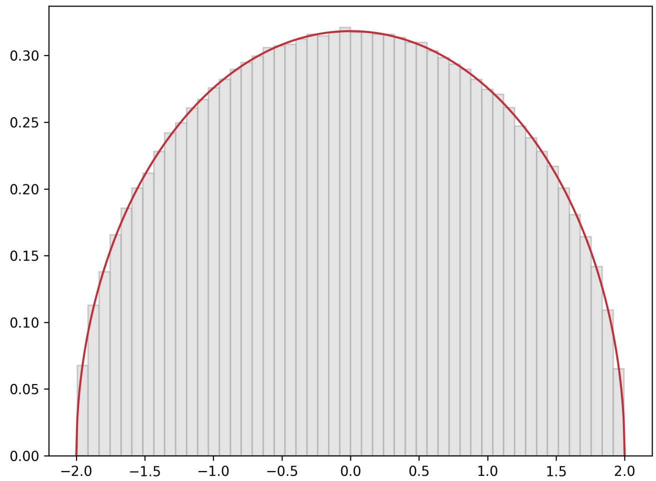
University of Bristol, MATH 30033

Random Matrix Theory - Part II

Instructor: Dr. Emma Bailey



Based on notes by Thomas Bothner, adapted by Emma Bailey (2025)

Spring 2025

Contents

References for this part of the course include

- G. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*
- L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*
- G. Szegő, *Orthogonal polynomials*

Please let me know about any typos in the text[1], any errors should be primarily attributed to the current author. This material is provided exclusively for educational purposes at the University of Bristol and is to be downloaded or copied for your private study only.

[1] *E-mail address:* e.c.bailey@bristol.ac.uk

15. Preliminaries

15.1. Spectral statistics

In the first part of the module, we saw *exact results*, i.e. those that are valid for matrices with a fixed dimension. Whilst it would be very informative to be able to answer all questions for finite matrix size, this is often infesible. Therefore, in this second part of the module we shall be concerned with the *infinite size* limit of random matrices and related asymptotic regimes.

At various stages within these notes, we will draw from other areas of mathematics, notably Probability and Combinatorics. A reminder of the results that will comprise our toolkit will be collected in the appendix of these notes. If you are so inclined, many of the results that you will see in this course (both this part and the preceding) can be relatively easily coded up in e.g. Python or Mathematica, as evidenced by the graphs throughout the notes. I encourage you to try to recreate them! A good reference is 'Computational Random Matrix Theory' by Robert Sweeney-Blanco.

Quantities

We collect here several quantities that are widely studied in random matrix theory. To begin with, it turns out that a number of spectral properties of a random matrix can be formulated and studied in terms of the counting measure of its eigenvalues.

Given a matrix (random or deterministic) of dimension N, say M_N, we will write $\lambda_j(M_N)$, $j = 1, \ldots, N$ for its eigenvalues. Whenever it is clear from context, we will write simply λ_j to clean up notation.

Definition 15.1. *Let $(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N$ denote the eigenvalues of a real symmetric or complex Hermitian matrix M_N of size $N \times N$. The **counting measure**, $\mathcal{N}_N$, of its eigenvalues is*

$$\mathcal{N}_N(\Delta) := \left|\left\{\ell \in \{1, \ldots, N\} : \ \lambda_\ell \in \Delta\right)\right|,$$

*where Δ is an interval of $\mathbb{R}$, and the **normalized counting measure**, $\overline{\mathcal{N}}_N$, of its eigenvalues is*

$$\overline{\mathcal{N}}_N(\Delta) := \frac{1}{N}\mathcal{N}_N(\Delta). \tag{15.1}$$

If M_N is a random matrix then the counting measure $\mathcal{N}_N(\Delta)$ of its eigenvalues is a random variable (a measurable function) for a given interval Δ and we shall index our eigenvalues in nondecreasing order

$$-\infty < \lambda_1(M_N) \leq \ldots \leq \lambda_N(M_N) < \infty. \tag{15.2}$$

Example 15.2.

1. *Take* $\begin{pmatrix} 7 & 6 \\ -4 & -3 \end{pmatrix}$ *and so* $\lambda_1 = 1, \lambda_2 = 3$. *In this case we have*
   - $\mathcal{N}_2([0,2]) = 1$; $\overline{\mathcal{N}}_2([0,2]) = 1/2$,
   - $\mathcal{N}_2((-\infty, 0.5] \cup [\pi, 101]) = 0$; $\overline{\mathcal{N}}_2([-\infty, 0.5] \cup [\pi, 101]) = 0$,
   - $\mathcal{N}_2(\mathbb{R}) = 2$; $\overline{\mathcal{N}}_2(\mathbb{R}) = 1$.
2. *Take* $\begin{pmatrix} B_1 & B_2 \\ B_2 & B_1 \end{pmatrix}$ *where* B_1, B_2 *are independent standard Bernoulli random variables (i.e. taking the values* 0, 1 *each with probability* $1/2$*). Then the possible spectra are* $\{0\}$, $\{1\}$, *and* $\{-1, 1\}$. *We emphasise that since this matrix is random, the associated counting measure is also random. Indeed, by considering cases we have that if* $\Delta_1, \Delta_{-1} \subset \mathbb{R}$ *such that* $1 \notin \Delta_1$ *and* $-1 \notin \Delta_{-1}$, *then*

$$\mathbb{P}(\mathcal{N}_2(\Delta_1) = 1) = 1/2, \qquad \mathbb{P}(\mathcal{N}_2(\Delta_{-1}) = 1) = 3/4.$$

We will need a generalisation of the counting measure $\mathcal{N}_N(\Delta)$. Rather than considering $\mathcal{N}_N(\cdot)$ as a function on an interval, we could instead consider it as a function on the *indicator function* χ_Δ (i.e. given $\Delta \subset \mathbb{R}$, $\chi_\Delta : \mathbb{R} \to \mathbb{C}$ is $\chi_\Delta(x) = 1$ if $x \in \Delta$ and $\chi_\Delta(x) = 0$ if $x \notin \Delta$). This then permits us to take a more general view of $\mathcal{N}_N$ as a function on a wider class of functions known as *test functions*. To avoid the fully technical definition of a test function, one can think of a test function $\varphi : \mathbb{R} \to \mathbb{C}$ as a complex valued, smooth (infinitely differentiable) function whose support $\mathrm{supp}(\varphi)$ is compact. These functions are 'nice' in that we can cook-up examples that closely approximate the indicator function but have the additional property that they are infinitely differentiable, cf. Figure 15.1.

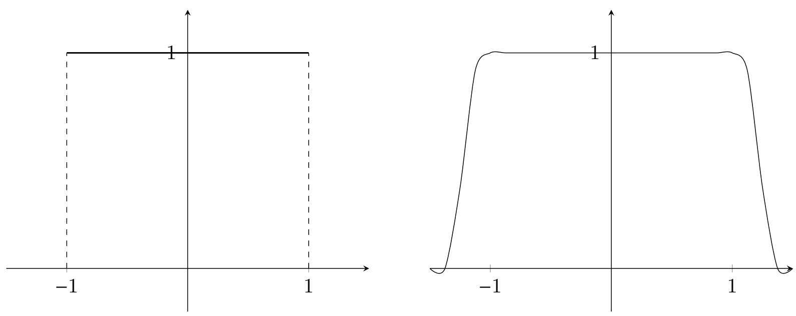


Figure 15.1.: Comparison of an indicator function and an example of a test function.

Definition 15.3. *Given a real or complex Hermitian matrix* M_N *with eigenvalues* $\lambda_1, \ldots, \lambda_N$ *and a* test function $\varphi : \mathbb{R} \to \mathbb{C}$, *the corresponding* ***linear statistic*** $\mathcal{N}_N[\varphi]$ *is*

$$\mathcal{N}_N[\varphi] := \sum_{j=1}^{N} \varphi(\lambda_j) = \mathrm{Tr}(\varphi(M_N)). \tag{15.3}$$

Remark. *The last equality in (15.3), involving the matrix trace, is justified via the spectral mapping theorem (cf. Theorem A.5).*

If $\mathbb{P}_N$ is a probability measure, determining a random matrix, and $\mathbb{E}_N$ denotes the corresponding expectation, then we consider further the *expectation* $\mathbb{E}_N\big[\mathcal{N}_N[\varphi]\big]$ of a linear statistic.

Of particular interest in the forthcoming section will be the expectation $\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]$ of the normalized counting measure (15.1); and the **occupation probabilities** $\mathbb{P}_N\big(\mathcal{N}_N(\Delta) = \ell\big), \ell \in \{0, 1, 2, \ldots, N\}$ of the counting measure for a given interval Δ. In the case $l = 0$ – i.e. the interval contains no eigenvalues – this will be referred to as the **gap probability** $\mathbb{P}_N\big(\mathcal{N}_N(\Delta) = 0\big)$.

The linear statistic (15.3), as well as many other functions of matrix eigenvalues that we are going to study, are symmetric functions of those eigenvalues. We also saw in the previous part of the course that this was true of certain joint probability density functions (e.g. Haar measure on $U(N)$). There, the *correlation function* was also defined. We recall the definition again here, written generally to allow for any joint probability density function $P_N(\cdot)$ on N real points

$$R_\ell(x_1, \ldots, x_\ell) = \frac{N!}{(N-\ell)!} \int_{\mathbb{R}^{N-\ell}} P_N(x_1, \ldots, x_N) dx_{\ell+1} \cdots dx_N \tag{15.4}$$

defined for any $\ell = 1, \ldots N$. Without the pre-factor, these integrals give the **marginal densities** of P_N

$$P_{N,\ell}(x_1, \ldots, x_\ell) := \int_{\mathbb{R}^{N-\ell}} P_N(x_1, \ldots, x_N) dx_{\ell+1} \cdots dx_N, \qquad \ell = 1, 2, \ldots, N. \tag{15.5}$$

It transpires that these three preceeding quantities (occupation probabilities, marginal densities, correlation functions) are all connected. Clearly the marginal densities and the correlation functions are trivially related. To see how they relate to the *occupation probabilities*, first notice that we can write the gap probability (the $\ell = 0$ case for the occupation probabilities) as

$$\mathbb{P}_N(\mathcal{N}_N(\Delta) = 0) = \mathbb{E}_N\left[\prod_{j=1}^{N}\Big(1 - \chi_\Delta(\lambda_j)\Big)\right]. \tag{15.6}$$

Again we can define the corresponding more general quantity, replacing χ_Δ in (15.6) by a test function $\varphi : \mathbb{R} \to \mathbb{C}$, compare (15.3),

$$\mathbb{E}_N\left[\prod_{j=1}^{N}\Big(1 - \varphi(\lambda_j)\Big)\right]. \tag{15.7}$$

By expanding the product in (15.7) and using the marginal densities (15.5), we can obtain

$$\begin{aligned}\mathbb{E}_N\left[\sum_{\Omega\subset\{1,\dots,N\}}(-1)^{|\Omega|}\prod_{j\in\Omega}\varphi(\lambda_j)\right] &= \mathbb{E}_N\left[\sum_{\ell=0}^{N}(-1)^{\ell}\binom{N}{\ell}\prod_{j=1}^{\ell}\varphi(\lambda_j)\right]\\ &= \sum_{\ell=0}^{N}(-1)^{\ell}\binom{N}{\ell}\mathbb{E}_N\left[\prod_{j=1}^{\ell}\varphi(\lambda_j)\right]\\ &= \sum_{\ell=0}^{N}(-1)^{\ell}\binom{N}{\ell}\int_{\mathbb{R}^\ell}P_{N,\ell}(x_1,\dots,x_\ell)\prod_{j=1}^{\ell}\varphi(x_j)dx_j, \quad (15.8)\end{aligned}$$

where the term corresponding to $\Omega=\emptyset$ in the first equality, respectively to $\ell=0$ in the second and all later equalities, is 1.

Now, it is possible to define a generalization of a directional derivative known as a Gateaux derivative or a *variational derivative*, $\frac{\delta}{\delta\varphi}\Phi$. It is defined implicitly via

$$\frac{d}{d\varepsilon}\Phi[\varphi+\varepsilon\psi]\bigg|_{\varepsilon=0} = \lim_{\varepsilon\to0}\frac{\Phi[\varphi+\varepsilon\psi]-\Phi[\varphi]}{\varepsilon} = \int\psi(x)\frac{\delta}{\delta\varphi(x)}\Phi[\varphi]dx,$$

for a functional Φ, acting on a certain set of functions φ. Applying this operator to (15.8) one finds for example

$$-\frac{\delta}{\delta\varphi(x_1)}\mathbb{E}_N\left[\prod_{j=1}^{N}\left(1-\varphi(\lambda_j)\right)\right]\bigg|_{\varphi=0} = NP_{N,1}(x_1).$$

The following proposition is the general case.

Proposition 15.4. *For any $\ell=1,2,\dots,N$, we have for any collection of distinct $(x_1,\dots,x_\ell)$,*

$$(-1)^{\ell}\frac{\delta^{\ell}}{\delta\varphi(x_1)\cdots\delta\varphi(x_\ell)}\mathbb{E}_N\left[\prod_{j=1}^{N}\left(1-\varphi(\lambda_j)\right)\right]\bigg|_{\varphi=0} = N(N-1)\cdot\ldots\cdot(N-\ell+1)P_{N,\ell}(x_1,\dots,x_\ell). \quad (15.9)$$

Notice that the right hand side of (15.9) is indeed the lth correlation function (15.4). This establishes the claimed connection to the gap probabilities. It is now not a far leap to tie-in the occupation probabilities $\mathbb{P}_N(\mathcal{N}_N(\Delta)=\ell)$. Choose $\varphi=a\cdot\chi_\Delta$ for some $a\in\mathbb{R}$. Then, similarly to Proposition 15.4, we have the following lemma.

Lemma 15.5. *For any $\ell\in\{0,1,2,\dots,N\}$ and any interval $\Delta\subset\mathbb{R}$, we have*

$$\frac{(-1)^{\ell}}{\ell!}\frac{d^{\ell}}{da^{\ell}}\mathbb{E}_N\left[\prod_{j=1}^{N}(1-a\chi_\Delta(\lambda_j))\right]\bigg|_{a=1} = \mathbb{P}_N\left(\mathcal{N}_N(\Delta)=\ell\right). \quad (15.10)$$

Proof. The claim is vacuous for $\ell = 0$, see (15.6), so let's begin with $\ell = 1$. In that case, by product rule, and symmetry of P_N,

$$\begin{aligned}
-\frac{d}{da}\mathbb{E}_N\left[\prod_{j=1}^N(1-a\chi_\Delta(\lambda_j))\right]\bigg|_{a=1} &= \int_{\mathbb{R}^N}\left[\sum_{k=1}^N \chi_\Delta(x_k)\prod_{\substack{j=1\\ j\neq k}}^N (1-\chi_\Delta(x_j))\right]P_N(x_1,\dots,x_N)\prod_{\ell=1}^N dx_\ell \\
&= N\int_{\mathbb{R}^N}\left[\chi_\Delta(x_1)\prod_{j=2}^N(1-\chi_\Delta(x_j))\right]P_N(x_1,\dots,x_N)\prod_{\ell=1}^N dx_\ell \\
&= \binom{N}{1}\int_\Delta\left[\int_{(\mathbb{R}\setminus\Delta)^{N-1}} P_N(x_1,\dots,x_N)\prod_{\ell=2}^N dx_\ell\right]dx_1 \\
&= \mathbb{P}_N(\mathcal{N}_N(\Delta)=1).
\end{aligned}$$

Next, for $\ell = 2$, again by the product role and symmetry of P_N,

$$\begin{aligned}
&\frac{(-1)^2}{2!}\frac{d^2}{da^2}\mathbb{E}_N\left[\prod_{j=1}^N(1-a\chi_\Delta(\lambda_j))\right]\bigg|_{a=1} \\
&\quad= \frac{N}{2}\frac{d}{da}\int_{\mathbb{R}^N}\left[\chi_\Delta(x_1)\prod_{j=2}^N(1-a\chi_\Delta(x_j))\right]P_N(x_1,\dots,x_N)\prod_{\ell=1}^N dx_\ell\bigg|_{a=1} \\
&\quad= \frac{N}{2}\int_{\mathbb{R}^N}\chi_\Delta(x_1)\left[\sum_{k=2}^N\chi_\Delta(x_k)\prod_{\substack{j=2\\ j\neq k}}^N(1-\chi_\Delta(x_j))\right]P_N(x_1,\dots,x_N)\prod_{\ell=1}^N dx_\ell \\
&\quad= \frac{N}{2}(N-1)\int_{\mathbb{R}^N}\left[\chi_\Delta(x_1)\chi_\Delta(x_2)\prod_{j=3}^N(1-\chi_\Delta(x_j))\right]P_N(x_1,\dots,x_N)\prod_{\ell=1}^N dx_\ell \\
&\quad= \binom{N}{2}\int_{\Delta^2}\left[\int_{(\mathbb{R}\setminus\Delta)^{N-2}} P_N(x_1,\dots,x_N)\prod_{\ell=3}^N dx_\ell\right]dx_1dx_2 \\
&\quad= \mathbb{P}_N(\mathcal{N}_N(\Delta)=2),
\end{aligned}$$

and the general case follows similarly by induction on ℓ. This completes our proof of (15.10). □

Remark. *Because of* (15.9) *and* (15.10), *the functional* $\mathbb{E}_N\left[\prod_{j=1}^N\left(1-\varphi(\lambda_j)\right)\right]$ *plays the role of the **generating functional** for both the marginal densities of the random variables* $\{\lambda_j\}_{j=1}^N$ *and for the probability distribution of the counting measure of eigenvalues. Indeed, recall from classical probability theory that, when it exists, the* moment generating function $\mathbb{E}[\exp(t\mathcal{X})]$ *for a random variable* $\mathcal{X}$ *is so-called since it generates the moments:*

$$\frac{d^\ell}{dt^\ell}\mathbb{E}[\exp(t\mathcal{X})]|_{t=0} = \mathbb{E}[\mathcal{X}^\ell].$$

Comparing now to (15.9) from Proposition 15.4 which states

$$(-1)^\ell \frac{\delta^\ell}{\delta\varphi(x_1)\cdots\delta\varphi(x_\ell)} \mathbb{E}_N\left[\prod_{j=1}^N \Big(1-\varphi(\lambda_j)\Big)\right]\Bigg|_{\varphi=0} = R_\ell(x_1,\ldots,x_\ell), \tag{15.11}$$

we see that the correlation functions R_ℓ are a kind of analogue of moments of random variables. Finally we record that by (15.8) and (15.4),

$$\mathbb{E}_N\left[\prod_{j=1}^N (1-\varphi(x_j))\right] = 1 + \sum_{\ell=1}^N \frac{(-1)^\ell}{\ell!}\int_{\mathbb{R}^\ell} R_\ell(x_1,\ldots,x_\ell)\prod_{j=1}^l \varphi(x_j)dx_j. \tag{15.12}$$

In practice: an application to i.i.d. random variables

We now discuss a standard object of probability theory, but embedded in the framework of random matrix theory: collections of independent, identically distributed (i.i.d.) random variables. We will repeatedly revisit this example as a comparison against other random matrices. To that end, let $(\lambda_j)_{j=1}^N$ be a collection of (possibly N-dependent) i.i.d. real-valued random variables with common law F_N. Define the *diagonal* $N\times N$ random matrix M_N as follows:

$$M_N = \big(\xi_{j,k}\big)_{j,k=1}^n, \qquad \xi_{j,k} := \begin{cases}\lambda_j, & j=k\\ 0, & j\neq k\end{cases}.$$

Clearly, the normalized counting measure of eigenvalues of M_N, compare (15.1), equals

$$\overline{\mathcal{N}}_N(\Delta) = \frac{1}{N}\sum_{j=1}^N \chi_\Delta(\lambda_j), \qquad \Delta\subset\mathbb{R},$$

and we find, exploiting independence and symmetry

$$\begin{aligned}\mathbb{E}_N\left[\overline{\mathcal{N}}_N(\Delta)\right] &= \mathbb{E}_N\left[\chi_\Delta(\lambda_1)\right] = \mathbb{P}_N(\lambda_1\in\Delta)\equiv F_N(\Delta),\\ \mathrm{Var}_N\left(\overline{\mathcal{N}}_N(\Delta)\right) &= \mathbb{E}_N\left[\left(\overline{\mathcal{N}}_n(\Delta)-\mathbb{E}_N\left[\overline{\mathcal{N}}_N(\Delta)\right]\right)^2\right]\\ &= \frac{1}{N}F_N(\Delta)\big(1-F_N(\Delta)\big).\end{aligned}$$

Similarly, the occupation probabilities (including the gap probability) are

$$\mathbb{P}_N(\mathcal{N}_N(\Delta)=\ell) = \binom{N}{\ell}\big(F_N(\Delta)\big)^\ell\big(1-F_N(\Delta)\big)^{N-\ell}, \quad \ell\in\{0,1,2,\ldots,N\}.$$

Moving ahead, let us further assume that F_N has a density ρ_N, i.e. $F_N(\Delta)=\int_\Delta \rho_N(x)dx$, then also

$$\mathbb{E}_N\left[\prod_{j=1}^N(1-\varphi(\lambda_j))\right] = \left(1-\int_{\mathbb{R}}\rho_N(x)\varphi(x)dx\right)^N. \tag{15.13}$$

Recall that a motivation for this part of the course is understanding how various statistical quantities behave in a limiting sense. In order to study the large N behaviour of the above, we now put in place the following assumption on the underlying probability measures.

Assumption. *The law F_N converges (weakly) to some limiting law F with density ρ, that is to say for every continuous and bounded function $\varphi : \mathbb{R} \to \mathbb{R}$ we have that*

$$\lim_{N\to\infty} \int_{\mathbb{R}} \varphi(x)\rho_N(x)dx = \int_{\mathbb{R}} \varphi(x)\rho(x)dx,$$

or equivalently, with $F_N(x) := F_N((-\infty, x])$, resp. $F(x) := F((-\infty, x])$, denoting the distribution function of F_N, resp. F, we have convergence of the distribution functions $\lim_{N\to\infty} F_N(x) = F(x)$ at every continuity point x of F.

Consequently, with this assumption in place on the law of λ_j, we have as $N \to \infty$,

$$\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)] = F_N(\Delta) = \int_\Delta \rho_N(x)dx \to \int_\Delta \rho(x)dx = F(\Delta), \tag{15.14}$$

for any $\Delta \subset \mathbb{R}$. Moreover, by Chebyshev's inequality (A.4) and (15.14), for any $\varepsilon > 0, \Delta \subset \mathbb{R}$ and $N \geq N_0$ sufficiently large,

$$\mathbb{P}_N\left(\left|\overline{\mathcal{N}}_N(\Delta) - \mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]\right| > \varepsilon\right) \leq \frac{4}{\varepsilon^2}\mathrm{Var}_N\left(\overline{\mathcal{N}}_N(\Delta)\right) = \frac{4}{\varepsilon^2 N} F_N(\Delta)(1 - F_N(\Delta)), \tag{15.15}$$

so the right hand side vanishes as $N \to \infty$ by (15.14). This shows that $\overline{\mathcal{N}}_N(\Delta)$ converges (in probability) to $\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)] = F(\Delta)$, i.e. the same counting measure satisfies a *weak law of large numbers*, cf. Theorem A.12.

In the same spirit, applying the classical Lindeberg central limit theorem, cf. Theorem A.14, and writing $\sigma^2 := F(\Delta)\big(1 - F(\Delta)\big)$,

$$\lim_{N\to\infty} \mathbb{P}_N\left(\sqrt{N}\big(\overline{\mathcal{N}}_N(\Delta) - F_N(\Delta)\big) \leq x\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}(y/\sigma)^2} dy, \tag{15.16}$$

valid for any fixed $x \in \mathbb{R}, \Delta \subset \mathbb{R}$. Interpreting (15.16), this says that the normalized counting measure, once properly centered and normalized, converges (in distribution) to the Gaussian random variable of zero mean and variance $F(\Delta)(1 - F(\Delta))$.

Consider now the gap probability for our diagonal matrix M_N

$$\mathbb{P}_N(\mathcal{N}_N(\Delta) = 0) = (1 - F_N(\Delta))^N. \tag{15.17}$$

We have

$$\lim_{N\to\infty} \mathbb{P}_N(\mathcal{N}(\Delta) = 0) = \begin{cases} 0, & F(\Delta) > 0 \\ 1, & F(\Delta) = 0 \end{cases}, \tag{15.18}$$

so in the limiting case, there is probability zero that *no eigenvalues* of M_N fall in Δ if $F(\Delta) > 0$. In this case that the limiting distribution function is positive on Δ, then we expect that Δ contains of the order $O(NF(\Delta))$ eigenvalues, for N sufficiently large. Note that this observation is consistent with the limit (15.14) for the average normalized number of eigenvalues in Δ.

More detailed information on the asymptotic behaviour of the distribution, or probability mass function $\mathbb{P}_N(\mathcal{N}_N(\Delta) = \ell)$, of the number of eigenvalues $\mathcal{N}_n(\Delta)$ falling in Δ can be obtained by considering N-dependent intervals.

Example 15.6. *Suppose our diagonal matrix is*

$$M_N = \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \lambda_N \end{pmatrix}$$

where the diagonal entries $\lambda_j \sim \text{Bin}(N, \frac{1}{N})$ *are i.i.d. Binomial random variables with unit mean. Then for example*

$$\overline{\mathcal{N}}_N([0, N]) = \frac{1}{N}\sum_{j=1}^{N} \chi_{[0,N]}(\lambda_j) = 1, \qquad \overline{\mathcal{N}}_N([-2, -1]) = 0.$$

Now consider $\Delta = [a, b]$ *where* $0 \leq a < b \leq \infty$. *We have as above*

$$\begin{aligned} \mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)] &= \mathbb{P}_N\left(\text{Bin}(N, \frac{1}{N}) \in [a, b]\right) \\ &= \sum_{\lceil a \rceil \leq j \leq \lfloor b \rfloor} \binom{N}{j} \frac{1}{N^j}\left(1 - \frac{1}{N}\right)^{N-j} \\ \mathbb{P}(\mathcal{N}_N(\Delta) = 0) &= \left(1 - \mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]\right)^N. \end{aligned}$$

An 'application': the law of large numbers (15.15) *tells us that as* N *gets larger, the normalized counting measure should be close to the expected value, which in turn is the probability that a* $\text{Bin}(N, 1/N)$ *lies in the interval. Further, we know that as* $N \to \infty$, *the* $\text{Bin}(N, 1/N)$ *distribution converges to the Poisson distribution* $\text{Pois}(1)$. *So the above tells us that for large* N, *the numbers of eigenvalues in* Δ *should be close to* $N \cdot \mathbb{P}(\text{Pois}(1) \in [a, b])$. *Let's test this pictorially, see Figure 15.2.*

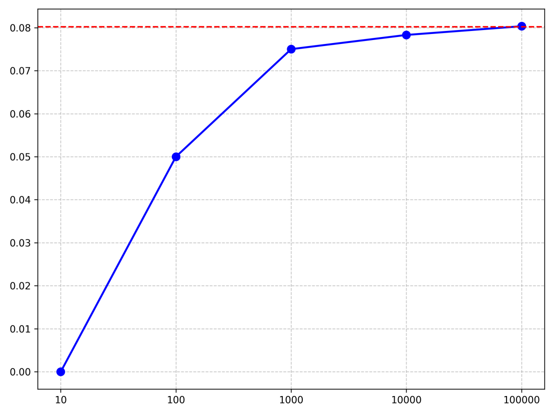


Figure 15.2.: Plot of the normalized counts $\overline{\mathcal{N}}_N([2.4, 6.1])$ found for different values of N (ranging from 10 to 100,000), where $\lambda_j \sim \text{Bin}(N, 1/N)$. Each point represents the normalized count for one diagonal matrix of size N. The experimental values are indeed approaching $\mathbb{P}(\text{Pois}(1) \in [2.4, 6.1]) = 0.08022\ldots$.

15.2. Spectral statistics for $\beta = 2$ ensembles

In this section, we will review and subsequently generalise some results from the first part of the course, including the Gaussian ensembles and their joint probability densities, and orthgonal polynomial techniques. To conclude we will apply *orthogonal polynomials* to understand various spectral statistics (e.g. the expected value of the normalised counting measure, occupation probabilities) for $\beta = 2$ *invariant ensembles*, for example GUE.

Joint eigenvalue probability density

Recall that the Gaussian Orthogonal Ensemble, $\mathrm{GOE}_N \equiv \mathrm{GOE}$, is the collection of all $N \times N$ symmetric matrices $M_N = (m_{ij})_{1\le i,j\le N}$ whose upper triangular entries are independently

$$m_{jj} \sim \mathcal{N}(0,1), \qquad m_{ij} \sim \mathcal{N}(0,1/2), \quad 1 \le i < j \le N. \tag{15.19}$$

Since M_N is symmetric, the number of independent parameters is $\binom{N+1}{2}$. Recall that we can view the GOE as a probability space on $\mathbb{R}^{N(N+1)/2}$ with entries distributed according to the joint p.d.f.

$$\begin{aligned} P_N^{(1)}(M_N) &= \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-m_{jj}^2/2} \prod_{1\le j<k\le N} \frac{1}{\sqrt{\pi}} e^{-m_{j,k}^2} \\ &= 2^{-N/2}\pi^{-N(N+1)/4} \exp\left(-\frac{1}{2}\operatorname{Tr} M_N^2\right). \end{aligned} \tag{15.20}$$

We also saw that we can also find the joint p.d.f. of the *eigenvalues* of M_N,

$$\begin{aligned} P_N^{(1)}(\lambda_1,\dots,\lambda_N) &= Z_N^{(1)} \prod_{1\le j<k\le N} |\lambda_k - \lambda_j| \cdot \exp\left(-\frac{1}{2}\sum_{j=1}^{N}\lambda_j^2\right) \\ &= Z_N^{(1)} |\Delta_N(\lambda_1,\dots,\lambda_N)| \cdot \exp\left(-\frac{1}{2}\sum_{j=1}^{N}\lambda_j^2\right) \end{aligned} \tag{15.21}$$

where recall $\Delta_N(x_1,\dots,x_N)$ is the Vandermonde determinant[1] (exactly equal to the product over j,k in the preceeding line). The prefactor $C_N^{(1)}$ is the appropriate normalising constant to make this a probability measure[2]. The relevance of the superscript (1) will be made clear shortly.

