SIAM Review
Vol. 9, No. 1, January, 1967

RANDOM MATRICES IN PHYSICS*

EUGENE P. WIGNER†

Introduction. It has been observed repeatedly that von Neumann made important contributions to almost all parts of mathematics with the exception of number theory. He had a particular interest in those parts of mathematics which formed cornerstones of other, more empirical sciences, such as physics or economics. A whole new discipline grew out of his theory of games, and it is hard to conjure up a picture of modern United States industry without the computing machines which he espoused. The subject about which I wish to talk to you today is at the crossroads of two of von Neumann's principal interests: it deals with matrices of very large dimensions in which he became interested in connection with his development of computers, and it resembles statistical mechanics, to which he contributed most among the physical theories. This closeness of my subject to von Neumann's interests is also the reason for my choosing it for today's discussion. This discussion will contain very little that is new. As for earlier reviews, there are at least two very good ones: one by Charles Porter, forming an introduction and summary to a collection of papers on the role of random matrices in physics [1], the second a more elaborate one by M. L. Mehta, based on his lectures at the Indian Institute of Technology in Kanpur [2].

There is another reason for my choice of subject. The theory of random matrices, though initiated by mathematicians and in particular statisticians [3], [4], [5], [6], [7], has made large strides in the hands of physicists. The names of Mehta, Gaudin and Dyson come to one's mind most easily. Reading these papers gave me much pleasure—they contain beautiful, though old-fashioned, mathematics. I would like to share some of this pleasure with you. Second, however, a number of problems has turned up, apparently too difficult for us amateur mathematicians. I would like to share these problems with you also.

Origin of the problem. For reasons which I hope will become evident in the course of the discussion, I will proceed in my review pretty much in the antihistoric order. However, the reason for the interest of physicists in random matrices should be stated first. This was articulated, most eloquently, by Dyson [8]:

> "Recent theoretical analyses have had impressive success in interpreting the detailed structure of the low-lying excited states of complex nuclei. Still, there must come a point beyond which such analyses of individual levels cannot usefully go. For example, observations of levels of heavy nuclei in the neutron-capture region give precise information concerning a stretch of levels from number N to number $(N + n)$, where N is an integer

* The seventh John von Neumann Lecture delivered at the 1966 SIAM Summer Meeting on August 31, 1966, in New Brunswick, New Jersey. Received by the editors September 22, 1966.

† Department of Physics, Princeton University, Princeton, New Jersey.

of the order of 10^6. It is improbable that level assignments based on shell structure and collective or individual-particle quantum numbers can ever be pushed as far as the millionth level. It is therefore reasonable to inquire whether the highly excited states may be understood from the diametrically opposite point of view, assuming as a working hypothesis that all shell structure is washed out and that no quantum numbers other than spin and parity remain good. The result of such an inquiry will be a statistical theory of energy levels. The statistical theory will not predict the detailed sequence of levels in 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.

"In ordinary statistical mechanics a comparable renunciation of exact knowledge is made. By assuming all states of a very large ensemble to be equally probable, one obtains useful information about the over-all behavior of a complex system, when the observation of the state of the system in all its detail is impossible. This type of statistical mechanics is clearly inadequate for the discussion of nuclear energy levels. We wish to make statements about the fine detail of the level structure, and such statements cannot be made in terms of an ensemble of states. What is here required is a new kind of statistical mechanics, in which we renounce exact knowledge not of the state of a system but 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. The problem then is to define in a mathematically precise way an ensemble of systems in which all possible laws of interaction are equally probable."

This last point can, perhaps, bear some elaboration. A system in quantum mechanics can be characterized by a self-adjoint linear operator in Hilbert space, its Hamilton operator. We think of this as a hermitian matrix of infinitely many dimensions, having somehow introduced a coordinate system in Hilbert space. Hence, the ensemble of systems can be thought of as an ensemble of hermitian matrices, and we think of matrices of very high dimensionality rather than infinite matrices. To this point I will return later. However, as Dyson said, the problem is still with us what ensemble of such matrices to consider. In this regard, there is a profound difference between the ensembles of statistical mechanics and our ensembles. In statistical mechanics one considers a system: that is, particles with definite masses, and forces acting between them. The state of such a system can be specified, in classical mechanics, by the coordinates q_i and momenta p_i of the particles, all functions of time. One then asks for the time average of continuous functions of the coordinates and momenta,

$$\lim_{T\to\infty} \frac{1}{T} \int_t^{t+T} f(q_1(\tau), q_2(\tau), \cdots, p_1(\tau), p_2(\tau), \cdots)\, d\tau. \tag{1}$$

Since the coordinates and momenta are, by Newton's equations of motion, completely determined as functions of time by their initial values, the averaging process is an entirely definite one. That, apart from rare exceptions, the average

exists, that it is a function only of the constants of motion, such as energy, but independent of the other initial conditions, is a theorem first proved by von Neumann, Koopman and G. D. Birkhoff [9], [10], [11], [12], [13]. What I wish to emphasize, however, is only that the averaging process is entirely definite; it is a time average, and equal time intervals have equal weights.

In contrast to this, the averaging process in the physical applications of the theory of random processes is not defined. One again deals with a specific system, with its proper (though in many cases unknown) Hamiltonian, yet pretends that one deals with a multitude of systems, all with their own Hamiltonians, and averages over the properties of these systems. Evidently, such a procedure can be meaningful only if it turns out that the properties in which one is interested are the same for the vast majority of the admissible Hamiltonians. The first question, then, is what are the admissible Hamiltonians, and what is the proper measure in the ensemble of these Hamiltonians. The second question is, of course, whether, given the ensemble of admissible Hamiltonians with a proper measure, the properties in which we are interested are common for the vast majority of them.

The experimental situation which prompted the interest of physicists in random matrices is illustrated in Figs. 1 and 2. The first of these shows the situation which did *not* prompt it. This is a level scheme of three nuclei with mass number 10. The horizontal lines represent stationary states, the vertical position is the characteristic value of the Hamilton operator to which the characteristic vector belongs. It is the energy of the stationary state. As the figure shows, some of the energy values are common to all three nuclei; some are present only in the Hamiltonian of one of the nuclei, that at the center. J is the angular momentum quantum number of the state. The diagram of Fig. 1 was obtained experimentally; a physicist interested in these matters knows most of the numbers of this table, and their relations to each other, by heart. The diagram, to repeat it, shows the low, i.e., small, characteristic values of the three Hamiltonians, those of Be^{10}, B^{10} and C^{10}. The energy difference between the two lowest states of B^{10} is 0.717 Mev.

Fig. 2 is the level diagram of another nucleus, U^{239}, in the energy interval

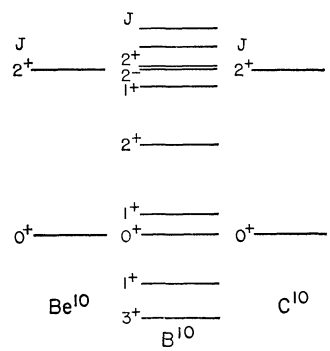

FIG. 1

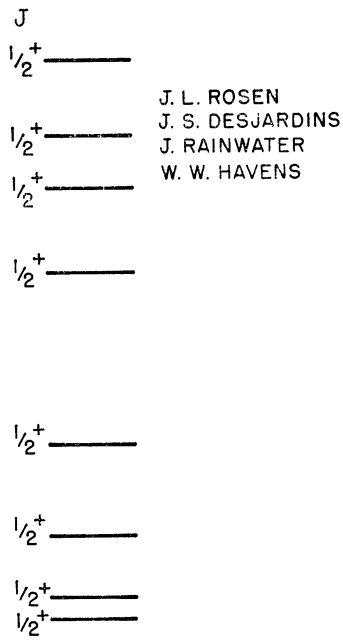

FIG. 2

between 4.7834 Mev from the lowest level and 4.7835 Mev therefrom. Some physicists, myself included, know a few of these energy values by heart because they happen to play an important role in nuclear chain reactors. However, no one is familiar with the levels in the next interval of the same width. As Dyson said in the passage quoted above, only the statistical properties of these levels are of interest.

