An Elementary Look at the Eigenvalues of Random Symmetric Real Matrices

John Coffey, Cheshire, UK.

2018

Key words: symmetric random matrix, maximum eigenvalue, Rayleigh distribution of eigenvalue spacing, Wigner's semi-circle distribution theorem, Catalan numbers, tree graph, Coulomb log-gas analogy, Hermite polynomials, Tracy-Widom distribution, universality

1 Introduction

Random matrix theory has been a lively topic of mathematical study for the last 30 years or so, and it continues to find new applications. This present article is an account of my intuitive attempts to explore and understand something of this ubiquitous but complicated subject. It is written by an amateur with other amateur's in mind.

Let me comment on 'random' in this context. The word means that the elements a_{ij} of the matrix are selected in a random fashion from a population which has known statistical properties, such as probability density function (pdf) and hence mean and variance. The principal distribution studied has been the so-called normal or Gaussian. In most cases all the elements are drawn from the same population – so are independent and identically distributed (i.i.d.) – though some important distributions in physics have a slightly larger variance for the diagonal elements. Bear in mind that in a diagonalised matrix, in which all but the diagonal elements are zero, the diagonal consists of the eigenvalues. If the off-diagonal elements are random but small compared with the diagonal ones, the eigenvalues will still be close to the diagonal values. Thus 'random' implies that the diagonal is not dominant. Also we deal with real elements in a symmetric matrix $a_{ij} = a_{ji}$ for which all the eigenvalues are real even though much of the literature is about complex Hermetian matrices. Interest in random matrices is mainly in the statistics of their eigenvalues.

The first application of random matrices was published in 1928 by the Scottish statistician John Wishart. He had been studying agricultural and veterinary data and was looking for correlations between the many input variables and output variables. Values of the input variables $I_k, 1 \le k \le m$ could be written as a vector and were assumed to be related to the outputs $O_j, 1 \le j \le n$ by linear relationships, so typically

$$O_j \;=\; a_{j1}I_1 + a_{j2}I_2 + a_{j3}I_3 + \cdots + a_{jn}I_m \,.$$

The collected relations can be written as a matrix multiplication $O = AI$ where O is the output vector and A the $n \times m$ matrix of elements a_{jk}. In Wishart's data these elements were so complicated that there appeared almost no correlation amongst them. He therefore compared them with matrices with purely random elements and developed statistical tests to see whether the real-world data were actually significantly different from random. He thereby developed a multi-variable generalisation

of the well known χ^2 (chi-squared) test used to assess whether data sets with only one variable are significantly different from each other.

Random matrices were then largely forgotten until 1956 when the theoretical physicist Eugene Wigner used random symmetric matrices to model the quantum mechanical Hamiltonians of heavy nuclei. This is a many-body system in which every neutron and proton interacts with every other. The eigenvalues of these Hamiltonian matrices give the energy levels of a nucleus, and the gap between two levels is the energy needed to excite the nucleus from the lower state to the higher. Heavy nuclei are so complicated that the elements of the true Hamiltonian matrix could not be calculated. Moreover, there seemed no reason to assume correlation amongst them. Wigner therefore modelled the Hamiltonian by a symmetric matrix of random elements, and found that their eigenvalues were is sufficient agreement with experimental values of nuclear excitation energies. The agreement with experiment was not exact, but had similar statistics; in particular, the distribution in excitation energies was quite well modelled.

The subject reappeared in a different guise in 1973 following a chance meeting over tea in the common room at Princeton between the mathematician Hugh Montgomery and the theoretical physicist Freeman Dyson. Montgomery has been studying the statistical distribution of zeros of the Riemann zeta function and mentioned to Dyson a formula he had found. Dyson pointed out that this was precisely the function giving the correlations between eigenvalues of a random matrix with a certain symmetry; the formula was in a 1967 textbook on random matrices by Mehta. Dyson had spotted a connection between two apparently unconnected fields of knowledge – quantum physics and number theory. Since then the statistical properties of random matrices have been studied extensively and applications found in many diverse areas of theoretical physics, number theory and even biology.

We may understand something of wide relevance of random matrices by noting first that matrices in general arise is many branches of mathematics and physics where they link some multi-valued input I via a linear transformation into a multi-valued output O. Perhaps the most important features of a square matrix are its eigenvalues and eigenvectors. Many matrices can be converted to diagonal form by a similarity transformation, equivalent to rotations and reflections in n-dimensional space, to bring the eigenvectors to lie along orthogonal axes. Then the eigenvalues lie down the main diagonal of the matrix and all other elements are zero. This places the matrix in its most simple, 'natural' state. The matrix elements a_{jk} are essentially the direction cosines of the rotations between two sets of axes, one pre-transformation, the other post-transformation. If the input and output have very large dimension, and the interactions represented by the matrix are so complicated that no pattern can be discerned amongst the matrix elements a_{ij} and they appear uncorrelated, then it can be quite a good approximation to let the a_{ij} be random numbers provided they respect the symmetry of the true matrix. By this I mean that if the true matrix is a real symmetric matrix, the approximating random matrix must also be so, and similarly if it is complex Hermitian. In fact Dyson identified three important types of symmetry in physics, characterised by a 'temperature' parameter β and related to the unitary symmetry groups (arising from Hermetian matrices, $\beta = 2$), the orthogonal groups (real orthogonal matrices, $\beta = 1$) and the symplectic groups (quaternion matrices, $\beta = 4$) respectively. Much of the analysis is linked to the statistical mechanics of large ensembles, but I do not delve into this at all. There is also a similarly with the central limit theorem. Recall that this states that the probability distribution of a process which has many contributing sub-processes tends to be Gaussian – also called normal – whatever the distributions of the sub-processes. It happens that the statistics of the output from large random matrices tend to be insensitive of the precise statistics of the a_{ij} provided the symmetry is correct. This highly useful property, called

'universality', is probably the main reason why random matrices have such wide application as models in physics and statistics. If you know that a system – physical, biological, financial, whatever – is subject to universality, its large scale features and behaviour can be modelled quite accurately without having to know the inner workings.

