1

INTRODUCTION

Random matrices first appeared in mathematical statistics in the 1930s but did not attract much attention at the time. In the theory of random matrices one is concerned with the following question. Consider a large matrix whose elements are random variables with given probability laws. Then what can one say about the probabilities of a few of its eigenvalues or of a few of its eigenvectors? This question is of pertinence for the understanding of the statistical behaviour of slow neutron resonances in nuclear physics, where it was proposed in the 1950s and intensively studied by the physicists. Later the question gained importance in other areas of physics and mathematics, such as the characterization of chaotic systems, elastodynamic properties of structural materials, conductivity in disordered metals, the distribution of the values of the Riemann zeta function on the critical line, enumeration of permutations having certain particularities, counting of certain knots and links, quantum gravity, quantum chromo dynamics, string theory, and others (cf. *J. Phys. A* **36** (2003), special issue: random matrices). The reasons of this pertinence are not yet clear. The impression is that some sort of a law of large numbers is in the back ground. In this chapter we will try to give reasons why one should study random matrices.

1.1 Random Matrices in Nuclear Physics

Figure 1.1 shows a typical graph of slow neutron resonances. There one sees various peaks with different widths and heights located at various places. Do they have any definite statistical pattern? The locations of the peaks are called nuclear energy levels, their widths the neutron widths and their heights are called the transition strengths.

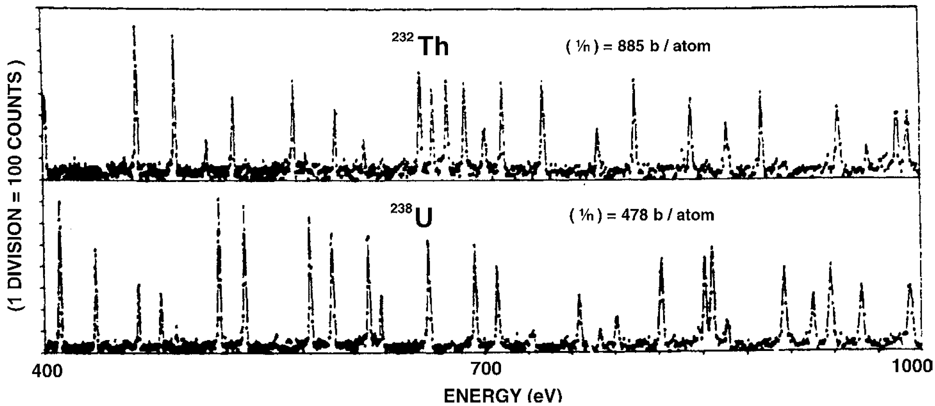

Figure 1.1. Slow neutron resonance cross-sections on thorium 232 and uranium 238 nuclei. Reprinted with permission from The American Physical Society, Rahn et al., Neutron resonance spectroscopy, X, *Phys. Rev. C* 6, 1854–1869 (1972).

The experimental nuclear physicists have collected vast amounts of data concerning the excitation spectra of various nuclei such as shown on Figure 1.1 (Garg et al., 1964, where a detailed description of the experimental work on thorium and uranium energy levels is given; (Rosen et al., 1960; Camarda et al., 1973; Liou et al., 1972b). The ground state and low lying excited states have been impressively explained in terms of an independent particle model where the nucleons are supposed to move freely in an average potential well (Mayer and Jensen, 1955; Kisslinger and Sorenson, 1960). As the excitation energy increases, more and more nucleons are thrown out of the main body of the nucleus, and the approximation of replacing their complicated interactions with an average potential becomes more and more inaccurate. At still higher excitations the nuclear states are so dense and the intermixing is so strong that it is a hopeless task to try to explain the individual states; but when the complications increase beyond a certain point the situation becomes hopeful again, for we are no longer required to explain the characteristics of every individual state but only their average properties, which is much simpler.

The statistical behaviour of the various energy levels is of prime importance in the study of nuclear reactions. In fact, nuclear reactions may be put into two major classes—fast and slow. In the first case a typical reaction time is of the order of the time taken by the incident nucleon to pass through the nucleus. The wavelength of the incident nucleon is much smaller than the nuclear dimensions, and the time it spends inside the nucleus is so short that it interacts with only a few nucleons inside the nucleus. A typical example is the head-on collision with one nucleon in which the incident nucleon hits and ejects a nucleon, thus giving it almost all its momentum and energy. Consequently in such cases the coherence and the interference effects between incoming and outgoing nucleons is strong.

Another extreme is provided by the slow reactions in which the typical reaction times are two or three orders of magnitude larger. The incident nucleon is trapped and all its energy and momentum are quickly distributed among the various constituents of the target nucleus. It takes a long time before enough energy is again concentrated on a single nucleon to eject it. The compound nucleus lives long enough to forget the manner of its formation, and the subsequent decay is therefore independent of the way in which it was formed.

In the slow reactions, unless the energy of the incident neutron is very sharply defined, a large number of neighboring energy levels of the compound nucleus are involved, hence the importance of an investigation of their average properties, such as the distribution of neutron and radiation widths, level spacings, and fission widths. It is natural that such phenomena, which result from complicated many body interactions, will give rise to statistical theories. We shall concentrate mainly on the average properties of nuclear levels such as level spacings.

According to quantum mechanics, the energy levels of a system are supposed to be described by the eigenvalues of a Hermitian operator H, called the Hamiltonian. The energy level scheme of a system consists in general of a continuum and a certain, perhaps a large, number of discrete levels. The Hamiltonian of the system should have the same eigenvalue structure and therefore must operate in an infinite-dimensional Hilbert space. To avoid the difficulty of working with an infinite-dimensional Hilbert space, we make approximations amounting to a truncation keeping only the part of the Hilbert space that is relevant to the problem at hand and either forgetting about the rest or taking its effect in an approximate manner on the part considered. Because we are interested in the discrete part of the energy level schemes of various quantum systems, we approximate the true Hilbert space by one having a finite, though large, number of dimensions. Choosing a basis in this space, we represent our Hamiltonians by finite dimensional matrices. If we can solve the eigenvalue equation

$$H\Psi_i = E_i\Psi_i,$$

we shall get all the eigenvalues and eigenfunctions of the system, and any physical information can then be deduced, in principle, from this knowledge. In the case of the nucleus, however, there are two difficulties. First, we do not know the Hamiltonian and, second, even if we did, it would be far too complicated to attempt to solve the corresponding equation.

Therefore from the very beginning we shall be making statistical hypotheses on H, compatible with the general symmetry properties. Choosing a complete set of functions as basis, we represent the Hamiltonian operators H as matrices. The elements of these matrices are random variables whose distributions are restricted only by the general symmetry properties we might impose on the ensemble of operators. And the problem is to get information on the behaviour of its eigenvalues. "The statistical theory will not predict the detailed level sequence of any one nucleus, but it will describe the general appearance and the degree of irregularity of the level structure that is expected to occur in any nucleus which is too complicated to be understood in detail" (Dyson, 1962a).

In classical statistical mechanics a system may be in any of the many possible states, but one does not ask in which particular state a system is. Here we shall renounce knowledge of the nature of the system itself. "We picture a complex nucleus as a black box in which a large number of particles are interacting according to unknown laws. As in orthodox statistical mechanics we shall consider an ensemble of Hamiltonians, each of which could describe a different nucleus. There is a strong logical expectation, though no rigorous mathematical proof, that an ensemble average will correctly describe the behaviour of one particular system which is under observation. The expectation is strong, because the system might be one of a huge variety of systems, and very few of them will deviate much from a properly chosen ensemble average. On the other hand, our assumption that the ensemble average correctly describes a particular system, say the U^{239} nucleus, is not compelling. In fact, if this particular nucleus turns out to be far removed from the ensemble average, it will show that the U^{239} Hamiltonian possesses specific properties of which we are not aware. This, then will prompt one to try to discover the nature and the origin of these properties" (Dyson, 1962b).

Wigner was the first to propose in this connection the hypothesis alluded to, namely that the local statistical behaviour of levels in a simple sequence is identical with the eigenvalues of a random matrix. A simple sequence is one whose levels all have the same spin, parity, and other strictly conserved quantities, if any, which result from the symmetry of the system. The corresponding symmetry requirements are to be imposed on the random matrix. There being no other restriction on the matrix, its elements are taken to be random with, say, a Gaussian distribution. Porter and Rosenzweig (1960a) were the early workers in the field who analyzed the nuclear experimental data made available by Harvey and Hughes (1958), Rosen et al. (1960) and the atomic data compiled by Moore (1949, 1958). They found that the occurrence of two levels close to each other in a simple sequence is a rare event. They also used the computer to generate and diagonalize a large number of random matrices. This Monte Carlo analysis indicated the correctness of Wigner's hypothesis. In fact it indicated more; the density and the spacing distribution of eigenvalues of real symmetric matrices are independent of many details of the distribution of individual matrix elements. From a group theoretical analysis Dyson found that an irreducible ensemble of matrices, invariant under a symmetry group G, necessarily belongs to one of three classes, named by him orthogonal, unitary and symplectic. We shall not go into these elegant group theoretical arguments but shall devote enough space to the study of the circular ensembles introduced by Dyson. As we will see, Gaussian ensembles are equivalent to the circular ensembles for large orders. In other words, when the order of the matrices goes to infinity, the limiting correlations extending to any finite number of eigenvalues are identical in the two cases. The spacing distribution, which depends on an infinite number of correlation functions, is also identical in the two cases. It is remarkable that standard thermodynamics can be applied to obtain certain results which otherwise require long and difficult analysis to derive. A theory of Brownian motion of matrix elements has also been created by Dyson (1962b) thus rederiving a few known results.

Various numerical Monte Carlo studies indicate, as Porter and Rosenzweig (1960a) noted earlier, that a few level correlations of the eigenvalues depend only on the overall symmetry requirements that a matrix should satisfy and they are independent of all other details of the distribution of individual matrix elements. The matrix has to be Hermitian to have real eigenvalues, the diagonal elements should have the same distribution and the off-diagonal elements should be distributed symmetrically about the zero mean and the same mean square deviation for all independent parts entering in their definition. What is then decisive is whether the matrix is symmetric or self-dual or something else or none of these. In the limit of large orders other details are not seen. Similarly, in the circular ensembles, the matrices are taken to be unitary to have the eigenvalues on the circumference of the unit circle, what counts then is whether they are symmetric or self-dual or none of these. Other details are washed out in the limit of large matrices.

This independence is expected; but apart from the impressive numerical evidence, some heuristic arguments of Wigner and the equivalence of Gaussian and circular ensembles, no rigorous derivation of this fact has yet been found. Its generality seems something like that of the central limit theorem.

1.2 Random Matrices in Other Branches of Knowledge

The physical properties of metals depend characteristically on their excitation spectra. In bulk metal at high temperatures the electronic energy levels lie very near to one another and are broad enough to overlap and form a continuous spectrum. As the sample gets smaller, this spectrum becomes discrete, and as the temperature decreases the widths of the individual levels decrease. If the metallic particles are minute enough and at low enough temperatures, the spacings of the electronic energy levels may eventually become larger than the other energies, such as the level widths and the thermal energy, kT. Under such conditions the thermal and the electromagnetic properties of the fine metallic particles may deviate considerably from those of the bulk metal. This circumstance has already been noted by Fröhlich (1937) and proposed by him as a test of the quantum mechanics. Because it is difficult to control the shapes of such small particles while they are being experimentally produced, the electronic energy levels are seen to be random and the theory for the eigenvalues of the random matrices may be useful in their study.

