Lectures 5 and 6: Wick's Theorem and the Semicircle Law

Pair partitions, the spectrum of a random matrix, and two proofs of the same theorem

---

OUTLINE Wick's theorem $\longrightarrow$ the Gaussian Wigner matrix $\longrightarrow$ $\|Mu\|^2$ for a fixed direction

The empirical spectral measure $\longrightarrow$ traces are moments $\longrightarrow$ $\mathbb{E}\operatorname{tr} M^2$ and $\mathbb{E}\operatorname{tr} M^4$ $\longrightarrow$ non-crossing pair partitions $\longrightarrow$ the Catalan count $\longrightarrow$ the semicircle law

HOW TO READ Sections marked $*$ are not lectured. They are complete, they are examinable, and they carry most of the exercises. Exercises are typed A, B and C; each week's quiz has one of each.

1 Wick's theorem, and Gaussian random matrices

Q. What can be computed about a random matrix without touching its eigenvalues?

Every moment of a Gaussian vector is a counting problem, and the entries of a symmetric Gaussian matrix are a Gaussian vector. So anything polynomial in the entries — the length of Mu for a fixed vector u, for instance — can be computed exactly.

1.1 Pair partitions

Definition 1.1 (Pair partition)**.** A *pair partition* of $\{1,\dots,2q\}$ is a partition into q blocks of size two.

Theorem 1.2 (Wick)**.** *Let $X \sim \mathcal{N}(0,\Sigma)$ on $\mathbb{R}^d$ and $a_1,\dots,a_m \in \{1,\dots,d\}$. If m is odd then $\mathbb{E}[X_{a_1}\cdots X_{a_m}] = 0$. If $m = 2q$ then*

$$\boxed{\mathbb{E}\big[X_{a_1}X_{a_2}\cdots X_{a_{2q}}\big] \;=\; \sum_{\text{pair partitions } P \text{ of } \{1,\dots,2q\}} \;\prod_{\{r,s\}\in P} \Sigma_{a_r a_s}}$$

Proof. By Corollary 1.7 of Lectures 1–2 every linear functional of X is a one-dimensional Gaussian, so X has moments of every order, φ_X is infinitely differentiable, and its derivatives may be taken inside the expectation. Since $\partial_{\xi_j} e^{-i\langle x,\xi\rangle} = -ix_j e^{-i\langle x,\xi\rangle}$, multiplication by X_j acts on the transform as $i\,\partial_{\xi_j}$:

$$\mathbb{E}\big[X_{a_1}\cdots X_{a_m}\big] \;=\; i^m\,\partial_{\xi_{a_1}}\cdots\partial_{\xi_{a_m}}\varphi_X(\xi)\big|_{\xi=0}. \tag{1}$$

By Definition 1.3 of Lectures 1–2, $\varphi_X(\xi) = \exp\big(-\tfrac12\xi^\top\Sigma\xi\big)$ — for every $\Sigma \succeq 0$, degenerate cases included — and differentiating it once,

$$\partial_{\xi_i}\varphi_X(\xi) \;=\; -\sum_j \Sigma_{ij}\,\xi_j\,\varphi_X(\xi). \tag{2}$$

Apply $\partial_{\xi_{a_2}}\cdots\partial_{\xi_{a_m}}$ to (2) taken at $i = a_1$. The factor ξ_j is linear, so Leibniz leaves terms of two kinds: those in which every derivative falls on φ_X, which still carry a factor ξ_j; and, for each $s \in \{2,\dots,m\}$, the one in which $\partial_{\xi_{a_s}}$ falls on ξ_j and produces δ_{ja_s}. Setting $\xi = 0$ kills the first kind, leaving

$$\partial_{\xi_{a_1}}\cdots\partial_{\xi_{a_m}}\varphi_X\big|_0 \;=\; -\sum_{s=2}^m \Sigma_{a_1a_s}\;\partial_{\{\xi_{a_r}:\,r\neq 1,s\}}\varphi_X\big|_0.$$

Rewrite both sides with (1) and multiply through by i^m; since $i^2 = -1$ the two minus signs cancel and

$$\mathbb{E}\big[X_{a_1}\cdots X_{a_m}\big] \;=\; \sum_{s=2}^{m} \Sigma_{a_1 a_s}\; \mathbb{E}\Big[\prod_{r\neq 1,s} X_{a_r}\Big] : \tag{3}$$

pair the first factor with each of the others in turn, and recurse. Induct on m. Both sides vanish for $m = 1$, and for $m = 2$ the claim is the definition of Σ. For $m > 2$, (3) drops the number of factors by two, so parity is preserved: an odd product reduces to $\mathbb{E}[X_a] = 0$, and an even one to a sum over pair partitions of the remaining $2q - 2$ indices, each prefixed by $\Sigma_{a_1 a_s}$ recording the block containing 1. Summing over s enumerates every pair partition of $\{1, \ldots, 2q\}$ once. ■

Remark 1.3. Reading (3) with $F(X) = X_{a_2}\cdots X_{a_m}$ and noting that $\partial_j F$ deletes the factor X_j gives

$$\mathbb{E}\big[X_i\, F(X)\big] \;=\; \sum_j \Sigma_{ij}\, \mathbb{E}\big[\partial_j F(X)\big],$$

Gaussian integration by parts, which for $d = 1$ and $\Sigma = 1$ is $\mathbb{E}[gF(g)] = \mathbb{E}[F'(g)]$ — Lemma 1.10 of Lectures 1–2, and Stein's lemma in Lecture 4. It is Exercise 2.3 of Lectures 3–4; here it falls out of Wick's recursion for free.

1.2 Three examples

Corollary 1.4. *There are* $(2q-1)!! = (2q-1)(2q-3)\cdots 3\cdot 1$ *pair partitions of* $2q$ *points, and* $\mathbb{E}[g^{2q}] = (2q-1)!!$ *for* $g \sim \mathcal{N}(0,1)$.