Another aspect of random matrices which has universality is the distribution in value of the largest eigenvalues, $|\lambda_{max}|$. This is known as the Tracy-Widom distribution after its two discoverers, and has three versions, one for each of the symmetry types β = 1, 2 or 4, identified by Dyson. Phenomena explained by the Tracy-Widom distribution were first described by biologists as they tried to model the stability of large complex eco-systems in which the many components interact[1]. The interactions can be represented by a matrix. It seemed that below a critical level of connectedness, the system would be stable over time, but if the number or strength of connections increased further, some species would multiply and dominate and others be wiped out. The critical point corresponds to the peak in the Tracy-Widom probability distribution. Since then the Tracy-Widom distribution has turned up in many seemingly unrelated places. It seems to be related to highly interacting, cooperative phenomena, and the switch in behaviour as the peak of the probability distribution curve is crossed is like a phase transformation from a phase with weak inter-element coupling to one with strong.

The whole subject of random matrices is now very large and well studied, as has been documented in several major books. One of the first, by Madan Mehta, is now in its third edition; Mehta spent almost his whole academic career studying these mathematical objects. The fairly recent thick volume entitled 'The Oxford Handbook of Random Matrix Theory', 2011, is a comprehensive set of review articles[2]. However none of the books I have located is light reading and none is an easy introduction for the amateur.

Before going into details let me give some results for random matrices which, in a simple visual way, show some of their characteristic features. Figure 1 plots as short radial lines the eigenvalues of a 32×32 symmetric matrix with real elements drawn from the Gaussian population N[0,1]; that is, a normal distribution[3] with mean zero and standard deviation 1. They are fanned around a circle such that each unit is represented by a 15° increase in angle. We expect the eigenvalues to be roughly equally split between positive and negative ones, as shown here by those above and below the horizontal reference line. Observe the similarity in position of these positive and negative eigenvalues, almost mirrored in the horizontal axis. The spacing of adjacent marks is somewhat the same to the right of the plot where values are closer to zero, but the last few high values to the left are further apart, as if they had been constrained by some force near zero which relaxed at the two ends of the sequence.

The two panels in Figure 2 let us see by eye how different the spacing is of the eigenvalues of a random matrix from the spacing of random numbers. The left panel plots around a spiral the 512 eigenvalues of a 512×512 random matrix whose elements are all from N[0,1]. The right panel

[1] See for example Robert May, Nature, Vol 238, p413, 18 August 1972. Also the length of the longest increasing sequence in a sample of N random numbers has the Tracy-Widom distribution

[2] Mehta's thick book *Random Matrices*, pub. Academic Press, had its third edition in 2004. 'The Oxford Handbook of Random Matrix Theory', is edited by Akemann, Bail and Di Franceso. Two other notable books are 'An introduction to random matrices' by Anderson, Guionnet and Zeitouni, CUP, 2010, and 'Log-gases and random matrices' by Peter Forrester, Princeton Univ Press, 2010. The recent books are for the specialist. In one the Introduction explains that the theory and application of random matrices has had input from so many disparate branches or maths and physics that few people will have the width of knowledge and expertise to understand all aspects.

[3] The probability density function (pdf) of a variable x drawn from N$[\mu, \sigma]$ is $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2/(2\sigma^2)\right]$.

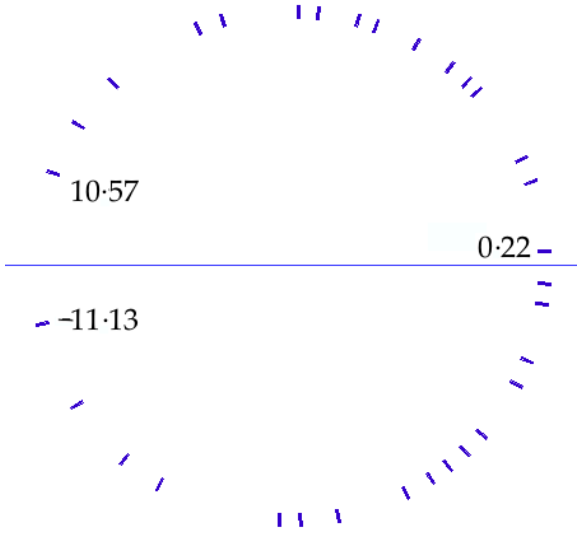

Figure 1: Plot of eigenvalues of a 32×32 symmetric matrix.

plots 512 random numbers from a uniform distribution over the same interval from −45 to +45. The eigenvalues are more evenly spaced than the uniformly random numbers; the latter have bunching in some places and wide gaps in others. There is not the 'relaxation' at the two ends that we saw first in Figure 1. The excitation energies of heavy nuclear are known from experiment not to bunch closely together. Another example of somewhat even spacing is seen in Figure 3, which plots the first 700 complex zeros of the Riemann zeta function on the scale of 1 unit is 3°. Again there are no great gaps and few very small gaps. You might believe that Figure 3 looks quite like the left panel of Figure 2 (apart from at from its two ends). This is what Dyson pointed out to Montgomery; there is a previously unsuspected statistical similarity between the eigenvalues of a large random matrix and the non-trivial zeros of the zeta function. Some writers have even suggested that the zeros of the zeta function and hence the prime numbers might actually be the eigenvalues of some infinite symmetric matrix which is fundamental to nature – the cosmic matrix at the end of the Universe!

