Topics in Random matrices
Preliminary version

Michael Anshelevich and Gregory Berkolaiko

January 11, 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.

$E =$ expectation functional,

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

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

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

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

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

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

and so

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

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

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

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

$$E[x_N] \to a.$$

$x_N \to a$ ***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

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

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,

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

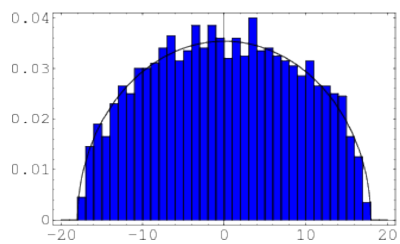

Figure 1.1: Semicircle law

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

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

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

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

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

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

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

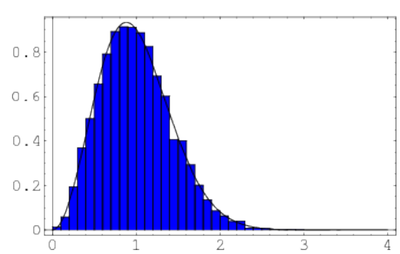

Figure 1.2: The Tracy-Widom distribution

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,

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

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

$\operatorname{Var} \nu_1 < \infty$, $\operatorname{Var} \nu_2 < \infty$, 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,

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

More generally, may look at

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

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

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

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

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

$X = \frac{1}{N} \hat{Y}^T \hat{Y}$ (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*

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

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,*

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

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

$E[Y_{ij}] = 0$, $\operatorname{Var} \nu_1 = 1$, *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.

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

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

$$S_{\vec{u}} = \{u(1), u(2), \ldots, u(k)\}.$$

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

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

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

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

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

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

$$|\{\vec{u} : |S_{\vec{u}}| = s\}| \leq \binom{N}{s} s^k \leq s^k N^s.$$

Therefore

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

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

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

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

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
$$\frac{1}{N}E[\mathrm{Tr}[X_N^k]] = \frac{N(N-1)\dots(N-k/2)}{N^{1+k/2}}\,\mathrm{Var}^k[\nu_1]c_{k/2} \to c_{k/2}$$
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 $\delta > 0$*
$$P(U \geq \delta) \leq \frac{1}{\delta}E[U].$$

Lemma 2.6 (Chebyshev inequality)**.** *Let V be a random variable with finite variance. Then for any $\delta > 0$*
$$P(|V - E[V]| \geq \delta) \leq \frac{1}{\delta^2}\,\mathrm{Var}[V].$$

Proof of Theorem 2.1 for convergence in probability. Since

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

we have just shown that $\left|\frac{1}{N}E[\operatorname{Tr}[X^k]]-m_k(\sigma)\right|\to 0$ 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

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

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

$$Y_{\vec{u}}=Y_{u(1)u(2)}Y_{u(2)u(3)}\ldots Y_{u(k)u(1)},$$

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

Again we have a multigraph with the vertex set $S_{\vec{u}}\cup S_{\vec{v}}$, 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 $S_{\vec{u}}$ and $S_{\vec{v}}$ if these are disjoint, or one component if these intersect. By the same arguments as above, we can conclude that

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

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

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

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

- $|S_{\vec{u}}\cup S_{\vec{v}}|=2+k$ and the graph has two components, each of which is a tree with double edges;
- or $|S_{\vec{u}}\cup S_{\vec{v}}|=1+k$, $|S_{\vec{u}}\cap S_{\vec{v}}|=1$, and the graph is a tree with double edges;
- or $|S_{\vec{u}}\cup S_{\vec{v}}|=1+k$, $|S_{\vec{u}}\cap S_{\vec{v}}|=0$, and the graph has two components, one a tree with double edges, the other with double edges and a single cycle;
- or $|S_{\vec{u}}\cup S_{\vec{v}}|=1+k$, $|S_{\vec{u}}\cap S_{\vec{v}}|=0$, 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 $|S_{\vec{u}} \cup S_{\vec{v}}| = 1 + k$, $|S_{\vec{u}} \cap S_{\vec{v}}| = 0$, and the graph has two components, each of which is a tree with double edges, and a single quadruple edge.