In the mathematical literature the Riemann zeta function is quite important. It is suspected that all its non-real zeros lie on a line parallel to the imaginary axis; it is also suspected (Montgomery, 1974) that the local fluctuation properties of these zeros on this line are identical to those of the eigenvalues of matrices from a unitary ensemble.

Random matrices are also encountered in other branches of physics. For example, glass may be considered as a collection of random nets, that is, a collection of particles with random masses exerting random mutual forces, and it is of interest to determine the distribution of frequencies of such nets (Dyson, 1953). The most studied model of glass is the so called random Ising model or spin glass. On each site of a 2- or 3-dimensional regular lattice a spin variable σ_i is supposed to take values $+1$ or -1, each with a probability equal to $1/2$. The interaction between neighboring spins σ_i and σ_j is $J_{ij}\sigma_i\sigma_j$, and that between any other pair of spins is zero. If $J_{ij} = J$ is fixed, we have the Ising model, whose partition function was first calculated by Onsager (cf. McCoy and Wu, 1973). If J_{ij} is a random variable, with a symmetric distribution around zero mean, we have the random Ising model, the calculation of the partition function of which is still an open problem.

A problem much studied during the 1980s is that of characterizing a chaotic system. Classically, a system is chaotic if small differences in the initial conditions result in large differences in the final outcome. A polygonal billiard table with incommensurate sides and having a circular hole inside is such an example; two billiard balls starting on nearby paths have diverging trajectories. According to their increasing chaoticity, systems are termed classically as ergodic, mixing, a K-system or a Bernoulli shift. Quantum mechanically, one may consider, for example, a free particle confined to a finite part of the space (billiard table). Its possible energies are discrete; they are the eigenvalues of the Laplace operator in the specified finite space. Given the sequence of these discrete energy values, what can one say about its chaoticity; whether it is ergodic or mixing or . . .; i.e. is there any correspondence with the classical notions? A huge amount of numerical evidence tells us that these energies behave as if they were the eigenvalues of a random matrix taken from the Gaussian orthogonal ensemble (GOE). Finer recognition of the chaoticity has not yet been possible.

Weaver (1989) found that the ultrasonic resonance frequencies of structural materials, such as aluminium blocks, do behave as the eigenvalues of a matrix from the GOE. This is expected of the vibrations of complex structures at frequencies well above the frequencies of their lowest modes. Thus random matrix theory may be of relevance in the non-destructive evaluation of materials, architectural acoustics and the decay of ultrasound in heterogeneous materials.

Le Caër (1989) considered the distribution of trees in the Scandinavian forests and found that their positions look like the eigenvalues of a random complex matrix while the distribution of capital cities of districts of mainland France (Le Caër and Delannay, 1993) looks statistically different.

The series of nuclear energy levels, or any sequence of random numbers for that matter, can be thought to have two distinct kinds of statistical properties, which may be called global and local. A global property varies slowly, its changes being appreciable only on a large interval. The average number of levels per unit of energy or the mean level density, for example, is a global property; it changes appreciably only on intervals containing thousands of levels. Locally it may be treated as a constant. A local property on the other hand fluctuates from level to level. The distance between two successive levels, for example, is a local property. Moreover, global and local properties of a complex system seem to be quite disconnected. Two systems having completely different global properties may have identical local fluctuation properties, and inversely two systems having the same global properties may have different local fluctuation properties.

The random matrix models studied here will have quite different global properties, none of them corresponding exactly to the nuclear energy levels or to the sequence of zeros on the critical line of the Riemann zeta function. One may even choose the global properties at will (Balian, 1968)! However, the nice and astonishing thing about them is that their local fluctuation properties are always the same and determined only by the over-all symmetries of the system. From the extended numerical experience (cf. Appendix A.1) one might state the kind of central limit theorem, referred to earlier, as follows.

Conjecture 1.2.1. Let H be an $N \times N$ real symmetric matrix, its off-diagonal elements H_{ij}, for $i < j$, being independent identically distributed (i.i.d.) random variables with mean zero and variance $\sigma > 0$, i.e. $\langle H_{ij} \rangle = 0$, and $\langle H_{ij}^2 \rangle = \sigma^2 \neq 0$. Then in the limit of large N the statistical properties of n eigenvalues of H become independent of the probability density of the H_{ij}, i.e. when $N \to \infty$, the joint probability density (j.p.d.) of arbitrarily chosen n eigenvalues of H tends, for every finite n, with probability one, to the n-point correlation function of the Gaussian orthogonal ensemble studied in Chapter 7.

Note that the off-diagonal elements H_{ij} for $i > j$ are determined from symmetry, and the diagonal elements H_{ii} may have any distributions.

For Hermitian matrices we suspect the following

Conjecture 1.2.2. Let H be an $N \times N$ Hermitian matrix with complex numbers as elements. Let the real parts of H_{ij} for $i < j$ be i.i.d. random variables with mean zero and variance $\sigma_1 > 0$, while let the imaginary parts of H_{ij} for $i < j$ be i.i.d. random variables with mean zero and variance $\sigma_2 > 0$. Then as the order $N \to \infty$, the j.p.d. of n arbitrarily chosen eigenvalues of H, for every finite n, tends with probability one to the n-point correlation function of the Gaussian unitary ensemble studied in Chapter 6.

Note that, as for real symmetric matrices, the elements H_{ij} for $i > j$ are determined from Hermiticity and the distributions of the diagonal elements H_{ii} seem irrelevant.

A similar result is suspected for self-dual Hermitian quaternion matrices; all finite correlations being identical to those for the Gaussian symplectic ensemble studied in Chapters 8 and 11.

In other words, the local statistical properties of a few eigenvalues of a large random matrix seem to be independent of the distributions of individual matrix elements. What matters is, whether the matrix is real symmetric, or self-dual (quaternion) Hermitian or only Hermitian. The rest does not seem to matter. And this seems to be true even under less restrictive conditions; for example, the probability law for different matrix elements (their real and imaginary parts) may be different.

For level density, i.e. the case $n = 1$, we have some arguments to say that it follows Wigner's "semi-circle" law (cf. Chapter 4). Except for this special case we have only numerical evidence in favor of the above conjectures.

Among the Hermitian matrices, the case of the Gaussian distributions of matrix elements is the one treated analytically by Hsu, Selberg, Wigner, Mehta, Gaudin, Dyson, Rosenzweig, Bronk, Ginibre, Pandey, des Cloizeaux, and others. Circular ensembles of unitary matrices have similarly been studied. We will describe these developments in great detail in the following pages.

1.3 A Summary of Statistical Facts about Nuclear Energy Levels

1.3.1 Level Density. As the excitation energy increases, the nuclear energy levels occur on the average at smaller and smaller intervals. In other words, level density increases with the excitation energy. The first question we might ask is how fast does this level density increase for a particular nucleus and what is the distribution of these levels with respect to spin and parity? This is an old problem treated by Bethe (1937). Even a simple model in which the nucleus is taken as a degenerate Fermi gas with equidistant single-particle levels gives an adequate result. It amounts to determining the number of partitions $\lambda(n)$ of a positive integer n into smaller positive integers $\nu_1, \nu_2, \ldots$

$$n = \nu_1 + \nu_2 + \cdots + \nu_\ell, \quad \nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_\ell > 0.$$

For large n this number, according to the Hardy–Ramanujan formula, is given by

$$\lambda(n) \sim \exp\bigl[(\theta\pi^2 n/3)^{1/2}\bigr],$$

where θ is equal to 1 or 2 according to whether the ν_i are all different or whether some of them are allowed to be equal. With a slight modification due to later work (Lang and Lecouteur, 1954; Cameron, 1956). Bethe's result gives the level density as

$$\rho(E, j, \pi) \propto (2j+1)(E-\Delta)^{-5/4} \exp\bigl[-j(j+1)/2\sigma^2\bigr] \exp\bigl[2a(E-\Delta)^{1/2}\bigr],$$

where E is the excitation energy, j is the spin and π is the parity. The dependence of the parameters σ, a and Δ on the neutron and proton numbers is complicated and only imperfectly understood. However, for any particular nucleus a few measurements will suffice to determine them all; the formula will then remain valid for a wide range of energy that contains thousands and even millions of levels.

1.3.2 Distribution of Neutron Widths. An excited level may decay in many ways; for example, by neutron ejection or by giving out a quantum of radiation. These processes are characterized by the corresponding decay widths of the levels. The neutron reduced widths $\Gamma_n^0 = \Gamma_n/E^{1/2}$, in which Γ_n is the neutron width and E is the excitation energy of the level, show large fluctuations from level to level. From an analysis of the available data Scott (1954) and later Porter and Thomas (1956) concluded that they had a χ^2-distribution with $\nu = 1$ degree of freedom:

$$\begin{aligned} P(x) &= (\nu/2)\big[\Gamma(\nu/2)\big]^{-1}(\nu x/2)^{(\nu/2)-1}e^{-\nu x/2} \\ &= (2\pi x)^{-1/2}e^{-x/2}, \\ x &= \Gamma_n^0/\bar{\Gamma}_n^0, \end{aligned}$$

where $\bar{\Gamma}_n^0$ is the average of Γ_n^0 and $P(x)$ is the probability that a certain reduced width will lie in an interval dx around the value x. This indicates a Gaussian distribution for the reduced width amplitude

$$\left(\frac{2}{\pi}\right)^{1/2} \exp\left[-\frac{1}{2}(\sqrt{x})^2\right] d(\sqrt{x})$$

expected from the theory. In fact, the reduced width amplitude is proportional to the integral of the product of the compound nucleus wave function and the wave function in the neutron-decay channel over the channel surface. If the contributions from the various parts of the channel surface are supposed to be random and mutually independent, their sum will have a Gaussian distribution with zero mean.

1.3.3 Radiation and Fission Widths. The total radiation width is almost a constant for particular spin states of a particular nucleus. The total radiation width is the sum of partial radiation widths

$$\Gamma = \sum_{i=1}^{m} \Gamma_i.$$

If we assume that each of these $\Gamma_i/\bar{\Gamma}_i$ ($\bar{\Gamma}_i$ denoting the average of Γ_i), has a χ^2-distribution with one degree of freedom like the neutron widths and all the $\bar{\Gamma}_i$ are the same, then $\Gamma/\bar{\Gamma}$ ($\bar{\Gamma}$ being the average of Γ), will have a χ^2-distribution with m degrees of freedom. For (moderately) large m, this is a narrow distribution. This conclusion remains valid even when the $\bar{\Gamma}_i$ are not all equal.

It is difficult to measure the partial radiation widths.

Little is known about the fission-width distributions. Some known fission widths of U^{235} have been analyzed (Bohr, 1956) and a χ^2-distribution with 2 to 3 degrees of freedom has been found to give a satisfactory fit.

From now on we shall no longer consider neutron, radiation, or fission widths.

1.3.4 Level Spacings. Let us regard the level density as a function of the excitation energy as known and consider an interval of energy δE centered at E. This interval is much smaller compared to E, whereas it is large enough to contain many levels; that is,

$$E \gg \delta E \gg D,$$

where D is the mean distance between neighboring levels. How are the levels distributed in this interval? On Figure 1.2 a few examples of level series are shown. In all these cases the level density is taken to be the same, i.e. the scale in each case is chosen so that the average distance between the neighboring levels is unity. It is evident that these different level sequences do not look similar. There are many coincident pairs or sometimes even triples of levels as well as large gaps when the levels have no correlations, i.e. the Poisson distribution; whereas the zeros of the Riemann zeta function are more or less equally spaced. The case of prime numbers, the slow neutron resonance levels of erbium 166 nucleus and the possible energy values of a free particle confined to a billiard table of a specified shape are far from either regularly spaced uniform series or the completely random Poisson series with no correlations.

