Lectures on Random Matrices (Spring 2025)
Lecture 1: Moments of random variables and random matrices

Leonid Petrov

Monday, January 13, 2025*

Contents

*Course webpage • TeX Source • Updated at 03:34, Wednesday 22nd January, 2025

1 Why study random matrices?

On the history. Random matrix theory (RMT) is a fascinating field that studies properties of matrices with randomly generated entries, focusing (at least initially) on the statistical behavior of their eigenvalues. This theory finds its roots in the domain of nuclear physics through the pioneering work of Wigner, Dyson, and others [Wig55], [Dys62a], [Dys62b], who utilized it to analyze the energy levels of complex quantum systems. Other, earlier roots include statistics [Dix05] and classical Lie groups [Hur97]. Today, RMT has evolved to span a wide array of disciplines, from pure mathematics, including areas such as integrable systems and representation theory, to practical applications in fields like data science and engineering.

Classical groups and Lie theory. Random matrices are deeply connected to *classical Lie groups*, particularly the orthogonal, unitary, and symplectic groups. This connection emerges primarily due to the invariance properties of these groups, such as those derived from the Haar measure. Random matrices significantly impact representation theory, linking to integrals over matrix groups through character expansions. The symmetry classes of random matrix ensembles, like the Gaussian Orthogonal (GOE), Unitary (GUE), and Symplectic (GSE), correspond to respective symmetry groups.

Toolbox. RMT utilizes a broad range of tools ranging across all of mathematics, including probability theory, combinatorics, analysis (classical and modern), algebra, representation theory, and number theory. The theory of random matrices is a rich source of problems and techniques for all of mathematics.

The main content of this course is to explore the toolbox around random matrices, including going into discrete models like dimers and statistical mechanics. Some of this will be included in the lectures, and some other topics will be covered in the reading course component, which is individualized.

Applications. Random matrix theory finds applications across a diverse set of fields. In nuclear physics, random matrix ensembles serve as models for complex quantum Hamiltonians, thereby explaining the statistics of energy levels. In number theory, connections have been drawn between random matrices and the Riemann zeta function, particularly concerning the distribution of zeros on the critical line. Wireless communications benefit from random matrix theory through the analysis of eigenvalue distributions, which helps in understanding channel capacity in multi-antenna (MIMO) systems. In the burgeoning field of machine learning, random weight matrices and their spectra are key to analyzing neural networks and their generalization capabilities. High-dimensional statistics and econometrics also draw on random matrix tools for tasks such as principal component analysis and covariance estimation in large datasets. Additionally, combinatorial random processes exhibit connections to random permutations, random graphs, and partition theory, all through the lens of matrix integrals.

2 Recall Central Limit Theorem

2.1 Central Limit Theorem and examples

We begin by establishing the necessary groundwork for understanding and proving the Central Limit Theorem. The theorem's power lies in its remarkable universality: it applies to a wide variety of probability distributions under mild conditions.

Definition 2.1. A sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to be *independent and identically distributed (iid)* if:

- Each X_i has the same probability distribution as every other X_j, for all i, j.
- The variables are mutually independent, meaning that for any finite subset $\{X_1, X_2, \ldots, X_n\}$, the joint distribution factors as the product of the individual distributions:

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1)\,\mathbb{P}(X_2 \leq x_2) \cdots \mathbb{P}(X_n \leq x_n).$$

Theorem 2.2 (Classical Central Limit Theorem)**.** *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$. Define the normalized sum*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)\,. \tag{2.1}$$

Then, as $n \to \infty$, the distribution of Z_n converges in distribution to a normal random variable with mean 0 and variance σ^2, i.e.,

$$Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Convergence in distribution means

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \mathbb{P}(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\, dt \qquad \text{for all } x \in \mathbb{R}, \tag{2.2}$$

where $Z \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian random variable.

Remark 2.3. For a general random variable instead of $Z \sim \mathcal{N}(0, \sigma^2)$, the convergence in distribution (2.2) holds only for x at which the cumulative distribution function of Z is continuous. Since the normal distribution is absolutely continuous (has density), the convergence holds for all x.

Example 2.4. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid Bernoulli random variables with parameter p, meaning that each X_i takes the value 1 with probability p and 0 with probability $1 - p$. The mean and variance of each X_i are given by:

$$\mu = \mathbb{E}[X_i] = p, \quad \sigma^2 = \mathrm{Var}(X_i) = p(1-p).$$

We also have the distribution of $X_1 + \cdots + X_n$:

$$\mathbb{P}\left(X_1 + \cdots + X_n = k\right) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

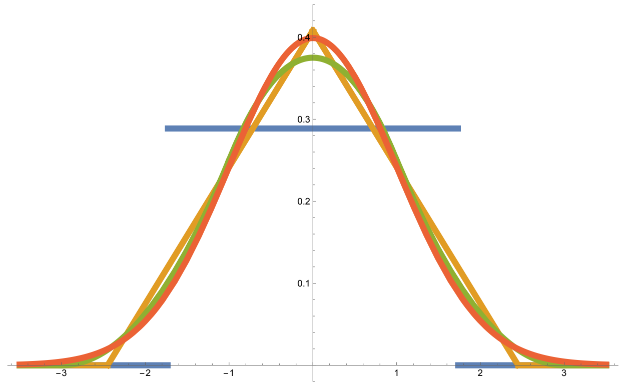

Figure 1: Densities of U_1, U_1+U_2, $U_1+U_2+U_3$ (where U_i are iid uniform on $[0,1]$), and $\mathcal{N}(0,1)$, normalized to have the same mean and variance.

Introduce the normalized quantity

$$z = \frac{k-np}{\sqrt{np(1-p)}}, \tag{2.3}$$

and assume that throughout the asymptotic analysis, this quantity stays finite.

Our aim is to show that, for k such that z remains bounded as $n\to\infty$, the following holds:

$$\mathbb{P}(S_n = k) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Bigl(-\frac{z^2}{2}\Bigr)(1+o(1)).$$

For large n, Stirling's formula gives

$$m! \sim \sqrt{2\pi m}\, m^m e^{-m}, \quad \text{as } m\to\infty.$$

Apply Stirling's approximation to $n!$, $k!$, and $(n-k)!$:

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}, \quad k! \sim \sqrt{2\pi k}\, k^k e^{-k}, \quad (n-k)! \sim \sqrt{2\pi(n-k)}\,(n-k)^{n-k} e^{-(n-k)}.$$

Thus,

$$\binom{n}{k} \sim \frac{\sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi k}\, k^k e^{-k} \sqrt{2\pi(n-k)}\,(n-k)^{n-k} e^{-(n-k)}} = \frac{n^n}{k^k (n-k)^{n-k}} \frac{1}{\sqrt{2\pi\, k(n-k)/n}}.$$

More precisely, one often writes

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Bigl(n\ln n - k\ln k - (n-k)\ln(n-k)\Bigr),$$

where $p \approx k/n$ thanks to the fact that z (2.3) is assumed to be finite.

We have

$$k = np + z\sqrt{np(1-p)}.$$

Then, consider the second-order Taylor expansion. We have

$$n \ln n - k \ln k - (n-k)\ln(n-k) \sim nH - \frac{z^2}{2},$$

where $H = -[p \ln p + (1-p)\ln(1-p)] + c(z;p)/\sqrt{n}$ (for an explicit function $c(z;p)$) is the "entropy" term which exactly cancels with the prefactors coming from $p^k(1-p)^{n-k}$.

After combining the approximations from the binomial coefficient and the probability weights, one arrives at

$$\mathbb{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{z^2}{2}\right),$$

as desired.

(Note that this is a *local* CLT as opposed to the convergence (2.2) in the classical CLT; but one can get the latter from the local CLT by integration.)

2.2 Moments of the normal distribution

Proposition 2.5. *The moments of a random variable $Z \sim \mathcal{N}(0, \sigma^2)$ are given by:*

$$\mathbb{E}[Z^k] = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ \sigma^k (k-1)!! = \sigma^k \cdot (k-1)(k-3)\cdots 1, & \text{if } k \text{ is even.} \end{cases} \tag{2.4}$$

Proof. We just compute the integrals. Assume k is even (for odd, the integral is zero by symmetry). Also assume $\sigma = 1$ for simplicity. Then

$$\mathbb{E}[Z^k] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^k e^{-z^2/2}\, dz.$$

Applying integration by parts (putting $ze^{-z^2/2}$ under d), we get

$$\mathbb{E}[Z^k] = \frac{1}{\sqrt{2\pi}} \left[-z^{k-1} e^{-z^2/2}\right]_{-\infty}^{\infty} + \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{k-2} e^{-z^2/2}\, dz.$$

The first term vanishes at infinity (you can verify this using L'Hôpital's rule), leaving us with:

$$\mathbb{E}[Z^k] = (k-1)\,\mathbb{E}[Z^{k-2}].$$

This gives us a recursive formula, and completes the proof. □

2.3 Moments of sums of iid random variables

Let us now show the CLT by moments. For example, the source is [Bil95, Section 30] or [Fil10].

Remark 2.6. This proof requires an additional assumption that all moments of the random variables are finite. This is quite a strong assumption, and while the CLT holds without it, this proof by moments is more algebraic, and will translate to random matrices more directly.

2.3.1 Computation of moments

Denote $Y_i = X_i - \mu$, these are also iid, but have mean 0. We consider

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right].$$

Expanding the k-th power using the multinomial theorem, we obtain:

$$\left(\sum_{i=1}^{n} Y_i\right)^k = \sum_{j_1+j_2+\cdots+j_n=k} Y_{j_1}Y_{j_2}\ldots Y_{j_n}.$$

Taking the expectation and using linearity, we have:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = \sum_{j_1+j_2+\cdots+j_n=k} \mathbb{E}\left[Y_{j_1}Y_{j_2}\ldots Y_{j_n}\right].$$

The sum over all $j_1, \ldots, j_n$ with $j_1 + \ldots + j_n = k$ is the number of ways to partition k into n non-negative integers. We can order these integers, and thus obtain the sum over all partitions of k into $\le n$ parts. Since n is large, we simply sum over all partitions of k. For each partition λ of k (where $k = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ and $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n \ge 0$), we must count the number of distinct multisets of indices $(j_1, j_2, \ldots, j_n)$ that yield the same collection $\{\lambda_1, \lambda_2, \ldots\}$. Then,

$$\mathbb{E}\left[Y_{j_1}Y_{j_2}\ldots Y_{j_n}\right] = \mathsf{m}_{\lambda_1}\mathsf{m}_{\lambda_2}\ldots\mathsf{m}_{\lambda_n},$$

where $\mathsf{m}_j = \mathbb{E}[Y^j]$ (recall the identical distribution of Y_i). Note that $\mathsf{m}_0 = 1$ and $\mathsf{m}_1 = 0$. Let us illustrate this with an example.

Example 2.7. For $k = 4$, there are only two partitions which have no parts equal to 1: $\lambda = (4)$ and $\lambda = (2, 2)$. The number of ways to get (4) (so that $\mathbb{E}[Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4}] = \mathsf{m}_4$) is to just assign one of the j_p to be 4, this can be done in n ways.

