j}$, where $k \in \mathbb{Z}$. Then

$$B_N^c(f) := B_N(f) - \mathbb{E}B_N(f) = \sum_{k\ne 0} \frac{1}{\sqrt{2\pi}N}\hat{f}\left(\frac{k}{N}\right) T_N^{(k)}\mathcal{T}_N^{(-k)}.$$

We consider the case (i) first. Using independence and $\mathbb{E}|T_N^{(k)}|^2 = \min(|k|, N)$, one has

$$\mathrm{Var}(B_N(f)) = \sum_{k\neq 0} \frac{1}{2\pi N^2} \left|\hat{f}\left(\frac{k}{N}\right)\right|^2 \min(|k|, N)^2 = \frac{N}{2\pi}\int_{\mathbb{R}} |\hat{f}(t)|^2 \min(|t|, 1)^2 dt(1+o(1)). \tag{4.4}$$

To study higher moments, we use cumulant bounds and power counting. For $l \geq 2$ we have:

$$\mathbb{E}(B_N^c(f))^l = (2\pi)^{-l/2} N^{-l} \sum_{k_1,\dots,k_l\neq 0} \prod_{i=1}^{l} \hat{f}\left(\frac{k_i}{N}\right) \mathbb{E}\prod_{i=1}^{l} T_N^{(k_i)} \mathbb{E}\prod_{j=1}^{l} \mathcal{T}_N^{(-k_j)}. \tag{4.5}$$

To evaluate the moments we use Lemma 5.2 from [1] that allows us to estimate joint cumulants of the traces of powers. It was shown that for any $n \geq 1$, $\kappa_n^{(N)}(k_1,\dots,k_n)$, the n-th joint cumulant of $T_N^{(k)}$'s is $O(N)$, uniformly in $k_1,\dots k_n$. In addition, $\kappa_2^{(N)}(k_1,k_2) = \min(N, |k_1|)1_{k_1+k_2=0}$. This implies that for odd values of $l = 2m+1$

$$\mathbb{E}(B_N^c(f))^{2m+1} = O(N^m), \tag{4.6}$$

and for even values $l = 2m$ the main contribution to (4.5) comes from the l-tuples $(k_1,\dots,k_m)$ that could be split into pairs $(k,-k)$. By power counting we then obtain

$$\mathbb{E}(B_N^c(f))^{2m} = \sigma^{2m}(2m-1)!!N^m(1+o(1)), \tag{4.7}$$

and the moment convergence implies CLT. The considerations in the case (ii) are very similar. In particular,

$$\mathrm{Var}(B_N(f)) = \sum_{k\neq 0} \frac{1}{2\pi N^2} \left|\hat{f}\left(\frac{k}{N}\right)\right|^2 \min(|k|, N)N = \frac{N}{2\pi}\int_{\mathbb{R}} |\hat{f}(t)|^2 \min(|t|, 1) dt(1+o(1)). \tag{4.8}$$

We leave higher order estimates to the reader. Finally, we turn our attention to the case (iii). In this case we have:

$$B_N^c(f) = \sum_{l\neq 0} \frac{1}{\sqrt{2\pi}} \hat{f}(l) T_N^{(lN)}. \tag{4.9}$$

This readily implies that:

$$\mathrm{Var}(B_N(f)) = N\frac{1}{2\pi}\sum_{l\neq 0} |\hat{f}(l)|^2. \tag{4.10}$$

The Central Limit Theorem again follows from the cumulant bounds and power counting. It should be noted that random variables $T_N^{(lN)}$, $l \in \mathbb{Z}\backslash\{0\}$, are not independent but are identically distributed – they have the same distribution as $\sum_{j=1}^{N} e^{i\tau_j}$. □

