CHAPTER 1

Introduction

1.1. Objectives and Problems

1.1.1. Ensembles. As was indicated in the Preface, we will deal mostly with spectral aspects of random matrix theory. One of the main subjects of this part of the theory is the large-n asymptotic form of various spectral characteristics of $n \times n$ matrices, whose probability distribution is given in terms of the matrix elements. In other words, the goal of the theory is to "transfer" the probabilistic information from matrix elements to eigenvalues and eigenvectors. Formulated in so general a form, the goal of random matrix theory is similar to that of random operator theory (see e.g. **[396]**), in particular the spectral theory of Schrödinger operators with random potential. However, in the latter the emphasis is put on the analysis of spectral types (pure point, absolutely continuous, etc.), i.e., in fact, on the spatial behavior of eigenfunctions (solutions of corresponding differential or finite-difference equations), while in the former we are mainly interested in the asymptotic behavior of eigenvalues as $n \to \infty$, although statistical properties of eigenvectors are also of considerable interest for a number of applications.

The goal of the theory, seen from the point of view of an analyst, is the study of integrals of the form

$$\int_{\mathcal{E}_n} \Phi_n(M_n) \mathbf{P}_n(dM_n),$$

where

- $\mathcal{E}_n$ is a set of $n \times n$ matrices, for instance
  - real symmetric $\mathcal{S}_n$,
  - hermitian $\mathcal{H}_n$,
  - unitary $\mathcal{U}_n$, etc.;
- Φ_n is a function from $\mathcal{E}_n$ to $\mathbb{R}$ or $\mathbb{C}$, which is often orthogonal or unitary invariant. For example, in the case of $\mathcal{S}_n$,

$$\Phi_n(O_n M_n O_n^T) = \Phi_n(M_n), \ \forall O_n \in O(n);$$

- $\mathbf{P}_n$ is a probability measure on $\mathcal{E}_n$.

One is often interested in the asymptotic behavior of integrals as the size n of matrices tends to infinity.

From the probabilistic point of view one is interested in the asymptotic properties of random variables of the form $\Phi_n(M_n)$, defined on the probability space $(\mathcal{E}_n, \mathbf{P}_n)$ and invariant in the above sense.

We will often call the sequence $\{(\mathcal{E}_n, \mathbf{P}_n)\}_n$ the *random matrix ensemble* or simply a random matrix.

Note that to simplify the notation we do not write in what follows the subindex n in matrices, probability measures, and corresponding expectations if this is clear from the context.

EXAMPLE 1.1.1. Consider the space $\mathcal{S}_n$ of the real symmetric matrices and define the measure

$$d_1 M = \prod_{j=1}^{n} dM_{jj} \prod_{1\le j<k\le n} dM_{jk}. \tag{1.1.1}$$

It can be viewed as an analogue of the Lebesgue measure for $\mathcal{S}_n$.

For the space $\mathcal{H}_n$ of the hermitian matrices this analog is

$$d_2 M = \prod_{j=1}^{n} dM_{jj} \prod_{1\le j<k\le n} d\Re M_{jk} d\Im M_{jk}. \tag{1.1.2}$$

We define the invariant Gaussian laws on $\mathcal{S}_n$ and $\mathcal{H}_n$ as

$$\mathbf{P}_{\beta,n}(d_\beta M) = Z_{\beta,n}^{-1} \exp\left(-n\beta \mathrm{Tr}\, M^2/4w^2\right) d_\beta M, \quad \beta = 1, 2, \tag{1.1.3}$$

where $Z_{\beta,n}$ is a normalizing constant. The case $\beta = 1$ of real symmetric matrices is known as the *Gaussian Orthogonal Ensemble* (GOE); the case $\beta = 2$ of hermitian matrices is known as the *Gaussian Unitary Ensemble* (GUE).

Introduce the matrix $W = \{W_{jk}\}_{j,k=1}^n$ by the relation

$$M = n^{-1/2} W. \tag{1.1.4}$$

Then in the case of the GOE the matrix W is distributed according to the law

$$\widetilde{\mathbf{P}}_{1,n}(d_1 W) = \prod_{1\le j\le k\le n} F_{jk}(dW_{jk}), \tag{1.1.5}$$

where

$$\begin{aligned} F_{jk}(dW) &= (2\pi\sigma_{jk}^2)^{-1/2} \exp\left\{-W^2 \diagup 2\sigma_{jk}^2\right\} dW, \\ \sigma_{jk}^2 : &= \mathbf{E}\{W_{jk}^2\} = (1+\delta_{jk}) w^2. \end{aligned} \tag{1.1.6}$$

Consider now the set $\mathcal{S}_\infty$ (in fact $\mathbb{R}^\infty$) of all infinite real symmetric matrices

$$\{W_{jk}, W_{jk} = W_{kj} \in \mathbb{R}\}_{j,k=1}^{\infty}$$

and define the product measure on $\mathcal{S}_\infty$:

$$\widetilde{\mathbf{P}}_{1,\infty}(d_1 W) = \prod_{1\le j\le k\le \infty} F_{jk}(dW_{jk}), \tag{1.1.7}$$

where F_{jk} is defined in (1.1.6). We obtain an infinite-dimensional probability space $(\mathcal{S}_\infty, \widetilde{\mathbf{P}}_{1,\infty})$, whose important property is that the GOE's, corresponding to all $1 \le n < \infty$, are simultaneously defined on $(\mathcal{S}_\infty, \widetilde{\mathbf{P}}_{1,\infty})$. This will allow us to formulate a number of asymptotic properties of the eigenvalue distribution of the GOE as valid with probability 1 in $(\mathcal{S}_\infty, \widetilde{\mathbf{P}}_{1,\infty})$.

The same construction with evident changes is valid for the GUE. We will denote the corresponding probability space $(\mathcal{H}_\infty, \widetilde{\mathbf{P}}_{2,\infty})$.

REMARK 1.1.2. Random matrix theory deals with one more invariant Gaussian Ensemble, the *Gaussian Symplectic Ensemble* (GSE), corresponding to $\beta = 4$. We will not discuss this ensemble, referring the reader to the works **[217, 356, 484]** and to references therein.

The terms *orthogonal* and *unitary* indicate the invariance of the density of the corresponding measure with respect to all orthogonal $M \to OMO^T$, $\forall O \in O(n)$ and all unitary $M \to UMU^*$, $\forall U \in U(n)$ transformations, respectively. It can also be shown (see **[370]** and Problem 1.3.5) that the measures $d_\beta M$, $\beta = 1, 2$, are also invariant with respect to the same transformations. Hence, the measure (1.1.3) is orthogonal ($\beta = 1$) and unitary ($\beta = 2$) invariant. One can also say that the terms refer to the subgroup of unitary matrices that diagonalize the matrices of the corresponding ensemble.

The invariance arguments play an important role in random matrix theory and its applications since the early 1960s when they were introduced by Dyson **[181, 183, 356]** to determine the basic Gaussian Ensembles: GOE ($\beta = 1$), GUE ($\beta = 2$), and GSE ($\beta = 4$), and their circular analogs (COE, CUE, and CSE; see Section 9.2). A rather complete form of these arguments can be found in **[217, 302, 356, 526]**; see also Theorem 1.1.3 and Sections 9.1.1 and 9.2.1.

Besides, since for $M = \{M_{jk}\}_{j,k=1}^n$

$$\operatorname{Tr} M^2 = \sum_{1 \le j \le n} M_{jj}^2 + 2 \sum_{1 \le j < k \le n} M_{jk}^2$$

in the case of real symmetric matrices and

$$\operatorname{Tr} M^2 = \sum_{1 \le j \le n} M_{jj}^2 + 2 \sum_{1 \le j < k \le n} (\Re M_{jk})^2 + 2 \sum_{1 \le j < k \le n} (\Im M_{jk})^2$$

in the case of hermitian matrices, we conclude that for the Gaussian Ensembles the functionally independent entries of corresponding matrices are statistically independent.

The following theorem shows that these two properties determine uniquely the corresponding ensembles. The theorem is an analog of the well-known Maxwell theorem of statistical physics for random vectors. We will formulate and prove the theorem in the technically simpler case of real symmetric matrices. The proof for hermitian matrices is analogous.

THEOREM 1.1.3. *Consider an ensemble of $n \times n$ real symmetric matrices. Then it has the form*

$$\mathbf{P}(d_1 M) = Z^{-1} \exp\{-\operatorname{Tr}(M - \alpha I)^2/\gamma\} d_1 M, \quad \alpha \in \mathbb{R},\ \gamma > 0, \tag{1.1.8}$$

if and only if the ensemble possesses the following properties:

(i) *it is orthogonal invariant: for any Borel set $\mathcal{A} \in \mathcal{S}_n$*

$$\mathbf{P}(\mathcal{A}) = \mathbf{P}(T_O^{-1}\mathcal{A}),\ \forall O \in O(n),$$

where T_O is the automorphism of $\mathcal{S}_n$ defined as $M \to OMO^T$;

(ii) *functionally independent matrix elements are statistically independent in any orthonormal basis, i.e.,*

$$\mathbf{P}(d_1 M) = \prod_{1 \le j \le k \le n} F_{jk}(dM_{jk}),$$

where F_{jk}, $1 \le j \le k \le n$, are probability measures on $\mathbb{R}$.

PROOF. It is evident that properties (i) and (ii) are necessary for (1.1.8). Let us prove then that the properties are sufficient.

Let $\Phi : \mathcal{S}_n \to \mathbb{C}$ be the matrix Fourier transform of a measure, satisfying the hypotheses of the theorem:

$$\Phi(X) = \int_{\mathcal{S}_n} e^{i\mathrm{Tr}\, XM} \mathbf{P}(d_1 M), \quad X \in \mathcal{S}_n.$$

The function Φ is the matrix analog of the characteristic function in probability theory.

Property (i) implies that for any $X \in \mathcal{S}_n$ we have

$$\Phi(OXO^T) = \Phi(X), \ \forall O \in O(n), \tag{1.1.9}$$

and property (ii) implies that Φ has the form

$$\Phi(X) = \prod_{1 \le j \le k \le n} \Phi_{jk}(X_{jk}), \tag{1.1.10}$$

where $X = \{X_{jk}\}_{j,k=1}^n$,

$$\begin{aligned} \Phi_{jj}(x) &= \int e^{ix\mu} F_{jj}(d\mu), \quad 1 \le j \le n, \\ \Phi_{jk}(x) &= \int e^{2ix\mu} F_{jk}(d\mu), \quad 1 \le j < k \le n, \end{aligned} \tag{1.1.11}$$

and

$$\Phi_{jk}(0) = 1, \quad 1 \le j \le k \le n. \tag{1.1.12}$$

Here and everywhere below integrals without limit denote integrals over the whole $\mathbb{R}$.

Take in (1.1.9) the real symmetric matrix X whose nonzero entries are $X_{jj} = x,\ X_{kk} = z,\ X_{jk} = X_{kj} = y$, for some pair $(j,k),\ j < k$. Then, making the rotation OXO^T on $\pi/2$ in the (j,k)-plane of X with $y = z = 0$, we obtain from (1.1.9) and (1.1.10) the equality $\Phi_{jj}(x) = \Phi_{kk}(x)$, implying that all the "diagonal" functions coincide. We denote the corresponding function by f. Note that f is continuous as the Fourier transform of a finite measure. Likewise, the rotation of X with $x = z$ on $\pi/4$ in the (j,k)-plane yields

$$f^2(x)\Phi_{jk}(y) = f(x+y)f(x-y). \tag{1.1.13}$$

For $x = 0$ we obtain, in view of (1.1.12), that $\Phi_{jk}(y) = f(y)f(-y)$, i.e., that all "off-diagonal" functions Φ_{jk} coincide. Denoting this function by g, we can rewrite the previous relation as

$$g(y) = f(y)f(-y). \tag{1.1.14}$$

This relation and (1.1.13) lead to the functional equation for f:

$$f^2(x)f(y)f(-y) = f(x+y)f(x-y). \tag{1.1.15}$$

Denote $|f(x)| = r(x) \ge 0,\ \forall x \in \mathbb{R}$. It follows from (1.1.15) and from the equality $f(-x) = \overline{f(x)}$ (see (1.1.11)) that r is even and satisfies the functional equation

$$r^2(x)r^2(y) = r(x+y)r(x-y). \tag{1.1.16}$$

Setting $x = y$ here and taking into account (1.1.12), we have the relations

$$r(x) = r^4(x/2) = \cdots = r^{4^k}(x/2^k),$$

implying that r is never zero, for if it were yes, we obtain a contradiction with (1.1.12) and with the continuity of f for sufficiently big k.

Furthermore, it follows from (1.1.16) that for any positive integer p and any real y we have $r(py) = r^{p^2}(y)$. In particular, if $y = 1/q$ and q is a positive integer, then $r(p/q) = r^{(p/q)^2}(1)$. Since r is continuous, we conclude that $r(x) = e^{-\gamma x^2/4}$, where $\gamma = -4\log r(1),\ 0 \le \gamma < \infty$. It follows from the positivity of r for all x that we can define the phase factor $e(x) = f(x)/r(x)$ and obtain from (1.1.15) and (1.1.16) the functional equation for e : $e^2(x) = e(x+y)e(x-y)$. The equation and the continuity of e imply that $e = e^{i\alpha x}$ with a real α. We obtain that $f(x) = e^{i\alpha x - \gamma x^2/4}$. This form of f and (1.1.14) yield $g = e^{-\gamma x^2/2}$.