The number of ways to get $(2, 2)$ (so that $\mathbb{E}[Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4}] = \mathsf{m}_2^2$) is to assign two of the j_p to be 2 and the other two to be 0, this can be done in $\binom{n}{2}$ ways. Moreover, there are also 6 permutations of the indices $j_p = (i, j)$ which give the same partition $(2, 2)$: (i, i, j, j), (j, j, i, i), (i, j, i, j), (j, i, j, i), (i, j, j, i), (j, i, i, j). Thus, the total number of ways to get $(2, 2)$ is $6\binom{n}{2} \sim 3n^2$.

So, we see that there is an n-dependent factor, and a "combinatorial" factor for each partition.

2.3.2 n-dependent factor

Consider first the n-dependent factor. In the case k is even and $\lambda = (2, 2, \ldots, 2)$, the power of n is $n^{k/2}$. In the case k is even and λ has at least one part ≥ 3, the power of n is at most $n^{k/2-1}$, which is subleading in the limit $n \to \infty$. When k is odd, the "best" we can do (without parts equal to 1) is going to be $\lambda = (3, 2, \ldots, 2)$ with $(k-1)/2$ parts, so the power of n is $n^{(k-1)/2}$. This is also subleading in the limit $n \to \infty$.

2.3.3 Combinatorial factor

Now, we see that we only need to consider the case when k is even and all parts of λ are 2. Then, the n-dependent factor is $\binom{n}{k/2} \sim n^{k/2}/(k/2)!$. The combinatorial factor is equal to the number of ways to partition k into pairs, which is the double factorial:

$$(k-1)!! = (k-1)(k-3)\dots 1,$$

times the number of permutations of the $k/2$ indices which are assigned to the pairs, so $(k/2)!$. In particular, for $k=4$ this is 6.

2.3.4 Putting it all together

We have as $n \to \infty$:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = n^{k/2}\frac{(k-1)!!}{(k/2)!} \cdot (k/2)!\sigma^k + o(n^{k/2}) = n^{k/2}(k-1)!!\sigma^k + o(n^{k/2}).$$

Now, we need to consider the normalization of the sum $\sum_{i=1}^{n} Y_i$ by $\sqrt{n}$:

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right)^k\right] = \frac{1}{n^{k/2}}\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = (k-1)!!\sigma^k + o(1).$$

Therefore, the moments of Z_n (2.1) converge to the moments of the standard normal distribution.

2.4 Convergence in distribution

Is convergence of moments enough to imply convergence in distribution? Not necessarily. First, note that the functions $x \mapsto x^k$ are not even bounded on $\mathbb{R}$.

A sufficient condition for convergence in distribution is found in the classical method of moments in probability theory [Bil95, Theorem 30.2]. This theorem states that if the limiting distribution X is uniquely determined by its moments, then convergence in moments implies convergence in distribution.

The normal distribution is indeed uniquely determined by its moments (Problem A.5), so the CLT holds in this case, provided that the original iid random variables X_i have finite moments of all orders.

3 Random matrices and semicircle law

We now turn to random matrices.

3.1 Where can randomness in a matrix come from?

The study of random matrices begins with understanding how randomness can be introduced into matrix structures. We consider three primary sources:

1. **iid entries:** The simplest form of randomness comes from filling matrix entries independently with samples from a fixed probability distribution. For an $n \times n$ matrix, this gives us n^2 independent random variables. If we do not impose any additional structure on the matrix, then the eigenvalues will be complex. So, often we consider real symmetric, complex Hermitian, or quaternionic matrices with symplectic symmetry.[1]

2. **Correlated Entries:** In many physical systems, especially those modeling local interactions, matrix entries are not independent but show correlation patterns. Common examples include:

   - Band matrices, where entries become negligible far from the diagonal
   - Matrices with correlation decay based on the distance between indices
   - Structured random matrices arising from specific physical models
   - Sparse matrices, where most entries are zero

3. **Haar Measure on Matrix Groups:** Randomness can come from considering matrices sampled according to the Haar measure on compact matrix group, for example, the orthogonal $O(n)$, unitary $U(n)$, or symplectic group $Sp(n)$.[2] One can think of this as a generalization of the uniform distribution (Lebesgue measure) on the unit circle in $\mathbb{C}$, or a unit sphere in $\mathbb{R}^n$. One can also mix and match: one of the most interesting families of random matrices is the one with constant eigenvalues, but random eigenvectors:

$$A = UD_\lambda U^\dagger, \qquad U \in U(n), \quad U \sim \text{Haar}.$$

   Here D_λ is a diagonal matrix with constant eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$. The random matrix A is the "uniform" random variable taking values in the set of all Hermitian matrices with fixed real eigenvalues λ. Here we may assume that $\lambda_1 \geq \ldots \geq \lambda_n$, since the unitary conjugation can permute the eigenvalues.

3.2 Real Wigner matrices

Definition 3.1 (Real Wigner Matrix)**.** An $n \times n$ random matrix $W = W_n = (X_{ij})_{1 \leq i,j \leq n}$ is called a *real Wigner matrix* if:

1. W is symmetric: $X_{ij} = X_{ji}$ for all i, j;

2. The upper triangular entries $\{X_{ij} : 1 \leq i \leq j \leq n\}$ are independent;

[1] Real symmetric means $A^\top = A$, complex Hermitian means $A^\dagger = A$ (conjugate transpose). Let us briefly discuss the quaternionic case. It can be modeled over $\mathbb{C}$. A quaternion $q = a + b\,\mathbf{i} + c\,\mathbf{j} + d\,\mathbf{k}$ can be represented by the complex 2×2 matrix

$$q \longmapsto \begin{pmatrix} a + \mathbf{i}b & c + \mathbf{i}d \\ -c + \mathbf{i}d & a - \mathbf{i}b \end{pmatrix}.$$

The entries a, b, c, d for the quaternion matrix case must be real, and the matrix A of size $2n \times 2n$ should also be Hermitian in the usual complex sense.

[2] The orthogonal and unitary groups are defined in the usual way, by $OO^\top = O^\top O = I$ and $UU^\dagger = U^\dagger U = I$, respectively. The group $Sp(n)$ is the compact real form of the full symplectic group $Sp(2n, \mathbb{C})$, consisting of $2n \times 2n$ matrices A such that $A^\top JA = J$, where J is the skew-symmetric form.

3. The diagonal entries $\{X_{ii}\}$ are iid real random variables with mean 0 and variance σ_d;
4. The upper triangular entries $\{X_{ij} : i < j\}$ are iid (possibly with a distribution different from the diagonal entries) real random variables with mean 0 and variance σ;
5. (optional, but we assume this) All entries have finite moments of all orders.

Example 3.2 (Gaussian Wigner Matrices, Gaussian Orthogonal Ensemble (GOE))**.** Let W be a real Wigner matrix where:

- Diagonal entries $X_{ii} \sim \mathcal{N}(0, 2)$;
- Upper triangular entries $X_{ij} \sim \mathcal{N}(0, 1)$ for $i < j$.

We can model W as $(Y + Y^\top)/\sqrt{2}$, where Y is a matrix with iid Gaussian entries $Y_{ij} \sim \mathcal{N}(0, 1)$. The matrix distribution of W is called the *Gaussian Orthogonal Ensemble* (*GOE*).

Remark 3.3 (Wishart Matrices)**.** There are other ways to define random matrices, most notably, *sample covariance matrices.* Let $A = [a_{i,j}]_{i,j=1}^{n,m}$ be an $n \times m$ matrix ($n \leq m$), where entries are iid real random variables with mean 0 and finite variance. Then $M = AA^\top$ is a positive symmetric random matrix of size $n \times n$. It almost surely has full rank.

3.3 Empirical spectral distribution

For an arbitrary random matrix of size $n \times n$ with real eigenvalues, the *empirical spectral distribution* (*ESD*) is defined as the random probability measure on $\mathbb{R}$:

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}, \tag{3.1}$$

which puts point masses of size $1/n$ at the eigenvalues λ_i of the matrix.

If you sample the ESD for a large real Wigner matrix, and take a histogram (to cluster the eigenvalues into boxes), you will see the semi-circular pattern. This pattern does not change over several samples. Hence, one can conjecture that the ESD (3.1) converges to a nonrandom measure, after rescaling.

We can guess the rescaling by looking at the first two moments of the ESD. The first moment is

$$\int_{\mathbb{R}} x\, \mu_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i = \frac{1}{n} \operatorname{Tr}(W) = \frac{1}{n} \sum_{i=1}^{n} X_{ii}, \tag{3.2}$$

and this sum has mean zero (and small variance), so it converges to zero. The second moment is

$$\int_{\mathbb{R}} x^2\, \mu_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 = \frac{1}{n} \operatorname{Tr}(W^2) = \frac{1}{n} \sum_{i,j=1}^{n} X_{ij}^2. \tag{3.3}$$

This sum has mean $\sim \sigma^2 n^2$, so even normalized by n, it still goes to infinity. But, if we normalize the matrix as $\frac{1}{\sqrt{n}} W$, then the second moment becomes bounded, and one can convince oneself that the ESD of the normalized Wishart matrix has a limit. Indeed, this is the case:

Theorem 3.4 (Wigner's Semicircle Law)**.** *Let W be a real Wigner matrix of size $n \times n$ (with off-diagonal entries having a fixed variance σ^2, independent of n). Then as $n \to \infty$, the ESD of $W/(\sigma\sqrt{n})$ converges in distribution to the semicircular law:*

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \longrightarrow \mu_{\mathrm{sc}}, \tag{3.4}$$

where μ_{sc} is the semicircular distribution with density with respect to the Lebesgue measure:

$$\mu_{\mathrm{sc}}(dx) := \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{|x|\le 2}\,dx. \tag{3.5}$$

Remark 3.5. The convergence in (3.4) may mean either *weakly in probability* or *weakly almost surely.* The first notion, weak convergence in probability, means that for every bounded continuous function f, we have

$$\int_{\mathbb{R}} f(x)\,\nu_n(dx) \longrightarrow \int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx), \qquad n \to \infty, \tag{3.6}$$

where in (3.6) the convergence is in probability. Indeed, the left-hand side of (3.6) is a random variable, so we need to qualify which sense of convergence we mean.

The weakly almost sure convergence means that the convergence in (3.6) holds for almost all realizations of the random matrix W, that is, for every bounded continuous function f, the random variable $\int_{\mathbb{R}} f(x)\,\nu_n(dx)$ converges almost surely to $\int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx)$.

Remark 3.6. There exists a version of the limiting ESD for the Wishart matrices (Remark 3.3). In this case, the limiting distribution is the *Marchenko-Pastur law* [MP67].

3.4 Expected moments of traces of random matrices

The main computation in the proof of Theorem 3.4 is the computation of expected moments of the ESD. This computation of moments is somewhat similar to the one in the proof of the CLT by moments, but has its own random matrix flavor.

Definition 3.7 (Normalized Moments)**.** For each $k \ge 1$, the normalized k-th moment of the empirical spectral distribution of $W_n/\sqrt{n}$ is given by

$$m_k^{(n)} = \int_{\mathbb{R}} x^k\,\nu_n(dx) = \frac{1}{n^{k/2+1}}\,\mathrm{Tr}(W^k).$$

Our first goal is to study the asymptotic behavior of $\mathbb{E}[m_k^{(n)}]$ as $n \to \infty$ for each fixed $k \ge 1$, just like we did in (3.2)–(3.3) for $k = 1, 2$:

$$\mathbb{E}[m_1^{(n)}] = 0, \qquad \mathbb{E}[m_2^{(n)}] \to \sigma^2.$$

Note that $\mathbb{E}[m_2^{(n)}]$ is not exactly equal to σ^2 because of the presence of the diagonal elements which have a different distribution. In general, we will see that the contribution of the diagonal elements to the moments is negligible in the limit $n \to \infty$.

