Leonid Petrov

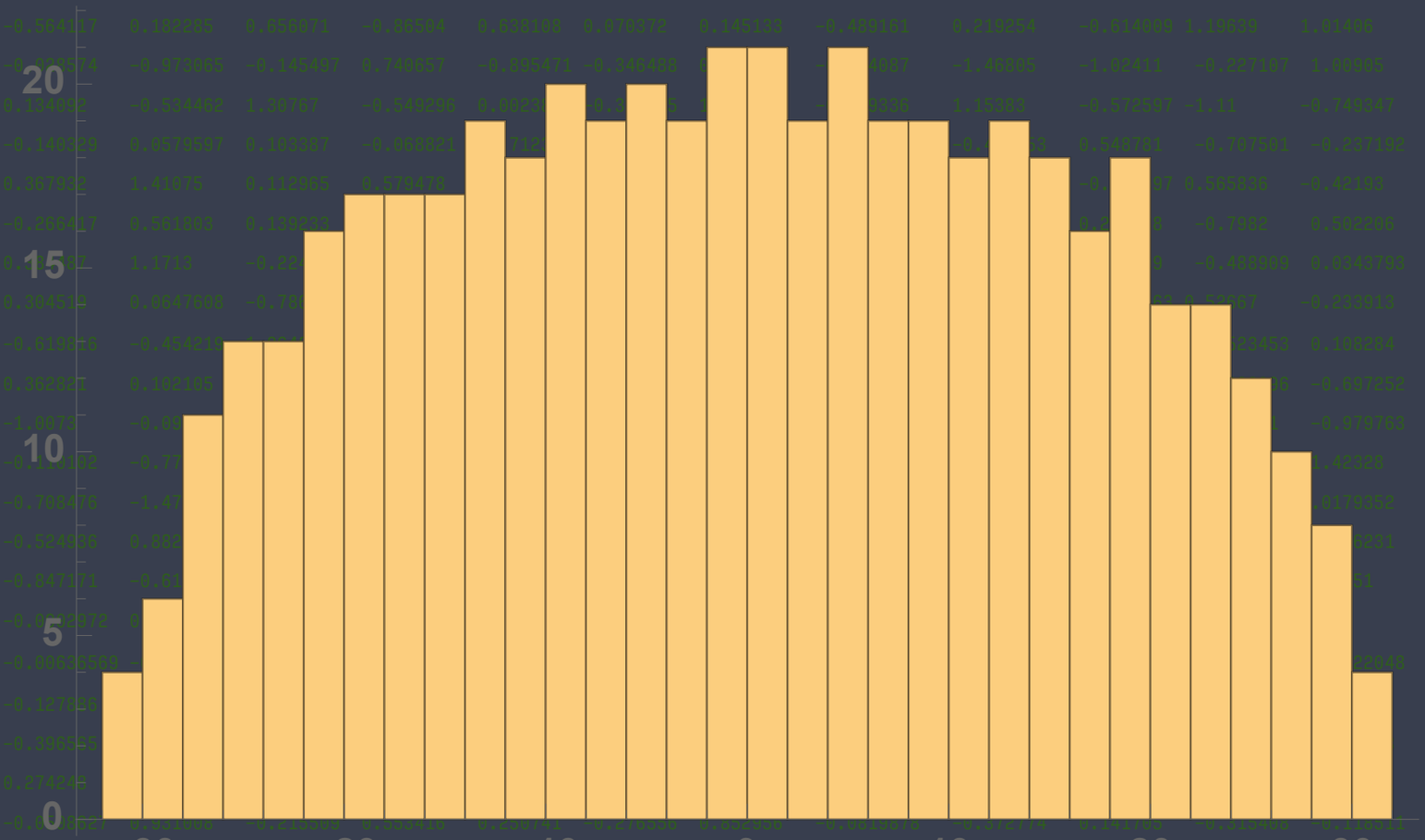
RANDOM MATRICES

Course Notes

University of Virginia

Fall 2019

https://lpetrov.cc/rmt19/



REMARKS

- These are notes for the graduate course MATH 8380 Random Matrices at University of Virginia.
- Exercises on the green background are expected to be turned in after 2.5 or 3 weeks of the corresponding lecture (see details in the syllabus).
- All images (unless otherwise specified with a link to the source) are generated by the author.

Books on random matrices

I. Mehta, M.L. *Random Matrices*

II. Anderson, G.W., Guionnet, A. and Zeitouni, O. *An Introduction to Random Matrices*

III. Pastur, L. and Shcherbina, M. *Eigenvalue Distribution of Large Random Matrices*

IV. Tao, T. *Topics in random matrix theory*

V. Forrester, P. *Log-gases and random matrices*

Chapter 1

SOME HISTORY

- Volumes of compact classical groups
- Statistics
- Nuclear physics
- Number theory
- Back to statistics

Random matrix theory is a rich and lively subject with connections to many areas of pure mathematics, mathematical physics, statistics (within mathematical sciences), and numerous applications. Examples of applications include biology, neuroscience, machine learning, and others. New applications are discovered weekly. This makes random matrices one of the frontiers of modern probability theory. Because of the diverse nature of the subject, purely probabilistic methods are not enough in the study of random matrices, and one has to learn other tools from algebra, combinatorics, representation theory, and classical analysis.

There is a rich toolbox to study random matrices, too. In the first lecture I will discuss origins of random matrix theory which motivate some of the tools.

Early on, random matrices and related objects were rediscovered / revitalized at least three times - first by Hurwitz in the study of compact classical groups around 1900, then by Wishart in statistics in the 1920s, and then in physics by Wigner and others since the 1950s. Shortly after the third, a boost in the development of random matrix theory came from connections to number theory (namely, to zeroes of the Riemann Zeta Functions).

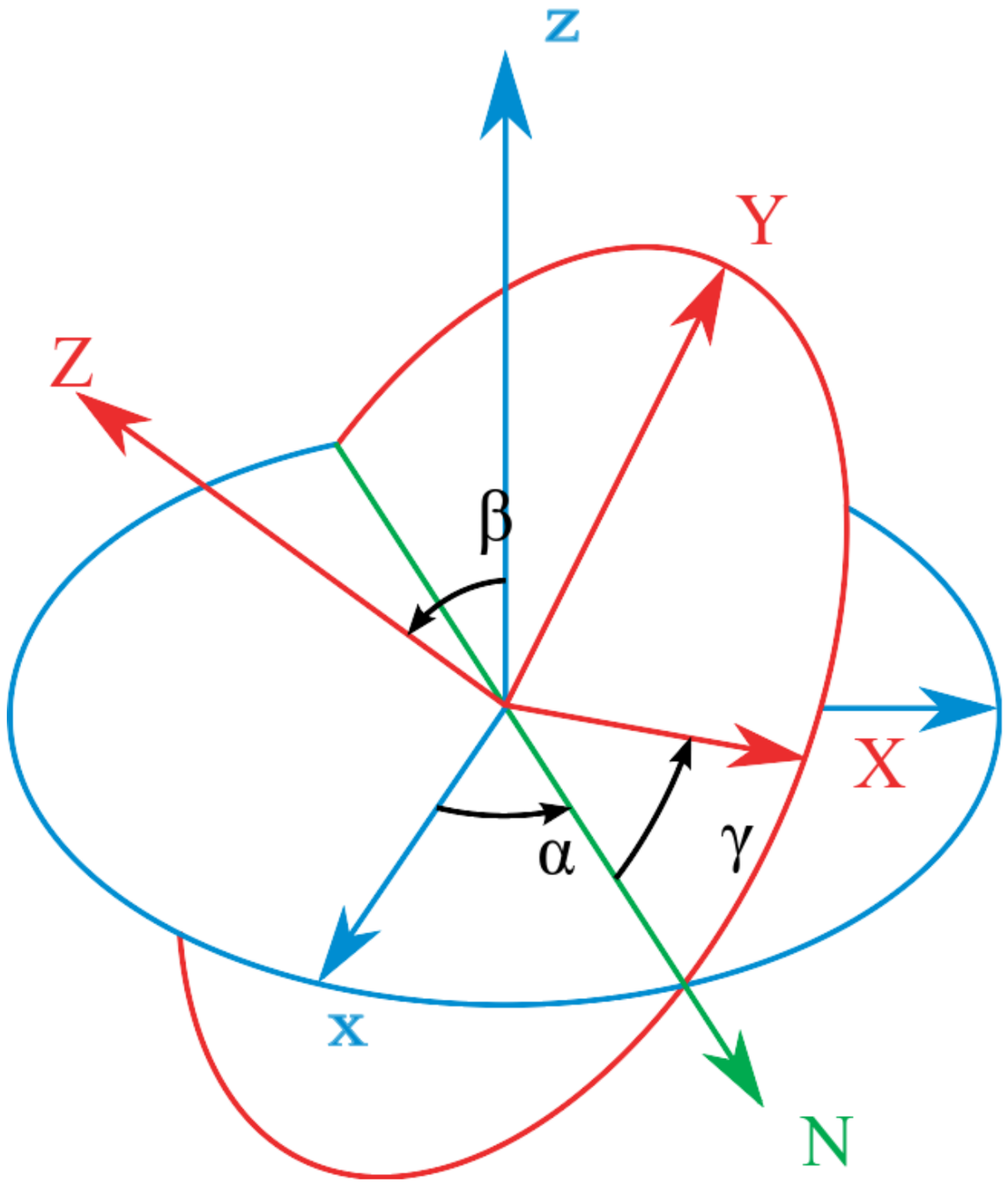
Since mid-20th century, random matrix theory entered into pure mathematics in full, and continues to be its vital part.

VOLUMES OF COMPACT CLASSICAL GROUPS

Compact classical groups are the Lie groups such as the orthogonal $O(N)$, unitary $U(N)$, or symplectic $Sp(N)$, or their special subgroups, etc. These are generally defined as groups of transformations of linear spaces (over $\mathbb{R}, \mathbb{C}, \mathbb{H}$) which preserve certain bilinear forms.

Each of the groups $O(N), U(N), Sp(N)$ is compact, and has a manifold structure (whose tangent space has a structure of the corresponding Lie algebra). As such, one can define natural invariant measures on these groups - called Haar measures (a Haar measure – name is from the 1930s – exists and is unique on any separable compact topological group). The "volume" of a group is then the total mass of this measure (which is finite because the group is compact).

The "initial" normalization of the Haar measure is taken in certain coordinates on the groups – Euler angles.



Euler angles

(by Lionel Brits https://commons.wikimedia.org/wiki/File:Eulerangles.svg)

INVARIANT VOLUME ELEMENT ON $SO(N)$

Recall that $SO(N)$ is the group of matrices with determinant one such that $VV^T = V^TV = Id$.

Let us discuss the invariant measure on $SO(N)$, that is, which satisfies $d\mu(V_0V) = d\mu(V)$ and $d\mu(VV_0) = d\mu(V)$ for each fixed $V_0 \in SO(N)$.

First, consider the toy example of the multiplicative group of positive real numbers. The invariant measure on this group must satisfy $d\mu(cx) = d\mu(x)$, and has the form $d\mu(x) = \frac{dx}{x}$.

So, for $SO(N)$ we need to mimic this, so let's consider something like $V^{-1}dV = V^TdV$. However, dV is not an independent differential, so we need to let $d\mu$ be the product of independent differentials of V^TdV.

Note that $V^TV = Id$ implies $d(V^T)V + V^T(dV) = 0$, so V^TdV is anti-symmetric. For an antisymmetric real matrix $A = [A_{ij}]$, its independent entries are all entries above the diagonal. So, the product of the differentials is $\text{const}\prod_{i<j} dA_{ij}$.

Therefore, we conclude that the invariant measure is $d\mu(V) = \text{const}\prod_{i<j} dA_{ij}$, where $A = V^TdV$ is antisymmetric.

The fact that $d\mu(V_0V) = d\mu(V)$ is clear from the definition. To check $d\mu(VV_0) = d\mu(V)$, use the next exercise.

Exercise 1.1. Show that for A antisymmetric and X real, we have $(X^T\,dA\,X) = (\det X^TX)^{(N-1)/2}dA$.
Hint: represent X as a product of elementary transformation matrices: (1) row transpositions, (2) row multiplications by real numbers, and (3) adding one row to another. Check what happens to the determinant in front, too.

Then we have $d\mu(VV_0) = (VV_0)^Td(VV_0) = V_0^T(V^TdV)V_0 = V^TdV$, because the determinant of $V_0^TV_0$ is one. This completes the proof.

EULER ANGLES ON $SO(N)$. VOLUME

Define $R_j(\theta) = \begin{bmatrix} I_{j-1} & & & \\ & \cos\theta & \sin\theta & \\ & -\sin\theta & \cos\theta & \\ & & & I_{N-j-1} \end{bmatrix}$

(where I_k means the $k \times k$ identity matrix), and

$E_j = R_j(\theta_{j,j+1})R_{j-1}(\theta_{j-1,j+1})\cdots R_1(\theta_{1,j+1})$.

Hurwitz in the 1890s proved that any generic matrix $V_N \in SO(N)$ (by "generic" we mean Haar-almost-every) can be represented uniquely as a product $V_N = E_1E_2\ldots E_{N-1}$. The amount of Euler angles is

$0 \le \theta_{1,j+1} < 2\pi,\ 0 \le \theta_{i,j+1} \le \pi\ (2 \le i \le j \le N-1)$.

A general $V_3 \in SO(3)$ has the decomposition

$$V_3 = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where $0 \le \theta \le \pi,\ 0 \le \phi, \psi < 2\pi$. Geometrically, this corresponds to rotations about the z-axis, the transformed x-axis, then the transformed z-axis, as first identified by Euler (1770).

For general N, Hurwitz computed that the invariant measure has the following density with respect to the Lebesgue measure in the Euler angles coorinates: $d\mu = 2^{N(N-1)/4} \prod_{1 \le j < k \le N} (\sin\theta_{j,k})^{j-1} d\theta_{j,k}$. Then Hurwitz computed the volume of the group, which is the integral of this density over the group. It is given by

$$\mathrm{vol}\big(SO(N)\big) = \frac{1}{2}\ 2^{N(N+3)/4} \prod_{k=1}^{N} \frac{\pi^{k/2}}{\Gamma(k/2)}.$$

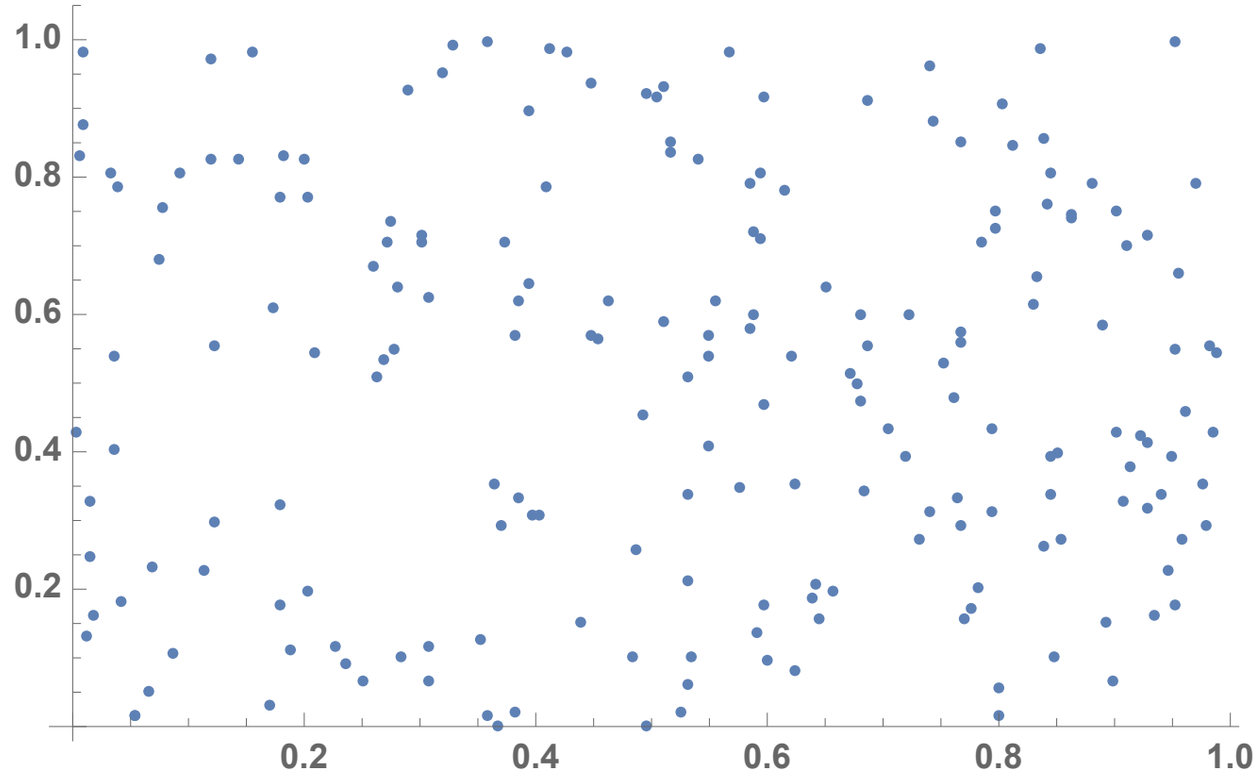
Exercise 1.2. Using this formula, write down the volume of $O(N)$.

The same computation can be performed for other compact classical groups. For the unitary group, we get

$$\mathrm{vol}(U(N)) = 2^{N(N+1)/2} \prod_{k=1}^{N} \frac{\pi^k}{\Gamma(k)}.$$

(Recall that unitary matrices are complex matrices with $UU^* = U^*U = Id$, where star means conjugate transpose.)

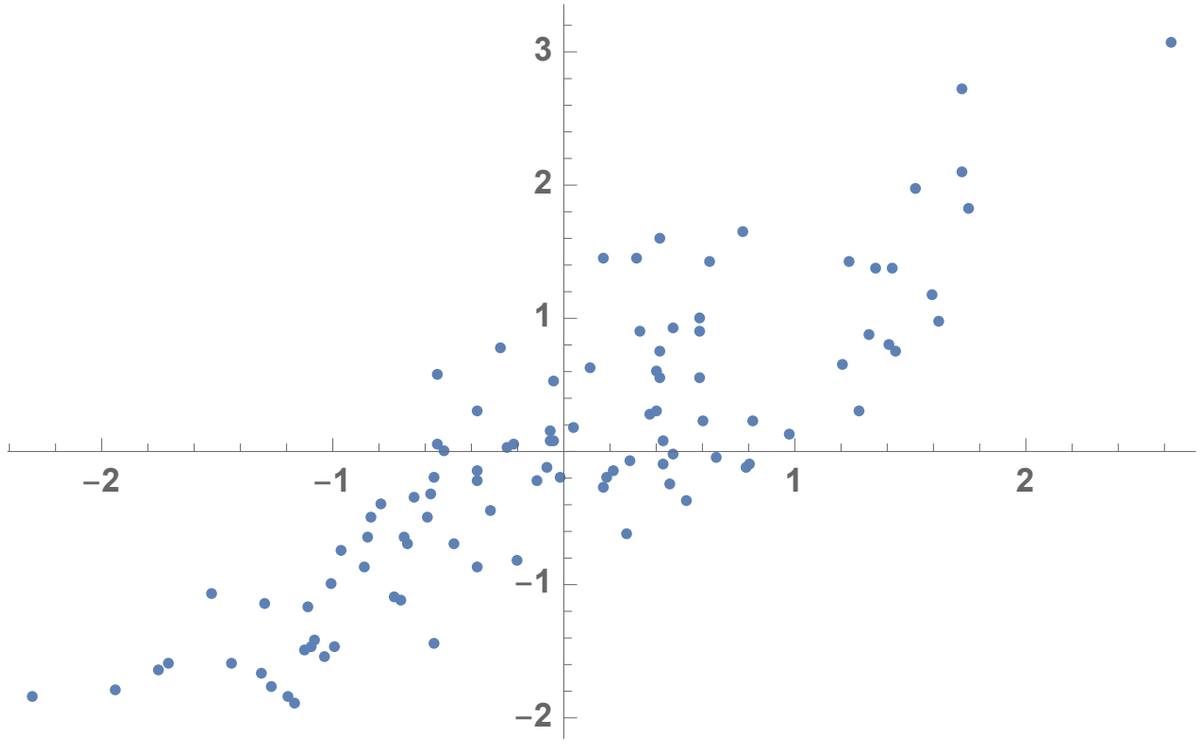
RANDOM COVARIANCE MATRICES



In statistics, Wishart distribution is a probability distribution on nonnegative-definite matrices which is a multidimensional analogue of the Gamma distribution. This arises as the distribution of the sample covariance matrix for a sample from a multivariate normal distribution. It occurs frequently in likelihood-ratio tests in multivariate statistical analysis. This distribution was introduced by Wishart in 1928.

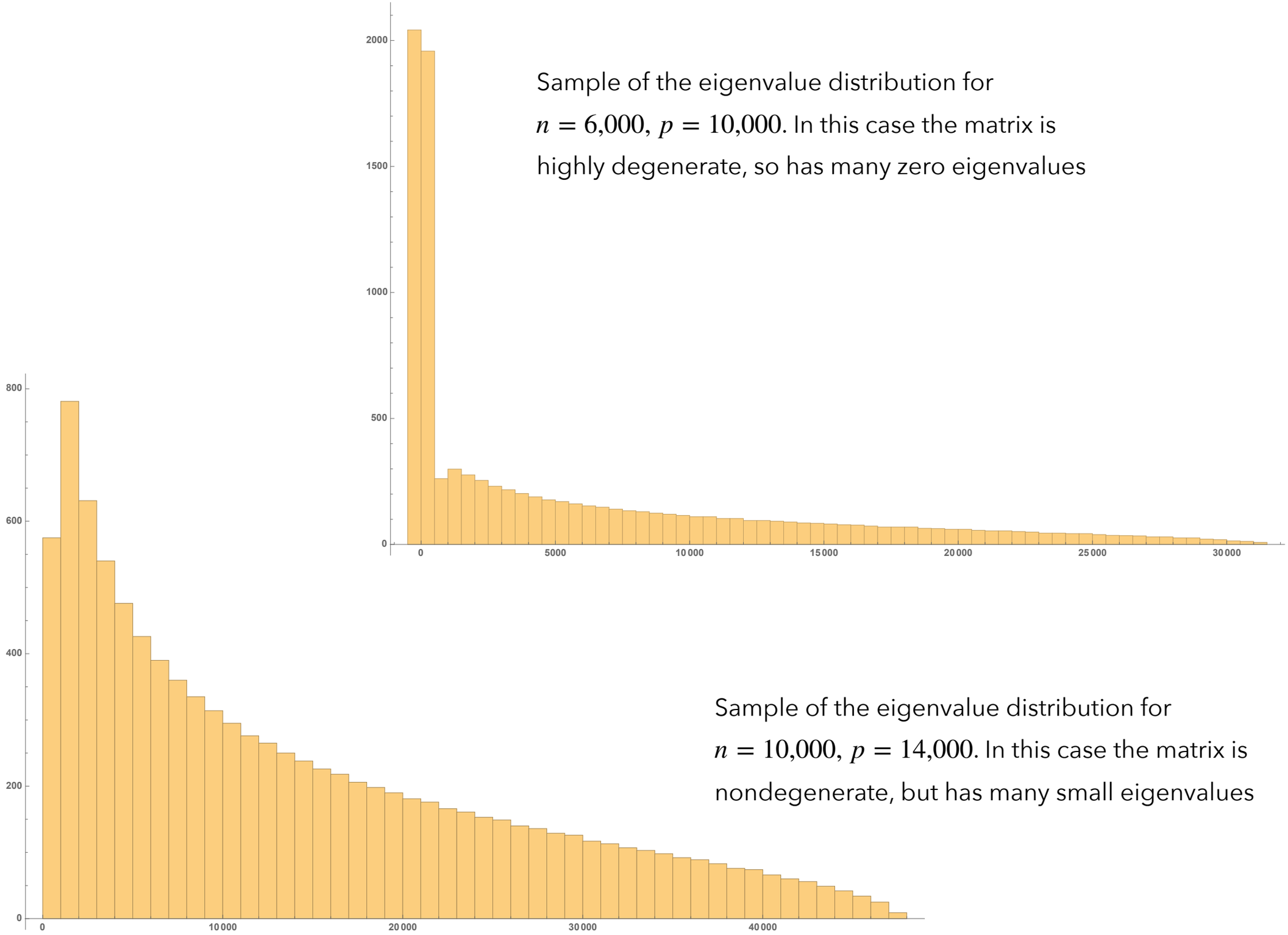
To define a Wishart random matrix, take X to be a $p \times n$ matrix of iid standard normal random variables. Define $W = XX^T$. This is the Wishart random matrix.

Note that if $n < p$, then the matrix W is almost surely degenerate, and we do not want that. So we assume that $n \geq p$.



Exercise 1.3. Do matrix elements W_{11} and W_{12} have the same distribution? How about W_{11} and W_{22}?

WISHART MATRICES. SIMULATIONS



Sample of the eigenvalue distribution for $n = 6{,}000$, $p = 10{,}000$. In this case the matrix is highly degenerate, so has many zero eigenvalues

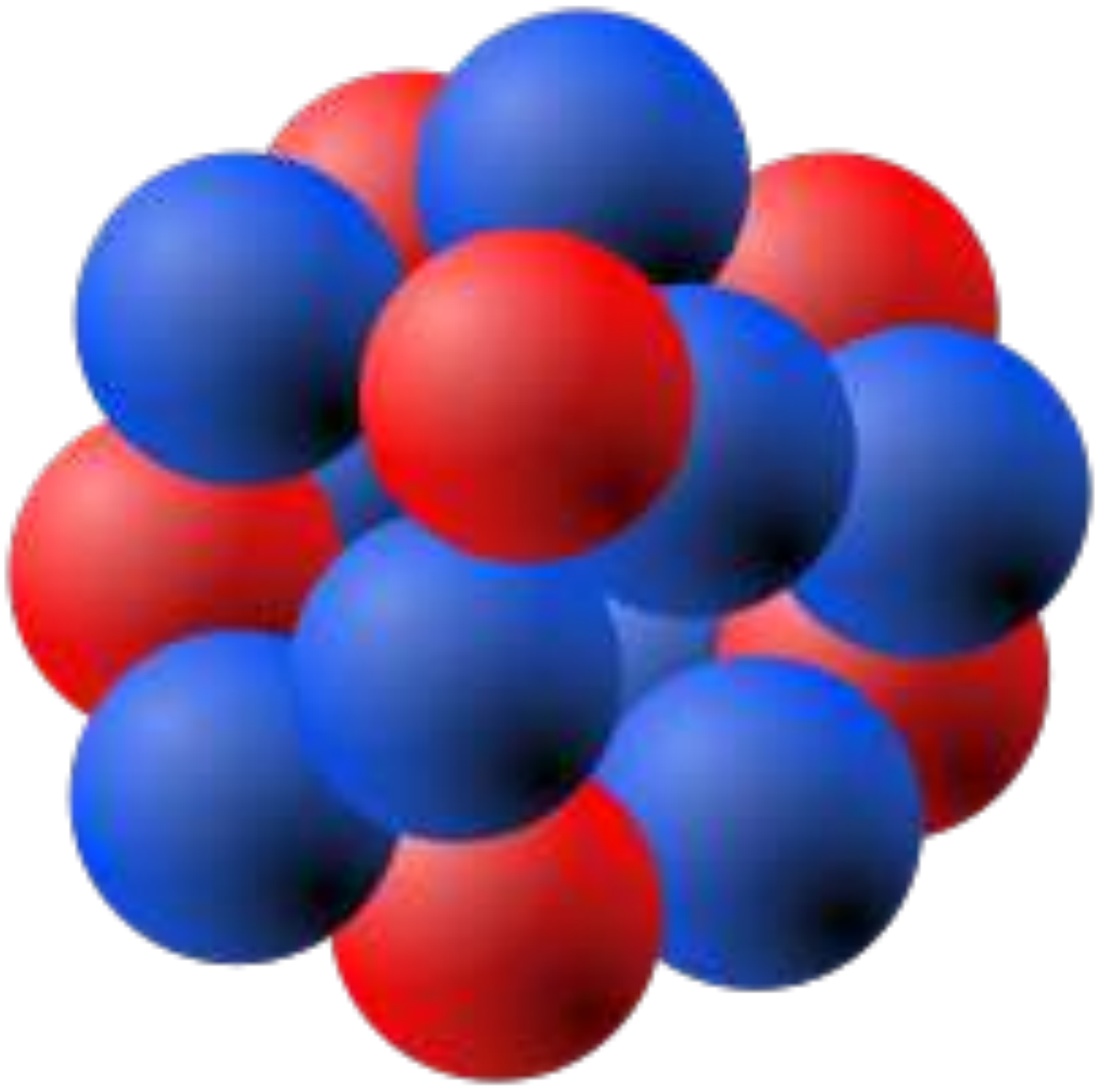
Sample of the eigenvalue distribution for $n = 10{,}000$, $p = 14{,}000$. In this case the matrix is nondegenerate, but has many small eigenvalues

NUCLEAR PHYSICS

Nuclear physics studies atomic nuclei viewed as quantum systems. As such, a quantum system is described by an operator in a Hilbert space (a "quantum Hamiltonian"). Of great interest are eigenvalues of this operator - they describe possible energy levels of the quantum system.

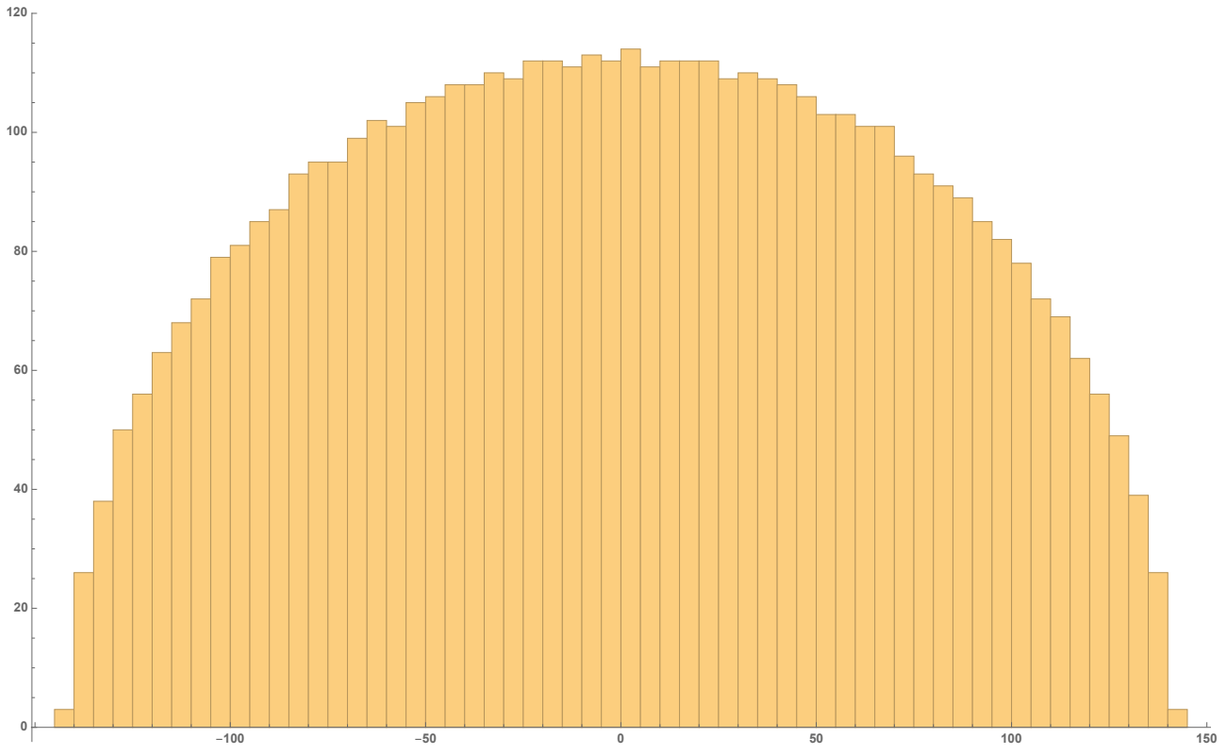
For example, the eigenvalue equation for the simplest, hydrogen, atom is the Schrodinger equation: $\left(-\frac{\hbar^2}{2\mu}\nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}\right)\psi(r,\theta,\phi) = E\psi(r,\theta,\phi)$. Here E are the possible energy levels.