Besides, we have in view of (1.1.11)

$$f(x) = \int e^{ix\mu} F_{11}(d\mu), \qquad g(x) = \int e^{2ix\mu} F_{12}(d\mu),$$

and by uniqueness of the Fourier transformation it follows that

$$F_{11}(d\mu) = (\pi\gamma)^{-1/2} e^{-\frac{(\mu-\alpha)^2}{\gamma}} d\mu, \qquad F_{12}(d\mu) = (\pi\gamma/2)^{-1/2} e^{-\frac{2\mu^2}{\gamma}} d\mu.$$

This implies (1.1.8). □

Remark 1.1.4. *(1)* The theorem has been known since the 1960s **[356, 416]**, although its proof in these works requires the continuous differentiability of the density of the matrix measure. The proof above makes clear that to obtain the Gaussian form (1.1.8) of the matrix measure we need not to assume either its absolute continuity or its invariance with respect to all orthogonal transformations. The last observation has to be compared with the Skitovich-Darmois theorem in statistics **[299]**, allowing one to obtain another proof of the theorem.

(2) For another argument, leading to the Gaussian distribution (1.1.8), see Problem 1.3.4.

Gaussian Ensembles, determined by (1.1.1) – (1.1.3), as well as random matrices, assuming values in classical compact groups and corresponding symmetric spaces, are among the most known and studied in random matrix theory. They are considered in Chapters 2, 3, and 8, respectively. There is, of course, a considerable amount of other ensembles that random matrix theory deals with. For instance, omitting the requirement of the invariance in the definition of Gaussian Ensembles (see Theorem 1.1.3), we obtain random matrices with statistical independent (modulo symmetry conditions) but not necessarily Gaussian entries, whose variances correspond to (1.1.3) for $\beta = 1, 2, 4$, in particular, distributions (1.1.9) but without (1.1.6) in the case of real symmetric matrices. These are *Wigner Ensembles* that are considered in Chapter 18.

On the other hand, omitting the condition of statistical independence of functionally independent entries in the definition of Gaussian Ensembles, we obtain a large class of invariant ensembles, i.e., real symmetric matrices with orthogonal invariant and hermitian random matrices with unitary invariant probability law. In recent decades an important subclass of this class, known as *Matrix Models*, has been actively studied. We consider Matrix Models in Chapters 11 – 16.

One more class of random matrices is well known in mathematical statistics (in fact, these are historically the first random matrices). These are the sample (or empirical) covariance matrices (see e.g. **[370**, Sections 1.2 and 3.1]). In the simplest case they are constructed from the real or complex random matrices with i.i.d. Gaussian entries and are known as the *Wishart* and the *Laguerre Ensembles*,

respectively (see Chapter 7). They can be viewed as matrix analogs of the χ^2 random variable. In more general cases sample covariance matrices are constructed from random matrices with independent but not necessarily Gaussian entries. They are considered in Chapter 19.

There is a wide variety of other ensembles that appear in random matrix theory and its numerous applications. We refer the reader to review works [**41, 217, 252, 356, 492**] and references therein for examples and results.

1.1.2. Quantities and Problems. We discuss here several typical quantities and problems that are widely studied in random matrix theory. There are, of course, a number of others that are considered in this book and in numerous works.

A number of spectral properties of random matrices can be formulated and studied in terms of the Counting Measure of eigenvalues and the of matrices.

Definition 1.1.5. Let $(\lambda_1^{(n)}, \dots, \lambda_n^{(n)})$ be eigenvalues of a real symmetric or hermitian matrix of order n. The *Counting Measure* $\mathcal{N}_n$ of its eigenvalues is

$$\mathcal{N}_n(\Delta) = \sharp\{l \in \{1, \dots, n\} : \lambda_l^{(n)} \in \Delta\}, \tag{1.1.17}$$

where Δ is an interval of $\mathbb{R}$, and the *Normalized Counting Measure* N_n of its eigenvalues is

$$N_n(\Delta) = n^{-1}\mathcal{N}_n(\Delta). \tag{1.1.18}$$

If M is a random matrix, then the Counting Measure $\mathcal{N}_n(\Delta)$ of its eigenvalues is a random variable (measurable function) for a given Δ. Indeed, assume that eigenvalues are indexed in the nondecreasing order

$$-\infty < \lambda_1^{(n)} \le ... \le \lambda_n^{(n)} < \infty. \tag{1.1.19}$$

Then each of them is a continuous (even a piecewise analytic) function of matrix elements (see e.g. [**301**, Section II.6.4]). Hence, $\mathcal{N}_n(\Delta)$ is a measurable functions of matrix elements for any Δ.

The Counting Measure is a particular case of a *linear statistic* (the term of mathematical statistics) or an *additive observable* (the term of statistical mechanics) $\mathcal{N}_n[\varphi]$ of random variables $\{\lambda_l^{(n)}\}_{l=1}^n$, defined by a *test function* $\varphi : \mathbb{R} \to \mathbb{C}$:

$$\mathcal{N}_n[\varphi] := \sum_{l=1}^{n} \varphi(\lambda_l^{(n)}) = \int \varphi(\lambda)\mathcal{N}_n(d\lambda) = \operatorname{Tr}\varphi(M). \tag{1.1.20}$$

We obtain $\mathcal{N}_n(\Delta)$ as $\mathcal{N}_n[\chi_\Delta]$, where χ_Δ is the indicator of $\Delta \subset \mathbb{R}$.

One can also consider the multilinear statistics, defined by the multiple sums instead of (1.1.20); see e.g. [**350**].

If $\mathbf{P}$ is a probability measure, determining a random matrix, and $\mathbf{E}\{...\}$ denotes the corresponding expectation, then we can consider

(i) the expectation

$$\overline{\mathcal{N}}_n[\varphi] = \mathbf{E}\{\mathcal{N}_n[\varphi]\}, \tag{1.1.21}$$

of a linear statistic, in particular, the expectation

$$\overline{N}_n(\Delta) = \mathbf{E}\{N_n(\Delta)\} \tag{1.1.22}$$

of the Normalized Counting Measure;

(ii) the covariance

$$\begin{aligned}(1.1.23)\qquad &\mathbf{Cov}\{\mathcal{N}_n[\varphi_1],\mathcal{N}_n[\varphi_2]\}\\ &\quad=\mathbf{E}\{\mathcal{N}_n[\varphi_1]\mathcal{N}_n[\varphi_2]\}-\mathbf{E}\{\mathcal{N}_n[\varphi_1]\}\mathbf{E}\{\mathcal{N}_n[\varphi_2]\}\end{aligned}$$

of linear statistics, corresponding to two test functions;

(iii) the probability law of $\mathcal{N}_n[\varphi]$, in particular, the distribution

$$(1.1.24)\qquad E_n(l;\Delta)=\mathbf{P}\{\mathcal{N}_n(\Delta)=l\},\ l=0,1,2,\ldots,n,$$

of the Counting Measure for a given interval $\Delta\subset\mathcal{R}$, in particular, the *gap probability*

$$(1.1.25)\qquad E_n(\Delta):=E_n(0;\Delta)=\mathbf{P}\{\mathcal{N}_n(\Delta)=0\}.$$

Linear statistics (1.1.20), as well as many other functions of eigenvalues that we are going to study, are symmetric functions of eigenvalues. This allows us to use the symmetric extension of the joint probability law of eigenvalues, defined initially on the simplex (1.1.19) of $\mathbb{R}^n$ to the whole $\mathbb{R}^n$. Moreover, most of the random matrices that we will discuss possess the probability law that is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^n$. Denote by $p_n(\lambda_1,...,\lambda_n)$ the symmetric extension of the corresponding probability density from (1.1.19) to $\mathbb{R}^n$ (see e.g. Corollary 4.1.3 and Proposition 9.1.1), and let

(1.1.26)

$$p_l^{(n)}(\lambda_1,\ldots,\lambda_l)=\int p_n(\lambda_1,\ldots,\lambda_l,\lambda_{l+1},\ldots,\lambda_n)d\lambda_{l+1}\ldots d\lambda_n,\ l=1,...,n,$$

be the *marginals* of p_n.

Write the gap probability as

$$(1.1.27)\qquad E_n(\Delta)=\mathbf{E}\left\{\prod_{l=1}^{n}\left(1-\chi_\Delta\left(\lambda_l^{(n)}\right)\right)\right\}.$$

This suggests considering a more general quantity, replacing χ_Δ by a function $\varphi:\mathbb{R}\to\mathbb{C}$ (cf. (1.1.20)):

$$(1.1.28)\qquad E_n[\varphi]=\mathbf{E}\left\{\prod_{l=1}^{n}\left(1-\varphi\left(\lambda_l^{(n)}\right)\right)\right\}.$$

Now, by using the marginals (1.1.26) of the symmetric version of the joint probability density of eigenvalues and denoting by the same symbol $\mathbf{E}\{...\}$ the corresponding expectation, we obtain

$$\begin{aligned}(1.1.29)\qquad E_n[\varphi]&=\sum_{l=0}^{n}(-1)^l\binom{n}{l}\mathbf{E}\left\{\varphi\left(\lambda_1^{(n)}\right)...\varphi\left(\lambda_l^{(n)}\right)\right\}\\ &=\sum_{l=0}^{n}(-1)^l\binom{n}{l}\int p_l^{(n)}(\lambda_1,\ldots,\lambda_l)\varphi\left(\lambda_1\right)\ldots\varphi\left(\lambda_l\right)d\lambda_1...d\lambda_l,\end{aligned}$$

where the term corresponding to $l=0$ is 1. This implies the formula

$$(1.1.30)\quad (-1)^l\frac{\delta^l}{\delta\varphi(\lambda_1)...\delta\varphi(\lambda_l)}E_n[\varphi]\Big|_{\varphi=0}=n(n-1)...(n-l+1)p_l^{(n)}(\lambda_1,\ldots,\lambda_l),$$

valid for any collection of distinct $(\lambda_1, ..., \lambda_l)$. Note that we understand the variational derivative

$$\frac{\delta}{\delta\varphi(\lambda)}\Phi[\varphi]$$

of functional Φ, defined on a certain set of functions, as the function that determines the linear functional in ψ

$$\lim_{\varepsilon\to 0}\varepsilon^{-1}(\Phi[\varphi+\varepsilon\psi]-\Phi[\varphi]) = \int \psi(\lambda)\frac{\delta}{\delta\varphi(\lambda)}\Phi[\varphi]d\lambda, \tag{1.1.31}$$

defined on a certain set of functions. In the case (1.1.29) – (1.1.30) it suffices to assume that φ and ψ are continuous functions of compact support.

The r.h.s. of formula (1.1.30) is known as the *correlation functions* (marginal probability densities of eigenvalues regardless of their labeling)

$$R_l^{(n)}(\lambda_1, ..., \lambda_l) = n(n-1)...(n-l+1)p_l^{(n)}(\lambda_1, \dots, \lambda_l), \quad l = 1, \dots, n, \tag{1.1.32}$$

of the collection $\{\lambda_l^{(n)}\}_{l=1}^n$ by analogy with statistical mechanics and the theory of point random processes. It can also be defined by the equality

(1.1.33)

$$\mathbf{E}\left\{\sum_{j_1\neq\cdots\neq j_l}\varphi(\lambda_{j_1}, \dots, \lambda_{j_l})\right\} = \int \varphi(\lambda_1, \dots, \lambda_l)R_l^{(n)}(\lambda_1, \dots, \lambda_l)d\lambda_1 \dots d\lambda_l,$$

valid for any continuous and bounded function $\varphi : \mathbb{R}^l \to \mathbb{C}$. The summation in the l.h.s. of the last formula is over all l-tuples of distinct integers $j_1, \dots, j_l = 1, \dots, n$.

Setting $\varphi = g\chi_\Delta,\ g \in \mathbb{R}$, in (1.1.28), we obtain

$$(-1)^l\left(\frac{\partial}{\partial g}\right)^l E_n[g\chi_\Delta]\Big|_{g=0} = l!E_n(l;\Delta), \tag{1.1.34}$$

where $E_n(l;\Delta)$ is defined in (1.1.24).

We see that (1.1.28) plays the role of the *generating functional* for both the marginals (correlation functions) of the random variables $\{\lambda_l^{(n)}\}_{l=1}^n$ and the probability distribution (1.1.24) of the Counting Measure of eigenvalues (1.1.17).