Although the level density varies strongly from nucleus to nucleus, the fluctuations in the precise positions of the levels seem not to depend on the nucleus and not even on the excitation energy. As the density of the levels is nearly constant in this interval, we might think that they occur at random positions without regard to one another, the only condition being that their density be a given constant. However, as we see on Figure 1.2(c), such is not the case. It is true that nuclear levels with different spin and parity or atomic levels with different sets of good quantum numbers seem to have no influence on each other. However, levels with the same set of good quantum numbers show a large degree of regularity. For instance, they rarely occur close together.

A more detailed discussion of the experimental data regarding the above quantities as well as the strength functions may be found in the review article by Brody et al. (1981).

1.4 Definition of a Suitable Function for the Study of Level Correlations

To study the statistical properties of the sequence of eigenvalues one defines suitable functions such as level spacings or correlation and cluster functions, or some suitable quantity called a statistic such as the Δ-, Q-, F- or the Λ-statistic of the literature. We will consider a few of them in due course. For the spacings let $E_1, E_2, \ldots$ be the positions of the successive levels in the interval δE $(E_1 \leqslant E_2 \leqslant \cdots)$ and let $S_1, S_2, \ldots$ be their distances apart, $S_i = E_{i+1} - E_i$. The average value of S_i is the mean spacing D. We define the relative spacing $s_i = S_i/D$. The probability density function $p(s)$ is defined by the condition that $p(s)\,ds$ is the probability that any s_i will have a value between s and $s + ds$.

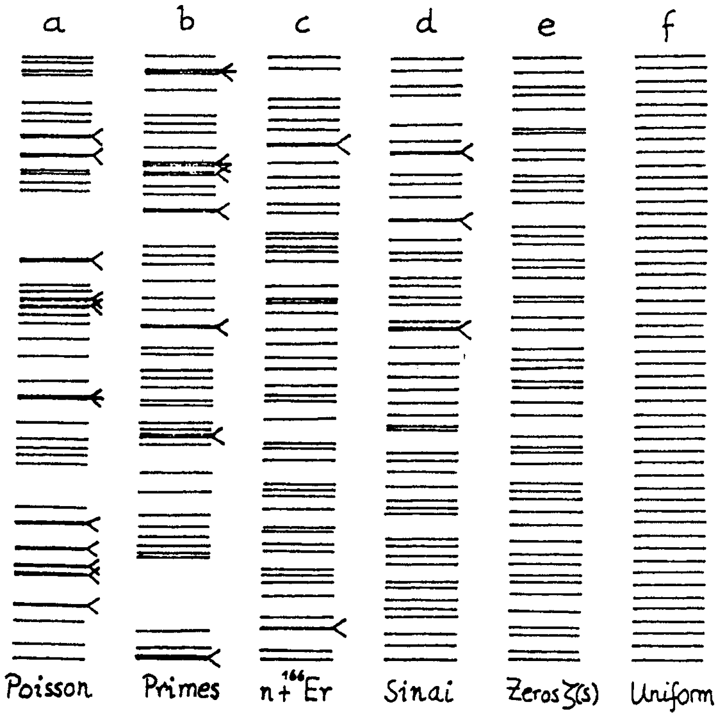

Figure 1.2. Some typical level sequences. From Bohigas and Giannoni (1984). (a) Random levels with no correlations, Poisson series. (b) Sequence of prime numbers. (c) Slow neutron resonance levels of the erbium 166 nucleus. (d) Possible energy levels of a particle free to move inside the area bounded by $1/8$ of a square and a circular arc whose center is the mid point of the square; i.e. the area specified by the inequalities, $y \geqslant 0$, $x \geqslant y$, $x \leqslant 1$, and $x^2 + y^2 \geqslant r$. (Sinai's billiard table.) (e) The zeros of the Riemann zeta function on the line $\operatorname{Re} z = 1/2$. (f) A sequence of equally spaced levels (Bohigas and Giannoni, 1984).

For the simple case in which the positions of the energy levels are not correlated the probability that any E_i will fall between E and $E + dE$ is independent of E and is simply $\rho\, dE$, where $\rho = D^{-1}$ is the average number of levels in a unit interval of energy. Let us determine the probability of a spacing S; that is, given a level at E, what is the probability of having no level in the interval $(E, E + S)$ and one level in the interval $(E + S, E + S + dS)$. For this we divide the interval S into m equal parts.

$$E \qquad E + S/m \qquad E + 2S/m \qquad \ldots \qquad E + (m-1)S/m \qquad E + S \qquad E + S + dS$$

Because the levels are independent, the probability of having no level in $(E, E + S)$ is the product of the probabilities of having no level in any of these m parts. If m is large, so that S/m is small, we can write this as $(1-\rho S/m)^m$, and in the limit $m \to \infty$,

$$\lim_{m\to\infty}\left(1-\rho\frac{S}{m}\right)^m = e^{-\rho S}.$$

Moreover, the probability of having a level in dS at $E+S$ is $\rho\, dS$. Therefore, given a level at E, the probability that there is no level in $(E, E+S)$ and one level in dS at $E+S$ is

$$e^{-\rho S}\rho\, dS,$$

or in terms of the variable $s = S/D = \rho S$

$$p(s)\, ds = e^{-s}\, ds. \tag{1.4.1}$$

This is known as the Poisson distribution or the spacing rule for random levels.

That (1.4.1) is not correct for nuclear levels of the same spin and parity or for atomic levels of the same parity and orbital and spin angular momenta is clearly seen by a comparison with the empirical evidence (Figures 1.3 and 1.4). It is not true either for the eigenvalues of a matrix from any of the Gaussian ensembles, as we will see.

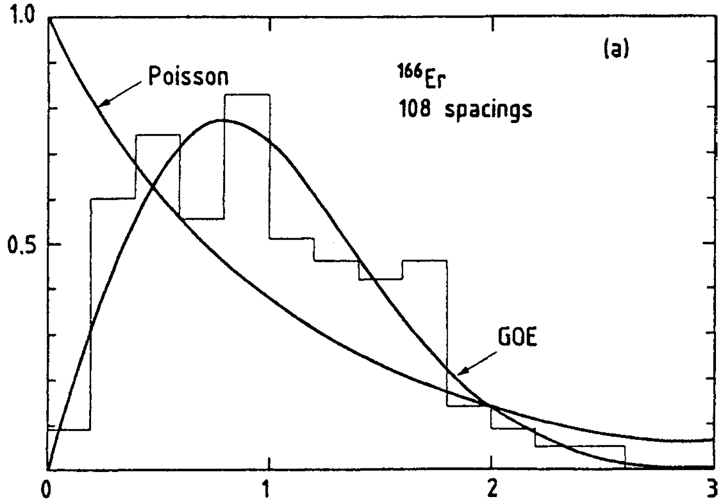

Figure 1.3. The probability density for the nearest neighbor spacings in slow neutron resonance levels of erbium 166 nucleus. The histogram shows the first 108 levels observed. The solid curves correspond to the Poisson distribution, i.e. no correlations at all, and that for the eigenvalues of a real symmetric random matrix taken from the Gaussian orthogonal ensemble (GOE). Reprinted with permission from The American Physical Society, Liou et al., Neutron resonance spectroscopy data, *Phys. Rev. C* 5 (1972) 974–1001.

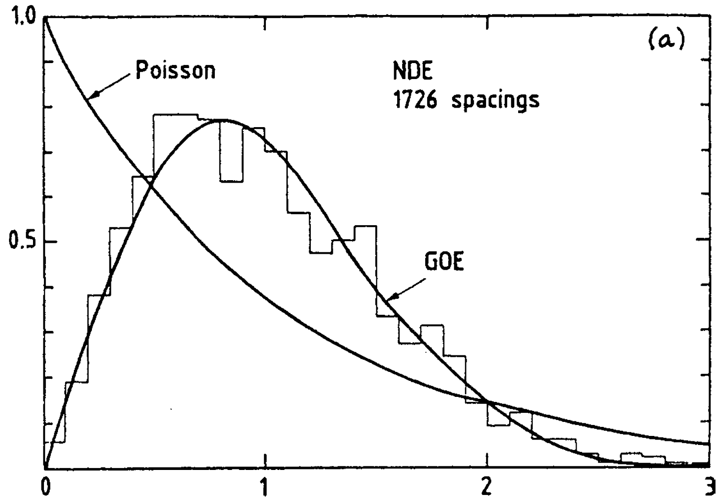

Figure 1.4. Level spacing histogram for a large set of nuclear levels, often referred to as nuclear data ensemble. The data considered consists of 1407 resonance levels belonging to 30 sequences of 27 different nuclei: (i) slow neutron resonances of Cd(110, 112, 114), Sm(152, 154), Gd(154, 156, 158, 160), Dy(160, 162, 164), Er(166, 168, 170), Yb(172, 174, 176), W(182, 184, 186), Th(232) and U(238); (1146 levels); (ii) proton resonances of Ca(44) ($J = 1/2+$), Ca(44) ($J = 1/2-$), and Ti(48) ($J = 1/2+$); (157 levels); and (iii) (n, γ)-reaction data on Hf(177) ($J = 3$), Hf(177) ($J = 4$), Hf(179) ($J = 4$), and Hf(179) ($J = 5$); (104 levels). The data chosen in each sequence is believed to be complete (no missing levels) and pure (the same angular momentum and parity). For each of the 30 sequences the average quantities (e.g. the mean spacing, spacing/mean spacing, number variance μ_2, etc., see Chapter 16) are computed separately and their aggregate is taken weighted according to the size of each sequence. The solid curves correspond to the Poisson distribution, i.e. no correlations at all, and that for the eigenvalues of a real symmetric random matrix taken from the Gaussian orthogonal ensemble (GOE). Reprinted with permission from Kluwer Academic Publishers, Bohigas O., Haq R.U. and Pandey A., Fluctuation properties of nuclear energy levels and widths, comparison of theory with experiment, in: *Nuclear Data for Science and Technology*, Bökhoff K.H. (Ed.), 809–814 (1983).

1.5 Wigner Surmise

When the experimental situation was not yet conclusive, Wigner proposed the following rules for spacing distributions:

(1) In the sequence of levels with the same spin and parity, called a simple sequence, the probability density function for a spacing is given by

$$p_W(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi}{4}s^2\right), \quad s = \frac{S}{D}. \tag{1.5.1}$$

(2) Levels with different spin and parity are not correlated. The function $p(s)$ for a mixed sequence may be obtained by randomly superimposing the constituent simple sequences (cf. Appendix A.2).

