NOTES ON RANDOM MATRICES

LEONID PETROV

(notes by Kristin Courtney; Stephen Hardy; Bryce Terwilliger; Daniel Franz; . . .)

ABSTRACT. These are notes for the MATH 8380 "Random Matrices" course at the University of Virginia in Spring 2016. The notes are constantly updated, and the latest version can be found at the git repository https://github.com/lenis2000/RMT_Spring_2016

(BEFORE TEXING, PLEASE FAMILIARIZE YOURSELF WITH STYLE SUGGESTIONS AT HTTPS://GITHUB.COM/LENIS2000/RMT_SPRING_2016/BLOB/MASTER/TEXING.MD)

CONTENTS

Lecture #1 on 1/20/2016 ——————————————————————

1. INTRODUCTION

1.1. **Matrices and eigenvalues.** The study of random matrices as a field is a patchwork of many fields. The main object we study is a probability distribution on a certain subset of the set of matrices $\mathrm{Mat}(N \times N, \mathbb{R}$ or $\mathbb{C})$, thus giving us a random matrix A.

Definition 1.1. An *eigenvalue* λ of the matrix A is a root of the polynomial $f(\lambda) = \det(A - \lambda I)$, where I is the identity matrix (we use this notation throughout the notes). Equivalently, λ is an eigenvalue if A if the matrix $A - \lambda I$ is not invertible. This second way of defining eigenvalues in fact works even when A is not a finite size matrix, but an operator in some infinite-dimensional space.

We will largely be only concerned with real eigenvalues. That is the eigenvalues of a real symmetric matrix over $\mathbb{R}$ or Hermitian over $\mathbb{C}$ that is where $A^* = A$ (here and everywhere below A^* means $\overline{A^T}$, i.e., transposition and complex conjugation).

Remark 1.2. The case when eigenvalues can be complex is also studied in the theory of random matrices, sometimes under the keyword *complex random matrices*. This area is more modern and is actively developing now. See, for example, [GT10] for a discussion of the law of large numbers.

Proposition 1.3. *Every eigenvalue of a Hermitian matrix is real.*

Date: March 21, 2016.

Proof. Let A be a Hermitian matrix, so that $A^* = A$. Let λ be an eigenvalue of A. Let v be a nonzero vector in the null space of $A - \lambda I$. Let $a = \overline{v^T} v = |v|^2$, so that a is a positive real number. Let $b = \overline{v^T} A v$. Then $\bar{b} = \overline{b^T} = \overline{\overline{v^T} A v}^T = \overline{v^T} \bar{A}^T v = \overline{v^T} A v = b$, so b is real. Since $b = \lambda a$, λ must be real. $\square$

Let $\mathcal{H}_N$ be the set of $N \times N$ Hermitian matrices. For each N, let μ_N be a probability measure on $\mathcal{H}_N$ (it can be supported not by the whole $\mathcal{H}_N$, but by a subset of it, too). Then for each matrix $A \in \mathcal{H}_N$ we may order the real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ of A (the collection of eigenvalues is called the *spectrum* of A).

A collection of probability measures μ_N on $\mathcal{H}_N$ for each $N \geq 1$ is said to be a *random matrix ensemble*. For such an ensemble, the eigenvalues $\lambda_1^{(N)} \geq \cdots \geq \lambda_N^{(N)}$ of matrices N form random point configurations on $\mathbb{R}$ with growing numbers of points. Our main goal is to study the asymptotic properties of these collections of points on $\mathbb{R}$, as $N \to \infty$.

1.2. **Why study random matrices?**
Let us briefly discuss five possible motivations to study random matrices and asymptotic distributions of random matrix spectra.

1.2.1. Matrices are a natural generalization of real numbers, so studying them would seem natural from a pure probability point of view. However, the development of the theory of Random Matrices was much application driven.

1.2.2. *Hurwitz and theory of invariants.* A. Hurwitz in the 1890s [Hur97] computed the volume of orthogonal and unitary groups of matrices.[1] For example, $U(1)$, the set of unitary 1×1 unitary matrices — the unit circle — has volume 2π. For general N, the volume of $U(N)$ is the normalization constant $Z_N = \int_{U(N)} 1 \cdot d(\text{Haar}_N)$ in probabilistic integrals over the Haar measure on the unitary group,

$$Z_N = 2^{N(N+1)/2} \prod_{k=1}^{N} \frac{\pi^k}{\Gamma(k)} = 2^{N(N+1)/2} \prod_{k=1}^{N} \frac{\pi^k}{(k-1)!}.$$

See [DF15] for a recent survey.

1.2.3. *Statistics.* J. Wishart in 1928 [Wis28] considered random covariance matrices of vector-valued data. For testing the hypothesis of independence of components of the vector, it is natural to study the distribution of the random covariance matrix of the uncorrelated vector (the null-hypothesis distribution). Let us assume that the components of the vector are identically distributed.

This latter matrix ensemble (called the *Wishart ensemble*) can be constructed by taking a rectangular matrix Y with independent (or uncorrelated) identically distributed entries, and taking $A = Y^T Y$. Then A is a square matrix which is said to have the (*real*) *Wishart distribution*.

For the purposes of statistics, the distribution of the Wishart matrix A should be compared with the distribution under the alternate hypothesis that the entries of the vector are correlated. For certain assumed nature of the correlation structure, this leads to considering *spiked random matrices* of the form $A + R$, where A is Wishart and R is a finite-rank perturbation. It turns out that sometimes the presence of a nonzero matrix R may be detected by looking at the spectrum of $A + R$, which again leads to considering spectra of random matrices. One reference (among many others which are not mentioned) relevant for the current research on spiked random matrices is [BBAP05].

---

[1]The group $O(N)$ of orthogonal $N \times N$ matrices consists of matrices O with real entries, for which $O^T O = OO^T = I$. The group $U(N)$ of unitary $N \times N$ matrices consists of matrices U with complex entries, for which $UU^T = U^T U = I$. Both groups are *compact*, and so possess finite Haar measures, i.e., measures μ which are invariant under left and right shifts on the group.

1.2.4. *Nuclear physics.* Active development of the theory of random matrices begins in the 1950s when Wigner, Dyson, Mehta, and their collaborators explored nuclear physics applications. In nuclear quantum physics a state of a system is an operator on an L^2 space of functions; its eigenvalues are the energy levels of the system. For large nuclei it is difficult to analyze the operator in L^2 directly, but Wigner postulated that differences in energy levels could be modeled by differences in eigenvalues of certain classes of matrices under appropriate probability measures. That is, the collections $\{\Delta E_i\}$ and $\{\lambda_i - \lambda_{i+1}\}$ should be statistically close. Moreover, the random matrix ensemble should have the same symmetry as that quantum system. The symmetry classes of random matrices are discussed in detail in a recent survey [Zir10]. Dyson proposed a model of stochastic dynamics of energies (eigenvalues of random matrices). We will study the Dyson's Brownian motion later.

Section 1.1 of the book [Meh04] contains a nice outline of physical applications of random matrices.

1.2.5. *Number theory.* Dyson and Montgomery uncovered number theoretic applications of random matrices in the 1970's [Mon73], with Odlyzko in the 1980's [Odl87] providing powerful numerical simulations.

Consider the Riemann Zeta Function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad \text{for } s \in \mathbb{C} \text{ with the real part of } s > 1$$

Riemann showed that $\zeta(s)$ can be analytically continued to a function on $\mathbb{C}$ with a pole at $s = 1$. The famous Riemann hypothesis is that all the zeroes of the Zeta function with real part greater than 0 lie on the *critical line* $\frac{1}{2} + it$. It turns out that the distribution of the zeroes on the critical line can be linked to the distribution of eigenvalues of random matrices. Consider the zeros $\frac{1}{2} + it_n$ of the zeta function with $t_n \in \mathbb{R}$. Let us define

$$w_n = \frac{t_n}{2\pi} \log\left(\frac{|t_n|}{2\pi}\right),$$

then $\lim_{L \to \infty} \frac{1}{L} \# \{w_n \in [0, L]\} = 1$, i.e., the average density of the w_n's is 1. The theorem/conjecture of Montgomery[2] states that the pair correlations of the zeroes of the zeta function have the form

$$\lim_{L \to \infty} \frac{1}{L} \# \left\{ \begin{array}{c} w_n \in [0, L] \\ \alpha \leq w_n - w_m \leq \beta \end{array} \right\} \sim \int_{\alpha}^{\beta} \left(\delta(x) + 1 - \frac{\sin^2(\pi x)}{\pi^2 x^2} \right) dx, \tag{1.1}$$

where $\delta(x)$ is the Dirac delta. Further details on this and other connections between number theory and random matrices can be found in [KS00], [Kea06].

Remark 1.4. There are accounts of Montgomery meeting Dyson at teatime at the IAS; the latter pointed out the connection between Montgomery's formula and the eigenvalue distributions of random matrices. A quick Internet search lead to the following links containing details: `https://www.ias.edu/articles/hugh-montgomery` and `http://empslocal.ex.ac.uk/people/staff/mrwatkin//zeta/dyson.htm`.

1.3. **Course outline.** The course will consist of five main parts, with the last part being optional:

1. Limit shape results for random matrices (such as Wigner's Semicircle Law). Connections to Free Probability.
2. Concrete ensembles of random matrices (GUE, circular, and Beta ensembles). Bulk and edge asymptotics via exact computations. Connection to determinantal point processes.
3. Dyson's Brownian Motion and related stochastic calculus.
4. Universality of random matrix asymptotics.
5. (optional, depending on time available) Discrete analogues of random matrix models: random permutations, random tilings, interacting particle systems.

---

[2]Depending in part on the Riemann hypothesis and in part on how strong is the assumed convergence in (1.1).

2. Wigner's Semicircle Law and its combinatorial proof

After discussing the object and motivations for studying random matrices, let us proceed to the first part of the course — the laws of large numbers for the eigenvalue distributions of random matrices. The first of these laws of large numbers is the *Wigner's Semicircle Law.* It dates back to [Wig55].

2.1. **Real Wigner matrices.** A particular ensemble of random matrices is the *real Wigner matrices.* Let $A \in \mathrm{Mat}(N \times N, \mathbb{R})$ with $A = (a_{ij})_{i,j=1}^{N}$ such that $a_{ij} = a_{ji}$. To describe the distribution of the random matrix A we only need to describe the upper triangular portion of A.

Definition 2.1. The law of the real Wigner $N \times N$ matrix is described as follows:

- $\{a_{ij}\}_{i \le j}$ is an independent collection of random variables
- $\{a_{ii}\}_{i=1}^{N}$ is iid[3], and $\{a_{ij}\}_{i<j}$ is iid.
- $\mathbb{E}\, a_{ij} = 0$ for all i, j; $\mathbb{E}\, a_{ij}^2 = 2$ for $i = j$; and $\mathbb{E}\, a_{ij}^2 = 1$ for $i \ne j$.
- all moments of a_{ij} are finite.

The last condition greatly simplifies technicalities of the proofs, but most results on real Wigner matrices hold under weaker assumptions.

Example 2.2. A large class of Wigner random matrices (which helps justify why in A the variances on the diagonal must be twice the off-diagonal variances) can be constructed as follows. Suppose the collection of random variables x_{ij} for $1 \le i, j \le N$ is iid with $\mathbb{E}\, x_{ij} = 0$ and $\mathbb{E}\, x_{ij}^2 = 1$. Let $X = (x_{ij})$ be an $N \times N$ matrix. Define

$$A := \frac{X + X^T}{\sqrt{2}}.$$

One readily sees that A is real Wigner. Namely, for example, $a_{11} = \frac{x_{11} + x_{11}}{\sqrt{2}} = \sqrt{2} x_{11}$, so $\mathbb{E}\, a_{11} = 0$ and $\mathbb{E}\, a_{11}^2 = 2\, \mathbb{E}\, x_{11}^2 = 2$. If $N \ge 2$ then $a_{12} = a_{21}$ with $a_{12} = \frac{x_{12} + x_{21}}{\sqrt{2}}$, and we have $\mathbb{E}\, a_{12} = 0$ and $\mathrm{Var}\, a_{12} = \frac{1}{2} \mathrm{Var}(x_{12} + x_{21}) = 1$ because x_{12} and x_{21} are independent.

2.2. **Gaussian Orthogonal Ensemble.** A special case of real Wigner matrices is when each a_{ij} is Gaussian. This case is called the *Gaussian Orthogonal Ensemble (GOE)*.

Lemma 2.3. *The distribution of the GOE is orthogonally invariant, that is, if A has the GOE distribution and $O \in O(N)$ is a fixed orthogonal matrix, then OAO^T has the same probability distribution as A.*

Proof. It is not hard to check that the probability density of A with respect to the Lebesgue measure on $\mathrm{Mat}(N \times N, \mathbb{R})$ (this space is isomorphic to $\mathbb{R}^{N(N+1)/2}$ by considering the upper triangular part) has the form

$$f(A) = c \exp(-\mathrm{tr}(A^2)),$$

where c is a normalization constant.[4] Since the matrix trace is invariant under cyclical permutations,

$$\mathrm{tr}(OA^2O^T) = \mathrm{tr}(A^2 O^T O) = \mathrm{tr}(A^2).$$

Thus, $OA^2O^T \overset{\mathcal{D}}{=} A$. □

We will discuss the GOE (and its close relative GUE, Gaussian Unitary Ensemble) in detail in the course later, but for now we will focus on properties of real Wigner matrices with general entry distribution.

---

[3]Independent identically distributed.

[4]Here and below $\mathrm{tr}(A) = a_{11} + a_{22} + \ldots + a_{NN}$ is the trace of a matrix.

2.3. **Formulation of the Wigner Semicircle Law.** For a real Wigner matrix $A_N \in \mathrm{Mat}(N \times N)$ let $\lambda_1^{(N)} \geq \cdots \geq \lambda_N^{(N)}$ be the eigenvalues of A_N. The *empirical distribution of the eigenvalues* is

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{N^{-1/2}\lambda_i^{(N)}}. \tag{2.1}$$

That is, we put delta masses of size $1/N$ into the N positions of rescaled eigenvalues $\lambda_i^{(N)}/\sqrt{N}$. This rescaling will turn out to be appropriate for the law of large numbers. Note that L_N is a probability measure on $\mathbb{R}$.

Remark 2.4. For the purposes of asymptotic statements, we will always assume that the off-diagonal entries of real Wigner matrices $A = A_N$ have the same fixed distribution independent of N, and similarly the diagonal entries have the same fixed (but different) distribution.
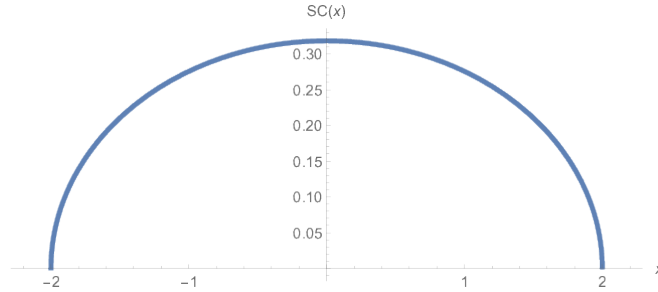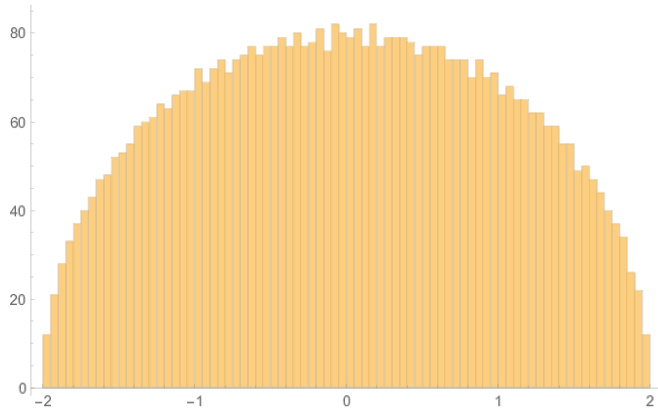

FIGURE 1. Semicircle density $\mathsf{SC}(x)$.