Lemma 3.8 (Convergence of Expected Moments)**.** *For each fixed $k \geq 1$, we have*

$$\lim_{n\to\infty} \mathbb{E}[m_k^{(n)}] = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \sigma^k C_{k/2} & \text{if } k \text{ is even,} \end{cases}$$

where $C_m = \frac{1}{m+1}\binom{2m}{m}$ is the m-th Catalan number.

The even moments are scaled by powers of σ just as in the case $k = 2$, while the odd moments vanish due to the symmetry of the limiting distribution around zero. As we will see, the appearance of Catalan numbers is not accidental, but it is due to the underlying combinatorics.

Proof of Lemma 3.8. The trace of W^k expands as a sum over all possible index sequences:

$$\mathrm{Tr}(W^k) = \sum_{i_1,\ldots,i_k=1}^{n} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}. \tag{3.7}$$

Due to independence and the fact that $\mathbb{E}[X_{ij}] = 0$ for all i, j, the only nonzero contributions come from index sequences where each matrix element appears least twice.

As in the CLT proof, there is a power-n factor and a combinatorial factor.

For k odd, let us count the power of n first. As in the CLT proof, the maximum power comes from index sequences where all matrix elements appear exactly twice except for one which appears three times. Indeed, this corresponds to the maximum freedom of choosing k indices among the large number n of indices, and thus to the maximum power of n. This maximum power of n is $n^{1+\lfloor k/2 \rfloor}$ (note that there is an extra factor n compared to the CLT proof, as now we have $\sim n^2$ random variables in the matrix instead of n). Since this is strictly less than the normalization $n^{k/2+1}$ in $m_k^{(n)}$, the term with odd k vanish in the limit $n \to \infty$.

Assume now that k is even. Then the maximum power of n comes from index sequences where each matrix element appears exactly twice. This power of n is $n^{k/2+1}$, which exactly matches the normalization in $m_k^{(n)}$.

It remains to count the combinatorial factor, assuming that k is even. For each term in the trace expansion, we can represent the sequence of indices $(i_1, \ldots, i_k)$ as a directed closed path with vertices $\{1, \ldots, n\}$ and edges given by the matrix entries $X_{i_a i_{a+1}}$. For example, if $k = 4$ and we have a term $X_{12}X_{23}X_{34}X_{41}$, this corresponds to the path $1 \to 2 \to 3 \to 4 \to 1$. Recall that our path must have each matrix entry exactly twice (within the symmetry $X_{ij} = X_{ji}$), and the path must be closed. The condition that each edge appears exactly twice means that if we forget the direction of the edges and the multiplicities, we must get a *tree*, with $k/2$ edges and $k/2 + 1$ vertices. The complete justification of this counting is the problem in Problem A.9.

The n-powers counting implies that the combinatorial factor (for even k) is equal to σ^k times the number of *rooted* (*planar*) *trees* with $k/2$ edges. The rooted condition comes from the fact that we are free to choose fix the starting point of the path to be 1 (this ambiguity is taken into account by the power-n factor).

In Problem A.10, we show that the number of these rooted trees is the $k/2$-th Catalan number $C_{k/2}$. This completes the proof of Lemma 3.8. $\square$

3.5 Immediate next steps

The proof of Theorem 3.4 is continued in the next Lecture 2. Immediate next steps are:

1. Show that the number of rooted trees with $k/2$ edges is the $k/2$-th Catalan number, and give the exact formula for the Catalan numbers.
2. Compute the moments of the semicircular distribution.
3. Make sure that the moment computation suffice to show the weak in probability convergence of the ESD to the semicircular law.

A Problems (due 2025-02-13)

Each problem is a subsection (like Problem A.1), and is may have several parts.

A.1 Normal approximation

1. In Figure 1, which color is the normal curve and which is the sum of three uniform random variables?
2. Show that the sum of 12 iid uniform random variables on $[-1, 1]$ (without normalization) is approximately standard normal.
3. Find (numerically is okay) the maximum discrepancy between the distribution of the sum of 12 iid uniform random variables on $[-1, 1]$ and the standard normal distribution:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\sum_{i=1}^{12} U_i \le x\right) - \mathbb{P}\left(Z \le x\right) \right|.$$

A.2 Convergence in distribution

Convergence in distribution $X_n \to X$ for real random variables X_n and X means, by definition, that

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$$

for all bounded continuous functions f. Show that convergence in distribution is equivalent to the condition outlined in (2.2):

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$$

for all x at which the cumulative distribution function of X is continuous.

A.3 Moments of sum justification

Justify the computations of the power of n in Section 2.3.2.

A.4 Distribution not determined by moments

Show that the log-normal random variable e^Z (where $Z \sim \mathcal{N}(0,1)$) is not determined by its moments.

A.5 Uniqueness of the normal distribution

Show that the normal distribution is uniquely determined by its moments.

A.6 Quaternions

Show that the 2×2 matrix representation of a quaternion given in Footnote 1 indeed satisfies the quaternion multiplication rules. Hint: Use linearity and distributive law.

A.7 Ensemble $UD_\lambda U^\dagger$

Let U be the random Haar-distributed unitary matrix of size $N \times N$. Let D_λ be the diagonal matrix with constant real eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \geq \ldots \geq \lambda_N$. Let us fix λ to be, say, $\lambda = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, for some proportion of 1's and 0's (you can start with half ones and half zeros).

Use a computer algebra system to sample the eigenvalues of the matrix obtained from $UD_\lambda U^\dagger$ by taking only its top-left corner of size $k \times k$, where $k = 1, 2, \ldots, N$. For a fixed k, let $\lambda_1^{(k)} \geq \ldots \geq \lambda_k^{(k)}$ be the eigenvalues of the top-left corner of size $k \times k$. Plot the two-dimensional array

$$\left\{(\lambda_i^{(k)}, k) \colon i = 1, \ldots, k\right\} \subset \mathbb{R} \times \mathbb{Z}_{\geq 1}.$$

A.8 Invariance of the GOE

Show that the distribution of the GOE is invariant under conjugation by orthogonal matrices:

$$\mathbb{P}(OWO^\top \in \mathrm{A}) = \mathbb{P}(W \in \mathrm{A})$$

for all orthogonal matrices O and Borel sets A.

A.9 Counting n-powers in the real Wigner matrix

Show that in the expansion of the expected trace of the k-th power of the real Wigner matrix, the maximum power of n is $k/2 + 1$ for even k and less for odd k. For even k, the power $k/2 + 1$ comes from index sequences where each off-diagonal matrix element appears exactly twice, and no diagonal elements are present.

A.10 Counting trees

Show that the number of rooted trees with m edges is the m-th Catalan number:

$$C_m = \frac{1}{m+1}\binom{2m}{m}.$$

References

[Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. ↑5, 7

[Dix05] A. L. Dixon, *Generalization of Legendre's Formula* $KE' - (K - E)K' = \frac{1}{2}\pi$, Proc. Lond. Math. Soc. **2** (1905), no. 1, 206–224. ↑2

[Dys62a] F.J. Dyson, *A Brownian motion model for the eigenvalues of a random matrix*, Jour. Math. Phys. **3** (1962), no. 6, 1191–1198. ↑2

[Dys62b] F.J. Dyson, *Statistical Theory of the Energy Levels of Complex Systems. I-III*, Jour. Math. Phys. **3** (1962), no. 166, 140–175. ↑2

[Fil10] Y. Filmus, *Two proofs of the central limit theorem*, 2010. http://www.cs.toronto.edu/~yuvalf/CLT.pdf. ↑5

[Hur97] A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachr. Ges. Wiss. Göttingen (1897), 71–90. ↑2

[MP67] V. Marchenko and L. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb. **1** (1967), 457–483. ↑10

[Wig55] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics (1955), 548–564. ↑2

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)
Lecture 2: Wigner semicircle law

Leonid Petrov

Wednesday, January 15, 2025*

Contents

*Course webpage • TeX Source • Updated at 03:51, Wednesday 22nd January, 2025

1 Recap

We are working on the Wigner semicircle law.

1. Wigner matrices W: real symmetric random matrices with iid entries X_{ij}, $i > j$ (mean 0, variance σ^2); and iid diagonal entries X_{ii} (mean 0, some other variance and distribution).

2. Empirical spectral distribution (ESD)

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}},$$

which is a random probability measure on $\mathbb{R}$.

3. Semicircle distribution μ_{sc}:

$$\mu_{\text{sc}}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2}\, dx, \qquad x \in [-2, 2].$$

4. Computation of expected traces of powers of W (with variance 1). We showed that

$$\int_{\mathbb{R}} x^k \, \nu_n(dx) \to \# \left\{ \text{rooted planar trees with } k/2 \text{ edges} \right\}.$$

Remark 1.1. If the off-diagonal elements of the matrix have variance σ^2, then the semicircle distribution should be scaled to be supported on $[-2\sigma, 2\sigma]$. We assume that the variance of the off-diagonal elements is 1 in most arguments throughout the lecture.

2 Two computations

First, we finish the combinatorial part, and match the limiting expected traces of powers of W to moments of the semicircle law.

2.1 Moments of the semicircle law

We also need to match the Catalan numbers to the moments of the semicircle law. Let $k = 2m$, and we need to compute the integral

$$\int_{-2}^{2} x^{2m} \frac{1}{2\pi} \sqrt{4 - x^2}\, dx.$$

By symmetry, we write:

$$\int_{-2}^{2} x^{2m} \rho(x)\, dx = \frac{2}{\pi} \int_0^2 x^{2m} \sqrt{4 - x^2}\, dx.$$

Using the substitution $x = 2\sin\theta$, we have $dx = 2\cos\theta\, d\theta$. The integral becomes:

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

Using $\cos^2\theta = 1 - \sin^2\theta$, we split the integral:

$$\frac{2^{2m+2}}{\pi}\left(\int_0^{\pi/2} \sin^{2m}\theta\,d\theta - \int_0^{\pi/2} \sin^{2m+2}\theta\,d\theta\right).$$

Using the standard formula (cf. Problem B.1)

$$\int_0^{\pi/2} \sin^{2n}\theta\,d\theta = \frac{\pi}{2}\frac{(2n)!}{2^{2n}(n!)^2}, \tag{2.1}$$

we compute each term:

$$\frac{2^{2m+2}}{\pi}\left(\frac{\pi}{2}\frac{(2m)!}{2^{2m}(m!)^2} - \frac{\pi}{2}\frac{(2m+2)!}{2^{2m+2}((m+1)!)^2}\right).$$

After simplification, this becomes C_m, the m-th Catalan number.

2.2 Counting trees and Catalan numbers

Throughout this section, for a random matrix trace moment of order k, we use $m = k/2$ as our main parameter. Note that m can be arbitrary (not necessarily even).

Definition 2.1 (Dyck Path)**.** A *Dyck path* of semilength m is a sequence of $2m$ steps in the plane, each step being either $(1,1)$ (up step) or $(1,-1)$ (down step), starting at $(0,0)$ and ending at $(2m,0)$, such that the path never goes below the x-axis. We denote an up step by U and a down step by D.

