STATISTICAL PROPERTIES OF REAL SYMMETRIC MATRICES WITH MANY DIMENSIONS

E. P. WIGNER, *Princeton University*

1. Introduction. Mathematically, the energy levels are characteristic values (also called "roots") of hermitian operators; the stationary states are the corresponding characteristic vectors or eigenfunctions. However, one's attention is focused on very different qualities of these characteristic values and characteristic functions,

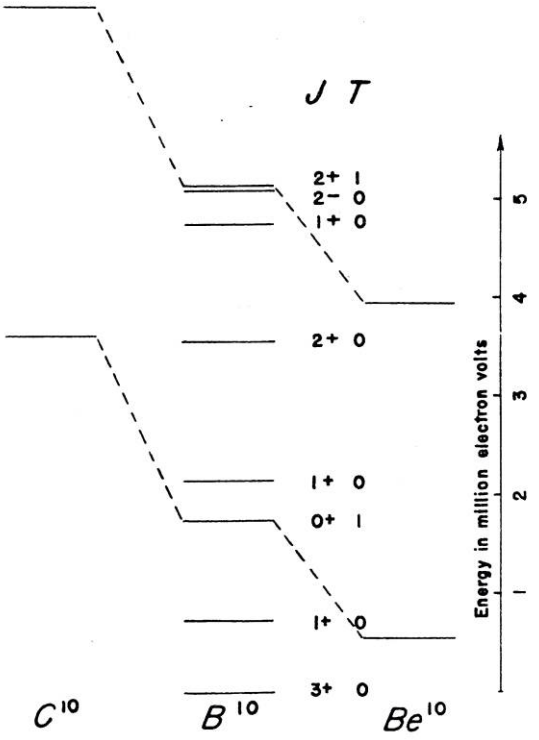

FIGURE 1

depending on the region of the spectrum. In the low energy region, near the smallest characteristic value, one would like to have a rather complete description of the stationary states and as complete an explanation of the exact values of the energy levels as possible. In the higher energy region there are so many energy levels that their exact position is difficult to catalogue, let alone explain. In this region, where the density of energy levels is high, that is, where the hermitian operator has many characteristic values per unit energy interval, one is interested only in the statistical properties of the spectrum.

Three diagrams are shown to illustrate the situation. The first of these refers to the energy levels of three nuclei, Be^{10}, B^{10}, and C^{10}. The diagram shows the eight lowest energy levels of B^{10} and the lowest two energy levels of Be^{10} and C^{10}. It gives the position of these energy levels **(1)**, their total angular momenta and parities. It indicates a connection among the energy levels of these nuclei. Not shown, but also of considerable interest, are transition probabilities between these levels. Such transition probabilities can be calculated if the characteristic functions associated with the characteristic values are known. Conversely, agreement between the observed transition probabilities, and the calculated values of these quantities, gives an indication of the accuracy of the calculated characteristic function.

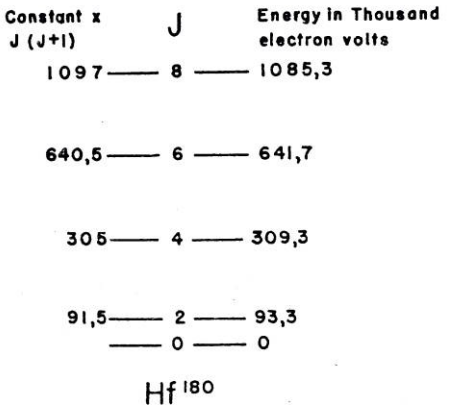

FIGURE 2

The second diagram shows five energy levels of Hf^{180}. This nucleus has a rotational band **(2)**: the angular momenta of the states shown are $J = 0, 2, 4, 6, 8$ units of $h/2\pi$. The energy values of these states should be proportional to $J(J + 1)$ where J is the angular momentum quantum number. In the present case, this formula is so accurate that the deviation could not be shown in the diagram.