Proof. Take $d = 1$ and $\Sigma = 1$ in Theorem 1.2: every product of covariances equals 1, so the moment is the number of pair partitions. That number is $(2q-1)!!$ because the point 1 has $2q-1$ choices of partner and deleting that block leaves $2q - 2$ points. ■

Example 1.5 (Four indices)**.** For $q = 2$ the three pair partitions of $\{1, 2, 3, 4\}$ are $\{12\}\{34\}$, $\{13\}\{24\}$, $\{14\}\{23\}$, so

$$\mathbb{E}\big[X_1X_2X_3X_4\big] = \Sigma_{12}\Sigma_{34} + \Sigma_{13}\Sigma_{24} + \Sigma_{14}\Sigma_{23}.$$

Setting all four indices equal recovers $\mathbb{E}[X_1^4] = 3\Sigma_{11}^2$.

Example 1.6 (A quadratic form)**.** Let $X \sim \mathcal{N}(0,\Sigma)$ on $\mathbb{R}^d$ and let A be symmetric. Then $\mathbb{E}[X^\top AX] = \sum_{ij} A_{ij}\mathbb{E}[X_iX_j] = \mathrm{tr}(A\Sigma)$, no Wick needed. For the second moment,

$$\mathbb{E}\big[(X^\top AX)^2\big] = \sum_{ijkl} A_{ij}A_{kl}\,\mathbb{E}\big[X_iX_jX_kX_l\big] = \sum_{ijkl} A_{ij}A_{kl}\big(\Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}\big)$$

by Example 1.5. The first term is $(\mathrm{tr}\, A\Sigma)^2$; the other two are each $\mathrm{tr}(A\Sigma A\Sigma)$, by symmetry of A and Σ. Hence

$$\mathrm{Var}\,\big(X^\top AX\big) = 2\,\mathrm{tr}\,\big((A\Sigma)^2\big), \qquad \text{and in particular} \qquad \mathrm{Var}\,\big(\|X\|^2\big) = 2\,\mathrm{tr}\,\Sigma^2.$$

1.3 The Gaussian Wigner matrix

Definition 1.7 (Gaussian Wigner matrix)**.** Let A be $N \times N$ with i.i.d. $\mathcal{N}(0,1)$ entries and put

$$M = \frac{1}{\sqrt{2N}}\big(A + A^\top\big).$$

Then M is symmetric, its entries $\{M_{ij}\}_{i \le j}$ are independent centred Gaussians with $\mathbb{E}M_{ij}^2 = \frac{1}{N}$ for $i < j$ and $\mathbb{E}M_{ii}^2 = \frac{2}{N}$, and for *all* four indices

$$\mathbb{E}\big[M_{ij}M_{kl}\big] = \tfrac{1}{N}\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big). \tag{4}$$

This ensemble is the *Gaussian orthogonal ensemble*, or GOE, for the reason in Remark 1.8. Everything below is (4) and Theorem 1.2; the construction itself never appears again. The two delta terms are the two ways of matching $\{i,j\}$ with $\{k,l\}$, and they are there because $M_{ij} = M_{ji}$: an entry may be met twice under either name. The diagonal variance $\frac{2}{N}$ comes out of the symmetrization (Exercise 1.9), and it is exactly what makes (4) hold at $i = j = k = l$ without a special case. The N diagonal entries are a vanishing fraction of the N^2.

Remark 1.8 (Why "orthogonal"). For any fixed orthogonal O, $OAO^\top$ again has i.i.d. $\mathcal{N}(0,1)$ entries — each entry is a unit linear functional of A, and Theorem 1.5 of Lectures 1–2 makes the collection jointly Gaussian with the same covariance — so $OMO^\top \overset{d}{=} M$. The law of the spectrum therefore cannot depend on the basis. This is the matrix form of the rotational invariance of Corollary 1.6 of Lectures 1–2.

Exercise 1.9 (TYPE A)**.** Verify Definition 1.7 from the construction: with A having i.i.d. $\mathcal{N}(0,1)$ entries and $M = \frac{1}{\sqrt{2N}}(A + A^\top)$, compute $\mathbb{E}[M_{ij}M_{kl}]$ directly and obtain (4). Check the case $i = j = k = l$ separately and confirm that the diagonal convention $\mathbb{E}M_{ii}^2 = \frac{2}{N}$ is forced, not chosen.

1.4 One direction at a time

Before the eigenvalues, a question with the same shape but no linear algebra in it: fix a unit vector u and ask how long Mu is.

Proposition 1.10. *Let M be as in Definition 1.7 and let $u \in \mathbb{R}^N$ with $\|u\| = 1$. Then Mu is a centred Gaussian vector with covariance*

$$C \;=\; \mathbb{E}\big[(Mu)(Mu)^\top\big] \;=\; \frac{1}{N}\big(I + uu^\top\big),$$

and consequently

$$\mathbb{E}\big\|Mu\big\|^2 = 1 + \frac{1}{N}, \qquad \operatorname{Var}\big\|Mu\big\|^2 = \frac{2}{N} + \frac{6}{N^2}.$$

Proof. Each coordinate $(Mu)_i = \sum_j M_{ij}u_j$ is a linear functional of the entries of M, so by Theorem 1.5 of Lectures 1–2 the vector Mu is centred Gaussian. Its covariance is (4) contracted with u twice:

$$\mathbb{E}\big[(Mu)_i(Mu)_k\big] = \sum_{j,l} u_j u_l \cdot \tfrac{1}{N}\big(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\big) = \tfrac{1}{N}\big(\delta_{ik}\|u\|^2 + u_i u_k\big),$$

which is C. Then $\mathbb{E}\|Mu\|^2 = \operatorname{tr} C = \frac{1}{N}(N+1)$. For the variance, Example 1.6 with $A = I$ and $\Sigma = C$ gives $\operatorname{Var}\|Mu\|^2 = 2\operatorname{tr} C^2$, and since $u^\top u = 1$,

$$\big(I + uu^\top\big)^2 = I + 2uu^\top + u(u^\top u)u^\top = I + 3uu^\top, \qquad \operatorname{tr} C^2 = \tfrac{1}{N^2}\big(N+3\big). \qquad \blacksquare$$

Remark 1.11. Two things are worth reading off. First, the answer does not depend on u — it cannot, because $OMO^\top \stackrel{d}{=} M$ by Remark 1.8, so the law of $\|Mu\|^2$ is the same in every direction. Second, the standard deviation is $\sqrt{2/N}$ against a mean of 1: for large N a Gaussian Wigner matrix maps *every* unit vector to a vector of length very close to 1. That is already a statement about eigenvalues, since $\|Mu\|^2 = u^\top M^2 u$ is a weighted average of the λ_i^2 with weights $\langle u, \phi_i\rangle^2$ summing to 1, where ϕ_i are the eigenvectors. Averaging over directions instead of fixing one leads to $\frac{1}{N}\operatorname{tr} M^2$, which is taken up in §2.3.