We then saw the Gaussian Unitary Ensemble, $\mathrm{GUE}_N = \mathrm{GUE}$, which is the 'complex' version of the GOE. It is the set of all $N \times N$ matrices $M_N = (m_{ij})$ that are Hermitian

[1] This notation is ubiquitous across RMT, so we will keep it although it might initially be confusing with the notation for a subset $\Delta \subset \mathbb{R}$ used in the preceeding section.

[2] In the case of the Gaussian ensembles, this number (dependent on N) is known by a famous integral equation called Selberg's integral, see (15.28).

($m_{ij} = \overline{m_{ji}}$) with entries independently (up to the self-adjoint condition) distributed according to

$$m_{jj} \sim \mathcal{N}(0,1/2), \qquad m_{ij} \sim \mathcal{N}(0,1/4) + i\mathcal{N}(0,1/4), \quad 1 \le i < j \le N. \tag{15.22}$$

The distribution of the off-diagonal entries is a centred complex Gaussian random variable (of variance 1/2). The joint p.d.f. for the matrix entries is

$$\begin{aligned} P_N^{(2)}(M_N) &= \prod_{j=1}^{N} \frac{1}{\sqrt{\pi}} e^{-m_{jj}^2} \prod_{1\le j<k\le N} \frac{2}{\pi} e^{-2|m_{j,k}|^2} \\ &= 2^{N(N-1)/2} \pi^{-N^2/2} \prod_{j=1}^{N} e^{-m_{jj}^2} \prod_{1\le j<k\le N} e^{-2(\Re m_{ij}^2 + \Im m_{ij}^2)} \\ &= 2^{N(N-1)/2} \pi^{-N^2/2} \exp\left(-\operatorname{Tr} M_N^2\right). \end{aligned} \tag{15.23}$$

Recall that since we are now dealing with complex matrix entries, when calculating with the density we integrate against

$$dM_N = \prod_{j=1}^{N} dm_{jj} \prod_{1\le j<k\le N} d(\Re m_{ij}) d(\Im m_{ij}).$$

Again we can write the joint p.d.f. of the eigenvalues explicitly

$$\begin{aligned} P_N^{(2)}(\lambda_1,\dots,\lambda_N) &= Z_N^{(2)} \prod_{1\le j<k\le N} |\lambda_k - \lambda_j|^2 \cdot \exp\left(-\sum_{j=1}^{N} \lambda_j^2\right) \\ &= Z_N^{(2)} |\Delta_N(\lambda_1,\dots,\lambda_N)|^2 \cdot \exp\left(-\sum_{j=1}^{N} \lambda_j^2\right) \end{aligned} \tag{15.24}$$

where once again $Z_N^{(2)}$ is the appropriate normalizing constant (cf. (15.28)).

Comparing for example (15.21) and (15.24), it might not be surprising that (again!) these are both specialisations of a generalised framework of *invariant ensembles*.

Definition 15.7. *Let $V : \mathbb{R} \to \mathbb{R}_{\ge 0}$ be a sufficiently nice continuous function*[3]*. Let $\mathcal{S}_N$ be the set of real symmetric $N \times N$ matrices, and $\mathcal{H}_N$ the same for Herimitian matrices.*

*The **invariant ensembles** are defined on the set $\mathcal{S}_N$, resp. $\mathcal{H}_N$, as the sets of random matrices whose joint probability density function on the entries is*

$$P_N^{(\beta)}(M_N) = C_N^{(\beta)} \exp\left\{-\frac{\beta}{2} \operatorname{Tr}\left(V(M_N)\right)\right\}. \tag{15.25}$$

The normaization constant $C_N^{(\beta)} > 0$ changes depending on the function V; for $V(x) = x^2$ the constant $C_N^{(\beta)}$ is shown in (15.20) and (15.23) for $\beta = 1, 2$ respectively.

[3]We need V to grow sufficiently fast at $\pm\infty$ to ensure integrability of $\mathbb{R} \ni x \mapsto e^{-V(x)}$.

Remark. *A word of caution: invariant ensembles* (15.25) *do not in general originate from constructions of the type* (15.19) *and* (15.22)*, i.e. by filling up real symmetric or complex Hermitian matrices with families of i.i.d. random variables. Still, constructions of type* (15.19) *and* (15.22) *for families other than Gaussian random variables will appear again in Chapter 17 below. The corresponding random matrices are known as real or complex* ***Wigner matrices****.*

Proposition 15.8. *Consider the invariant ensembles, i.e. $N \times N$ random matrices whose joint density is given by* (15.25) *with $\beta = 1$ on $\mathcal{S}_N$ and with $\beta = 2$ on $\mathcal{H}_N$. Then the joint density for the* eigenvalues *can be written via*

$$P_N^{(\beta)}(\lambda_1, \ldots, \lambda_N) = Z_N^{(\beta)} \exp\left\{ -\frac{\beta}{2} \sum_{j=1}^{N} V(\lambda_j) \right\} \left|\Delta(\lambda_1, \ldots, \lambda_N)\right|^{\beta}, \qquad \beta = 1, 2,$$

where $Z_N^{(\beta)} > 0$ is the normalization constant.

We saw, for example, in the first part of the course via an explicit calculation (cf. Exercise 4 there), that for $V(x) = x^2$, $Z_2^{(1)} = 1/(4\sqrt{\pi})$. For more general N see (15.28) below.

As we will often deal with orthogonal invariant functions of real symmetric matrices[4] or with unitary invariant functions of complex Hermitian matrices, we emphasize that such (class) functions will depend only on eigenvalues! These functions can always be extended from the set defined by (15.2) to the whole of $\mathbb{R}^N$ as symmetric functions.

Example 15.9. *Take the counting measure $\mathcal{N}_2([a,b]) : \mathcal{S}_2 \to \mathbb{C}$ on an interval $[a,b]$ acting on the set*

$$\mathcal{S}_2 = \left\{ \begin{pmatrix} x & y \\ y & x \end{pmatrix}, x, y \in \mathbb{R} \right\}.$$

Since this function only depends on the eigenvalues of a given matrix, it is orthogonally invariant. Given $\mathcal{S} \in \mathcal{S}_2$ with eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ then for $\mathcal{S}$ we have

$$\mathcal{N}_2([a,b]) = \begin{cases} 0 & \lambda_1, \lambda_2 \notin [a,b] \\ 1 & \lambda_1 \in [a,b], \lambda_2 \notin [a,b] \text{ or } \lambda_2 \in [a,b], \lambda_1 \notin [a,b] \\ 2 & \lambda_1, \lambda_2 \in [a,b]. \end{cases}$$

Therefore we could define $f : \mathbb{R}^2 \to \mathbb{C}$ to be the symmetric function such that

$$f(x,y) = \begin{cases} 0 & x, y \notin [a,b] \\ 1 & x \in [a,b], y \notin [a,b] \text{ or } y \in [a,b], x \notin [a,b] \\ 2 & x, y \in [a,b]. \end{cases}$$

The function f is pictured in Figure 15.3 for the interval $[a,b] = [0,1]$.

[4] This is to say, a function $\Phi : \mathcal{S}_N \to \mathbb{C}$ where $\Phi(M_N) = \Phi(O_N M_N O_N^T)$ for any $N \times N$ orthogonal matrix O_N. Since $\operatorname{spec}(O_N M_N O_N^T) = \operatorname{spec}(M_N)$, these Φ only depend on the eigenvalues of M_N.

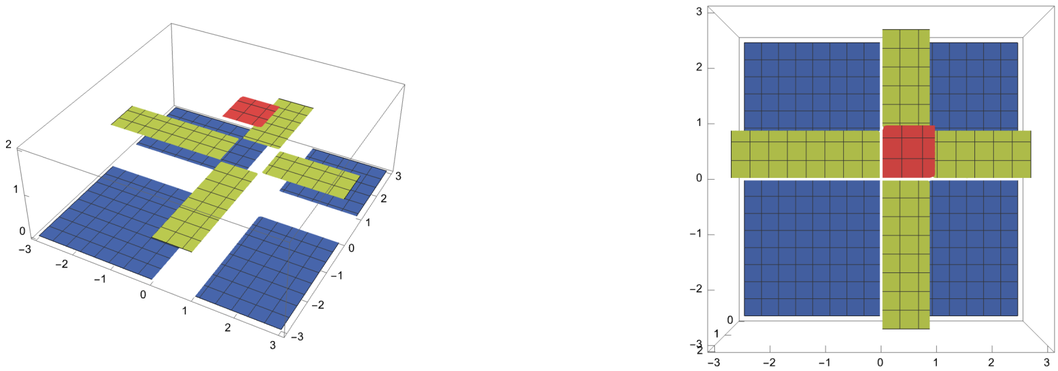


Figure 15.3.: The symmetric extension of $\mathcal{N}_2([0,1])$ (on say $\mathcal{S}_2$) to $\mathbb{R}^2$, shown on the subset $U \subset \mathbb{R}^2$, $U = [-3,3]^2$. Colours are to aid in differentiating the values that f takes.

This allows us to state the following version of Proposition 15.8.

Corollary 15.10. *Let $\Phi : \mathcal{S}_N \to \mathbb{C}$ be an orthogonal invariant function or let $\Phi : \mathcal{H}_N \to \mathbb{C}$ be a unitary invariant function, integrable with respect to (15.25). Denote by $f : \mathbb{R}^N \to \mathbb{C}$ the symmetric function whose restriction to (15.2) is Φ, and by $\mathbb{E}_{\beta,N}[\cdot]$ the expectation with respect to (15.25). Then*

$$\mathbb{E}_{\beta,N}[\Phi] = \int_{\mathbb{R}^N} f(x_1,\ldots,x_N) P_N^{(\beta)}(x_1,\ldots,x_N) \prod_{j=1}^{N} dx_j, \tag{15.26}$$

where

$$P_N^{(\beta)}(x_1,\ldots,x_N) = Z_N^{(\beta)} \exp\left\{-\frac{\beta}{2}\sum_{j=1}^{N} V(x_j)\right\} \left|\Delta(x_1,\ldots,x_N)\right|^{\beta}, \tag{15.27}$$

with a normalizing constant $Z_N^{(\beta)} > 0$.

We re-emphasise that the choice of quadratic $V(x) = x^2$ in (15.27) corresponds to the Gaussian invariant ensembles; the Gaussian orthogonal and unitary ensembles in the case of $\beta = 1,2$. In these cases the normalizing constant $Z_N^{(\beta)}$ is known explicitly,

$$Z_N^{(\beta)} = (2\pi)^{-\frac{N}{2}} \left(\frac{\beta}{2}\right)^{\frac{1}{4}(2N+\beta N(N-1))} \prod_{j=1}^{N} \frac{\Gamma(\beta/2)}{j\Gamma(\beta j/2)} \tag{15.28}$$

where $\Gamma(z)$ is the Gamma function, i.e. $\Gamma(n) = (n-1)!$ for integer n and otherwise it is defined for general $z \in \mathbb{C}$ with positive real part via

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt, \quad \Re(z) > 0.$$

(There is an analytic continuation to the whole complex plane, where $\Gamma(z)$ has simple poles at $-z \in \mathbb{N} \cup \{0\}$.) The expression in (15.28) is found using an explicit integral formula known as **Selberg's integral** (see also Mehta's integral).

Exercise. *Confirm that (15.28) indeed returns $1/(4\sqrt{\pi})$ for the values of $\beta = 1$ and $N = 2$ as previously found in the first part of the course.*

Orthogonal polynomial techniques

Recall that we interested mostly in the asymptotic properties of the eigenvalue distribution of $N \times N$ random matrices as N tends to infinity. This means that we need an efficient method to compute (at least asymptotically) integrals

$$\int_{\mathbb{R}^N} f(x_1, \dots, x_N) P_N^{(\beta)}(x_1, \dots, x_N) dx_1 \cdots dx_N$$

cf. (15.26) for a sufficiently big class of symmetric functions $f : \mathbb{R}^N \to \mathbb{C}$. Think in particular those that determine the basic quantities (15.1), (15.5), (15.6). Achieving this goal is technically most involved for real symmetric matrices, i.e. for $\beta = 1$. The case $\beta = 2$ of complex Hermitian matrices is more feasible. In both cases, $\beta = 1$ and $\beta = 2$, one can use techniques based on *orthogonal polynomials* and on asymptotic formulæ for classical orthogonal polynomials in the case of the GOE and GUE. Some of these techniques were discussed towards the end of the first part of this module and we shall review the essential results below - focusing throughout on the case $\beta = 2$.

We saw that, given a **weight function**[5] $w : \mathbb{R} \to \mathbb{R}_+$, there exists an infinite sequence $(p_j)_{j=0}^{\infty} \subset \mathbb{R}[x]$ of polynomials with real coefficients that satisfy various properties. Here we have collected those most relevant henceforth.

1. The polynomial $p_j(x) = \gamma_{j,j} x^j + \gamma_{j-1,j} x^{j-1} + \cdots + \gamma_{0,j}$ has degree j with leading coefficient $\gamma_j = \gamma_{j,j} > 0$.

2. The collection $(p_j)_{j=0}^{\infty}$ is a sequence of *orthonormal polynomials*[6] with respect to the given weight:

$$(p_j, p_k) = \int_{\mathbb{R}} p_j(x) p_k(x) w(x) dx = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}, \qquad \text{for all } j, k \in \mathbb{Z}_{\geq 0}. \tag{15.29}$$

3. The function

$$K_N(x, y) := \sqrt{w(x)w(y)} \sum_{j=0}^{N-1} \frac{p_j(x) p_j(y)}{(p_j, p_j)} = \sqrt{w(x)w(y)} \sum_{j=0}^{N-1} p_j(x) p_j(y) \tag{15.30}$$

[5]The weight function should additionally satisfy $\int_{\mathbb{R}} |x|^k w(x) dx < \infty$ for all non-negative integers k (think: why?).

[6]In the first part of the course, the choice was made to consider *monic* polynomials (p_j), and hence the polynomials were orthogonal but not necessarily orthonormal with respect to the weight, i.e. $(p_j, p_k) = N_k \delta_{j,k}$ without scaling p_j so that $N_k = 1$. For this part of the course, it is more often useful to have the polynomials scaled to be *orthonormal*, so we henceforth will assume (15.29). Of course, it is just a rescaling of the polynomials to switch between the two standards.

usig orthonormality (called the **reproducing kernel** of the family $(p_j)_{j=0}^{\infty}$) satisfies Gaudin's lemma

$$\int_{\mathbb{R}} K_N(x,x)dx = N, \tag{15.31}$$

$$\int_{\mathbb{R}} K_N(x,z)K_N(z,y)dz = K_N(x,y), \tag{15.32}$$

for all $x, y \in \mathbb{R}$, and hence both the joint eigenvalue density and corresponding correlation functions R_n (hence also the marginals) for a j.p.d.f. of particular shape (think: $\beta = 2$)

$$P_N(x_1,\dots,x_N) = \frac{1}{N!}\prod_{j=1}^{N} w(x_j)|\Delta(x_1,\dots,x_N)|^2 \tag{15.33}$$

can be expressed as a determinant involving the kernel for w:

$$P_N(x_1,\dots,x_N) = \frac{1}{N!}\det_{N\times N}(K_N(x_j,x_k)) \tag{15.34}$$

$$R_n(x_1,\dots,x_n) = \frac{N!}{(N-n)!}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}} P(x_1,\dots,x_N)dx_{n+1}\cdots dx_N = \det_{n\times n}(K_N(x_j,x_k)). \tag{15.35}$$

4. There exists a sequence $(b_j)_{j=0}^{\infty}$ of real numbers and a sequence $(a_j)_{j=0}^{\infty}$ of nonnegative real numbers such that the following three-term recurrence relation is valid for $j \in \mathbb{Z}_{\geq 0}$,

$$xp_j(x) = a_{j+1}p_{j+1}(x) + b_jp_j(x) + a_jp_{j-1}(x), \tag{15.36}$$

where

$$\begin{aligned} a_0 &= 0, \\ a_j &= \int_{\mathbb{R}} xp_{j-1}(x)p_j(x)w(x)dx, && j \in \mathbb{Z}_{\geq 1}; \\ b_j &= \int_{\mathbb{R}} p_j^2(x)w(x)dx, && j \in \mathbb{Z}_{\geq 0}. \end{aligned}$$

The kernel similarly satisfies

$$K_N(x,y) = a_N\sqrt{w(x)w(y)}\frac{p_N(x)p_{N-1}(y) - p_{N-1}(x)p_N(y)}{x-y}. \tag{15.37}$$

This last relation is known as the **Christoffel-Darboux identity**. On the 'diagonal' we have, taking $y \to x$

$$K_N(x,x) = a_N\left(\sqrt{w(x)p_N(x)}\cdot\frac{d}{dx}\{w(x)p_{N-1}(x)\} - \sqrt{w(x)p_{N-1}(x)}\cdot\frac{d}{dx}\{w(x)p_N(x)\}\right).$$

Example 15.11. *In this example, we will see that we can use orthogonal polynomial techniques to show* (15.34)*:*

$$P_N(x_1,\ldots,x_N) = \frac{1}{N!}\det_{N\times N}(K_N(x_j,x_k))$$

for the joint p.d.f. $P(\underline{x})$ *from* (15.33)*. Previously (in the first part of the course) this relationship was found via the Transposing lemma. Here we explicitly appeal to orthogonality (note also here our polynomials are orthonormal and not necessarily monic).*

Let $w:\mathbb{R}\to\mathbb{R}_+$ *be a weight function and* $(p_j)_{j=0}^{\infty}$ *be the corresponding family of orthonormal polynomials, so* $p_j(x) = \gamma_j x^j + \cdots$*. Then, manipulating the Vandermonde determinant in* (15.33) *we can establish via row operations*

$$\begin{aligned}\Delta(x_1,\ldots,x_N) &= \det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{N-1} & x_2^{N-1} & \cdots & x_N^{N-1}\end{pmatrix} \\ &= \frac{1}{\prod_{j=0}^{N-1}\gamma_j}\begin{pmatrix} p_0(x_1) & p_0(x_2) & \cdots & p_0(x_N) \\ p_1(x_1) & p_1(x_2) & \cdots & p_1(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ p_{N-1}(x_1) & p_{N-1}(x_2) & \cdots & p_{N-1}(x_N)\end{pmatrix}.\end{aligned}$$

Thus

$$P(x_1,\ldots,x_N) = \frac{1}{N!}\left[\prod_{j=0}^{N-1}\frac{1}{\gamma_j^2}\right]\left(\det_{N\times N}\left(\sqrt{w(x_k)}p_{j-1}(x_k)\right)\right)^2 = \frac{1}{N!}\left[\prod_{j=0}^{N-1}\frac{1}{\gamma_j^2}\right]\det_{N\times N}(K_N(x_j,x_k))$$

using the definition of the reproducing kernel (15.30)*. By Andréief's identity (cf.* (A.3)*) with* $f_j(x) = g_j(x) = \sqrt{w(x)}p_{j-1}(x)$ *we find*

$$\begin{aligned}\frac{1}{N!}\int_{\mathbb{R}^N}\det_{N\times N}(f_j(x_k))\cdot\det_{N\times N}(g_j(x_k))dx_1\cdots dx_N &= \det_{N\times N}\left(\int_{\mathbb{R}} f_j(x)g_k(x)dx\right) \\ &= \det_{N\times N}\left(\int_{\mathbb{R}} w(x)p_{j-1}(x)p_{k-1}(x)dx\right) \\ &= 1.\end{aligned}$$

Therefore

$$\begin{aligned}\int_{\mathbb{R}^N} P(x_1,\ldots,x_N)dx_1\cdots dx_N &= \frac{1}{N!}\left[\prod_{j=0}^{N-1}\frac{1}{\gamma_j^2}\right]\int_{\mathbb{R}^N}\left(\det_{N\times N}\left(\sqrt{w(x_k)}p_{j-1}(x_k)\right)\right)^2 dx_1\cdots dx_N \\ &= \prod_{j=0}^{N-1}\frac{1}{\gamma_j^2}.\end{aligned}$$

Finally, we see that the left hand side of the last set of equations is the density function integrated over the whole sample space and therefore is equal to 1*. Hence* $\gamma_0 \cdot \dots \cdot \gamma_{N-1} = 1$ *and*

$$P(x_1, \dots, x_N) = \frac{1}{N!} \det_{N \times N} (K_N(x_j, x_k)).$$

Next is the notion of the *Fredholm determinant* of a continuous kernel function.

Definition 15.12. *Let* $J \subset \mathbb{R}$ *be a compact interval, let* $K : J \times J \to \mathbb{C}$ *be a continuous function (kernel), and let* K_J *be the associated integral operator:*

$$(K_J f)(x) \coloneqq \int_J K(x, y) f(y) dy, \qquad x \in J,$$

acting on functions $f : J \to \mathbb{C}$ *that are square integrable, i.e.*

$$\int_J |f(x)|^2 dx < \infty.$$

The ***Fredholm determinant*** *of the operator* K_J *(or the kernel* K*) is*

$$\det(I - K_J) \coloneqq 1 + \sum_{\ell=1}^{\infty} \frac{(-1)^\ell}{\ell!} \left\{ \int_{J^\ell} \det \left(K(x_j, x_k) \right)_{j,k=1}^{\ell} \prod_{m=1}^{\ell} dx_m \right\}. \tag{15.38}$$

Remark. *Now, it is not clear apriori that* (15.38)*, which involves an infinite sum, will converge and hence describe a uesful quantity. We will often use the Fredholm determinant of the reproducing kernel* K_N *of* (15.30)*. In this case* K_J *has rank* N*, and so the sum in* (15.38) *truncates at* $\ell = N$*, i.e. convergence in* (15.38) *is ensured. In the general case of a continuous kernel* K *the convergence of the series in* (15.38) *must be deduced by other means, for instance by* ***Hadamard's inequality*** *(cf.* (A.2)*).*

The next Proposition displays the utility of orthogonal polynomials in the study of unitary invariant matrix models, i.e. when it comes to the calculation of the averages (15.26) for $\beta = 2$. For example, as we will see, it will allow for another method of showing (15.35) (previously established by Gaudin's lemma).

Proposition 15.13. *Consider the invariant ensemble* (15.25) *with* $\beta = 2$ *and continuous*[7] $V : \mathbb{R} \to \mathbb{R}_{\geq 0}$*. Let* $(p_j)_{j=0}^{\infty}$ *be the family of orthonormal polynomials with respect to the weight* $w(x) = \exp(-V(x))$*:*

$$\int_{\mathbb{R}} p_j(x) p_k(x) e^{-V(x)} dx = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases},$$

for all $j, k \in \mathbb{Z}_{\geq 0}$ *and all* $n \in \mathbb{N}$*.*

[7] Recall we also need V to grow sufficiently fast at $\pm\infty$. The precise condition we need is $V(x) \geq (2 + \varepsilon) \ln(1 + |x|)$ for all $x \in \mathbb{R}$, for some $\varepsilon > 0$.

Then, the generating function of the marginal densities for $P_N^{(2)}$ (cf. (15.7), (15.11)) can be expressed as a Fredholm determinant:

$$\mathbb{E}_N\left[\prod_{j=1}^{N}(1-\varphi(\lambda_j))\right] = \det(I - K_{N,J}M_\varphi), \tag{15.39}$$

where M_φ is the operator of multiplication by the test function φ of compact support $J = \operatorname{supp}(\varphi)$, and the integral operator $K_{N,J}$ is defined by the reproducing kernel (15.30).

Proof. By definition with $\beta = 2$ (cf. (15.27)),

$$P_N^{(2)}(x_1,\ldots,x_N) = Z_N^{(2)}\exp\left\{-\sum_{j=1}^{N}V(x_j)\right\}\left|\Delta(x_1,\ldots,x_N)\right|^2.$$

Set $f_j(x) = (1-\varphi(x))\sqrt{w(x)}p_{j-1}(x)$, $g_j(x) = \sqrt{w(x)}p_{j-1}(x)$. Using (15.34) and subsequently Andréief's identity (A.3) we have

$$\begin{aligned}
\mathbb{E}_N\left[\prod_{j=1}^{N}(1-\varphi(x_j))\right] &= \int_{\mathbb{R}^N}\prod_{j=1}^{N}(1-\varphi(x_j))P_N^{(2)}(x_1,\ldots,x_N)dx_1\cdots dx_N \\
&= \frac{1}{N!}\int_{\mathbb{R}^N}\prod_{j=1}^{N}(1-\varphi(x_j))\left(\det_{N\times N}(\exp(-V(x_k)/2)p_{j-1}(x_k))\right)^2 dx_1\cdots dx_N \\
&= \frac{1}{N!}\int_{\mathbb{R}^N}\det(f_j(x_k))\det(g_j(x_k))dx_1\cdots dx_N \\
&= \det\left(\int_{\mathbb{R}}(1-\varphi(x))w(x)p_{j-1}(x)p_{k-1}(x)dx\right) \\
&= \det\left(\delta_{j,k}-\xi_{j,k}\right)_{j,k=1}^{N}
\end{aligned}$$

where

$$\xi_{j,k} = \int_{\mathbb{R}}\varphi(x)w(x)p_{j-1}(x)p_{k-1}(x)dx.$$

The next task is to unpick this determinant. This can be directly done using an expansion in terms of the matrix minors (cf. A.4), yielding

$$\det(\delta_{j,k}-\xi_{j,k})_{j,k=1}^{N} = 1 + \sum_{\ell=1}^{N}\frac{(-1)^\ell}{\ell!}\sum_{m_1,\ldots,m_\ell=1}^{N}\det(\xi_{j,k})_{j,k\in\{m_1,\ldots,m_\ell\}}, \tag{15.40}$$

and using the explicit expression of $\xi_{j,k}$

$$\begin{aligned}
\det(\xi_{j,k})_{j,k\in\{m_1,\dots,m_\ell\}} &= \det(\xi_{m_j,m_k})_{j,k=1}^{\ell} \\
&= \sum_{\sigma\in S_\ell} \operatorname{sgn}(\sigma) \prod_{j=1}^{\ell} \xi_{m_j,\sigma(m_j)} \\
&= \int_{\mathbb{R}^\ell} \left[\sum_{\sigma\in S_\ell} \operatorname{sgn}(\sigma) \prod_{j=1}^{\ell} p_{m_j-1}(x_j) p_{\sigma(m_j)-1}(x_j) \right] \prod_{m=1}^{\ell} w(x_m)\varphi(x_m) dx_m \\
&= \int_{\mathbb{R}^\ell} \left[\sum_{\sigma\in S_\ell} \operatorname{sgn}(\sigma) \prod_{j=1}^{\ell} p_{m_j-1}(x_j) p_{m_j-1}(x_{\sigma(j)}) \right] \prod_{m=1}^{\ell} w(x_m)\varphi(x_m) dx_m \\
&= \int_{\mathbb{R}^\ell} \det\left[p_{m_j-1}(x_j) p_{m_j-1}(x_k) \right]_{j,k=1}^{\ell} \prod_{m=1}^{\ell} w(x_m)\varphi(x_m) dx_m.
\end{aligned}$$

Hence, combining the above and using multilinearity of the determinant,

$$\begin{aligned}
&\mathbb{E}_N\left[\prod_{j=1}^{N} (1-\varphi(x_j)) \right] \\
&\quad = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{\mathbb{R}^\ell} \sum_{m_1,\dots,m_\ell=1}^{N} \det\left(p_{m_j-1}(x_j) p_{m_j-1}(x_k) \right)_{j,k=1}^{\ell} \prod_{m=1}^{\ell} w(x_m)\varphi(x_m) dx_m. \\
&\quad = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{\mathbb{R}^\ell} \det\left(\sum_{r=1}^{N} p_{r-1}(x_j) p_{r-1}(x_k) \right)_{j,k=1}^{\ell} \prod_{m=1}^{\ell} w(x_m)\varphi(x_m) dx_m \\
&\quad = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{J^\ell} \det\left(\sum_{r=1}^{N} \sqrt{w(x_j)w(x_k)}\, p_{r-1}(x_j) p_{r-1}(x_k) \right)_{j,k=1}^{\ell} \prod_{m=1}^{\ell} \varphi(x_m) dx_m \\
&\quad = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{J^\ell} \det\left(K_N(x_j,x_k)\varphi(x_k) \right)_{j,k=1}^{\ell} \prod_{m=1}^{\ell} dx_m \\
&\quad \overset{(15.38)}{=} \det(I - K_{N,J} M_\varphi),
\end{aligned} \tag{15.41}$$

as claimed. □

Remark. *Notice that in Proposition 15.13 we didn't ever use the actual form of the weight other than that the density function could be written as a constant multiple of*

$$\prod_{j=1}^{N} w(x_j) \left|\Delta(x_1,\dots,x_N)\right|^2.$$

Therefore the statement holds more widely than just $w(x) = \exp(-V(x))$ provided the density function has this shape (cf. also (15.33)–(15.35)). However, we will be particularly interested in the case of $w(x) = \exp(-x^2)$ in the forthcoming.

Example 15.14 (Determinantal formulae via orthogonal polynomials). *Recall that above we saw a reproof using orthogonal polynomials (previously seen via the Transposing lemma) that the joint probability density function for eigenvalues of $\beta = 2$ invariant ensembles has a determinantal form:*

$$P_N^{(\beta)}(x_1, \ldots, x_N) = \frac{1}{N!} \det_{N \times N}(K_N(x_j, x_k)).$$

Earlier in the course, we have also already seen (via Gaudin's lemma) that as well the marginal densities and correlation functions can be expressed via determinants

$$P_{N,\ell}(x_1, \ldots, x_\ell) = \frac{(N-\ell)!}{N!} \det(K_N(x_j, x_k))_{j,k=1}^{\ell}$$
$$R_\ell(x_1, \ldots, x_\ell) = \det(K_N(x_j, x_k))_{j,k=1}^{\ell}$$

for $\ell = 1, \ldots, N$.