Definition 2.2 (Rooted Plane Tree)**.** A *rooted plane tree* is a tree with a designated root vertex where the children of each vertex have a fixed left-to-right ordering. The size of such a tree is measured by its number of edges, which we denote by m.

Definition 2.3 (Catalan Numbers)**.** The sequence of *Catalan numbers* $\{C_m\}_{m\geq 0}$ is defined recursively by:

$$C_0 = 1, \quad C_{m+1} = \sum_{j=0}^{m} C_j C_{m-j} \quad \text{for } m \geq 0. \tag{2.2}$$

Alternatively, they have the closed form[1]

$$C_m = \frac{1}{m+1}\binom{2m}{m} = \binom{2m}{m} - \binom{2m}{m+1}. \tag{2.3}$$

These numbers appear naturally in the moments of random matrices, where $m = k/2$ for trace moments of order k.

Lemma 2.4. *Formulas* (2.2) *and* (2.3) *are equivalent.*

[1] See Problem B.4 for a combinatorial proof of the second inequality.

Proof. One can check that the closed form satisfies the recurrence relation by direct substitution. The other direction involves generating functions. Namely, (2.2) can be rewritten for the generating function

$$C(z) = \sum_{m=0}^{\infty} C_m z^m$$

as

$$C(z) = 1 + zC(z)^2.$$

Solving for $C(z)$, we get

$$C(z) = \frac{1 \pm \sqrt{1-4z}}{2z}. \tag{2.4}$$

We need to pick the solution which is nonsingular at $z = 0$, and it corresponds to the minus sign. Taylor expansion of the right-hand side of (2.4) at $z = 0$ gives the closed form. $\square$

Remark 2.5. Catalan numbers enumerate many (too many!) combinatorial objects. For a comprehensive treatment, see [Sta15].

Proposition 2.6 (Dyck Path–Rooted Tree Correspondence)**.** *For any m, there exists a bijection between the set of Dyck paths of semilength m and the set of rooted plane trees with m edges.*

Proof. Given a Dyck path of semilength m, we build the corresponding rooted plane tree as follows (see Figure 1 for an illustration):

1. Start with a single root vertex
2. Read the Dyck path from left to right:
   - For each up step (U), add a new child to the current vertex
   - For each down step (D), move back to the parent of the current vertex
3. The order of children is determined by the order of up steps

This is clearly a bijection, and we are done. $\square$

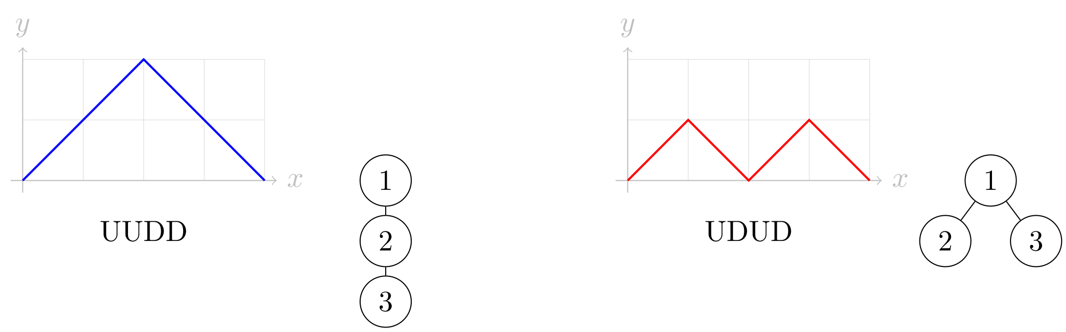

Figure 1: The two possible Dyck paths of semilength $m = 2$ and their corresponding rooted plane trees.

It remains to show that the Dyck paths or rooted plane trees are counted by the Catalan numbers, by verifying the recursion (2.2) for them. By Proposition 2.6, it suffices to consider only Dyck paths.

Proposition 2.7. *The number of Dyck paths of semilength m satisfies the Catalan recurrence* (2.2).

Proof. We need to show that the number of Dyck paths of semilength $m+1$ is given by the sum in the right-hand side of (2.2). Consider a Dyck path of semilength $m+1$, and let the *first* time it returns to zero be at semilength $j+1$, where $j = 0, \ldots, m$. Then the first and the $(2j+1)$-st steps are, respectively, U and D. From 0 to $2j+2$, the path does not return to the x-axis, so we can remove the first and the $(2j+1)$-st steps, and get a proper Dyck path of semilength j. The remainder of the Dyck path is a Dyck path of semilength $m-j$. This yields the desired recurrence. □

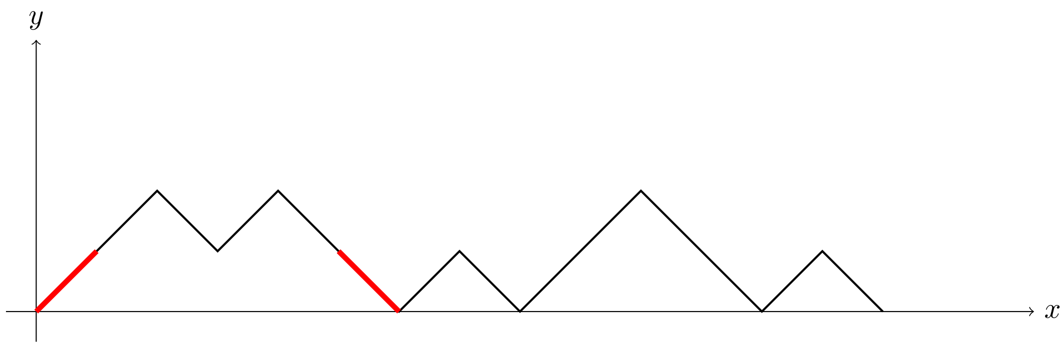

Figure 2: Illustration of a Dyck path decomposition for the proof of Proposition 2.7.

3 Analysis steps in the proof

We are done with combinatorics, and it remains to justify that the computations lead to the desired semicircle law from Lecture 1.

Let us remember that so far, we showed that

$$\lim_{n\to\infty} \frac{1}{n^{k/2+1}} \mathbb{E}\left[\operatorname{Tr} W^k\right] = \begin{cases} \sigma^{2m} C_m & \text{if } k = 2m \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

Here, W is real Wigner (unnormalized) with mean 0, where its off-diagonal entries are iid with variance σ^2.

3.1 The semicircle distribution is determined by its moments

We use (without proof) the known Carleman's criterion for the uniqueness of a distribution by its moments.

Proposition 3.1 (Carleman's criterion [ST43, Theorem 1.10], [Akh65]). *Let X be a real-valued random variable with moments $m_k = \mathbb{E}[X^k]$ of all orders. If*

$$\sum_{k=1}^{\infty} (m_{2k})^{-1/(2k)} = \infty, \tag{3.1}$$

then the distribution of X is uniquely determined by its moments $(m_k)_{k\geq 1}$.

Remark 3.2. Note that we do not assume that the measure is symmetric, but use only even moments for the Carleman criterion. Indeed, in determining uniqueness, the decisive aspect is how the distribution mass "escapes" to $\pm\infty$. Since $\int |x|^n d\mu(x)$ can be bounded by twice $\int x^{2\lfloor n/2 \rfloor} d\mu(x)$ (roughly speaking), controlling $\int x^{2n} d\mu(x)$ also controls $\int |x|^n d\mu(x)$. Thus, one does not need to worry about positive or negative signs in x; the even powers handle both sides of the real line at once.

Moreover, the convergence of (3.1), as for any infinite series, is only determined by arbitrarily large moments, for the same reason.

Remark 3.3. By the Stone-Wierstrass theorem, the semicircle distribution on $[-2, 2]$ is unique among distributions with an arbitrary, but fixed compact support with the moments $\sigma^{2k} C_k$. However, we need to guarantee that there are no distributions on $\mathbb{R}$ with the same moments.

Now, the moments satisfy the asymptotics

$$m_{2k} = C_k \sigma^{2k} \sim \frac{4^k}{k^{3/2}\sqrt{\pi}} \sigma^{2k},$$

so

$$\sum_{k=1}^{\infty} (m_{2k})^{-1/(2k)} \sim \sum_{k=1}^{\infty} \left(\frac{k^{3/2}\sqrt{\pi}}{4^k} \right)^{1/2k} \sigma^{-1}.$$

The k-th summands converges to $1/(2\sigma)$, so the series diverges.

Remark 3.4. See also Problem A.4 from Lecture 1 on an example of a distribution not determined by its moments.

3.2 Convergence to the semicircle law

Recall [Bil95, Theorem 30.2] that convergence of random variables in moments plus the fact that the limiting distribution is uniquely determined by its moments implies convergence in distribution. However, we need weak convergence in probability or almost surely (see the previous Lecture 1). which deals with random variables

$$\int_{\mathbb{R}} f(x)\, \nu_n(dx), \qquad f \in C_b(\mathbb{R}),$$

and we did not compute the moments of these random variables.

To complete the argument, let us show that for each fixed integer $k \geq 1$, we have almost sure convergence of the moments (of a random distribution, so that the $Y_{n,k}$'s are random variables):

$$Y_{n,k} := \int_{\mathbb{R}} x^k\, \nu_n(dx) \xrightarrow[n\to\infty]{\text{a.s.}} m_k, \qquad n \to \infty,$$

where m_k are the moments of the semicircle distribution, and ν_n is the ESD corresponding to the scaling of the eigenvalues as $\lambda_i/\sqrt{n}$.

As typical in asymptotic probability, we not only need the expectation of $Y_{n,k}$, but also their variances, to control the almost sure convergence. Recall that we showed $\mathbb{E}(Y_{n,k}) \to m_k$. Let us assume the following:

Proposition 3.5 (Variance bound)**.** *For each fixed integer $k \geq 1$ and large enough n, we have*

$$\mathrm{Var}(Y_{n,k}) \leq \frac{m_k}{n^2}.$$

We will prove Proposition 3.5 in Section 4 below. Let us finish the proof of convergence to the semicircle law modulo Proposition 3.5.

3.2.1 A concentration bound and the Borel–Cantelli lemma

From Chebyshev's inequality,

$$\mathbb{P}\Big(\big|Y_{n,k} - \mathbb{E}[Y_{n,k}]\big| \geq n^{-\frac{1}{4}}\Big) \leq \mathrm{Var}[Y_{n,k}]\sqrt{n} = O(n^{-\frac{3}{2}}),$$

where in the last step we used Proposition 3.5.

Hence the probability that $|Y_{n,k} - \mathbb{E}[Y_{n,k}]| > n^{-\frac{1}{4}}$ is summable in n. By the Borel–Cantelli lemma, with probability 1 only finitely many of these events occur. Since $\mathbb{E}[Y_{n,k}] \to m_k$, we conclude

$$\big|Y_{n,k} - m_k\big| \;\leq\; \big|Y_{n,k} - \mathbb{E}[Y_{n,k}]\big| + \big|\mathbb{E}[Y_{n,k}] - m_k\big| \;\xrightarrow[n\to\infty]{}\; 0 \quad \text{almost surely.}$$

3.2.2 Tightness of $\{\nu_n\}$ and subsequential limits

Since $|Y_{n,k}| = \big|\int x^k \,\nu_n(dx)\big|$ stays almost surely bounded for each k, one readily checks (Problem B.5) that almost surely, for each fixed k,

$$\nu_n\big(\{x : |x| > M\}\big) \;\leq\; \frac{C}{M^k}. \tag{3.2}$$

By choosing k large, we see that ν_n puts arbitrarily little mass outside any large interval $[-m, m]$. Thus, the sequence of probability measures $\{\nu_n\}$ is *tight*. By Prokhorov's theorem [Bil95, Theorem 25.10], there exists a subsequence ν_{n_j} converging weakly to some probability measure ν^*. We will now characterize all subsequential limits ν^* of ν_n.