Exercise 1.12 (TYPE A)**.** Let $u, v \in \mathbb{R}^N$ be unit vectors. Show from (4) alone that $u^\top M v$ is centred Gaussian with

$$\operatorname{Var}\big(u^\top M v\big) = \frac{1}{N}\big(1 + \langle u, v\rangle^2\big), \qquad \text{and} \qquad \operatorname{Cov}\big(u^\top M u,\, v^\top M v\big) = \frac{2}{N}\,\langle u, v\rangle^2.$$

In particular $u^\top M u$ has variance $\frac{2}{N}$, and for $u \perp v$ the variable $u^\top M v$ has variance $\frac{1}{N}$: the diagonal and off-diagonal variances of Definition 1.7, seen in any orthonormal basis containing u and v. Remark 1.8 predicted exactly this.

Exercise 1.13 (TYPE B)**.** Let $u, v \in \mathbb{R}^N$ be unit vectors. Show that

$$\operatorname{Cov}\big(\|Mu\|^2, \|Mv\|^2\big) = \frac{2}{N^2}\Big(1 + (N+2)\langle u, v\rangle^2\Big),$$

and check that $u = v$ recovers Proposition 1.10. *Mu and Mv are jointly Gaussian; compute their cross-covariance $\mathbb{E}[(Mu)(Mv)^\top]$ from (4), then apply Theorem 1.2 to $\mathbb{E}[(Mu)_i^2(Mv)_k^2]$.* For $u \perp v$ the correlation of the two lengths is $\frac{1}{N+3}$: a Gaussian Wigner matrix stretches orthogonal directions almost independently.

1.5 $*$ Cumulants

Not lectured. Theorem 1.2 is the Gaussian instance of a formula that holds for any random vector with moments. This section states that formula; the exercises develop the properties that make it useful.

Definition 1.14 (Joint cumulants)**.** For $Y_1, \dots, Y_m$ with all moments finite,

$$\kappa\big(Y_1, \dots, Y_m\big) \;=\; \frac{\partial^m}{\partial t_1 \cdots \partial t_m} \log \mathbb{E}\big[e^{t_1 Y_1 + \cdots + t_m Y_m}\big]\bigg|_{t=0}.$$

Moments are derivatives of the transform; cumulants are derivatives of its logarithm. For a single variable write $\kappa_m(Y) = \kappa(Y, \dots, Y)$.

Stated without proof: the moment–cumulant formula. For any $Y_1, \dots, Y_m$ with moments,

$$\mathbb{E}\big[Y_1 Y_2 \cdots Y_m\big] \;=\; \sum_{\pi \in \mathcal{P}(m)} \prod_{B \in \pi} \kappa\big(Y_i : i \in B\big),$$

where $\mathcal{P}(m)$ is the set of *all* partitions of $\{1, \dots, m\}$ into non-empty blocks, of which there are $1, 2, 5, 15, 52, \dots$

For $m = 2$ this reads $\mathbb{E}[Y_1 Y_2] = \mathbb{E}Y_1 \mathbb{E}Y_2 + \operatorname{Cov}(Y_1, Y_2)$, the two partitions of $\{1, 2\}$. If the Y_i are coordinates of a centred Gaussian vector, every joint cumulant of order other than two vanishes

(Exercise 1.16), so a partition contributes nothing unless all of its blocks have size two. Those are the pair partitions, and the moment–cumulant formula becomes Theorem 1.2. Of the fifteen partitions of a four-element set, three survive, which is Example 1.5.

Remark 1.15. For a non-Gaussian vector the larger blocks do not vanish, $\log \mathbb{E}[e^{\langle t,X\rangle}]$ is no longer a polynomial, and the recursion (3) acquires a tail that never terminates. Cumulants then measure the distance from Gaussian, and because mixed cumulants of independent variables vanish, they are the natural bookkeeping for sums of independent variables.

Exercise 1.16 (TYPE B)**.** Prove the following directly from Definition 1.14.

(a) **Low orders.** $\kappa(Y) = \mathbb{E}Y$ and $\kappa(Y_1, Y_2) = \mathrm{Cov}(Y_1, Y_2)$.

(b) **Multilinearity.** κ is symmetric in its arguments and linear in each: $\kappa(aY+bZ, Y_2, \dots, Y_m) = a\,\kappa(Y, Y_2, \dots, Y_m) + b\,\kappa(Z, Y_2, \dots, Y_m)$. For $m \geq 2$ it is unchanged when a constant is added to any argument. In particular $\kappa_m(cY) = c^m \kappa_m(Y)$.

(c) **Independence.** If $\{Y_1, \dots, Y_m\}$ splits into two non-empty groups independent of one another, then $\kappa(Y_1, \dots, Y_m) = 0$. Deduce $\kappa_m(Y+Z) = \kappa_m(Y) + \kappa_m(Z)$ for independent Y, Z.

(d) **The Gaussian is where they stop.** For $X \sim \mathcal{N}(0, \Sigma)$, $\kappa(X_a, X_b) = \Sigma_{ab}$ and every joint cumulant of order $\neq 2$ vanishes.