However, by. (15.12), *we have*

$$\mathbb{E}_N\left[\prod_{j=1}^{N}(1 - \varphi(x_j))\right] = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{\mathbb{R}^\ell \ell} R_\ell(x_1, \ldots, x_\ell) \prod_{j=1}^{\ell} \varphi(x_j) dx_j.$$

Therefore the determinantal formula for $R_\ell(x_1, \ldots, x_\ell)$ *(and hence for* $P_{N,\ell}$*) also follows immediately upon comparison with the equation preceeding* (15.41) *(which is underpinned by the orthogonal polynomial structure).*

Proposition 15.13 allows us to express the quantities of primary interest, e.g. the expectation of the normalized counting measure $\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]$, the occupation probabilities $\mathbb{P}(\mathcal{N}_N(\Delta) = \ell)$, the gap probability $\mathbb{P}(\mathcal{N}_N(\Delta) = 0)$ and the variance of the linear statistic $\mathrm{Var}_N(\mathcal{N}_N[\phi])$ via the reproducing kernel (15.30). This is because, as we saw in the first part of this chapter, all the aforementioned quantities can be connected via the generating function $\mathbb{E}_N[\prod(1 - \varphi(x_j))]$.

Before we state the relationships, we record the following rescaling of the orthogonal polynomials,

$$\psi_j(x) \coloneqq \sqrt{w(x)} p_j(x) \tag{15.42}$$

where $(p_j)_{j\geq 0}$ are orthonormal with respect to the weight w. This rewriting will simplify notation in the following.

Theorem 15.15. *Consider an invariant ensemble defined by* (15.25) *with* $\beta = 2$. *Then*

1. *For any bounded function* $\varphi : \mathbb{R} \to \mathbb{C}$ *the expectation of the linear eigenvalue statistic equals*

$$\mathbb{E}_N\left[\mathcal{N}_N[\varphi]\right] = N \int_{\mathbb{R}} \varphi(x) \rho_N(x) dx, \tag{15.43}$$

where, for any $x \in \mathbb{R}$,

$$\begin{aligned} \rho_N(x) \coloneqq \frac{1}{N} K_N(x, x) &= \frac{1}{N} \sum_{j=0}^{N-1} (\psi_j(x))^2 \\ &= \frac{a_N}{N}\Big(\psi_{N-1}(x) \frac{d}{dx} \psi_N(x) - \psi_N(x) \frac{d}{dx} \psi_{N-1}(x)\Big). \end{aligned} \tag{15.44}$$

In particular, the expectation of the normalized counting measure $\overline{\mathcal{N}}_N(\Delta)$ *admits the representation, for* $\Delta \subset \mathbb{R}$

$$\mathbb{E}_N\left[\overline{\mathcal{N}}_N(\Delta)\right] = \int_\Delta \rho_N(x)dx. \tag{15.45}$$

Further, we have the inequality

$$\left|K_N(x,y)\right|^2 \leq K_N(x,x)K_N(y,y) = N^2\rho_N(x)\rho_N(y) \tag{15.46}$$

which holds for all $x, y \in \mathbb{R}$.

2. *The variance of the linear eigenvalue statistic is*

$$\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) = \frac{1}{2}\int_{\mathbb{R}^2}\left|\varphi(x)-\varphi(y)\right|^2 K_N^2(x,y)dxdy. \tag{15.47}$$

3. *The probability* $\mathbb{P}_N(\mathcal{N}_N(\Delta)=\ell)$ *that the interval* Δ *contains exactly* ℓ *eigenvalues is*

$$\mathbb{P}_N(\mathcal{N}_N(\Delta)=\ell) = \frac{(-1)^\ell}{\ell!}\frac{d^\ell}{da^\ell}\det(I-K_{N,\Delta}M_{\Phi_a})\Big|_{a=1}, \tag{15.48}$$

where $\det(I-K_{N,\Delta}M_{\Phi_a})$ *is the Fredholm determinant (cf. (15.39)) with* $\Phi_a = a\cdot\chi_\Delta$ *for* $a \in \mathbb{R}$ *and* $\mathrm{supp}(\varphi) = J = \Delta$, *and* $K_{N,\Delta}$, *as in (15.39), the reproducing kernel (15.30). In particular, the gap probability of the ensemble is*

$$\mathbb{P}_N(\mathcal{N}_N(\Delta)=0) = \det(I-K_{N,\Delta}) = 1+\sum_{\ell=1}^{N}\frac{(-1)^\ell}{\ell!}\int_{\Delta^\ell}\det(K_N(x_j,x_k))_{j,k=1}^{\ell}\prod_{m=1}^{\ell}dx_m. \tag{15.49}$$

Proof. For (15.43) and (15.45)

$$\begin{aligned}
\mathbb{E}_N\left[\mathcal{N}_N[\varphi]\right] &= \sum_{j=1}^{N}\mathbb{E}_N[\varphi(\lambda_j)] \\
&= N\int_{\mathbb{R}^N}\varphi(x_1)P_N(x_1,\ldots,x_N)\prod_{j=1}^{N}dx_j \\
&= N\int_{\mathbb{R}}\varphi(x)P_{N,1}(x)dx \\
&= \int_{\mathbb{R}}\varphi(x)K_N(x,x)dx
\end{aligned}$$

which proves (15.43), (15.44), (15.45). For (15.46) we simply use Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|K_N(x,y)\right| \leq \sum_{\ell=0}^{N-1}\left|\psi_\ell(x)\right|\left|\psi_\ell(y)\right| &\leq \sqrt{\sum_{\ell=0}^{N-1}\left(\psi_\ell(x)\right)^2}\sqrt{\sum_{\ell=0}^{N-1}\left(\psi_\ell(y)\right)^2} \\
&= \sqrt{K_N(x,x)}\sqrt{K_N(y,y)} \\
&= N\sqrt{\rho_N(x)\rho_N(y)}.
\end{aligned}$$

To prove (15.47), we unpack the variance

$$\begin{aligned}\mathrm{Var}_N(\mathcal{N}_N[\varphi]) &= \mathbb{E}_N[|\mathcal{N}_N[\varphi]|^2] - \left|\mathbb{E}_N[\mathcal{N}_N[\varphi]]\right|^2 \\ &= \int_{\mathbb{R}} |\varphi(x)|^2 R_1(x)dx + \int_{\mathbb{R}^2} \varphi(x)\overline{\varphi(y)}(R_2(x,y) - R_1(x)R_1(y))dxdy.\end{aligned}$$

Then, writing $T_2(x,y) = R_1(x)R_1(y) - R_2(x,y)$ (this is also known as the 'second cluster function'), we have that

$$\int_{\mathbb{R}} T_2(x,y)dy = R_1(x)$$

so

$$\int_{\mathbb{R}} |\varphi(x)|^2 R_1(x)dx = \int_{\mathbb{R}^2} |\varphi(x)|^2 T_2(x,y)dxdy$$

and hence

$$\mathrm{Var}_N(\mathcal{N}_N[\varphi]) = \frac{1}{2}\int_{\mathbb{R}^2} |\varphi(x) - \varphi(y)|^2 T_2(x,y)dxdy. \tag{15.50}$$

Now,

$$T_2(x,y) = K_N(x,x)K_N(y,y) - \left(K_N(x,x)K_N(y,y) - K_N(x,y)K_N(y,x)\right) = K_N^2(x,y),$$

which establishes (15.47). To prove (15.48) and (15.49) we use Lemma 15.5 together with (15.39) with $\varphi = g\chi_\Delta$. □

Remark. *The determinantal formulæ for $R_n(x_1,\ldots,x_n)$ and $P_{N,n}(x_1,\ldots,x_n)$ are the key ingredient of the orthogonal polynomial technique. They play an important role in the analysis of the $\beta = 2$ invariant ensembles since they reduce the analysis of the correlation functions of any order $\ell \in \mathbb{Z}_{\geq 1}$ to that of the kernel K_N and then to the polynomials p_N and p_{N-1} in view of the Christoffel-Darboux identity (15.37). We will use the same formulæ in Chapter 16 below. Various versions of such formulæ including discrete degenerations or matrix generalizations thereof have found applications in probability theory, combinatorics, group representation theory and statistical physics.*

We close out the current subsection with the following useful variance estimates:

Corollary 15.16. *Let $\mathcal{N}_N[\varphi]$ be a linear statistic corresponding to an invariant matrix model with $\beta = 2$. Then we have the following three estimates:*

1. *For any bounded continuous function $\varphi : \mathbb{R} \to \mathbb{C}$,*

$$\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) \leq 2N \sup_{x\in\mathbb{R}} |\varphi(x)|^2. \tag{15.51}$$

2. *For any Lipschitz continuous function $\varphi : \mathbb{R} \to \mathbb{C}$ that satisfies*

$$|\varphi(x) - \varphi(y)| \leq c|x-y|^\alpha$$

with $c > 0$ and some $0 < \alpha \leq 1$, we have

$$\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) \leq C_{\alpha,N} N^{1-\alpha} \tag{15.52}$$

with $C_{\alpha,N} = c^2 2^{\alpha-1}(a_N)^{2\alpha}$ and a_N as the coefficient in front of p_N in the three-term recurrence relation (15.36) for $j = N-1$.

3. *For any continuous function $\varphi : \mathbb{R} \to \mathbb{C}$ of compact support and any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) \leq N\varepsilon^2 + 4\left(\frac{a_N}{\delta}\right)^2 \sup_{x\in\mathbb{R}} |\varphi(x)|^2. \tag{15.53}$$

In particular, if a_N is bounded (again, the coefficient from the three-term recurrence (15.36)), then

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) = 0.$$

Proof. By using (15.47) and the properties (15.31) and (15.32) of the reproducing kernel, we find

$$\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) = \frac{1}{2}\int_{\mathbb{R}^2} \left|\varphi(x) - \varphi(y)\right|^2 K_N^2(x,y)dxdy \leq 2\|\varphi\|_\infty^2 \int_{\mathbb{R}^2} K_N^2(x,y)dydx = 2N\|\varphi\|_\infty^2,$$

where $\|\varphi\|_\infty := \sup_{x\in\mathbb{R}} |\varphi(x)|$. This establishes (15.51) and we emphasize that the same bound is independent of the choice of function V in the density. Moving now to (15.52), let $\alpha = 1$ and start again from (15.47). Then by (15.37), the Lipschitz constraint and orthonormality,

$$\begin{aligned}\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) &\leq \frac{c^2}{2}\int_{\mathbb{R}^2} (x-y)^2 K_N^2(x,y)dxdy \\ &= \frac{c^2}{2}a_N^2 \int_{\mathbb{R}^2} \left(\psi_N(x)\psi_{N-1}(y) - \psi_{N-1}(x)\psi_N(y)\right)^2 dxdy = (c\cdot a_N)^2.\end{aligned}$$

This is (15.52) for $\alpha = 1$. For $\alpha \in (0,1)$ we obtain from (15.47), the Lipschitz constraint and Hölder's inequality (A.1) (used here with $p = 1/\alpha > 1$ and $q = 1/(1-\alpha) > 1$),

$$\begin{aligned}\mathrm{Var}_N\left(\mathcal{N}_N[\varphi]\right) &\leq \frac{c^2}{2}\int_{\mathbb{R}^2} |x-y|^{2\alpha} K_N^2(x,y)dxdy \\ &\leq \frac{c^2}{2}\left(\int_{\mathbb{R}^2} |x-y|^2 K_N^2(x,y)dxdy\right)^{\alpha}\left(\int_{\mathbb{R}^2} K_N^2(x,y)dxdy\right)^{1-\alpha} \\ &= \frac{c^2}{2}2^\alpha a_N^{2\alpha} N^{1-\alpha},\end{aligned}$$

where we evaluated the first integral as before when $\alpha = 1$ and the second one via the reproducing property (15.32). The last estimate yields (15.52) for $\alpha \in (0,1)$. Moving finally to (15.53), since φ is continuous, for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $|\varphi(x) - \varphi(y)| < \varepsilon$ whenever $x, y \in \mathrm{supp}(\varphi)$ and $|x-y| < \delta$. Hence we write the integral in (15.47) as the sum of the integral over $|x-y| < \delta$ and $|x-y| \geq \delta$. In the first integral we then replace $|\varphi(x) - \varphi(y)|$ by ε and afterwards extend the integral to all of $\mathbb{R}^2$. This and the reproducing property (15.32) yield the bound $N\varepsilon^2/2$ for the first integral. In the second integral we instead replace $K_N^2(x,y)$ by $\delta^{-2}(x-y)^2 K_N^2(x,y)$ and then extend the obtained integral to all of $\mathbb{R}^2$. Then the Christoffel-Darboux identity (15.37) and the argument used in the proof of (15.52) yield (15.53). □

Remark. *The coefficients a_N are bounded in many interesting cases. For the* GUE *this follows from the formula* (15.56) *below. In turn, assertions* (15.51), (15.52) *and* (15.53) *yield in these interesting cases*

$$\lim_{N\to\infty} \mathrm{Var}_N\left(\overline{\mathcal{N}}_N[\varphi]\right) = 0,$$

for the normalized linear statistic. In turn, by Chebyshev's inequality (A.4),

$$\lim_{N\to\infty} \mathbb{P}_N\left(\left|\overline{\mathcal{N}}_N[\varphi] - \mathbb{E}_N\left[\overline{\mathcal{N}}_N[\varphi]\right]\right| > \varepsilon\right) = 0, \tag{15.54}$$

for any $\varepsilon > 0$. The limit law (15.54) *constitutes a generalization of the previous weak law of large numbers* (15.15) *for $\overline{\mathcal{N}}_N(\Delta)$ in the context of diagonal random matrices.*

Hermite polynomials

At the end of the first part of this module, you met some 'classical' weight functions

$$w(x) = \begin{cases} e^{-x^2} & \text{for } \lambda \in \mathbb{R}, \quad \text{Hermite} \\ x^a e^{-x} & \text{for } \lambda > 0, a \in \mathbb{R}, \quad \text{Laguerre} \\ (1-x)^a(1+x)^b & \text{for } \lambda \in (0,1), a, b > -1, \quad \text{Jacobi} \\ (1+x^2)^{-a} & \text{for } x \in \mathbb{R}, a \geq 1/2, \quad \text{Cauchy.} \end{cases}$$

which each subsequently generate a family (p_j) of orthogonal polynomials. We here focus on the **Hermite** case as it relates to the GUE ($\beta = 2$ and $V(x) = x^2$ in (15.25)).

Define the classical **Hermite polynomials** $(H_j)_{j=0}^{\infty} \subset \mathbb{R}[x]$ via the contour integral

$$H_j(x) := \frac{j!}{2\pi i} \int_{\Gamma} e^{2xz - z^2} \frac{dz}{z^{j+1}}, \tag{15.55}$$

where the contour $\Gamma = S_1 = \{z \in \mathbb{C} : |z| = 1\}$, the unit circle in the complex plane is traversed once counterclockwise.

Example 15.17. *Let's see the Hermite polynomials explicitly. Firstly, for $j = 0$,*

$$H_0(x) = \frac{1}{2\pi i} \int_{\Gamma} \frac{e^{2xz-z^2}}{z} dz = \lim_{z\to 0} e^{2xz-z^2} = 1$$

by the residue theorem. Similarly

$$H_1(x) = \frac{1}{2\pi i} \int_{\Gamma} \frac{e^{2xz-z^2}}{z^2} dz = \lim_{z\to 0} \frac{d}{dz} e^{2xz-z^2} = \lim_{z\to 0} (2x - 2z) e^{2xz-z^2} = 2x.$$

Theorem 15.18. *Consider the* GUE *given by* (15.25), (15.27) *for $V(x) = x^2$ and $\beta = 2$. Then the orthonormal polynomials $(p_j)_{j=0}^{\infty} \subset \mathbb{R}[x]$ with respect to the weight $w(x) = \exp(-V(x))$ are*

$$p_j(x) = \gamma_j H_j(x).$$

where $\gamma_j^{-2} := \sqrt{\pi}\, 2^j j!$. *In particular, the coefficient* a_j *in the three-term recurrence for* $(p_j)_{j=0}^{\infty}$ *equals*

$$a_j = \sqrt{\frac{j}{2}} \tag{15.56}$$

for $j \in \mathbb{N}$.

Proof. The Hermite polynomials obey

$$\int_{\mathbb{R}} H_j(x) H_k(x) e^{-x^2} dx = \sqrt{\pi}\, 2^j j! \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

which after rescaling yields the expression for $p_j(x)$. By using the three-term recurrence relation

$$x H_j(x) = \frac{1}{2} H_{j+1}(x) + j H_{j-1}(x),$$

for $j = 0, 1, 2, \ldots$ and $x \in \mathbb{R}$, we obtain (15.56). $\square$

16. Gaussian unitary ensemble

In this chapter we use the orthogonal polynomial techniques of Section 15.2, in particular Theorem 15.15, in the asymptotic analysis of eigenvalues of the Gaussian Unitary Ensemble. We first address the global scaling regime and then pass to the study of the local bulk and edge regimes. This is possible because in the GUE case, the corresponding orthogonal polynomials can be expressed via Hermite polynomials whose asymptotic properties are well known.

16.1. Rescaling the GUE

We emphasise again that we will be interested in taking the matrix size $N \to \infty$. It thus transpires that it is more natural to consider a *rescaling* of our matrix from the GUE. Take such an $N \times N$ matrix $A = (a_{j,k})$ and assume it has eigenvalues $\alpha_1 < \cdots < \alpha_N$. Then for example

$$\sum_{j=1}^{N} \alpha_j^2 \le N \max_{j=1,\dots,N} \{\alpha_j^2\} = N \max\{\alpha_1^2, \alpha_N^2\}.$$

The left hand side (after diving by N) is the linear statistic $\frac{1}{N}\sum_{j=1}^{N} \alpha_j^2$. By the spectral mapping theorem,

$$\frac{1}{N}\sum_{j=1}^{N} \alpha_j^2 = \frac{1}{N}\operatorname{Tr}(A^2) = \frac{1}{N}\sum_{j,k=1}^{N} a_{j,k}a_{kj} = \frac{1}{N}\sum_{j,k=1}^{n} |a_{j,k}|^2, \tag{16.1}$$

since A is Hermitian and using (15.23),

$$\mathbb{E}_N\left[\frac{1}{N}\sum_{j,k=1}^{N} |a_{j,k}|^2\right] = \frac{1}{N}\left(\sum_{j=1}^{N}\mathbb{E}\left[a_{11}^2\right] + 2\sum_{\substack{j,k=1\\ j<k}}^{N}\mathbb{E}\left[|a_{12}|^2\right]\right) = \mathbb{E}\left[a_{11}^2\right] + (N-1)\mathbb{E}\left[|a_{12}|^2\right].$$

Thus

$$\mathbb{E}\left[a_{11}^2\right] + (N-1)\mathbb{E}\left[|a_{12}|^2\right] \le \mathbb{E}\left[\max\{\alpha_1^2, \alpha_N^2\}\right].$$

Consequently, if we take $N \to \infty$ then the left hand side will diverge, forcing the right hand side to diverge (i.e. the biggest – in magnitude – eigenvalue will have infinite variance). So, in order to get a non-trivial scaling limit in some sense, we should scale the entries $a_{j,k}$ by $1/\sqrt{N}$, compensating for the linear growth in N on the off diagonal second moment.

To this end, we will henceforth be considering *rescaled* matrices $\mathrm{M_N} = (1/\sqrt{N})M_N$ of the GUE. The diligent reader can comb through the previous section and see how this scaling will affect the previous results, but we collect here the most pertinent:

- the j.p.d.f. of the rescaled matrix entries $\mathrm{M_N} = (m_{j,k})$ and eigenvalues $\lambda_1, \dots, \lambda_N$ are respectively

$$\mathrm{P}_N(\mathrm{M_N}) \propto \exp\left\{-N \operatorname{Tr}(\mathrm{M_N}^2)\right\} \tag{16.2}$$

$$\mathrm{P}_N(\lambda_1, \dots, \lambda_N) \propto \exp\left\{-N \sum_{j=1}^{N} \lambda_j^2\right\} \left|\Delta(\lambda_1, \dots, \lambda_N)\right|^2. \tag{16.3}$$

- the associated weight $w_N(x) = \exp(-Nx^2)$is now known also as a *varying weight* (since it depends on N) and consequently the jth *orthonormal polynomial* for the rescaled GUE is (now making the N-dependence explicit in the superscript notation)

$$p_j^{(N)}(x) = N^{1/4} \cdot \gamma_j \cdot H_j(\sqrt{N}x) \tag{16.4}$$

where $H_j(x)$ is the jth Hermite polynomial (15.55), and $\gamma_j^{-2} = \sqrt{\pi}2^j \cdot j!$ is the same factor as previously. The equivalent form of (15.42) here is

$$\psi_j^{(N)}(x) = \sqrt{w_N(x)}p_j^{(N)}(x). \tag{16.5}$$

The coefficient $a_j^{(N)}$ in the three-term recurrence for $(p_j^{(N)})_{j\geq 0}$ is

$$a_j^{(N)} = \sqrt{\frac{j}{2N}}. \tag{16.6}$$

Hence, for example, $p_0^{(N)}(x) = (N/\pi)^{1/4}$ and $p_1^{(N)}(x) = (4N^3/\pi)^{1/4}x$.

Henceforth, for the rest of this chapter (unless explicitly otherwise indicated), we assume that we are working with $\mathrm{M_N}$. Since we will want to analyse e.g. the quantities in Theorem 15.15 as $N \to \infty$, it will serve us well to first understand how the orthonormal polynomials $(p_j^{(N)})_{j\geq 0}$ behave in this limit. The proofs of these statements can be found in, e.g. the book of Szegő.

Proposition 16.1 (Plancherel-Rotach asymptotics)**.** *Denote by $H_\ell(x)$ the classical Hermite polynomial (15.55) and let ε and ω be fixed positive numbers. We have, as $\ell \to \infty$,*

1. *for $x = \sqrt{2\ell+1}\cos\phi$ with $\varepsilon \leq \phi \leq \pi - \varepsilon$,*

$$e^{-\frac{1}{2}x^2} H_\ell(x) = 2^{\frac{\ell}{2}+\frac{1}{4}} \sqrt{\ell!} (\pi\ell)^{-\frac{1}{4}} \frac{1}{\sqrt{\sin\phi}} \times \left\{ \sin\left(\left[\frac{\ell}{2}+\frac{1}{4}\right]\left(\sin(2\phi) - 2\phi\right) + \frac{3\pi}{4}\right) + O\left(\ell^{-1}\right)\right\}. \tag{16.7}$$

2. *for $x = \sqrt{2\ell+1}\cosh\phi$ with $\varepsilon \leq \phi \leq \omega$,*

$$e^{-\frac{1}{2}x^2} H_\ell(x) = 2^{\frac{\ell}{2}-\frac{3}{4}} \sqrt{\ell!} (\pi\ell)^{-\frac{1}{4}} \frac{1}{\sqrt{\sinh\phi}} \times \exp\left(\left[\frac{\ell}{2}+\frac{1}{4}\right]\left(2\phi - \sinh(2\phi)\right)\right)\left\{1 + O\left(\ell^{-1}\right)\right\}. \tag{16.8}$$

3. *for* $x = \sqrt{2\ell+1} - 2^{-\frac{1}{2}}3^{-\frac{1}{3}}\ell^{-\frac{1}{6}}t$ *with* $t \in \mathbb{C}$ *bounded,*

$$e^{-\frac{1}{2}x^2}H_\ell(x) = \pi^{\frac{1}{4}}2^{\frac{\ell}{2}+\frac{1}{4}}\sqrt{\ell!}\,\ell^{-\frac{1}{12}}\left\{A(t) + O\big(\ell^{-\frac{7}{3}}\big)\right\}, \tag{16.9}$$

with $A(t) = \mathrm{Ai}(-3^{-\frac{1}{3}}t)$ *in terms of the standard Airy function*

$$\mathrm{Ai}(x) \coloneqq \frac{1}{\pi}\int_0^\infty \cos\left(\tfrac{1}{3}t^3 + xt\right)dt, \qquad x \in \mathbb{R}. \tag{16.10}$$

One can use (16.7) and (16.8) to derive the leading terms of the large N-asymptotics of the functions

$$\begin{aligned} \psi_{N-k}^{(N)}(x) &= \sqrt{w_N(x)}\,p_{N-k}^{(N)}(x) \\ &= e^{-\frac{N}{2}x^2}N^{\frac{1}{4}}\gamma_{N-k}H_{N-k}\big(\sqrt{N}\,x\big) \end{aligned} \tag{16.11}$$

for integer $k \geq 0$.

Corollary 16.2. *Consider* $\psi_{N-k}^{(N)}$ *as in* (16.11) *for fixed* $k \in \mathbb{Z}_{\geq 0}$*. We have, as* $N \to \infty$*,*

1. *for any* $x \in (-\sqrt{2}, \sqrt{2})$*, parametrised as* $x = \sqrt{2}\cos\theta, \theta \in (0, \pi)$*,*

$$\psi_{N-k}^{(N)}(x) = \sqrt{\frac{2}{\pi\sqrt{2-x^2}}}\cos\left(N\beta(\theta) - \left(k - \frac{1}{2}\right)\theta - \frac{\pi}{4}\right) + O\big(N^{-1}\big), \tag{16.12}$$

where $\beta(\theta) \coloneqq \theta - \frac{1}{2}\sin(2\theta)$*.*

2. *for* $|x| > \sqrt{2}$ *parametrised as* $x = \sqrt{2}\cosh\theta, \theta > 0$*,*

$$\psi_{N-k}^{(N)}(x) = \sqrt{\frac{1}{2\pi\sqrt{x^2-2}}}\exp\left(-N\alpha(\theta) - \left(k - \frac{1}{2}\right)\theta\right)\left\{1 + O\big(N^{-1}\big)\right\}, \tag{16.13}$$

where $\alpha(\theta) \coloneqq \frac{1}{2}\sinh(2\theta) - \theta$*.*

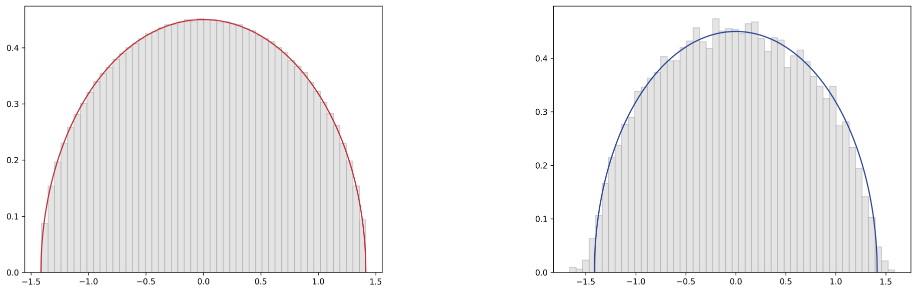
3. *for* $t \in \mathbb{C}$ *fixed and setting* $x = \sqrt{2} + t/(\sqrt{2}N^{2/3})$*,*

$$\psi_{N-k}^{(N)}(x) = 2^{\frac{1}{4}}N^{\frac{1}{6}}\left\{\mathrm{Ai}(t) + O\big(N^{-\frac{3}{4}}\big)\right\}. \tag{16.14}$$

Remark. *As* $x \to \infty$*, since* $\theta = \mathrm{arccosh}(x/\sqrt{2})$ *also goes to infinity, we can use* (16.13) *and find* $\lim_{x\to\infty}\psi_{N-k}^{(N)}(x) = 0$*.*

16.2. Bulk of the spectrum

We begin to use the formulæ of Theorem 15.15 by computing the 'limiting density of states', i.e. by obtaining the **semicircle law** visible in the figures 16.1 (and on the front of these notes).



(a) Density plot of the eigenvalues of one GUE matrix of size $N = 10,000$.

(b) Density plot of the eigenvalues of 100 independently drawn GOE matrices of size $N = 100$.

Figure 16.1.: Eigenvalue density plots.

This section is concerned with formalising the following idea. For a matrix drawn from the GUE (with the $1/\sqrt{N}$ scaling), consider $\overline{\mathcal{N}}_N(\Delta)$, the normalised count of its eigenvalues in Δ, so $\overline{\mathcal{N}}_N(\mathbb{R}) = 1$ certainly. Examining figure 16.1a implies that it is highly likely that $\overline{\mathcal{N}}_N([-\sqrt{2}, \sqrt{2}]) = 1$, and even more that if we looked at any $[a, b] \subset [-\sqrt{2}, \sqrt{2}]$, the proporition of eigenvalues in this range should be close to the area under the red curve between $[a, b]$. This curve is derived from the semi-circle of radius $\sqrt{2}$ (though since this is a density plot it is normalized): $\sigma(x) = (1/\pi)\sqrt{2 - x^2}$ for $-\sqrt{2} \leq x \leq \sqrt{2}$. Another perspective is that in terms of the gap probability $\mathbb{P}_N(\mathcal{N}_N(\Delta) = 0)$, one could infer that for $\Delta = \mathbb{R} \backslash [-\sqrt{2}, \sqrt{2}]$, the probability should approach 1.

This observation is proved in the next chapter in the case of a wider class of matrices (Wigner matrices), of which the GUE is specialisation. For this section and the next, we will prove related results concerning the *mean* normalised counting measure $\mathbb{E}[\overline{\mathcal{N}}_N(\Delta)]$ – note this is *not* a random quantity! Consult figure 16.1b. Here we are drawing 100 GOE matrices[1] of size 100 and plotting the density of the resulting 10,000 eigenvalues. This is capturing the idea of $\mathbb{E}[\overline{\mathcal{N}}_{100}(\Delta)]$ as we're getting an average picture by looking over 100 experiments. This average looks also close to the semi-circle plot (there seem to be more discrepencies compared to figure 16.1a, but this is attributable to the difference in matrix size; as we will see as matrix size increases, the agreement between the histogram and the semi-circle improves). It also typically looks close to the histogram from just *one*

[1] GOE has been chosen here only to demonstrate that this phenomenon might hold more widely. Indeed we will see that it does in the next chapter. You can freely replace GOE with GUE when reading this section.