What are these statistical properties? First, one would like to know how many energy levels there are per unit energy interval, that is, the density of the characteristic values of the Hamiltonian. Second, one would like to be able to describe what might be called subtleties of the arrangement of the characteristic values: the probability of a given distance to the nearest neighbor, the probability of two consecutive distances to assume definite values, and so on. Third, though this is not suggested by the diagram, one would like to obtain the probability of a transition rate to assume a definite value. The most important transition is the emission of a neutron by the nucleus, and it is on this that experimental data are available most abundantly. Altogether, the choice of questions is strongly influenced by the possibility of experimental study—a situation not unusual in physics. However, at least the first two questions, the average spacing of the characteristic values as function of the characteristic value itself and the distribution of the spacings around their average, are questions which everyone would naturally ask. The three questions now will be considered in succession.

The density of characteristic values. It must be admitted at the outset that

it is in this case that the independence of the result from the ensemble of Hamiltonians, and from the measure chosen therefor, is most nearly demonstrated, and that it is also the case in which the conclusion is least satisfactory, that is, agrees least with experimental observation.

It seems to be firmly established that the Hamiltonians of physics are, in the usual coordinate systems in Hilbert space, real. Hence, the admissible Hamiltonians are real symmetric matrices. An N-dimensional real symmetric matrix can be characterized by $N(N-1)/2$ real numbers H_{ik} with $i \leqq k$, and the measure in ensemble space is, therefore, a positive real-valued function of these variables,

$$P(H_{11}, H_{12}, H_{13}, \cdots, H_{22}, H_{23}, \cdots, H_{NN}), \tag{2}$$

which gives the number of matrices in the ensemble, of which the matrix elements are within the unit interval around the corresponding argument of P. This P function, therefore, defines the ensemble, and every positive-real-valued P function defines an ensemble. Needless to say, the problem of spectral density has not been solved for the general ensemble, characterized by an arbitrary P. Rather, in every case the independence of the distribution of some of the variables of P is assumed. If these variables are the matrix elements H_{11}, H_{12}, H_{13}, $\cdots$ of H themselves, P assumes the form

$$P(H_{11}, H_{12}, \cdots, H_{22}, H_{23}, \cdots, H_{NN}) = \prod_{i \leqq k} p_{ik}(H_{ik}). \tag{3}$$

In this case, and if the average values of all H_{ik} are zero, their second moments equal

$$\int p_{ik}(H')H'\,dH' = 0 \quad \text{and} \quad \int p_{ik}(H')H'^2\,dH' = v^2, \tag{4}$$

and all higher moments exist, the density of the characteristic values for very large N is given by the so-called semicircle law [14], [15],

$$\sigma(\lambda) = \begin{cases} \dfrac{(4Nv^2 - \lambda^2)^{\frac{1}{2}}}{4\pi v^2} & \text{if } \lambda^2 < 4Nv^2, \\ 0 & \text{if } \lambda^2 > 4Nv^2. \end{cases} \tag{5}$$

This distribution is very different from that of the real roots of an algebraic equation of order N. The ensemble in this case is obtained by considering the coefficients to be components of a vector of definite length which has equal probabilities for directions within equal solid angles [16]. Fig. 3 is a histogram of $\sigma(\lambda)$, due to N. Rosenzweig, obtained by diagonalizing 20 by 20 matrices, selected at random from an ensemble which I will discuss later. Not very surprisingly, the distribution approached a semiellipse—semicircle is actually a misnomer; the two axes do not even have the same dimension.

What is distressing about this distribution is that it shows no similarity to the observed distribution in spectra. The behavior at large positive λ is not relevant—what is known and what could be hoped to be reproduced by the

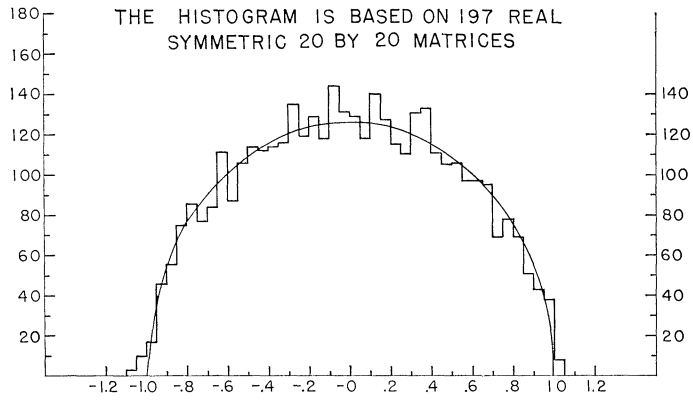

FIG. 3

ensemble is the distribution in the neighborhood of the lowest energy level. This appears to show, in nuclei, an exponential increase with energy. The density in the neighborhood of the lowest state is such that there are few levels per million electron volts. Around 5 Mev, on the other hand, there were several levels in an interval of 100 ev. At any rate, the density of the levels, as function of the energy, is convex from below, whereas the semicircle or semiellipse is concave. It could be surmised that the convex distribution applies only in the neighborhood of the lower range of the asymptotic formula, in the region where the asymptotic formula does not hold. The density in the range of the semicircle law is proportional to $\sqrt{N}$, the square root of the dimension of the random matrix. If it were proportional to a lower power of N outside the ellipse, this would not show in the asymptotic law but might explain the region in which, in actual nuclei, the density of levels increases fast. Hence, this region was more closely investigated by B. Bronk [17] for the so-called Wishart ensemble. These are ensembles in which the matrix elements are independent of each other and each shows a Gaussian distribution. Bronk found, much to the dismay of everyone, that the semicircle law is too accurate: there are, on the average, only about two levels outside its range.

Bronk's calculation applies only for the Wishart ensemble—that is, for a Gaussian distribution of the matrix elements—but, as we shall see, there are good reasons for preferring that distribution among all those in which the distributions of the various matrix elements are uncorrelated. Quite apart from this, it is clear that the existence of a reasonably large region in which the second energy-derivative of the density of levels is positive does not follow from the assumptions which we have made.

This, then, raises the problem of the proper ensemble of matrices to give a density of characteristic values approximating the observed distribution of these in nuclear spectra. Evidently, the ensemble that appears to be simplest from the mathematical point of view does not satisfy this criterion, and some further reference to the physical problem is needed.

There are two characteristics of the ensemble suggested by physical considerations which I wish to mention even though recourse to them alone has not proved successful. The first is the postulate that the matrix ensemble be invariant with respect to orthogonal transformations in Hilbert space. This is not a necessary postulate, but it appears to be a reasonable one. Incidentally, if this postulate is added to those made earlier, one is led to the Wishart-like ensemble in which every matrix element has the same Gaussian distribution around 0. This was known to statisticians Hsu [18], Nanda [19], and perhaps others,[1] but was rediscovered by the physicists Porter and Rosenzweig [20]. Hence, if one adopts this invariance, the assumption of the statistical independence of the matrix elements must be abandoned. On the other hand, simply omitting this postulate will not do: if an ensemble is invariant under orthogonal transformations, the frequency of matrices can yet be multiplied with any function

$$f(\text{trace } H, \text{ trace } H^2, \text{ trace } H^3, \cdots).$$

By a proper choice of this function, any desired density distribution of the levels can be obtained.

In order to reduce the freedom in the choice of the ensemble, one may recall that the Hamiltonian operator is unbounded above but bounded from below. This is the second characteristic I referred to. If the spectrum of the Hamiltonian were not bounded from below, there would be no lowest characteristic value. It should not matter much where the lower bound is; if it is assumed at zero, one is tempted to substitute

$$H = A^\dagger A, \quad A \text{ real, Wishart distributed.} \tag{6}$$

The last statement means that there are no statistical correlations between the matrix elements of A and that they all have the same Gaussian distribution. However, the matrix elements of H are not statistically independent in this case. Unfortunately, it turns out that the characteristic values of this ensemble, all positive, are distributed according to a quarter-circle law so that the density is quite large at the lower bound. I mention this unsuccessful attempt because the use of the aforementioned distribution is suggestive and because I do not recall having read an evaluation thereof before. Bronk has considered a similar problem and obtained the density of levels as function of energy—it was grossly unsatisfactory also [17].