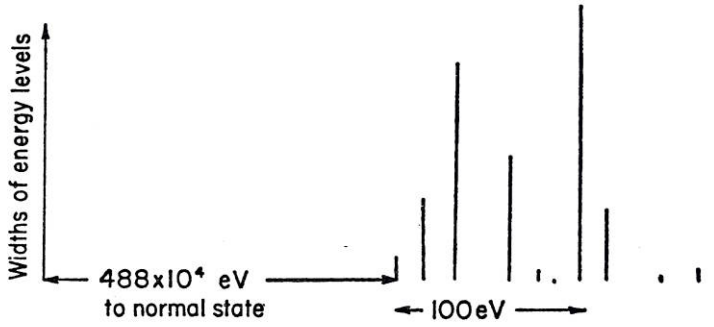

FIGURE 3

The third diagram shows those energy levels **(3)** of U^{239} the angular momentum quantum number of which is $\frac{1}{2}$. The diagram extends over 200 ev and its lowest point is about 4.88 million electron volts over the lowest energy level. It would be of little interest, and it would be quite impossible, to calculate the exact position of these energy levels. Their position is known with the accuracy shown only because the addition of a low energy neutron to a U^{238} nucleus gives a U^{239} nucleus with an energy of about 4.88 million electron volts. It is clear that, in the energy region illustrated in the third diagram, one will be interested principally in statistical statements, such as the density of the energy levels, their average "width," that is, the square of the wave function at the nuclear boundary, and so on. In addition to the average density (that is, the reciprocal of the average distance of adjacent energy levels) one is also interested in the probability for a certain spacing. This includes the question of whether the levels are, on the whole, equidistant or distributed according to a probability law. Besides the average width of the levels, one is interested in the distribution of the widths, that is,

the fraction of levels the widths of which are in unit interval at a certain width.

From the point of view of pure mathematics, the statistical questions may be even more interesting than the question of the exact properties of the low-lying energy levels because it is very likely that the statistical properties of a large class of real symmetric operators are in many respects identical. They should depend then on only a few parameters which are characteristic of the problem. One can restrict one's attention to the class of real symmetric operators because the energy operator is not only hermitian but real; this follows from the time inversion symmetry.

2. The distribution of the widths. One might have thought that the distribution of the widths of the energy levels—that is, of the square of the wave function at the nuclear surface—is much more difficult to determine than the distribution of the spacing of levels—that is, the intervals between adjacent characteristic values. This, however, did not prove to be the case; although a detailed theoretical argument is yet lacking, Scott **(4)**, and Porter and Thomas **(5)**, found, on the basis of experimental information furnished particularly by Hughes and Harvey **(6)** that the probability that the *value* of the wave function be between γ and $\gamma + d\gamma$ is

$$(2\pi\overline{\gamma^2})^{-\frac{1}{2}} \exp[-\gamma^2/2\overline{\gamma^2}]\, d\gamma. \tag{1}$$

Here $\overline{\gamma^2}$ denotes the average value of γ^2. Surely, (1) is the most simple distribution law that could be postulated. It gives for the distribution of the reduced width $\Gamma_0 = \gamma^2$

$$(8\pi\overline{\Gamma}_0\Gamma_0)^{-\frac{1}{2}} \exp[-\Gamma_0/2\overline{\Gamma}_0]\, d\Gamma_0. \tag{1a}$$

where $\overline{\Gamma}_0 = \overline{\gamma^2}$. The experimental data are compared with (1a), the so-called Porter-Thomas distribution, in Figure 4.