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

$$E[Y_{\vec{u}} Y_{\vec{v}}] - E[Y_{\vec{u}}]\, E[Y_{\vec{v}}] = 0.$$

Finally, suppose $|S_{\vec{u}} \cup S_{\vec{v}}| = 1 + k$, 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

$$E[Y_{\vec{u}} Y_{\vec{v}}] - E[Y_{\vec{u}}]\, E[Y_{\vec{v}}] = 0.$$

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

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

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

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

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

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

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, $E[xyx] = E[x^2]E[y]$, but in general $e[xyxy] \neq E[x^2]E[y^2]$.

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

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

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

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

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. $\square$

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

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

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 $E[Z_N^k]$ 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

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

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. □

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

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

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, $\{X_{ij} : 1 \leq i \leq j \leq N\}$ are independent,*

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

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

$E[Y_{ij}] = 0$, $\operatorname{Var} \nu_1 = 1$, 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 $C > 0$*

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

where $\sigma_{ij}(C)^2 = \operatorname{Var}(Y_{ij} \mathbf{1}_{|Y_{ij}| \leq C})$ 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

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

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

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

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

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

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

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

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

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 $A = \frac{1}{2}(A+B) + \frac{1}{2}(A-B)$. $\square$

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

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

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

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

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

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

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

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

Therefore

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

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

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

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 $\|A\|_F \leq \sqrt{2}\,\|A\|$. 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) $\|A\|_{op} \le \|A\|_F$. 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 $\|f_{\mathrm{Tr}}\|_{\mathrm{Lip}} \le \sqrt{2N}\|f\|_{\mathrm{Lip}}$.

Proof. For part (a),

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

Similarly, for part (b),

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

□

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 $\lambda \in \mathbb{R}$,*

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

Therefore for all $\delta > 0$,

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

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,

$$E[E[Z|Y]] = E[Z].$$

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

$$E[f(X)g(Y)|Y] = E[f(X)]g(Y).$$

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

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

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 $E[F(X)] = 0$.

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

$$E[\exp(-\lambda F(Y))] \geq \exp(-\lambda E[F(Y)]) = 1$$

and thus (by independence of X and Y)

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

It thus suffices to estimate $E[\exp(\lambda(F(X) - F(Y)))]$, 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

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

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

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

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,

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

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

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

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

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

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 $F \in \mathrm{Lip}(\mathbb{R}^n)$. Then for all $\lambda \in \mathbb{R}$,

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

and so for all $\delta > 0$,

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

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

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

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

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

Similarly,

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

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

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

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 $F \in \mathrm{Lip}(\mathbb{R}^n)$,*

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

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

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

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

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

Therefore

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

and so

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

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 $\mathrm{Hess}(\Phi)_{ij} = \partial_i\partial_j\Phi$*. If for all x,*

$$\mathrm{Hess}(\Phi)(x) \ge \frac{1}{c}I$$

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 $\frac{1}{Z_N}e^{-N\,\mathrm{Tr}[V(X)]}\,dX$ with $V''(x) \ge \frac{1}{c} > 0$. Then

for any Lipschitz f,

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

and for any k,

$$P\left(|\lambda_k(X_N) - E\left[\lambda_k(X_N)\right]| \ge \delta\right) \le 2\exp\left(-N\delta^2/4c\right)$$

The corollary applies for example to $V(x) = |x|^a$, $a \ge 2$, but not for $a < 2$. For $1 \le a < 2$, 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

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

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

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

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 $F \in \mathrm{Lip}(\mathbb{R}^n)$,*

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

where K is determined by m.