3.2.3 Characterizing the limit measure

We claim that $\nu^* = \mu_{\mathrm{sc}}$, the semicircle distribution (and in particular, this measure is not random). Indeed, fix k. Since x^k is a bounded function on a sufficiently large interval, and $\nu_{n_j} \to \nu^*$ weakly, we have

$$\int_{\mathbb{R}} x^k \,\nu_{n_j}(dx) \;\to\; \int_{\mathbb{R}} x^k \,\nu^*(dx).$$

On the other hand, we have already shown

$$\int_{\mathbb{R}} x^k \, \nu_{n_j}(dx) \;=\; Y_{n_j,k} \;\xrightarrow[j\to\infty]{\text{a.s.}}\; m_k \;=\; \int_{\mathbb{R}} x^k \, \mu_{\text{sc}}(dx).$$

Thus

$$\int_{\mathbb{R}} x^k \, \nu^*(dx) \;=\; m_k \;=\; \int_{\mathbb{R}} x^k \, \mu_{\text{sc}}(dx) \qquad \text{for all } k \geq 1.$$

By Proposition 3.1, the measure ν^* is uniquely determined by its moments. Hence ν^* must coincide with μ_{sc}.

Remark 3.6. In Sections 3.2.2 and 3.2.3 we tacitly assumed that we choose an elementary outcome ω, and view ν_n as measures depending on ω. Then, since the convergence of moments is almost sure, ω belongs to a set of full probability. The limiting measure ν^* must coincide with μ_{sc} for this ω, and thus, ν^* is almost surely nonrandom.

Any subsequence of $\{\nu_n\}$ has a further sub-subsequence convergent to ν. By a standard diagonal argument, this forces $\nu_n \to \nu$ in the weak topology (almost surely). This completes the proof that the ESD of our Wigner matrix (rescaled by $\sqrt{n}$) converges to the semicircle distribution weakly almost surely, modulo Proposition 3.5. (See also Problem B.6 for the weakly in probability convergence.)

4 Proof of Proposition 3.5: bounding the variance

There is one more "combinatorial" step in the proof of the semicircle law: we need to show that the variance of the moments of the ESD is bounded by m_k/n^2.

Recall that

$$Y_{n,k} = \int_{\mathbb{R}} x^k \, \nu_n(dx) = \frac{1}{n^{1+\frac{k}{2}}} \sum_{i_1,\ldots,i_k=1}^{n} X_I, \quad \text{where} \quad X_I = X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}.$$

Here we use the notation I for the multi-index $(i_1, \ldots, i_k)$, and throughout the computation below, we use the notation $I \in [n]^k$, where $[n] = \{1, \ldots, n\}$. We have

$$\operatorname{Var}\big(Y_{n,k}\big) = \frac{1}{n^{2+k}} \operatorname{Var}\Big(\sum_{I\in[n]^k} X_I\Big) = \frac{1}{n^{2+k}} \sum_{I,J\in[n]^k} \operatorname{Cov}\big(X_I, X_J\big).$$

We claim that the sum of all covariances is bounded by a constant times n^k, which then implies $\operatorname{Var}\big(Y_{n,k}\big) \leq \text{const} \cdot n^k/n^{2+k} = O\big(\frac{1}{n^2}\big)$.

Step 1. Identifying when $\operatorname{Cov}\big(X_I, X_J\big)$ can be nonzero. For each k-tuple $I = (i_1, i_2, \ldots, i_k) \in [n]^k$, the product

$$X_I = X_{i_1 i_2} \, X_{i_2 i_3} \, \ldots \, X_{i_k i_1}$$

is the product of the entries of our Wigner matrix corresponding to the directed "edges" $(i_1 \to i_2), (i_2 \to i_3), \ldots, (i_k \to i_1)$. Similarly, X_J is determined by the edges of another closed directed walk J.

1. If I and J use disjoint collections of matrix entries, then X_I and X_J are independent, and hence $\mathrm{Cov}(X_I, X_J) = 0$.
2. If there is an edge (say, $X_{i_1 i_2}$) which appears *only once* in exactly one of I or J but not both, then that edge factor is independent and forces $\mathrm{Cov}(X_I, X_J) = 0$ since $\mathbb{E}[X_{i_1 i_2}] = 0$. Indeed, for example if $X_{i_1 i_2}$ appears only in X_I, then

$$\mathbb{E}\left[X_I\right] = \mathbb{E}\left[X_{i_1 i_2}\right] \cdot \mathbb{E}\left[\text{other factors}\right] = 0, \qquad \mathbb{E}\big[X_I X_J\big] = \mathbb{E}[X_{i_1 i_2}] \cdot \mathbb{E}\big[\text{other factors}\big] = 0.$$

Thus, the only way we could get a nonzero covariance is if *every* edge that appears in $I \cup J$ appears at least twice overall. Graphically, let us represent each k-tuple I by a directed closed walk in the complete graph on $[n]$. The union $I \cup J$ must be a connected subgraph in which every directed edge has total multiplicity ≥ 2.

Step 2. Counting the contributions to the sum. Denote by $q = |V(I \cup J)|$ the number of distinct vertices involved in the union $I \cup J$. In principle, there are $O(n^q)$ ways to choose q vertices from $[n]$. Then we need to specify how the edges form two closed walks of length k.

We split into two cases:

1. $\boldsymbol{q \leq k}$**.** Then the n-power in the sum over I, J is at most n^k, which yields the overall contribution $O(n^{-2})$, as desired.
2. $\boldsymbol{q \geq k+1}$**.** Ignoring directions and multiplicities, we see that the subgraph corresponding to $I \cup J$ contains at most k edges. Since $q \geq k+1$, we must have $q = k+1$ (by connectedness). Thus, $I \cup J$ is a double tree. Since I and J are subsets of this double tree and $q = k+1$, they also must be double trees. Thus, there exists an edge which appears in both I and J, and at least twice in I and twice in J, so four times in $I \cup J$. This contradicts the assumption that $I \cup J$ is a double tree.

   This implies that there are no leading contributions to the sum when $q \geq k+1$.

Combining these two cases, we conclude that the total number of pairs (I, J) with nonzero covariance is of order at most n^k, This yields the desired bound on the variance, and completes the proof of Proposition 3.5.

With that, we are done with the Wigner semicircle law proof for real Wigner matrices (with weakly almost sure convergence; see Lecture 1 for the definitions).

Also, see Problem B.7 for the complex case of the Wigner semicircle law.

5 Remark: Variants of the semicircle law

Let us briefly outline a few examples of the semicircle law for real/complex Wigner matrices which relax the iid conditions and the conditions that all moments of the entries must be finite. This list is not comprehensive, it is presented as an illustration of the universality / robustness of the semicircle law.

Theorem 5.1 (Gaussian β-Ensembles [Joh98], [For10])**.** *Let $\beta > 0$, and consider an $n \times n$ random matrix ensemble with joint eigenvalue density:*

$$p_n(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_{n,\beta}} \exp\left(-\frac{\beta}{4} \sum_{i=1}^{n} \lambda_i^2\right) \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \tag{5.1}$$

where $Z_{n,\beta}$ is the normalization constant.[2] *Then the ESD of the normalized eigenvalues $\lambda_i/\sqrt{n}$ converges weakly almost surely to the semicircle law.*

Theorem 5.2 (Correlated entries [SSB05]). *Let $W_n = \left(\frac{1}{\sqrt{n}} X_{pq}\right)_{1 \le p,q \le n}$ be a sequence of $n \times n$ Hermitian random matrices where:*

1. *The entries X_{pq} are complex random variables that are:*
   - *Centered:* $\mathbb{E}[X_{pq}] = 0$,
   - *Unit variance:* $\mathbb{E}[|X_{pq}|^2] = 1$,
   - *Moment bound:* $\sup_n \max_{p,q=1,\ldots,n} \mathbb{E}\left[|X_{pq}|^k\right] < \infty$ *for all* $k \in \mathbb{N}$.
2. *There exists an equivalence relation $\sim_n$ on pairs of indices (p,q) in $\{1,\ldots,n\}^2$ such that:*
   - *Entries $X_{p_1q_1}, \ldots, X_{p_jq_j}$ are independent when $(p_1,q_1), \ldots, (p_j,q_j)$ belong to distinct equivalence classes.*
   - *The relation satisfies the following bounds:*

     (a) $\max_p \#\{(q,p',q') \in \{1,\ldots,n\}^3 \mid (p,q) \sim_n (p',q')\} = o(n^2)$,

     (b) $\max_{p,q,p'} \#\{q' \in \{1,\ldots,n\} \mid (p,q) \sim_n (p',q')\} \le B$ *for some constant* B,

     (c) $\#\{(p,q,p') \in \{1,\ldots,n\}^3 \mid (p,q) \sim_n (q,p')$ *and* $p \ne p'\} = o(n^2)$.
3. *The matrices are Hermitian: $X_{pq} = \overline{X_{qp}}$. In particular, $(p,q) \sim_n (q,p)$, and this is consistent with the conditions on the equivalence relation.*

Then, as $n \to \infty$, the ESD of W_n converges to the semicircle law.

There are variants of this theorem without the assumption that all moments of the entries are finite.

Theorem 5.3 ([BGK16]). *Let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with random entries such that:*

- *The off-diagonal elements X_{ij}, for $i < j$, are i.i.d. random variables with $\mathbb{E}[X_{ij}] = 0$ and $\mathbb{E}[X_{ij}^2] = 1$.*
- *The diagonal elements X_{ii} are i.i.d. random variables with $\mathbb{E}[X_{ii}] = 0$ and a finite second moment, $\mathbb{E}[X_{ii}^2] < \infty$, for $1 \le i \le n$.*

Then the ESD of M_n, normalized by $\sqrt{n}$, converges to the semicircle law.

Theorem 5.4. *For each $n \in \mathbb{Z}_+$, let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with real random entries satisfying the following conditions:*

- *The entries X_{ij} are independent (but not necessarily identically distributed) random variables with $\mathbb{E}[X_{ij}] = 0$ and $\mathbb{E}[X_{ij}^2] = 1$.*

[2] For $\beta = 1, 2, 4$, this is the joint eigenvalue density of the Gaussian Orthogonal, Unitary, and Symplectic Ensembles, respectively. For general β, there is no invariant random matrix distribution (while the eigenvalue density (5.1) makes sense), and we can still treat all the β cases in a unified manner.

- *There exists a constant C such that* $\sup_{i,j,n} \mathbb{E}\left[|X_{ij}|^4\right] < C$.

Then the ESD of M_n, normalized by $\sqrt{n}$, converges to the semicircle law almost surely. The second condition can also be replaced by a uniform integrability condition on the variances.

