

Lectures on Random Matrices (Spring 2025)

Lecture 1: Moments of random variables and random matrices

Leonid Petrov

Monday, January 13, 2025*

Notes for the lecturer

“If you find something nice or beautiful or efficient in RMT or computing/numerics around RMT, please share with me.”

Course Meetings:

- Classes are scheduled for 2:00-3:15 PM in Kerchof 326 on Wednesdays and some Mondays.
- Lecture notes, including problem sets, will be available online.

Lecture Schedule:

- (Mon) January 13: Moments of random variables and matrices.
- (Wed) January 15: Wigner’s semicircle law.
- (Wed) January 22: Eigenvalue densities, Dyson Brownian motion.
- Additional lectures are scheduled for the following dates:
 - (Wed) January 29
 - (Wed) February 5
 - (Wed) February 12
 - (Wed) February 19
 - (Wed) February 26
 - (Wed) March 5
 - (Mon) March 24
 - (Wed) March 26
 - (Wed) April 2
 - (Wed) April 9

*[Course webpage](#) • [TeX Source](#) • Updated at 15:37, Sunday 12th January, 2025

- (Wed) April 16
- (Wed) April 23

Student Presentations:

- To be held on Mondays in April:
 - (Mon) April 7
 - (Mon) April 14
 - (Mon) April 21
 - (Mon) April 28 (tentative - dependent on student enrollment)

Weekly Individual Meetings:

- The course covers a span of 12 full weeks, excluding the first week, the final week, and the week following Spring break, during which I will be traveling.
- Students will have one-on-one meetings at least 8 times during this period, each lasting approximately 45 minutes. While in-person meetings are preferred, a Zoom option is available.
- These meetings are vital to the reading course component, allowing discussions on lectures, homework problems, student presentations, and research topics related to random matrices.

During the first week, a regular weekly meeting time will be set up for each student. Consistency throughout the semester is highly encouraged, although some flexibility is possible.

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 $\{X_1, X_2, \dots, X_n\}$, the joint distribution factors as the product of the individual distributions:

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

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 $\sigma^2 = \text{Var}(X_i)$. Define the normalized sum*

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

Then, as $n \rightarrow \infty$, 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

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

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 .

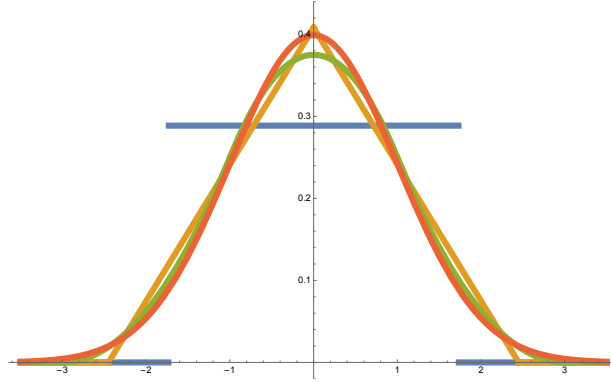


Figure 1: Densities of U_1 , $U_1 + U_2$, $U_1 + U_2 + U_3$ (where U_i are iid uniform on $[0, 1]$), and $\mathcal{N}(0, 1)$, normalized to have the same mean and variance.

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 $X_1 + \dots + X_n$:

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

Introduce the normalized quantity

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

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 $n \rightarrow \infty$, 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

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

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

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

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:*

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

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. \square

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:

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

Taking the expectation and using linearity, we have:

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{i=1}^n Y_i \right)^k \right] &= \sum_{j_1+j_2+\dots+j_n=k} \mathbb{E} [Y_1^{j_1}] \mathbb{E} [Y_2^{j_2}] \dots \mathbb{E} [Y_n^{j_n}] \\ &= \sum_{j_1+j_2+\dots+j_n=k} \mathbf{m}_{j_1} \dots \mathbf{m}_{j_n}, \end{aligned}$$

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

The sum over all j_1, \dots, j_n with $j_1 + \dots + j_n = k$ 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 $k = \lambda_1 + \lambda_2 + \dots + \lambda_n$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$), we must count the number of distinct multisets of indices (j_1, j_2, \dots, j_n) that yield the same collection $\{\lambda_1, \lambda_2, \dots\}$. 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 $\lambda = (2, 2, \dots, 2)$, 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 $n \rightarrow \infty$. When k is odd, the “best” we can do (without parts equal to 1) is going to be $\lambda = (3, 2, \dots, 2)$ with $(k-1)/2$ parts, so the power of n is $n^{(k-1)/2}$. This is also subleading in the limit $n \rightarrow \infty$.

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 $n \rightarrow \infty$:

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

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

3.1 Real Wigner matrices

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:

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

A.2 Convergence in distribution

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

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

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

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

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.

References

- [Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. [↑6, 8](#)
- [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>. [↑6](#)
- [Hur97] A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachr. Ges. Wiss. Göttingen (1897), 71–90. [↑2](#)
- [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