Two simple arguments give rise to Rule 1. As pointed out by Wigner (1951) and by Landau and Smorodinski (1955), it is reasonable to expect that, given a level at E, the probability that another level will lie around $E + S$ is proportional to S for small S. Now if we extrapolate this to all S and, in addition, assume that the probabilities in various intervals of length S/m obtained by dividing S into m equal parts are mutually independent, we arrive at

$$p(S/D)\,dS = \lim_{m\to\infty} \prod_{r=0}^{m-1} \left(1 - \frac{Sr}{m}\frac{S}{m}a\right) aS\,dS = aSe^{-aS^2/2}\,dS. \tag{1.5.2}$$

The constant a can be determined by the condition that the average value of $s = S/D$ is unity:

$$\int_0^\infty sp(s)\,ds = 1. \tag{1.5.3}$$

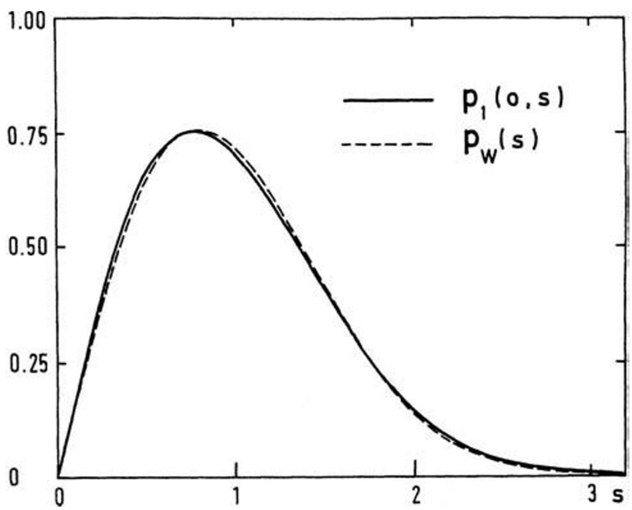

Figure 1.5. The probability density $p_1(0; s)$ of the nearest neighbor spacings for the eigenvalues of a random real symmetric matrix from the Gaussian orthogonal ensemble, Eq. (7.4.18) and the Wigner surmise $p_W(s)$; Eq. (1.5.1). Reprinted with permission from Elsevier Science Publishers, Gaudin M., Sur la loi limite de l'espacements des valeurs propres d'une matrice aléatoire, *Nucl. Phys.* 25, 447–458 (1961).

Let us, at this point, define the n-point correlation function $R_n(E_1, \ldots, E_n)$ so that $R_n \, dE_1 \, dE_2 \cdots dE_n$ is the probability of finding a level in each of the intervals $(E_1, E_1 + dE_1), \ldots, (E_n, E_n + dE_n)$ all other levels being unobserved. The two simple arguments of Wigner given in the derivation of Rule 1 are equivalent to the following. The two-point correlation function $R_2(E_1, E_2)$ is linear in the variable $|E_1 - E_2|$, and three and higher order correlation functions are negligibly small.

We shall see in Chapter 7 that both arguments are inaccurate, whereas Rule 1 is very near the correct result (cf. Figure 1.5). It is surprising that the two errors compensate so nearly each other.

1.6 Electromagnetic Properties of Small Metallic Particles

Consider small metallic particles at low temperatures. The number of electrons in a volume V is $n \approx 4\pi p_0^3 V/3h^3$, where p_0 is the Fermi momentum and h is Planck's constant. The energy of an excitation near the Fermi surface is $E_0 \approx p_0^2/2m^*$, where m^* is the effective mass of the electron. The level density at zero excitation is therefore $\sigma = dn/dE_0 \approx 4\pi p_0 V m^*/h^3$, and the average level spacing is the inverse of this quantity $D \approx \sigma^{-1}$. For a given temperature we can easily estimate the size of the metallic particles for which $D \gg kT$, where k is Boltzmann's constant and T is the temperature in degrees Kelvin. For example, the number of electrons in a metallic particle of size 10^{-6}–10^{-7} cm may be as low as a few hundred and, at $T \approx 10$ K, $D \approx 1$ eV, whereas $kT \approx 10^{-3}$ eV. It is possible to produce particles of this size experimentally and then to sort them out according to their size (e.g., by centrifuging and sampling at a certain radial distance). Thus we have a large number of metallic particles, each of which has a different shape and therefore a different set of electronic energy levels but the same average level spacing, for the volumes are equal. It would be desirable if we could separate (e.g., by applying a non uniform magnetic field) particles containing an odd number of conduction electrons from those containing an even number. The energy-level schemes for these two types of particles have very different properties.

Given the position of the electron energies, we can calculate the partition function in the presence of a magnetic field and then use thermodynamic relations to derive various properties such as electronic specific heat and spin paramagnetism. Fröhlich (1937) assumed that the energies were equally spaced and naturally obtained the result that all physical quantities decrease exponentially at low temperatures as $\exp(-D/kT)$ for $1 \ll D/kT$. Kubo (1969) repeated the calculation with the assumption that the energies were random without correlations and that their spacings therefore follow a Poisson law. He arrived at a linear law for the specific heat $\sim kT/D$. The constants are different for particles containing an odd number of electrons from those containing an even number. For spin paramagnetism even the dependence on temperature is different for the two sets of particles. Instead of Fröhlich's equal spacing rule or Kubo's Poisson law, it would perhaps be better to suppose with Gorkov and Eliashberg (1965), that these energies behave as the eigenvalues of a random matrix. This point of view may be justified as follows. The energies are the eigenvalues of a fixed Hamiltonian with random boundary conditions. We may incorporate these boundary conditions into the Hamiltonian by the use of fictitious potentials. The energies are thus neither equally spaced, nor follow the Poisson law, but they behave as the eigenvalues of a random matrix taken from a suitable ensemble. In contrast to nuclear spectra, we have the possibility of realizing in practice all three ensembles considered in various sections of this book. They apply in particular when (a) the number of electrons (in each of the metallic particles) is even and there is no external magnetic field, (b) the number of electrons (in each of the metallic particles) is odd and there is no external magnetic field, (c) there is an external magnetic field much greater than D/μ, where μ is the magnetic moment of the electron.

As to which of the three assumptions is correct should be decided by the experimental evidence. Unfortunately, such experiments are difficult to perform neatly and no clear conclusion is yet available. See the discussion in Brody et al. (1981).

1.7 Analysis of Experimental Nuclear Levels

The enormous amount of available nuclear data was analyzed in the 1980s by French and coworkers, specially in view of deriving an upper bound for the time reversal violating part of the nuclear forces (cf. Haq et al. (1982), Bohigas et al. (1983, 1985)). Actually, if nuclear forces are time reversal invariant, then the nuclear energy levels should behave as the eigenvalues of a random real symmetric matrix; if they are not, then they should behave as those of a random Hermitian matrix. The level sequences in the two cases have different properties; for example, the level spacing curves are quite distinct. Figures 1.6 and 1.7 indicate that nuclear and atomic levels behave as the eigenvalues of a random real symmetric matrix and the admixture of a Hermitian anti-symmetric part, if any, is small. Figure 1.4 shows the level spacing histogram for a really big sample of available nuclear data from which French et al. (1985) could deduce an upper bound for this small admixture. Figure 1.8 shows how the probability density of the nearest neighbor spacings of atomic levels changes as one goes from one long period to another in the periodic table of elements. For further details on the analysis of nuclear levels and other sequences of levels, see Chapter 16.

1.8 The Zeros of The Riemann Zeta Function

The zeta function of Riemann is defined for $\operatorname{Re} z > 1$ by

$$\zeta(z) = \sum_{n=1}^{\infty} n^{-z} = \prod_{p} (1 - p^{-z})^{-1}, \tag{1.8.1}$$

and for other values of z by its analytical continuation. The product in Eq. (1.8.1) above is taken over all primes p, and is known as the Euler product formula.

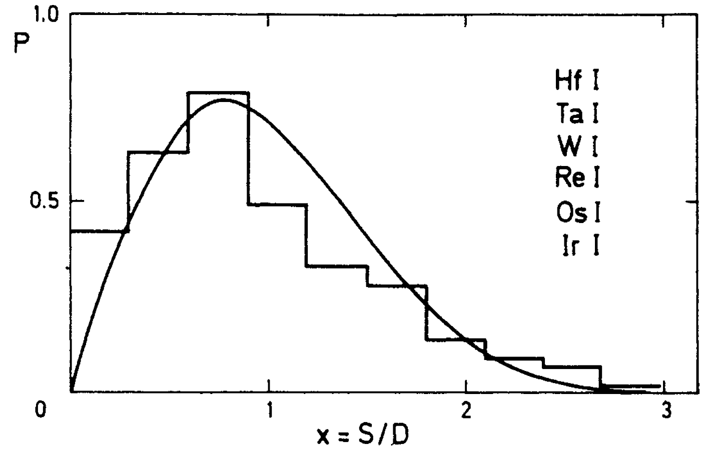

Figure 1.6. Plot of the density of nearest neighbor spacings between odd parity atomic levels of a group of elements in the region of osmium. The levels in each element were separated according to angular momentum, and separate histograms were constructed for each level series, and then combined. The elements and the number of contributed spacings are HfI, 74; TaI, 180; WI, 262; ReI, 165; OsI, 145; IrI, 131 which lead to a total of 957 spacings. The solid curve corresponds to the Wigner surmise, Eq. (1.5.1). Reprinted with permission from Annales Academiae Scientiarum Fennicae, Porter C.E. and Rosenzweig N., Statistical properties of atomic and nuclear spectra, *Annale Academiae Scientiarum Fennicae, Serie A VI, Physica* 44, 1–66 (1960).

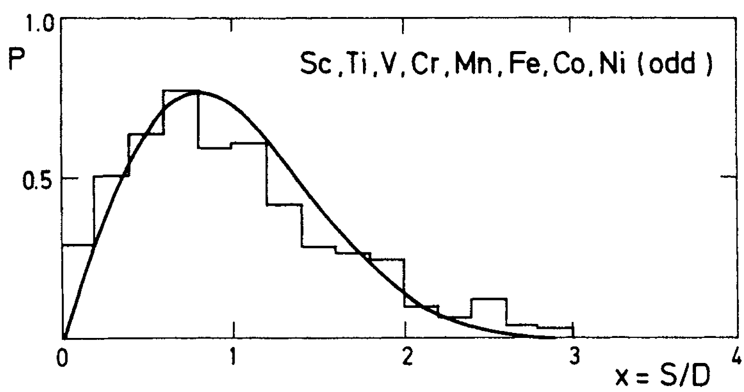

Figure 1.7. Plot of the density of nearest neighbor spacings between odd parity levels of the first period corresponding to a separation of the levels into sequences each of which is labeled by definite values of S, L and J. Comparison with the Wigner surmise (solid curve) shows a good fit if the (approximate) symmetries resulting from (almost) complete absence of spin-orbit forces are taken into consideration. Reprinted with permission from Annales Academiae Scientiarum Fennicae, Porter C.E. and Rosenzweig N., Statistical properties of atomic and nuclear spectra, *Annale Academiae Scientiarum Fennicae, Serie A VI, Physica* 44, 1–66 (1960).