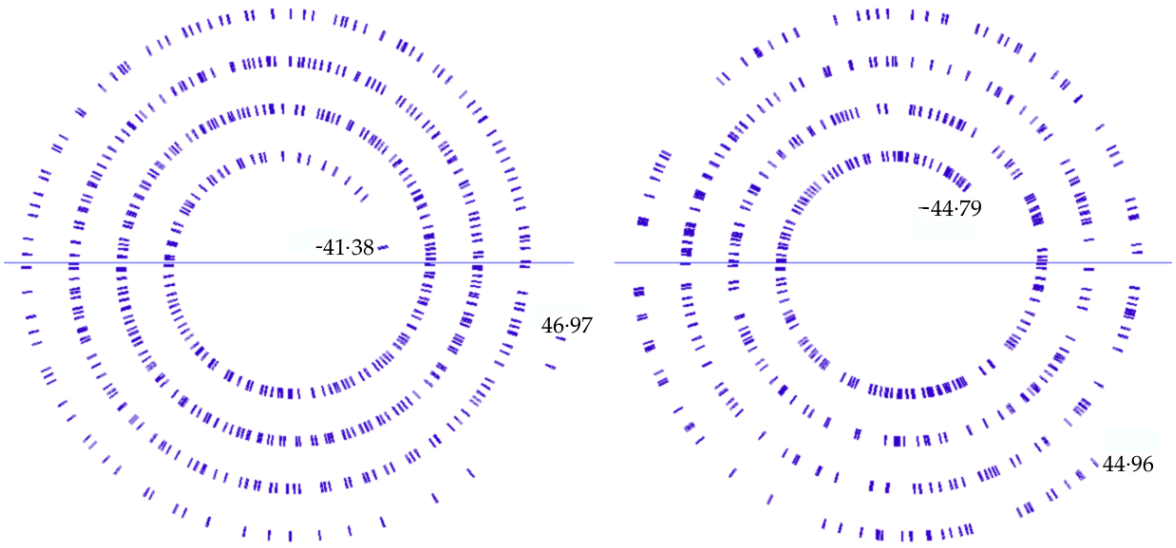

Figure 2: Left panel: Eigenvalues of a 512×512 real symmetric matrix plotted around a spiral. Right: 512 numbers between −45 and +45 from a uniform distribution plotted similarly.

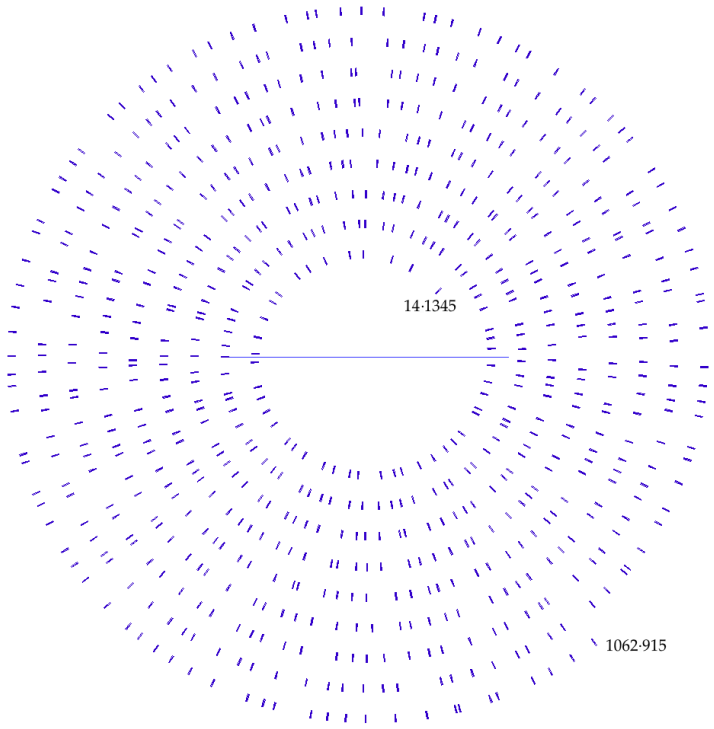

Figure 3: The first 700 complex zeros of the Riemann zeta function plotted around a spiral.

2 Numerical survey of some Gaussian random matrices

This section gives an account of some of my numerical experiments with symmetric matrices whose elements are all real and drawn from the Gaussian population $\mathrm{N}[0, 1]$. The elements are said to be 'i.i.d.', meaning independent and identically distributed.

2.1 Method

Throughout this investigation my approach has been to examine small matrices and infer from them the behaviour as N, the number of rows and columns, become arbitrarily large.

My first step was to write a computer program to create random numbers with a normal distribution. I used the standard uniform pseudo-random number generator provided within the BBC Basic for Windows language (at *www.rtrussell.co.uk*) and map it to a normal distribution using the inverse error function[4]. The function RND(1) in BBC Basic for Windows supplies a 40-bit floating point number in the range $0{\cdot}0$ to $1{\cdot}0$, exclusive of the end points. The cycle length is $2^{33}-1 \approx 8{\cdot}59\times10^9$, which is sufficient for the largest matrices which my home computer can handle. The generator is given a new seed each time using the function RND(−N) where N is a new random integer. The inverse error function of x is calculated to adequate accuracy by the following formula given in Wikipedia:

$$L = \ln(1-x^2), \qquad a = 0{\cdot}147, \qquad A = 2/(\pi a) + L/2, \qquad B = \sqrt{A^2 - L/a}\ , \qquad erfinv(x) = \sqrt{B-A}\ .$$

I created a significant number of square symmetric matrices with real elements drawn from the same distribution with mean zero, standard deviation 1. The size N of the matrices was scaled up in steps of $\times\sqrt{2}$ up to $2^{11} = 2048$, supplemented by some intermediates plus the largest I could handle which was $2^{11{\cdot}25} = 2435$ to the nearest integer. Thus N = 2, 4, 6, 8, 16 to 1448, 2048 and 2435. Mostly there were either 4 or 8 matrices of each size, with 95 in all. Each was entered into Mathematica 10 to calculate its eigenvalues. As expected from the symmetry of the normal distribution, these too appeared roughly symmetrically distributed about zero as in Figure 1 Using a spreadsheet I examined how the largest positive and negative eigenvalues vary with N, and the spacing between adjacent eigenvalues once placed in ascending order. I also determined the characteristic polynomial of selected matrices.

Throughout this study, where early results have indicted fruitful or merely interesting lines of enquiry, I have reported the early results as well as the more refined and later conclusions. Hence the text contains several conjectured and hence provisional formulae.