3. The distribution of the spacings. It is well known that the characteristic values of a real symmetric matrix (and also of a complex hermitian matrix) "repel" each other **(7)**. To be more precise: if the matrix elements depend on a number of continuous parameters, the dimensionality of the domain in the space of the parameters, for which a real symmetric matrix has a double root,

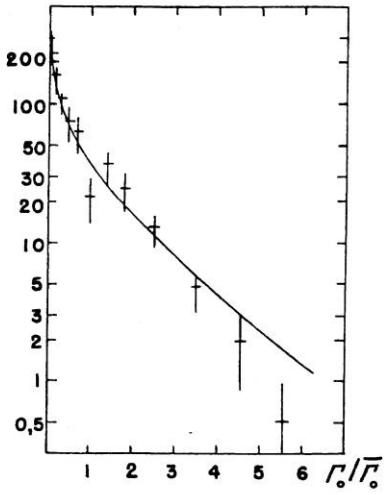

FIGURE 4. Distribution of the widths of energy levels. The curve is (1a), the crosses represent the experimental points. The vertical stem of the cross gives the probable experimental error.

is in general lower by *two* than the dimensionality of the parameter space itself. This can be demonstrated already in the case of a two-dimensional real symmetric matrix

$$\left\| \begin{matrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{matrix} \right\| \tag{2}$$

using the three matrix elements, a_{11}, a_{12}, a_{22}, themselves as parameters. The roots (characteristic values) of the matrix (2) are

$$\lambda = \tfrac{1}{2}(a_{11} + a_{22}) \pm \tfrac{1}{2}[(a_{11} - a_{22})^2 + 4a_{12}^2]^{\frac{1}{2}}. \tag{2a}$$

The two roots given by (2a) will be equal only if $a_{11} = a_{22}$ *and* $a_{12} = 0$. This represents a line in the three-dimensional space of the parameters. It follows from this that the probability for a spacing S (interval between adjacent roots) is proportional to S itself if S is very small as compared with the average spacing which is denoted usually by D. Some time ago, I surmised **(8)** that the probability law is

$$(\pi S/2D^2)e^{-\pi S^2/4D^2}dS. \tag{3}$$

It seems, however, that the experimental data agree better with the distribution

$$(4S/D^2)e^{-2S/D}dS. \tag{3a}$$

which has been suggested by Harvey and Hughes **(9)**. The comparison of the distributions (3) and (3a) with the observed distribution of four nuclei is given in Figures 5 and 6. The abscissa of the figures is $x = S/D$, the ordinate the actual number of spacings observed per unit x. The discrepancy does not seem large in either case. However, if one computes the probability of finding four spacings among 37 which exceed the average spacing by more than a factor 2.5, one finds that this is too unlikely (around 10^{-4}). Actually, the argument which leads to (3) is vague and it is not at all surprising that it fails. It does seem surprising, on the other hand, that no simple argument has yet been found to derive the distribution of the spacings and that one still has to resort to guesswork.

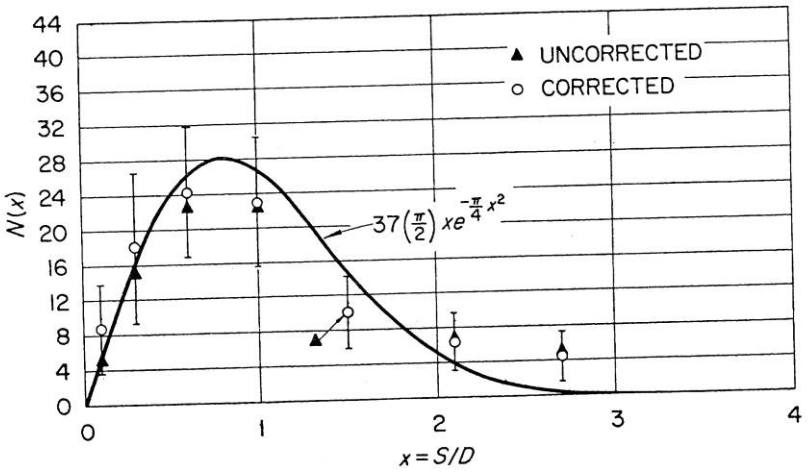

FIGURE 5.