5 Appendix

In this appendix we discuss the details in Remark 1.1 of [10] that justify the statement of Proposition 2.1 for arbitrary $\beta > 0$. In Theorem 1.5 of [10] Johansson and Lambert provide the bound:

$$\mathcal{W}_2(X, G) \leq \mathcal{O}\left(\frac{d^{3/2}}{N}\right),$$

where

$$X = \sqrt{\frac{\beta}{2k}} T_d \qquad G = \sqrt{\frac{\beta}{2k}} G_d.$$

Now define the map $X : \mathbb{T}^n \to \mathbb{R}^{2d}$ by $X_{2k-1} = \Re T_N^{(k)}$ and $X_{2k} = \Re T_N^{(k)}$, $1 \leq k \leq d$, and set $\mathbf{\Gamma}$ to be a $2d$-dimensional square matrix with the entries $\mathbf{\Gamma}_{k,l} = \nabla X_k \cdot \nabla X_l$. We also define

$$\mathbf{K} = N \cdot \mathrm{diag}(1, 1, 2, 2, \cdots, d, d)$$
$$\xi = \big(\Re\zeta_1, \Im\zeta_1, \Re\zeta_2, \Im\zeta_2, \cdots, \Re\zeta_d, \Im\zeta_d\big),$$

where for $k \geq 1$:

$$\zeta_k = \sqrt{\frac{k}{2}} \sum_{\ell=1}^{k-1} \sqrt{\ell(k-\ell)} T_N^{(\ell)} T_N^{(k-\ell)}.$$

We refer to Section 7 of [10] (specifically Lemmas 7.2 and 7.3) for full details of the following lemma.

Lemma 5.1. *For all $N, d \in \mathbb{N}$ and for any positive definite diagonal matrix $\mathbf{K}$ of size $2d \times 2d$, we have*

$$\mathcal{W}_2(X, G) \leq \sqrt{\mathbb{E}_N \left[|\mathbf{K}^{-1}\xi|^2\right]} + \sqrt{\mathbb{E}_N \left[\|\mathbf{I} - \mathbf{K}^{-1}\mathbf{\Gamma}\|^2\right]}, \tag{5.1}$$

where $\|\cdot\|_{HS}$ denotes the Hilbert–Schmidt norm.

We arrive at the desired bound with the following lemma. This corresponds to lemmas 7.4 and 7.5 of [10] where we used the moment estimates of Jiang-Matsumoto instead.

Lemma 5.2. *With $\mathbf{K}, \mathbf{\Gamma}$ and ξ as above we have:*

$$\mathbb{E}_N|\mathbf{K}^{-1}\xi|^2 = \mathcal{O}\left(\frac{d^3}{N^2}\right) \qquad \mathbb{E}_N[||\mathbf{I} - \mathbf{K^{-1}\Gamma}||^2_{HS}] = \mathcal{O}\left(\frac{d^3}{N^2}\right). \tag{5.2}$$

Proof. From page 37 of [10] we have the identity:

$$||\mathbf{I} - \mathbf{K^{-1}\Gamma}||^2_{HS} = \frac{\beta}{N^2} \sum_{1\leq k<l\leq d} |T_N^{(l-k)}|^2 + |T_N^{(l+k)}|^2 \quad + \frac{5\beta}{2N^2} \sum_{k=1}^{d} |\Re(T_N^{(k)})|^2$$

Since $k \leq L_N$, Proposition 2.3 gives that $\mathbb{E}|T_N^{(k)}|^2 \leq C_\beta k$ and thus

$$\mathbb{E}_N\left[\|\mathbf{I} - \mathbf{K}^{-1}\mathbf{\Gamma}\|^2_{HS}\right] \leq C_\beta \times \left(\sum_{1\leq k<\ell\leq d} \frac{\ell}{N^2} + \sum_{k=1}^{d} \frac{k}{N^2} \right) = \mathcal{O}\left(\frac{d^3}{N^2}\right).$$

From page 36 of [10] we have the identity:

$$|\mathbf{K}^{-1}\xi|^2 = \sum_{1\leq \ell,\ell'<k\leq d} \frac{\beta^2}{8kN^2} T_N^{(\ell)} T_N^{(k-\ell)} \overline{T_N^{(\ell')} T_N^{(k-\ell')}}.$$