FIGURE 2. Histogram of the empirical distribution L_N for $N = 5000$.

Definition 2.5. The semicircle distribution SC is a fixed probability distribution on $\mathbb{R}$ supported on $[-2, 2]$ which is absolutely continuous with respect to the Lebesgue measure and has the density

$$\mathsf{SC}(x) := \frac{1}{2\pi} \sqrt{4 - x^2}, \qquad -2 \leq x \leq 2. \tag{2.2}$$

See Figure 1.

Note that slightly abusing the notation, by SC we will denote both the semicircle distribution and its probability density (2.2).

Theorem 2.6 (Wigner's Semicircle Law). *As $N \to \infty$, the empirical distributions L_N converge weakly, in probability to the semicircle distribution* SC.

Let us explain what we mean by convergence "weakly in probability". Formally this means that for any bounded continuous function f on $\mathbb{R}$ ($f \in C_B(\mathbb{R})$) and each $\varepsilon > 0$ we have

$$\lim_{N\to\infty} \mathbb{P}\left(\left| \int_{\mathbb{R}} f \, dL_N - \int_{\mathbb{R}} f \, d\mathsf{SC} \right| > \varepsilon \right) = 0. \tag{2.3}$$

That is, "in probability" means the usual convergence in probability of random elements L_N to a (non-random) element SC. On the other hand, "weakly" specifies the metric on the space of probability measures on $\mathbb{R}$ (to which all L_N and SC belong). Convergence of probability measures in this metric simply means weak convergence of probability measures on $\mathbb{R}$.

In other words, let us use a convenient notation for the pairing $\langle f, \mu \rangle = \int_{\mathbb{R}} f \, d\mu = \int_{\mathbb{R}} f(x) \, \mu(dx)$ for a given function f and measure μ. If μ is a random measure (such as L_N, since L_N depends on A_N which is random), then $\langle f, \mu \rangle$ is a random element of $\mathbb{R}$ (usually we say random variable). Since SC is not random, the pairing $\langle f, \mathsf{SC} \rangle$ is a fixed number for a given function f. The Semicircle Law thus states that for any given $f \in C_B(\mathbb{R})$ the random variable $\langle f, L_N \rangle$ converges in probability to the constant $\langle f, \mathsf{SC} \rangle$ which may be written as

$$\forall \varepsilon > 0, \qquad \lim_{N\to\infty} \mathbb{P}\left(|\langle f, L_N \rangle - \langle f, \mathsf{SC} \rangle| > \varepsilon \right) = 0, \tag{2.4}$$

which is the same as (2.3).

Remark 2.7. This type of convergence is reminiscent of the classical weak law of large numbers: for $\{X_i\}_{i=1}^{\infty}$ iid random variables with $\mathbb{E}\,|X_1| < \infty$, the random variables $\frac{1}{N} \sum_{i=1}^{N} X_i$ converge to the constant $\mathbb{E}\,X_1$ in probability as $N \to \infty$.

Lecture #2 on 1/25/2016

2.4. **Strategy of the proof.** We will employ the following strategy in our proof of the Wigner's semi-circle law. This is only the first of the proofs we will consider, and it is based on the computation of moments and on the related combinatorics. Recall that for a probability measure μ the quantities $\langle x^k, \mu \rangle$, $k = 0, 1, 2, \ldots$, are called the *moments* of μ.

First, in Section 2.5 we will compute the moments $m_k := \langle x^k, \mathsf{SC} \rangle$ of the limiting semicircle distribution, and identify the answer in terms of the Catalan numbers. Our second step in the proof is to show the convergence $\lim_{N\to\infty} \mathbb{E}\langle x^k, L_N \rangle = m_k$ for each k. We do this in Section 2.7 below. The third step (in Section 2.9) is to show that the variance of $\langle x^k, L_N \rangle$ goes to zero as $N \to \infty$ for each k. Finally, to complete the proof we will need to justify that the convergence (2.4) for any function $f(x)$ follows from the case of $f(x) = x^k$, $k = 0, 1, 2, \ldots$. This is done in Section 2.10.

2.5. **Moments of the semicircle distribution.** Here we will compute the moments of the semicircle distribution:

$$m_k = \langle x^k, \mathsf{SC} \rangle = \int_{-2}^{2} x^k \, \mathsf{SC}(x) \, dx = \int_{-2}^{2} x^k \left(\frac{\sqrt{4 - x^2}}{2\pi} \right) dx.$$

Clearly, by symmetry we have $m_k = 0$ for k odd. If k is even, let us perform a trigonometric substitution $x = 2\sin\theta$, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$, in the above integral:

$$m_{2k} = \frac{2^{2k+1}}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k}\theta \cos^2\theta \, d\theta. \tag{2.5}$$

Lemma 2.8. *We have*

$$\frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k}\theta \, d\theta = \frac{(2k-1)!!}{2^k k!},$$

where recall that $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdots (2k-3)(2k-1)$.

Proof. Denote the integral in the right-hand side by α_k. Observe that $\alpha_0 = 1$. Integrating by parts for $k \geq 1$, we have

$$\alpha_k = -\frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k-1}\theta \, d(\cos\theta) = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (2k-1) \sin^{2k-2}\theta \cos^2\theta \, d\theta = (2k-1)\alpha_{k-1} - (2k-1)\alpha_k.$$

Therefore, the quantities α_k satisfy

$$\frac{\alpha_k}{\alpha_{k-1}} = \frac{2k-1}{2k},$$

which is the same relation as for the quantities $\frac{(2k-1)!!}{2^k k!}$. This completes the proof. □

By relating m_{2k} (2.5) and α_k in the above lemma, we see that the even moments of the semicircle distribution are given by

$$m_{2k} = \frac{1}{k+1}\binom{2k}{k}, \qquad k = 0, 1, 2, \ldots. \tag{2.6}$$

These quantities are called the *Catalan numbers*.

2.6. **Catalan numbers.** The Catalan numbers Cat_k are defined as

$$\mathrm{Cat}_k := \frac{1}{k+1}\binom{2k}{k}, \qquad k = 0, 1, 2, \ldots. \tag{2.7}$$

The first twenty one of them are

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, 742900, 2674440,$$
$$9694845, 35357670, 129644790, 477638700, 1767263190, 6564120420.$$

They are ubiquitous in combinatorics: for example, there are more than 200 families of objects enumerated by the Catalan numbers [Sta15]. A list of references and properties of Catalan numbers may be found at [Cat].

Here we will list a number of properties of the Catalan numbers which will be important for our proof of the semicircle law.

2.6.1. *Dyck paths.*

Definition 2.9. A *Dyck path* of length $2n$ is a sequence $d_0, d_1, \ldots, d_{2n}$ such that $d_0 = d_{2n} = 0$, $d_{i+1} - d_i = \pm 1$ for all i, and that $d_i \geq 0$ for all i. Graphically Dyck paths can be represented as in Figure 3.
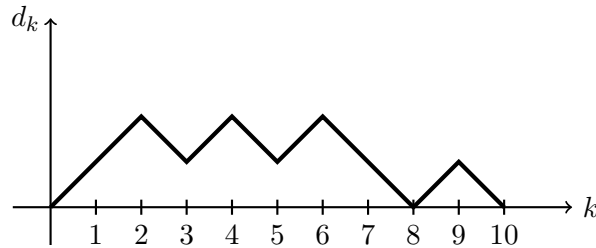

FIGURE 3. A Dyck path of length $2n = 10$.

FIGURE 4. All five Dyck paths of length $2n = 6$. The first two paths first return to zero at time $2j = 2$, the third path first returns to zero at time $2j = 4$, and the last two paths first return to zero at time $2j = 6$.

Exercise 2.10. The number of Dyck paths of length $2n$ is equal to the Catalan number Cat_n.

Idea. Use the *reflection principle* — a tool used in the study of random walks and Brownian motion. See `https://en.wikipedia.org/wiki/Catalan_number#Second_proof` for details.

Another way to count the Dyck paths is to first establish the recurrence (2.8), and then use generating functions to solve the recurrence (see Remark 2.12 below). □

2.6.2. *Recurrence.*

Lemma 2.11. *The Catalan numbers satisfy the recurrence relation*

$$\mathrm{Cat}_0 = 1, \qquad \mathrm{Cat}_n = \sum_{j=0}^{n} \mathrm{Cat}_{j-1} \mathrm{Cat}_{n-j}. \tag{2.8}$$

Proof. The easiest way to see this is by counting the Dyck paths: let the first time a Dyck path reaches 0 be $2j$, then j can be any number from 1 to n (see Figure 4). The part of the Dyck path after time $2j$ is independent from the part before $2j$. The number of paths from $2j$ to $2n$ is exactly Cat_{n-j}. The number of paths from 0 to $2j$ (with the condition that they do not get to 0 in the meantime) can be seen to be Cat_{j-1} by cutting out the first up and the last down steps. This implies the recurrence. □

Remark 2.12. The recurrence (2.8) provides a way to get the explicit formula (2.7). Namely, considering the generating function $G(z) = \sum_{n=0}^{\infty} \mathrm{Cat}_n z^n$, we see that (2.8) implies

$$G(z) = zG(z)^2 + 1.$$

This equation on $G(z)$ has two solutions $\frac{1 \pm \sqrt{1-4z}}{2z}$, of which we should pick $\frac{1 - \sqrt{1-4z}}{2z}$ because the other solution diverges as $z \to 0$. The Taylor expansion of this function is

$$\frac{1 - \sqrt{1 - 4z}}{2z} = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} z^n = 1 + z + 2z^2 + 5z^3 + 14z^4 + 42z^5 + 132z^6 + \dots,$$

which converges for $|z| < \frac{1}{4}$. This shows that the Dyck paths are enumerated by the Catalan numbers.

2.6.3. *Trees.* As was mentioned before, the Catalan numbers enumerate numerous families of combinatorial objects. We will need one more family of such objects — *rooted ordered trees.* An ordered tree is a rooted tree (i.e., a tree with a distinguished vertex R called the *root*) in which children of every vertex are linearly ordered. On pictures this ordering will be represented from left to right (see Figure 5).

Lemma 2.13. *The number of rooted ordered trees with $n+1$ vertices (including the root) is equal to the Catalan number Cat_n.*

Proof. Assume that the leftmost subtree contains j vertices (without the root), then the rest of the tree including the root contains $n-j+1$ vertices. This readily implies the recurrence (2.8), which establishes the claim. □

FIGURE 5. These trees are isomorphic as rooted trees, but are different as rooted ordered trees. A beginning of the walk of Exercise 2.14 is displayed for the second tree.

FIGURE 6. Dyck paths corresponding to the rooted ordered trees in Figure 5 (see Exercise 2.14).

Exercise 2.14. By comparing the proof of Lemmas 2.11 and 2.13, come up with a bijection between Dyck paths and ordered rooted trees.

Idea. Consider the walk around the tree (such that the tree is always to the left of the walker), which starts to the left of the root. Let d_k be the distance of the walker from the root. The Dyck paths corresponding to trees in Figure 5 are given in Figure 6. $\qquad\square$

2.7. **Convergence of expectations** $\mathbb{E}\langle x^k, L_N\rangle \to m_k$. With the Catalan preparations in place, let us return to the semicircle law. We would like to show that

$$\lim_{N\to\infty} \mathbb{E}\langle x^k, L_N\rangle = m_k = \begin{cases} 0, & k \text{ odd;} \\ \text{Cat}_{k/2}, & k \text{ even.} \end{cases} \tag{2.9}$$

First, observe that the left-hand side has the form

$$\mathbb{E}\langle x^k, L_N\rangle = \mathbb{E}\int_{\mathbb{R}} x^k\, L_N(dx)$$

$$= \mathbb{E}\int_{\mathbb{R}} x^k \frac{1}{N}\sum_{i=1}^{N} \delta_{N^{-1/2}\lambda_i}(dx)$$

$$= \mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x^k \delta_{N^{-1/2}\lambda_i}(dx)$$

$$= \mathbb{E}\frac{1}{N}\sum_{i=1}^{N}(N^{-1/2}\lambda_i)^k$$

$$= N^{-1-k/2}\,\mathbb{E}\sum_{i=1}^{N}\lambda_i^k.$$

Since A is diagonalizable (as an $N\times N$ real symmetric matrix), we have $\sum_{i=1}^{N}\lambda_i^k = \text{tr}(A^k)$. We may express the trace of the kth power of a matrix by a k-fold sum of cyclic products

$$\text{tr}(A^k) = \sum_{i_1,i_2,\ldots,i_k=1}^{N} a_{i_1 i_2} a_{i_2 i_3}\cdots a_{i_{k-1}i_k} a_{i_k i_1}.$$

So we have

$$\mathbb{E}\langle x^k, L_N \rangle = N^{-1-k/2} \sum_{i_1, i_2, \ldots, i_k = 1}^{N} \mathbb{E}(a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}). \tag{2.10}$$

Our goal now is to understand the combinatorial structure of the above big sum.

Definition 2.15. Each term of the sum can be encoded by a *closed word* $i_1 \ldots i_k i_1$ of length $k+1$ ("closed" in the sense that the word starts and ends with the same letter). For example, 123241 is a closed word of length 6. The *support* of a closed word is the set of all letters participating in this word. The support of 123241 is $\{1, 2, 3, 4\}$.

To each closed word w we associate an undirected graph G_w with vertices labeled by the support of the word, edges $(i_1, i_2), (i_2, i_3), \ldots, (i_k, i_1)$ connecting each consecutive pair of letters in the word. For example, if $w = 123241$, then G_w has four vertices $\{1, 2, 3, 4\}$ and five edges $\{(1, 2), (2, 3), (3, 2), (2, 4), (4, 1)\}$ (see Figure 7). Notice each graph G_w is connected. These (and similar) graphs are sometimes referred to as Feynman diagrams.
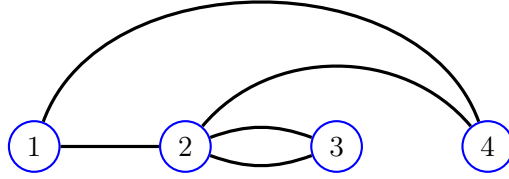

FIGURE 7. Graph G_w corresponding to the word $w = 123241$.

Let $N_{i_1 i_2}^w$ be the number of distinct edges connecting i_1 to i_2 in G_w. In our running example we have $N_{12}^w = 1$ and $N_{23}^w = 2$. Each edge may be a *self* edge such as $(1, 1)$, or it can be an edge *connecting* distinct vertices such as $(2, 3)$.

