Topics in Random matrices
Preliminary version

Michael Anshelevich

August 13, 2018

Contents

Chapter 1

Overview of asymptotic random matrix results.

Brief review of probability theory.

Probability space (Ω, Σ, P). Here

$\Omega =$ set.

$\Sigma = \sigma$-algebra of measurable subsets of Ω.

$P =$ probability measure on (Ω, Σ), $P(\Omega) = 1$.

$X =$ random variable $=$ (real-valued) measurable function on Ω.

$\mathbb{E} =$ expectation functional,

$$\mathbb{E}[X] = \int X \, dP = \int X(\omega) \, dP(\omega)$$

whenever defined.

$\mu_X =$ distribution of X = probability measure on $\mathbb{R}$,

$$\mu_X(A) = P(X \in A) = P(\omega \in \Omega : X(\omega) \in A).$$

Also for $f \in C_b(\mathbb{R})$,

$$\mathbb{E}[f(X)] = \int f(x) \, d\mu_X(x).$$

A *random matrix* is an $N \times N$ matrix of random variables = M_N-valued random variable.

These come up in a variety of models and settings.

Remark 1.1. *In this course*, our main interest is in the behavior of $N \times N$ random X as $N \to \infty$. So often a "random matrix X" really means a sequence $(X_N)_{N=1}^{\infty}$, each X_N $N \times N$. There are also many exact results for finite N, which we will omit.

1.1 Gaussian orthogonal ensemble GOE$_N$.

Fix $N \geq 1$. For $1 \leq i \leq j \leq N$, let $B_{ij} \sim \mathcal{N}(0,1)$ be independent standard normal variables. Define X_N an $N \times N$ matrix by

$$[X_N]_{ij} = [X_N]_{ji} = \frac{1}{\sqrt{N}} B_{ij}, \qquad i < j$$

(so that X_N is *symmetric*), and

$$[X_N]_{ii} = \frac{\sqrt{2}}{\sqrt{N}} B_{ii}.$$

Usually the Gaussian Orthogonal Ensemble is defined without the $\frac{1}{\sqrt{N}}$ normalization. We include this normalization from the very beginning, to have

$$\frac{1}{N}\operatorname{Tr}[X_N] = \frac{1}{N}\sum_{i=1}^{N} X_{ii}$$

and

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}[X_N]\right] = 0.$$

$$\frac{1}{N}\operatorname{Tr}[X_N^2] = \frac{1}{N}\sum_{i,j=1}^{N} X_{ij}X_{ji} = \frac{1}{N}\left(2\sum_{i<j}\frac{1}{N}B_{ij}^2 + \sum_{i}\frac{2}{N}B_{ii}^2\right) = \frac{2}{N^2}\sum_{i\leq j}^{N} B_{ij}^2,$$

and so

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}[X_N^2]\right] = \frac{2}{N^2}\frac{N(N+1)}{2} \to 1.$$

Theorem (Wigner's Theorem I)**.** *Let $X_N \sim GOE_N$. Then as $N \to \infty$,*

$$\underbrace{\frac{1}{N}\operatorname{Tr}[X^{2k}]}_{random} \to \underbrace{c_k}_{number} = \textit{ Catalan number } = \frac{1}{k+1}\binom{2k}{k}$$

and

$$\frac{1}{N}\operatorname{Tr}[X^{2k+1}] \to 0.$$

We will see a combinatorial interpretation of c_k soon.

Convergence in what sense?

Definition 1.2. $(x_N)_{N=1}^{\infty}$ random variables.

$x_N \to a$ *in expectation* if

$$\mathbb{E}[x_N] \to a.$$

$x_N \to a$ ***in probability*** if $\forall \delta > 0$,

$$P(|x_N - a| \geq \delta) \to 0.$$

$x_N \to a$ ***a.s. (almost surely)*** if

$$P(x_N \not\to a) = 0.$$

Review the relation between these modes of convergence.

The theorem above holds in all three senses (Wigner 1955, 1958, Grenander ?, Arnold 1967).

Remark 1.3 (Second point of view: matrix-valued distribution)**.** Take $X \sim \text{GOE}_N$ as before. Forgetting the matrix structure, we may think of X as an $R^{N(N+1)/2}$-valued random variable, jointly Gaussian with joint density

$$\begin{aligned}\frac{1}{Z}\prod_{i<j}\exp(-\frac{x_{ij}^2}{2(1/N)})\prod_i \exp(-\frac{x_{ii}^2}{2(2/N)})\prod_{i\leq j} dx_{ij} &= \frac{1}{Z}\prod_{i,j}\exp(-N x_{ij}x_{ji}/4)\prod_{i,j} dx_{ij} \\ &= \frac{1}{Z}\exp\left(-\frac{N}{4}\operatorname{Tr}[X^2]\right)\,dX.\end{aligned}$$

Note that if U is an orthogonal matrix,

$$\frac{1}{Z}\exp\left(-\frac{N}{4}\operatorname{Tr}[(UXU^T)(UXU^T)]\right)\,d(UXU^T) = \frac{1}{Z}\exp\left(-\frac{N}{4}\operatorname{Tr}[X^2]\right)\,dX.$$

So this ensemble is orthogonally invariant. Since X is symmetric,

$$X = U\Lambda U^T,$$

where U is a random orthogonal matrix, and Λ is a random diagonal matrix. From orthogonal invariance it follows (after some work) that U is a *Haar orthogonal matrix*, with a uniform distribution over the orthogonal group, and so eigenvectors of X are uniformly distributed on a sphere. What about eigenvalues?

Third point of view: eigenvalues and spectral measure.

X_N is symmetric, so diagonalizable, with (random!) eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$. Combine these into an *empirical spectral measure*

$$\hat{\mu}_X = \frac{1}{N}(\delta_{\lambda_1} + \ldots + \delta_{\lambda_N})$$

which is a random measure.

Note:

$$\frac{1}{N}\operatorname{Tr}[X^k] = \frac{1}{N}\operatorname{Tr}[(U\Lambda U^T)^k] = \frac{1}{N}\operatorname{Tr}[\Lambda^k] = \frac{1}{N}\sum_{i=1}^N \lambda_i^k = \int x^k\,d\hat{\mu}(x),$$

the *k'th moment of $\hat{\mu}$*.

Theorem (Wigner's Theorem II). *Let $X_N \sim$ GOE. Then as $N \to \infty$,*

$$\hat{\mu}_N \to \sigma$$

weakly, meaning that for any $f \in C_b(\mathbb{R})$,

$$\int f \, d\hat{\mu}_N \to f \, d\sigma,$$

and σ is the (Wigner) semicircle law,

$$d\sigma(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{[-2,2]} \, dx.$$

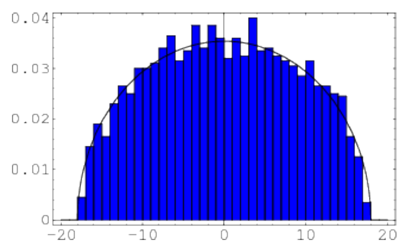

Figure 1.1: Semicircle law

Again, $\hat{\mu}_N$ random*, so:* $\hat{\mu}_N \to \sigma$ weakly in expectation *if* $\forall f \in C_b(\mathbb{R})$,

$$\mathbb{E}\left[\int f \, d\hat{\mu}_N\right] \to f \, d\sigma;$$

weakly in probability *if*

$$P\left(\left|\int f \, d\hat{\mu}_N - f \, d\sigma\right| \geq \delta\right) \to 0;$$

weakly a.s. *if*

$$P\left(\int f \, d\hat{\mu}_N \not\to f \, d\sigma\right) = 0.$$

Remark 1.4. It is not hard to check that

$$\int x^{2k+1} \, d\sigma(x) = 0, \qquad \int x^{2k} \, d\sigma(x) = c_k.$$

So Wigner I says

$$\int x^k \, d\hat{\mu}_N \to \int x^k \, d\sigma(x).$$

Of course this function is not in $C_b(\mathbb{R})$.

Results so far are about (weighted) averages of eigenvalues. What about individual eigenvalues? Recall that we defined $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$, and showed

$$\frac{1}{N}(\delta_{\lambda_1} + \ldots + \delta_{\lambda_N}) \to \sigma.$$

with support $[-2, 2]$, figure omitted.

Theorem. *Let $X \sim GOE_N$. Then $\lambda_N(X_N) \to 2$ in probability.*

(Füredi, Komlos 1981, Bai, Yin 1988)

Fluctuations.

Recall

$$\frac{1}{N}\,\mathrm{Tr}[X_N^{2k}] \to c_k \quad \text{in probability,}$$
$$\lambda_N(X_N) \to 2 \quad \text{in probability.}$$

These are analogs of the laws of large numbers. What about the analogs of the Central Limit Theorem?

Theorem.

$$N\left(\frac{1}{N}\,\mathrm{Tr}[X_N^{2k}] - c_k\right) \to \mathcal{N}(0, ?) \quad \textit{in distribution.}$$

In contrast,

$$N^{2/3}(\lambda_N(X_N) - 2) \to \textit{ Tracy-Widom distribution.}$$

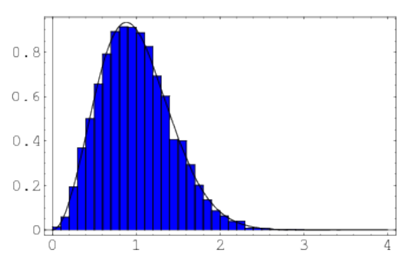

Figure 1.2: The Tracy-Widom distribution

Remark 1.5 (Large deviations)**.** Recall that for large N, $\hat{\mu}_N \approx \sigma$ with large probability. What are the chances that it is far from σ? **Very roughly**, for a probability measure ν,

$$\mathrm{Prob}(\hat{\mu}_N \approx \nu) \sim e^{-N^2 I(\nu)},$$

where

$$I(\nu) = \frac{1}{4}\int x^2\, d\nu(x) - \frac{1}{2}\iint \log|x-y|\; d\nu(x)\, d\nu(y)$$

$=$ logarithmic energy $=$ free entropy. Here $I(\sigma)$ minimizes I.

Spacing distributions.

Recall

$$\hat{\mu}_N = \frac{1}{N}(\delta_{\lambda_1} + \ldots + \delta_{\lambda_N}) \to \rho(x)\, dx,$$

$$\rho(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{[-2,2]}(x).$$

So intuitively,

$$\int_{-2}^{\lambda_j} \rho(x)\, dx \approx \frac{j}{N}$$

(in fact true) and

$$\frac{1}{N} \approx \int_{\lambda_j}^{\lambda_{j+1}} \rho(x)\, dx \approx (\lambda_{j+1} - \lambda_j)\rho(\lambda_j).$$

Thus

$$\lambda_{j+1} - \lambda_j \approx \frac{1}{N\rho(\lambda_j)}.$$

So renormalize

$$s_j = N\rho(\lambda_j)(\lambda_{j+1} - \lambda_j).$$

For $0 \ll j \ll N$, independently of j, $s \sim$ Gaudin distribution. Figure omitted.

This is of interest because of Wigner's original model: X models the Hamiltonian of a large atom, in which case λ_j's are the energy levels. Physically what is observed are not λ_j's but $(\lambda_i - \lambda_j)$'s. Figure omitted.

These properties can also be stated in terms of k-point correlation functions.

1.2 Other ensembles.

Other Gaussian ensembles.

Gaussian unitary ensemble GUE$_N$.

For $1 \le i, j \le N$, let $B_{ij} \sim \mathcal{N}(0,1)$ be independent. Define X_N by

$$[X_N]_{ij} = \frac{1}{\sqrt{2N}}(B_{ij} + \sqrt{-1}B_{ji}), \quad i < j,$$

$$[X_N]_{ji} = \overline{[X_N]_{ij}} = \frac{1}{\sqrt{2N}}(B_{ij} - \sqrt{-1}B_{ji}), \quad i < j,$$

$$[X_N]_{ii} = \frac{1}{\sqrt{N}}B_{ii}.$$

Thus X_N is complex Hermitian, its off-diagonal entries are complex Gaussian, and its diagonal entries are real Gaussian. Its distribution is invariant under conjugation by unitary matrices.

Gaussian symplectic ensemble GSE$_N$.

Recall that the algebra $\mathbb{H}$ of quaternions is

$$\{a + bi_1 + ci_2 + di_3 : a, b, c, d \in \mathbb{R}\}$$

subject to the relations $i_1^2 = i_2^2 = i_3^2 = i_1 i_2 i_3 = -1$. For $q \in \mathbb{H}$, we may define the quaternion conjugate $\overline{q}$ as in the complex case. The dual Q^* of a quaternion matrix is its conjugate transpose. A matrix $Q = Q^*$ is self-dual. Finally, the symplectic group consists of quaternion matrices such that $S^*S = SS^* = I$. The Gaussian symplectic ensemble consists of self-dual quaternionic matrices whose entries are properly normalized independent quaternionic Gaussians. Its distribution is invariant under conjugation by symplectic matrices.

Most results which hold for GOE hold, either exactly or with appropriate modification, for GUE and GSE (in fact the results for GUE are often neater). Moreover we can include all these in the family of Gaussian β-ensembles, with $\beta = 1$ real/orthogonal, $\beta = 2$ complex/unitary, and $\beta = 4$ quaternionic/symplectic.

GOE satisfies two properties:

a. Symmetric with independent entries.

b. Orthogonally invariant.

These two properties in fact characterize GOE. So have two natural directions to generalize.

Wigner ensembles.

$[X_N]_{ij}$ symmetric, independent,

$$X_{ij} = X_{ji} \sim \frac{1}{\sqrt{N}} Y_1, \quad Y_1 \sim \nu_1,$$

$$X_{ii} \sim \frac{1}{\sqrt{N}} Y_2, \quad Y_2 \sim \nu_2,$$

$\mathrm{Var}\, \nu_1 < \infty$, $\mathrm{Var}\, \nu_2 < \infty$, and possibly with assumptions on higher moments.

Then $\hat{\mu}_N \to \sigma$ still! *Under extra assumptions*, also $\lambda_N(X_N) \to 2$ and $N^{2/3}(\lambda_N - 2) \to$ TW.

Orthogonally invariant ensembles.

Recall for GOE,

$$X_N \sim \frac{1}{Z} \exp\left(-\frac{N}{4} \mathrm{Tr}[X^2]\right) dX.$$

More generally, may look at

$$X \sim \frac{1}{Z} \exp\left(-N \,\mathrm{Tr}[V(X)]\right) dX$$

for V a nice function. Then $\hat{\mu}_N \to \mu_V$, the equilibrium measure for the potential V (different from σ), which can be described using $I_V(\nu)$. However the spacing distributions, for nice V, do not depend on V, and so are universal, as is the convergence to the Tracy-Widom distribution.

Non-symmetric ensembles.

$$X_{ij} \sim \frac{1}{\sqrt{N}} Y, \quad Y \sim \nu$$

all independent.

X diagonalizable a.s.

$$\hat{\mu}_N = \frac{1}{N}(\delta_{\lambda_1} + \ldots + \delta_{\lambda_N})$$

is a measure on $\mathbb{C}$. For nice ν, $\hat{\mu}_N$ converges to the *circular law* (Girko 1984, Tao, Vu 2008), figure omitted.

Can ask similar questions in this context.

Wishart ensembles.

The oldest appearance of asymptotic theory of random matrices (Wishart 1928).

Consider a k-component Gaussian vector $Y \sim \mathcal{N}(0, \Sigma)$ with the covariance matrix Σ. How to estimate Σ?

Take N independent samples $Y^{(1)}, Y^{(2)}, \ldots Y^{(N)}$. The sample covariance estimate is

$$\mathbb{E}[Y_i Y_j] \approx \frac{1}{N} \sum_{n=1}^{N} Y_i^{(n)} Y_j^{(n)}.$$

Let $\hat{Y} = (Y_1 | Y_2 | \ldots | Y_N)$, a $k \times N$ matrix. Then

$$\mathbb{E}[Y_i Y_j] \approx \frac{1}{N} \sum_{n=1}^{N} \hat{Y}_{in} \hat{Y}_{jn} = \frac{1}{N} (\hat{Y} \hat{Y}^T)_{ij}.$$

$X = \frac{1}{N} \hat{Y}^T \hat{Y}$ (note order) is the $N \times N$ Wishart(k, N, Σ) matrix. If k is fixed, as $N \to \infty$, $\frac{1}{N} \hat{Y} \hat{Y}^T \to \Sigma$. What if both k and N are large? Note that X is orthogonally invariant (check), so only its eigenvalues matter, and they are closely related to the eigenvalues of $\frac{1}{N} \hat{Y} \hat{Y}^T$. For example if $\Sigma = I$, and $\frac{k}{N} \to p$, then $\hat{\mu}_{X_N}$ converges to the *Marchenko-Pastur distribution*.

Connections to other fields.

Wigner, Tracy-Widom, Gaudin, Marchenko-Pastur distributions appear in unexpected contexts with no a priori connection to random matrices. We only give two examples.

Example 1.6 (Ulam problem)**.** Let $\alpha \in S(N)$ be a permutation. Reorder $\{1, 2, \ldots, N\}$ according to α, and let $L(\alpha)$ be the length of the *longest increasing subsequence* in it. Pick α uniformly at random. What can be said about L?

Theorem. *(Vershik, Kerov 1977, Baik, Deift, Johansson 1999) As $N \to \infty$, the average length of the longest increasing subsequence*

$$\mathbb{E}[L] \approx 2\sqrt{N},$$

and

$$\frac{L - 2\sqrt{N}}{N^{1/6}} \to \textit{Tracy-Widom distribution}.$$

Example 1.7 (Riemann zeta function)**.** $\zeta(z) =$ analytic continuation of $\sum_{n=1}^{\infty} \frac{1}{n^z}$. The Riemann Hypothesis states that all zeros of ζ lie on the critical line $z = \frac{1}{2} + iy$. Denote the imaginary parts of the zeros by $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$. The prime number theorem implies that

$$\lambda_n \sim \frac{2\pi n}{\log n}$$

and so

$$\lambda_{n+1} - \lambda_n \sim \frac{2\pi}{\log n}.$$

So renormalize: **roughly**,

$$v_n = \frac{\log n}{2\pi}(\lambda_{n+1} - \lambda_n)$$

Then for large n, v appears to follow the (GUE version of the) Gaudin distribution. Extensive numerical and some theoretical evidence (Montgomery 1973, Odlyzko 1987). No proof!

Chapter 2

Wigner's theorem by the method of moments.

The techniques in this chapter go all the way back to Wigner (1955), but continue to be used with great success.

2.1 Convergence of moments.

Theorem 2.1. *Let X be a Wigner matrix with finite moments. That is, for each N, $\{X_{ij} : 1 \leq i \leq j \leq N\}$ are independent,*

$$X_{ij} = X_{ji} \sim \frac{1}{\sqrt{N}} Y_{ij}, \quad Y_{ij} \sim \nu_1,$$

$$X_{ii} \sim \frac{1}{\sqrt{N}} Y_{ii}, \quad Y_{ii} \sim \nu_2,$$

$\mathbb{E}[Y_{ij}] = 0$, $\operatorname{Var} \nu_1 = 1$*, and all the higher moments of ν_1 and ν_2 are finite. Then for $k \geq 1$,*

$$\frac{1}{N} \operatorname{Tr}[X_N^{2k}] \to c_k$$

and

$$\frac{1}{N} \operatorname{Tr}[X_N^{2k-1}] \to 0$$

in expectation, in probability, and (as long as all the random variables live on the same probability space) almost surely.

Remark 2.2. The condition $\operatorname{Var} \nu_1 = 1$ is there purely to simplify the normalization. The condition that entries are identically distributed can easily be removed as long as the moments of the entries are uniformly bounded. The condition of equal variances is absolutely essential. The independence condition can be weakened, but the proof becomes significantly more complicated.

Proof of Theorem 2.1 for convergence in expectation.

$$\frac{1}{N}\operatorname{Tr}[X_N^k] = \frac{1}{N^{1+k/2}} \sum_{u(1),\dots,u(k)=1}^{N} Y_{u(1)u(2)} Y_{u(2)u(3)} \dots Y_{u(k)u(1)}.$$

Fix $\mathbf{u} = (u(1), u(2), \dots, u(k))$. Let $S_{\mathbf{u}}$ be the set

$$S_{\mathbf{u}} = \{u(1), u(2), \dots, u(k)\}.$$

Consider the multigraph with the vertex set S, and the number of undirected edges between $u(i)$ and $u(j)$ equal to the multiplicity of the factor $Y_{u(i)u(j)} = Y_{u(j)u(i)}$ in the product above; multiplicity zero means no edge. Note that this multigraph comes equipped with an Eulerian circuit: the path

$$u(1), u(2), u(3), \dots u(k), u(1)$$

passes through each edge of the graph exactly as many times as its multiplicity. Finally, if we forget the multiplicities, we end up with the underlying (simple) graph. The Eulerian condition implies in particular that this graph is connected.

We decompose the sum above according to

$$\frac{1}{N}\mathbb{E}[\operatorname{Tr}[X_N^k]] = \frac{1}{N^{1+k/2}} \sum_{s=1}^{k} \sum_{\mathbf{u}:|S_{\mathbf{u}}|=s} \mathbb{E}[Y_{u(1)u(2)} Y_{u(2)u(3)} \dots Y_{u(k)u(1)}].$$

Note that each expectation on the right-hand side is independent of N.

First suppose that $s < 1 + k/2$. Then

$$|\{\mathbf{u} : |S_{\mathbf{u}}| = s\}| \le \binom{N}{s} s^k \le s^k N^s.$$

Therefore

$$\frac{1}{N^{1+k/2}} \sum_{s=1}^{k/2} \sum_{\mathbf{u}:|S_{\mathbf{u}}|=s} \mathbb{E}[Y_{u(1)u(2)} Y_{u(2)u(3)} \dots Y_{u(k)u(1)}] \to 0$$

as $N \to \infty$.

Next, note that if some edge in the graph appears with multiplicity 1, then since entries of the matrix are independent and centered, the corresponding expectation is zero. But if a connected multigraph has k edges and each edge has multiplicity at least 2, it can have at most $1 + k/2$ vertices. Therefore

$$\frac{1}{N^{1+k/2}} \sum_{s=2+k/2}^{N} \sum_{\mathbf{u}:|S_{\mathbf{u}}|=s} \mathbb{E}[Y_{u(1)u(2)} Y_{u(2)u(3)} \dots Y_{u(k)u(1)}] = 0$$

and

$$\frac{1}{N}\,\mathbb{E}[\mathrm{Tr}[X_N^k]] = \frac{1}{N^{1+k/2}} \sum_{\mathbf{u}:|S_{\mathbf{u}}|=1+k/2} \mathbb{E}[Y_{u(1)u(2)}Y_{u(2)u(3)}\dots Y_{u(k)u(1)}].$$

In particular this is zero for k odd; from now on we assume k to be even. In that case the argument above shows that the non-zero terms in the sum correspond to graphs with $1 + k/2$ vertices and $k/2$ edges, each of multiplicity 2. This means that each underlying simple graph is a tree, and the sum is taken over precisely all labeled ordered rooted trees with $1 + k/2$ vertices, with a root (corresponding to $u(1)$), an order of leaves at each vertex (corresponding to the order in which they are traversed by the Eulerian circuit), and $1 + k/2$ distinct numbers between 1 and N (labels of the vertices). The number of such ordered rooted trees is the Catalan number $c_{k/2}$ (see the lemma below). Note also that a tree cannot have self-edges, so no terms of the form Y_{ii} appear. Thus using independence of entries

$$\frac{1}{N}\,\mathbb{E}[\mathrm{Tr}[X_N^k]] = \frac{N(N-1)\dots(N-k/2)}{N^{1+k/2}}\,\mathrm{Var}^k[\nu_1]c_{k/2} \to c_{k/2}$$

as $N \to \infty$. □

Lemma 2.3. *The number of ordered trees rooted with $k+1$ vertices is the Catalan number c_k.*

Proof. Note that a tree with a fixed Eulerian circuit and root can be identified with an ordered tree, since drawing the tree with the circuit on the outside corresponds to a unique way to define a depth-first order on it. Let t_k be the number of such trees. By removing the edge $(u(1), u(2))$, we see that these numbers satisfy the Catalan recursion

$$t_k = \sum_{i=0}^{k-1} t_i t_{k-i-1},$$

with $t_0 = 1$, $t_1 = 1$. So $t_k = c_k$. □

Exercise 2.4. Prove that the Catalan numbers satisfy the Catalan recursion. Here is one possible approach. Suppose $b_0 = b_1 = 1$ and the b_k's satisfy the Catalan recursion. Let $F(z) = \sum_{k=0}^{\infty} b_k z^k$ be their generating function. Show that F satisfies a quadratic equation. Solve this equation to find a formula for F. Finally, use the generalized binomial theorem to expand F into a power series, to see that its coefficients are the Catalan numbers.

To upgrade convergence in expectation to convergence in probability, we recall

Lemma 2.5 (Markov inequality)**.** *Let U be a positive random variable with a finite expectation. Then for any $\delta > 0$*

$$P(U \geq \delta) \leq \frac{1}{\delta}\,\mathbb{E}[U].$$

Lemma 2.6 (Chebyshev inequality)**.** *Let V be a random variable with finite variance. Then for any $\delta > 0$*

$$P(|V - \mathbb{E}[V]| \geq \delta) \leq \frac{1}{\delta^2}\,\mathrm{Var}[V].$$

Proof of Theorem 2.1 for convergence in probability. Since

$$P\left(\left|\frac{1}{N}\operatorname{Tr}[X^k]-m_k(\sigma)\right|\geq\delta\right)\leq P\left(\left|\frac{1}{N}\operatorname{Tr}[X^k]-\frac{1}{N}\mathbb{E}[\operatorname{Tr}[X^k]]\right|\geq\delta-\left|\frac{1}{N}\mathbb{E}[\operatorname{Tr}[X^k]]-m_k(\sigma)\right|\right),$$

we have just shown that $\left|\frac{1}{N}\mathbb{E}[\operatorname{Tr}[X^k]]-m_k(\sigma)\right|\to 0$ as $N\to\infty$, and using Chebyshev's inequality, it suffices to show that

$$\operatorname{Var}\left[\frac{1}{N}\operatorname{Tr}[X^k]\right]\to 0.$$

This is

$$\frac{1}{N^2}\mathbb{E}\left[\operatorname{Tr}[X^k]\operatorname{Tr}[X^k]\right]-\frac{1}{N^2}\mathbb{E}[\operatorname{Tr}[X^k]]\,\mathbb{E}[\operatorname{Tr}[X^k]].$$

We refine our analysis in the previous proof, noting for future reference the speed of decay of various terms. Denoting

$$Y_{\mathbf{u}}=Y_{u(1)u(2)}Y_{u(2)u(3)}\dots Y_{u(k)u(1)},$$

$$\operatorname{Var}\left[\frac{1}{N}\operatorname{Tr}[X^k]\right]=\frac{1}{N^{2+k}}\sum_{s,t=1}^{k}\sum_{\mathbf{u},\mathbf{v}:|S_{\mathbf{u}}|=s,|S_{\mathbf{v}}|=t}\left(\mathbb{E}[Y_{\mathbf{u}}Y_{\mathbf{v}}]-\mathbb{E}[Y_{\mathbf{u}}]\,\mathbb{E}[Y_{\mathbf{v}}]\right).$$

Again we have a multigraph with the vertex set $S_{\mathbf{u}}\cup S_{\mathbf{v}}$, this time covered by a pair of paths which together traverse each edge according to its multiplicity. So it has at most two connected components with vertex sets $S_{\mathbf{u}}$ and $S_{\mathbf{v}}$ if these are disjoint, or one component if these intersect. By the same arguments as above, we can conclude that

the terms with $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|\leq k$ go to zero at least as fast as $\frac{1}{N^2}$ with $N\to\infty$, while the terms with $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|=1+k$ go to zero as $\frac{1}{N}$.

Thus assume $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|\geq 1+k$.

$\mathbb{E}[Y_{\mathbf{u}}]\,\mathbb{E}[Y_{\mathbf{v}}]=0$ unless both $|S_{\mathbf{u}}|,|S_{\mathbf{v}}|\leq 1+k/2$ and the subgraphs restricted to $S_{\mathbf{u}},S_{\mathbf{v}}$ are trees with double edges.

$\mathbb{E}[Y_{\mathbf{u}}Y_{\mathbf{v}}]=0$ unless

- $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|=2+k$ and the graph has two components, each of which is a tree with double edges;
- or $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|=1+k$, $|S_{\mathbf{u}}\cap S_{\mathbf{v}}|=1$, and the graph is a tree with double edges;
- or $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|=1+k$, $|S_{\mathbf{u}}\cap S_{\mathbf{v}}|=0$, and the graph has two components, one a tree with double edges, the other with double edges and a single cycle;
- or $|S_{\mathbf{u}}\cup S_{\mathbf{v}}|=1+k$, $|S_{\mathbf{u}}\cap S_{\mathbf{v}}|=0$, and the graph has two components, each of which is a tree with double edges, and the total of two triple edges (note that these two edges have to lie in the same sub-graph);

- or $|S_{\mathbf{u}} \cup S_{\mathbf{v}}| = 1 + k$, $|S_{\mathbf{u}} \cap S_{\mathbf{v}}| = 0$, and the graph has two components, each of which is a tree with double edges, and a single quadruple edge.

In the case when $|S_{\mathbf{u}} \cup S_{\mathbf{v}}| = 2 + k$ and the graph has two components, each of which is a tree with double edges, it follows that $S_{\mathbf{u}}$ and $S_{\mathbf{v}}$ are disjoint of size $1 + k/2$, and

$$\mathbb{E}[Y_{\mathbf{u}} Y_{\mathbf{v}}] - \mathbb{E}[Y_{\mathbf{u}}]\ \mathbb{E}[Y_{\mathbf{v}}] = 0.$$

Finally, suppose $|S_{\mathbf{u}} \cup S_{\mathbf{v}}| = 1 + k$, and the graph is a tree with double edges. We have two non-empty paths whose union traverses each edge exactly twice. Since the graph is a tree, each edge must be traversed by each path either zero times or twice. So the paths are actually edge-disjoint, although they may contain common vertices. Then independence again implies that

$$\mathbb{E}[Y_{\mathbf{u}} Y_{\mathbf{v}}] - \mathbb{E}[Y_{\mathbf{u}}]\ \mathbb{E}[Y_{\mathbf{v}}] = 0.$$

The same conclusion follows in the other sub-cases. We conclude that

$$\mathrm{Var}\left[\frac{1}{N}\,\mathrm{Tr}[X^k]\right] = \frac{1}{N^{2+k}} \sum_{s,t=1}^{k} \sum_{\mathbf{u},\mathbf{v}:|S_{\mathbf{u}}|=s,|S_{\mathbf{v}}|=t} \left(\mathbb{E}[Y_{\mathbf{u}} Y_{\mathbf{v}}] - \mathbb{E}[Y_{\mathbf{u}}]\ \mathbb{E}[Y_{\mathbf{v}}]\right) \to 0$$

at least as fast as $\frac{1}{N^2}$. □

To upgrade convergence in probability to almost sure convergence, we recall

Lemma 2.7 (The Borel-Cantelli Lemma). *Let $\{E_N\}_{N=1}^{\infty}$ be events (measurable subsets) such that*

$$\sum_{N=1}^{\infty} P(E_N) < \infty.$$

Then $P(\omega : \omega$ lies in infinitely many $E_N) = 0$.

Corollary 2.8. *Let $\{x_N\}_{N=1}^{\infty}$ be a sequence of random variables. If*

$$\sum_{N=1}^{\infty} P(|x_N - a| \geq \delta) < \infty$$

for all $\delta > 0$, then $x_N \to a$ a.s. In particular, this conclusion follows from the stronger assumption that

$$\sum_{N=1}^{\infty} \mathrm{Var}[x_N] < \infty.$$

Proof of Theorem 2.1 for almost sure convergence. We note that since the variances decay at least as fast as $\frac{1}{N^2}$,

$$\sum_{N=1}^{\infty} \mathrm{Var}\left[\frac{1}{N}\,\mathrm{Tr}[X_N^k]\right] < \infty. \qquad \square$$

Remark 2.9. One can use more complicated versions of the moment method to prove the Gaussian fluctuations for moments, and convergence of the largest eigenvalue to 2.

Exercise 2.10. A complex Wigner matrix X_N has the form $X_N = \frac{1}{\sqrt{N}} Y_N$. Here Y_N is a complex Hermitian random matrix, such that the random variables $\{Y_{ij} : 1 \leq i \leq j \leq N\}$ are independent,

$$Y_{ij} = \overline{Y}_{ji}, \quad i < j$$

are identically distributed complex random variables with mean zero and variance

$$\mathbb{E}\left[|Y_{ij}|^2\right] = \mathbb{E}\left[(\Re Y_{ij})^2 + (\Im Y_{ij})^2\right] = 1,$$

and Y_{ii} are identically distributed real random variables with mean zero and finite variance. Prove the analog of Theorem 2.1 for these matrices. You do not need to repeat all the arguments from this section, just indicate where and how they need to be modified.

What about the quaternionic Wigner matrices, defined similarly? How does non-commutativity of entries affect the argument? Note that for independent but non-commuting variables x, y, $\mathbb{E}[xyx] = \mathbb{E}[x^2]\,\mathbb{E}[y]$, but in general $\mathbb{E}[xyxy] \neq \mathbb{E}[x^2]\,\mathbb{E}[y^2]$.

Sketch of a solution. The argument, at least for convergence in expectation, basically goes through until the last step, when the moments are reduced to a sum of terms over trees with Eulerian circuits traversing each edge exactly twice. For a general graph, it is possible for an edge to be traversed twice in the same direction (which causes problems in the complex case), and the Y terms corresponding to the same edge may not be adjacent (which causes problems in the quaternionic case). However for a tree, each edge traversed twice has to be traversed in opposite directions, and the Y terms corresponding to the same edge may always be taken to be adjacent by "pruning the leaves." Instead of a proof, we illustrate these statements with an example. Consider the term

$$\mathbb{E}[Y_{u(1)u(2)}Y_{u(2)u(3)}Y_{u(3)u(2)}Y_{u(2)u(4)}Y_{u(4)u(5)}Y_{u(5)u(4)}Y_{u(4)u(6)}Y_{u(6)u(4)}Y_{u(4)u(2)}Y_{u(2)u(1)}]$$

corresponding to the path $1, 2, 3, 2, 4, 5, 4, 6, 4, 2, 1$ (draw the corresponding tree!). Then using the very weak form of independence called *singleton independence*, the term above is equal to

$$\begin{aligned}
&\mathbb{E}[Y_{u(1)u(2)}Y_{u(2)u(4)}Y_{u(4)u(2)}Y_{u(2)u(1)}]\,\mathbb{E}[Y_{u(2)u(3)}Y_{u(3)u(2)}]\,\mathbb{E}[Y_{u(4)u(5)}Y_{u(5)u(4)}]\,\mathbb{E}[Y_{u(4)u(6)}Y_{u(6)u(4)}] \\
&\quad = \mathbb{E}[Y_{u(1)u(2)}Y_{u(2)u(1)}]\,\mathbb{E}[Y_{u(2)u(3)}Y_{u(3)u(2)}]\,\mathbb{E}[Y_{u(2)u(4)}Y_{u(4)u(2)}]\,\mathbb{E}[Y_{u(4)u(5)}Y_{u(5)u(4)}]\,\mathbb{E}[Y_{u(4)u(6)}Y_{u(6)u(4)}]
\end{aligned}$$

This argument in fact shows that, as long as they satisfy appropriate joint moment bounds and singleton independence, the entries of the matrix Y can be taken from any non-commutative (operator) algebra, and the corresponding moments of X will still converge in expectation to the Catalan numbers. □

Exercise 2.11. Let Y_N be an $N \times N$ matrix with independent identically distributed entries, with mean zero, variance 1, and finite moments. Let $X_N = \frac{1}{\sqrt{N}} Y_N$ and $Z_N = X_N X_N^T$. Then Z_N is a (generalized) Wishart matrix. For each k, show that

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}[Z_N^k]\right] \to c_k$$

as $N \to \infty$. Thus the moments of the asymptotic empirical spectral distribution of Z_N are equal to the *even* moments of the semicircular distribution. Use this to conclude that this asymptotic distribution is the quarter-circle law

$$d\mu(x) = \frac{\sqrt{x(4-x)}}{2\pi x} \mathbf{1}_{[0,4]}\, dx.$$

More generally, in the construction above we may start with Y_N an $K \times N$ matrix, and assume that both K and N go to infinity in such a way that $K/N \to \alpha \in (0,1]$. For each k, show that $\mathbb{E}[Z_N^k]$ converges as $N \to \infty$, and express the answer in terms of the number of certain combinatorial objects. Hint: the answer involves directed bi-partite graphs. In fact the "combinatorial objects" can be enumerated, showing that

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}[Z_N^k]\right] \to \sum_{j=0}^{k-1} \frac{\alpha^{j+1}}{j+1}\binom{k}{j}\binom{k-1}{j}.$$

Here the coefficients of α^j are called the Narayana numbers. The distribution with these moments is the Marchenko-Pastur distribution with parameter α,

$$d\mu(x) = \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{2\pi\alpha x} \mathbf{1}_{[\lambda_-,\lambda_+]}\, dx,$$

where $\lambda_\pm = (1 \pm \sqrt{\alpha})^2$.