4. The density of the characteristic values. The determination of the density of the characteristic values of a real symmetric matrix is much simpler than the determination of the distribution of the spacing. There are, to my knowledge, two approaches to this

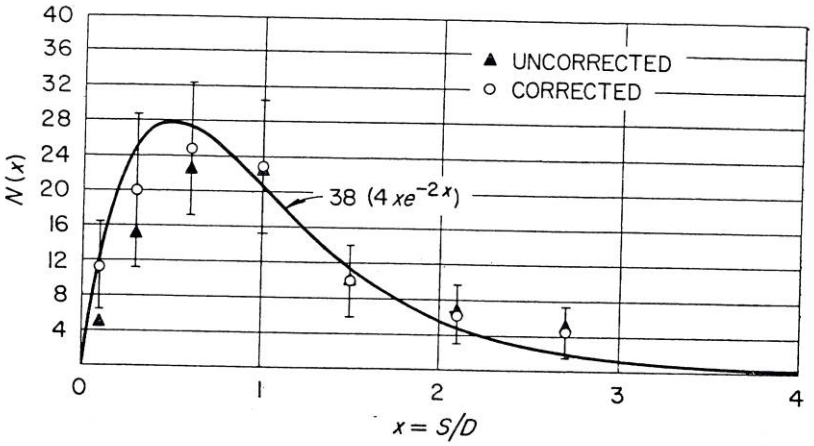

FIGURE 6.

problem. The first of these is based on the theory of the Wishart distribution **(10)** and was initiated, as far as I know, by Bargmann and Von Neumann. The second method determines the distribution by calculating its moments.

The method based on the Wishart distribution considers a particular set of real symmetric n dimensional matrices $||a_{ik}||$. The number of matrices in the set in which the $\frac{1}{2}n(n+1)$ matrix elements a_{ik} with $i \leqslant k$ have, within unit interval, the values specified by these numbers is proportional to

$$(4) \quad \exp[-\tfrac{1}{4}(a_{11}^2 + \ldots + a_{nn}^2) - \tfrac{1}{2}(a_{12}^2 + a_{13}^2 + \ldots + a_{n-1n}^2)].$$

In other words, the distribution of the $\frac{1}{2}n(n+1)$ matrix elements a_{ik} with $i \leqslant k$ are independent of each other and each a_{ik} has a normal (Gaussian) distribution. The root mean square of the distribution of the diagonal elements is $\sqrt{2}$, that of the off-diagonal elements is 1. If it is true that the relevant statistical properties of the characteristic values depend little on the set of matrices considered, the Wishart set, given by (4), can be considered representative.

The remarkable point about the Wishart set is that the probability for the characteristic values to be equal to $\lambda_1, \lambda_2, \ldots, \lambda_n$ (within unit interval) can be given in a closed form **(11)**. It is proportional to

$$(5)\quad P(\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{const} \prod_{i<k} |\lambda_i - \lambda_k| \exp[-\tfrac{1}{4} \sum \lambda_i^2].$$

One might conclude that it is easy to calculate, from (5), not only the number of characteristic values in unit interval at λ:

$$(5a)\qquad \sigma(\lambda) = n \int_{-\infty}^{\infty} \ldots \int P(\lambda, \lambda_2, \ldots, \lambda_n)\, d\lambda_2 \ldots d\lambda_n$$

but also the distribution of the intervals between adjacent roots. The latter could be done by integrating (5) over all but *two* of the λ. However, even the integration (5a), over the very simple domain indicated, is prohibitively difficult if n is a large number. It is necessary, therefore, to use less direct methods to obtain the quantities of interest. This was done by Bargmann and Von Neumann who obtained expressions for the distribution of the smallest, and of the largest, characteristic value.[1]