With this notation we have

$$\mathbb{E}\, a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1} = \prod_{\substack{\text{self } e \\ e \in G_{i_1 \ldots i_k i_1}}} \mathbb{E}\, a_{11}^{N_e} \prod_{\substack{\text{connecting } e \\ e \in G_{i_1 \ldots i_k i_1}}} \mathbb{E}\, a_{12}^{N_e}, \tag{2.11}$$

since all diagonal elements are iid, and all the elements above the diagonal are iid. Here the product runs over all possible distinct edges in the graph of the word.

In order for the expectation (2.11) to be nonzero, we must have the following properties:

- Since $\mathbb{E}\, a_{ij} = 0$, each edge in $G_{i_1 \ldots i_k i_1}$ must have $N_e \geq 2$..
- The graph $G_{i_1 \ldots i_k i_1}$ has $k+1$ edges, and so it can have at most $1 + k/2$ vertices.

Now let us look at the sum (2.10) as a whole. Call two graphs *equivalent* if they differ only by relabeling the vertices. Note that the expectations of the form (2.11) coming from equivalent graphs are the same. If a graph has t vertices, then there are $N^{\downarrow t} := N(N-1) \ldots (N-t+1)$ ways to relabel the vertices to get an equivalent graph. This implies that the sum (2.10) can be rewritten as

$$\mathbb{E}\langle x^k, L_N \rangle = \sum_{t=0}^{1+\lfloor k/2 \rfloor} \frac{N^{\downarrow t}}{N^{1+k/2}} \underbrace{\sum_{G_w \in \text{EqClass}_t} \prod_{\substack{\text{self } e \\ e \in G_w}} \mathbb{E}\, a_{11}^{N_e} \prod_{\substack{\text{connecting } e \\ e \in G_w}} \mathbb{E}\, a_{12}^{N_e}}_{(*)}, \tag{2.12}$$

where by EqClass_t we have denoted the set of equivalence classes of graphs G_w corresponding to closed word, having t vertices and $k+1$ edges, and also having $N_e \geq 2$ for each edge.

Clearly, for fixed t and k, the expression $(*)$ above does not depend on N and is finite. Also, since $N^{\downarrow t} = O(N^t)$, the sum (2.12) vanishes as $N \to \infty$ unless $t = 1 + k/2$. Because $t \leq \lfloor k/2 \rfloor$, this is possible only for k even. Therefore, $\mathbb{E}\langle x^k, L_N \rangle$ converges to zero if k is odd.

Now consider the case when k is even and $t = 1 + k/2$. Then the graph corresponding to each word $i_1 \ldots i_k i_1$ has $k + 1$ edges, $1 + k/2$ vertices, and $N_e \geq 2$ for each edge. Hence, gluing together pairs of edges connecting the same vertices, we see that the graph $G_{i_1 \ldots i_k i_1}$ must be a *tree* (see Figure 8).[5] In particular, there are no self edges and $N_e = 2$ for each connecting edge. This implies that

$$\lim_{N \to \infty} \mathbb{E}\langle x^k, L_N \rangle = \left| \text{EqClass}_{1+k/2} \right|.$$
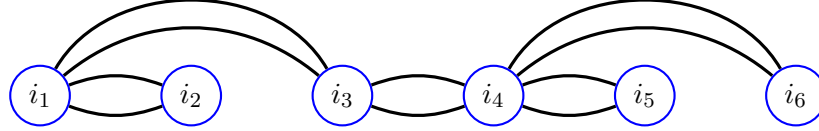

FIGURE 8. A graph G_w corresponding to a Wigner word $w = i_1 i_3 i_4 i_5 i_4 i_6 i_4 i_3 i_1 i_2 i_1$ which nontrivially contributes to the expansion (2.12). Here $k = 10$.
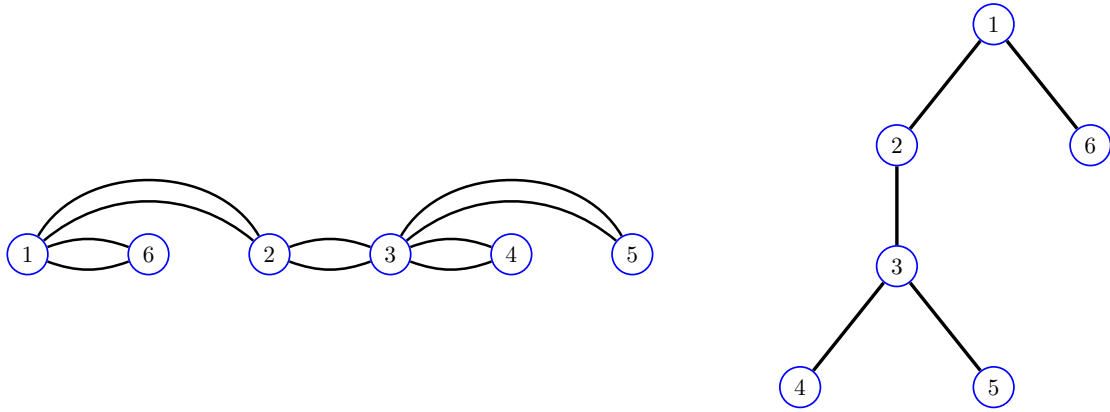

FIGURE 9. A representative graph $G_w \in \text{EqClass}_{1+k/2}$ corresponding to the graph as in Figure 8 (left), and its representation as a rooted ordered tree (right).

To count the number of trees $G_w \in \text{EqClass}_{1+k/2}$, let us choose representatives $w = v_1 \ldots v_{k+1}$, such that for each $i = 1, \ldots, k + 1$, the set $\{1, 2, \ldots, v_i\}$ is an interval in $\{1, 2, \ldots, N\}$ beginning at 1 (thus, $v_1 = v_{k+1} = 1$).

Exercise 2.16. There is a unique way of choosing such representatives.

Let the vertex 1 be the root R, and clearly the order coming from the word defines an order on this rooted tree (see Figure 9). This implies that $\left| \text{EqClass}_{1+k/2} \right| = \text{Cat}_k$, and finally proves the desired convergence (2.9).

Lecture #3 on 1/27/2016 ———————————————————————

---

[5] The words which correspond to such trees contributing to (2.12) are sometimes referred to as *Wigner words*.

2.8. **An example of counting terms in the expansion of eq. (2.10).** Before proceeding to finish the proof, let us consider one example how expansion (2.10) works for $k = 6$.

Exercise 2.17. How go get $\text{Cat}_3 = 5$ from $\mathbb{E}(\text{tr}(A^6))$?

Solution. We want to show how

$$\mathbb{E}(\text{tr}(A^6)) = N^{-1-3} \sum_{i_1,\dots,i_6=1}^{N} \mathbb{E}\left(a_{i_1,i_2} \cdot a_{i_2,i_3} \cdots a_{i_5,i_6} \cdot a_{i_6,i_1}\right) \xrightarrow[N\to\infty]{} 5.$$

We need to determine which terms are nonzero and how many such terms there are. If there are 5 or 6 independent indices, we get a product of expected values of independent random variables with expected value zero, so these terms do not contribute. If there are 3 or fewer independent indices, there are not enough terms to overcome the factor of N^{-4}, so these terms do not contribute in the limit. Thus we are interested in nonzero terms with 4 independent indices. One can check that there are 5 types of such terms:

(1)
$$\underbrace{(i_1,i_2),(i_2,i_1)}, \quad \underbrace{(i_1,i_3),(i_3,i_1)}, \quad \underbrace{(i_1,i_4),(i_4,i_1)}$$

(2)
$$(i_1,i_2), \quad \underbrace{(i_2,i_3),(i_3,i_2)}, \quad (i_2,i_1), \quad \underbrace{(i_1,i_4),(i_4,i_1)}$$

(3)
$$(i_1,i_2), \quad \underbrace{(i_2,i_3),(i_3,i_2)}, \quad \underbrace{(i_2,i_4),(i_4,i_2)}, \quad (i_2,i_1)$$

(4)
$$(i_1,i_2), \quad (i_2,i_3), \quad \underbrace{(i_3,i_4),(i_4,i_3)}, \quad (i_3,i_2), \quad (i_2,i_1)$$

(5)
$$\underbrace{(i_1,i_2),(i_2,i_1)}, \quad (i_1,i_3), \quad \underbrace{(i_3,i_4),(i_4,i_3)}, \quad (i_3,i_1)$$

These sequences bijectively correspond to noncrossing pair partitions of 6 elements. These partitions are in a clear bijection with Dyck paths of length 6 (shown in Figure 4), and are enumerated by the Catalan number $\text{Cat}_3 = 5$.

For each of these patterns, there are $N(N-1)(N-2)(N-3) \sim N^4$ terms in the sum, and each term is a product of the expected value of the squares of three independent off-diagonal random variables with expected value 0 and variance 1, like

$$\mathbb{E}(a_{i_1,i_2} \cdot a_{i_2,i_1} \cdot a_{i_1,i_3} \cdot a_{i_3,i_1} \cdot a_{i_1,i_4} \cdot a_{i_4,i_1}) = \mathbb{E}(a_{i_1,i_2}^2) \cdot \mathbb{E}(a_{i_1,i_3}^2) \cdot \mathbb{E}(a_{i_1,i_4}^2) = 1.$$

So in the limit we get the Catalan number $\text{Cat}_3 = 5$. $\qquad\qquad\square$

2.9. **Variances of $\langle x^k, L_N \rangle$.** Let us now show that the variances vanish in the limit:

$$\mathbb{E}(\langle x^k, L_N \rangle^2) - \left(\mathbb{E}(\langle x^k, L_N \rangle)\right)^2 \xrightarrow[N\to\infty]{} 0. \tag{2.13}$$

Recall that

$$\langle x^k, L_N \rangle = N^{-1-k/2} \sum_{\vec{i}=i_1,\dots,i_k=1}^{n} a_{i_1,i_2} \cdots a_{i_k,i_1}.$$

Now, writing $a_{\vec{i}}$ for $a_{i_1,i_2} \cdots a_{i_k,i_1}$, we have

$$\mathbb{E}(\langle x^k, L_N \rangle^2) - \Big(\mathbb{E}(\langle x^k, L_N \rangle) \Big)^2 = N^{-2-k} \sum_{\vec{i},\vec{j}} \Big(\mathbb{E}\Big(a_{\vec{i}} \cdot a_{\vec{j}}\Big) - \mathbb{E}(a_{\vec{i}}) \cdot \mathbb{E}(a_{\vec{j}}) \Big).$$

If the graphs $G_{\vec{i}}$ and $G_{\vec{j}}$ (corresponding to the words $i_1 \ldots i_k i_1$ and $j_1 \ldots j_k j_1$, respectively) do not share common edges, then the corresponding random variables $a_{\vec{i}}$ and $a_{\vec{j}}$ are independent, and so $\mathbb{E}(a_{\vec{i}} \cdot a_{\vec{j}}) = \mathbb{E}(a_{\vec{i}}) \cdot \mathbb{E}(a_{\vec{j}})$. Thus we are only interested in the terms for which edges of the graphs $G_{\vec{i}}$ and $G_{\vec{j}}$ overlap.

Example 2.18. For instance, if $\vec{i} = (1,2,3,2,1)$ and $\vec{j} = (1,2,1,1,1)$, then

$$\mathbb{E}(a_{\vec{i}}) = \mathbb{E}(a_{i_1,i_2} \cdot a_{i_2,i_3} \cdot a_{i_3,i_2} \cdot a_{i_2,i_1}) = \mathbb{E}(a_{1,2}^2)^2 = 1;$$
$$\mathbb{E}(a_{\vec{j}}) = \mathbb{E}(a_{i_1,i_2} \cdot a_{i_2,i_1} \cdot a_{i_1,i_1} \cdot a_{i_1,i_1}) = \mathbb{E}(a_{1,2}^2) \cdot \mathbb{E}(a_{1,1}) = 2;$$
$$\mathbb{E}(a_{i_1,i_2} \cdot a_{i_2,i_3} \cdot a_{i_3,i_2} \cdot a_{i_2,i_1} \cdot a_{i_1,i_2} \cdot a_{i_2,i_1} \cdot a_{i_1,i_1} \cdot a_{i_1,i_1}) = \mathbb{E}(a_{i_1,i_2}^4) \cdot \mathbb{E}(a_{i_2,i_3}^2) \cdot \mathbb{E}(a_{i_1,i_1}^2) = 2\,\mathbb{E}(a_{1,2}^4).$$
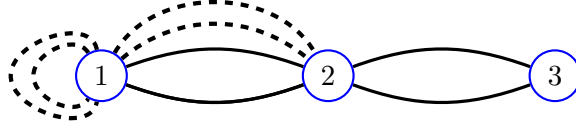
The corresponding graphs are given in Figure 10.

FIGURE 10. Graphs $G_{\vec{i}}$ (solid lines) and $G_{\vec{j}}$ (dashed lines) in Example 2.18.

We now argue similarly to the proof given in Section 2.7 (for the convergence of the first moments). Namely, in order for $\mathbb{E}(a_{\vec{i}} \cdot a_{\vec{j}}) - \mathbb{E}(a_{\vec{i}})\,\mathbb{E}(a_{\vec{j}})$ to be nonzero we must have the following:

- Since $\mathbb{E}(a_{ij}) = 0$, the graphs need to have $N_e \geq 2$;
- The graphs $G_{\vec{i}}$ and $G_{\vec{j}}$ need to share some edges.

If the combined graph has t vertices, there are $N^{\downarrow t} = N(N-1)\cdots(N-t+1)$ equivalent classes of graphs. Thus, the variance takes the form

$$\mathbb{E}(\langle x^k, L_N \rangle^2) - \Big(\mathbb{E}(\langle x^k, L_N \rangle) \Big)^2 = N^{-2-k} \sum_{t=1}^{2k} N^{\downarrow t} \underbrace{\left[\sum_{\substack{\text{equiv. classes} \\ \text{of graphs with} \\ 2k \text{ vertices}}} (\text{finite products of finite moments}) \right]}_{\text{finite and independent of } N}$$

Thus, we must have $t \geq k+2$ in order to have a nonzero contribution as $N \to \infty$. The associated graphs have $N_e \geq 2$ and are connected (since $G_{\vec{i}}$ and $G_{\vec{j}}$ are connected and overlap). There are totally $2k$ edges with multiplicities, thus $\leq k$ double edges. We conclude that there are no such graphs, and so there are no nonzero contributions to the variance in the limit as $N \to \infty$. This completes the proof of (2.13).

Remark 2.19. Remark: by a similar argument, $t = k+1$ also cannot contribute. Indeed, the combined graph of $G_{\vec{i}}$ and $G_{\vec{j}}$ has $\leq k$ double edges and $k+1$ vertices so it must be a tree (in the same sense of gluing edges as in Section 2.7 above). However, as $G_{\vec{i}}$ and $G_{\vec{j}}$ must also overlap (i.e., share common edges), there are no such trees. This implies a better estimate on the variance:

$$\mathbb{E}\Big(\langle x^k, L_N \rangle^2 \Big) - \mathbb{E}\Big(\langle x^k, L_N \rangle \Big)^2 = O(N^{-2}), \qquad N \to \infty.$$

This estimate can in fact be used to show almost-sure convergence to the semi-circular law.