This is a disappointing situation. However, perhaps we should not have expected otherwise. Operators in Hilbert space have properties which no finite-dimensional matrix has—in particular, a continuous spectrum. The true Hamiltonians all have a continuous spectrum, in addition to the discrete one, but the Hamiltonians considered in the theory of virtual levels are modified and have only a discrete spectrum. Somehow, these properties should play a role in the ensemble one chooses: this should contain only matrices which converge, in the

[1] Mehta [27] quotes S. N. Roy, Sankhyā (Dec. 1943), but without further details.

limit of infinite dimensionality, to operators with discrete spectra. It is not known what the spectrum of the infinite Wishart matrices is; it is not even known whether the self-adjoint operators with only discrete spectra form a manifold of zero measure. Von Neumann [21] has proved that there is, in any neighborhood of an operator with a continuous spectrum, an operator with a purely discrete spectrum, whereas the opposite is not true. However, this does not decide the question of which is the more "natural" situation: that of a discrete or that of a continuous spectrum. My guess goes for the continuous one and, if this should be correct, the ensembles which we have considered cannot be the right ones.

In order to avoid a misunderstanding, I should like to state that there is a model which does give the observed energy dependence of the level density. In fact, the exponential formula was obtained originally less from experiment than on the basis of this model [22], [23]. The model Hamiltonian is part of the Kronecker product of a large number of identical Hamiltonians; and it is true that the details of the common spectra of these are not very important as long as the common spectrum is a point spectrum and is bounded from below. The part of the Kronecker product that is considered is the antisymmetric one. The model is, in the language of the physicist, an independent particle model, or a model with small interactions. I would hesitate to call the resulting matrices random, but they do give an energy dependence of the level density which is at least similar to the observed one. What I hope is that there are more truly random ensembles which give a similar density of levels—the independent particle model was always considered to be a very special one.

Let me state, finally, before leaving the subject of level densities, that physicists have contributed to our mathematical knowledge of level densities more than I have just reviewed. First, the analogue of Wishart distributions was considered not only for real symmetric, but also for general hermitian matrices and for real, complex and quaternion matrices without any symmetry restriction. By quaternion matrices we mean matrices the elements of which are real quaternions. I wish to draw particular attention to the problem of the density of the characteristic values of complex matrices without symmetry. The solution of this problem by Ginibre [24] contains a number of shortcuts which I found fascinating. The result is, for very large N, a constant density in the complex plane within a circle of radius $(2N)^{1/2}2v$ centered at the origin, zero density outside. N is the dimension of the matrix, v the mean root square of the real and of the imaginary parts of its matrix elements.

Second, the results, as far as the Wishart-like distributions are concerned, were obtained by first calculating the joint distribution function of all characteristic values, i.e., the probability $P(\lambda_1, \cdots, \lambda_N)$ that the characteristic values be in unit intervals at $\lambda_1, \lambda_2, \cdots, \lambda_N$. In the case of complex matrices without symmetry, the probabilities refer to unit area in the complex plane. For Wishart-like distributions, these joint distribution functions all had the form

$$C \prod_{i>j} (\lambda_i - \lambda_j)^\beta \, e^{-\Sigma \lambda_i^2/2v^2}, \tag{7}$$

where the characteristic values are supposed to be ordered in increasing magnitude; $\beta = 1$ for real symmetric matrices, $\beta = 2$ for general hermitian matrices, $\beta = 4$ for symmetric quaternion matrices; $\beta = 2$ also for unrestricted complex matrices, but in this case the absolute value of λ^2 appears in the exponent and the absolute values of the differences $|\lambda_i - \lambda_j|$ appear as factors. C is a normalizing constant. From the joint distribution function the density of the characteristic values can be obtained by integration over all variables but one. The more general, not necessarily Wishartian, case was not handled in this way but by calculating the moments of the level density as function of energy [14], [15].

Statistics of spacings. This is the second problem raised by the appearance of the spectra at high energies, such as that of U^{239}. It is the problem which stimulated the most interesting, the most elegant, and the most successful work on the theory of random matrices, particularly in the hands of Mehta, Gaudin, and Dyson. The specific question which was in the foreground right from the start concerned the probability of a succession of spacings $S_1, S_2, \cdots, S_r$ between adjacent energy levels or, synonymously, adjacent characteristic values of a random matrix. These functions, which we shall denote by

$$\Psi_1(S_1), \quad \Psi_2(S_1, S_2), \quad \Psi_3(S_1, S_2, S_3), \cdots, \tag{8}$$

can all be obtained by integration with respect to the last variable from the next one of the series and can also be obtained by repeated integration from the joint distribution function (7) of all the characteristic values. However, as we well know from the similar problem in classical statistical mechanics, this is by no means an easy procedure. The most important among the spacing functions is the first one, probably because it requires least data to check it. All experimental and most theoretical work is directed toward the determination of this function, Ψ_1.

It was assumed, from the beginning, that the aforementioned distributions depend only on a crude characterization of the underlying matrix ensemble, that it will be the same for all "reasonable" ensembles of real symmetric matrices. It will also be the same for all reasonable ensembles of general hermitian matrices, though the Ψ for these will be different from the Ψ for ensembles of real symmetric matrices. When Dyson drew attention to quaternion matrices, it was natural to assume the same for these. Similarly, it was assumed that, if the actual spacings are measured in units of the average spacing, the distribution will be independent of the average spacing. In this sense, the functions $\Psi_1, \Psi_2, \cdots$ for real symmetric matrices are definite functions, and the same applies for these functions for hermitian matrices, and so on. One can probably even define such functions for the real, complex, and quaternion matrix ensembles without symmetry restrictions, except that the variables S should be complex numbers.

The postulate of the existence of an average spacing already implies that the energy range in which the average spacing of the levels is essentially constant is much larger than the average spacing itself. The same applies to the

reciprocal of the average spacing, the density. We have seen before that, in all ensembles considered, the density is proportional to the square root of the dimension of the matrix wherever the density is not zero. Since the logarithmic derivative of the spacing or of the density is inversely proportional to this square root, the condition of the existence of an average spacing will be satisfied for ensembles of matrices of high dimensionality wherever the density of the levels is not zero. The assumption mentioned then implies also that the Ψ are independent of energy if their variables are measured in terms of the local average spacing.

What is the evidence for the uniqueness of the functions Ψ? Certainly, the independence of Ψ_1 from the matrix ensemble chosen has not been proved for ensembles of as great generality as the semicircle law. The most general result, outside of the realm of Wishart-like distributions, refers to the boundary condition of the truncated Hamiltonian and shows, indeed, that Ψ_1 is invariant with respect to this boundary condition even though all characteristic values are changed by replacing the most usual boundary condition—zero derivative of the wave function at the nuclear surface—by postulating an arbitrary but fixed real ratio between the value and the derivative of the wave function [25]. This theorem, as stated, applied to a one-dimensional problem but it can be generalized to many dimensions by introducing a closed surface in Schrödinger's configuration space. On this surface, one can introduce a complete orthonormal set χ_ν and specify boundary conditions for each of them:

$$(\chi_\nu , \varphi) = \alpha_\nu(\chi_\nu , \operatorname{grad}_n \varphi). \tag{9}$$

If these equations are valid with constant (that is, energy independent) α_ν, then φ satisfies the boundary conditions. The usual boundary conditions set all $\alpha_\nu = \infty$, but the statistical distribution $\Psi_1(S)$ of the spacings can be shown, by a slight extension of an argument given before, to be independent of the α_ν. This applies, at sufficiently high energy, to any continuously differentiable closed surface and almost any orthogonal set χ_ν thereupon.

This is a rather suggestive result but comes nowhere near in generality to what one would like to have. I believe that it has, in fact, little to do with the conviction that, for instance, Ψ_1 for ensembles of real symmetric matrices is unique. I think the argument which is in our minds is that the joint distribution function for all characteristic values necessarily contains a factor $|\lambda_i - \lambda_j|$ corresponding to each pair (i, j) of characteristic values. This follows from the fact, also due at least partially to von Neumann [26], that two real constraints must be satisfied for two characteristic values of a real symmetric matrix to coincide.[2] It is believed that the other factors in the expression for the joint distribution function vary slowly with the distance of two roots, which is—as long as this distance remains in the interesting region—inversely proportional to

[2] Incidentally, the exponents of $|\lambda_i - \lambda_j|$ in the expressions for the joint distribution of the roots of hermitian and quaternion matrices can be understood in the same way: three real constraints must be satisfied for two roots of a general hermitian matrix to coincide, and five such constraints in the case of a quaternion matrix.