2.2 The largest eigenvalues, with a partial explanation

Figure 4 shows how the largest positive and negative eigenvalues increase with N. Call these λ^+_{max} and λ^-_{max} respectively. For the larger matrices the ratio of these is close to −1; for instance, for one $N = 2435$ matrix $\lambda^+_{max} = 99{\cdot}378$ and $\lambda^-_{max} = -98{\cdot}996$. Their ratio has more scatter for the smaller matrices, but the mean appears to be −1 as one would expect. To gain some idea of the functional dependence on N, consider Figure 5. In this the absolute values of λ^+_{max} and λ^-_{max} have been plotted for each matrix against size, as $\ln|\lambda^\pm_{max}|$ versus $\log_2 N$. The light green dotted curve is Excel's quartic trend line showing how the points tend to follow a straight line as N increases. The dark green solid straight line approximates the upper limit of these maximum eigenvalues. It has the form '$y = 0{\cdot}346x + 0{\cdot}742$'. When allowance has been made for the log to base 2, this corresponds to

$$|\lambda^\pm_{max}| \ \lessapprox\ 2{\cdot}1\sqrt{N}\,, \qquad N > 100\,. \tag{1}$$

[4] I subsequently occasionally used the following recipe within Mathematica 10:
`g = RandomVariate[NormalDistribution[0, 1], {N, N}];`
`A = ((g + Transpose[g]) - DiagonalMatrix[Diagonal[g]]*0.585786)*0.707107;`
where N is the size of the matrix, $0{\cdot}585786 = 2-\sqrt{2}$ and $0{\cdot}707107 = 1/\sqrt{2}$.

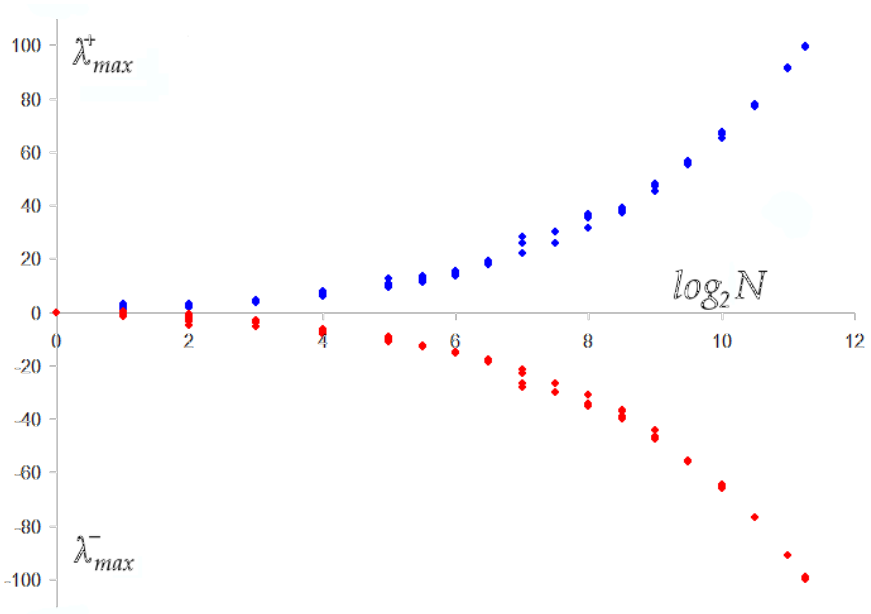

Figure 4: The greatest positive and negative eigenvalues in each of 85 symmetric real matrices plotted against size, as the logarithm of N to base 2.

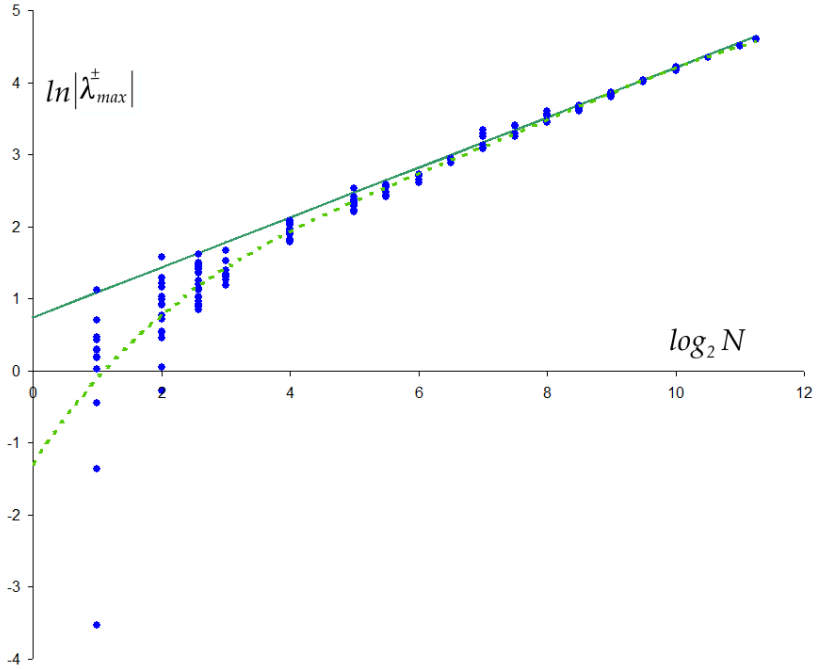

Figure 5: Log-log plot of absolute values of largest eigenvalues, $|\lambda_{max}|$, for each of 95 matrices against size N, with two trend lines.

In words, the most positive and most negative eigenvalues have about the same magnitude and this varies as $\sqrt{N}$, N being the matrix dimension.