2.2 Generalities about weak convergence.

Let $C_0(\mathbb{R})$ be the space of continuous functions going to zero at infinity, with the uniform norm. Riesz Representation Theorem states that the dual Banach space $C_0(\mathbb{R})'$ is isometrically isomorphic to the Banach space of finite (Radon, complex) measures, with the total variation norm. By definition, a sequence of finite measures $\nu_N \to \nu$ in the weak* topology if for all $f \in C_0(\mathbb{R})$,

$$\int f\, d\nu_N \to \int f\, d\nu.$$

For this particular Banach space, this topology is also called the *vague topology*. According to the Banach-Alaoglu theorem, the unit ball of the dual space is compact in the weak* topology. Since it is also metrizable in this topology, this unit ball is also sequentially compact. Putting all these results together, we get

Proposition 2.12. *Any sequence of probability measures has a subsequence which converges vaguely to a finite measure.*

The limit need not be a probability measure. However the weak limit of a sequence of probability measures is again a probability measure. To upgrade vague to weak convergence, we need the following notion.

Definition 2.13. A family of measures $\{\nu_N\}_{N=1}^{\infty}$ is *tight* if $\forall \varepsilon > 0\ \exists C\ \forall N$

$$\nu_N(|x| \geq C) \leq \varepsilon.$$

Note that the set $\left\{\int x^2\, d\nu_N\right\}_{N=1}^{\infty}$ being bounded is a sufficient condition for tightness.

Exercise 2.14. Let $\{\nu_N\}_{N=1}^{\infty}$ be a sequence of probability measures. The following are equivalent.

a. The sequence is tight and converges vaguely.

b. The sequence converges vaguely to a probability measure.

c. The sequence converges weakly.

Corollary 2.15. *Any tight sequence of probability measures has a subsequence converging weakly to a probability measure.*

Lemma 2.16. *In a metric space, a sequence $\{x_N\}_{N=1}^{\infty}$ converges to a if and only any of its subsequences has a further subsequence converging to a.*

Lemma 2.17. *Suppose g, h are continuous functions such that $g \geq 0$ and $\lim_{x\to\infty} |h(x)|/g(x) = 0$. Suppose $\nu_n \to \nu$ weakly and $C = \sup\left\{\int g(x)\, d\nu_N(x)\right\}_{N=1}^{\infty} < \infty$. Then*

$$\int h\, d\nu_N \to \int h\, d\nu.$$

Proof. Fix $\varepsilon > 0$, and choose $I = [-K, K]$ so that $|h(x)|/g(x) < \varepsilon$ on I^c. Let $J = [-K-1, K+1]$, and let φ be a continuous function such that $0 \leq \varphi \leq 1$, $\varphi \equiv 1$ on I, and $\varphi \equiv 0$ on J^c. Then

$$\int h\varphi\, d\nu_N \to \int h\varphi\, d\nu$$

while

$$\int |h|\,(1-\varphi)\, d\nu_N = \int \frac{|h|}{g} g(1-\varphi)\, d\nu_N \leq \varepsilon \int g\, d\nu_N \leq \varepsilon C$$

and by Fatou's lemma also

$$\int |h|\,(1-\varphi)\, d\nu \leq \varepsilon C.$$

The result follows. $\square$

Corollary 2.18. *If for all k, $\int x^k\, d\nu_N \to \int x^k\, d\nu$ and ν is uniquely determined by its moments, then $\nu_N \to \nu$ weakly.*

Proof. Any subsequence of $\{\nu_N\}_{N=1}^{\infty}$ has a further subsequence converging weakly to a probability measure. Call the limit $\tilde{\nu}$. It suffices to show that $\tilde{\nu} = \nu$. Indeed, since the sequence $\left\{\int x^k \, d\nu_N\right\}_{N=1}^{\infty}$ converges, it is bounded; and

$$\int |x|^{k+1} \, d\nu_N \le \sqrt{\int x^2 \, d\nu_N \quad \int x^{2k} \, d\nu_N}.$$

Taking $g(x) = |x|^{k+1}$ and $h(x) = x^k$ in the preceding lemma, we conclude that $\int x^k \, d\nu_N \to \int x^k \, d\tilde{\nu}$ and so $\int x^k \, d\tilde{\nu} = \int x^k \, d\nu$. Since ν is uniquely determined by its moments, the measures are equal. □

Lemma 2.19. *A compactly supported measure is uniquely determined by its moments.*

The idea of the proof is that for a compactly supported μ, its Fourier transform (characteristic function) $\mathcal{F}(\theta) = \int e^{ix\theta} \, d\theta$ is an analytic function with the power series expansion

$$\mathcal{F}(\theta) = \sum_{n=0}^{\infty} \frac{i^n m_n(\mu)}{n!} \theta^n,$$

and that any μ is uniquely determined by its Fourier transform. Alternatively, we could use Stieltjes transforms as in Remark 4.3.

Theorem 2.20. *Let X be a Wigner matrix with finite moments as in Theorem 2.1. Then*

$$\hat{\mu}_{X_N} \to \sigma$$

weakly almost surely.

Proof. Fix ω such that for all k,

$$\int x^k \, d\hat{\mu}_{X_N(\omega)} \to \int x^k \, d\sigma.$$

We know that the set of ω where this is false has measure zero. Since σ is compactly supported, for such ω, $\hat{\mu}_{X_N(\omega)} \to \sigma$ weakly. □

2.3 Removing the moment assumptions.

Theorem 2.21. *Let X be a Wigner matrix. That is, for each N, $\{X_{ij} : 1 \le i \le j \le N\}$ are independent,*

$$X_{ij} = X_{ji} \sim \frac{1}{\sqrt{N}} Y_{ij}, \quad Y_{ij} \sim \nu_1,$$

$$X_{ii} \sim \frac{1}{\sqrt{N}} Y_{ii}, \quad Y_{ii} \sim \nu_2,$$

$\mathbb{E}[Y_{ij}] = 0$, $\operatorname{Var} \nu_1 = 1$, *and* $\operatorname{Var} \nu_2 < \infty$. *Then*

$$\hat{\mu}_{X_N} \to \sigma$$

weakly in probability.

The result is obtained by combining the lemmas below. For the proofs we will follow very closely the presentation in Section 5 of Todd Kemp's notes, and so omit them here.

Lemma 2.22. *In the notation from the preceding theorem, define for a (large) constant $C > 0$*

$$\tilde{Y}_{ij} = \frac{1}{\sigma_{ij}(C)}(Y_{ij}\mathbf{1}_{|Y_{ij}|\leq C} - \mathbb{E}[Y_{ij}\mathbf{1}_{|Y_{ij}|\leq C}]),$$

where $\sigma_{ij}(C)^2 = \mathrm{Var}(Y_{ij}\mathbf{1}_{|Y_{ij}|\leq C})$ for $i \neq j$, and $\sigma_{ii}(C) = 1$ (see Remark 2.29 for an alternative). Then $Y_{ij} - \tilde{Y}_{ij} \to 0$ in L^2 as $C \to \infty$.

Definition 2.23. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz* if

$$\|f\|_{\mathrm{Lip}} = \sup_{x\neq y} \frac{|f(x) - f(y)|}{\|x - y\|} + \sup_x |f(x)| < \infty.$$

The space of Lipschitz functions is denoted by $\mathrm{Lip}(\mathbb{R}^n)$. We only put in the second term to have $\mathrm{Lip}(\mathbb{R}^n) \subset C_b(\mathbb{R}^n)$.

Lemma 2.24. *If $\int f\, d\mu_n - \int f\, d\nu_n \to 0$ for all $f \in \mathrm{Lip}(\mathbb{R})$, then $\int f\, d\mu_n - \int f\, d\nu_n \to 0$ for all $f \in C_b(\mathbb{R})$.*

Lemma 2.25. *Let A and B be $N \times N$ complex Hermitian (or in particular, real symmetric) matrices. Denote by $\lambda_1^A \leq \ldots \leq \lambda_N^A$ and $\lambda_1^B \leq \ldots \leq \lambda_N^B$ their eigenvalues, and by $\hat{\mu}_A$ and $\hat{\mu}_B$ their empirical spectral measures. Then for any $f \in \mathrm{Lip}(\mathbb{R})$,*

$$\left| \int f\, d\hat{\mu}_A - \int f\, d\hat{\mu}_B \right| \leq \|f\|_{\mathrm{Lip}} \left(\frac{1}{N} \sum_{i=1}^N (\lambda_i^A - \lambda_i^B)^2 \right)^{1/2}.$$

Lemma 2.26 (Hoffman-Wielandt inequality)**.** *For A, B as in the preceding lemma,*

$$\sum_{i=1}^N (\lambda_i^A - \lambda_i^B)^2 \leq \mathrm{Tr}[(A - B)^2].$$

We will follow Todd Kemp's notes for the proof, but also outline the proof of the Birkhoff-von Neumann theorem.

Theorem 2.27 (Birkhoff-von Neumann)**.** *Let $\mathcal{D}$ be the space of $N \times N$ doubly stochastic matrices. The extreme points of $\mathcal{D}$ are the permutation matrices.*

Proof. It is easy to check that $\mathcal{D}$ is convex. Let A *not* be a permutation matrix. We will show that A is not an extreme point, that is, it is a convex combination of two matrices in $\mathcal{D}$.

Since A is not a permutation matrix, it has an entry $A_{u(1)u(2)}$ with $0 < A_{u(1)v(1)} < 1$. Since columns add up to 1, there is another entry $A_{u(2)v(1)}$ in the same column with the same property. Since rows add up to

1, there is another entry $A_{u(2)v(2)}$ in the same row with the same property. Continue in this fashion until we arrive in a row or column previously encountered. By possibly removing the beginning of this path, we arrive at the family of entries

$$S = \left\{A_{u(1)v(1)}, A_{u(2)v(1)}, A_{u(2)v(2)}, \dots, A_{u(k)v(k)}, A_{u(1)v(k)}\right\}$$

all of which are strictly between 0 and 1. Note that there is necessarily an even number of them. Let $\varepsilon = \min\{a, 1-a : a \in S\}$. Let B be the matrix whose entries are ε for even numbered elements of S, $-\varepsilon$ for odd numbered elements of S, and 0 otherwise. Then $A+B$ and $A-B$ are both doubly stochastic, and $A = \frac{1}{2}(A+B) + \frac{1}{2}(A-B)$. □

Exercise 2.28. Let $x_1 \le x_2 \le \dots \le x_N$ and $y_1 \le y_2 \le \dots y_N$. Then for any permutation α,

$$\sum x_i y_{\alpha(i)} \le \sum x_i y_i.$$

Proof of Theorem 2.21. Fix $f \in \mathrm{Lip}(\mathbb{R})$ and $\varepsilon, \delta > 0$. By Lemma 2.24, it suffices to show that

$$P\left(\left|\int f\, d\hat{\mu}_{X_N} - \int f\, d\sigma\right| \ge \delta\right) \le \varepsilon$$

for sufficiently large N. For $\tilde{Y}$ as in Lemma 2.22, denote $\tilde{X} = \frac{1}{\sqrt{N}}\tilde{Y}$. Then the entries of $\tilde{X}$ satisfy the assumptions of Theorem 2.20, and so

$$P\left(\left|\int f\, d\hat{\mu}_{\tilde{X}_N} - \int f\, d\sigma\right| \ge \delta/2\right) \le \varepsilon/2 \tag{2.1}$$

for sufficiently large N. On the other hand, combining Lemma 2.25 with the Hoffman-Wielandt inequality,

$$\left|\int f\, d\hat{\mu}_{X_N} - \int f\, d\hat{\mu}_{\tilde{X}_N}\right| \le \|f\|_{\mathrm{Lip}} \left(\mathrm{Tr}\left[\frac{1}{N}(X_N - \tilde{X}_N)^2\right]\right)^{1/2}.$$

Therefore

$$\begin{aligned}
P\left(\left|\int f\, d\hat{\mu}_{X_N} - \int f\, d\hat{\mu}_{\tilde{X}_N}\right| \ge \delta/2\right) &\le P\left(\|f\|_{\mathrm{Lip}}\left(\mathrm{Tr}\left[\frac{1}{N}(X_N - \tilde{X}_N)^2\right]\right)^{1/2} \ge \delta/2\right) \\
&\le \frac{4\|f\|_{\mathrm{Lip}}^2}{\delta^2}\, \mathbb{E}\left[\mathrm{Tr}\left[\frac{1}{N}(X_N - \tilde{X}_N)^2\right]\right] \\
&= \frac{4\|f\|_{\mathrm{Lip}}^2}{\delta^2}\frac{1}{N^2}\, \mathbb{E}\left[\mathrm{Tr}\left[(Y_N - \tilde{Y}_N)^2\right]\right] \\
&= \frac{4\|f\|_{\mathrm{Lip}}^2}{\delta^2}\frac{1}{N^2}\, \mathbb{E}\left[\sum_{i,j=1}^N (Y_{ij} - \tilde{Y}_{ij})^2\right] \\
&= \frac{4\|f\|_{\mathrm{Lip}}^2}{\delta^2}\frac{1}{N^2}\left(N(N-1)\,\mathbb{E}[(Y_{12} - \tilde{Y}_{12})^2] + N\,\mathbb{E}[(Y_{11} - \tilde{Y}_{11})^2]\right) \\
&\le \frac{4\|f\|_{\mathrm{Lip}}^2}{\delta^2}\left(\mathbb{E}[(Y_{12} - \tilde{Y}_{12})^2] + \mathbb{E}[(Y_{11} - \tilde{Y}_{11})^2]\right) \le \varepsilon/2,
\end{aligned}$$

where by Lemma 2.22 the last quantity can be made arbitrarily small by choosing a sufficiently large C. The result follows by combining with the inequality (2.1). □

Remark 2.29. In our cutoff, we could also have taken $\tilde{Y}_{ii} = 0$, and the argument would still work.

Chapter 3

Concentration of measure techniques.

Concentration inequalities are estimates on quantities of the form

$$F(x_1, \dots, x_n) - \mathbb{E}[F(x_1, \dots, x_n)],$$

for (almost) independent and (almost) identically distributed random variables x_i with distributions drawn from some class, and sufficiently nice functions F. Typically, this means that $F \in \mathrm{Lip}(\mathbb{R}^n)$. Our main interest is in the random variables being entries of a random matrix. The lemma following the remark contains natural examples of Lipschitz functions of such entries.

Remark 3.1 (Norms)**.** The *Frobenius norm* of a real matrix A is

$$\|A\|_F = \sqrt{\mathrm{Tr}[AA^T]} = \sqrt{\sum_{i,j=1}^{N} a_{ij}^2}.$$

For a symmetric matrix we may re-write this as

$$\|A\|_F = \sqrt{2\sum_{i<j} a_{ij}^2 + \sum_i a_{ii}^2}.$$

On the other hand, in the arguments below we will need to identify A with a vector in $\frac{N(N+1)}{2}$-dimensional space, with norm

$$\|A\| = \sqrt{\sum_{i<j} a_{ij}^2 + \sum_i a_{ii}^2}$$

Clearly $\|A\|_F \le \sqrt{2}\,\|A\|$. We will also occasionally use the operator norm, defined as

$$\|A\|_{op} = \sup_{\|v\|\neq 0} \frac{\|Av\|}{\|v\|} = \sup_{\|u\|,\|v\|\neq 0} \frac{|\langle Av, u\rangle|}{\|u\|\,\|v\|}.$$

Exercise 3.2. Prove that for any matrix (symmetric or not) $\|A\|_{op} \leq \|A\|_F$. Clearly this implies that the map $A \mapsto \|A\|_{op}$ is Lipschitz.

Lemma 3.3. *Let X be a symmetric $N \times N$ matrix.*

a. For each k, the map $X \mapsto \lambda_k(X)$ is Lipschitz of norm at most $\sqrt{2}$.

b. Let $f \in \mathrm{Lip}(\mathbb{R})$. Extend f to a map on symmetric matrices by

$$f_{\mathrm{Tr}}(X) = \sum_{i=1}^{N} f(\lambda_i(X)) = \mathrm{Tr}[f(X)].$$

Then f_{Tr} is Lipschitz and $\|f_{\mathrm{Tr}}\|_{\mathrm{Lip}} \leq \sqrt{2N}\|f\|_{\mathrm{Lip}}$.

Proof. For part (a),

$$|\lambda_k(A) - \lambda_k(B)| \leq \sqrt{\sum_{i=1}^{N} |\lambda_i(A) - \lambda_i(B)|^2} \leq \|A - B\|_F \leq \sqrt{2}\,\|A - B\|\,.$$

Similarly, for part (b),

$$\begin{aligned}
|f_{\mathrm{Tr}}(A) - f_{\mathrm{Tr}}(B)| &= \left|\sum_{i=1}^{N} f(\lambda_i(A)) - f(\lambda_i(B))\right| \\
&\leq \sum_{i=1}^{N} |f(\lambda_i(A)) - f(\lambda_i(B))| \\
&\leq \|f\|_{\mathrm{Lip}} \sum_{i=1}^{N} |\lambda_i(A) - \lambda_i(B)| \\
&\leq \|f\|_{\mathrm{Lip}} \sqrt{N} \sqrt{\sum_{i=1}^{N} |\lambda_i(A) - \lambda_i(B)|^2} \\
&\leq \|f\|_{\mathrm{Lip}} \sqrt{N}\,\|A - B\|_F \\
&\leq \|f\|_{\mathrm{Lip}} \sqrt{2N}\,\|A - B\|\,.
\end{aligned}$$

□

3.1 Gaussian concentration.

Theorem 3.4. *Let $X = (X_1, \ldots, X_n)$ be i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, and $F \in \mathrm{Lip}(\mathbb{R}^n)$. Then for all $\lambda \in \mathbb{R}$,*

$$\mathbb{E} \exp\left(\lambda(F(X) - \mathbb{E}[F(X)])\right) \leq \exp\left(\pi^2 \lambda^2 \sigma^2 \|F\|_{\mathrm{Lip}}^2 / 8\right).$$

Therefore for all $\delta > 0$,

$$P\left(|F(X) - \mathbb{E}[F(X)]| \geq \delta\right) \leq 2\exp\left(-2\delta^2/\pi^2\sigma^2\|F\|_{\text{Lip}}^2\right).$$

We follow the "duplication argument" of Maurey and Pisier as presented in Theorem 2.1.12 of Terry Tao's book. For a more conceptual approach using the Ornstein-Uhlenbeck semigroup, see Section 10 in Todd Kemp's notes. First we note two basic properties of the multivariate normal distribution.

Exercise 3.5. Let $X = (X_1, \ldots, X_n)$ be a vector of independent normal random variables, with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

a.
$$c \cdot X = c_1 X_1 + \ldots + c_n X_n$$
is also Gaussian, with mean $\sum_{i=1}^n c_i \mu_i$ and variance $\sum_{i=1}^n |c_i|^2 \sigma_i^2$.

b. Assume in addition that all X_i's are i.i.d. normal with $X_i \sim \mathcal{N}(0, \sigma^2)$. Let U be an orthogonal matrix. Then
$$UX = \left(\sum_{j=1}^N U_{ij} X_j\right)_{i=1}^n$$
has the same distribution as X, so that its components are independent standard normals. Hint: recall that for jointly normal variables, uncorrelated implies independent.

Remark 3.6. We briefly recall the notion of conditional expectation. Instead of giving the definition, we only list two key properties. First, for random variables Y and Z,

$$\mathbb{E}[\mathbb{E}[Z|Y]] = \mathbb{E}[Z].$$

Second, if f, g are functions and X, Y are independent random variables,

$$\mathbb{E}[f(X)g(Y)|Y] = \mathbb{E}[f(X)]g(Y).$$

Proof of the theorem. **Step I.** We first show how the second part of the theorem follows from the first. By Markov inequality,

$$\begin{aligned} P\left(|F(X) - \mathbb{E}[F(X)]| \geq \delta\right) &= P\left(\exp(\lambda\,|F(X) - \mathbb{E}[F(X)]|) \geq e^{\lambda\delta}\right) \\ &\leq e^{-\lambda\delta}\,\mathbb{E}[\exp(\lambda\,|F(X) - \mathbb{E}[F(X)]|)] \leq 2e^{-\lambda\delta}\exp\left(\pi^2\lambda^2\sigma^2\|F\|_{\text{Lip}}^2/8\right) \end{aligned}$$

where we use $e^{|x|} \leq e^x + e^{-x}$ and apply the first part of the theorem to both F and $-F$. By taking

$$\lambda = \frac{4\delta}{\pi^2\sigma^2\|F\|_{\text{Lip}}^2},$$

we get the result.

Step II. We assume for now that F is smooth. By definition of the gradient and of the Lipschitz norm, for a vector u,

$$|\nabla F(x) \cdot u| = \lim_{h \to 0} \left| \frac{F(x+hu) - F(x)}{h} \right| \leq \|F\|_{\text{Lip}} \|u\| .$$

Taking $u = \nabla F(x)$, we conclude that $\|\nabla F(x)\| \leq \|F\|_{\text{Lip}}$ for all x.

By subtracting a constant from F (which does not change its gradient), we may assume that $\mathbb{E}[F(X)] = 0$.

Step III. (Duplication trick) Let Y be an independent copy of X. Since $\mathbb{E}[F(Y)] = 0$, we see from Jensen's inequality that

$$\mathbb{E}[\exp(-\lambda F(Y))] \geq \exp(-\lambda\, \mathbb{E}[F(Y)]) = 1$$

and thus (by independence of X and Y)

$$\mathbb{E}[\exp(\lambda F(X))] \leq \mathbb{E}[\exp(\lambda F(X))]\, \mathbb{E}[\exp(-\lambda F(Y))] = \mathbb{E}[\exp(\lambda(F(X) - F(Y)))].$$

It thus suffices to estimate $\mathbb{E}[\exp(\lambda(F(X) - F(Y)))]$, which is natural for Lipschitz F.

We first use the fundamental theorem of calculus along a circular arc to write

$$F(X) - F(Y) = \int_0^{\pi/2} \frac{d}{d\theta} F(Y \cos\theta + X \sin\theta)\, d\theta.$$

Note that $X_\theta = Y\cos\theta + X \sin\theta$ is another gaussian random variable equivalent to X, as is its derivative $X'_\theta = -Y\sin\theta + X\cos\theta$; furthermore, and crucially, these two random variables are independent. Applying Jensen's inequality for the probability density $\frac{2}{\pi}\mathbf{1}_{[0,\pi/2]}$, we get

$$\exp(\lambda(F(X) - F(Y))) = \exp\left(\lambda \frac{2}{\pi} \int_0^{\pi/2} \frac{\pi}{2} \frac{d}{d\theta} F(X_\theta)\, d\theta\right) \leq \frac{2}{\pi} \int_0^{\pi/2} \exp\left(\lambda \frac{\pi}{2} \frac{d}{d\theta} F(X_\theta)\right) d\theta.$$

Applying the chain rule and taking expectations, we have

$$\mathbb{E}[\exp(\lambda(F(X) - F(Y)))] \leq \frac{2}{\pi} \int_0^{\pi/2} \mathbb{E}\left[\exp\left(\frac{\lambda\pi}{2} \nabla F(X_\theta) \cdot X'_\theta\right)\right] d\theta.$$

Let us first condition X_θ to be fixed. Recalling that X'_θ is equidistributed with X, we conclude that $\frac{\lambda\pi}{2}\nabla F(X_\theta) \cdot X'_\theta$ is normally distributed with standard deviation at most

$$\frac{\lambda\pi}{2} \sqrt{\sum_{i=1}^{N} (\nabla F(X_\theta))_i^2 \sigma^2} \leq \frac{\pi}{2} \lambda\sigma \|F\|_{\text{Lip}}.$$

Therefore its moment generating function

$$\mathbb{E}\left[\exp\left(\frac{\lambda\pi}{2} \nabla F(X_\theta) \cdot X'_\theta\right) \middle| X_\theta\right] \leq \exp\left(\pi^2 \lambda^2 \sigma^2 \|F\|_{\text{Lip}}^2 / 8\right).$$

Taking now the expectation with respect to X_θ, the result follows.

Step IV. To approximate a general Lipschitz function F by smooth functions, we follow the argument of Todd Kemp in the proof of his Theorem 11.1. Let $\{\psi_\varepsilon : \varepsilon > 0\}$ be a smooth compactly supported approximate identity on $\mathbb{R}^n$. That is, $\psi \in C_c^\infty(\mathbb{R}^n)$ is a non-negative function with support in the unit ball B_1 and total integral 1, and $\psi_\varepsilon(x) = \varepsilon^{-n}\psi(x/\varepsilon)$. Then ψ_ε is also non-negative, has total integral 1, and is supported in B_ε. Let

$$F_\varepsilon(x) = (F * \psi_\varepsilon)(x) = \int_{\mathbb{R}^n} F(x-y)\psi_\varepsilon(y)\,dy.$$

Then F_ε is smooth, and so the theorem has been proven for it. Also,

$$\begin{aligned} |F(x) - F_\varepsilon(x)| &= \left| \int F(x)\psi_\varepsilon(y)\,dy - \int F(x-y)\psi_\varepsilon(y)\,dy \right| \\ &\le \int |F(x) - F(x-y)|\,\psi_\varepsilon(y)\,dy \\ &\le \|F\|_{\text{Lip}} \int |y|\,\psi_\varepsilon(y)\,dy \le \varepsilon \|F\|_{\text{Lip}} \end{aligned}$$

since F_ε is supported in B_ε. Thus $F_\varepsilon \to F$ uniformly as $\varepsilon \to 0$. Then $\mathbb{E}[F_\varepsilon(X)] \to \mathbb{E}[F(X)]$ and $\mathbb{E}\exp\left(\lambda(F_\varepsilon(X) - \mathbb{E}[F_\varepsilon(X)])\right) \to \mathbb{E}\exp\left(\lambda(F(X) - \mathbb{E}[F(X)])\right)$ as $\varepsilon \to 0$ by the bounded convergence theorem (since the distribution of X is a probability measure). Finally, by similar reasoning

$$\begin{aligned} |F_\varepsilon(x) - F_\varepsilon(y)| &= \left| \int F(x-z)\psi_\varepsilon(z)\,dz - \int F(y-z)\psi_\varepsilon(z)\,dz \right| \\ &\le \int |F(x-z) - F(y-z)|\,\psi_\varepsilon(z)\,dz \\ &\le \|F\|_{\text{Lip}} \int |x-y|\,\psi_\varepsilon(z)\,dy = \|F\|_{\text{Lip}}\,\|x-y\| \end{aligned}$$

and so $\|F_\varepsilon\|_{\text{Lip}} \le \|F\|_{\text{Lip}}$. Therefore

$$\begin{aligned} \mathbb{E}\exp\left(\lambda(F(X) - \mathbb{E}[F(X)])\right) &= \lim_{\varepsilon \downarrow 0} \mathbb{E}\exp\left(\lambda(F_\varepsilon(X) - \mathbb{E}[F_\varepsilon(X)])\right) \\ &\le \lim_{\varepsilon \downarrow 0} \exp\left(\pi^2\lambda^2\sigma^2\|F_\varepsilon\|_{\text{Lip}}^2/8\right) \\ &\le \exp\left(\pi^2\lambda^2\sigma^2\|F\|_{\text{Lip}}^2/8\right). \end{aligned}$$ □

Exercise 3.7. Let $Z = (Z_1, \ldots, Z_n)$ be i.i.d. $\mathcal{N}(0,1)$ random variables. Let Σ be a positive definite matrix, and define $X = \Sigma^{1/2}Z$. Then X is a jointly normal vector with mean zero and covariance matrix Σ. Let $F \in \text{Lip}(\mathbb{R}^n)$. Then for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\exp\left(\lambda(F(X) - \mathbb{E}[F(X)])\right) \le \exp\left(\pi^2\lambda^2\,\|\Sigma\|_{op}\,\|F\|_{\text{Lip}}^2/8\right),$$

and so for all $\delta > 0$,

$$P\left(|F(X) - \mathbb{E}[F(X)]| \geq \delta\right) \leq 2\exp\left(-2\delta^2/\pi^2 \|\Sigma\|_{op} \|F\|_{\text{Lip}}^2\right).$$

In particular if $\Sigma_{ij} = \delta_{ij}\sigma_i$,

$$P\left(|F(X) - \mathbb{E}[F(X)]| \geq \delta\right) \leq 2\exp\left(-2\delta^2/\pi^2 \max_i(\sigma_i^2)\|F\|_{\text{Lip}}^2\right).$$

3.2 Concentration results for GOE.

Now let X_N be a GOE matrix, $f \in \text{Lip}(\mathbb{R})$, and $F = f_{\text{Tr}}$. Note that

$$F(X_N) = N \int f \, d\hat{\mu}_{X_N}$$

and for each matrix entry, the variance is at most $\frac{2}{N}$. Then

$$\begin{aligned} P\left(\left|\int f \, d\hat{\mu}_{X_N} - \mathbb{E}\left[\int f \, d\hat{\mu}_{X_N}\right]\right| \geq \delta\right) &= P\left(|F(X_N) - \mathbb{E}[F(X_N)]| \geq N\delta\right) \\ &\leq 2\exp\left(-N^2 2\delta^2/\pi^2\sigma^2\|F\|_{\text{Lip}}^2\right) \\ &\leq 2\exp\left(-N\delta^2/\pi^2\sigma^2\|f\|_{\text{Lip}}^2\right) \\ &\leq 2\exp\left(-N^2\delta^2/2\pi^2\|f\|_{\text{Lip}}^2\right). \end{aligned}$$

Similarly,

$$P\left(|\lambda_k(X_N) - \mathbb{E}[\lambda_k(X_N)]| \geq \delta\right) \leq 2\exp\left(-\delta^2/\pi^2\sigma^2\right) = 2\exp\left(-N\delta^2/2\pi^2\right)$$

Thus linear statistics concentrate at the rate of $\delta \sim \frac{1}{N}$ (which is consistent with our moment method results) while the eigenvalues appear to concentrate only at the rate of $\delta \sim \frac{1}{\sqrt{N}}$.

Since the operator norm $\|A\|_{op}$ is less than the Frobenius norm, it also has Lipschitz constant at most $\sqrt{2}$, and

$$P\left(\left|\|X_N\|_{op} - \mathbb{E}[\|X_N\|_{op}]\right| \geq \delta\right) \leq 2\exp\left(-N\delta^2/2\pi^2\right).$$

This last inequality holds also for non-symmetric Gaussian matrices.

3.3 Other concentration inequalities.

The arguments in the preceding section only worked for Gaussian entries. Here are some alternative conditions on the matrix entries leading to roughly the same conclusions.

Definition 3.8. Let μ be a probability measure on $\mathbb{R}^n$. The μ-entropy of a function f is

$$\mathrm{Ent}_\mu(f) = \int f \log f \, d\mu - \int f \, d\mu \cdot \log \int f \, d\mu.$$

μ satisfies the *logarithmic Sobolev inequality* with constant c if, for any continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(f^2) \leq 2c \int \|\nabla f\|^2 \, d\mu.$$

Gaussian measure satisfies LSI, as does any density of the form $\frac{1}{Z}e^{-V}$ for sufficiently smooth potential V (see below), as does the joint distribution of independent random variables each satisfying the LSI.

Lemma 3.9 (Herbst)**.** *Suppose the joint distribution of random variables μ_X satisfies LSI on $\mathbb{R}^n$ with constant c. For $F \in \mathrm{Lip}(\mathbb{R}^n)$,*

$$P(|F(X) - \mathbb{E}[F(X)]| \geq \delta) \leq 2\exp(-\delta^2/2c\|F\|_{\mathrm{Lip}}^2).$$

Proof. As in the proof of Theorem 3.4, we may assume that $\mathbb{E}[F(X)] = 0$, F is smooth, and it suffices to show that for all λ,

$$\mathbb{E}\exp\left(\lambda F(X)\right) \leq \exp\left(c\lambda^2\|F\|_{\mathrm{Lip}}^2/2\right).$$

Let $f(X) = e^{\lambda F(X)/2}$, $f^2(X) = e^{\lambda F(X)}$ and $\varphi(\lambda) = \mathbb{E}[e^{\lambda F(X)}]$. Then

$$\mathrm{Ent}_\mu(f^2) = \int e^{\lambda F(X)}\lambda F(X) \, d\mu - \int e^{\lambda F(X)} \, d\mu \cdot \log \int e^{\lambda F(X)} \, d\mu = \lambda\varphi'(\lambda) - \varphi(\lambda)\log\varphi(\lambda)$$

while

$$2c\int \|\nabla f\|^2 \, d\mu = 2c\int e^{\lambda F(X)}\frac{\lambda^2}{4}\|\nabla F\|^2(X) \, d\mu \leq \frac{c\lambda^2}{2}\|F\|_{\mathrm{Lip}}^2 \int e^{\lambda F(X)} \, d\mu = \frac{c\lambda^2}{2}\varphi(\lambda)\|F\|_{\mathrm{Lip}}^2.$$

Applying the LSI and dividing both sides by $\lambda^2\varphi(\lambda)$, we get

$$\frac{\varphi'(\lambda)}{\lambda\varphi(\lambda)} - \frac{\log\varphi(\lambda)}{\lambda^2} \leq \frac{c}{2}\|F\|_{\mathrm{Lip}}^2.$$

Note that for $\lambda > 0$, the left-hand side is precisely $\frac{d}{d\lambda}\frac{\log\varphi(\lambda)}{\lambda}$. Thus

$$\frac{d}{d\lambda}\frac{\log\varphi(\lambda)}{\lambda} \leq \frac{c}{2}\|F\|_{\mathrm{Lip}}^2.$$

Moreover

$$\lim_{\lambda\to 0}\frac{\log\varphi(\lambda)}{\lambda} = \lim_{\lambda\to 0}\frac{\log\varphi(\lambda) - \log\varphi(0)}{\lambda} = \frac{\varphi'(0)}{\varphi(0)} = \mathbb{E}[F(X)] = 0.$$

Therefore

$$\frac{1}{\lambda_0}\log\varphi(\lambda_0) - 0 = \int_0^{\lambda_0} \left(\frac{d}{d\lambda}\frac{\log\varphi(\lambda)}{\lambda}\right)\, d\lambda \leq \frac{c\lambda_0 \|F\|_{\text{Lip}}^2}{2}$$

and so

$$\varphi(\lambda) \leq \exp\left(\frac{c\lambda^2\|F\|_{\text{Lip}}^2}{2}\right),$$

which is the desired result. □

Proposition 3.10 (Corollary of the Bakry-Emery criterion)**.** *Let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be at least twice continuously differentiable growing sufficiently fast so that the probability measure*

$$\mu_\Phi(dx) = \frac{1}{Z}\exp(-\Phi(x_1,\ldots,x_n))\, dx_1 \ldots\, dx_n$$

is well defined. Write $\text{Hess}(\Phi)_{ij} = \partial_i\partial_j\Phi$. If for all x,

$$\text{Hess}(\Phi)(x) \geq \frac{1}{c}I$$

as matrices, then μ_Φ satisfies the LSI with constant c.

Corollary 3.11. *Suppose X is*

either a Wigner matrix with un-normalized entries satisfying the LSI with constant c

or is drawn from an orthogonally invariant ensemble $\frac{1}{Z_N}e^{-N\,\text{Tr}[V(X)]}\, dX$ with $V''(x) \geq \frac{1}{c} > 0$. Then

for any Lipschitz f,

$$P\left(\left|\int f\, d\hat{\mu}_N - \mathbb{E}\left[\int f\, d\hat{\mu}_N\right]\right| \geq \delta\right) \leq 2\exp\left(-N^2\delta^2/4c\|f\|_{\text{Lip}}^2\right)$$

and for any k,

$$P\left(|\lambda_k(X_N) - \mathbb{E}\left[\lambda_k(X_N)\right]| \geq \delta\right) \leq 2\exp\left(-N\delta^2/4c\right)$$

The corollary applies for example to $V(x) = |x|^a$, $a \geq 2$, but not for $a < 2$. For $1 \leq a < 2$, we may still get a weaker form of concentration using the following ideas.