$\sqrt{N}$, hence very small. In fact, Dyson [8] replaced the ensemble of real symmetric matrices with the ensemble of unitary symmetric matrices. In itself, this appears harmless because a one-to-one correspondence can be established between real symmetric and unitary symmetric matrices S. However, he also assumed that the ensemble is invariant under the transformation $S \to W^T S W$ (the ^{T} denotes transpose), where W is any unitary matrix. This condition, if translated into a measure for the real symmetric matrices, is not the most natural one. Nevertheless, his joint distribution function for all roots is so similar to that of the Wishart-like ensemble that it can surely substitute for the latter.

Before embarking on a more serious discussion of the calculation of Ψ_1, let us first make a crude guess. If the location of the roots were independent, the probability ν that there is no root within the distance S from a given root would obey the equation

$$\nu(S + h) = \nu(S) - \nu(S)hc \tag{10}$$

for very small h, where c is the probability of a root in the unit interval. This leads to the differential equation

$$\frac{d\nu}{dS} = -c\nu, \tag{10a}$$

and hence $\nu = \exp(-cS)$ from which $\Psi_1 = -d\nu/dS = c \exp(-cS)$, i.e., the simple exponential law would follow. Actually, we know that the probability of a root right next to another one is proportional to the distance therefrom. This suggests, instead of (10),

$$\nu(S + h) = \nu(S) - \nu(S)hcS \tag{11}$$

or

$$\frac{d\nu}{dS} = -cS\nu, \quad \nu = e^{-cS^2/2}. \tag{11a}$$

From this, Ψ_1 could be obtained again by differentiation. This gives

$$\Psi_1(S) = \frac{\pi S}{2D} e^{-\pi S^2/4D^2}, \tag{12}$$

where c was expressed in terms of the mean spacing $D = \int \Psi_1(S) S \, dS$.

The preceding surmise for Ψ_1—a Gaussian multiplied by S—was made by me when asked, in the course of a meeting, to "guess" Ψ_1. Not much later, the calculation of the slope at $S = 0$ of the real Ψ_1 convinced me that the guess (12) was incorrect: the slope, $\pi^2/6D$, of the real Ψ_1 is, at $S = 0$, larger than that of (12) by a factor $\pi/3$. It was, therefore, quite surprising when, years later, Mehta [27] and Gaudin [28] succeeded in calculating Ψ_1 for the Wishart-like ensemble. They found that it differed so little from (12) that this was, after all, usable for practical comparisons. This is illustrated in Fig. 4. This shows, in addition to (12) and the true Ψ_1 calculated by Mehta and Gaudin, also the function $(4/D^2)S \cdot$

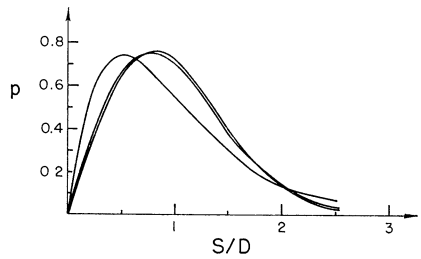

FIG. 4

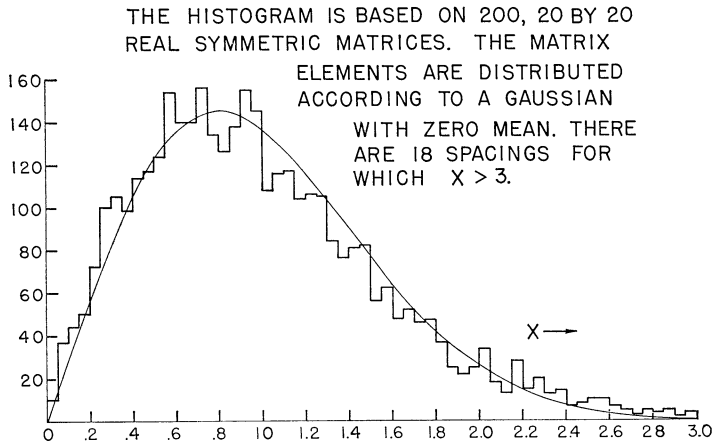

FIG. 5

exp $(-2S/D)$, demonstrating that a function which is proportional to S at small S and whose first moment is D can differ considerably from Ψ_1. The subsequent comparisons are all made with the inaccurate guess rather than Gaudin's accurate Ψ_1 function.

Fig. 5 shows the comparison between the empirical distribution for a Wishart-like ensemble and the approximate formula. The agreement is not surprising because the calculation was made with the ensemble underlying the work of Mehta and Gaudin. This does not, however, apply to the calculation illustrated in Fig. 6: the ensemble consisted here of matrices all the elements of which had random signs but the same absolute value. The semicircle law for the density of this ensemble has been proved, but it has not been proved that the spacing statistics is identical with that of the Wishart ensemble. This, nevertheless, seems to be true, confirming the surmise concerning the generality of the distribution of spacings. The calculations of the level spacing distribution which form the basis for these figures are due to N. Rosenzweig; they were done on a computing machine.

Calculation of the joint distribution function. Before turning to our last subject, the statistical distribution of matrix elements, it may be worth while to carry

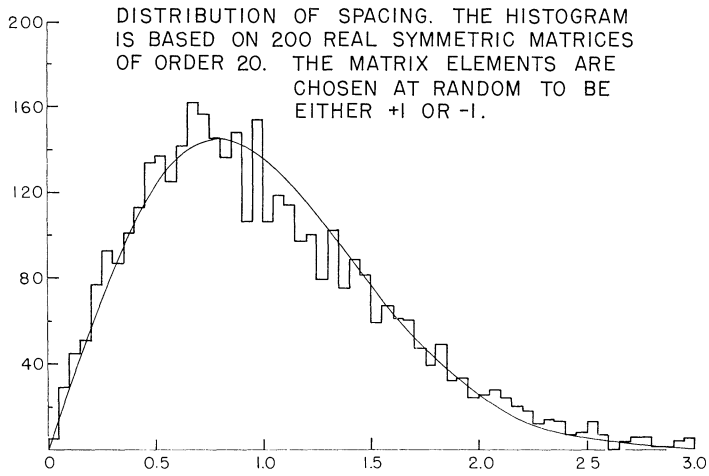

FIG. 6

out two of the important calculations concerning level spacings in detail. The present one concerns the joint distribution function of the roots of real symmetric matrices of high dimensionality. The ensemble considered is the same in the present and the next calculation: the probability function P of (2) is

$$(13) \qquad P(H_{11}, H_{12}, \cdots, H_{NN}) = C \exp\{-\tfrac{1}{2}\sum H_{ii}^2 - \sum_{i<k} H_{ik}^2\} \prod_{i \leqq k} dH_{ik}.$$

We shall not calculate constant factors such as the normalization constant C in intermediate expressions, because the final result can be normalized by the requirement that the integral over a probability function is 1.

The calculation of the joint distribution function of the characteristic values consists in introducing these, and some other variables, instead of the H_{ik}, into P and integrating over the "other variables". We start from

$$(14) \qquad H_{ik} = \sum_j \lambda_j R_{ji} R_{jk}.$$

Here the λ_j are the characteristic values, which should be assumed to be ordered, $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_N$, and R is a real orthogonal matrix. This is supposed to be specified by $N(N-1)/2$ parameters p over which we will want to integrate. We first note that, because of $H_{ik} = H_{ki}$,

$$(15) \qquad \frac{1}{2}\sum_i H_{ii}^2 + \sum_{i<k} H_{ik}^2 = \frac{1}{2}\sum_i H_{ii}^2 + \frac{1}{2}\sum_{i \neq k} H_{ik}^2 = \frac{1}{2}\sum H_{ik}^2 = \frac{1}{2}\sum \lambda_j^2,$$

so that the exponent in (13) does not depend on R or the parameters p. Hence, all we have to do is to calculate the Jacobian

$$(16) \qquad J = \frac{\partial(H_{11}, H_{12}, \cdots, H_{1N}, H_{22}, H_{23}, \cdots, H_{NN})}{\partial(\lambda_1, \lambda_2, \cdots, \lambda_N, p_1, p_2, \cdots, p_{N(N-1)/2})}$$

and integrate it over the p. We shall see, however, that J is a product,

$$J = \prod_{i>k} (\lambda_i - \lambda_k) j(p_1, \cdots, p_{N(N-1)/2}), \tag{17}$$

so that integration over the variables p need not be carried out since it only gives a constant factor.