2.10. **Estimates and completing the proof.** We want to show that for any continuous bounded function $f \in C_B(\mathbb{R})$, the random variables $\langle f, L_N \rangle$ converge in probability to $\langle f, \mathsf{SC} \rangle$ (this is further detailed in (2.4)). We have already shown that

- The moments converge: $\mathbb{E}\langle x^k, L_N \rangle \to \langle x^k, \mathsf{SC} \rangle$.
- The variances vanish: $\mathbb{E}(\langle x^k, L_N \rangle^2) - (\mathbb{E}\langle x^k, L_N \rangle)^2 \to 0$.

We will also need the following a priori estimate that the empirical distributions L_N are concentrated around zero:

Lemma 2.20. *For all $\varepsilon > 0$ there exists $B > 0$ so that*

$$\mathbb{P}\left(\left\langle |x|^k \mathbf{1}_{|x|>B}, L_N \right\rangle > \varepsilon \right) \xrightarrow[N\to\infty]{} 0.$$

Here an everywhere below $\mathbf{1}_A$ denotes the indicator function of an event A, so the function $|x|^k \mathbf{1}_{|x|>B}$ looks as in Figure 11.

Proof. We will use the Markov (sometimes also called Chebyshev) inequality:

$$\mathbb{P}(|X| > a) < \frac{\mathbb{E}\,|X|}{a} \qquad \text{for any } a > 0.$$

Note that for $|x| \geq B \geq 1$, $x^{2k} = |x|^k |x|^k \geq B^k |x|^k$, hence $|x|^k \leq x^{2k}/B^k$. Now we have by the Markov inequality:

$$\mathbb{P}\left(\left\langle |x|^k \mathbf{1}_{|x|>B}, L_N \right\rangle > \varepsilon \right) < \frac{1}{\varepsilon} \mathbb{E}\left(\left\langle |x|^k \mathbf{1}_{|x|>B}, L_N \right\rangle \right) \leq \frac{\mathbb{E}\left(\langle x^{2k}, L_N \rangle \right)}{\varepsilon B^k}.$$

We know that

$$\mathbb{E}\left(\langle x^{2k}, L_N \rangle \right) \to \mathbb{E}\left(\langle x^{2k}, \mathsf{SC} \rangle \right) = \mathrm{Cat}_k.$$

An easy (and in fact exact in the exponential order) estimate for the Catalan numbers is

$$\mathrm{Cat}_k = \frac{1}{k+1}\binom{2k}{k} \leq \sum_{j=0}^{2k} \binom{2k}{j} = 2^{2k} = 4^k.$$

Thus

$$\limsup_{N\to\infty} \mathbb{P}\left(\left\langle |x|^k \mathbf{1}_{|x|>B}, L_N \right\rangle > \varepsilon \right) \leq \frac{\mathrm{Cat}_k}{\varepsilon B^k} \leq \frac{4^k}{\varepsilon B^k}.$$

As k grows, the left hand side grows. However, for $B > 4$ the right hand side decays to zero. Thus if we set

$$\alpha_k = \limsup_{N\to\infty} \mathbb{P}\left(\left\langle |x|^k \mathbf{1}_{|x|>B}, L_N \right\rangle > \varepsilon \right),$$

then

$$0 \leq \alpha_1 \leq \alpha_2 \leq \ldots \leq \frac{4^k}{\varepsilon B^k} \to 0.$$

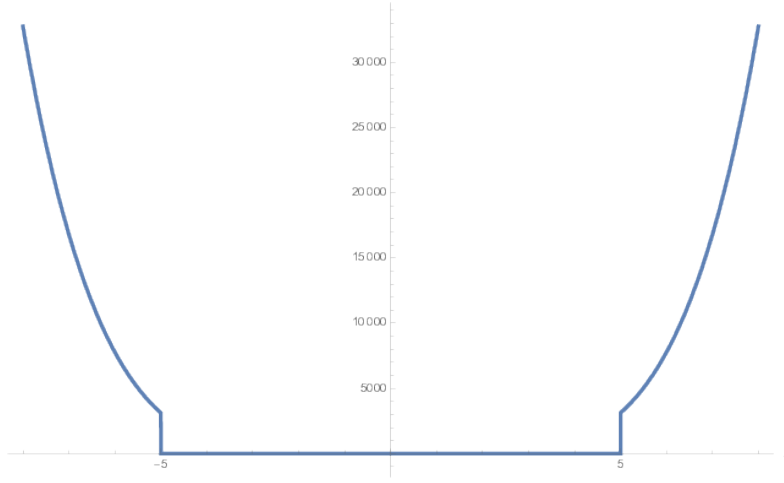Thus, all the α_k are zero. Since the probabilities are nonnegative, the desired result follows. $\square$

Now, fix $B > 4$ (say, $B = 5$), and uniformly approximate the function $f\mathbf{1}_{|x| \leq B}$ (a continuous function on a compact interval) by a polynomial. That is, by the Weierstrass Approximation Theorem, for every $\delta > 0$ there is a polynomial $Q_\delta(x)$ such that

$$\sup_{|x| \leq B} |f(x) - Q_\delta(x)| < \delta.$$

Therefore, we can estimate

$$|\langle f, L_N \rangle - \langle f, \mathsf{SC} \rangle| \leq |\langle f, L_N \rangle - \langle Q_\delta, \mathsf{SC} \rangle| + |\langle Q_\delta, \mathsf{SC} \rangle - \langle f, \mathsf{SC} \rangle|$$
$$\leq \left| \langle f\mathbf{1}_{|x| \leq B}, L_N \rangle - \langle Q_\delta, \mathsf{SC} \rangle \right| + \left| \langle f\mathbf{1}_{|x|>B}, L_N \rangle \right| + |\langle Q_\delta, \mathsf{SC} \rangle - \langle f, \mathsf{SC} \rangle|$$

FIGURE 11. The function $|x|^5 \mathbf{1}_{|x|>5}$.

$$\leq \left|\langle Q_\delta \mathbf{1}_{|x|\leq B}, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle\right|$$
$$+ \left|\langle f\mathbf{1}_{|x|\leq B}, L_N\rangle - \langle Q_\delta \mathbf{1}_{|x|\leq B}, L_N\rangle\right| + \left|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\right| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle|$$
$$\leq |\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| + \left|\langle Q_\delta \mathbf{1}_{|x|>B}, L_N\rangle\right|$$
$$+ \left|\langle f\mathbf{1}_{|x|\leq B}, L_N\rangle - \langle Q_\delta \mathbf{1}_{|x|\leq B}, L_N\rangle\right| + \left|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\right| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle|$$
$$\leq |\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| + |\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| + \left|\langle Q_\delta \mathbf{1}_{|x|>B}, L_N\rangle\right|$$
$$+ \left|\langle f\mathbf{1}_{|x|\leq B}, L_N\rangle - \langle Q_\delta \mathbf{1}_{|x|\leq B}, L_N\rangle\right| + \left|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\right| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle|$$
$$\leq |\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| + |\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle|$$
$$+ \left|\langle Q_\delta \mathbf{1}_{|x|>B}, L_N\rangle\right| + \left|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\right| + 2\delta.$$

Therefore, we can estimate the probabilities as follows (given that δ is sufficiently small):

$$\mathbb{P}\left(|\langle f, L_N\rangle - \langle f, \mathsf{SC}\rangle| > \varepsilon\right) \leq \mathbb{P}\left(|\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| > \varepsilon/5\right) + \mathbb{P}\left(|\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| > \varepsilon/5\right)$$
$$+ \mathbb{P}\left(\left|\langle Q_\delta \mathbf{1}_{|x|>B}, L_N\rangle\right| > \varepsilon/5\right) + \mathbb{P}\left(\left|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\right| > \varepsilon/5\right).$$

The first summand above convergences to zero by Chebyshev inequality:

$$\mathbb{P}\left(|\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| > \varepsilon/5\right) \leq \frac{\mathbb{E}\left(\langle Q_\delta, L_N\rangle^2\right) - \left(\mathbb{E}\langle Q_\delta, L_N\rangle\right)^2}{(\varepsilon/5)^2},$$

which goes to zero because variances go to zero (Section 2.9). The second summand convergences to zero because the moments converge (Section 2.7). The last two summands converge to zero by Lemma 2.20 (note that f is bounded and so can be bounded by a polynomial). This completes our first proof of the Wigner's semicircle law (formulated above as Theorem 2.6).

2.11. **Related laws of large numbers for random matrix spectra.** Let us mention two relatives of the Wigner semicircle law which can be proven by similar methods of moments and trees counting.

2.11.1. *Complex Wigner matrices.* The first is the law of large numbers for complex (Hermitian) random Wigner matrices $A = (a_{ij})_{i,j=1}^N$, in which

- $\overline{a_{ij}} = a_{ji}$, $i \leq j$, are complex-valued independent random variables.
- The diagonal elements a_{ii} are iid real valued with mean 0 and variance 2.

- a_{ij} with $i < j$ are iid complex random variables with expected value 0 and (complex) variance 1. On other words, $a_{ij} = x_{ij} + y_{ij}$, where x_{ij} and y_{ij} are independent real random variables with mean 0 and variance $\frac{1}{2}$. This implies that $\mathbb{E}\, a_{ij}^2 = 0$ and $\mathbb{E}\, |a_{ij}|^2 = 1$.
- All moments $\mathbb{E}\, |a_{ij}|^k$ are finite.

Defining L_N in in the same way by (2.1) (note that the eigenvalues of A are real because it is Hermitian), we still have the semicircle law: $L_N \to \mathsf{SC}$ weakly in probability. The proof of this result involves counting of graphs similarly to the real case described above, but in the case of Hermitian Wigner matrices the graphs will be directed.

There also exist laws of large numbers for complex eigenvalues of random matrices, and typical is the *circular law* stating that the eigenvalues of a random matrix with iid entries are distributed uniformly inside a unit disc on the complex plane (cf. Remark 1.2).

2.11.2. *Marchenko–Pastur law.* The second relative is the Marchenko–Pastur law [MP67] for Wishart matrices (random sample covariance matrices). The Wishart ensemble is defined as follows. Let Y be an $N \times M$ matrix of iid real-valued random variables with mean 0 and variance 1. Then, by definition, $W = YY^{\mathrm{T}}$ is called a *Wishart* $N \times N$ random matrix. It is a symmetric random matrix which is, moreover, positive definite. Therefore, all its eigenvalues are real and nonnegative.

Remark 2.21. If the entries of Y are Gaussian, then the entries of W are sums of squares of Gaussians, so they have chi square distributions — a type of distribution arising when studying sample covariance matrices of Gaussian random vectors.
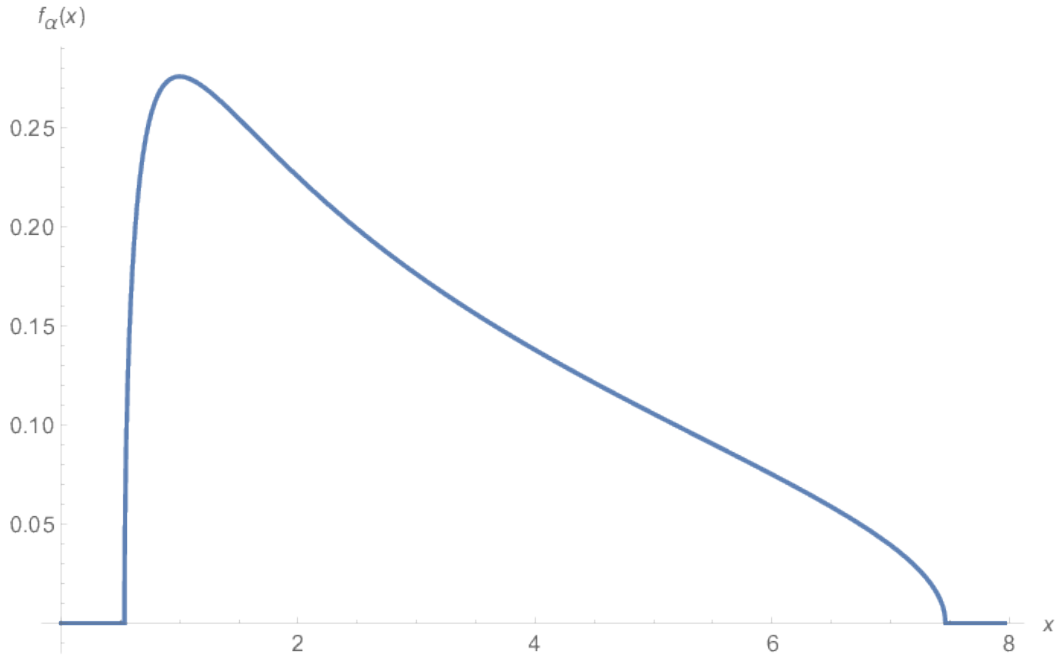

FIGURE 12. The Marchenko–Pastur density $f_\alpha(x)$ for $\alpha = 3$.

Suppose $M/N \to \alpha \in [1, \infty)$. The Marchenko–Pastur law states that

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i/\sqrt{N}} \to S_\alpha$$

weakly in probability. Here S_α, $\alpha \in [1, \infty)$ are the Marchenko–Pastur distributions defined as follows. Set

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

Then the density function of S_α is

$$f_\alpha(x) = \frac{\sqrt{(x - b_-)(b_+ - x)}}{2\pi x}, \qquad b_- \le x \le b_+,$$

which looks as in Figure 12.

Note that if $M \gg N$, then the random matrices W are likely to be nondegenerate, so the distribution does not hit zero. In fact, the distribution S_1 corresponding to $\alpha = 1$ is the image of the semicircle law SC under the squaring map $x \mapsto x^2$.

2.12. **Notes and references.** The combinatorial proof of the Wigner's semicircle law was essentially given by Wigner in [Wig55]. In our proof we closely follow Section 2.1 of [AGZ10].

Lecture #4 on 2/1/2016

3. Elements of Free Probability

3.1. **Motivating question.** We will discuss certain ideas from Free Probability applied to spectra of random matrices. Our main goal is to motivate and describe the operation of free convolution of two probability distributions with compact support. We will not discuss all the technical aspects and proofs, but will focus on certain concrete examples.

The motivating question we will be interested in is the following. Suppose A_N and B_N are two ensembles of (real symmetric or Hermitian) $N \times N$ random matrices, and assume

$$L_N(A_N) \to \mu, \qquad L_N(B_N) \to \nu,$$

where L_N is the empirical spectral distribution (2.1), and μ and ν are limiting probability measures on $\mathbb{R}$ with compact support. The above convergence is the weak convergence in probability, as in (2.3) and (2.4). The question is, how to understand the convergence of the sum:

$$L_N(A_N + B_N) \to \,?$$

Of course, the answer will involve the dependence structure of the ensembles A_N and B_N, and here we will consider one choice of this dependence — *free independence*, which informally means that the eigenbases corresponding to the matrices A_N and B_N are in generic position. Then the empirical spectral distributions of $A_N + B_N$ converge to the *free convolution* $\mu \boxplus \nu$ of the measures μ and ν.