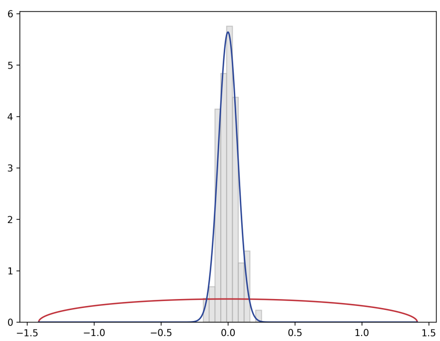
GUE matrix of the same size – i.e. for a given Δ and large enough N, $\overline{\mathcal{N}}_N(\Delta)$ (random) is not typically too far from $\mathbb{E}[\overline{\mathcal{N}}_N(\Delta)]$ (non-random). This is a statement in line with the law of large numbers (cf. 15.15 and A.12).

Compare this to example 15.6. There we had an $N \times N$ matrix $M_N = (\xi_{j,k})$ with the only (possibly) non-zero entries lie on the diagonal: $\xi_{j,k} = 0$ for $j \neq k$ and to be explicit we take $\sqrt{N} \cdot \xi_{j,j} \sim \mathcal{N}(0, 1/2)$. Therefore M_N is real and symmetric with a diagonal distributed exactly as the diagonal for the normalized GUE (but clearly the off-diagonal entries have a different distribution). The version of the density plots 16.1 for M_N are given in figure 16.2. There, we take in figure 16.2a a single 100×100 matrix with independent Gaussians on the diagonal (and zero elsewhere), and in figure 16.2b we take 100 such matrices. In this case,

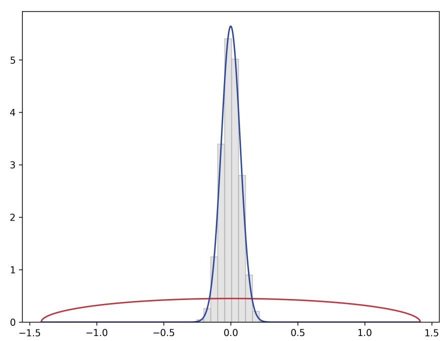
$$\overline{\mathcal{N}}_N(\Delta) = \frac{1}{N} \sum_{j=1}^{N} \chi_\Delta(\xi_{j,j}),$$

$$\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)] = \frac{1}{N} \sum_{j=1}^{N} \mathbb{P}(\xi_{j,j} \in \Delta) = \mathbb{P}(\mathcal{N}(0, 1/2N) \in \Delta) = \frac{1}{\sqrt{4\pi N}} \int_\Delta e^{-x^2/4N} dx$$

so patently $\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]$ does *not* converge to the semi-circle seen in the previous figure, instead since $\overline{\mathcal{N}}_N(\Delta)$ is the sum of independent random variables, as we saw in example 15.6, it will satisfy a central limit theorem (cf. the blue curve in figure 16.2). Therefore something *different* is going on in the case of the GUE due to the interactions of the off-diagonal entries.



(a) Density plot of the eigenvalues of one diagonal matrix $\mathrm{diag}([\mathcal{N}(0, 1/200), \ldots, \mathcal{N}(0, 1/200)])$ of size $N = 100$.



(b) Density plot of the eigenvalues of 100 independently drawn diagonal matrices $\mathrm{diag}([\mathcal{N}(0, 1/200), \ldots, \mathcal{N}(0, 1/200)])$ of size $N = 100$.

Figure 16.2.: Eigenvalue density plots for i.i.d. random variables.

Recall from Theorem 15.15 that, for the GUE, the mean normalized counting measure

$\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)]$ can be expressed in terms of orthogonal polynomials (cf. 15.44):

$$\mathbb{E}_N[\overline{\mathcal{N}}_N(\Delta)] = \int_\Delta \rho_N(x)dx = \frac{1}{N}\int_\Delta K_N(x,x)dx. \tag{16.15}$$

We can therefore interpret $\rho_N(x)$ as a density function for the mean normalized counting measure, giving the appropriate weight for the expected proportion of GUE_N eigenvalues lying in an infinitesimally small interval around x. Hence, given any $x \in \Delta \subset [-\sqrt{2}, \sqrt{2}]$, we may expect, following say figure 16.1b, that $\rho_N(x) \to \sigma(x) = (1/\pi)\sqrt{2-x^2}$ as N grows. This is part of the statement of Theorem 16.4. First we first prove two useful representations for ρ_N using orthogonal polynomials.

Proposition 16.3. *Let ρ_N be the density of the mean normalized counting measure for the* GUE*, see* (16.15)*. Then, for any $x \in \mathbb{R}$ and any integer $N \geq 2$,*

$$\rho_N(x) = \left(\psi_{N-1}^{(N)}(x)\right)^2 - \sqrt{\frac{N-1}{N}}\psi_N^{(N)}(x)\psi_{N-2}^{(N)}(x), \tag{16.16}$$

and

$$\rho_N(x) = \sqrt{2}\int_x^\infty \psi_N^{(N)}(y)\psi_{N-1}^{(N)}(y)dy \tag{16.17}$$

where $\psi_j^{(N)}(x) = \exp(-(N/2)x^2)\cdot N^{1/4}\cdot\gamma_j\cdot H_j(\sqrt{N}x)$ (cf. (16.5)*).*

Proof. To prove the first formula (16.16), we use the relation $\frac{d}{dx}H_\ell(x) = 2\ell H_{\ell-1}(x)$ for the Hermite polynomials (cf. (15.55)) to find

$$\begin{aligned}\frac{d}{dx}\psi_\ell^{(N)}(x) &= N^{1/4}\cdot\gamma_\ell\frac{d}{dx}\exp(-(N/2)x^2)\cdot H_\ell(\sqrt{N}x)\\ &= -Nx\psi_\ell^{(N)}(x) + \sqrt{2N\ell}\,\psi_{\ell-1}^{(N)}(x),\end{aligned} \tag{16.18}$$

and subsequently (16.16) follows from the third equality in (15.44), from (16.6) and (16.18). To obtain (16.17) we use the differential equation

$$\frac{d^2}{dx^2}H_\ell(x) - 2x\frac{d}{dx}H_\ell(x) + 2\ell H_\ell(x) = 0$$

for the Hermite polynomials (which can also be deduced from (15.55)) and hence

$$\frac{d^2}{dx^2}\psi_\ell^{(N)}(x) = N\left(Nx^2 - (2\ell+1)\right)\psi_\ell^{(N)}(x). \tag{16.19}$$

Then, using the third equality in (15.44) combined with (16.19), we find

$$\frac{d}{dx}\rho_N(x) = -\sqrt{2}\psi_N^{(N)}(x)\psi_{N-1}^{(N)}(x)$$

for all $x \in \mathbb{R}$.

By (16.13), $\lim_{x\to\infty}\psi_\ell^{(N)}(x) = 0$, see (16.13) and the remark thereafter. Hence integrating the final equation above from ∞ to x we find identity (16.17). This completes the proof. □

Theorem 16.4 (Wigner's semicircle law (GUE))**.** *Let ρ_N be the density of the mean normalized counting measure for the* GUE*. Then its pointwise limit indeed converges to the semi-circle:*

$$\lim_{N\to\infty} \rho_N(x) = \frac{1}{\pi}\sqrt{(2-x^2)_+}, \qquad x_+ := \max\{x,0\}, \quad x \in \mathbb{R}. \tag{16.20}$$

This convergence is uniform on any closed interval of $\mathbb{R} \setminus \{\pm\sqrt{2}\}$.

Proof: (Non-examinable). We use formula (16.16): from (16.13), once $|x| \geq \sqrt{2}+\varepsilon$ with $\varepsilon > 0$, the right hand side in the same identity (16.13) vanishes exponentially as $N \to \infty$, matching onto the right hand side in (16.20). For $|x| \leq \sqrt{2}-\varepsilon$, we obtain from (16.12) for $N \to \infty$,

$$\big(\psi_{N-1,N}(x)\big)^2 = \frac{2}{\pi\sqrt{2-x^2}}\cos^2\left(N\beta(\theta) - \frac{\theta}{2} - \frac{\pi}{4}\right) + O\big(N^{-1}\big),$$

(recall $x = \sqrt{2}\cos\theta$ in this range) and likewise

$$\psi_{N,N}(x)\psi_{N-2,N}(x) = \frac{2}{\pi\sqrt{2-x^2}}\cos\left(N\beta(\theta) + \frac{\theta}{2} - \frac{\pi}{4}\right)\cos\left(N\beta(\theta) - \frac{3\theta}{2} - \frac{\pi}{4}\right) + O\big(N^{-1}\big).$$

Hence, using trigonometric identities,

$$\begin{aligned}
&\big(\psi_{N-1,N}(x)\big)^2 - \sqrt{\frac{N-1}{N}}\psi_{N,N}(x)\psi_{N-2,N}(x)\\
&\qquad = \frac{1}{\pi\sqrt{2-x^2}}\left\{1 + \cos\left(2N\beta(\theta) - \theta - \frac{\pi}{2}\right)\right\}\\
&\qquad\qquad\qquad - \frac{1}{\pi\sqrt{2-x^2}}\left\{\cos\left(2N\beta(\theta) - \theta - \frac{\pi}{2}\right) + \cos(2\theta)\right\} + O\big(N^{-1}\big)\\
&\qquad = \frac{1-\cos(2\theta)}{\pi\sqrt{2-x^2}} + O\big(N^{-1}\big),
\end{aligned}$$

and since $1-\cos(2\theta) = \sin^2\theta = 1 - \frac{1}{2}x^2$, the pointwise limit (16.20) follows. □

The next result describes the *bulk case* of the local regime for the GUE. This scenario concerns the statistical properties of its eigenvalues in an $O(N^{-1})$-neighborhood of $x_0 \in (-\sqrt{2},\sqrt{2})$: a point of the support of the limiting normalized counting measure of eigenvalues, where its density (16.20) does not vanish. Generally, we have the following definition.

Definition 16.5. *Suppose the normalized counting measure $\overline{N}_N$ of eigenvalues of a given random matrix M_N converges (in probability) to some limit law N that has density ρ. Then the* ***bulk of the spectrum*** *of N is defined as*

$$\mathrm{bulk}(N) := \Big\{x \in \mathrm{supp}(N):\ \exists\,\delta > 0 \ \text{ such that } \inf_{y\in[x-\delta,x+\delta]}\rho(y) > 0\Big)$$

Points of the spectrum that do not belong to its bulk are called ***special points****.*

Theorem 16.6. *Consider the* GUE *whose joint eigenvalue density is given by* (15.27) *with* $\beta = 2$ *and* $V(x) = x^2$*. Assume that* x_0 *belongs to the interior of the support* $[-\sqrt{2}, \sqrt{2}]$ *(bulk of the spectrum) of the semicircle law with density* (16.20)*. We have the following:*

1. *If* $P_{N,\ell} \equiv P^{(2)}_{N,\ell}$ *is the* ℓ*th marginal of the joint eigenvalue density (cf.* (15.5) *and exercise 15.14) and* $\rho_N(x)$ *remains the density of the mean normalized counting measure, then for any* $\ell \in \mathbb{N}$*,*

$$\lim_{n\to\infty} \frac{1}{(\rho_N(x_0))^\ell} P_{N,\ell}\left(x_0 + \frac{x_1}{N\rho_N(x_0)}, \ldots, x_0 + \frac{x_N}{N\rho_N(x_0)}\right) = \det\left(\frac{\sin\pi(x_j - x_k)}{\pi(x_j - x_k)}\right)_{j,k=1}^{\ell}. \tag{16.21}$$

   The kernel

$$K_N(x,y) = \frac{\sin\pi(x-y)}{\pi(x-y)}$$

   is known as the ***Sine kernel****.*

2. *If we consider* $\mathbb{P}_N(\mathcal{N}_N(\Delta_N) = 0)$ *the probability that a matrix from the* GUE_N *has no eigenvalues in*

$$\Delta_N := \left(x_0, x_0 + \frac{s}{N\rho_N(x_0)}\right) \subset \mathbb{R},$$

   then in the large matrix limit this has determinantal form

$$\mathcal{E}_0(s) := \lim_{N\to\infty} \mathbb{P}_N(\mathcal{N}_N(\Delta_N) = 0) = \det(I - Q_s), \tag{16.22}$$

   uniformly in s *on any compact set in* $\mathbb{R}_{>0}$*, where* Q_s *is the integral operator*

$$(Q_s f)(x) := \int_0^s \frac{\sin\pi(x-y)}{\pi(x-y)} f(y)dy, \tag{16.23}$$

   for any $x \in (0,s)$*, acting on square-integrable functions* $f : (0,s) \to \mathbb{C}$*.*

3. *If we consider* $\mathbb{P}_N(\mathcal{N}_N(\Delta_n) = \ell)$*, the probability for the* GUE *to have* $\ell \in \mathbb{Z}_{\geq 0}$ *eigenvalues in* Δ_N*, then as* $s \to 0$ *from the right*[2]

$$\mathcal{E}_\ell(s) := \lim_{N\to\infty} \mathbb{P}_N(\mathcal{N}_N(\Delta_n) = \ell) = C_\ell s^{\ell^2}(1 + o(1)), \tag{16.24}$$

   where

$$C_0 = 1, \quad C_\ell = \frac{1}{(2\pi)^\ell \ell!} \frac{1}{\left(\prod_{j=1}^{\ell-1} j!\right)^2} \int_{[-\pi,\pi]^\ell} \Delta^2(x_1,\ldots,x_\ell) \prod_{j=1}^{\ell} dx_j, \quad \ell \in \mathbb{N}; \qquad A_\ell = \prod_{j=1}^{\ell-1} j!.$$

[2]The notation $(1 + o(1))$ here just means that the expression to the left of the bracket is the *dominating contribution* in this limit. For example, if $f(x) = 1/x + x$ then as $x \to 0$ from the right, we could write $f(x) = (1/x)(1 + o(1))$.

Proof: (Highlights, if you are interested in the details, consult the lecturer or the quoted texts.) By the determinantal expression for the marginal (cf. exercise 15.14) we have

$$\frac{1}{(\rho_N(x_0))^\ell} P_{N,\ell}\left(x_0 + \frac{x_1}{N\rho_N(x_0)}, \ldots, x_0 + \frac{x_N}{N\rho_N(x_0)}\right)$$
$$= \prod_{j=1}^{\ell-1}\left(1 - \frac{j}{N}\right)^{-1} \det\left(\frac{1}{N\rho_N(x_0)} K_N\left(x_0 + \frac{x_j}{N\rho_N(x_0)}, x_0 + \frac{x_k}{N\rho_N(x_0)}\right)\right)_{j,k=1}^{\ell},$$

so it suffices to prove that for any x_0 such that $\rho_N(x_0) > 0$, i.e. for $|x_0| < \sqrt{2}$ when N is sufficiently large, we have uniformly in x_j and x_k, varying in a compact set of $\mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{N\rho_N(x_0)} K_N\left(x_0 + \frac{x_j}{N\rho_N(x_0)}, x_0 + \frac{x_k}{N\rho_N(x_0)}\right) = \frac{\sin\pi(x_j - x_k)}{\pi(x_j - x_k)}, \tag{16.25}$$

where K_N and $\psi_N(x)$ are given by (15.30) and (16.5). One can then manipulate the expression for K_N using the Plancherel-Rotach asymptotics (Proposition 16.1). After some (involved) trigonometric manipulations, one indeed arrives at (16.21).

To find the expressions for $\mathbb{P}_N(\mathcal{N}_N(\Delta_N) = \ell)$, $\ell = 0, \ldots, N$, one uses the expressions (15.48) and (15.38), and justifies taking the limit $N \to \infty$ termwise in (15.38), using (16.25). This can be done via Hadamard's inequality (cf. Theorem A.7). From there we see (16.22). The small-s asymptotics (16.24) require more careful handling, cf. Pastur-Shcherbina for details. □

Remark.

1. *If eigenvalues in the* GUE *were* independent *random variables, then the probability of finding ℓ eigenvalues in the interval $[0, s]$ is proportional to s^ℓ. Thus, formula (16.24) is a manifestation of **eigenvalue repulsion** of large random matrices in the local regime.*

2. *Throughout the theorem, we are concerned with spacings normalised by $N\rho_N(x)$. Recall that*

$$\begin{aligned}\frac{1}{N}\mathbb{E}_N[\mathcal{N}_N([-\sqrt{2}, \lambda])] &= \int_{-\sqrt{2}}^{\lambda} \rho_N(x)dx \\ &\to \int_{-\sqrt{2}}^{\lambda} \frac{1}{\pi}\sqrt{(2 - x^2)_+}dx \\ &= \frac{\lambda}{2\pi}\sqrt{2 - \lambda^2} + \frac{1}{\pi}\arcsin\left(\frac{\lambda}{\sqrt{2}}\right) + \frac{1}{2}.\end{aligned}$$

Therefore the expected number of eigenvalues between $-\sqrt{2}$ and λ is

$$\mathbb{E}_N[\mathcal{N}_N([-\sqrt{2}, \lambda])] = \begin{cases} 0 & \lambda \le -\sqrt{2} \\ \frac{N\lambda}{2\pi}\sqrt{2 - \lambda^2} - \frac{N}{\pi}\arcsin\left(\frac{\lambda}{\sqrt{2}}\right) + \frac{N}{2} & \lambda \in (-\sqrt{2}, \sqrt{2}) \\ N & \lambda \ge \sqrt{2}. \end{cases} \tag{16.26}$$

Therefore the expected position of the jth eigenvalue λ_j (ordered) is the right hand side of (16.26) at $\lambda = \lambda_j$. If we rescale and call $\overline{\lambda}_j = \mathbb{E}_N[\mathcal{N}_N([-\sqrt{2}, \lambda_j])]$, then (check!) the expected spacing between $\overline{\lambda}_j$ and $\overline{\lambda}_{j+1}$ is 1. (Cf. also the motivation in Section 9, 10 in the first part of this course.) Figure 16.3 captures this idea. There we have drawn 100 *times from the* GUE *of size* 1000*. For each matrix, we found the eigenvalues, rescaled $\lambda_j \mapsto \overline{\lambda}_j$ according to (16.26), and then recorded the* 999 *consecutive spacings $\overline{\lambda}_j - \overline{\lambda}_{j-1}$. The histogram pictured is the frequency of the* 99, 900 *experimental spacings. Indeed, we see that the average spacing is approximately* 1*. The curve plotted over the top is the Wigner surmise, which you will have seen in the first part of the course.*

3. *Let's examine point (2). From the previous point, Δ_N is structured so that we expect to find one eigenvalue in this interval (rather than rescaling the eigenvalues so that in an interval of length* 1 *we expect to see one eigenvalue, we could instead 'zoom in' to focus on an interval where we expect to see only one unscaled eigenvalue). The gap result then says that the limiting probability to find* no eigenvalues *in Δ_N is given by the Fredholm determinant* $\det(I - Q_s)$*. However, it might not be clear (cf. (15.38)) how to leverage this fact; i.e. given an s actually compute this probability. It transpires (beyond the scope of this course) that there is a connection between the Fredholm determinant and a particular non-linear differential equation, yielding*
$$\det(I - Q_s) = \exp\left(\int_0^{\pi s} \frac{\nu(x)}{x} dx\right)$$
where $\nu(x) \sim -(x/\pi) - x^2/(\pi^2)$ near zero solves the differential equation referenced. Thus, the probability that in the bulk an interval of length $s = 0.1$ contains no rescaled eigenvalues is
$$\mathcal{E}_0(0.1) \approx \exp\left(-\int_0^{0.1\pi} \frac{1}{x}\left(\frac{x}{\pi} + \frac{x^2}{\pi^2}\right) dx\right) \approx 0.90003.$$
(Or alternatively an interval $(x, x + 0.1/N\rho_N(x))$ contains no eigenvalues is also approximately $\mathcal{E}_0(0.1)$ for large N.)

16.3. Edges of the spectrum

In the previous section we presented the basic properties of GUE bulk eigenvalues, i.e. for $|x| < \sqrt{2}$, where the limiting density of states (16.20) is strictly positive. Theorem 16.4 treats the global regime, and Theorem 16.6 the local one. We now consider analogous results concerning the edges of the support, i.e. the points $x = \pm\sqrt{2}$. Firstly, we have that as the matrix size grows, the largest (resp. smallest) eigenvalue converges with probability equal to one[3] to $\sqrt{2}$ (resp. $-\sqrt{2}$).

[3] This is a probabilistic statement since there might be matrices having the largest eigenvalue $> \sqrt{2}$, but these are very unlikely (and become less and less likely as $N \to \infty$).

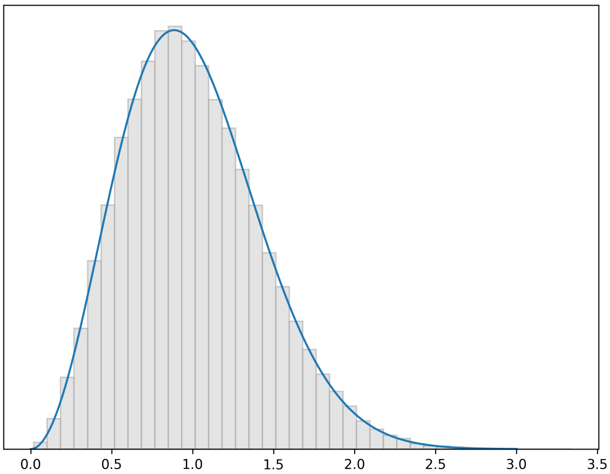


Figure 16.3.: Histogram of a collection of 100 consecutive renormalised eigenvalue spacings of 1000×1000 GUE matrices. The blue curve is the Wigner surmise.

Theorem 16.7. *Consider* $\mathrm{M_N}$ *drawn from the* GUE *(rescaled) and denote by* $\lambda_N = \lambda_N(\mathrm{M_N})$*, resp.* $\lambda_1 = \lambda_1(\mathrm{M_N})$*, its maximum, resp. minimum, eigenvalue. Then with probability one,*

$$\lim_{N\to\infty} \lambda_N = \sqrt{2}, \qquad \lim_{N\to\infty} \lambda_1 = -\sqrt{2}. \tag{16.27}$$

Proof. We focus on the first equality as the proof of the second one is analogous. We begin with one half of the statement, namely that

$$\mathbb{P}_N(\limsup_{N\to\infty} \lambda_N \le \sqrt{2}) = 1. \tag{16.28}$$

Indeed, if we manage to show that for any $\varepsilon > 0$,

$$\sum_{N=1}^{\infty} \mathbb{P}_N\big(\lambda_N \ge \sqrt{2}(1+\varepsilon)\big) < \infty, \tag{16.29}$$

then inequality (16.28) follows from the Borel-Cantelli lemma, see Theorem A.11. So, to prove that the sum in (16.29) is finite we use Markov's inequality (A.5):

$$\begin{aligned}
\mathbb{P}_N\big(\lambda_N \ge \sqrt{2}(1+\varepsilon)\big) &= \mathbb{P}_N\left(\mathcal{N}_N\left([\sqrt{2}(1+\varepsilon), \infty)\right) \ge 1\right) \\
&\le N\mathbb{E}_N\left[\overline{\mathcal{N}}_N\left([\sqrt{2}(1+\varepsilon), \infty)\right)\right] && \text{(Markov's ineq.)} \\
&= N \int_{\sqrt{2}(1+\varepsilon)}^{\infty} \rho_N(x)dx
\end{aligned}$$

with the final equality using the density for the mean normalized counting measure (15.45). To conclude we need a bound on the density $\rho_N(x)$. One can show (via (16.16), the Plancherel-Rotach asymptotics 16.13, the Cauchy-Schwarz inequality, and Laplace's method of integral approximation) that the bound

$$\rho_N\left(\sqrt{2}+\frac{\sqrt{2}s}{N^{\frac{2}{3}}}\right) \le c_1 N^{-\frac{1}{3}}\frac{1}{s}e^{-c_2 s^{\frac{3}{2}}}, \tag{16.30}$$

is valid for some constants $c_1, c_2 > 0$ and as $s, N \to \infty$.

Hence

$$\begin{aligned}
\mathbb{P}_N\left(\lambda_N \ge \sqrt{2}(1+\varepsilon)\right) &\le N\int_{\sqrt{2}(1+\varepsilon)}^{\infty}\rho_N(x)dx \\
&= \frac{N^{\frac{1}{3}}}{\sqrt{2}}\int_{\varepsilon N^{2/3}}^{\infty}\rho_N\left(\sqrt{2}+\frac{y}{\sqrt{2}N^{\frac{2}{3}}}\right)dy \\
&\le c_1\int_{\varepsilon N^{2/3}}^{\infty}\frac{1}{y}e^{-c_2 y^{\frac{3}{2}}}dy \\
&\le c_1\varepsilon^{-\frac{3}{2}}\frac{1}{N}e^{-Nc_2\varepsilon^{\frac{3}{2}}}
\end{aligned}$$

for N sufficiently large and where $c_1, c_2 > 0$ are some constants independent of ε and N. For the last inequality, try integration by parts.

The last estimate guarantees convergence in (16.29) since it is rapidly summable. The other half of the statement, namely

$$\mathbb{P}\left(\liminf_{N\to\infty}\lambda_N \ge \sqrt{2}\right) = 1$$

is harder and we refer the interested reader to Pastur-Shcherbina (cf. Corollary 2.2.8 in that text for details of its proof). □

For the remainder of this section, for the sake of definiteness, we confine ourselves to the point $x = \sqrt{2}$. Then, the limits (16.27) imply the relation

$$\lim_{N\to\infty}\mathbb{P}_N\left(\lambda_N \le x\right) = \lim_{N\to\infty}\mathbb{P}_N\left(\mathcal{N}((x,\infty)) = 0\right) = \begin{cases} 0, & x \le \sqrt{2} \\ 1, & x > \sqrt{2} \end{cases}$$

for the *cumulative distribution function* of the maximal eigenvalue in the GUE. This is a global regime result concerning the edge of the support of the limiting normalized counting measure of eigenvalues. We now discuss the corresponding local regime results, zooming in around the edge.

Theorem 16.8. *Take* $\mathrm{M_N}$ *from the* GUE *(rescaled). Then we have the following limits:*

1. *For the density* ρ_N *of the mean normalized counting measure* (16.15)*:*

$$\lim_{N\to\infty}\frac{1}{\sqrt{2}}N^{\frac{1}{3}}\rho_N\left(\sqrt{2}+\frac{s}{\sqrt{2}N^{\frac{2}{3}}}\right) = \int_s^{\infty}\mathrm{Ai}^2(y)dy = -s\mathrm{Ai}^2(s) + \left(\mathrm{Ai}'(s)\right)^2, \tag{16.31}$$

where $\mathrm{Ai}(x)$ *is the standard Airy function, cf.* (16.10)*.*

2. *If* $P_{N,\ell} \equiv P^{(2)}_{N,\ell}$ *is the* ℓ*th marginal of* (15.27) *(with* $\beta = 2$ *and* $V(x) = x^2$*), then for any* $\ell \in \mathbb{N}$,

$$\lim_{N\to\infty} \left(\frac{1}{\sqrt{2}} N^{\frac{1}{3}}\right)^{\ell} P_{N,\ell}\left(\sqrt{2} + \frac{x_1}{\sqrt{2}N^{\frac{2}{3}}}, \ldots, \sqrt{2} + \frac{x_\ell}{\sqrt{2}N^{\frac{2}{3}}}\right) = \det\left(K_{\mathrm{Ai}}(x_j, x_k)\right)_{j,k=1}^{\ell}, \tag{16.32}$$

where

$$K_{\mathrm{Ai}}(x, y) := \int_0^\infty \mathrm{Ai}(x+u)\mathrm{Ai}(u+y)du = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x-y},$$

is known as the ***Airy kernel****.*

3. *For the gap probability* $\mathbb{P}_N(\mathcal{N}_N(\Delta_N) = 0)$ *corresponding to* $\Delta_N = \sqrt{2} + J/(\sqrt{2}N^{2/3})$*, where* $J \subset (a, b)$ *with* $-\infty < a < b \le \infty$*,*

$$\lim_{N\to\infty} \mathbb{P}_N(\mathcal{N}_N(\Delta_N) = 0) = \det(I - K_J). \tag{16.33}$$

Here, K_J *is the integral operator defined by the Airy kernel* K_{Ai}*, acting on square-integrable functions* $f : J \to \mathbb{C}$*.*

Proof. Take as a starting point the integral (16.17) at $x = \sqrt{2} + s/(\sqrt{2}N^{2/3})$:

$$\begin{aligned}\rho_N(x) &= \sqrt{2}\int_x^\infty \psi_N^{(N)}(y)\psi_{N-1}^{(N)}(y)dy \\ &= \sqrt{2}\int_{\sqrt{2}+s/(\sqrt{2}N^{2/3})}^{A} \psi_N^{(N)}(y)\psi_{N-1}^{(N)}(y)dy + \sqrt{2}\int_A^\infty \psi_N^{(N)}(y)\psi_{N-1}^{(N)}(y)dy,\end{aligned}$$

where we have split the integral at some arbitrary point A. This will help with analysis.. In the first integral (with the prefactor from (16.31)) we use the asymptotic formula (16.14) from the Plancherel-Rotach asymptotics (and write simply $\psi_j^{(N)} = \psi_j$),

$$\begin{aligned}N^{\frac{1}{3}}\int_{\sqrt{2}+s/(\sqrt{2}N^{2/3})}^{A} \psi_N(y)\psi_{N-1}(y)dy &= \frac{1}{\sqrt{2}N^{\frac{1}{3}}}\int_s^A \psi_N\left(\sqrt{2} + \frac{t}{\sqrt{2}N^{\frac{2}{3}}}\right)\psi_{N-1}\left(\sqrt{2} + \frac{t}{\sqrt{2}N^{\frac{2}{3}}}\right)dt \\ &= \int_s^A \mathrm{Ai}^2(t)dt + o(1).\end{aligned} \tag{16.34}$$

as $N \to \infty$, using (16.14) for the last expression.