In order to calculate the over-all density $\sigma(\lambda)$ of the characteristic values, one may observe that the problem is, mathematically, very similar to problems encountered in statistical mechanics. One may use, therefore, the approximate methods of that discipline. If the density of the roots at λ is $\sigma(\lambda)$, the logarithm of the probability P is given by

$$(6)\quad \ln P(\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{const} - \sum_i \tfrac{1}{4}\lambda_i^2 + \sum_{i<k} \ln|\lambda_i - \lambda_k|.$$

It can be approximated by the following functional of σ

$$(6a)\quad [\sigma] = \text{const} - \tfrac{1}{4}\int d\lambda\, \lambda^2\sigma(\lambda) + \tfrac{1}{2}\int d\lambda \int d\mu\, \sigma(\lambda)\sigma(\mu) \ln|\lambda - \mu|.$$

All integrations have to be extended from $-\infty$ to ∞ and σ is so normalized that

$$(7)\qquad \int \sigma(\lambda) d\lambda = n.$$

The first integral in (6a) reproduces the first sum of (6) accurately if the number of dimensions n is sufficiently high. This is not true

[1]Personal communication of Dr. V. Bargmann. The smallest and largest characteristic values of the matrices of the Wishart set lie in the energy regions in which the density of the energy levels tends to zero as n increases. Hence, $\sigma(\lambda) = 0$ for both lowest and highest roots considered by Bargmann and Von Neumann. See also H. H. Goldstine and J. V. Neumann, Bull. Amer. Math. Soc., *53* (1947), 1021, and particularly Proc. Amer. Math. Soc., *2* (1951), 188.

of the second integral: this neglects correlations between the positions of the roots. However, these correlations can be expected to extend only over a few neighbouring roots and since the total number of roots is large and since the second integral in (6a) converges, one can expect that the effect of the correlations between the positions of the roots is negligible. The factor $\frac{1}{2}$ of the second integral in (6a) stems from the condition $i < k$ in (6).

One may now proceed, as in statistical mechanics, by postulating that the actual density makes the expression $[\sigma]$ a maximum consistent with the requirement (7) and the condition $\sigma(\lambda) \geqslant 0$. This leads to the integral equation

$$-\tfrac{1}{4}\lambda^2 + \int d\mu\, \sigma(\mu) \ln|\lambda - \mu| = C \tag{8}$$

where C is independent of λ. Actually, (8) has to hold only for λ for which $\sigma(\lambda) > 0$; one cannot add a negative increment to $\sigma(\lambda)$ where $\sigma(\lambda) = 0$ and (8) can not be derived for such λ. It is not difficult to solve (8). This will not be done, however, but the solution given and then verified.

Differentiation of (8) with respect to λ eliminates C. Before carrying it out, one must replace the integral by

$$\lim_{\epsilon \to 0} \left(\int_{-\infty}^{\lambda - \epsilon} d\mu + \int_{\lambda + \epsilon}^{\infty} d\mu \right) \sigma(\mu) \ln|\lambda - \mu|.$$

When this is differentiated with respect to λ, the terms arising from the differentiation of the limits drop out and only the derivative of $\ln|\lambda - \mu|$ remains. The integral becomes a principal value integral, or the arithmetic mean of two integrals in the complex plane, one contour C_+ above, the other contour C_- below the singularity at $\mu = \lambda$. Hence, (8) becomes

$$\int_{C+} \frac{d\mu\, \sigma(\mu)}{\lambda - \mu} + \int_{C-} \frac{d\mu\, \sigma(\mu)}{\lambda - \mu} = \lambda. \tag{8a}$$

Conversely, if (8a) is satisfied by some σ, (8) also will follow if σ is an even function. We try

$$\begin{aligned} \sigma(\mu) &= c(A^2 - \mu^2)^{\frac{1}{2}} && |\mu| < A \\ &= 0 && |\mu| > A. \end{aligned} \tag{9}$$

The sum of the two integrals on the left of (8a) can now be united to an integral going around all three singularities, at $\mu = -A, \lambda, A$.