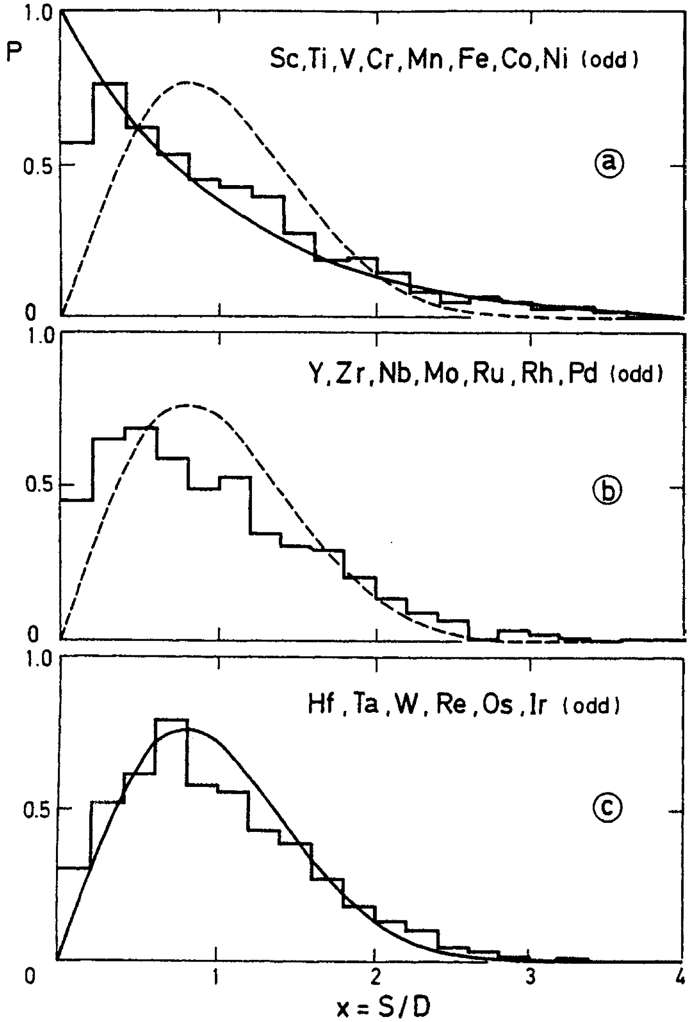

Figure 1.8. Empirical density of nearest neighbor spacings between odd parity levels of elements in the first, second and third long periods (histograms a, b and c respectively). To obtain these figures separate histograms were constructed for the J sequences of each element and then the results were combined. Comparison with Poisson (or exponential) distribution and Wigner surmise (also shown) indicates that curves go from Poisson to Wigner curves as one goes from the first to the second and then finally to the third period. This variation can be understood in terms of the corresponding increase in strength of the spin dependent forces. Reprinted with permission from Annales Academiae Scientiarum Fennicae, Porter C.E. and Rosenzweig N., Statistical properties of atomic and nuclear spectra, *Annale Academiae Scientiarum Fennicae, Serie A VI, Physica* 44, 1–66 (1960).

It is well-known that $\zeta(z)$ is zero for $z = -2n$, $n = 1, 2, \ldots$; these are called the "trivial zeros". All the other zeros of $\zeta(z)$ lie in the strip $0 < \operatorname{Re} z < 1$, and are symmetrically situated with respect to the critical line $\operatorname{Re} z = 1/2$. It is conjectured that they actually lie on the critical line $\operatorname{Re} z = 1/2$ (Riemann hypothesis, 1876). Since $\zeta(z^*) = \zeta^*(z)$, if z is a zero of $\zeta(z)$, then z^* is also a zero of it. Assuming the truth of the Riemann hypothesis (RH), let $z = 1/2 \pm i\gamma_n$, γ_n real and positive, be the "non-trivial" zeros of $\zeta(z)$. How the γ_n are distributed on the real line?

These questions have their importance in number theory. Many, as yet unsuccessful, attempts have been made to prove or disprove the RH. With the advent of electronic computers of ever faster speeds and larger memories, a series of efforts have also been made to verify the truth of the RH numerically and to evaluate some statistical properties of the γ_n.

Until the 1970s few persons were interested in the distribution of the γ_n. The main reason seems to be the feeling that if one cannot prove (or disprove!) the γ_n to be all real (RH), then there is no point in asking even harder questions. Montgomery (1973) took the lead in investigating questions about the distribution of the γ_n. It was known (Titchmarsh, 1951, Chapter 10) that the number $N(T)$ of the γ_n with $0 < \gamma_n \leqslant T$ is

$$N(T) = \frac{T}{2\pi} \ln\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + S(T) + \frac{7}{8} + \mathrm{O}(T^{-1}), \tag{1.8.2}$$

as $T \to \infty$, where

$$S(t) = \frac{1}{\pi} \arg \zeta\left(\frac{1}{2} + it\right). \tag{1.8.3}$$

Maximum order of $S(T)$ remains unknown. Probably,

$$S(T) = \mathrm{O}\left(\left(\frac{\ln T}{\ln\ln T}\right)^{1/2}\right), \tag{1.8.4}$$

or more likely (Montgomery, private communication)

$$S(T) = \mathrm{O}\big((\ln T \ln\ln T)^{1/2}\big). \tag{1.8.4$'$}$$

Assuming the RH Montgomery studied $D(\alpha, \beta)$, the number of pairs γ, γ' such that $\zeta(1/2 + i\gamma) = \zeta(1/2 + i\gamma') = 0$, $0 < \gamma \leqslant T$, $0 < \gamma' \leqslant T$, and $2\pi\alpha/(\ln T) \leqslant \gamma - \gamma' \leqslant 2\pi\beta/(\ln T)$. Or taking the Fourier transforms, it amounts to evaluate the function

$$F(\alpha) = \frac{2\pi}{T \ln T} \sum_{0<\gamma,\gamma'\leqslant T} T^{i\alpha(\gamma-\gamma')} \frac{4}{4 + (\gamma - \gamma')^2} \tag{1.8.5}$$

for real α. Since F is symmetric in γ, γ', it is real and even in α. Montgomery (1973) showed that if RH is true then $F(\alpha)$ is nearly non-negative, $F(\alpha) \geqslant -\varepsilon$ uniformly in α

for $T > T_0(\varepsilon)$, and that

$$F(\alpha) = (1 + \mathrm{o}(1))T^{-2\alpha} \ln T + \alpha + \mathrm{o}(1) \tag{1.8.6}$$

uniformly for $0 \leqslant \alpha \leqslant 1$. For $\alpha > 1$, the behaviour changes. He also gave heuristic arguments to suggest that for $\alpha \geqslant 1$,

$$F(\alpha) = 1 + \mathrm{o}(1). \tag{1.8.7}$$

And from this conjecture he deduced that

$$\frac{2\pi}{T \ln T} D(\alpha, \beta) \approx \int_\alpha^\beta \left(1 - \left(\frac{\sin r}{r}\right)^2 + \delta(\alpha, \beta)\right) dr \tag{1.8.8}$$

for any real α, β with $\alpha < \beta$. Here $\delta(\alpha, \beta) = 1$ if $\alpha < 0 < \beta$, and $\delta(\alpha, \beta) = 0$ otherwise. This $\delta(\alpha, \beta)$ is there because for $\alpha < 0 < \beta$, $D(\alpha, \beta)$ includes terms with $\gamma = \gamma'$.

Equation (1.8.8) says that the two-point correlation function of the zeros of the zeta function $\zeta(z)$ on the critical line is

$$R_2(r) = 1 - (\sin(\pi r)/(\pi r))^2. \tag{1.8.9}$$

As we will see in Chapter 6.2, this is precisely the two-point correlation function of the eigenvalues of a random Hermitian matrix taken from the Gaussian unitary ensemble (GUE) (or that of a random unitary matrix taken from the circular unitary ensemble, cf. Chapter 11.1). This is consistent with the view (quoted to be conjectured by Polya and Hilbert) that the zeros of $\zeta(z)$ are related to the eigenvalues of a Hermitian operator. As the two-point correlation function seems to be the same for the zeros of $\zeta(z)$ and the eigenvalues of a matrix from the GUE, it is natural to think that other statistical properties also coincide. With a view to prove (or disprove!) this expectation, Odlyzko (1987) computed a large number of zeros $1/2 + i\gamma_n$ of $\zeta(z)$ with great accuracy. Taking γ_n to be positive and ordered, i.e. $0 < \gamma_1 \leqslant \gamma_2 \leqslant \cdots$ with $\gamma_1 = 14.134\ldots$, $\gamma_2 = 21.022\ldots$ etc. he computed sets of 10^5 zeros γ_n with a precision $\pm 10^{-8}$, and $N + 1 \leqslant n \leqslant N + 10^5$, for $N = 0$, $N = 10^6$, $N = 10^{12}$, $N = 10^{18}$ and $N = 10^{20}$. (This computation gives, for example, $\gamma_n = 15202440115920747268.6290299\ldots$ for $n = 10^{20}$, cf. Odlyzko, 1989 or Cipra, 1989.) This huge numerical undertaking was possible only with the largest available supercomputer CRAY X-MP and fast efficient algorithms devised for the purpose. The numerical evidence so collected is shown on Figures 1.9 to 1.14.

Note that as one moves away from the real axis, the fit improves for both the spacing as well as the two point correlation function. The convergence is very slow; the numerical curves become indistinguishable from the GUE curves only far away from the real axis, i.e. when $N \geqslant 10^{12}$. In contrast, for Hermitian matrices or for the unitary

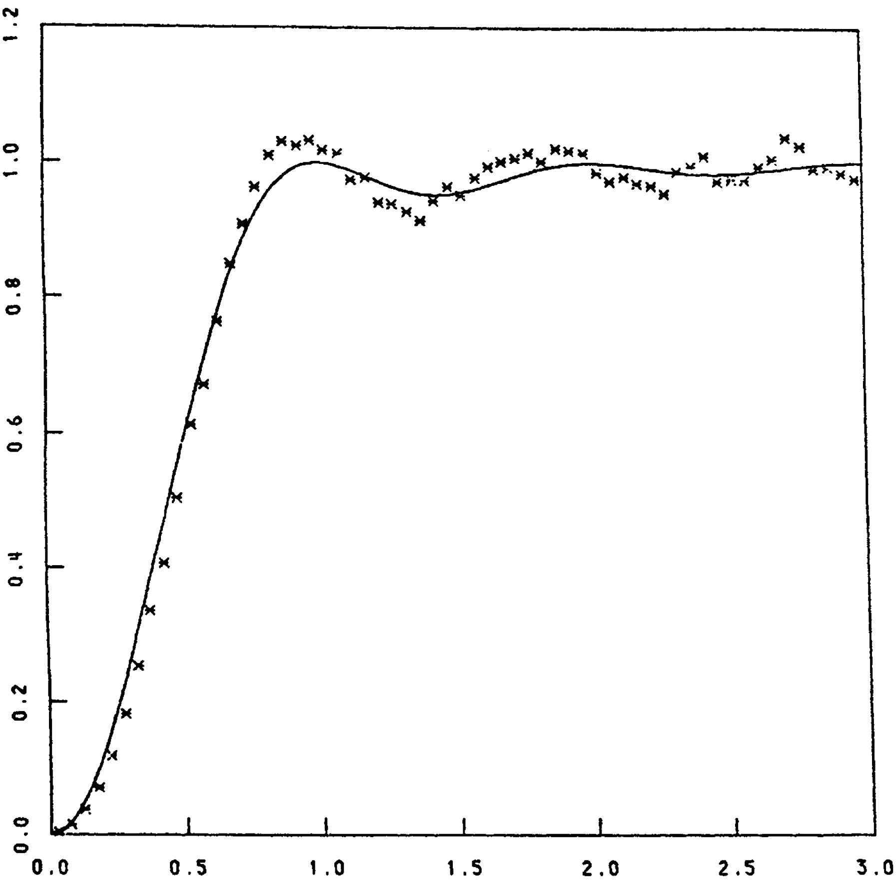

Figure 1.9. Two point correlation function for the zeros $0.5 \pm i\gamma_n$, γ_n real, of the Riemann zeta function; $1 < n < 10^5$. The solid curve is Montgomery's conjecture, Eq. (1.8.9). Reprinted from "On the distribution of spacings between zeros of the zeta function," A.M. Odlyzko, *Mathematics of Computation* pages 273–308 (1987), by permission of The American Mathematical Society.

matrices the limit is practically reached for matrix orders 100×100 and 20×20 respectively. This can be understood by the fact that the density of the zeros of the zeta function around $N \approx 10^{12}$ is comparable to the eigenvalue density for the Hermitian or the unitary matrices of orders 100 or 20.