For the second integral

$$\sqrt{2}\int_A^\infty \psi_N(y)\psi_{N-1}(y)dy = \rho_N(A),$$

we can show this vanishes as $A \to \infty$. Indeed, it is possible (similarly to how one establishes (16.30) in the global regime) to show the following bound: for any $\delta > 0$ and $|x| > \sqrt{2}(1+\delta)$,

$$\rho_N(x) \le c_1 e^{-c_2 N x^2}$$

where $c_1, c_2 > 0$ are two constants possibly dependent on δ. Therefore, if we take $A \to \infty$, the second integral vanishes.

Hence, carrying out the successive limits $N \to \infty$ and then $A \to \infty$, we obtain the first equality in (16.31).

To obtain the second equality in (16.31), we use that the Airy function satisfies the differential equation $y'' - xy = 0$ for $y = \mathrm{Ai}(x)$. From this we can find the relation

$$\int_0^\infty \mathrm{Ai}(x+u)\mathrm{Ai}(u+y)du = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x-y}.$$

The limit of this relation as $x \to y$ and the same differential equation again yield

$$\int_0^\infty \mathrm{Ai}(x+u)\mathrm{Ai}(u+x)du = \left(\mathrm{Ai}'(x)\right)^2 - \mathrm{Ai}''(x)\mathrm{Ai}(x) = \left(\mathrm{Ai}'(x)\right)^2 - x\mathrm{Ai}^2(x).$$

This proves the second equality in (16.31).

Next, in view of the determinantal formula for the ℓth marginal in e.g. Exercies 15.14, the proof of (16.32) is reduced to the proof of the limit

$$\lim_{N\to\infty} \frac{1}{\sqrt{2}N^{\frac{2}{3}}} K_N\left(\sqrt{2} + \frac{x_j}{\sqrt{2}N^{\frac{2}{3}}}, \sqrt{2} + \frac{x_k}{\sqrt{2}N^{\frac{2}{3}}}\right) = K_{\mathrm{Ai}}(x_j, x_k). \tag{16.35}$$

This is shown using the Christoffel-Darboux identity (cf. (15.37)) and Plancherel-Rotach asymptotics.

To prove (16.33), we first change variables $y_k = \sqrt{2}N^{2/3}(x_k - \sqrt{2})$ in the ℓth term in the Fredholm determinant sum (15.41), (15.49) and obtain

$$\mathbb{P}_N(\mathcal{N}_N(\Delta_N) = 0) = 1 + \sum_{\ell=1}^{N} \frac{(-1)^\ell}{\ell!} \int_{J^\ell} \det\left(\frac{1}{\sqrt{2}N^{\frac{2}{3}}} K_N\left(\sqrt{2} + \frac{y_j}{\sqrt{2}N^{\frac{2}{3}}}, \sqrt{2} + \frac{y_k}{\sqrt{2}N^{\frac{2}{3}}}\right)\right]_{j,k=1}^{\ell} \prod_{m=1}^{\ell} dy_m. \tag{16.36}$$

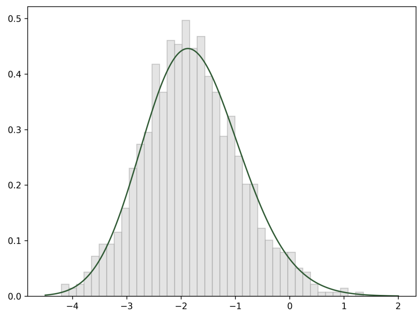
Consulting (16.35) we see that the integrand will converge as $N \to \infty$, but we need to justify that the infinite sum still converges so that the Fredholm determinant is well-defined. It is indeed possible to do so (with the help of appealing to the determinant bound from the appendix (A.2)). $\square$

Finally for this chapter, let λ_N^* be the maximum eigenvalue of a GUE_N matrix $\mathrm{M_N} = (1/\sqrt{N})M_N$ (so this is a random variable). Setting the rescaling

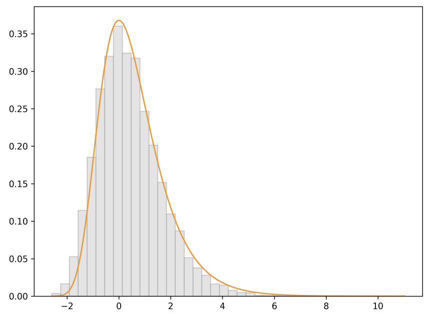
$$\lambda_N^* = \sqrt{2} + \frac{\Lambda_N^*}{\sqrt{2}N^{2/3}},$$

we find from (16.33) that the limiting cumulative distribution function for this rescaled maximum eigenvalue is

$$F(s) := \lim_{N\to\infty} \mathbb{P}_N\left(\Lambda_N^* \le s\right) = \lim_{N\to\infty} \mathbb{P}_N\left(\mathcal{N}_N\left(\left(\sqrt{2} + \frac{s}{\sqrt{2}N^{2/3}}, \infty\right)\right)\right) = \det\left(I - K_{(s,\infty)}\right). \tag{16.37}$$



(a) A histogram of the largest eigenvalue (rescaled) of 1000 different GUE matrices of size 1000. The Tracy-Widom ($\beta = 2$) distribution is plotted in green.



(b) A histogram of the largest eigenvalue (rescaled) of 1000 different diagonal matrices of size 1000 whose diagonal is $(\mathcal{N}(0, 1/2000), \ldots, \mathcal{N}(0, 1/2000))$. The standard Gumbel distribution is plotted in orange.

Figure 16.4.: Comparison of the distribution of the maximum (rescaled) eigenvalue in two random matrix models.

This distribution (currently with a non-explicit density) is known as the ($\beta = 2$) **Tracy-Widom distribution**. Its analogues for the GOE, and other ensembles, occur in a number of interesting problems of analysis, probability, asymptotic combinatorics, and statistical physics. In particular: the asymptotic of generalized *Toeplitz determiants*, the length of the longest increasing subsequence of random permutations, the asymptotic shape of Young tableaux, the last passage of time in certain percolation models, all those relate to the Tracy-Widom distribution. Figure 16.4a shows a histogram of the largest eigenvalue after rescaling of 1000 different GUE matrices $\mathrm{M_N}$ of size 1000, with the corresponding Tracy-Widom distribution plotted.

Remark. *Let's again compare to the case of independent random variables. Let* $(\lambda_j)_{j=1}^N$ *be a collection of i.i.d. real-valued random variables where* $\lambda_j \sim \mathcal{N}(0, 1/2N)$*. Define the* diagonal $N \times N$ *random matrix* $M_N = (\xi_{j,k})_{j,k=1}^N$ *where* $\xi_{j,j} = \lambda_j$ *and* $\xi_{j,k} = 0$ *for* $j \neq k$*. Therefore* M_N *is real and symmetric with a diagonal distributed exactly as the diagonal for the normalized* GUE *(but clearly the off-diagonal entries have a different distribution). What is the distribution of the maximum eigenvalue* λ_N^* *of* M_N*? Equivalently: what is the distribution of the maximum of a collection of* N *i.i.d Gaussian random variables? It transpires that, after appropriate rescaling, this will follow a* Gumbel *distribution with density function* $\exp(-(x + \exp(-x)))$*. See figure 16.4b.*

17. Real Wigner matrices

In this chapter we readjust our focus: instead of working with a random matrix model whose eigenvalue joint distribution can be written down explicitly for finite N and manipulated (such as in the case of the GUE), we shall consider the broader class of real matrices called Wigner matrices, and establish Wigner's semicircle law (as in (15.54)) for them. The techniques we will use will differ significantly from the orthogonal polynomial techniques we used in Sections 16.2 and 16.3 for the GUE (which required the form of the density (15.27))– instead we will follow Wigner's original idea and follow a beautiful combinatorial approach.

17.1. Real Wigner matrices: traces, moments and combinatorics

The concept of a real Wigner matrix is a generalization of the construction for the GOE or GUE (cf. (15.19), (15.22)) in that we no longer impose a concrete distribution on our matrix entries.

Definition 17.1 (Real Wigner matrices)**.** *Start with two independent families of i.i.d. real-valued random variables $(Z_{j,k})_{1\le j<k\le N}$ and i.i.d. real-valued random variables $(Y_j)_{j=1}^N$ such that*

$$\begin{aligned} \mathbb{E}[Z_{1,2}] &= \mathbb{E}[Y_1] = 0, \\ \mathbb{E}\left[Z_{1,2}^2\right] &= t > 0, \end{aligned}$$

and for all $k \in \mathbb{N}$,

$$r_k \coloneqq \max\left\{\mathbb{E}\left[|Z_{1,2}|^k\right], \mathbb{E}\left[|Y_1|^k\right]\right\} < \infty. \tag{17.1}$$

Then the ***real Wigner ensemble*** *is the set of all real symmetric matrices $\mathrm{M_N} = (\xi_{j,k})_{j,k=1}^N$ of size $N \times N$ with entries*

$$\xi_{j,k} = \frac{1}{\sqrt{N}}\begin{cases} Z_{j,k}, & \text{if } j < k \\ Z_{k,j}, & \text{if } j > k \\ Y_j, & \text{if } j = k \end{cases}.$$

As before, compare (15.2), we let $\{\lambda_j(\mathrm{M_N})\}_{j=1}^N \subset \mathbb{R}$ denote the eigenvalues of a real Wigner matrix $\mathrm{M_N}$, indexed in nondecreasing order

$$-\infty < \lambda_1(\mathrm{M_N}) \le \ldots \le \lambda_N(\mathrm{M_N}) < \infty.$$

The original result for real Wigner matrices, written down by Wigner in 1955, asserts convergence of the normalized counting measure $\overline{\mathcal{N}}_N$ (cf. (15.1)) of the same eigenvalues $\{\lambda_j(\mathrm{M_N})\}_{j=1}^N$ to the **semicircle distribution** (or **law**), that is to the probability distribution on $\mathbb{R}$ with density

$$\sigma_t(x) := \frac{1}{2\pi t}\sqrt{(4t-x^2)_+}\,, \qquad x_+ := \max\{x,0\}, \quad x \in \mathbb{R} \tag{17.2}$$

Recall t/N is the off-diagonal entry variance. In the GUE case (rescaled), $t = 1/2$.

More precisely, we have the following famous result.

Theorem 17.2 (Wigner's semicircle law)**.** *Let $\Delta \subset \mathbb{R}$ be an interval and $\mathrm{M_N}$ a real Wigner matrix of size $N \times N$, with eigenvalues $-\infty < \lambda_1 \leq \ldots \leq \lambda_N < \infty$. Then the normalized counting measure (15.1) satisfies, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\overline{\mathcal{N}}_N(\Delta) - \int_\Delta \sigma_t(x)dx\right| > \varepsilon\right) = 0. \tag{17.3}$$

In greater detail, for any continuous bounded function $\varphi : \mathbb{R} \to \mathbb{C}$, and each $\varepsilon > 0$, the linear statistic (15.3) obeys

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\frac{1}{N}\mathcal{N}_N[\varphi] - \int_\mathbb{R} \varphi(x)\sigma_t(x)dx\right| > \varepsilon\right) = 0. \tag{17.4}$$

Remark. *The assumption (17.1) regarding the finiteness of all moments is not really needed, see Anderson-Guionnet-Zeitouni, Section* 2.1.5*. The conclusions of Theorem 17.2 still hold when only $r_2 < \infty$ is assumed.*

Remark. *The statements (17.3) and (17.4) should be compared to the law of large numbers type result for the* GUE, *(15.54). They are not as strong as the pointwise limit (16.20). Still, the importance and wide applicability of Theorem 17.2 stems from the fact that (17.3), (17.3) are **universality results**. This means that (17.3), (17.4) are* universally *true for* any *real Wigner matrix $\mathrm{M_N}$ of the type described in Definition 17.1.*

An important special case of (17.4), which we will use for the eventual proof of Theorem 17.2, stems from choosing

$$\varphi(x) = x^k$$

in (17.4), for real x and $k \in \mathbb{N}$. Although such φ are unbounded we still have the following result:

Theorem 17.3 (Wigner's semicircle law for matrix moments)**.** *Let $\mathrm{M_N}$ be a real Wigner matrix of dimension N. Then for any fixed $k \in \mathbb{N}$ and $\varepsilon > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}\left(\mathrm{M_N}^k\right) - \int_\mathbb{R} x^k\sigma_t(x)dx\right| > \varepsilon\right) = 0. \tag{17.5}$$

Going back from (17.5) to (17.4) is, in principal, a matter of approximating any continuous bounded functions by polynomials on the supports of $\overline{\mathcal{N}}_N$ and σ_t. However, even though $\overline{\mathcal{N}}_N$ has finite support, there is no obvious bound on $\bigcup_N \operatorname{supp}(\overline{\mathcal{N}}_N)$, and this makes such an approximation scheme tricky. In fact, when we eventually prove Theorem 17.2 in Section 17.3, we will take a different approach entirely. For now, however, it will be useful to have some intuition for (17.4), so we begin by presenting a scheme for proving Theorem 17.3. The place to begin is calculating the *moments* of the semicircle law.

Lemma 17.4. *For $k \in \{0, 1, 2, 3, \dots\}$, define the* ***moments*** *of the semicircle law as*

$$m_k := \int_{\mathbb{R}} x^k \sigma_1(x) dx. \tag{17.6}$$

Then the following is true:

1. *For any $t > 0$,*
$$\int_{\mathbb{R}} x^k \sigma_t(x) dx = t^{k/2} m_k.$$

2. *We have*
$$\begin{aligned} m_0 &= 1; \\ m_{2k-1} &= 0, \\ m_{2k} &= \frac{2(2k-1)}{k+1} m_{2k-2}, \end{aligned}$$
the latter holding for $k \in \mathbb{N}$. Therefore for any non-negative integer k,
$$m_{2k} = \frac{1}{k+1}\binom{2k}{k} =: C_k. \tag{17.7}$$
The numbers C_k are the ***Catalan numbers****.*

Proof. The first assertion follows from a suitable change of variables, the second from symmetry, a trigonometric substitution and integration by parts. Indeed, $m_{2k-1} = 0$ by evenness of $x \mapsto \sigma_t(x)$ and for the even moments we compute

$$\begin{aligned} m_{2k} &= \int_{-2}^{2} x^{2k} \sigma_1(x) dx \\ &= \frac{2^{2k+1}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) \cos^2(\theta) d\theta \\ &= \frac{2^{2k+1}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) d\theta - \frac{2^{2k+1}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k+1}(\theta) \sin(\theta) d\theta \\ &= \frac{2^{2k+1}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) d\theta - (2k+1) m_{2k}, \end{aligned}$$

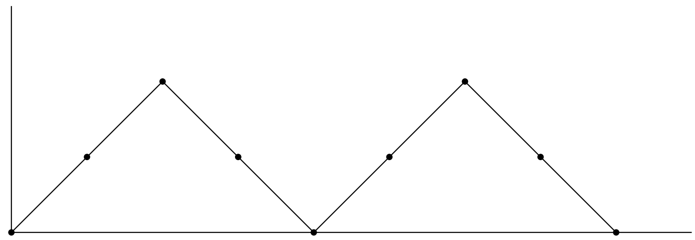


Figure 17.1.: An example of a Dyck path of length 8 (so $k = 4$). Each step to the right corresponds to either a unit step up or down.

so that, for any $k \in \mathbb{N}$, we can relate the $2k$th moment to the previous even integer moment

$$m_{2k} = \frac{2^{2k}}{\pi(k+1)} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) d\theta = \frac{2^{2k}}{\pi(k+1)} \int_{-\pi/2}^{\pi/2} \sin^{2k-1}(\theta) \sin(\theta) d\theta = \frac{2(2k-1)}{k+1} m_{2k-2}.$$

Note that the last recursion for m_{2k}, together with $m_0 = 1$, completely determines the even moments. It remains to check that $m_{2k} = C_k$ as in (17.7) indeed solves the same recursion. □

The Catalan numbers (17.7) possess several combinatorial interpretations and appear in many other applications. One that will feature prominently in our upcoming workings is related to *simple walks* on $\mathbb{Z}$.

Definition 17.5. *An integer-valued sequence $(S_n)_{n=0}^{\ell}$ is called a* ***Bernoulli walk*** *of length $\ell \in \mathbb{N}$, if $S_0 = 0$ and $|S_{n+1} - S_n| = 1$ for $n \leq \ell - 1$. A* ***Dyck path*** *of length $2k$ is a nonnegative Bernoulli walk of length $2k$ that terminates at $S_{2k} = 0$. See figure 17.1 for an example.*

We can count the number of possible Dyck paths of length $2k$. This is a nice exercise in combinatorics.

Lemma 17.6. *Let D_k denote the number of Dyck paths of length $2k$ with $k \in \mathbb{N}$. Then*

$$D_k = C_k \leq 4^k.$$

Further, the generating function

$$\hat{D}(z) := 1 + \sum_{k=1}^{\infty} D_k z^k$$

satisfies, for $|z| < \frac{1}{4}$,

$$\hat{D}(z) = \frac{1}{2z}\left(1 - (1-4z)^{\frac{1}{2}}\right). \tag{17.8}$$

(Technicality: in (17.8), the parameter z can be complex and the square-root function in the complex plane is multi-valued. We are using 'principal branch' such that $\hat{D}(0) = 1$.)

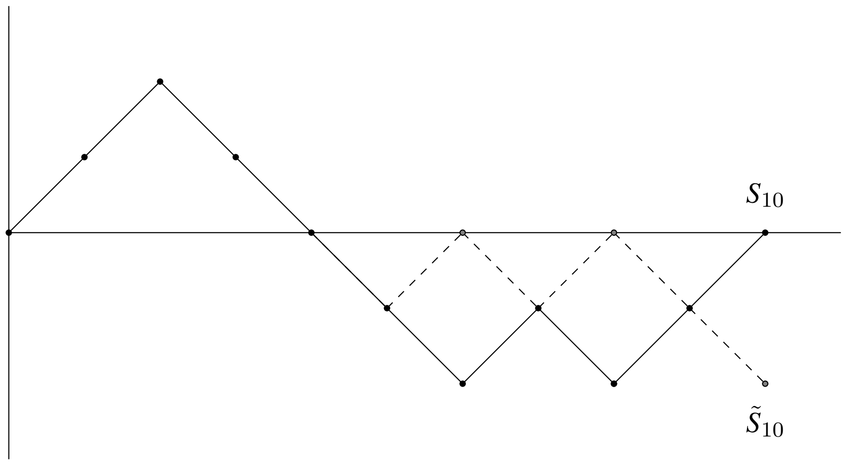


Figure 17.2.: An example of reflecting a Bernoulli walk on the first hitting of $S_m = -1$. The dashed path is the reflection, ending at -2.

Proof. We let B_k denote the number of Bernoulli walks $(S_n)_{n=0}^{2k}$ of length $2k$ that satisfy $S_{2k} = 0$ and β_k the number of Bernoulli walks $(S_n)_{n=0}^{2k}$ of length $2k$ that satisfy $S_{2k} = 0$ and $S_m < 0$ for some $m < 2k$ (so these are not Dyck paths). Evidently,

$$D_k = B_k - \beta_k$$

and $B_k = \binom{2k}{k}$, as one needs as many up and down steps to return to zero when starting from $S_0 = 0$.

On the other hand, perform the following procedure. Take a path $(S_n)_{n=0}^{2k}$ that crosses the x-axis at some point. Let the first crossing be at the step m, so $S_m < 0$. Define a reflected copy of the path $(\tilde{S}_n)_{n=0}^{2k}$ that takes the reflected step to $(S_n)_{n=0}^{2k}$ after step m, see figure 17.2 for an example). Hence $\tilde{S}_{2k} = -2$ and β_k equals the number of Bernoulli walks $(S_n)_{n=0}^{2k}$ of length $2k$ terminating at -2.

In turn, $\beta_k = \binom{2k}{k+1} = \binom{2k}{k-1}$ as one needs two more down steps to end up at $S_{2k} = -2$ when starting from $S_0 = 0$. All together,

$$D_k = \binom{2k}{k} - \binom{2k}{k-1} = \frac{1}{k+1}\binom{2k}{k} = C_k$$

for any $k \in \mathbb{N}$.

Turning to the evaluation of $\hat{D}(z)$, at each step of a Bernoulli walk we can either go up or down, so clearly $D_k \leq 2^{2k} = 4^k$ and thus the power series for $\hat{D}(z)$ is absolutely convergent for $|z| < 1/4$ (and so $\hat{D}(z)$ is analytic in the open disk $|z| < 1/4$).

Now, if a Bernoulli walk starts and ends at 0, then we could equally consider ‘breaking’ the walk at step $2p$ for some $1 \leq p \leq k$ and count all the ways we could return to 0 in $2p$ steps, multiplied by all the ways we could return to the ‘breakpoint’ in $2k - 2p$ steps (this is a similar principle to the Chapman-Kolmogorov equations you may have seen in

Probability 2). This gives the relations

$$D_0 = 1,$$
$$D_k = \sum_{p=1}^{k} D_{p-1} D_{k-p}, \quad k \in \mathbb{N}.$$

Substituting those into the definition of $\hat{D}(z), |z| < \frac{1}{4}$, we find after some algebra,

$$\hat{D}(z) - 1 = \sum_{k=1}^{\infty}\left(\sum_{p=1}^{k} D_{p-1}D_{k-p}\right)z^k = z\sum_{k=0}^{\infty}\left(\sum_{p=0}^{k} D_p D_{k-p}\right)z^k = z\left(\sum_{k=0}^{\infty} D_k z^k\right)\left(\sum_{p=0}^{\infty} D_p z^p\right),$$

and thus the quadratic $z(\hat{D}(z))^2 - \hat{D}(z) + 1 = 0$. Solving the same quadratic equation, while choosing the correct sign to match onto $\hat{D}(0) = 1$, results in (17.8). This completes our proof. □

Remark. *For $z \in \mathbb{C}$ such that $z \notin [-2, 2] \subset \mathbb{R}$, the* ***Stieltjes transform*** *$S(z)$ of the semicircle law ($t = 1$) equals*

$$S(z) := \int_{\mathbb{R}} \frac{\sigma_1(\lambda)d\lambda}{\lambda - z} = -\frac{1}{2}\left(z - (z^2 - 4)^{\frac{1}{2}}\right), \tag{17.9}$$

where the principal branch for the complex square root is chosen such that $S(z) \to 0$ as $|z| \to \infty$. We will return to integrals of the type (17.9) in Chapter 18 below.

17.2. Proof of Theorem 17.3

We now commence the proof of Theorem 17.3, i.e. the proof of Wigner's semicircle law for matrix moments. Suppose to that end we manage to complete the following *two-step outline*:

Step 1. Show that, as $N \to \infty$,

$$\frac{1}{N}\mathbb{E}_N\left[\,\mathrm{Tr}\left(\mathrm{M_N}^{\ell}\right)\right] \to \begin{cases} 0 & \ell = 2k-1, \\ t^k C_k & \ell = 2k \end{cases}, \tag{17.10}$$

for $k \in \mathbb{N}$, with the Catalan numbers C_k as in (17.7).

Step 2. Show that, as $N \to \infty$,

$$\mathrm{Var}\left(\frac{1}{N}\,\mathrm{Tr}\left(\mathrm{M_N}^{k}\right)\right) \to 0 \tag{17.11}$$

for all $k \in \mathbb{N}$.

Then Theorem 17.3 follows from Chebyshev's inequality (A.4) and Lemma 17.4:

Proof of Theorem 17.3. Abbreviate $X_{N,k} \coloneqq \frac{1}{N}\operatorname{Tr}(\mathrm{M_N}^k)$ with $N, k \in \mathbb{N}$. We have

$$\mathbb{P}_N\left(\left|X_{N,k} - t^{\frac{k}{2}} m_k\right| > \varepsilon\right) \le \mathbb{P}\left(\left|X_{N,k} - \mathbb{E}_N[X_{N,k}]\right| > \varepsilon/2\right) + \mathbb{P}\left(\left|\mathbb{E}_N[X_{N,k}] - t^{\frac{k}{2}} m_k\right| > \varepsilon/2\right). \tag{17.12}$$

By (17.10) and Lemma 17.4, as $N \to \infty$,

$$\mathbb{E}_N[X_{N,k}] \to t^{\frac{k}{2}} m_k \tag{17.13}$$

for $k \in \mathbb{N}$. Hence, for sufficiently large N, the second summand in the right hand side of (17.12) will be zero.

To conclude we examine the first summand in the right hand side of (17.12). By (A.4), as $N \to \infty$,

$$\begin{aligned}\mathbb{P}_N\left(\left|X_{N,k} - \mathbb{E}_N[X_{N,k}]\right| > \varepsilon\right) &\le \frac{1}{\varepsilon^2}\operatorname{Var}(X_{N,k}) \\ &\xrightarrow{(17.11)} 0 \end{aligned}\tag{17.14}$$

for all $k \in \mathbb{N}$. Thus, Theorem 17.3 follows. ◻

Remark. *Observe how the assumption $r_k < \infty$ for all $k \in \mathbb{N}$, see (17.1), is used in (17.10) as we take expectations of terms that involve kth powers of the variables $\xi_{j,k}$. Still, as previously mentioned, this is a technical assumption that can be removed after the fact with appropriate truncation arguments.*

We now follow through with the *two-step outline* (i) and (ii) above, using combinatorial arguments and beginning with a proof of (17.14) about the convergence of the averaged matrix moments.

The proofs of the two-step outline are a beautiful application of the (relatively!) simple combinatorial objects we've just encountered.

Proof of step 1.

First we have for any $\mathrm{M_N} = (\xi_{j,k})_{j,k=1}^N$, by linearity of expectation (dropping the subscript N in the expectation notation),

$$\frac{1}{N}\mathbb{E}\left[\operatorname{Tr}\left(\mathrm{M_N}^k\right)\right] = \frac{1}{N}\mathbb{E}\left[\sum_{j_1,\ldots,j_k=1}^{N} \xi_{j_1,j_2}\xi_{j_2,j_3}\cdot\ldots\cdot\xi_{j_{k-1},j_k}\xi_{j_k,j_1}\right] \eqqcolon \frac{1}{N}\sum_{\underline{j}\in\{1,\ldots,N\}^k}\mathbb{E}[\xi_{\underline{j}}], \tag{17.15}$$

where given the vector $\underline{j} \coloneqq (j_1, \ldots, j_k)$ we define $\xi_{\underline{j}}\xi_{j_1,j_2}\xi_{j_2,j_3}\cdot\ldots\cdot\xi_{j_{k-1},j_k}\xi_{j_k,j_1}$. The sum in the right hand side of (17.15) is determined by $\underline{j}$, but it will be advantageous to reindex by a collection of *walks on graphs* arising from indices.