Moreover, replacing φ in (1.1.28) by $1 - e^{-\varphi}$, we obtain the *Laplace characteristic (moment generating) functional* of the linear statistics (1.1.20):

$$\begin{aligned} Z_n[\varphi] := \mathbf{E}\left\{e^{-\mathcal{N}_n[\varphi]}\right\} &= \mathbf{E}\left\{\prod_{l=1}^n e^{-\varphi(\lambda_l^{(n)})}\right\} \\ &= E_n[1-e^{-\varphi}]. \end{aligned} \tag{1.1.35}$$

Write (1.1.30) – (1.1.32) as

$$R_l^{(n)}(\lambda_1, \dots, \lambda_l) = (-1)^l\frac{\delta^l}{\delta\varphi(\lambda_1)\dots\delta\varphi(\lambda_l)}E_n[\varphi]\Big|_{\varphi=0}. \tag{1.1.36}$$

The formula shows that the correlation functions are analogs of moments of random variables. Recall that another collection of parameters characterizing random variables is the collection of their cumulants (semi-invariants) (see (18.1.16) – (18.1.17) for their definition). Analogs of cumulants in statistical mechanics and the theory of random processes are the *cluster (or Ursell) functions*. This makes natural

the introduction of analogous functions for the collection of eigenvalues of random matrices as

$$T_l^{(n)}(\lambda_1,\dots,\lambda_l) = -\frac{\delta^l}{\delta\varphi(\lambda_1)\dots\delta\varphi(\lambda_l)} \log E_n[\varphi]\Big|_{\varphi=0}. \tag{1.1.37}$$

Here are the expressions of the first three cluster functions via the correlation functions:

$$\begin{aligned} T_1^{(n)}(\lambda_1) &= R_1^{(n)}(\lambda_1), \\ T_2^{(n)}(\lambda_1,\lambda_2) &= R_1^{(n)}(\lambda_1)R_1^{(n)}(\lambda_2) - R_2^{(n)}(\lambda_1,\lambda_2), \\ T_3^{(n)}(\lambda_1,\lambda_2,\lambda_3) &= R_3^{(n)}(\lambda_1,\lambda_2,\lambda_3) - R_1^{(n)}(\lambda_1)R_2^{(n)}(\lambda_2,\lambda_3) \\ &\quad -R_1^{(n)}(\lambda_2)R_2^{(n)}(\lambda_3,\lambda_1) - R_1^{(n)}(\lambda_3)R_2^{(n)}(\lambda_1,\lambda_2) \\ &\quad +2R_1^{(n)}(\lambda_1)R_1^{(n)}(\lambda_2)R_1^{(n)}(\lambda_3). \end{aligned} \tag{1.1.38}$$

Cluster functions describe the correlation properties of a single cluster of a given number of eigenvalues, isolated from the effects of lower order correlations. Here is a simple but useful example.

PROPOSITION 1.1.6. *Let $\mathcal{N}_n[\varphi]$ be a linear eigenvalue statistic (1.1.20) with bounded test function $\varphi : \mathbb{R} \to \mathbb{C}$. Then*

$$\begin{aligned} \mathbf{Var}\{\mathcal{N}_n[\varphi]\} : &= \mathbf{E}\{|\mathcal{N}_n[\varphi]|^2\} - |\mathbf{E}\{\mathcal{N}_n[\varphi]\}|^2 \\ &= \frac{1}{2}\int\int |\varphi(\lambda_1)-\varphi(\lambda_2)|^2 T_2^{(n)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2, \end{aligned} \tag{1.1.39}$$

and more generally

$$\begin{aligned} &\mathbf{Cov}\{\mathcal{N}_n[\varphi_1],\mathcal{N}_n[\varphi_2]\} \\ &= \frac{1}{2}\int\int (\varphi_1(\lambda_1)-\varphi_1(\lambda_2))(\varphi_2(\lambda_1)-\varphi_2(\lambda_2)) T_2^{(n)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2. \end{aligned} \tag{1.1.40}$$

PROOF. It follows from (1.1.23), (1.1.20), and (1.1.33) that

$$\mathbf{E}\{\mathcal{N}_n[\varphi]\} = \int \varphi(\lambda_1)R_1^{(n)}(\lambda_1)d\lambda_1, \tag{1.1.41}$$

$$\begin{aligned} \mathbf{E}\{|\mathcal{N}_n[\varphi]|^2\} &= \int |\varphi(\lambda_1)|^2 R_1^{(n)}(\lambda_1)d\lambda_1 \\ &\quad + \int\int \varphi(\lambda_1)\overline{\varphi(\lambda_2)}R_2^{(n)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2, \end{aligned} \tag{1.1.42}$$

and then (1.1.38) implies that

$$\begin{aligned} \mathbf{Var}\{\mathcal{N}_n[\varphi]\} &= \int |\varphi(\lambda_1)|^2 R_1^{(n)}(\lambda_1)d\lambda_1 \\ &\quad - \int\int \varphi(\lambda_1)\overline{\varphi(\lambda_2)}T_2^{(n)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2. \end{aligned} \tag{1.1.43}$$

Now we use (1.1.32) to write

$$\int R_2^{(n)}(\lambda_1,\lambda_2)d\lambda_2 = (n-1)R_1^{(n)}(\lambda_1),$$

and then (1.1.38) yields

$$R_1^{(n)}(\lambda_1) = T_1^{(n)}(\lambda_1) = \int T_2^{(n)}(\lambda_1,\lambda_2)d\lambda_2.$$

This and (1.1.43) lead to (1.1.39). The proof of (1.1.40) is analogous. □

Equivalent definitions of correlation and cluster functions are given by the formulas, involving the generating functional (1.1.28) – (1.1.29):

$$E_n[\varphi] = 1 + \sum_{l=1}^{n} \frac{(-1)^l}{l!} \int \dots \int R_l^{(n)}(\lambda_1, \dots, \lambda_l)\varphi(\lambda_1) \dots \varphi(\lambda_l) d\lambda_1 \dots d\lambda_l, \tag{1.1.44}$$

and

$$\log E_n[\varphi] = -\sum_{l=1}^{\infty} \frac{1}{l!} \int \int T_l^{(n)}(\lambda_1, \dots, \lambda_l)\varphi(\lambda_1) \dots \varphi(\lambda_l) d\lambda_1 \dots d\lambda_l. \tag{1.1.45}$$

It follows from (1.1.36) and (1.1.37) (or (1.1.44) and (1.1.45)) that (cf. (1.1.38))

$$T_l^{(n)}(\lambda_1, \dots, \lambda_l) = \sum_{\mathcal{P}} (-1)^{l-|\mathcal{P}|}(|\mathcal{P}| - 1)! \prod_{P \in \mathcal{P}} R_{|P|}^{(n)}(\{\lambda_j\}_{j \in P}), \tag{1.1.46}$$

and

$$R_l^{(n)}(\lambda_1, \dots, \lambda_l) = \sum_{\mathcal{P}} (-1)^{l-|\mathcal{P}|} \prod_{P \in \mathcal{P}} T_{|P|}^{(n)}(\{\lambda_j\}_{j \in P}), \tag{1.1.47}$$

where the sum in both formulas is over the partitions $\mathcal{P}$ of $(1, \dots, n)$ into disjoint sets, and $|\dots|$ denotes the cardinality of the corresponding set.

It is important that (1.1.46) and (1.1.47) establish a one-to-one correspondence between the correlation and cluster functions via a finite number of additions and multiplications. This will allow us to find the form of $T_l^{(n)}$ for a given l in various asymptotic regimes as $n \to \infty$ via corresponding asymptotics of $R_k^{(n)}$, $k \le l$, and vice versa.

One more important characteristic of the eigenvalue distribution is the gap probability (1.1.25). In particular, it is closely related to the distribution of spacings between pairs of adjacent eigenvalues. Indeed, consider the conditional probability distribution of an eigenvalue given its left-hand neighbor in $[a - h, a)$, $h > 0$. The limit of this distribution as $h \to 0^+$ can be identified with the probability that the distance between an eigenvalue at a and its nearest neighbor to the right is bigger than $b - a$. Denoting this probability $q_a^{(n)}(b - a)$, we have

$$\begin{aligned} q_a^{(n)}(b-a) &= \lim_{h \to 0^+} \mathbf{P}\{\mathcal{N}_n((a,b]) = 0 \,|\, \mathcal{N}_n((a-h,a]) > 0\} \\ &= \lim_{h \to 0^+} \mathbf{P}\{\mathcal{N}_n((a,b]) = 0,\ \mathcal{N}_n((a-h,a]) > 0\} / \mathbf{P}\{\mathcal{N}_n((a-h,a]) > 0\}. \end{aligned} \tag{1.1.48}$$

In view of (1.1.17) the events on the r.h.s. of the formula are symmetric in eigenvalues. Hence we can use the symmetric extension p_n of the joint probability density of eigenvalues and write

$$q_a^{(n)}(b-a) = (\rho_n(a))^{-1} \int_{(\mathbb{R} \setminus (a,b])^{n-1}} p_n(a, \lambda_2, \dots, \lambda_n) d\lambda_2 \dots d\lambda_n, \tag{1.1.49}$$

where

$$\rho_n(a) := p_1^{(n)}(a) = \int p_n(a, \lambda_2, \dots, \lambda_n) d\lambda_2 \dots d\lambda_n$$

is the first marginal of p_n and also the density of the measure $\overline{N}_n$ of (1.1.22):

$$\overline{N}_n(\Delta) = \int_{\Delta} \rho_n(\lambda) d\lambda. \tag{1.1.50}$$

On the other hand, the gap probability (1.1.25) for $\Delta = (a, b]$ is

$$E_n((a,b]) = \int_{(\mathbb{R}\setminus(a,b])^n} p_n(\lambda_1, \dots, \lambda_n) d\lambda_1 \dots d\lambda_n.$$

This formula and (1.1.49) imply

$$\frac{\partial}{\partial a} E_n((a,b]) = n\rho_n(a) q_a^{(n)}(b-a) \tag{1.1.51}$$

or, since $E_n((b,b]) = 1$,

$$1 - E_n((a,b]) = \int_a^b q_{a'}^{(n)}(b-a') n\rho_n(a') da'. \tag{1.1.52}$$

The last formula admits a simple interpretation. Indeed, its l.h.s. is the probability of the event $\{\mathcal{N}_n((a,b]) \geq 1\}$, and if we look "backward" from a point b, there will be some point $a' < b$, for which $\mathcal{N}_n((a',b]) = 0$ but $\mathcal{N}_n([a',b]) = 1$. Thus we have with probability 1

$$\{\mathcal{N}_n((a,b]) \geq 1\} = \bigcup_{a \leq a' \leq b} \{\mathcal{N}_n((a',b]) = 0,\ \mathcal{N}_n(\{a'\}) = 1\}, \tag{1.1.53}$$

and observing that the r.h.s. is the union of mutually exclusive events, we can add their probabilities. This yields the r.h.s. of (1.1.52).

It follows from (1.1.51) that the density of the spacing distribution is

$$p_a^{(n)}(t) = \frac{1}{n\rho_n(a)} \frac{\partial^2}{\partial a \partial b} E_n((a,b]) \Big|_{b=a+t}. \tag{1.1.54}$$

In particular, the expression for $p_a^{(n)}$ via the joint probability density of eigenvalues is

$$p_a^{(n)}(t) = \frac{n-1}{\rho_n(a)} \int_{(\mathbb{R}\setminus(a,a+t])^{n-2}} p_n(a, a+t, \lambda_3, \dots, \lambda_n) d\lambda_3 \dots d\lambda_n. \tag{1.1.55}$$

These formulas and their limits as $n \to \infty$ will often be used below while discussing the spacing distribution of various random matrices. Note that (1.1.51) – (1.1.52) is a simple version of the Palma-Khintchin formula in the theory of point processes (see [**147**, Section 3.4]). For another derivation of the above formulas see [**302**, Sections 1.0 – 1.1 and 2.0 – 2.1].

The gap probability is also related to the probability law of the minimum $\lambda_{min}^{(n)} = \lambda_1^{(n)}$ (maximum $\lambda_{max}^{(n)} = \lambda_n^{(n)}$) eigenvalues of real symmetric and hermitian random matrices. Namely, we have

$$\mathbf{P}\{\lambda_{min}^{(n)} \geq \lambda\} = E_n((-\infty, \lambda)); \quad \mathbf{P}\{\lambda_{max}^{(n)} \leq \lambda\} = E_n((\lambda, \infty)). \tag{1.1.56}$$

The gap probability $E_n(\Delta)$ of (1.1.25) can be viewed as the first term of the sequence $\{E_n(m;\Delta)\}_{m=0}^{\infty}$, when $E_n(0;\Delta) = E_n(\Delta)$ and $E_n(m;\Delta)$ is the probability (1.1.24) to have exactly m eigenvalues in Δ. We obviously have

$$E_n(m;\Delta) = \frac{n!}{m!(n-m)!} \int_{\Delta^m} d\lambda_1 \dots d\lambda_m \int_{(\mathbb{R}\setminus\Delta)^m} p_n(\lambda_1, \dots, \lambda_n) d\lambda_{m+1} \dots d\lambda_n, \tag{1.1.57}$$

where p_n is the symmetrized probability density of eigenvalues and

$$E_n(m;\Delta) = \frac{(-1)^m}{m!} \left(\frac{\partial}{\partial t}\right)^m E_n[t\chi_\Delta] \Big|_{t=0}, \tag{1.1.58}$$

where $E_n[\varphi]$ is defined in (1.1.28).