Large nuclei are described by much more complicated Hamiltonians. In the 1950s, Wigner suggested that instead of studying complicated Hamiltonians, one could get some information about heavy nuclei from *random* Hamiltonians. In particular, he predicted *universality of energy spacings*. Random Hamiltonians are then approximated by random Hermitian matrices.



https://upload.wikimedia.org/wikipedia/commons/f/f0/Nucleus_drawing.svg

WIGNER'S SEMICIRCLE LAW



In one of the first papers of Wigner, he also obtained the *Semicircle Law* - joint eigenvalue distribution of large random Hermitian matrices with independent entries tends to the semicircle.

In particular, let X, Y be real matrices with iid mean 0 variance 1 entries (and some other moment conditions). Define $G = ((X + iY) + (X - iY)^T)/2$. This is a random Hermitian matrix.

Then the histogram of the eigenvalues of G looks like a (normalized) semicircle from $-2\sqrt{N}$ to $2\sqrt{N}$, where N is the size of the matrix.

This also suggests how the spacing of the eigenvalues in the "bulk" of the spectrum should look like. Namely, the eigenvalues have typical distance between them of order $1/\sqrt{N}$, and on that scale the distance depends on the location in the spectrum through the density.

Later, connections between eigenvalue spacing and number theory (spacing of zeroes of the Riemann Zeta function on the critical line) have been discovered - see note [4] for details.

BACK TO STATISTICS

Random matrix theory continues to be widely applicable in domains where it was originally discovered. Let me mention one statistical application based on a 2004 (fairly recent) work of Baik-Ben Arous-Peche.

Consider the goal of estimating a "spike" in the presence of a random matrix. Namely, let W be a Wishart (random covariance) matrix, and X be a rank-one matrix (a "spike") of "size" 1.

Looking at the eigenvalues of $W + aX$, for which values of a is it possible to recover X? In other words, can we statistically detect the null-case (random covariance matrix of a mean-0 iid vector) from the "spiked" case?

Turns out that we should look at the edge of the spectrum. The edge is on scale $\sqrt{n}$, and there exists a critical value c (independent of n but depending on $\gamma = p/n < 1$ which is assumed constant), such that

- If $a < c$, the edge of the spectrum of $W + aX$ is distributed the same as for W, so the "spike is lost in the bulk". This limit is not Gaussian, by the way, but is given by the so-called *Tracy-Widom* (TW) distribution.
- If $a > c$, the largest eigenvalue has the Gaussian distribution, so we can detect the spike
- At $a = c - \tilde{a}n^{-1/3}$, there is a phase transition, and the distribution of the largest eigenvalue of $W + aX$ tends to yet another law called the BBP distribution. Depending on the parameter $\tilde{a}$, BBP interpolates between Gaussian and TW.

NOTES AND REFERENCES

1. Diaconis, Forrester, *A. Hurwitz and the Origins of Random Matrix Theory in Mathematics* (2015) gives an account of the first matrix integrals (invariant group integrals) which appeared over 100 ago. The original paper of Hurwitz is *Uber die Erzeugung der Invarianten durch Integration* [Gott. Nachrichten (1897), 71-90].

2. The decomposition $V_N = E_1 E_2 \ldots E_{N-1}$ of $SO(N)$ matrices as products of rotation matrices is an analogue of the fact that every permutation can be written as a product of elementary transpositions (in a particular order). One can think that in this case all Euler angles are 0 or π.

3. J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika (1928) – one of the first appearances of random matrices in statistics.

4. Wigner's original work on random matrices in nuclear physics: Wigner, E. (1955). *Characteristic vectors of bordered matrices with infinite dimensions*. Annals of Mathematics. 62 (3): 548–564.

5. An account of the meeting of Dyson and Montgomery, when the applicability of random matrices to number theory was discovered: *https://www.ias.edu/ideas/2013/primes-random-matrices*

6. *Phase transition of the largest eigenvalue for non-null complex sample covariance matrices* - Jinho Baik, Gerard Ben Arous, Sandrine Peche. *https://arxiv.org/abs/math/0403022*

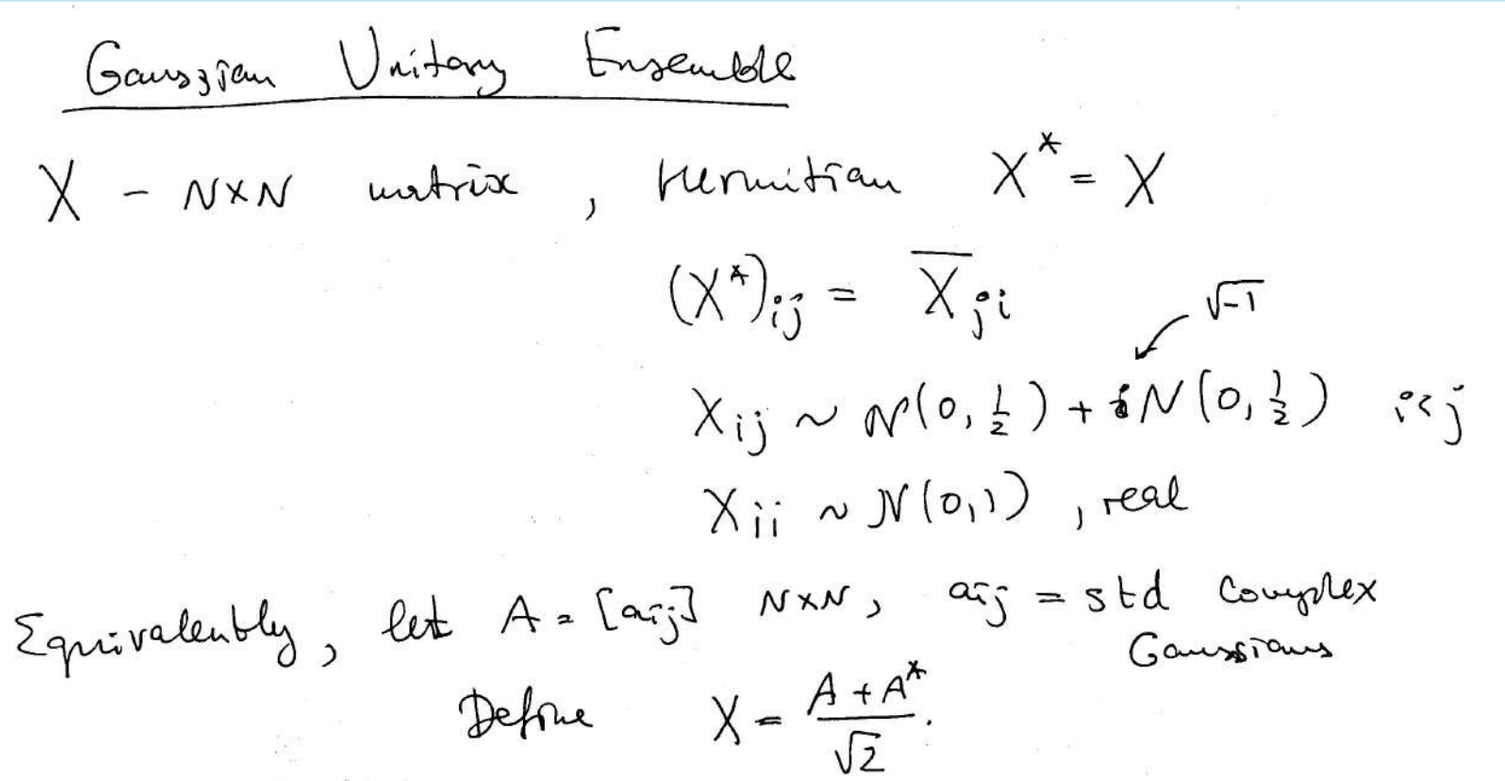
Chapter 2

GUE AND DBM

- Gaussian distribution and Brownian motion
- Gaussian Unitary Ensemble (GUE)
- GUE: some properties
- GUE: distribution of eigenvalues
- Dyson Brownian Motion (DBM)

In this chapter we discuss the most basic ensemble of random matrices – the *Gaussian Unitary Ensemble*. We will obtain its spectral distribution.

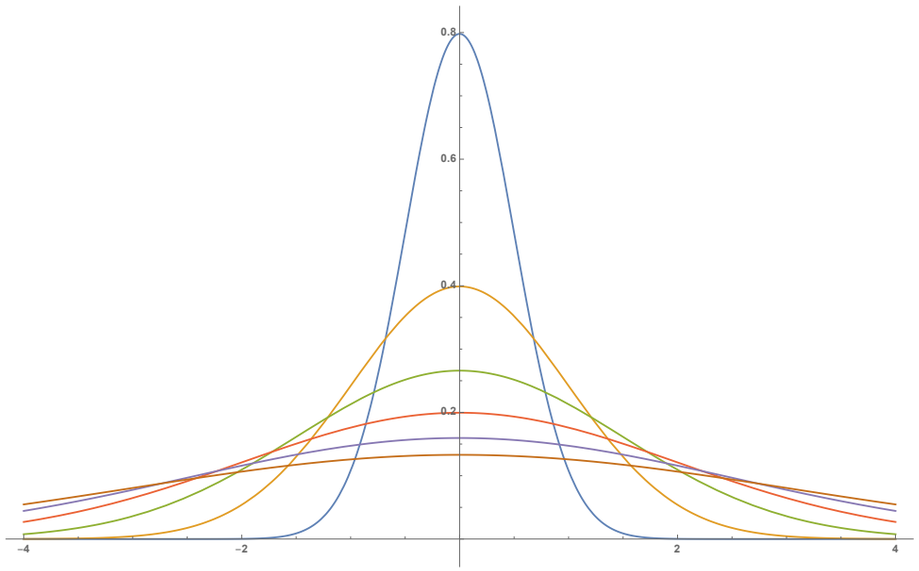
We will also connect the GUE to a dynamical model – the Dyson Brownian motion (DBM). This connection is a direct analogue of the link between the Gaussian (normal) distribution and the usual Brownian motion.



Exercise 2.1. Prove the equivalency between these two definitions of the GUE

GAUSSIAN (=NORMAL) DISTRIBUTION

Let us recall basic facts about the Gaussian distribution and the Brownian motion.



$N(0,\sigma^2)$: density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$, $x \in \mathbb{R}$

($\sigma > 0$ std. deviation)

(σ^2 = variance)

$N(\mu,\sigma^2)$: replace x^2 by $(x-\mu)^2$

μ = mean (= expectation)

This was the real Gaussian. We will also need the complex Gaussian.

The standard complex Gaussian random variable is given by $Z = X + iY$, where $i = \sqrt{-1}$ and X, Y are *independent* mean 0 normal random variables with variance $\frac{1}{2}$. Here we need variance $\frac{1}{2}$ and not 1 because the variance of a complex random variable is $Var(Z) = E(Z - EZ|^2)$, which is $E(|Z|^2) = E(X^2) + E(Y^2)$ in our case.

Exercise 2.2. For Z a standard complex Gaussian random variable, let $Z = Re^{i\Theta}$ be its representation in terms of absolute value and argument. What are the distributions of the random variables R and Θ?

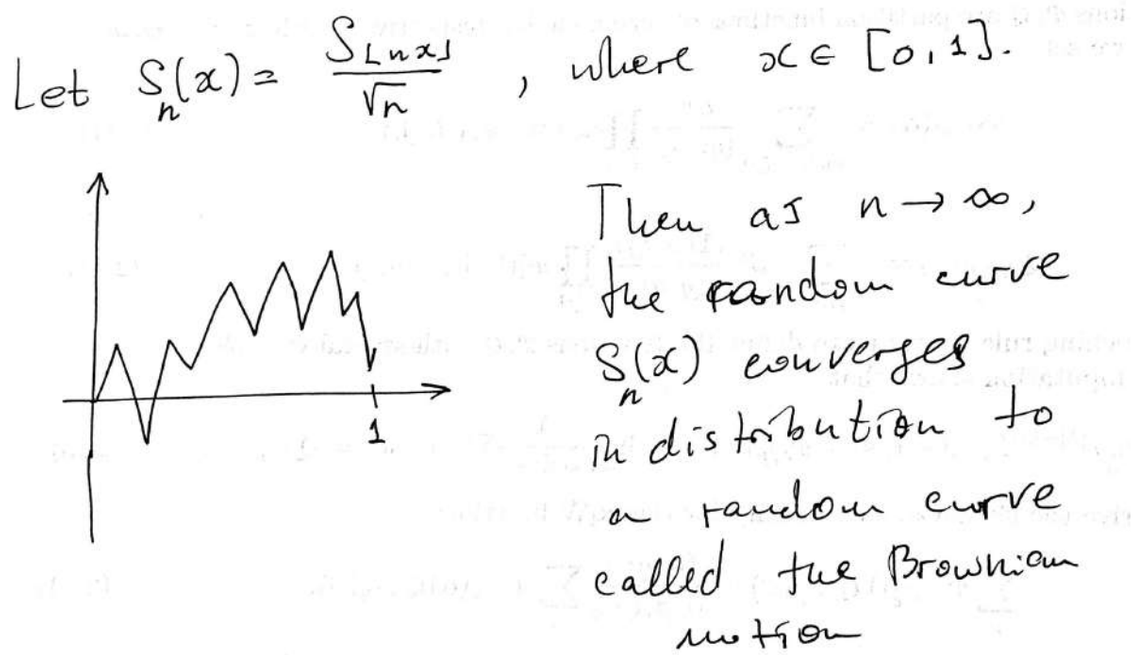
CENTRAL LIMIT THEOREM

Let $X_1, X_2, \ldots$ be independent identically distributed random variables with mean 0 and variance 1. Denote $S_n = X_1 + \ldots + X_n$. Then $\frac{S_n}{n} \to 0$ as $n \to \infty$ (this is the *Law of Large Numbers, LLN*). The convergence is almost sure, that is, $P(\omega\colon S_n(\omega)/n \nrightarrow 0) = 0$.

The next order approximation to the Law of Large Numbers is given by the *Central Limit Theorem (CLT)*: $\frac{S_n}{\sqrt{n}} \to \xi$, where ξ has the standard normal distribution. The convergence here is in distribution, that is, for each $r \in \mathbb{R}$ we have $P(S_n \le r\sqrt{n}) \to P(\xi \le r)$.

Note. In general, the convergence in distribution is valid only for r being points of continuity of the cumulative distribution function (cdf) fo the limiting random variable. However, as for the standard normal random variable the cdf is continuous everywhere, the convergence holds for all r.

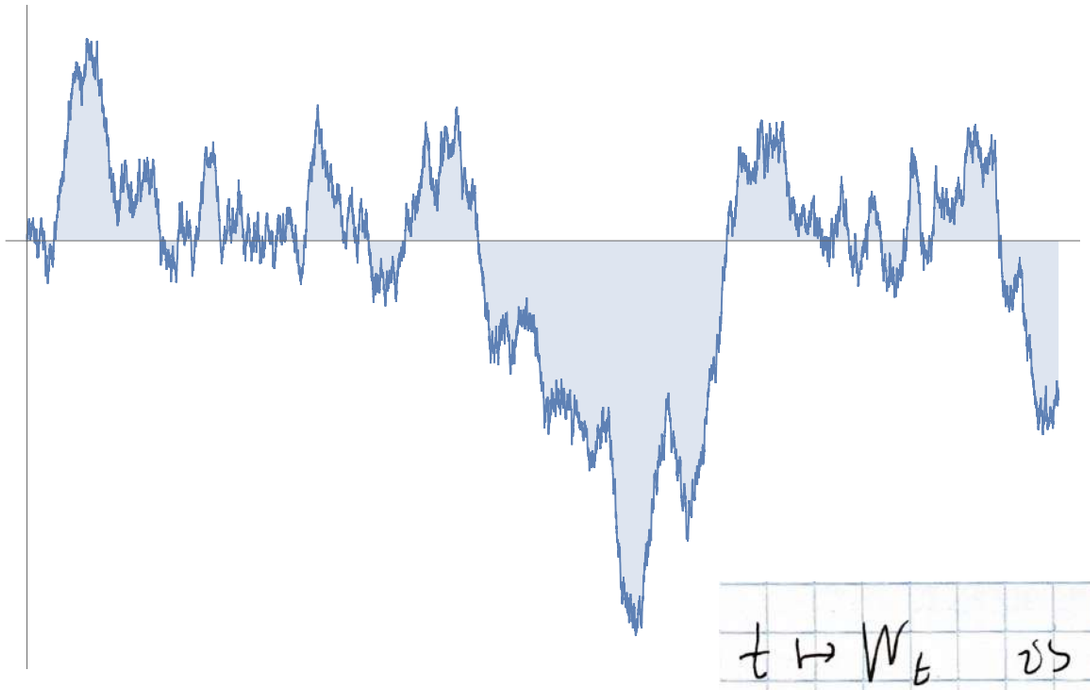
In fact, there is a much more detailed CLT which was established in the mid-20th century (Donsker invariance principle). In a relatively weak form it states:



The convergence here is in distribution, in Skorokhod space (space of cadlag functions). The metric in Skorokhod space is defined as $d(f,g) = \inf_{\lambda}(\|\lambda\| + \sup_{0\le t\le 1} |f(t) - g(\lambda(t))|)$, where the infimum is taken over all strictly increasing continuous mappings $[0,1] \to [0,1]$, and $\|\lambda\| = \sup_{s<t} \log \frac{\lambda(t)-\lambda(s)}{t-s}$.

BROWNIAN MOTION

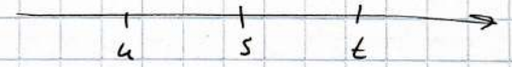
Let us discuss the limiting object – the Brownian motion (= Wiener process) – in detail.



W_t – Brownian motion

- $W_0 = 0$
- W_t is continuous
- increments are independent

$$W_t - W_s \perp\!\!\!\perp W_s - W_u$$



- increments are normal:

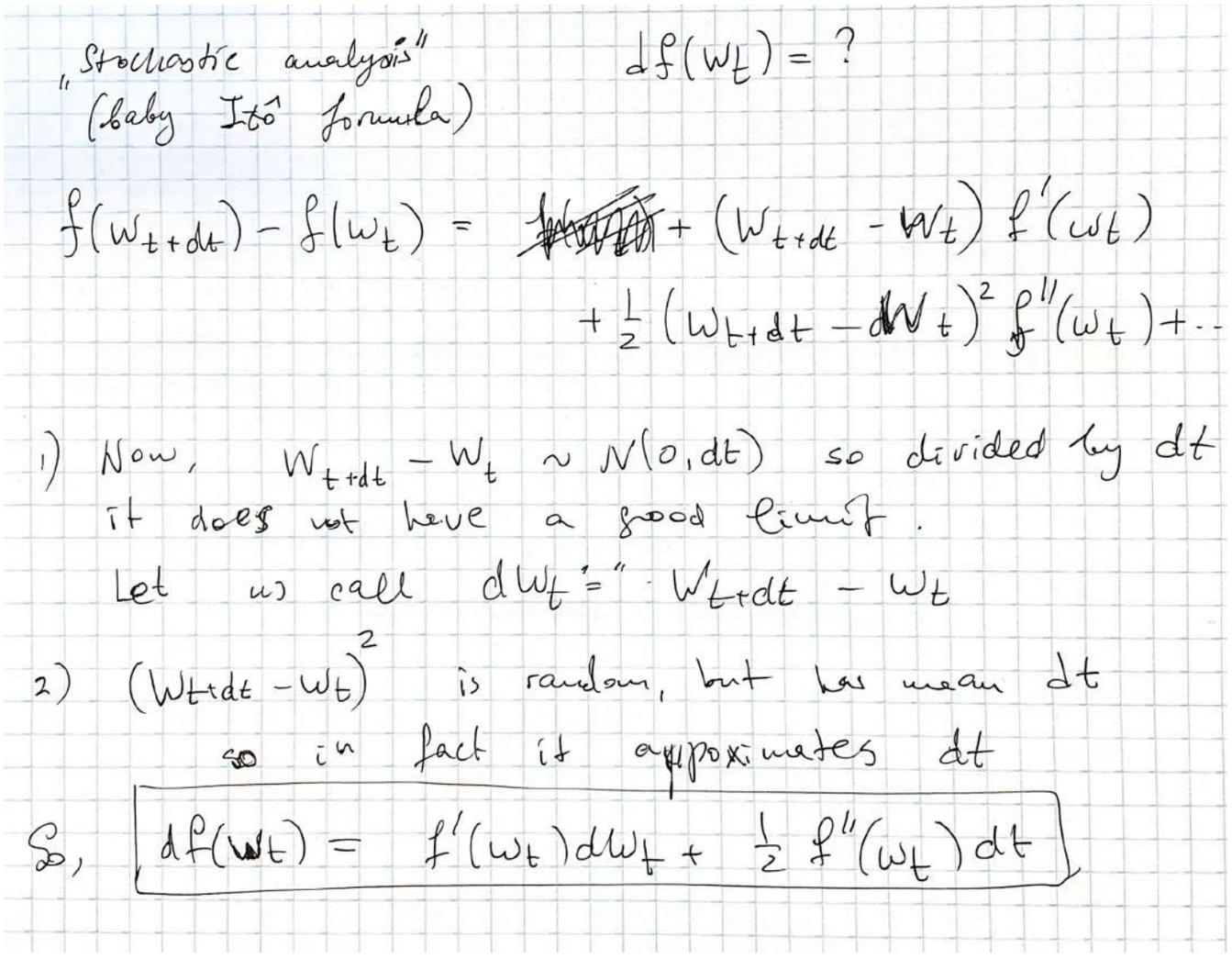
$$W_t - W_s \sim \mathcal{N}(0, t-s), \quad s \le t$$

Exercise 2.3. For W_t Brownian motion, show that $X_t = tW_{1/t}$ is also a Brownian motion. Hint: use the fact that the covariance function $\rho(t,s) = Cov(Y_t, Y_s)$ uniquely determines a mean zero Gaussian process. (A process is Gaussian if all of its multitime joint distributions are Gaussian.)

$t \mapsto W_t$ is not differentiable:
(in fact, a.s. nowhere diff.)

$\frac{W_t - W_s}{t-s}$ is normal with mean 0 and variance $\frac{1}{t-s}$ } informally:
so does not converge to any number.

W_t is Hölder $\frac{1}{2} - \varepsilon$ for all $\varepsilon > 0$.



If instead of W_t we take a diffusion process $dX_t = \mu_t dt + \sigma_t dW_t$, then we obtain the full Ito's formula:

$$df(X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{\sigma_t^2}{2} \frac{\partial^2 f}{\partial x^2} \right) dt + \sigma_t \frac{\partial f}{\partial x} dW_t.$$

Brownian motion is also a *Markov process*, in the sense that the future location of the moving particle is determined only by its present location, and not by the past history. Mathematically this is expressed as follows.

There is a transition function
$P_t(x, y) = P(W_{s+t} = y \mid W_s = x)$, for all $s \geq 0,\ x, y \in \mathbb{R}$.
The Brownian motion is time-homogeneous, i.e., its transition function does not depend on s. The Markov property is then
$P(W_{s+t} = y \mid W_s = x\, W_{s_1} = x_1, \dots, W_{s_k} = x_k) = P_t(x, y)$, for all x_i and all previous time moments $s_i < s$.

For the Brownian motion the transition function has a very simple density:

$$P_t(x, y)\, dy = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/(2t)}\, dy.$$

The transition function satisfies the heat equation

$$\frac{\partial}{\partial t} P_t(x, y) = \frac{1}{2} \frac{\partial^2}{\partial x^2} P_t(x, y)$$

and is also known as the *heat kernel*.

IDEA OF STOCHASTIC DIFFERENTIAL EQUATIONS

A diffusion is a process satisfying $dX_t = \mu_t dt + \sigma_t dW_t$ (where μ_t, σ_t are functions of time). This means in integral form that $X_t - X_0 = \int_0^t \mu_u du + \int_0^t \sigma_u dW_u$. The second integral is a stochastic integral (such as Ito integral).

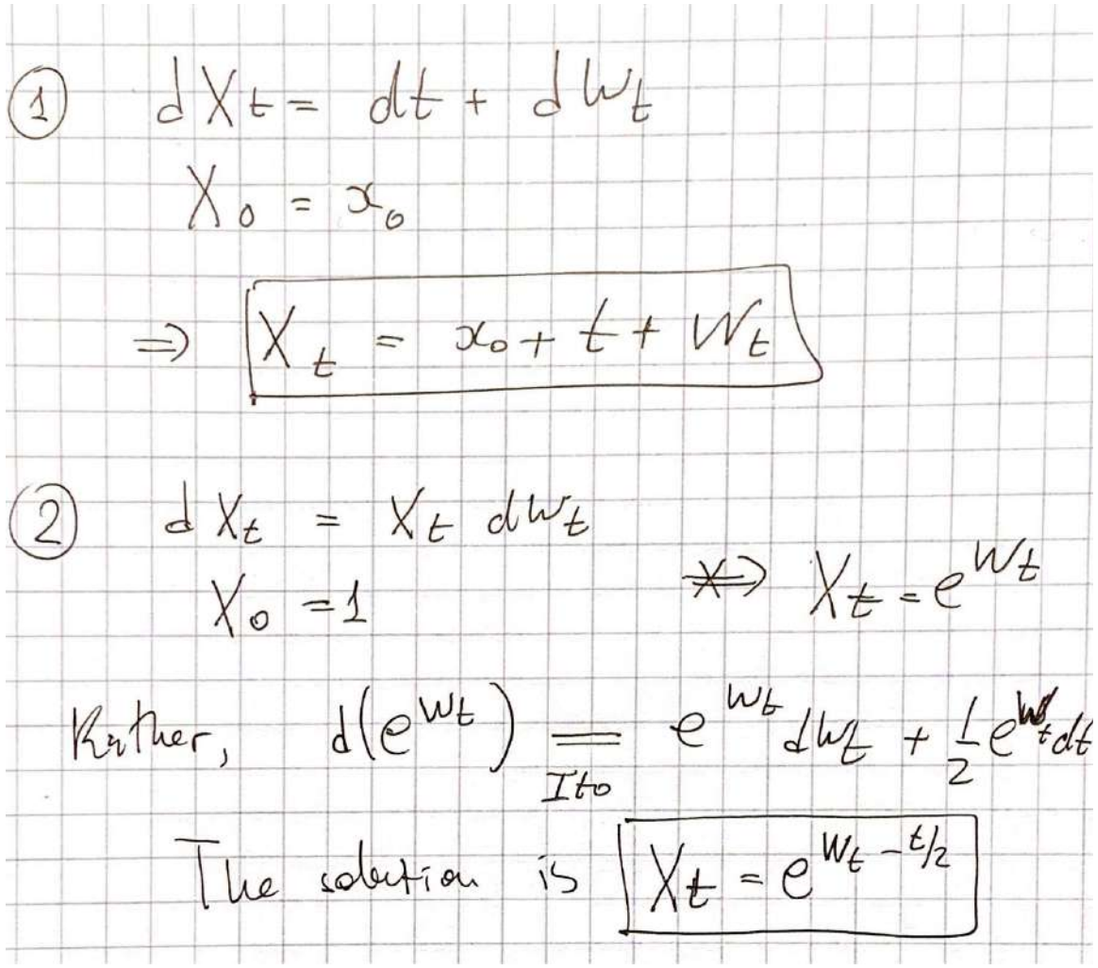
On the other hand, we can describe processes implicitly, requiring that the drift μ and diffusion coefficient σ depend not only on time, but on the value of X_t (e.g., location of the particle at a given time). This dependence still leads to a Markov process, and is called a *stochastic differential equation* (*SDE*): $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$. This equation is also understood in integral form.

Under mild conditions on μ, σ a solution to the SDE exists and it unique.

One should be careful with stochastic integrals. For example, for the Ito integral, $\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$.

We will use SDEs to write down PDEs for probability distributions of their solutions.

Examples.



Exercise 2.5. (1) In example 2, check that this X_t is indeed the solution of the SDE.

(2) Write the general solution of $dX_t = aX_t dt + bX_t dW_t$, $X_0 = x$, where a, b are constants.

GAUSSIAN UNITARY ENSEMBLE (GUE)

X – $N \times N$ matrix, Hermitian $X^* = X$

$(X^*)_{ij} = \overline{X_{ji}}$

$X_{ij} \sim \mathcal{N}(0, \frac{1}{2}) + \sqrt{-1}\,\mathcal{N}(0, \frac{1}{2})$ $i<j$

$X_{ii} \sim \mathcal{N}(0,1)$, real

Equivalently, let $A = [a_{ij}]$ $N \times N$, a_{ij} = std complex Gaussians

Define $X = \frac{A + A^*}{\sqrt{2}}$.

$X_{ii} \sim \mathcal{N}(0,1)$

$\mathrm{Re}\, X_{ij}$, $\mathrm{Im}\, X_{ij} \sim \mathcal{N}(0, \frac{1}{2})$

Density

$$\text{const} \cdot \prod_{i=1}^{N} e^{-x_{ii}^2/2} \prod_{i<j} e^{-(\mathrm{Re}\, x_{ij})^2 - (\mathrm{Im}\, x_{ij})^2}$$

$$\mathrm{Tr}(X^2) = \sum_{i,j=1}^{N} X_{ij} X_{ji} = \sum_{i,j=1}^{N} |X_{ij}|^2$$

Let us obtain some basic properties of the GUE. First, compute the joint density of the matrix elements with respect to the Lebesgue measure on the space of $N \times N$ Hermitian matrices. The Lebesgue measure itself is

$$dX = \prod_{i=1}^{N} dX_{ii} \prod_{1 \le i < j \le N} d(\Re X_{ij}) d(\Im X_{ij}),$$

Here there are $N + 2\dfrac{N(N-1)}{2} = N^2$ real parameters in the space. In other words, the space of $N \times N$ Hermitian matrices can be identified with the Euclidean space $\mathbb{R}^{N^2}$.

Therefore, we see that the GUE density of matrix entries has the form

$$d\mathrm{GUE}(X) = \text{const}\, e^{-\frac{1}{2}\mathrm{Tr}(X^2)} dX,$$

which is the (simplest) random matrix analogue of the Gaussian distribution.

Exercise 2.6. Show that the GUE distribution is invariant under conjugations by arbitrary unitary matrices (i.e., under the action $X \mapsto UXU^*$, where $U \in U(N)$ is fixed).

MATRIX BROWNIAN MOTION

When we have a normal distribution, we also have a Brownian motion. GUE is a Gaussian measure on the space of $N \times N$ Hermitian matrices. This space is identified with $\mathbb{R}^{N^2}$, and the coordinates X_{ii}, $\Re X_{ij}$, $\Im X_{ij}$ $(i < j)$ become independent random variables. Taking independent Brownian motions in these coordinates, we get a Brownian motion on the space of Hermitian matrices. We call is the *matrix Brownian motion* X_t.

More precisely, for X_{ii} the variance of the motion should be 1, and for $\Re X_{ij}$, $\Im X_{ij}$ $(i < j)$ the variance is $\frac{1}{2}$.

Exercise 2.7. Write down the transition function $P_t(X, Y) = Prob(X_{s+t} = Y \mid X_s = X)$ of the matrix Brownian motion.

Exercise 2.8. Show that the matrix Brownian motion is unitary invariant (you also need to define what this means exactly).

In one dimension, the Brownian motion does not have to start from 0. A Brownian motion starting from any $x \in \mathbb{R}$ simply has the form $x + W_t$, where W_t is the standard Brownian motion starting from 0.

In the matrix case, we can start the matrix Brownian motion from any fixed (or even random) matrix A, and it has the form $A + X_t$.

If A is diagonalized, $A = VD_aV^*$, where D_a is diagonal (with eigenvalues $a = (a_1, \ldots, a_N)$ of A on the diagonal), then we have $VD_aV^* + X_t$. But since X_t has the GUE distribution, it is unitary invariant, so $VD_aV^* + X_t$ has the same distribution as $V(D_a + X_t)V^*$.

In other words, all unitary invariant properties of $A + X_t$ (such as eigenvalues of $A + X_t$) depend on A only through its eigenvalues a.

This implies the following result:

Theorem. On eigenvalues, the matrix Brownian motion (started from any fixed matrix) reduces to a *Markov* process on the space $\{\lambda_1 \geq \ldots \geq \lambda_N : \lambda_i \in \mathbb{R}\}$ of ordered N -tuples of reals.

This Markov process is called the *Dyson Brownian Motion*. We will study some of its properties.