Example 3.1. Suppose A_N and B_N are independent and real Wigner. Then by Theorem 2.6 we have $L_N(A_N) \to$ SC and $L_N(B_N) \to$ SC. However, $\dfrac{A_N + B_N}{\sqrt{2}}$ is also clearly real Wigner, and so

$$L_N\left(\frac{A_N + B_N}{\sqrt{2}}\right) \to \mathsf{SC}.$$

We will see that this corresponds to the following rule for the free convolution:

$$\frac{\mathsf{SC} \boxplus \mathsf{SC}}{\sqrt{2}} = \mathsf{SC}. \tag{3.1}$$

In classical probability, if X and Y are iid random variables such that for any constants a, b we have $aX + bY \stackrel{D}{=} cX + d$ for some other constants c, d, then X is called *stable*. An important such example is Gaussian random variables. Indeed, if X, Y are iid Gaussian random variables with mean 0 and variance 1, then $X + Y = \sqrt{2}X$. This suggests that the semicircle distribution should be the free analogue of the Gaussian distribution.

3.2. **Classical moments and cumulants.** Given a random variable X, the nth *moment* of X is

$$m_n(X) = \mathbb{E}(X^n), \qquad n \geq 0.$$

Definition 3.2. The *moment* (*exponential*[6]) *generating function* for a random variable X is given by

$$M(z) = \mathbb{E}(e^{zX}) = \sum_{n=0}^{\infty} \frac{m_n(X)z^n}{n!}. \tag{3.2}$$

Let us define a new exponential generating function $C(z) := \log(M(z))$, and call its expansion coefficients $c_n(X)$, $n \geq 1$, the *cumulants* of X:

$$C(z) = \sum_{n=1}^{\infty} \frac{c_n(X)z^n}{n!} = \log(M(z)). \tag{3.3}$$

Remark 3.3. When speaking of moments and cumulants, we will assume that the series (3.2) converges for sufficiently small z, and that the moments define the distribution of X uniquely. A sufficient assumption which we will employ in this section unless otherwise noted is that all random variables (and probability distributions) have compact support.

The cumulants (sometimes also called *half/semi-invariants*) of X form a sequence $(c_n)_{n \geq 1}$, where c_n is a homogeneous polynomial of moments m_k, $k \leq n$, with the leading term m_n:

$$
\begin{aligned}
c_1 &= m_1 = \mathbb{E}\,X, \\
c_2 &= \text{Var}(X) = m_2 - m_1^2, \\
c_3 &= \text{skewness} = m_3 - 3m_2m_1 + 2m_1^3, \\
c_4 &= \text{kurtosis} = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4, \\
&\ \ \vdots
\end{aligned}
\tag{3.4}
$$

Example 3.4 (Cumulants of the Gaussian random variable)**.** The moment generating function (3.2) of $\mathcal{N}(0,1)$, is

$$
\begin{aligned}
M(z) &= \int_{-\infty}^{\infty} e^{zx} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \\
&= e^{z^2/2} \int_{-\infty}^{\infty} \frac{e^{-(x-z)^2/2}}{\sqrt{2\pi}}\, dx \\
&= e^{z^2/2}.
\end{aligned}
$$

Hence, the cumulant generating function (3.3) is

$$C(z) = z^2/2.$$

This implies that the the cumulant sequence of $\mathcal{N}(0,1)$ is

$$(0, 1, 0, 0, ...). \tag{3.5}$$

Now, if X and Y are independent random variables, then

$$c_2(X + Y) = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = c_2(X) + c_2(Y),$$

so their second cumulants add up. In fact, this holds for all higher cumulants, too:

Proposition 3.5. *If X and Y are independent random variables, then for all $n \geq 1$,*

$$c_n(X + Y) = c_n(X) + c_n(Y). \tag{3.6}$$

---

[6]The word "exponential" refers to the presence of the factorials in the series, which resembles the Taylor series for the exponent e^z.

Proof. We have $M_X(z) = \mathbb{E}\, e^{zX}$, so $\mathbb{E}(e^{zX+zY}) = \mathbb{E}\, e^{zX}\, \mathbb{E}\, e^{zY}$ by the usual product rule for expectations of independent random variables e^{zX} and e^{zY}, which implies $\log M_{X+Y}(z) = \log M_X(z) + \log M_Y(z)$, as desired. $\qquad\square$

So, the cumuluants "linearize" addition of independent random variables. Note also that for $\alpha \in \mathbb{R}$,

$$c_n(\alpha X) = \alpha^n c_n(X). \tag{3.7}$$

To showcase the utility of the cumulants, we use them to prove the classical Central Limit Theorem (CLT):

Theorem 3.6 (Central Limit Theorem). *Let $X_1, X_2, \dots$ be iid random variables with $\mathbb{E}\, X_i = 0$ (recall that we assume compact support). Then,*

$$S_N = \frac{X_1 + \dots + X_N}{\sqrt{N}} \xrightarrow{D} \mathcal{N}(0, \mathrm{Var}(X_1)).$$

Proof sketch. Note that, by (3.6) and (3.7), for each n and N,

$$c_n(S_N) = N^{-n/2} N c_n(X_1).$$

In particular, we have:[7]

$$\begin{aligned} n = 1: &\quad c_1(S_N) = \mathbb{E}(X_1) = 0, \\ n = 2: &\quad c_2(S_N) = c_2(X_1) = \mathrm{Var}(X_1), \\ n \geq 3: &\quad c_n(S_N) \xrightarrow{N \to \infty} 0. \end{aligned}$$

So, the cumulant sequence of $\lim_{N \to \infty} S_N$ is

$$(0, \mathrm{Var}(X_1), 0, 0, \dots). \tag{3.8}$$

By comparing (3.8) and (3.5), we see that the cumulant sequences of S_N and $\mathcal{N}(0, 1)$ are the same up to a constant, which implies the CLT. $\qquad\square$

3.3. Moments and cumulants of Gaussian and semicircle distributions.
From Example 3.4 we see that the moment generating function of $\mathcal{N}(0, 1)$ is

$$M(z) = e^{z^2/2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{2^n n!} = \sum_{k=0}^{\infty} \frac{m_k z^k}{k!}.$$

Hence, $\mathcal{N}(0, 1)$ has moments

$$m_{2n+1} = 0,$$

$$m_{2n} = \frac{(2n)!}{2^n n!} = (2n-1)!! := (2n-1)(2n-3)\cdots 3 \cdot 1.$$

Furthermore, we know from (2.6) that the semicircle distribution SC distribution has moments

$$m_{2n+1} = 0,$$

$$m_{2n} = \mathrm{Cat}_n = \frac{1}{n+1}\binom{2n}{n},$$

where Cat_n is the nth Catalan number. Now, we have the following correspondence:

Classical:	moments	$\longleftrightarrow$	cumulants	$\longleftrightarrow$	$\mathcal{N}(0, 1)$ has simplest cumulants	
Random matrices:	moments	$\longleftrightarrow$	?	$\longleftrightarrow$	"cumulants" s.t. SC has simplest?	

---

[7]The argument for $c_n(S_N)$ vanishing in the limit is similar to the one from the proof of the Wigner's Semicircle Law, see (2.12) in particular.

A natural question is how the cumulants of SC look like. Turns out they are not so nice. In fact,

$$\sum_{n=0}^{\infty} z^n \mathrm{Cat}_n = \frac{1 - \sqrt{1 - 4z}}{2z},$$

which is a good algebraic function. However, the moment exponential generating function (3.2) for SC is more complicated:

$$M(z) = \sum_{n=0}^{\infty} \frac{\mathrm{Cat}_n z^n}{n!} = e^{2z}(I_0(2z) - I_1(2z)),$$

where the I_k's are the modified Bessel functions of the first kind:

$$I_\alpha(z) = \sum_{n=0}^{\infty} \frac{(z/2)^{\alpha+2n}}{n!\, \Gamma(n + 1 + \alpha)}.$$

We thus will need a nicer analogue for cumulants.

3.4. **Free cumulants.** To explain the definition of free cumulants which replace the usual cumulants in the context of random matrices in connection with the semicircle distribution, let us first investigate the combinatorial relation between the moments and the classical cumulants in detail. That is, we want to generalize the formulas (3.4) to arbitrary n. We will use the following combinatorial statement which can be found in [Sta01, Section 5.1]:

Proposition 3.7 (Exponential formula). *The moments can be expressed through cumulants as follows:*

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{B \in \pi} c_{|B|}, \tag{3.9}$$

where $\mathcal{P}(n)$ is the set of all partitions of the set $\{1, ..., n\}$, and $B \in \pi$ are the blocks of the partition π. See Figure 13 for an example.
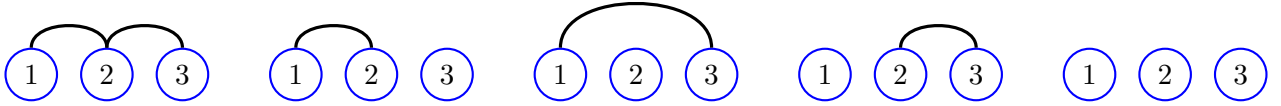

FIGURE 13. All partitions of $\{1, 2, 3\}$, yielding the expansion $m_3 = c_3 + 3c_2 c_1 + c_1^3$.

This also implies that the cumulants can be recursively determined by the moments:

$$c_n = m_n - \sum_{\substack{\pi \in \mathcal{P}(n) \\ \# \text{ of blocks} \geq 2}} \prod_{B \in \pi} c_{|B|}. \tag{3.10}$$

Proof. We start with the identity of generating functions

$$M(z) = \exp(C(z)), \tag{3.11}$$

that is,

$$\sum_{n=0}^{\infty} \frac{m_n z^n}{n!} = \sum_{k=0}^{\infty} \frac{(C(z))^k}{k!}.$$

Our aim is to use the definition of the exponential generating function $C(z)$ (3.3) to get the coefficient by z^n from the right-hand side of the above identity of the generating functions. Write

$$\frac{1}{k!}(C(z))^k = \frac{1}{k!}\left(\sum_{i=0}^{\infty} \frac{c_i z^i}{i!}\right)^k.$$

Take a partition $i_1, ..., i_k$ of n, i.e. $i_1 + ... + i_k = n$, and take the i_j-th term from each bracket above. Then, we can write

$$m_n = \sum_{\substack{k \\ i_1,...,i_k}} \frac{c_{i_1} \cdots c_{i_k} n!}{k! i_1! \cdots i_k!}. \tag{3.12}$$

Notice that $\dfrac{n!}{i_1! \cdots i_k!}$ is the multinomial coefficient, that is, the number of ways to put n labeled objects into k labeled boxes, such that there are i_j elements in the jth box. Since in the set partition of $\{1, 2, \ldots, n\}$ the blocks are unlabeled, we have to divide this multinomial coefficient by $k!$ to get the number of the corresponding set partitions. With this perspective, we can rewrite the nth moment (3.12) as (3.9), where k is the number of the blocks of the partition π, and $i_1, \ldots, i_k$ are the sizes of these blocks. $\qquad\square$

Applying (3.9) to the Gaussian distribution $\mathcal{N}(0, 1)$, we have

$$m_{2n} = (2n - 1)!! = \sum_{\pi \in \mathcal{P}(2n)} \prod_{B \in \pi} c_{|B|}. \tag{3.13}$$

However, since $c_n = 0$ for $n \neq 2$ and $c_2 = 1$, the sum is only over pair partitions (i.e. partitions in which all blocks have size 2). Hence,

$$m_{2n} = (2n - 1)!! = \#\Big\{\text{pair partitions of } \{1, 2, \ldots, 2n\}\Big\}.$$
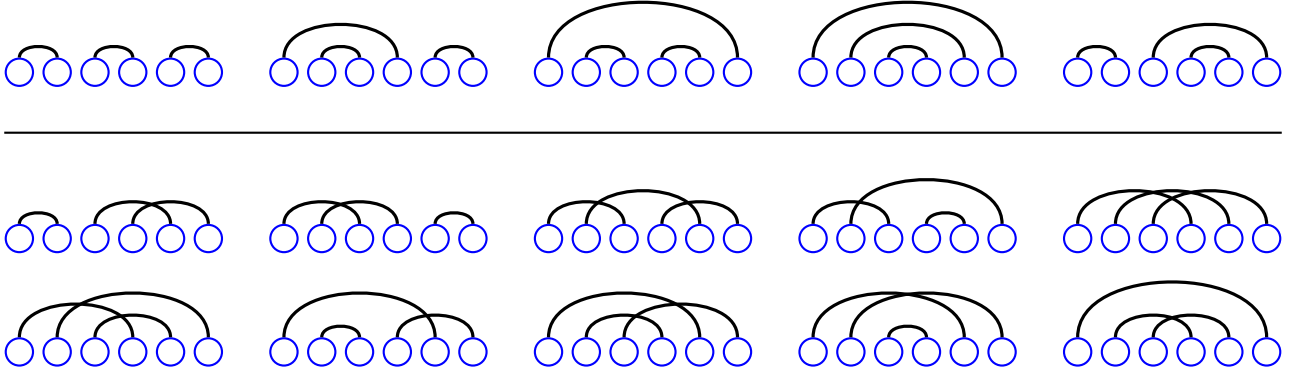
See Figure 14.

FIGURE 14. All pair partitions of the set $\{1, 2, 3, 4, 5, 6\}$. There are $5 = \text{Cat}_3$ noncrossing such partitions (the top 5), and $15 = 5!! = 5 \cdot 3 \cdot 1$ pair partitions in total.

Mimicking the classical Gaussian case, we would like to find a class $\mathcal{P}'(2n)$ of partitions of the set $\{1, 2, \ldots, 2n\}$, such that the semicircle moments have the form

$$\text{Cat}_n = m_{2n}(\mathsf{SC}) = \sum_{\pi \in \mathcal{P}'(2n)} \prod_{B \in \pi} \kappa_{|B|}, \tag{3.14}$$

where $(\kappa_n)_{n \geq 1}$ are the would-be free cumulants. We want the semicircle distribution to have the simplest free cumulants, that, is, $\kappa_2 = 1$ and $\kappa_n = 0$ for $n \neq 2$. Therefore, for the semicircle distribution the sum in (3.14) must be over pair partitions from $\mathcal{P}'(n)$.

But we already know that the Catalan number Cat_n counts the number of *noncrossing* pair partitions of $\{1, 2, \ldots, 2n\}$ — they are in a natural bijection with Dyck paths of length $2n$ (cf. Section 2.8).

The noncrossing pair partitions of $\{1, 2, 3, 4, 5, 6\}$ are given in Figure 14, top.

This motivates the following general definition:

Definition 3.8 (Free cumulants). Given a random variable X with compact support and moments $m_n = \mathbb{E}(X^n)$, its *free cumulants* are defined via the following expansion of moments

$$m_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{B \in \pi} \kappa_{|B|}, \tag{3.15}$$

where $\mathcal{NC}(n)$ is the set of noncrossing partitions of $\{1, 2, \ldots, n\}$ (cf. Figure 15).

Clearly, rewriting (3.15) similarly to (3.10) as

$$\kappa_n = m_n - \sum_{\substack{\pi \in \mathcal{NC}(n) \\ \# \text{ of blocks} \geq 2}} \prod_{B \in \pi} \kappa_{|B|}, \tag{3.16}$$

we see that the sequence of free cumulants $(\kappa_1, \kappa_2, \ldots)$ is uniquely determined by the sequence of moments.
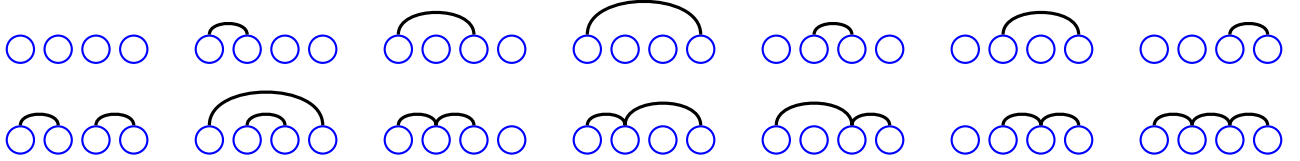

FIGURE 15. All noncrossing partitions of $\{1, 2, 3, 4\}$, yielding the expansion $m_4 = \kappa_4 + 4\kappa_3\kappa_1 + 2\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4$.