This relationship, assuming it continues for even larger N, is so simple that we must suspect a simple reason. In a previous article on *www.mathstudio.co.uk* entitled 'Iterative numerical methods

for real eigenvalues and eigenvectors of matrices' I have listed some properties of eigenvalues, including three estimates for their magnitudes. In §2, item 14 of that article I cite the simple rule by Alfred Brauer, 1946, which states that if R is the largest of the sums of absolute values of the elements along the rows, and C the largest sum taken down the columns, then for all eigenvalues λ, $|\lambda| \leq min(R, C)$. For a matrix with Gaussian distribution of elements, mean $\mu = 0$, variance $\sigma^2 = 1$, the absolute values (according to Wikipedia) have a so-called folded normal distribution with mean

$$\mu_f \;=\; \sigma\sqrt{\frac{2}{\pi}}\exp\left(\frac{-\mu^2}{2\sigma^2}\right) \;-\; \mu\, erf\left(\frac{-\mu}{\sqrt{2\sigma^2}}\right) \;=\; \sqrt{\frac{2}{\pi}} \;=\; 0\cdot 798.$$

and variance

$$\sigma_f^2 \;=\; \mu^2 + \sigma^2 - \mu_f^2 \;=\; 1 - \frac{2}{\pi} \;=\; 0\cdot 363.$$

The largest sum of these absolute values along a row or column might therefore have a value not far from $0\cdot 9N$. Unfortunately for our purposes this linear dependence on N does not account for the $\sqrt{N}$ in Eq. 1, and $\lambda^{\pm}_{max}$ are grossly over-estimated.

Another set of bounds on eigenvalues has been given by Wolkowicz and Styan[5]. This too fails to explain the $\sqrt{N}$ trend seen in numerical examples. For large N is merely places $|\lambda^{\pm}_{max}|$ between 1 and N, which seems true but is not helpful. Other ideas seem necessary.

Here is my own explanation. Following Figure 1 I assume that the eigenvalues are symmetrically spread either side of zero; that is their values are $\pm\lambda_k$. Suppose also that the position and spacing between adjacent eigenvalues are reckoned as multiples of a unit q. If N is even, its characteristic polynomial will then be

$$[\lambda^2 - (b_1 q)^2][\lambda^2 - (b_2 q)^2][\lambda^2 - (b_3 q)^2]....[\lambda^2 - (b_{\frac{N}{2}} q)^2], \quad N/2 \text{ factors.} \tag{2}$$

If N is odd, Eq 2 will simply be multiplied by λ since the central eigenvalue will be 0. The b_k count the position of each pair of eigenvalues, these being at $\pm b_k q$. We want the largest value of this.

I need to give some justification for the simplifications made in this approximate model. Figures 1 and 4 show by eye that each positive eigenvalue has another with roughly negative its value. If you take a small real symmetric matrix A and evaluate its characteristic polynomial algebraically, a pattern can be seen in the values of some of the coefficients of λ^k. Here is a 5×5 symmetric matrix

$$\begin{pmatrix} a & b & c & d & e \\ b & f & g & h & j \\ c & g & m & p & r \\ d & h & p & s & t \\ e & j & r & t & u \end{pmatrix}$$

and the coefficients of some of its powers:

$$\lambda^N : \quad 1$$

$$\lambda^{N-1} : \quad -(a+f+m+s+u)q \;=\; q.Tr(A)$$

$$\lambda^{N-2} : \quad -(b^2+c^2+d^2+e^2+g^2+h^2+j^2+p^2+r^2+t^2 - af - am - fm - as - fs - ms - au - fu - mu - su)q^2$$

$$\lambda^{N-3} : \quad -(-c^2 f - d^2 f - e^2 f + 2bcg - ag^2 + 2bdh - ah^2 + 2bej - aj^2 - b^2 m - d^2 m - e^2 m + afm.....)q^3 .$$

[5] 'Bounds on eigenvalues using traces' in Linear Algebra and its Applications, Vol 29, p 471-506, 1980.

In a random N[0,1] matrix of independent elements (apart from being symmetric) the expectation value of each element is 0, and of a product of different elements is also 0. Hence the coefficient of λ^{N-1} tends to zero. In the coefficient of λ^{N-2} the only non-zero terms are the squares of all 10 off-diagonal elements in the upper triangle. The average value of each of these squares is $\sigma^2 = 1$, so their average sum is 10. For an $N \times N$ matrix this coefficient will on average be $-N(N-1)/2$. The coefficient of λ^{N-3} has many 3-factor terms, but every one averages to zero. The coefficient of λ^{N-4} has even more 4-factor terms, many of which average to 0, but some do not; these are the 15 products of two squares of the form b^2p^2, c^2h^2, etc., each of which is on average equal to 1. The characteristic polynomial of a 5×5 matrix is therefore on average

$$\lambda^5 - 10\lambda^3 + 15\lambda$$

and of an $N \times N$ matrix will start

$$\lambda^N - \tfrac{N(N-1)}{2}\lambda^{N-2} + B\lambda^{N-4}.... \tag{3}$$

where B is the mean sum of products like b^2p^2.

I carried out a similar procedure with matrices up to $N = 11$ and find that their average characteristic polynomials $P_N(\lambda)$ are

$$N=2: \quad \lambda^2 - 1 \qquad\qquad N=3: \quad \lambda^3 - 3\lambda \tag{4}$$

$$N=4: \quad \lambda^4 - 6\lambda^2 + 3, \qquad\qquad N=5: \quad \lambda^5 - 10\lambda^3 + 15\lambda,$$

$$N=6: \quad \lambda^6 - 15\lambda^4 + 45\lambda^2 - 15, \qquad\qquad N=7: \quad \lambda^7 - 21\lambda^5 + 105\lambda^3 - 105\lambda,$$

$$N=8: \quad \lambda^8 - 28\lambda^6 + 210\lambda^4 - 420\lambda^2 + 105,$$

$$N=9: \quad \lambda^9 - 36\lambda^7 + 378\lambda^5 - 1260\lambda^3 + 945\lambda,$$

$$N=10: \quad \lambda^{10} - 45\lambda^8 + 630\lambda^6 - 3150\lambda^4 + 4725\lambda^2 - 945,$$

$$N=11: \quad \lambda^{11} - 55\lambda^9 + 990\lambda^7 - 6930\lambda^5 + 17325\lambda^3 - 10395\lambda.$$

I shall show in §5 below that these are cousins of the Hermite polynomials. In calculating the expected values of the coefficients, I have used the independence of all the elements of A_N to conclude that the expectation E() of the product of any x, y is $\mathrm{E}(xy) = \mathrm{E}(x)\,\mathrm{E}(y)$ where x, y represent any two distinct matrix elements a_{ij} or their powers. In the coefficients there were no cubes a_{ij}^3, only squares, which average to 1, and single powers of the elements, which average to 0.