DYSON BROWNIAN MOTION

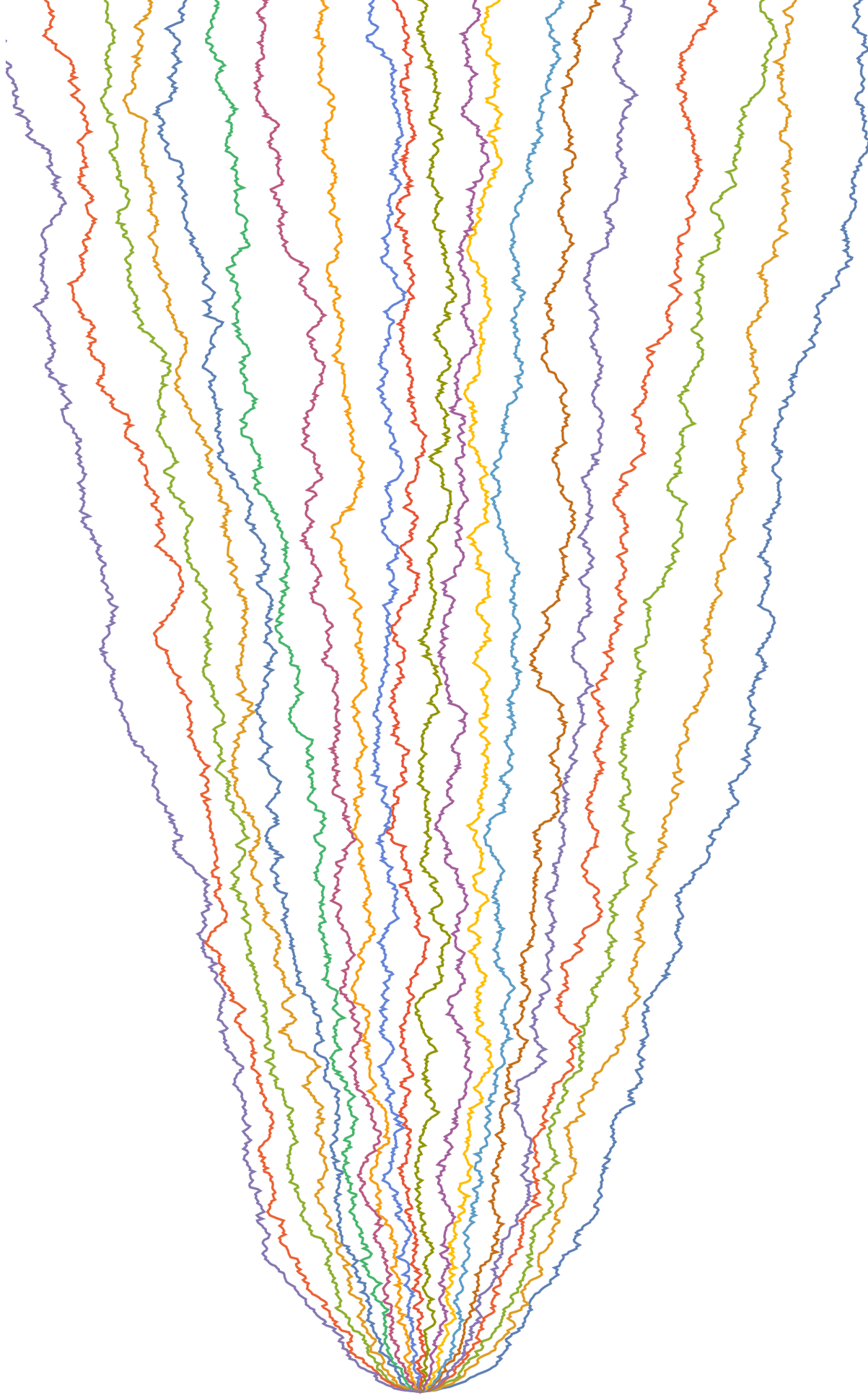
Let $\lambda_i = \lambda_i(t)$ be the eigenvalues of the matrix $A + X_t$, where X_t is the matrix Brownian motion started at zero (also, X_t has GUE distribution with variance t).

Theorem. These eigenvalues satisfy an SDE

$$d\lambda_i = dB_i + \sum_{j \neq i} \frac{dt}{\lambda_i - \lambda_j}, \quad i = 1, \dots, N,$$

where B_i are independent standard Brownian motions.

This result is essentially due to Dyson (1962). We will prove it by following *T. Tao's notes*, which in a sense repeat Dyson's argument in a more modern language. The argument is based on understanding how eigenvalues and eigenvectors change (in first order) if we change the matrix.



HADAMARD VARIATIONS

A – Hermitian with simple spectrum

- $\lambda_i(A)$ = zeroes of $\det(\lambda - A) = 0$

  $\Rightarrow A \mapsto \lambda_i(A)$ locally smooth maps (in v.f.th.)

- eigenvectors are determined by

$$\begin{cases} (\lambda_i(A) - A)\, u_i(A) = 0 \\ u_i(A)^* \, u_i(A) = 1 \end{cases}$$

$\Rightarrow A \mapsto u_i(A)$ are also locally smooth

If $A = A(t)$ is smooth, then

① $\dot\lambda_i = u_i^* \dot A\, u_i$

② $\ddot\lambda_i = u_i^* \ddot A\, u_i + 2 \sum\limits_{j \neq i} \dfrac{|u_j^* \dot A\, u_i|^2}{\lambda_i - \lambda_j}$

(here we assume that A is Hermitian and has simple spectrum)

$$\begin{cases} A u_i = \lambda_i u_i \\ u_i^* u_i = 1 \end{cases} \Rightarrow \dot u_i^* u_i + u_i^* \dot u_i = 0$$

$$\dot A u_i + A \dot u_i = \dot\lambda_i u_i + \lambda_i \dot u_i$$

$$u_i^* \dot A u_i + u_i^* A \dot u_i = \dot\lambda_i + \lambda_i u_i^* \dot u_i \qquad \square$$

equal

Exercise 2.9. Why here we have $u_i^* A \dot u_i = \lambda_i u_i^* \dot u_i$?

$$\ddot\lambda_i = u_i^* \ddot A\, u_i + 2\mathrm{Re}\,(u_i^* \dot A \dot u_i)$$

Compute $u_j^* \dot A u_i = -\lambda_j u_j^* \dot u_i + 0 + \lambda_i u_j^* \dot u_i$

$$= (\lambda_i - \lambda_j)\, u_j^* \dot u_i \quad (*)$$

Need $u_i^* \dot A \dot u_i$ or $\dot u_i^* \dot A u_i$

$\leftrightarrow$ product of $\dot A u_i$ & $\dot u_i$ written in basis of u_j

$$\Rightarrow \sum_j (\lambda_i - \lambda_j) |u_j^* \dot u_i|^2 \underset{\text{by } (*)}{=} \sum_{j : j \neq i} \frac{|u_j^* \dot A u_i|^2}{\lambda_i - \lambda_j}$$

SDE FOR DBM

Now, back to the Dyson Brownian motion. Modulo technicalities (related to non-simple eigenvalues) which can be resolved, the derivation of the SDE looks as follows.

Let the matrix Brownian motion at time t be in state A. Then, taking G to be the standard GUE random matrix, we see that the matrix Brownian motion at time $t + dt$ is in state $A + (dt)^{1/2}G$.

We have by the Ito's formula that $d\lambda_i$ equals the first derivative times dW_t, plus $\frac{1}{2}$ of the second derivative times dt. The first derivative is $u_i^* G u_i$, and the second is $\sum_{j \neq i} \frac{|u_j^* G u_i|^2}{\lambda_i - \lambda_j}$ (the "2" from the Hadamard variational formula disappears because of $\frac{1}{2}$ in Ito's formula).

Since G is unitary invariant, we may assume that u_i form an orthonormal basis of the N-dimensional space, and so $u_i^* G u_i$ simply corresponds to independent real Brownian motions dB_i, while each term $|u_j^* G u_i|^2$ has mean 1 and small variance (and the variance is small even after summing over j).

This completes the sketch of the proof of the SDE for the Dyson Brownian Motion.

Exercise 2.10. Instead of the matrix Brownian motion we can take the matrix Ornstein-Uhlenbeck process $e^{-t} W_{e^{2t}}$ which satisfies the SDE $dX_t = -X_t\, dt + dW_t$. Then the motion of eigenvalues is also Markovian. Repeat the previous argument and write down an SDE for the eigenvalues under the matrix Ornstein-Uhlenbeck process.

NOTES AND REFERENCES

1. Freeman J. Dyson, *A Brownian−Motion Model for the Eigenvalues of a Random Matrix,* Journal of Mathematical Physics **3**, 1191 (1962). This paper contains a derivation of the equations of motion of eigenvalues under the matrix Brownian motion.

2. One of the (many) interviews of Dyson *https://www.youtube.com/playlist?list=PLVV0r6CmEsFzDA6mtmKQEgWfcIu49J4nN*

3. The GUE is extended in at least two directions. One can either keep the unitary invariance (invariant ensembles with density $\propto e^{\mathrm{Tr}(V(X))}$), or keep the independence of the entries (Wigner matrices). The unitary symmetry / complex field can also be replaced by orthogonal symmetry / reals (the Gaussian ensemble is called GOE), or by symplectic symmetry / quaternions (leading to GSE).

4. At the level of eigenvalue distribution, the symmetry classes can be unified by including an interpolation parameter β, which is equal to 1,2, or 4 for real, complex, or quaternionic ensembles.

Interlude

MARKOV CHAINS AND SDE

On a finite state space X, a time-homogeneous Markov chain with discrete time corresponds to a stochastic transition matrix $P(x, y)$, whose elements are nonnegative and sum to one along each row: $\sum_{y \in X} P(x, y) = 1$ for all x. In diffusion setting, an SDE is an analogue of this transition matrix $P(x, y)$.

When $x \in X$ is random with some distribution $\mu(x)$, and we apply to it one step of a Markov chain, then the distribution becomes $(\mu P)(y) = \sum_{x \in X} \mu(x) P(x, y)$. In other words, in finite setting probability distributions are modeled by row vectors. Multiplication by transition matrices corresponds to time evolution of measures under Markov chains.

On the other hand, if f is a function on X viewed as a column vector, and we consider the function $(Pf)(x) = \sum_{y \in X} P(x, y) f(y)$, then $(Pf)(x)$ is the expectation of $f(\xi)$, where ξ is the location of the Markov chain started at x after one step.

Thus, we see that the Markov chain transition matrix P describes both the evolution of measures and the evolution of functions. An SDE does essentially the same.

EQUATIONS FOR THE BROWNIAN MOTION

Our goal now is to mimic the finite Markov chain setting for the Bronian motion.

Exercise 2.11. Let W_t be the standard Brownian motion. Show that $\frac{d}{dt}E\left(F(W_t)\right) = \frac{1}{2}E\left(F_{xx}(W_t)\right)$ for any smooth function, where F_{xx} is the second derivative. (You can take expectation in Ito's formula.)

This observation allows to obtain a PDE for the density for W_t. Namely, if the density of W_t is $\rho(t,x)$, then $EF(W_t) = \int_{-\infty}^{\infty} F(x)\rho(t,x)dx$. From the previous exercise, integrating by parts, one checks that $\frac{\partial}{\partial t}\rho(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\rho(t,x)$. The fact that the transition density of the Brownian motion satisfies this PDE was mentioned earlier. In fact, combined with the initial condition $\rho(0,x) = \delta_x$, the solution $\rho(t,x) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/(2t)}$ is unique. This is the density of the normal distribution with mean 0 and variance t.

Exercise 2.12. Let X_1 be a real random variable with mean 0 and variance 1. Define for $t \geq 0$, $X_t := e^{-t}(X_1 + W_{e^{2t}-1})$, where W_t is the usual Brownian motion.

- Show that X_t has mean 0 and variance 1
- Show that X_t converges in distribution to the standard normal random variable as $t \to +\infty$
- If F is smooth with all bounded derivatives, then show that $\frac{d}{dt}EF(X_t) = E\left(LF(X_t)\right)$, where L is the differential operator $LF := F_{xx} - xF_x$.
- Show that the density function $\rho(t,x)$ of X_t satisfies PDE $\partial_t\rho = L^*\rho$, where the adjoint operator is $L^*\rho = \rho_{xx} + \partial_x(x\rho)$.

Chapter 3

GUE EIGENVALUE DISTRIBUTION

- Eigenvalue density of GUE
- Possible extensions
- Derivation via Dyson Brownian motion
- Derivation by computing Jacobian of $A \mapsto \lambda(A)$

In this chapter we present two derivations of the joint eigenvalue density of the GUE:

- A stochastic processes computation involving the Dyson Brownian motion
- A more classical computation of the joint density of eigenvalues of the GUE random matrix,

We will also discuss extensions of the eigenvalue density formula which relates to other random matrix ensembles.

$$c_N \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_i e^{-\lambda_i^2/2}$$

GUE EIGENVALUE DENSITY AND GENERALIZATIONS

Let $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i \in \mathbb{R}$, be the eigenvalues of the GUE. We want to show that their joint density is given by

$$\rho(\lambda_1, \ldots, \lambda_N) = \frac{1}{Z_2} \prod_{i=1}^{N} e^{-\lambda_i^2/2} \prod_{j<k} (\lambda_j - \lambda_k)^2$$

Here $Z_2 = (2\pi)^{N/2} 0!1!2!\ldots(N-1)!$ is the normalizing constant.

Interestingly, the eigenvalue density is generalized in at least two different directions

Beta ensembles. The GUE comes from unitarily invariant Gaussian complex matrices. If we consider real symmetric Gaussian matrices (GOE), then the eigenvalue density has a similar form, except for the factor $\prod_{j<k} |\lambda_j - \lambda_k|$. For GSE, the factor is $\prod_{j<k} (\lambda_j - \lambda_k)^4$. These ensembles are unified in the Gaussian beta ensemble, GβE, which for $\beta = 1,2,4$ becomes the GOE/GUE/GSE.

The GβE does not have a similar invariant matrix model, but has a tridiagonal model [Forrester, (1.159)]:

$$\begin{bmatrix}
\mathrm{N}[0,1] & \tilde{\chi}_{(N-1)\beta} & & & \\
\tilde{\chi}_{(N-1)\beta} & \mathrm{N}[0,1] & \tilde{\chi}_{(N-2)\beta} & & \\
 & \tilde{\chi}_{(N-2)\beta} & \mathrm{N}[0,1] & \tilde{\chi}_{(N-3)\beta} & \\
 & \ddots & \ddots & \ddots & \\
 & & \tilde{\chi}_{2\beta} & \mathrm{N}[0,1] & \tilde{\chi}_{\beta} \\
 & & & \tilde{\chi}_{\beta} & \mathrm{N}[0,1]
\end{bmatrix}$$

Here $\tilde{\chi}_k$ is the square root of the chi square random variable, with density $\frac{2}{\Gamma(k/2)} u^{k-1} e^{-u^2}$, $u > 0$.

Invariant ensembles. The GUE is unitary-invariant and Gaussian, and the matrix has density $\propto e^{-\frac{1}{2}\mathrm{Tr}(X^2)}$. If instead of $-X^2/2$ we take an even polynomial which decays at infinity, then we obtain an invariant ensemble with matrix density $\propto e^{-\frac{1}{2}\mathrm{Tr}(V(X))}$. The eigenvalue density will be replaced by $\propto \prod_i e^{-V(\lambda_i)} \prod_{j<k} (\lambda_j - \lambda_k)^2$.

FROM DBM TO EIGENVALUE DENSITY

First, we know that $d\lambda_i = dB_i + \sum_{j \neq i} \frac{dt}{\lambda_i - \lambda_j}$.

Then, taking a smooth function $F(\lambda_1, \ldots, \lambda_N)$ and applying the multivariable Ito formula

$$dF(X_t)_k = \sum_i \partial_i F dX_{t,i} + \frac{1}{2} \sum_{i,j} \partial_i \partial_j F \, dX_{i,t} dX_{j,t'}$$

we get a formula for $dF(\lambda_1, \ldots, \lambda_N)$. Because the vector $d\lambda$ involves only independent Brownian motions, we see that in the second term we get $d\lambda_{i,t} d\lambda_{j,t} = \delta_{ij} dt$. Recall that the expectation of the Brownian motion (or of dB_i) is zero, so we have, averaging the SDE,

$$\partial_t EF(\lambda) = E\left[\frac{1}{2} \Delta F + \sum_{i,j:\ i \neq j} \frac{\partial_i F}{\lambda_i - \lambda_j} \right].$$

Therefore, the density of the eigenvalues $\rho = \rho(t, x)$ of the GUE at time t (where $x = (x_1 \geq \ldots \geq x_N)$) should satisfy the PDE with the adjoint of the above operator that acted on functions.

Exercise 3.1. Obtain the multivariable Ito formula for $dF(W_t)$, where W_t is the multivariable Brownian motion, by using Taylor expansions.

Namely, we obtain $\partial_t \rho = D\rho$, where

$$D = \frac{1}{2} \Delta \rho - \sum_{i \neq j} \partial_i \left(\frac{\rho}{\lambda_i - \lambda_j} \right).$$

Let us now check that the density $\propto \prod e^{-\lambda_i^2/(2t)} V(\lambda)^2$, where $V(\lambda) = \prod_{i<j} (\lambda_i - \lambda_j)$ is the Vandermonde determinant, satisfies this PDE.

Exercise 3.2. Show that $\partial_i V = V \sum_{j:\ j \neq i} \frac{1}{\lambda_i - \lambda_j}$.

Exercise 3.3. Show that V is harmonic, that is, $\Delta V = 0$.

Exercise 3.4. Define $u = \rho / V$, where ρ solves the above PDE. Then u solves the heat equation $\partial_t u = \frac{1}{2} \Delta u$.

Now, extend ρ to the whole space $\mathbb{R}^N$ in a symmetric way (from the Weyl chamber $\lambda_1 \geq \ldots \geq \lambda_N$). This means that u extends antisymmetrically to $\mathbb{R}^N$.

Assume that the initial eigenvalues for the DBM are distinct $\nu = (\nu_1 \geq \ldots \geq \nu_N)$. Then, by antisymmetric extension, $u(0,\lambda) = \frac{1}{V(\nu)} \sum_\sigma (-1)^\sigma \delta_{\lambda - \sigma(\nu)}$. The solution to the heat equation with this initial data is immediate:

$$u(t,\lambda) = \frac{1}{V(\nu)} \frac{1}{(2\pi t)^{N/2}} \sum_\sigma (-1)^\sigma \exp(-|\lambda - \sigma(\nu)|^2/(2t)),$$

which equals the determinant

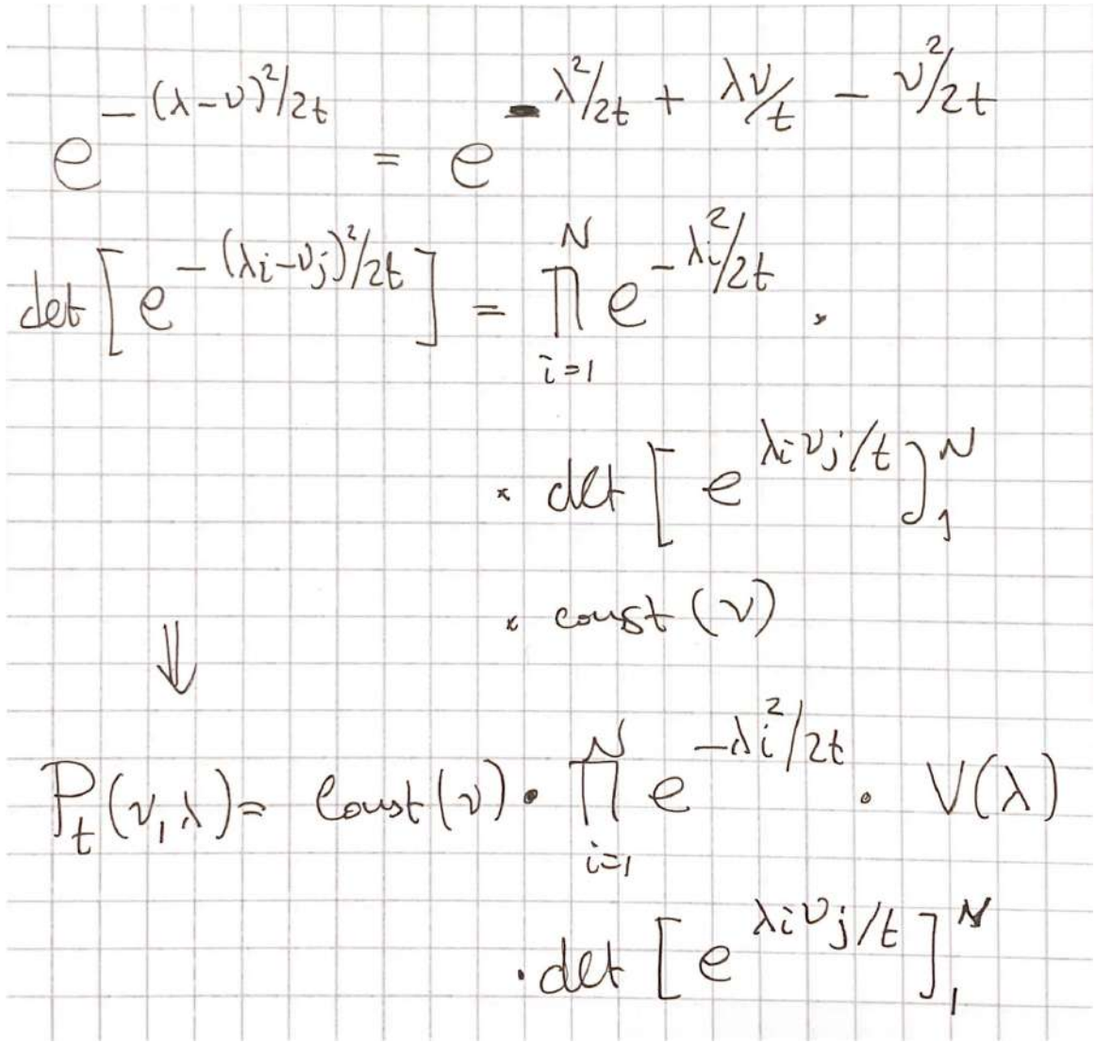
$$u(t,\lambda) = \frac{1}{V(\nu)} \frac{1}{(2\pi t)^{N/2}} \det\left[e^{-(\lambda_i - \nu_j)^2/(2t)}\right].$$

This leads to the formula for the transition density of DBM, which is due to Brezin-Hikami-Johansson (around 2000):

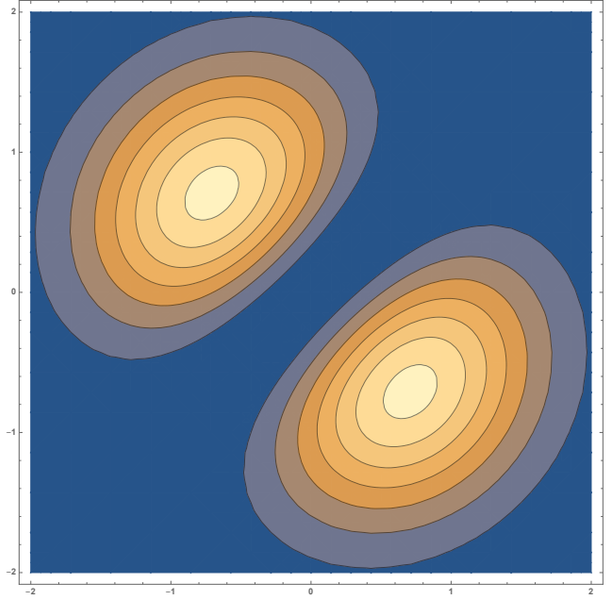
$$P_t(\nu \to \lambda) = \frac{1}{(2\pi t)^{N/2}} \frac{V(\lambda)}{V(\nu)} \det\left[e^{-(\lambda_i - \nu_j)^2/(2t)}\right].$$

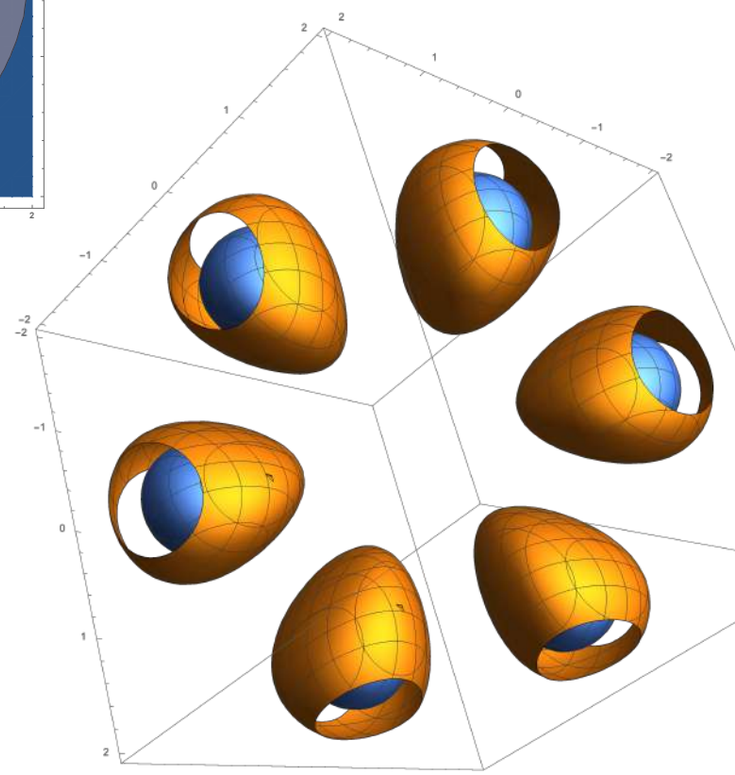
Taking limit $\nu \to 0$ gives the desired eigenvalue density of the GUE.

Exercise 3.5. Finish the proof of the eigenvalue density formula of the GUE by taking $\nu \to 0$ limit of the transition density. Hint: consider the case when ν is an arithmetic progression, and express $\det\left[e^{-(\lambda_i - \nu_j)^2/(2t)}\right]$ as another Vandermonde determinant.



“JACOBIAN” PROOF OF THE EIGENVALUE DENSITY





Here we prove again that the joint density of eigenvalues of the GUE of size N is given by

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

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

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

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 $2\pi r = \text{const} \cdot r$ 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 $U^* = U^{-1} = 1 - X + \text{smaller order terms}$. Since $A = U\Lambda U^*$, we can write

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

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 $f(A) = e^{-\frac{1}{2}\text{tr}(A^2)}$. □

EVALUATION OF THE NORMALIZATION CONSTANT

The normalization constant C_N in the GUE eigenvalue density is equal to $\frac{1}{(2\pi)^{N/2} 0!1!\ldots(N-1)!}$, if the integral is taken over the ordered set (Weyl chamber) $\{\lambda_1 \geq \ldots \geq \lambda_N\}$.

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

Such multidimensional integrals are quite interesting and important. A more general (but by far not in the full generality of known results) integral is known as the **Selberg integral:**

$$\int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} t_i^{\alpha-1}(1-t_i)^{\beta-1} \prod_{1\leq i<j\leq n} |t_i - t_j|^{2\gamma}\, dt_1 \cdots dt_n = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}.$$

The Selberg integral itself has several proofs (e.g., see [Forrester, Ch. 4]).

The Gaussian version (in fact, limit) of the Selberg integral is

$$\int_{\mathbb{R}^N} |V(x)|^{2c} \prod_i e^{-x_i^2/2} dx_i = N!(2\pi)^{N/2} \prod_{j=0}^{N-1} \frac{\Gamma(c(j+1))}{\Gamma(c)}.$$

Exercise 3.6. (you can volunteer for a talk)

Read on and present (any) proof of the Selberg integral formula.

The way we will compute the $c = 1$ integral as above would follow from representation theory plus a simple limit. The general Selberg integral follows from more involved representation theory around Macdonald symmetric functions.

NOTES AND REFERENCES

1. Tridiagonal model for GβE was discovered by Dumitriu and Edelman (*https://arxiv.org/abs/math-ph/0206043*)

2. Most of the discussion of the proof of the GUE eigenvalue density using DBM is taken from this *post* by T. Tao.

3. The second derivation of the GUE eigenvalue density (not using DBM) is more classical, and follows the idea around volume elements on orthogonal or unitary groups. This proof is present in many notes on random matrix theory.

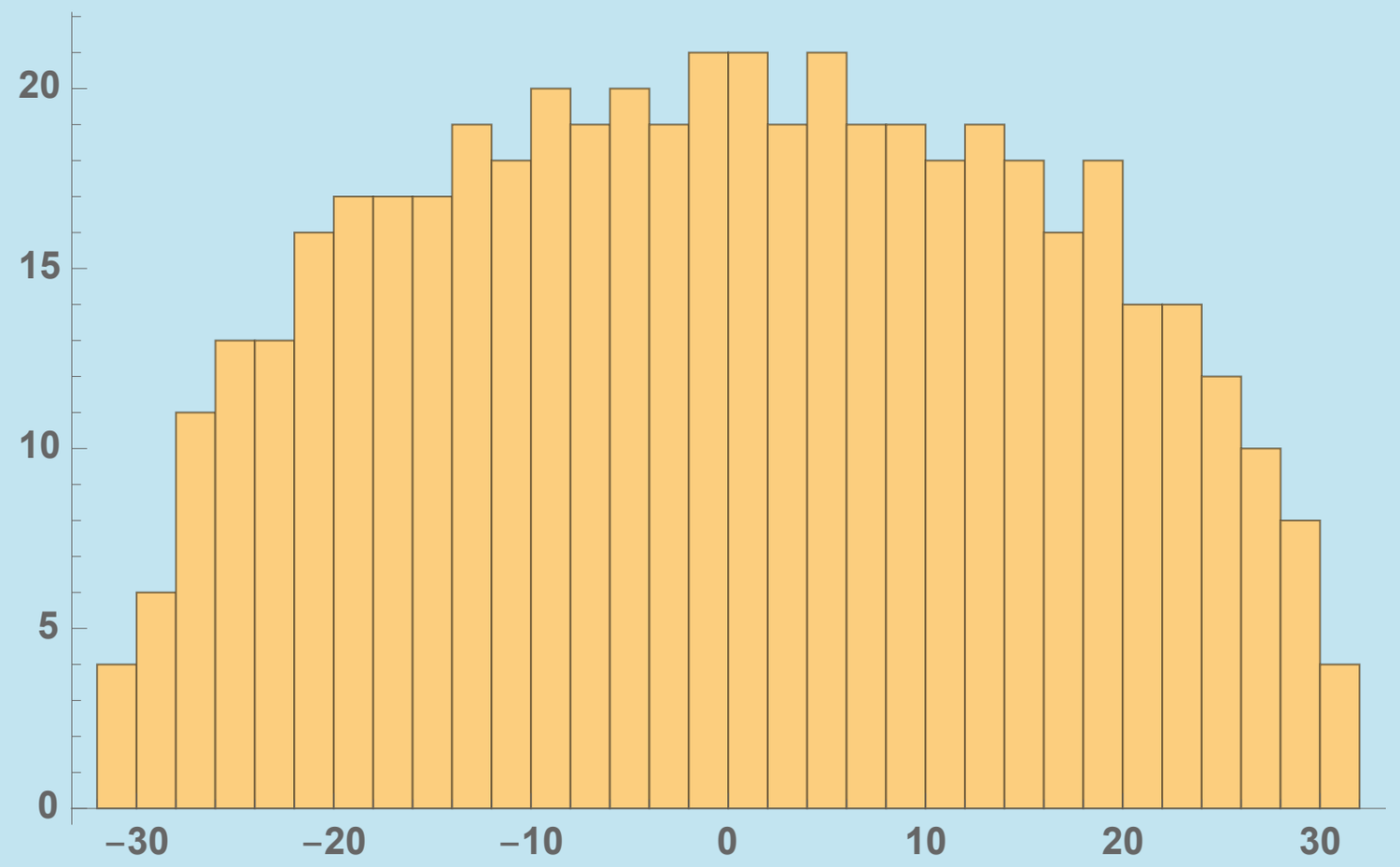
Chapter 4

SEMICIRCLE LAW

We discussed two generalizations of GUE with explicit eigenvalue densities - GβE and invariant ensembles.

We can also consider random real (or Hermitian, or quaternionic) matrices with independent entries. There is no explicit eigenvalue density in this case, but the limiting density of eigenvalues (as size of the matrix grows) obeys a universal law - the *Wigner Semicircle Law*. This chapter proves this law for real Wigner matrices, under the condition that all moments of the elements are finite.

The notes on Wigner Semicircle Law follow my previous incomplete notes (*https://github.com/lenis2000/RMT_Spring_2016/blob/master/Random_Matrices_Notes.pdf*)



REAL WIGNER MATRICES

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 $i \neq j$.
- 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.

2.3. **Formulation of the Wigner Semicircle Law.** For a real Wigner matrix $A_N \in \mathrm{Mat}(N \times N)$ let $\lambda_1^{(N)} \ge \cdots \ge \lambda_N^{(N)}$ 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.