The first N rows of the Jacobian matrix have the form

$$J_{ik;j} = R_{ji} R_{jk} \tag{18}$$

and are independent of the λ. The last $N(N-1)/2$ rows consist of

$$J_{ik;\nu} = \sum_j \lambda_j \left(\frac{\partial R_{ji}}{\partial p_\nu} R_{jk} + R_{ji} \frac{\partial R_{jk}}{\partial p_\nu} \right). \tag{18a}$$

They are linear in the λ. The Jacobian determinant is, therefore, a homogeneous polynomial of the λ, of order $N(N-1)/2$. In order to prove that it has the product form given before, we have to show only that it vanishes if any two λ coincide. We shall show that it vanishes for $\lambda_1 = \lambda_2$. Before proceeding with the calculation, we note that such vanishing is independent of the choice of the parameters p in the region in which both sets are unique, because the transition from one set of parameters to a new set merely multiplies J with the Jacobian of the old parameters with respect to the new ones. Hence, we can choose for one of the parameters, p_1, the angle of the rotation in the plane of the first two coordinates, such that multiplication of R with this rotation renders $R_{12} = 0$. If we denote this rotation by $R^{12}(p_1)$, the rotation R will assume the form

$$R^{12}(-p_1) C^{12}(p_2, p_3, \cdots, p_{N(N-1)/2}). \tag{19}$$

The parameters $p_2, p_3, \cdots, p_{N(N-1)/2}$ label the cosets of the subgroup of rotations in the 1–2 plane. Since, however, for $\lambda_1 = \lambda_2$, the H_{ik} become independent of p_1, the row in which the derivatives with respect to p_1 appear becomes 0 and the determinant vanishes. This shows that the Jacobian determinant contains a factor $\lambda_1 - \lambda_2$, and one can show in a similar way that it contains all $N(N-1)/2$ factors $\lambda_i - \lambda_k$. Since it is only of order $N(N-1)/2$, the whole λ dependence of the Jacobian is given by $\prod_{i>k} (\lambda_i - \lambda_k)$, and the joint probability function P in terms of the parameters λ, p becomes

$$P = e^{-\Sigma \lambda_i^2/2} \prod_{i>k} (\lambda_i - \lambda_k) j(p_1, \cdots, p_{N/2-1}). \tag{20}$$

Integration over the p now gives the joint distribution function of the roots except for its normalization factor C:

$$P_\lambda(\lambda_1, \lambda_2, \cdots, \lambda_N) = C \prod_{i>k} (\lambda_i - \lambda_k) e^{-\Sigma \lambda_i^2/2}, \quad \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_N. \tag{21}$$

Naturally, the same result can be obtained also by direct calculation.

Calculation of the probability $\Psi_1(S)$ that the spacing between adjacent levels is in the unit interval at S. As has been mentioned before, the function $\Psi_1(S)$ can be calculated from the joint distribution function P_λ by calculating the probability $\nu(S)$ that there be no root within an interval S beyond a given root; $\Psi_1(S)$ is, then, the negative derivative of this function. However, in order to

obtain $\nu(S)$, one has to carry out many integrations, and such multiple integrations are quite difficult. It was, therefore, a considerable accomplishment on the part of Mehta [27] and Gaudin [28] to calculate Ψ_1. Actually, their calculation proceeds by evaluating, first, the probability $p(\vartheta)$ that there is no root between $-\vartheta$ and ϑ. They assume that this is also the probability that there is no root in an interval of the length 2ϑ centered at random. The connection between Ψ_1 and p then can be obtained from the observation that an interval S between two roots does not permit any interval of length 2ϑ to be free of roots if $S < 2\vartheta$; it gives an interval $S - 2\vartheta$ for the center of a root-free interval of length 2ϑ, if $S > 2\vartheta$. Since the number of intervals between S and $S + dS$ in a long stretch L is $L\sigma\Psi_1(S)\,dS$, where σ is the density of the roots, we have

$$L\sigma \int_{2\vartheta}^{\infty} \Psi_1(S)(S - 2\vartheta)\,dS = Lp(\vartheta)\,. \tag{22}$$

Differentiating this twice with respect to ϑ one obtains

$$4\sigma\Psi_1(2\vartheta) = p''(\vartheta), \tag{22a}$$

so that the calculation of $p(\vartheta)$ gives Ψ_1 rather directly.

In order to calculate $p(\vartheta)$, Mehta and Gaudin first calculate the functional of u,

$$p_u = \int P_\lambda(\lambda_1, \cdots, \lambda_N) \prod_i u(\lambda_i)\,d\lambda_1 \cdots d\lambda_N\,. \tag{23}$$

We shall assume that u is an even function of λ. Evidently, for the function

$$\begin{aligned} u(\lambda) &= 1 \quad \text{if} \quad |\lambda| > \vartheta, \\ u(\lambda) &= 0 \quad \text{if} \quad |\lambda| < \vartheta, \end{aligned} \tag{23a}$$

the functional $p_u = p(\vartheta)$.

The product in the expression (21) for P_λ is the Vandermonde determinant, so that the integrand can be written as a determinant:

$$\begin{aligned} P_\lambda(\lambda_1, \cdots, \lambda_N) \prod_i u(\lambda_i) &= |\,q_\nu(\lambda_i)\,|, \qquad \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_N, \\ q_\nu(\lambda) &= e^{-\lambda^2/2} u(\lambda)\lambda^\nu. \end{aligned} \tag{24}$$

In fact, by multiplying the determinant $|\,q_\nu(\lambda_i)\,|$ on the left by a numerical nonsingular matrix (which introduces only a constant factor), the λ^ν can be replaced by polynomials a_ν, so that we can also write, instead of (24),

$$q_\nu(\lambda) = e^{-\lambda^2/2} u(\lambda) a_\nu(\lambda), \tag{24a}$$

where the $a_0, a_1, a_2, \cdots, a_{N-1}$ are any N linearly independent polynomials of degree not higher than $N - 1$. We shall make use of this possibility, but the a_ν will remain even or odd depending on whether ν is even or odd.

The difficulty of integrating the determinant (24) is caused by the form $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_N$ of the domain of integration. It would be much easier if

the integration could be extended over all values of the variables. This would be possible if P_λ were a symmetric function of its variables, but it is not.

Mehta and Gaudin overcome this difficulty by first integrating over all odd variables in the proper interval. The first such integration, over λ_1, must be extended from $-\infty$ to λ_2, and if we write

$$Q_\nu(\lambda) = \int_{-\infty}^{\lambda} q_\nu(\lambda_i')\, d\lambda_i', \tag{25}$$

it simply replaces the $q_\nu(\lambda_1)$ of the first column by $Q_\nu(\lambda_2)$. Integration over λ_3 will replace the $q_\nu(\lambda_3)$ of the third column by $Q_\nu(\lambda_4) - Q_\nu(\lambda_2)$—the limits of integration are λ_2 and λ_4. Adding the first column to this gives $Q_\nu(\lambda_4)$. Similarly, the $(2j-1)$th column will change, as a result of the integration over λ_{2j-1}, into a column of $Q_\nu(\lambda_{2j})$. One avoids some unessential complications by assuming that the dimension of the original matrix, $N = 2m$, is even. Then, the integrand becomes, after integration over $\lambda_1, \lambda_3, \cdots, \lambda_{2m-1}$,

$$C \begin{vmatrix} Q_0(\lambda_2) & q_0(\lambda_2) & Q_0(\lambda_4) & q_0(\lambda_4) & \cdots & q_0(\lambda_{2m}) \\ Q_1(\lambda_2) & q_1(\lambda_2) & Q_1(\lambda_4) & q_1(\lambda_4) & \cdots & q_1(\lambda_{2m}) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ Q_{2m-1}(\lambda_2) & q_{2m-1}(\lambda_2) & Q_{2m-1}(\lambda_4) & q_{2m-1}(\lambda_4) & \cdots & q_{2m-1}(\lambda_{2m}) \end{vmatrix}. \tag{26}$$