(e) For centred Y, $\kappa_3(Y) = \mathbb{E}[Y^3]$ and $\kappa_4(Y) = \mathbb{E}[Y^4] - 3(\mathbb{E}[Y^2])^2$.

For (b), give Y and Z separate variables s, s' in the log-transform and apply the chain rule once. For (d), use Definition 1.3 of Lectures 1–2 with $\xi = it$.

Exercise 1.17 (TYPE C)**.** Let $X_1, X_2, \dots$ be i.i.d. with mean μ, variance σ^2 and all moments finite, and put

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(X_i - \mu\big).$$

Show that $\kappa_1(S_n) = 0$ and

$$\kappa_m(S_n) = n^{1-m/2}\,\kappa_m(X_1) \qquad (m \geq 2).$$

Deduce that $\kappa_2(S_n) = \sigma^2$ for every n while every cumulant of order ≥ 3 tends to 0, and conclude from the moment–cumulant formula that $\mathbb{E}[S_n^m] \to \mathbb{E}[(\sigma g)^m]$ for every m, where $g \sim \mathcal{N}(0,1)$.

This is the central limit theorem seen through cumulants: the $\sqrt{n}$ normalization is exactly the rescaling that freezes κ_2 and kills everything above it. The factor $n^{-1/2}$ in κ_3 is the rate in Theorem 2.6 of Lectures 3–4.

ASSIGNED **Lecture 5**, due at the start of Lecture 6: Exercises 1.9, 1.12, 1.13, 1.16, 1.17 — two A, two B, one C. *If you are short of time, do the* C *problem.*

2 The empirical spectral measure and the semicircle law

Q. Where are the eigenvalues of a large random symmetric matrix?

A trace of a power is at once a moment of the spectrum and a polynomial in the entries, and Theorem 1.2 computes the latter. What remains is to say what object these numbers are moments *of*, to compute the first two, and then to do every power at once.

2.1 Where the eigenvalues are: the empirical measure

M is symmetric, so it has N real eigenvalues $\lambda_1, \dots, \lambda_N$, listed with multiplicity. We want to say where they are, and the object that says it is a measure rather than a function.

Write $\langle \mu, f \rangle = \int f \, d\mu$ for a measure μ and a bounded continuous f.

Definition 2.1 (Point mass)**.** For $x \in \mathbb{R}$, the *point mass* δ_x is the probability measure determined by

$$\langle \delta_x, f \rangle = f(x) \qquad \text{for every bounded continuous } f.$$

It puts mass 1 at the single point x and none anywhere else. It has no density: there is no function p with $\int fp = f(x)$ for all f, since such a p would have to vanish off $\{x\}$ and still integrate to 1. Measures are the right objects here precisely because we want to place mass at points.

Definition 2.2 (Empirical spectral measure)**.** For a symmetric $N \times N$ matrix M with eigenvalues $\lambda_1, \dots, \lambda_N$,

$$\hat{\rho}_M = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}, \qquad \text{so} \qquad \langle \hat{\rho}_M, f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i).$$

It is a probability measure: total mass $\frac{1}{N} \cdot N = 1$. It records the eigenvalues and nothing else — not which eigenvector goes with which, not the order.

Now let M be random. Then $\hat{\rho}_M$ is a *random measure*: for each fixed f, the number $\langle \hat{\rho}_M, f \rangle$ is a random variable. There is one sensible way to average it.

Definition 2.3 (Mean measure)**.** $\mathbb{E}\hat{\rho}$ is the measure determined by

$$\langle \mathbb{E}\hat{\rho}, f \rangle \;=\; \mathbb{E} \, \langle \hat{\rho}_M, f \rangle \qquad \text{for every bounded continuous } f.$$

The left side is a number attached to a deterministic measure; the right side is the expectation of a random variable. Nothing is being integrated twice — the definition simply says that averaging a measure means averaging what it does to each test function.

Lemma 2.4. *$\mathbb{E}\hat{\rho}$ is the law of λ_J, where J is uniform on $\{1, \dots, N\}$ and independent of M: it is the distribution of one eigenvalue picked at random.*

Proof. $\mathbb{E}[f(\lambda_J)] = \mathbb{E}\big[\frac{1}{N} \sum_i f(\lambda_i)\big] = \mathbb{E}\langle \hat{\rho}_M, f \rangle$. ∎

Remark 2.5 (Two levels of randomness, kept apart). $\hat{\rho}_M$ is random; $\mathbb{E}\hat{\rho}$ is not. A statement about $\mathbb{E}\hat{\rho}$ is a statement about averages over many independent matrices. A statement about $\hat{\rho}_M$ is a statement about one matrix, and getting from the first to the second is exactly a variance estimate. Both appear below, and they are not the same theorem.

2.2 Traces are moments

Lemma 2.6. *For every $k \geq 1$,*

$$\langle \hat{\rho}_M, x^k \rangle \;=\; \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k \;=\; \frac{1}{N} \operatorname{tr} M^k, \qquad \textit{hence} \qquad \langle \mathbb{E}\hat{\rho}, x^k \rangle = \frac{1}{N} \, \mathbb{E} \operatorname{tr} M^k.$$

Proof. The first equality is Definition 2.2 with $f(x) = x^k$. For the second, diagonalize: $M = O\Lambda O^\top$ with O orthogonal and Λ diagonal, so $M^k = O\Lambda^k O^\top$ and $\operatorname{tr} M^k = \operatorname{tr}\Lambda^k = \sum_i \lambda_i^k$ by cyclicity of the trace. The last claim is Definition 2.3. ■

The left side of Lemma 2.6 is a question about eigenvalues, which are a complicated function of the entries of M. The right side is a polynomial in the entries, whose expectation Theorem 1.2 computes. The next subsection does this for $k = 2$ and $k = 4$.