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

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

See Figure 1.

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

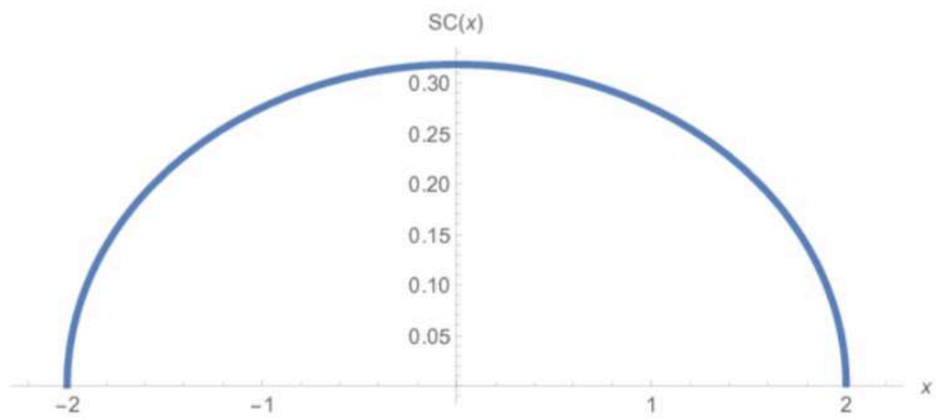


FIGURE 1. Semicircle density $\mathsf{SC}(x)$.

FORMULATION OF THE SC LAW

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 (nonrandom) 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 $N \to \infty$.

COMBINATORICS OF THE SEMICIRCLE DISTRIBUTION

First, computing the moments of the semicircle distribution, we have

$$m_{2k} = \int_{-2}^{2} x^{2k}(\frac{1}{2\pi}\sqrt{4-x^2})dx = \frac{1}{k+1}\binom{2k}{k}.$$

This is done by change the variables as $x = 2\sin\theta$, and computing recursively the integral of $\sin^{2k}\theta$.

Exercise 4.1. Prove the formula for the moments of the semicircle distribution

These quantities are the well-known *Catalan numbers*. Catalan numbers enumerate many interesting families of objects, including Dyck paths and trees. Many of the counting facts follows from the recurrence relation

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

Exercise 4.2. Prove that the Catalan recurrence (plus obvious initial values) leads to $\mathrm{Cat}_k = \frac{1}{k+1}\binom{2k}{k}$. Hint: from recurrence, get a quadratic equation for the generating function of Cat_k. Solve this quadratic equation, and expand the answer in a Taylor series.

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.

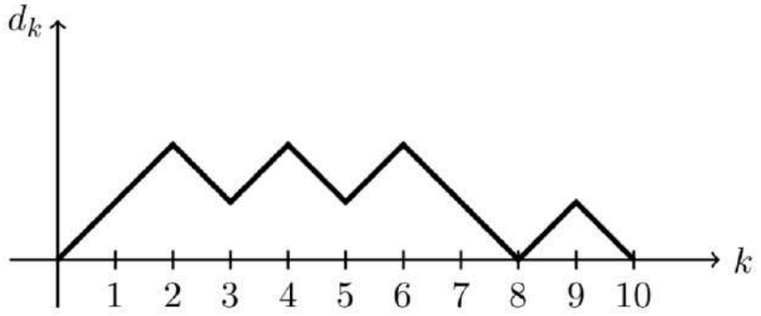


FIGURE 3. A Dyck path of length $2n = 10$.

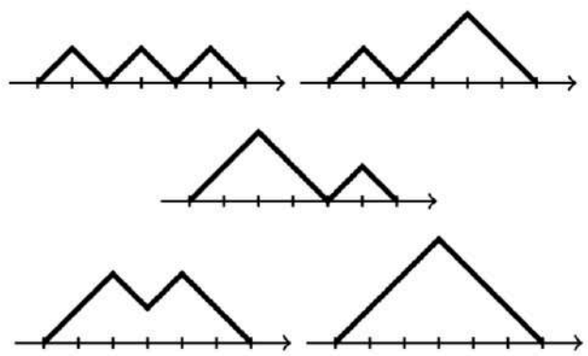


FIGURE 4. All five Dyck paths of length $2n = 6$. The first two paths first return to zero at time $2j = 2$, the third path first returns to zero at time $2j = 4$, and the last two paths first return to zero at time $2j = 6$.

Exercise 4.3. Prove that the number Dyck paths satisfies the Catalan recurrence

TREES

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).

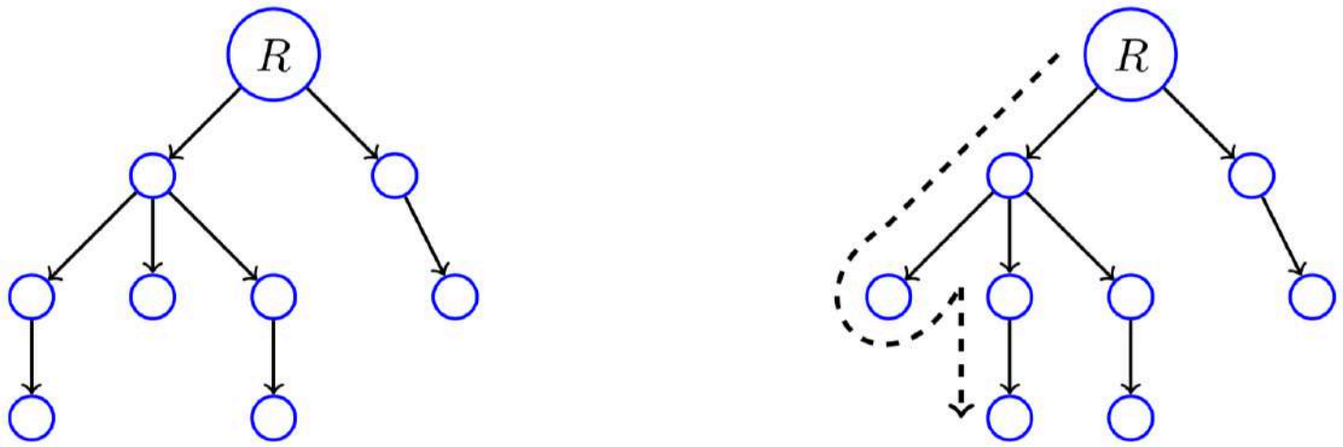


FIGURE 5. These trees are isomorphic as rooted trees, but are different as rooted ordered trees. A beginning of the walk of Exercise 2.14 is displayed for the second tree.

Exercise 4.4. Prove that the number of rooted ordered trees with $n+1$ vertices (including the root) is equal to the Catalan number Cat_n. Hint: use recurrence.

Exercise 4.5. Find and justify a bijection between rooted ordered trees and Dyck paths.

CONVERGENCE IN EXPECTATION

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

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

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

$$\begin{aligned} \mathbb{E}\langle x^k, L_N\rangle &= \mathbb{E}\int_{\mathbb{R}} x^k \, L_N(dx) \\ &= \mathbb{E}\int_{\mathbb{R}} x^k \frac{1}{N}\sum_{i=1}^{N} \delta_{N^{-1/2}\lambda_i}(dx) \\ &= \mathbb{E}\,\frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x^k \delta_{N^{-1/2}\lambda_i}(dx) \\ &= \mathbb{E}\,\frac{1}{N}\sum_{i=1}^{N}(N^{-1/2}\lambda_i)^k \\ &= N^{-1-k/2}\,\mathbb{E}\sum_{i=1}^{N}\lambda_i^k. \end{aligned}$$

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

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

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* $i_1 \dots i_k i_1$ 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 $(i_1, i_2), (i_2, i_3), \dots, (i_k, i_1)$ 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.

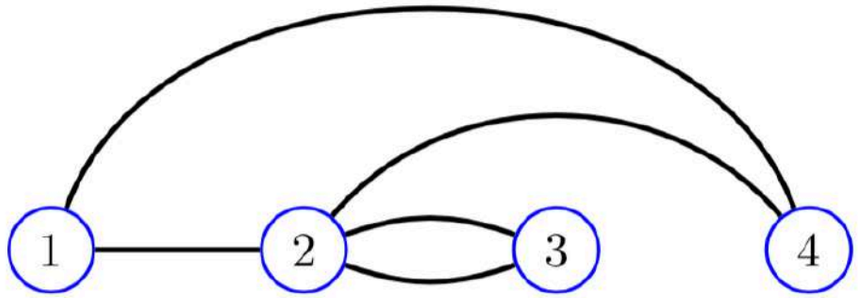


FIGURE 7. Graph G_w corresponding to the word $w = 123241$.

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

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

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 $G_{i_1 \dots i_k i_1}$ must have $N_e \geq 2$..
- The graph $G_{i_1 \dots i_k i_1}$ 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 $N^{\downarrow t} := N(N-1)\dots(N-t+1)$ ways to relabel the vertices to get an equivalent graph. This implies that the sum (2.10) can be rewritten as

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

where by $\mathrm{EqClass}_t$ 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 $i_1 \dots i_k i_1$ 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 $G_{i_1 \dots i_k i_1}$ 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

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

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

Exercise 4.6. There is a unique way of choosing representatives in the tree counting ➡️

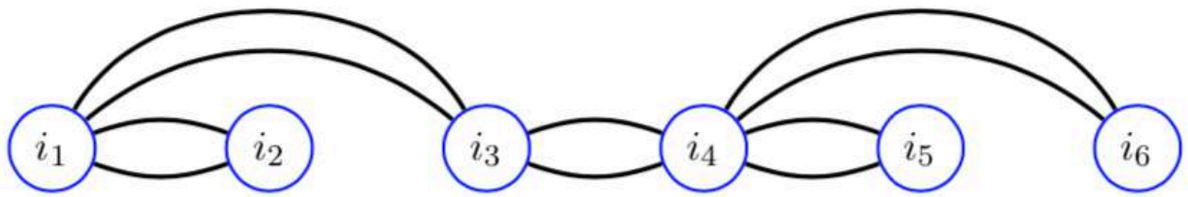
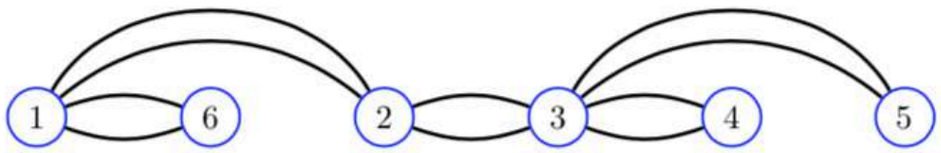


FIGURE 8. A graph G_w corresponding to a Wigner word $w = i_1 i_3 i_4 i_5 i_4 i_6 i_4 i_3 i_1 i_2 i_1$ which nontrivially contributes to the expansion (2.12). Here $k = 10$.



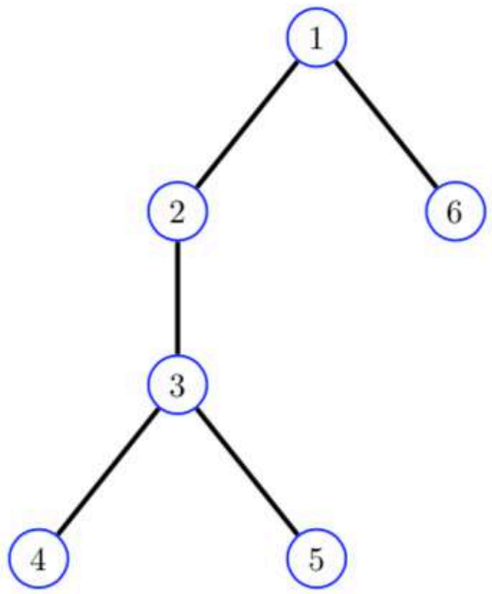


FIGURE 9. A representative graph $G_w \in \mathrm{EqClass}_{1+k/2}$ corresponding to the graph as in Figure 8 (left), and its representation as a rooted ordered tree (right).

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 $\left|\mathrm{EqClass}_{1+k/2}\right| = \mathrm{Cat}_k$, and finally proves the desired convergence (2.9).

Exercise 4.7. Use the tree counting to explicitly show that

$$\lim_{N\to\infty} \mathbb{E}(\mathrm{tr}(A^6)) = 5 = \mathrm{Cat}_3.$$

(see also section 2.8 in *these 2016 notes*)

VARIANCES

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

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

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

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

If the graphs $G_{\vec{i}}$ and $G_{\vec{j}}$ (corresponding to the words $i_1\dots i_k i_1$ and $j_1\dots j_k j_1$, 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.

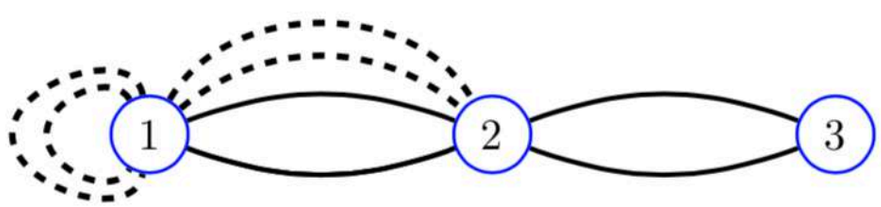


FIGURE 10. Graphs $G_{\vec{i}}$ (solid lines) and $G_{\vec{j}}$ (dashed lines) in Example 2.18.

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:

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

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

ESTIMATES AND LAST STEPS

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*

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

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.

Now, fix $B > 4$ (say, $B = 5$), and uniformly approximate the function $f\mathbf{1}_{|x|\le B}$ (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

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

Therefore, we can estimate

$$\begin{aligned}
|\langle f, L_N\rangle - \langle f, \mathsf{SC}\rangle| &\le |\langle f, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle| \\
&\le \big|\langle f\mathbf{1}_{|x|\le B}, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle\big| + \big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle| \\
&\le \big|\langle Q_\delta\mathbf{1}_{|x|\le B}, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle\big| \\
&\quad + \big|\langle f\mathbf{1}_{|x|\le B}, L_N\rangle - \langle Q_\delta\mathbf{1}_{|x|\le B}, L_N\rangle\big| + \big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle| \\
&\le |\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| + \big|\langle Q_\delta\mathbf{1}_{|x|>B}, L_N\rangle\big| \\
&\quad + \big|\langle f\mathbf{1}_{|x|\le B}, L_N\rangle - \langle Q_\delta\mathbf{1}_{|x|\le B}, L_N\rangle\big| + \big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle| \\
&\le |\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| + |\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| + \big|\langle Q_\delta\mathbf{1}_{|x|>B}, L_N\rangle\big| \\
&\quad + \big|\langle f\mathbf{1}_{|x|\le B}, L_N\rangle - \langle Q_\delta\mathbf{1}_{|x|\le B}, L_N\rangle\big| + \big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| + |\langle Q_\delta, \mathsf{SC}\rangle - \langle f, \mathsf{SC}\rangle| \\
&\le |\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| + |\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| \\
&\quad + \big|\langle Q_\delta\mathbf{1}_{|x|>B}, L_N\rangle\big| + \big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| + 2\delta.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}\Big(|\langle f, L_N\rangle - \langle f, \mathsf{SC}\rangle| > \varepsilon\Big) &\le \mathbb{P}\Big(|\langle Q_\delta, L_N\rangle - \mathbb{E}\langle Q_\delta, L_N\rangle| > \varepsilon/5\Big) + \mathbb{P}\Big(|\mathbb{E}\langle Q_\delta, L_N\rangle - \langle Q_\delta, \mathsf{SC}\rangle| > \varepsilon/5\Big) \\
&\quad + \mathbb{P}\Big(\big|\langle Q_\delta\mathbf{1}_{|x|>B}, L_N\rangle\big| > \varepsilon/5\Big) + \mathbb{P}\Big(\big|\langle f\mathbf{1}_{|x|>B}, L_N\rangle\big| > \varepsilon/5\Big).
\end{aligned}$$

The first summand above convergences to zero by Chebyshev inequality:

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

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).

NOTES AND REFERENCES

1. Catalan numbers count over 200 various families of objects (*http://www-math.mit.edu/~rstan/ec/catadd.pdf*)

2. My incomplete 2016 lecture notes are available at *https://github.com/lenis2000/RMT_Spring_2016/blob/master/Random_Matrices_Notes.pdf*

Chapter 5

ORBITAL MEASURES AND FREE OPERATIONS

- Limiting spectral distribution
- Operations on random matrices
- Orbital measures
- Moments and limiting spectral distribution
- Freeness
- Free convolution and related operations

The Wigner semicircle law can be applied to GUE matrices, too. (Well, we proved it for real matrices, but it also holds for complex ones.) Moreover, we know that GUE is infinitely divisible: $G \sim G_1 + G_2 + \ldots + G_k$, where G has normalized variance 1, and each G_i is an independent GUE, and has normalized variance $1/k$ (k is assumed finite). ("Normalized variance" is normalized by $1/N$ since we want the matrices themselves to have a limiting spectral density.)

In fact, this relation for GUE determines the limiting spectral distribution. Because all the G_j's are unitary invariant, the limiting spectrum of G "should not" depend on the eigenbases of these pieces, and only depend on their limiting spectra (which are the same, up to scale).

Therefore, it should be $\mathsf{SC} = \mathsf{SC}_1 \boxplus \mathsf{SC}_2 \boxplus \ldots \boxplus \mathsf{SC}_k$, where $\boxplus$ is a yet mysterious operation on spectra ("free convolution"), and SC_j are the rescaled copies of SC. The identity should hold for every k, and this determines the semicircle distribution.

This is a roundabout way of proving the semicircle law for the GUE (which we already almost established, modulo real/complex analogy), but the operation $\boxplus$ is very nice in its own right. This chapter discusses free convolution and related operations on random matrix spectra.

SPECTRAL DISTRIBUTION. ORBITAL MEASURES

Let $A = \{A_N\}$ be a family of Hermitian random matrices. We say that A has a limiting spectral distribution μ if $\frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(A_N)} \to \mu$, where the convergence is weak convergence in probability (like in the semicircle law).

We will only consider the compact case, i.e., when the limiting spectral distribution is supported on a compact interval. Then μ is a probability distribution on this compact interval.

Examples of existence of limiting spectral distributions include

- GUE matrices, properly normalized
- Complex Wigner matrices (Hermitian matrices with iid entries with all moments finite), properly normalized

These examples have SC as the limiting spectral distribution. Are there any other examples?

Exercise 5.1. Let μ be any probability distribution on a compact interval. Give an example of a family of random matrices which have μ as the limiting spectral distribution.

Let A_N be a fixed Hermitian matrix of size $N \times N$. Define $Orb_N(A_N) = UA_NU^*$, where $U \in U(N)$ is uniformly random (that is, distributed according to the Haar measure). We call the distribution of $Orb_N(A_N)$ the **orbital measure**.

The name comes from the fact that the unitary group $U(N)$ acts on the space $\mathscr{H}_N$ of $N \times N$ Hermitian matrices by conjugation, and orbital measures are precisely the "uniform" measures on orbits of this action.

Clearly, the spectrum of $Orb_N(A_N)$ is the same as that of A_N. However, the conjugation by a random unitary matrix rotates the eigenbasis of the Hermitian matrix at random, while keeping the eigenvalues.

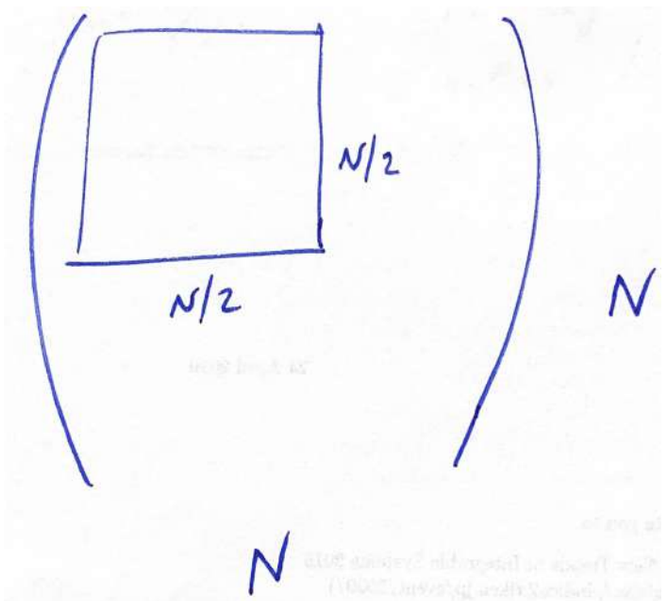
Call $Orb(A)$ the family of random matrices $\{Orb_N\}$ constructed from $A = A_N$ in this way.

OPERATIONS ON RANDOM MATRICES

Let A, B be two families of random matrices, such that entries of A are independent from those of B. Let A, B have the limiting spectral distributions μ, ν, respectively.

Exercise 5.2. Give examples of families A, B and families A', B' as above, such that the limiting spectral distributions of $A + B = \{A_N + B_N\}$ and $A' + B' = \{A'_N + B'_N\}$ are different.

The basic operation which we will consider is the addition. One can also multiply matrices and look at their spectra, or take one family A and look at "projections" - eigenvalue distributions of corners of A (usually of size proportional to N):



If we take arbitrary families of random matrices A, B as above (i.e., with entries independent from each other, and possessing limiting spectral densities), and consider sum, product, or projection of $Orb(A)$ and B, then it turns out that the limiting spectral distributions of the resulting family **depends** only on μ, ν, and can be described from them efficiently.

The idea is that $Orb(A)$ forgets all the information about relative positions of the eigenvectors of A, B, and only looks at eigenvalues.

Before we proceed to describing this in detail, let us discuss finite orbital measures (they are nice objects). This will open a direct path to the projection question.

In short, *what can we say about the random matrix from the orbital measure?*

FOURIER TRANSFORM OF PROBABILITY MEASURES

μ — measure on $\mathbb{R}^n$ (Borel, probability)

Fourier

$$\mathcal{F}: \mu \longmapsto \int e^{i\langle x, y\rangle} \mu(dy) =: f_\mu(x)$$

① f_μ characterizes μ

② Bochner's theorem, what can be said about f_μ if μ is a Borel probab. measure?

- $\mathcal{F}$: $\mathrm{Meas}(\mathbb{R}^n) \longleftrightarrow$ { complex-valued f on $\mathbb{R}^n$ s.t.
  - f is continuous
  - f is normalized, $f(0)=1$
  - f is nonneg. def.

  i.e.

$\overline{f(x)} = f(-x)$

$\forall x_i \in \mathbb{R}^n$, $c_i \in \mathbb{C}$, $\sum_{jk} c_j \bar{c}_k f(x_j - x_k) \geq 0$

Here we take the Euclidean space $\mathbb{R}^n$.

Exercise 5.3. Show that the space $\mathscr{H}_N$ of $N \times N$ Hermitian matrices is a real Euclidean space with the inner product $\langle A, B\rangle = \mathrm{tr}(AB)$.

Exercise 5.4. Show that the Fourier transform of a probability measure on $\mathbb{R}^n$ is nonnegative definite.

ORBITAL MEASURES, HCIZ INTEGRAL

Let $\lambda = (\lambda_1 \geq \ldots \geq \lambda_N)$ be fixed real numbers, and let D_λ be the (nonrandom) diagonal matrix with these eigenvalues. The orbital random matrix corresponding to D_λ, denoted by $Orb(D_\lambda) = UD_\lambda U^*$, where $U \in U(N)$ is random Haar, defines a probability distribution on $\mathscr{H}_N$. The goal is to compute its Fourier transform:

$$f_\lambda(A) = \int\limits_{U \in U(N)} e^{i \operatorname{Tr}(AUD_\lambda U^{-1})} dU,$$

where the integral is over the normalized (probability) distribution dU on $U(N)$.

Exercise 5.5. The integral $f_\lambda(A)$ depends only on the eigenvalues of A.

Theorem (Harish-Chandra, Itsykson, Zuber, HCIZ**).** For each $\lambda, a \in \mathbb{C}^n$ we have

$$\int\limits_{U \in U(N)} e^{\operatorname{Tr}(D_a U D_\lambda U^{-1})} dU = 0!1! \ldots (N-1)! \frac{\det\left[e^{a_j \lambda_k}\right]_{j,k=1}^N}{V(\lambda)V(a)}.$$

Here $V(a) = \prod\limits_{i<j} (a_i - a_j)$ is the Vandermonde determinant,

and same for $V(\lambda)$.

The proof of HCIZ will be given later in the course.

This formula is valid if all the a_i and λ_j are distinct. Otherwise, the value of the right-hand side should be defined by continuity.

How does this relate to projection?

$P = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}$,

$PUD_\lambda U^{-1}P$ - projected matrix $= \begin{pmatrix} \square & 0 \\ 0 & 0 \end{pmatrix}$

$\operatorname{Tr}(D_a P U D_\lambda U^{-1} P)$

$= \operatorname{Tr}(\underbrace{P D_a P}\, U^* D_\lambda U^{-1})$

$\| \quad \begin{pmatrix} a_1 & & a_k & & 0 \\ & \ddots & & & \\ 0 & & & & 0 \end{pmatrix}$

$\Rightarrow$ viewing $f_\lambda(a_1, \ldots a_k, 0, \ldots 0)$ as a Fourier transform on $\mathcal{H}_k$ provides distribution of e.v. of the $k \times k$ corner of $UD_\lambda U^{-1}$.

So, asymptotics of projections would follow from analysis of HCIZ integral with a proportion of a_i's set to zero, and when the size of $\det$ goes to infinity. Sounds *hard*.

HCIZ INTEGRAL AND RELATED STATEMENTS

- The proof of HCIZ integral can be deduced from the Dyson Brownian motion formulas considered earlier - if done carefully enough.
- There is one more computational proof of HCIZ using Schur symmetric functions and a couple of facts about them (this would be in the “discrete analogues” part)
- For a completely different proof using Duistermaat-Heckman theorem see *this post* by T. Tao.
- The right-hand side of HCIZ integral satisfies a “branching rule”, which allows to split off the last a_N and expand $f_\lambda(A)$ as an integral over the smaller space of eigenvalues of $(N-1)\times(N-1)$ matrices. This is checked by expanding the determinants.
- Thus, $f_\lambda(1)$ is related to the volume of the space of corners eigenvalues
- Corners eigenvalues interlace
- From this, we can obtain the following fundamental property: the corners eigenvalues of $UD_\lambda U^{-1}$ are distributed uniformly over all possible configurations (satisfying the interlacing).
- This space of allowed corners eigenvalues is called the Gelfand-Tsetlin polytope, and HCIZ allows to compute its volume.

Exercise 5.6. Obtain the “branching rule”, that is,
$f_\lambda(a_1,\dots,a_{N-1},0) = c_N \int_\mu f_\mu(a_1,\dots,a_{N-1})d\mu$, where the integral is taken over the Lebesgue measure on all configurations μ of size $N-1$ interlacing with λ. The interlacing means
$\lambda_N \le \mu_{N-1} \le \lambda_{N-1} \le \dots \le \lambda_2 \le \mu_1 \le \lambda_1$.

Use the right-hand side of HCIZ formula for f_λ and f_μ.

CONNECTION TO SPLINES

Let us set all $a_2 = \ldots = a_N = 0$, and keep only $a_1 = a$. Then $f_\lambda(A)$ is a Fourier transform of a probability measure on $\mathbb{R}$, which depends on the λ_i's. There are not many "good" measures on $\mathbb{R}$ depending on a large number N of parameters, and the one coming out of orbital measures has some very nice properties.

First, let us understand what happens when we set all these variables to zero. It turns out that there are better ways of going to zero: first take an arithmetic progression, and then send the difference to zero.

Then one can see a connection to splines.

Here is a definition of a B-spline, and a formula for its density.

A (fundamental, or B-) spline with knots $y_1 < \ldots < y_N$ is a unique function $x \mapsto M(x)$ such that

- The function is C^{N-3} everywhere
- The function vanishes outside (y_1, y_N)
- The function is equal to a polynomial of degree at most $N-2$ on each $[y_i, y_{i+1}]$
- $\int M(x)\,dx = 1.$

The formula for the density is

$$M(x) = (N-1) \sum_{i:\ y_i > x} \frac{(y_i - x)^{N-2}}{\prod_{r:\ r \neq i} (y_i - y_r)}.$$

Define divided differences by

$$f[y_1, y_2] = \frac{f(y_2) - f(y_1)}{y_2 - y_1}, \quad f[y_1, y_2, y_3] = \frac{f[y_2, y_3] - f[y_1, y_2]}{y_3 - y_1},$$

and so on.

Then for any function whose derivative is piecewise C^{N-1} we have

$$f[y_1, \ldots, y_N] = \frac{1}{(N-1)!} \int M(x) f^{(N-1)}(x)\,da.$$

In the exercises 5.7-5.10 below you can use these two properties of B-splines (explicit formula and connection to the derivative), but you can also prove them yourself beforehand.

CONNECTION TO SPLINES. EXERCISES

Exercise 5.7. Let $\phi_\lambda(a_1, \ldots, a_N)$ denote the HCIZ integral (you can use the determinantal formula for it). Compute the value $\phi_\lambda(0,0,\ldots,0)$ by setting $a_1 = 0, a_2 = \xi, \ldots, a_N = \xi(N-1)$ and noting a Vandermonde determinant on top. (Note: there is a much easier way of computing $\phi_\lambda(0)$, but suppose we didn't know that the determinantal formula represents an orbital integral.)

Exercise 5.8. Setting $a_1 = a, a_2 = 0, a_3 = \xi, \ldots, a_N = \xi(N-2)$ and using an expansion of the determinant along the first row or column, compute $\phi_\lambda(a,0,0,\ldots,0)$.

Exercise 5.9. Show that the divided difference can alternatively be written as

$$f[y_1, \ldots, y_N] = \sum_j \frac{f(y_j)}{\prod_{i:\, i \neq j}(y_j - y_i)}.$$

Exercise 5.10. Apply this to get the formula for the Fourier transform of a B-spline. Match this to $\phi_\lambda(ia,0,0,\ldots,0)$.

In particular, for an orbital random matrix $UD_\lambda U^*$, the distribution of any of its diagonal elements a_{ii} contains all the information about the eigenvalues λ_i (as knots of the spline which is the density of a_{ii}).

Next, we get back to the original questions of moments,
and leave the detailed study of corners spectra for later.

BACK TO SPECTRA. CAUCHY-STIELTJES

Using the explicit HCIZ integral turns out to be an "overkill" for dealing with the projection problem. It also does not help much for the other two problems, addition and multiplication of the matrices with limiting spectral distributions (and how to characterize the spectrum of the result).

First, let us take moments as characterizations of the eigenvalue distribution. We did this for the semicircle law. We also saw there that the variance goes to zero. So, to understand the limiting eigenvalue distribution, it is enough to look at

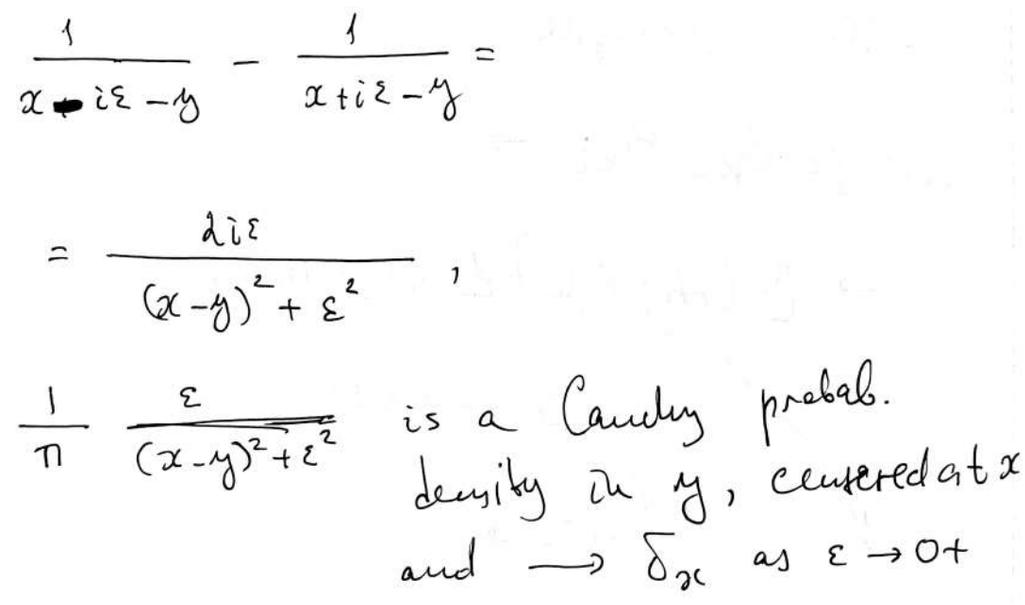
$\frac{1}{N}\mathbb{E}(\mathrm{tr}(A^k)) = \frac{1}{N}(\lambda_1^k + \ldots + \lambda_N^k) = \mathbb{E}\int x^k \, dL_N$, where

$L_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(A)}$ is the empirical spectral distribution of A.