The integrand is now symmetric in the remaining variables and the integration can be extended over all of them from $-\infty$ to ∞ if one also divides by $m!$

Let us now look at the integral of the subdeterminants of the first two columns:

$$I_{\mu\nu} = \int_{-\infty}^{\infty} [Q_\mu(\lambda_2) q_\nu(\lambda_2) - Q_\nu(\lambda_2) q_\mu(\lambda_2)]\, d\lambda_2. \tag{27}$$

If μ and ν are both odd, Q_μ and Q_ν will be even, and since q_μ and q_ν are odd, $I_{\mu\nu}$ will vanish. The same is true, however, also if μ and ν are both even: in this case we can write

$$\begin{gathered} Q_\mu(\lambda) = i_\mu + Q_{\mu 0}(\lambda), \\ i_\mu = \int_{-\infty}^{0} q_\mu(\lambda)\, d\lambda, \qquad Q_{\mu 0}(\lambda) = \int_0^{\lambda} q_\mu(\lambda')\, d\lambda'. \end{gathered} \tag{27a}$$

$Q_{\mu 0}$ and the similar $Q_{\nu 0}$ are odd. Hence, for even μ and ν,

$$I_\mu = \int_{-\infty}^{\infty} i_u q_\nu(\lambda) - i_\nu q_\mu(\lambda)\, d\lambda + \int_{-\infty}^{\infty} [Q_{\mu 0}(\lambda) q_\nu(\lambda) - Q_{\nu 0}(\lambda) q_\mu(\lambda)]\, d\lambda.$$

The last integral vanishes because the Q therein are odd, the q even. The first integral is $i_\mu 2i_\nu - i_\nu 2i_\mu = 0$. Hence, $I_{\mu\nu}$ is different from zero only if μ is even, ν odd, or conversely. Since, furthermore, $I_{\mu\nu} = -I_{\nu\mu}$, we can restrict ourselves to the case that the first index is odd, the second even. We note, finally, that in the only interesting case, in which one of the indices is even, the other odd, the two terms for $I_{\mu\nu}$ in (27) are equal: since $q_\mu(\lambda) = Q_\mu'(\lambda)$, we have

$$\int_{-\infty}^{\infty} Q_\mu Q_\nu'\, d\lambda = [Q_\mu Q_\nu]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} Q_\mu' Q_\nu\, d\lambda = -\int_{-\infty}^{\infty} Q_\nu q_\mu\, d\lambda.$$

The integrated term vanishes because either Q_μ or Q_ν is even and thus vanishes not only for $\lambda = -\infty$ but also for $\lambda = \infty$. We can write, therefore,

$$I_{\mu\nu} = 2\int_{-\infty}^{\infty} Q_\mu(\lambda)q_\nu(\lambda)\,d\lambda \quad (\mu \text{ odd}, \nu \text{ even}). \tag{28}$$

Since μ is odd, Q_μ is even.

Let us return now to integrating (26). The determinant of (26) can be expanded into a sum of products of subdeterminants of two successive columns, and the integral of each of the subdeterminants gives a factor $I_{\mu\nu}$. Hence, the whole integral will be a sum of terms

$$I_{\mu_1\nu_1}I_{\mu_2\nu_2}\cdots I_{\mu_m\nu_m} \tag{29}$$

with the same coefficients but with appropriate signs. In these terms, the μ are all possible permutations of the numbers 1, 3, 5, $\cdots$, $2m-1$, the ν are all permutations of the numbers 0, 2, 4, $\cdots$, $2m-2$. This looks like an $m \times m$ determinant with rows labeled with the odd, columns with the even, numbers. One can check the signs and verify that it is a determinant. One finds, therefore, that

$$\begin{aligned}&\int_{-\infty}^{\infty} P_\lambda(\lambda_1, \lambda_2, \cdots, \lambda_{2m}) \prod_i u(\lambda_i)\,d\lambda_1\cdots d\lambda_{2m}\\ &= C_1\begin{vmatrix} I_{10} & I_{12} & \cdots & I_{1\,2m-2}\\ I_{30} & I_{32} & \cdots & I_{3\,2m-2}\\ \vdots & \vdots & & \vdots\\ I_{2m-1\,0} & I_{2m-1\,2} & \cdots & I_{2m-1\,2m-2}\end{vmatrix}.\end{aligned} \tag{30}$$

We shall carry the calculation one step further, restricting ourselves from now on explicitly to the $u(\lambda)$ of (23a). For this purpose, we have to make a choice of the polynomials $a_\nu(\lambda)$ which appear in (24a). The simplest choice is

$$\begin{aligned}a_0(\lambda) &= 1,\\ e^{-\lambda^2/2}a_\nu(\lambda) &= (-)^\nu h'_{\nu-1}(\lambda) \quad \text{for} \quad \nu \geqq 1,\end{aligned} \tag{31}$$

where the prime denotes differentiation and h_ν is the νth Hermite orthogonal function [29]

$$h_\nu(\lambda) = e^{\lambda^2/2}\frac{(-)^\nu}{[\sqrt{\pi}\,2^\nu\nu!]^{1/2}}\frac{d^\nu}{d\lambda^\nu}e^{-\lambda^2}. \tag{31a}$$

Then

$$\begin{aligned}q_\nu(\lambda) &= (-)^\nu u(\lambda)h'_{\nu-1}(\lambda) \quad \text{for} \quad \nu \geqq 1,\\ q_0(\lambda) &= e^{-\lambda^2/2}u(\lambda),\\ Q_\mu(\lambda) &= (-)^\mu h_{\mu-1}(\lambda) \quad \text{for} \quad \mu \text{ odd}, \quad |\lambda| \geqq \vartheta.\end{aligned} \tag{32}$$

Q_μ need not be calculated for even μ because we have an expression (28) for $I_{\mu\nu}$ in which only Q with an odd index appears. Similarly, $Q_\mu(\lambda)$ need not be calculated for $|\lambda| < \vartheta$ because all q_ν are zero between $-\vartheta$ and ϑ; actually

$(-)^{\mu}Q_{\mu}(\lambda) = h_{\mu-1}(\vartheta) = h_{\mu-1}(-\vartheta)$. One sees that q_ν is indeed an even function for even ν and an odd for odd ν.

The integration in (28) can be replaced by integration from ϑ to ∞, so that

$$I_{\mu\nu} = 4(-)^{\mu+\nu}\int_{\vartheta}^{\infty} h_{\mu-1}(\lambda)h'_{\nu-1}(\lambda)\,d\lambda \tag{33}$$

(μ odd, ν even), except that $I_{\mu 0} = 2\delta_{\nu-1\,0}$. This last integral can be decomposed into one, $I^0_{\mu\nu}$, from 0 to ∞ from which the integral $I^1_{\mu\nu}$ from 0 to ϑ must be subtracted: $I_{\mu\nu} = I^0_{\mu\nu} - I^1_{\mu\nu}$. The former integral is a well-known one:

$$I^0_{\mu\nu} = -4\int_0^{\infty} h_{\mu-1}(\lambda)h'_{\nu-1}(\lambda)\,d\lambda = 2^{1/2}(\delta_{\nu\,\mu-1}(\mu-1)^{1/2} - \delta_{\nu\,\mu+1}\mu^{1/2}). \tag{33a}$$

The corresponding part of (30) is

$$\| I^0_{\mu\nu} \| = C_1 \begin{Vmatrix} 2 & -2^{1/2} & 0 & 0 & \cdots & 0 \\ 0 & 4^{1/2} & -6^{1/2} & 0 & \cdots & 0 \\ 0 & 0 & 8^{1/2} & -10^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (4m-4)^{1/2} \end{Vmatrix}.$$