If $\lambda_j^{(n)}$ is the jth lowest eigenvalue (see (1.1.19)), then (cf. (1.1.56))

$$\mathbf{P}\{\lambda_j^{(n)} \geq \lambda\} = \sum_{l=0}^{j-1} E_n(l; (-\infty, \lambda)), \tag{1.1.59}$$

because the event $\{\lambda_j^{(n)} \geq \lambda\}$ is the disjoint union of sets of matrices with exactly l eigenvalues in $(-\infty, \lambda)$ for $l = 0, \dots, j-1$.

We mention the interpretation of the functional $E_n[t\chi_\Delta]$, $t \in [0,1]$, appearing in (1.1.58) [**72**]. Let us remove at random $1-t$ eigenvalues from the spectrum $\{\lambda_l^{(n)}\}_{l=1}^n$; i.e., remove every $\lambda_l^{(n)}$ with probability $1-t$ independently of other eigenvalues. In order to keep the mean density fixed, as $n \to \infty$, we contract (rescale) the spectral axis by t. Denoting the spectral characteristics resulting from this "rarefication" procedure by the same symbol with the superscript (t), we obtain

$$\rho_n^{(t)}(\lambda) = \rho_n(\lambda t^{-1}) t^{-1} \tag{1.1.60}$$

and

$$E_n^{(t)}(\Delta) = \sum_{m=0}^{n} E_n(m; \Delta/t).$$

The last formula is clear from the fact that $1-t$ is the probability of dropping the eigenvalue. Combining the formula with (1.1.58), we obtain

$$E_n^{(t)}(\Delta) = E_n[t\chi_{\Delta/t}]. \tag{1.1.61}$$

This is used in Problems 1.3.8 and 5.4.12. Various advanced versions of decimation (or alternating) as well as the superposition (a kind of inversion to the decimation) allow one to relate different ensembles (see e.g. [**356**, Section 11.6], [**218, 221**] and references therein).

We will first be interested in the asymptotic properties of these quantities in the *global regime*, in which the weak or tight convergence of the sequence $\{N_n\}$ of random measures is studied. Typically N_n converges weakly in probability or almost surely to a nonrandom measure N.

Recall that we say that a sequence $\{m_n\}$ of nonnegative normalized to unity (or probability) measures on $\mathbb{R}$ *converges weakly* to a probability measure m if for every continuous and bounded function $\varphi : \mathbb{R} \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int \varphi(x) m_n(dx) = \int \varphi(x) m(dx). \tag{1.1.62}$$

Furthermore, if m is a probability measure on $\mathbb{R}$, then the nonnegative and nondecreasing function

$$F(\lambda) = m((-\infty, \lambda]) \tag{1.1.63}$$

is called the *distribution function* of m. The weak convergence of $\{m_n\}$ to m is equivalent to the convergence

$$\lim_{n\to\infty} F_n(\lambda) = F(\lambda) \tag{1.1.64}$$

of respective distribution functions at every continuity point of F, and

$$F(-\infty) = 0, \qquad F(\infty) = 1. \tag{1.1.65}$$

A drawback of the above notion is that not every sequence $\{m_n\}$ of probability measures contains a weakly convergent subsequence. This justifies the following

DEFINITION 1.1.7. We say that a sequence $\{m_n\}$ of probability measures converges *vaguely* to a nonnegative measure m if their distribution functions satisfy (1.1.64) at every continuity point of F (but F does not necessarily satisfy (1.1.65)) or if (1.1.62) holds for all bounded continuous functions of compact support.

A drawback of this notion is that the limiting measure is not necessarily a probability measure, because a vaguely convergent sequence of probability measures can lose mass, so that $0 \leq m(\mathbb{R}) \leq 1$. An advantage is that any sequence of probability measures contains a subsequence, converging vaguely to a nonnegative measure (Helly's selection principle). To prove that a vaguely converging sequence $\{m_n\}$ of probability measures is weakly convergent, we have to prove that the limiting measure possesses the property $m(\mathbb{R}) = 1$ or (1.1.65). Thus we have to prevent the "escape" of mass to infinity. This is provided by the notion of *tightness*.

A sequence $\{m_n\}$ of probability measures is said to be *tight* if for any $\varepsilon > 0$ there exists a finite interval I, such that $m_n(I) \geq 1- \varepsilon$ for all n. A convenient sufficient condition for a sequence of probability measures to be tight is the uniform boundedness in n of a moment of positive order of m_n:

$$\sup_n \int |\lambda|^a m_n(d\lambda) < \infty, \text{ for some } a > 0. \tag{1.1.66}$$

If $\{m_n\}$ is a tight sequence of probability measures, then there exists a weakly convergent subsequence $\{m_{n_k}\}$ **[62, 440]**.

The notion of weak convergence can be paraphrased as the notion of the convergence in distribution of random variables. Namely, we say that a sequence $\{\xi_n\}$ *converges in distribution* to the random variable ξ if the sequence of the probability laws of ξ_n converges weakly to the probability law of ξ.

1.2. Example

This section deals in fact with standard objects of probability theory, collections of independent identically distributed (i.i.d.) random variables. We will present, however, corresponding arguments and results in a form that is natural from the point of view of random matrix theory.

Let $\{\lambda_j^{(n)}\}_{j=1}^n$ be a collection (a triangular array) of i.i.d. random variables with the common law F_n. Define the diagonal $n \times n$ random matrix M as follows:

$$M = \{M_{jk}\}_{j,k=1}^n, \quad M_{jk} = \delta_{jk}\lambda_j^{(n)}. \tag{1.2.1}$$

Thus the Normalized Counting Measure of eigenvalues of M, the empirical measure of $\{\lambda_j^{(n)}\}_{=1}^n$ in statistical terms, is

$$N_n(\Delta) = n^{-1} \sum_{j=1}^n \chi_\Delta(\lambda_j^{(n)}), \tag{1.2.2}$$

where χ_Δ is the indicator of $\Delta \subset \mathbb{R}$, and we easily obtain for (1.1.22) and (1.1.23) with $\varphi_1 = \varphi_2 = \chi_\Delta$:

$$\overline{N}_n(\Delta) = F_n(\Delta), \qquad \mathbf{Var}\{N_n(\Delta)\} = F_n(\Delta)(1 - F_n(\Delta))/n, \tag{1.2.3}$$

and for (1.1.24) and (1.1.25)

$$E_n(l;\Delta) = \binom{n}{l} F_n^l(\Delta)(1-F_n(\Delta))^{n-l}, \tag{1.2.4}$$

$$E_n(\Delta) = (1-F_n(\Delta))^n. \tag{1.2.5}$$

More generally, we have in this case for (1.1.23) and (1.2.7)

$$\begin{aligned}\mathbf{Cov}\{\mathcal{N}_n[\varphi_1],\mathcal{N}_n[\varphi_2]\} \\ = \frac{n}{2}\int(\varphi_1(\lambda_1)-\varphi_1(\lambda_2))(\varphi_2(\lambda_1)-\varphi_2(\lambda_2))F_n(d\lambda_1)F_n(d\lambda_2),\end{aligned} \tag{1.2.6}$$

$$E_n[\varphi] = \left(1-\int\varphi(\lambda)F_n(d\lambda)\right)^n. \tag{1.2.7}$$

In particular, if F_n is absolutely continuous, i.e., $F_n(d\lambda) = \rho_n(\lambda)d\lambda$, then we have from (1.1.37) and (1.1.38)

$$T_2^{(n)}(\lambda_1,\lambda_2) = n\rho_n(\lambda_1)\rho_n(\lambda_2) \tag{1.2.8}$$

in accordance with Proposition 1.1.6 and (1.2.6).

Assume that F_n converges weakly to some F. Then $\overline{N}_n = F_n$ converges weakly to $N = F$:

$$\lim_{n\to\infty}\overline{N}_n = N. \tag{1.2.9}$$

Furthermore, since for any $\Delta\subset\mathbb{R}$, $\mathbf{Var}\{N_n(\Delta)\} = O(1/n)$, $n\to\infty$, $N_n(\Delta)$ converges in probability to $N(\Delta)$. Moreover, computing the 4th central moment $\mathbf{E}\{|N_n(\Delta)-\mathbf{E}\{N_n(\Delta)\}|^4\}$ of N_n, we find that it is of the order $O(1/n^2)$, $n\to\infty$. Thus, if the collections $\{\lambda_j^{(n)}\}_{j=1}^n$ are defined on the same probability space for all n, for example, if $\lambda_1^{(n)},\dots,\lambda_n^{(n)}$ are the first n terms of an infinite sequence $\{\lambda_j\}_{j=1}^\infty$ of i.i.d. random variables, then, by applying the Borel-Cantelli lemma, we conclude that $N_n(\Delta)$ converges to $N(\Delta)$ with probability 1 for any Δ.

Taking a set of intervals with rational endpoints, we obtain that the random measures $\{N_n\}$ converge weakly to the nonrandom measure N with probability 1, and taking $\Delta = (-Q,Q)$ such that $F(\mathbb{R}\setminus\Delta)\le\varepsilon$, we obtain that $\{N_n\}$ is tight with probability 1.

This proof of the tightness of the sequence $\{N_n\}$ is rather an exception than a rule. As a rule, one proves this fact by using the uniform boundedness in n of a certain moment of N_n. Namely, assume for instance that $\mathbf{E}\{|\lambda_1|\}<\infty$. Then, according to the strong Law of Large Numbers, the random variable

$$\int|\lambda|N_n(d\lambda) = n^{-1}\sum_{j=1}^n|\lambda_j|$$

converges with probability 1 to $\mathbf{E}\{|\lambda_1|\}$, hence is bounded with probability 1. Then the Tchebyshev inequality yields

$$N_n((-Q,Q)) \ge 1 - Q^{-1}n^{-1}\sum_{j=1}^n|\lambda_j|,$$

and we conclude that with probability 1, $N_n((-Q,Q))$ is arbitrarily close to 1 if Q is big enough, i.e., $\{N_n\}$ is tight with probability 1.

The above is a simple consequence of the Law of Large Numbers. Likewise, the Central Limit Theorem for the i.i.d. random variables implies that the random variable $n^{-1/2}\mathcal{N}_n(\Delta)$ converges in distribution to the Gaussian random variable of zero mean and variance $F(\Delta)(1-F(\Delta))$ (see (1.2.3)). One can also find the *large deviations* of $N_n(\Delta)$ (see Comment 11.4.1 for a brief discussion).

Consider now the gap probability (1.1.25) for the diagonal random matrix (1.2.1). We have from (1.2.5)

$$\lim_{n\to\infty} E_n(\Delta) = \begin{cases} 0, & F(\Delta) > 0, \\ 1, & F(\Delta) = 0. \end{cases} \tag{1.2.10}$$

This is in accordance with (1.2.9), implying that if $F(\Delta) > 0$, then Δ contains $O(nF(\Delta))$ eigenvalues of M in Δ as $n \to \infty$.

More detailed information on the asymptotic behavior of the distribution (1.2.4) of the number of eigenvalues $\mathcal{N}_n(\Delta)$ falling in Δ, the gap probability in particular, can be obtained by considering n-dependent intervals. Suppose that $\overline{N}_n = F_n$ and $N = F$ have continuous densities ρ_n and ρ.

Definition 1.2.1. Let N be the limiting Normalized Counting Measure of eigenvalues of a given random matrix. Then the *spectrum* is the support of N. Assume that N has the density ρ. Then the *bulk of spectrum* is

$$\text{bulk } N = \{\lambda \in \text{supp } N : \exists \delta > 0, \inf_{\mu\in[\lambda-\delta,\lambda+\delta]} \rho(\mu) > 0\}. \tag{1.2.11}$$

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

Choose $\lambda_0 \in \text{bulk } N$ and set

$$\Delta_n = (\lambda_0, \lambda_0 + s/n\rho_n(\lambda_0)), \quad s \geq 0. \tag{1.2.12}$$

Then we have that

$$\lim_{n\to\infty} E_n(l; \Delta_n) = s^l e^{-s}/l!, \tag{1.2.13}$$

i.e., that in the case of diagonal random matrices the law of $\mathcal{N}_n(\Delta_n)$, renormalized according to (1.2.12), converges to the *Poisson law* . Moreover, if

$$E_n(l_1, \ldots, l_p; \Delta^{(1)}, \ldots, \Delta^{(p)}) = \mathbf{P}\{\mathcal{N}_n(\Delta^{(1)}) = l_1, \ldots, \mathcal{N}_n(\Delta^{(p)}) = l_p\}, \tag{1.2.14}$$

then we have for $\Delta_n^{(m)} = (\lambda_m, \lambda_m + s_m/n\rho_n(\lambda_m))$, $\lambda_m \in \text{bulk } N$, $m = 1, \ldots, p$:

$$\lim_{n\to\infty} E_n(l_1, \ldots, l_p; \Delta_n^{(1)}, \ldots, \Delta_n^{(p)}) = \prod_{m=1}^{p} s_m^{l_m} e^{-s_m}/l_m!, \tag{1.2.15}$$

i.e., the multivariate Poisson law. In particular, we have for the gap probability

$$\lim_{n\to\infty} E_n(\Delta_n) = \lim_{n\to\infty} E_n(0; \Delta_n) = e^{-s}. \tag{1.2.16}$$

Note that the local regimes in the neighborhood of distinct points are independent.