2.3 Traces of powers

Return to Proposition 1.10. Averaging $u^\top M^2 u$ over u uniform on the sphere — equivalently, over the vectors of any orthonormal basis — replaces $\|Mu\|^2$ by $\frac{1}{N}\operatorname{tr} M^2$, since $\mathbb{E}[uu^\top] = \frac{1}{N}I$. Traces of higher powers are no harder: $\operatorname{tr} M^k$ is a polynomial in the entries of M,

$$\operatorname{tr} M^k = \sum_{i_1,\dots,i_k} M_{i_1 i_2} M_{i_2 i_3} \cdots M_{i_k i_1},$$

and Theorem 1.2 computes the expectation of any polynomial in jointly Gaussian variables.

Proposition 2.7. $\mathbb{E}\operatorname{tr} M^2 = N + 1$.

Proof. $\operatorname{tr} M^2 = \sum_{i,j} M_{ij} M_{ji}$, and (4) gives $\mathbb{E}[M_{ij}M_{ji}] = \frac{1}{N}(\delta_{ij}\delta_{ji} + \delta_{ii}\delta_{jj}) = \frac{1}{N}(1 + \delta_{ij})$, so

$$\mathbb{E}\operatorname{tr} M^2 = \frac{1}{N}\sum_{i,j}\bigl(1 + \delta_{ij}\bigr) = \frac{1}{N}\bigl(N^2 + N\bigr) = N + 1.$$ ■

Only one pair partition was available there. The fourth power is where the partitions start to disagree.

Definition 2.8 (Crossing)**.** Draw the $2q$ points of a pair partition on a line and join each block by an arc above it. Two blocks $\{a, b\}$ and $\{c, d\}$ *cross* if $a < c < b < d$. A pair partition with no crossing blocks is *non-crossing*; there are $1, 2, 5, 14, \dots$ of them on $2, 4, 6, 8$ points, against $1, 3, 15, 105$ pair partitions in total.

Proposition 2.9. $\mathbb{E}\operatorname{tr} M^4 = 2N + 5 + \dfrac{5}{N}$.

Proof. Expanding, $\operatorname{tr} M^4 = \sum_{i,j,k,l} M_{ij}M_{jk}M_{kl}M_{li}$, and Theorem 1.2 splits each term into the three pair partitions of the four factors. Take them in turn.

$\{12\}\{34\}$, *non-crossing.* By (4), $\mathbb{E}[M_{ij}M_{jk}] = \frac{1}{N}(\delta_{ij}\delta_{jk} + \delta_{ik})$ and $\mathbb{E}[M_{kl}M_{li}] = \frac{1}{N}(\delta_{kl}\delta_{li} + \delta_{ki})$. Multiplying out gives four terms; summing each over i, j, k, l counts the index tuples it allows:

$$\underbrace{\delta_{ij}\delta_{jk}\delta_{kl}\delta_{li}}_{\text{all equal: } N} + \underbrace{\delta_{ij}\delta_{jk}\delta_{ki}}_{i=j=k:\ N^2} + \underbrace{\delta_{ik}\delta_{kl}\delta_{li}}_{i=k=l:\ N^2} + \underbrace{\delta_{ik}}_{i=k:\ N^3},$$

so the contribution is $\frac{1}{N^2}(N^3 + 2N^2 + N) = N + 2 + \frac{1}{N}$.

$\{14\}\{23\}$, *non-crossing.* The same computation with $\mathbb{E}[M_{ij}M_{li}] = \frac{1}{N}(\delta_{il}\delta_{ij} + \delta_{jl})$ and $\mathbb{E}[M_{jk}M_{kl}] = \frac{1}{N}(\delta_{jk}\delta_{kl} + \delta_{jl})$ gives $N + 2 + \frac{1}{N}$ again.

$\{13\}\{24\}$, *crossing.* Here

$$\mathbb{E}[M_{ij}M_{kl}]\,\mathbb{E}[M_{jk}M_{li}] = \tfrac{1}{N^2}\bigl(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\bigr)\bigl(\delta_{jl}\delta_{ki} + \delta_{ji}\delta_{kl}\bigr),$$

whose four terms allow N^2, N, N and N tuples respectively — the first is $\delta_{ik}\delta_{jl}$, two constraints, and each of the other three forces all four indices equal. The contribution is $\frac{1}{N^2}(N^2+3N) = 1+\frac{3}{N}$. Adding, $\mathbb{E}\operatorname{tr} M^4 = 2(N+2+\frac{1}{N}) + 1 + \frac{3}{N} = 2N+5+\frac{5}{N}$. ■

Remark 2.10 (What a crossing costs). Each non-crossing partition left one constraint on four indices, so N^3 tuples survived and the contribution was of order N. The crossing one left two constraints, N^2 tuples, and a contribution of order 1. Concretely: $i = k$ and $j = l$ make $M_{jk} = M_{ji} = M_{ij}$, $M_{kl} = M_{ij}$ and $M_{li} = M_{ij}$, so all four factors collapse to the *same* entry of M. A crossing forces a coincidence, and coincidences are rare among N^4 index tuples.

Remark 2.11 (Real versus complex). The two delta terms in (4) come from $M_{ij} = M_{ji}$. A complex Hermitian matrix has $\mathbb{E}[H_{ij}H_{kl}] = \frac{1}{N}\delta_{il}\delta_{jk}$, one term instead of two, and then the crossing partition leaves three constraints rather than two and costs N^{-2}. One extra way of gluing two entries together is the whole difference between the two ensembles.

Dividing by N,

$$\frac{1}{N}\mathbb{E}\operatorname{tr} M^2 \longrightarrow 1, \qquad \frac{1}{N}\mathbb{E}\operatorname{tr} M^4 \longrightarrow 2 :$$

one non-crossing pair partition on two points, two on four points, and the crossing one contributing nothing in the limit. By Lemma 2.6 these are $\langle \mathbb{E}\hat\rho, x^2\rangle$ and $\langle \mathbb{E}\hat\rho, x^4\rangle$. The pattern is a theorem, and §2.4 proves it.