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

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

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

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

so that

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

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

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

Iterating,

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

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

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

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

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

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

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,

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

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

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

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 $z \in \mathbb{C}^+$,

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

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. □

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 $E[Y_{ij}^4] = m_4 < \infty$. 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

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

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,

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

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,

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

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

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

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

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

Therefore

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

and so

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

It follows that

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

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), $\left| \Im E[V_k \mid Y_N^{(k)}] \right| \geq |\Im z|$ a.s. Then

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

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

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

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

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

Write

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

where

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

and

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

On the one hand, from (4.6)

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

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

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

Thus using equation (4.8), we obtain the estimate

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

Let

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

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

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

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

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

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

$$\sup_{\Im z \geq 2}\operatorname{Var}[S_N(z)] \leq \frac{C'}{N}. \tag{4.11}$$

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

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

Combining with equation (4.4), we obtain

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

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. $\square$

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 $E[\chi_\beta^2] = \beta$,

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

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

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

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

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

and so

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

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 Topics for the second part of the course.

Exact eigenvalue distribution for β-ensembles.

We will prove that the joint density of ordered eigenvalues of an un-normalized β-ensemble matrix is

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

Along the way we will develop some general results on the spectral theory of finite Jacobi matrices and orthogonal polynomials.

Asymptotic distributions and asymptotic freeness.

Question. Let $(A_N)_{N=1}^{\infty}$ and $(B_N)_{N=1}^{\infty}$ be ensembles of self-adjoint random matrices at least one of which is unitarily invariant. Suppose their empirical spectral distributions converge weakly almost surely:

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

Since A_N and B_N do not commute, one cannot speak directly of their joint distribution. What are their asymptotic joint moments? What is the asymptotic distribution of A_N+B_N and (if it makes sense) of $A_N B_N$?

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

$$\frac{1}{N}E[\mathrm{Tr}[A_N^{u(1)}B_N^{v(1)}\ldots A_N^{u(k)}B_N^{v(k)}]]\to\varphi[a^{u(1)}b^{v(1)}\ldots a^{u(k)}b^{v(k)}]. \tag{5.1}$$

Moveover, using free independence, the joint moments of a and b can be computed in terms of individual moments of a and b. There are also formulas for expressing μ_{a+b} and (in the appropriate setting) μ_{ab} in terms of μ, ν.

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

Fluctuations and second order freeness.

Question. Convergence (5.1) is a version of the law of large numbers. One may then ask a central limit type question: what is the asymptotic distribution of

$$E[\mathrm{Tr}[A_N^{u(1)} B_N^{v(1)} \ldots A_N^{u(k)} B_N^{v(k)}]] - N\varphi[a^{u(1)} b^{v(1)} \ldots a^{u(k)} b^{v(k)}]$$

(note the factor of N).

Answer. Under appropriate assumptions, these fluctuations are asymptotically Gaussian. Under stronger assumptions, one can compute their joint covariances for any $\mathbf{u}, \mathbf{v}$. Under stronger assumptions, one may explicitly diagonalize these covariances. One approach is to show that A_N and B_N are not just asymptotically free but asymptotically *second order free*.

Reference: (Mingo, Speicher 2016) Chapter 5.

Band and block matrices and operator-valued freeness.

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 Gaussian with mean zero and covariance

$$E[Y_{rp} Y_{qs}] = \delta_{rs} \delta_{pq} \sigma(r/N, p/N).$$

Here $\sigma(x, y) = \sigma(y, x)$ is a sufficiently nice function.

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 Gaussian entries with mean zero, $(Y^{(ij)})^* = Y^{(ji)}$, and covariance

$$E[Y_{rp}^{(ij)} Y_{qs}^{(kl)}] = \delta_{rs} \delta_{pq} \sigma(i, j; k, l).$$

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.

Reference: (Mingo, Speicher 2016) extracts from Chapters 9, 10.

Spiked models, subordination, and infinitesimal freeness.