Definition 3.12. Let μ be a probability measure on $\mathbb{R}^n$. The μ-variance of a function f is

$$\text{Var}_\mu[f] = \int \left(f - \int f\, d\mu\right)^2 d\mu$$

μ satisfies the *Poincaré inequality* with constant m if, for any continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$,

$$\text{Var}_\mu[f] \leq \frac{1}{m}\int \|\nabla f\|^2\, d\mu.$$

Exercise 3.13. If μ satisfies the logarithmic Sobolev inequality with constant c, show that it satisfies the Poincaré inequality with an appropriate constant. Hint: apply LSI to $f = 1 + \varepsilon g$.

Proposition 3.14. *Suppose the joint distribution of random variables μ_X satisfies the PI with constant m on $\mathbb{R}^n$. For $F \in \mathrm{Lip}(\mathbb{R}^n)$,*

$$P\left(|F(X) - \mathbb{E}[F(X)]| \geq \delta\right) \leq 2K \exp\left(-\sqrt{m}\delta/\sqrt{2}\|F\|_{\mathrm{Lip}}\right),$$

where K is determined by m.

Proof. We may again assume that F is smooth, and it suffices to show that for sufficiently small $|\lambda|$,

$$\mathbb{E}\left[\exp\left(\lambda(F(X) - \mathbb{E}[F(X)])\right)\right] \leq K.$$

Apply the PI to $f(X) = e^{\lambda F(X)/2}$. We get

$$\mathbb{E}[e^{\lambda F(X)}] - \mathbb{E}[e^{\lambda F(X)/2}]^2 \leq \frac{1}{4m}\lambda^2\|F\|_{\mathrm{Lip}}^2\,\mathbb{E}[e^{\lambda F(X)}],$$

so that

$$\mathbb{E}[e^{\lambda F(X)}] \leq \left(1 - \frac{1}{4m}\lambda^2\|F\|_{\mathrm{Lip}}^2\right)^{-1}\mathbb{E}[e^{\lambda F(X)/2}]^2$$

for sufficiently small $|\lambda|$. That is,

$$\log\mathbb{E}[e^{\lambda F(X)}] \leq -\log\left(1 - \frac{1}{4m}\lambda^2\|F\|_{\mathrm{Lip}}^2\right) + 2\,\mathbb{E}[e^{\lambda F(X)/2}].$$

Iterating,

$$\log\mathbb{E}[e^{\lambda F(X)}] \leq -\sum_{j=1}^{n} 2^{j-1}\log\left(1 - \frac{1}{4^j m}\lambda^2\|F\|_{\mathrm{Lip}}^2\right) + 2^n\,\mathbb{E}[e^{\lambda F(X)/2^n}]$$

Since $\lim_{n\to\infty} 2^n\,\mathbb{E}[e^{\lambda F(X)/2^n}] = \mathbb{E}[F(X)]$, it follows that

$$\log\mathbb{E}[e^{\lambda(F(X)-\mathbb{E}[F(X)])}] \leq -\sum_{j=1}^{\infty} 2^{j-1}\log\left(1 - \frac{1}{4^j m}\lambda^2\|F\|_{\mathrm{Lip}}^2\right)$$

Since the right-hand side is an increasing function of λ, by taking $\lambda = \sqrt{m}/\|F\|_{\mathrm{Lip}}^2$ we get an upper estimate

$$\log\mathbb{E}[e^{\lambda(F(X)-\mathbb{E}[F(X)])}] \leq -\sum_{j=1}^{\infty} 2^{j-1}\log\left(1 - \frac{1}{4^j}\right) = \log K < \infty$$

since $-\sum_{j=1}^{\infty} 2^{j-1}\log\left(1 - \frac{1}{4^j}\right) \sim \sum_{j=1}^{\infty} 2^{j-1}\frac{1}{4^j}$. $\square$

For Wigner matrices, having un-normalized entries satisfying the PI with a uniform constant leads to concentration of the empirical spectral distribution at the rate e^{-NC}.

Proposition 3.15 (Talagrand)**.** *Let the i.i.d. random variables be bounded, with $|X_i| \leq K/2$. Suppose that F is a* convex *Lipschitz function. Then*

$$P(|F(X) - MF(X)| \geq \delta) \leq 4\exp\left(-\delta^2/16K^2\|F\|_{\text{Lip}}^2\right),$$

where $MF(X)$ is the median of $F(X)$.

Remark 3.16. Note that in this case,

$$\begin{aligned}|\mathbb{E}[F(X)] - MF(X)| &\leq \mathbb{E}[|F(X) - MF(X)|] \\ &= \int_0^\infty P(|F(X) - MF(X)| > t)\,dt \\ &\leq \int_0^\infty 4\exp\left(-t^2/16K^2\|F\|_{\text{Lip}}^2\right) = 8\sqrt{\pi}K\|F\|_{\text{Lip}},\end{aligned}$$

which is small if $K\|F\|_{\text{Lip}}$ is. It follows that

$$\begin{aligned}P(|F(X) - \mathbb{E}[F(X)]| \geq N\delta) &\leq P(|F(X) - MF(X)| + |\mathbb{E}[F(X)] - MF(X)| \geq N\delta) \\ &\leq P(|F(X) - MF(X)| \geq N\delta - 8\sqrt{\pi}K\|F\|_{\text{Lip}}) \\ &\leq 4\exp\left(-(N\delta - 8\sqrt{\pi}K\|F\|_{\text{Lip}})^2/16K^2\|F\|_{\text{Lip}}^2\right) \\ &= 4e^{-4\pi}e^{N\delta\sqrt{\pi}/K\|F\|_{\text{Lip}}}\exp\left(-N^2\delta^2/16K^2\|F\|_{\text{Lip}}^2\right).\end{aligned}$$

For a Wigner matrix with bounded entries and a convex $F = f_{\text{Tr}}$, $K \sim \frac{1}{\sqrt{N}}$ and $F \sim \sqrt{N}$, so we have Gaussian concentration with N.

Obviously (any) matrix norm is a convex function of the matrix.

Exercise 3.17. Let A be a symmetric matrix, with the largest eigenvalue $\lambda_N(A)$.

a. Prove that $\lambda_N(A) = \sup\{\langle Av, v\rangle : \|v\| = 1\}$.

b. Prove that λ_N is a convex function of A.

c. Prove that the smallest eigenvalue λ_1 is a concave function of A. Hint: use $-A$.

Proposition 3.18 (Klein's Lemma)**.** *If f is a convex function, then so is f_{Tr}.*

Proof. By approximation, we may assume that f is twice differentiable and $f'' \geq c > 0$. Then

$$R_f(x, y) = f(x) - f(y) - (x - y)f'(y) \geq \frac{c}{2}(x - y)^2.$$

Let X have eigenvalues $\{\lambda_j(X)\}$ with unit eigenvectors $\{\xi_i(X)\}$, and similarly for Y. Denote $c_{ij} = |\langle \xi_i(X), \xi_j(Y)\rangle|^2$. Then

$$\begin{aligned}
\langle \xi_i(X), R_f(X,Y)\xi_i(X)\rangle &= \langle \xi_i(X), f(X) - f(Y) - (X-Y)f'(Y)\rangle\, \xi_i(X) \\
&= f(\lambda_i(X)) + \sum_j \left(-c_{ij} f(\lambda_j(Y)) - c_{ij}\lambda_i(X) f'(\lambda_j(Y)) + c_{ij}\lambda_j(Y) f'(\lambda_j(Y))\right) \\
&= \sum_j c_{ij} \left(f(\lambda_i(X)) - f(\lambda_j(Y)) - \lambda_i(X) f'(\lambda_j(Y)) + \lambda_j(Y) f'(\lambda_j(Y))\right) \\
&= \sum_j c_{ij} R_f(\lambda_i(X), \lambda_j(Y)) \geq \frac{c}{2} \sum_j c_{ij} (\lambda_i(X) - \lambda_j(Y))^2,
\end{aligned}$$

where we used $\sum_j c_{ij} = 1$. Now summing over i, we obtain

$$\mathrm{Tr}[f(X) - f(Y) - (X-Y)f'(Y)] \geq \frac{c}{2} \sum_{i,j} c_{ij} (\lambda_i(X) - \lambda_j(Y))^2$$

Applying this argument to $f(x) = x^2$ with $R_{x^2}(x,y) = (x-y)^2$, we see that

$$\sum_{i,j} c_{ij} (\lambda_i(X) - \lambda_j(Y))^2 = \mathrm{Tr}[(X-Y)^2].$$

Thus finally,

$$\mathrm{Tr}[f(X) - f(Y) - (X-Y)f'(Y)] \geq \frac{c}{2}\, \mathrm{Tr}[(X-Y)^2] \geq 0.$$

For $(X,Y) = (A, \frac{1}{2}(A+B))$ this gives

$$\mathrm{Tr}\left[f(A) - f\left(\frac{1}{2}(A+B)\right) - \frac{1}{2}(A-B) f'\left(\frac{1}{2}(A+B)\right)\right] \geq 0$$

while for $(X,Y) = (B, \frac{1}{2}(A+B))$ this gives

$$\mathrm{Tr}\left[f(B) - f\left(\frac{1}{2}(A+B)\right) - \frac{1}{2}(B-A) f'\left(\frac{1}{2}(A+B)\right)\right] \geq 0.$$

Adding these inequalities, we obtain

$$\mathrm{Tr}\left[f(A) + f(B) - 2f\left(\frac{1}{2}(A+B)\right)\right] \geq 0$$

and so

$$\frac{1}{2} f_{\mathrm{Tr}}(A) + \frac{1}{2} f_{\mathrm{Tr}}(B) \geq f_{\mathrm{Tr}}\left(\frac{1}{2}A + \frac{1}{2}B\right). \qquad \square$$

By cutoff arguments as in Section 2.3 we can thus obtain concentration results (for convex f) for quite general Wigner-type matrices.

Chapter 4

The Stieltjes transform methods.

Stieltjes transform methods in random matrix theory were introduced by Leonid Pastur and collaborators (1967–) in their study of Wishart matrices. They have been developed by many contributors, and are used throughout the theory, in the study of many other classes, such as band matrices and spiked models.

4.1 General properties.

The transform method, using Fourier transforms or moment generating functions, is a standard technique in probability theory. The transform most appropriate for random matrix theory is the *Stieltjes transform*.

Complex-analytic properties.

For a probability measure μ on $\mathbb{R}$, its Stieltjes transform is the function

$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\mu(x).$$

Sometimes it is called the Cauchy or the Borel transform, or is defined as $\int_{\mathbb{R}} \frac{1}{z-x}\, d\mu(x)$. Note that for any $z \in \mathbb{C} \setminus \mathbb{R}$,

$$\left|\frac{1}{x-z}\right| \leq \frac{1}{|\Im z|}.$$

So the function $x \mapsto \frac{1}{x-z}$ is bounded, and $S_\mu(z)$ is well defined on this set (in fact it can also be extended to $\mathbb{R} \setminus \operatorname{supp}(\mu)$). It is also clear that $S_\mu(\overline{z}) = \overline{S_\mu(z)}$. Moreover, we may differentiate under the integral to obtain $S'_\mu(z) = \int_{\mathbb{R}} \frac{1}{(x-z)^2}\, d\mu(x)$, so S_μ is analytic on $\mathbb{C} \setminus \mathbb{R}$. For later use, we record that for any $z \in \mathbb{C}^+$,

$$\left\| x \mapsto \frac{1}{z-x} \right\|_{\text{Lip}} \leq \frac{1}{(\Im z)^2}. \tag{4.1}$$

Next, we note that

$$|iyS_\mu(iy)+1| = \left|\int_{\mathbb{R}} \frac{iy}{x-iy}\,d\mu(x)+1\right| \le \int_{\mathbb{R}} \frac{|x|}{\sqrt{x^2+y^2}}\,d\mu(x) \to 0$$

as $y \to \infty$ by Dominated Convergence. Thus

$$\lim_{y\to\infty} iyS_\mu(iy) = -1. \tag{4.2}$$

Finally, we compute

$$S_\mu(x+iy) = \int_{\mathbb{R}} \frac{1}{t-x-iy}\,d\mu(t) = \int_{\mathbb{R}} \frac{t-x}{(x-t)^2+y^2}\,d\mu(t) + i\int_{\mathbb{R}} \frac{y}{(x-t)^2+y^2}\,d\mu(t).$$

In particular, we note that S_μ maps $\mathbb{C}^+$ to itself. Moreover for $\varepsilon > 0$, denote

$$P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2}.$$

Then $\Im S_\mu(x+iy) = \pi(\mu * P_y)(x)$. The family $\{P_y : y > 0\}$ is called the Poisson kernel for $\mathbb{C}^+$. Note that $P_y(x)\,dx = P_1(x/y)\,d(x/y)$.

Theorem 4.1 (Stieltjes)**.** *Any analytic function $S : \mathbb{C}^+ \to \mathbb{C}^+$ satisfying equation* (4.2) *is a Stieltjes transform of some probability measure.*

We will not prove this theorem, but the measure corresponding to S is identified through the

Lemma 4.2 (Stieltjes Inversion Formula)**.** *For any probability measure μ, the measures*

$$\mu_y(dx) = \frac{1}{\pi}\Im S_\mu(x+iy)\,dx$$

converge weakly to μ as $y \downarrow 0$.

Proof. The Poisson kernel is an approximate identity: $P_y \ge 0$, $\int_{\mathbb{R}} P_y(x)\,dx = 1$, and for any $\varepsilon, \delta > 0$, for sufficiently small y, $\int_{|x|>\delta} P_y(x)\,dx < \varepsilon$. Then by general theory (cf. the proof of Theorem 3.4; recall details?) $\mu * P_y \to \mu$ vaguely. Since μ is a probability measure, we automatically get weak convergence. □

Remark 4.3. Suppose μ is compactly supported in $[-a, a]$. Then it is easy to see that $zS_\mu(z) \to -1$ as $z \to \infty$ and not just along the imaginary axis. Moreover the moments $|m_k(\mu)| \le a^k$, and so by the series version of the Dominated Convergence Theorem

$$-\sum_{k=0}^{\infty} \frac{m_k(\mu)}{z^{k+1}} = -\sum_{k=0}^{\infty} \int \frac{x^k}{z^{k+1}}\,d\mu(x) = -\int \sum_{k=0}^{\infty} \frac{x^k}{z^{k+1}}\,d\mu(x) = \int \frac{1}{x-z}\,d\mu(x) = S_\mu(z).$$

So $S_\mu(z)$ is an (ordinary) generating function for moments of μ.

Proposition 4.4. *A sequence of probability measures $\mu_N \to \mu$ converges weakly to a probability measure if and only if $S_{\mu_N}(z) \to S_\mu(z)$ pointwise for every $z \in \mathbb{C}^+$. It suffices to require convergence on a set which has an accumulation point.*

Proof. Since for every $z \in \mathbb{C}^+$, the function $x \mapsto \frac{1}{x-z}$ is in $C_0(\mathbb{R})$, one direction is clear. Now let $A \subset \mathbb{C}^+$ be a set with an accumulation point. Suppose $S_{\mu_N}(z) \to S_\mu(z)$ for all $z \in A$. Choose a subsequence such that $\mu_{N_k} \to \nu$ vaguely for some measure ν. Then by the other direction of the argument, $S_\nu(z) = S_\mu(z)$ for all $z \in A$. By analytic continuation, it follows that $S_\mu = S_\nu$ on $\mathbb{C}^+$, and so $\nu = \mu$ (and in particular it is a probability measure). Since this is true for any convergent subsequence, the result follows. $\square$

The Stieltjes transform of the empirical distribution of a matrix.

For a symmetric or Hermitian $N \times N$ matrix A,

$$S_{\hat{\mu}_A}(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\hat{\mu}_A(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i(A) - z} = \frac{1}{N}\operatorname{Tr}\left[(A - zI)^{-1}\right].$$

Here the operator $(A - zI)^{-1}$ is the *resolvent* of A.

4.2 Convergence of Stieltjes transforms for random matrices.

In this section, we will give another proof of weak convergence of empirical spectral distributions for general Wigner matrices under a slightly stronger assumption. One approach involves concentration inequalities. For example, we could assume that the matrix entries satisfy LSI, and then apply Herbst's method; or we could apply the cutoff procedure from Lemma 2.22 to suppose that the matrix entries are uniformly bounded, and apply Talagrand's inequality. We instead choose to avoid any sophisticated concentration techniques by assuming the finiteness of the fourth moments.

Theorem 4.5. *Let $X_N = \frac{1}{\sqrt{N}} Y_N$ be Wigner matrices as in Theorem 2.21. Thus Y_N is symmetric and otherwise has independent entries, $\{Y_{ij} : i < j\}$ are identically distributed with mean zero and variance 1, and $\{Y_{ii}\}$ are identically distributed with mean zero and variance at most $m_2 \geq 1$. We will additionally assume that the fourth moment $\mathbb{E}[Y_{ij}^4] = m_4 < \infty$. Then*

$$\hat{\mu}_{X_N} \to \sigma$$

weakly almost surely.

The rest of the section constitutes the proof of this theorem.

Let $Y_N^{(k)}$ be Y_N with the k'th row and column removed, and u_k be its k'th column with its k't entry removed. So for example

$$Y_N = \begin{pmatrix} (Y_N)_{11} & u_1^T \\ u_1 & Y_N^{(1)} \end{pmatrix}.$$

Also, let $\tilde{Y}_N^{(1)}$ be the $N \times N$ matrix obtained by adjoining to $Y_N^{(1)}$ a zero row and column. This matrix has the same eigenvalues as $Y_N^{(1)}$ plus an extra zero eigenvalue. Thus

$$\begin{aligned} S_{\tilde{Y}_N^{(1)}/\sqrt{N}}(z) &= \frac{1}{N}\sum_{j=1}^{N-1} \frac{1}{\lambda_j(Y_N^{(1)})/\sqrt{N} - z} - \frac{1}{N}\frac{1}{z} = \frac{1}{N}\sum_{j=1}^{N-1} \frac{1}{\lambda_j(Y_{N-1})/\sqrt{N} - z} - \frac{1}{N}\frac{1}{z} \\ &= \frac{1}{N}\frac{\sqrt{N}}{\sqrt{N-1}}\sum_{j=1}^{N-1} \frac{1}{\lambda_j(Y_{N-1})/\sqrt{N-1} - \frac{\sqrt{N}}{\sqrt{N-1}}z} - \frac{1}{N}\frac{1}{z} \\ &= \frac{1}{N-1}\frac{\sqrt{N-1}}{\sqrt{N}}\sum_{j=1}^{N-1} \frac{1}{\lambda_j(X_{N-1}) - \frac{\sqrt{N}}{\sqrt{N-1}}z} - \frac{1}{N}\frac{1}{z} \\ &= \frac{\sqrt{N-1}}{\sqrt{N}} S_{X_{N-1}}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right) - \frac{1}{N}\frac{1}{z}. \end{aligned} \tag{4.3}$$

Denote the Stieltjes transform of the empirical spectral distribution of X_N

$$S_N(z) = \int \frac{1}{x-z}\, d\hat{\mu}_{X_N}(x) = \frac{1}{N}\operatorname{Tr}\left[(X_N - zI)^{-1}\right]$$

and its average

$$\overline{S}_N(z) = \mathbb{E}\left[\int \frac{1}{x-z}\, d\hat{\mu}_{X_N}(x)\right] = \frac{1}{N}\mathbb{E}\left[\operatorname{Tr}\left[(X_N - zI)^{-1}\right]\right].$$

Our goal is to prove that S_N converges pointwise a.s., and to obtain an equation satisfied by the limiting function. The main strategy will be to relate S_N and S_{N-1}, in two ways. For the first relation, recall that for $N \times N$ matrices,

$$\left|\mathbb{E}\left[\int f\, d\hat{\mu}_A\right] - \mathbb{E}\left[\int f\, d\hat{\mu}_B\right]\right| \le \mathbb{E}\left[\left|\int f\, d\hat{\mu}_A - \int f\, d\hat{\mu}_B\right|\right] \le \|f\|_{\text{Lip}}\, \mathbb{E}\left[\left(\operatorname{Tr}\left[\frac{1}{N}(A-B)^2\right]\right)^{1/2}\right].$$

Applying this to $f(x) = \frac{1}{x-z}$, $A = X_N$ and $B = \tilde{Y}_N^{(1)}/\sqrt{N}$, and using equations (4.3) and (4.1), and Jensen's inequality,

$$\begin{aligned} &\left|\overline{S}_N(z) - \frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right) + \frac{1}{N}\frac{1}{z}\right| = \left|\mathbb{E}[S_{X_N}(z)] - \mathbb{E}[S_{\tilde{Y}_N^{(1)}/\sqrt{N}}(z)]\right| \\ &\le \frac{1}{(\Im z)^2}\mathbb{E}\left[\left(\frac{1}{N}\frac{2}{N}\sum_{j=1}^{N} Y_{j1}^2\right)^{1/2}\right] \le \frac{1}{(\Im z)^2}\frac{1}{N}\left(\mathbb{E}\left[2\sum_{j=1}^{N} Y_{j1}^2\right]\right)^{1/2} \le \frac{\sqrt{2m_2}}{(\Im z)^2}\frac{1}{\sqrt{N}}. \end{aligned} \tag{4.4}$$

Now we derive the second relation. Write

$$S_N(z) = \frac{1}{N}\operatorname{Tr}[(X_N - zI)^{-1}] = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{V_k},$$

where

$$V_k = \frac{1}{((X_N - zI)^{-1})_{kk}} = \frac{1}{\left(\left(\frac{1}{\sqrt{N}}Y_N - zI\right)^{-1}\right)_{kk}}.$$

Lemma 4.6 (Schur complement)**.** *Let*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

If D is invertible, then

$$\det M = \det(A - BD^{-1}C)\cdot\det D.$$

Proof. It suffices to note that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & B \\ 0 & D \end{pmatrix}\begin{pmatrix} 1 & 0 \\ D^{-1}C & 1 \end{pmatrix}.$$

□

Applying the lemma to

$$\frac{1}{\sqrt{N}}Y_N - zI = \begin{pmatrix} \left(\frac{1}{\sqrt{N}}Y_N - zI\right)_{11} & \frac{1}{\sqrt{N}}u_1^T \\ \frac{1}{\sqrt{N}}u_1 & \left(\frac{1}{\sqrt{N}}Y_N^{(1)} - zI\right) \end{pmatrix}$$

etc. and using Cramer's rule,

$$V_k = \frac{\det\left(\frac{1}{\sqrt{N}}Y_N - zI\right)}{\det\left(\frac{1}{\sqrt{N}}Y_N^{(k)} - zI\right)} = \frac{1}{\sqrt{N}}(Y_N)_{kk} - z - \frac{1}{N}u_k^T\left(\frac{1}{\sqrt{N}}Y_N^{(k)} - zI\right)^{-1}u_k. \tag{4.5}$$

Exercise 4.7. Let u be a vector of independent real random variables with mean zero and variance 1, and A a deterministic complex matrix. Then

$$\mathbb{E}[u^TAu] = \operatorname{Tr}[A]$$

If moreover A is symmetric and $\mathbb{E}[u_i^4] \le m_4$ for all i, then

$$\operatorname{Var}[u^TAu] = \mathbb{E}[(\overline{u^TAu})(u^TAu)] - \overline{\mathbb{E}[u^TAu]}\,\mathbb{E}[u^TAu] \le (2 + m_4)\operatorname{Tr}[\overline{A}A].$$

Therefore

$$
\begin{aligned}
\mathbb{E}[V_1 \mid Y_N^{(1)}] &= \frac{1}{\sqrt{N}}\,\mathbb{E}[(Y_N)_{11} \mid Y_N^{(1)}] - z - \frac{1}{N}\,\mathbb{E}\left[u_1^T \left(\frac{1}{\sqrt{N}}Y_N^{(1)} - zI\right)^{-1} u_1 \;\middle|\; Y_N^{(1)}\right] \\
&= -z - \frac{1}{N}\operatorname{Tr}\left[\left(\frac{1}{\sqrt{N}}Y_N^{(1)} - zI\right)^{-1}\right] \\
&= -z - \frac{1}{N}\frac{\sqrt{N}}{\sqrt{N-1}}\operatorname{Tr}\left[\left(\frac{1}{\sqrt{N-1}}Y_N^{(1)} - \frac{\sqrt{N}}{\sqrt{N-1}}zI\right)^{-1}\right] \\
&= -z - \frac{\sqrt{N-1}}{\sqrt{N}}S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right).
\end{aligned}
\tag{4.6}
$$

and so

$$
\mathbb{E}[V_k] = \mathbb{E}[V_1] = -z - \frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right).
$$

It follows that

$$
\begin{aligned}
S_N(z) &= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{V_k} = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{V_k} - \frac{1}{\mathbb{E}[V_k]}\right) + \frac{1}{\mathbb{E}[V_1]} \\
&= \frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{V_k} - \frac{1}{\mathbb{E}[V_k]}\right) - \frac{1}{z + \frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)}.
\end{aligned}
$$

Our eventual goal is to conclude from this that

$$
S_N(z) \approx -\frac{1}{z + S_N(z)}.
$$

We thus want to bound the first term above. Note that since the Stieltjes transform preserves the sign of the imaginary part, from equation (4.6), $\left|\Im\,\mathbb{E}[V_k \mid Y_N^{(k)}]\right| \geq |\Im z|$ a.s. Then

$$
\begin{aligned}
\mathbb{E}\left[\left(S_N(z) - \frac{1}{\mathbb{E}[V_1]}\right)^2\right] &\leq \mathbb{E}\left[\left(\frac{1}{N}\sum_{k=1}^{N}\left|\frac{1}{V_k} - \frac{1}{\mathbb{E}[V_k]}\right|\right)^2\right] \\
&\leq \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[\left|\frac{1}{V_k} - \frac{1}{\mathbb{E}[V_k]}\right|^2\right] \\
&= \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[\mathbb{E}\left[\frac{(V_k - \mathbb{E}[V_k])^2}{V_k^2\,\mathbb{E}[V_k]^2} \;\middle|\; Y_N^{(k)}\right]\right] \\
&\leq \frac{1}{(\Im z)^4}\,\mathbb{E}[(V_1 - \mathbb{E}[V_1])^2] = \frac{1}{(\Im z)^4}\operatorname{Var}[V_1].
\end{aligned}
\tag{4.7}
$$

Exercise 4.8. For any number c and a random variable x,

$$\mathrm{Var}[x] + (\mathbb{E}[x] - c)^2 = \mathbb{E}[(x-c)^2]$$

and so each term on the left-hand side is $\leq$ the right-hand side.

It follows that

$$\mathrm{Var}[S_N(z)] \leq \frac{1}{(\Im z)^4} \mathrm{Var}[V_1] \tag{4.8}$$

and

$$\left| \overline{S}_N(z) - \frac{1}{\mathbb{E}[V_1]} \right|^2 \leq \frac{1}{(\Im z)^4} \mathrm{Var}[V_1]. \tag{4.9}$$

Write

$$\mathrm{Var}[V_1] = \mathbb{E}\left[\mathrm{Var}[V_1 | Y_N^{(1)}]\right] + \mathrm{Var}\left[\mathbb{E}[V_1 | Y_N^{(1)}]\right],$$

where

$$\mathbb{E}\left[\mathrm{Var}[V_1 | Y_N^{(1)}]\right] = \mathbb{E}\left[\mathbb{E}[V_1^2 | Y_N^{(1)}]\right] - \mathbb{E}\left[\mathbb{E}[V_1 | Y_N^{(1)}]^2\right]$$

and

$$\mathrm{Var}\left[\mathbb{E}[V_1 | Y_N^{(1)}]\right] = \mathbb{E}\left[\mathbb{E}[V_1 | Y_N^{(1)}]^2\right] - \left(\mathbb{E}\left[\mathbb{E}[V_1 | Y_N^{(1)}]\right]\right)^2$$

On the one hand, from (4.6)

$$\mathrm{Var}\left[\mathbb{E}[V_1 | Y_N^{(1)}]\right] = \frac{N-1}{N} \mathrm{Var}\left[S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}} z\right)\right] \leq \mathrm{Var}\left[S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}} z\right)\right]. \tag{4.10}$$

On the other hand,

$$\begin{aligned}
\mathrm{Var}[V_1 | Y_N^{(1)}] &= \mathrm{Var}\left[\frac{1}{\sqrt{N}}(Y_N)_{11} - z - \frac{1}{N} u_1^T \left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)^{-1} u_1 \,\middle|\, Y_N^{(1)}\right] \\
&= \frac{1}{N} \mathrm{Var}[(Y_N)_{11}] + \frac{1}{N^2} \mathrm{Var}\left[u_1^T \left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)^{-1} u_1 \,\middle|\, Y_N^{(1)}\right] \\
&\leq \frac{1}{N} m_2 + \frac{1}{N^2}(2 + m_4) \,\mathrm{Tr}\left[\left(\frac{1}{\sqrt{N}} Y_N^{(1)} - \overline{z} I\right)^{-1} \left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)^{-1}\right] \\
&= \frac{1}{N} m_2 + \frac{1}{N^2}(2 + m_4) \sum_{j=1}^{N-1} \left| \lambda_j \left(\left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)^{-1}\right)\right|^2 \\
&\leq \frac{1}{N} m_2 + \frac{1}{N}(2 + m_4) \frac{1}{(\Im z)^2}
\end{aligned}$$

since

$$\left| \lambda_j \left(\left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)^{-1}\right)\right| = \frac{1}{\left|\lambda_{N-j+1}\left(\frac{1}{\sqrt{N}} Y_N^{(1)} - zI\right)\right|} \leq \frac{1}{|\Im z|}.$$

Taking expectations preserves this estimate. Combining with equation (4.10), we get

$$\mathrm{Var}[V_1] \le \mathrm{Var}\left[S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)\right] + \frac{1}{N}m_2 + \frac{1}{N}(2+m_4)\frac{1}{(\Im z)^2}.$$

Thus using equation (4.8), we obtain the estimate

$$\mathrm{Var}[S_N(z)] \le \frac{1}{(\Im z)^4}\mathrm{Var}\left[S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)\right] + \frac{1}{N}m_2\frac{1}{(\Im z)^4} + \frac{1}{N}(2+m_4)\frac{1}{(\Im z)^6}$$

Let

$$C_N = \sup\left\{\mathrm{Var}[S_N(z)] : \Im z \ge 2\right\}.$$

Then denoting $b = m_2\frac{1}{2^4} + (2+m_4)\frac{1}{2^6}$,

$$\begin{aligned} C_N &\le \frac{1}{2^4}\mathrm{Var}\left[S_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)\right] + \frac{1}{N}m_2\frac{1}{2^4} + \frac{1}{N}(2+m_4)\frac{1}{2^6} \\ &\le \frac{1}{16}C_{N-1} + \frac{1}{N}b \end{aligned}$$

since $\Im\frac{\sqrt{N}}{\sqrt{N-1}}z \ge \Im z$ for $z \in \mathbb{C}^+$. Recursively,

$$C_N \le \frac{1}{N}b\sum_{j=0}^{N-2}\frac{1}{16^j} + \frac{1}{16^{N-1}}C_1 \le \frac{2b}{N} + \frac{1}{16^{N-1}}C_1.$$

We conclude that for large N, $C_N \le C'/N$,

$$\sup_{\Im z \ge 2}\mathrm{Var}[S_N(z)] \le \frac{C'}{N}. \tag{4.11}$$

Also, $\mathrm{Var}[V_1] \le \frac{1}{16}C_{N-1} + \frac{1}{N}b$, so from equation (4.9),

$$\sup_{\Im z\ge 2}\left|\overline{S}_N(z) + \frac{1}{z + \frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)}\right|^2 = \sup_{\Im z\ge 2}\left|\overline{S}_N(z) - \frac{1}{\mathbb{E}[V_1]}\right|^2 \le \frac{C''}{N}.$$

Combining with equation (4.4), we obtain

$$\sup_{\Im z\ge 2}\left|\frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right) - \frac{1}{N}\frac{1}{z} + \frac{1}{z + \frac{\sqrt{N-1}}{\sqrt{N}}\overline{S}_{N-1}\left(\frac{\sqrt{N}}{\sqrt{N-1}}z\right)}\right| \le \frac{C'''}{\sqrt{N}} \tag{4.12}$$

First we prove weak convergence of $\hat{\mu}_{X_N}$ in expectation. Since their variances are uniformly bounded, this family is tight. Given any subsequence, we may choose a further subsequence (N_k) such that $\hat{\mu}_{X_{N_k}} \to \mu$ weakly in expectation, for some probability measure μ. Then S_μ satisfies

$$S_\mu(z) + \frac{1}{z + S_\mu(z)} = 0.$$

Then $S_\mu(z)^2 + zS_\mu(z) + 1 = 0$ (compare with Exercise 2.4), and

$$S_\mu(z) = \frac{-z + \sqrt{z^2 - 4}}{2},$$

where we chose the branch of the square root so that $S_\mu(z) \sim -\frac{1}{z}$ at infinity. By Stieltjes inversion,

$$d\mu(x) = \lim_{y \downarrow 0} \frac{1}{\pi} \Im \frac{-(x+iy) + \sqrt{(x+iy)^2 - 4}}{2}\, dx = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{[-2,2]}(x)\, dx,$$

that is, $\mu = \sigma$. Since this is true for any initial subsequence, we conclude that $\hat{\mu}_N \to \sigma$ weakly in expectation.

Finally, we prove weak convergence almost surely. Given any increasing subsequence of positive integers, choose a further subsequence (N_k) so that $\sum_{k=1}^{\infty} \frac{1}{N_k} < \infty$. Then using equation (4.11) and the Borel-Cantelli lemma, for any fixed $z \in \mathbb{C}^+ + 2i$, $S_{N_k}(z) - \overline{S}_{N_k} \to 0$ a.s. Since $\overline{S}_{N_k}(z) \to S_\sigma(z)$, it follows $S_{N_k}(z) \to S_\sigma(z)$ a.s. By a diagonal argument, we may assume this to hold for all z in a countable set $A \subset \mathbb{C}^+ + 2i$ which has an accumulation point. Therefore $\hat{\mu}_{N_k} \to \sigma$ weakly a.s. Since this is true for any initial subsequence, we conclude that $\hat{\mu}_N \to \sigma$ weakly almost surely.

Chapter 5

Joint eigenvalue distributions for orthogonally invariant ensembles.

The exact joint distribution of eigenvalues for the orthogonally/unitarily/simplectically invariant ensembles takes some work to compute, primarily because there is no natural bijective parametrization of such matrices by eigenvalues and eigenvectors. See various sources on the web page for (different) ways to do this. In the following section we will use this result without proof, and in the later section will derive it (in the Gaussian case), by first reducing the matrix to a tridiagonal form, where such a bijective parametrization is in fact available.

5.1 Mean field approximation.

The following is another, heuristic, derivation of the convergence to semicircle law for GOE. It already appears in the work of Brézin, Itzykson, Parisi, and Zuber (1978). The joint eigenvalue density for a normalized orthogonally invariant ensemble with potential V is

$$\begin{aligned}
\rho(\lambda_1, \ldots, \lambda_N) &= \frac{1}{Z_N} \prod_{i<j} |\lambda_j - \lambda_i|^\beta \exp\left(-N \sum_{i=1}^N V(\lambda_i)\right) \\
&= \frac{1}{Z_N} \exp\left[\beta \sum_{i<j} \log|\lambda_j - \lambda_i| - N \sum_{i=1}^N V(\lambda_i)\right] \\
&= \frac{1}{Z_N} \exp\left[\frac{\beta}{2} N^2 \iint \log|x-y| \; d\hat{\mu}_N(x)\, d\hat{\mu}_N(y) - N^2 \int V(x)\, d\hat{\mu}_N(x)\right] \\
&= \frac{1}{Z_N} \exp\left[-N^2 I_V(\hat{\mu})\right],
\end{aligned}$$

where

$$I_V(\mu) = \int V(x)\, d\mu(x) - \frac{\beta}{2} \iint \log|x-y|\; d\mu(x)\, d\mu(y).$$

For large N, we expect $\hat{\mu}_N$ to concentrate around the measure μ which minimises I_V. Looking at the perturbations $\mu_\varepsilon = \mu + \varepsilon\nu$ for ν a signed measure of integral zero, if μ is an extremum of I_V then

$$V(x) - \beta \int \log|x-y|\; d\mu(y) = C, \quad x \in \operatorname{supp}(\mu).$$

So (at least formally)

$$\frac{1}{\beta} V'(x) = p.v. \int \frac{1}{x-y}\, d\mu(y) = \pi H_\mu(x), \quad x \in \operatorname{supp}(\mu)$$

the Hilbert transform of μ. For example, for $V(x) = \frac{\beta}{4}x^2$, $\pi H_\mu(x) = x/2$, and $\mu = \sigma$ on $[-2,2]$. Indeed, recall that for $d\mu(x) = \rho(x)\, dx$,

$$S_\mu(x+0i) = -p.v. \int \frac{1}{x-t}\, d\mu(t) + \pi i \rho(x).$$

Thus

$$0 = \left(\Re S_\mu(x+0i) + \frac{x}{2}\right) \Im S_\mu(x+0i) = \frac{1}{2} \Im \left(G_\mu(x+0i)^2 + (x+i0) S_\mu(x+0i)\right).$$

Therefore the function $S_\mu(z)^2 + zS_\mu(z) + 1$ analytically extends to $\mathbb{R}$ and so, since $S_\mu(\overline{z}) = \overline{S_\mu(z)}$, to all of $\mathbb{C}$, i.e. it is entire. Moreover, assuming μ is compactly supported, $S_\mu(z) \sim -\frac{1}{z}$ as $z \to \infty$, and so this function is bounded. So by Liouville's theorem it is constant, and by the asymptotics above the constant is zero. Thus finally, $S_\mu(z)^2 + zS_\mu(z) + 1 = 0$, which we know characterizes σ.