Theorem (Cauchy-Stieltjes transform)**.** Suppose that $A = (A_N)$ is a sequence of $N \times N$ Hermitian matrices which have a compactly supported spectral distribution μ with density with respect to the Lebesgue measure. Let $\alpha_k := \lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(A^k)$. Then we have the following connection between μ and the sequence $\{\alpha_k\}$:

- $\sum_{k=0}^{\infty} \frac{\alpha_k}{z^{k+1}} = \int_{\mathbb{R}} \frac{\mu(dx)}{z-x}$, where z does not belong to the support of μ.
- In another direction, the density has the form $\mu(x) = \lim_{\epsilon \to 0+} \frac{G(x-i\epsilon) - G(x+i\epsilon)}{2\pi i}$.

The first is obvious. For the second, note that



So, $\frac{G(x-i\epsilon) - G(x+i\epsilon)}{2\pi i}$ is a convolution of μ with this δ sequence, and hence in the limit $\epsilon \to 0+$ we get the density.

MIXED MOMENTS

Cauchy-Stieltjes transform allows to reconstruct the limiting spectral density of matrices from their moments, and the moments are the limits of $\frac{1}{N}\mathbb{E}\mathrm{tr}(A^k)$. For example, in the first step of the proof of Wigner's semicircle law, we showed that for $A \sim$ GOE, $\frac{1}{N}\mathbb{E}\mathrm{tr}(A^k)$ converges to either $\mathrm{Cat}_{k/2}$ for k even, or 0 if k is odd (same holds for GUE, which we accept without proof).

From now on, we will ignore technical questions which were the other steps of the proof of Wigner's semicircle law, and focus on the "main computation" of the expectations of traces of powers.

Consider the question of adding random matrices, whose eigenspaces are in a "generic position" - i.e., $A_N + B_N$, whereA_N, B_N are families of (deterministic or random) matrices with limiting spectral distributions μ, ν, respectively, and B_N is unitary invariant (to ensure generic position of the eigenspaces).

To understand moments, we need to study $\frac{1}{N}\mathbb{E}\mathrm{tr}((A+B)^k)$, which expands into a sum of noncommutative monomials.

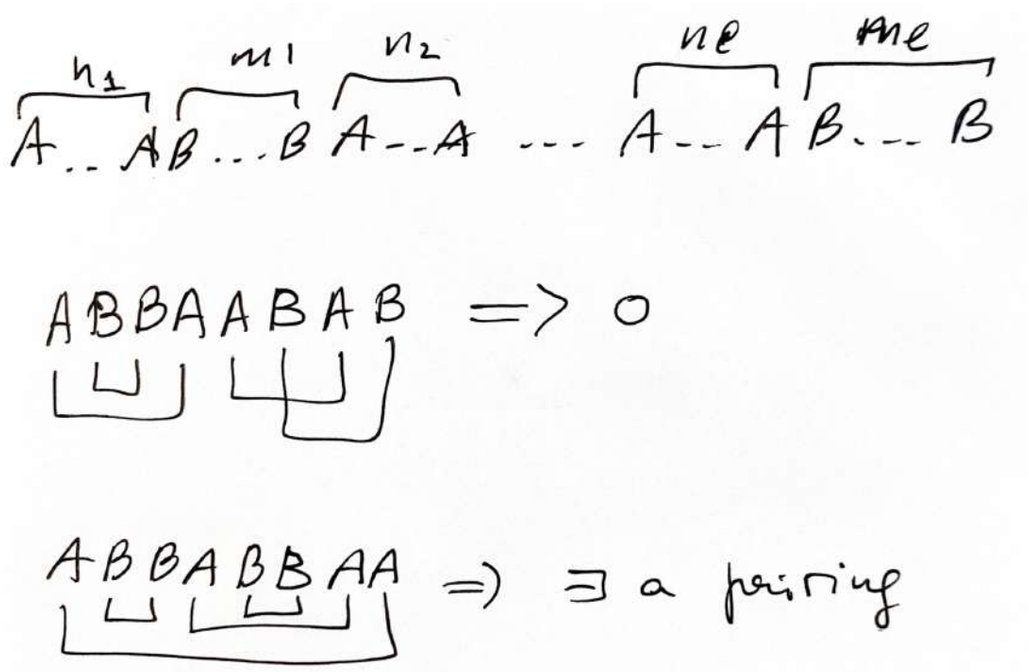
Definition. A *mixed moment* is an expression of the form $\frac{1}{N}\mathbb{E}\mathrm{tr}(A^{n_1}B^{m_1}\ldots A^{n_l}B^{n_l})$, or (depending on the context) its limit as $N \to \infty$.

The goal is to describe the situation in which mixed moments are determined by the "pure" moments $\lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(A^k)$ and $\lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(B^k)$. The way how they are determined is also nontrivial and interesting.

We start with the simple example, when A_N, B_N are two independent copies of GUE (or GOE, the analysis is basically the same), whose elements have variance $1/N$ (so that we do not need to normalize L_N, and L_N has the semicircle limit). In this case we know that their sum $A_N + B_N$ is also Gaussian, and so has the same semicircle density. However, the underlying combinatorics is going to be illuminating.

MIXED GAUSSIAN MOMENTS

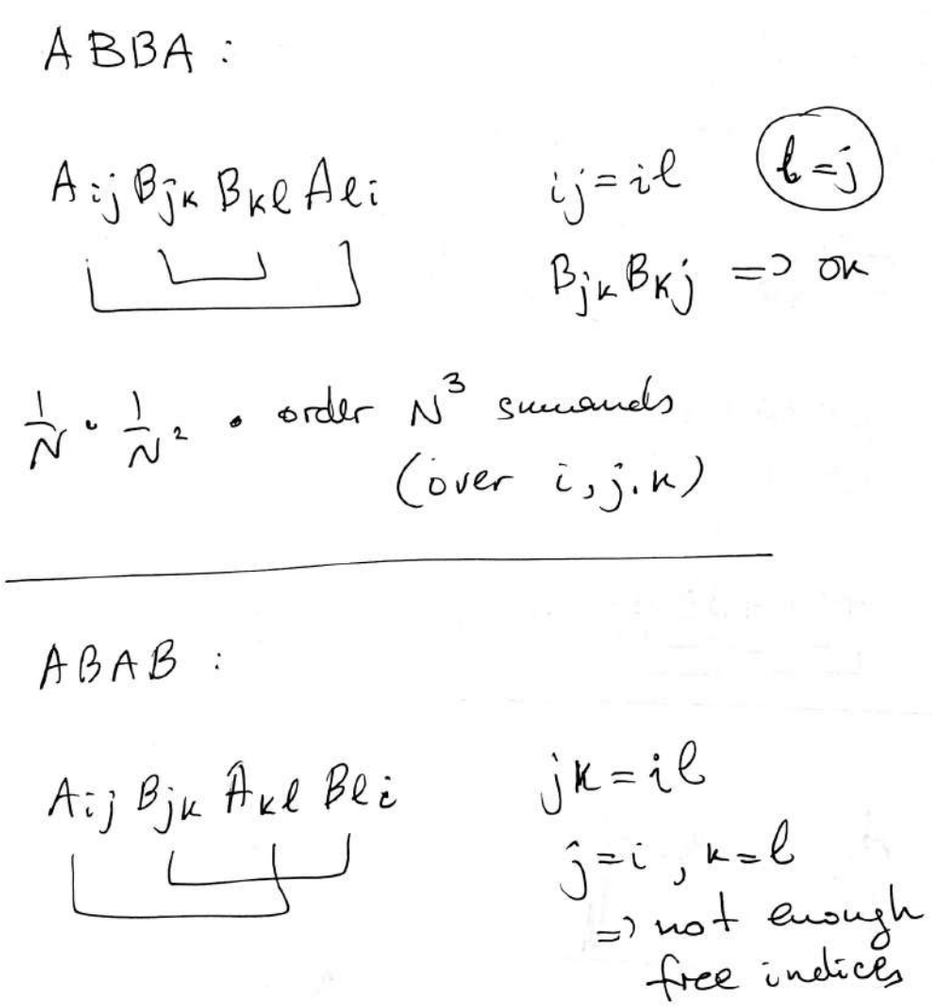
Theorem. Let A, B be families of independent Gaussian (GOE/GUE) random matrices. Then $\lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(A^{n_1}B^{m_1}\dots A^{n_l}B^{m_l})$ is equal to the number of noncrossing pairings in the word, without edges pairing an A to a B.



In particular, for words only with one letter (say, A), the number of noncrossing pairings is equal to the corresponding Catalan number.

Exercise 5.11. Finish the proof of the theorem about mixed Gaussian moments.

The **proof** is very similar to the first step in the Wigner semicircle law's proof. Namely, to avoid expectations of odd powers, each matrix element must have a "twin". They together have expectation $1/N$. To avoid convergence to zero, we need to have the maximal possible number of free indices. This is in bijection with noncrossing pairings. Finally, pairings cannot connect A to B as due to independence, they cannot be "twins".



MIXED GAUSSIAN MOMENTS

For words with one letter (i.e., A^k), the number of noncrossing pairings (the Catalan number) is also equal to the limit $\lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(A^k)$. Denote $\tau(M) := \lim_N \frac{1}{N}\mathbb{E}\mathrm{tr}(M)$.
Clearly, $\tau(1) = 1$, where 1 is the identity matrix.

Exercise 5.12. Show that the previous theorem implies that $\tau[(A^{n_1} - \tau(A^{n_1}))(B^{m_1} - \tau(B^{m_1}))\ldots(A^{n_l} - \tau(A^{n_l}))(B^{m_l} - \tau(B^{m_l}))]$ is equal to the number of noncrossing pairings of the same word, in which each group of A's and B's is connected to some other group. Here by subtracting $\tau(\cdot)$ we mean subtracting the corresponding multiple of the identity matrix.

Since there are no such noncrossing pairings which connect between groups, we have $\tau[(A^{n_1} - \tau(A^{n_1}))(B^{m_1} - \tau(B^{m_1}))\ldots(A^{n_l} - \tau(A^{n_l}))(B^{m_l} - \tau(B^{m_l}))] = 0$ for all possible words.

Turns out that the vanishing of "mixed centered moments" is more powerful than the noncrossing pairing interpretation of the mixed moments. This lies at the center of the definition of *asymptotic freeness*.

Definition. Two sequences of matrices A, B are called *asymptotically free* if $\tau[(A^{n_1} - \tau(A^{n_1}))(B^{m_1} - \tau(B^{m_1}))\ldots(A^{n_l} - \tau(A^{n_l}))(B^{m_l} - \tau(B^{m_l}))] = 0$ for all $l \geq 1$ and all $n_i, m_i \geq 1$.

- The previous computation shows that the pair of independent GUE sequences is asymptotically free.
- Voiculescu showed that if A, B have asymptotic spectral densities and B is unitary invariant, then these are asymptotically free.
- Let X_N be a real or complex Wigner matrix (iid entries), with mean zero and all moments, and A_N be a random matrix independent of X_N. Let A_N have an asymptotic spectral distribution, and $\sup_N \|A_N\| < \infty$. Then A_N, X_N are asymptotically free. (Thus, unitary invariance is not necessary for freeness.)

FREE AND USUAL CONVOLUTION

If we have two asymptotically free sequences of matrices, then

$\tau[(A^{n_1} - \tau(A^{n_1}))(B^{m_1} - \tau(B^{m_1}))\ldots(A^{n_l} - \tau(A^{n_l}))(B^{m_l} - \tau(B^{m_l}))] = 0,$

where τ is the limit of the expected normalized trace. Therefore, we in principle can reconstruct $\tau((A+B)^k)$ from $\tau(A^k)$ and $\tau(B^k)$. This is done by means of the *free convolution* – an operation that *linearizes* summation of asymptotically free random matrices.

Naive computations of moments $\tau((A+B)^k)$ using freeness are tedious, and we need a certain toolbox to tackle them.

So, before discussing free convolution, let us focus on how the usual convolution works in terms of moments, and what is our toolbox in the usual situation.

Let X, Y be two usual random variables. Assume they are independent and all their moments are finite:

$$\mathbb{E}(X^k) = \alpha_k, \qquad \mathbb{E}(Y^k) = \beta_k$$

Define $Z = X + Y$. How to express moments of Z, $\gamma_k = \mathbb{E}(Z^k)$, through the moments α_k, β_k?

Let us consider the first several moments. We have

$$\gamma_1 = \mathbb{E}(X+Y) = \alpha_1 + \beta_1;$$
$$\gamma_2 = \mathbb{E}(X^2 + 2XY + Y^2) = \alpha_2 + 2\alpha_1\beta_1 + \beta_2;$$
$$\begin{aligned}\gamma_3 &= \mathbb{E}(X^3 + 3X^2Y + 3XY^2 + Y^3) \\ &= \alpha_3 + 3\alpha_2\beta_1 + 3\alpha_1\beta_2 + \beta_3,\end{aligned}$$

and so on.

We see that these formulas are not so complicated, but still, γ_k is expressed not only through α_k, β_k, but also through all their lower combinations.

(USUAL) CUMULANTS

There is a better approach to the addition of the usual random variables through moments. The key is to pass from moments to cumulants. The cumulants of a random variable X are defined as follows. Let

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

be the moment generating function of X. Take its log, and expand:

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

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

Exercise 5.13. For the standard Gaussian random variable, we have $C(z) = z^2/2$, so only the second cumulant is nonzero (it is equal to 1).

Theorem. If $C_X(z)$ denotes the cumulant generating function of the variable X, then for adding independent random variables we have: $C_{X+Y}(z) = C_X(z) + C_Y(z)$. In other words, for independent r.v.'s cumulants simply add up.

We say that cumulants *linearize* addition of independent random variables.

For compactly supported r.v., convergence of cumulants clearly implies convergence of r.v.'s.

Exercise 5.14. Prove the Central Limit Theorem by cumulants. That is, if $X_1, X_2, \ldots$ are independent identically distributed random variables with all moments, and $Z_n = (X_1 + \ldots + X_n - n\mathbb{E}(X_1))/(\sqrt{nVar(X_1)})$, then show using the previous theorem that the cumulant sequence of Z_n converges to that of the standard normal random variable.

COMBINATORICS OF USUAL CUMULANTS

Define also the joint moments by $m_5(X, X, Y, X, Y) := \mathbb{E}(XXYXY)$, and so on.

The connection between the moment and cumulant generating functions for one random variable leads to the following definition of the joint cumulants:

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

This exercise implies that for independent random variables, we have $c_n((X+Y)^n) = c_n(X^n) + c_n(Y^n)$, as it should be. (Mixed cumulants appear in the expansion of the left-hand side, but only the pure cumulants survive.)

Thus, passing from moments to cumulants (via taking the log of the moment generating function) encodes the usual convolution (= addition of independent r.v.).

Both m and c are clearly multilinear functions.

Exercise 5.15. For independent random variables, all joint cumulants $c_n(X, X, Y, X, \ldots, Y)$ vanish

Combinatorially, this vanishing also means that we can compute any joint moments, for example:

$$\begin{aligned} m_4(X, X, Y, Y) = {} & c_2(X, X)c_2(Y, Y) + c_2(X, X)c_1(Y)c_1(Y) + c_2(Y, Y)c_1(X)c_1(X) \\ & + c_1(X)c_1(X)c_1(Y)c_1(Y), \end{aligned}$$

(This implies $\mathbb{E}(X^2Y^2) = \mathbb{E}(X^2)\mathbb{E}(Y^2)$.)

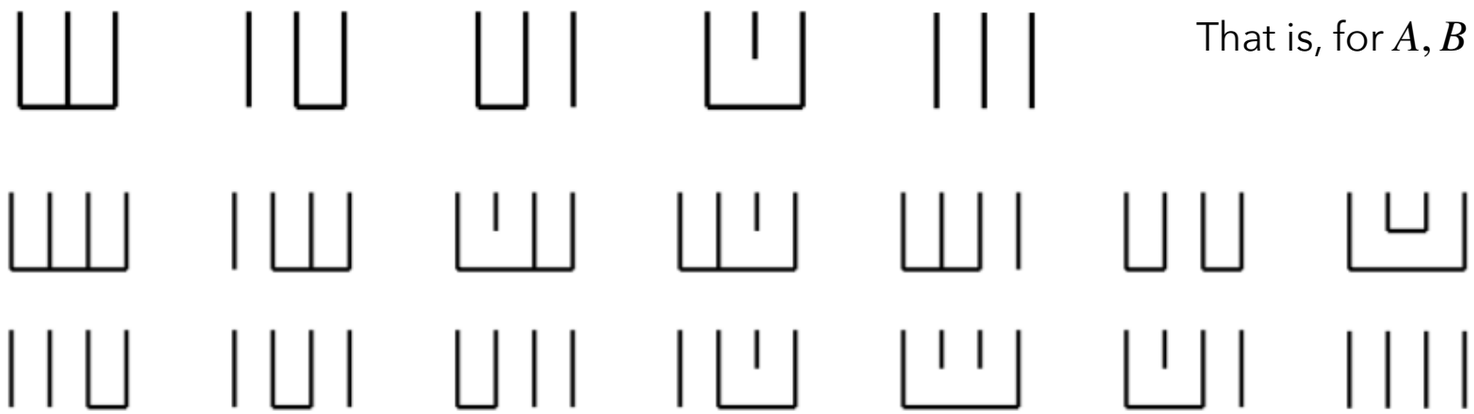
FREE CUMULANTS

The free joint cumulants are defined through the usual moments by

$$m_n(X_1, \ldots, X_n) = \sum_{\pi \in NC(n)} \prod_{B \in \pi} \kappa_{|\pi|}(X_i : i \in B),$$

where the sum goes over the set of noncrossing partitions of $\{1, \ldots, n\}$. Free cumulants are multilinear, too.

Here are examples of noncrossing partitions for $n = 3,4$:



Exercise 5.16. Find the "free Gaussian" distribution - a random variable X whose all pure free cumulants $\kappa_n(X, X, \ldots, X)$ vanish except $n = 2$.

Theorem (Speicher 1994**).** All free joint mixed cumulants $\kappa_n(A, A, B, A, \ldots, B)$, $n \geq 2$, vanish if and only if A, B are free as in the random matrix definition.

[To describe this in general we would need an algebra with a trace τ, but let us not go there.]

Therefore, free cumulants *linearize* addition of freely independent random variables (in particular, asymptotically free random matrices).

That is, for A, B asymptotically free, we have

$$\kappa_n(A + B) = \kappa_n(A) + \kappa_n(B).$$

The remaining goal is to turn this observation into a basis for computations with limiting spectral distributions. The main problem is to turn the definition of κ_n into a connection between free cumulants and moments.

R TRANSFORM

We need a power series identity equivalent to the identities $m_n = \sum_{\pi \in NC(n)} \kappa_{\pi_1} \dots \kappa_{\pi_{\ell(\pi)}}$. In fact, this identity is equivalent to the fact that the two power series

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

satisfy the relation $L(z) = K(z\,L(z))$. This is a formal power series identity which is an analytic identity if $|z|$ is small, and the random variables involved are compactly supported.

The function $K - 1$ linearizes free convolution (addition of free independent random matrices), that is, $K_{X \boxplus Y} - 1 = K_X - 1 + K_Y - 1$.

Recall the Cauchy transform $G(z) = \int \frac{d\mu(x)}{z - x} = \frac{1}{z} L(1/z)$.

We have from the relation: $K(G(z))/G(z) = z$.

Denote $V(z) = K(z)/z$, we have $V(G(z)) = z$. Therefore, V is an inverse function of the Cauchy transform of μ.

Finally, define the *Voiculescu R-transform* by

$$R(z) := V(z) - \frac{1}{z} = G^{(-1)}(z) - \frac{1}{z}.$$

Then we have $R_{X \boxplus Y}(z) = R_X(z) + R_Y(z)$.

Exercise 5.17. Compute the R-transform of the semicircle distribution.

Exercise 5.18. Let X, Y be free independent random variables, each of which takes values ± 1 with probabilities $\frac{1}{2}$. Compute the free convolution $X \boxplus Y$.

Note that the last problem is equivalent to computing the limiting spectrum of $A + UBU^*$, where U is Haar, and A, B are matrices with half of the eigenvalues 1 and the other half -1.

NOTES AND REFERENCES

1. Free probability for random matrices and beyond is discussed in surveys:

    1. “Free Probability Theory” by Roland Speicher (arXiv: 0911.0087)

    2. “Three Lectures on Free Probability” by Jon Novak (arXiv:1205.2097)

    3. “Free probability and random matrices” by Roland Speicher (arXiv:1404.3393)

2. Freeness (vanishing of mixed centrered moments) is equivalent to vanishing of mixed free cumulants, see this *paper by Speicher*.

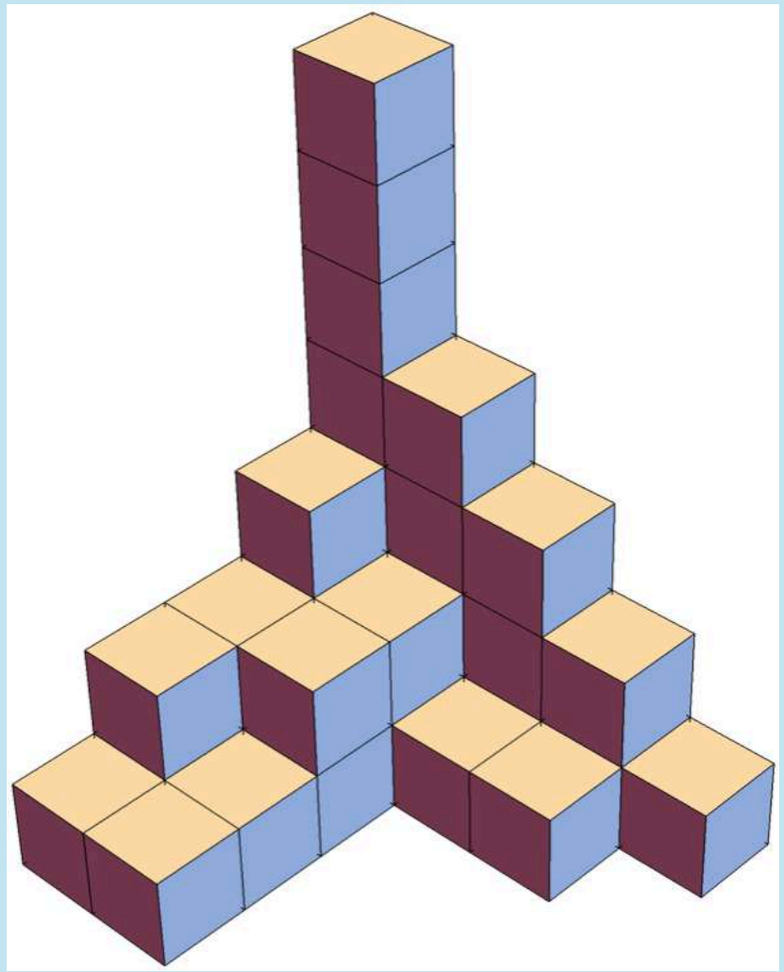
Chapter 6

DISCRETE ANALOGUES

- Representation theory background
- Schur-Weyl duality
- Counting dimensions in Schur-Weyl duality
- Robinson-Schensted-Knuth
- Joint distribution
- Limit of the joint distribution to GUE

A lot of formulas in unitary invariant random matrix theory (around GUE) follows from certain classical and natural representation theoretic constructions. The goal of this chapter is to explain how to use this connection to evaluate the normalization constant in the GUE eigenvalue density (which most commonly is derived from the Selberg integral).

The alternative derivation presented here could seem somewhat cumbersome to some, but it is very natural to others - likely this is a matter of taste.



NECESSARY BACKGROUND IN REP. THEORY

Let E be a finite-dimensional space over $\mathbb{C}$, and $\mathfrak{a} \subset \mathrm{End}(E)$ be an associative unital algebra ("unital" means that it contains the identity operator). In other words, E is a $\mathfrak{a}$-module.

A model for such an algebra is an algebra generated by all operators coming from a representation of a group in the space E.

Let us give three definitions.

- The algebra $\mathfrak{a}$ is *semisimple* if any submodule $E' \subset E$ is a direct summand. That is, if there is a basis of E such that every matrix from $\mathfrak{a}$ has a corner of zeroes in this basis, then in fact every matrix in $\mathfrak{a}$ has a block-diagonal form (maybe in some other "compatible" basis).

- The algebra $\mathfrak{a} \subset \mathrm{End}(V)$ is *irreducible* if E does not contain nontrivial $\mathfrak{a}$-invariant subspaces.

If $\mathfrak{a}$ is semisimple, then $E = \oplus_i E_i$ is a direct sum of spaces, where each E_i is a sum of equivalent copies of some irreducible modules. In other words, for a semisimple $\mathfrak{a}$ there exists a basis of E such that each of the matrices from $\mathfrak{a}$ in it has the form

$$A = \begin{pmatrix} A_1 & & & 0 & \vdots \\ & \ddots & & & \\ & & A_1 & & \\ & 0 & & A_2 & \\ & & & \ddots & \vdots \\ & & & & A_2 \\ & \dots & & \dots & \ddots \end{pmatrix}$$

In other words,

$$\mathfrak{a} = \bigoplus_i \mathrm{End}(V_i),$$

and

$$E = \bigoplus_i V_i \otimes W_i,$$

where $\dim W_i$ is equal to the multiplicity of the corresponding irreducible component. We have

$$\mathfrak{a} = \left\{ \bigoplus_i A_i \otimes I_{W_i} : A_i \in \mathrm{End}(V_i) \right\}$$

Exercise 6.1. (Burnside lemma) Show that if $\mathfrak{a}$ is irreducible, then $\mathfrak{a} = \mathrm{End}(E)$.

NECESSARY BACKGROUND IN REP. THEORY

For $\mathfrak{b} \subset \mathrm{End}(E)$, the *commutant* $\mathfrak{b}'$ is defined as $\mathfrak{b}' = \{A \in \mathrm{End}(E) \colon AB = BA \text{ for all } B \in \mathfrak{b}\}$.

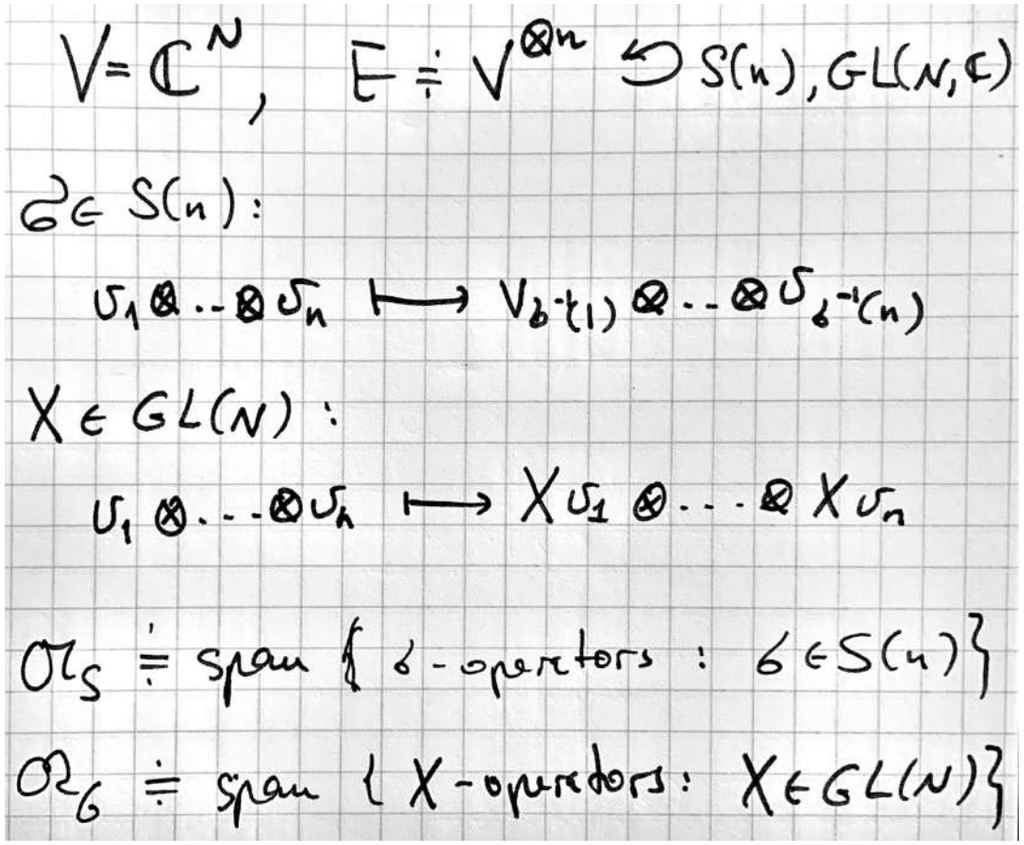
Exercise 6.2. For a semisimple $\mathfrak{a}$, its commutant has the form $\mathfrak{a}' = \left\{ \bigoplus_i I_{V_i} \otimes B_i \colon B_i \in \mathrm{End}(W_i) \right\}$. (Hint: consider one summand: $\mathfrak{a} = \mathrm{End}(V)$ and $E = V \otimes W$ and use the description of matrices which commute with the whole $\mathrm{End}(V)$.)

Then we see that the whole space E decomposes as

$$E = \bigoplus_i \mathrm{End}(V_i) \otimes \mathrm{End}(W_i).$$

Exercise 6.3. For a semisimple $\mathfrak{a}$, we have $(\mathfrak{a}')' = \mathfrak{a}$.

SCHUR-WEYL DUALITY



By the above "abstract" theory of matrix algebras, we thus see that the $(S(n), GL(N))$-bimodule $V^{\otimes n}$ decomposes as

$$V^{\otimes n} = \bigoplus_{\lambda} V_\lambda^{S(n)} \otimes V_\lambda^{GL(N)},$$

where λ replaces i (it is still some finite set of indices), and $V_\lambda^{S(n)}$, $V_\lambda^{GL(N)}$ are irreducible representations of the corresponding groups.

We have $(\mathfrak{a}_S)' = \mathfrak{a}_G$.

To prove this, observe $\mathrm{End}(V^{\otimes n}) = (\mathrm{End}(V))^{\otimes n}$, and so $(\mathfrak{a}_S)' = S^n(\mathrm{End}(V))$ is the space of symmetric tensors of order n over $\mathrm{End}(V)$.

Exercise 6.4. Show *polarization*, that is, $S^n(W) = \mathrm{span}\{X^{\otimes n} : X \in W\}$ for any finite-dimensional W.

We thus have

$$(\mathfrak{a}_S)' = \mathrm{span}\left\{X^{\otimes s} : X \in \mathrm{End}(V)\right\},$$

but this is the same as if we only took matrices from $GL(N, \mathbb{C})$.

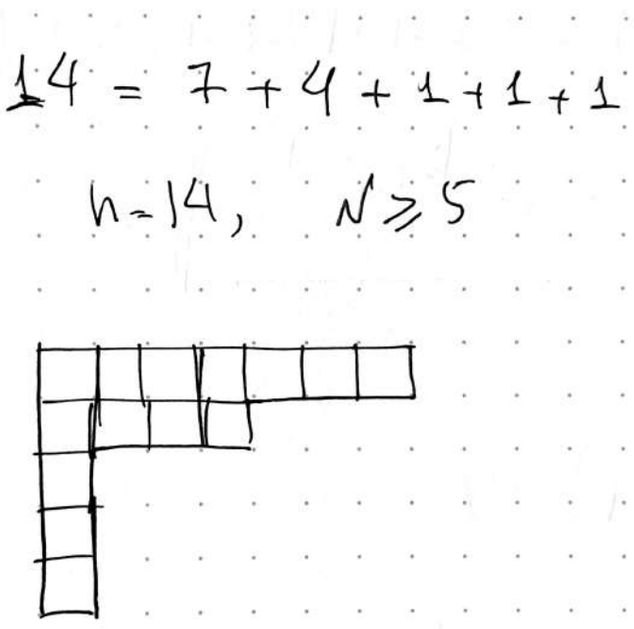
Exercise 6.5. Show the last assertion in the proof, that it suffices to only take matrices from $GL(N)$, i.e., invertible matrices for X.