The determinant of this gives the integral of P_λ over all values of the variables λ. Since this is 1, we have

$$C_1 = 2^{-m}[(m-1)!]^{-1/2}. \tag{30a}$$

For the evaluation of the second part of the integral in (33), extending from 0 to ϑ, it is good to recall that the distance of the next root approaches 0 as $N^{-1/2}$ or $m^{-1/2}$ as N and m become very large. Hence, the ϑ of interest approaches 0 in the same manner. This should make it permissible to use the expression for h which is valid in the neighborhood of $\lambda = 0$. Such an expression can be derived by omitting in the differential equation for h,

$$-h_\nu'' + \lambda^2 h_\nu = (2\nu+1)h_\nu, \tag{34}$$

the $\lambda^2 h_\nu$ term. This makes $h_{\mu-1}$ and $h'_{\nu-1}$ proportional to $\cos(2\mu-1)^{1/2}\lambda$ and $\cos(2\nu-1)^{1/2}\lambda$, respectively, the proportionality constant to be determined from $h_{\mu-1}(0)$ and $h'_{\nu-1}(0)$. Since, furthermore, all the terms additional to (33a), resulting from the integral between 0 and ϑ, are small, they matter only in the aggregate, and one can use, for $h_{\mu-1}$ and $h'_{\nu-1}$, the expressions valid for large μ and ν. This is important only in the proportionality constants of the aforementioned cosines. One thus obtains for the part of $I_{\mu\nu}$ additional to (33a)

$$\begin{aligned} I^1_{\mu\nu} &= \frac{4}{\pi}(-)^{(\mu+\nu-1)/2}\left(\frac{\nu}{\mu}\right)^{1/4}\left\{\frac{\sin s\vartheta}{s} + \frac{\sin t\vartheta}{t}\right\}, \\ s &= (2\mu-1)^{1/2} + (2\nu-1)^{1/2}, \\ t &= (2\mu-1)^{1/2} - (2\nu-1)^{1/2}. \end{aligned} \tag{33b}$$

Since $|\, s^{-1}\sin s\vartheta\,| < \vartheta$, the part of $I_{\mu\nu}$ represented by $I^1_{\mu\nu}$ is indeed very small

as long as $\vartheta \sim N^{-1/2}$. Their aggregate effect is, of course, large nevertheless. Mehta and Gaudin evaluate the determinant by first subtracting suitable multiples of every column from the next one so as to render $\| I^0_{\mu\nu} \|$ diagonal, then multiplying the rows or columns with the corresponding factors of C_1 so that $\| I^0_{\mu\nu} \|$ is replaced by the unit matrix. These operations also alter $\| I^1_{\mu\nu} \|$. However, Gaudin succeeded in diagonalizing the modified $I^1_{\mu\nu}$. If its characteristic values are denoted by η_i, the value of the determinant (30) becomes

$$\int_{-\infty}^{\infty} P_\lambda(\lambda_1, \cdots, \lambda_{2m}) \prod_i u(\lambda_i)\, d\lambda_1 \cdots d\lambda_N = \prod_i (1 - \eta_i). \tag{35}$$

This was evaluated numerically and Ψ_1 obtained by (22a). This is the origin of the curve in Fig. 4 which is so close to that given by (12).

I have perhaps described the calculation of Ψ_1 in unnecessarily much detail. However, I do consider it a major accomplishment. At the same time, it is evident that there are several steps in the calculation which are not carried out in rigorous detail. For this reason, at least, one would like to see a simpler derivation.

We now turn to the last of our three items, the statistical distribution of the matrix elements.

Matrix elements. Mathematically, the calculation of the statistics of the matrix elements is the easiest of our three problems—in fact, it is an easy problem—but inhibitions prevented its discovery longer than now appears reasonable. It was, therefore, a major breakthrough when Scott [30] and Porter and Thomas [31] proposed the now generally accepted rule, the "Porter-Thomas distribution", actually essentially without any proof. According to this, the matrix elements in complex spectra show a Gaussian distribution. This is also well confirmed experimentally.

We have to do here with two self-adjoint operators, the Hamiltonian H, which defines the coordinate system, and the other operator M representing the physical quantity, such as dipole moment, in the matrix elements of which one is interested. We shall assume, first, that both are real. This means, more precisely, that there is a coordinate system in which all permissible operators of the physical quantity in which we are interested have real matrix elements. If we choose the coordinate axes in such a way that the states represented by them are time-inversion invariant, the matrix elements of the Hamiltonian will be real. As to the physical quantity in which we are interested, the matrix elements of its operator will also be real if the quantity—such as electric dipole moment—is also time-inversion invariant. This means that it retains its value if all the velocities are reversed. Most physically important quantities do have this property, or the opposite one, of reversing their signs if the directions of all the velocities are reversed. The operator of these is, naturally, also hermitian but purely imaginary and hence skew-symmetric. The operator of the magnetic moment is of this nature. However, though this will not be explained in detail, the calculation of the distribution function of the matrix elements of time-reversal invariant operators can be modified so as to be applicable for anti-invariant operators also.

Hence, the same distribution function can be expected for the matrix elements of these as for real, time-inversion invariant operators. It will be assumed, further, that the density of the characteristic values of the physical quantity in question is an even function of the characteristic value—a condition naturally fulfilled for operators which are anti-invariant with respect to time inversion. However, all known time-inversion invariant transition operators also share this property.

The following calculation uses the coordinate system whose axes are the characteristic functions of M, rather than of H. The characteristic values of this will be denoted by μ. The matrix element in question is, then,

$$m = \sum \mu_i x_i y_i , \tag{36}$$

where x_i and y_i are the coordinates of the states between which the matrix element is taken. These are characteristic vectors of H, hence x_i and y_i can be assumed to be real and m is real. The problem is to calculate the distribution of m for the Hamiltonians of the ensemble chosen. If this is rotationally invariant in Hilbert space, it amounts to calculating the distribution of the above expression for m when the vectors x and y move over a sphere but remain perpendicular to each other:

$$\sum x_i^2 = 1, \qquad \sum y_i^2 = 1, \qquad \sum x_i y_i = 0. \tag{36a}$$

This calculation can be carried out in detail, but since the individual terms of m are of the order $|\lambda|/N$, where N is the dimension of the space of the matrices, and since they have similar orders of magnitude and alternate in sign, one can infer already from the central limit theorem that the distribution is Gaussian:

$$(2\pi\overline{\mu^2})^{-1/2} e^{-m^2/2\overline{\mu^2}}.$$

If one introduces the transition rate, $\Gamma = m^2$, one obtains for this the Porter-Thomas distribution

$$(2\pi\overline{\mu^2}\Gamma)^{-1/2} e^{-\Gamma/2\overline{\mu^2}}. \tag{37}$$

Fig. 7 shows that this is quite well confirmed experimentally in the case of neutron emission. It has also been applied to various moments, not in nuclear but in atomic physics, and the agreement is satisfactory in these cases also. (Porter's book [1] contains articles applying the Porter-Thomas distribution to various types of transition rates.)

It may be worthwhile to add a few words about the basis and limitations of this formula. First, there is, of course, the same reservation which we had to make in all other cases: we used a finite-dimensional space, though a space of high dimensionality, rather than a Hilbert space. It appears that this problem is not quite as serious in this case as in the case of level densities, and Rosenzweig made it at least plausible that the replacement of Hilbert space by a space of high dimensionality is justified in this case [32].