5.2 Beta ensembles.

The initial ideas in this and the next section are due to Hale Trotter (1984) who tridiagonalized the GOE, and researchers who studied the β-eigenvalue distributions before the β-ensembles were defined. They were combined and developed in much greater depth by Ioana Dumitriu in her thesis (2002).

Tridiagonalization of Gaussian ensembles for $\beta = 1, 2, 4$.

Lemma 5.1 (Householder transformation)**.** *Let Y_N be an complex Hermitian (or in particular real symmetric) $N \times N$ matrix, and write it as*

$$Y_N = \begin{pmatrix} y & v^* \\ v & Y_{N-1} \end{pmatrix},$$

where $y \in \mathbb{R}$ and Y_{N-1} is $(N-1)\times(N-1)$. There is a unitary (or in particular orthogonal) transformation $\tilde{U}_{N-1}$ such that

$$\tilde{U}_{N-1}v = \begin{pmatrix} \|v\| \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Therefore denoting

$$U_N = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{U}_{N-1} \end{pmatrix},$$

we have

$$U_N Y_N U_N^T = \begin{pmatrix} y & \|v\| & 0 & \dots & 0 \\ \|v\| & & & & \\ 0 & & & & \\ \vdots & & \tilde{U}_{N-1} Y_{N-1} \tilde{U}_{N-1}^* & & \\ 0 & & & & \end{pmatrix}.$$

Proof. In fact one can choose $\tilde{U}_{N-1}$ a reflection in the hyperplane orthogonal to the vector $w = v - \|v\|\, e_1$,

$$\tilde{U}_{N-1}x = x - 2\frac{\langle x, w\rangle}{\|w\|^2}w,$$

which is automatically unitary and Hermitian (or orthogonal in the real case). Indeed,

$$\tilde{U}_{N-1}v = \tilde{U}_{N-1}\left(\frac{1}{2}w + \frac{1}{2}(v + \|v\|\, e_1)\right) = -\frac{1}{2}w + \frac{1}{2}(v + \|v\|\, e_1) - 0 = \|v\|\, e_1. \qquad \square$$

A random variable Z has the χ_k *distribution* if Z^2 has χ_k^2 distribution, that is, the same distribution as $X_1^2 + \ldots + X_k^2$ for $X_1, \ldots, X_k$ independent standard normals.

Theorem 5.2. *Let Y_N be an un-normalized GOE matrix. Then the eigenvalue distribution of Y_N is the same as for the random tridiagonal matrix*

$$\tilde{Y}_N = \begin{pmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & \ddots & b_{N-1} \\ & & & b_{N-1} & a_N \end{pmatrix},$$

whose entries are, except for the symmetry, independent with distributions

$$\begin{pmatrix} \mathcal{N}(0,2) & \chi_{N-1} & & & \\ \chi_{N-1} & \mathcal{N}(0,2) & \chi_{N-2} & & \\ & \chi_{N-2} & \ddots & \ddots & \\ & & \ddots & \ddots & \chi_1 \\ & & & \chi_1 & \mathcal{N}(0,2) \end{pmatrix}.$$

Proof. Start with an un-normalized GOE_N matrix

$$Y_N = \begin{pmatrix} Y_{11} & v_{N-1}^T \\ v_{N-1} & Y_{N-1} \end{pmatrix},$$

Recall that its entries are independent (except for symmetry), $Y_{ii} \sim \mathcal{N}(0,2)$ and $Y_{ij} \sim \mathcal{N}(0,1)$. Using the transformation from the lemma, we may choose an orthogonal $\tilde{U}_{N-1}$ so that

$$U_N Y_N U_N^T = \begin{pmatrix} Y_{11} & \|u_{N-1}\| & 0 & \dots & 0 \\ \|u_{N-1}\| & & & & \\ 0 & & & & \\ \vdots & & \tilde{U}_{N-1} Y_{N-1} \tilde{U}_{N-1}^T & & \\ 0 & & & & \end{pmatrix}.$$

Here

$$\|U_{N-1}\| = \sqrt{Y_{12}^2 + \ldots + Y_{1N}^2},$$

so it is independent of all the other entries of the matrix (except for symmetry) and has χ_{N-1} distribution. Moreover $\tilde{U}_{N-1} Y_{N-1} \tilde{U}_{N-1}^T$ is a GOE_{N-1} matrix. Applying the same procedure recursively, we end up with a tridiagonal matrix with the claimed entry distributions which is unitarily equivalent to Y_N. □

Exercise 5.3. Show that a similar procedure works for the GUE matrices, except the distributions of entries become

$$\frac{1}{\sqrt{2}} \begin{pmatrix} \mathcal{N}(0,2) & \chi_{2(N-1)} & & & \\ \chi_{2(N-1)} & \mathcal{N}(0,2) & \chi_{2(N-2)} & & \\ & \chi_{2(N-2)} & \ddots & \ddots & \\ & & \ddots & \ddots & \chi_2 \\ & & & \chi_2 & \mathcal{N}(0,2) \end{pmatrix}.$$

Since the χ^2 distribution is infinitely divisible, χ_β can actually be defined for any real positive β, with the density

$$\frac{2^{1-\beta/2}}{\Gamma(\beta/2)} x^{\beta-1} e^{-x^2/2}.$$

Definition 5.4. For $\beta > 0$, the (un-normalized) β-ensemble consists of random symmetric tridiagonal matrices

$$\tilde{Y}_N = \begin{pmatrix} a_1 & b_1 & & & \\ b_1 & a_2 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & \ddots & b_{N-1} \\ & & & b_{N-1} & a_N, \end{pmatrix}$$

whose entries are, except for the symmetry, independent with distributions

$$\frac{1}{\sqrt{\beta}} \begin{pmatrix} \mathcal{N}(0,2) & \chi_{(N-1)\beta} & & & \\ \chi_{(N-1)\beta} & \mathcal{N}(0,2) & \chi_{(N-2)\beta} & & \\ & \chi_{(N-2)\beta} & \ddots & \ddots & \\ & & \ddots & \ddots & \chi_\beta \\ & & & \chi_\beta & \mathcal{N}(0,2) \end{pmatrix}.$$

Trotter's proof of convergence to the semicircle law.

Theorem 5.5. *The normalized β-ensemble matrices have, for any β, the same asymptotic spectral distribution as the sequence of deterministic matrices*

$$T_N = \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & \sqrt{N-1} & & & \\ \sqrt{N-1} & 0 & \sqrt{N-2} & & \\ & \sqrt{N-2} & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & 0 \end{pmatrix}.$$

In particular this is the case for normalized GOE/GUE/GSE. The asymptotic spectral distribution of T_N will be shown to be semicircular in the following proposition.

Proof. We note first that since $\chi_\beta \geq 0$ and $\mathbb{E}[\chi_\beta^2] = \beta$,

$$\begin{aligned} \mathbb{E}[\beta(\chi_\beta - \sqrt{\beta})^2] &\leq \mathbb{E}[(\chi_\beta + \sqrt{\beta})^2(\chi_\beta - \sqrt{\beta})^2] \\ &= \mathbb{E}[(\chi_\beta^2 - \beta)^2] = \mathrm{Var}[\chi_\beta^2] = \beta\,\mathrm{Var}[\chi_1^2] = \beta\,\mathbb{E}[\chi_1^4 - 2\chi_1^2 + 1] = 2\beta. \end{aligned}$$

Thus $\mathbb{E}[(\chi_\beta - \sqrt{\beta})^2] \leq 2$. So in an un-normalized β-matrix $\tilde{Y}_N$, $\mathbb{E}[a_k^2] = 2$ and $\mathbb{E}[(b_k - \sqrt{(N-k)\beta})^2] \leq 2$.

Then for $\tilde{X}_N = \frac{1}{\sqrt{N}}\tilde{Y}_N$,

$$\begin{aligned}\left|\int f\, d\hat{\mu}_{\tilde{X}_N} - \int f\, d\sigma\right| &\leq \|f\|_{\mathrm{Lip}} \frac{1}{\sqrt{N}} \left\|\tilde{X}_N - T_N\right\|_F \\ &= \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{\beta}} \frac{1}{N} \sqrt{\sum_{k=1}^{N} a_k^2 + 2\sum_{k=1}^{N-1} (b_k - \sqrt{(N-k)\beta})^2} \\ &= \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{\beta}} \frac{1}{\sqrt{N}} \sqrt{\frac{\sum_{k=1}^{N} a_k^2}{N} + 2\frac{\sum_{k=1}^{N-1}(b_k - \sqrt{(N-k)\beta})^2}{N}},\end{aligned}$$

and so using Markov's inequality,

$$P\left(\left|\int f\, d\hat{\mu}_{\tilde{X}_N} - \int f\, d\sigma\right| \geq \delta\right) \leq \frac{1}{\delta}\frac{\|f\|_{\mathrm{Lip}}}{\sqrt{\beta}}\frac{1}{\sqrt{N}}\sqrt{6} \to 0$$

as $N \to \infty$. □

To compute the asymptotic spectral distribution of T_N, we use the following theorem.

Theorem 5.6 (Kac, Murdock, Szegő 1953, Trotter 1984, particular case)**.** *Let H be the Hilbert space of sequences $h = \{h_j : j \in \mathbb{Z}, h_j \in\subset L^2([0,1])\}$, with the norm $\|h\|^2 = \sum_j \|h_j\|^2 < \infty$. Denote*

$$\sigma(h) = \sum_{j\in\mathbb{Z}} h_j(x) e^{2\pi i j t},$$

a function in $L^2([0,1]^2)$. Note that the map σ is an isometry.

For any square $N \times N$ matrix A, define $\eta(A) \in H$ as follows. Consider A as included in an infinite matrix. h_j is a step function, with steps of length $\frac{1}{N}$, and heights given by values of A in the j'th diagonal (where the main diagonal corresponds to $j = 0$).

If each A_N is normal and $\eta(A_N) \to h$ in H, then the spectral distribution of A_N converges weakly to the distribution of $\sigma(h)$.

Proposition 5.7. *The asymptotic spectral distribution of T_N is the semicircular distribution.*

Proof. T_N has zero diagonal entries, and

$$T_{k,k+1} = T_{k+1,k} = \sqrt{1 - \frac{k}{N}}.$$

Thus clearly $h_1(x) = h_{-1}(x) = \sqrt{1-x}$, and we have

$$\sigma(x,t) = \sqrt{1-x}\,2\cos(2\pi t), \quad (x,t) \in [0,1]^2$$

It remains to compute its distribution. It is clearly symmetric. For $a \geq 2$, $|\{\sigma(x,t) \leq a\}| = 1$. Finally, for $0 \leq a < 2$, let $\bar{t} \in [0, 1/4]$ satisfy $\cos(2\pi t) = a/2$. Then

$$\begin{aligned}
&\left|\left\{(x,t) \in [0,1]^2 : \sqrt{1-x}2\cos(2\pi t) \leq a\right\}\right| \\
&\quad = \frac{1}{2} + 2\left|\left\{(x,t) \in [0,1] \times [0,1/4] : 1 - \frac{1}{4}a^2 \sec^2(2\pi t) \leq x \leq 1\right\}\right| \\
&\quad = \frac{1}{2} + 2\int_0^{1/4} \left(\frac{1}{4}a^2 \sec^2(2\pi t)\mathbf{1}_{[0,\bar{t}]} + \mathbf{1}_{[\bar{t},1/4]}\right) dt \\
&\quad = \frac{1}{2} + \frac{1}{4\pi}a^2 \tan(2\pi t)|_0^{\bar{t}} + \frac{1}{2} - 2\bar{t} \\
&\quad = 1 + \frac{1}{4\pi}a^2 \frac{\sqrt{1-a^2/4}}{a/2} - \frac{1}{\pi}\arccos(a/2) \\
&\quad = 1 + \frac{1}{4\pi}\left(a\sqrt{4-a^2} - 4\arccos(a/2)\right).
\end{aligned}$$

Differentiating with respect to a, we get

$$\frac{1}{4\pi}\left(\sqrt{4-a^2} - \frac{a^2}{\sqrt{4-a^2}} + 2\frac{1}{\sqrt{1-a^2/4}}\right) = \frac{1}{2\pi}\sqrt{4-a^2}.$$

□

Remark 5.8. We have proved Wigner's theorem by four different methods (under various assumptions), which ultimately reduce to four different characterizations of the semicircle law:

- Its moments are the Catalan numbers.
- Its Stieltjes transform satisfies the quadratic equation $S_\mu(z)^2 + zS_\mu(z) + 1 = 0$.
- It is the minimizer of the logarithmic energy $I_V(\mu) = \int x^2 \, d\mu(x) - 2\iint \log|x-y| \, d\mu(x)\, d\mu(y)$.
- It is the distribution of $\sigma(x,t) = \sqrt{1-x}2\cos(2\pi t)$, $(x,t) \in [0,1]^2$.

Remark 5.9. The spectral distribution of T_N is the uniform distribution on its eigenvalues, in other words on the roots of its characteristic polynomial $Q_N(\lambda) = \det(\lambda I_N - \sqrt{N}T_N)$. By Lemma 5.11, the polynomials $Q_N(\lambda)$ satisfy the recursion

$$Q_{N+1}(\lambda) + NQ_{N-1}(\lambda) = \lambda Q_{N-1}(\lambda),$$

so they are the (monic) Hermite polynomials. So we have also proved that the density of the re-scaled roots of the Hermite polynomials converges to the semicircle law.

This is the end of the 2017 course notes.

Exercise 5.10. (unverified) Let Y_N be a $K \times N$ matrix with independent $\mathcal{N}(0,1)$ entries, and assume that both K and N go to infinity in such a way that $K/N \to \alpha \in (0,1]$. Let $X_N = \frac{1}{\sqrt{N}} Y_N$ and $Z_N = X_N X_N^T$. Then Z_N is a $K \times K$ Wishart matrix (compare with Exercise 2.11).

a. Show that we may choose orthogonal matrices $V \in O(K)$, $U \in O(N)$ such that

$$VY_NU^* \sim \begin{pmatrix} \chi_K & \chi_{N-1} & & & & \\ & \chi_{K-1} & \chi_{N-2} & & & \\ & & \ddots & \ddots & & \\ & & & \chi_1 & \chi_{N-K} & \cdots \end{pmatrix}.$$

and so

$$VY_NY_N^TV^T \sim \begin{pmatrix} \chi_K^2 + \chi_{N-1}^2 & \chi_{N-1}\chi_{K-1} & & & \\ \chi_{N-1}\chi_{K-1} & \chi_{K-1}^2 + \chi_{N-2}^2 & \chi_{N-2}\chi_{K-2} & & \\ & \chi_{N-2}\chi_{K-2} & \ddots & \ddots & \\ & & \ddots & \ddots & \chi_{N-K+1}\chi_1 \\ & & & \chi_{N-K+1}\chi_1 & \chi_1^2 + \chi_{N-K}^2 \end{pmatrix}.$$

b. Show that the asymptotic spectral distribution of Z_N is the same as of

$$\frac{1}{N}\begin{pmatrix} K+N-1 & \sqrt{(N-1)(K-1)} & & & \\ \sqrt{(N-1)(K-1)} & K+N-3 & \sqrt{(N-2)(K-2)} & & \\ & \chi_{N-2} & \ddots & \ddots & \\ & & \ddots & \ddots & \sqrt{N-K+1} \\ & & & \sqrt{N-K+1} & 1+N-K \end{pmatrix}$$

c. Show that this distribution is the same as the distribution of

$$\sigma(x,t) = 1 + \alpha - 2\alpha x + 2\cos(2\pi t)\sqrt{(1-\alpha x)(\alpha - \alpha x)}.$$

5.3 Spectral theory of finite Jacobi matrices.

An $N \times N$ *Jacobi matrix* is a tridiagonal matrix of the form

$$J_N = \begin{pmatrix} a_1 & b_1 & & & 0 \\ b_1 & a_2 & & & \\ & \ddots & \ddots & \ddots & \\ & & & & b_{N-1} \\ 0 & & & b_{N-1} & a_N \end{pmatrix},$$

where $a_1, \ldots a_N \in \mathbb{R}$ and $b_1, \ldots, b_{N-1} > 0$.

Spectral bijection.

Lemma 5.11. *Let $Q_n(\lambda)$ be the characteristic polynomial of J_n, $Q_n(\lambda) = \det(\lambda I_n - J_n)$. Then for all $n \geq 2$*

$$Q_n(\lambda) + a_n Q_{n-1}(\lambda) + b_{n-1}^2 Q_{n-2}(\lambda) = \lambda Q_{n-1}(\lambda) \tag{5.1}$$

while setting $Q_0 = 1$, also $Q_1(\lambda) + a_1 Q_0 = \lambda Q_0$.

Proof. Expand the determinant with respect to the last row, and then the last column, to obtain

$$Q_n(\lambda) = (\lambda - a_n) Q_{n-1}(\lambda) - b_{n-1}^2 Q_{n-2}(\lambda).$$ □

Exercise 5.12. Let X be a general real symmetric (or complex Hermitian) $N \times N$ matrix, with eigenvalues $\lambda_1(X) \leq \ldots \leq \lambda_N(X)$.

a. The eigenvalues of X have the following *minimax* description:

$$\lambda_N(X) = \max_{u \neq 0} \frac{\langle Xu, u \rangle}{\|u\|^2}$$

and for $n < N$,

$$\lambda_n(X) = \min_{V : \dim V = n} \max_{u \in V, u \neq 0} \frac{\langle Xu, u \rangle}{\|u\|^2}$$

Hint: diagonalize the matrix, and recall that its eigenvectors are orthogonal.

b. Let $\tilde{X}$ be X with the last row and column removed. Then the eigenvalues of X and $\tilde{X}$ *interlace*:

$$\lambda_1(X) \leq \lambda_1(\tilde{X}) \leq \lambda_2(X) \leq \lambda_2(\tilde{X}) \leq \ldots \leq \lambda_{N-1}(X) \leq \lambda_{N-1}(\tilde{X}) \leq \lambda_N(X).$$

Corollary 5.13. *A Jacobi matrix with positive b_i's has distinct eigenvalues.*

Proof. We need to show that Q_N has distinct roots. Suppose $\lambda_i(J_N) = \lambda_{i+1}(J_N)$. Then from the interlacing property, also $Q_{N-1}(\lambda_i(J_N)) = 0$. From the recursion (5.1) it follows that $Q_{N-2}(\lambda_i(J_N)) = 0$. Applying the recursion repeatedly, by induction we get that $Q_0(\lambda_i(J_N)) = 0$, and so obtain a contradiction. □

Since the matrix J_N is symmetric, it can be diagonalized, so that $J = U \Lambda U^T$, in other words

$$J_N U = U \Lambda.$$

Here Λ is the diagonal matrix whose diagonal entries are the eigenvalues $\lambda_1 \leq \ldots \leq \lambda_N$, and U is an orthogonal matrix whose columns $\mathbf{u}_1, \ldots \mathbf{u}_N$ are the normalized eigenvectors. Denote

$$p_i = |U_{1i}|^2,$$

so that $p_1 + \ldots + p_N = 1$.

Proposition 5.14. *The map*

$$\varphi : \mathbb{R}^N \times \mathbb{R}_+^{N-1} \to \mathbb{R}^N \times \left\{ (p_1, \dots, p_{N-1}) : \textit{ all } p_i > 0, \sum_{i=1}^{N-1} p_i < 1 \right\}$$

given by

$$\varphi : (a_1, \dots a_N, b_1, \dots b_{N-1}) \mapsto (\lambda_1, \dots, \lambda_N, p_1, \dots, p_{N-1})$$

is a bijection.

For $(p_1, \dots, p_{N-1})$ as in the proposition, denote $p_N = 1 - \sum_{i=1}^{N-1} p_i$, and

$$\nu_N = \sum_{i=1}^{N} p_i \delta_{\lambda_i}.$$

Then ν_N is a probability measure, such that $\int f \, d\nu_N = \sum_{i=1}^N p_i f(\lambda_i)$.

Lemma 5.15. *For all k,*

$$m_k(N) = (J^k)_{11} = \int x^k \, d\nu_N.$$

Proof.

$$(J^k)_{11} = (U \Lambda^k U^T)_{11} = \sum_i U_{1i} \lambda_i^k U_{1i} = \sum_i p_i \lambda_i^k = \int x^k \, d\nu_N.$$ □

Exercise 5.16. Recall that for the empirical spectral measure, we had

$$\frac{1}{N} \operatorname{Tr}[J^k] = \int x^k \, d\hat{\mu}_{J_N},$$

and the measure also had atoms at the eigenvalues of J_N, except the weights of all the atoms were equal to $\frac{1}{N}$. Show that for any unit vector ξ, the probability measure with moments $\left\langle J^k \xi, \xi \right\rangle$ is also atomic with atoms at the eigenvalues. What is its precise form? Which ξ corresponds to the empirical spectral measure?

Remark 5.17. For a sequence of random (for example, Wigner) or deterministic (for example, Jacobi) matrices X_N, one can ask whether the sequence of measures from the preceding exercise corresponding to some vectors ξ_N converges weakly as $N \to \infty$. Some natural choices for ξ_N are $\xi_N = e_1$ (the first basis vector), $\xi_N = e_N$ (the last basis vector), the trace case from the exercise, or ξ_N random uniformly distributed on the unit sphere.

Remark 5.18. Define a family of polynomials as follows: $P_0 = 1$,

$$a_1 P_0 + b_1 P_1(x) = x P_0,$$

for $1 \le n \le N-1$,

$$b_{n-1}P_{n-2}(x) + a_nP_{n-1}(x) + b_nP_n(x) = xP_{n-1}(x),$$

and

$$b_{N-1}P_{N-2}(x) + a_NP_{N-1}(x) + P_N(x) = xP_{N-1}(x).$$

Then each P_n, $0 \le n \le N$, is a polynomial of degree n. Since all b_i's are positive, each P_n has a positive leading coefficient. It is easy to see that

$$Q_n(x) = b_n \dots b_1 P_n(x)$$

for $0 \le n \le N-1$ and

$$Q_N(x) = b_{N-1} \dots b_1 P_N(x).$$

In particular $P_N(\lambda_i) = 0$ for all $1 \le i \le N$.

Moreover by definition,

$$J_N \begin{pmatrix} P_0(x) \\ P_1(x) \\ \vdots \\ P_{N-1}(x) \end{pmatrix} = x \begin{pmatrix} P_0(x) \\ P_1(x) \\ \vdots \\ P_{N-1}(x) \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \vdots \\ P_N(x) \end{pmatrix}.$$

So for each i, $\begin{pmatrix} P_0(\lambda_i) \\ P_1(\lambda_i) \\ \vdots \\ P_{N-1}(\lambda_i) \end{pmatrix}$ is an eigenvector of J_N with eigenvalue λ_i. Since all of the eigenspaces are one-dimensional, and $P_0 = 1$, while the first entry of $\mathbf{u}_i$ is U_{1i}, it follows that

$$U_{ji} = U_{1i}P_{j-1}(\lambda_i).$$

Lemma 5.19. $\{P_0, \dots, P_{N-1}\}$ *are the orthonormal polynomials with respect to the measure* ν_N *with positive leading coefficients.*

Proof.

$$\begin{aligned}\int P_{j-1}(x)P_{n-1}(x)\, d\nu_N(x) &= \sum_{i=1}^N p_i P_{j-1}(\lambda_i)P_{n-1}(\lambda_i) \\ &= \sum_{i=1}^N U_{1i}P_{j-1}(\lambda_i)U_{1i}P_{n-1}(\lambda_i) = \sum_{i=1}^N U_{ji}U_{ni} = (UU^T)_{jn} = \delta_{j=n}. \qquad \square\end{aligned}$$

Lemma 5.20. *The orthonormal polynomials with respect to any measure* ν *with positive leading coefficients satisfy a three-term recursion as above (which may not terminate, and the* b *coefficients may not be strictly positive). If* ν *is supported on at least* N *points, then* $b_1, \dots, b_{N-1} > 0$.

Proof. Let $\{P_n : n \geq 0\}$ be orthonormal polynomials with respect to a measure ν. Since they are obtained by a Gram-Schmidt procedure from the basis $\{x^n : n \geq 0\}$,

$$xP_{n-1}(x) = \sum_{i=0}^{n} \alpha_{n,i} P_i(x)$$

for some coefficients $\alpha_{n,i}$. Since

$$\langle P_i, xP_{n-1} \rangle_\nu = \int P_i(x) x P_{n-1}(x)\, d\nu(x) = \langle xP_i, P_{n-1} \rangle_\nu = 0$$

for $n-1 > i+1$, $\alpha_{n,i} = 0$ for $i < n-2$. Denote $b_n = \alpha_{n,n}$, $a_n = \alpha_{n,n-1}$ and $c_n = \alpha_{n,n-2}$. Then

$$c_n = \langle xP_{n-1}, P_{n-2} \rangle_\nu = \langle P_{n-1}, xP_{n-2} \rangle_\nu = b_{n-1}.$$

Thus finally,

$$xP_{n-1}(x) = b_{n-1}P_{n-2}(x) + a_n P_{n-1}(x) + b_n P_n(x).$$

Since the leading coefficients of P_n and P_{n-1} are positive, $b_n \geq 0$. If $b_N = 0$, then xP_{N-1}, and so all polynomials of degree N, are in the linear span of $\{P_0, \ldots, P_{N-1}\}$. It follows that the space of polynomials on the support of ν has dimension at most N, and so this support contains at most N points. □

Proof of Proposition 5.14. It suffices to prove a bijection between matrices J_N and measures ν_N. Starting with J_N, its $\nu_N = \sum_{i=1}^{N} p_i \delta_{\lambda_i}$ is uniquely determined by its eigenvalues and eigenvectors. Conversely, start with ν_N. Use Gram-Schmidt orthogonalization to construct orthonormal polynomials with positive leading coefficients $P_0, P_1, \ldots, P_{N-1}$. They satisfy a three-term recursion relation, whose coefficients are determined by

$$b_n = \langle xP_{n-1}, P_n \rangle_{\nu_N}$$

and

$$a_n = \langle xP_{n-1}, P_{n-1} \rangle_{\nu_N}.$$ □

The Jacobian of the spectral bijection.

Now that we have proved that φ is a bijection, we want to compute its Jacobian determinant. We do this in two steps, using an intermediate bijection with $(m_1(N), \ldots, m_{2N-1}(N))$.

Proposition 5.21. *The Jacobian determinant of the transformation*

$$(a_1, b_1, a_2, b_2, \ldots, b_{N-1}, a_N) \mapsto (m_1(N), \ldots, m_{2N-1}(N))$$

is

$$2^{N-1} \frac{\prod_{k=1}^{N-1} b_k^{4(N-k)}}{\prod_{k=1}^{N-1} b_k}.$$

Proof. Recall that $m_k(N) = (J^k)_{11}$. Using Motzkin paths, we can observe that

$$m_{2k-1}(N) = a_k b_{k-1}^2 \dots b_1^2 + \text{ Polynomial}(a_{k-1}, \dots, a_1, b_{k-1}, \dots, b_1)$$

and

$$m_{2k}(N) = b_k^2 b_{k-1}^2 \dots b_1^2 + \text{ Polynomial}(a_k, \dots, a_1, b_{k-1}, \dots, b_1).$$

It follow that the Jacobian matrix of the transformation

$$(a_1, b_1, a_2, b_2, \dots, b_{N-1}, a_N) \mapsto (m_1(N), \dots, m_{2N-1}(N)$$

is upper triangular, and its Jacobian determinant is

$$\prod_{k=2}^{N} (b_{k-1}^2 \dots b_1^2) \prod_{k=1}^{N-1} (2b_k b_{k-1}^2 \dots b_1^2) = 2^{N-1} \frac{\prod_{k=1}^{N-1} b_k^{4(N-k)}}{\prod_{k=1}^{N-1} b_k}.$$

□

Proposition 5.22. *The Jacobian determinant of the transformation*

$$(p_1, \dots, p_{N-1}, \lambda_1, \dots, \lambda_N) \mapsto (m_1(N), \dots, m_{2N-1}(N))$$

is, up to a sign,

$$\left(\prod_{i=1}^{N} p_i \right) \Delta(\lambda_1, \dots, \lambda_N)^4,$$

where

$$\Delta(\lambda_1, \dots, \lambda_N) = \prod_{1 \le i < j \le N} (\lambda_j - \lambda_i)$$

is the Vandermonde determinant.

Proof.

$$m_k(N) = \sum_{i=1}^{N} p_i \lambda_i^k = \sum_{i=1}^{N-1} p_i \lambda_i^k + (1 - p_1 - \dots - p_{N-1}) \lambda_N^k$$

So

$$\frac{\partial m_k}{\partial p_i} = \lambda_i^k - \lambda_N^k,$$
$$\frac{\partial m_k}{\partial \lambda_i} = k p_i \lambda_i^{k-1}, \qquad i < N,$$
$$\frac{\partial m_k}{\partial \lambda_N} = k(1 - p_1 - \dots - p_{N-1}) \lambda_N^{k-1} = k p_N \lambda_N^{k-1}.$$

and the Jacobian matrix of the transformation

$$(p_1, \dots, p_{N-1}, \lambda_1, \dots, \lambda_N) \mapsto (m_1(N), \dots, m_{2N-1}(N)$$

is

$$\begin{pmatrix} \lambda_1-\lambda_N & \dots & \lambda_{N-1}-\lambda_N & p_1 & \dots & p_{N-1} & p_N \\ \lambda_1^2-\lambda_N^2 & \dots & \lambda_{N-1}^2-\lambda_N^2 & 2p_1\lambda_1 & \dots & 2p_{N-1}\lambda_{N-1} & 2p_N\lambda_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ \lambda_1^{2N-1}-\lambda_N^{2N-1} & \dots & \lambda_{N-1}^{2N-1}-\lambda_N^{2N-1} & (2N-1)p_1\lambda_1^{2N-2} & \dots & (2N-1)p_{N-1}\lambda_{N-1}^{2N-2} & (2N-1)p_N\lambda_N^{2N-2} \end{pmatrix}.$$

Factoring out $p_1 \dots p_N$, we get the matrix

$$\begin{pmatrix} \lambda_1-\lambda_N & \dots & \lambda_{N-1}-\lambda_N & 1 & \dots & 1 & 1 \\ \lambda_1^2-\lambda_N^2 & \dots & \lambda_{N-1}^2-\lambda_N^2 & 2\lambda_1 & \dots & 2\lambda_{N-1} & 2\lambda_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ \lambda_1^{2N-1}-\lambda_N^{2N-1} & \dots & \lambda_{N-1}^{2N-1}-\lambda_N^{2N-1} & (2N-1)\lambda_1^{2N-2} & \dots & (2N-1)\lambda_{N-1}^{2N-2} & (2N-1)\lambda_N^{2N-2} \end{pmatrix}.$$

We need to compute its determinant. Up to a sign,

$$\det\begin{pmatrix} \lambda_1-\lambda_N & \dots & \lambda_{N-1}-\lambda_N & 1 & \dots & 1 & 1 \\ \lambda_1^2-\lambda_N^2 & \dots & \lambda_{N-1}^2-\lambda_N^2 & 2\lambda_1 & \dots & 2\lambda_{N-1} & 2\lambda_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots \\ \lambda_1^{2N-1}-\lambda_N^{2N-1} & \dots & \lambda_{N-1}^{2N-1}-\lambda_N^{2N-1} & (2N-1)\lambda_1^{2N-2} & \dots & (2N-1)\lambda_{N-1}^{2N-2} & (2N-1)\lambda_N^{2N-2} \end{pmatrix}$$

$$= \det\begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ \lambda_1-\lambda_N & \dots & \lambda_{N-1}-\lambda_N & \lambda_N & 1 & \dots & 1 & 1 \\ \lambda_1^2-\lambda_N^2 & \dots & \lambda_{N-1}^2-\lambda_N^2 & \lambda_N^2 & 2\lambda_1 & \dots & 2\lambda_{N-1} & 2\lambda_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \lambda_1^{2N-1}-\lambda_N^{2N-1} & \dots & \lambda_{N-1}^{2N-1}-\lambda_N^{2N-1} & \lambda_N^{2N-1} & (2N-1)\lambda_1^{2N-2} & \dots & (2N-1)\lambda_{N-1}^{2N-2} & (2N-1)\lambda_N^{2N-2} \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & \dots & 1 & 1 & 0 & \dots & 0 & 0 \\ \lambda_1 & \dots & \lambda_{N-1} & \lambda_N & 1 & \dots & 1 & 1 \\ \lambda_1^2 & \dots & \lambda_{N-1}^2 & \lambda_N^2 & 2\lambda_1 & \dots & 2\lambda_{N-1} & 2\lambda_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \lambda_1^{2N-1} & \dots & \lambda_{N-1}^{2N-1} & \lambda_N^{2N-1} & (2N-1)\lambda_1^{2N-2} & \dots & (2N-1)\lambda_{N-1}^{2N-2} & (2N-1)\lambda_N^{2N-2} \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & \dots & 1 & 1 & 0 & \dots & 0 & 0 \\ \lambda_1 & \dots & \lambda_{N-1} & \lambda_N & 1 & \dots & 1 & 1 \\ \lambda_1^2 & \dots & \lambda_{N-1}^2 & \lambda_N^2 & 2\tau_1 & \dots & 2\tau_{N-1} & 2\tau_N \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \lambda_1^{2N-1} & \dots & \lambda_{N-1}^{2N-1} & \lambda_N^{2N-1} & (2N-1)\tau_1^{2N-2} & \dots & (2N-1)\tau_{N-1}^{2N-2} & (2N-1)\tau_N^{2N-2} \end{pmatrix}\Bigg|_{\tau_1=\lambda_1,\dots,\tau_N=\lambda_N}$$

$$= \frac{\partial^N}{\partial\tau_1\dots\partial\tau_N}\Bigg|_{\tau_1=\lambda_1,\dots,\tau_N=\lambda_N} \det\begin{pmatrix} 1 & \dots & 1 & 1 & 1 & \dots & 1 & 1 \\ \lambda_1 & \dots & \lambda_{N-1} & \lambda_N & \tau_1 & \dots & \tau_{N-1} & \tau_N \\ \lambda_1^2 & \dots & \lambda_{N-1}^2 & \lambda_N^2 & \tau_1^2 & \dots & \tau_{N-1}^2 & \tau_N^2 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \lambda_1^{2N-1} & \dots & \lambda_{N-1}^{2N-1} & \lambda_N^{2N-1} & \tau_1^{2N-1} & \dots & \tau_{N-1}^{2N-1} & \tau_N^{2N-1} \end{pmatrix}$$

The determinant is a Vandermonde determinant

$$\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i,j}(\tau_j-\lambda_i)\prod_{i<j}(\tau_j-\tau_i)=\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i>j}(\tau_j-\lambda_i)\prod_{i}(\tau_i-\lambda_i)\prod_{i<j}(\tau_j-\lambda_i)\prod_{i<j}(\tau_j-\tau_i)$$

So the expression above is

$$\frac{\partial^N}{\partial\tau_1\ldots\partial\tau_N}\bigg|_{\tau_1=\lambda_1,\ldots,\tau_N=\lambda_N}\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i>j}(\tau_j-\lambda_i)\prod_{i}(\tau_i-\lambda_i)\prod_{i<j}(\tau_j-\lambda_i)\prod_{i<j}(\tau_j-\tau_i)$$
$$=\lim_{h_1,\ldots,h_N\to 0}\frac{\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i>j}(\lambda_j+h_j-\lambda_i)\prod_i(\lambda_i+h_i-\lambda_i)\prod_{i<j}(\lambda_j+h_j-\lambda_i)\prod_{i<j}(\lambda_j+h_j-\lambda_i-h_i)-0}{h_1\ldots h_N}$$
$$=\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i>j}(\lambda_j-\lambda_i)\prod_{i<j}(\lambda_j-\lambda_i)\prod_{i<j}(\lambda_j-\lambda_i)$$
$$=\pm\Delta(\lambda_1,\ldots,\lambda_N)^4.$$

□

Corollary 5.23. *The Jacobian determinant of φ is, up to a sign,*

$$2^{N-1}\frac{\prod_{k=1}^{N-1}b_k^{4(N-k)}}{\Delta(\lambda_1,\ldots,\lambda_N)^4\prod_{i=1}^N p_i\prod_{k=1}^{N-1}b_k}$$

Proposition 5.24.

$$\prod_{k=1}^{N-1}b_k^{N-k}=\left(\prod_{i=1}^N p_i\right)^{1/2}\Delta(\lambda_1,\ldots,\lambda_N).$$

Consequently the Jacobian determinant of φ is, up a sign,

$$2^{N-1}\frac{\prod_{i=1}^N p_i}{\prod_{k=1}^{N-1}b_k}$$

Proof. Denote by e_1 the first basis vector. Let A be the matrix with columns

$$A=(e_1,Je_1,\ldots,J^{N-1}e_1).$$

Then A is upper-triangular, with entries $1,b_1,\ldots,\prod_{k=1}^{N-1}b_k$ on the diagonal. So

$$\det A=\prod_{n=0}^{N-1}\prod_{k=1}^{n}b_k=\prod_{k=1}^{N-1}b_k^{N-k}.$$

On the other hand, denote $\mathbf{q}=(q_1,\ldots,q_N)^T=U^Te_1$ the first row of U, so that $q_i^2=p_i$. Then

$$A=(UU^Te_1,U\Lambda U^Te_1,\ldots,U\Lambda^{N-1}U^Te_1)=U(U^Te_1,\Lambda U^Te_1,\ldots,\Lambda^{N-1}U^Te_1)$$
$$=U(\mathbf{q},\Lambda\mathbf{q},\ldots,\Lambda^{N-1}\mathbf{q})=U\begin{pmatrix}q_1 & \lambda_1 q_1 & \ldots & \lambda_1^{N-1}q_1\\ \vdots & \vdots & \ldots & \vdots\\ q_N & \lambda_N q_N & \ldots & \lambda_N^{N-1}q_N\end{pmatrix}$$

and so

$$\det A = \left(\prod_{i=1}^{N} q_i\right) \det \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{N-1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & \lambda_N & \dots & \lambda_N^{N-1} \end{pmatrix} = \left(\prod_{i=1}^{N} q_i\right) \Delta(\lambda_1, \dots, \lambda_N).$$

We conclude that the Jacobian determinant of φ is

$$2^{N-1} \frac{\left(\left(\prod_{i=1}^N p_i\right) \Delta(\lambda_1, \dots, \lambda_N)\right)^4}{\Delta(\lambda_1, \dots, \lambda_N)^4 \prod_{k=1}^{N-1} b_k} = 2^{N-1} \frac{\left(\prod_{i=1}^N p_i\right)^4}{\prod_{k=1}^{N-1} b_k}. \qquad \square$$

Exact eigenvalue distribution for β-ensembles.