Since $k \leq L_N$, Proposition 2.3 gives us (see Theorem 1 part (b) of [11]):

$$\mathbb{E}_N\left[T_N^{(\ell)} T_N^{(k-\ell)} \overline{T_N^{(\ell')} T_N^{(k-\ell')}}\right]$$
$$\leq C_\beta \left(\ell(k-\ell)\mathbf{1}_{\{\ell=\ell', \ell\neq k/2\}} + \ell(k-\ell)\mathbf{1}_{\{\ell=k-\ell', \ell\neq k/2\}}) + \frac{k^2}{4}\mathbf{1}_{\ell=\ell'=k/2} \right).$$

Taking expectations we have:

$$\mathbb{E}_N\left[|\mathbf{K}^{-1}\xi|^2\right] \leq \frac{K_\beta}{N^2} \sum_{1\leq \ell<k\leq d} \frac{\ell(k-\ell)}{k} = \mathcal{O}\left(\frac{d^3}{N^2}\right). \qquad \square$$

References

[1] Aguirre A., Soshnikov A. and Sumpter J.: Pair Dependent Linear Statistics for CβE. *Random Matrices: Theory and Appl.* https://doi.org/10.1142/S2010326321500350 MR4379540

[2] Aguirre A. and Soshnikov A.: A Note on Pair Dependent Linear Statistics with Slowly Growing Variance. *Theor. Math. Phys.* **205** no. 3, (2020), 502–512. MR4181088

[3] Diaconis, P, Shahshahani, M., *On Eigenvalues of Random Matrices.* J. Appl. Probab. **31A**, (1994), 49–62 MR1274717

[4] Döbler, C, Stolz, M., *Stein's Method and the Multivariate CLT for Traces of Powers on the Classical Compact Groups.* Electron. J. Probab, **16**, (2011), paper no. 86, 2375–2405. MR2861678

[5] Dyson, F.J.: Statistical theory of the energy levels of complex systems. I. *J. Math. Phys.* **3**, (1962), 140–156. MR0143556

[6] Dyson, F.J.: Statistical theory of the energy levels of complex systems. II. *J. Math. Phys.* **3**, (1962),157–165. MR0143557

[7] Dyson, F.J.: Statistical theory of the energy levels of complex systems. III. *J. Math. Phys.* **3**, (1962), 166–175. MR0143558

[8] Dyson, F.J.: Correlations between the eigenvalues of a random matrix. *Comm. Math. Phys.* **19**, (1970), 235–250. MR0278668

[9] Johansson, K.: On Szego's Asymptotic Formula for Toeplitz Determinants and Generalizations. Bull. Sci. Math. (2) 112 (1988), no. 3, 257–304 MR0975365

[10] Johansson,K., Lambert, G. *Multivariate Normal Approximation for Traces of Random Unitary Matrices*, arXiv:2002.01879. MR4348683

[11] Jiang, T. and Matsumoto, S.: Moments of Traces of Circular β-ensembles. *Ann. Probab.* **43** no 6., (2015), 3279–3336 MR3433582

[12] Killip, R., Nenciu, I.: Matrix models for circular ensembles. *Int. Math. Res. Not.* no. 50, (2004), 2665–2701. MR2127367

[13] Lambert, G.: Mesoscopic central limit theorem for the circular beta-ensembles and applications. *Electronic Journal Probab.* **26** no. 7, (2021), 33pp. MR4216520

[14] Leadbetter, M.R., Lindgren, G., Rootze, H.: Extremes and Related Properties of Random Sequences and Processes, *Springer*, New York, (1983). MR0691492

[15] Montgomery, H.L.: On pair correlation of zeros of the zeta function. *Proc. Sympos. Pure Math.* **24**, (1973), 181–193. MR0337821

[16] Montgomery, H.L.: Distribution of the zeros of the Riemann zeta function. *Proc. Internat. Congr. Math.* **1**, Vancouver, BC (1974), 379–381. MR0419378

[17] Soshnikov, A.: Central Limit Theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.* **28**, (2000), 1353–1370. MR1797877

[18] Webb, C.: Linear statistics of the circular ensemble, Stein's method, and circular Dyson Brownian motion. *Elec. J. Probab.* **20**, (2015), No. 104, 21pp. MR3485367