One may compare the Riemann zeta function on the critical line to the characteristic function of a random unitary matrix. In other words, consider the variations of the Riemann zeta function as one moves along the critical line $\mathrm{Re}\, z = 1/2$ far away from the real axis as well as the variations of the characteristic function of an $N \times N$ random

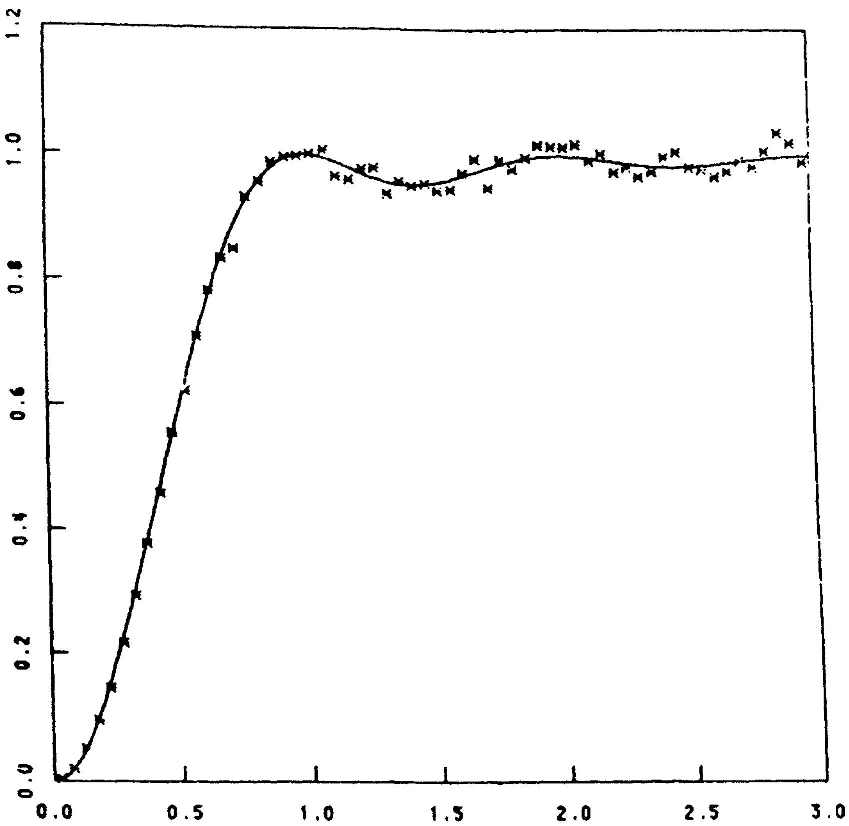

Figure 1.10. The same as Figure 1.9 with $10^{12} < n < 10^{12} + 10^5$; Note that the curves fit much better. Reprinted from "On the distribution of spacings between zeros of the zeta function," A.M. Odlyzko, *Mathematics of Computation* pages 273–308 (1987), by permission of The American Mathematical Society.

unitary matrix,

$$f(x) = \zeta\left(\frac{1}{2} + it\right), \quad t = T + \frac{x}{2\pi}\ln\left(\frac{T}{2\pi}\right), \tag{1.8.10}$$

$$g(x) = \det[e^{i\theta} \cdot \mathbf{1} - U], \quad \theta = x \cdot \frac{N}{2\pi}, \tag{1.8.11}$$

then in the limit of large T and N empirically the zeros of $f(x)$ behave as if they were the zeros of $g(x)$; they have the same m-point correlation functions for $m = 2, 3, 4, \ldots$; they have the same spacing functions $p(m, s)$ for $m = 0, 1, \ldots$ (cf. Chapters 6 and 16).

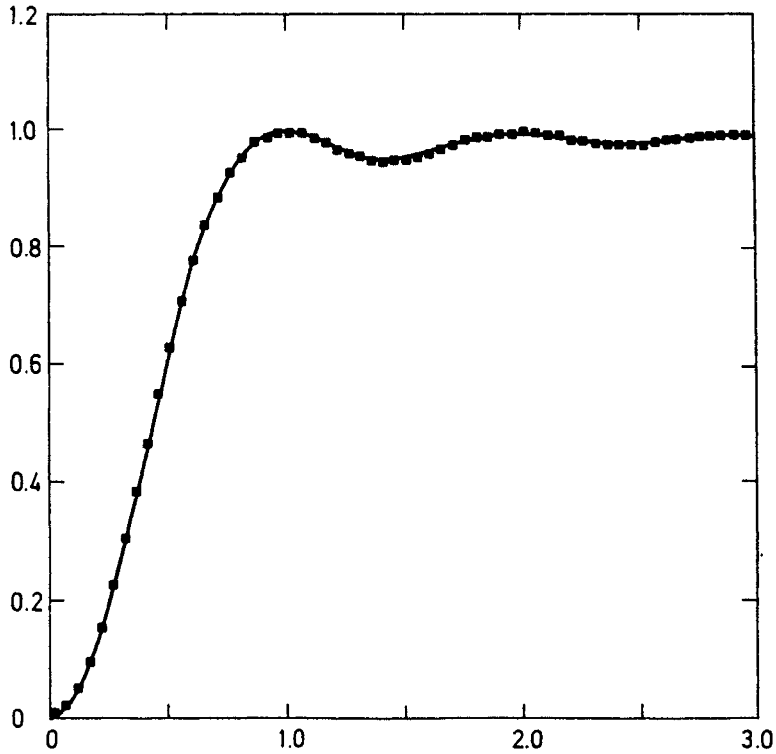

Figure 1.11. The same as Figure 1.9, but for 79 million zeros around $n \approx 10^{20}$. From Odlyzko (1989). Copyright © 1989 American Telephone and Telegraph Company, reprinted with permission.

Quantities other than the zeros of these functions have been considered and found to match in all the known cases. For example, Odlyzko computed the density of values of the real and imaginary parts of the logarithm of $\zeta(1/2 + it)$ for large values of t and noticed that their convergence to the asymptotic limit, known to be Gaussian, is quite slow. Keating and Snaith (2000a) computed the density of values of the real and imaginary parts of the logarithm of $\zeta(1/2 + it)$ for t around the 10^{20}th zero of the zeta function from the data of Odlyzko and compared it with the probability density of the real and imaginary parts of the logarithm of the characteristic function of a random 42×42 unitary matrix (since 10^{20}th zero of the zeta function is at $t \approx 1.52 \times 10^{19}$ and $\ln[1.52 \times 10^{19}/(2\pi)] \approx 42$). The two curves are indistinguishable and visibly differ from their asymptotic limits, known to be Gaussian. See Figures 1.15 and 1.16 (taken from Keating and Snaith (2000a), *Comm. Math. Phys.* **214**, 57–89).

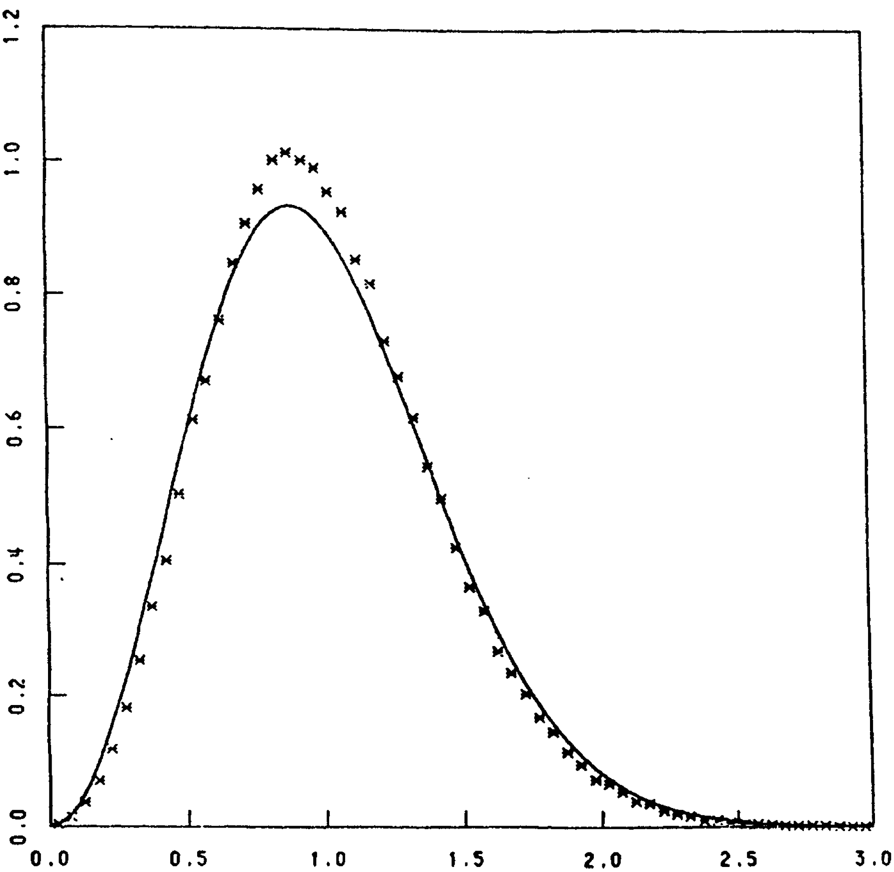

Figure 1.12. Plot of the density of normalized spacings for the zeros $0.5 \pm i\gamma_n$, γ_n real, of the Riemann zeta function on the critical line. $1 < n < 10^5$. The solid curve is the spacing probability density for the Gaussian unitary ensemble, Eq. (6.4.32). From Odlyzko (1987). Reprinted from "On the distribution of spacings between zeros of the zeta function," *Mathematics of Computation* (1987), pages 273–308, by permission of The American Mathematical Society.

Also the distribution of the zeros of $d\zeta(1/2 + it)/dt$ and those of the derivative of the characteristic function of a random unitary matrix has been investigated and found empirically to be the same when centered and normalized properly (Mezzadri, 2003).

It is hard to imagine the zeros of the zeta function as the eigenvalues of some unitary or Hermitian operator. It is even harder to imagine the zeta function on the critical line as the characteristic function of a unitary operator.