Theorem 5.25. *The joint density of ordered eigenvalues of an un-normalized β-ensemble matrix is*

$$\frac{1}{Z_N} \exp\left(-\frac{\beta}{4} \sum_{i=1}^{N} \lambda_i^2\right) \Delta(\lambda_1, \dots, \lambda_N)^\beta.$$

Proof. Recall that for the β-ensembles, a_i and b_i are independent, with distributions

$$\sqrt{\beta} a_k \sim \mathcal{N}(0, 2)$$

and

$$\sqrt{\beta} b_k \sim \chi_{(N-k)\beta}.$$

So their individual densities are $\frac{1}{Z} e^{-(\beta/4)a_k^2}$ and $\frac{1}{Z} b_k^{(N-k)\beta-1} e^{-(\beta/2)b_k^2}$, and their joint density is

$$\frac{1}{Z} \prod_{k=1}^{N} e^{-(\beta/4)a_k^2} \prod_{k=1}^{N-1} b_k^{k\beta-1} e^{-(\beta/2)b_k^2} = \frac{1}{Z} \exp\left(-\frac{\beta}{4} \sum_{k=1}^{N} a_k^2 - \frac{\beta}{2} \sum_{k=1}^{N-1} b_k^2\right) \prod_{k=1}^{N-1} b_k^{(N-k)\beta-1}.$$

We want to express this in terms of λ_i's and p_i's. We note that all $b_k > 0$ a.s. (so the results from earlier in the section apply),

$$\sum_{k=1}^{N} a_k^2 + 2 \sum_{k=1}^{N-1} b_k^2 = \text{Tr}[J^T J] = \sum_{i=1}^{N} \lambda_i^2,$$

and

$$\prod_{k=1}^{N-1} b_k^{(N-k)\beta} = \left(\prod_{i=1}^{N} p_i\right)^{\beta/2} \Delta(\lambda_1, \dots, \lambda_N)^\beta.$$

Since

$$\left|\text{Jac}_{(a,b)\to(\lambda,p)}\right| = 2^{N-1} \frac{\prod_{i=1}^N p_i}{\prod_{k=1}^{N-1} b_k}$$

we obtain the density

$$\frac{1}{Z_N} \exp\left(-\frac{\beta}{4}\sum_{i=1}^N \lambda_i^2\right) \left(\prod_{i=1}^N p_i\right)^{\beta/2-1} \Delta(\lambda_1, \dots, \lambda_N)^\beta.$$

Thus the joint densities of λ_i's and p_j's are independent, and the joint distribution of the p_i's may be integrated out. □

To obtain the precise normalization constant in the joint density of eigenvalues we need to trace the constants carefully throughout the proof, and at the end use the *Dirichlet integral*

$$\int_0^1 \int_0^{1-p_1} \dots \int_0^{1-\sum_{i=1}^{N-2}} \left(\prod_{i=1}^N p_i\right)^{\beta/2-1} dp_{N-1} \dots dp_1 = \frac{\Gamma(\beta/2)^N}{\Gamma(n\beta/2)}.$$

Another approach to compute

$$\begin{aligned} Z_N &= \iint_{\lambda_1 \le \dots \le \lambda_N} \exp\left(-\frac{\beta}{4}\sum_{i=1}^N \lambda_i^2\right) \Delta(\lambda_1, \dots, \lambda_N)^\beta \, d\lambda_1 \dots d\lambda_N \\ &= \frac{1}{N!}\int_{\mathbb{R}^N} \exp\left(-\frac{\beta}{4}\sum_{i=1}^N \lambda_i^2\right) |\Delta(\lambda_1, \dots, \lambda_N)|^\beta \, d\lambda_1 \dots d\lambda_N \end{aligned}$$

is to deduce it as a limiting case of the *Selberg integral*

$$\frac{1}{N!}\int_0^1 \dots \int_0^1 \prod_{i=1}^N \lambda_i^{a-1}(1-\lambda_i)^{b-1} |\Delta(\lambda_1, \dots, \lambda_N)|^{2c} \, d\lambda_1 \dots d\lambda_N = \prod_{j=0}^{N-1} \frac{\Gamma(a+jc)\Gamma(b+jc)\Gamma((j+1)c)}{\Gamma(a+b+(N+j-1)c)\Gamma(c)}.$$

Chapter 6

Asymptotic distributions, asymptotic freeness, and free convolution.

Motivating question 1. Let $(A_N)_{N=1}^{\infty}$ and $(B_N)_{N=1}^{\infty}$ be (families of) self-adjoint random matrices with the distribution of (B_N) unitarily invariant. Suppose their empirical spectral distributions converge weakly almost surely:

$$\hat{\mu}_{A_N} \to \mu, \quad \hat{\mu}_{B_N} \to \nu.$$

Note that, roughly speaking, we are picking B_N "uniformly at random" from all self-adjoint matrices with distribution $\approx \nu$. Suppose each A_N, B_N have mutually independent entries.

Since A_N and B_N do not commute, one cannot speak directly of their joint distribution. What are their asymptotic joint moments? Are they determined purely by μ and ν?

Answer. Under mild assumptions, A_N and B_N are *asymptotically free*. This means that there exists an operator algebra $\mathcal{A}$ with a positive linear functional φ, and two operators $a, b \in \mathcal{A}$ which are *freely independent*, such that almost surely

$$\frac{1}{N} \operatorname{Tr}[A_N^{u(1)} B_N^{v(1)} \dots A_N^{u(k)} B_N^{v(k)}] \to \varphi[a^{u(1)} b^{v(1)} \dots a^{u(k)} b^{v(k)}]. \tag{6.1}$$

Motivating question 2. Fix a large N. Let A be an $N \times N$ (non-random) Hermitian matrix, with spectral distribution $\hat{\mu}_A \approx \mu$. We interpret A as the "true signal". Let B be an $N \times N$ normalized GUE matrix, with spectral distribution $\hat{\mu}_B \approx \sigma$, the semicircular distribution. We interpret B as "noise". What is the approximate spectral distribution of "signal plus noise" $A + B$?

Answer. Under mild assumptions, the joint distribution of A and B is approximately the same as for freely independent operators a and b above, and the spectral distribution of $A + B$ is approximately the distribution of $a + b$. This distribution is the *additive free convolution* $\mu \boxplus \sigma$, and can be computed by combinatorial and complex-analytic methods.

Reference: (Mingo, Speicher 2016) extracts from Chapters 1, 3, 4.

6.1 Freeness.

Spaces and distributions.

Definition 6.1. A *non-commutative probability space* (ncps) is a pair $(\mathcal{A}, \varphi)$. Here $\mathcal{A}$ is a complex unital algebra with an involution $*$, and φ is a *state* on $\mathcal{A}$: a complex linear functional which is unital ($\varphi[1] = 1$), self-adjoint ($\varphi[a^*] = \overline{\varphi[a]}$), and positive ($\varphi[a^*a] \geq 0$).

A state is *tracial* (or *a trace*) if for all $a, b \in \mathcal{A}$, $\varphi[ab] = \varphi[ba]$.

There are other versions of this definition, some with more, some with less conditions. In particular, we may sometimes assume that $\mathcal{A}$ consists of bounded operators on some Hilbert space, and is either norm closed (a C^*-algebra) or weakly closed (a von Neumann algebra). In this case we would put extra continuity assumptions on φ.

Example 6.2. In all cases below, the conditions on φ are easy to verify.

a. Matrices: $\mathcal{A} = M_N(\mathbb{C})$, a^* is the usual adjoint, $\varphi = \frac{1}{N}\operatorname{Tr}$ or more generally $\varphi[a] = \langle a\xi, \xi\rangle$ for a unit vector ξ.

b. Commutative probability space: $\mathcal{A} = L^\infty(\Omega, \Sigma, P)$, $a^* = \bar{a}$, $\varphi[a] = \mathbb{E}[a]$.

c. Random matrices: $\mathcal{A} = M_N(\mathbb{C}) \otimes L^\infty(\Omega, \Sigma, P) = M_N(L^\infty(\Omega, \Sigma, P))$, $\varphi = \frac{1}{N}\mathbb{E} \circ \operatorname{Tr}$. Note that here we only consider bounded random entries; see remarks below for the discussion of affiliated operators.

d. Infinite-dimensional spaces: $\mathcal{A} = \mathcal{B}(H)$, the space of all bounded operators on a Hilbert space H, and $\varphi[a] = \langle a\xi, \xi\rangle$ for a unit vector $\xi \in H$. Note that none of these states are tracial, and in fact $\mathcal{B}(H)$ has no continuous tracial states. So it is important to note that there exist *finite von Neumann algebras*: infinite-dimensional, weakly closed $*$-aubalgebras of $\mathcal{B}(H)$ which do have continuous tracial states.

Definition 6.3. An element $a \in \mathcal{A}$ is *self-adjoint* if $a^* = a$. For a self-adjoint element, the sequence of its moments

$$(m_n(a))_{n=0}^\infty = (\varphi[a^n])_{n=0}^\infty$$

is positive definite. Therefore by Bochner's theorem, there exists a probability measure μ_a on $\mathbb{R}$ (*the distribution* of a) such that

$$m_n(a) = m_n(\mu) = \int_{\mathbb{R}} x^n \, d\mu_a(x).$$

Remark 6.4. In many situations of interest to us, $\mathcal{A}$ is a von Neumann algebra, so that its elements are bounded operators on a Hilbert space. Their distributions are thus compactly supported. If we want to realize a general probability measure on $\mathbb{R}$ as a distribution, we need to consider unbounded self-adjoint operators. For such an operator a, we have a well-defined (bounded, self-adjoint) operator $f(a)$ for any bounded continuous function $f \in C_b(\mathbb{R})$. If all of these operators lie in $\mathcal{A}$, we say that a is *affiliated* to $\mathcal{A}$, and define its distribution by

$$\int f \, d\mu_a = \varphi \left[f(a) \right].$$

One can also define μ_a using the spectral measure corresponding to a, assuming that all the spectral projections of a are in $\mathcal{A}$. One can define the "affiliated" relation for non-self-adjoint operators. If $\mathcal{A}$ is a von Neumann algebra and φ is a tracial state, then operators affiliated to $\mathcal{A}$ themselves form an algebra.

Definition 6.5. Star-subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ in an ncps $(\mathcal{A}, \varphi)$ are *free* (or *freely independent*) if whenever $\varphi[a_1] = \varphi[a_2] = \ldots = \varphi[a_n] = 0$, $a_i \in \mathcal{A}_{u(i)}$, $u(1) \neq u(2) \neq u(3) \neq \ldots$ (neighbours distinct), then also

$$\varphi \left[a_1 a_2 \ldots a_n \right] = 0.$$

Elements $a_1, a_2, \ldots, a_k$ are free if the star-subalgebras they generate are free.

Note that for self-adjoint a_i, these are conditions on their joint moments: for any polynomials $P_1, \ldots, P_n$,

$$\varphi \left[\prod_{i=1}^{n} \Big(P_i(a_{u(i)}) - \varphi \left[P_i(a_{u(i)}) \right] \Big) \right] = 0.$$

For general a_i we need to consider more general polynomials $P_i(a, a^*)$.

Theorem 6.6. *Given ncps $(\mathcal{A}_i, \varphi_i)_{i=1}^k$, there exists an ncps $(\mathcal{A}, \varphi) = *_{i=1}^k (\mathcal{A}_i, \varphi_i)$, called their* reduced free product*, such that*

- *there are embeddings $J_i : \mathcal{A}_i \to \mathcal{A}$ such that $\varphi \circ J_i = \varphi_i$*
- *$(J_i(\mathcal{A}_i))_{i=1}^k$ are free with respect to φ.*

Thus we may construct free copies of given nc random variables.

Example 6.7. Let $a, b \in (\mathcal{A}, \varphi)$ be free. How to compute $\varphi \left[abab \right]$?

Write $a^\circ = a - \varphi \left[a \right]$. Note

$$\varphi \left[(a^\circ)^2 \right] = \varphi \left[a^2 - 2a\varphi \left[a \right] + \varphi \left[a \right]^2 \right] = \varphi \left[a^2 \right] - \varphi \left[a \right]^2.$$

Then

$$\varphi \left[abab \right] = \varphi \left[(a^\circ + \varphi \left[a \right])(b^\circ + \varphi \left[b \right])(a^\circ + \varphi \left[a \right])(b^\circ + \varphi \left[b \right]) \right].$$

Using freeness and linearity, this reduces to

$$\begin{aligned} \varphi \left[abab \right] &= \varphi \left[a \right] \varphi \left[b \right] \varphi \left[a \right] \varphi \left[b \right] + \varphi \left[a \right]^2 \varphi \left[(b^\circ)^2 \right] + \varphi \left[b \right]^2 \varphi \left[(a^\circ)^2 \right] \\ &= \varphi \left[a \right] \varphi \left[b \right] \varphi \left[a \right] \varphi \left[b \right] + \varphi \left[a \right]^2 (\varphi \left[b^2 \right] - \varphi \left[b \right]^2) + \varphi \left[b \right]^2 (\varphi \left[a^2 \right] - \varphi \left[a \right]^2) \\ &= \varphi \left[a^2 \right] \varphi \left[b \right]^2 + \varphi \left[a \right]^2 \varphi \left[b^2 \right] - \varphi \left[a \right]^2 \varphi \left[b \right]^2. \end{aligned}$$

Moral: not a good way to compute.

Free cumulants.

Definition 6.8. For an ncps $(\mathcal{A}, \varphi)$, the n'th *moment functional* is the $\mathbb{C}$-multi-linear functional on $\mathcal{A}$,

$$M_n^\varphi(a_1, \ldots, a_n) = \varphi[a_1 \ldots a_n].$$

Note that

$$m_n(a) = M_n(a, a, \ldots, a).$$

Definition 6.9. Let π be a (set) partition of n, that is, a decomposition of $\{1, 2, \ldots, n\}$ into disjoint non-empty subsets, called *blocks*. Denote all such partitions by $\mathcal{P}(n)$. For a family of multi-linear functionals $(F_i)_{i=1}^\infty$, denote by F_π the n-linear functional

$$F_\pi(a_1, \ldots, a_n) = \prod_{V \in \pi} F_{|V|}(a_i : i \in V).$$

For example, for the partition $\pi = \{(1,3), (2,4), 5\}$,

$$F_\pi(a_1, a_2, a_3, a_4, a_5) = F_2(a_1, a_3) F_2(a_2, a_4) F_1(a_5).$$

Definition 6.10. For an ncps $(\mathcal{A}, \varphi)$, define the n'th *cumulant* functional C_n implicitly by

$$M_n^\varphi(a_1, a_2, \ldots, a_n) = \sum_{\pi \in \mathcal{P}(n)} C_\pi^\varphi(a_1, a_2, \ldots, a_n).$$

For example,

$$M_1(a_1) = C_1(a_1)$$

is the mean,

$$M_2(a_1, a_2) = C_2(a_1, a_2) + C_1(a_1) C_1(a_2), \qquad C_2(a_1, a_2) = M_2(a_1, a_2) - M_1(a_1) M_1(a_2)$$

is the covariance. The implicit relation can be inverted using Möbius inversion. Again, the n'th cumulant of a single element is

$$c_n(a) = C_n(a, a, \ldots, a).$$

Since it depends only on the moments, we may define $c_n(\mu) = c_n(a)$ where $\mu = \mu_a$.

Definition 6.11. A partition is *non-crossing* if its blocks in a natural graphical representation do not cross. For example, $\{(1,3), (2,4)\}$ is crossing while $\{(1,4), (2,3)\}$ is non-crossing. Denote all non-crossing partitions of n by $\mathcal{NC}(n)$.

For an ncps $(\mathcal{A}, \varphi)$, define the n'th *free cumulant* functional R_n^φ implicitly by

$$M_n^\varphi(a_1, a_2, \ldots, a_n) = \sum_{\pi \in \mathcal{NC}(n)} R_\pi^\varphi(a_1, a_2, \ldots, a_n)$$

and the n'th free cumulant of a single element by $r_n(a) = R_n(a, a, \ldots, a)$. Then the first free cumulant is still the mean and the second the covariance, but free cumulants differ from cumulants starting with order $n = 4$.

Exercise 6.12. Let a be a bounded random variable with distribution μ. Show that

$$\sum_{n=1}^{\infty} \frac{(i\theta)^n}{n!} c_n(a) = \log \int_{\mathbb{R}} e^{i\theta x} \, d\mu(x),$$

the logarithm of the characteristic function of μ. More generally, for a k-tuple of (commuting) random variables $a_1, \ldots, a_k$ with the joint distribution μ on $\mathbb{R}^k$, in the series expansion of

$$\log \int_{\mathbb{R}^k} e^{i \sum \theta_j x_j} \, d\mu(x_1, \ldots, x_k),$$

the coefficient of $(i\theta_1)^{u(1)} \ldots (i\theta_k)^{u(k)}$ is

$$\frac{1}{u(1)!u(2)! \ldots u(k)!} C_n(\underbrace{a_1}_{u(1) \text{ times}}, \ldots, \underbrace{a_k}_{u(k) \text{ times}})$$

(note that unlike free cumulants, cumulants depend only on the quantity of each of their arguments and not on their order). Conclude that these random variables are independent if and only if their mixed cumulants vanish, in the sense of the theorem below.

Theorem 6.13. *In a ncps $(\mathcal{A}, \varphi)$, star-subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ are free if and only if "their mixed free cumulants vanish":*

$$R_n[a_1, a_2, \ldots, a_n] = 0$$

unless all a_i belong to the same $\mathcal{A}_j$.

Proof. (Easy direction) Suppose all mixed free cumulants vanish. Take $(a_i \in \mathcal{A}_{u(i)})_{i=1}^n$, $u(1) \neq u(2) \neq \ldots$, with $\varphi[a_i] = 0$. Then

$$\varphi[a_1 a_2 \ldots a_n] = \sum_{\pi \in \mathcal{NC}(n)} R_\pi[a_1, a_2, \ldots, a_n].$$

It remains to note that any non-crossing partition π contains a block which is an interval, that is, consists of several consecutive elements. So each term on the right-hand side contains a factor of the form $R_\ell[a_i, a_{i+1}, \ldots, a_{i+\ell-1}]$. If $\ell = 1$, this factor is zero since $R_1[a_i] = \varphi[a_i] = 0$. If $\ell > 1$, this factor is a mixed free cumulant, and so vanishes as well. If follows that the left-hand side is zero.

(Hard direction: sketch) The difficulty is that in the free cumulant formula, we neither assume that elements are centered nor that they are alternating. We first show that if $n \geq 2$ and any of $a_1, a_2, \ldots, a_n$ is a scalar, then

$$R[a_1, a_2, \ldots, a_n] = 0.$$

Decomposing each $a_i = (a_i - \varphi[a_i]) + \varphi[a_i]$ and using multi-linearity, we may then assume that each a_i is centered. Next, group a_i's into consecutive products so that different consecutive products lie in different algebras. Using the combinatorial formula for free cumulants with products as entries, we can then reduce to the case when a_i's are alternating. But for this case the result is easy.

For the full proof, see Theorem 5.24 in the Free Probability Notes. $\square$

Example 6.14. For a $\mathcal{N}(0,t)$ random variable, the characteristic function is $e^{t\theta^2/2}$. Thus according to Exercise 6.12, its cumulants are $c_2 = t$, $c_n = 0$ for $n \neq 2$.

Exercise 6.15. Recall that a family $(x_i)_{i=1}^k$ of random variables is jointly Gaussian with mean zero and covariance matrix Σ if the characteristic function of their joint distribution is $\exp(\theta \cdot \Sigma\theta)$. Compute all the joint cumulants of this family. Deduce the *Wick formula*

$$M\left[x_{u(1)}, x_{u(2)}, \ldots, x_{u(2n)}\right] = \sum_{\{(i_1,j_1),\ldots,(i_n,j_n)\} \in \mathcal{P}_2(2n)} \Sigma_{u(i_1),u(j_1)} \ldots \Sigma_{u(i_n),u(j_n)},$$

where $\mathcal{P}_2$ denotes the pair partitions (perfect matchings).

Example 6.16. Which distribution μ_t has free cumulants $r_2 = t$, $r_n = 0$ for $n \neq 2$? Note that its moments are

$$m_{2n+1}(\mu_t) = 0, \quad m_{2n} = t\left|\mathcal{NC}_2(2n)\right|,$$

where $\mathcal{NC}_2$ denotes the non-crossing *pairings*. It is not hard to construct a bijection with trees and use Lemma 2.3 to show that $|\mathcal{NC}_2(2n)|$ is the n'th Catalan number. It follows that μ_t is the semicircular distribution with variance t,

$$d\mu_t(x) = \frac{1}{2\pi t}\sqrt{4t - x^2}\,\mathbf{1}_{[-2\sqrt{t},2\sqrt{t}]}(x)\,dx.$$

Exercise 6.17. Derive the result in Example 6.7 using free cumulants.

Exercise 6.18. A *standard free semicircular system* is a family $s_1, \ldots, s_k \in (\mathcal{A}, \varphi)$ of self-adjoint free random variables each of which has the standard semicircular distribution. Show that

$$\varphi\left[s_{u(1)} s_{u(2)} \ldots s_{u(n)}\right] = \left|\left\{\pi \in \mathcal{NC}_2(n) \mid i \overset{\pi}{\sim} j \Rightarrow u(i) = u(j)\right\}\right|.$$

Exercise 6.19. Let s_1 and s_2 be two free standard semicircular variables. Denote $c = \frac{1}{\sqrt{2}}(s_1 + is_2)$. c is called a (standard) circular variable, and is the free analog of a complex Gaussian variable.

a. Prove that

$$R[c, c^*] = R[c^*, c] = 1,$$

and the rest of the joint cumulants of c and c^* (including those of order 2) are zero.

b. Let $\varepsilon = (\varepsilon(1), \ldots, \varepsilon(n))$, where each $\varepsilon(i)$ is either blank or $*$. Write down the expression (in terms of non-crossing partitions and ε) for

$$M[c^{\varepsilon(1)}, c^{\varepsilon(2)}, \ldots, c^{\varepsilon(n)}].$$

6.2 Free convolution

Definition 6.20. Let μ and ν be probability measures on $\mathbb{R}$. Let a and b be freely independent operators in (or affiliated to) an ncps, with distributions μ and ν, respectively. The distribution of $a + b$ depends only on μ and ν and not on the choice of a and b, and is $\mu \boxplus \nu$, the (additive) *free convolution* of μ and ν.

There is a related notion of multiplicative free convolution $\mu \boxtimes \nu$, arising from a product of freely independent operators. We will not pursue the corresponding theory.

This section addresses various approaches to computing $\mu \boxplus \nu$.

Cumulant approach

Corollary 6.21. *If a, b are free, then*

$$r_n[a + b] = r_n[a] + r_n[b].$$

Thus

$$r_n[\mu \boxplus \nu] = r_n[\mu] + r_n[\mu]$$

Proof. Using multi-linearity and the vanishing of mixed free cumulants,

$$\begin{aligned} r_n[a+b] &= R[a+b, a+b, \ldots, a+b] \\ &= \sum R[a \text{ or } b, \ldots, a \text{ or } b] = R[a, \ldots, a] + R[b, \ldots, b] = r_n[a] + r_n[b]. \quad \square \end{aligned}$$

Example 6.22. From Example 6.14 it follows that $\sigma_{t+s} = \sigma_t \boxplus \sigma_s$, so semicircular distributions form a free convolution semigroup.

We can thus, in principle, compute arbitrary free cumulants, and so arbitrary moments, of $\mu \boxplus \sigma_t$. But the computation may involve complicated combinatorics. For example: what is

$$\left(\frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a \right)^{\boxplus 2} ?$$

Cauchy transform approach

Definition 6.23. For a probability measure μ on $\mathbb{R}$, its *Cauchy transform* is the function

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - x} \, d\mu(x).$$

Note that $S_\mu(z) = -G_\mu(z)$ is the Stieltjes transform discussed in Chapter 4. So from the results in that chapter, we deduce that G_μ is an analytic function from $\mathbb{C}^+$ to $\mathbb{C}^-$ and

$$\lim_{y\to+\infty} iyG_\mu(iy) = 1;$$

that we may recover μ as a weak limit

$$\mu = \lim_{y\downarrow 0} \frac{1}{\pi}\Im G_\mu(x+iy)\,dx$$

(Stieltjes inversion formula); and that if μ is compactly supported, then $zG_\mu(z) \to 1$ as $z \to \infty$, and

$$\sum_{k=0}^{\infty} \frac{m_k(\mu)}{z^{k+1}} = G_\mu(z).$$

Exercise 6.24. Let μ be compactly supported. Denote

$$R_\mu(z) = \sum_{n=1}^{\infty} r_n(\mu) z^{n-1}.$$

Show that G_μ and $\frac{1}{z} + R_\mu(z)$ are inverses of each other under composition (as formal power series). Hint: show that this is equivalent to the relation

$$m_n = \sum_{k=1}^{n} r_k \sum_{i(1),\ldots,i(k)=0}^{i(1)+\ldots+i(k)=n-k} m_{i(1)} \ldots m_{i(k)}.$$

It will sometimes be more convenient to work with functions $F_\mu(z) = \frac{1}{G_\mu(z)}$ (the F-transform) and $\varphi_\mu(z) = R_\mu(1/z)$ (sometimes called the Voiculescu transform). For compactly supported measures, it follows that F and $z + \varphi(z)$ are compositional inverses as formal power series. Note also that F_μ is an analytic function from $\mathbb{C}^+$ to itself, such that $\lim_{y\to+\infty} \frac{F_\mu(iy)}{iy} = 1$.

Theorem 6.25. *Let F be an analytic self-map of $\mathbb{C}^+$. Then F has a* Nevanlinna representation*:*

$$F(z) = \alpha + \beta z + \int_{\mathbb{R}} \frac{1+uz}{u-z}\,d\tau(u),$$

where $\alpha \in \mathbb{R}$, $\beta \geq 0$, and τ is a finite positive measure on $\mathbb{R}$.

It is not hard to deduce this result from the Herglotz representation

$$\eta(z) = i\alpha + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\tau(\theta),$$

where η is a self-map of the unit disk. This in turn is the Poisson kernel formula for the disk (compare with the proof of the Stieltjes inversion formula in Chapter 4).

Exercise 6.26. Let F be an analytic self-map of $\mathbb{C}^+$.

a. Show that the limit $\lim_{y\to+\infty} \frac{F(iy)}{iy}$ exists (compare with the properties of the Stieltjes transform in the beginning of Chapter 4).

b. Suppose the limit in part (a) is 1. Show that F increases the imaginary part,

$$\Im F(z) \geq \Im z,$$

and $F(z) = z(1 + o(1))$ as $z \to \infty$ nontangentially to $\mathbb{R}$, meaning in such a way that $\Re z / \Im z$ remains bounded.

Solution for the second claim in part (b). From the Nevanlinna representation, we need to show that

$$\frac{1}{z} \int_{\mathbb{R}} \frac{1 + uz}{u - z} \, d\tau(u) \to 0$$

as $z \to \infty$ nontangentially. That is, the limit is taken over $z = x + iy$ with $\left|\frac{x}{y}\right| \leq M$. For any fixed $u \in \mathbb{R}$,

$$\frac{1}{z}\frac{1 + uz}{u - z} \to 0$$

as $z \to \infty$ (nontangentially or not). Also,

$$\left|\frac{1 + uz}{z(u - z)}\right| = \left|1 + \frac{1 + z^2}{z(u - z)}\right| \leq 1 + \frac{|1 + z^2|}{y^2} \leq 1 + \frac{1}{y^2} + (1 + M^2)$$

is uniformly bounded over all u and $z \to \infty$ nontangentially. The result follows from the dominated convergence theorem. □

It follows that for a general probability measure μ, F_μ has a compositional inverse F_μ^{-1} on a truncated Stolz angle

$$\Gamma_{\alpha,\beta} = \{z : \alpha \Im z > |\Re z|, \Im z > \beta\}.$$

So we may always define $\varphi_\mu = F_\mu^{-1}(z) - z$ on such a region. We may then attempt to define $\mu \boxplus \nu$ by

$$\varphi_{\mu \boxplus \nu} = \varphi_\mu + \varphi_\nu$$

on their common domain. For compactly supported measures, the free cumulant approach implies that this can be done. That is, for any compactly supported μ and ν, the function $z + \varphi_\mu(z) + \varphi_\nu(z)$ has a compositional inverse which can be analytically continued to a function from $\mathbb{C}^+$ to $\mathbb{C}^+$, and is an F-transform of a probability measure (which can be recovered via Stieltjes inversion). In a beautiful paper, (Bercovici, Voiculescu 1993) proved that this can be done in general, using unbounded operator techniques.

Example 6.27. Let $\mu = \frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$ be a Bernoulli distribution. Its Cauchy transform is

$$G_\mu(z) = \frac{1}{2}\frac{1}{z-a} + \frac{1}{2}\frac{1}{z+a} = \frac{z}{z^2-a^2},$$

and its F transform $F_\mu(z) = z - \frac{a^2}{z}$, so that

$$F_\mu^{-1}(w) = \frac{w + \sqrt{w^2+4a^2}}{2}$$

(the branch of the square root chosen so that $F^{-1}(w) = w + o(w)$ at infinity), and

$$\varphi_\mu(w) = \frac{\sqrt{w^2+4a^2} - w}{2}.$$

Thus

$$\varphi_{\mu^{\boxplus t}} = \frac{t\sqrt{w^2+4a^2} - tw}{2}$$

and

$$F^{-1}_{\mu\boxplus\sigma_t} = \frac{t\sqrt{w^2+4a^2} - (t-2)w}{2}.$$

It follows that

$$F_{\mu^{\boxplus t}}(z) = \frac{(t-2)z + t\sqrt{z^2 - 4a^2(t-1)}}{2(t-1)}$$

and

$$G_{\mu^{\boxplus t}}(z) = \frac{(t-2)z - t\sqrt{z^2 - 4a^2(t-1)}}{2(a^2t^2 - z^2)}.$$

By Stieltjes inversion it follows that

$$\mu^{\boxplus t} = \frac{1}{2\pi}\frac{t\sqrt{(4a^2(t-1)-x^2)_+}}{2(a^2t^2-x^2)}\,dx + \left(1 - \frac{t}{2}\right)_+ (\delta_{-at} + \delta_{at})$$

for $t \geq 1$. Note that these convolution powers are thus defined for all $t \geq 1$! In particular, for $t = 2$,

$$G_{\mu\boxplus\mu}(z) = \frac{z - 2\sqrt{z^2-4a^2}}{2(4a^2-z^2)}$$

and so

$$\mu \boxplus \mu = \frac{1}{2\pi}\frac{1}{\sqrt{4a^2-x^2}}\,dx,$$

the standard arcsine distribution.

Subordination function approach

Proposition 6.28. *(Voiculescu 1993, Biane 1998, Voiculescu 2000, Belinschi, Bercovici 2007) Let μ_1, μ_2 be probability measures on $\mathbb{R}$. Then $F_{\mu_1 \boxplus \mu_2}$ is analytically subordinate to F_{μ_1}: there is an analytic map $\omega_1 : \mathbb{C}^+ \to \mathbb{C}^+$ (the* subordination function*) such that*

$$F_{\mu_1 \boxplus \mu_2}(z) = F_{\mu_1}(\omega_1(z)).$$

Similarly, there is a function ω_2 such that $F_{\mu_1 \boxplus \mu_2}(z) = F_{\mu_2}(\omega_2(z))$. Both functions satisfy $\Im \omega(z) \geq \Im z$ and $\omega(z) = z(1 + o(1))$ as $z \to \infty$ nontangentially to $\mathbb{R}$.

Note that the same subordination relation will then hold for the Cauchy transforms.

Of course, on a domain

$$\omega_1(z) = F_{\mu_1}^{-1}(F_{\mu_1 \boxplus \mu_2}(z)) = F_{\mu_1 \boxplus \mu_2}(z) + \varphi_{\mu_1}(F_{\mu_1 \boxplus \mu_2}(z)).$$

The point is that ω_1 can be analytically continued to a function from $\mathbb{C}^+$ to itself. Instead of proving this directly, we take a different approach, which in particular will allow us to (if necessary) compute ω_j numerically, thereby leading to numerical valued of $G_{\mu_1 \boxplus \mu_2}$ and $\mu_1 \boxplus \mu_2$. Subordination functions have also proven useful in the study of qualitative properties of free convolution, limit theorems etc.

The following is a version of the Denjoy-Wolff theorem.

Theorem 6.29. *Let f be an analytic self map of the unit disk $\mathbb{D}$, neither a constant nor an automorphism of the disk. Then for any initial $z_0 \in \mathbb{D}$, the iterates $f^{\circ n}(z_0)$ converge (uniformly on compact sets) to the* Denjoy-Wolff *point z in the closed unit disk $\overline{\mathbb{D}}$. If $z \in \mathbb{D}$, then z is the unique fixed point of f, and $|f'(z)| < 1$.*

The case of $z \in \mathbb{D}$ is easy to prove using Schwarz Lemma. By using the Cayley transform $i\frac{1+z}{1-z}$, we may translate this result to $\mathbb{C}^+$ and its closure.

To simplify notation, we will denote $F_j = F_{\mu_j}$ and $H_j(z) = F_j(z) - z$.

Proposition 6.30. *Let*

$$g_z(w) = g_1(z, w) = z + H_2(z + H_1(w)).$$

Then for any $z \in \mathbb{C}^+$, the map g_z has a unique fixed point $\omega_1 \in \mathbb{C}^+$, to which its iterations converge. As a function of z, ω_1 is analytic, and increases the imaginary part.