Furthermore, since in this case the joint symmetrized probability density of eigenvalues is

$$p_n(\lambda_1, \ldots, \lambda_n) = \rho_n(\lambda_1) \ldots \rho_n(\lambda_n),$$

we obtain from (1.1.55)

$$q_a^{(n)}(t) = (1 - F_n(a, a+t))^{n-1},$$

and taking $a = \lambda_0 \in \operatorname{bulk} N,\ t = s/n\rho(\lambda_0)$ in (1.1.55), we obtain for the spacing probability density

$$p^{(n)}_{\lambda_0}(s/n\rho_n(\lambda_0)) = \frac{(n-1)\rho_n(\lambda_0 + s/n\rho_n(\lambda_0))}{n\rho_n(\lambda_0)}(1 - F_n(\lambda_0, \lambda_0 + s/n\rho_n(\lambda_0)))^{n-2};$$

hence

$$p_{\lambda_0}(s) := \lim_{n\to\infty} p^{(n)}_{\lambda_0}(s/n\rho_n(\lambda_0)) = e^{-s}. \tag{1.2.17}$$

The shrinking rate $1/n$ of intervals (1.2.12) is easy to understand. Indeed, if $\Delta \subset \mathbb{R}$ is such that $N(\Delta) > 0$, then the number of eigenvalues falling in Δ is $O(nN(\Delta))$ with probability 1 as $n \to \infty$ in view of (1.2.9). This is why we have (1.2.10). If, however, we would like to find in this situation the probability of having an n-independent number of eigenvalues in a given interval (the gap probability in particular), we have to change the scale of the spectral axis, fixed initially by the requirement of having a well-defined limit of the Normalized Counting Measure (1.1.18), to the scale on which the typical spacing between eigenvalues will be $O(1)$ as $n \to \infty$. The typical initial spacing is $O(1/nN(\Delta))$. Thus, the order of magnitude of the scale that provides the "resolution" of a finite number of eigenvalues as $n \to \infty$ has to be determined by the condition

$$nN(\Delta_n) = O(1), \quad n \to \infty. \tag{1.2.18}$$

This shows that corresponding intervals have to shrink to a point of the spectral axis, and if Δ_n is an infinitesimal neighborhood of a point λ_0 of the bulk of the support of the limiting Normalized Counting Measure of eigenvalues, then we obtain

$$n\rho_n(\lambda_0)|\Delta_n| = O(1), \quad n \to \infty. \tag{1.2.19}$$

Thus, s in (1.2.12) can be viewed as the length of interval, "measured" in the typical spacing units.

The initial scale on which we have the weak convergence (1.2.9) is called *global* (or macroscopic), and the corresponding limiting and asymptotic results (e.g. (1.2.9) itself, (1.2.10), etc.) belong to the *global (or macroscopic) regime* of random matrix theory. The scale of the spectral axis, corresponding to intervals (1.2.12) (or (1.2.19)) is called the local (or microscopic) bulk scale and corresponding results (e.g. (1.2.13) – (1.2.16)) are in the *local bulk (or microscopic) regime*.

The local bulk regime can also be written in terms of the generating functionals E_n of (1.1.28). To this end we use n-dependent test functions (cf. (1.2.12))

$$\varphi_n(\lambda) = \varphi((\lambda - \lambda_0)n\rho_n(\lambda_0)), \tag{1.2.20}$$

where $\varphi : \mathbb{R} \to \mathbb{C}$ is a bounded piecewise continuous function of compact support.

We have from (1.2.7) in the local regime, i.e., for (1.2.20),

$$\lim_{n\to\infty} E_n[\varphi_n] = \exp\Big\{ - \int \varphi(x)dx \Big\}. \tag{1.2.21}$$

In particular, if $\varphi = g\chi_\Delta$, where χ_Δ is the indicator of $(0, s]$, then the r.h.s. of the last formula is e^{-gs}, the generating function of (1.2.13) according to (1.1.34).

Consider now the test function

$$\varphi_n = \varphi^{(1)}_n + \varphi^{(2)}_n, \qquad \varphi^{(1,2)}_n = \varphi_{1,2}((\lambda - \lambda_{1,2})n\rho_n(\lambda_{1,2})), \tag{1.2.22}$$

where $\varphi_{1,2}$ are bounded piecewise continuous functions of compact support, $\lambda_{1,2} \in$ bulk N, and $\lambda_1 \neq \lambda_2$. We obtain in this case

$$\lim_{n\to\infty} E_n[\varphi_n] = \exp\Big\{ -\int \varphi_1(x)dx\Big\} \exp\Big\{ -\int \varphi_2(x)dx\Big\}. \tag{1.2.23}$$

We conclude, in particular, that the random variables $\mathcal{N}_n[\varphi_n^{(1)}]$ and $\mathcal{N}_n[\varphi_n^{(2)}]$ tend in distribution to independent random variables, in particular, the numbers of eigenvalues $\mathcal{N}_n(\Delta_n^{(1)})$ and $\mathcal{N}_n(\Delta_n^{(2)})$ in the intervals $\Delta_n^{(1,2)} = (\lambda_{1,2}, \lambda_{1,2} + s_{1,2}/n\rho_n(\lambda_{1,2}))$, $\lambda_1 \neq \lambda_2$, are independent Poisson random variables in the limit $n \to \infty$ (cf. (1.2.14) – (1.2.15)).

One can also consider one more scaling of shrinking neighborhoods of points of bulk N. This can be seen from the analysis of the validity of the Central Limit Theorem for linear eigenvalue statistics (1.1.20). Assume for simplicity that F_n does not depend on n. Then we write the Laplace characteristic function (1.1.35) of the *centered linear statistics*

$$\mathcal{N}_n^\circ[\varphi] = \mathcal{N}_n[\varphi] - \mathbf{E}\{\mathcal{N}_n[\varphi]\} \tag{1.2.24}$$

as

$$Z_n[\varphi] := \mathbf{E}\{e^{-n^{-b}\mathcal{N}_n^\circ[\varphi]}\} = \left(\int e^{-n^{-b}\varphi^\circ(\lambda)} F(d\lambda)\right)^n, \tag{1.2.25}$$

where

$$\varphi^\circ(\lambda) = \varphi(\lambda) - \int \varphi(\mu)F(d\mu), \tag{1.2.26}$$

and we introduced the normalizing factor n^{-b}, $b > 0$, to be chosen later.

The standard probabilistic situation corresponds to an n-independent φ and $b = 1/2$, where

$$\lim_{n\to\infty} Z_n[\varphi] = \exp\{\mathbf{var}\{\varphi\}/2\} \tag{1.2.27}$$

is the Laplace transform of the Gaussian law and

$$\begin{aligned} \mathbf{var}\{\varphi\} =& \int \varphi^2(\mu)F(d\mu) - \left(\int \varphi(\mu)F(d\mu)\right)^2 \\ =& \frac{1}{2}\int\int (\varphi(\lambda) - \varphi(\mu))^2 F(d\lambda)F(d\mu). \end{aligned} \tag{1.2.28}$$

It is easy to check that

$$\mathbf{var}\{\varphi\} = \lim_{n\to\infty} \mathbf{Var}\{\mathcal{N}_n[\varphi]\}$$

(see (1.2.3)) in complete correspondence to the Central Limit Theorem of probability theory.

Consider now n-dependent test functions

$$\varphi_n(\lambda) = \varphi((\lambda - \lambda_0)n^a), \quad a > 0, \tag{1.2.29}$$

where φ is a piecewise continuous function of compact support. In this case an easy computation shows that if $a + 2b = 1$, then we have for $a > 0,\ b > 0$,

$$\mathbf{var}_{int}\{\varphi\} := \lim_{n\to\infty} \mathbf{Var}\{n^{-b}\mathcal{N}_n[\varphi_n]\} = \rho(\lambda_0)\int \varphi^2(\mu)F(d\mu), \tag{1.2.30}$$

(cf. (1.2.28)), and

$$\lim_{n\to\infty} Z_n[\varphi_n] = \exp\{\mathbf{var}_{int}\{\varphi\}/2\}. \tag{1.2.31}$$

We conclude that the one-parameter family of sequences

$$\left\{n^{-(1-a)/2}\mathcal{N}_n[\varphi_n(\cdot-\lambda_0)n^{-a}]\right\}_{n\geq 1}, \quad a\in(0,1), \tag{1.2.32}$$

of random variables converges in distribution to the Gaussian random variable of zero mean and variance (1.2.30). On the other hand, if $a = 1$ in (1.2.29), then it follows from (1.2.25) that $b = 0$ and

$$\lim_{n\to\infty} Z_n[\varphi_n] = \exp\Big\{-\int\Big(1-e^{-\varphi(x)}\Big)\,dx\Big\}. \tag{1.2.33}$$

This is obviously not the Laplace transform of the Gaussian law. In fact, this is the Laplace characteristic function of linear statistics of the Poisson point process. In particular, for $\varphi = \chi_{(0,s]}$, the indicator of $(0, s]$, the r.h.s. of the last formula is $\exp\{-(1-e^{-s})\}$, the Laplace transform of the Poisson distribution (1.2.16).

Formulas (1.2.27) and (1.2.33) correspond to the global and the local regimes. Their comparison with (1.2.31), which differs from both, makes reasonable the introduction of the *intermediate regime*, dealing with shrinking spectral intervals, whose rate (or length) L_n^{-1} is such that (cf. (1.2.18))

$$L_n\to\infty, \quad L_n/n\to 0, \quad n\to\infty. \tag{1.2.34}$$

The intermediate regime differs from the global regime in the form of the limiting law of fluctuation of linear statistics but has the same as the global regime limiting form (1.2.10) of the gap probability. This is easy to see from the limit of (1.2.5) in which Δ is replaced by $(\lambda_0, \lambda_0 + x/L_n)$.

There is one more important case where one has to use shrinking intervals whose rate is less than n^{-1}, corresponding to the local bulk regime. Assume that λ_0, belonging to the support of F, is such that $\rho(\lambda) = C|\lambda-\lambda_0|^\alpha(1+o(1))$, $\lambda\to\lambda_0$, $\alpha > -1$, $\alpha\neq 0$; hence, $\lambda_0\notin$ bulk N. Consider the intervals

$$\Delta_n = (\lambda_0, \lambda_0 + s/Bn^{1/(1+\alpha)}), \tag{1.2.35}$$

where $B = (C/(1+\alpha))^{1/(1+\alpha)}$. Then we obtain from (1.2.5)

$$\lim_{n\to\infty} E_n(\Delta_n) = e^{-s^{(1+\alpha)}}. \tag{1.2.36}$$

In particular, if $\alpha = 1/2$, we obtain the intervals

$$\Delta_n = (\lambda_0, \lambda_0 + s/Bn^{2/3}). \tag{1.2.37}$$

It will be shown in what follows that there exist analogs of the above scaling in random matrix theory. The scaling is the case in a neighborhood of endpoints of supp N and in the points of its interior, where the density of N is either zero or infinity. We will call all of them the *special points* .

We have seen that for diagonal random matrices with i.i.d. entries and for intervals of the form (1.2.12), i.e., in the bulk of the spectrum, the limiting gap probability is independent of the particular form of the probability distribution of the matrix and is the Poisson distribution. The same distribution is also pertinent for the much less trivial case of random differential and finite-difference operators of the second order **[360, 368, 360, 140, 227]** for λ_0 belonging to the bulk of their pure point spectrum. As for the gap probability in the bulk of the spectrum

of many random matrices, Gaussian Ensembles (GE) first of all, it is different from (1.2.16). It is given by Theorems 5.2.9, 7.5.4, 12.1.1 below and is known as the *Wigner-Dyson-Gaudin* distribution. The distribution is also independent of the form of the ensemble, at least in a certain class of cases. This is known as the universality of the local bulk regime of random matrix theory (see Comment 1.3.1).

If λ_0 is the right-hand endpoint of supp N, then $E_n(\Delta_n)$ is the probability law of the maximum eigenvalue of M_n of (1.2.1). The r.h.s. of (1.2.36) is the Weibull distribution of the extreme value theory of statistics. A similar question on the maximum eigenvalue of Gaussian matrices (1.1.1) – (1.1.3) corresponds to the scaling (1.2.37), leading to the Tracy-Widom distribution, different from (1.2.36) (see (5.3.6) – (5.3.7), (5.3.12) – (5.3.17)). In this case we are talking about the *local soft edge regime* (see Remark 5.3.4 *(1)* for the explanation of the term. The regime as well as other special point regimes is also universal (see Comment 1.3.1).