From samples of 20 $N = 5$ and 40 $N = 6$ matrices I found the average coefficients of the characteristic polynomials to be as in Table 1. The agreement with theory is not too bad.

The roots of these mean polynomials are readily found and give the expectation values of the eigenvalues for i.i.d. N[0,1] matrices with $N = 2$ to 11:

$N=2$: $\lambda = \pm 1{\cdot}000$,
$N=3$: $\lambda = \pm 1{\cdot}732$,
$N=4$: $\lambda = \pm 2{\cdot}334, \ \pm 0{\cdot}742$,
$N=5$: $\lambda = \pm 2{\cdot}857, \ \pm 1{\cdot}356, \ 0$,
$N=6$: $\lambda = \pm 3{\cdot}324, \ \pm 1{\cdot}889, \ \pm 0{\cdot}617$.
$N=7$: $\lambda = \pm 3{\cdot}750, \ \pm 2{\cdot}367, \ \pm 1{\cdot}154, \ 0$.
$N=8$: $\lambda = \pm 4{\cdot}146, \ \pm 2{\cdot}802, \ \pm 1{\cdot}637, \ \pm 0{\cdot}539$.
$N=9$: $\lambda = \pm 4{\cdot}513, \ \pm 3{\cdot}205, \ \pm 2{\cdot}077, \ \pm 1{\cdot}023, \ 0$.

Power	$k=5$ Average Coeff.	$k=5$ Theory	$k=6$ Average Coeff.	$k=6$ Theory
0	$3\cdot56$	0	$-6\cdot71$	−15
1	$14\cdot95$	15	$8\cdot72$	0
2	$-1\cdot55$	0	$31\cdot61$	45
3	$-8\cdot19$	−10	$-5\cdot12$	0
4	$0\cdot24$	0	$-13\cdot00$	−15
5	1	1	$0\cdot63$	0
6			1	1

Table 1: Average coefficients of the characteristic polynomial of 5×5 and 6×6 symmetric matrices with elements in N[0, 1]. Comparison of, respectively, 20 and 40 numerical values with theory.

$N=10:\ \lambda = \pm4{\cdot}859,\ \pm3{\cdot}582,\ \pm2{\cdot}484,\ \pm1{\cdot}466,\ \pm0{\cdot}485.$
$N=11:\ \lambda = \pm5{\cdot}188,\ \pm3{\cdot}936,\ \pm2{\cdot}865,\ \pm1{\cdot}876,\ \pm0{\cdot}929,\ 0.$

From samples of 24 $N=4$, and the 20 $N=5$ and 40 $N=6$ matrices featured in Table 1, I found average ordered eigenvalues as follows:
$N=4:\ \lambda = -2{\cdot}58,\ -0{\cdot}76,\ 0{\cdot}73,\ 2{\cdot}38,$
$N=5:\ \lambda = -2{\cdot}84,\ -1{\cdot}45,\ -0{\cdot}05,\ 1{\cdot}38,\ 2{\cdot}73,$
$N=6:\ \lambda = -3{\cdot}38,\ -1{\cdot}92,\ -0{\cdot}78,\ 0{\cdot}52,\ 1{\cdot}70,\ 3{\cdot}23$
all of which agree passably well with the theoretical polynomial roots. Taken together with the average coefficients in Table 1, these average eigenvalues give some support to my argument.

Now we return to the characteristic polynomial in Eq 2. The coefficient of λ^{N-2} is $-(b_1^2 + b_2^2 + b_3^2 + ... + b_{\frac{N}{2}}^2)\,q^2$. We want to match this to Eq 3 and to do so some values must be set on the spacings $b_k q$. Let us set aside for the moment that the largest eigenvalues appear to 'expand' apart, and treat all eigenvalues as equally spaced by $2q$ units. Then $b_1=1$, $b_2=3$, $b_3=5$, ... $b_k=2k-1$, ... $b_{\frac{N}{2}}=N-1$. The sum can be obtained in closed form as

$$\sum_{k=1}^{N/2}(2k-1)^2 \;=\; \tfrac{1}{6}N(N^2-1)\,.$$

This allows q to be found:

$$\tfrac{1}{2}N(N-1) \;=\; \tfrac{1}{6}N(N^2-1)q^2 \quad \text{from which} \quad q \;=\; \sqrt{\frac{3}{N+1}}\,. \tag{5}$$

The mean spacing is $2q$ and the extreme eigenvalues are at $\pm(N-1)q = (N-1)\sqrt{\frac{3}{N+1}}$, which tends to $\sqrt{3N}$ as $N\to\infty$. This explains Figure 5 and Eq. 1.