Proof. (Sketch) Recall that F_j increases the imaginary part. Therefore H_j maps $\mathbb{C}^+$ to itself. It follows that for each z, g_z maps $\mathbb{C}^+$ to $z + \mathbb{C}^+$. In particular, it is not an automorphism of $\mathbb{C}^+$, and its iterations cannot converge to a point in $\mathbb{R}$. We omit the argument showing they do not converge to infinity. The existence of the fixed point then follows from the Denjoy-Wolff theorem. Since

$$\left| \frac{\partial(w - g_1(z, w))}{\partial w} \right| = 1 - |g_z'(w)| > 0,$$

analyticity follows from the implicit function theorem. □

Theorem 6.31. *There are analytic functions $\omega_1, \omega_2 : \mathbb{C}^+ \to \mathbb{C}^+$ such that*

$$F_1(\omega_1(z)) = F_2(\omega_2(z))$$

and

$$\omega_1(z) + \omega_2(z) = z + F_1(\omega_1(z)).$$

Moreover such functions are unique. They both increase the imaginary part, and satisfy

$$\lim_{y \to +\infty} \frac{\omega_j(iy)}{iy} = 1.$$

Proof. Define $\omega_1(z)$ as the unique fixed point

$$g_1(z, \omega_1(z)) = \omega_1(z),$$

and define

$$\omega_2(z) = z + F_1(\omega_1(z)) - \omega_1(z).$$

Then both functions are analytic, map $\mathbb{C}^+$ to itself, and increase the imaginary part. The second equation is clearly satisfied. Also,

$$\begin{aligned} F_2(\omega_2(z)) - F_1(\omega_1(z)) &= F_2(z + H_1(\omega_1)) - F_1(\omega_1(z)) \\ &= z + F_2(z + H_1(\omega_1)) - (z + H_1(\omega_1(z))) + \omega_1(z) = g(z, \omega_1(z)) - \omega_1(z) = 0, \end{aligned}$$

so the first equation is satisfied as well. Conversely, the second equation implies the relation between ω_1 and ω_2, while the first equation implies that ω_1 is the fixed point of $g_1(z, \cdot)$, and so is unique. Exchanging 1 and 2, ω_2 is the fixed point of

$$g_2(z, \cdot) = z + H_1(z + H_2(\cdot)),$$

and so is unique as well. Finally, by the Nevanlinna representation, the limit $\ell = \lim_{y \to +\infty} \frac{\omega_2(iy)}{iy}$ exists. Since ω_1 increases the imaginary part, $\lim_{y \to +\infty} \frac{\omega_1(iy)}{iy} \geq 1$, and $\omega_1(iy)$ goes to infinity nontangentially as $y \to \infty$. On the other hand,

$$\begin{aligned} \ell + \lim_{y \to +\infty} \frac{\omega_1(iy)}{iy} = \lim_{y \to +\infty} \frac{\omega_1(iy) + \omega_2(iy)}{iy} &= \limsup_{y \to +\infty} \frac{iy + F_2(\omega_2(iy))}{iy} \\ &= 1 + \limsup_{y \to +\infty} \frac{F_2(\omega_2(iy))}{\omega_2(iy)} \cdot \frac{\omega_2(iy)}{iy} = 1 + \ell, \end{aligned}$$

which implies that the limit for ω_1 (and so by symmetry also for ω_2) is 1. □

Finally, we verify that $F = F_1 \circ \omega_1 = F_2 \circ \omega_2$ is in fact $F_{\mu_1 \boxplus \mu_2}$. Since both functions are analytic, it suffices to show they, or their inverses, coincide on a domain. Indeed, from the second equation in the theorem,

$$F^{-1}(z) = \omega_1(F^{-1}(z)) + \omega_2(F^{-1}(z)) + z = F_1^{-1}(z) + F_2^{-1}(z) - z,$$

that is,

$$F^{-1}(z) - z = \varphi_{\mu_1}(z) + \varphi_{\mu_2}(z).$$

6.3 Asymptotic freeness.

Definition 6.32. Let $(a_1, \ldots, a_k) \subset (\mathcal{A}, \varphi)$. Their $*$-distribution is the state $\mu_{a_1,\ldots,a_k}$ on the space of non-commutative polynomials $\mathbb{C}\langle x_1, x_1^*, \ldots, x_k, x_k^* \rangle$ defined by

$$\mu_{a_1,\ldots,a_k}\left[P(x_1, x_1^*, \ldots, x_k, x_k^*)\right] = \varphi\left[P(a_1, a_1^*, \ldots, a_k, a_k^*)\right]$$

(check that this is a state!). If a_i's are self-adjoint, their joint distribution is the state $\mu_{a_1,\ldots,a_k}$ on the space of non-commutative polynomials $\mathbb{C}\langle x_1, \ldots, x_k \rangle$ (where the formal variables x_i are now considered self adjoint).

Note that the joint distribution may be identified with the collection of all joint moments, and the $*$-distribution with the collection of all joint $*$-moments.

Definition 6.33. Let $a_1^{(N)}, a_2^{(N)}, \ldots, a_k^{(N)} \in (\mathcal{A}_N, \varphi_N)$ be self-adjoint.

a. We say that
$$(a_1^{(N)}, \ldots, a_k^{(N)}) \to (a_1, \ldots, a_k) \subset (\mathcal{A}, \varphi)$$
in distribution if for each $\mathbf{u}$,
$$\varphi_N\left[a_{u(1)}^{(N)} a_{u(2)}^{(N)} \ldots a_{u(n)}^{(N)}\right] \to \varphi\left[a_{u(1)} a_{u(2)} \ldots a_{u(n)}\right]$$
as $N \to \infty$. Equivalently,
$$R^{\varphi_N}\left[a_{u(1)}^{(N)}, a_{u(2)}^{(N)}, \ldots, a_{u(n)}^{(N)}\right] \to R^{\varphi}\left[a_{u(1)}, a_{u(2)}, \ldots, a_{u(n)}\right]$$
or for all $P \in \mathbb{C}\langle x_i \rangle_{i=1}^k$,
$$\varphi_N\left[P(a_1^{(N)}, a_2^{(N)}, \ldots, a_k^{(N)})\right] \to \varphi\left[P(a_1, a_2, \ldots, a_k)\right].$$
In this case we say that $(a_1^{(N)}, a_2^{(N)}, \ldots, a_k^{(N)})$ have the asymptotic distribution
$$\mu_{a_1,a_2,\ldots,a_k}.$$
Note that limits of tracial joint distributions are tracial. Convergence in $*$-distribution (for non-self-adjoint elements) is defined similarly.

b. $a_1^{(N)}, a_2^{(N)}, \ldots, a_k^{(N)}$ are *asymptotically free* if $(a_1, \ldots, a_k)$ from part (a) are free in $(\mathcal{A}, \varphi)$. Equivalently,
$$R^{\varphi_N}\left[a_{u(1)}^{(N)}, a_{u(2)}^{(N)}, \ldots, a_{u(n)}^{(N)}\right] \overset{N \to \infty}{\longrightarrow} 0$$
unless all $u(1) = u(2) = \ldots = u(n)$.

c. Random matrices $X_1^{(N)}, X_2^{(N)}, \dots, X_k^{(N)}$ are *almost surely asymptotically free* if for each $\mathbf{u}$,

$$\frac{1}{N} \operatorname{Tr}\left[X_{u(1)}^{(N)} X_{u(2)}^{(N)} \dots X_{u(n)}^{(N)}\right] \to \varphi\left[a_{u(1)} a_{u(2)} \dots a_{u(n)}\right]$$

almost surely as $N \to \infty$, and $(a_1, \dots, a_k) \subset (\mathcal{A}, \varphi)$ are free.

d. Suppose all $\mathcal{A}_N$, $\mathcal{A}$ are actually C^*-algebras, and so have norms. Note that this is the case for random matrices. We say that

$$(a_1^{(N)}, \dots, a_k^{(N)}) \to (a_1, \dots, a_k) \subset (\mathcal{A}, \varphi)$$

strongly in distribution if they converge in distribution and in addition, for each non-commutative polynomial P in k variables,

$$\left\| P(a_1^{(N)}, a_2^{(N)}, \dots, a_k^{(N)}) \right\|_{\mathcal{A}_N} \to \| P(a_1, a_2, \dots, a_k) \|_{\mathcal{A}}$$

as $N \to \infty$. $a_1^{(N)}, a_2^{(N)}, \dots, a_k^{(N)}$ are strongly asymptotically free if $(a_1, \dots, a_k) \subset (\mathcal{A}, \varphi)$ are free.

Remark 6.34. Consider the case of a single Hermitian random matrix X_N. Fix $\varepsilon > 0$. If $X_N \to \sigma$ in distribution, in particular, the proportion of eigenvalues of X_N in $(2 + \varepsilon, \infty)$ goes to zero. But this interval may still contain $o(N)$ eigenvalues, for any N: X_N may have *outliers*. If $X_N \to \sigma$ strongly in distribution, then for sufficiently large N, X_N has no eigenvalues larger than $2 + \varepsilon$ (no outliers). As noted on page 7 of the notes (Füredi, Komlos 1981, Bai, Yin 1988), GOE/GUE matrices have no outliers. In the last chapter, we will analyse matrix models which do have outliers.

Theorem 6.35. *(Voiculescu 1991, 1998) Let $X_1, \dots, X_k$ be entry-wise independent (normalized) GUE matrices, and $D_1, \dots, D_q$ be non-random matrices with an asymptotic (tracial) joint distribution $\mu_{d_1, \dots, d_q}$.*

a. $X_1, \dots, X_k$ are asymptotically freely independent. Consequently, in distribution

$$(X_1, \dots, X_k) \to (s_1, \dots, s_k),$$

where $(s_1, \dots, s_k)$ is a free semicircular system.

b. More generally, in distribution

$$(X_1, \dots, X_k, D_1, \dots, D_q) \to (s_1, \dots, s_k, d_1, \dots, d_q),$$

where $s_1, \dots, s_k$ are free semicircular elements free from $d_1, \dots, d_q$.

Proof of part (a). We follow the ideas from the proof of Theorem 2.1. For $X^{(i)} = \frac{1}{\sqrt{N}} Y^{(i)}$, consider

$$\frac{1}{N} \mathbb{E}\left[\operatorname{Tr}\left[X^{u(1)} \dots x^{u(n)}\right]\right] = \frac{1}{N^{1+n/2}} \sum_{i(1), i(2), \dots, i(n)=1}^{N} \mathbb{E}\left[Y_{i(1)i(2)}^{u(1)} \dots Y_{i(n)i(1)}^{u(n)}\right].$$

Using the Wick formula from Exercise 6.15, this is zero if n is odd. Replacing n by $2n$, as in the proof of Theorem 2.1, the only terms which survive in the large N limit are those corresponding to trees with $n+1$ vertices, each edge traversed exactly twice. Consider instead a $2n$-gon, with vertices labeled by $i(1), \ldots, i(2n)$, entries corresponding to the factors in the product above. Wick formula gives a pairing between edges of this $2n$-gon. It is not hard to see that the corresponding graph is a tree as above if any only if the corresponding pairing is non-crossing. Recalling additionally that different matrices have independent entries, the expression above is asymptotically equal to

$$\left| \left\{ \pi \in \mathcal{NC}_2(n) \mid i \overset{\pi}{\sim} j \ \Rightarrow u(i) = u(j) \right\} \right| .$$

Now compare with Exercise 6.18. □

Exercise 6.36. For the full proof of part (b), see Theorem 11.41 in the Free Probability Notes. We only give an illustrative example. Let X be GUE, and $D^{(1)}, \ldots, D^{(4)}$ (for each N) fixed (non-random) matrices. Consider

$$\frac{1}{N} \mathbb{E} \circ \mathrm{Tr} \left[X D^{(1)} X D^{(2)} X D^{(3)} X D^{(4)} \right] .$$

According to the Wick formula, this expression is again a sum of three terms, corresponding to the pair partitions $\{(1,2),(3.4)\}$, $\{(1,4),(2,3)\}$, and $\{(1,3)(2,4)\}$. The first one is

$$\begin{aligned} &\frac{1}{N} \mathbb{E} \left[X_{i(1)i(2)} D_{i(2)i(2)} X_{i(2)i(1)} D_{i(1)i(3)} X_{i(3)i(4)} D_{i(4)i(4)} X_{i(4)i(3)} D_{i(3)i(1)} \right] \\ &\quad = \frac{1}{N} \frac{1}{N^2} \sum_{i(2)} D_{i(2)i(2)} \sum_{i(4)} D_{i(4)i(4)} \sum_{i(1),i(3)} D_{i(1)i(3)} D_{i(3)i(1)} \\ &\quad = \left(\frac{1}{N} \mathrm{Tr}[D^{(1)}] \right) \left(\frac{1}{N} \mathrm{Tr}[D^{(3)}] \right) \left(\frac{1}{N} \mathrm{Tr}[D^{(2)} D^{(4)}] \right) . \end{aligned}$$

Similarly, the second term is $\left(\frac{1}{N} \mathrm{Tr}[D^{(2)}] \right) \left(\frac{1}{N} \mathrm{Tr}[D^{(4)}] \right) \left(\frac{1}{N} \mathrm{Tr}[D^{(1)} D^{(3)}] \right)$. In the third term,

$$\begin{aligned} &\frac{1}{N} \mathbb{E} \left[X_{i(1)i(2)} D_{i(2)i(3)} X_{i(3)i(4)} D_{i(4)i(2)} X_{i(2)i(1)} D_{i(1)i(4)} X_{i(4)i(3)} D_{i(3)i(1)} \right] \\ &\qquad\qquad = \frac{1}{N^2} \left(\frac{1}{N} \mathrm{Tr} \left[D^{(1)} D^{(4)} D^{(3)} D^{(2)} \right] \right) . \end{aligned}$$

Note the order! However this last term disappears in the large N limit. Now let s be a standard semicircular variable, free from d_1, d_2, d_3, d_4. Show (probably by expanding in terms of free cumulants) that

$$\varphi \left[s d_1 s d_2 s d_3 s d_4 \right] = \varphi[d_1] \varphi[d_3] \varphi[d_2 d_4] + \varphi[d_2] \varphi[d_4] \varphi[d_1 d_3].$$

Remark 6.37. If $(D_1, \ldots, D_q)$ converge almost surely in distribution, so do $(X_1, \ldots, X_k, D_1, \ldots, D_q)$. If $(D_1, \ldots, D_q)$ converge strongly in distribution, so do $(X_1, \ldots, X_k, D_1, \ldots, D_q)$ (Haagerup, Thorbjørnsen 2005, Male 2011). The same conclusions hold if X_i's are GOE matrices (Schultz 2005). We will see in the last chapter that there are differences between GUE and GOE matrices in the subleading order in N.

Similar results hold for Wigner matrices with finite fourth moment (Dykema 1993, Capitaine, Donati-Martin 2007, Anderson 2013), at least in the case when $\left\| D_i^{(N)} \right\|$ are uniformly bounded, and random permutation matrices (Nica 1993, Bordenave, Collins 2018). They fail for heavy Wigner matrices (Ryan 1998, Male 2017).

Exercise 6.38. Let X_N be a normalized $N \times N$ complex Ginibre matrix, that is, $X_N = \frac{1}{\sqrt{N}} Y_N$, where entries of Y_N are independent complex Gaussian variables (no symmetry). Show that X_N converges in $*$-distribution to a circular variable. That is, if c and ε are as in Exercise 6.19, then

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E}\left[\operatorname{Tr}\left[X_N^{\varepsilon(1)} \dots X_N^{\varepsilon(n)} \right] \right] = \varphi\left[c^{\varepsilon(1)} \dots x^{\varepsilon(n)} \right].$$

Exercise 6.39. In a ncps $(\mathcal{A}, \varphi)$, let $\{s_i\}_{i=1}^N$ be free standard semicircular variables, and $\{c_{i,j}\}_{1 \le i < j \le N}$ be free circular variables free from them. Form a matrix S with entries

$$S_{ii} = s_i, \quad S_{ij} = c_{ij} \text{ if } i < j, \quad S_{ij} = c_{ij}^* \text{ if } i > j;$$

thus S is a free analog of a GUE matrix. Show that as an element of the ncps $(M_N(\mathcal{A}), \frac{1}{N} \varphi \circ \operatorname{Tr})$, S has the semicircular distribution (exactly, for any N).

Definition 6.40. An element $u \in (\mathcal{A}, \varphi)$ is a *Haar unitary* if $uu^* = u^*u = 1$ and

$$\varphi[u^n] = \varphi[(u^*)^n] = \delta_{n=0}.$$

For example, a CUE matrix (of any size) has this property.

Theorem 6.41. *Let $U_1, \dots, U_k$ be independent $N \times N$ CUE (Haar unitary) matrices, and $(D_1, \dots, D_q)$ non-random matrices converging in $*$-distribution to $(d_1, \dots, d_q) \subset (\mathcal{A}, \varphi)$. Then in $*$-distribution,*

$$(U_1, \dots, U_k, D_1, \dots, D_q) \to (u_1, \dots, u_k, d_1, \dots, d_q),$$

where $u_1, \dots, u_k$ are free Haar unitaries free from $\{d_1, \dots, d_q\}$.

Methods of proof. The original proof of Voiculescu used the fact that (ignoring invertibility issues of the moment), for X_N a GUE matrix, $U_N = X_N (X_N^* X_N)^{-1/2}$ is a Haar unitary matrix. Approximating $x/|x|$ by bounded continuous functions, asymptotic freeness of Haar unitary matrices follows from the asymptotic freeness of GUE matrices. Alternatively, one may use the Weingarten calculus methods discussed in class. □

Remark 6.42. If $(D_1, \dots, D_q)$ converge strongly in distribution, so do $(U_1, \dots, U_k, D_1, \dots, D_q)$ (Collins, Male 2014) The conditions on U_i in the theorem can be considerably weakened, as shown for example by (Anderson, Farrell 2014).

Exercise 6.43. Suppose $\{u_1, \dots, u_k\}$ are free Haar unitaries. Suppose the family $(a_i)_{i=0}^k$ is freely independent from the family $(u_i)_{i=1}^k$ (note that a_i's need not be free among themselves). Then the family $a_0, u_1 a_1 u_1^*, \dots, u_k a_k u_k^*$ is free. In fact the full strength of the assumption that u_i are Haar unitaries is not necessary; what weaker condition suffices?

Corollary 6.44. *Let A_N, B_N be $N \times N$ (non-random) matrices such that A_N and B_N converge in distribution as $N \to \infty$. Let U_N be an $N \times N$ CUE matrix . Then $U_N A_N U_N^*$ and B_N are asymptotically free. Consequently,*

$$\hat{\mu}_{A_N+B_N} \to \mu \boxplus \nu$$

weakly.

Remark 6.45. If A_N and B_N (with a finite asymptotic absolute first moment) converge in distribution almost surely, so does $A_N + B_N$ (Speicher 1993, Pastur and Vasilchuk 2000).

Example 6.46. Let A_{2N} be a $2N \times 2N$ matrix

$$A_{2N} = \begin{pmatrix} I_N & 0 \\ 0 & -I_N \end{pmatrix},$$

and U_{2N} be a CUE matrix. Note that

$$\hat{\mu}_{A_{2N}} = \hat{\mu}_{U_{2N} A_{2N} U_{2N}^*} = \frac{1}{2}\left(\delta_{-1} + \delta_1\right).$$

Then according to Example 6.27,

$$\hat{\mu}_{A_{2N}+U_{2N} A_{2N} U_{2N}^*} \to \frac{1}{\pi} \frac{1}{\sqrt{(4-x^2)_+}}\, dx$$

weakly almost surely as $N \to \infty$.

6.4 Extra section: proofs of asymptotic freeness

Unlike the rest of the notes, this section is not self-contained, but relies on the topics covered in Dr. Berkolaiko's part of the course. Our goal is to prove

Proposition 6.47. *Let A_N, B_N be $N \times N$ (non-random) Hermitian matrices such that A_N and B_N converge in distribution as $N \to \infty$. That is, for some $a, b \in (\mathcal{A}, \varphi)$,*

$$\frac{1}{N} \operatorname{Tr}[A_N^k] \to \varphi[a^k], \quad \frac{1}{N} \operatorname{Tr}[B_N^k] \to \varphi[b^k].$$

Let U_N be an $N \times N$ CUE (Haar unitary) matrix. Then $U_N A_N U_N^*$ and B_N are asymptotically free.

To prove asymptotic freeness of $U_N A_N U_N^*$ and B_N, we want to show that for any (polynomial) functions f_j, g_j with $\varphi[f_j(a)] = \varphi[g_j(b)] = 0$, the expression

$$\begin{aligned}\frac{1}{N} &\mathbb{E} \circ \operatorname{Tr}\left[f_1(U_N A_N U_N^*) g_1(B_N) \ldots g_{n-1}(B_N) f_n(U_N A_N U_N^*) g_n(B_n)\right] \\ &= \frac{1}{N} \mathbb{E} \circ \operatorname{Tr}\left[U_N f_1(A_N) U_N^* g_1(B_N) \ldots g_{n-1}(B_N) U_N f_n(A_N) U_N^* g_n(B_N)\right]\end{aligned}$$

goes to zero. To simplify notation, denote $A^{(i)} = f_i(A_N)$, $B^{(i)} = g_i(B_N)$. Thus we need to compute

$$\begin{aligned}
&\frac{1}{N}\,\mathbb{E}\circ\operatorname{Tr}\left[UA^{(1)}U^*B^{(1)}UA^{(2)}U^*B^{(2)}\dots UA^{(n)}U^*B^{(n)}\right] \\
&\quad = \frac{1}{N}\sum_{\substack{\mathbf{i},\mathbf{j}\\ \mathbf{u},\mathbf{v}}}\mathbb{E}\left[U_{i(1)j(1)}A^{(1)}_{j(1)v(1)}\bar U_{u(1)v(1)}B^{(1)}_{u(1)i(2)}\dots U_{i(n)j(n)}A^{(n)}_{j(n)v(n)}\bar U_{u(n)v(n)}B^{(n)}_{u(n)i(1)}\right] \\
&\quad = \frac{1}{N}\sum_{\substack{\mathbf{i},\mathbf{j}\\ \mathbf{u},\mathbf{v}}} A^{(1)}_{j(1)v(1)}B^{(1)}_{u(1)i(2)}A^{(2)}_{j(2)v(2)}B^{(2)}_{u(2)i(3)}\dots A^{(n)}_{j(n)v(n)}B^{(n)}_{u(n)i(1)} \\
&\qquad\quad \mathbb{E}\left[U_{i(1)j(1)}\bar U_{u(1)v(1)}U_{i(2)j(2)}\bar U_{u(2)v(2)}\dots U_{i(n)j(n)}\bar U_{u(n)v(n)}\right]
\end{aligned}$$

Using the notation from class and the definition of the Weingarten function, this equals

$$\begin{aligned}
&\frac{1}{N}\sum_{\substack{\mathbf{i},\mathbf{j}\\ \mathbf{u},\mathbf{v}}} A^{(1)}_{j(1)v(1)}B^{(1)}_{u(1)i(2)}A^{(2)}_{j(2)v(2)}B^{(2)}_{u(2)i(3)}\dots A^{(n)}_{j(n)v(n)}B^{(n)}_{u(n)i(1)}I^N_{\mathbf{i},\mathbf{u};\mathbf{j},\mathbf{v}} \\
&\quad = \frac{1}{N}\sum_{\substack{\mathbf{i},\mathbf{j}\\ \mathbf{u},\mathbf{v}}} A^{(1)}_{j(1)v(1)}B^{(1)}_{u(1)i(2)}A^{(2)}_{j(2)v(2)}B^{(2)}_{u(2)i(3)}\dots A^{(n)}_{j(n)v(n)}B^{(n)}_{u(n)i(1)}\sum_{\substack{\beta\in S_n:\mathbf{i}_\beta=\mathbf{u}\\ \alpha\in S_n:\mathbf{j}_\alpha=\mathbf{v}}}\operatorname{Wg}^N_n(\beta\alpha^{-1}) \\
&\quad = \frac{1}{N}\sum_{\substack{\mathbf{i},\mathbf{j}\\ \mathbf{u},\mathbf{v}}} A^{(1)}_{j(1)v(1)}B^{(1)}_{u(1)i(2)}A^{(2)}_{j(2)v(2)}B^{(2)}_{u(2)i(3)}\dots A^{(n)}_{j(n)v(n)}B^{(n)}_{u(n)i(1)} \\
&\qquad\quad \sum_{\alpha,\beta\in S_n}\delta_{i(\beta(1))=u(1)}\dots\delta_{i(\beta(n))=u(n)}\delta_{j(\alpha(1))=v(1)}\dots\delta_{j(\alpha(n))=v(n)}\operatorname{Wg}^N_n(\beta\alpha^{-1}) \\
&\quad = \frac{1}{N}\sum_{\alpha,\beta\in S_n}\operatorname{Wg}^N_n(\beta\alpha^{-1})\sum_{\mathbf{j}}A^{(1)}_{j(1)j(\alpha(1))}\dots A^{(n)}_{j(n)j(\alpha(n))}\sum_{\mathbf{u}}B^{(1)}_{u(1)u(\beta^{-1}\gamma(1))}B^{(n)}_{u(n)u(\beta^{-1}\gamma(n))},
\end{aligned}$$

where $\gamma = (12\dots n)$ is the long cycle. For a permutation α and matrices $\{X_i\}$, we can write

$$\operatorname{Tr}_\alpha[X_1,\dots,X_n] = \prod_{\text{cycles of }\alpha}\operatorname{Tr}\left[\prod_{i\in\text{ cycle}}X_i\right].$$

For example,

$$\operatorname{Tr}_{(152)(34)}[X_1,X_2,X_3,X_4,X_5] = \operatorname{Tr}[X_1X_5X_2]\operatorname{Tr}[X_3X_4].$$

Note that this is well-defined because Tr is cyclically symmetric. Then the expression above is

$$\frac{1}{N}\sum_{\alpha,\beta\in S_n}\operatorname{Wg}^N_n(\beta\alpha^{-1})\operatorname{Tr}_\alpha\left[A^{(1)},\dots,A^{(n)}\right]\operatorname{Tr}_{\beta^{-1}\gamma}\left[B^{(1)},\dots,B^{(n)}\right].$$

Using the notation from class that $|\alpha|$ is the number of cycles of α, this equals to

$$\frac{1}{N}\sum_{\alpha,\beta\in S_n}\operatorname{Wg}^N_n(\beta\alpha^{-1})N^{|\alpha|+\left|\beta^{-1}\gamma\right|}\left(\frac{1}{N^{|\alpha|}}\operatorname{Tr}_\alpha\left[A^{(1)},\dots,A^{(n)}\right]\right)\left(\frac{1}{N^{|\beta^{-1}\gamma|}}\operatorname{Tr}_{\beta^{-1}\gamma}\left[B^{(1)},\dots,B^{(n)}\right]\right).$$

Recalling the Weingarten function asymptotics

$$\mathrm{Wg}_n^N(\alpha) = \mu(\alpha)\frac{1}{N^{2n-|\alpha|}} + O\left(\frac{1}{N^{2n-|\alpha|+2}}\right)$$

and the limiting distributions of A_N, B_N, the expression above is, asymptotically

$$\sum_{\alpha,\beta\in S_n} \mu(\beta\alpha^{-1})\frac{1}{N^{2n+1-|\beta\alpha^{-1}|-|\alpha|-|\beta^{-1}\gamma|}}\left(1+o(1)\right)\varphi_\alpha\left[f_1(a),\ldots,f_n(a)\right]\varphi_{\beta^{-1}\gamma}\left[g_1(b),\ldots,g_n(b)\right]. \quad (6.2)$$

So the question is, is it true that for all α, β,

$$2n+1-\left|\beta\alpha^{-1}\right|-|\alpha|-\left|\beta^{-1}\gamma\right| \geq 0, \quad (6.3)$$

and for what α, β do we have equality?

Lemma 6.48. *Let $\alpha \in S_n$, and $\tau = (ab)$ be a transposition. Then*

$$|\alpha\tau| = \begin{cases} |\alpha|+1, & \text{if } a,b \text{ are in the same cycle of } \alpha, \\ |\alpha|-1, & \text{if } a,b \text{ are in different cycles of } \alpha. \end{cases}$$

Proof. The result follows from the observation that

$$\begin{aligned}(u(1)\ldots u(i)\ldots u(j)\ldots u(k))\cdot(u(i)u(j)) =&(u(1)\ldots u(i-1)u(i)u(j+1)\ldots u(k))\\ &(u(i+1)\ldots u(j-1)u(j))\end{aligned}$$

and

$$(u(1)\ldots u(k))(v(1)\ldots v(m))\cdot(u(i)v(j)) = (u(1)\ldots u(i)v(j+1)\ldots v(m)v(1)\ldots v(j)u(i+1)\ldots u(k)).$$

□

Lemma 6.49. *For $\alpha, \beta \in S_n$, define*

$$d(e,\alpha) = \min\left\{k : \alpha = \tau_1\tau_2\ldots\tau_k \text{ for } \tau_i \text{ transpositions}\right\},$$

and

$$d(\alpha,\beta) = d(e,\alpha^{-1}\beta).$$

a. d is a metric on S_n. (In fact, it is a distance in a certain Cayley graph of S_n.)

b.

$$d(e,\alpha) = n - |\alpha|.$$

c. $|\alpha\beta| = |\beta\alpha|$.

Proof. For part (a), $d(\alpha, \alpha) = 0$; since $\alpha^{-1}\nu = (\alpha^{-1}\beta)(\beta^{-1}\nu)$,

$$d(\alpha, \nu) = d(e, \alpha^{-1}\nu) \leq d(e, \alpha^{-1}\beta) + d(e, \beta^{-1}\nu) = d(\alpha, \beta) + d(\beta, \nu);$$

and

$$\alpha^{-1}\beta = \tau_1\tau_2 \ldots \tau_k \quad \Leftrightarrow \quad \beta^{-1}\alpha = \tau_k \ldots \tau_2\tau_1.$$

For part (b), note first that we may decompose a k-cycle

$$(u(1)u(2) \ldots u(k)) = (u(1)u(2)) \cdot (u(2)u(3)) \cdot \ldots \cdot (u(k-2)u(k-1)) \cdot (u(k-1)u(k))$$

as a product of $k-1$ transpositions. So

$$n - |\alpha| = \sum_{V \text{ a cycle in } \alpha} (\#(V) - 1) \geq d(e, \alpha).$$

On the other hand, since $|e| = n$, by Lemma 6.48

$$|\tau_1 \ldots \tau_k| \geq n - k,$$

and so $|\alpha| \geq n - d(e, \alpha)$. The result follows. Part (c) is equivalent to proving that in general, $|\alpha^{-1}\beta\alpha| = |\beta|$, which is true since conjugation preserves the cycle structure. □

Now observe that the expression (6.3) is

$$\begin{aligned} 2n + 1 - \left|\beta\alpha^{-1}\right| - |\alpha| - \left|\beta^{-1}\gamma\right| &= (n - |\alpha|) + (n - \left|\alpha^{-1}\beta\right|) + (n - \left|\beta^{-1}\gamma\right|) - (n-1) \\ &= d(e, \alpha) + d(\alpha, \beta) + d(\beta, \gamma) - d(e, \gamma). \end{aligned}$$

Thus this expression is always non-negative. When is it zero?

Remark 6.50. Partitions are *partially ordered* by reverse refinement: if

$$\pi = \{B_1, \ldots, B_k\}, \quad \sigma = \{C_1, \ldots, C_r\}$$

then

$$\pi \leq \sigma \quad \Leftrightarrow \quad \forall i \, \exists j : B_i \subset C_j.$$

For example, if $\pi = \{(1, 3, 5)(2)(4, 6)\}$ and $\sigma = \{(1, 3, 5)(2, 4, 6)\}$, then $\pi \leq \sigma$. This partial order restricts to $\mathcal{NC}(n)$.

Remark 6.51. We have a natural embedding

$$P : \mathcal{P}(n) \hookrightarrow S_n, \quad \pi \mapsto P_\pi,$$

where a block of a partition is mapped to a cycle of a permutation, in increasing order.

Denote by $S_{\mathcal{NC}}(n)$ the image of $\mathcal{NC}(n)$ in S_n under the embedding above.

Theorem 6.52. *(Biane 1997)*

a. $\alpha \in S_{\mathcal{NC}}(n)$ *if and only if*

$$d(e,\alpha) + d(\alpha,\gamma) = n - 1 = d(e,\gamma),$$

that is, α *lies on a geodesic from* e *to* γ*.*

b. *For* $\alpha, \beta \in S_{\mathcal{NC}}(n)$*, denote* $\alpha \leq \beta$ *if*

$$d(e,\alpha) + d(\alpha,\beta) + d(\beta,\gamma) = n - 1.$$

That is, α *and* β *lie on the same geodesic from* e *to* γ*, with* α *closer to* e*. Then the map* $\mathcal{NC}(n) \to S_{\mathcal{NC}}(n)$ *is a lattice isomorphism.*

Proof. Suppose

$$d(e,\alpha) + d(\alpha,\gamma) = d(e,\alpha) + d(e,\alpha^{-1}\gamma) = n - 1,$$

so that

$$\alpha = \tau_1\tau_2\ldots\tau_k, \quad \alpha^{-1}\gamma = \tau_{k+1}\ldots\tau_{n-1}.$$

Then $\gamma = \tau_1\tau_2\ldots\tau_{n-1}$. Recall that e has n cycles, γ has 1 cycle, and each multiplication by τ_j changes the number of cycles by 1. It follows that

$$|\tau_1\ldots\tau_j| = n - j.$$

Therefore going from $\tau_1\ldots\tau_j$ to

$$\tau_1\ldots\tau_{j-1} = (\tau_1\ldots\tau_j)\tau_j,$$

τ_j cuts one cycle into two, which by the proof of Lemma 6.48 happens in a non-crossing way. The converse is similar. For part (b), we note that $\alpha \leq \beta$ if and only if we can write

$$\alpha = \tau_1\ldots\tau_j, \quad \beta = \tau_1\ldots\tau_j\ldots\tau_k, \quad \gamma = \tau_1\ldots\tau_j\ldots\tau_k\ldots\tau_{n-1}.$$ □

It follows that expression (6.3),

$$2n + 1 - \left|\beta\alpha^{-1}\right| - |\alpha| - \left|\beta^{-1}\gamma\right| = d(e,\alpha) + d(\alpha,\beta) + d(\beta,\gamma) - d(e,\gamma),$$

is zero if and only if $\alpha = P_\sigma$, $\beta = P_\pi$, and $\sigma \leq \pi$.

Remark 6.53. Under the identification $\mathcal{NC}(n) \leftrightarrow S_{\mathcal{NC}}(n)$, $P_\pi^{-1}\gamma$ corresponds to the *Kreweras complement*: $K[\pi] \in \mathcal{NC}(n)$ is the largest partition in $\mathcal{NC}(\{\bar{1},\bar{2},\ldots,\bar{n}\})$ such that

$$\pi \cup K[\pi] \in \mathcal{NC}(\{1,\bar{1},2,\bar{2},\ldots,n,\bar{n}\}).$$

For example,

$$K[\{(1,2,5,6),(3,4),(7)\}] = \{(1),(2,4),(3),(5),(6,7)\}$$

and

$$((1,2,5,6)(3,4)(7))^{-1}(1,2,3,4,5,6,7) = (1,6,5,2)(3,4)(7)(1,2,3,4,5,6,7) = (1)(2,4)(3)(5)(6,7).$$

So in equation (6.2) we get

$$\sum_{\substack{\sigma,\pi\in\mathcal{NC}(n)\\ \sigma\leq\pi}} \mu(P_\pi P_\sigma^{-1})\,(1+o(1))\,\varphi_\sigma\left[f_1(a),\ldots,f_n(a)\right]\varphi_{K[\pi]}\left[g_1(b),\ldots,g_n(b)\right].$$

Recall that $\varphi[f_j(a)] = \varphi[g_j(b)] = 0$. It remains to note that by the lemma below, each term in the sum above involves a factor of either $\varphi[f_i(a)]$ or $\varphi[g_j(b)]$, and so equals zero. Asymptotic freeness of $U_N A_N U_N^*$ and B_N follows.

Lemma 6.54. *For $\sigma \leq \pi$ in $\mathcal{NC}(n)$, the partition $\sigma \cup K[\pi] \in \mathcal{NC}(2n)$ has at least one singleton block.*

Proof. Since $\sigma \leq \pi$, it suffices to show the result for $\pi \cup K[\pi]$. Recall that since π is non-crossing, at least one of its blocks is an interval. So if π has no singletons, then it has a block containing two neighboring elements. But then $K[\pi]$ contains a singleton (the one between these elements). □

Theorem 6.55. *Let X_N be a (normalized) GUE matrix, and $B_N^{(1)},\ldots,B_N^{(k)}$ be non-random Hermitian matrices with an asymptotic joint distribution. That is, there are self-adjoint elements $b_1,\ldots,b_k$ in some ncps $(\mathcal{A},\varphi)$ such that for all $\mathbf{u}$,*

$$\frac{1}{N}\operatorname{Tr}\left[B_N^{u(1)}\ldots B_N^{u(n)}\right] \to \varphi\left[b_{u(1)}\ldots b_{u(n)}\right].$$

Then X_N and the family $\left\{B_N^{(1)},\ldots,B_N^{(k)}\right\}$ are asymptotically free.