Theorem 5.5 (For example, see [SB95])**.** *Let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with random entries. Assume that the expected matrix $\mathbb{E}[M_n]$ has rank $r(n)$, where*

$$\lim_{n\to\infty} \frac{r(n)}{n} = 0.$$

Additionally, suppose $\mathbb{E}[X_{ij}] = 0$, $\operatorname{Var}(X_{ij}) = 1$, and

$$\sup_{i,j,n} \mathbb{E}\left[|X_{ij} - \mathbb{E}[X_{ij}]|^4\right] < \infty.$$

Then the ESD of M_n, normalized by $\sqrt{n}$, converges to the semicircle law almost surely.

B Problems (due 2025-02-15)

B.1 Standard formula

Prove formula (2.1):

$$\int_0^{\pi/2} \sin^{2n}\theta \, d\theta = \frac{\pi}{2} \frac{(2n)!}{2^{2n}(n!)^2}.$$

B.2 Tree profiles

Show that the expected height of a uniformly random Dyck path of semilength m is of order $\sqrt{m}$.

B.3 Ballot problem

Suppose candidate A receives p votes and candidate B receives q votes, where $p > q \geq 0$. In how many ways can these votes be counted such that A is always strictly ahead of B in partial tallies?

B.4 Reflection principle

Show the equality

$$C_m = \binom{2m}{m} - \binom{2m}{m-1},$$

where C_m counts the number of lattice paths from $(0,0)$ to $(2m,0)$ with steps $(1,1)$ and $(1,-1)$ that never go below the x-axis, and binomial coefficients count arbitrary lattice paths from $(0,0)$ to $(2m,0)$ or to $(2m,2)$ with steps $(1,1)$ and $(1,-1)$. In other words, show that the difference between the number of paths to $(2m,0)$ and to $(2m,2)$ is C_m, the number of paths that never go below the x-axis.

B.5 Bounding probability in the proof

Show inequality (3.2).

B.6 Almost sure convergence and convergence in probability

Show that in Wigner's semicircle law, the weakly almost sure convergence of random measures ν_n to μ_{sc} implies weak convergence in probability.

B.7 Wigner's semicircle law for complex Wigner matrices

Complex Wigner matrices are Hermitian symmetric, with iid complex off-diagonal entries, and real iid diagonal entries (all mean zero). Each complex random variable has independent real and imaginary parts.

1. Compute the expected trace of powers of a complex Wigner matrix.
2. Outline the remaining steps in the proof of Wigner's semicircle law for complex Wigner matrices.

B.8 Semicircle law without the moment condition

Prove Theorem 5.3.

References

[Akh65] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*, Translated by N. Kemmer, Hafner Publishing Co., New York, 1965. ↑6

[BGK16] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for Wigner matrices*, Advanced Topics in Random Matrices, 2016. arXiv:1601.04055 [math.PR]. ↑10

[Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. ↑6, 7

[For10] Peter J. Forrester, *Log-gases and random matrices*, Princeton University Press, 2010. ↑9

[Joh98] K. Johansson, *On fluctuations of eigenvalues of random hermitian matrices*, Duke Math. J. **91** (1998), no. 1, 151–204. ↑9

[SB95] J. W. Silverstein and Z. D. Bai, *On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices*, J. Multivar. Anal. **54** (1995), no. 2, 175–192. ↑11

[SSB05] J. Schenker and H. Schulz-Baldes, *Semicircle law and freeness for random matrices with symmetries or correlations*, Math. Res. Lett. **12** (2005). arXiv:math-ph/0505003. ↑10

[ST43] J. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society (AMS), 1943. ↑6

[Sta15] R. Stanley, *Catalan numbers*, Cambridge University Press, 2015. ↑4

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)
Lecture 3: Gaussian and tridiagonal matrices

Leonid Petrov

Wednesday, January 22, 2025*

Contents

*Course webpage • TeX Source • Updated at 04:14, Wednesday 22nd January, 2025

1 Recap

We have established the semicircle law for real Wigner random matrices. If W is an $n \times n$ real symmetric matrix with independent entries X_{ij} above the main diagonal (mean zero, variance 1), and mean zero diagonal entries, then the empirical spectral distribution of $W/\sqrt{n}$ converges to the semicircle law as $n \to \infty$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} = \mu_{\mathrm{sc}}, \tag{1.1}$$

where

$$\mu_{\mathrm{sc}}(dx) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}\, dx, & \text{if } |x| \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

The convergence in (1.1) is weakly almost sure.

Today, we are going to focus on Gaussian ensembles. The plan is:

- Definition and spectral density for real symmetric Gaussian matrices (GOE).
- Tridiagonalization and general beta ensemble.
- Wigner's semicircle law via tridiagonalization.

2 Gaussian Ensembles

Recall that a real Wigner matrix W can be modeled as

$$W = \frac{Y + Y^\top}{\sqrt{2}},$$

where Y is an $n \times n$ matrix with independent entries Y_{ij}, $1 \leq i, j \leq n$, such that Y_{ij} are mean zero, variance 1. Then

$$\mathrm{Var}\left(W_{ii}\right) = \mathrm{Var}\left(\sqrt{2}Y_{ii}\right) = 2, \qquad \mathrm{Var}\left(W_{ij}\right) = \mathrm{Var}\left(\frac{Y_{ij}+Y_{ji}}{\sqrt{2}}\right) = 1,$$

3 Joint Eigenvalue Distribution for GOE ($\beta = 1$)

In this section, we give a very explicit and detailed derivation of the joint probability density for the eigenvalues of a real-symmetric Gaussian matrix, commonly known as the *Gaussian Orthogonal Ensemble* (GOE). Our primary goal is to show:

Theorem 3.1 (GOE Joint Eigenvalue Density)**.** *Let M be an $N \times N$ real-symmetric matrix with distribution defined by*

- *Off-diagonal entries M_{ij}, $i < j$, i.i.d. $\mathcal{N}(0, \sigma^2)$.*
- *Diagonal entries M_{ii} i.i.d. $\mathcal{N}(0, 2\sigma^2)$.*

Then its ordered real eigenvalues $\lambda_1 \le \cdots \le \lambda_N$ have a joint probability density function given by:

$$p(\lambda_1, \ldots, \lambda_N) \;=\; \frac{1}{Z_{N,\sigma}} \prod_{1 \le i < j \le N} \bigl|\lambda_i - \lambda_j\bigr| \, \exp\Bigl(-\frac{1}{4\sigma^2} \sum_{k=1}^{N} \lambda_k^2\Bigr),$$

where $Z_{N,\sigma}$ is a constant (depending on N and σ but not on λ_i) ensuring the density integrates to 1.

Remark 3.2. Often one takes $\sigma^2 = 1/2$, in which case the exponent becomes $\exp(-\frac{1}{2}\sum \lambda_k^2)$. We will keep a general σ for clarity.

We break the proof into four major steps:

3.1 Step A: Joint Density of Matrix Entries

Let us label all independent entries of M:

$$\{\underbrace{M_{12}, M_{13}, \ldots}_{\text{above diag}}, \underbrace{M_{22}, M_{33}, \ldots}_{\text{diag}}, \ldots\}.$$

There are $\frac{N(N-1)}{2}$ off-diagonal entries and N diagonal entries. By definition:

$$M_{ij} = M_{ji}, \quad M_{ij} \sim \mathcal{N}(0, \sigma^2) \text{ for } i < j, \quad M_{ii} \sim \mathcal{N}(0, 2\sigma^2).$$

Thus the joint density of these entries (ignoring normalization for a moment) is

$$f(M) \;\propto\; \exp\Bigl(-\frac{1}{2\sigma^2} \sum_{i<j} M_{ij}^2 - \frac{1}{4\sigma^2} \sum_{i=1}^{N} M_{ii}^2\Bigr).$$

One can rewrite $\sum_{i<j} M_{ij}^2 + \frac{1}{2}\sum_i M_{ii}^2$ as $\frac{1}{2}\sum_{i,j} M_{ij}^2$. Indeed,

$$\mathrm{Tr}(M^2) = \sum_{i,j} M_{ij}^2 \quad \text{for real-symmetric } M.$$

But each off-diagonal term M_{ij}^2 for $i < j$ appears exactly once in $\sum_{i,j}$, while each diagonal term M_{ii}^2 appears once. Hence

$$\sum_{i<j} M_{ij}^2 + \tfrac{1}{2} \sum_{i} M_{ii}^2 \;=\; \tfrac{1}{2} \sum_{i,j} M_{ij}^2 \;=\; \tfrac{1}{2}\, \mathrm{Tr}(M^2).$$

Thus

$$f(M) \;=\; (\text{constant}) \times \exp\Bigl(-\tfrac{1}{4\sigma^2}\, \mathrm{Tr}(M^2)\Bigr).$$

Including the correct normalization for Gaussians, one arrives at

$$f(M)\, dM \;=\; \bigl(2\pi\sigma^2\bigr)^{-\frac{N(N-1)}{4}} \, \bigl(4\pi\sigma^2\bigr)^{-\frac{N}{4}} \, \exp\Bigl(-\tfrac{1}{4\sigma^2}\, \mathrm{Tr}(M^2)\Bigr)\, dM,$$

where dM is the product measure over the $\frac{N(N+1)}{2}$ independent entries.

3.2 Step B: Spectral Decomposition $M = Q\Lambda Q^T$

Since M is real-symmetric, it can be orthogonally diagonalized:

$$M = Q\,\Lambda\,Q^T, \quad Q \in O(N),$$

where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ has the eigenvalues. Then:

$$\mathrm{Tr}(M^2) = \mathrm{Tr}\big(Q\,\Lambda\,Q^T Q\,\Lambda\,Q^T\big) = \mathrm{Tr}(\Lambda^2) = \sum_{k=1}^{N} \lambda_k^2.$$

So

$$f(M) \;=\; (\text{constant})\,\exp\Big(-\tfrac{1}{4\sigma^2}\sum_{k=1}^{N}\lambda_k^2\Big).$$

3.3 Step C: The Orthogonal-Group Volume dQ and Jacobian Calculation

We now examine how the measure dM in the space of real-symmetric matrices factors into a piece depending on $\{\lambda_i\}$ and a piece depending on Q. Formally,

$$dM = \Big|\det\big(\tfrac{\partial M}{\partial(\Lambda,Q)}\big)\Big|\,d\Lambda\,dQ,$$

where dQ is the (right) Haar measure on $O(N)$, and $d\Lambda$ is the Lebesgue measure on $\mathbb{R}^N$ (restricted to $\lambda_1 \le \cdots \le \lambda_N$ if we want an ordering).

A standard result states:

Theorem 3.3 (Jacobian for Spectral Decomposition)**.** *For real-symmetric $M = Q\Lambda Q^T$, one has*

$$\big|\det\big(\tfrac{\partial M}{\partial(\Lambda,Q)}\big)\big| \;=\; \prod_{1\le i<j\le N} \big|\lambda_i - \lambda_j\big|.$$

Remark 3.4. Equivalently, one often writes

$$dM \;=\; \big|\Delta(\lambda_1,\dots,\lambda_N)\big|\,d\Lambda\,dQ, \quad \text{where } \Delta(\lambda_1,\dots,\lambda_N) = \prod_{i<j}(\lambda_j - \lambda_i)$$

is the *Vandermonde determinant*.

Below is one detailed proof, using the idea of "infinitesimal variations" of Q.