Exercise 2.12 (TYPE B)**.** Show that $\operatorname{Var}(\operatorname{tr} M^2) = 4+\frac{4}{N}$. *Expand the square, apply Theorem 1.2 to the four-fold products, and note that the partition* $\{12\}\{34\}$ *reproduces* $(\mathbb{E}\operatorname{tr} M^2)^2$ *and is subtracted off, while the other two are equal. The remaining sum is* $\frac{2}{N^2}\sum_{ijkl}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})^2$*, and* $\delta^2 = \delta$*.* The mean is of order N and the variance is $O(1)$: the fluctuations of $\operatorname{tr} M^2$ do not grow with N at all. That is what §2.4 needs in order to turn a statement about averages into a statement about one matrix.

Exercise 2.13 (TYPE A)**.** Show $\mathbb{E}\operatorname{tr} M^{2k+1} = 0$ for every $k \geq 0$, with no computation — say why. Then draw all 15 pair partitions of 6 points, mark the crossings, and check that exactly 5 are non-crossing.

Exercise 2.14 (TYPE C)**.** Drop the Gaussian hypothesis. Let ξ_{ij}, $i \leq j$, be i.i.d. centred with variance 1 and finite fourth moment, and build a symmetric M by

$$M_{ij} = M_{ji} = \frac{\xi_{ij}}{\sqrt{N}} \quad (i<j), \qquad M_{ii} = \frac{\sqrt{2}\,\xi_{ii}}{\sqrt{N}},$$

so that (4) still holds but Theorem 1.2 is false. Show that

$$\mathbb{E}\operatorname{tr} M^4 \;=\; 2N + 5 + \frac{5}{N} + \kappa_4(\xi)\Big(1 + \frac{3}{N}\Big),$$

which is Proposition 2.9 again when ξ is Gaussian, and whose leading term is $2N$ whatever ξ is. *The leading order does not know what the entries are. That is universality, and with §1.5* in hand it costs one computation. Check the formula numerically with* $\xi_{ij} = \pm 1$*, for which* $\kappa_4 = -2$*.*

2.4 The semicircle law by counting

Propositions 2.7 and 2.9 contain the whole mechanism. What remains is bookkeeping for general k and one identification. The lecture gives the argument in outline; the exercises fill in steps 3, 5 and 6.

Theorem 2.15 (Wigner). *For the GOE, $\hat{\rho}_M \Rightarrow \rho_{sc}$ in probability, where*

$$\rho_{sc}(x) = \frac{1}{2\pi}\sqrt{4-x^2} \quad \text{on } [-2,2].$$

Stated without proof: the method of moments. A probability law on $\mathbb{R}$ with bounded support is determined by its moments, and if $\langle \mu_N, x^k \rangle \to \langle \mu, x^k \rangle$ for every k with μ compactly supported, then $\mu_N \Rightarrow \mu$. We use this and do not prove it.

The proof runs in six steps. Steps 1 and 2 are Lemma 2.6 and Theorem 1.2.

1. **Trace to moment.** $\langle \hat{\rho}_M, x^{2k} \rangle = \frac{1}{N} \operatorname{tr} M^{2k}$, which is Lemma 2.6.
2. **Wick.** Expanding the trace and applying Theorem 1.2 turns $\frac{1}{N}\mathbb{E} \operatorname{tr} M^{2k}$ into a sum over pair partitions of the $2k$ factors, exactly as in Proposition 2.9.
3. **Count indices, by reading the trace as a walk.** An index tuple $(i_1, \dots, i_{2k})$ is a closed walk of length $2k$, its rth step traversing the edge $\{i_r, i_{r+1}\}$. Since $\mathbb{E}[M_{ab}M_{cd}] \neq 0$ only when $\{a,b\} = \{c,d\}$, a surviving pair partition matches steps traversing the *same* edge, so the walk uses at most k edges, each twice. Its edge graph is connected, so $v \leq e+1 \leq k+1$, and a walk with v free indices contributes N^{v-k-1}: only walks along a tree with k edges survive. The counting is nothing but *a tree is the sparsest connected graph.*
4. **Limit in expectation.** Hence $\frac{1}{N}\mathbb{E} \operatorname{tr} M^{2k} \to C_k$, the number of such tree walks up to relabelling — equivalently the number of non-crossing pair partitions of $2k$ points. Odd moments vanish.
5. **Identify C_k.** Condition on the block containing the point 1, which gives the Catalan recursion $C_k = \sum_{j<k} C_j C_{k-1-j}$, hence $C_k = \frac{1}{k+1}\binom{2k}{k}$. Separately, $\int_{-2}^{2} x^{2k} \rho_{sc}(x)\, dx = \frac{1}{k+1}\binom{2k}{k}$, so the two sides agree.
6. **From expectation to probability, and from moments to the law.** $\operatorname{Var}\left(\frac{1}{N} \operatorname{tr} M^{2k}\right) = O(N^{-2})$, as in Exercise 2.12, so the empirical moments converge in probability and not merely on average; the method of moments then converts convergence of all moments into weak convergence.

Remark 2.16. Gluing the edges of a $2k$-gon according to a pair partition builds a surface whose faces are the surviving indices, and Euler's formula makes that count maximal exactly for planar gluings. This is the usual proof of step 3. We will not use it.

Exercise 2.17 (TYPE B). *(Step 3, first half.)* Expand $\frac{1}{N}\mathbb{E} \operatorname{tr} M^{2k}$ and apply Theorem 1.2. Using $\mathbb{E}[M_{ab}M_{cd}] \neq 0 \Rightarrow \{a,b\} = \{c,d\}$, show that a pair partition contributes nothing unless each of its blocks pairs two steps traversing the **same edge**. Deduce that a contributing walk uses at most k distinct edges, each traversed exactly twice in the leading terms. *Steps with $i_r = i_{r+1}$, and edges used four or more times, both lower the vertex count and are lower order.*

Exercise 2.18 (TYPE B). *(Step 3, second half.)* Let G be the edge graph of a contributing walk, with v vertices and e edges. The walk is closed, so G is connected and $e \geq v-1$; combine with $e \leq k$ to get $v \leq k+1$, with equality exactly when G is a tree with k edges each traversed twice. Then count: $\sim N^v$ index tuples, k covariances of size $1/N$, and a $1/N$ in front, so the contribution is N^{v-k-1} and only tree walks survive. *Check the verdict against Proposition 2.9: at $k=2$ the two non-crossing partitions are tree walks and the crossing one is not.*