Let us give a couple of remarks:

Remark 3.9. For the "simplest" random variables — Gaussian in the classical case and the semicircle in the free case — the sequence of cumulants or free cumulants, respectively, has the form

$$(0, 1, 0, 0, \ldots).$$

Therefore, the expansions (3.13) and (3.14) involve only pair partitions. For general random variables such expansions should involve all partitions with arbitrary part sizes.

Remark 3.10. Because all partitions are noncrossing for $n \leq 3$, we have

$$c_1 = \kappa_1,$$
$$c_2 = \kappa_2,$$
$$c_3 = \kappa_3.$$

3.5. **Example. Counting graphs.** It is not essential that the quantities m_n and c_n in Proposition 3.7 are moments and cumulants of a certain random variable. Indeed, the exponential formula (3.9) holds for coefficients of any exponential generating series related as (3.11).

The combinatorial meaning of the exponential formula (3.9) is the following. If the quantities m_n enumerate certain kinds of objects on n labeled vertices, then the quantities c_n related to m_n via (3.9) or (3.10) count such objects which are *connected*.

In particular, for counting undirected graphs on n labeled vertices:

$$m_n = \# \text{ of graphs } = 2^{\binom{n}{2}},$$
$$c_n = \# \text{ of connected graphs}.$$

This property that the c_n's count connected objects is clear from (3.9): we count the total number of objects by splitting each object into connected components (this corresponds to picking a partition π), and then counting the numbers of connected objects on vertices inside these connected components.

Arguing in a similar way with the noncrossing cumulant formula (3.15), we see that if the moments $m_n = 2^{\binom{n}{2}}$ count graphs, then the free cumulants κ_n count geometrically connected graphs on n labeled vertices. The geometrically connectedness condition naturally leads to splitting into noncrossing partitions. See Figure 16.
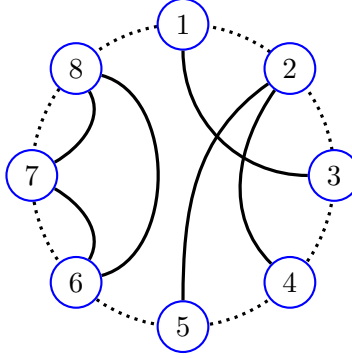

FIGURE 16. The geometric connectedness means that the graph on n labeled vertices located on the circle is connected as a geometric figure if not as a graph. The displayed graph has 3 connected and 2 geometrically connected components.

Lecture #5 on 2/3/2016

3.6. **Free independence.** Recall that in the classical case, for independent random variables X and Y and each $n \geq 1$,

$$c_n(X + Y) = c_n(X) + c_n(Y).$$

So, we would like to define free independence such that the same holds for free cumulants. That is, if X and Y are free independent random variables, then for all $n \geq 1$,

$$\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y).$$

To make a more precise definition of the free independence that we will work with, we need joint cumulants (classical and free). That is, we want to speak not only about the variance and its higher analogues, but also about the covariance and its higher analogues (for two or more random variables). This construction is rather straightforward: the joint moments are defined as

$$m_n(X_1, ..., X_n) = \mathbb{E}[X_1, ..., X_n],$$

and the joint cumulants — by the expansion

$$m_n(X_1, ..., X_n) = \sum_{\pi \in \mathcal{P}(n)} \prod_{B \in \pi} c_{|B|}(X_i \colon i \in B).$$

To get the definition of the joint free cumulants, simply replace $\mathcal{P}(n)$ in the above formula by $\mathcal{NC}(n)$, and $c_{|B|}(X_i \colon i \in B)$ by $\kappa_{|B|}(X_i \colon i \in B)$:

$$m_n(X_1, ..., X_n) = \sum_{\pi \in \mathcal{NC}(n)} \prod_{B \in \pi} \kappa_{|B|}(X_i \colon i \in B). \tag{3.17}$$

Note that from (3.17) (more precisely, rewriting (3.17) in a recursive form like (3.16)) we see that κ_n is a symmetric multilinear function of its arguments.

Definition 3.11 (Free independence of classical random variables)**.** Two random variables X and Y with $\mathbb{E}\,X = \mathbb{E}\,Y = 0$ are said to be *free independent* if all their joint mixed free cumulants vanish for all $n \geq 2$:

$$\kappa_n(X, Y, Y, \ldots, Y, Y) = \kappa_n(X, X, Y, \ldots, Y, Y) = \ldots = \kappa_n(X, X, X, \ldots, X, Y) = 0.$$

Exercise 3.12. Two random variables X and Y with $\mathbb{E}\,X = \mathbb{E}\,Y = 0$ are classically independent if and only if all their joint mixed classical cumulants are zero: $c_n(X, X, \ldots X, Y, \ldots, Y) = 0$.

Example 3.13. Note that this will not be the same as regular independence. Indeed, if you have two free independent SC random variables X and Y, then

$$X + Y \overset{D}{=} \sqrt{2} \cdot \mathsf{SC}.$$

For classically independent X and Y, their sum $X + Y$ clearly does not have the distribution $\sqrt{2} \cdot \mathsf{SC}$ (in particular, SC is not a stable distribution).

It turns out that free independence of classical random variables is problematic. To see this, consider free independent classical random variables X and Y. Then,

$$m_4(X, X, Y, Y) = m_4(X, Y, X, Y),$$

where

$$m_4(X, X, Y, Y) = \kappa_2(X, X)\kappa_2(Y, Y) + \kappa_2(X, X) + \kappa_1(Y)^2 + \kappa_2(Y, Y)\kappa_1(X)^2 + \kappa_1(X)^2\kappa_1(Y)^2,$$

and

$$m_4(X, Y, X, Y) = \kappa_2(X, X)\kappa_1(Y)^2 + \kappa_1(X)^2\kappa_2(Y, Y) + \kappa_1(X)^2\kappa_2(Y)^2.$$

Hence $\kappa_2(X, X)\kappa_2(Y, Y) = 0$, which means that, for classical random variables, free independence requires one of the random variables to be constant.

The solution is to introduce noncommutativity, whence

$$m_4(X, X, Y, Y) \neq m_4(X, Y, X, Y),$$

and, in fact, $XXYY \neq XYXY$. This is the motivation for the definitions in the following subsection.

3.7. **Noncommutative spaces of random variables.** The idea of free random variables was invented by Voiculescu in the beginning of the 1980s, cf. [VDN92]. The space $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ of bounded random variables is a commutative von Neumann algebra. Hence, noncommutative random variables would live in a noncommutative von Neumann algebra.

To define a von Neumann algebra, first, let H be a Hilbert space (complete inner product space over $\mathbb{C}$) and $\mathcal{B}(H)$ be the space of all bounded linear operators on H (bounded with respect to the norm induced by the inner product $(\cdot|\cdot)$ on H). The space $\mathcal{B}(H)$ has a natural involution given by

$$(Tx|y) = (x|T^*y)$$

for any $T \in \mathcal{B}(H)$ and $x, y \in H$. Elements of $\mathcal{B}(H)$ of the form T^*T are called *nonnegative*.

In addition to the topology on $\mathcal{B}(H)$ induced by the norm, there is a *weak operator topology*. We say that for $T_n, T \in \mathcal{B}(H)$, $T_n \to T$ as $n \to \infty$ in the weak operator topology if for any $x, y \in H$,

$$(T_n x|y) \to (Tx|y), \qquad n \to \infty.$$

Definition 3.14. A *von Neumann algebra* $\mathcal{A}$ is a unital (i.e., $1 \in \mathcal{A}$) $*$-closed subalgebra of $\mathcal{B}(H)$, which is closed with respect to the weak operator topology.

We would like to replace $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ with a von Neumann algebra $\mathcal{A}$. To complete the picture, we would need an analogue of the expectation which in the commutative case comes from the probability distribution. In lieu of the classical expectation, we have a notion of tracial state on our von Neumann algebra.

We will not consider the most general definition of a tracial state on a von Neumann algebra, but will deal with functionals $\tau \colon \mathcal{A} \to \mathbb{C}$ satisfying

(1) $\tau(1) = 1$;
(2) $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}_+$ (here $\mathcal{A}_+$ is the subset of nonnegative elements in $\mathcal{A}$);
(3) $\tau(a^*a) \geq 0$ for all $a \in \mathcal{A}$.

The tracial state is *faithful* if $\tau(a^*a) = 0$ implies $a = 0$.

So, we have the following correspondences:

Classical	Noncommutative
$L^\infty(\Omega, \mathcal{F}, P)$	a von Neumann algebra $\mathcal{A}$
$\mathbb{E}[\cdot]$	$\tau[\cdot]$
$m_n(X)$	$m_n(a) = \tau(a^n), \ a \in \mathcal{A}_+$

Now, we define free independence for noncommutative random variables.

Definition 3.15. Elements a and b in a von Neumann algebra $\mathcal{A}$ are *free independent* if for any polynomials $f_1, g_1, ..., f_k, g_k$ such that $\tau(f_i(a)) = \tau(g_j(b)) = 0$, we have

$$\tau(f_1(a)g_1(b)f_2(a)g_2(b) \cdots g_k(b)) = 0.$$

Note that this is essentially mimicking Definition 3.11 stating that joint mixed cumulants are zero.

Compare this to the classical independence:

Definition 3.16. Elements a and b in a von Neumann algebra $\mathcal{A}$ are *classically independent* if $ab = ba$ and for any polynomials f and g such that $\tau(f(a)) = \tau(g(b)) = 0$, we have

$$\tau(f(a)g(b)) = 0.$$

If the moments $m_n(a)$ of an element $a \in \mathcal{A}$ can be identified with moments of a compactly supported probability distribution on $\mathbb{R}$, we will denote this distribution by μ_a. That is, talking about a single random variable yields the same results in commutative or noncommutative setting.

Definition 3.17 (Free convolution). If $a, b \in \mathcal{A}$ are free independent and correspond to compactly supported probability distributions μ_a and μ_b on $\mathbb{R}$, respectively, then their sum $a + b$ corresponds to a new compactly supported probability distribution μ_{a+b} on $\mathbb{R}$. This distribution μ_{a+b} is called the *free convolution* of the distributions μ_a and μ_b, and is denoted by

$$\mu_{a+b} = \mu_a \boxplus \mu_b.$$

As we will see below in Proposition 3.19 and Section 3.9, working with free convolution does not require von Neumann algebras.

To apply free probability to random matrices, we will need the following definition:

Definition 3.18. For a sequence $(\mathcal{A}_N, \tau_N)$ of von Neumann algebras with tracial states, and sequences (a_N) and (b_N) with $a_N, b_N \in \mathcal{A}_N$ for each N, we say that (a_N) and (b_N) are *asymptotically free* if for all polynomials $f_1, g_1, ..., f_k, g_k$ such that $\tau(f_i(a_N)) \to 0$ and $\tau(g_j(b_N)) \to 0$, we have

$$\tau(f_1(a_N)g_1(b_N)f_2(a_N)g_2(b_N) \cdots g_k(b_N)) \to 0.$$

An example of a sequence $(\mathcal{A}_N, \tau_N)$ of von Neumann algebras is given by random matrices. Take $\mathcal{A}_N$ to be the von Neumann algebra of Hermitian random matrices with almost surely bounded entries.[8] Equip each $\mathcal{A}_N$ with a tracial state

$$\tau_N(A) := \frac{1}{N} \mathbb{E} \operatorname{tr}(A), \qquad A \in \mathcal{A}_N,$$

---

[8]One can also take real symmetric random matrices as we did in Section 2. However, from now on we will mostly work with Hermitian complex matrices, so we are switching to this setting.

coming from the usual matrix trace. Recall that quantities like $\tau_N(A)$ played a major role in the proof of the Wigner's semicircle law in Section 2.

Proposition 3.19. *Consider sequences $(A_N), (B_N)$ of $N \times N$ Hermitian random matrices from $\mathcal{A}_N$ with*

$$L_N(A_N) \to \mu, \qquad L_N(B_N) \to \nu,$$

as discussed in Section 3.1. Assume that μ and ν are compactly supported. For each N, let U_N be the random Haar-distributed unitary matrix that is independent of A_N and B_N.

Then the random matrices A_N and $U_N B_N U_N^*$ from $\mathcal{A}_N$ are asymptotically free.

Proof. We will not prove this statement as it requires a graph enumeration somewhat similar to the proof of the semicircle law. See, e.g., [NL12, §3.6] for an outline of the proof and more references. □

This proposition shows that the abstract definition of the free convolution (Definition 3.17) can be realized more concretely using summation of random matrices, one of which is randomly rotated, so that the eigenbases of A_N and $U_N B_N U_N^*$ are in "completely generic" position.

3.8. **Semicircle law again.** Using Proposition 3.19, we can give another proof of the Wigner's semicircle law, now for the Gaussian Unitary Ensemble (GUE). First, note that a free analogue of the Central Limit theorem states that if $X_1, X_2, \ldots$ is a family of free independent identically distributed elements of $(\mathcal{A}, \tau)$ with mean 0 and variance 1,[9] then

$$\frac{X_1 + \ldots + X_N}{\sqrt{N}} \to \mathsf{SC}, \qquad N \to \infty,$$

in the sense that for any $n \geq 0$ we have

$$\tau\left[\left(\frac{X_1 + \ldots + X_N}{\sqrt{N}}\right)^n\right] \to m_n(\mathsf{SC}) = \begin{cases} 0, & n \text{ odd}; \\ \text{Cat}_{n/2}, & n \text{ even}. \end{cases} \tag{3.18}$$

This is proven using free cumulants exactly in the same way as in the proof of the classical Central Limit Theorem using the classical cumulants (Theorem 3.6).

Let us now sketch a proof of the semicircle law for the Gaussian Hermitian matrices. Let G_N be an $N \times N$ GUE random matrix (i.e., an Hermitian random matrix with iid standard real Gaussians on the diagonal, and iid standard complex Gaussians — with independent real and imaginary parts having Gaussian distribution with variance $\frac{1}{2}$ — above the diagonal). We would like to show that $L_N(G_N) \to \mathsf{SC}$ weakly in probability. We have seen in Section 2 that for this it is enough to show the convergence of the moments as in (3.18). Let $G_N^{(1)}, \ldots, G_N^{(N)}$ be independent copies of G_N. By the unitary invariance of the GUE (see Section 4.1 below), for for $U_N^{(k)}, 1 \leq k \leq N$ Haar-distributed unitary matrices independent of everything else, we have $U_N^{(k)} G_N^{(k)} (U_N^{(k)})^* \overset{D}{=} G_N$. Furthermore, we clearly have by the Gaussian nature of the elements of the matrix:

$$G_N \overset{D}{=} \frac{U_N^{(1)} G_N^{(1)} (U_N^{(1)})^* + \ldots + U_N^{(N)} G_N^{(N)} (U_N^{(N)})^*}{\sqrt{N}}.$$

Finally, by Proposition 3.19, the summands above are asymptotically free. Thus, by the free CLT, $L_N(G_N)$ converges to the semicircle distribution SC.