Detailed Proof of the Jacobian

We will consider small perturbations of Λ and Q. Write

$$M = Q\,\Lambda\,Q^T, \quad \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N).$$

Let δM be an infinitesimal change in M. We want to see how δM depends on $\delta\Lambda$ and δQ.

Parametrizing δQ. Since $Q \in O(N)$, any small variation of Q can be written as $Q\,\exp(B) \approx Q(I+B)$ where B is an infinitesimal skew-symmetric matrix: $B^T = -B$. Indeed, the $\dim(O(N)) = \frac{N(N-1)}{2}$, matching the dimension of the space of skew-symmetric matrices.

Compute δM. Under an infinitesimal change, say

$$Q \mapsto Q\,(I+B), \quad \Lambda \mapsto \Lambda + \delta\Lambda,$$

we have

$$M = Q\,\Lambda\,Q^T \implies \delta M = Q\,(\delta\Lambda)\,Q^T \;+\; Q\,\Lambda\,(I+B)^T\,Q^T \;-\; Q\,\Lambda\,Q^T$$

to first order in small quantities. Simplify $(I+B)^T = I+B^T = I-B$ because B is skew-symmetric. Thus

$$\delta M = Q\,(\delta\Lambda)\,Q^T + Q\,\Lambda\,(I-B)\,Q^T - Q\,\Lambda\,Q^T = Q\,(\delta\Lambda)\,Q^T + Q\,\Lambda\,(-B)\,Q^T = Q\,(\delta\Lambda)\,Q^T - Q\,\Lambda\,B\,Q^T.$$

So

$$\delta M = Q\,(\delta\Lambda)\,Q^T - Q\,\Lambda\,Q^T\,(Q\,B\,Q^T),$$

since $Q^T Q = I$. But keep in mind that Λ is diagonal, so ΛB is simpler in some sense.

Orthogonal Decomposition of δM. Now we want to separate the part of δM that corresponds to changes in the eigenvalues from the part that corresponds to changes in Q. One can write $\delta\Lambda = \mathrm{diag}(\delta\lambda_1, \ldots, \delta\lambda_N)$. Also note that ΛB is a matrix that has certain off-diagonal structure, since Λ is diagonal but B is skew-symmetric.

If we track the rank-1 changes $\delta\lambda_i$ and the $\frac{N(N-1)}{2}$ parameters in B carefully, one obtains that the Jacobian is precisely the product of all eigenvalue gaps $\lambda_i - \lambda_j$. A fully coordinate-based approach would assign a local parameter system to $O(N)$ near a fixed Q, solve for $\delta\lambda_i$ and the $\frac{N(N-1)}{2}$ independent components of δQ, and then match to the $\frac{N(N+1)}{2}$ differentials in δM. The resulting determinant from that coordinate transformation is the Vandermonde product $\prod_{i<j} |\lambda_i - \lambda_j|$.

One can find many standard treatments of this in random matrix textbooks (e.g., Mehta's *Random Matrices*, Forrester's *Log-Gases and Random Matrices*, or Tao's *Topics in Random Matrix Theory*). This completes the proof of Theorem 3.3.

3.4 Step D: Integration Over $O(N)$ and Final Form of the PDF

Putting Steps A–C together, we find:

$$dM \;=\; \Big(\prod_{i<j} |\lambda_i - \lambda_j|\Big)\, d\Lambda\, \big(\underbrace{\text{Haar measure on } O(N)}_{\text{does not depend on } \lambda_i}\big).$$

Hence, the joint density of $\{\lambda_1, \ldots, \lambda_N\}$ is (up to a global constant):

$$\prod_{i<j} |\lambda_i - \lambda_j| \;\exp\Big(-\tfrac{1}{4\sigma^2} \sum_{k=1}^{N} \lambda_k^2\Big).$$

Finally, there is a constant factor from $\int_{O(N)} dQ$ (the volume of the orthogonal group) and the earlier normalizing Gaussians, yielding the claim:

$$p(\lambda_1, \ldots, \lambda_N) \;=\; \frac{1}{Z_{N,\sigma}} \prod_{i<j} |\lambda_i - \lambda_j| \; \exp\Big(-\tfrac{1}{4\sigma^2} \sum_{k=1}^{N} \lambda_k^2\Big).$$

This completes the detailed proof of the GOE joint eigenvalue distribution.

Remark 3.5 (Ordering of Eigenvalues)**.** Often we incorporate the ordering $\lambda_1 \leq \cdots \leq \lambda_N$ by restricting Λ to the "chamber" $\{\lambda_1 \leq \cdots \leq \lambda_N\}$ and multiplying by $N!$. One can do either approach: the above formula typically assumes ordered eigenvalues and includes a factor $\prod_{i<j} |\lambda_i - \lambda_j|$. The differences are routine normalizing constants.

4 Tridiagonal (Householder) Form for Real-Symmetric Matrices

We now give a step-by-step procedure (and proof) of how any real-symmetric matrix can be orthogonally transformed into a tridiagonal matrix. This is a standard topic in numerical linear algebra (the "Householder reduction") but is also central in random matrix theory (especially the Dumitriu–Edelman approach to the Gaussian ensembles).

4.1 Statement

Theorem 4.1 (Real-Symmetric Tridiagonalization)**.** *Any real-symmetric matrix $A \in \mathbb{R}^{N\times N}$ can be represented as*

$$A = Q^T \, T \, Q, \quad \textit{where } Q \in O(N) \textit{ and } T \textit{ is real-symmetric tridiagonal.}$$

That is, T has nonzero entries only on the main diagonal and the first sub- and super-diagonals:

$$T = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & d_2 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & d_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \alpha_{N-1} \\ 0 & 0 & \cdots & \alpha_{N-1} & d_N \end{pmatrix}.$$

4.2 Householder Reflections: A Detailed Algorithm

Householder Reflection (Definition). A *Householder reflection* in $\mathbb{R}^N$ is a matrix H of the form

$$H = I - 2\,\frac{v\,v^T}{\|v\|^2},$$

where $v \in \mathbb{R}^N$ is nonzero. One can check:

$$H^T = H, \quad H^2 = I, \quad H \text{ is orthogonal, i.e. } H^T H = I.$$

Geometrically, H reflects vectors across the hyperplane orthogonal to v.

Goal. We want to apply successive Householder reflections to "zero out" all sub-subdiagonal (and super-subdiagonal by symmetry) entries of A, leaving only the main diagonal and the first super-/sub-diagonal possibly nonzero.

1. **Start with** $A^{(0)} = A$.

2. **Step** $k = 1$**.** We aim to zero out entries $A^{(0)}_{2,1}, A^{(0)}_{3,1}, \ldots, A^{(0)}_{N,1}$, except for one to remain on the first subdiagonal if needed. Specifically, define the vector

$$x = (A^{(0)}_{2,1},\, A^{(0)}_{3,1},\, \ldots,\, A^{(0)}_{N,1})^T \;\in\; \mathbb{R}^{N-1}.$$

   We want a Householder H_1 such that

$$H_1\, A^{(0)}\, H_1 = A^{(1)}$$

   has zeros in the first column (and row, by symmetry) except possibly $A^{(1)}_{2,1}$.

   Concretely, embed x into $\tilde{x} \in \mathbb{R}^N$ by placing a 0 in the top slot:

$$\tilde{x} = (\,0,\, A^{(0)}_{2,1}, \ldots, A^{(0)}_{N,1}\,)^T.$$

   Choose

$$v = \tilde{x} \;+\; \alpha e_1 \;\in\; \mathbb{R}^N,$$

   with α chosen so that $\|v\| \neq 0$ and $(I - 2vv^T/\|v\|^2)\,\tilde{x}$ is a scalar multiple of e_1. A common choice is

$$\alpha = \pm\,\|\tilde{x}\|,$$

   picking a sign that avoids cancellation. Define

$$H_1 = I - 2\,\frac{v\,v^T}{\|v\|^2}.$$

   Then H_1 is an orthogonal, symmetric matrix that kills the sub-subdiagonal entries in column 1.

3. **Step** $k = 2, \ldots, N-2$**.** Inductively, we zero out the $(k+2)$-th to N-th entries in the k-th column (and by symmetry, in the k-th row). Each step uses a smaller Householder reflection H_k acting nontrivially in the lower-right $(N-k+1) \times (N-k+1)$ submatrix. Then set

$$A^{(k)} = H_k\, A^{(k-1)}\, H_k.$$

4. **End result.** After $N-2$ steps, we get $A^{(N-2)}$, which is tridiagonal, and

$$A^{(N-2)} = (H_{N-2} \cdots H_1)\, A\, (H_1 \cdots H_{N-2}).$$

   Define

$$Q = H_1 \cdots H_{N-2}.$$

   Since each H_k is orthogonal, $Q \in O(N)$. Moreover,

$$A^{(N-2)} = Q\, A\, Q^T$$

   has the desired tridiagonal form.

Remark 4.2. This procedure is also used in numerical methods for eigenvalue computations: once you reduce to tridiagonal form, one can apply specialized algorithms (like the QR algorithm) more efficiently.

Proof of Theorem 4.1. It is essentially just the algorithmic outline above. Each step is valid because Householder transformations preserve symmetry: if B is symmetric, then

$$(HBH)_{ij} = \sum_{r,s} H_{ir} B_{rs} H_{sj}.$$

But since H is symmetric itself, (HBH) remains symmetric. Also, each step zeroes out the sub-subdiagonal entries in the appropriate column and row, thus eventually forcing a tridiagonal shape. Finally, the product of all Householder reflections used is an orthogonal matrix. This completes the argument. □

5 Wigner's Semicircle Law via Tridiagonalization

We now present a *detailed* outline of how one proves the Wigner semicircle law for the GOE by using its *random tridiagonal model*. This method is due to Dumitriu and Edelman (2002) and is often considered more direct than Wigner's original moment method.

5.1 Dumitriu–Edelman Tridiagonal Model

Theorem 5.1 (Tridiagonal Representation of GOE)**.** *Let M be an $N \times N$ GOE matrix (real-symmetric) with variance chosen so that the off-diagonal entries have variance $\frac{1}{2}$ and diagonal entries have variance 1. Then there exists an orthogonal matrix Q such that*

$$M = Q^T \, T \, Q,$$

where T is a real-symmetric tridiagonal matrix of the special form

$$T = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots \\ \alpha_1 & d_2 & \alpha_2 & \ddots \\ 0 & \alpha_2 & d_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

and the random variables $\{d_i, \alpha_j\}$ are mutually independent with

$$d_i \sim \mathcal{N}(0,1), \quad \alpha_j = \sqrt{\frac{\chi^2_{N-j}}{2}},$$

where χ^2_ν is a chi-square distribution with ν degrees of freedom, and equivalently $\sqrt{\frac{\chi^2_\nu}{2}}$ is half the norm of a Gaussian vector in $\mathbb{R}^\nu$.

Remark 5.2. - In short, the diagonal entries d_i are i.i.d. $\mathcal{N}(0,1)$. - The subdiagonal entries $\alpha_1, \ldots, \alpha_{N-1}$ are independent with each α_j distributed like $\sqrt{\frac{\chi^2_{N-j}}{2}}$. - Off-diagonal entries above the first superdiagonal are all zero, so T has only $2N-1$ nontrivial entries (the N diagonal + $(N-1)$ sub-/super-diagonal).