Proof. Since odd moments of both GUE matrices and semicircular operators are zero, it suffices to consider only even moments. Let $D^{(j)} = f_j(B^{(1)},\ldots,B^{(k)})$, $j=1,\ldots,2n$, f_j a polynomial. We compute

$$\begin{aligned}&\frac{1}{N}\mathbb{E}\circ\operatorname{Tr}\left[XD^{(1)}XD^{(2)}\ldots XD^{(2n)}\right]\\ &\qquad= \sum_{\mathbf{i},\mathbf{u}}\mathbb{E}\left[X_{i(1)u(1)}D_{u(1)i(2)}X_{i(2)u(2)}D_{u(2)i(3)}\ldots X_{i(2n)u(2n)}D_{u(2n)i(1)}\right].\end{aligned}$$

From the Wick formula for complex Gaussians, the expression above equals

$$\begin{aligned}&\sum_{\mathbf{i},\mathbf{u}}\sum_{\substack{\pi\in\mathcal{P}_2(2n)\\ \mathbf{i}_\pi=\mathbf{u},\mathbf{u}_\pi=\mathbf{i}}}\frac{1}{N^n}D^{(1)}_{u(1)i(2)}D^{(2)}_{u(2)i(3)}\ldots D^{(2n)}_{u(2n)i(1)}\\
&= \sum_{\pi\in\mathcal{P}_2(2n)}\frac{1}{N^{n+1}}\sum_{u(j)}D^{(1)}_{u(1)u(\pi(2))}D^{(2)}_{u(2)u(\pi(3))}\ldots D^{(2n)}_{u(2n)u(\pi(1))}\\
&= \sum_{\pi\in\mathcal{P}_2(2n)}\frac{1}{N^{n+1}}\sum_{u(j)}D^{(1)}_{u(1)u(\pi\gamma(1))}D^{(\pi\gamma(1))}_{u(\pi\gamma(1))u(\pi\gamma\pi\gamma(1))}D^{(\pi\gamma)^2(1)}_{u((\pi\gamma)^2(1))u((\pi\gamma)^3(1))}\cdots\\
&= \sum_{\pi\in\mathcal{P}_2(2n)}\frac{1}{N^{n+1}}\operatorname{Tr}_{\pi\gamma}\left[D^{(1)},D^{(2)},\ldots,D^{(2n)}\right]\\
&= \sum_{\pi\in\mathcal{P}_2(2n)}\frac{1}{N^{n-|\pi\gamma|+1}}\frac{1}{N^{|\pi\gamma|}}\operatorname{Tr}_{\pi\gamma}\left[D^{(1)},D^{(2)},\ldots,D^{(2n)}\right],\end{aligned}$$

where we identify π with P_π. Noting that $|\pi| = n$,

$$n - |\pi\gamma| + 1 = n + (2n - \left|\pi^{-1}\gamma\right|) - (2n-1) = d(e,\pi) + d(\pi,\gamma) - d(e,\gamma).$$

Thus as before, the limit of the expression above is

$$\sum_{\pi \in \mathcal{NC}_2(2n)} \varphi_{K[\pi]}\left[d_1, d_2, \ldots, d_{2n}\right],$$

where $d_j = f_j(b_1, \ldots, b_k)$. It remains to show that we obtain the same expression for the corresponding moment of a standard semicircular element s free from $\{d_1, \ldots, d_{2n}\}$. Indeed, using the fact that mixed free cumulants are zero, and that the semicircular element has only non-zero free cumulants of order two, which are equal to 1,

$$\begin{aligned}
\varphi\left[sd_1 sd_2 \ldots sd_{2n}\right] &= \sum_{\pi \in \mathcal{NC}_2(\{1,3,\ldots,4n-1\})} \sum_{\substack{\sigma \in \mathcal{NC}(\{2,4,\ldots,4n\}) \\ \pi \cup \sigma \in \mathcal{NC}(4n)}} R_\sigma(d_1, \ldots, d_{2n}) \\
&= \sum_{\pi \in \mathcal{NC}_2(2n)} \sum_{\sigma \leq K[\pi]} R_\sigma(d_1, \ldots, d_{2n}) = \sum_{\pi \in \mathcal{NC}_2(2n)} \varphi_{K[\pi]}\left[d_1, d_2, \ldots, d_{2n}\right],
\end{aligned}$$

where in the last step we used the free moment-cumulant formula. □

Chapter 7

Band and block matrices and operator-valued freeness.

Motivating question. In many applications one encounters random matrix ensembles which are not unitarily invariant and whose entries are not independent or identically distributed. Two commonly occurring generalizations of GOE/GUE matrices are the following.

A *Gaussian band random matrix* is an Hermitian $N \times N$ matrix $X = \frac{1}{\sqrt{N}} Y$, where the entries of Y are jointly complex Gaussian with mean zero and covariance

$$\mathbb{E}[Y_{rp} Y_{qs}] = \delta_{rs} \delta_{pq} \sigma(r/N, p/N).$$

Here $\sigma(x, y) = \sigma(y, x)$ is a sufficiently nice function. The choice of $\sigma(x, y) = \mathbf{1}_{|x-y|<\delta}$ leads to an actual band matrix.

A *Gaussian block random matrix* is an $Nd \times Nd$ matrix $X = \frac{1}{\sqrt{N}} Y$ considered as a $d \times d$ matrix of $N \times N$ blocks, such that the blocks $(Y^{(ij)})_{i,j=1}^{d}$ have jointly complex Gaussian entries with mean zero, $(Y^{(ij)})^* = Y^{(ji)}$, and covariance

$$\mathbb{E}[Y_{rp}^{(ij)} Y_{qs}^{(kl)}] = \delta_{rs} \delta_{pq} \sigma(i, j; k, l).$$

Note that for $d = 1$ we get a GOE matrix, while for $\sigma(i, j; k, l) = \delta_{il} \delta_{jk} \sigma(i, j)$, we get a special band matrix.

One can study asymptotic (joint) distributions of such matrices using *operator-valued* free probability.

7.1 Generalities

Definition 7.1. Let $\mathcal{B}$ be a C^*-algebra (see below for examples). A *$\mathcal{B}$-valued (non-commutative) probability space* is a triple $(\mathcal{A}, \Phi, \mathcal{B})$. Here $\mathcal{A}$ is a complex unital $*$-algebra containing $\mathcal{B}$, and $\Phi : \mathcal{A} \to \mathcal{B}$ is a *conditional expectation*: a $\mathbb{C}$-linear $\mathcal{B}$-bimodule map ($\Phi[b_1 a b_2] = b_1 \Phi[a] b_2$ for $a \in \mathcal{A}$ and $b_1, b_2 \in \mathcal{B}$) which is unital ($\Phi[1_{\mathcal{B}}] = 1_{\mathcal{B}}$), self-adjoint ($\Phi[a^*] = \varphi[a]^*$), and positive ($\Phi[a^* a]$ is a positive element in $\mathcal{B}$).

Note that if $\mathcal{B} = \mathbb{C}$, we obtain an ordinary ncps.

Example 7.2. In all cases below, the conditions on Φ are easy to verify (do so!).

a. Let $\mathcal{B} = M_d(\mathbb{C})$, and let $(\mathcal{A}_0, \varphi)$ be an ordinary ncps. Define $\mathcal{A} = M_d(\mathcal{A}_0)$ and $(\Phi[a])_{ij} = \varphi[a_{ij}]$. We may also write $\mathcal{A} = M_d(\mathbb{C}) \otimes \mathcal{A}_0$ and $\Phi = I \otimes \varphi$.

b. In the preceding example, take $(\mathcal{A}_0, \varphi) = (M_N(L^\infty(\Omega, \Sigma)), \frac{1}{N} \mathbb{E} \circ \mathrm{Tr})$. Then we may identify $\mathcal{A}$ with $M_{Nd}(L^\infty(\Omega, \Sigma))$. Here the map on $M_{Nd}(\mathbb{C})$ corresponding to $I_d \otimes \mathrm{Tr}$ in its identification with $M_d(\mathbb{C}) \otimes M_N(\mathbb{C})$ is the *partial trace*.

c. Let $\mathcal{B} = \mathrm{Diag}_N(\mathbb{C})$, the diagonal $N \times N$ matrices over $\mathbb{C}$. Note that we may identity $\mathcal{B}$ simply with $\mathbb{C}^N$. Let $\mathcal{A} = M_N(\mathbb{C})$, and

$$\Phi[a] = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{NN}),$$

the diagonal matrix with the same diagonal as a. More generally, we may let $\mathcal{A} = M_N(L^\infty(\Omega, \Sigma))$, and $\Phi[a] = \mathrm{diag}(\mathbb{E}[a_{11}], \mathbb{E}[a_{22}], \ldots, \mathbb{E}[a_{NN}])$, the expectation of the diagonal of a.

d. Let $\mathcal{B}$ be a C^*-algebra, and x a (self-adjoint) symbol. Let $\mathcal{A} = \mathcal{B}\langle x \rangle$ be the algebra of non-commutative polynomials,

$$\mathcal{B}\langle x \rangle = \mathbb{C} - \mathrm{Span}\,(\mathcal{B}, \mathcal{B}x\mathcal{B}, \mathcal{B}x\mathcal{B}x\mathcal{B}, \ldots).$$

It has a natural involution $(b_0 x b_1 x \ldots x b_n)^* = b_n^* x \ldots b_1^* x b_0^*$. A *$\mathcal{B}$-valued distribution* is a $\mathcal{B}$-valued conditional expectation on $\mathcal{B}\langle x \rangle$. Multivariate versions $\mathcal{B}\langle x_1, \ldots, x_k \rangle$ are defined similarly.

Definition 7.3. For a family $a_1, \ldots, a_k$ of self-adjoint elements in $(\mathcal{A}, \Phi, \mathcal{B})$, their joint distribution is a conditional expectation

$$\mu_{a_1, \ldots, a_k} : \mathcal{B}\langle x_1, \ldots, x_k \rangle \to \mathcal{B}$$

defined by

$$\mu_{a_1, \ldots, a_k}\left[b_0 x_{u(1)} b_1 x_{u(2)} \ldots x_{u(n)} b_n\right] = \Phi\left[b_0 a_{u(1)} b_1 a_{u(2)} \ldots a_{u(n)} b_n\right].$$

Check that it is in fact a conditional expectation!

Definition 7.4. Star-subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k \supset \mathcal{B}$ in a $\mathcal{B}$-valued probability space $(\mathcal{A}, \Phi, \mathcal{B})$ are *$\mathcal{B}$-free* if whenever $\Phi[a_1] = \Phi[a_2] = \ldots = \Phi[a_n] = 0$, $a_i \in \mathcal{A}_{u(i)}$, $u(1) \neq u(2) \neq u(3) \neq \ldots$ (neighbours distinct), then also for any $b_0, b_1, \ldots, b_n \in \mathcal{B}$,

$$\Phi[b_0 a_1 b_1 a_2 \ldots b_{n-1} a_n b_n] = 0.$$

Elements $a_1, a_2, \ldots, a_k$ are $\mathcal{B}$-free if the star-subalgebras they generate are $\mathcal{B}$-free. The terms “free with amalgamation over $\mathcal{B}$” and “ conditionally free over $\mathcal{B}$” are also used.

One can again construct $\mathcal{B}$-free copies of given nc random variables, using the reduced free product amalgamated over $\mathcal{B}$.

Definition 7.5. For a $\mathcal{B}$-valued probability space $(\mathcal{A}, \Phi, \mathcal{B})$, the n'th *moment functional* is the $\mathbb{C}$-multi-linear functional on $\mathcal{A}$,

$$\begin{aligned} M^{\Phi}(b_0 a_1 b_1, a_2 b_2, \ldots, a_{n-1} b_{n-1}, a_n b_n) &= b_0 M^{\Phi}(a_1, b_1 a_2, \ldots, b_{n-2} a_{n-1}, b_{n-1} a_n) b_n \\ &= \Phi[b_0 a_1 b_1 a_2 b_2 \ldots a_{n-1} b_{n-1} a_n b_n], \end{aligned}$$

where $a_i \in \mathcal{A}$, $b_i \in \mathcal{B}$. Equivalently, M^{Φ} is a linear functional on the bimodule amalgamated tensor product ${}_{\mathcal{B}}\mathcal{A} \otimes_{\mathcal{B}} \mathcal{A} \otimes_{\mathcal{B}} \ldots \otimes_{\mathcal{B}} \mathcal{A}_{\mathcal{B}}$.

Definition 7.6. Let π be a non-crossing partition of n. For a family of multi-linear functionals $(F_i)_{i=1}^{\infty}$, denote by F_π the n-linear functional obtained by nesting the functionals $(F_i)_{i=1}^{\infty}$ according to partition π. For example, for the partition $\pi = \{(1,2,6),(3,5),(4),(7),(8,10),(9)\}$,

$$F_\pi(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}) = F_3(a_1, a_2 F_2(a_3 F_1(a_4), a_5)) F_1(a_7) F_2(a_8 F_1(a_9), a_{10}).$$

Definition 7.7. For a $\mathcal{B}$-valued probability space $(\mathcal{A}, \Phi, \mathcal{B})$, define the n'th *free cumulant* functional R_n^{Φ} implicitly by

$$M_n^{\Phi}(a_1, a_2, \ldots, a_n) = \sum_{\pi \in \mathcal{NC}(n)} R_\pi^{\Phi}(a_1, a_2, \ldots, a_n)$$

Note that like M^{Φ}, it also have the property that

$$R^{\Phi}(b_0 a_1 b_1, a_2 b_2, \ldots, a_{n-1} b_{n-1}, a_n b_n) = b_0 R^{\Phi}(a_1, b_1 a_2, \ldots, b_{n-2} a_{n-1}, b_{n-1} a_n) b_n.$$

Example 7.8.

$$R^{\Phi}(b_0 a_1 b_1) = b_0 \Phi[a_1] b_1,$$

and so the mean $R^{\Phi}[a_1]$ is an element of $\mathcal{B}$. On the other hand,

$$R^{\Phi}(b_0 a_1 b_1, a_2 b_2) = b_0 \Big(\Phi[a_1 b_1 a_2] - \Phi[a_1] b_1 \Phi[a_2] \Big) b_2,$$

and so the variance should not be thought of as an element of $\mathcal{B}$, but as a (completely) positive map

$$b \mapsto R^{\Phi}(a_1 b, a_2) = \Phi[a_1 b a_2] - \Phi[a_1] b \Phi[a_2].$$

Theorem 7.9. *Star-subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k \supset \mathcal{B}$ in a $\mathcal{B}$-valued probability space $(\mathcal{A}, \Phi, \mathcal{B})$ are $\mathcal{B}$-*free *if and only if their mixed $\mathcal{B}$-valued free cumulants vanish:*

$$R^{\Phi}[a_1, a_2, \ldots, a_n] = 0$$

unless all a_i belong to the same $\mathcal{A}_j$.

7.2 Operator-valued semicircular elements

Example 7.10. As noted above, the variance of an element is actually a completely positive map on $\mathcal{B}$. For any such map $\eta : \mathcal{B} \to \mathcal{B}$, there is a $\mathcal{B}$-valued distribution with $\mathcal{B}$-valued free cumulants

$$R_2[b] = \eta(b), \quad R_n = 0 \text{ for } n \neq 2.$$

It is natural to call this distribution the $\mathcal{B}$-valued semicircular distribution with variance η. The low order even moments of an η-semicircular element a are

$$\Phi[ab_1a] = \eta(b_1),$$

$$\Phi[ab_1ab_2ab_3a] = \eta(b_1)b_2\eta(b_3) + \eta(b_1\eta(b_2)b_3),$$

$$\begin{aligned}\Phi[ab_1ab_2ab_3ab_4ab_5a] &= \eta(b_1)b_2\eta(b_3)b_4\eta(b_5) + \eta(b_1\eta(b_2)b_3)b_4\eta(b_5) + \eta(b_1)b_2\eta(b_3\eta(b_4)b_5) \\ &+ \eta(b_1\eta(b_2)b_3\eta(b_4)b_5) + \eta(b_1\eta(b_2\eta(b_3)b_4)b_5).\end{aligned}$$

Gaussian band and block matrices

Example 7.11. Band matrices fit in the setting of Example 7.2(c). For $X_N \in (M_N(L^\infty(\Omega, \Sigma)), \Phi)$ and $B \in \mathcal{B}_N = \mathrm{Diag}_N(\mathbb{C}) \simeq \mathbb{C}^N$, the variance of X_N is the map $\eta_N : \mathbb{C}^N \to \mathbb{C}^N$ given by

$$\eta_N(B) = \Phi[XBX] = \left(\sum_{j=1}^N \mathbb{E}[X_{ij}B_jX_{ji}]\right)_{i=1}^N = \left(\frac{1}{N}\sum_{j=1}^N \sigma(i/N, j/N)B_j\right)_{i=1}^N$$

Thus η_N is implemented by the matrix $(\frac{1}{N}\sigma(i/N, j/N))_{i,j=1}^N$.

For later reference, we consider the following construction. Let $\mathcal{B} = L^\infty[0,1)$. Embed $\mathcal{B}_N$ in $\mathcal{B}$ by

$$i_N : B \mapsto \sum_{i=1}^N B_i \mathbf{1}_{[\frac{i-1}{N}, \frac{i}{N})}.$$

Also define the function on $[0,1) \times [0,1)$ by

$$\sigma_N = \sum_{i,j=1}^N \sigma(i/N, j/N)\mathbf{1}_{[\frac{i-1}{N}, \frac{i}{N})\times[\frac{j-1}{N}, \frac{j}{N})}$$

and the map $\tilde{\eta}_N : \mathcal{B} \to \mathcal{B}$ by

$$\tilde{\eta}_N(f)(x) = \int_0^1 \sigma_N(x,y)f(y)\,dy.$$

Clearly $\tilde{\eta}_N \circ i_N = i_N \circ \eta_N$. Then for nice σ (for example, if its discontinuities belong to a finite set of rectifiable curves), $\tilde{\eta}_N$ converges in L^∞ to the map

$$\eta(f)(x) = \int_0^1 \sigma(x, y) f(y)\, dy,$$

the integral operator with kernel σ. Note that the standard state $\frac{1}{N}\operatorname{Tr}$ on $\mathcal{B}_N$ considered as acting on $i_N(\mathcal{B}_N)$ also converges, to the integration with respect to the Lebesgue measure.

For the case of genuine band matrix, $\sigma(x, y) = \mathbf{1}_{|x-y|<\delta}$, so that

$$\eta(f)(x) = \int_{\max(0,x-\delta)}^{\min(x+\delta,1)} f(y)\, dy.$$

Example 7.12. Block matrices fit in the setting of Example 7.2(b). By using Wigner's theorem and Exercise 6.38, the asymptotic distribution of such a normalized block matrix is the exact distribution of the element $S \in (M_d(\mathcal{A}), \frac{1}{d}\varphi \circ \operatorname{Tr})$. Here S_{ii} are semicircular, S_{ij} are circular for $i \neq j$, $S_{ij}^* = S_{ji}$, and the covariance between S_{ij} and S_{kl} is $\sigma(i, j; k, l)$.

If $\sigma(i, j; k, l) = \delta_{il}\delta_{jk}$, by Exercise 6.39 the distribution of S is (scalar-valued) semicircular. We now consider the case of general σ. The $M_d(\mathbb{C})$-valued variance of S is the entry-wise application

$$\varphi\left[SBS\right] = \left(\sum_{j,k=1}^{d} \varphi\left[S_{ij} B_{jk} S_{kl}\right]\right)_{i,l=1}^{d} = \left(\sum_{j,k=1}^{d} \sigma(i, j; k, l) B_{jk}\right)_{i,l=1}^{d}.$$

Thus it can be identified with an $d^2 \times d^2$ matrix with entries $\sigma(i, j; k, l)$ acting on $M_d(\mathbb{C}) \simeq \mathbb{C}^{d^2}$.

Theorem 7.13. *In the setting of Example 7.11, the asymptotic $\mathcal{B}_N$-valued distribution of X_N is the $\mathcal{B} = L^\infty[0, 1)$-valued semicircular distribution with covariance*

$$\eta(f)(x) = \int_0^1 \sigma(x, y) f(y)\, dy.$$

In the setting of Example 7.12, the precise $M_d(\mathbb{C})$-valued distribution of S (which is the asymptotic distribution of the corresponding block matrix) is semicircular with covariance

$$\eta(B) = \left(\sum_{j,k=1}^{d} \sigma(i, j; k, l) B_{jk}\right)_{i,l=1}^{d}.$$

Proof. We will discuss the setting of block matrices; the argument for band matrices is similar. Consider

$$\begin{aligned} &\left((I \otimes \varphi)\left[S B^{(1)} S B^{(2)} \ldots B^{(n-1)} S\right]\right)_{ab} \\ &= \sum_{\substack{u(0),u(1),\ldots,u(n-1)=1 \\ v(1),v(2),\ldots,v(n-1),v(n)=1}}^{d} \delta_{u(0)=a} \delta_{v(n)=b} \varphi\left[S_{u(0),v(1)} B^{(1)}_{v(1),u(1)} S_{u(1),v(2)} \ldots B^{(n-1)}_{v(n-1),u(n-1)} S_{u(n-1),v(n)}\right]. \end{aligned}$$

Each moment in the sum can be decomposed as a sum over non-crossing pairings. Let $\pi \in \mathcal{NC}_2(n)$ be such a pairing. Note that if n is odd, all the terms are zero, so from now on we assume that n is even. Since π is non-crossing, it contains an interval, that is, some i is paired off with $i+1$. Then the sum over $v(i)$ and $u(i)$ involves only the terms

$$\sum_{u(i),v(i)=1}^{d} \varphi\left[S_{u(i-1)v(i)} B^{(i)}_{v(i)u(i)} S_{u(i)v(i+1)}\right]$$
$$= \sum_{u(i),v(i)=1}^{d} \sigma(u(i-1),v(i);u(i),v(i+1)) B^{(i)}_{v(i)u(i)} = \eta(B^{(i)})_{u(i-1)v(i+1)}$$

by definition of σ and η. Thus we may remove from π the block $(i, i+1)$ and obtain the corresponding term in

$$\sum_{\substack{u(0),u(1),\ldots,\widehat{u(i)},\ldots,u(n-1)=1 \\ v(1),v(2),\ldots,\widehat{v(i)},\ldots,v(n-1),v(n)=1}}^{d} \delta_{u(0)=a}\delta_{v(n)=b}\varphi\Big[S_{u(0),v(1)} B^{(1)}_{v(1),u(1)} S_{u(1),v(2)} \ldots$$
$$S_{u(i-2),v(i-1)} B^{(i-1)}_{v(i-1),u(i-1)} \eta(B^{(i)}_{v(i),u(i)}) B^{(i+1)}_{v(i+1)u(i+1)} S_{u(i+1),v(i+2)} \ldots B^{(n-1)}_{v(n-1),u(n-1)} S_{u(n-1),v(n)}\Big].$$

Removing intervals of π recursively, we arrive precisely at the expansion of the moments of η-semicircular variable in terms of its free cumulants. □

Operator-valued Cauchy transform

Lemma 7.14. *Let $\mathcal{B}$ be a C^*-algebra. For $b \in \mathcal{B}$, write*

$$\Re b = \frac{1}{2}(b + b^*), \quad \Im b = \frac{1}{2i}(b - b^*),$$

which are both self-adjoint operators such that $b = \Re b + i\Im b$. Define the upper half-plane *of $\mathcal{B}$*

$$\mathbb{H}^+(\mathcal{B}) = \{b \in \mathcal{B} \mid \exists \varepsilon > 0 \text{ s.t. } \Im b - \varepsilon \text{ is a positive operator}\}.$$

Then every $b \in \mathbb{H}^+(\mathcal{B})$ is invertible.

Proof. Since $\Im b$ is self-adjoint and has spectrum in (ε, ∞), to prove that b is invertible it suffices to prove that $(\Im b)^{-1/2}(\Re b)(\Im b)^{-1/2} + i$ is invertible. Indeed, $(\Im b)^{-1/2}(\Re b)(\Im b)^{-1/2}$ is self-adjoint and so has real spectrum, therefore the spectrum of the operator above is in $\mathbb{R} + i$. So this operator is invertible. □

Definition 7.15. For a self-adjoint $a \in (\mathcal{A}, \Phi, \mathcal{B})$, we define its $\mathcal{B}$-valued Cauchy transform as a function on $\mathbb{H}^+(\mathcal{B})$

$$G_a(b) = b^{-1} + b^{-1}\Phi[a]b^{-1} + b^{-1}\Phi\left[ab^{-1}a\right]b^{-1} + \ldots = b^{-1}\left(\sum_{n=0}^{\infty} \Phi\left[(ab^{-1})^n\right]\right).$$

If $\mathcal{A}$ is a C^*-algebra (for example, $\mathcal{A} = M_N(L^\infty(\Omega, \Sigma))$) a is self-adjoint, we may define instead

$$G_a(b) = \Phi\left[(b-a)^{-1}\right].$$

If $\|a\| < \|b^{-1}\|^{-1}$, the series converges and the two definitions coincide.

Remark 7.16. Note that G_a provides information only about the moments of a of the form $\Phi[aba \ldots ba]$ and not about more general moments $\Phi[ab_1a \ldots b_{n-1}a]$, so G_a does not determine the $\mathcal{B}$-valued distribution of a. To obtain all moments, one needs to consider a *fully matricial* version of G_a; such functions have also been called non-commutative functions. This approach is an active area of research; we will not pursue these ideas further.

Proposition 7.17. *Suppose $a \in (\mathcal{A}, \Phi, \mathcal{B})$ is an η-semicircular element. Then*

$$G_a(b)^{-1} + \eta(G_a(b)) = b.$$

Proof. We need to show that

$$b^{-1} + b^{-1}\eta(G_a(b))G_a(b) = G_a(b).$$

Expanding both sides and replacing b^{-1} with b, we need to show that

$$b + b\eta\left(b\left(\sum_{i=0}^{\infty} \Phi\left[(ab)^i\right]\right)\right) b\left(\sum_{j=0}^{\infty} \Phi\left[(ab)^j\right]\right) = b\left(\sum_{n=0}^{\infty} \Phi\left[(ab)^n\right]\right).$$

Recalling that all the odd moments of a are zero, and cancelling b and the constant terms, this says

$$\eta\left(b\left(\sum_{i=0}^{\infty} \Phi\left[(ab)^{2i}\right]\right)\right) b\left(\sum_{j=0}^{\infty} \Phi\left[(ab)^{2j}\right]\right) = \left(\sum_{n=1}^{\infty} \Phi\left[(ab)^{2n}\right]\right).$$

Comparing coefficients, we need for every $n \geq 1$

$$\sum_{i=0}^{n-1} \eta\left(b\left(\Phi\left[(ab)^{2i}\right]\right)\right) b\Phi\left[(ab)^{2(n-i-1)}\right] = \left(\Phi\left[(ab)^{2n}\right]\right).$$

This follows from expanding all moments as sums over non-crossing pairs of free cumulants, and letting $i+1$ be the index of the element paired with 1 (compare with Exercise 6.24). ☐

Remark 7.18. Let $(\mathcal{A}, \varphi)$ be a ncps, and $x \in \mathcal{A}$ be self-adjoint. To compute the scalar-valued distribution of x, it is enough to compute its scalar-valued Cauchy transform G_x. Now suppose that there is a C^*-subalgebra $\mathcal{B} \subset \mathcal{A}$ and a conditional expectation $\Phi : \mathcal{A} \to \mathcal{B}$ which is φ-preserving: for any $a \in \mathcal{A}$, $\varphi[\Phi[a]] = \varphi[a]$. Consider a as an element of the probability space $(\mathcal{A}, \Phi, \mathcal{B})$, and denote $\psi = \varphi|_{\mathcal{B}}$, which is a state on $\mathcal{B}$ (ignoring continuity issues). Then we may also define the $\mathcal{B}$-valued Cauchy transform $\tilde{G}_x$ of x. It may be easier to compute than G_x. For example, we may have $x = a_1 + a_2$, where a_1 and a_2 are not free with respect to φ, but are free with respect to Φ. Then we can still find the scalar-valued distribution of x, by observing that $G_x(z) = \psi\left[\tilde{G}_x(z1_{\mathcal{B}})\right]$.

Example 7.19. Let X_N be a Gaussian band matrix as in Example 7.11. Let A_N be a diagonal $N \times N$ matrix such that

$$\sum_{i=1}^{N} (A_N)_{ii} \mathbf{1}_{[\frac{i-1}{N}, \frac{i}{N})} \to f \in L^\infty[0,1)$$

uniformly. What is the asymptotic empirical spectral distribution of $A_N + X_N$? The following computation is from (Shlyakhtenko 1996).

Let $b \in \mathcal{B} = L^\infty[0,1)$. Then the $\mathcal{B}$-valued Cauchy transform of asymptotic distribution of X_N satisfies

$$\frac{1}{\tilde{G}_s(b)} + \eta(\tilde{G}_s(b)) = b,$$

that is

$$\tilde{G}_s(b)(x) = (b(x) - \eta(\tilde{G}_s(b))(x))^{-1} = \left(b(x) - \int_0^1 \sigma(x,y)\tilde{G}_s(b)(y)\, dy\right)^{-1}.$$

Moreover, A_N and X_N are asymptotically $\mathcal{B}$-free (the proof is similar to those in the previous chapter). So their asymptotic free cumulant generating function is $f + \eta(b)$, that is

$$\frac{1}{\tilde{G}_{a+s}(b)} + f + \eta(\tilde{G}_{a+s}(b)) = b$$

and

$$\tilde{G}_{a+s}(b)(x) = (b(x) - f(x) - \eta(\tilde{G}_{a+s}(b))(x))^{-1} = \left(b(x) - f(x) - \int_0^1 \sigma(x,y)\tilde{G}_{a+s}(b)(y)\, dy\right)^{-1}.$$

Recall that ψ is the integration on $[0,1]$ with respect to the Lebesgue measure. Thus the Cauchy transform of the asymptotic (scalar-valued) distribution of $A_N + X_N$ is

$$G_{a+s}(z) = \int_0^1 \tilde{G}_{a+s}(z1)(x)\, dx.$$

Denote $a(x,z) = \tilde{G}_{a+s}(z1)(x)$. Then

$$G_{a+s}(z) = \int_0^1 a(x,z)\, dx,$$

where

$$a(x,z) = \left(z - f(x) - \int_0^1 \sigma(x,y) a(y,z)\, dy \right)^{-1}.$$

To compute the operator-valued Cauchy transform $G_{a_1+a_2}$ of a sum of two operator-free random variables more generally, we may use the operator-valued version of the subordination machinery. Denote

$$F_j(b) = F_{a_j}(b) = \left(G_{a_j}(b)\right)^{-1}, \quad H_j(b) = F_j(b) - b.$$

Theorem 7.20. *(Belinschi, Mai, Speicher 2017) Let a_1, a_2 be two $\mathcal{B}$-free self-adjoint elements in $(\mathcal{A}, \Phi, \mathcal{B})$. Then there exists a unique pair of Frechet analytic maps $\omega_1, \omega_2 : \mathbb{H}^+(\mathcal{B}) \to \mathbb{H}^+(\mathcal{B})$ such that for $b \in \mathbb{H}^+(\mathcal{B})$,*

$$\Im(\omega_j(b)) \geq \Im(b),$$

$$F_1(\omega_1(b)) + b = \omega_1(b) + \omega_2(b),$$

and

$$G_{a_1}(\omega_1(b)) = G_{a_2}(\omega_2(b)) = G_{a_1+a_2}(b).$$

Moreover, for a fixed b, $\omega_1(b)$ is the unique fixed point of the map

$$f_b : \mathbb{H}^+(\mathcal{B}) \to \mathbb{H}^+(\mathcal{B}), \quad f_b(c) = b + H_2(b + H_1(c)),$$

and is the limit of iterations $\lim_{n\to\infty} f_b^{\circ n}(c)$ for any initial $c \in \mathbb{H}^+(\mathcal{B})$.

7.3 Linearization trick

This section will likely be omitted in the course.

The following is a somewhat unexpected application of matrix-valued free probability to a fundamental problem in scalar-valued free probability. Let $a_1, a_2 \in (\mathcal{A}, \varphi)$ be free. We now know how to compute the distribution of $a_1 + a_2$ (if they are self-adjoint), and one can similarly compute the distribution of $a_1 a_2$ (if they are unitary or positive). But what about other functions of a_1, a_2? For example, it was a significant achievement of (Nica, Speicher 1998) to be able to compute the distribution of $i(a_1 a_2 - a_2 a_1)$. The machinery below can, in principle, be used to compute the distribution of any self-adjoint polynomial (or even rational) function of a_1, a_2. Using asymptotic freeness, this result can then be used to approximate the empirical spectral distribution of such a polynomial in random matrices. First we take a linear algebra digression.

Remark 7.21. Let $(\mathcal{A}, \varphi)$ be an ncps, and $M \in M_d(\mathcal{A})$. Let $U \in M_{1\times(d-1)}(\mathcal{A})$ be the first row of M with the first entry removed, $V \in M_{(d-1)\times 1}(\mathcal{A})$ the first column with the first entry removed, and $Q \in M_{d-1}(\mathcal{A})$ be M with the first row and columns removed, so that

$$M = \begin{pmatrix} M_{11} & U \\ V & Q \end{pmatrix}.$$

Then assuming all the inverses in the expression below exist,

$$\left((M^{-1})_{11}\right)^{-1} = M_{11} - UQ^{-1}V.$$

This is a consequence of the general Schur complement formula, see Lemma 4.6 and the discussion following it for the scalar version. Now let $a_1, a_2 \in (\mathcal{A}, \varphi)$ be self-adjoint and p be a self-adjoint polynomial in a_1, a_2. Suppose for some M as above, $p(a_1, a_2) = M_{11} - UQ^{-1}V$. Then

$$G_{p(a_1,a_2)}(z) = \varphi\left[(z - p(a_1, a_2))^{-1}\right] = \varphi\left[(z - (M_{11} - UQ^{-1}V))^{-1}\right] = \varphi\left[\left(\left(\begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} - M\right)^{-1}\right)_{11}\right]$$

Now define $\mathcal{B} = M_d(\mathbb{C})$, $\Phi = I \otimes \varphi$, so that we think of M as an element of $(M_d(\mathcal{A}), \Phi, \mathcal{B})$, and ψ a state on $\mathcal{B} = M_d(\mathbb{C})$ given by $\psi[A] = A_{11}$. Then the last expression above is

$$\psi \circ \Phi\left[\left(\begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} - M\right)^{-1}\right] = \psi\left[\tilde{G}_M(b_z)\right],$$

where $b_z \in M_d(\mathbb{C})$ is the matrix with the $(1, 1)$ entry z and zero elsewhere. Note that $b_z \notin \mathbb{H}^+(\mathcal{B})$, so to be more precise, in what follows we should use $b_z(\varepsilon) = b_z + i\varepsilon I$, and then let $\varepsilon \downarrow 0$.

Thus to compute the scalar-valued Cauchy transform $G_{p(a_1,a_2)}(z)$ it suffices to find the operator-valued Cauchy transform $\tilde{G}_M(b_z)$ for the appropriate M. The point is that M may be chosen so that $\tilde{G}_M$ can be found using free convolution.

The following idea has been developed by a number of people, notably (Haagerup, Thorbjørnsen 2005, Anderson 2013). Note that it has precursors as far back as (Higman 1940), and is still under development (Pisier 2018).

Proposition 7.22. *Let $p \in \mathbb{C}\langle x_1, \ldots, x_k\rangle$. Then p has a* linearization $\hat{p} \in M_d(\mathbb{C}\langle x_1, \ldots, x_k\rangle) = M_d(\mathbb{C}) \otimes \mathbb{C}\langle x_1, \ldots, x_k\rangle$ *(for some d) such that*

$$\hat{p} = \begin{pmatrix} 0 & U \\ V & Q \end{pmatrix},$$

$$p = -UQ^{-1}V,$$

and

$$\hat{p} = A_0 + A_1 \otimes x_1 + \ldots + A_k \otimes x_k,$$

where $A_0, \ldots, A_k \in M_d(\mathbb{C})$. If p is a self-adjoint polynomial, $\hat{p}$ may be chosen to be self-adjoint.

Therefore for $\{a_1, \ldots, a_k\} \in (\mathcal{A}, \varphi)$,

$$G_{p(a_1,\ldots,a_k)}(z) = \lim_{\varepsilon \downarrow 0}\left(\tilde{G}_{\hat{p}(a_1,\ldots,a_k)}(b_z(\varepsilon))\right)_{11}.$$

In the case when $a_1, \ldots, a_k$ are free in $(\mathcal{A}, \varphi)$, the latter $M_d(\mathbb{C})$-valued Cauchy transform may be computed using $M_d(\mathbb{C})$-valued free convolution and $M_d(\mathbb{C})$-valued subordination functions.