---

[9]That is, $\tau(X_1) = 0$ and $\tau(X_1^2) = 1$.

3.9. **Free convolution and Voiculescu's algorithm.** Let us now describe the operation of the free convolution $(\mu, \nu) \mapsto \mu \boxplus \nu$ of Definition 3.17 (where μ and ν are probability distributions on $\mathbb{R}$ with compact support) in more concrete terms.

We would like to describe an algorithm for computing $\mu \boxplus \nu$ given μ and ν. Recall that under free convolution we have $\kappa_n(\mu \boxplus \nu) = \kappa_n(\mu) + \kappa_n(\nu)$ for each $n \geq 1$. We will use the following fact relating (ordinary, not exponential) moment generating function of a measure and its free cumulant generating function:

$$L(z) = 1 + \sum_{n=1}^{\infty} m_n z^n, \qquad K(z) = 1 + \sum_{n=1}^{\infty} \kappa_n z^n, \tag{3.19}$$

Proposition 3.20. *Thus defined generating series are related as follows:*

$$L(z) = K\big(zL(z)\big). \tag{3.20}$$

Proof. This is the free analogue of the exponential formula (Proposition 3.7), which can be proven by combinatorial manipulations with generating series. We will not give its proof here. □

Remark 3.21. Relation (3.20) between infinite power series can be used to compute coefficients of $L(z)$ if the coefficients of $K(z)$ are known, and vice versa. This is done in a purely formal way, by recursively looking at coefficients by powers of z in both sides. Thus, (3.20) shows that moments and free cumulants uniquely determine each other. Of course, this uniqueness already follows from the combinatorial relations (3.15) and (3.16).

Now, take the *Cauchy transform* (sometimes also called the *Stieltjes transform*) of each measure:

$$
\begin{aligned}
G_\mu(z) &= \int_{\mathbb{R}} \frac{d\mu(x)}{z - x} \\
&= \int_{\mathbb{R}} \frac{z^{-1} d\mu(x)}{1 - x/z} \\
&= \int_{\mathbb{R}} \sum_{n=0}^{\infty} z^{-1} \frac{x^n}{z^n} d\mu(x) \\
&= \sum_{n=0}^{\infty} \int_{\mathbb{R}} z^{-1} \frac{x^n}{z^n} d\mu(x) \\
&= \sum_{n=0}^{\infty} \frac{m_n(\mu)}{z^{n+1}}
\end{aligned}
$$

for large enough $|z|$, and similarly for $G_\nu(z)$ (recall that the measures have finite support). Using (3.19) and Proposition 3.20, define

$$V_\mu(z) := \frac{1}{z} K_\mu(z),$$

and similarly for ν. Proposition 3.20 readily implies that the functions G_μ and V_μ are mutual inverses:

$$V_\mu\big(G_\mu(z)\big) = G_\mu\big(V_\mu(z)\big) = z. \tag{3.21}$$

Example 3.22 (Semicircle distribution)**.** We can compute G_{SC} using the Cauchy Transform:

$$
\begin{aligned}
G_{\mathsf{SC}}(z) &= \int_{-2}^{2} \frac{\frac{1}{2\pi}\sqrt{1 - 4x^2}}{z - x} dx \\
&= \sum_{n=0}^{\infty} \frac{\mathrm{Cat}_n}{z^{2n+1}}
\end{aligned}
$$

$$= \frac{1}{z}\left(\frac{1 - \sqrt{1 - 4/z^2}}{2}\right)$$

$$= \frac{1}{2}z - \frac{1}{2}\sqrt{z^2 - 4}.$$

We have computed this using the explicit formula for the Catalan numbers (2.7). This expansion can be verified by simply Taylor expanding $\frac{1}{2}z - \frac{1}{2}\sqrt{z^2 - 4}$ at $z = \infty$.

On the other hand, since we know that the free cumulant sequence for SC is $(0, 1, 0, 0, ...)$, we can compute

$$V_{\mathsf{SC}}(z) = \frac{1}{z} + z.$$

Hence, we can regain G_{SC} from (3.21):

$$\frac{1}{G} + G = z \Rightarrow G^2 - zG + 1 = 0$$

$$\Rightarrow G = \frac{1}{2}(z - \sqrt{z^2 - 4}).$$

Therefore, we have established the following algorithm for computing the free convolution of compactly supported measure μ and ν on $\mathbb{R}$:

Theorem 3.23 (Voiculescu's Algorithm for Free Convolution). *To compute the free convolution $\mu \boxplus \nu$, follows these steps:*

(1) *Compute the Cauchy transforms $G_\mu(z) = \mathbb{E}(z - X)^{-1}$ and $G_\nu(z) = \mathbb{E}(z - Y)^{-1}$ (here X and Y are distributed as μ and ν, respectively).*
(2) *Solve $G_\mu(V_\mu(z)) = z$ and $G_\nu(V_\nu(z)) = z$ subject to $V(z) \sim 1/z$ at 0.*
(3) *Compute Voiculescu's R-transforms*

$$R_\mu(z) = V_\mu(z) - 1/z, \qquad R_\nu(z) = V_\nu(z) - 1/z.$$

These R-transforms linearize the free convolution:

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z),$$

so $V_{\mu \boxplus \nu}(z) = V_\mu(z) + V_\nu(z) - 1/z$.
(4) *Finally, solve $V_{\mu \boxplus \nu}(G_{\mu \boxplus \nu}(z)) = z$ subject to $G(z) \sim 1/z$ at ∞.*
(5) *Recover the distribution $\mu \boxplus \nu$ from its Cauchy transform $G_{\mu \boxplus \nu}$.*

This algorithm essentially involves only formal manipulations with power series and rational functions in one variable z, and thus can be implemented to compute an arbitrary number of first terms of the Cauchy transform $G_{\mu \boxplus \nu}$ for arbitrary input $G_\mu(z)$ and $G_\nu(z)$.

Lecture #6 on 2/8/2016

To recover a distribution from its Cauchy transform requires one more step. In principle, having a distribution μ and its Cauchy transform

$$G_\mu(z) = \sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}}, \tag{3.22}$$

one can take the moments m_n and construct the exponential generating function $\mathbb{E}\, e^{zX} = \int_{\mathbb{R}} e^{zx}\mu(dx)$. The exponential generating function can then be inverted using the usual inverse Fourier transform. However, there is a more direct approach:

Proposition 3.24 (Inversion of Cauchy transform). *Let μ be an absolutely continuous distribution (i.e., it has density $\rho(t)$) supported inside a finite interval $I \subset \mathbb{R}$. Then we have*

$$\rho(x) = \lim_{y \to 0^+} \frac{1}{\pi} Im(G(x + iy)),$$

where $i = \sqrt{-1}$ and Im means the imaginary part of a complex number.

Proof. Setting $z = x + iy$, we get

$$G(z) = \int_I \frac{\rho(t)\, dt}{z - t}$$
$$= \int_I \rho(t) \left[\frac{x - t}{(x - t)^2 + y^2} - i\frac{y}{(x - t)^2 + y^2} \right] dt.$$

Now for any $a < b$ we can find the following integral by interchanging the integrations (allowed by Fubini's theorem):

$$\int_a^b Im(G(x + iy))\, dx = \int_I \rho(t)\, dt \int_a^b \frac{-y}{y^2 + (x - t)^2}\, dx$$
$$= \int_I \rho(t)\, dt\, \arctan\left(\frac{x - t}{y}\right) \Big|_{x=a}^{x=b}$$
$$= \int_I \rho(t)\, dt \left[\arctan\left(\frac{b - t}{y}\right) - \arctan\left(\frac{a - t}{y}\right) \right].$$

As $y \to 0^+$, we have $\arctan\left(\frac{b-t}{y}\right) - \arctan\left(\frac{a-t}{y}\right) = 0$ for $t \notin [a, b]$. Moreover,

$$\arctan\left(\frac{b - t}{y}\right) - \arctan\left(\frac{a - t}{y}\right) \to \frac{\pi}{2} - \frac{-\pi}{2} = \pi, \qquad y \to 0^+.$$

Hence,

$$\int_a^b Im(G(x + iy))\, dx \to \int_I \pi\rho(t)\, dt, \qquad y \to 0^+.$$

This completes the proof. $\qquad \square$

Exercise 3.25. Derive a generalization of the inversion formula of Proposition 3.24 in the case when μ is not necessarily absolutely continuous (but still compactly supported).

3.10. **Examples of free convolution.** Let us conclude with a couple of concrete examples.

Example 3.26 (Bernoulli $\boxplus$ Bernoulli). Suppose $\mu = \nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, that is, the corresponding random variables take values ± 1 with probabilities $\frac{1}{2}$. Note that classical convolution yields

$$\mu * \nu = \frac{1}{4}\delta_{-2} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_2.$$

To find $\mu \boxplus \nu$, we follow Voiculescu's Algorithm (Theorem 3.23):

(1) $G(z) = \mathbb{E} \frac{1}{z - x} = \frac{1}{2}\left(\frac{1}{z + 1} + \frac{1}{z - 1}\right) = \frac{z}{z^2 - 1}$.

(2) Solving $z = G(V(z)) = \frac{V}{V^2 - 1}$ yields $zV^2 - V - z = 0$. So,

$$V(z) = \frac{1 \pm \sqrt{1 + 4z^2}}{2z}.$$

Since $V(z) = \frac{K(z)}{z} \sim \frac{1}{z}$ at $z = 0$, we choose

$$V(z) = \frac{1 + \sqrt{1 + 4z^2}}{2z}.$$

(3) $V_{\mu \boxplus \nu}(z) = V_\mu(z) + V_\nu(z) - \frac{1}{z} = \frac{\sqrt{1 + 4z^2}}{z}.$

(4) Solving $z = V_{\mu \boxplus \nu}(G_{\mu \boxplus \nu}(z)) = \frac{\sqrt{1 + 4G^2}}{G}$ yields

$$G_{\mu \boxplus \nu}(z) = \pm(z^2 - 4)^{-1/2}.$$

Since $zG_{\mu \boxplus \nu}(z) \to 1$ at $z = 0$, we choose

$$G_{\mu \boxplus \nu}(z) = \frac{1}{\sqrt{z^2 - 4}}.$$

(5) From Proposition 3.24, we see that the probability density $\rho(x)$ for $\mu \boxplus \nu$ is

$$\rho(x) = \frac{1}{\pi} \lim_{y \to 0^+} Im \left(\frac{1}{\sqrt{(x + iy)^2 - 4}} \right)$$

$$= \begin{cases} 0, & |x| > 2; \\ \dfrac{1}{\pi\sqrt{4 - x^2}}, & |x| \leq 2. \end{cases}$$

This is the *arcsine distribution* (Figure 17). We see that this is quite different from the classical convolution.
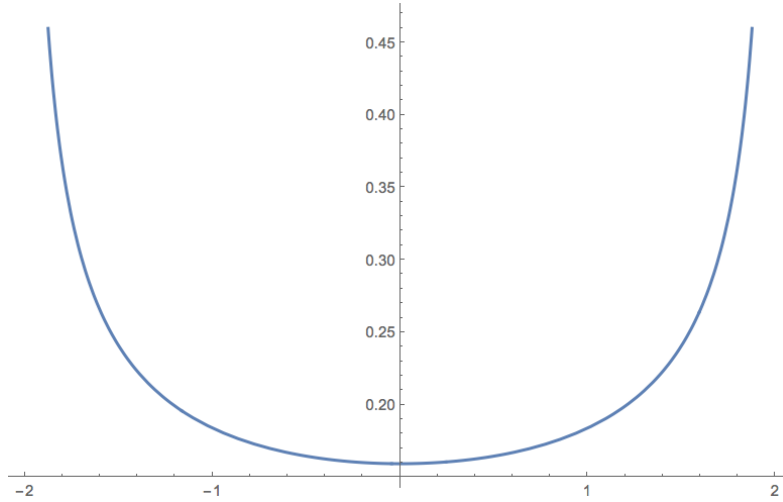

FIGURE 17. The arcsine density.

Exercise 3.27 (Free Poisson Theorem)**.** Let $\alpha > 0, \lambda > 1$, and consider the Bernoulli density

$$\mu_N = (1 - \lambda/N)\delta_0 + \lambda/N\delta_\alpha$$

supported by 0 and α. Classically,

$$\mu_N^{*N} \to \text{Poisson}(\lambda) \text{ on } \{0, \alpha, 2\alpha, \dots\}, \qquad N \to \infty.$$

In the free situation, by computing the R-transform, multiplying it by N, and then inverting, show that $\mu_N^{\boxplus N}$ converges to the density $\frac{1}{2\pi\alpha t}\sqrt{4\lambda\alpha^2 - (t - \alpha(1 + \lambda))^2}$, which is the Marchenko–Pastur distribution

(Section 2.11.2) under a change of parameters. If $\lambda \leq 1$, then the limiting distribution is different. Namely, it has an atom of size $1 - \lambda$ at 0, and the rest of the mass λ is distributed according to the same Marchenko–Pastur law.

Exercise 3.28 (Cauchy Density)**.** Show that for the Cauchy density $C = \frac{1}{\pi(1+x^2)}$ on $\mathbb{R}$,

$$\mu * C = \mu \boxplus C,$$

i.e., the free convolution with the Cauchy density is the same as the classical convolution. See [AGZ10, Exercise 5.3.41] for more hints.

3.11. **Notes and references.** Free Probability was invented by Voiculescu in the 1980s to study operator algebras (e.g., [VDN92] gives an early introduction). In particular, the word "free" comes from the free groups. Studying operator algebras associated with the free groups was one of the motivations for developing Free Probability. In our discussions we mainly follow Chapter 5 of [AGZ10], and also lecture notes [NL12] which provide a lot of combinatorial insights into Free Probability.

4. Gaussian Unitary Ensemble and its asymptotic behavior

4.1. **Definition and basic properties.** Recall that by $\mathcal{H}_N$ we denote the set of $N \times N$ Hermitian matrices over $\mathbb{C}$.

Definition 4.1. The law of the Gaussian Unitary Ensemble (GUE) is a probability distribution on $\mathcal{H}_N$, and the law of a random matrix $A = (a_{ij})_{i,j=1}^N \in \mathcal{H}_N$ is described as follows:

- $a_{ii} \sim \mathcal{N}(0,1)$ for each i.
- $a_{ij} = \bar{a}_{ji} \sim \mathcal{N}(0,1/2) + \sqrt{-1}\,\mathcal{N}(0,1/2)$ for $i < j$.

All real normal random variables involved above are assumed to be independent. Note that $\mathbb{E}\,a_{ij} = 0$ for all i, j.

Now fix any (nonrandom) $A \in \mathcal{H}_N$. Then we can write

$$\mathrm{tr}(A^2) = \mathrm{tr}(AA) = \sum_{i=1}^N a_{ii}^2 + \sum_{i \neq j} a_{ij}a_{ji} = \sum_{i=1}^N a_{ii}^2 + 2\sum_{i<j} |a_{ij}|^2.$$

Identify $\mathcal{H}_N$ with $\mathbb{R}^{N^2}$ with the standard Lebesgue measure which in the matrix coordinates can be written as

$$dA = \prod_i da_{ii} \prod_{i<j} da_{ij}^{\Re} da_{ij}^{\Im}.$$

Here the superscripts $\Re$ and $\Im$ refer to the real and imaginary parts.