In Figure 6 I have replotted all the data of Figure 5 and added the trend equation

$$|\lambda^{\pm}_{max}| \;\approx\; \beta\,\frac{N-1}{\sqrt{N+1}}\,. \tag{6}$$

β is $\sqrt{3}$ in the lower curve which is my theoretical average largest eigenvalue, and $2{\cdot}1$ in the upper which approximates an upper envelope to the largest eigenvalues. Note how both curves show the downward trend seen in the data and sketched by the dashed green curve of Figure 5. Increasing β

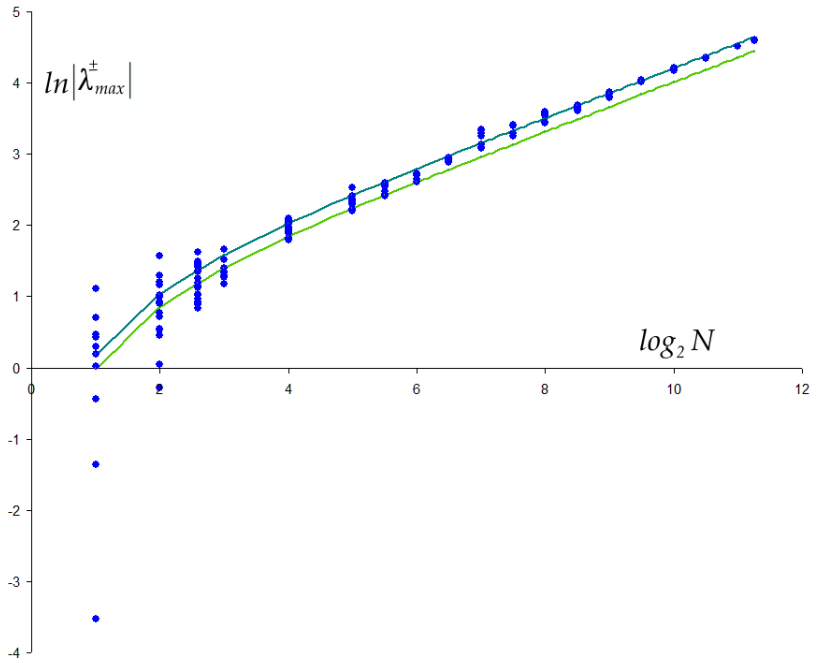

Figure 6: Log-log plot of Figure 5 repeated with two trend lines using the relation of Eq 6. Lower line β factor is $\sqrt{3}$. Upper line $\beta = 2 \cdot 1$.

from 1·73 to 2·1 simulates the extreme eigenvalues being more widely spaced by a factor of about 1·2, to give the behaviour of Eq 1.

To add further numerical evidence to this analysis I have examined the coefficients in the characteristic polynomials of 12 random N[0,1] 32×32 matrices. Their statistics are

Mean $\lambda^{\pm}_{max}$: $-10 \cdot 377$ and $10 \cdot 629$

Average spacing over all 32 eigenvalues : $0 \cdot 678$, to compare with $2q = 2\sqrt{(3/33)} = 0 \cdot 603$,

Average coefficient of λ^{30} : $-501 \cdot 8$, to compare with -496 by theory,

Average constant term : $-2 \cdot 096 \times 10^{17}$.

The actual average spacing is about 12% up on the theoretical because of the expanding spacing of the extreme eigenvalues. The mean constant term also allows a second estimate of q, albeit not a reliable one since I found a large variation in constant terms across the 12 sample matrices. The constant term of the characteristic polynomial in Eq 2 equals the product of all the roots $(\pm b_j q)^2$. For eigenvalues placed at $b_1 = 1$, $b_2 = 3$, ... $b_k = 2k - 1$, ... $b_{\frac{N}{2}} = N - 1$, the product evaluates to

$$\left(\frac{31\,!}{2^{15}\,15\,!}\right)^2 q^{32} \;=\; 3 \cdot 6825 \times 10^{34}\, q^{32}.$$

Setting this equal to the observed $-2 \cdot 096 \times 10^{17}$ gives $2q = 0 \cdot 58$, a slightly lower estimate but still showing consistency in the approach.

In summary, the square root dependence of Gaussian N[0,1] matrices on the matrix size seems to be due to two competing effects: the number of eigenvalues increasing as N and the separation of them decreasing as $1/\sqrt{N}$. To give a sense of scale to this, suppose that you were to write out the matrix on a large square of paper or card so that each element was printed in small type in a square cell with side length 2 cm. A 100×100 matrix would then require a 2 m square of paper – the size

of a double doorway. The eigenvalues, marked on a number line to the same scale, would stretch 84 cm end-to-end, and the spacing of adjacent eigenvalues along most of this line would be about 7 mm, with thickness of a pencil. A matrix with N a million (10^6) would cover a square 20 kilometres by 20 kilometres – the distance between towns – but the eigenvalues would be marked along a line only 84 m long – less than the length of a football field – and the space between eigenvalues would be only $0 \cdot 07$ mm, the width of a human hair. Taking N to infinity would cause the eigenvalues to become densely packed, with vanishing separation. We might consider, therefore, that even if the zeros of the zeta function are the eigenvalues of some infinite primordial matrix, it cannot be of the i.i.d. N[0,1] type, since the Riemann zeros are not packed to infinite density.

To close this section let us use the listed expectation values of the largest eigenvalues above to refine the estimate of β, now regarding as a function of N, in Eq 6,

$$|\lambda^{\pm}_{max}| \;\approx\; \beta \frac{N-1}{\sqrt{N+1}} \,.$$

This can be rearranged to give a value for $\beta(N)$, and I find that a log-log plot of $\beta(N)$ against N is close to a straight line. Taking the coefficients of this from the graph through the higher points N = 8 to 11, gives

$$\beta(N) \;\approx\; 1 \cdot 64526\, N^{0 \cdot 0368} \,, \qquad N \leq 200 \,. \tag{7}$$

This reaches the value 2 when N = 200, which seems to fit with the points plotted in Figure 6. As working values for the time being, therefore, take β from Eq 7 up to N = 200 and β = 2 above. §6.2 will present a further refinement and also an alternative formula for the dependence of λ_{max} on N.