∎

DIMENSION COUNTING

The Schur-Weyl duality leads to the following identity between dimensions: $N^n = \sum_\lambda \dim \lambda \cdot \dim_N \lambda$. Here the sum is over labels of irreducible representations of symmetric/linear groups, and $\dim \lambda, \dim_N \lambda$ are their respective dimensions (the first one for symmetric group, the second - for $GL(N)$).

It turns out that the labels are **partitions of n with at most N parts**. These objects are discrete analogues of random matrix spectra. Partitions are represented by Young diagrams. Notation: $\lambda = (\lambda_1 \geq \ldots \geq \lambda_N)$.



We will not prove this identification of labels.

On the other hand, will provide combinatorial interpretations of $\dim \lambda, \dim_N \lambda$ which will imply the following formulas:

$$\dim \lambda = \frac{n! \prod_{1 \leqslant i < j \leqslant N} (\lambda_i - \lambda_j + j - i)}{\prod_{i=1}^N (\lambda_i + N - i)!},$$

where N is arbitrary number greater than the number of nonzero parts in λ (the formula does not depend on N).

$$\mathrm{Dim}_N \lambda = \prod_{1 \leqslant i < j \leqslant N} \frac{\lambda_i - \lambda_j + j - i}{j - i}.$$

In other words, we will work with the formulas, without connecting them to dimensions of irreducible representations.

SCHUR-WEYL MEASURES ON PARTITIONS

The dimension counting identity $N^n = \sum_\lambda \dim \lambda \cdot \dim_N \lambda$ implies the definition of a probability distribution which is called the Schur-Weyl measure:

$$P_{N,n}(\lambda) = \frac{\dim \lambda \cdot \dim_N \lambda}{N^n}.$$

Because dimensions are nonnegative, this is indeed a probability distribution.

From the previous formulas for $\dim \lambda, \mathrm{Dim}_N \lambda$ we see that the probability weight $P_{N,n}(\lambda)$ contains a square of the Vandermonde in the shifted partition coordinates $l_i = \lambda_i + N - i, i = 1, \dots, N$. Thus, it is very natural to view $P_{N,n}$ as a **discrete analogue of a random matrix spectrum**.

There is one obstacle in getting a complete analogy, though. The GUE distribution has the form $\mathrm{const}(N) \prod_{i<j} (x_i - x_j)^2 \prod_{i=1}^N w(x_i)$ (where x_i are the eigenvalues and w is the Gaussian weight). There is no parameter n in the GUE

Exercise 6.6. $P_{N,n}(\lambda)$ cannot be written in the form $\mathrm{const}(N) \prod_{i<j} (l_i - l_j)^2 \prod_{i=1}^N w_N(l_i)$, where the constant and the function $w_N(l)$ do not depend on n.

This problem can be addressed by slightly modifying the measure $P_{N,n}$ by means of *randomizing* the parameter n, and this leads to better formulas in the end.

ROBINSON-SCHENSTED-KNUTH

The aim now is to give a combinatorial proof of the dimension counting formula $N^n = \sum_\lambda \dim \lambda \cdot \dim_N \lambda$, after

which the quantities $\dim \lambda, \dim_N \lambda$ will take their combinatorial meaning.

(Again, let me emphasize that we're not going to connect them to dimensions of irreducible representations.)

Consider a word $w = w_1 \dots w_n$ of length n from the alphabet $\{1,\dots,N\}$ (that is, each $w_i \in \{1,\dots,N\}$).

An *increasing subsequence* in this word is a subword $w_{i_1} \dots w_{i_{k'}}$ where $i_1 < \dots < i_{k'}$ and $w_{i_1} \leq \dots \leq w_{i_k}$. The length of the longest increasing subsequence in w, $LIS(w)$, is defined as the maximum over the lengths of all increasing subsequences in w. Note that this maximum can be achieved on more than one subwords.

For example, in the word $w = 2133215454$, each of $13345, 13355, 13344$ forms a longest increasing subsequence, and $LIS(w) = 5$.

How to compute $LIS(w)$ in general?

Note that we're not interested in the subsequence itself - only its length $LIS(w) \in \{1,\dots,n\}$.

There are algorithms for finding the subsequence, too, but we will present a linear time algorithm for getting $LIS(w)$.

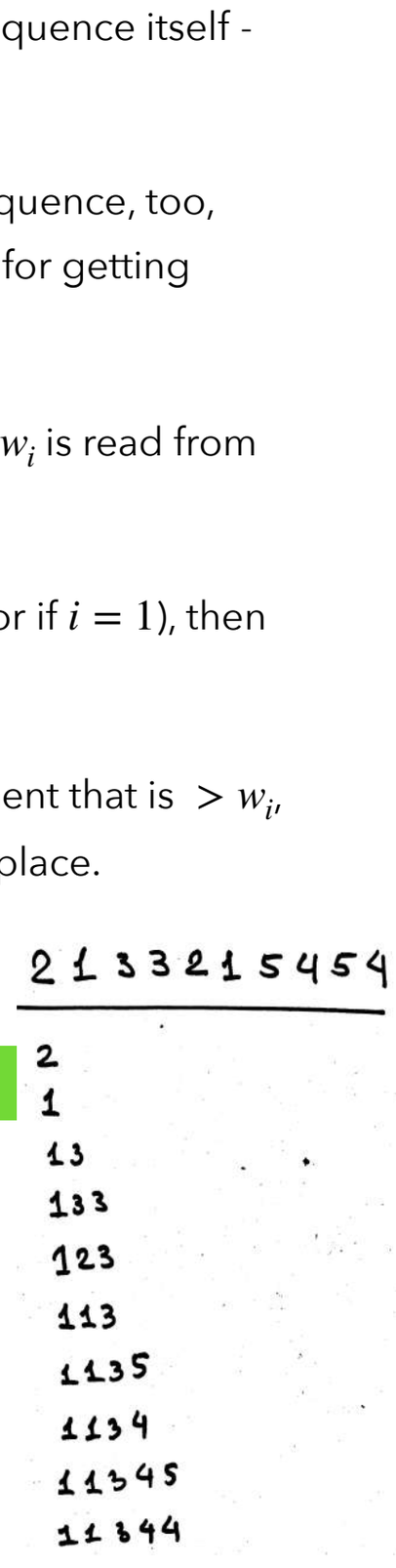
The algorithm has one buffer. A new letter w_i is read from the word at each step. Then:

1. If w_i is $\geq$ than everything in the buffer (or if $i = 1$), then put w_i at the end of the buffer
2. Otherwise, w_i bumps the smallest element that is $> w_i$, out from the buffer, and stands into its place.

The final length of the buffer is $LIS(w)$.

Exercise 6.7. Prove the last statement.

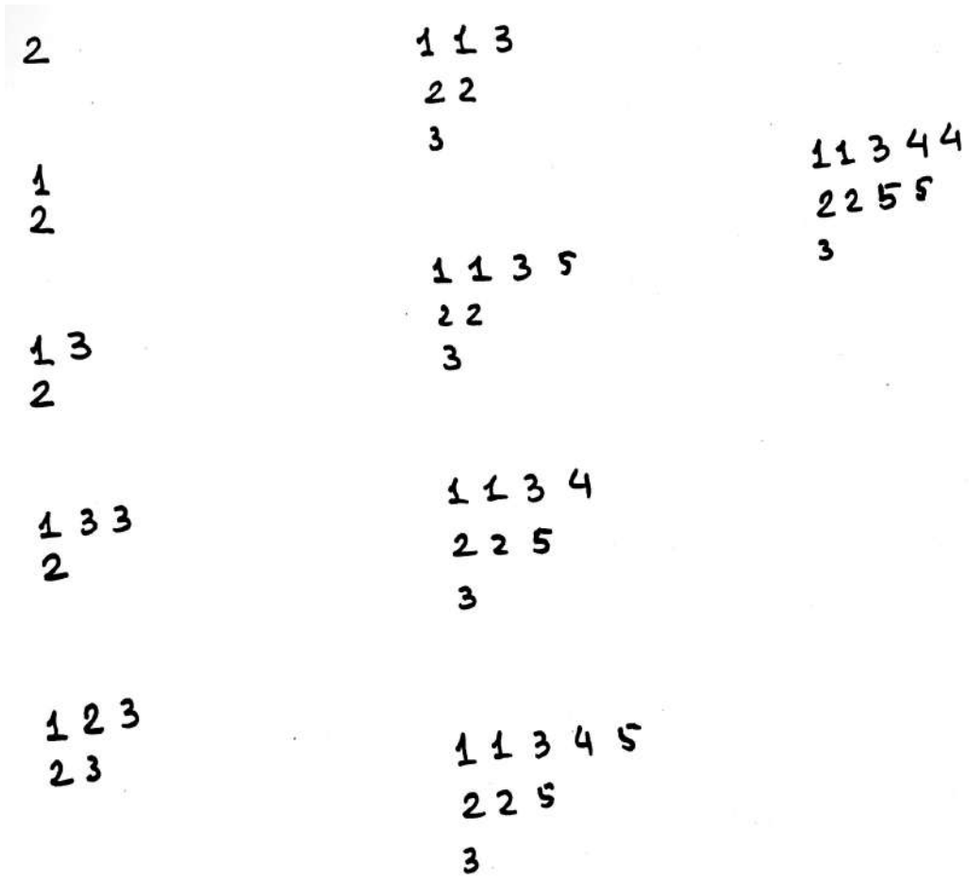
Example of the buffer's evolution for the word $w = 2133215454$:



ROBINSON-SCHENSTED-KNUTH

The RSK (Robinson-Schensted-Knuth) correspondence is a refinement of the above "LIS" algorithm. Namely, consider several buffers $1,2,3,\ldots$. Each new letter arrives into the first buffer and follows the rules as before. However, when a letter is bumped from some buffer j, it is not discarded as before, but rather it is inserted in the next buffer $j+1$.

Here is the example:

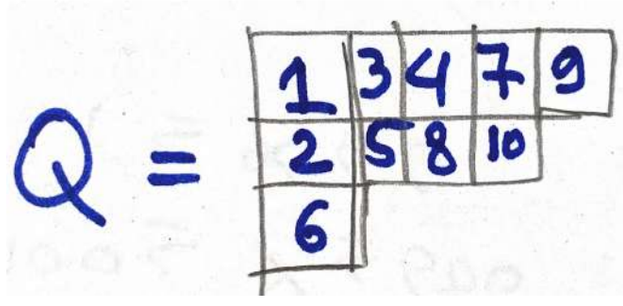


This collection of buffers is referred to as the *P-tableau*. It is a filling of a Young diagram ($\lambda = (5,4,1)$ in the example) by numbers from $\{1,\ldots,N\}$ such that the numbers weakly increase in the horizontal direction and strictly increase in the vertical direction. Such a filling is called a *semistandard Young tableau* of shape λ.

Exercise 6.8. Show that the map $w \mapsto$ P-tableau is not one-to-one.

To record the full information about the word w, let us also encode the place where the last letter settles in the construction of the P-tableau, at each step. This can be encoded by placing the step's number into the corresponding box. The result is a so-called *Q-tableau* - a semistandard filling of boxes of λ by distinct numbers from 1 to $|\lambda|$ (where $|\lambda|$ is the number of boxes). This filling is called a *standard Young tableau* of shape λ.

In the example we get:



ROBINSON-SCHENSTED-KNUTH

By definition, RSK is a map from the set of words to the set of pairs of Young tableaux (P, Q) of same shape, where P is semistandard with entries $\{1,2,\ldots,N\}$, and Q is standard.

Exercise 6.9. Show that the RSK map $w \mapsto (P, Q)$ is one-to-one. (Hint: construct its inverse, which can be done inductively step-by-step).

Comparing the sizes of the sets, we have

$$N^n = \sum_{\lambda} \#SYT(\lambda) \cdot \#SSYT_N(\lambda).$$

The sum is over all Young diagrams λ with n boxes and $\leq N$ rows. Thus, we have combinatorially (bijectively!) proven a formula corresponding to dimension counting in the Schur-Weyl duality.

In fact, $\#SYT(\lambda) = \dim \lambda$, and $\#SSYT_N(\lambda) = \mathrm{Dim}_N \lambda$.

From now on we will use the notations $\dim \lambda, \mathrm{Dim}_N \lambda$ to denote these combinatorial quantities, the numbers of standard / semistandard tableaux. We will proceed to compute them.

To summarize, $P_{N,n}(\lambda) = \dfrac{\dim \lambda \cdot \dim_N \lambda}{N^n}$ is a probability distribution on the set of Young diagrams λ with n boxes and $\leq N$ rows, which is obtained by taking a uniformly random word of length n from the alphabet $\{1,2,\ldots,N\}$, applying RSK to it, and reading off the shape of the tableaux.

In the next two steps we will generalize the input of the RSK to a certain Poisson random input, which will have random (and not fixed) length of the word. This generalization is convenient for the connection to random matrices.

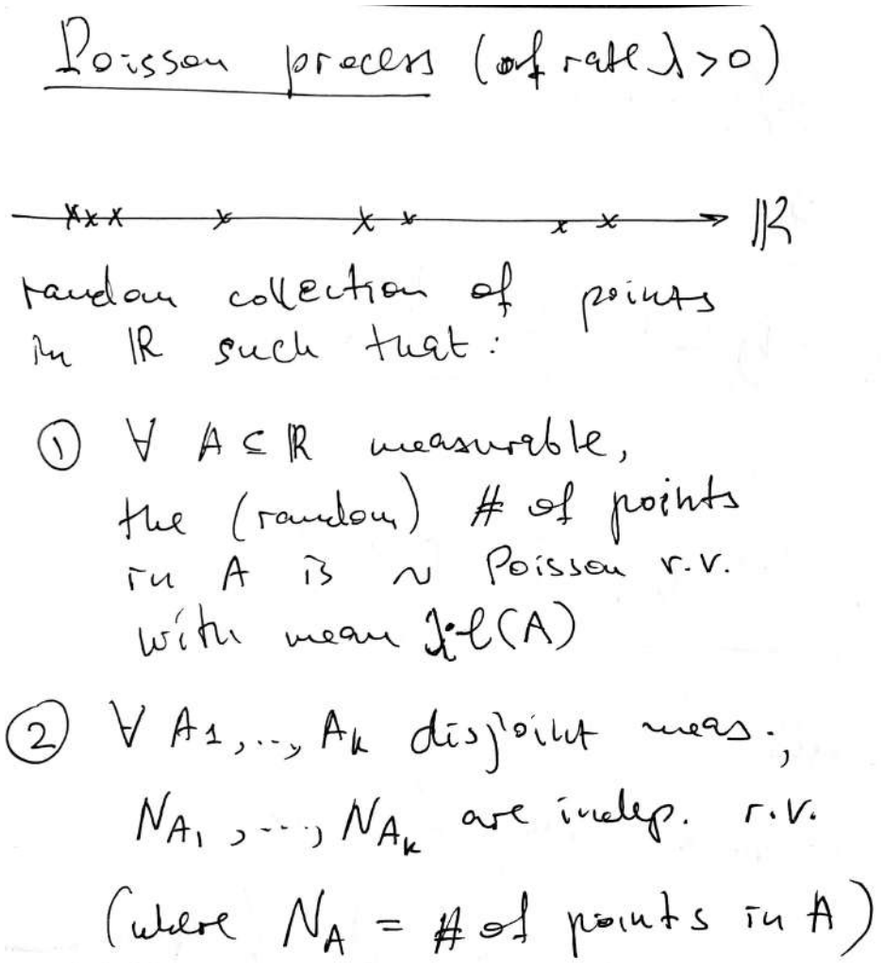
POISSONIZATION

Let us recall the Poisson process (on $\mathbb{R}$, but this construction works on any measurable space with a Radon measure).

The Poisson random variable of rate μ is

$$P(X=k)=\frac{\mu^k}{k!}e^{-\mu}.$$

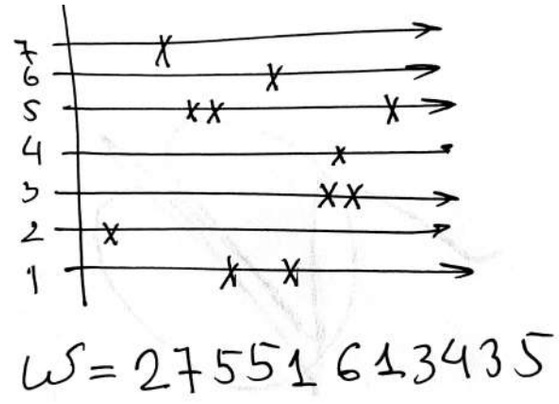
Poisson process is uniquely determined by:



Exercise 6.10. Find the distribution of the distance between two consecutive points in a Poisson process on $\mathbb{R}$ of rate λ.

We assume basic familiarity with Poisson processes.

Now take N independent Poisson processes of rate 1, located one under another. Start at 0 and fix "time" t. Then the configuration of the points in the Poisson processes can be read a word from the alphabet $\{1,\dots,N\}$. The length of this word is random, and has Poisson distribution with rate Nt.



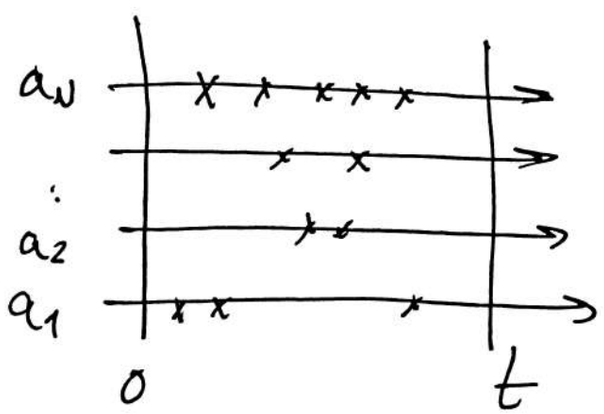
(Because there are only finitely many Poisson processes, the probability that there is a point in two or more of them at the same time is zero.)

Exercise 6.11. Conditioned on the length, the distribution of the word is uniform among all words of this length.

Proposition. Applying RSK to the Poisson random word we get measure $P_{N,t}(\lambda):=e^{-Nt}\frac{(Nt)^{|\lambda|}}{|\lambda|!}\frac{\dim\lambda\cdot\dim_N\lambda}{N^{|\lambda|}}$.

GENERALIZATION. PARAMETRIC MODEL

Now, let us generalize the distribution of the input word, and the measure $P_{N,t}$. Fix arbitrary parameters $a_1, \dots, a_N > 0$. Consider independent Poisson point processes of rates a_i, and fix a "time" parameter t. View the points of these Poisson processes as producing a random word from the alphabet $\{1,2,\dots,N\}$. The goal is to understand what is the distribution of the Young diagram obtained by applying the RSK to this random word.



The difference with $P_{N,t}$ is that now the random word will have different proportions of different letters. Let us fix these proportions, that is, the numbers $k_1, \dots, k_N$ of letters $1,\dots,N$ in a random word. Because these numbers are point-counts in independent Poisson processes, their joint distribution is simply $e^{-(a_1+\dots+a_N)t}\dfrac{a_1^{k_1}\dots a_N^{k_N}t^{k_1+\dots+k_N}}{k_1!\dots k_N!}$.

Exercise 6.12. Show that:

1. Conditioned on the number $n = k_1 + \dots + k_N$, the distribution of the random letter counts $k_1, \dots, k_N$ is multinomial $\binom{n}{k_1,\dots,k_N}\dfrac{a_1^{k_1}\dots a_N^{k_N}}{(a_1+\dots+a_N)^n}$
2. Conditioned on all letter counts $k_1, \dots, k_N$, the distribution of the word is uniform.

Using RSK, we see that the distribution of the standard tableau Q is still uniform (conditioned on the shape λ), while the distribution of the semistandard tableaux P depends on the parameters a_i. Namely, fix λ and set

$$s_\lambda(\vec{a}) = s_\lambda(a_1, \dots, a_N) := \sum_{P:\ sh(P)=\lambda} a_1^{k_1}\dots a_N^{k_N}.$$

Exercise 6.13. Show that the distribution of λ is

$$P_{N,\vec{a}}(\lambda) = e^{-t(a_1+\dots+a_N)}t^{|\lambda|}s_\lambda(\vec{a})\frac{\dim\lambda}{|\lambda|!}.$$

s_λ is called the **Schur polynomial**.

We have $s_\lambda(1,\dots,1) = \mathrm{Dim}_N\lambda$ (where 1 appears N times).

SCHUR POLYNOMIALS

To compute $\mathrm{Dim}_N\lambda$, $\dim\lambda$, let us first look at the Schur polynomials $s_\lambda(\vec{a})$ which arise in the parametric model. We will prove

Theorem. $s_\lambda(x_1,\dots,x_N)=\dfrac{\det[x_i^{\lambda_j+N-j}]_{i,j=1}^N}{V(\vec{x})}$.

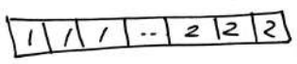
Here $V(\vec{x})=\prod_{1\le i<j\le N}(x_i-x_j)=\det[x_i^{N-j}]_{i,j=1}^N$ is the Vandermonde determinant.

① Example $\lambda=(n,0)$

$$\det\begin{bmatrix} x^{n+1} & y^{n+1}\\ 1 & 1\end{bmatrix}=x^{n+1}-y^{n+1}$$

$$S_{(n,0)}(x,y)=\frac{x^{n+1}-y^{n+1}}{x-y}.$$

② SSYT:



$$\sum_T = x^n + x^{n-1}y + x^{n-2}y^2 + \cdots + xy^{n-1} + y^n$$

To prove this theorem, we will show that both the determinantal formula for s_λ, and the definition of s_λ through the sum over semistandard tableaux satisfy the same recursion.

Def. Interlacing $\lambda=(\lambda_1\cdots\lambda_N)$, $\mu=(\mu_1,\dots\mu_{N-1})$

$\mu\prec\lambda$:

$$\lambda_N \overset{N}{\wedge} \mu_{N-1} \le \lambda_{N-1} \overset{N-1}{\wedge} \mu_{N-2} \le \cdots \le \lambda_2 \overset{2}{\wedge} \mu_1 \le \lambda_1$$

Recurrence:

$$S_\lambda(x_1,\dots,x_N)=\sum_{\mu:\mu\prec\lambda} S_\mu(x_1,\dots,x_{N-1})\,x_N^{|\lambda|-|\mu|}$$

Exercise 6.14. Show that the Schur polynomial defined as the sum over semistandard tableaux (in the previous page) satisfies the recurrence.

The next goal is to show that the determinantal formula satisfies the recurrence, too.

Exercise 6.15. Show that to prove the recurrence formula for general $x_1,\dots,x_N$, it suffices to prove it for $x_N=1$.

SCHUR POLYNOMIALS. RECURRENCE FOR DETERM'S

Let us take $N = 4$ for simplicity. Set
$\ell_i = \lambda_i + N - i = \lambda_i + 4 - i$, and
$m_i = \mu_i + N - 1 - i = \mu_i + 3 - i$.

We will perform the following operations with the determinant in the numerator:

- Subtract row j from row $j-1$ for all $j = 2, \ldots, N$.
- The resulting determinant's last column contains only one 1 and all other elements are zero, so we can reduce order of the determinant by 1.
- The i-th column then is divisible by $x_i - 1$, this is how the Vandermonde drops in order, too.
- After the division, use the multilinearity of the determinant to get the desired recurrence.

$$\frac{1}{V_4(x_1, x_2, x_3, 1)} \det \begin{bmatrix} x_1^{\ell_1} & x_2^{\ell_1} & x_3^{\ell_1} & 1 \\ x_1^{\ell_2} & x_2^{\ell_2} & x_3^{\ell_2} & 1 \\ x_1^{\ell_3} & x_2^{\ell_3} & x_3^{\ell_3} & 1 \\ x_1^{\ell_4} & x_2^{\ell_4} & x_3^{\ell_4} & 1 \end{bmatrix}$$

$$= \frac{1}{V_3(x_1, x_2, x_3)(x_1 - 1)(x_2 - 1)(x_3 - 1)} \det \begin{bmatrix} x_1^{\ell_1} - x_1^{\ell_2} & x_2^{\ell_1} - x_2^{\ell_2} & x_3^{\ell_1} - x_3^{\ell_2} \\ x_1^{\ell_2} - x_1^{\ell_3} & x_2^{\ell_2} - x_2^{\ell_3} & x_3^{\ell_2} - x_3^{\ell_3} \\ x_1^{\ell_3} - x_1^{\ell_4} & x_2^{\ell_3} - x_2^{\ell_4} & x_3^{\ell_3} - x_3^{\ell_4} \end{bmatrix}$$

$$= \frac{1}{V_3(x_1, x_2, x_3)} \det \begin{bmatrix} \sum_{\ell_2 \le m_1 < \ell_1} x_1^{m_1} & \sum_{\ell_2 \le m_1 < \ell_1} x_2^{m_1} & \sum_{\ell_2 \le m_1 < \ell_1} x_3^{m_1} \\ \sum_{\ell_3 \le m_2 < \ell_2} x_1^{m_2} & \sum_{\ell_3 \le m_2 < \ell_2} x_2^{m_2} & \sum_{\ell_3 \le m_2 < \ell_2} x_3^{m_2} \\ \sum_{\ell_4 \le m_3 < \ell_3} x_1^{m_3} & \sum_{\ell_4 \le m_3 < \ell_3} x_2^{m_3} & \sum_{\ell_4 \le m_3 < \ell_3} x_3^{m_3} \end{bmatrix}$$

∎

DIMENSION FORMULAS

First, we compute $\mathrm{Dim}_N\lambda = s_\lambda(1,\dots,1)$ (where 1 is repeated N times). Because s_λ is a polynomial, this substitution is valid. On the other hand, both the numerator and the denominator in $\dfrac{\det[x_i^{\lambda_j+N-j}]}{V(\vec{x})}$ are singular when the x_i's are equal. We will compute the limit as $\vec{x}\to(1,\dots,1)$ of the ratio along a particular direction. But since the limit exists, this will give us the answer.

Set $x_i = q^{i-1}$ (we'll send $q\to 1$). We have

$$\begin{aligned}\det[x_i^{\lambda_j+N-j}] &= \det[q^{(i-1)(\lambda_j+N-j)}]\\ &= \det[(q^{\lambda_j+N-j})^{i-1}]\\ &= (-1)^{\binom{N}{2}}\prod_{i<j}(q^{\lambda_i+N-i}-q^{\lambda_j+N-j}),\end{aligned}$$

as we recognize the Vandermonde determinant.

Lemma. $\lim\limits_{q\to1}\dfrac{q^a-q^b}{q^c-q^d}=\dfrac{a-b}{c-d}$.

Thus,

$$\begin{aligned}\mathrm{Dim}_N\lambda &= \lim_{q\to1} s_\lambda(1,q,\dots,q^{N-1})\\ &= (-1)^{\binom{N}{2}}\lim_{q\to1}\prod_{1\le i<j\le N}\frac{q^{\lambda_i+N-i}-q^{\lambda_j+N-j}}{q^{i-1}-q^{j-1}}\\ &= \prod_{i<j}\frac{\lambda_i-\lambda_j+j-i}{j-i}.\end{aligned}$$

This computes the dimension $\mathrm{Dim}_N\lambda$, which is the number of semistandard Young tableaux of shape λ. To compute the number of standard Young tableaux, we "just need to take N very large". In other words, if λ is fixed and N is very large, then most of the diagrams will have distinct entries. Namely:

Exercise 6.16. Show that $\lim\limits_{N\to\infty}\dfrac{\mathrm{Dim}_N\lambda}{N^{|\lambda|}}=\dfrac{\dim\lambda}{|\lambda|!}$.

We would like to show that (here $n=|\lambda|$, and $N\ge\ell(\lambda)$ is arbitrary):

$$\dim\lambda=\frac{n!\prod_{1\leqslant i<j\leqslant N}(\lambda_i-\lambda_j+j-i)}{\prod_{i=1}^N(\lambda_i+N-i)!},$$

DIMENSION FORMULAS

We have

$$\prod_{1\le i<j\le N}\frac{\lambda_i-\lambda_j+j-i}{j-i}=\prod_{1\le i<j\le \ell(\lambda)}\frac{\lambda_i-\lambda_j+j-i}{j-i}$$
$$\times\prod_{i\le\ell(\lambda),\ j>\ell(\lambda)}\frac{\lambda_i+j-i}{j-i}$$
$$\times\prod_{\ell(\lambda)<i<j\le N}\frac{j-i}{j-i}$$
$$=\frac{\prod_{1\le i<j\le\ell(\lambda)}(\lambda_i-\lambda_j+j-i)}{0!1!\dots(\ell(\lambda)-1)!}$$
$$\times\prod_{i\le\ell(\lambda),\ j>\ell(\lambda)}\frac{\lambda_i+j-i}{j-i}.$$

The first of the factors has a limit, and so we need to consider the second factor. Actually, the product over i is finite, so it suffices to take a single i.

$$\prod_{j=\ell(\lambda)+1}^{N}\frac{\lambda_i-i+j}{j-i}=\frac{(\ell(\lambda)-i)!}{(\lambda_i+\ell(\lambda)-i)!}\frac{\Gamma(\lambda_i-i+N)}{\Gamma(N-i)}.$$

The first factor is not growing, and for the Gamma function there is a lemma:

Lemma. For all $\alpha\in\mathbb{C}$ we have $\lim\limits_{|z|\to\infty}\dfrac{\Gamma(z+\alpha)}{\Gamma(z)}z^{-\alpha}=1$,

where z is complex of argument $\neq\pi$.

Exercise 6.17. Prove this lemma.

Using this lemma, we see that the ratio of the Gamma functions produces N^{λ_i}, which cancels the factor $N^{-|\lambda|}$ in Exercise 6.16.

Putting all together, we obtain the desired formula for $\dim\lambda$. We proved the following representation, where $\vec{\ell}=(\lambda_1+N-1,\lambda_2+N-2,\dots,\lambda_N)$:

$$P_{N,t}(\lambda)=c_N V(\vec{\ell})^2\prod_{i=1}^{N}w(\ell_i),\ \text{where } c_N=\frac{t^{-\binom{N}{2}}}{0!1!\dots(N-1)!},$$

and $w(\ell)=\dfrac{e^{-t}t^{\ell}}{\ell!}$, $\ell\in\mathbb{Z}_{\ge0}$ is the Poisson distribution.

LIMIT TO GUE, AND SELBERG INTEGRAL AGAIN

The distribution $P_{N,t}$ looks very similar to GUE. The latter has the density of eigenvalues $\overrightarrow{x} = (x_1 \geq \ldots \geq x_N)$:

$P_{GUE}(x_1, \ldots, x_N) = c_N^{GUE} V(\overrightarrow{x})^2 \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2}$. Let us obtain the GUE as a limit of the Schur-Weyl measures $P_{N,t}$, which will also bring us the proof of the value of the constant c_N^{GUE}. We begin with $N = 1$:

Exercise 6.18. Using Stirling's formula, show the local central limit theorem for the Poisson random variable:

$e^{-t}\frac{t^k}{k!} \sim \frac{1}{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ as $t \to \infty, k = \lfloor t + z\sqrt{t} \rfloor$

The division by the square root of t is not surprising - it corresponds to the scaling of the discrete space to continuous.

Exercise 6.19. Show that for general N, we have $P_{N,t}(\lambda_1, \ldots, \lambda_N) \sim t^{-N/2} P_{GUE}(x_1, \ldots, x_N)$, where $\lambda_i = t + x_i\sqrt{t}$ and $t \to +\infty$.

This representation of the GUE eigenvalue density as a limit of the discrete Schur-Weyl distributions is useful because it proves the value of the GUE normalizing constant c_N^{GUE}, which is the (inverse of) the value of the Selberg integral:

$$\iint\limits_{x_1 \geq \ldots \geq x_N} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \, dx_1 \ldots dx_N = 0!1! \ldots (N-1)!$$

That is, $c_N^{GUE} = \frac{1}{0!1! \ldots (N-1)!}$.

NOTES AND REFERENCES

1. There are many good books on representation theory, including H. Weyl, “The Classical Groups: Their Invariants and Representations” (1946, reprinted many times).

2. Combinatorics of Young tableaux is discussed in several books, including

    1. R. Stanley “Enumerative Combinatorics”

    2. B. Sagan “The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions”

    3. W. Fulton “Young tableaux”

3. Exercise 6.7 is the first case of the Greene’s Theorem. For one of its proofs see, for example, *these video lectures* by X. Viennot (at the mark “increasing and decreasing subsequences f maximal size”).