The other assumption, that of the real nature of the matrices considered, is, however, relevant. If we had chosen an ensemble of complex hermitean rather than real symmetric matrices, the result would have been, for the transition

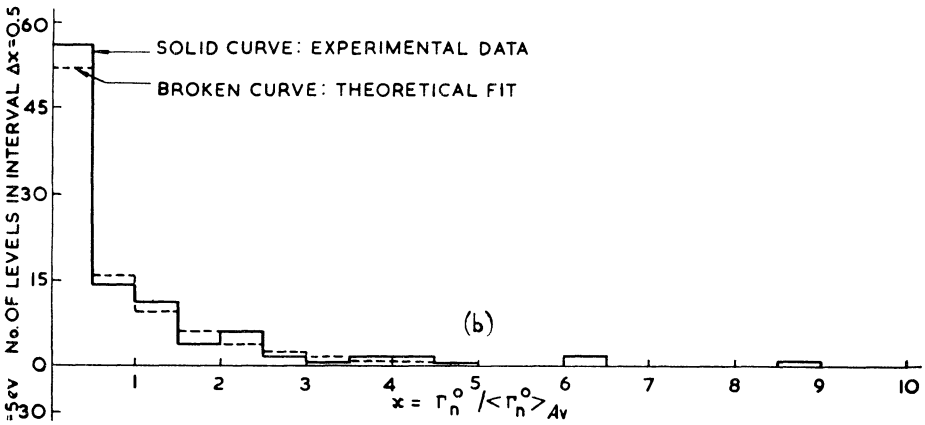

FIG. 7

probability Γ (which is proportional to the absolute square of the matrix element),

$$(\overline{\mu^2})^{-1} e^{-\Gamma/2\overline{\mu^2}}. \tag{38}$$

This would be easily distinguishable from the former distribution. One can derive the latter formula by decomposing H into a real symmetric and an imaginary skew-symmetric part. Both, separately, give a Porter-Thomas distribution, and the distribution for the sum of these is as given above.

This last point is, I believe, significant. There are experiments which indicate that the actual Hamiltonian is not strictly time-inversion invariant, that it has a small anti-invariant part. Nowhere would this manifest itself in the character of the wave functions as strongly as in the region where the levels are close to each other. Hence, an experimental check on the Porter-Thomas distribution may give an indication of the magnitude, or at least an upper limit, of the not-time-inversion invariant part of the Hamiltonian. Of course, the physical operator M, the matrix elements of which we consider, must be carefully chosen, and it must be ascertained that the observed transitions are due entirely either to an invariant or to an anti-invariant operator with respect to time inversion. I might have mentioned that there is an effect in the level-spacing distribution similar to the one discussed here: the not-time-inversion invariant part of the Hamiltonian would manifest itself in this case in an added repulsion of the levels.

This last observation will conclude the review of the recent contributions of physicists to the theory of random matrices and of the role of random matrices in the statistical theory of spectra. There remains, in the solution of almost every problem which we have tackled, a good deal that should be clarified, and almost all our derivations should be made more precise. More important than this would be the clarification of a number of rather general questions, some of which I wish to enumerate.

(a) Is the many-dimensional space as we use it a fair approximation to the Hilbert space? In particular, can one claim that, for any reasonable measure for the operators in Hilbert space, most operators have a pure continuous spectrum, or do they have a pure discrete spectrum?

(b) In all ensembles of matrices considered so far, the distributions of the matrix elements of some basic matrix are independent, i.e., there are no correlations between these matrix elements. Can one define more general ensembles, in particular ensembles of operators in Hilbert space rather than of many-dimensional matrices? Of particular interest would be ensembles of operators with a lower bound but with no upper bound.

(c) Under what conditions does the so-called semicircle law for the density of the characteristic values hold? Are there simple ensembles for which it is not valid and the second derivative of the density with respect to the characteristic value is positive over a range containing many characteristic values?

(d) What are the conditions for the validity of the Mehta-Gaudin distribution law for spacings? It has been derived, so far, only for the Wishart ensemble but seems to be valid much more generally.

I hope that mathematicians will help to clear up at least the more conceptual problems—von Neumann's example shows that the investigation of the exact foundation of the bases of physical theories can be fruitful for both the physical and the mathematical disciplines.

REFERENCES

[1] C. E. Porter, *Fluctuations of quantal spectra*, Statistical Theories of Spectra: Fluctuations, C. E. Porter, ed., Academic Press, New York, 1965.

[2] M. L. Mehta, *Lectures on Random Matrices and Applications*, Academic Press, New York, to be published.

[3] J. Wishart, *Generalized product moment distribution in samples*, Biometrika, 20 A (1928), pp. 32–52. (Correction: Ibid., 20 A (1928), p. 425.)

[4] R. A. Fisher, *The sampling distribution of some statistics obtained from nonlinear equations*, Ann. Eugenics (now Ann. Human Genetics), 9 (1939), pp. 238–249.

[5] A. T. James, *Normal multivariate analysis and the orthogonal group*, Ann. Math. Statist., 25 (1954), pp. 40–75.

[6] S. S. Wilks, *Mathematical Statistics*, John Wiley, New York, 1962, Chap. XI.

[7] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946, §29.9.

[8] F. J. Dyson, *Statistical theory of the energy levels of complex systems, I, II & III*, J. Mathematical Phys., 3 (1962), pp. 140–175.

[9] B. O. Koopman, *Hamiltonian systems and transformations in Hilbert space*, Proc. Nat. Acad. Sci. U.S.A., 17 (1931), pp. 315–318.

[10] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Ibid., 18 (1932), pp. 70–82.

[11] G. D. Birkhoff, *Proof of a recurrence theorem for strongly transitive systems*, Ibid., 17 (1931), pp. 650–655.

[12] ———, *Proof of the ergodic theorem*, Ibid., 17 (1931), pp. 656–660.

[13] G. D. Birkhoff and B. O. Koopman, *Recent contributions to the ergodic theory*, Ibid., 18 (1932), pp. 279–282.

[14] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math., 62 (1955), pp. 548–564.

[15] ———, *On the distribution of the roots of certain symmetric matrices*, Ibid., 67 (1958), pp. 325–326.

[16] M. KAC, *Probability and Related Topics in Physical Sciences*, vol. I, Interscience, New York, 1959, p. 11.

[17] B. V. BRONK, *Accuracy of the semicircle approximation for the density of eigenvalues of random matrices*, J. Mathematical Phys., 5 (1964), pp. 215–220. (Also Princeton dissertation, 1963.)

[18] P. L. HSU, *On the distribution of roots of certain determinantal equations*, Ann. Eugenics (now Ann. Human Genetics), 9 (1939), pp. 250–258.

[19] D. N. NANDA, *Distribution of a root of a determinantal equation*, Ann. Math. Statist., 19 (1948), pp. 47–57.

[20] C. E. PORTER AND N. ROSENZWEIG, *Statistical properties of atomic and nuclear spectra*, Ann. Acad. Sci. Fenn. Ser. A VI, no. 44 (1960). (Also reprinted in [1] above.)

[21] J. VON NEUMANN, *Charakterisierung des Spectrums eines Integraloperators*, Actualités Sci. Indust. (229), no. 13 (1935). (Reprinted in John von Neumann Collected Works, vol. IV, A. H. Taub, ed., Pergamon Press, New York, 1962.)

[22] J. M. BLATT AND V. F. WEISSKOPF, *Theoretical Nuclear Physics*, John Wiley, New York, 1952, §VIII-6.

[23] T. D. NEWTON, *Shell effects on the spacing of nuclear levels*, Canad. J. Phys., 34 (1956), pp. 804–829.

[24] J. GINIBRE, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys., 6 (1965), pp. 440–449.

[25] E. P. WIGNER, *On a class of analytic functions from the quantum theory of collisions*, Ann. of Math., 53 (1951), pp. 36–67.

[26] J. VON NEUMANN AND E. P. WIGNER, *Behavior of eigenvalues in adiabatic processes*, Phys. Zeitschr., 30 (1929), pp. 467–470.

[27] M. L. MEHTA, *On the statistical properties of the level-spacings in nuclear spectra*, Nuclear Phys., 18 (1960), pp. 395–419.

[28] M. GAUDIN, *Sur la loi limite de l'espacement des valeurs propres d'une matrice aléatoire*, Ibid., 25 (1961), pp. 447–458.

[29] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23, American Mathematical Society, New York, 1959, pp. 104–105.

[30] J. M. C. SCOTT, *Neutron widths and the density of nuclear levels*, Philos. Mag. (7), 45 (1954), pp. 1322–1331.

[31] C. E. PORTER AND R. G. THOMAS, *Fluctuations of nuclear reaction widths*, Phys. Rev., 104 (1956), pp. 483–491.

[32] N. ROSENZWEIG, *Anomalous statistics of partial radiation widths*, Phys. Lett., 6 (1963), pp. 123–125.