If the path of integration is deformed into a circle of very large radius R, one can set

$$\mu = iRe^{-i\phi} \qquad d\mu = Re^{-i\phi}\, d\phi. \tag{10}$$

The left side of (8a) then becomes

$$\int_0^{2\pi} \frac{Re^{-i\phi} d\phi}{\lambda - iRe^{-i\phi}} c(A^2 + R^2 e^{-2i\phi})^{\frac{1}{2}} = \int_0^{2\pi} icRe^{-i\phi} \frac{(1 + A^2 e^{2i\phi}/R^2)^{\frac{1}{2}}}{1 + i\lambda e^{i\phi}/R} d\phi.$$

Since R is large, the integrand can be expanded into a power series of $1/R$. As was to be expected, only the term independent of R gives a non-zero contribution and this is $2\pi c\lambda$. Hence, (8a) gives

$$c = 1/2\pi \tag{9a}$$

and (7) gives $(1/2\pi)(\pi/2)A^2 = n$. Thus

$$\begin{aligned} \sigma(\mu) &= (2\pi)^{-1}(4n - \mu^2)^{\frac{1}{2}} && \mu^2 < 4n \\ &= 0 && \mu^2 > 4n \end{aligned} \tag{9b}$$

indeed satisfies (8a). Since it represents an even function of μ, it also satisfies (8) wherever $\sigma(\mu) \neq 0$. Note that the preceding evaluation of the left side of (8a) is valid only if $|\lambda| < A$, as only in this case does the path which surrounds the $\mu = -A$ and $\mu = A$ points contain the singularity $\mu = \lambda$. If $|\lambda| > A$, the integral on the left side of (8a) contains an additional term, from the singularity at λ and is not proportional to λ any more. As a result, (9b) does not satisfy (8a) or (8) for $\lambda^2 > A^2 = 4n$; but this is not necessary because $\sigma(\lambda) = 0$ in this case.

5. Generalization of (9b). The expression (9b) for the density of the characteristic values of the matrices of the Wishart set is identical with the expression obtained before, by another method, for the density of the characteristic values of another set of real symmetric matrices **(12)**. This latter set will be called random sign set. The diagonal elements of the matrices of this latter set all vanish, the non-diagonal elements are, with equal probabilities, $+1$ or -1. This suggests that the density of the characteristic values is given by (9b) under more general conditions, embracing both sets as special cases.

An analysis of the derivation of (9b) for the random sign set shows that this is indeed the case **(12)**. The conditions under which the earlier derivation given for (9b) is valid are:

If the path of integration is deformed into a circle of very large radius R, one can set

$$\mu = iRe^{-i\phi} \qquad d\mu = Re^{-i\phi}\, d\phi. \tag{10}$$

The left side of (8a) then becomes

$$\int_0^{2\pi} \frac{Re^{-i\phi} d\phi}{\lambda - iRe^{-i\phi}} c(A^2 + R^2 e^{-2i\phi})^{\frac{1}{2}} = \int_0^{2\pi} icRe^{-i\phi} \frac{(1 + A^2 e^{2i\phi}/R^2)^{\frac{1}{2}}}{1 + i\lambda e^{i\phi}/R} d\phi.$$

Since R is large, the integrand can be expanded into a power series of $1/R$. As was to be expected, only the term independent of R gives a non-zero contribution and this is $2\pi c\lambda$. Hence, (8a) gives

$$c = 1/2\pi \tag{9a}$$

and (7) gives $(1/2\pi)(\pi/2)A^2 = n$. Thus

$$\begin{aligned} \sigma(\mu) &= (2\pi)^{-1}(4n - \mu^2)^{\frac{1}{2}} && \mu^2 < 4n \\ &= 0 && \mu^2 > 4n \end{aligned} \tag{9b}$$

indeed satisfies (8a). Since it represents an even function of μ, it also satisfies (8) wherever $\sigma(\mu) \neq 0$. Note that the preceding evaluation of the left side of (8a) is valid only if $|\lambda| < A$, as only in this case does the path which surrounds the $\mu = -A$ and $\mu = A$ points contain the singularity $\mu = \lambda$. If $|\lambda| > A$, the integral on the left side of (8a) contains an additional term, from the singularity at λ and is not proportional to λ any more. As a result, (9b) does not satisfy (8a) or (8) for $\lambda^2 > A^2 = 4n$; but this is not necessary because $\sigma(\lambda) = 0$ in this case.