4. On algorithms for finding the longest increasing subsequence itself see, for example,

    1. *https://stackoverflow.com/questions/2631726/how-to-determine-the-longest-increasing-subsequence-using-dynamic-programming/2631810*

    2. *https://en.wikipedia.org/wiki/Longest_increasing_subsequence*

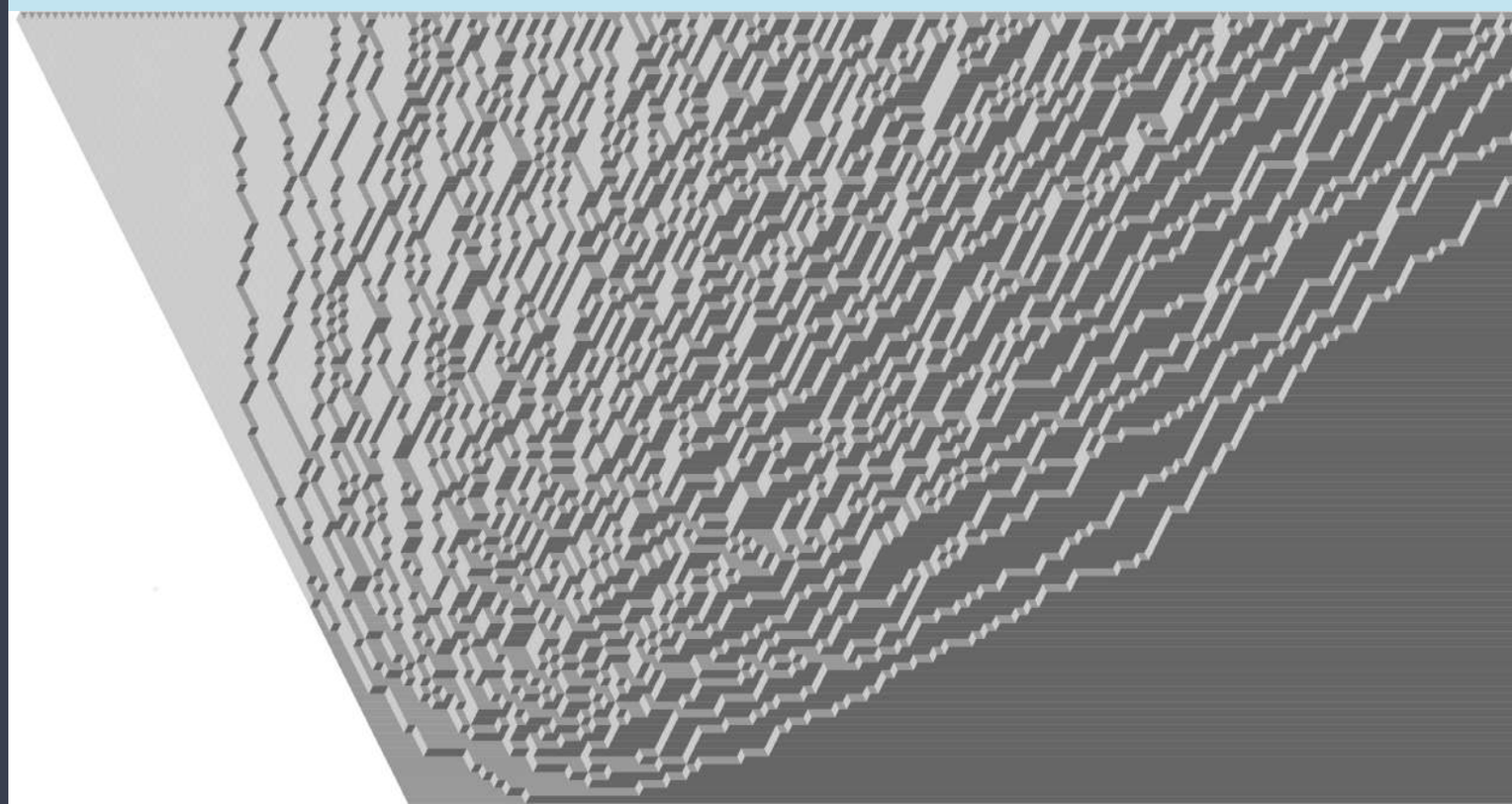
5.

Chapter 7

DETERMINANTAL POINT PROCESSES

- Definitions
- Biorthogonal and orthogonal ensembles
- Determinantal structure of GUE and Schur-Weyl measures

Our next goal is to describe local Asymptotics fo the GUE eigenvalue distribution. This can be done using *determinantal point processes* - a powerful algebraic framework for studying random point configurations featuring “free fermionic” structure. We will discuss basic definitions, certain interesting subclasses of determinantal processes, and then apply this theory to Schur-Weyl measures and the GUE eigenvalue distribution.



7.1 Point processes

Let $\mathfrak{X}$ be a locally compact separable topological space. A *point configuration* X in $\mathfrak{X}$ is a locally finite collection of points of the space $\mathfrak{X}$. Any such point configuration is either finite or infinite. For our purposes it suffices to assume that the points of X are always pairwise distinct. The set of all point configurations in $\mathfrak{X}$ will be denoted as $\mathrm{Conf}(\mathfrak{X})$.

A relatively compact Borel subset $A \subset \mathfrak{X}$ is called *a window*. For a window A and $X \in \mathrm{Conf}(\mathfrak{X})$, set $N_A(X) = |A \cap X|$ (number of points of X in the window). Thus, N_A can be viewed as a function on $\mathrm{Conf}(\mathfrak{X})$. We equip $\mathrm{Conf}(\mathfrak{X})$ with the Borel structure generated by functions N_A for all windows A. That is, all these functions N_A are assumed measurable.

A *random point process* on $\mathfrak{X}$ is a probability measure on $\mathrm{Conf}(\mathfrak{X})$.

For algebraic purposes it is always enough to assume that the space $\mathfrak{X}$ is finite, and then $\mathrm{Conf}(\mathfrak{X}) = 2^{\mathfrak{X}}$ is simply the space of all subsets of $\mathfrak{X}$. The same space of configurations $\mathrm{Conf}(\mathfrak{X}) = 2^{\mathfrak{X}}$ may be taken for countable discrete $\mathfrak{X}$, then $\mathrm{Conf}(\mathfrak{X})$ is compact.

7.2 Correlation measures and correlation functions

Given a random point process, one can usually define a sequence $\{\rho_n\}_{n=1}^{\infty}$, where ρ_n is a symmetric measure on $\mathfrak{X}^n$ called the nth *correlation measure*. Under mild conditions on the point process, the correlation measures exist and determine the process uniquely, cf. [Len73].

The correlation measures are characterized by the following property: For any $n \geq 1$ and a compactly supported bounded Borel function f on $\mathfrak{X}^n$ one has

$$\int_{\mathfrak{X}^n} f\rho_n = \left\langle \sum_{x_{i_1},\dots,x_{i_n}\in X} f(x_{i_1},\dots,x_{i_n}) \right\rangle_{X\in \mathrm{Conf}(\mathfrak{X})} \tag{7.1}$$

where the sum on the right is taken over all n-tuples of pairwise distinct points of the random point configuration X.

Often one has a natural measure μ on $\mathfrak{X}$ (called the *reference measure*) such that the correlation measures have densities with respect to $\mu^{\otimes n}$, $n = 1, 2, \dots$. Then the density of ρ_n is called the nth *correlation function* and it is usually denoted by the same symbol "ρ_n".

If $\mathfrak{X} \subset \mathbb{R}$ and μ is absolutely continuous with respect to the Lebesgue measure, then the probabilistic meaning of the nth correlation function is that of the density of probability to find an eigenvalue in each of the infinitesimal intervals around points $x_1, x_2, \dots x_n$:

$$\rho_n(x_1, x_2, \dots x_n)\mu(dx_1)\cdots\mu(dx_n)$$
$$= \mathbb{P}\{\text{there is a particle in each interval } \ (x_i, x_i + dx_i)\}.$$

On the other hand, if μ is supported by a discrete set of points, then

$$\rho_n(x_1, x_2, \dots x_n)\mu(x_1)\cdots\mu(x_n)$$
$$= \mathbb{P}\{\text{there is a particle at each of the points } \ x_i\}.$$

7.3 Determinantal processes

Assume that we are given a point process $\mathcal{P}$ and a reference measure such that all correlation functions exist. The process $\mathcal{P}$ is called *determinantal* if there exists a function $K : \mathfrak{X} \times \mathfrak{X} \to \mathbb{C}$ such that

$$\rho_n(x_1, \dots, x_n) = \det[K(x_i, x_j)]_{i,j=1}^n, \qquad n = 1, 2, \dots . \tag{7.2}$$

The function K is called a *correlation kernel* of $\mathcal{P}$.

The determinantal form of the correlation functions (7.2) implies that many natural observables for $\mathcal{P}$ can be expressed via the kernel K. The determinantal structure of a point process allows to study its asymptotics in various regimes, which lead to many interesting results.

- Kernel is not unique. For example, $\dfrac{f(x)}{f(y)}K(x,y)$ is also a kernel for the same process.

Exercise 7.1. Prove this proposition

- **(complementation principle)**

  If the space $\mathfrak{X}$ is discrete, there is an important operation on point processes which preserves the determinantal structure.

For any subset $\mathfrak{Y} \subset \mathfrak{X}$ one can define an involution on point configurations $X \subset \mathfrak{X}$ by $X \mapsto X \triangle \mathfrak{Y}$ (here $\triangle$ is the symbol of symmetric difference). This map leaves intact the "particles" of X outside of $\mathfrak{Y}$, and inside $\mathfrak{Y}$ it picks the "holes" (points of $\mathfrak{Y}$ free of particles). This involution is called the *particle-hole involution* on $\mathfrak{Y}$.

Given an arbitrary discrete state space $\mathfrak{X}$, a kernel $K(x, y)$ on $\mathfrak{X} \times \mathfrak{X}$, and a subset $\mathfrak{Y}$ of $\mathfrak{X}$, consider another kernel,

$$K^\circ(x,y) = \begin{cases} K(x,y), & x \notin \mathfrak{Y}, \\ \delta_{xy} - K(x,y), & x \in \mathfrak{Y}, \end{cases}$$

where δ_{xy} is the Kronecker symbol.

Proposition 7.1. *Let $\mathcal{P}$ be a determinantal point process with correlation kernel K on a discrete space $\mathfrak{X}$, and let $\mathcal{P}^\circ$ be the image of $\mathcal{P}$ under the particle–hole involution on $\mathfrak{Y} \subset \mathfrak{X}$. Then $\mathcal{P}^\circ$ is also a determinantal point process with correlation kernel $K^\circ(x, y)$ as defined above:*

$$\rho_m(x_1, \dots, x_m \mid \mathcal{P}^\circ) = \det\left[K^\circ(x_i, x_j)\right]_{i,j=1}^m, \quad m = 1, 2, \dots .$$

- **(processes with Hermitean symmetric kernel)**

  There exists a characterization of point processes with Hermitean symmetric kernels [Sos00]. Interpret K as an integral operator in $L^2(\mathfrak{X}, \mu)$ (where μ is the reference measure with respect to which the kernel is defined):

$$(Kf)(x) := \int_{\mathfrak{X}} K(x, y) f(y) \mu(dy).$$

  The fact that the correlation functions are nonnegative means that $K \geq 0$, i.e., the operator K is nonnegative definite (all its diagonal minors which are the correlation functions are nonnegative). In fact, the operator $1 - K$ (here and below 1 is the identity operator) is also nonnegative definite.

  > Hermitian locally trace class[1] operator K in $L^2(\mathfrak{X}, \mu)$ defines a determinantal random point process if and only $0 \leq K$ and $0 \leq 1 - K$. If the corresponding random point process exists then it is unique.

A large subclass of processes with Hermitean symmetric kernel is formed by orthogonal polynomial ensembles whose kernels are finite-dimensional orthogonal projections. In fact, any Hermitean symmetric operator K with $0 \leq K \leq 1$ may be approximated by finite-dimensional projections.

7.4 Biorthogonal ensembles

Definition 7.1. Consider a state space $\mathfrak{X}$ with a reference measure μ. An N-point **biorthogonal ensemble** on $\mathfrak{X}$ is an N-point point process (= probability measure on $\mathfrak{X}^N$) of the form

$$P_N(\mathrm{d}x_1 \dots \mathrm{d}x_n) = c_N \det(\varphi_i(x_j))_{i,j=1}^N \det(\psi_i(x_j))_{i,j=1}^N \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n)$$

where c_N is some constant and φ_i, ψ_i are arbitrary functions on $\mathfrak{X}$.

Proposition 7.2. *Any biorthogonal ensemble is a determinantal point process. Its correlation kernel has the form*

$$K(x,y) = \sum_{i,j=1}^N (G^{-T})_{i,j} \varphi_i(x) \psi_j(y)$$

where $G_{i,j} = \int_{\mathfrak{X}} \varphi_i(x)\psi_j(x)\mu(\mathrm{d}x)$ is the Gram matrix. The matrix G^{-T} is the inverse transposed matrix.

This was first considered by F. Dyson in 1962.

Proof. First let us obtain an expression for the normalizing constant.

$$\begin{aligned}
\int_{\mathfrak{X}^N} \det(\varphi_i(x_j))_{i,j=1}^N \det(\psi_i(x_j))_{i,j=1}^N \mathrm{d}x &= \int_{\mathfrak{X}^N} \sum_{\sigma,\tau \in S_N} \operatorname{sign}(\sigma\tau) \prod_{i=1}^N \varphi_{\sigma(i)}(x_i)\psi_{\tau(i)}(x_i)\mathrm{d}x \\
&= \sum_{\sigma,\tau \in S_N} \prod_{i=1}^N G_{\sigma(i),\tau(i)} \\
&= N! \sum_{\rho \in S_N} \operatorname{sign}(\rho) \prod_{i=1}^N G_{i,\rho(i)} = N! \det(G_{i,j})_{i,j=1}^N.
\end{aligned}$$

This implies that G is invertible, and we will have $c_N = (N! \det(G))^{-1}$. Now we have[1]

$$\rho_n(x_1, \dots, x_n) = \frac{N!}{(N-n)!} \int_{(x_{n+1}, \dots, x_N) \in \mathfrak{X}^{N-n}} \det(\varphi_i(x_j))_{i,j=1}^N \det(\psi_i(x_j))_{i,j=1}^N \mathrm{d}x \frac{1}{N! \det(G)}.$$

Take matrices A and B such that $AGB^T = 1$. Set

$$\Phi_k = \sum_{\ell=1}^N A_{k\ell} \varphi_\ell, \quad \Psi_k = \sum_{\ell=1}^N B_{k\ell} \psi_\ell.$$

The result is that

$$P_N(\mathrm{d}x_1 \cdots \mathrm{d}x_N) = \frac{1}{N!} \det(\Phi_i(x_j))_{i,j=1}^N \det(\Psi_i(x_j))_{i,j=1}^N,$$

Exercise 7.2. ⬅️ Prove this identity

and

$$\begin{aligned}
\rho_n(x_1, \dots, x_n) &= \frac{1}{(N-n)!} \int_{x_{n+1}, \dots, x_N} \det(\varphi_i(x_j))_{i,j=1}^N \det(\psi_i(x_j))_{i,j=1}^N \mathrm{d}x \\
&= \frac{1}{(N-n)!} \int_{x_{n+1}, \dots, x_N} \sum_{\sigma, \tau \in S_N} \operatorname{sign}(\sigma\tau) \prod_{i=1}^N \Phi_{\sigma(i)}(x_i) \Psi_{\tau(i)}(x_i) \mathrm{d}x_i \\
&= \frac{1}{(N-n)!} \int_{x_{n+1}, \dots, x_N} \sum_{\substack{\sigma, \tau \in S_N \\ \sigma(k) = \tau(k) \\ \text{for } k = n+1, \dots, N}} \operatorname{sign}(\sigma\tau) \prod_{i=1}^n \Phi_{\sigma(i)}(x_i) \Psi_{\tau(i)}(x_i) \mathrm{d}x_i \\
&= \sum_{1 \le j_1 < \cdots < j_n \le N} \det \Phi^{j_1, \dots, j_n} \det \Psi^{j_1, \dots, j_n},
\end{aligned}$$

where $\Phi^{j_1, \dots, j_n}$ is the submatrix of $[\varphi_i(x_j)]$ with columns $j_1, \dots, j_n$ and rows $i = 1, \dots, n$, and smae for $\Psi^{j_1, \dots, j_n}$.

Now using the Cauchy–Binet theorem, this last expression becomes

$$\det(\Phi\Psi^T)_{i,j=1}^n = \det\left(\sum_{k=1}^N \Phi_k(x_i)\Psi_k(j)\right)_{i,j=1}^N.$$

Now define $K(x_i, x_j) = \sum_{k=1}^N \Phi_k(x_i)\Psi_k(x_j)$. We can write this as

$$\begin{aligned}\sum_{k=1}^N \Phi_k(x)\Psi_k(y) &= \sum_{k,\ell,m} A_{k\ell}\varphi_\ell(x)B_{km}\psi_m(y) \\ &= \sum_{\ell,m} \varphi_\ell(x)\psi_m(y)\sum_k A_{k\ell}B_{km} = A^T B = G^{-T}. \qquad \square\end{aligned}$$

Exercise 7.3. Formulate and prove the Cauchy-Binet theorem.

7.5 Orthogonal polynomial ensembles

Take $\mathfrak{X} = \mathbb{R}$. Let $w(\mathrm{d}x)$ be a positive measure on $\mathbb{R}$ with finite moments, i.e., $\int_{\mathbb{R}} |x|^k w(\mathrm{d}x) < \infty$ for all $k \geq 0$.

Example 7.4. The natural map $\mathbb{C}[x]_{\leq N} \to L^2(\mathbb{R}, w(\mathrm{d}x))$ is an embedding if and only if $\#\mathrm{supp}(w) > N + 1$.

We will assume that $\#\mathrm{supp}(w) = \infty$.
Notation: $V_N(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq N}(x_i - x_j) = \det(x_i^{N-j})_{i,j=1}^N$.

Definition 7.5. The **N-particle orthogonal polynomial ensemble** with weight w is the N-point random point process with joint probability density

$$P_N(\mathrm{d}x_1 \cdots \mathrm{d}x_N) = \mathrm{const}_N \cdot (V_N(x_1, \ldots, x_N))^2 \prod_{i=1}^N w(\mathrm{d}x_i).$$

Example 7.6. The most well-known example is the **Gaussian unitary ensemble** (GUE(N)). In this case the space is $\{H \in \mathrm{Mat}_N(\mathbb{C}) \mid H = H^*\}$ with measure $ce^{-\mathrm{Tr}(H^2)}$ for some constant c. The eigenvalues of H form an N-point orthogonal polynomial ensemble with $w(\mathrm{d}x) = e^{-x^2}\mathrm{d}x$.

An orthogonal polynomial ensemble is a biorthogonal ensemble with $\varphi_i(x) = \psi_i(x) = x^{i-1}\sqrt{w(x)}$ where w denotes the density function of the measure w. A kernel K defines a linear operator $K\colon L^2 \to L^2$ by $(Kf)(x) = \int K(x,y)f(y)\mathrm{d}y$.

Proposition 7.7. *The correlation kernel $K(x,y)$ is the kernel of the orthogonal projection operator onto* $\mathrm{span}(\sqrt{w(x)}, x\sqrt{w(x)}, \ldots, x^{N-1}\sqrt{w(x)})$ *in* $L^2(\mathbb{R}, w)$.

Exercise 7.4. Prove Proposition 7.7

SCHUR MEASURES

A (particular case of a) *Schur measure* is a probability distribution on partitions $\lambda = (\lambda_1 \geq \ldots \geq \lambda_N \geq 0)$ depending on parameters $x_i, y_j \in [0,1)$ whose probability weights are

$$P(\lambda_1, \ldots, \lambda_N) = c_N s_\lambda(x_1, \ldots, x_N) s_\lambda(y_1, \ldots, y_N).$$

It is a biorthogonal ensemble. Denote $\ell_j := \lambda_j + N - j$, then

$$P(\lambda_1, \ldots, \lambda_N) = \tilde{c}_N \det[x_i^{\ell_j}] \det[y_i^{\ell_j}],$$

where we have incorporated the Vandermonde determinants into the constant.

This is a biorthogonal ensemble on $\mathbb{Z}_{\geq 0}$ with $\pi_i(\ell) = x_i^\ell$, $\psi_j(\ell) = y_j^\ell$. The matrix G has the form

$$G_{ij} = \sum_{\ell=0}^{\infty} (x_i y_j)^\ell = \frac{1}{1 - x_i y_j}.$$

Inverting this matrix would produce a correlation kernel for the Schur measure.

Exercise 7.5. 1. Find the inverse of the $N \times N$ matrix $G_{ij} = 1/(1 - x_i y_j)$. Hint: use the cofactor formula for the inverse matrix, and note that the cofactors are matrices of the same form as G_{ij}.

2. Use the inverse of G_{ij} to find the correlation kernel for the Schur measures. (Hint: See the links at the end of the chapter where you can check your answer.)

Definition 7.8. A **system of orthogonal polynomials** on $\mathbb{R}$ with weight w is a sequence $\{p_n(x)\}_{n\geq 0}$ with $p_n \in \mathbb{C}[x]$ and $\deg p_n = n$, such that $p_n \perp \mathbb{C}[x]_{\leq n-1}$ in $L^2(\mathbb{R}, w)$, i.e.,

$$\int_{\mathbb{R}} p_n(x)p_m(x)w(\mathrm{d}x) = \|p_n\|^2\delta_{m,n}.$$

Note that a system of orthogonal polynomials is an orthogonal basis in $\mathbb{C}[x]$ with inner product $\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)w(\mathrm{d}x)$, so one can construct such systems using the Gram–Schmidt orthogonalization algorithm. Note that the degree constraint $\deg p_n = n$ uniquely determines the p_n up to a constant.

Standard notation: Let k_n be the leading coefficient of p_n and set $h_n = \|p_n\|^2_{L^2(\mathbb{R},w)}$.

Exercise 7.6. Show that the system of orthogonal polynomials is indeed defined uniquely.

Proposition 7.9. *Let $\{p_n\}$ be the sequence of monic orthogonal polynomials with weight w. Then the correlation kernel of the N-point orthogonal polynomial ensemble has the form*

$$K_N(x, y) = \sum_{j=0}^{N-1} \frac{p_j(x)p_j(y)}{h_j}$$

with respect to the reference measure $w(\mathrm{d}x)$ on $\mathbb{R}$.

Proof. Let π_{j-1} be monic polynomials of degree $j-1$. Then

$$\begin{aligned} P_N(\mathrm{d}x_1 \cdots \mathrm{d}x_N) &= c\det(x_i^{j-1})\det(x_i^{j-1})\prod w(\mathrm{d}x_i) \\ &= \det(\pi_{j-1}(x_i))\det(\pi_{j-1}(x_i))\prod w(\mathrm{d}x_i), \end{aligned}$$

where the last equality is via row operations. Then $G^{-T} = \mathrm{diag}(h_0^{-1}, \ldots, h_{N-1}^{-1})$. □

Proposition 7.10 (Christoffel–Darboux)**.**

$$\sum_{j=0}^{N-1} \frac{p_j(x)p_j(y)}{h_j} = \frac{p_N(x)p_{N-1}(y) - p_{N-1}(x)p_N(y)}{h_{N-1}(x-y)}.$$

Proof 1. Consider the operator given by multiplication by x in $\mathbb{R}[x]$. This operator is self-adjoint:

$$(x \cdot f, g)_{L^2(\mathbb{R},\omega)} = \int xf(x)g(x)w(\mathrm{d}x) = (f, x \cdot g)_{L^2(\mathbb{R},\omega)}.$$

The matrix of a self-adjoint operator in any orthonormal basis is symmetric. For our basis, we will use $\{p_n/\|p_n\|\}_{n\geq 0}$. By degree considerations, this matrix must be 0 below the subdiagonal. By symmetry, it must be 0 above the superdiagonal. Since the p_n are monic, we have

$$xp_n = A_{n,n+1}p_{n+1} + A_{n,n}p_n + A_{n,n-1}p_{n-1}$$

where $A_{n,n+1} = 1$. Now multiply the desired identity by $(x-y)$ and use this recurrence relation and symmetry of the matrix to finish (the left-hand side is a telescoping sum). □

Exercise 7.7. Finish this first proof the Christoffel-Darboux formula.

Proof 2. Consider the average ($x_1, \dots, x_N$ distributed as the orthogonal polynomial ensemble). Then

$$\mathbb{E}(\prod_{i=1}^{N}(u-x_i)(v-x_i)) = \text{constant}\cdot\int\prod_{i=1}^{N}(u-x_i)(v-x_i)\prod_{i<j}(x_i-x_j)^2 w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N).$$

(If the x_i are eigenvalues of a random matrix X, then $\prod_i(u-x_i) = \det(u-X)$.)

In the simpler case, we have

$$\begin{aligned}
\mathbb{E}(\prod_{i=1}^{N}(u-x_i)) &= \text{constant}\cdot\int\prod_{i=1}^{N}(u-x_i)\prod_{i<j}(x_i-x_j)^2 w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N)\\
&= \text{constant}\cdot\int V_{N+1}(u,x_1,\dots,x_N)V_N(x_1,\dots,x_N)w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N)\\
&= \text{constant}\cdot\int \det\begin{bmatrix} p_N(u) & p_N(x_1) & \cdots & p_N(x_N)\\ p_{N-1}(u^{N-1}) & p_{N-1}(x_1^{N-1}) & \cdots & p_{N-1}(x_N^{N-1})\\ \vdots & & & \\ 1 & 1 & \cdots & 1\end{bmatrix}\cdot\\
&\qquad \det\begin{bmatrix} p_{N-1}(x_1^{N-1}) & \cdots & p_{N-1}(x_N^{N-1})\\ \vdots & & \\ 1 & \cdots & 1\end{bmatrix} w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N)\\
&= \text{constant}\cdot p_N(u)\int V_N(x_1,\dots,x_N)^2 w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N) = p_N(u).
\end{aligned}$$

(In the third equality, we have done row operations.)

Using this, we get

$$\begin{aligned}
\mathbb{E}(\prod_{i=1}^{N}(u-x_i)(v-x_i)) &= \text{constant}\cdot\int V_{N+1}(u,x_1,\dots,x_N)V_{N+1}(v,x_1,\dots,x_N)w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N)\\
&= \frac{\text{constant}}{u-v}\int V_{N+2}(u,v,x)V_N(x)w(\mathrm{d}x_1)\cdots w(\mathrm{d}x_N)
\end{aligned}$$

The first integral expression can be simplified as $\sum_{k=0}^{N} c_k p_k(u)p_k(v)$ for some coefficients c_k. The second integral expression can be simplified as $\frac{1}{u-v}\det\begin{bmatrix} p_{N+1}(u) & p_{N+1}(v)\\ p_N(u) & p_N(v)\end{bmatrix}$.

□

SCHUR-WEYL AND GUE: POLYNOMIALS

We will discuss basic properties of Hermite and Charlier polynomials, and derive their properties which will be useful for asymptotic analysis of the Christoffel-Darboux kernel.

We will talk about monic orthogonal polynomials with Poisson and Gaussian weights:

$$w_P(x) := e^{-t}\frac{t^x}{x!}, \quad , x \in \mathbb{Z}_{\geq 0},$$
$$w_G(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad x \in \mathbb{R}.$$

The polynomials will be denoted by $C_n(x)$ and $H_n(x)$, respectively.

Exercise 7.8. Prove the hypergeometric formulas ➡️ for the Charlier and Hermite polynomials. You may use all the statements proven later in this chapter.

If the weight is "nice" (like a named probability distribution occurring in a first undergraduate probability textbook), then the corresponding family of orthogonal polynomials is also known. Moreover, it is likely a member of the "Askey scheme" of hypergeometric orthogonal polynomials. Basically, these are polynomials which can be expressed through the *hypergeometric function*

$${}_rF_s[a_1, \dots, a_r; b_1, \dots, b_s; z] := \sum_{n=0}^{\infty} \frac{(a_1)_n (a_2)_n \dots (a_r)_n}{n!(b_1)_n \dots (b_s)_n} z^n,$$

where a_i, b_j, and z are parameters, aand we use the *Pochhammer symbols* $(a)_k := a(a+1)\dots(a+k-1)$.

In particular, we have for the monic Hermite and Charlier polynomials:

$$C_n(x) = (-t)^n {}_2F_0\left(-n, -x; -; -\frac{1}{t}\right),$$
$$H_n(x) = x^n {}_2F_0\left(-n/2, -(n-1)/2; -; -\frac{2}{x^2}\right).$$

RODRIGUES-TYPE FORMULAS AND NORMS

Proposition. We have

$$C_n(x) = (-t)^n \frac{x!}{t^x} \nabla^n \left(\frac{t^x}{x!} \right),$$
$$H_n(x) = (-1)^n e^{x^2/2} \left(\frac{d}{dx} \right)^n e^{-x^2/2},$$

where $\nabla f(x) = f(x) - f(x-1)$ is the discrete derivative.

Exercise 7.9. Prove this proposition.

Hint: it is enough to show that the right-hand sides are monic and orthogonal with respect to the corresponding weight. To show orthogonality, use the summation/integration by parts.

Proposition. We have

$$\sum_{x=0}^{\infty} e^{-t} \frac{t^x}{x!} (C_n(x))^2 = t^n n!,$$
$$\int_{-\infty}^{\infty} (H_n(x))^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = n!.$$

Exercise 7.10. Prove this proposition. Hint: again use integration/summation by parts.

INTEGRAL REPRESENTATIONS FOR HERMITE

Proposition. We have

$$\sum_{n=0}^{\infty} \frac{H_n(x)}{n!}(-z)^n = \exp\left\{-zx - \frac{z^2}{2}\right\}.$$

Proof. This is a statement about the Taylor expansion of $f = e^{-zx-z^2/2}$ at $z = 0$. We have $e^{-x^2/2}f = e^{-(x+z)^2/2}$, and all derivatives of this function at $z = 0$ are just the same as derivatives of $e^{-x^2/2}$ at x. These derivatives are readily connected to the Hermite polynomials. Therefore,

$$e^{-(x+z)^2/2} = \sum_{n=0}^{\infty} \frac{z^n}{n!}\left(\frac{d}{dx}\right)^n e^{-x^2/2},$$

which implies the result. □

By Cauchy's integral theorem, we thus can write

$$H_n(x) = \frac{(-1)^n n!}{2\pi i}\oint_0 e^{-zx-z^2/2}\frac{dz}{z^{n+1}},$$

where the integration is over a small contour around 0.

Another integral representation for the Hermite polynomials is obtained using Fourier transform. We start with the known fact that $e^{-z^2/2}$ is essentially invariant under the Fourier transform. That is,

$$e^{-z^2/2} = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-x^2/2+ixz}dx.$$

Differentiating n times in z and multiplying by the prefactors, we get

$$H_n(z) = (-1)^n e^{z^2/2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(ix)^n e^{-x^2/2+ixz}dx.$$

Finally, let us change the variables to $x = u/i$, so that

$$H_n(z) = \frac{(-1)^n}{i\sqrt{2\pi}}\int_{-i\infty}^{i\infty} e^{u^2/2+zu+z^2/2}u^n du.$$

INTEGRAL REPRESENTATIONS FOR CHARLIER

To get integral representations for Charlier, we employ Fourier series instead of Fourier transform. The Fourier series associates to a function $f(\xi)$, $\xi \in [0,1]$, a family of coefficients

$$\hat{f}(n) = \int_0^1 f(\xi) e^{2\pi i \xi n} d\xi, \qquad n \in \mathbb{Z}_{\geq 0},$$

so that the Fourier series

$$\sum_{n=0}^{\infty} \hat{f}(n) e^{-2\pi i \xi n}$$

converges to $f(\xi)$ under some conditions (also, there are different notions of convergence of this series).

We use Fourier series because the function f behaves well when we apply discrete derivatives ∇ to $\hat{f}$. Namely,

$$\nabla \hat{f}(n) = \int_0^1 f(\xi)(1 - e^{-2\pi i \xi}) e^{2\pi i \xi n} d\xi,$$

so the function f is replaced by $f(\xi)(1 - e^{-2\pi i \xi})$.

Exercise 7.11. Find a function whose Fourier coefficients are $\hat{f}(n) = t^n/n!$.

Exercise 7.12. Using the previous exercise, show that the generating function for the Charlier polynomials is $\sum_{n=0}^{\infty} z^n \frac{C_n(x)}{n!} = e^{-zt}(1+z)^x$.

INTEGRAL REPRESENTATIONS FOR CHARLIER

The generating fucntion for the Charlier polynomials implies that

$$C_n(x) = \frac{n!}{2\pi i} \oint_0 e^{-zt}(1+z)^x \frac{dz}{z^{n+1}},$$

where the integration contour is a small counterclockwise circle around 0.