Exercise 2.19 (TYPE C). *(Step 5.)* Let C_n be the number of non-crossing pair partitions of $2n$ points and condition on the block $\{1, b\}$ containing the point 1. Show the points strictly between 1 and b must pair among themselves, so $b = 2j + 2$ is even and the rest splits into intervals of $2j$ and $2(n-1-j)$; deduce

$$C_n = \sum_{j=0}^{n-1} C_j C_{n-1-j}, \qquad C_0 = 1,$$

and compute $1, 1, 2, 5, 14, 42$. Then show the generating function $C(w) = \sum_n C_n w^n$ satisfies $C = 1 + wC^2$, solve, and take the root finite at $w = 0$; expanding it gives $C_n = \frac{1}{n+1}\binom{2n}{n}$, which you may quote. Finally, substitute $x = 2\sin\theta$ and evaluate

$$\int_{-2}^{2} x^{2n}\, \rho_{sc}(x)\, dx = \frac{1}{n+1}\binom{2n}{n},$$

so that nothing is matched by eye.

Exercise 2.20 (TYPE A). Plot the empirical spectral measure at $N = 50$ and $N = 2000$ against ρ_{sc}, then once more with entries $\pm 1/\sqrt{N}$, which are not Gaussian and for which Theorem 1.2 is false. One figure. *Universality, weeks before the course explains it.*

2.5 $*$ The resolvent route

Not lectured. A second proof of Theorem 2.15, which produces the density rather than recognizing a sequence of numbers.

What the counting proof cannot do

(i) It never produced a **density**. It produced a sequence of numbers that we recognized; perturb the model and there is nothing to recognize.

(ii) It requires **all moments finite**.

(iii) It gives **no local information** — nothing about one eigenvalue, a gap, or the edge.

(iv) It does not survive **deformation**: unequal variances, or a rank-one spike, and the combinatorics does not adapt.

All four are repaired by replacing the polynomials x^k with a different family of test functions.

The Stieltjes transform

Definition 2.21. For a probability law μ on $\mathbb{R}$ and $z \in \mathbb{C}$ with $\Im z > 0$,

$$m_\mu(z) = \int \frac{d\mu(\lambda)}{\lambda - z} = \left\langle \mu, \tfrac{1}{x-z} \right\rangle.$$

The test function $\frac{1}{x-z}$ is bounded and continuous for each fixed z off the real axis, so Definition 2.21 makes sense for every probability law, with no moment hypothesis at all. It is analytic on the upper half plane, satisfies $|m_\mu(z)| \le 1/\Im z$, and $m_\mu(z) = -1/z + O(|z|^{-2})$ as $z \to \infty$ — the last of which will pin down a branch.

Stieltjes inversion. If μ has a continuous density ρ near λ then $\rho(\lambda) = \frac{1}{\pi} \lim_{\varepsilon \downarrow 0} \Im m_\mu(\lambda + i\varepsilon)$. Proved in Lectures 7–8, where the resolvent method is carried further.

The moment method has no analogue of this step: it is what turns a transform back into a picture of the spectrum. The bridge to matrices is immediate from Definition 2.2, since $\frac{1}{\lambda_i - z}$ are the eigenvalues of $(M-z)^{-1}$:

$$m_{\hat{\rho}_M}(z) = \frac{1}{N} \operatorname{tr} \left(M - z\right)^{-1}. \tag{5}$$

The matrix $(M-z)^{-1}$ is the *resolvent*. It exists for every z off the real axis because M is symmetric, and $\|(M-z)^{-1}\| \le 1/\Im z$.

Remark 2.22. Ridge regression in §2.5 of Lectures 1–2 is $(H^\top H + \lambda I)^{-1} H^\top y$, and the matrix inverted there is (5) at $z = -\lambda$.

The self-consistent equation

Write $m_N(z) = \frac{1}{N} \mathbb{E} \operatorname{tr}(M-z)^{-1}$. Let $M^{(1)}$ be M with its first row and column deleted, and let $a \in \mathbb{R}^{N-1}$ be the first column of M with its first entry deleted. The strategy is to compute one diagonal entry of the resolvent in terms of the same object one dimension down.

Lemma 2.23 (Schur complement)**.** $\left((M-z)^{-1}\right)_{11} = \dfrac{1}{M_{11} - z - a^\top \left(M^{(1)} - z\right)^{-1} a}.$

This is the block-inversion identity behind the Schur complement in §2.4 of Lectures 1–2.

Now estimate the three pieces. Put $R = (M^{(1)} - z)^{-1}$, which is a function of $M^{(1)}$ alone and therefore independent of a, and note $\mathbb{E}[a_r a_s] = \delta_{rs}/N$.

(a) $M_{11} = O_{\mathbb{P}}(N^{-1/2})$, since $\operatorname{Var} M_{11} = 2/N$.

(b) $\mathbb{E}\left[a^\top R a \,\middle|\, M^{(1)}\right] = \frac{1}{N} \operatorname{tr} R$, and its conditional variance is $\frac{2}{N^2} \operatorname{tr}(R^2)$. Both are Example 1.6 with $A = R$ and $\Sigma = \frac{1}{N} I$; the computation there is algebraic, so it applies verbatim to the complex symmetric R. Since $\|R\| \le (\Im z)^{-1}$ we have $|\operatorname{tr}(R^2)| \le N(\Im z)^{-2}$, so the variance is $O(1/N)$. This is the only place probability enters.

(c) $\frac{1}{N} \operatorname{tr} R$ differs from $m_N(z)$ by $O(1/N)$, since deleting one row and column of an $N \times N$ matrix moves a normalized resolvent trace by that much.