Definition 17.7. *Let $\underline{j} = (j_1, \ldots, j_k) \in \{1, \ldots, N\}^k$. The **graph** $G_{\underline{j}} = (V_{\underline{j}}, E_{\underline{j}})$ **associated with $\underline{j}$** is defined as follows: the set of vertices, $V_{\underline{j}}$, consists of the distinct elements of $\{j_1, j_2, \ldots, j_k\}$ and the set of edges, $E_{\underline{j}}$, consists of the distinct pairs amongst*

$\{j_1, j_2\}, \{j_2, j_3\}, \ldots, \{j_{k-1}, j_k\}, \{j_k, j_1\}$. *Furthermore, the* ***walk*** $w_{\underline{j}}$ ***associated with*** $\boldsymbol{j}$ *is given by the edge sequence*

$$w_{\underline{j}} := \big(\{j_1, j_2\}, \{j_2, j_3\}, \ldots, \{j_{k-1}, j_k\}, \{j_k, j_1\}\big).$$

Note that the walk $w_{\underline{j}}$ *visits each edge of* $G_{\underline{j}}$*, including any self-edges present, and it begins and ends at the vertex* j_1*. This means in particular that the graph* $G_{\underline{j}}$ *is connected. Also, the walk* $w_{\underline{j}}$ *encodes a labelling of the edges* $E_{\underline{j}}$ *in that it implicitly counts the number of times each edge is traversed. Then for each edge* $e \in E_{\underline{j}}\}$*, we write*

$$w_{\underline{j}}(e) = |\{e' = \{e_1, e_2\} \in E_{\underline{j}} : e' = e\}|$$

for the number of times the walk traverses the edge e *(equivalently the number of times the edge* e *appears in* $E_{\underline{j}}$*).*

Example 17.8. *Suppose* $\underline{j} = (1, 2, 2, 3, 5, 2, 4, 1, 4, 2) \in \{1, \ldots, 5\}^{10}$*, then the vertex set is*

$$V_{\underline{j}} = \{1, 2, 3, 4, 5\},$$

and the edge set is

$$E_{\underline{j}} = \big\{\{1, 2\}, \{2, 2\}, \{2, 3\}, \{3, 5\}, \{5, 2\}, \{2, 4\}, \{4, 1\}\big\}.$$

The corresponding walk

$$w_{\underline{j}} = (\{1, 2\}, \{2, 2\}, \{2, 3\}, \{3, 5\}, \{5, 2\}, \{2, 4\}, \{4, 1\}, \{1, 4\}, \{4, 2\}, \{2, 1\})$$

defines a walk on the graph $G_{\underline{j}} = (V_{\underline{j}}, E_{\underline{j}})$*, see Figure 17.3. Also, by symmetry and construction of* $M_N = (\xi_{j,k})_{j,k=1}^N$*,*

$$\begin{aligned} \xi_{\underline{j}} &= \xi_{1,2}\xi_{2,2}\xi_{2,3}\xi_{3,5}\xi_{5,2}\xi_{2,4}\xi_{4,1}\xi_{1,4}\xi_{4,2}\xi_{2,1} \\ & \frac{1}{(\sqrt{N})^{10}} Z_{1,2}^2 Z_{1,4}^2 Z_{2,4}^2 Y_2 Z_{2,3} Z_{2,5} Z_{3,5}. \end{aligned} \tag{17.16}$$

For the number of times edges are traversed, we have for example

$$\begin{aligned} w_{\underline{j}}(\{1, 2\}) &= w_{\underline{j}}(\{1, 4\}) = w_{\underline{j}}(\{2, 4\}) = 2, \\ w_{\underline{j}}(\{2, 2\}) &= w_{\underline{j}}(\{2, 3\}) = w_{\underline{j}}(\{3, 5\}) = w_{\underline{j}}(\{5, 2\}) = 1, \end{aligned}$$

and the same labels are evident in the expansion of $\xi_{\underline{j}}$*, compare the second equality in (17.16).*

The observations in Example 17.8 are, by definition, true in general, so if we exploit the independence and i.i.d. assumptions of Definition 17.1, the following compact expression for $\mathbb{E}[\xi_{\underline{j}}]$ in (17.15) emerges,

$$\mathbb{E}[\xi_{\underline{j}}] = \frac{1}{(\sqrt{N})^k} \prod_{e \in E_{\underline{j}}^s} \mathbb{E}\Big[Y_1^{w_{\underline{j}}(e)}\Big] \prod_{e \in E_{\underline{j}}^c} \mathbb{E}\Big[Z_{1,2}^{w_{\underline{j}}(e)}\Big], \tag{17.17}$$

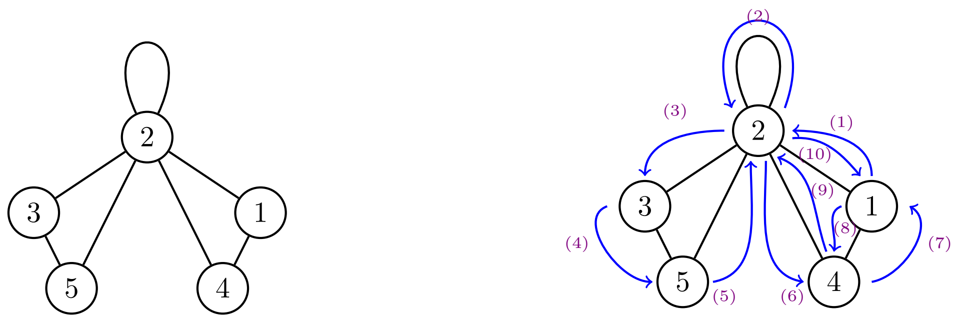


Figure 17.3.: Visual representation of the graph $G_{\underline{j}}$ and the walk on the graph in Example 17.8. The labels in the graph on the right are giving the order in which the edges are traversed.

where $E^s_{\underline{j}} := \{\{e_1, e_1\} \in E_{\underline{j}}\}$ are the *self-edges* in $E_{\underline{j}}$ and $E^c_{\underline{j}} := \{\{e_1, e_2\} \in E_{\underline{j}} : \ e_1 \neq e_2\}$ the *connecting edges* in $E_{\underline{j}}$. What (17.17) tells us, in particular, is that the value of $\mathbb{E}[\xi_{\underline{j}}]$ is determined by the pair $(G_{\underline{j}}, w_{\underline{j}})$, or better yet $(E_{\underline{j}}, w_{\underline{j}})$, and we henceforth abbreviate

$$\Pi(G_{\underline{j}}, w_{\underline{j}}) := \prod_{e \in E^s_{\underline{j}}} \mathbb{E}\Big[Y_1^{w_{\underline{j}}(e)}\Big] \prod_{e \in E^c_{\underline{j}}} \mathbb{E}\Big[Z_{1,2}^{w_{\underline{j}}(e)}\Big]. \tag{17.18}$$

Next, for any vector $\underline{j} \in \{1, \ldots, N\}^k$, the connected graph $G_{\underline{j}}$ has at most k vertices and for all $e \in E_{\underline{j}}$ we have in fact

$$|w_{\underline{j}}| := \sum_{e \in E_{\underline{j}}} w_{\underline{j}}(e) = k.$$

Motivated by these conditions, we consider now:

Definition 17.9. *Let $\mathcal{G}_k, k \in \mathbb{N}$ denote the set of all pairs (G, w) where $G = (V, E)$ is a connected graph*[1] *with at most k vertices given by the set V and edges given by the set E, and w is a closed walk covering G and satisfying $|w| = k$.*

Using all the above, i.e. (17.17), (17.18) and Definition 17.9, we finally reindex the sum in (17.15),

$$\begin{aligned} \frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\big(M_N^k\big)\Big] &= \frac{1}{N} \sum_{\underline{i} \in \{1,\ldots,N\}^k} \mathbb{E}[\xi_{\underline{i}}] \\ &= \sum_{(G,w) \in \mathcal{G}_k} \sum_{\substack{\underline{j} \in \{1,\ldots,N\}^k \\ (G_{\underline{j}}, w_{\underline{j}}) = (G,w)}} \frac{1}{N}\mathbb{E}[\xi_{\underline{j}}] \\ &= \frac{1}{N} \sum_{(G,w) \in \mathcal{G}_k} \Pi(G, w) \, \frac{|\{\underline{j} \in \{1, \ldots, N\}^k : \ (G_{\underline{j}}, w_{\underline{j}}) = (G, w)\}|}{(\sqrt{N})^k}, \end{aligned} \tag{17.19}$$

[1]We do not allow for multigraphs, i.e. those graphs with multiple edges.

and are now left to count the sets of vectors $\underline{j}$ in (17.19). For a fixed, but arbitrary, $(G, w) \in \mathcal{G}_k$, an index $\underline{j} \in \{1, \dots, N\}^k$ with that corresponding graph G and walk w is completely determined by assigning which distinct values from $\{1, \dots, N\}$ appear as the vertices of G. So, if $|G|$ denotes the number of vertices of the graph G, then there are

$$N(N-1)(N-2) \cdot \ldots \cdot (N - |G| + 1)$$

ways of doing this. Consequently we have

Lemma 17.10. *Given $(G, w) \in \mathcal{G}_k$ with $k \in \mathbb{N}$, denote by $|G|$ the number of vertices in G. Then*

$$\left|\left\{\underline{j} \in \{1, \dots, N\}^k : \ (G_{\underline{j}}, w_{\underline{j}}) = (G, w)\right\}\right| = N(N-1) \cdot \ldots \cdot (N - |G| + 1).$$

In turn, (17.19) simplifies to

$$\frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^k\right)\Big] = \sum_{(G,w)\in\mathcal{G}_k} \Pi(G, w) \frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k/2+1}}, \tag{17.20}$$

and since $k \in \mathbb{N}$ is fixed throughout, see (17.10), we are left to determine the values $\Pi(G, w)$. We begin with a simple observation: let $(G, w) \in \mathcal{G}_k$ and suppose there exists an edge e that is only traversed once in the walk, so $w(e) = 1$. This means that, in the expression (17.18) for $\Pi(G, w)$, a singleton term

$$\mathbb{E}\big[Y_1^{w(e)}\big] = \mathbb{E}[Y_1] \qquad \text{or} \qquad \mathbb{E}\big[Z_{1,2}^{w(e)}\big] = \mathbb{E}[Z_{1,2}]$$

appears. But by Definition 17.1, all variables are centered (i.e. have zero mean) and so those terms are zero. Hence the product $\Pi(G, w)$ vanishes for any such pair (G, w). This reduces the sum in (17.20) considerably, since we need only consider those walks w that cross each edge at least twice. We record this condition as $w \geq 2$, so (17.20) becomes

$$\frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^k\right)\Big] = \sum_{\substack{(G,w)\in\mathcal{G}_k \\ w\geq 2}} \Pi(G, w) \frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k/2+1}}. \tag{17.21}$$

The condition $w \geq 2$ restricts those graphs that can appear: since $|w| = k$, if each edge in G is traversed at least twice, then the number of edges cannot exceed $k/2$. A classical combinatorial result is the following.

Proposition 17.11. *Let $G = (V, E)$ be a finite connected[2] graph. Then $|V| \leq |E| + 1$ and $|V| = |E| + 1$ if and only if G is a tree, i.e. a connected graph not containing any cycles (i.e. not containing any walks over distinct edges from one vertex back to itself).*

Proof. If $|V| = 1$ then the claim is obvious, because $|E| \geq 0$ in general and $|E| = 0$ in this case if and only if G is a tree. Now use induction on the number of vertices to prove the inequality: if G has $v + 1$ vertices then consider any subgraph $G' = (V', E')$ consisting

[2]This is, a graph with a finite number of vertices and there exists a path between all pairs of vertices.

of ν vertices of G and all edges of G that do not connect to the remaining vertex. By hypothesis, $\nu = |V'| \le |E'| + 1$ and so

$$|V| = \nu + 1 = |V'| + 1 \le (|E'| + 1) + 1 \le |E| + 1,$$

as we need at least one more edge to link the remaining vertex to the vertices in V', seeing that G is connected. Lastly, if G is a tree, then any subgraph thereof is also a tree, so a close examination of the previous induction workings yields, by induction on the number of vertices, the equality $|V| = |E| + 1$. Conversely, if $|V| = |E| + 1$, then G cannot contain a cycle, for otherwise there is at least one vertex such that, once removed, the number of edges in G reduces by at least two. But this is impossible in view of $|V| = |E| + 1$. □

Using Proposition 17.11, we see that for any graph $G = (V, E)$ appearing in the sum in the right hand side of (17.21) we have $|G| \le k/2 + 1$ as the number of edges cannot exceed $k/2$, given that they need to be traversed at least twice. On the other hand, the product $N(N-1) \cdot \ldots \cdot (N - |G| + 1)$ in (17.21) is asymptotically equal to $N^{|G|}$ as $N \to \infty$, and hence

$$\left(\frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k/2+1}} \right)_{N=1}^{\infty}$$

is a bounded sequence. Hence $\frac{1}{N}\mathbb{E}[\mathrm{Tr}(M_N^k)]$ is bounded for all $k \in \mathbb{N}$. What's more, suppose that k is odd. Since $|G| \le k/2 + 1$ and $|G|$ is integer-valued, it follows that $|G| \le (k-1)/2 + 1 = k/2 + 1/2$. Hence, in this case, all the terms in the sum in (17.21) are of order $O(N^{-1/2})$ for fixed $k \in \mathbb{N}$, and consequently we have proven the first limit in (17.10), i.e., as $N \to \infty$,

$$\frac{1}{N}\mathbb{E}\Big[\mathrm{Tr}\left(M_N^{2k-1} \right) \Big] \to 0 \tag{17.22}$$

for all $k \in \mathbb{N}$.

Henceforth, we assume that k is even. In this case, it is still true that most of the terms in the sum in (17.21) are zero. Our next result testifies to this fact.

Proposition 17.12. *Let $(G, w) \in \mathcal{G}_k$ with $k \in \mathbb{N}$ even and $w \ge 2$. Then we have the following:*

1. *If there exists any self-edge e in G, then $|G| \le k/2$.*
2. *If there exists an edge e in G with $w(e) \ge 3$, then $|G| \le k/2$.*

Proof. If $G = (V, E)$ contains a self-edge, then it contains a cycle and so is not a tree. Hence, by Proposition 17.11, $|V| < |E| + 1$. But $w \ge 2$ implies that $|E| \le k/2$ and so $|V| < k/2 + 1$, i.e. $|G| \le k/2$ since $|G|$ is integer-valued. Next, if for some e' we have $w(e') \ge 3$, then $\sum_{e \in E \setminus \{e'\}} w(e) \le k - 3$ and thus

$$|E| \le 1 + \frac{1}{2}(k-3) = \frac{1}{2}(k-1)$$

since $w \ge 2$. By Proposition 17.11 then $|V| \le \frac{1}{2}(k-1) + 1 = \frac{1}{2}(k+1)$. But k is even, so in fact $|G| = |V| \le k/2$. This completes our proof. □

Combining Proposition 17.12 with (17.21) suggests we drastically refine the set $\mathcal{G}_k$ for even $k \in \mathbb{N}$.

Definition 17.13. *Take an even positive integer 2ℓ. Let $\mathcal{G}_{2\ell}^{\circ}$ be defined as the set of pairs $(G, w) \in \mathcal{G}_{2\ell}$ where G has $\ell + 1$ vertices, G contains no self-edges, and the walk w traverses every edge exactly twice.*

In turn,

$$\frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^k\right)\Big] = \sum_{(G,w)\in\mathcal{G}_k^{\circ}} \Pi(G,w)\,\frac{N(N-1)\cdot\ldots\cdot(N-|G|+1)}{N^{k/2+1}} + O_k\left(\frac{1}{N}\right), \tag{17.23}$$

where $O_k(1/N)$ means that the absolute difference of the two explicit terms in (17.23) is at most $c(k)/N$ for some N-independent term $c(k) > 0$. Since $|G| = k/2+1$ for the G appearing in the sum in (17.23), and so $N(N-1)\cdot\ldots\cdot(N-|G|+1) \sim N^{k/2+1}$ as $N \to \infty$, we deduce from (17.23),

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^{2k}\right)\Big] = \sum_{(G,w)\in\mathcal{G}_{2k}^{\circ}} \Pi(G,w) \tag{17.24}$$

for all $k \in \mathbb{N}$.

Thus, it remains to show that the outstanding sum in (17.24) is equal to $t^k C_k$, with C_k as in (17.7). Well, if $(G,w) \in \mathcal{G}_k^{\circ}$, then number of edges in G is $k/2$ since w traverses each edge exactly twice, and so G must be a tree by Proposition 17.11. In particular G cannot contain any self-edges and so the value $\Pi(G,w)$, see (17.18), is simply

$$\Pi(G,w) = \prod_{e\in E^c}\mathbb{E}\Big[Z_{1,2}^{w(e)}\Big] = \prod_{e\in E^c}\mathbb{E}\big[Z_{1,2}^2\big] = t^{|E|} = t^{\frac{k}{2}},$$

having used Definition 17.1. Consequently, from (17.24), for any $k \in \mathbb{N}$ we have

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^{2k}\right)\Big] = t^k\,|\mathcal{G}_{2k}^{\circ}|. \tag{17.25}$$

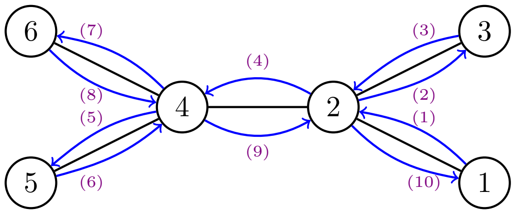
Finally, to complete the derivation of (17.10), we must enumerate the set $\mathcal{G}_{2k}^{\circ}$. To do so, we utilize the Dyck paths of Definition 17.5: given $(G,w) \in \mathcal{G}_{2k}^{\circ}, k \in \mathbb{N}$ where $w = (w_{j_1}, w_{j_2}, \ldots, w_{j_{2k}})$ with $w_{j_i} \in E$, we define a sequence $d = d(G,w) \in \{-1,+1\}^{2k}$ recursively as follows:

1. Set $d_1 \coloneqq +1$.
2. For $i \in \{2,3,\ldots,2k\}$, if $w_{j_i} \notin \{w_{j_1}, w_{j_2}, \ldots, w_{j_{i-1}}\}$, set $d_i \coloneqq +1$; otherwise, set $d_i \coloneqq -1$.
3. Put $d = d(G,w) = (d_i)_{i=1}^{2k}$.

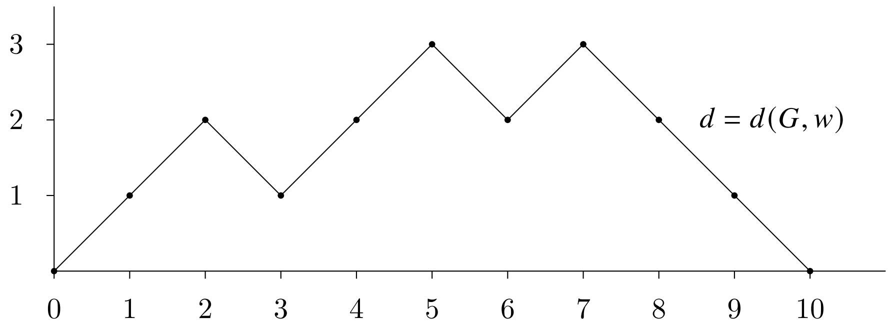
Example 17.14. *Consider $(G, w) \in \mathcal{G}_{10}^{\circ}$ where G is a tree with six vertices $\{1, 2, 3, 4, 5, 6\}$, with inner vertices $\{2, 4\}$, and w denotes the walk*

$$w = \big(\{1,2\}, \{2,3\}, \{3,2\}, \{2,4\}, \{4,5\}, \{5,4\}, \{4,6\}, \{6,4\}, \{4,2\}, \{2,1\}\big),$$

also denoted in the following figure.



Then, by the above recursion, $d = d(G, w) = (+1, +1, -1, +1, +1, -1, +1, -1, -1, -1)$, which we view as a lattice path on $\mathbb{Z}_{\geq 0} \times \mathbb{Z}$: set $P_0 = (0, 0)$ and $P_i = (i, d_1 + \ldots + d_i)$ where, by construction, $d_1 + \ldots + d_i \geq 0$ for all $i = 1, \ldots, 2k$ and $d_1 + \ldots + d_{2k} = 0$. In short, $d = d(G, w)$ is a Dyck path. See the following pictoral representation.



Using this construction, we can see that the map taking the pair (G, w) to the Dyck path $d(G, W)$ is actually a bijection between $\mathcal{G}_{2k}^{\circ}$ and the set of Dyck paths of length $2k$. Thus by Lemma 17.6, $|\mathcal{G}_{2k}^{\circ}| = \beta_k = C_k$. Combining the same with (17.25) finally results in

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E}\Big[\operatorname{Tr}\big(M_N^{2k}\big) \Big] = t^k\, C_k,$$

and completes, together with (17.22), our proof of (17.10). Step 1 is complete.

We now move to step 2 and a proof of (17.11).

Proof of step 2.

We proceed as in the previous subsection, expanding the variance in terms of the entries of the matrix M_N, compare (17.15),

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right) &= \mathbb{E}\left[\left(\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right]\right)^2 \\ &= \frac{1}{N^2}\sum_{\underline{i},\underline{j}\in\{1,\ldots,N\}^k}\left(\mathbb{E}[\xi_{\underline{i}}\xi_{\underline{j}}] - \mathbb{E}[\xi_{\underline{i}}]\mathbb{E}[\xi_{\underline{j}}]\right),\end{aligned}$$

where, as before, the values $\mathbb{E}[\xi_{\underline{i}}\xi_{\underline{j}}]$ and $\mathbb{E}[\xi_{\underline{i}}]\mathbb{E}[\xi_{\underline{j}}]$ only depend on the vectors $\underline{i},\underline{j}\in\{1,\ldots,N\}^k$ through a certain graph structure underlying the $2k$-tuple $(\underline{i},\underline{j})$. Indeed, if $G_{\underline{i}} = (V_{\underline{i}}, E_{\underline{i}})$, resp. $G_{\underline{j}} = (V_{\underline{j}}, E_{\underline{j}})$, are the graphs associated with $\underline{i}$, resp. $\underline{j}$, then consider the graph

$$G_{\underline{i}} \cup G_{\underline{j}} \coloneqq (V_{\underline{i}} \cup V_{\underline{j}}, E_{\underline{i}} \cup E_{\underline{j}}),$$

whose vertex set is the union of the vertex sets, and whose edge set is the union of the edge sets.

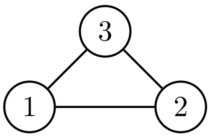
Example 17.15. *If* $\underline{i} = (1,2,1,3,2,3) \in \{1,\ldots,5\}^6$ *and* $\underline{j} = (2,4,5,1,4,3) \in \{1,\ldots,5\}^6$, *then*

$$\xi_{\underline{i}}\xi_{\underline{j}} = (\xi_{1,2}\xi_{2,1}\xi_{1,3}\xi_{3,2}\xi_{2,3}\xi_{3,1})(\xi_{2,4}\xi_{4,5}\xi_{5,1}\xi_{1,4}\xi_{4,3}\xi_{3,2})$$

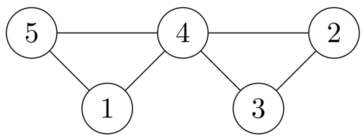
and so

$$G_{\underline{i}} \cup G_{\underline{j}} = \Big(\{1,2,3,4,5\}, \big\{\{1,2\},\{1,3\},\{3,2\},\{2,4\},\{4,5\},\{5,1\},\{1,4\},\{4,3\}\big\}\Big).$$

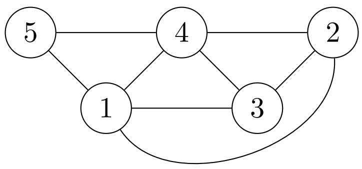
Pictorially:



(a) $G_{\underline{i}}$



(b) $G_{\underline{j}}$



(c) $G_{\underline{i}} \cup G_{\underline{j}}$

The product $\xi_{\underline{i}}\xi_{\underline{j}}$ also gives rise to the two walks $w_{\underline{i}}$ and $w_{\underline{j}}$, but, as Example 17.15 shows, they do not constitute a single walk of length $2k$, as there is a priori no reason for the endpoint of the first walk to coincide with the starting point of the second walk. Still, since we can recover $(G_{\underline{i}}, w_{\underline{i}})$ and $(G_{\underline{j}}, w_{\underline{j}})$ from $(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}})$, the same reasoning as back in (17.17), (17.18) shows that the value of $\mathbb{E}[\xi_{\underline{i}}\xi_{\underline{j}}] - \mathbb{E}[\xi_{\underline{i}}]\mathbb{E}[\xi_{\underline{j}}]$ is determined by the data $(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}})$. Denote this common value as

$$\mathbb{E}[\xi_{\underline{i}}\xi_{\underline{j}}] - \mathbb{E}[\xi_{\underline{i}}]\mathbb{E}[\xi_{\underline{j}}] =: \frac{1}{N^k}\Pi(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}}),$$

and then generalize Definition 17.9 to the current setting:

Definition 17.16. *Let $\mathcal{G}_{k,k}$, $k \in \mathbb{N}$ denote the set of all triples (G, w, w') where $G = (V, E)$ is a connected graph (with vertices V and edges E) with at most $2k$ vertices, and w, w' are closed walks whose union covers G and which satisfy $|w| = |w'| = k$.*

This means we can expand, as before in (17.19),

$$\operatorname{Var}\left(\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right) = \sum_{(G,w,w')\in\mathcal{G}_{k,k}} \Pi(G, w, w')\,\frac{|\{(\underline{i},\underline{j}) \in \{1,\ldots,N\}^{2k} :\ (G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}}) = (G, w, w')\}|}{N^{k+2}}.$$

Next, we let $E^s_{\underline{i}\underline{j}}$, resp. $E^c_{\underline{i}\underline{j}}$, denote the self-edges, resp. connecting edges, in $G_{\underline{i}} \cup G_{\underline{j}}$, and $w_{\underline{i}\underline{j}}(e)$ the number of times the edge $e \in E_{\underline{i}} \cup E_{\underline{j}}$ is traversed by either of the two walks $w_{\underline{i}}$ and $w_{\underline{j}}$. Then, by Definition 17.1,

$$\begin{aligned}
&\Pi(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}}) && (17.26)\\
&\quad= \prod_{e\in E^s_{\underline{i}\underline{j}}} \mathbb{E}\left[Y_1^{w_{\underline{i}\underline{j}}(e)}\right] \prod_{e\in E^c_{\underline{i}\underline{j}}} \mathbb{E}\left[Z_{1,2}^{w_{\underline{i}\underline{j}}(e)}\right]\\
&\qquad - \prod_{e\in E^s_{\underline{i}}} \mathbb{E}\left[Y_1^{w_{\underline{i}}(e)}\right] \prod_{e\in E^c_{\underline{i}}} \mathbb{E}\left[Z_{1,2}^{w_{\underline{i}}(e)}\right] \prod_{e\in E^s_{\underline{j}}} \mathbb{E}\left[Y_1^{w_{\underline{j}}(e)}\right] \prod_{e\in E^c_{\underline{j}}} \mathbb{E}\left[Z_{1,2}^{w_{\underline{j}}(e)}\right]. && (17.27)
\end{aligned}$$

The key thing is to note that

$$\sum_{e\in E_{\underline{i}}\cup E_{\underline{j}}} w_{\underline{i}\underline{j}}(e) = 2k = \sum_{e\in E_{\underline{i}}} w_{\underline{i}}(e) + \sum_{e\in E_{\underline{j}}} w_{\underline{j}}(e),$$

that is, the sum of all exponents in either term $\mathbb{E}[\xi_{\underline{i}}\xi_{\underline{j}}]$ or $\mathbb{E}[\xi_{\underline{i}}]\mathbb{E}[\xi_{\underline{j}}]$ is the total label (or length) of the walk, which is $2k$. Thus, using now crucially the bound (17.1), each of the terms in (17.27) is bounded above, in absolute value, by something of the form $r_{m_1} \cdot \ldots \cdot r_{m_\ell}$ for $m_1, \ldots, m_\ell \in \mathbb{Z}_{\geq 1}$ for which $m_1 + \ldots + m_\ell = 2k$. Precisely, we have the blunt upper bound

$$\left|\Pi(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}})\right| \leq 2 \prod_{\substack{j=1\\ \sum_{\underline{i}} m_{\underline{i}}=2k}}^{\ell} r_{m_{\underline{j}}} \leq 2c_{2k},$$

for some large, but undetermined, constant c_{2k} as there are only finitely many non-negative integers $(m_1, \ldots, m_\ell)$ such that $m_1 + \ldots + m_\ell = 2k$. That being said, we can show that many of these terms in (17.27) are identically zero, as before. Indeed: by construction, every edge in the union graph $G_{\underline{i}} \cup G_{\underline{j}}$ is traversed at least once by the union of the two walks $w_{\underline{i}}$ and $w_{\underline{j}}$. Suppose $e \in E_{\underline{i}} \cup E_{\underline{j}}$ is traversed only *once*. This means that $w_{\underline{i}\underline{j}}(e) = 1$, and so it follows that the two values $w_{\underline{i}}(e), w_{\underline{j}}(e)$ are in $\{0, 1\}$. Hence, (17.27) and the fact that our matrix entries are centered show that $\Pi(G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}})$ vanishes in this case. Consequently, the variance sum reduces to

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right) \\
&\quad = \sum_{\substack{(G,w,w') \in \mathcal{G}_{k,k} \\ w+w' \geq 2}} \Pi(G, w, w') \, \frac{\left|\{(\underline{i}, \underline{j}) \in \{1, \dots, N\}^{2k} : \ (G_{\underline{i}} \cup G_{\underline{j}}, w_{\underline{i}}, w_{\underline{j}}) = (G, w, w')\}\right|}{N^{k+2}}.
\end{aligned}$$

However, the enumeration of the number of $2k$-tuples yielding a certain graph with two walks is the same as in our previous workings, see Lemma 17.10: the structure (G, w, w') specifies the $2k$-tuple precisely once we select the $|G|$ distinct indices for the vertices. So, as before, the remaining ratio in the variance sum becomes

$$\frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k+2}}.$$

Now, we have the condition $w + w' \geq 2$, meaning every edge is traversed at least twice. Since there are k steps in each of the two walks, this means there are at most k edges. Consequently, by Proposition 17.11, it follows that $|G| \leq k + 1$ and hence

$$\frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k+2}} = O_k\left(N^{|G|-k-2}\right) = O_k\left(\frac{1}{N}\right)$$

as $N \to \infty$. In summary, we have proven that

$$\mathrm{Var}\left(\frac{1}{N}\operatorname{Tr}\left(M_N^k\right)\right) = \sum_{\substack{(G,w,w') \in \mathcal{G}_{k,k} \\ w+w' \geq 2}} \Pi(G, w, w') \, \frac{N(N-1) \cdot \ldots \cdot (N - |G| + 1)}{N^{k+2}} = O_k\left(\frac{1}{N}\right)$$

and the same yields (17.11), proving step 2. Hence, the proof of Theorem 17.3 is complete.

17.3. Proof of Theorem 17.2

We have proven Theorem 17.3, a weaker form of Theorem 17.2, in the previous section. Moving forward, we can now fairly easily prove the stronger result (17.4) and afterwards (17.3). Let us first remark that we may easily fix the variance of $Z_{1,2}$ to be $t = 1$: an elementary scaling argument then extends Theorem 17.2 to the general case $t > 0$. With this convention in hand, we begin with the following truncation lemma.

Lemma 17.17. *Let $k \in \mathbb{N}, \varepsilon > 0$ and $b \geq 0$ be fixed. Take $\Delta = \Delta_b \subset \mathbb{R}$ to be*

$$\Delta := (-\infty, b) \cup (b, \infty).$$

Define a function (dependent on Δ) $\varphi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ by

$$\varphi(x) := |x|^k \chi_\Delta(x) = \begin{cases} |x|^k, & x \in \Delta \\ 0, & x \notin \Delta \end{cases}.$$

Then for any $b > 4$

$$\limsup_{N\to\infty} \mathbb{P}\left(\frac{1}{N}\mathcal{N}_N[\varphi] > \varepsilon\right) = 0. \tag{17.28}$$

Proof. First, by Markov's inequality (cf. Theorem A.10), noting that $\mathcal{N}_n[\varphi] = \sum_{j=1}^N \varphi(\lambda_j) \geq 0$, we have

$$\mathbb{P}\left(\frac{1}{N}\mathcal{N}_N[\varphi] > \varepsilon\right) \leq \frac{1}{\varepsilon N}\mathbb{E}\big[\mathcal{N}_N[\varphi]\big],$$

where

$$\mathcal{N}_N[\varphi] = \sum_{j=1}^N |\lambda_j(M_N)|^k \chi_\Delta\big(\lambda_j(M_N)\big) \leq \frac{1}{b^k}\sum_{j=1}^N |\lambda_j(M_N)|^{2k} \chi_\Delta\big(\lambda_j(M_N)\big) \leq \frac{1}{b^k}\operatorname{Tr}\left(M_N^{2k}\right).$$

Hence,

$$\mathbb{P}\left(\frac{1}{N}\mathcal{N}_N[\varphi] > \varepsilon\right) \leq \frac{1}{\varepsilon N b^k}\mathbb{E}\Big[\operatorname{Tr}\left(M_N^{2k}\right)\Big],$$

where the right hand side converges to $C_k/\varepsilon b^k$ as $N \to \infty$ by (17.10). Hence it follows that, using also Lemma 17.6,

$$\limsup_{N\to\infty} \mathbb{P}\left(\frac{1}{N}\mathcal{N}_N[\varphi] > \varepsilon\right) \leq \frac{C_k}{\varepsilon b^k} \leq \frac{1}{\varepsilon}\left(\frac{4}{b}\right)^k. \tag{17.29}$$

On the other hand, when $|x| > b > 4 > 1$, the function $k \mapsto |x|^k$ is strictly increasing, which means that the sequence of lim sups in the left-hand side of (17.29) is increasing in k. But this sequence decays exponentially since $4/b < 1$. The only way this is possible is if the sequence of lim sups is identically zero, as claimed in (17.28). This completes our proof of the Lemma. □

We are now ready to prove Theorem 17.2.