Exercise 7.23. The proof is outlined in Exercise 10.3.1 in (Mingo, Speicher 2017). Prove the following properties.

Linearizations of monomials: x_i has a linearization

$$\hat{p} = \begin{pmatrix} 0 & x_i \\ 1 & -1 \end{pmatrix},$$

while $p = x_{i(1)} x_{i(2)} \dots x_{i(n)}$ has a linearization

$$\hat{p} = \begin{pmatrix} & & & x_{i(1)} \\ & & x_{i(2)} & -1 \\ & \iddots & \iddots & \\ x_{i(n)} & -1 & & \end{pmatrix}.$$

Linearization of sums: if each p_i has a linearization

$$\hat{p}_i = \begin{pmatrix} 0 & U_i \\ V_i & Q_i \end{pmatrix},$$

then $p = p_1 + \dots + p_k$ has a linearization

$$\hat{p} = \begin{pmatrix} 0 & U_1 & \dots & U_k \\ V_1 & Q_1 & & \\ \vdots & & \ddots & \\ V_k & & Q_k & \end{pmatrix}.$$

Self-adjoint linearization: if p has a linearization

$$\hat{p} = \begin{pmatrix} 0 & U \\ V & Q \end{pmatrix},$$

then $p + p^*$ has a linearization

$$\begin{pmatrix} 0 & U & V^* \\ U^* & 0 & Q^* \\ V & Q & 0 \end{pmatrix}.$$

Example 7.24. Consider $a_1^2 a_2 + a_1 a_2 + a_1 a_2^2$. The the summands have linearizations

$$\begin{pmatrix} 0 & 0 & a_1 \\ 0 & a_1 & -1 \\ a_2 & -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & a_1 \\ a_2 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & a_1 \\ 0 & a_2 & -1 \\ a_2 & -1 & 0 \end{pmatrix},$$

and the non-self-adjoint version of $\hat{p}$ is

$$\hat{p}(a_1, a_2) = \begin{pmatrix} 0 & 0 & a_1 & a_1 & 0 & a_1 \\ 0 & a_1 & -1 & 0 & 0 & 0 \\ a_2 & -1 & 0 & 0 & 0 & 0 \\ a_2 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_2 & -1 \\ a_2 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} a_1 + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} a_2.$$

Chapter 8

Spiked models, subordination, and infinitesimal freeness.

Motivating question. Suppose B_N is as in the beginning of Chapter 6, but A_N is a diagonal matrix with 5 non-zero eigenvalues (independently of N). Then clearly $\hat{\mu}_{A_N} \to \delta_0$ and $\hat{\mu}_{A_N+B_N} \to \mu$. Nevertheless, it turns out that one may identify certain eigenvalues of $A_N + B_N$ as coming from those of A_N. They can be studied using subordination functions and infinitesimal freeness.

Reference: (Shlyakhtenko 2015) and earlier work by Capitaine, Belinschi, Bercovici, Fevrier.

8.1 Random matrix results

The following type of results go back to (Baik, Ben Arous, Péché 2005). For a Hermitian matrix X_N, we denote its eigenvalues

$$\lambda_1(X_N) \geq \lambda_2(X_N) \geq \ldots \geq \lambda_N(X_N).$$

Note that this notation is the opposite of the one used in the early chapters.

The following is a particular case of a theorem from (Péché 2006).

Theorem 8.1. *Let X_N be an $N \times N$ GUE matrix. Let θ be a fixed real number, and P_N be a Hermitian rank 1 matrix with non-zero eigenvalue θ. Let*

$$\tilde{X}_N = X_N + P_N.$$

Then we have the following dichotomy. If $\theta \leq 1$, then almost surely

$$\lambda_1(\tilde{X}_N) \to 2,$$

the supremum of the support of the standard semicircular distribution. If $\theta > 1$, then almost surely

$$\lambda_1(\tilde{X}_N) \to \theta + \frac{1}{\theta}.$$

A similar result holds for the smallest eigenvalue.

The following is a particular case of a theorem from (Benaych-Georges, Nadakuditi 2011).

Theorem 8.2. *Let X_N be an $N \times N$ symmetric orthogonally invariant (or Hermitian unitarily invariant) random matrix such that its empirical spectral distribution $\hat{\mu}_{X_N}$ converges weakly almost surely to a compactly supported measure μ. Let $a = \inf \operatorname{supp}(\mu)$ and $b = \sup \operatorname{supp}(\mu)$, and assume that the smallest and largest eigenvalues of X_N converge almost surely to a and b.*

Let $\theta_1 \geq \ldots \geq \theta_s > 0 > \theta_{s+1} \geq \ldots \geq \theta_r$ be fixed real numbers (independently of N), and P_N an $N \times N$ symmetric (or Hermitian) matrix of rank r with eigenvalues $(\theta_i)_{i=1}^r$. Let

$$\tilde{X}_N = X_N + P_N.$$

Then for all $1 \leq i \leq s$, almost surely

$$\lambda_i(\tilde{X}_N) \to \begin{cases} F_\mu^{-1}(\theta_i) & \text{if } \theta_i > F_\mu(b^+), \\ b & \text{otherwise}. \end{cases}$$

A similar result holds for the smallest eigenvalues.

The following is a particular case of a theorem from (Belinschi, Bercovici, Capitaine, Février 2017).

Theorem 8.3. *Let μ, ν be compactly supported probability measures on $\mathbb{R}$, $\Theta = \{\theta_1 \geq \theta_2 \geq \ldots \theta_p\}$ fixed numbers outside of the support of μ, and $T = \{\tau_1 \geq \tau_2 \geq \ldots \geq \tau_q\}$ fixed numbers outside of the support of ν. Let A_N, B_N be deterministic Hermitian matrices such that $\hat{\mu}_{A_N} \to \mu$ and $\hat{\mu}_{B_N} \to \nu$ weakly. Assume also that for each $N \geq p$ and $\theta \in \{\theta_1, \ldots, \theta_p\}$,*

$$|\{n : \lambda_n(A_N) = \theta\}| = |\{i : \theta_i = \theta\}|$$

while the eigenvalues of A_N not in this set converge uniformly to $\operatorname{supp}(\mu)$:

$$\lim_{N \to \infty} \max_{\lambda_n(A_N) \notin \Theta} \operatorname{dist}(\lambda_n(A_N), \operatorname{supp}(\mu)) = 0.$$

Make the same assumption concerning B_N, T, and ν. Finally, let U_N be a CUE matrix, and denote

$$X_N = A_N + U_N^* B_N U_N.$$

Let ω_1, ω_2 be the subordination functions for $\mu, \nu, \mu \boxplus \nu$, $K = \operatorname{supp}(\mu \boxplus \nu)$,

$$K' = K \cup \omega_1^{-1}(\Theta) \cup \omega_2^{-1}(T),$$

and K'_ε its ε-neighborhood. Then for any given $\varepsilon > 0$, almost surely for large N, all the eigenvalues of X_N lie in K'_ε. Moreover, let $\rho \in K' \setminus K$, and choose $\varepsilon > 0$ so that $(\rho - 2\varepsilon, \rho + 2\varepsilon) \cap K' = \{\rho\}$. Then the number of eigenvalues of X_N in $(\rho - \varepsilon, \rho + \varepsilon)$ equals

$$|\{i : \omega_1(\rho) = \theta_i\}| + |\{j : \omega_2(\rho) = \tau_j\}|.$$

These theorems can be proved using Stieltjes transform methods related to those in Chapter 4. We take a different route, which will lead to weaker but equally interesting conclusions. First however we will follow (Benaych-Georges, Nadakuditi 2011) and give an outline of a proof for the following version of the result.

Proposition 8.4. *Let X_N be an $N \times N$ Hermitian unitarily invariant random matrix such that its empirical spectral distribution $\hat{\mu}_{X_N}$ converges weakly almost surely to a compactly supported measure μ. Let $a = \inf \operatorname{supp}(\mu)$ and $b = \sup \operatorname{supp}(\mu)$, and assume that the smallest and largest eigenvalues of X_N converge almost surely to a and b.*

Let θ be a fixed real number, and P_N an $N \times N$ Hermitian rank 1 matrix with non-zero eigenvalue θ.Let

$$\tilde{X}_N = X_N + P_N.$$

Then almost surely,

$$\lambda_1(\tilde{X}_N) \to \begin{cases} F_\mu^{-1}(\theta) & \text{if } \theta > F_\mu(b^+), \\ b & \text{otherwise}. \end{cases}$$

Outline of proof. Instead of assuming unitary invariance of X_N, we may assume instead that X_N is a non-random diagonal matrix with eigenvalues $\{\lambda_i\}_{i=1}^N$ but $P_N = U_N^* Q_N U_N$, where $(Q_N)_{ij} = \delta_{i=1}\delta_{j=1}\theta$ and U_N is CUE. It is easy to see that equivalently, $P_N = \theta u_N u_N^*$, where u_N is a vector uniformly distributed on the sphere in $\mathbb{C}^N$.

The eigenvalues of $X_N + P_N$ are solutions of the equation

$$\det(zI - (X_N + P_N)) = 0.$$

Since

$$\det(zI - (X_N + P_N)) = \det(zI - X_N)\det(I - (zI - X_N)^{-1}P_N),$$

It follows that z is an eigenvalue of $X_N + P_N$ and *not* an eigenvalue of X_N if and only if 1 is an eigenvalue of $(zI - X_N)^{-1}P_N$. Now $(zI - X_N)^{-1}P_N = (zI - X_N)^{-1}\theta u_N u_N^*$ has rank 1. So its only non-zero eigenvalue equals its trace

$$\operatorname{Tr}\left[(zI - X_N)^{-1}\theta u_N u_N^*\right] = \theta \sum_{i=1}^N (u_N)_i (z - \lambda_i)^{-1}\bar{u}_i = \theta \sum_{i=1}^N \frac{|(u_N)_i|^2}{z - \lambda_i} = \theta G_{\mu_N}(z),$$

where

$$\mu_N = \sum_{i=1}^N |(u_N)_i|^2 \delta_{\lambda_i}$$

(compare with Exercise 5.16 and the remark following it). Thus z outside of the spectrum of X_N is an eigenvalue of $X_N + P_N$ if and only if

$$\sum_{i=1}^N \frac{|(u_N)_i|^2}{z - \lambda_i} = G_{\mu_N}(z) = \frac{1}{\theta}.$$

Next we compare μ_N with

$$\hat{\mu}_{X_N} = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i}.$$

It is clear from symmetry that $\mathbb{E}\left[|(u_N)_i|^2\right] = \frac{1}{N}$. By using appropriate concentration of measure results (compare with Chapter 3), it follows that for large N, (morally) $|(u_N)|^2 \approx \frac{1}{N}$ and $\mu_N \approx \hat{\mu}_{X_N} \approx \mu$, and (precisely) $G_{\mu_n} \approx G_{\hat{\mu}_{X_N}} \approx G_\mu$. Thus the outlier eigenvalue of $X_N + P_N$, if any, is an approximate solution, if any, of the equation $G_\mu(z) = \frac{1}{\theta}$, or equivalently of $F_\mu(z) = \theta$. The dichotomy in the statement of the proposition is explained by the following lemma. □

Lemma 8.5. *Let μ be a compactly supported probability measure on $\mathbb{R}$ and $a = \inf\operatorname{supp}(\mu)$, $b = \sup\operatorname{supp}(\mu)$. Then F_μ is analytic on $\mathbb{R}\setminus[a,b]$, and*

$$F_\mu(a^-) = \lim_{z\uparrow a} F_\mu(z) \leq 0 \textit{ and } F_\mu(b^+) = \lim_{z\downarrow b} F_\mu(z) \geq 0$$

are well defined. Moreover, F_μ is an increasing homeomorphism from $(-\infty, a)$ onto $(-\infty, F_\mu(a^-))$ and from $(b, +\infty)$ onto $(F_\mu(b^+), +\infty)$.

Proof. It suffices to prove the corresponding statements for $G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-u}\, d\mu(u)$. The analyticity of G_μ on the complement of $\operatorname{supp}(\mu)$ is clear. $G_\mu(x) < 0$ for $x < a$ and $G_\mu(x) > 0$ for $x > b$, so $F_\mu(x)$ is defined on these intervals. For $x \in \mathbb{R}\setminus\operatorname{supp}(\mu)$,

$$G'_\mu(x) = -\int_{\mathbb{R}} \frac{1}{(x-u)^2}\, d\mu(u) \leq 0,$$

so G_μ decreases on each interval in $\mathbb{R}\setminus\operatorname{supp}(\mu)$. It follows that F_μ increases on each such interval. Therefore the limits defining $F_\mu(a^-)$ and $F_\mu(b^+)$ exist. □

8.2 Free convolution computations

Exercise 8.6. Show that if $A_t(z) = A(z) + t\tilde{A}(z) + o(t)$, then its inverse with respect to composition is

$$A_t^{-1}(w) = A^{-1}(w) - t\tilde{A}(A^{-1}(w))(A^{-1})'(w) + o(t).$$

Feel free to make the question more precise.

Proposition 8.7. *Let $\{\mu_t, \nu_t\}_{t\in[0,\varepsilon]}$ be probability measures such that*

$$\mu_t = \mu + t\tilde{\mu} + o(t), \quad \nu_t = \nu + t\tilde{\nu} + o(t).$$

Here μ, ν are probability measures, and $\tilde{\mu}, \tilde{\nu}$ are signed measures of total weight 0. Let ω_1, ω_2 be the subordination functions for $\mu, \nu, \mu \boxplus \nu$. Then

$$G_{\mu_t\boxplus\nu_t} = G_{\mu\boxplus\nu} + t[(G_{\tilde{\mu}}\circ\omega_1)\cdot\omega_1' + (G_{\tilde{\nu}}\circ\omega_2)\cdot\omega_2'] + o(t).$$

Note that the second term is, in general, not a Cauchy transform of a signed measure, but only of a distribution.

Proof. We compute

$$\frac{1}{w} + R_{\mu_t}(w) = \frac{1}{w} + R_\mu(w) - tG_{\tilde{\mu}}(G_\mu^{-1}(w))(G_\mu^{-1})'(w) + o(t)$$

and the same for ν_t. Adding the expressions, we get

$$\frac{1}{w} + R_{\mu_t \boxplus \nu_t}(w) = \frac{1}{w} + R_{\mu\boxplus\nu} - t[G_{\tilde{\mu}}(G_\mu^{-1}(w))(G_\mu^{-1})'(w) + G_{\tilde{\nu}}(G_\nu^{-1}(w))(G_\nu^{-1})'(w)] + o(t).$$

Inverting, we get

$$\begin{aligned} G_{\mu_t\boxplus\nu_t} &= G_{\mu\boxplus\nu} + t[G_{\tilde{\mu}} \circ G_\mu^{-1} \cdot (G_\mu^{-1})' + G_{\tilde{\nu}} \circ G_\nu^{-1} \cdot (G_\nu^{-1})'] \circ G_{\mu\boxplus\nu} \cdot G'_{\mu\boxplus\nu} + o(t) \\ &= G_{\mu\boxplus\nu} + t[(G_{\tilde{\mu}} \circ \omega_1) \cdot \omega_1' + (G_{\tilde{\nu}} \circ \omega_2) \cdot \omega_2'] + o(t) \end{aligned}$$

□

Lemma 8.8. *The size of the atom of a measure ρ at x is*

$$\rho(\{x\}) = \lim_{y\downarrow 0} iyG_\rho(x+iy).$$

Proof. Clearly if suffices to consider the atom at $x = 0$. Also, since

$$\lim_{y\downarrow 0} iyG_{\alpha\rho+\beta\delta_0}(iy) = \alpha \lim_{y\downarrow 0} iyG_\rho(iy) + \lim_{y\downarrow 0} iy\beta\frac{1}{(0+iy)-0} = \alpha \lim_{y\downarrow 0} iyG_\rho(iy) + \beta,$$

if suffices to consider the case when $\rho(\{0\}) = 0$. In this case,

$$\lim_{y\downarrow 0} iyG_\rho(iy) = \lim_{y\downarrow 0} \int_{\mathbb{R}} \frac{iy}{iy-u}\, d\rho(u).$$

As $y \downarrow 0$, $\frac{iy}{iy-u} \to 0$ for ρ-almost every u (since ρ has no atom at 0). Also, $\left|\frac{iy}{iy-u}\right| \leq 1$. So by the dominated convergence theorem, the limit above is zero. □

Corollary 8.9. *If $\tilde{\mu}$ (resp., $\tilde{\nu}$) has an atom at $\omega_1(a)$ (resp., $\omega_2(a)$) of size α then $\mu_t \boxplus \nu_t$ has an atom at a of size*

$$(\mu \boxplus \nu)(\{a\}) + t\alpha + o(t).$$

Proof. We compute

$$\lim_{y\downarrow 0} iyG_{\tilde{\mu}}(\omega_1(a+iy))\omega_1'(a+iy) = \lim_{y\downarrow 0} iyG_{\tilde{\mu}}(\omega_1(a) + iy\omega_1'(a)))\omega_1'(a+iy) = \lim_{y\downarrow 0} iyG_{\tilde{\mu}}(\omega_1(a)+iy).$$

Now apply the preceding lemma. □

Example 8.10. In the preceding proposition, let $\mu_t = \mu$ and $\nu_t = \delta_0 + t\tilde{\nu}$. Then

$$G_{\mu_t \boxplus \nu_t} = G_\mu + t(G_{\tilde{\nu}} \circ F_\mu) \cdot F'_\mu + o(t).$$

Next, suppose that $\alpha_0 = \sum_{j=1}^n \alpha_j$ and

$$\tilde{\nu} = \sum_{j=1}^n \alpha_j \delta_{\theta_j} - \alpha_0 \delta_0$$

is a purely atomic measure. Then

$$(G_{\tilde{\nu}} \circ F_\mu) \cdot F'_\mu = \sum_{j=1}^n \alpha_j \frac{F'_\mu(z)}{F_\mu(z) - \theta_j} - \alpha_0 \frac{F'_\mu(z)}{F_\mu(z)}.$$

The atoms of this measure are located at the points λ_ℓ satisfying $F_\mu(\lambda_\ell) = \theta_\ell$ (note that the solution need not be unique), with the weight α_ℓ.

If the support of μ

$$\text{supp}(\mu) = [a, b]$$

is connected, then the image of F_μ is $(-\infty, F_\mu(a^-)) \cup (F_\mu(b^+), \infty)$. If θ_ℓ lies in this set, it has a unique pre-image; if it lies in $(F_\mu(a^-), F_\mu(b^+))$, it has no pre-image.

Finally, restricting to the case $\mu = \sigma$, the semicircular distribution, and $\tilde{\nu} = \delta_\theta - \delta_0$, we get the atoms at

$$\lambda_\ell = \theta_\ell + \frac{1}{\theta_\ell}$$

for θ_ℓ in the complement of $(-1, 1)$.

8.3 Infinitesimal freeness

It thus appears that random matrices in the theorems at the beginning of the chapter are "asymptotically free up to order $o(\frac{1}{N})$". The following definition formalizes this notion.

Definition 8.11. An *infinitesimal non-commutative probability space* is a triple $(\mathcal{A}, \varphi, \varphi')$, where $\mathcal{A}$ is a $*$-algebra, φ is a linear functional such that $\varphi[1] = 1$ (and, perhaps, a state), and φ' is a linear functional such that $\varphi'[1] = 0$.

Subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ are *infinitesimally free* in $(\mathcal{A}, \varphi, \varphi')$ if for any $a_i \in \mathcal{A}_{u(i)}$, $u(1) \neq u(2)$, $u(2) \neq u(3)$, ... such that $\varphi[a_1] = \ldots = \varphi[a_n] = 0$, we have

$$\varphi[a_1 \ldots a_n] = 0$$

(thus they are free in $(\mathcal{A}, \varphi)$) and

$$\varphi'(a_1 \ldots a_n) = \sum_{j=1}^n \varphi[a_1 \ldots a_{j-1} \varphi'[a_j] a_{j+1} \ldots a_n]. \tag{8.1}$$

Remark 8.12. Let $t \geq 0$ and on $(\mathcal{A}, \varphi, \varphi')$, define $\varphi_t = \varphi + t\varphi'$. Then

$$\varphi_t \left[\prod_{i=1}^{n} (a_i - \varphi_t[a_i]) \right] = (\varphi + t\varphi') \left[\prod_{i=1}^{n} (a_i - \varphi[a_i] - t\varphi'[a_i]) \right]$$
$$= \varphi \left[\prod_{i=1}^{n} (a_i - \varphi[a_i]) \right]$$
$$+ t \left(\varphi' \left[\prod_{i=1}^{n} (a_i - \varphi[a_i]) \right] - \sum_{j=1}^{n} \varphi \left[\prod_{i<j} (a_i - \varphi[a_i]) \varphi'[a_j] \prod_{i>j} (a_i - \varphi[a_i]) \right] \right) + o(t).$$

Using the observation that $\varphi'[\varphi[a]] = 0$, this equals

$$\varphi \left[\prod_{i=1}^{n} (a_i - \varphi[a_i]) \right] + t \left(\varphi' \left[\prod_{i=1}^{n} (a_i - \varphi[a_i]) \right] - \sum_{j=1}^{n} \varphi \left[\prod_{i<j} (a_i - \varphi[a_i]) \varphi'[a_j - \varphi[a_j]] \prod_{i>j} (a_i - \varphi[a_i]) \right] \right) + o(t).$$

Thus $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ are infinitesimally free if and only if they are free in $(\mathcal{A}, \varphi_t)$ up to $o(t)$.

Lemma 8.13. *Let $(\mathcal{A}, \varphi, \varphi')$ be an infinitesimal ncps such that both φ and φ' are tracial. Let $\mathcal{E} \subset \mathcal{A}$ be a non-unital subalgebra such that $\varphi|_{\mathcal{E}} = 0$. Then a subalgebra $\mathcal{B} \subset \mathcal{A}$ and $\mathcal{E}$ are infinitesimally free if and only if for any $E_j \in \mathcal{E}$, and any $Q_j \in \mathcal{B}$ such that $\varphi[Q_j] = 0$, we have*

$$\varphi \left[E_1 Q_1 \ldots E_n Q_n \right] = 0$$

and

$$\varphi' \left[E_1 Q_1 \ldots E_n Q_n \right] = 0.$$

Proof. Since both infinitesimal freeness and these conditions determine all the joint moments, it suffices to prove one direction. So assume infinitesimal freeness. The first condition in the lemma follows directly from the definition of freeness with respect to φ. For the second, from the definition of infinitesimal freeness we will have two types of terms. For the first one,

$$\varphi \left[E_1 Q_1 \ldots E_k \varphi'[Q_k] E_{k+1} \ldots E_n Q_n \right] = \varphi'[Q_k] \varphi \left[E_1 Q_1 \ldots (E_k E_{k+1}) \ldots E_n Q_n \right] = 0$$

by freeness for $n \geq 2$. For the second one,

$$\varphi \left[E_1 Q_1 \ldots Q_{k-1} \varphi'[E_k] Q_k \ldots E_n Q_n \right] = \varphi'[E_k] \varphi \left[E_1 Q_1 \ldots \left(Q_{k-1} Q_k - \varphi \left[Q_{k-1} Q_k \right] \right) \ldots E_n Q_n \right]$$
$$+ \varphi'[E_k] \varphi \left[Q_{k-1} Q_k \right] \varphi \left[E_1 Q_1 \ldots (E_{k-1} E_{k+1}) \ldots E_n Q_n \right] = 0$$

by freeness for $n \geq 3$. The particular cases for $n \leq 3$ are easy to verify. □

Shlyakhtenko proved asymptotic infinitesimal freeness in the following setting. Note that it only covers the (Benaych-Georges, Nadakuditi) theorem, while the computation above suggests it holds also in the (Belinschi, Bercovici, Capitaine, Février) setting. Note also that our setup and proofs are a little different.

Theorem 8.14. *Let X_N be a random $N \times N$ matrix, either normalized GUE or a unitarily invariant matrix $X_N = U_N A_N U_N^*$ with $\hat{\mu}_{A_N} \to \mu_a$. Fix N_0, and denote $\mathcal{E} = M_{N_0}(\mathbb{C})$, considered as a subalgebra of any $M_N(\mathbb{C})$ for $N \geq N_0$. Denote $\mathcal{A} = \mathcal{E}\langle x \rangle$ a formal algebra generated by $\mathcal{E}$ and a self-adjoint symbol x. On $\mathcal{A}$, define*

$$\varphi\left[P(x)\right] = \lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\left[P(X_N)\right]$$

and

$$\varphi'\left[P(x)\right] = \lim_{N \to \infty} \Big(\operatorname{Tr}\left[P(X_N)\right] - N\varphi\left[P(x)\right] \Big),$$

so that

$$\frac{1}{N} \operatorname{Tr}\left[P(X_N)\right] = \varphi\left[P(x)\right] + \frac{1}{N}\varphi'\left[P(x)\right] + o(\frac{1}{N}).$$

Then both limits exist, and in $(\mathcal{A}, \varphi, \varphi')$, x and $\mathcal{E}$ are infinitesimally free.

Proof. Note that if they are well-defined, φ, φ' are tracial, and $\varphi|_{\mathcal{E}} = 0$. To conclude infinitesimal freeness, it remains to verify the two identities from the lemma.

First consider the case $X_N = U_N A_N U_N^*$. Denote $A^{(j)} = f_j(A_N)$, so that $f_j(X_N) = U_N A^{(j)} U_N^*$, and $E^{(j)} \in \mathcal{E}$. Then

$$\begin{aligned}
&\mathbb{E} \circ \operatorname{Tr}\left[f_1(X_N) E^{(1)} f_2(X_N) E^{(2)} \ldots f_n(X_N) E^{(n)}\right] \\
&\quad = \mathbb{E} \circ \operatorname{Tr}\left[U A^{(1)} U^* E^{(1)} U A^{(2)} U^* E^{(2)} \ldots U A^{(n)} U^* E^{(n)}\right] \\
&\quad = \sum_{\alpha, \beta \in S_n} \operatorname{Wg}_n^N(\beta\alpha^{-1}) \operatorname{Tr}_\alpha\left[A^{(1)}, \ldots, A^{(n)}\right] \operatorname{Tr}_{\beta^{-1}\gamma}\left[E^{(1)}, \ldots, E^{(n)}\right] \\
&\quad = \sum_{\alpha, \beta \in S_n} \mu(\beta\alpha^{-1}) \frac{1}{N^{2n - |\beta\alpha^{-1}| - |\alpha| - |\beta^{-1}\gamma|}} (1 + o(1)) \varphi_\alpha\left[f_1(a), \ldots, f_n(a)\right] \\
&\qquad \left(\frac{1}{N^{|\beta^{-1}\gamma|}} \operatorname{Tr}_{\beta^{-1}\gamma}\left[E^{(1)}, \ldots, E^{(n)}\right] \right) \\
&\quad = \sum_{\alpha, \beta \in S_n} \mu(\beta\alpha^{-1}) \frac{N}{N^{d(e,\alpha) + d(\alpha,\beta) + d(\beta,\gamma) - d(e,\gamma)}} (1 + o(1)) \varphi_\alpha\left[f_1(a), \ldots, f_n(a)\right] \\
&\qquad \left(\frac{1}{N^{|\beta^{-1}\gamma|}} \operatorname{Tr}_{\beta^{-1}\gamma}\left[E^{(1)}, \ldots, E^{(n)}\right] \right).
\end{aligned}$$

The only terms with non-zero limit correspond to $\alpha = P_\sigma$, $\beta = P_\pi$, $\sigma \leq \pi$, and $\left|P_{K[\pi]}\right| = |\beta^{-1}\gamma| = 1$. Thus $K[\pi] = \{(12 \ldots n)\}$ and $\pi = \sigma = \{(1), (2), \ldots, (n)\}$, so that the limit is

$$\prod_{i=1}^{n} \varphi\left[f_i(a)\right] \operatorname{Tr}\left[E^{(1)} \ldots E^{(n)}\right]$$

Thus φ on such monomials is always zero, and φ' is zero provided some $\varphi[f_i(a)] = 0$.

Now we consider the case when X_N is GUE. We will first prove that for f_j as above,

$$\mathbb{E} \circ \operatorname{Tr}\left[f_1(X_N)E^{(1)}f_2(X_N)E^{(2)}\dots f_n(X_N)E^{(n)}\right] \to \prod_{i=1}^{n} \varphi\left[f_i(s)\right] \operatorname{Tr}\left[E^{(1)}\dots E^{(n)}\right],$$

where s is standard semicircular. It suffices to prove the result for $f_j(x) = x^{p(j)}$. It then suffices to consider the following monomial, where $\tilde{E}^{(i)} \in \mathcal{E}$ for $i \in S = \{p(1), p(1)+p(2), \dots, \sum_{i=1}^{n} p(i) = 2k\}$, and $\tilde{E}^{(i)} = I_N$ otherwise:

$$\mathbb{E} \circ \operatorname{Tr}\left[X\tilde{E}^{(1)}X\tilde{E}^{(2)}\dots X\tilde{E}^{(2k)}\right] = \sum_{\pi \in \mathcal{P}_2(2k)} \frac{N}{N^{d(e,\pi)+d(\pi,\gamma)-d(e,\gamma)}} \frac{1}{N^{|\pi\gamma|}} \operatorname{Tr}_{\pi\gamma}\left[\tilde{E}^{(1)}, \tilde{E}^{(2)}, \dots, \tilde{E}^{(2k)}\right],$$

Again the only asymptotically non-zero terms arise from non-crossing pairings, so we may replace the expression above by

$$\sum_{\pi \in \mathcal{NC}_2(2k)} \frac{N}{N^{|K[\pi]|}} \operatorname{Tr}_{K[\pi]}\left[\tilde{E}^{(1)}, \tilde{E}^{(2)}, \dots, \tilde{E}^{(2k)}\right].$$

If a block of $K[\pi]$ contains an element of S, the corresponding trace contains an element of $\mathcal{E}$, and does not grow with N; if it is contained in S^c, the corresponding trace is the trace of I_N, and equals to N. So the power of $\frac{1}{N}$ in the expression above is $|\{V \in K[\pi] : V \cap S \neq \emptyset\}| - 1$. Thus the only asymptotically non-zero terms correspond to $\pi \in \mathcal{NC}_2(2n)$ such that S is a subset of a single block of $K[\pi]$. Equivalently,

$$\pi \leq \{(1, \dots, p(1)), (p(1)+1, \dots, p(1)+p(2)), \dots, \}$$

We conclude that

$$\begin{aligned}
\mathbb{E} \circ \operatorname{Tr}\left[X^{p(1)}E^{(1)}X^{p(2)}E^{(2)}\dots X^{p(n)}E^{(n)}\right] \to & \sum_{\substack{\pi \in \mathcal{NC}_2(2k) \\ \pi \leq \{(1,\dots,p(1)),(p(1)+1,\dots,p(1)+p(2)),\dots,\}}} \operatorname{Tr}\left[E^{(1)}\dots E^{(n)}\right] \\
&= \prod_{i=1}^{k} c_{p(i)/2} \operatorname{Tr}\left[E^{(1)}\dots E^{(n)}\right] \\
&= \prod_{i=1}^{n} \varphi\left[s^{p(i)}\right] \operatorname{Tr}\left[E^{(1)}\dots E^{(n)}\right],
\end{aligned}$$

where s is standard semicircular. The rest of the argument is as before. □

Remark 8.15. Let A_N, B_N be $N \times N$ (non-random) Hermitian matrices such that A_N and B_N converge in infinitesimal distribution as $N \to \infty$. That is, for some $a, b \in (\mathcal{A}, \varphi, \varphi')$,

$$\frac{1}{N} \operatorname{Tr}[A_N^k] \to \varphi[a^k], \quad \frac{1}{N} \operatorname{Tr}[B_N^k] \to \varphi[b^k]$$

and

$$\operatorname{Tr}[A_N^k] - N\varphi[a^k] \to \varphi'[a^k], \quad \operatorname{Tr}[B_N^k] - N\varphi[b^k] \to \varphi'[b^k].$$

Let U_N be an $N \times N$ CUE (Haar unitary) matrix. We know that $U_N A_N U_N^*$ and B_N are asymptotically free. Are they asymptotically infinitesimally free, that is, do their joint moments converge to those of a, b above which are infinitesimally free?

As before, choose (polynomial) functions f_j, g_j with $\varphi[f_j(a)] = \varphi[g_j(b)] = 0$, and denote $A^{(i)} = f_i(A_N)$, $B^{(i)} = g_i(B_N)$. Then

$$\begin{aligned}
&\frac{1}{N}\,\mathbb{E}\circ\mathrm{Tr}\left[f_1(U_N A_N U_N^*)g_1(B_N)\ldots g_{n-1}(B_N)f_n(U_N A_N U_N^*)g_n(B_n)\right] \\
&\quad= \frac{1}{N}\,\mathbb{E}\circ\mathrm{Tr}\left[UA^{(1)}U^*B^{(1)}UA^{(2)}U^*B^{(2)}\ldots UA^{(n)}U^*B^{(n)}\right] \\
&\quad= \frac{1}{N}\sum_{\alpha,\beta\in S_n}\mathrm{Wg}_n^N(\beta\alpha^{-1})\,\mathrm{Tr}_\alpha\left[A^{(1)},\ldots,A^{(n)}\right]\mathrm{Tr}_{\beta^{-1}\gamma}\left[B^{(1)},\ldots,B^{(n)}\right] \\
&\quad= \sum_{\alpha,\beta\in S_n}\mu(\beta\alpha^{-1})\frac{1}{N^{d(e,\alpha)+d(\alpha,\beta)+d(\beta,\gamma)-d(e,\gamma)}}\left(1+o(\frac{1}{N})\right) \\
&\qquad\quad \left(\frac{1}{N^{|\alpha|}}\,\mathrm{Tr}_\alpha\left[A^{(1)},\ldots,A^{(n)}\right]\right)\left(\frac{1}{N^{|\beta^{-1}\gamma|}}\,\mathrm{Tr}_{\beta^{-1}\gamma}\left[B^{(1)},\ldots,B^{(n)}\right]\right) \\
&\quad= \sum_{\alpha,\beta\in S_n}\mu(\beta\alpha^{-1})\frac{1}{N^{d(e,\alpha)+d(\alpha,\beta)+d(\beta,\gamma)-d(e,\gamma)}}\left(1+o(\frac{1}{N})\right) \\
&\qquad\quad \prod_{V\in\mathrm{cyc}(\alpha)}\left(\varphi[f_i(a):i\in V]+\frac{1}{N}\varphi'[f_i(a):i\in V]\right)\prod_{U\in\mathrm{cyc}(\beta^{-1}\gamma)}\left(\varphi[g_j(b):j\in U]+\frac{1}{N}\varphi'[g_j(b):j\in U]\right).
\end{aligned}$$

It is easy to see from the Cut-and-Join Lemma that if $d(e,\alpha)+d(\alpha,\beta)+d(\beta,\gamma)-d(e,\gamma)>0$ then it is at least 2. So the expression above is

$$\begin{aligned}
&\sum_{\substack{\sigma,\pi\in\mathcal{NC}(n)\\ \sigma\le\pi}}\mu(P_\pi P_\sigma^{-1})\prod_{V\in\sigma}\varphi[f_i(a):i\in V]\prod_{U\in K[\pi]}\varphi[g_j(b):j\in U] \\
&\quad+\frac{1}{N}\sum_{\substack{\sigma,\pi\in\mathcal{NC}(n)\\ \sigma\le\pi}}\mu(P_\pi P_\sigma^{-1})\sum_{\tilde V\in\sigma}\varphi'[f_i(a):i\in\tilde V]\prod_{\substack{V\in\sigma\\ V\neq\tilde V}}\varphi[f_i(a):i\in V]\prod_{U\in K[\pi]}\varphi[g_j(b):j\in U] \\
&\quad+\frac{1}{N}\sum_{\substack{\sigma,\pi\in\mathcal{NC}(n)\\ \sigma\le\pi}}\mu(P_\pi P_\sigma^{-1})\prod_{V\in\sigma}\varphi[f_i(a):i\in V]\sum_{\tilde U\in K[\pi]}\varphi'[f_i(a):i\in\tilde U]\prod_{\substack{U\in K[\pi]\\ U\neq\tilde U}}\varphi[g_j(b):j\in U]+o(\frac{1}{N}).
\end{aligned}$$

Since $\sigma\cup K[\pi]$ has at least one singleton block, we know that the first term is zero. But in fact this partition has at least two singleton blocks. So the second term is zero as well. It follows that

$$\varphi'[f_1(a)g_1(b)\ldots f_n(a)g_n(b)] = 0,$$

which should be compared with the value of the corresponding right-hand side of (8.1). See (Curran, Speicher 2011) and (Collins, Hasebe, Sakuma 2015) for related results.