All diagonal entries of $(M-z)^{-1}$ have the same law, so $m_N(z) = \mathbb{E}\left[((M-z)^{-1})_{11}\right]$. Feeding (a)–(c) into Lemma 2.23 and passing to the limit,

$$\boxed{m(z) = \frac{1}{-z - m(z)}, \qquad \text{that is} \qquad m^2 + zm + 1 = 0.} \tag{6}$$

Asserted. That $\frac{1}{N} \operatorname{tr}(M-z)^{-1}$ concentrates around its mean, so that the limit in (6) may be taken inside, is a concentration estimate we do not prove.

Equation (6) is a fixed-point equation for the answer, derived without knowing the answer. The moment method had to be told what to recognize.

Solving, and the comparison

Both roots of (6) are $\frac{-z\pm\sqrt{z^2-4}}{2}$. As $z\to\infty$, $\sqrt{z^2-4}=z-2/z+O(|z|^{-3})$, so only the + root satisfies $m\sim -1/z$:

$$m(z)=\frac{-z+\sqrt{z^2-4}}{2}.$$

For $\lambda\in(-2,2)$ we have $\sqrt{\lambda^2-4}=i\sqrt{4-\lambda^2}$, so $\Im m(\lambda+i0)=\frac{1}{2}\sqrt{4-\lambda^2}$, and Stieltjes inversion returns $\rho_{sc}(\lambda)=\frac{1}{2\pi}\sqrt{4-\lambda^2}$. That is Theorem 2.15 again, now with its density.

Remark 2.24 (The two proofs are the same equation). Expanding Definition 2.21 for large $|z|$ gives $m(z)=-\sum_{k\geq 0}C_k z^{-(2k+1)}$, so with $w=z^{-2}$ we have $-zm(z)=C(w)$. Substituting into the Catalan relation $C=1+wC^2$ of Exercise 2.19 returns $-zm=1+m^2$, which is (6). *The recursion that counts non-crossing pair partitions and the self-consistent equation for the resolvent are one equation written twice.*

Return to the four complaints above. The density came out directly. Only the second moment of the entries was used, so nothing needed all moments. Unequal variances $\mathbb{E}M_{ij}^2=S_{ij}/N$ turn (6) into a system of N coupled equations rather than one, which is a change of size and not of method. A rank-one spike is a rank-one perturbation of the resolvent. And shrinking $\Im z$ toward the real axis, rather than holding it fixed, is precisely what converts a global law into a local one.

Exercise 2.25 (TYPE A)**.** Compute m_μ for $\mu=\delta_0$, for $\mu=\frac{1}{2}(\delta_{-1}+\delta_1)$, and for μ uniform on $[0,1]$. In each case check $|m_\mu(z)|\leq 1/\Im z$ and $m_\mu(z)=-1/z+O(|z|^{-2})$, and recover the density of the third by Stieltjes inversion.

Exercise 2.26 (TYPE A)**.** Numerically, for a GOE matrix at $N=2000$, plot $\frac{1}{\pi}\Im\big[\frac{1}{N}\operatorname{tr}(M-\lambda-i\varepsilon)^{-1}\big]$ against $\lambda\in[-3,3]$ for $\varepsilon=1,\ 0.1,\ 0.01,\ 0.001$, with ρ_{sc} overlaid. *At $\varepsilon=1$ the curve is a smoothed semicircle; at $\varepsilon=0.001$ it is a forest of spikes at the individual eigenvalues. Somewhere in between is the global law. That crossover is at $\varepsilon\sim 1/N$, and it is the whole content of the word "local".*

Exercise 2.27 (TYPE B)**.** Prove Remark 2.24 carefully. Expand $\frac{1}{\lambda-z}=-\sum_{k\geq 0}\lambda^k z^{-(k+1)}$ for $|z|>2$, integrate against ρ_{sc} to get $m(z)=-\sum_k C_k z^{-(2k+1)}$ (odd moments vanish), substitute $w=z^{-2}$ to identify $-zm(z)=C(w)$, and check that $C=1+wC^2$ is (6).

Exercise 2.28 (TYPE B)**.** Prove estimate (c): if M is symmetric $N\times N$ and $M^{(1)}$ deletes one row and column, then

$$\left|\operatorname{tr}(M-z)^{-1}-\operatorname{tr}\left(M^{(1)}-z\right)^{-1}\right|\leq\frac{\pi}{\Im z}.$$

Cauchy interlacing gives $\lambda_i\leq\mu_i\leq\lambda_{i+1}$ for the eigenvalues of M and $M^{(1)}$, so the two eigenvalue counting functions differ by at most 1 at every point. Write both traces as integrals against those counting functions, integrate by parts, and use $\int_{\mathbb{R}}|x-z|^{-2}dx=\pi/\Im z$. Note that the bound does not depend on N, which is why dividing by N makes it $O(1/N)$.

Exercise 2.29 (TYPE C)**.** *Unequal variances.* Let M be symmetric Gaussian with $\mathbb{E}M_{ij}^2=S_{ij}/N$, where S is a fixed symmetric matrix with non-negative entries, and let $m_i(z)=\mathbb{E}\big[((M-z)^{-1})_{ii}\big]$. Redo the derivation above with the ith row and column deleted instead of the first, and derive the system

$$m_i(z)=\frac{1}{-z-\frac{1}{N}\sum_j S_{ij}\,m_j(z)},\qquad i=1,\dots,N.$$

Check that $S_{ij} \equiv 1$ returns (6). *Then take S block-constant with two blocks and solve the resulting pair of equations; the spectrum is no longer a semicircle, and no combinatorial recognition step could have produced it.*

ASSIGNED **Lecture 6**, due at the start of Lecture 7: Exercises 2.12, 2.13, 2.17, 2.18, 2.19, 2.20, 2.25, 2.29 — three A, three B, two C. Exercise 2.14 and every other exercise in §2.5* are pool problems: available and examinable, not collected. *If you are short of time, do the C problems.*

---

§2.5* proves the semicircle law a second time, by the resolvent, and is where the density comes from. Lectures 7–8 prove Stieltjes inversion, supply the estimates (b) and (c) used there, and rederive Marchenko–Pastur by the same route.