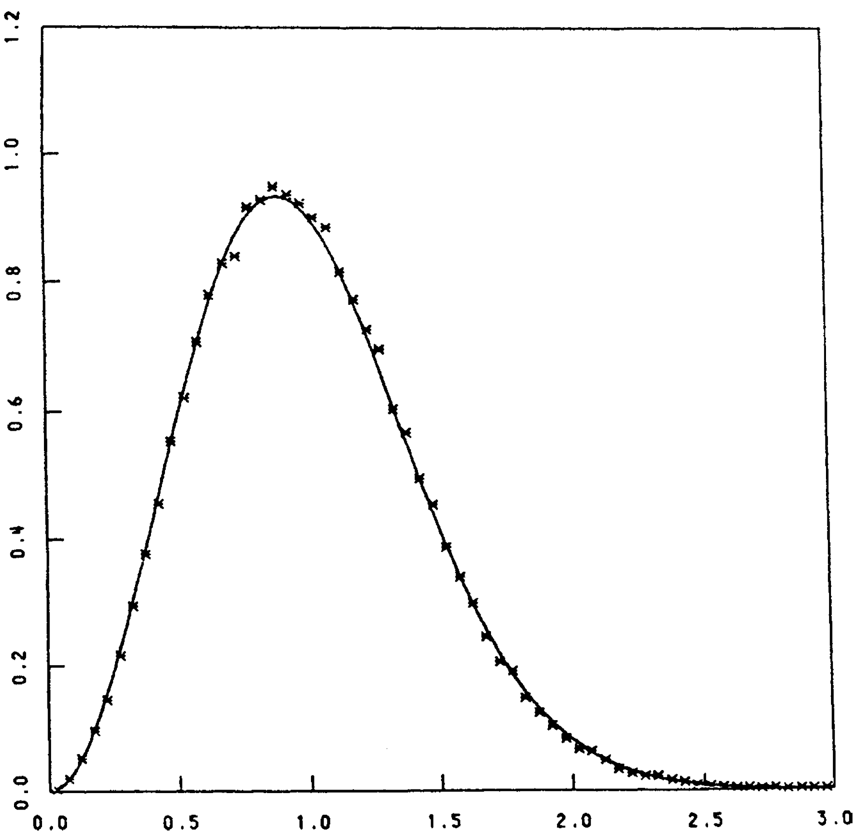

Figure 1.13. The same as Figure 1.12 with $10^{12} < n < 10^{12} + 10^5$. Note the improvement in the fit. From Odlyzko (1987). Reprinted from "On the distribution of spacings between zeros of the zeta function," *Mathematics of Computation* (1987), pages 273–308, by permission of The American Mathematical Society.

A generalization of the Riemann zeta function is the function $\zeta(z, a)$ defined for $\operatorname{Re} z > 1$, by

$$\zeta(z, a) = \sum_{n=0}^{\infty} (n + a)^{-z}, \quad 0 < a \leqslant 1, \tag{1.8.12}$$

and by its analytical continuation for other values of z. For $a = 1/2$ and $a = 1$, one has

$$\zeta(z, 1/2) = (2^z - 1)\zeta(z), \quad \zeta(z, 1) = \zeta(z), \tag{1.8.13}$$

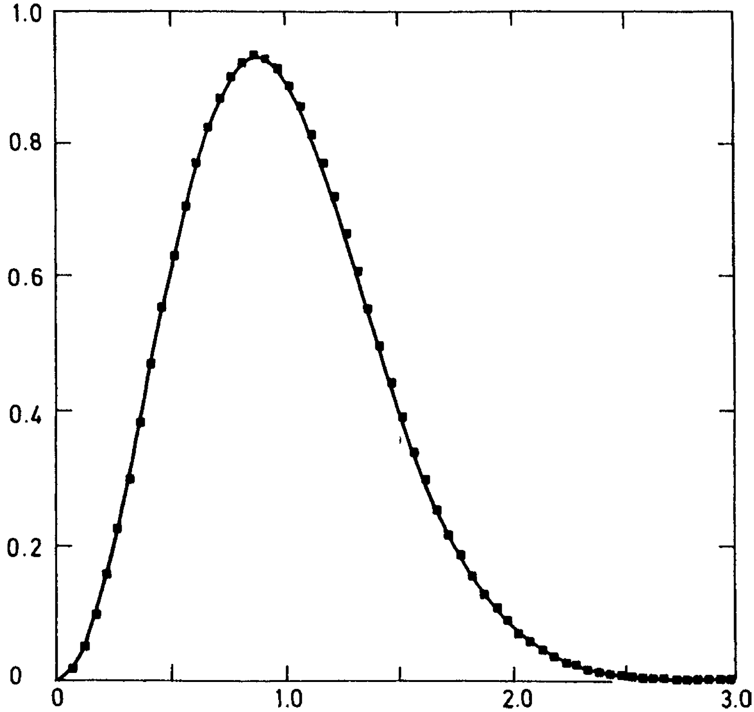

Figure 1.14. The same as Figure 1.12 but for the 79 million zeros around the 10^{20}th zero. From Odlyzko (1989). Copyright © 1989, American Telephone and Telegraph Company, reprinted with permission.

so there is nothing more about their zeros. For rational values of a other than $1/2$ or 1 or for transcendental values of a, it is known that $\zeta(z, a)$ has an infinity of zeros with $\operatorname{Re} z > 1$. For irrational algebraic values of a it is not known whether there are any zeros with $\operatorname{Re} z > 1$ (Davenport and Heilbronn, 1936).

A quadratic form (in two variables) $Q(x, y) = ax^2 + bxy + cy^2$, a, b, c integers, is positive definite if $a > 0$, $c > 0$ and the discriminant $d = b^2 - 4ac < 0$. It is primitive, if a, b, c have no common factor other than 1. Let the integers α, β, γ, δ be such that $\alpha\delta - \beta\gamma = \pm 1$. When x and y vary over all the integers, the set of values taken by the quadratic forms $Q(x, y)$ and

$$Q'(x, y) := Q(\alpha x + \beta y, \gamma x + \delta y) = a'x^2 + b'xy + c'y^2$$

are identical; the two forms have the same discriminant d, they are said to be equivalent. The number $h(d)$ of inequivalent primitive positive definite quadratic forms with a given discriminant d is finite and is called the class function. (See Appendix A.53.)

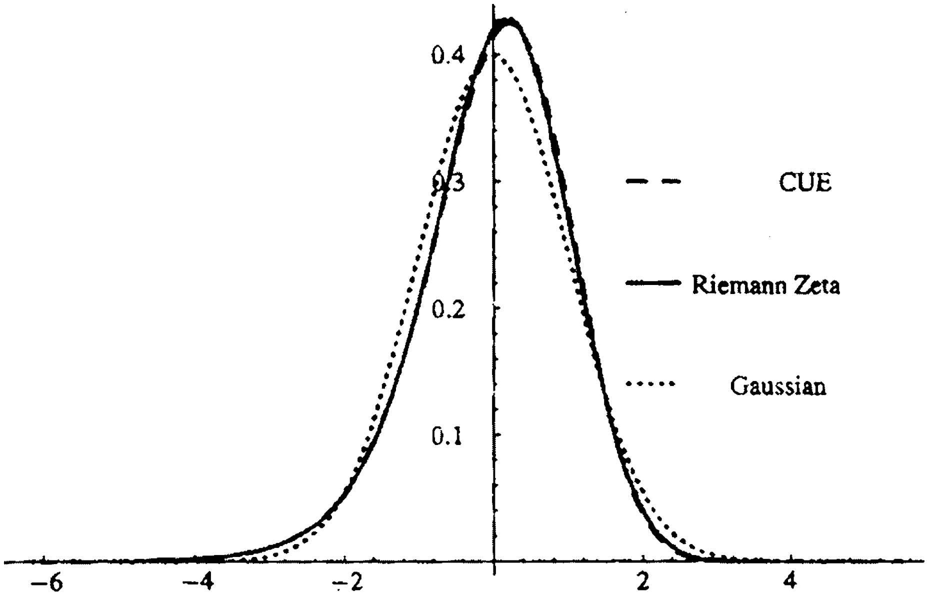

Figure 1.15. The density of values of $\log|f(x)|$ and the probability density of $\log|g(x)|$ for $T=$ the 10^{20}th zero of the Riemann zeta function and $N=42$ for matrices from the CUE; Eqs. (1.8.10), (1.8.11). The asymptotic limit, the Gaussian, is also shown. From Keating and Snaith (2000a), reprinted with permission.

For any positive definite quadratic form

$$Q(x,y)=ax^2+bxy+cy^2, \quad a,b,c \text{ integers}, \tag{1.8.14}$$

the Epstein zeta function is defined for $\operatorname{Re} z>1$ by

$$\zeta(z,Q)=\sum_{\substack{-\infty<m,n<\infty\\ m,n\neq 0,0}} Q(m,n)^{-z}, \tag{1.8.15}$$

and for other values of z by its analytical continuation. This Epstein zeta function satisfies the functional equation

$$\left(\frac{2\pi}{\sqrt{-d}}\right)^{-z}\Gamma(z)\zeta(z,Q)=\left(\frac{2\pi}{\sqrt{-d}}\right)^{z-1}\Gamma(1-z)\zeta(1-z,Q), \tag{1.8.16}$$

where d is the discriminant of Q, $d=b^2-4ac<0$. But in general it has no Euler product formula. Let $h(d)$ be the class function. If $h(d)=1$, then the RH seems to be

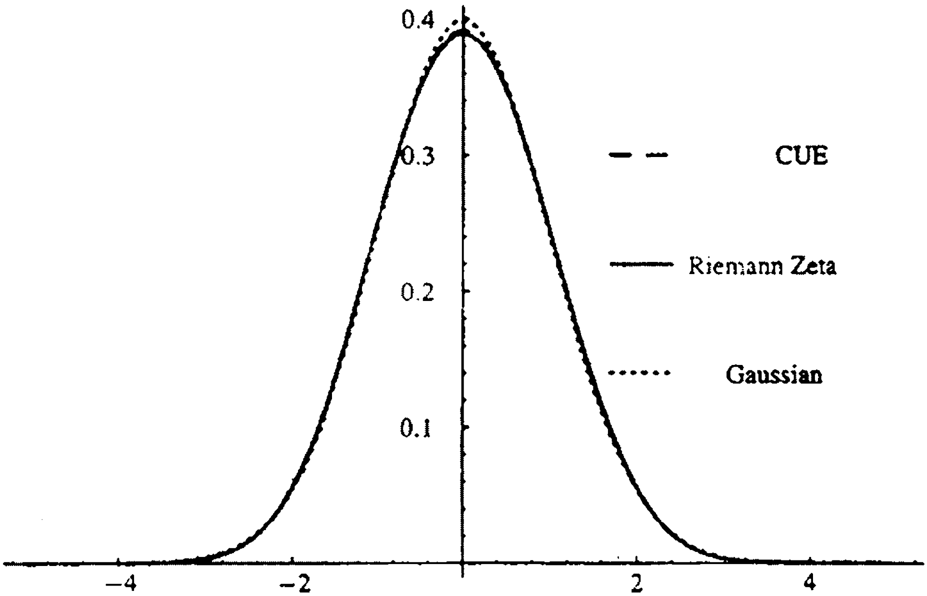

Figure 1.16. The density of values of the phase of $f(x)$ and the probability density of the phase of $g(x)$ for $T =$ the 10^{20}th zero of the Riemann zeta function and $N = 42$ for matrices from the CUE; Eqs. (1.8.10), (1.8.11). The asymptotic limit, the Gaussian, is also shown. From Keating and Snaith (2000a), reprinted with permission.

true, i.e. all the non-trivial zeros of $\zeta(z, Q)$ seem to lie on the critical line $\operatorname{Re} z = 1/2$. If $h(d) > 1$, then it is known that $\zeta(z, Q)$ has an infinity of zeros with $\operatorname{Re} z > 1$ (Potter and Titchmarsh, 1935).

In case $h(d) > 1$, it is better to consider the sum

$$\zeta_d(z) = \sum_{r=1}^{h(d)} \zeta(z, Q_r), \tag{1.8.17}$$

where the inequivalent primitive positive definite quadratic forms $Q_1, Q_2, \ldots, Q_r$ all have the same discriminant d. This function is proportional to a Dedekind zeta function, satisfies a functional equation and has an Euler product formula. It has no zeros with $\operatorname{Re} z > 1$, and it is believed that all its non-trivial zeros lie on the critical line $\operatorname{Re} z = 1/2$ (Potter and Titchmarsh, 1935).