In applications of random matrix theory, dealing with the local bulk regime (nuclear physics, quantum and wave chaos studies, condensed matter theory, number theory) one uses a series of m energy levels (zeros of the ζ-function), such that

$$1 << m << n, \tag{1.2.38}$$

where n is the total number of energy levels (zeros of the ζ-function) under study, and one analyzes the empirical spacing density (histogram of level spacing frequencies) in the series. To model the analysis, we introduce the random variable

$$\eta_n(t,\Delta) = \sharp\{j : \lambda_j^{(n)} \in \Delta,\ \lambda_{j+1}^{(n)} - \lambda_j^{(n)} > t\}, \tag{1.2.39}$$

where Δ is an interval of the spectral axis. Denote by $\chi_{t,\lambda}(d\lambda)$ the indicator of the event of having an eigenvalue in $(\lambda - d\lambda, \lambda]$ and no eigenvalue in $(\lambda, \lambda + t]$. Then we have with probability 1 (a version of (1.1.53))

$$\eta_n(t,\Delta) = \int_\Delta \chi_{t,\lambda}(d\lambda); \tag{1.2.40}$$

hence,

$$\mathbf{E}\{\chi_{t,\lambda}(d\lambda)\} = \left(n \int_{(\mathbb{R}\setminus(\lambda,\lambda+t])^{n-1}} p_n(\lambda, \lambda_2, \dots, \lambda_n) d\lambda_2 \dots d\lambda_n \right) d\lambda$$

and by (1.1.49)

$$\begin{aligned} \mathbf{E}\{\eta_n(t,\Delta)\} &= n \int_\Delta d\lambda \int_{(\mathbb{R}\setminus(\lambda,\lambda+t])^{n-1}} p_n(\lambda, \lambda_2, \dots, \lambda_n) d\lambda_2 \dots d\lambda_n \\ &= \int_\Delta q_\lambda^{(n)}(t) n \rho_n(\lambda) d\lambda, \end{aligned}$$

where

$$q_\lambda^{(n)}(t) = \mathbf{P}\{t_\lambda^{(n)} > t\} \tag{1.2.41}$$

(cf. (1.1.51) – (1.1.52)) and $t_\lambda^{(n)}$ is the distance between an eigenvalue at λ and its nearest next neighbor to the right.

Likewise, we have

$$\begin{aligned}\mathbf{E}\{\eta_n^2(t,\Delta)\} &= \int_{\{(\lambda_1,\lambda_2)\in\Delta\times\Delta:\ |\lambda_1-\lambda_2|>t\}} \mathbf{E}\{\chi_{t,\lambda_1}(d\lambda_1)\chi_{t,\lambda_2}(d\lambda_2)\} \\ &\quad + \int_\Delta \mathbf{E}\{\chi_{t,\lambda}(d\lambda)\},\end{aligned}$$

where we took into account the relations

$$\chi_{t,\lambda}^2(d\lambda)=\chi_{t,\lambda}(d\lambda),\quad \chi_{t,\lambda_1}(d\lambda_1)\chi_{t,\lambda_2}(d\lambda_2)=0,\quad 0<|\lambda_1-\lambda_2|<t.$$

The above formulas imply that

$$\begin{aligned}\mathbf{Var}\{\eta_n(t,\Delta)\} &= \int_{\{(\lambda_1,\lambda_2)\in\Delta\times\Delta:\ |\lambda_1-\lambda_2|>t\}} C_t^{(n)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2 \\ &\quad - \int_{\{(\lambda_1,\lambda_2)\in\Delta\times\Delta:\ 0<|\lambda_1-\lambda_2|<t\}} C_t^{(n)}(\lambda_1)C_t^{(n)}(\lambda_2)d\lambda_1 d\lambda_2 \\ &\quad + \int_\Delta C_t^{(n)}(\lambda)d\lambda,\end{aligned} \tag{1.2.42}$$

where for $\delta_{1,2}=(\lambda_{1,2},\lambda_{1,2}+t]$

$$\begin{aligned}C_t^{(n)}(\lambda_1,\lambda_2) &= n(n-1)\int_{(\mathbb{R}\setminus(\delta_1\cup\delta_2))^{n-2}} p_n(\lambda_1,\lambda_2,\lambda_3,\dots,\lambda_n)d\lambda_3\dots d\lambda_n \\ &\quad -C_t^{(n)}(\lambda_1)C_t^{(n)}(\lambda_2),\end{aligned}$$

and

$$C_t^{(n)}(\lambda_1)=n\int_{(\mathbb{R}\setminus\delta_1)^{n-1}} p_n(\lambda_1,\lambda_2,\dots,\lambda_n)d\lambda_2\dots d\lambda_n.$$

This is valid for any collection $\{\lambda_l^{(n)}\}_{l=1}^n$ of symmetrically distributed (exchangeable) random variables with an absolutely continuous probability law. In the case of diagonal random matrices, i.e., i.i.d. $\{\lambda_l^{(n)}\}_{l=1}^n$ we have

$$\begin{aligned}C_t^{(n)}(\lambda_1,\lambda_2) &= n(n-1)\rho_n(\lambda_1)\rho_n(\lambda_2)(1-(F_1+F_2))^{n-2}-C_t^{(n)}(\lambda_1)C_t^{(n)}(\lambda_2), \\ C_t^{(n)}(\lambda_{1,2}) &= n\rho_n(\lambda_{1,2})\rho_n(\lambda_{1,2})(1-F_{1,2})^{n-1},\end{aligned}$$

where $F_{1,2}=F_n(\delta_{1,2})$.

We now choose

$$\lambda_0\in\text{bulk }N,\quad \sigma_n=s/n\rho_n(\lambda_0), \tag{1.2.43}$$

and interval Δ_n such that

$$\lambda_0\in\Delta_n,\quad |\Delta_n|\to 0,\quad n|\Delta_n|\to\infty \tag{1.2.44}$$

to model (1.2.38). This leads to the asymptotic relations as $n\to\infty$

$$\mathbf{E}\{\eta_n(t_n,\Delta_n)\} = n|\Delta_n|\rho(\lambda_0)e^{-s}(1+o(1)), \tag{1.2.45}$$

$$\mathbf{Var}\{\eta_n(t_n,\Delta_n)\} = O(n^2|\Delta_n|). \tag{1.2.46}$$

Thus

$$\mathbf{Var}\{\eta_n(t_n,\Delta_n)/n|\Delta_n|\rho(\lambda_0)\} = O(1/n|\Delta_n|). \tag{1.2.47}$$

We conclude that under the condition (1.2.44) the random variable

$$\eta_n(s/n\rho_n(\lambda_0),\Delta_n)/n|\Delta_n|\rho(\lambda_0) \tag{1.2.48}$$

converges in probability to the tail of the Poisson distribution.

In terms of statistics (1.2.45) shows that (1.2.48) is an asymptotically unbiased estimator for the tail of the normalized spacing distribution given by (1.2.41) for $\lambda = \lambda_0$ and $t_\lambda = s/n\rho(\lambda_0)$ (see also (1.1.51) – (1.1.52)), and (1.2.46) – (1.2.47) show that (1.2.48) converges in probability to e^{-s}, the tail of the spacing distribution for the Poisson point process.

This can be viewed as a justification of use of the histograms of (1.2.48) to find the asymptotic form of its distribution. Analogous justifications for random matrices related to classical compact groups (see Chapter 8) are given in **[454, 302]**.

We will study in what follows the counterparts of the facts, given above for the diagonal random matrices as well as a number of other facts, concerning the eigenvalue distribution of various random matrices. Namely, the convergence of the Normalized Counting Measure of eigenvalues of various ensembles of random matrices will be studied in Sections 2.2, 7.2, 10.1, and 11.1, the Central Limit Theorem will be considered in Sections 3.2, 7.3, 8.2, 8.4, 10.2, 11.3, 14.2, 18.4, and 19.2 and the local properties of the eigenvalue distribution both at the bulk and at the singular points of supp N will be studied in Chapters 5, 7, 9, 12, 13, 15, and 16.

1.3. Comments and Problems

Comment 1.3.1. Universality. Universality is an important topic of random matrix theory. The term seems to be borrowed from statistical mechanics, where the description of critical phenomena and phase transitions of the second kind in a small neighborhood of critical or transition points is independent of all but few properties of the model, such as the dimension of physical space, symmetry of interaction, ... As a result, upon an appropriate rescaling of relevant variables, in fact, a *certain scaling limit*, the description can be given in terms of "universal" or scaling functions (see e.g. **[525]**) and the transitions as different as the liquid-gas and ferromagnetic are described by the same scaling functions and are said to belong to the same *universality class*.

The terms *universality* and *universality classes* are used in other contexts as well (e.g. in classical and quantum dynamical systems, random walks with constraints, percolation theory, etc.), allowing one to separate generic properties from the system dependent features. In view of this it seems reasonable to use the terms in order to interpret the asymptotic (output) results of any field in which one encounters their considerable independence on the input entities, a kind of reduction of description (again a statistical mechanics term).

We mention here probability theory, where the Law of Large Numbers for the sum of i.i.d. random variables depends only on their first moment and the Central Limit Theorem depends only on their first two moments. One can also recall the invariance principle. On the other hand, the large deviation principle does not demonstrate a similar reduction of description, since in this case there is essentially one-to-one correspondence between asymptotic formulas and the probability laws via the Legendre transform (see e.g. formula (11.4.6) and **[163**, Section 2.2]). These observations seem evident if not trivial, but it is worth mentioning that in the 1950s Linnik extended them quite nontrivially and found a series of limit theorems which involve any given number of moments and were called by him the *collective theorems*, since "one collects in the same *class* an infinite number of sequences

of random variables, satisfying certain conditions and such that the appropriately normalized and centered deviations have the same form" **[282]**, thus, again classes.

In random matrix theory universality in its "narrow sense" concerns the local regimes, i.e., the asymptotic properties of eigenvalue distributions on the scale of typical distance between the adjacent eigenvalues (see e.g. formula (1.2.18)). This requires, as in statistical mechanics, a certain rescaling of relevant parameters from the scale fixed by the large size (thermodynamic or macroscopic) limit to finer scales, and this is why the local regimes are also often called the *scaling limits* or *microscopic limits*.

There is, however, a serious difference between the underlying mechanisms of the universality of critical phenomena in statistical mechanics and of the local regimes in random matrix theory. Indeed, in statistical mechanics the approach to the critical point results in a considerable increase of correlations for relevant observables (order parameters), so that the range of correlations becomes much larger than the microscopic scale, determined by the interaction. Thus, we can go *up* from the microscopic scale to a much bigger scale (in fact, for any scale whose order is just less than the macroscopic scale) on which the relevant observables do not already depend on the particularities of interaction, except its overall properties (dimension, symmetry, etc.); thus they are governed by universal laws, after normalizing certain parameters.

On the other hand, in random matrix theory the situation is in a way the opposite. This is simpler to explain by considering the case of the local bulk regime of Matrix Models (4.1.1) – (4.1.2). In this case we are looking for statistical properties of eigenvalues on the scale $O(n^{-1})$ of the typical distance between eigenvalues (see (1.2.18)); i.e., we go *down* to the smallest possible scale on which the asymptotic properties are well defined. Since the logarithm function that plays the role of the interaction in random matrix theory (see e.g (11.1.1) – (11.1.3)) is scale invariant and the potential is practically constant on this scale, the relevant properties of local bulk regime become independent of the potential, i.e., of a particular model.

An analogous argument can be applied to other local regimes of Matrix Models. Thus, in statistical mechanics terms, universality of the local regimes in random matrix theory is the independence of results on the external field on the microscopic scale, provided that the interaction is scale invariant and the same in all the models in question. Unfortunately, it seems not too evident how to turn the above heuristic argument into a rigorous proof. Besides, we cannot apply the argument to the Wigner matrices of Chapter 18 and the sample covariance type matrices of Chapter 19, since their joint probability density cannot be written in the Gibbs form (11.1.1) – (11.1.3). However, in view of recent results on universality of local regimes for these matrices (see Sections 18.7 and 19.3) it is tempting to believe that the underlying heuristics of universality in these cases is analogous.

It is widely believed and proved in many cases that provided that the scale is appropriately chosen the eigenvalue distribution in local regimes is independent of the details of the random matrix in question (i.e., the corresponding matrix probability law), except the group that reduces the matrix to the diagonal form, the number of nonzero entries relative to the matrix size, ... Thus, again following statistical mechanics, one can consider the corresponding universality classes, i.e., the set of random matrices having the same local regime. Most known are the *Gaussian universality classes* for the bulk and the edge local regimes, pertinent to

the Gaussian Ensembles (1.1.1) – (1.1.3). For other universality classes of the local regime see e.g. Sections 7.5 and 13.2 and references therein and below.

Note, however, that random matrix theory possesses various results, arising in other asymptotic regimes and having the same form for sufficiently broad classes of random matrices (recall the semicircle law, also determined by the first two moments of entries of the Wigner (see Theorem 18.3.2) and band matrices (see Problem 2.4.13). In view of this, the active development of the theory, and the expanding range of its applications, where it is a source of insights, models, and results, it seems useful to extend the notion of *universality* beyond the local regimes and to quantify it by considering *universality classes* for sets of random matrices, for which certain properties are the same in a given asymptotic regime.

We discuss now certain results of random matrix theory from the point of view of this, more broadly understood, concept of universality.

There are also related results on the limit theorems for determinantal point processes **[276, 457]**.

(1) Global (or macroscopic) regime. This regime is similar to the macroscopic limit of statistical mechanics and to the standard setting for limit theorems of probability theory.