Changing the variables, we have

$$C_n(x) = \frac{n!}{2\pi i} \oint_1 e^{-t(w-1)} w^x \frac{dw}{(w-1)^{n+1}},$$

where now the integration is around 1.

Here we need to use the symmetry ("bispectrality") of the Charlier polynomials, which follows from their hypergeometric representation: $\frac{C_n(x)}{C_x(n)} = (-t)^{n-x}$. This, together with the previous integral formula, leads to the second integral representation for the Charlier polynomials:

$$C_n(x) = (-t)^{n-x} \frac{x!}{2\pi i} \oint_1 e^{-t(w-1)} w^n \frac{dw}{(w-1)^{x+1}}$$

KERNELS

In both Schur-Weyl and GUE ensembles, we can use the integral representations for the orthogonal polynomials to obtain integral formulas for the kernels which are suitable for asymptotic analysis. We use the formula

$K(x,y) = \sum_{n=0}^{N-1} \frac{p_n(x)p_n(y)}{h_n}\sqrt{w(x)w(y)}$ for the kernel. In the sum, we in both cases extend the summation to $n = -\infty, \ldots, N-1$ because one of the integrals vanishes for negative n.

Exercise 7.13. Prove the formula for K_N^{GUE}.

Exercise 7.14. Prove the formula for K_N^{SW}.

Putting this all together, we get the following formulas for the GUE and the Schur-Weyl kernels.

$$K_N^{GUE}(x,y) = \frac{e^{-\frac{x^2}{4}+\frac{y^2}{4}}}{(2\pi i)^2}\oint_0 dz \int_{-i\infty}^{i\infty} du \left(\frac{u}{z}\right)^N \frac{1}{u-z} e^{\frac{u^2}{2}+uy-\frac{z^2}{2}-zx}.$$

$$K_N^{SW}(x,y) = \frac{t^{x/2}}{t^{y/2}}\sqrt{\frac{y!}{x!}}\frac{(-1)^y}{(2\pi i)^2}\oint_0 dz \oint_1 dw\, e^{-t(z+w)} \frac{(1+z)^x}{(w-1)^{y+1}} \frac{(-1)(-w/z)^N}{w+z}.$$

In the first kernel, $x, y \in \mathbb{R}$, while in the second one $x, y \in \mathbb{Z}_{\geq 0}$.

NOTES AND REFERENCES

1. There are several surveys on determinantal point processes, including

   - A. Soshnikov. Determinantal random point fields. Russian Mathematical Surveys, 55(5):923--975, 2000. arXiv:math/0002099 [math.PR]

   - J.B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. Probability Surveys, 3:206--229, 2006. arXiv:math/0503110 [math.PR]

   - A. Borodin. *Determinantal point processes*. In G. Akemann, J. Baik, and P. Di Francesco, editors, Oxford Handbook of Random Matrix Theory. Oxford University Press, 2011. arXiv:0911.1153 [math.PR].

2. A book on orthogonal polynomials which is more modern than the more classical texts is *Classical and Quantum Orthogonal Polynomials in One Variable* by Mourad E.H. Ismail

3. Another valuable source on orthogonal polynomials is the online "encyclopedia" by Roelof Koekoek & René F. Swarttouw, available at *homepage.tudelft.nl/11r49/askey/* (PDF version at arXiv:math/9602214 [math.CA])

4. C. Krattenthaler. Advanced determinant calculus. Séminaire Lotharingien Combin, 42:B42q, 1999. arXiv:math/9902004 [math.CO].

5. Papers on Schur measures:

   - A. Okounkov. Infinite wedge and random partitions. Selecta Math., 7(1):57--81, 2001. arXiv:math/9907127 [math.RT].

   - K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. Ann. Math., 153(1):259--296, 2001. arXiv:math/9906120 [math.CO].

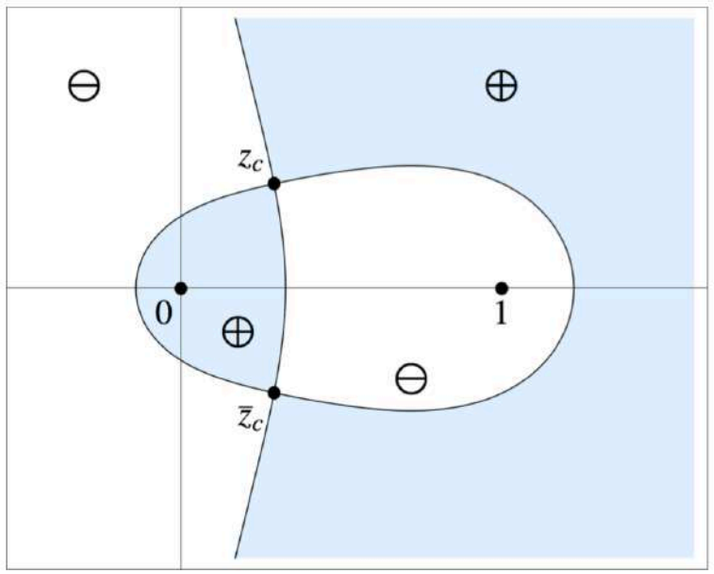
Chapter 8

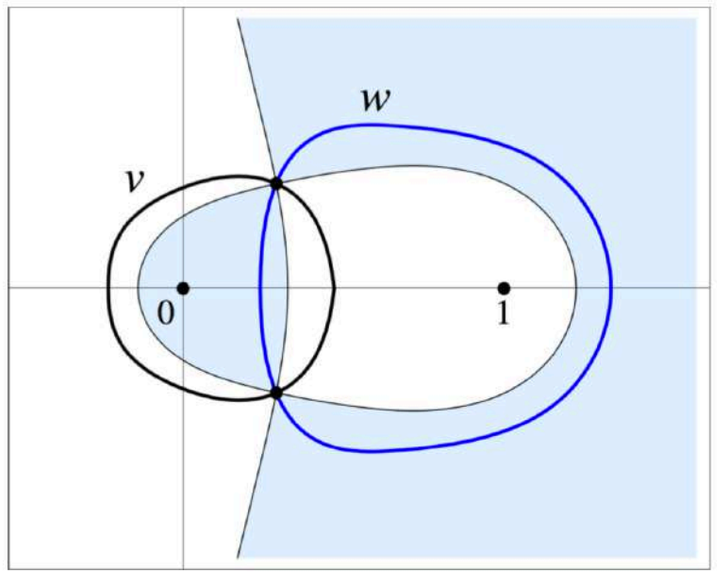
ASYMPTOTICS VIA CONTOUR INTEGRALS

- Steepest descent
- GUE asymptotics
- Schur-Weyl asymptotics
- From correlations to distributions: Fredholm determinants

Here we discuss asymptotics of the GUE spectrum and its discrete analogue (the Schur-Weyl measure) using determinantal point processes. The steepest descent method is the key to our proofs.

In the limit, we discover the universal objects: the continuous and discrete sine processes, and the Airy kernel/point process.





NOTES AND REFERENCES

1. Steepest descent method for determinantal point processes related to the Plancherel measure on partitions is outlined in *"Symmetric functions and random partitions" by Andrei Okounkov*.

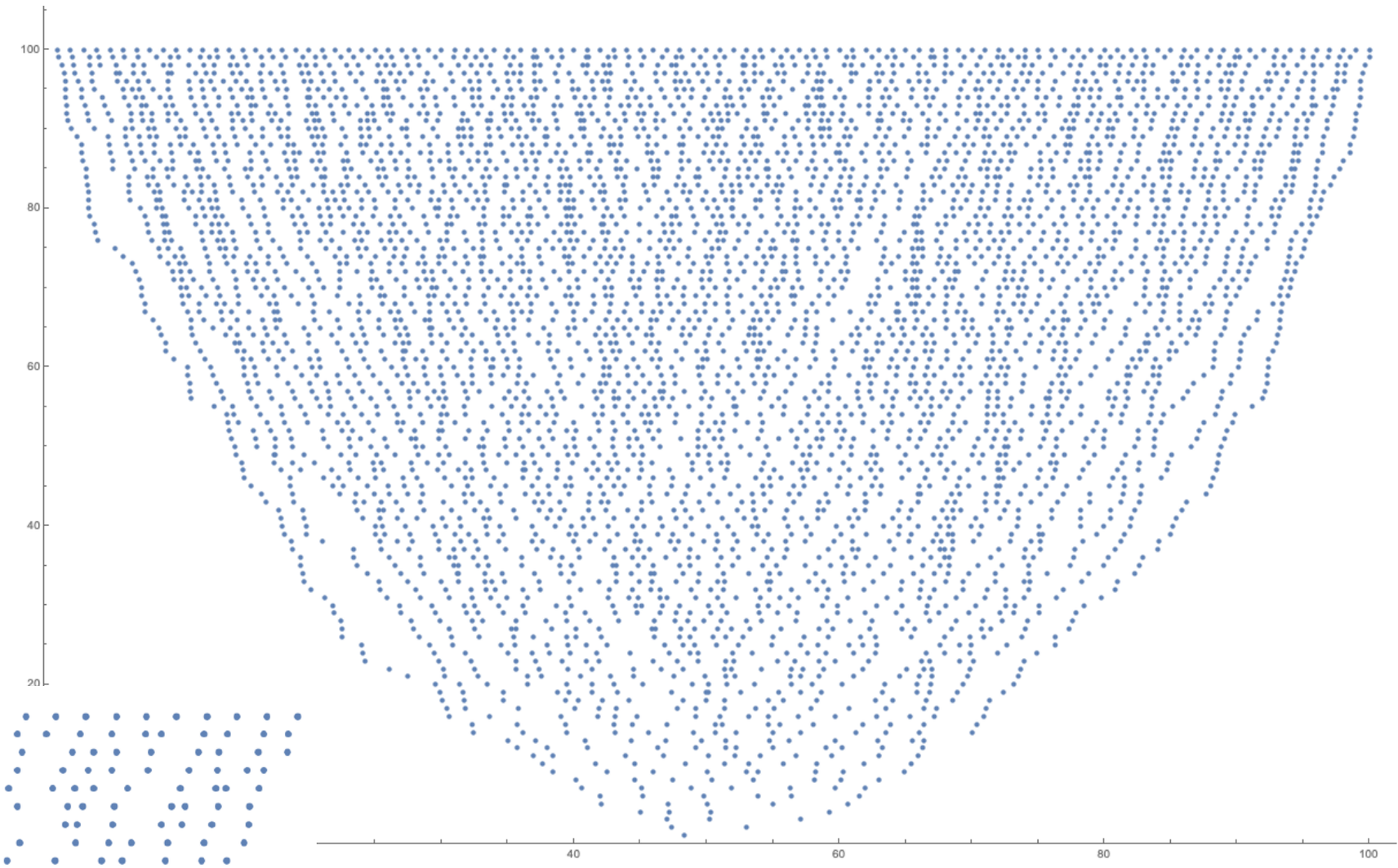
2.

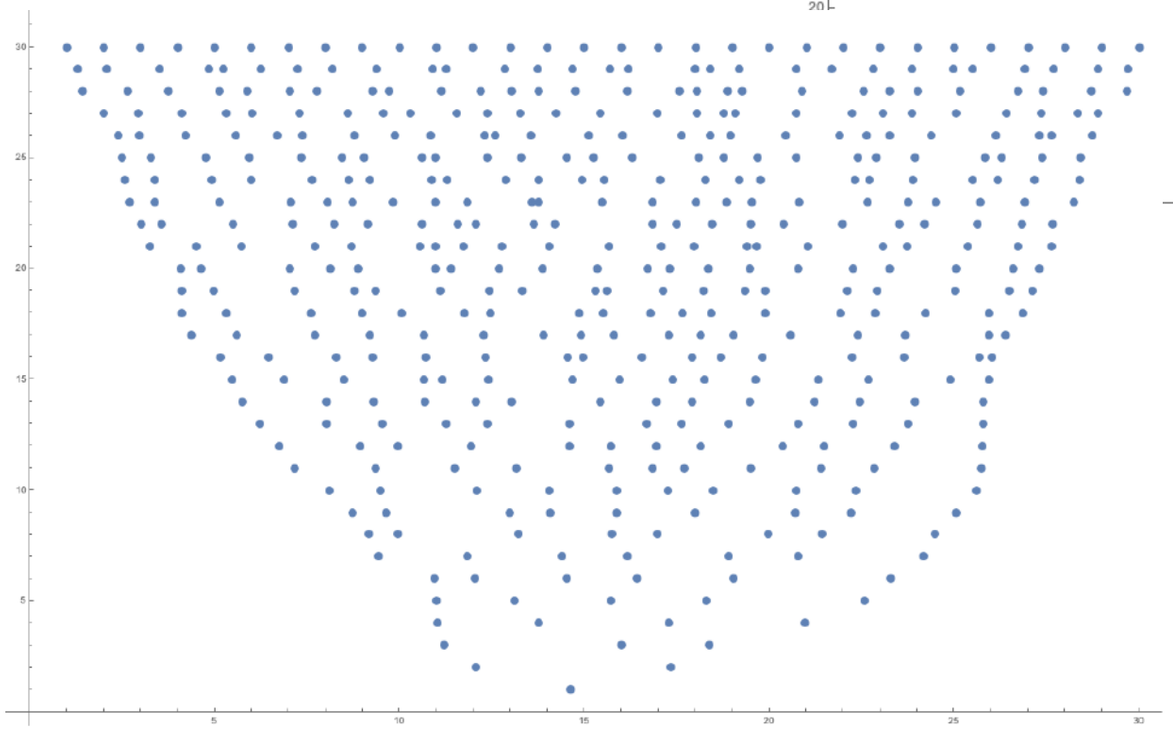
Chapter 9

HCIZ AND HERMITIAN ENSEMBLES

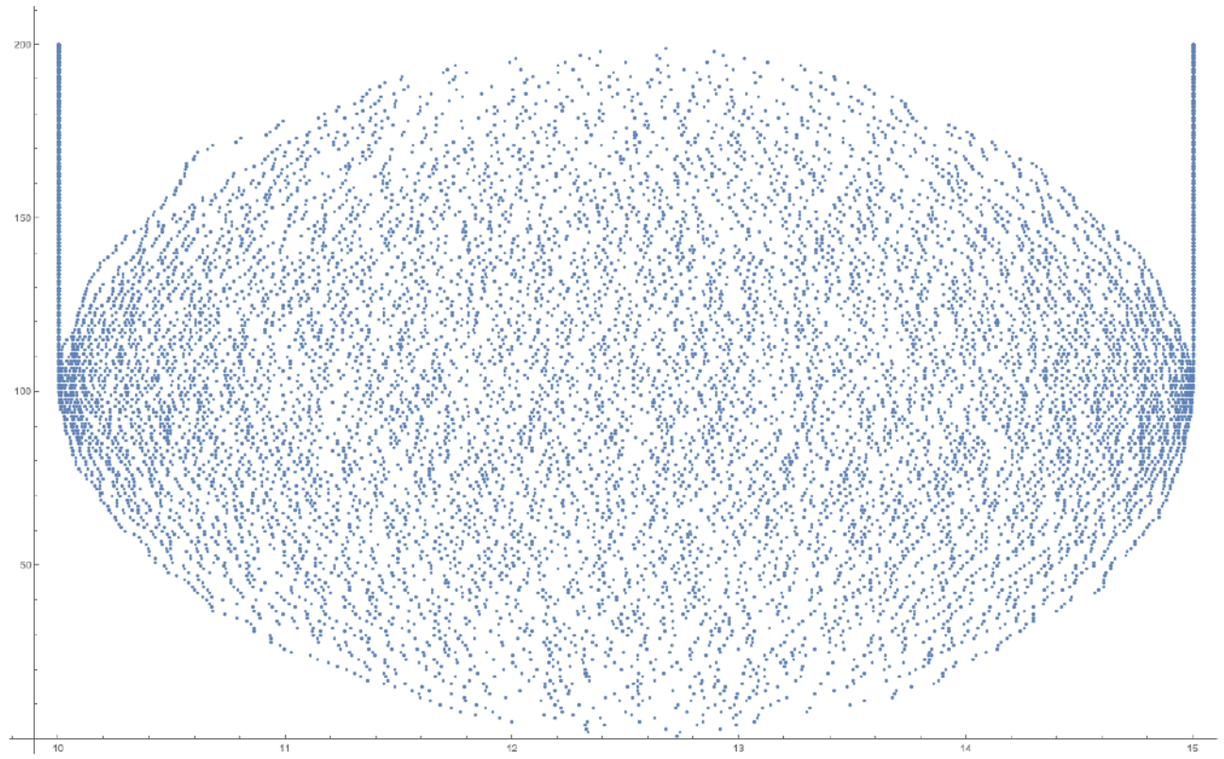
-

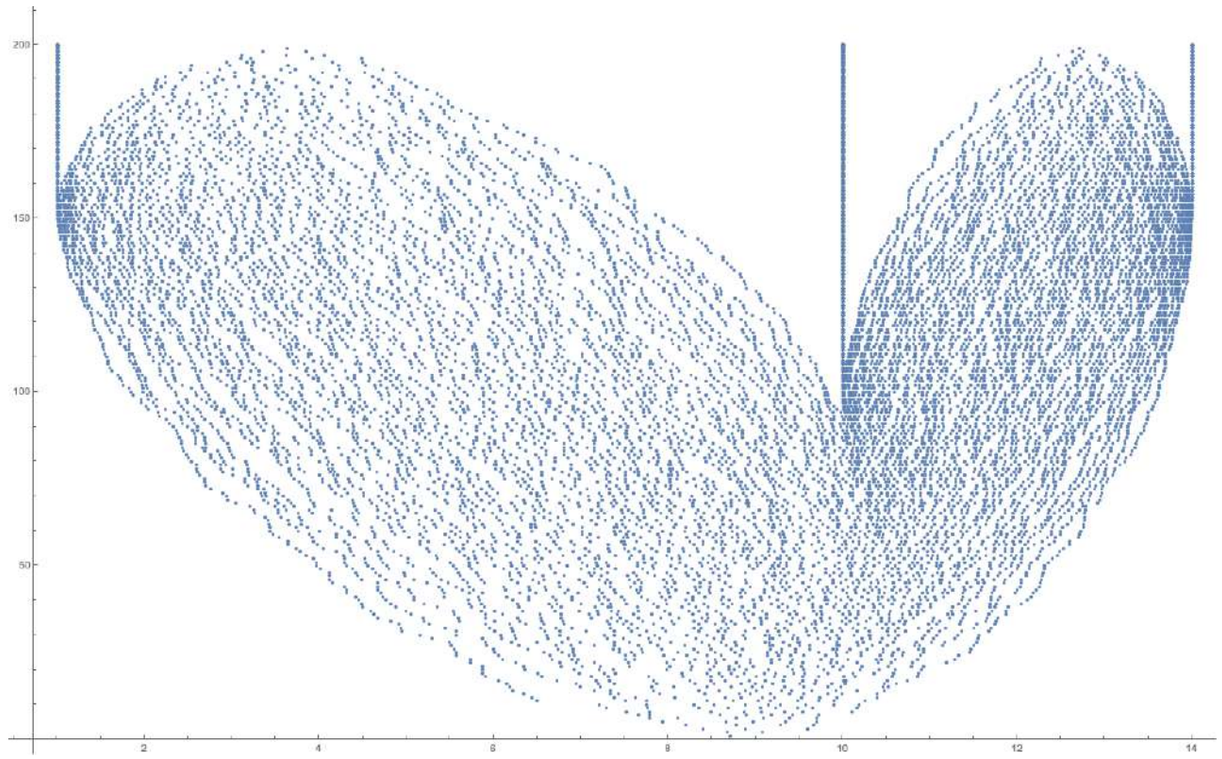
EXAMPLES OF ORBITAL CORNERS PROCESSES

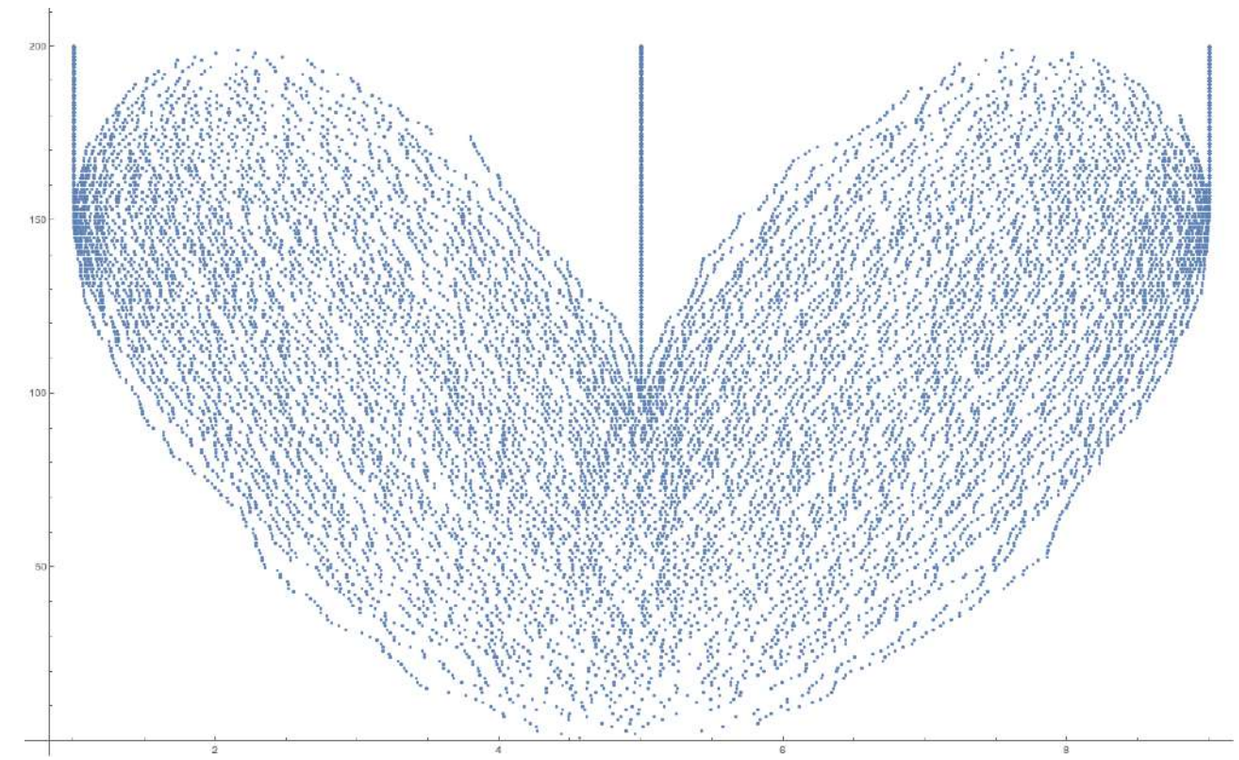


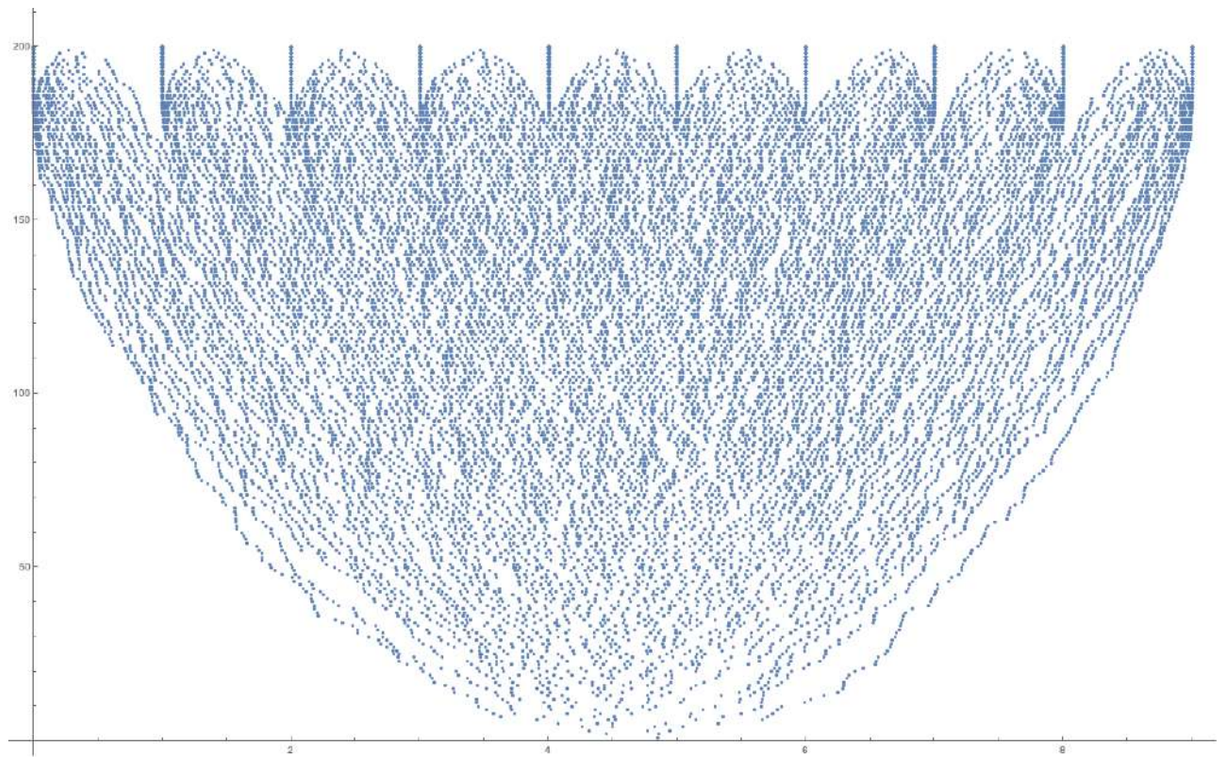


EXAMPLES OF ORBITAL CORNERS PROCESSES

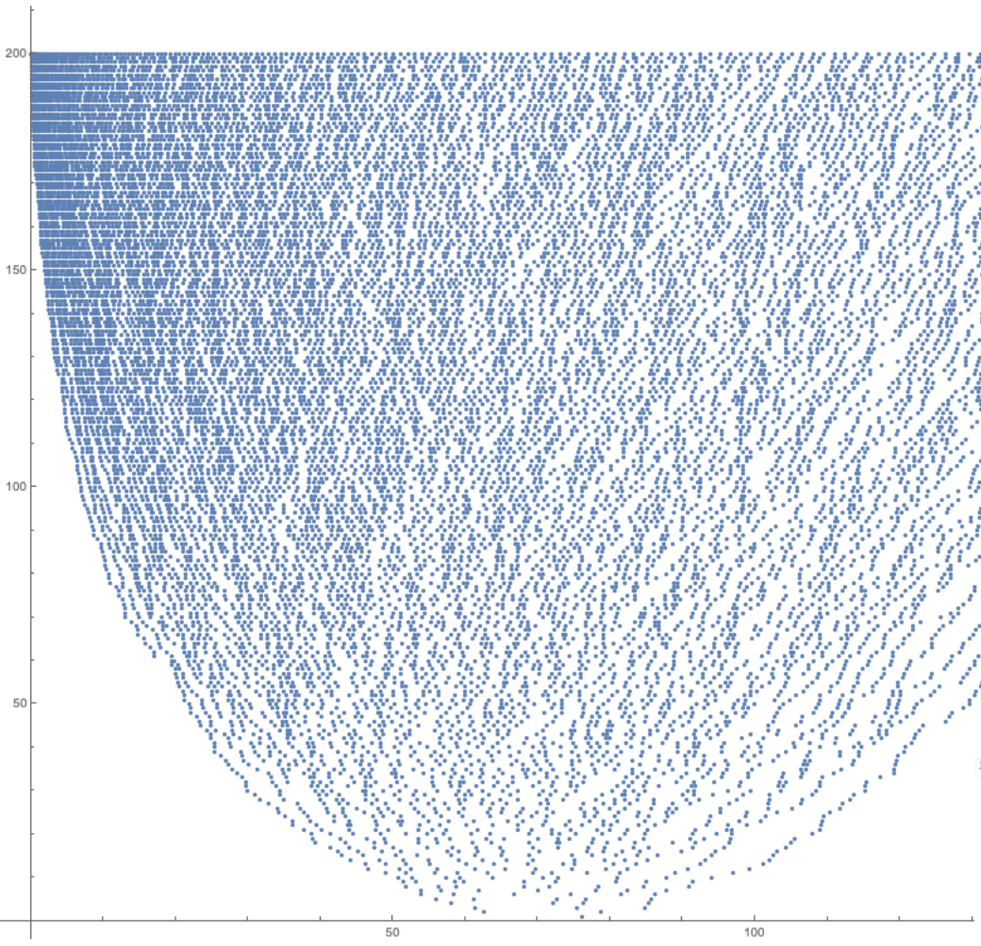


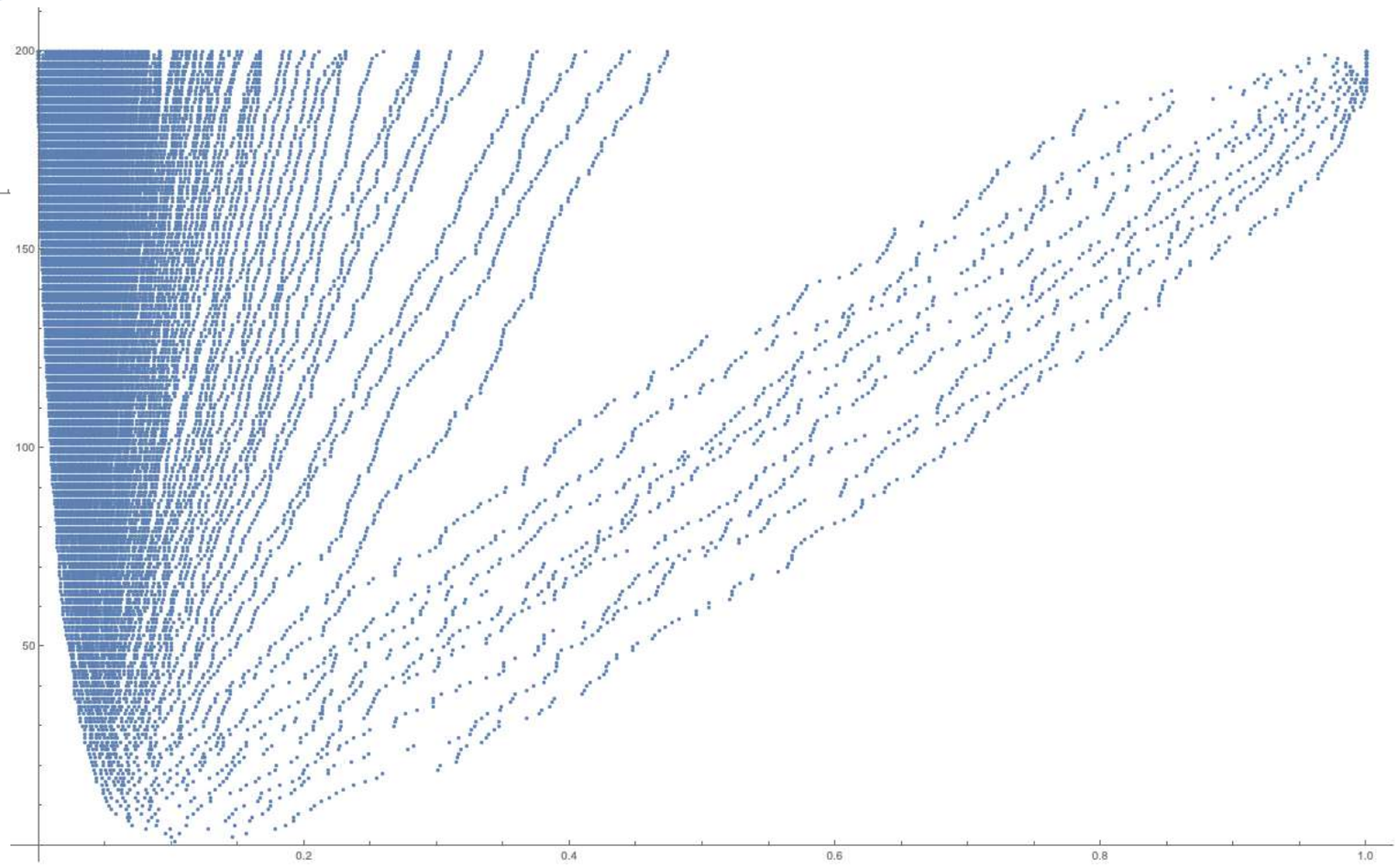






EXAMPLES OF ORBITAL CORNERS PROCESSES





Chapter 10

NONCOLLIDING BROWNIAN MOTIONS

-

Chapter 11

UNIVERSALITY

Chapter 12

MARKOV MAPS