Proof of Theorem 17.2. Fix a continuous bounded function $\varphi : \mathbb{R} \to \mathbb{C}$, fix $\varepsilon > 0$ and fix $b > 4$. By the Weierstrass approximation theorem, (cf. by Theorem A.1) there exists a polynomial $P_\varepsilon \in \mathbb{C}[x]$ such that

$$\sup_{|x|\leq b} |\varphi(x) - P_\varepsilon(x)| < \frac{\varepsilon}{6}.$$

Then, by triangle inequality estimates,

$$\begin{aligned} \left|\frac{1}{N}\mathcal{N}_N[\varphi] - \int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx\right| &\leq \left|\frac{1}{N}\mathcal{N}_N[\varphi] - \frac{1}{N}\mathcal{N}_N[P_\varepsilon]\right| + \left|\frac{1}{N}\mathcal{N}_N[P_\varepsilon] - \int_{\mathbb{R}} P_\varepsilon(x)\sigma_1(x)dx\right| \\ &\quad + \left|\int_{\mathbb{R}} P_\varepsilon(x)\sigma_1(x)dx - \int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx\right| \\ &=: A_1 + A_2 + A_3. \end{aligned} \tag{17.30}$$

Hence, the event $\{|\frac{1}{N}\mathcal{N}_N[\varphi] - \int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx| > \varepsilon\}$ is contained in the union of three events that each of the three terms in the right hand side of (17.30) is bigger than $\varepsilon/3$. This means

$$\mathbb{P}\left(\left|\frac{1}{N}\mathcal{N}_N[\varphi] - \int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx\right| > \varepsilon\right) \le \mathbb{P}\left(A_1 > \frac{\varepsilon}{3}\right) + \mathbb{P}\left(A_2 > \frac{\varepsilon}{3}\right) + \mathbb{P}\left(A_3 > \frac{\varepsilon}{3}\right).$$

By construction, $|f(x) - P_\varepsilon(x)| < \varepsilon/6$ on $[-b, b]$, which includes the support $[-2, 2]$ of $x \mapsto \sigma_1(x)$. Hence the term $\mathbb{P}(A_3 > \varepsilon/3)$ is identically zero. For the term $\mathbb{P}(A_1 > \varepsilon/3)$ we partition over $\Delta = [-b, b]$ and its complement,

$$\begin{aligned}\left|\frac{1}{N}\mathcal{N}_N[\varphi] - \frac{1}{N}\mathcal{N}_N[P_\varepsilon]\right| &\le \frac{1}{N}\sum_{j=1}^{N}\Big|\varphi(\lambda_j(M_N)) - P_\varepsilon\big(\lambda_j(M_N)\big)\Big|\chi_\Delta\big(\lambda_j(M_N)\big)\\ &\quad + \frac{1}{N}\sum_{j=1}^{N}\Big|\varphi(\lambda_j(M_N)) - P_\varepsilon\big(\lambda_j(M_N)\big)\Big|\chi_{\mathbb{R}\setminus\Delta}\big(\lambda_j(M_N)\big)\\ &=: B_1 + B_2.\end{aligned}$$

By the same reasoning as above, we then estimate

$$\begin{aligned}\mathbb{P}\left(A_1 > \frac{\varepsilon}{3}\right) &= \mathbb{P}\left(\left|\frac{1}{N}\mathcal{N}_N[\varphi] - \frac{1}{N}\mathcal{N}_N[P_\varepsilon]\right| > \frac{\varepsilon}{3}\right)\\ &\le \mathbb{P}\left\{B_1 > \frac{\varepsilon}{6}\right\} + \mathbb{P}\left\{B_2 > \frac{\varepsilon}{6}\right\},\end{aligned} \tag{17.31}$$

where, by construction, $|f(x) - P_\varepsilon(x)| < \varepsilon/6$ on $\Delta = [-b, b]$ and so the first term in the right hand side of (17.31) vanishes. Consequently, combining this information in (17.30), (17.31) we have

$$\text{LHS in (17.30)} \le \mathbb{P}\left(B_2 > \frac{\varepsilon}{6}\right) + \mathbb{P}\left(A_2 > \frac{\varepsilon}{3}\right). \tag{17.32}$$

That the second term in (17.32) tends to zero as $N \to \infty$ follows immediately from Theorem 17.3, for if $P_\varepsilon(x) = \sum_{k=0}^{d} f_k x^k$ then with some $c_k > 0$, as $N \to \infty$,

$$\mathbb{P}\left(A_2 > \frac{\varepsilon}{3}\right) \le \sum_{k=0}^{d} |f_k|\ \mathbb{P}\left(\left|\frac{1}{N}\operatorname{Tr}\left(M_N^k\right) - \int_{\mathbb{R}} x^k\sigma_1(x)dx\right| > \varepsilon c_k\right) \xrightarrow{(17.5)} 0.$$

So we are left only to estimate the first term in the right hand side of (17.32). We do this as follows: let $d = \deg(P_\varepsilon)$. Since φ is globally bounded, we have $|\varphi(x) - P_\varepsilon(x)| \le \|\varphi\|_\infty + |P_\varepsilon(x)|$ with $\|\varphi\|_\infty := \sup_{x\in\mathbb{R}} |\varphi(x)| < \infty$, which yields on the set $\mathbb{R} \setminus \Delta$ the bound

$$|f(x) - P_\varepsilon(x)| \le c|x|^d, \qquad x \in \mathbb{R} \setminus \Delta$$

for some constant $c > 0$. This implies that

$$\mathbb{P}\left(B_2 > \frac{\varepsilon}{6}\right) \le \mathbb{P}\left(\left|\frac{1}{N}\sum_{j=1}^{N}|\lambda_j(M_N)|^d\chi_{\mathbb{R}\setminus\Delta}\big(\lambda_j(M_N)\big)\right| > \frac{\varepsilon}{6c}\right)$$

and so by Lemma 17.17 the lim sup of the first term in the right hand side of (17.32), as $N \to \infty$, equals zero. This completes our proof of (17.4) in Theorem 17.3. For (17.3), with (17.4) established, one can cite a general theorem in measure theory, see Billingsley, *Probability and measure*, Theorem 25.8, using crucially that the density function $x \mapsto \sigma_t(x)$ is continuous. We leave these details to the dedicated reader and conclude with this comment our proof of Theorem 17.2. □

17.4. Maximal eigenvalues

The Wigner theorems 17.2 and 17.3 assert in particular the convergence of the normalized counting measure of eigenvalues to the compactly supported semicircle law. One is immediately led to suspect that the maximal eigenvalue of M_N should converge in some sense to the value $2\sqrt{t}$, the largest element of the support of the semicircle distribution. After all we have recorded a similar result for the GUE in Theorem 16.7. However, we have to be more careful in the case of real Wigner matrices as fluctuations and large deviations of the normalized counting measure of eigenvalues around the semicircle law are heavily dependent on the distribution of the matrix entries of M_N.

Lemma 17.18. *Let M_N be a real Wigner matrix as in Definition 17.1 with $t = 1$ and let $\lambda_N = \lambda_N(M_N)$ denote its largest eigenvalue. Then for any $\delta > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\big(\lambda_N < 2 - \delta\big) = 0. \tag{17.33}$$

Proof. Fix $\delta > 0$ and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous function supported on $[2-\delta, 2]$ with $\int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx = 1$. If $\lambda_N < 2 - \delta$, then

$$\frac{1}{N}\mathcal{N}_N[\varphi] = \frac{1}{N}\sum_{j=1}^{N}\varphi(\lambda_j) = \frac{1}{N}\sum_{j=1}^{N}\varphi(\lambda_j)\chi_{(-\infty,2-\delta)}(\lambda_j) = 0.$$

On the other hand, since $\int_{\mathbb{R}} \varphi(x)\sigma_1(x)dx = 1 > \frac{1}{2}$, we have

$$\mathbb{P}\big(\lambda_N < 2-\delta\big) \le \mathbb{P}\Big(\frac{1}{N}\mathcal{N}_N[\varphi] = 0\Big) \le \mathbb{P}\left(\left|\frac{1}{N}\mathcal{N}_N[\varphi] - \int_{\mathbb{R}}\varphi(x)\sigma_1(x)dx\right| > \frac{1}{2}\right)$$

and the last quantity tends to zero as $N \to \infty$ by Theorem 17.2. This completes our proof. □

To prove that for any $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\big(|\lambda_N - 2| > \varepsilon\big) = 0,$$

i.e. that $\lambda_N \to 2$ as $N \to \infty$ (in probability), it therefore suffices to prove a complementary estimate on the probability that λ_N is large, precisely that $\lambda_N > 2+\delta$. As it happens, *this can fail to be true.* If the entries of M_N have 'heavy tails' (informally: the density function has relatively slow decay), then the largest eigenvalue may fail to converge at all. See for example Figures 17.5 and 17.6. There, notice the 'tightness' of Figure 17.5b: whilst the majority of the eigenvalues still fit in the (red) semicircle, the support of the density is far wider as there are a few large (positive and negative) eigenvalues. Similarly, (perhaps after zooming!) notice the difference between the largest values present in Figures 17.6a and 17.6b.

However, if we assume in addition to (17.1) that all entries of M_N are bounded, then the convergence of $\lambda_N(M_N)$ (in probability) to 2 holds true.

Theorem 17.19. *Let M_N be a real Wigner matrix as in Definition 17.1 with $t = 1$, with bounded entries and let $\lambda_N = \lambda_N(M_N)$ denote its largest eigenvalue. Then for any $\delta > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\big(\lambda_N > 2 + \delta\big) = 0. \tag{17.34}$$

Thus, combined with (17.33), Theorem 17.19 shows that as $N \to \infty$

$$\lambda_N \to 2 \quad \text{in probability.}$$

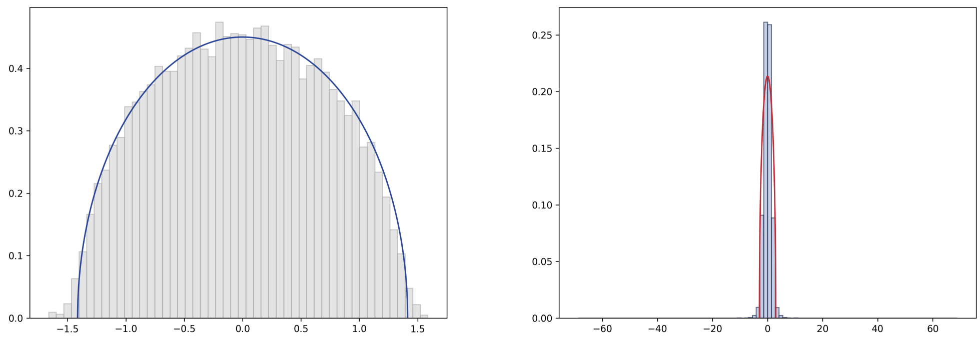
Proof idea for Theorem 17.19. We wish to estimate $\mathbb{P}(\lambda_N > 2 + \delta)$ for fixed $\delta > 0$. Well, for any $k \in \mathbb{N}$,

$$\lambda_N > 2 + \delta \quad \Rightarrow \quad \lambda_N^{2k} > (2+\delta)^{2k} \quad \Rightarrow \quad \mathrm{Tr}(M_N^{2k}) = \sum_{j=1}^{N} \lambda_j^{2k} > (2+\delta)^{2k},$$

and so by Markov's inequality

$$\mathbb{P}\big(\lambda_N > 2 + \delta\big) \le \mathbb{P}\big(\,\mathrm{Tr}(M_N^{2k}) > (2+\delta)^{2k}\big) \le \frac{N}{(2+\delta)^{2k}} \frac{1}{N} \mathbb{E}\Big[\,\mathrm{Tr}\,\big(M_N^{2k}\big)\Big]. \tag{17.35}$$

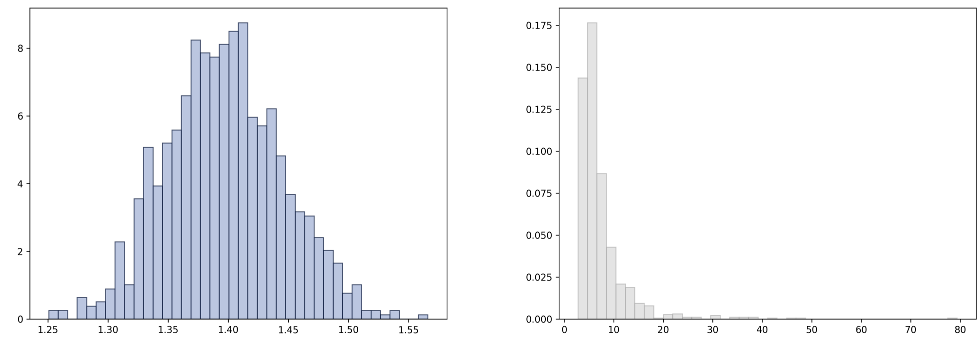
Since this estimate holds for any $k \in \mathbb{N}$, we can choose an appropriate k. The idea (due to Füredi and Komlós) is then to choose $k = k(N)$ as growing with N in a precisely controlled fashion. To that end one rewrites the factor on the far right of (17.35) as in (17.21), afterwards one decomposes the sum over $(G, w) \in \mathcal{G}_{2k}$ further and lastly one has to solve a (tricky!) counting problem. For details, see Anderson-Guionnet-Zeitouni, Section 2.1.6. ◻



(a) Density plot of the eigenvalues of a collection of GOE matrices of size 100.

(b) Density plot of the eigenvalues of a collection of Wigner matrices with heavy tails of size 250.

Figure 17.5.: Eigenvalue density plots for (a) GOE, (b) a Wigner matrix with entries with heavy tails.



(a) Density plot of maximum eigenvalue (unscaled) of a collection of GOE matrices of size 500.

(b) Density plot of maximum eigenvalue (unscaled) of a collection of Wigner matrices with heavy tails of size 500.

Figure 17.6.: Maximum eigenvalue (unscaled) density plots (a) GOE, (b) a Wigner matrix with entries with heavy tails.

18. Stieltjes transforms

In this chapter we shall encounter another method that can be used in the study of the normalized counting measure of eigenvalues of a random matrix. We have used orthogonal polynomials in Chapter 16 that rely on the explicit form of the joint probability law of eigenvalues (not extendable to general Wigner) and we have used combinatorial objects in Chapter 17. Now we shall introduce some purely *analytical* tools to study said eigenvalues.

18.1. Technical preparations

For $f : \mathbb{R} \to \mathbb{C}$ integrable (cf. Definition A.2), its *Fourier transform* is $\hat{f} : \mathbb{R} \to \mathbb{C}$ defined by

$$\hat{f}(x) = \int_{\mathbb{R}} f(y) e^{-2\pi i x y} dy. \tag{18.1}$$

Example 18.1. *Let $f(x) = \exp(-x^2)$. Then its Fourier transform is*

$$\begin{aligned} \hat{f}(x) &= \int_{\mathbb{R}} e^{-y^2 - 2\pi i x y} dy \\ &= \sqrt{\pi} e^{-\pi^2 x^2}, \end{aligned}$$

so for this choice of function, its transform is nearly identical.

Compare this to the *characteristic function* of a random variable Y with density function $\rho : \mathbb{R} \to [0, 1]$:

$$\mathbb{E}[e^{i\lambda Y}] = \int_{\mathbb{R}} \rho(y) e^{i\lambda y} dy. \tag{18.2}$$

Example 18.2. *Take a Gaussian random variable $Y \sim \mathcal{N}(0, 1/2)$. Then its characteristic function is*

$$\begin{aligned} \mathbb{E}[e^{i\lambda Y}] &= \int_{\mathbb{R}} e^{-y^2 + i\lambda y} \frac{dy}{\sqrt{\pi}} \\ &= e^{-\lambda^2/4} \end{aligned}$$

Clearly (18.1) and (18.2) are very similar functions.

Our main tool for this section is, in many ways, an analogue of the Fourier transform/characteristic function.

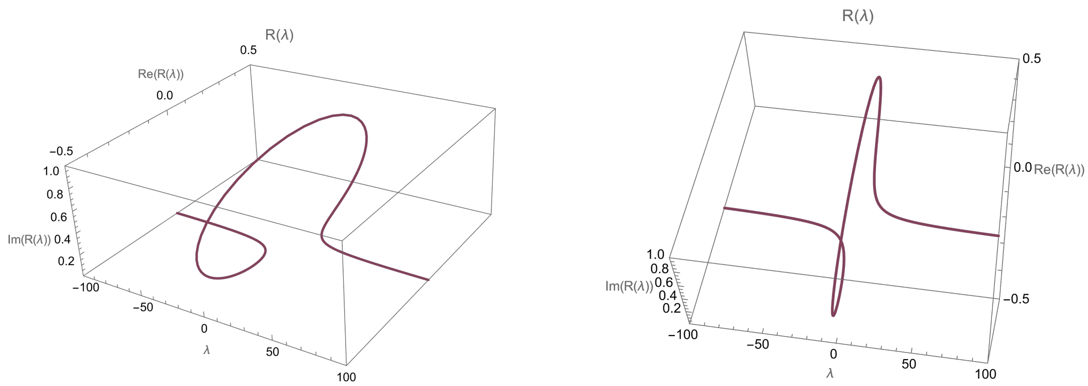


Figure 18.1.: The function $R(\lambda)$ for $z = i$ (from two different angles).

Definition 18.3. *Let $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a nonnegative integrable function. The function*

$$S(z) := \int_{\mathbb{R}} \frac{\rho(\lambda)}{\lambda - z} d\lambda, \tag{18.3}$$

*for z not purely real (i.e. $z \in \mathbb{C} \setminus \mathbb{R}$) is called the **Stieltjes transform** of ρ.*

Note that for fixed $z = x + iy$, with $y \neq 0$, we have the function $R \equiv R_z : \mathbb{R} \to \mathbb{C}$ defined by

$$R(\lambda) = \frac{1}{\lambda - z} = \frac{\lambda - x + iy}{(\lambda - x)^2 + y^2}$$

is a bounded continuous function, see for example Figure 18.1.

Hence, since ρ is integrable, the integral (18.3) exists for any $z \notin \mathbb{R}$. Other properties of interest of $S(z)$ are summarized below.

Proposition 18.4. *Let S be the Stieltjes transform (18.3) of an integrable function $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$. Then we have the following properties:*

1. *S is analytic in $\mathbb{C} \setminus \mathbb{R}$, and $\overline{S(z)} = S(\bar{z})$ on the same domain.*

2. *$\Im S(z) \cdot \Im z > 0$ for $\Im z \neq 0$.*

3. *$|S(z)| \cdot |\Im z| \leq \int_{\mathbb{R}} \rho(\lambda) d\lambda$ for $\Im z \neq 0$, in particular*

$$\lim_{\eta \to +\infty} \eta |S(i\eta)| = \int_{\mathbb{R}} \rho(\lambda) d\lambda.$$

4. *If for some $\lambda \in \mathbb{R}$ there exists the nontangential limit from the upper half-plane,*

$$\Im S_+(\lambda) := \lim_{\varepsilon \downarrow 0} \Im S(\lambda + i\varepsilon),$$

then we have the inversion identity

$$\rho(\lambda) = \frac{1}{\pi} \Im S_+(\lambda). \tag{18.4}$$

Morally speaking, S is a kind of *moment-generating function*: suppose that ρ is compactly-supported, and let

$$m_k(\rho) := \int_{\mathbb{R}} \lambda^k \rho(\lambda) d\lambda,$$

denote its moments, compare (17.6). Then, if $\operatorname{supp}(\rho) \subset [-R, R]$ with $R > 0$, we have $|m_k(\rho)| \le R^k \int_{\mathbb{R}} \rho(\lambda) d\lambda$, and so the generating function

$$z \mapsto \sum_{k=0}^{\infty} m_k(\rho) z^k, \qquad |z| < \frac{1}{R},$$

has a positive radius of convergence. In this case, using the geometric series expansion we have

$$S(z) = \int_{-R}^{R} \frac{\rho(\lambda)}{\lambda - z} d\lambda = -\frac{1}{z} \int_{-R}^{R} \sum_{k=0}^{\infty} \left(\frac{\lambda}{z}\right)^k \rho(\lambda) d\lambda,$$

and as long as $|z| > R$, the sum inside the integral is uniformly convergent, i.e. we may interchange the integral with the sum. What results is

$$S(z) = -\frac{1}{z} \sum_{k=0}^{\infty} \frac{1}{z^k} \int_{-R}^{R} \lambda^k \rho(\lambda) d\lambda = -\frac{1}{z} \sum_{k=0}^{\infty} \frac{m_k(\rho)}{z^k}, \tag{18.5}$$

i.e. a convergent Laurent series expansion of $S(z)$ in a neighborhood of $z = \infty$, in the variable $1/z$, whose coefficients are *almost* the moments of ρ – almost because of the extra minus and $1/z$ term in the right hand side of (18.5). This observation can be useful in calculating Stieltjes transforms.

Moving forward, we will also need the following elementary facts:

Proposition 18.5. *If $X \sim \frac{1}{2}(N(0,1) + iN(0,1))$ denotes the complex Gaussian random variable whose real and imaginary part are independent $N(0, \frac{1}{4})$ copies, then for any differentiable function $\varphi : \mathbb{C}^2 \to \mathbb{C}$ that grows at worst polynomially at infinity with the same property for its derivative,*

$$\mathbb{E}\big[\varphi(X, \overline{X}) X\big] = \mathbb{E}\big[|X|^2\big] \mathbb{E}\left[\frac{\partial}{\partial \bar{z}} \varphi(z, \bar{z})\bigg|_{z=X}\right]. \tag{18.6}$$

Proof. The left hand side in (18.6) equals, after integrating by parts in the complex plane $\mathbb{C}$,

$$\begin{aligned} \mathbb{E}\big[\varphi(X, \overline{X}) X\big] &= \frac{2}{\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(z, \bar{z}) z e^{-2|z|^2} d\Re(z) d\Im(z) \\ &= \frac{1}{2} \frac{2}{\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \left[\frac{\partial}{\partial \bar{z}} \varphi(z, \bar{z})\right] e^{-2|z|^2} d\Re(z) d\Im(z). \end{aligned}$$

However, since $\mathbb{E}[|X|^2] = \frac{1}{2}$, identity (18.6) now readily follows. □

Proposition 18.6. *Let $M \in \mathcal{H}_N$ be a complex Hermitian matrix of size $N \times N$ and let $R_M(z)$ denote its matrix* ***resolvent****, i.e. the mapping*

$$\begin{aligned} R_M : \mathbb{C} \setminus \mathbb{R} &\to \mathbb{C}^{N\times N} \\ z &\mapsto (M - zI)^{-1}, \end{aligned} \tag{18.7}$$

with $I = I_N$ is the $N \times N$ identity matrix. Then we have the following two properties:

1. *For any $A, B \in \mathcal{H}_N$ and $z \notin \mathbb{R}$,*

$$R_B(z) = R_A(z) - R_B(z)(B - A)R_A(z). \tag{18.8}$$

2. *For any $M \in \mathcal{H}_N$ and $z \notin \mathbb{R}$, with $R_M(z) = (R_{j,k}(z))_{j,k=1}^N$,*

$$\|R_M(z)\| \le \frac{1}{|\Im z|}; \qquad |R_{j,k}(z)| \le \frac{1}{|\Im z|}, \qquad j, k = 1, \ldots, N. \tag{18.9}$$

Here, $\|R_M(z)\|$ denotes the operator norm of the $N \times N$ matrix $R_M(z)$, i.e. the largest value vector norm for $R_M(z)$ on the the sphere

$$\begin{aligned} \|R_M(z)\| &:= \sup_{\|x\|=1} \|R_M(z)x\|, \\ \|y\| &:= \sqrt{\sum_{j=1}^N |y_j|^2} \qquad \forall\, y = (y_1, \ldots, y_N) \in \mathbb{C}^N. \end{aligned}$$

Proof. Identity (18.8) follows from matrix algebra by multiplying with $(A - zI)$ from the right and with $(B - zI)$ from the left. For (18.9) one needs to use the spectral theorem for Hermitian matrices and the Cauchy-Schwarz inequality, we leave the details to the reader.

□

Example 18.7. *Let $M = \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}$. Then its matrix resolvent is*

$$R_M(z) = \begin{pmatrix} 1 - z & i \\ -i & 2 - z \end{pmatrix}^{-1} = \frac{1}{z^2 - 3z + 1} \begin{pmatrix} 2 - z & -i \\ i & 1 - z \end{pmatrix}.$$

This inverse always exists since M is Hermitian so it has real eigenvalues, and so if $z \notin \mathbb{R}$ then $M - zI$ must have non-zero determinant and the corresponding matrix must be invertible. Figure 18.2 displays the property (18.9).

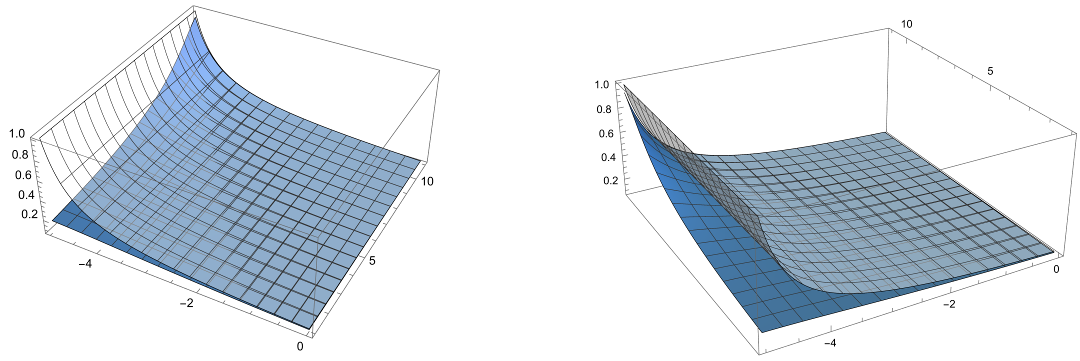


Figure 18.2.: Graphs showing the operator norm of $R_M(z)$ and $1/|\Im(z)|$ for $\Re(z) \in [-5, 0]$ and $\Im(z) \in [1, 11]$, for M in Example 18.7. The blue surface is $||R_M(z)||$ and the gray translucent surface is $1/|\Im(z)|$.

Remark. *On the operator norm: if M is a matrix with eigenvalue λ and corresponding eigenvector x, chosen with the normalization $\|x\| = 1$, then for any eigenvalue of M,*

$$\|\lambda\| = |Mx| \leq \sup_{\|v\|=1} |Mv| =: \|M\|.$$

So the operator norm of a matrix is always at least as large as its maximum (in modulus) eigenvalue (a.k.a. the 'spectral radius'). However, if we can diagonalize M by a unitary matrices U (as we can if M is Hermitian) then also $\|Mv\| = \|\overline{U}^T DUv\| = \|\Lambda Uv\|$ where Λ is the diagonal matrix comprising eigenvalues of M. Hence $\|\Lambda Uv\| \leq |\lambda^*|\|v\|$ where $\lambda*$ is the maximal (positive or negative) eigenvalue, using again that unitary matrices preserve vector length. Thus in the case of Hermitian M, the operator norm equals the spectral norm.

The utility of the Stieltjes transform (18.3) in random matrix theory stems from the following observation which allows us to derive global scaling results: let $\overline{N}_N$ denote the normalized counting measure (15.1) of the eigenvalues $\lambda_1, \ldots, \lambda_N$ of some random matrix $M_N \in \mathcal{H}_N$. Write

$$\overline{N}_N(\Delta) = \frac{1}{N}\sum_{j=1}^{N} \chi_\Delta(\lambda_j) = \int_{\mathbb{R}} \chi_\Delta(x) \left[\frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j}(x)\right] dx, \qquad \Delta \subset \mathbb{R},$$

in terms of the 'delta mass' $\delta_a, a \in \mathbb{R}$ with $\int_{\mathbb{R}} \delta_a(x)dx = 1$ and $\int_{\mathbb{R}} f(x)\delta_a(x)dx = f(a)$ (essentially δ_a puts all mass at a and nothing elsewhere). Then, we can view $\varrho_N = \sum_{j=1}^{N} \delta_{\lambda_j}/N$ as density of the random variable N_n and thus compute its Stieltjes transform,

$$S_N(z) := \int_{\mathbb{R}} \frac{\varrho_N(x)}{x - z} dx = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{\lambda_j - z} \overset{(18.7)}{=} \frac{1}{N} \operatorname{Tr}\left(R_{M_N}(z)\right). \tag{18.10}$$

Note that the last equality in (18.10) follows from the spectral theorem. Since $S_N(z)$ is then essentially a moment generating function for the normalized counting function (cf. (18.5)), we can try to study it an infer information back about the distribution of $\overline{\mathcal{N}}_N$. For example, we can study convergence properties of (the random variable) $S_N(z)$ in some domains of the complex plane, prove a limit theorem, show that the limit is again a Stieltjes transform of some density and ultimately compute the same density via the inversion identity (18.4). We carry out the necessary steps for the GUE in the upcoming section, recovering in a different manner results akin to those of Chapter 16.