For example, $h(-4) = 1$, has the unique primitive positive definite quadratic form up to equivalence $Q_1 = x^2 + y^2$, while the nonequivalent primitive positive definite quadratic forms $Q_2 = x^2 + 5y^2$ and $Q_3 = 3x^2 + 2xy + 2y^2$ both have $d = -20$, $h(-20) = 2$ (the quadratic form $3x^2 + 4xy + 3y^2$ taken by Potter and Titchmarsh is equivalent to our Q_3). Thus it is known that each of the functions

$$\zeta_2(z) = \sum_{\substack{-\infty<m,n<\infty \\ m,n\neq 0,0}} (m^2+5n^2)^{-z} \quad \text{and} \quad \zeta_3(z) = \sum_{\substack{-\infty<m,n<\infty \\ m,n\neq 0,0}} (3m^2+2mn+2n^2)^{-z}$$

has an infinity of zeros in the half plane $\operatorname{Re} z > 1$, while their sum $\zeta_2 + \zeta_3$ and the function

$$\zeta_1(z) = \sum_{\substack{-\infty<m,n<\infty \\ m,n\neq 0,0}} (m^2+n^2)^{-z}$$

are never zero in the same half plane. Moreover, it is believed that all the non-trivial zeros of $\zeta_1(z)$ and of $\zeta_2(z)+\zeta_3(z)$ lie on the critical line $\operatorname{Re} z = 1/2$.

What about their distribution? And in general, what about the distribution of the values of $\zeta_1(1/2+it)$ and of $\zeta_2(1/2+it)+\zeta_3(1/2+it)$ for real t? Do they have anything in common with the distribution of the values of the characteristic function of a random unitary matrix?

As another generalization one may consider for example,

$$Z_k(z) = \sum_{n_1,\ldots,n_k} (n_1^2+\cdots+n_k^2)^{-z}, \tag{1.8.18}$$

where the sum is taken over all integers $n_1,\ldots,n_k$, positive, zero and negative, except when $n_1=\cdots=n_k=0$. The zeta function, Eq. (1.8.18), is defined with respect to a hypercubic lattice in k dimensions. Equation (1.8.1) is the special case $k=1$ of Eq. (1.8.18). Instead of the hypercubic lattice one can take any other lattice and define a zeta function corresponding to this lattice. Or one could consider the Dirichlet series

$$L(z) = \sum_{n\in\mathcal{N}^r} P(n)^{-z}, \tag{1.8.19}$$

where $P(x)$ is a polynomial in r variables $x_1,\ldots,x_r$ with real non-negative coefficients, the sum is taken over all positive integers $n_1,\ldots,n_r$ except for the singular points of $P(x)$, if any. Thus if $P(x)=x$, we have $L(z)\approx\zeta(z)$; if $P(x)=a+bx$, a,b real and >0, then we have the Hurwitz zeta function

$$\zeta_{a,b}(z) = \sum_{n=0}^{\infty}(a+bn)^{-z}, \quad \operatorname{Re} z > 1. \tag{1.8.20}$$

Such Dirichlet series or general zeta functions, though do not have a functional equation in general, they have many important properties similar to those of the Riemann or Hurwitz zeta functions. For example, $\Gamma(z)\zeta_{a,b}(z)$ (with Hurwitz function $\zeta_{a,b}(z)$, Eq. (1.8.20), and the gamma function $\Gamma(z)$), has simple poles at $z = 1, 0, -1, -2, \dots$ with residues rational in (a, b). The "non-trivial" zeros of the hypercubic lattice zeta function $Z_k(z)$, Eq. (1.8.18), for example, for $k = 2, 4, 8, \dots$, may be expected to be on the critical line $\operatorname{Re} z = 1/2$. What about their distribution?

Little is known about the zeros of such general zeta functions or of the Dirichlet L-series, even empirically. It will be interesting to see what is the distribution of their (non-trivial) zeros and if they have any thing to do with the GUE results.

An eventual proof of the RH for the zeta function and its generalizations is important for at least two reasons; (i) it has remained a challenge for such a long time, (ii) it will imply a significant improvement in the estimation of various arithmetic functions, such as $\pi(x)$, the number of primes $\leqslant x$. However, one needs some further knowledge of the distribution of the γ_n to answer other questions like how the largest gap between two consecutive primes $\leqslant x$ increases with x. Or can one approximate real numbers by rational numbers whose numerator and denominator are both primes; more specifically, is it true that for every irrational number θ there are infinitely many prime numbers p, q such that $|\theta - p/q| < q^{-2+\varepsilon}$.

Katz and Sarnak (1999) found some particular L-series whose zeros would mimic not the eigenvalues of a random unitary matrix (circular unitary ensemble) but those of a random real orthogonal matrix or those of a random symplectic matrix, not related to the circular orthogonal ensemble (COE) or the circular symplectic ensemble (CSE). According to knowledgeable persons no number theoretic functions have yet been found that show COE or CSE statistics.

The fluctuation properties of the nuclear energy levels or that of the levels of a chaotic system are quite different, they behave, in the absence of a strong magnetic field, as the eigenvalues of a matrix from the orthogonal ensemble, cf. Chapter 7.

1.9 Things Worth Consideration, But Not Treated in This Book

Another much studied problem is that of percolation, or a random assembly of metals and insulators, or that of normal and super-conductors. Consider again a 2- or 3-dimensional lattice. Let each bond of the lattice be open with probability p and closed with probability $1 - p$. Or p and $1 - p$ may be considered as the respective probabilities of the bond being conducting or insulating or it being a normal and a super-conductor. The question is what is the probability that one can pass from one end to the other of a large lattice? Or how conducting or super-conducting a large lattice is? It is clear that if $p = 0$, no bond is open, and the probability of passage is zero. What is not so clear is that it remains zero for small positive values of p. Only when p increases and passes beyond a certain critical value p_c, this probability attains a non-zero value. It is known from numerical studies that the probability of passage is proportional to $(p - p_c)^\alpha$ for $p \geqslant p_c$, where the constant α, known as the critical index, is independent of the lattice and depends only on the dimension. The mathematical problem here can be characterized in its simplest form as follows. Given two fixed matrices A and B of the same order, let the matrix M_i for $i = 1, 2, \ldots, N$ be equal to A with probability p and equal to B with probability $1 - p$. How does the product $M = M_1 M_2 \ldots M_N$ behave when N is large? For example, what is the distribution of its eigenvalues? Of course, instead of just two matrices A and B and probabilities p and $1 - p$, we can take a certain number of them with their respective probabilities, but the problem is quite difficult even with two. No good analytical solution is known.

The actual problem of percolation or conduction is more complicated. Numerical simulation is performed on a ribbon of finite width, ≈ 10, for two dimensions (see Figure 1.17), or on a rod of finite cross-section, $\approx 5 \times 5$, for three dimensions. If all the resistances of the lattice lines are given, then we can explicitly compute the currents flowing through the open ends $B_1, B_2, \ldots, B_n$, when an electric field is applied between A and these ends. Adding one more layer in the length of the ribbon, the currents in the ends $B_1, \ldots, B_n$ change and this change can be characterized by an $n \times n$ matrix T, called the transfer matrix. This matrix T gives the new currents in terms of the old ones; it depends on the resistances of the lattice lines we have just added. As these resistances are random, say c with probability p and ∞ with probability $1 - p$, the matrix T is random; its matrix elements are constructed from the added random resistances according to known laws. For a length L of the ribbon, the currents in the ends $B_1, \ldots, B_n$ are determined by the product of L matrices $T_1 T_2 \ldots T_L$. Computing the spectrum of the product of L matrices for L large, $\approx 10^6$, and studying how it depends on L and on the order n, one can get a fair estimate for the critical index α.

Chaotic systems are simulated by a particle free to move inside a certain 2-dimensional domain. If the plane is Euclidean, then the shape of the domain is chosen so that the Laplace equation in the domain with Dirichlet boundary conditions has no analytic solution. This usually happens when the classical motion of the particle in the domain with elastic bouncing at the boundary has no, or almost no, periodic orbits. If the plane has a negative curvature, the boundary is usually taken to be polygonal with properly chosen angles. Sometimes one takes two quartic coupled oscillators with the potential

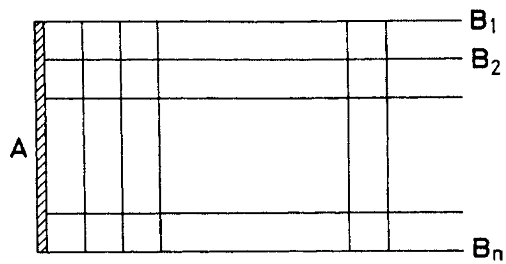

Figure 1.17. A model for the numerical simulation of conduction through a random assembly of conductors and insulators. At each step a new layer of randomly chosen conductors and insulators is added to the right and the subsequent change in the currents through the open ends $B_1, \ldots, B_n$ is studied.

energy $x^4 + y^4 + \lambda x^2 y^2$ as a model of a chaotic system. Why the spectrum of all these systems has to obey the predictions of the random matrix theory is not at all clear.

Still another curious instance has been reported (Balazs and Voros, 1989). Consider the matrix of the finite Fourier transform with matrix elements $A_{jk}(n) = n^{-1/2}\omega_n^{jk}$, $\omega_n = \exp(2\pi i/n)$. This $A(n)$ is unitary for every $n \geqslant 1$. Its eigenvalues are either ± 1 or $\pm i$. Let $B(n) = A^{-1}(2n)[A(n)\dot{+}A(n)]$ be the $2n \times 2n$ matrix, the product of the inverse of $A(2n)$ and the direct sum of $A(n)$ and $A(n)$. This $B(n)$ is also unitary, so that its eigenvalues lie on the unit circle. The matrix $B(n)$ is in some sense the quantum analog of the baker's transformation.

Now when n is very large, the eigenvalues of $B(n)$ and the eigenvalues of a random matrix taken from the circular unitary ensemble (see Chapters 10 and 11) have almost the same statistical properties. Instead of the above $B(n)$, we could have considered $B(n) = A^{-1}(3n)[A(n)\dot{+}A(2n)]$ or $B(n) = A^{-1}(3n)[A(n)\dot{+}A(n)\dot{+}A(n)]$, and the result would have been the same. Why this should be so, is not known.

Fluctuations in the electrical conductivity of heterogeneous metal junctions in a magnetic field and in general the statistical properties of transmission through a random medium is related to the theory of random matrices. In this connection see the articles by Al'tshuler, Mailly, Mello, Muttalib, Pichard, Zano and coworkers.

The hydrogen atom in an intense magnetic field has a spectrum characteristic of matrices from the Gaussian unitary ensemble, see Delande and Gay (1986).

In the last decade random matrix models have found applications in subjects as divers as quantum gravity, string theories, quantum chromo dynamics, counting of certain knots and links, counting of tree graphs (i.e. graphs having no closed paths) having special structures, and others. We will not deal with them either for lack of competence. Fortunately good review articles dealing with the vast literature on these topics are available. See for example, di Francesco et al. (1995), 2-D gravity and random matrix models, *Phys. Rep.* **254**, 1–133; Guhr et al. (1998), Random matrix theories in quantum physics: common concepts, *Phys. Rep.* **299**, 189–425; Katz and Sarnak (1999), *Random Matrices, Frobenius Eigenvalues and Monodromy*, Amer. Math. Soc., Providence, RI and many references therein.