It was mentioned above that the semicircle law (2.2.28) (more generally, the deformed semicircle law, defined by equation (2.2.4) for its Stieltjes transform) is valid for any real symmetric or hermitian Wigner matrix (18.1.1) with independent (modulo symmetry) entries, satisfying (18.1.2). Confining ourselves to matrices with zero expectation of entries, we find that the semicircle form of the limiting Normalized Counting Measure of eigenvalues is valid for an even broader class of band matrices (see Problem 2.4.13 and references therein). Thus, assuming that the variance of entries is 1 (i.e., just rescaling the spectral axis), we obtain the *semicircle universality class* with respect to the form of limiting Normalized Counting Measure of eigenvalues. For the deformed Wigner matrices (18.1.6), satisfying (18.1.2) and (18.3.10), we have the family of universality classes, indexed by the limiting Normalized Counting Measure of eigenvalues $N^{(0)}$ of "unperturbed matrices" $H^{(0)}$. It is also the case for the parametric limits of certain random operators (see Section 17.3).

An analogous situation is for the sample covariance type matrices (19.1.1) – (19.1.3) with $a^2 = 1$, where according to Theorem 19.1.10 the universality classes are indexed by parameter $c \in (0, \infty)$ of (7.2.9) and by limiting Normalized Counting Measures $N^{(0)}$ and σ of eigenvalues of matrices $H^{(0)}$ and T for the deformed version (19.1.6).

On the other hand, if the entries of the Wigner matrix have an infinite second moment (their probability laws have "heavy tails") or the entries are dependent (even being Gaussian), then the limiting Normalized Counting Measure is not the (deformed) semicircle law anymore. We refer the reader to **[19, 87, 90, 306, 313]** and Sections 11.1 and 17.2 for examples of random matrices with dependent entries, **[44, 63, 131]** for matrices with the "heavy tail" entries, **[522]** for matrices whose independent entries do not satisfy the condition of the uniform negligibility of tails, and **[69, 306, 309, 313, 367, 431]** and Problem 2.4.13 for the band and sparse matrices.

Likewise, in the case of Matrix Models (4.1.1), whose entries are always strongly dependent, the limiting Normalized Counting Measure of eigenvalues depends strongly on the model, i.e., on the potential. Indeed, it follows from the results of Sections 11.1 and 11.2, equation (11.2.6) in particular, that we have essentially a one-to-one correspondence between the density of the measure and the potential on the support of the measure, thus, practically no "reduction of description" while passing from the matrix law to the asymptotic result in this case.

Although we do not consider in this book the real (but not real symmetric or orthogonal) and complex (but not hermitian or unitary) matrices, it is worth noting that for them there exists an analog for the semicircle law, known as the *circular law*, according to which the limiting Normalized Counting Measure of eigenvalues in this case is the uniform measure on the unit disc if the entries of the matrix are i.i.d. random variables with zero mean and unit variance (see the review **[473]** and references therein).

The limiting Normalized Counting Measure of eigenvalues is the first one we find for any random matrix since the measure determines to a large extend other asymptotic spectral properties.

The above results are of the type of the Law of Large Numbers for linear eigenvalue statistics with bounded and continuous test functions. Passing to the fluctuations around the corresponding limits, hence to the Central Limit Theorem type results, we conclude in view of Theorems 5.2.7(iii) and 18.4.7 and their analogs for the hermitian matrices, assuming again without loss of generality that the variance of entries is one, that for linear eigenvalue statistics with C^1 test functions the Wigner matrices have two families of universality classes (for $\beta = 1, 2$, respectively), indexed by the fourth cumulant κ_4 of entries. Gaussian Ensembles (GOE and GUE) correspond to $\kappa_4 = 0$.

A similar situation is for the sample covariance matrices and the classical group ensembles; see Theorems 19.2.5, 8.4.3, and 8.4.5. In particular, according to Theorem 19.2.5 and its hermitian analog, there are two two-parameter families of universality classes with respect to the limiting laws of fluctuations of linear eigenvalue statistics with C^1 test functions, indexed by $\beta = 1, 2$, by $c \in (0, \infty)$ of (7.2.9), and by the fourth cumulant κ_4 of entries of the data matrix X of (19.1.1). The case $\kappa_4 = 0$ corresponds again to the classical ensembles, this time to the Wishart Ensemble and the Laguerre Ensemble.

Note that the fourth cumulant κ_4 of entries and/or their fourth moment seem to be important parameters for the Wigner and sample covariance type matrices. They also determine the large-n corrections for the mean of the linear eigenvalue statistics with smooth test functions, quite general bounds for the matrix norm **[332]**, and, although implicitly, the asymptotic behavior of extreme eigenvalues (see Theorem 18.6.1) and the local regime **[294, 477]**.

It is also worth noting that despite the coincidence of the above results on the limiting laws of fluctuations for linear eigenvalue statistics with C^1 test functions (or even with the Hardy class $H^{1/2}$ functions; see e.g. Theorems 8.4.3 and 8.4.5), i.e., results of the Central Limit Theorem type, there is a difference with respect to the sums of independent random variables. This is because the variance of linear eigenvalue statistics with regular test functions in the above sense is just bounded as $n \to \infty$, but not of order $O(n)$ as for i.i.d. random variables with finite variance (or at least tends to infinity as $n \to \infty$ for weakly dependent random variables

[**281**, Chapter 18]). As a result, the Central Limit Theorem for linear eigenvalue statistics is valid just upon its centering, i.e., the subtraction of its expectation, but without dividing the centered statistics by $n^{-1/2}$. This suggests, in view of the existence of the limiting Normalized Counting Measure of eigenvalues, implying that the eigenvalues are of order $O(1)$ as $n \to \infty$, that the asymptotic emergence of the Gaussian law may include certain cancellation effects between the terms of linear eigenvalue statistics. In certain cases the cancellation can be easily taken into account. Consider a simple example of the "linear" linear statistics of Wigner matrices, i.e., the sum $\lambda_1^{(n)} + \cdots + \lambda_n^{(n)}$. It follows from the spectral theorem and the basic formulas (18.1.1) – (18.1.3) that it is the sum $n^{-1/2}(W_{11} + \cdots + W_{nn})$; thus we are in the standard probabilistic situation and the Central Limit Theorem follows.

One can imagine, however, that the cancellation effects may not be sufficiently strong sometimes, especially in the case of strongly dependent entries, for instance, in the case of the Matrix Models (4.1.1). From this point of view it may seem a bit surprising that the Matrix Models with a single interval spectrum also belong to the "Gaussian" universality class with respect to the validity of the Central Limit Theorem for linear eigenvalue statistics with C^1 and even less regular test functions (see **[288]** and example (i) of Section 14.2.2). However, for the hermitian Matrix Models with a q-interval ($q \geq 2$) spectrum the situation is more complex. Here, first of all, the asymptotic probability laws of linear eigenvalue statistics depends on the subsequence $\{n_j\}$ of the matrix sizes and thus the universality classes are indexed by the points of a certain closure $\mathbb{H}_\beta^{q-1}$ of sequences (see (14.1.52)), by the edges of the spectrum, and by the variational derivatives (14.2.30) of masses $\{\beta_l\}_{l=1}^{q-1}$ of (14.1.42) with respect to the potential. If the derivatives are zero, for instance, if $\{\beta_l\}_{l=1}^{q-1}$ are rationally dependent (see e.g. Corollary 11.2.9 and (14.2.34)), then the limiting law is the generalized Central Limit Theorem of (14.2.23), but if not, then we have different limiting laws (see e.g. example (iii) of Section 14.2.2).

Note also that in the case of hermitian Matrix Models the universality classes with respect to the variance of linear eigenvalue statistics with C^1 test functions are different from those for the corresponding limiting laws of fluctuations of these statistics. Indeed, according to (14.2.10) – (14.2.12), the variance is determined by the edges of the spectrum and by $x \in \mathbb{H}_\beta^{q-1}$; thus the universality classes are indexed by a smaller number of parameters than that for the limiting laws.

It is worth mentioning that the variance and the Central Limit Theorem for the Counting Measure of the GUE matrices have the same form as those for the classical group ensembles, the Haar distributed unitary matrices in particular (see Theorem 3.1.1(ii) and Problem 5.4.14) for the GUE and Theorem 8.4.9 and Remark 9.1.11*(3)* for the classical group ensembles, respectively). This suggests the existence of the universality class for linear eigenvalue statistics corresponding to the piecewise constant test functions. The class has to include the GUE, the classical group ensembles of Chapter 8, and the unitary matrix models of Chapter 16, all of them possessing the unitary symmetry. Likewise, in view of Theorem 9.2.6(ii) there should be one more universality class for which the variance is twice more than that of the previous class and which has to include the Gaussian Orthogonal Ensemble, the Circular Orthogonal Ensemble, and other ensembles possessing the orthogonal symmetry, say, the real symmetric Matrix Models of (4.1.1) with

$\beta = 1$, and their analogs for unitary symmetric matrices, i.e., the analogs (16.1.1) for unitary symmetric matrices.

The global regime results have a number of applications, in particular, in combinatorics ([**333**]), functional analysis ([**150, 260, 263, 505**]), multivariate statistics ([**26, 233, 432**] and Comment 7.6.1), quantitative finances ([**84**]), and telecommunication theory ([**492**]).

(2) Intermediate (mesoscopic) regime. We discuss here the universality classes for the variance of linear eigenvalue statistics. The meaning of the regime is explained in Section 1.2 by using the toy model of diagonal random matrices, i.e., in fact, i.i.d. random variables. The corresponding scale is intermediate between those of the global and the bulk local regimes (see e.g. (5.2.25)). In this case, according to formulas (5.2.24) – (5.2.27) and (14.3.8), we have just one universality class for all hermitian Matrix Models, the GUE in particular. Analogous results are valid for the classical group ensembles (see Theorems 9.1.10 and 9.2.5(iii)). These asymptotic results match those for the zero scale (with respect to the intermediate) limit of the global regime results and the infinite scale (with respect to the intermediate) limit of the local regime results (see Remark 5.2.8*(4)* and Section 14.3).

The above concerns the random matrix ensembles with the unitary symmetry. In view of Theorem 9.2.5(iii) the random matrix ensembles with orthogonal symmetry have to comprise another universality class for which the variance is twice more than that of the previous class and which has to include the Gaussian Orthogonal Ensemble, the Circular Orthogonal Ensemble, the real symmetric Matrix Models of (4.1.1) with $\beta = 1$, and their analogs for unitary symmetric matrices, i.e., the analogs (16.1.1) for unitary symmetric matrices.

There are also related results on the determinantal point processes [**276, 457**] and the strong Szegö theorem for the *Toeplitz determinants* [**38**].

The intermediate regime is useful in explaining the universal conductance fluctuations of small metallic particles [**41, 252**].

(3) Local (microscopic) regimes.

Local bulk regime. This regime concerns the points of the spectrum (the support of the limiting Normalized Counting Measure of eigenvalues) at which the density of the measure (assumed to exist) is continuous and not zero, i.e., for the *bulk* of the spectrum; see Definition 1.2.1. It follows then from (1.2.18) that the regime is about the eigenvalue distribution in the $O(n^{-1})$ neighborhood of internal points of the spectrum. There is a general belief that (under reasonable hypotheses) the limiting distribution is universal, hence coincides with that for the classical ensembles of the same symmetry. The belief is essentially due to Wigner and Dyson and is often called the *universality conjecture*. According to the conjecture there exists a unique universality class for each of three basic sets (in fact spaces) of matrices: real symmetric, hermitian, and quaternion real (the "three fold way" by Dyson [**183**]) and the same for the three sets (in fact symmetric spaces) of unitary matrices (see Section 9.2 and [**356**]). They are often called the Wigner-Dyson-Gaudin classes, since Gaudin was the first to find explicit formulas for the gap probability (known now as the *sin*-kernel Fredholm determinant) for the reference Gaussian Unitary Ensemble. Note that in the last decades other reference Gaussian Ensembles were found [**526**].

It was a remarkable idea by Wigner [**518, 521**] to use large random matrices, i.e., the statistical approach, in order to describe rather complex spectra of heavy nuclei, i.e., to replace a complex object by a random one, as was done more than once in the history of probability theory. Thus, one has to find large random matrices, somehow related to the original complex infinite-dimensional operators, and to look for the statistical description of their spectra, however in range much smaller than the characteristic magnitude of eigenvalues (energy levels) of the operators in question, i.e., locally.

The genesis of Wigner's idea on the statistical approach can also be seen from his earlier works on the statistical theory of nuclear reactions (see e.g. [**71, 107, 252, 416**]). Likewise, the idea on the energy level repulsion that was the initial principal motivation of Wigner's interest in the problem (the repulsion was experimentally discovered at the end of the 1940s and was important to explain in terms of vanishing of the level spacing density at origin) goes back to his work with von Neumann of 1929 on the crossing of energy levels of quantum mechanical Hamiltonians depending on parameters (see e.g. [**331**, Section 79] and [**21**, Section 79]). Nowadays the eigenvalue repulsion (see e.g. formulas (5.2.58), (6.2.18), and (6.2.19)), more generally, the rigidity of the spectrum, since the level spacing density vanishes also at infinity (see (5.2.75) and (6.2.23)), is among the central topics of random matrix theory and its applications. This is because it has been found in various studies that the statistics of spectra of many complex quantum and wave systems, more generally, many sufficiently complex sequences of numbers, including the zeros of the Riemann ζ-function, exhibit a remarkable universality, provided that one is interested in numbers falling into an interval sufficiently small compared with the magnitude of numbers in question and is able to resolve the scale down to "local" scale given by the typical spacings between the numbers.