Suppose A_N is as before, but B_N is a diagonal matrix with 5 non-zero eigenvalues (independently of N). Then clearly $\hat{\mu}_{B_N} \to \delta_0$ 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 B_N. They can be studied using subordination functions and infinitesimal freeness.

Reference: (Shlyakhtenko 2015) and earlier work by Capitaine, Belinschi, Bercovici, Fevrier.

5.4 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 $n \geq 2$*

$$Q_n(\lambda) + a_n Q_{n-1}(\lambda) + b_{n-1}^2 Q_{n-2}(\lambda) = \lambda Q_{n-1}(\lambda) \tag{5.2}$$

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 symmetric $N \times N$ 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*:

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

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 $Q_{N-1}(\lambda(J_N)) = 0$. From the recursion (5.2) it follows that $Q_{N-2}(\lambda(J_N)) = 0$. Applying the recursion repeatedly, by induction we get that $Q_0(\lambda(J_N)) = 0$, 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 $\lambda_1 \le \ldots \le \lambda_N$, and U is an orthogonal matrix whose columns $\vec{u}_1, \ldots \vec{u}_N$ are the normalized eigenvectors. Denote

$$p_i = |U_{i1}|^2,$$

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

Proposition 5.14. *The map*

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

given by

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

is a bijection.

For $(p_1, \ldots, p_{N-1})$ 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.

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

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 $\langle J^k \xi, \xi \rangle$ 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 $1 \le n \le N-1$,

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

and

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

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 $\vec{u}_i$ is U_{i1}, it follows that

$$U_{ij} = U_{i1} P_{j-1}(\lambda_i).$$

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

Proof.

$$\begin{aligned}\int P_{j-1}(x) P_{n-1}(x)\, d\nu_N(x) &= \sum_{i=1}^{N} p_i P_{j-1}(\lambda_i) P_{n-1}(\lambda_i) \\ &= \sum_{i=1}^{N} U_{i1} P_{j-1}(\lambda_i) U_{i1} P_{n-1}(\lambda_i) = \sum_{i=1}^{N} U_{ij} U_{in} = (U^T U)_{jn} = \delta_{j=n}. \qquad \square\end{aligned}$$

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* $b_1, \ldots, b_{N-1} > 0$.

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 $\{x^n : n \geq 0\}$,

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

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

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

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

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

Thus finally,

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

Since the leading coefficients of P_n and P_{n-1} are positive, $b_n \geq 0$. If $b_N = 0$, then $x P_{N-1}$, 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. $\square$

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 $P_0, P_1, \dots, P_{N-1}$. They satisfy a three-term recursion relation, whose coefficients are determined by

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

and

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

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 $(m_1(N), \dots, m_{2N-1}(N))$.

Proposition 5.21. *The Jacobian determinant 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

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

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

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

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,

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

where

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

is the Vandermonde determinant.

Proof.

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

So

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

and the Jacobian matrix of the transformation

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

is

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

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

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

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

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

□

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

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

Proposition 5.24.

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

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

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

Then A is upper-triangular, with entries $1,b_1,\dots,\prod_{k=1}^{N-1}b_k$ 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 $\vec q=(q_1,\dots,q_N)^T=U^Te_1$ be the first row of U, so that $q_i^2=p_i$. Then

$$
\begin{aligned}
A&=(UU^Te_1,U\Lambda U^Te_1,\dots,U\Lambda^{N-1}U^Te_1)=U(U^Te_1,\Lambda U^Te_1,\dots,\Lambda^{N-1}U^Te_1)\\
&=U(\vec q,\Lambda\vec q,\dots,\Lambda^{N-1}\vec q)=U\begin{pmatrix}q_1&\lambda_1q_1&\dots&\lambda_1^{N-1}q_1\\ \vdots&\vdots&\dots&\vdots\\ q_N&\lambda_Nq_N&\dots&\lambda_N^{N-1}q_N\end{pmatrix}
\end{aligned}
$$

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

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

□

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),

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

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

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

we obtain the density

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

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*

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

Another approach to compute

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

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

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