5. Generalization of (9b). The expression (9b) for the density of the characteristic values of the matrices of the Wishart set is identical with the expression obtained before, by another method, for the density of the characteristic values of another set of real symmetric matrices **(12)**. This latter set will be called random sign set. The diagonal elements of the matrices of this latter set all vanish, the non-diagonal elements are, with equal probabilities, $+1$ or -1. This suggests that the density of the characteristic values is given by (9b) under more general conditions, embracing both sets as special cases.

An analysis of the derivation of (9b) for the random sign set shows that this is indeed the case **(12)**. The conditions under which the earlier derivation given for (9b) is valid are:

(1) The distributions for the matrix elements a_{ik} with $i \leqslant k$ are independent of each other.

(2) All these distribution functions, which may be different from each other, are even functions of the corresponding a_{ik}.

(3) The second moments of all these distribution functions is 1. This can be further relaxed to the condition that the second moments of *almost all* of these distribution functions is 1. A further, rather trivial generalization replaces 1 by another arbitrary number, independent of i and k; (9b) is then changed correspondingly.

(4) All moments of all distribution functions exist and are uniformly bounded in i and k. In other words, the nth moment of all the distributions is smaller than a number which may depend on n but is the same for the distribution functions of all a_{ik}.

Unfortunately, these results on the density of the characteristic values do not shed any light on the distribution of the *intervals between adjacent* characteristic values. They do show, however, that at least one statistical property, the density of the characteristic values, is the same for a great variety of real symmetric matrices.

ACKNOWLEDGMENT

I am much indebted to Dr. T. V. Narayana for a clarifying discussion of some of the papers in mathematical statistics which were referred to above.

REFERENCES

1. F. Ajzenberg and T. Lauritsen, Rev. Mod. Phys., *27* (1955), 77.
2. J. W. Mihelich, G. Scharff-Goldhaber, and M. McKeown, Phys. Rev., A94 (1954), 794.
3. D. J. Hughes and J. A. Harvey, *Neutron Cross Sections*, Brookhaven National Laboratory Report No. 325 (1955).
4. J. M. C. Scott, Phil. Mag., *45* (1954), 1322.
5. C. E. Porter and R. G. Thomas, Phys. Rev., *104* (1956), 483.
6. J. A. Harvey, D. J. Hughes, R. S. Carter, and V. E. Pilcher, Phys. Rev., *99* (1955), 10; D. J. Hughes and J. A. Harvey, Phys. Rev., *99* (1955), 1032.
7. J. Von Neumann and E. Wigner, Phys. Zeitschr., *30* (1929), 467.
8. *Conference on Neutron Phyiscs by Time-of-Flight*, Oak Ridge National Laboratory Report No. 2309 (1957), 59.
9. J. A. Harvey and D. J. Hughes, Phys. Rev., 40 (1951), 502. Meantime, further experiments on the question of spacing became available; they seem to be very close to the distribution (3). See W. W. Havens, Columbia University Report CU-174.
10. J. Wishart, Biometrika 20, *32* (1928). Cf. for example, also H. Cramer, Mathematical Methods of Statistics (Princeton, Princeton University Press, 1946) Section 29.9.
11. R. A. Fisher, Ann. Eug. London, *9* (1939), 238 and also A. T. James, Ann. Math. Stat., *25* (1954), 40.
12. E. P. Wigner, Ann. of Math., *62* (1955), 548, *65* (1957), 203.