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

Leonid Petrov

Monday, January 13, 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.

---

[*]Course webpage ● TeX Source ● Updated at 02:08, Tuesday 14th January, 2025

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.
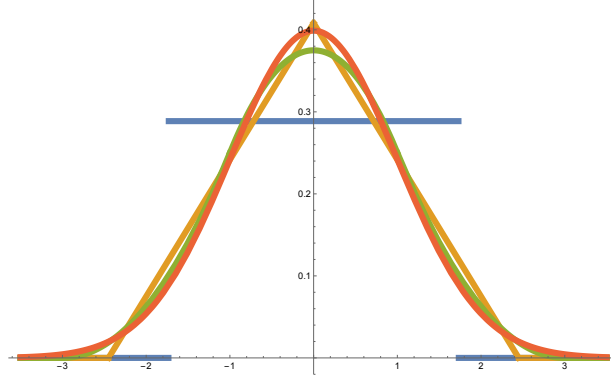

Figure 1: Densities of U_1, $U_1 + U_2$, $U_1 + U_2 + U_3$ (where U_i are iid uniform on $[0, 1]$), and $\mathcal{N}(0, 1)$, normalized to have the same mean and variance.

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:

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

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

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

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:

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

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 $(n - k)!$:

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

3

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

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

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. $\qquad\square$

4

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:

$$\left(\sum_{i=1}^n Y_i\right)^k = \sum_{j_1+j_2+\cdots+j_n=k} \frac{k!}{j_1!j_2!\ldots j_n!} Y_1^{j_1} Y_2^{j_2} \ldots Y_n^{j_n}.$$

Taking the expectation and using linearity, we have:

$$\mathbb{E}\left[\left(\sum_{i=1}^n Y_i\right)^k\right] = \sum_{j_1+j_2+\cdots+j_n=k} \mathbb{E}\left[Y_1^{j_1}\right] \mathbb{E}\left[Y_2^{j_2}\right] \ldots \mathbb{E}\left[Y_n^{j_n}\right]$$

$$= \sum_{j_1+j_2+\cdots+j_n=k} \mathsf{m}_{j_1} \ldots \mathsf{m}_{j_n},$$

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

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 $\leq n$ 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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$), we must count the number of distinct multisets of indices $(j_1, j_2, \ldots, j_n)$ that yield the same collection $\{\lambda_1, \lambda_2, \ldots\}$. Let us start by 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) 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 $(2, 2)$ 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, but 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:

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

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 $\mathbb{R}$ (cf. Appendix A.2).

In general, convergence of moments does *not* necessarily imply convergence in distribution. Indeed, the classical method of moments in probability theory [Bil95, Theorem 30.2] states that if a sequence of random variables X_n has moments (of all orders) converging to the moments of a random variable X, then X_n converges in distribution to X, provided that the distribution of X is uniquely determined by its moments (also known as being moment-determinate).

The normal distribution is indeed uniquely determined by its moments, 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:

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

   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.

---

[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 $A^\top J A = J$, where J is the skew-symmetric form.

3.2 Real Wigner matrices

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

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

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

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 $(Y + Y^\top)/2$, 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 ($n \le m$), 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 entries having a fixed variance, independent of n). Then as $n \to \infty$, the ESD $W/\sqrt{n}$ 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:

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

Remark 3.5. The convergence in (3.4) is *weakly in probability*, that is, 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.

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 $k \geq 1$, the normalized k-th moment of the empirical spectral distribution of $W_n/\sqrt{n}$ is given by

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

Our first goal is to study the asymptotic behavior of $\mathbb{E}[m_k^{(n)}]$ as $n \to \infty$ for each fixed $k \geq 1$, 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:

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

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 Appendix A.8.

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 Appendix A.9, 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. $\qquad\square$

A   Problems (due 2025-02-13)

Each problem is a subsection (like 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:

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

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):

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

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 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.6 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.7 Invariance of the GOE

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

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

for all orthogonal matrices O and Borel sets A.

A.8 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.9 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, 6

[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. ↑1

[Dys62a] F.J. Dyson, *A Brownian motion model for the eigenvalues of a random matrix*, Jour. Math. Phys. **3** (1962), no. 6, 1191–1198. ↑1

[Dys62b] F.J. Dyson, *Statistical Theory of the Energy Levels of Complex Systems. I-III*, Jour. Math. Phys. **3** (1962), no. 166, 140–175. ↑1

[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. ↑1

[MP67] V. Marchenko and L. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb. **1** (1967), 457–483. ↑9

[Wig55] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics (1955), 548–564. ↑1

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