Sketch of Construction. This is essentially a specialized version of the Householder procedure (Section 4), carefully arranged so that each step ends up with exactly the distributions described for α_j and d_i. One uses the fact that a Gaussian matrix is rotationally invariant in a suitable sense, ensuring that each step's "residual vector" has an isotropic Gaussian distribution. Then the norm of that vector yields χ^2 variables. Full details appear in [?DumitriuEdelman2002] or advanced RMT texts. □

Thus, to study the eigenvalues of the GOE matrix M, we can equivalently study the eigenvalues of the (much sparser) tridiagonal matrix T.

5.2 Characteristic Polynomial and Three-Term Recurrence

Consider $p_N(\lambda) = \det(T - \lambda I)$. Since T is tridiagonal, one has the well-known three-term recurrence:

$$p_0(\lambda) := 1, \quad p_1(\lambda) := (d_1 - \lambda),$$
$$p_{k+1}(\lambda) \;=\; (d_{k+1} - \lambda)\, p_k(\lambda) \;-\; \alpha_k^2\, p_{k-1}(\lambda), \quad (k = 1, \ldots, N-1).$$

The roots of $p_N(\lambda)$ are precisely the eigenvalues $\lambda_1, \ldots, \lambda_N$ of T.

5.3 Outline of the Semicircle Limit Proof

We now want to show that the empirical measure

$$L_N \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$$

converges weakly (almost surely) to the semicircle distribution

$$\mu_{\mathrm{sc}}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbf{1}_{|x| \le 2}\, dx.$$

A typical route has these ingredients:

1. **Law of Large Numbers for α_j.** Notice that $\alpha_j^2 = \frac{1}{2}\chi^2_{N-j}$ has mean $\frac{N-j}{2}$. For large N, it is typically of order N. More precisely, $\alpha_j \approx \sqrt{\frac{N-j}{2}}$ in a probabilistic sense as $N \to \infty$.

2. **Scale invariance.** One usually rescales T by $\sqrt{N}$. That is, consider $\frac{1}{\sqrt{N}} T$. Its subdiagonal entries become
$$\frac{\alpha_j}{\sqrt{N}} \;\approx\; \sqrt{\frac{N-j}{2N}} \;\approx\; \sqrt{\frac{1 - j/N}{2}} \quad (\text{for large } N).$$
Meanwhile, the diagonal entries become $\frac{d_i}{\sqrt{N}}$, which are $\mathcal{O}(\frac{1}{\sqrt{N}})$. Hence the subdiagonal terms set the main scale for the "bulk" of the spectrum, while the diagonal is negligible in the large N limit.

3. **Asymptotic Analysis of Recurrence.** A known fact from orthogonal polynomial theory (or from direct PDE-like arguments on the discrete recurrence) is that the location of the roots of $p_N(\lambda)$ concentrate where the effective continuum limit of the recurrence matches a certain "Stieltjes equation" whose solution is the semicircle density.

   In more elementary terms, one can check that the *moment generating function* or *Stieltjes transform* of the measure L_N converges to that of μ_{sc}. Alternatively, one can do a direct argument on the polynomials $p_k(\lambda)$ by bounding their growth and linking it to an integral equation reminiscent of

   $$g(z) = \int \frac{1}{x - z} \, d\mu_{\mathrm{sc}}(x),$$

   which leads to a quadratic equation solved by the semicircle's Cauchy transform.

   For details, see [?DumitriuEdelman2002] or [?TaoTopics], as the full proof is somewhat technical but completely rigorous.

The net result is that, *with probability 1*, as $N \to \infty$, the empirical spectral measure of $\frac{1}{\sqrt{N}} M$ (equivalently of $\frac{1}{\sqrt{N}} T$) converges to the semicircle distribution on $[-2, 2]$:

$$\mu_{\mathrm{sc}}(dx) \;=\; \frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbf{1}_{|x| \le 2} \, dx.$$

This is precisely *Wigner's semicircle law.*

Remark 5.3 (Extensions)**.** A very similar approach works for the Gaussian Unitary Ensemble ($\beta = 2$), yielding a random *complex Hermitian* tridiagonal (or banded) matrix. And for $\beta = 4$, there is an analogous construction with quaternionic entries, usually leading to a block-tridiagonal matrix. All roads lead to the semicircle law for the limiting global spectrum.

6 Eigenvalue Distributions for Classical Ensembles

We begin by studying eigenvalue distributions for the three fundamental classes of random matrices. These distributions arise from matrices with different symmetry properties and correspond to the real, complex, and quaternionic cases.

6.1 Matrix Ensembles with Different Symmetries

Let X be an $N \times N$ matrix. We consider three cases of random matrices with i.i.d. matrix elements:

a) **Real case:** $X_{ij} \sim \mathcal{N}(0,1)$

b) **Complex case:** $X_{ij} \sim \mathcal{N}(0,1) + i\mathcal{N}(0,1)$

c) **Quaternion case:** $X_{ij} \sim \mathcal{N}(0,1) + i\mathcal{N}(0,1) + j\mathcal{N}(0,1) + k\mathcal{N}(0,1)$

For each case, we form a self-adjoint matrix:

$$M = \frac{1}{2}(X + X^*)$$

where X^* denotes the appropriate adjoint. This construction ensures real eigenvalues and proper spectral properties.

Theorem 6.1 (Joint Eigenvalue Distribution)**.** *The eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ of the matrix M have joint probability density:*

$$\frac{1}{Z} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^{N} \exp\left(-\frac{\lambda_i^2}{2}\right)$$

where:

- $\beta = 1, 2, 4$ *for cases (a), (b), (c) respectively*
- Z *is the normalization constant given by:*

$$Z = \frac{(2\pi)^{N/2}}{N!} \prod_{j=1}^{N-1} \frac{\Gamma(1 + \beta(j+1)/2)}{\Gamma(1 + \beta/2)}$$

This density is often called the "multivariate Gaussian" distribution in this context.

6.2 Proof Strategy

We will prove this theorem for $\beta = 1$ (the real case) and outline the modifications needed for other cases. The proof proceeds in three main steps.

Step 1: Matrix Density. The probability density of the matrix M is proportional to:

$$\exp\left(-\frac{1}{2}\operatorname{Tr}(M^2)\right)$$

Indeed, we can expand the trace:

$$\operatorname{Tr}(M^2) = \sum_{i,j} |M_{ij}|^2 = \sum_{i=1}^{N} M_{ii}^2 + 2\sum_{i<j} |M_{ij}|^2$$

Each element of M is formed from the corresponding elements of X according to the self-adjointness condition. $\square$

Step 2: Eigenvalue Transformation. Using the spectral decomposition $M = ODO^*$ where D is diagonal with eigenvalues λ_i and O is orthogonal/unitary/symplectic (depending on β), we have:

$$\exp\left(-\frac{1}{2}\operatorname{Tr}(M^2)\right) = \exp\left(-\frac{1}{2}\sum_{i=1}^{N} \lambda_i^2\right) = \prod_{i=1}^{N} \exp\left(-\frac{\lambda_i^2}{2}\right)$$

$\square$

Step 3: Jacobian Calculation. The key step is computing the Jacobian of the transformation from matrix elements to eigenvalues and eigenvectors. Consider the map:

$$\Pi : W_N \times \mathcal{G}(N) \to \mathfrak{sl}_N$$

where:

- W_N is the space of diagonal matrices with ordered eigenvalues
- $\mathcal{G}(N)$ is $O(N)$, $U(N)$, or $Sp(N)$ depending on β
- $\mathfrak{sl}_N$ is the space of self-adjoint matrices

This map is given by:

$$(\lambda, g) \mapsto g \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix} g^*$$

Near the identity element of $\mathcal{G}(N)$, we can write:

$$g = \exp(B) \approx I + B + \frac{B^2}{2} + \cdots$$

where B is skew-symmetric/skew-Hermitian/skew-quaternionic.

The Jacobian computation yields:

$$\prod_{i<j} |\lambda_i - \lambda_j|^\beta$$

which explains the appearance of this term in the joint density. $\square$

7 Laguerre/Wishart Ensemble

Consider a matrix X of size $N \times M$ with $N < M$ having singular value decomposition:

$$X = U \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_N \end{pmatrix} V$$

where U and V are orthogonal/unitary/symplectic matrices of appropriate sizes.

Theorem 7.1 (Wishart Distribution)**.** *Let X be an $N \times M$ matrix with i.i.d. Gaussian elements as in Theorem 6.1. Then the eigenvalues $\lambda_i = s_i^2$ of XX^* have joint density proportional to:*

$$\prod_{i<j} (\lambda_i - \lambda_j)^\beta \prod_{i=1}^{N} \lambda_i^{\frac{\beta}{2}(M-N+1)-1} e^{-\lambda_i/2}$$

This is known as the "multivariate Γ-distribution."

8 Jacobi/MANOVA/CCA Ensemble

Consider two rectangular arrays:

$$\begin{aligned} &X : N \times T \\ &Y : K \times T \qquad N \le K \le T \end{aligned}$$

Define:

- P_X = projector onto N-dimensional subspace spanned by rows of X
- P_Y = projector onto K-dimensional subspace spanned by rows of Y

The squared canonical correlations are $\min(N, K)$ non-zero eigenvalues of $P_X P_Y$.

Theorem 8.1 (Canonical Correlations)**.** *Assume X and Y are independent with i.i.d. Gaussian elements. Then the eigenvalues of $P_X P_Y$ have density proportional to:*

$$\prod_{i<j} (\lambda_i - \lambda_j)^\beta \prod_{i=1}^{N} \lambda_i^{\frac{\beta}{2}(K-N+1)-1} (1 - \lambda_i)^{\frac{\beta}{2}(T-K+1)-1}$$

where $0 \le \lambda_i \le 1$. This is the "multivariate Beta distribution."

9 General Pattern

A remarkable feature emerges across these classical ensembles. The eigenvalue distributions consistently take the form:

$$\prod_{i<j} |\lambda_j - \lambda_i|^\beta \prod_{i=1}^{N} V(\lambda_i)$$

where:

- The first term represents logarithmic pairwise interaction
- $V(\lambda)$ is an appropriate potential function
- β represents the symmetry class (1, 2, or 4)

This structure appears in various contexts in random matrix theory and is often referred to as a "log-gas" or "β-ensemble" system.

C Problems (due 2025-02-22)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)
Lecture 4: Title TBD

Leonid Petrov

DATE, 2025*

D Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 5: Title TBD

Leonid Petrov

DATE, 2025*

E Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 6: Title TBD

Leonid Petrov

DATE, 2025*

F Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 7: Title TBD

Leonid Petrov

DATE, 2025*

G Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 8: Title TBD

Leonid Petrov

DATE, 2025*

H Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 9: Title TBD

Leonid Petrov

DATE, 2025*

I Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 10: Title TBD

Leonid Petrov

DATE, 2025*

J Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 11: Title TBD

Leonid Petrov

DATE, 2025*

K Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:45, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 12: Title TBD

Leonid Petrov

DATE, 2025*

L Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:45, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 13: Title TBD

Leonid Petrov

DATE, 2025*

M Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:45, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 14: Title TBD

Leonid Petrov

DATE, 2025*

N Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:45, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 15: Title TBD

Leonid Petrov

DATE, 2025*

O Problems (due DATE)

References

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA
E-mail: lenia.petrov@gmail.com

*Course webpage • TeX Source • Updated at 13:45, Wednesday 15th January, 2025