18.2. The GUE semicircle law

Our proof of Corollary 18.9 below is based on the following ingredient:

Proposition 18.8. *The Stieltjes transform $S_N(z)$, see (18.10), of the normalized eigenvalue counting measure for the* GUE *has the following properties:*

1. *For all $\Im z \geq 2$,*
$$\lim_{N\to\infty} \mathbb{E}_N\left[S_N(z)\right] = f(z) \tag{18.11}$$
exists, where $f = f(z)$ is the unique solution of the equation
$$f^2 + 2zf + 2 = 0 \tag{18.12}$$
subject to
$$\begin{cases} \Im f(z) \cdot \Im z > 0 \quad \text{for} \quad \Im z \geq 2, \\ \text{and} \qquad \lim\limits_{\eta\to+\infty} \eta |f(i\eta)| = 1. \end{cases} \tag{18.13}$$

2. *For all $\Im z \geq 2$ and $N \in \mathbb{N}$,*
$$\mathrm{Var}_N\left(S_N(z)\right) = \mathbb{E}_N\left[|\gamma_N(z)|^2\right] \leq \frac{1}{24N^2} \tag{18.14}$$
with $\gamma_N(z)$ the centred version of the random variable $S_N(z)$
$$\gamma_N(z) := S_N(z) - \mathbb{E}_N\left[S_N(z)\right].$$

Proof. (Non-examinable). For $p \in \mathbb{N}$ and $z_1, \ldots, z_p \in \mathbb{C} \setminus \mathbb{R}$, define the multivariate moments
$$m_p(z_1, \ldots, z_p) := \mathbb{E}_N\left[\prod_{j=1}^{p} S_N(z_j)\right].$$
We first apply (18.8) for $B = M_N, A = 0$. Since $M_N = (\zeta_{j,k})_{j,k=1}^N$ and $R_{M_N}(z) = (R_{j,k}(z))_{j,k=1}^N$,
$$R_{M_N}(z) = -\frac{1}{z} I + \frac{1}{z} R_{M_N}(z) M_N$$
and hence (writing $\mathbb{1}_{j,k} = 1$ if $j = k$ and 0 otherwise)
$$\mathbb{E}_N\left[R_{j,k}(z)\right] = -\frac{1}{z}\mathbb{1}_{j,k} + \frac{1}{z}\sum_{\ell=1}^{N} \mathbb{E}_N\left[R_{j,\ell}(z)\zeta_{\ell,k}\right]. \tag{18.15}$$

For any fixed $z \notin \mathbb{R}$, we can view the entry $R_{j,k}(z)$ as a complex-valued function of the matrix M_N. To make this dependence clear, in the following we write (for some j, k and still for fixed non-real z) $R_{j,k}(z) \equiv \Phi(M_N)$. We thus need to calculate

$$\mathbb{E}_N\left[\Phi(M_N)\zeta_{\ell,k}\right] = \frac{1}{2N}\begin{cases} \mathbb{E}_N\left[\dfrac{\partial\Phi}{\partial\zeta_{\ell,\ell}}(M_N)\right], & \ell = 1,\ldots,N \\ \mathbb{E}_N\left[\dfrac{\partial\Phi}{\partial\overline{\zeta_{\ell,k}}}(M_N)\right], & 1 \leq \ell < k \leq N\,, \\ \mathbb{E}_N\left[\dfrac{\partial\Phi}{\partial\zeta_{k,\ell}}(M_N)\right], & 1 \leq k < \ell \leq N \end{cases} \tag{18.16}$$

where the equality follows exactly as in the proof of (18.6): integrating by parts while exploiting (16.2) (and the fact that Φ and its derivatives grow at worst polynomially at infinity). On the other hand, by differentiating the identity $R_{M_N}(M_N - zI) = I$ we also find (returning to the entry-notation $R_{j,k}(z)$)

$$\frac{\partial R_{j,k}(z)}{\partial\zeta_{ab}} = -R_{j,a}(z)R_{b,k}(z) \tag{18.17}$$

for $a, b, j, k = 1,\ldots,N$. Thus (18.16), (18.17) yield together

$$\mathbb{E}_N\left[R_{j,\ell}(z)\zeta_{\ell,k}\right] = -\frac{1}{2N}\mathbb{E}_N\left[R_{j,k}(z)R_{\ell,\ell}(z)\right]$$

for all $j, k, \ell = 1,\ldots,N$. Consequently, upon return to (18.15),

$$\begin{aligned} \mathbb{E}_N\left[R_{j,k}(z)\right] &= -\frac{1}{z}\mathbb{1}_{j,k} - \frac{1}{2Nz}\sum_{\ell=1}^{N}\mathbb{E}_N\left[R_{j,k}(z)R_{\ell,\ell}(z)\right] \\ &\overset{(18.10)}{=} -\frac{1}{z}\mathbb{1}_{j,k} - \frac{1}{2z}\mathbb{E}_N\left[R_{j,k}(z)S_N(z)\right], \end{aligned} \tag{18.18}$$

and so, taking $j = k$ in (18.18) and averaging over all j afterwards, we find a relationship between the first and second moments:

$$m_1(z) = \mathbb{E}_N\left[S_N(z)\right] = \mathbb{E}_N\left[\frac{1}{N}\sum_{j=1}^{N}R_{j,j}(z)\right] = -\frac{1}{z} - \frac{1}{2z}m_2(z,z). \tag{18.19}$$

Equality (18.19) constitutes our *first main identity* for the multivariate moments $m_p(z_1,\ldots,z_p)$. A second identity can be found by returning to the first equality in (18.15), by taking matrix entries in it and multiplying through with $R_{a,b}(w), w \notin \mathbb{R}$, afterwards evaluating the expectation,

$$\mathbb{E}_N\left[R_{j,k}(z)R_{a,b}(w)\right] = -\frac{1}{z}\mathbb{1}_{j,k}\mathbb{E}_N\left[R_{a,b}(w)\right] + \frac{1}{z}\sum_{\ell=1}^{N}\mathbb{E}_N\left[R_{j,\ell}(z)R_{a,b}(w)\zeta_{\ell,k}\right]$$

for $a, b, j, k = 1, \ldots, N$. Here, $R_{j,\ell}(z)R_{a,b}(w)$ is another function of the matrix M_N, so we can apply (18.16) once more and since (18.17) yields by product rule,

$$\frac{\partial}{\partial \zeta_{c,d}}\big(R_{j,k}(z)R_{a,b}(w)\big) = -R_{j,c}(z)R_{d,k}(z)R_{a,b}(w) - R_{j,k}(z)R_{a,c}(w)R_{d,b}(w),$$

we obtain all together

$$\begin{aligned}\mathbb{E}_N\left[R_{j,k}(z)R_{a,b}(w)\right] = -\frac{1}{z}\mathbb{1}_{j,k}\mathbb{E}_N\left[R_{a,b}(w)\right] - \frac{1}{2z}\mathbb{E}_N\left[R_{j,k}(z)S_N(z)R_{a,b}(w)\right] \\ - \frac{1}{2Nz}\sum_{\ell=1}^{N}\mathbb{E}_N\left[R_{j,\ell}(z)R_{a,k}(w)R_{\ell,b}(w)\right].\end{aligned}$$

Taking $a = b, j = k$ in the last equality and afterwards averaging over a and j finally results in

$$\begin{aligned}m_2(z_1, z_2) &= \mathbb{E}_N\left[S_N(z_1)S_N(z_2)\right] \\ &= \mathbb{E}_N\left[\frac{1}{N}\sum_{j=1}^{N}R_{j,j}(z_1)\frac{1}{N}\sum_{a=1}^{N}R_{a,a}(z_2)\right] \\ &= -\frac{1}{z_1}m_1(z_2) - \frac{1}{2z_1}m_3(z_1, z_1, z_2) - \frac{1}{2N^2 z_1}\mathbb{E}_N\left[\frac{1}{N}\operatorname{Tr}\left(R_{M_N}(z_1)R^2_{M_N}(z_2)\right)\right],\end{aligned} \tag{18.20}$$

the *second main identity*. One can, in principle, derive an infinite system of such identities, but we will work with (18.19) and (18.20) only. To that end, take $z_1 = z$ and $z_2 = \bar{z}$ and recall Proposition 18.4 en route. We find

$$\mathbb{E}_N\left[S_N(z)\right] = -\frac{1}{z} - \frac{1}{2z}\mathbb{E}_N\left[S_N^2(z)\right], \tag{18.21}$$

$$\mathbb{E}_N\left[|S_N(z)|^2\right] = -\frac{1}{z}\mathbb{E}_N\left[\overline{S_N(z)}\right] - \frac{1}{2z}\mathbb{E}_N\left[S_N^2(z)\overline{S_N(z)}\right] + r_N(z, \bar{z}). \tag{18.22}$$

Here,

$$r_N(z, \bar{z}) \coloneqq -\frac{1}{2N^2 z}\mathbb{E}_N\left[\frac{1}{N}\operatorname{Tr}\left(R_{M_N}(z)R^2_{M_N}(\bar{z})\right)\right].$$

Using (18.21) to replace $-1/z$ in the first term in the right hand side of (18.22), we then obtain

$$\mathbb{E}_N\left[|\gamma_N(z)|^2\right] = -\frac{1}{2z}\mathbb{E}_N\left[S_N^2(z)\overline{\gamma_N(z)}\right] + r_N(z, \bar{z}), \tag{18.23}$$

where, exploiting $S_N(z) = \gamma_N(z) + \mathbb{E}_N[S_N(z)]$ and the fact that $\gamma_N(z)$ is centered,

$$\begin{aligned}\mathbb{E}_N\left[S_N^2(z)\overline{\gamma_N(z)}\right] &= \mathbb{E}_N\left[S_N(z)|\gamma_N(z)|^2\right] + \mathbb{E}_N\left[S_N(z)\right]\mathbb{E}_N\left[S_N(z)\overline{\gamma_N(z)}\right] \\ &= \mathbb{E}_N\left[S_N(z)|\gamma_N(z)|^2\right] + \mathbb{E}_N\left[S_N(z)\right]\mathbb{E}_N\left[|\gamma_N(z)|^2\right].\end{aligned} \tag{18.24}$$

It now remains to insert (18.24) for $\mathbb{E}_N[S_N^2(z)\overline{\gamma_N(z)}]$ in (18.23) and estimate by triangle inequality,

$$\mathbb{E}_N\big[|\gamma_N(z)|^2\big] \leq \frac{1}{2|\Im z|}\bigg[\mathbb{E}_N\big[|S_N(z)|\,|\gamma_N(z)|^2\big] + \mathbb{E}_N\big[|S_N(z)|\big]\mathbb{E}_N\big[|\gamma_N(z)|^2\big]\bigg] + \big|r_n(z,\bar{z})\big|.$$

However, for any $N\times N$ matrix P, we have $|\operatorname{Tr}(P)| \leq N\|P\|$ with the operator norm of P on the far right (see the remark following Proposition 18.6), and so by (18.10) and (18.9), $|S_N(z)| \leq \|R_{M_N}(z)\| \leq 1/|\Im z|$, leading to

$$\left(1-\frac{1}{|\Im z|^2}\right)\mathbb{E}_N\big[|\gamma_N(z)|^2\big] \leq \big|r_N(z,\bar{z})\big| \leq \frac{1}{2N^2|\Im z|^4}. \tag{18.25}$$

Once $\Im z \geq 2$, then (18.25) yields precisely (18.14) and we now use the same to derive (18.11) and (18.12): return to (18.21) and rewrite the same as

$$\mathbb{E}_N\big[S_N(z)\big] = -\frac{1}{z} - \frac{1}{2z}\Big(\mathbb{E}_N\big[S_N(z)\big]\Big)^2 - \frac{1}{2z}\mathbb{E}_N\big[\gamma_N^2(z)\big],$$

so that by (18.14),

$$\left|\mathbb{E}_N\big[S_N(z)\big] + \frac{1}{z} + \frac{1}{2z}\Big(\mathbb{E}_N\big[S_N(z)\big]\Big)^2\right| \leq \frac{1}{96N^2} \qquad \forall\,\Im z \geq 2,\ N\in\mathbb{N}. \tag{18.26}$$

Since $|\mathbb{E}_N[S_N(z)]| \leq 1/|\Im z|$, the sequence $(\mathbb{E}_N[S_N(z)])_{N=1}^\infty$ consists of functions that are analytic and uniformly bounded in N and z, so long $\Im z \geq 2$. Hence, there exists an analytic in $\{z\in\mathbb{C}:\ \Im z \geq 2\}$ function f and a subsequence $(\mathbb{E}_{N_m}[S_{N_m}(z)])_{m=1}^\infty$ that converges to f uniformly on any compact set of $\{z\in\mathbb{C}:\ \Im z\geq 2\}$. From (18.26), we can see that the limiting f, after passing to the aforementioned subsequence in (18.26), will solve the quadratic equation written in (18.12). But the same equation has only one solution

$$f(z) = (z^2-2)^{\frac{1}{2}} - z, \qquad z\in\mathbb{C}\setminus[-\sqrt{2},\sqrt{2}], \tag{18.27}$$

with the indicated two properties (given that the complex square root is chosen with its principal branch). This completes our proof of the Proposition. □

We are now prepared to state the Wigner semicircle law for the GUE in the following version, which is slightly different from (16.20) and (17.3):

Corollary 18.9. *Consider the* GUE *and let* $\overline{N}_N$ *be the normalized counting measure* (15.1) *of its eigenvalues* (15.2)*. Then for* $\sigma(x)=\sigma_{1/2}(x)$ *the semicircle density with* $t=1/2$ *(cf.* (17.2)*),*

$$\mathbb{P}_N\Big(\lim_{N\to\infty}\overline{N}_N(\Delta) = \int_\Delta \sigma(x)dx\Big) = 1 \tag{18.28}$$

for any interval $\Delta\subset\mathbb{R}$.

Proof. (Non-examinable). We require a few more details about Stieltjes transforms than what is summarized in Proposition 18.4, those can be found for example in Pastur-Shcherbina, Proposition 2.1.2 and Theorem 2.2.11, with pointers to their proofs in the same source. Namely, the solution $z \mapsto f(z)$ of (18.12), see (18.27), is actually analytic for $\Im z \neq 0$, not just for $\Im z \geq 2$, and it satisfies

$$\Im f(z) \cdot \Im z > 0 \quad \text{for} \quad \Im z \neq 0 \qquad \text{as well as} \qquad \lim_{\eta \to +\infty} \eta |f(i\eta)| = 1.$$

Moreover, $\overline{f(z)} = f(\bar{z})$ for $\Im z \neq 0$. As such, and since the (nontangential) limit

$$\Im f_+(x) = \lim_{\varepsilon \downarrow 0} \Im f(x + i\varepsilon) \overset{(18.27)}{=} \sqrt{(2 - x^2)_+}\,,$$

exists for all $x \in \mathbb{R} \setminus \{\pm\sqrt{2}\}$, the function (18.27) can itself be written as Stieltjes transform,

$$f(z) = \int_{\mathbb{R}} \frac{\rho(x)dx}{x - z}, \qquad z \in \mathbb{C} \setminus [-\sqrt{2}, \sqrt{2}],$$

with $\rho(x) = \frac{1}{\pi} \Im f_+(x)$, as in (18.28), compare also (18.4). What we are exploiting here is the fact that any function S satisfying the first three properties in Proposition 18.4 and the fourth one for almost all $\lambda \in \mathbb{R}$ is of type (18.3) with the underlying ρ coming from (18.4). Moving forward, by using Chebyshev's inequality (A.4) and the variance bound (18.14), we have for any $\varepsilon > 0$ that

$$\mathbb{P}_N\big(|S_N(z) - \mathbb{E}_N[S_N(z)]| > \varepsilon\big) \leq \frac{1}{\varepsilon^2} \operatorname{Var}_N\big(S_N(z)\big) \overset{(18.14)}{\leq} \frac{1}{24\varepsilon^2 N^2},$$

valid for all $\Im z \geq 2$ and all $N \in \mathbb{N}$. Hence, the series

$$\sum_{N=1}^{\infty} \mathbb{P}_N\big(|S_N(z) - \mathbb{E}_N[S_N(z)]| > \varepsilon\big)$$

converges for $\Im z \geq 2$, i.e. by the Borel-Cantelli lemma, see Theorem A.11, with probability one,

$$\lim_{N \to \infty} S_N(z) = f(z) \qquad \overset{(18.10)}{\Leftrightarrow} \qquad \lim_{N \to \infty} \int_{\mathbb{R}} \frac{\varrho_N(x)dx}{x - z} = \int_{\mathbb{R}} \frac{\rho(x)dx}{x - z}, \tag{18.29}$$

for any fixed $z \in \mathbb{C} : \Im z \geq 2$, in fact uniformly in z chosen from any compact set of the domain $\Im z \geq 2$ by the analytic properties of S_N and f. The uniform convergence (18.29) of the Stieltjes transform on compact sets in $\Im z \geq 2$, with probability one, is sufficient to conclude the limit law (18.28) for the normalized counting measure, also with probability one. For details of this last conclusion we refer the dedicated reader to Pastur-Shcherbina, Proposition 2.1.2, part (vi). This concludes our proof. □

A. Appendix

A.1. Notation and Terminology

Here is a reminder of the notation used throughout the notes. We write

$$\mathbb{R} = \text{real numbers} \qquad \mathbb{Q} = \text{rationals} \qquad \mathbb{Z} = \text{integers}$$

$$\mathbb{C} = \text{complex numbers} = \{x + \mathrm{i}y : x, y \in \mathbb{R}\}$$

with their sums and products. For $z = x + \mathrm{i}y \in \mathbb{C}$, we write $\Re z = x$, $\Im z = y$, $|z| = \sqrt{x^2 + y^2}$. We also use

$$\mathbb{N} = \text{natural numbers} = \{1, 2, 3, \ldots\}.$$

Given two sets A and B, a function, f, from A into B is denoted by $f : A \to B$, or $x \mapsto f(x)$ where for each $x \in A$, we denote by $f(x)$ the member of B to which x is assigned. For two functions f and g on a region Ω and $z_0 \in \Omega$, we write

$$f = O(g) \quad \text{at } z_0$$

to mean that $f(z)/g(z)$ is bounded for z near z_0 and $z \in \Omega$, and

$$f = o(g) \quad \text{at } z_0$$

if $f(z)/g(z) \to 0$ as $z \to z_0$ and $z \in \Omega$. Lastly, for a nonempty set E of real numbers we use

$$\inf, \qquad \text{resp.} \qquad \sup,$$

to denote its greatest lower, resp. least upper, bound.

A.2. Mathematical toolkit

The following is a collection, grouped in to themes, of useful results used throughout these notes. You will (hopefully!) have encountered them during your mathematical training to date, possibly during the first part of the course.

A.2.1. Analysis toolkit

Theorem A.1 (Weierstrass approximation)**.** *Let $f : [a, b] \to \mathbb{C}$ be a continuous function on a closed, bounded interval $[a, b] \subset \mathbb{R}$. Then for each $\varepsilon > 0$, there is a polynomial p for which*

$$|f(x) - p(x)| < \varepsilon \qquad \text{for all } x \in [a, b].$$

Definition A.2. *Let $f : \mathbb{R} \to \mathbb{C}$ denote a continuous function and $1 \leq p < \infty$. If*

$$\|f\|_p := \left(\int_{\mathbb{R}} |f(x)|^p dx \right)^{\frac{1}{p}} < \infty$$

*then we say that f is **p-integrable** (so $f \in \mathcal{L}^p(\mathbb{R})$). The statement that f is **integrable** is short-hand for $p = 1$-**integrable**, i.e. the integral over $\mathbb{R}$ of the absolute value of f is finite.*

Theorem A.3 (Hölder inequality)**.** *Let $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. If f is p-integrable and g is q-integrable, then fg is integrable (i.e. $fg \in \mathcal{L}^1(\mathbb{R})$) and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{A.1}$$

A.2.2. (Random) matrix toolkit

Lemma A.4 (Determinant expansion)**.** *Let A, B be $N \times N$ matrices. Then*

$$det(A + B) = \sum_{r=0}^{N} \sum_{\alpha,\beta} (-1)^{\sum \alpha_j + \sum \beta_j} \det(A_{(\alpha,\beta)}) \det(B_{(\alpha^c,\beta^c)})$$

where the inner sum is over $\alpha = (\alpha_1, \ldots, \alpha_r)$, $\beta = (\beta_1, \ldots, \beta_r)$, all possible length r subsets of $\{1, \ldots, N\}$ with $\alpha_j < \alpha_{j+1}$ and similarly for β. The matrix $A_{(\alpha,\beta)}$ is the $r \times r$ submatrix of A lying at the intersection of rows α_j, and columns β_j. The matrix $B_{(\alpha^c,\beta^c)})$ is the $(N - r) \times (N - r)$ submatrix of B lying in the intersection of the rows complementary to α and similarly for the columns.

Proof. This is a generalisation of Laplace's expansion, using multilinearity of the determinant. □

Theorem A.5 (Spectral Mapping Theorem[1])**.** *Let $A \in M_N(\mathbb{C})$ (i.e. an $N \times N$ matrix with complex entries). Let $p(x) \in \mathbb{C}[x]$ be any polynomial. Then the spectrum of the polynomial operator $p(A)$ is precisely the image of the spectrum under p,*

$$\sigma(p(A)) = p(\sigma(A)).$$

Example A.6. *Let $A = \begin{pmatrix} 5 & 6 \\ 2 & 1 \end{pmatrix}$, so $\sigma(A) = \{-1, 7\}$. Set $p(x) = x^2 + 3$. Then*

$$p(A) = A^2 + 3 \cdot I = \begin{pmatrix} 37 & 36 \\ 12 & 13 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 40 & 36 \\ 12 & 16 \end{pmatrix}$$

so $\sigma(p(A)) = \{4, 52\}$. Clearly $p(\sigma(A)) = \{(-1)^2 + 3, 7^2 + 3\} = \{4, 52\}$.

[1] This is a simplified, specialised case of a more general theorem. In particular it applies also to test functions.

Theorem A.7 (Hadamard's inequality)**.** *Suppose $C = [C_{jk}]_{j,k=1}^n$ is an arbitrary $n \times n$ matrix with complex entries. We have*

$$|\det C| \le \prod_{k=1}^{n} \sqrt{\sum_{k=1}^{n} |C_{jk}|^2},$$

and so in particular, for any positive definite $n \times n$ matrix $A = [A_{jk}]_{j,k=1}^n$,

$$\det A \le \prod_{j=1}^{n} A_{jj}. \tag{A.2}$$

Theorem A.8 (Andréief's identity)**.** *Let $\{f_j\}_{j=1}^n$ and $\{g_j\}_{j=1}^n$ be two collections of square integrable functions on $\mathbb{R}$. Then*

$$\int_{\mathbb{R}^n} \det\left[f_j(x_k)\right]_{j,k=1}^n \det\left[g_j(x_k)\right]_{j,k=1}^n \prod_{\ell=1}^{n} dx_\ell = n!\,\det\left[\int_{\mathbb{R}} f_j(x) g_k(x) dx\right]_{j,k=1}^n. \tag{A.3}$$

A.2.3. Probability toolkit

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. Unless specifically stated otherwise the following statements concern random variables $X : \Omega \to \mathbb{R}$ on this space, measurable with respect to $\mathcal{F}$.

Firstly we recall the different types of convergence of random variables.

Definition A.9. *Let $X_1, \ldots, X_n$ have distribution functions $F_1, \ldots, F_n$. The sequence $X_1, X_2, \ldots$* ***converges in distribution*** *(also 'converges in law') to a random variable X with distribution function F if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ where F is continuous. We sometimes will write $X_n \xrightarrow{d} X$ or perhaps $X_n \xrightarrow{l} X$. Think: the finer our random experiment is, the closer it will approximate the limiting distribution. E.g. taking more and more samples in a histogram will improve the resemblence to a bell-curve if that is the limiting distribution.

The sequence $X_1, X_2, \ldots$ ***converges in probability*** *to X if, for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

We write $X_n \xrightarrow{p} X$. Think: the likelihood of a X_n differing by some fixed value from X becomes less and less likely through the seqeunce. This is stronger information than just the overall shape of the histogram.

The sequence $X_1, X_2, \ldots$ ***converges almost surely*** *to X if*

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

We write $X_n \xrightarrow{a.s.} X$. Think: this tells us that the places where X_n differs from X in the limit have probability zero. This is most like the 'pointwise' convergence in analysis as it is equivalent to asking $\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

Remark. *We have the hierarchy*

$$X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{p} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X.$$

The implication directions are strict. For a simple counter-example to the claim $X_n \xrightarrow{d} X \implies X_n \xrightarrow{p} X$*, consider the following. Let* $X, X_1, X_2, \ldots \sim \text{Bernoulli}(1/2)$ *be i.i.d. random variables. Clearly* $X_n \xrightarrow{d} X$*, but we have that the random variable* $Y \coloneqq |X_n - X|$ *is* 0 *with probability* $1/2$ *and* 1 *with probability* $1/2$*, so again* $Y \sim \text{Bernoulli}(1/2)$*. Therefore* $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(Y > \varepsilon) = 1/2$ *for say* $\varepsilon \in (0,1)$*, and we cannot have convergence in probability.*

Theorem A.10 (Markov's and Chebyshev's inequality)**.** *If* $X : \Omega \to \mathbb{R}$ *is a scalar random variable with finite* k*th moment,* $\mathbb{E}\{|X|^k\} < \infty$*. Then, for any* $\varepsilon > 0$*,*

$$\mathbb{P}\big\{|X| \geq \varepsilon\big) \leq \mathbb{E}\big\{|X|^k\big)\varepsilon^{-k},$$

and so in particular, for a scalar random variable $X : \Omega \to \mathbb{R}$ *with finite second moment,*

$$\mathbb{P}\big\{|X - \mathbb{E}\{X\}| \geq \varepsilon\big) \leq \frac{\text{Var}\{X\}}{\varepsilon^2}, \qquad \textit{Chebyshev} \tag{A.4}$$

and for one with finite first moment,

$$\mathbb{P}\big\{|X| \geq \varepsilon\big) \leq \frac{\mathbb{E}\{|X|\}}{\varepsilon}, \qquad \textit{Markov.} \tag{A.5}$$

Theorem A.11 (Borel-Cantelli)**.** *Let* $(E_n)_{n=1}^{\infty} \subset \mathcal{F}$ *be a sequence of events such that* $\sum_{n=1}^{\infty} \mathbb{P}\{E_n\} < \infty$*. Then, with probability one, at most finitely many of the events* E_n *occur at once.*

Theorem A.12 ((Weak) Law of large numbers)**.** *Let* $X_1, \ldots, X_n$ *be i.i.d. random variables with finite mean. Write* $\overline{X}_n$ *for the sample mean* $(1/n)\sum_j X_j$*. Then for any* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mathbb{E}[X_1]| \geq \varepsilon) = 0.$$

Remark. *This statement says that for large enough* n*, the likelihood that the sample average of the random variables differs from the mean of the random variable is vanishingly small. Another way of writing this is that the sample mean* converges in probability *to the mean of* X_1*.*

Theorem A.13 ((Strong) Law of large numbers)**.** *Let* $X_1, \ldots, X_n$ *be i.i.d. random variables with finite mean. Write* $\overline{X}_n$ *for the sample mean* $(1/n)\sum_j X_j$*. Then*

$$\mathbb{P}(\lim_{n\to\infty} \overline{X}_n = \mathbb{E}[X_1]) = 1.$$

Remark. *This statement, stronger than the weak LLN, says that for large enough* n*, the likelihood that the sample mean will converge to the mean of the summand is* 1*. Another way of writing this is that the sample mean* converges almost surely *to the mean of* X_1*. This is* stronger *than the WLLN in the same sense that almost sure convergence implies convergence in probability, but not vice versa.*

One might ask: what is the point of the WLLN if we have the SLLN? Indeed, if this was the extent of the story then the weak law would merely be a corollary of the strong law. However, we can state a stronger version of the strong law for independent (but not identically distributed) random variables:

Kolmogorov's SLLN: *If* $X_1, X_2, \ldots$ *is a sequence of independent random variables with finite means* $\mu_1, \mu_2, \ldots$ *and variances* $\sigma_1^2, \sigma_2^2, \ldots$ *such that*

$$\sum_{k\geq 1} \frac{\sigma_k^2}{k^2} < \infty.$$

Then

$$\mathbb{P}\left(\lim_{n\to\infty}\left(\overline{X}_n - \frac{1}{n}\sum_{k\geq 1}\mu_k\right) = 0\right) = 1.$$

That is, the strong law of large numbers holds for $(X_k)_{k\geq 1}$*.*

There are certainly sequences of independent random variables $(X_k)_{k\geq 1}$ *with finite means that satisfy the weak law but* not *the strong law.*

Theorem A.14 (Lindeberg Central Limit Theorem)**.** *Let* $X_1, \ldots, X_n$ *be independent random variables with finite first and second moment. Set* $\mathbb{E}[X_j] = \mu_j$*,* $\mathrm{Var}[X_j] = \sigma_j^2$ *and* $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$*. If for all* $\varepsilon > 0$ *the random variables* $X_1, \ldots, X_n$ *are such that*

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{j=1}^{n}\mathbb{E}\big[(X_j-\mu_j)^2\mathbb{1}\{|X_j-\mu_j| \geq \varepsilon s_n\}\big] = 0, \tag{A.6}$$

then for any $x \in \mathbb{R}$

$$\mathbb{P}\left(\frac{1}{s_n}\sum_{j=1}^{n}(X_j-\mu_j) \leq x\right) \to_{n\to\infty} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du, \tag{A.7}$$

i.e. $\frac{1}{s_n}\sum_{j=1}^{n}(X_j-\mu_j)$ *converges in distribution (law) to a standard Gaussian.*

Remark. *The condition* (A.6) *is called Lindeberg's condition, and is sufficient but not necessary for the central limit theorem* (A.7) *to hold.*

Subject Index