The existence of the eigenvalue repulsion was justified by explicit calculations by Gaudin, Mehta, and by Dyson in the early 1960s for the invariant Gaussian and Circular Ensembles of random matrices (see [**356**] and Sections 5.2, 6.2, 9.2 of this book). It was also found that the resulting expressions for the correlation functions and the gap probability for three Gaussian Ensembles, whose spectra are on the real axis, and for the corresponding three Circular Ensembles, whose spectra are on the unit circle, coincide in the local bulk limit, thereby suggesting its universality.

As for establishing various forms of universality of local regimes in random matrix theory, i.e., the independence of results by Gaudin, Mehta, and Dyson on the random matrix in question, it has been demonstrated in a number of physics papers during the 1960s – 1990s and the process of the rigorous proof of universality for various classes of random matrices started in the middle of the 1990s and is quite active till now. This provides a justification for the use of random matrices in the description of spectra of a wide variety of complex systems [**68, 71, 107, 264, 511**], transport phenomena in disordered solids [**16, 41, 191, 206**], various results and conjectures in number theory [**302, 358**]. Large random matrices allow one to reproduce quantitatively a number of important properties of the corresponding complex sets of numbers, independent of their specific aspects and origins.

The limiting form of correlation functions and their generating functional, the gap probability in particular, dating back to works by Gaudin, Mehta, and Dyson (see e.g. [**356**] for the corresponding references), are given by Theorems 5.2.9 and 6.2.1 for the hermitian and real symmetric matrices, respectively, and are

expressed via the trigonometric functions. This can be called the strong form of the conjecture, or the *strong universality*, since it requires uniform convergence on compact intervals of correlation functions and the gap probability (and the corresponding convergence of the point process, determined by them).

In this form the universality conjecture for the bulk local regime is proved by now for the large class of Matrix Models (see **[152, 154, 162]** and Chapters 12 and 15 of this book, Theorems 12.1.1 and 15.3.1 in particular, and references therein). The conjecture is also proved for the Laguerre Ensemble (see Theorem 7.5.4), deformed Gaussian Unitary and Laguerre Ensembles (see Subsections 18.7.4 and 19.3.2 for discussions and references), the Laguerre-type ensembles of real symmetric, hermitian, and quaternion real matrices **[161]**, and for classical group ensembles (see **[167]**, **[302**, Chapter 1] and Sections 9.1.3, 9.2.3, and 16.2 of this book).

There are also recent results by Erdös et al. **[199]** and Tao and Vu **[474, 476]** (see also Sections 18.7 and 19.3) on the validity of the universality conjecture for the Wigner and sample covariance matrices, where one uses the weak convergence of correlation functions, i.e., upon integrating them with smooth functions of compact support with respect to the rescaled eigenvalues $\{x_j\}_{j=1}^{l}$ and even the location λ_0 of the $O(n^{-1})$-window (see e.g. formulas (5.2.52) and (12.1.1) for their definition). Note that the integration with respect to $\{x_j\}_{j=1}^{l}$ is indispensable in the case of discrete distributed entries of the matrices in question.

Note that the local bulk regime and the corresponding universality classes are also the case beyond random matrix theory, for instance, in the determinantal point processes, their discrete analogs, and related combinatorial and statistical mechanics problems **[76, 102, 214, 213, 276, 292, 324, 381, 383, 384, 385, 457]**.

Local regime for special points. Recall that the special points of the spectrum are those at which the density of the measure is either zero or infinity, thus requiring another scaling than that for the local bulk regime (see e.g. (1.2.18)). The most common is the case of generic edges of the spectrum, at which its density has either square root zero (soft edges; see Remark 5.3.4*(1))* or square root infinity (hard edges; see Remark 7.5.9); hence the appropriate scales are $O(n^{-2/3})$ and $O(n^{-2})$ according to the criterion (1.2.18). The reference ensembles are the Gaussian Ensembles, where both edges are soft because of the semicircle law (2.2.28) and the Wishart and Laguerre Ensembles, where for $c = 1$ the lower edge of support is hard because of the quarter-circle law (7.2.39). The form of limiting correlation functions and the gap probabilities are given by Theorems 5.3.3, 6.3.1, and 7.5.8 (see also **[217**, Chapters 8 and 14] and references therein) and are expressed via the Airy and Bessel functions, respectively.

The soft edge gap probability, in particular, the maximum eigenvalue distribution for Gaussian Ensembles was found and extensively studied by Tracy and Widom **[485, 488, 489]**. While the local bulk regime is widely used to describe the spectra of complex operators (the corresponding field is often called the quantum and wave chaos studies) and other interesting number sequences, the local edge regimes prove to be important for a wide variety of interesting and seemingly unrelated problems of combinatorics, integrable systems, quantum field theory, large order groups and their representations, random point processes, random walks with

constraints, one-dimensional particle systems, and multivariate statistics. We refer the reader to review works [**153, 169, 213, 292, 296, 324, 352, 462, 491**].

The universality classes for generic soft and hard edges include the deformed Gaussian and Laguerre Ensembles [**294, 45**], the Matrix Models (see Chapter 13 and [**162**, Part 2]), the modified Jacobi Ensembles [**327**], and the Wigner and sample covariance matrices under various conditions on their entries (see the review [**199**] and Sections 18.7 and 19.3).

Note that in the case of hermitian Matrix Models with real analytic potential one can tune the potential to obtain nongeneric soft edges where the density of the limiting Normalized Counting Measure has the zero of order $2k+1/2,\ k \in \mathbb{N}$, hence a nongeneric local soft edge regimes for $k > 0$. These regimes arise in certain toy models of gauge field theory and string theory (see e.g. [**55, 95, 148, 164, 169, 212**]) as the result of the so-called *double scaling limit*. It has this name because the corresponding potentials, known also as the multi-critical potentials, are obtained there by the simultaneous n-dependent tuning of the potential parameters and the spectral parameter. There are a number of interesting links of the regimes with integrable systems, Painlevé transcendents, topology, statistical mechanics of surfaces, etc. (see e.g. [**133, 208, 250, 333**] for recent results and references).

Other cases of special points are given by internal zeros of the density. Such zeros are absent in classical ensembles. On the other hand, in the case of Matrix Models even the simplest nontrivial quartic potential produces such a point (see (11.2.53), and Theorem 11.2.7 for polynomial potentials, producing an arbitrary number of internal zeros). If the potential is real analytic, then the generic internal zero is of multiplicity 2 and leads to an interesting case of the local regime, hence to one more universality class, described in Section 13.2 and references therein. In all these cases of special points the corresponding local regime is expressed via special solutions of the *Painlevé equations*, playing the role of the Airy and Bessel functions of generic soft and hard edges.

Other local special point regimes for various versions of hermitian matrix models are treated in [**53, 132, 135, 136, 366**].

A different universality class appears at an interior zero of the density of the limiting Normalized Counting Measure of eigenvalues of the deformed Gaussian Ensemble. Here a generic internal zero of the density (at the closure of the gap) is of order 1/3 (see Problem 2.4.16), and the local regime is described via so-called Pearcey functions and corresponding point processes, appearing also in related problems [**67, 98, 385, 490**].

One more local edge universality class describes the soft edge regime of band matrices in the case where the band width b_n of matrices grows sufficiently slow as $n \to \infty$ (see [**452**] and Section 18.7.2). The class is of interest for several reasons, in particular, because it requires a more sophisticated heuristic criterion than (1.2.18), suggesting the scaling $n^{-4/5}$ for $b_n/n^{5/6} \to 0$ instead of $n^{-2/3}$ for $b_n/n^{5/6} \to \infty$ for the generic soft edge in particular (recall that the density of the limiting Normalized Counting Measure of band matrices is the semicircle for any rate of $b_n \to \infty$ [**69, 367**] and Problem 2.4.13). Correspondingly, one has the generic local soft edge in the latter case and a new regime in the former.

The criterion is a version of the Thouless criterion in localization theory [**339, 226**], the theoretical physics counterpart of the spectral theory of random Schrödinger operators. Its use demonstrates the role of the number of nonzero entries of random matrices in determining the corresponding universality class. It is worth noting that according to the criterion one expects a similar crossover in the local bulk regime of band matrices: the standard Wigner-Dyson-Gaudin class if $b_n/n^{1/2} \to \infty$ and the Poisson local regime (1.2.12)–(1.2.16) otherwise [**124, 225, 443**]. Note that the same regime is the case for the pure point spectrum of the random Schrödiger operator [**360, 140, 227**], having nonzero entries only on the principal and the adjacent diagonals, while for the absolutely continuous spectrum one expects the standard local bulk regime (Wigner-Dyson-Gaudin class) or, possibly, another non-Poissonian regime [**25**].

PROBLEM 1.3.2. Let M be a real symmetric (hermitian) random matrix whose probability law is absolutely continuous with respect to the density d_1M of (1.1.1) (d_2M of (1.1.2)). Prove that the eigenvalues of M are distinct with probability 1.

Hint. If (1.1.19) are eigenvalues of M, then the product $\prod_{1\le j<k\le n}(\lambda_j^{(n)} - \lambda_k^{(n)})^2$ is a polynomial in matrix entries.

PROBLEM 1.3.3. Let G be an $n \times n$ random matrix whose entries are i.i.d. Gaussian random variables of zero mean and variance w^2. Then $(G+G^T)(2n)^{-1/2}$ is the GOE matrix, defined by (1.1.3) with $\beta = 1$ and (1.1.1). Likewise, if $G = \{G_{jk}\}_{j,k=1}^n$ is an $n \times n$ random matrix, whose entries are complex i.i.d. Gaussian random variables such that $\mathbf{E}\{G_{jk}\} = \mathbf{E}\{G_{jk}^2\} = 0$, $\mathbf{E}\{|G_{jk}|^2\} = w^2$, then $(G + G^*)(2\sqrt{n})^{-1}$ is the GUE matrix, defined by (1.1.3) with $\beta = 2$ and (1.1.2).

PROBLEM 1.3.4. Show that the Gaussian distribution (1.1.8) can be obtained as the minimizer of the "entropy" functional

$$-\int p(M)\log p(M) d_\beta M$$

in the class of probability measures $P(d_\beta M) = p(M)d_\beta M$ on the spaces $\mathcal{S}_n$ of real symmetric or $\mathcal{H}_n$ of hermitian matrices absolute continuous with respect to (1.1.1) or (1.1.2) and satisfying the conditions

$$\mathbf{E}\{\operatorname{Tr} M\} = \alpha/\gamma, \quad \mathbf{E}\{\operatorname{Tr} M^2\} = 1/\gamma.$$

This can be interpreted as the fact that the Gaussian distribution (1.1.8) corresponds to the minimum of associated information subject to the above constraints.

PROBLEM 1.3.5. Prove that the measures $d_\beta M$ of (1.1.1) – (1.1.2) are invariant with respect to the change $M \to OMO^T$, $\beta = 1$ and $M \to UMU^*$, $\beta = 2$.

Hint. Observe that the Jacobian of this change of variables does not depend on M and consider the integral

$$\int_{\mathcal{S}_n} e^{-\operatorname{Tr} M^2} d_1 M$$

and the analogous integral over $\mathcal{H}_n$ to conclude that the Jacobian is 1. For another proof see [**370**, Theorem 2.1.6].

PROBLEM 1.3.6. Let $\{A_n\}_{n\geq 1}$ be an infinite matrix sequence, A_n being an $n \times n$ matrix. Assume that the sequence $\{N_{A_n}\}_{n\geq 1}$ of the Normalized Counting Measures of their eigenvalues converges weakly to a probability measure N. Prove that if $\{B_n\}_{n\geq 1}$ is a sequence of $n \times n$ hermitian matrices and the rank r_n of B_n is such that $\lim_{n\to\infty} r_n/n = 0$, then the sequence $\{N_{A_n+B_n}\}_{n\geq 1}$ of the Normalized Counting Measures of eigenvalues of $A_n + B_n$ converges weakly to N.

Hint. Use the min-max principle of linear algebra, implying that

$$|N_{A_n+B_n} - N_{A_n}| \leq r_n/n.$$

PROBLEM 1.3.7. Let $\mathcal{N}_n(\Delta)$ and $R_l^{(n)}$ be defined in (1.1.17) and (1.1.32). Prove the formula

$$\begin{aligned}\mathbf{E}\{(\mathcal{N}_n(\Delta))(\mathcal{N}_n(\Delta)-1)\dots(\mathcal{N}_n(\Delta)-l+1)\}\\ = \int_{\Delta^k} R_l^{(n)}(\lambda_1,\dots,\lambda_l)d\lambda_1\dots d\lambda_l.\end{aligned}$$

PROBLEM 1.3.8. Prove that the rarefication procedure described by (1.1.60) – (1.1.61) does not change the local regime of diagonal random matrices of Section 1.2 (i.e., the Poisson point process).

Hint. Use (1.1.61) and (1.2.21).