The GUE distribution μ on $\mathcal{H}_N$ has probability density (with respect to the above Lebesgue measure)

$$
\begin{aligned}
d\mu(A) &= \frac{1}{Z_N} e^{-\sum_i a_{ii}^2/2} e^{-\sum_{i<j}(a_{ij}^{\Re})^2 - \sum_{i<j}(a_{ij}^{\Im})^2} dA \\
&= \frac{1}{Z_N} e^{-\frac{1}{2}(\sum_i a_{ii}^2 + 2\sum_{i<j}|a_{ij}|^2)} dA \\
&= \frac{1}{Z_N} e^{-\frac{1}{2}\mathrm{tr}(A^2)} dA,
\end{aligned}
$$

where $Z_N = (\sqrt{\pi})^{N^2}(\sqrt{2})^N$ is the normalization constant (its value follows from the normalization of the Gaussian distribution).

This computation shows that the GUE density is *unitary invariant*. That is, if A is a random matrix with GUE distribution and $U \in U(N)$ is a fixed unitary matrix, then $UAU^{-1} \overset{\mathcal{D}}{=} A$ (equality in distribution).

4.2. **Joint eigenvalue density.** Let $A \in \mathcal{H}_N$ have the GUE distribution. Denote its eigenvalues (which are necessarily real) by $\lambda_1 \geq \cdots \geq \lambda_N$.

Exercise 4.2. Show that under the GUE distribution the eigenvalues of the matrix are almost surely pairwise distinct.

We will now derive the joint density of eigenvalues of the GUE random matrix:

Theorem 4.3. *Let $A = A_N \in \mathcal{H}_N$ have the GUE distribution. Then the joint density of the eigenvalues of A with respect to the Lebesgue measure on $\mathbb{R}^N$ is given by*

$$C_N \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{N} e^{-\lambda_i^2/2}.$$

We will not compute the value of the normalization constant C_N right now, but later in **??** we will see that it is equal to $\dfrac{(2\pi)^{-N/2}}{0!1!\cdots(N-1)!}$.

Proof. We will in fact prove a more general result. Namely, let f be a function on the space $\mathcal{H}_N$ of Hermitian $N \times N$ matrices which is unitary invariant, i.e. $f(UAU^*) = f(A)$ for any fixed unitary matrix U. (This implies that f depends only on the eigenvalues $\lambda_1, \ldots, \lambda_N$ of A, and moreover, in a symmetric way.) We will show that

$$\int_{\mathcal{H}_N} f(A)dA = \mathrm{const}_N \int_{\mathbb{R}^N} f(\lambda_1, \ldots, \lambda_N) \prod_{i<j} |\lambda_i - \lambda_j|^2 \, d\lambda_1 \cdots d\lambda_N, \tag{4.1}$$

where both integrals are with respect to the corresponding Lebesgue measures. In other words, we aim to find the Jacobian of the transformation mapping from A to the set of its eigenvalues.

Example 4.4. As an example, consider the change of coordinates in $\mathbb{R}^2$ from Cartesian to polar: $(x, y) \to (r, \theta)$. If $f(x, y)$ is independent of θ, i.e. it is radially symmetric, then

$$\iint_{\mathbb{R}^2} f(x, y) \, dxdy = 2\pi \int_0^\infty f(r) r dr.$$

The factor r comes from the Jacobian, and so $2\pi r = \mathrm{const} \cdot r$ is the *radial part* of the Lebesgue measure on $\mathbb{R}^2$. What we are about to do now is a generalization of this procedure. That is, we will find the radial part of the Lebesgue measure on the space of $N \times N$ Hermitian matrices.

Continuing with the proof, view $\mathcal{H}_N$ as a space with an action of the unitary group $U(N)$ by conjugations: $A \mapsto UAU^*$. A typical matrix[10] $A \in \mathcal{H}_N$ can be written as $U\Lambda U^*$, where Λ is the diagonal matrix of eigenvalues of A. This gives rise to a change of variables

$$A \leftrightarrow (\vec{\lambda}, \dot{U}),$$

where $\vec{\lambda}$ is the vector of eigenvalues and $\dot{U} = U \mod T$, where T is the torus of diagonal matrices which stabilize $\vec{\lambda}$. The matrix $\dot{U}$ represents the rotation to the eigenbasis of A.

The Jacobian we aim to compute is $dA = (?)d\vec{\lambda}d\dot{U}$, where $d\dot{U}$ is Haar measure. We note that the Jacobian must be independent of U since the Lebesgue measure is rotationally invariant.

Let us write $U = 1 + X$ where X is "small". In other words, we are applying a Lie algebraic perspective. Since U is unitary, the matrix X must be skew-Hermitian: $X + X^* = 0$. This leaves N^2 real parameters in X (which corresponds to the well-known fact that $U(N)$ is N^2-dimensional over $\mathbb{R}$). Then $U^* = U^{-1} = 1 - X +$ smaller order terms. Since $A = U\Lambda U^*$, we can write

$$A + dA = (1 + X)(\Lambda + d\Lambda)(1 - X) = \Lambda + d\Lambda + X\Lambda - \Lambda X + \text{smaller order terms}.$$

---

[10]That is, almost every matrix with respect to the GUE (or, equivalently, the Lebesgue) measure on $\mathcal{H}_N$.

Hence small changes in diagonal terms in A are

$$da_{ii} = d\lambda_i,$$

while small changes in the off-diagonal terms have the form (here $j < k$)

$$da_{jk} = x_{jk}\lambda_k - \lambda_j x_{jk} = (\lambda_k - \lambda_j)x_{jk}.$$

Since both da_{jk} and x_{jk} are complex, the contribution $(\lambda_k - \lambda_j)$ affects both $da_{jk}^{\Re}$ and $da_{jk}^{\Im}$, so overall we see that the the Jacobian is $\prod_{j<k}|\lambda_j - \lambda_k|^2$.

The theorem then follows from (4.1) by setting $f(A) = e^{-\frac{1}{2}\mathrm{tr}(A^2)}$. $\qquad\square$

4.3. **Examples of other random matrix ensembles with explicit eigenvalue density.** The eigenvalue density result of Theorem 4.3 opens a way to generalize the Gaussian Unitary Ensemble in several directions. We can express the joint eigenvalue density of a GUE using the Vandermonde determinant

$$V(\lambda) = \prod_{i<j}|\lambda_i - \lambda_j|.$$

Exercise 4.5. Show that $V(\lambda) = \pm\det[\lambda_i^{j-1}]_{i,j=1}^N$.

The density of eigenvalues of Theorem 4.3 is

$$C_N^{GUE}V(\lambda)^2 \prod_{i=1}^N e^{-\frac{1}{2}\lambda_i^2}.$$

4.3.1. *Beta ensembles.* Certain other Gaussian random matrix ensembles defined similarly to the GUE have eigenvalue densities also involving $V(\lambda)$.

Recall that a real symmetric matrix A has Gaussian Orthogonal Ensemble (GOE) distribution if $a_{ii} \sim N(0,2)$ and $a_{ij} = a_{ji} \sim N(0,1)$ for $i < j$. Then the GOE eigenvalue density is

$$C_N^{GOE}|V(\lambda)| \prod_{i=1}^N e^{-\lambda_i^2/4}$$

with respect to the Lebesgue measure on $\mathbb{R}^N$.

The powers of $V(\lambda)$ can be generalized, leading to Gaussian β Ensembles (GβE), with eigenvalue density (with respect to Lebesgue measure) given by

$$C_N^\beta|V(\lambda)|^\beta \prod_{i=1}^N e^{-(\lambda_i^2/4)\beta}. \tag{4.2}$$

For certain special values of β the above density is a probability density of eigenvalues of random matrices:

$\beta = 1$	GOE	$\mathbb{R}$
$\beta = 2$	GUE	$\mathbb{C}$
$\beta = 4$	G Symplectic E (GSE)	$\mathbb{H}$

(We will not define the GSE here — see, for example, [Meh04].) No other values of β correspond to similar invariant random matrix ensembles. However, there are still ways to study general GβE. In fact, the spectral distributions of these ensembles still converge to the Wigner semicircle law (but the local limiting distributions are substantially different).

One way to treat the Gaussian β ensembles is to introduce a completely different random matrix model for the spectral distribution (4.2). This was first observed in [DE02] by applying the standard

tridiagonalization procedure from linear algebra to Gaussian random matrices, and reading off the corresponding distributions. This tridiagonal matrix ensemble for GβE looks as follows. Let χ_t be a χ random variable of parameter t, that is, with density

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

on the real line. Let $y_i \sim \chi_{i\beta}$, $\xi_i \sim N(0,1)$ be independent random variables. Then the tridiagonal ensemble

$$A_N^\beta = \begin{bmatrix} \sqrt{\frac{2}{\beta}}\xi_1 & y_{N-1}/\sqrt{\beta} & \cdots & 0 \\ y_{N-1}/\sqrt{\beta} & \sqrt{\frac{2}{\beta}}\xi_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & y_1/\sqrt{\beta} \\ 0 & \cdots & y_1/\sqrt{\beta} & \sqrt{\frac{2}{\beta}}\xi_N \end{bmatrix}$$

has the eigenvalue density GβE.

One can view the above matrix A_N^β as a linear operator, in this case it will be a *random difference operator*. As $N \to \infty$ it converges to a random differential operator. This analysis is performed in, e.g., [RRV11].

Remark 4.6. Eigenvalue densities are also accessible for Gaussian random matrices without Hermitian or real symmetry. That is, if the elements of a random matrix $a_{ij} \sim N(0,1)$ are i.i.d. for all i, j, then the spectrum of this matrix (which is complex) is the so-called Ginibre ensemble. The limiting spectral distribution in this case is uniform on the unit disc in the complex plane. This result is known under the name *Girko's Circular Law*.

4.3.2. *Circular ensembles.* Another example is the Circular Dyson's ensemble (CUE) of random matrices. Consider the Haar measure μ on the compact group of unitary matrices $U(N)$. This measure is shift invariant, i.e.

$$\forall \mathcal{A} \subseteq U(N), \ \forall V \in U(N), \qquad \mu(V\mathcal{A}) = \mu(\mathcal{A}V) = \mu(\mathcal{A}),$$

and is normalized so $\mu(U(N)) = 1$.

The eigenvalues of a random matrix $U_N \in U(N)$ lie on the unit circle, so they are of the form $e^{i\theta_k}, k = 1, \ldots, N$. The joint density of the eigenvalues is

$$C_N^{CUE} \prod_{k<j} |e^{i\theta_k} - e^{i\theta_j}|^2$$

with respect to the Lebesgue measure on the N-dimensional torus $\{|z| = 1\}^N$. This ensemble does not have any edge behavior, so we will not consider it. The local (bulk) distribution in the CUE case is similar to that of the Gaussian Unitary Ensemble with real eigenvalues.

It is also possible to define general β CUE ensembles by changing the power of the Vandermonde determinant.

4.3.3. *Invariant ensembles.* (LOG-GAS INTERPRETATION OF THE EIGENVALUE DENSITY, AND INVARIANT ENSEMBLES — TO BE FILLED IN HERE)

Lecture #7 on 2/10/2016 ───────────────────────────

4.4. **Determinantal structure of the GUE.**

4.5. **Hermite polynomials.**

4.6. **Non-rigorous asymptotics via spectral projections.**

4.7. **Rigorous asymptotics via contour integrals.**

4.8. **Edge asymptotics and the Tracy–Widom distribution.**

4.9. **Notes and references.** The material in Section 4 represents a classical chapter in random matrix theory. The Gaussian Unitary Ensemble is a random matrix ensemble for which all the computations can be carried out explicitly. Section 4.6 closely follows the book [Tao12].

The asymptotic results for the GUE presented here should be *universal* — that is, they should not depend on the concrete structure of the ensemble and hold in much greater generality. Getting such universal results for random matrices is an important ongoing program, see, e.g., [Joh01], [TV11], [TV12], [ESY11]. Addressing some universality results is the next big goal of the course.

References

[AGZ10]   G.W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge University Press, 2010.

[BBAP05] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Annals of Probability (2005), 1643–1697, arXiv:math/0403022 [math.PR].

[Cat]   OEIS (The On-Line Encyclopedia of Integer Sequences), http://oeis.org, sequence A000108.

[DE02]   I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, Journal of Mathematical Physics **43** (2002), no. 11, 5830–5847, arXiv:math-ph/0206043.

[DF15]   P. Diaconis and P.J. Forrester, *A. Hurwitz and the origins of random matrix theory in mathematics*, arXiv:1512.09229 [math-ph].

[ESY11]   L. Erdős, B. Schlein, and H.-T. Yau, *Universality of random matrices and local relaxation flow*, Invent. Math. **185** (2011), no. 1, 75–119, arXiv:0907.5605 [math-ph].

[GT10]   F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491, arXiv:0709.3995 [math.PR].

[Hur97]   A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachr. Ges. Wiss. Göttingen (1897), 71–90.

[Joh01]   K. Johansson, *Universality of the Local Spacing Distribution in Certain Ensembles of Hermitian Wigner Matrices*, Comm. Math. Phys. **215** (2001), no. 3, 683–705, arXiv:math-ph/0006020.

[Kea06]   J. Keating, *Random matrices and number theory*, Applications of random matrices in physics, Springer, 2006, pp. 1–32.

[KS00]   J. Keating and N. Snaith, *Random matrix theory and ζ (1/2+ it)*, Communications in Mathematical Physics **214** (2000), no. 1, 57–89.

[Meh04]   M.L. Mehta, *Random matrices*, Academic press, 2004.

[Mon73]   H. Montgomery, *The pair correlation of zeros of the zeta function*, Proc. Symp. Pure Math, vol. 24, 1973, pp. 181–193.

[MP67]   V. Marchenko and L. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb. **1** (1967), 457–483.

[NL12]   J. Novak and M. LaCroix, *Three lectures on free probability*, arXiv:1205.2097 [math.CO].

[Odl87]   A. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Mathematics of Computation **48** (1987), no. 177, 273–308.

[RRV11]   J. Ramirez, B. Rider, and B. Virag, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, J. Amer. Math. Soc. **24** (2011), no. 4, 919–944, arXiv:math/0607331 [math.PR].

[Sta01]   R. Stanley, *Enumerative Combinatorics. Vol. 2*, Cambridge University Press, Cambridge, 2001, With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[Sta15]   _____, *Catalan numbers*, Cambridge University Press, 2015.

[Tao12]   T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, Amer. Math. Soc., 2012.

[TV11]   T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta mathematica **206** (2011), no. 1, 127–204, arXiv:0906.0510 [math.PR].

[TV12]   _____, *Random matrices: The Universality phenomenon for Wigner ensembles*, arXiv:1202.0068 [math.PR].

[VDN92]   D. Voiculescu, K. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series, 1992.

[Wig55]   E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics (1955), 548–564.

[Wis28]   J. Wishart, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika **20A** (1928), 32–43.

[Zir10]   Martin R. Zirnbauer, *Symmetry classes*, arXiv:1001.0722 [math-ph].