2.3 Spacing of eigenvalues

We now move the focus from the largest eigenvalues to the spacings between any adjacent pair. Eq 1 implies that the average spacing falls at $1/\sqrt{N}$ because all N eigenvalues must fit into the interval $[\lambda^{-}_{max}, \lambda^{+}_{max}]$ which is $\lessapprox 4 \cdot 2\sqrt{N}$. The way in which the extreme positive and negative eigenvalues expand away from the others – or the innermost ones are squashed together – is seen in all matrices, but clearest when N is large. This behaviour can be pictured in more than one way. First, Figure 7 plots the separation distance δ between adjacent eigenvalues for an N = 2435 matrix with i.i.d. elements from N[0, 1]. Second, for the same matrix I have counted the number of eigenvalues between −100 and −99, −99 and −98, and so on in bins of unit width up to 99 to 100. These frequencies are plotted at the midpoints $-99 \cdot 5$, $-98 \cdot 5$ etc. in Figure 8. The curves drawn through the points in Figures 7 and 8 are essentially reciprocals of each other. The red and green approximate bounding curves in Figure 8 are both semi-ellipses. With suitable scaling these ellipses can be transformed into circles and present us with an example of Wigner's semi-circular asymptotic law for eigenvalues, described in §4.

A third type of presentation is the frequency distribution of separation distances δ between adjacent eigenvalues. The left panel of Figure 9 gives a frequency plot for the N = 2435 matrix of Figures 7 and 8. It records the number of gaps between adjacent eigenvalues (over the span from $\lambda^{-}_{max} \approx -100$ to $\lambda^{+}_{max} \approx +100$) lying within intervals of widths 0·01 units. Note how few gaps are smaller than 0·02 or wider than 0·2. The smooth green curve has the form $6170\,\delta\,\exp(-130\delta^2)$ which is a Rayleigh distribution. The right panel in Figure 9 simulates what this distribution in gap widths would have looked like if the 2435 eigenvalues were uniformly distributed over $(-100, 100)$. The two panels are therefore quantitative versions of the two spiral plots in Figure 2.

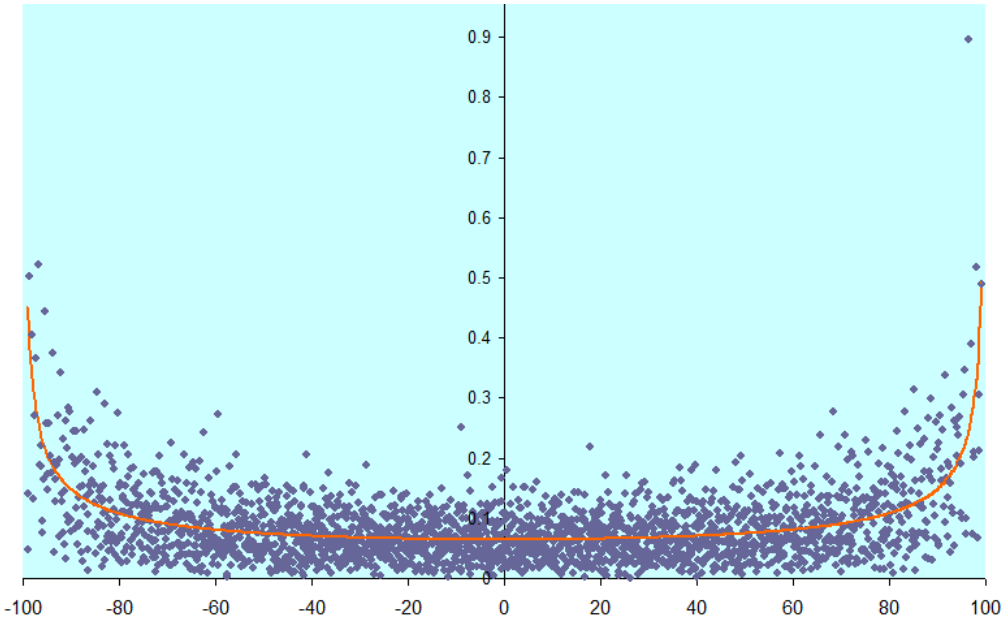

Figure 7: Separation of adjacent eigenvalues for one $N = 2435$ real symmetric matrix with elements from N[0, 1]. Distance δ between eigenvalues is plotted against the midway position of each pair. Orange curve is the reciprocal of a semi-ellipse with semi-axes 100 and $15 \cdot 5$.

The green curve in the right panel of Figure 9 is a binomial function. To see how this arises consider that in a uniform distribution over $-h$ to $+h$ the probability that any one point lies in an interval of width w is $w/2h$, and the probability that the same point does *not* lie in this interval is $1 - w/2h$. We have N independently, randomly chosen points and the probability that all N do not lie in the one same interval – that is, that there is a gap of width w – is $(1 - w/2h)^N$. There are $N - 1$ gaps in all of which the widest possible in $2h$. In our case $N = 2435$, $h = 100$ and w advances in steps of $\delta = 0 \cdot 01$, so $w = k\delta$, $k = 1, 2, 3, ..., 2h/\delta$. Introducing a normalising constant C, the number of gaps of width w is

$$C(1 - w/2h)^N \quad \text{where} \quad C \sum_{k=1}^{2h/\delta} (1 - k\delta/2h)^N \;=\; N - 1\,.$$

The sum evaluates to $7 \cdot 72032$ so $C = 315 \cdot 3$. For large N this binomial curve can be closely approximated by a simple decaying exponential function by noting that

$$(1 - v)^N \;=\; 1 - Nv + \tfrac{1}{2!}N(N-1)v^2 - \tfrac{1}{3!}N(N-1)(N-2)v^3 + ...$$

$$\approx\; 1 - Nv + \tfrac{1}{2!}N^2v^2 - \tfrac{1}{3!}N^3v^3 + ... \;=\; e^{-Nv}, \qquad v = \tfrac{k\delta}{2h}$$

provided v remains finite as $N \to \infty$. C is approximated by C' such that

$$C' \int_{1/2}^{\infty} \exp\left(-\frac{Nk\delta}{2h}\right) dk \;=\; N - 1\,.$$

Thus $C' = 2434/7 \cdot 7285 = 314 \cdot 94$. When plotted, the exponential lies on top of the binomial curve with difference $0 \cdot 1\%$ or less.

That the gaps between eigenvalues seem Rayleigh distributed was previously observed by Wigner, and the so-called Wigner surmise is that eigenvalues of all such random matrices show this

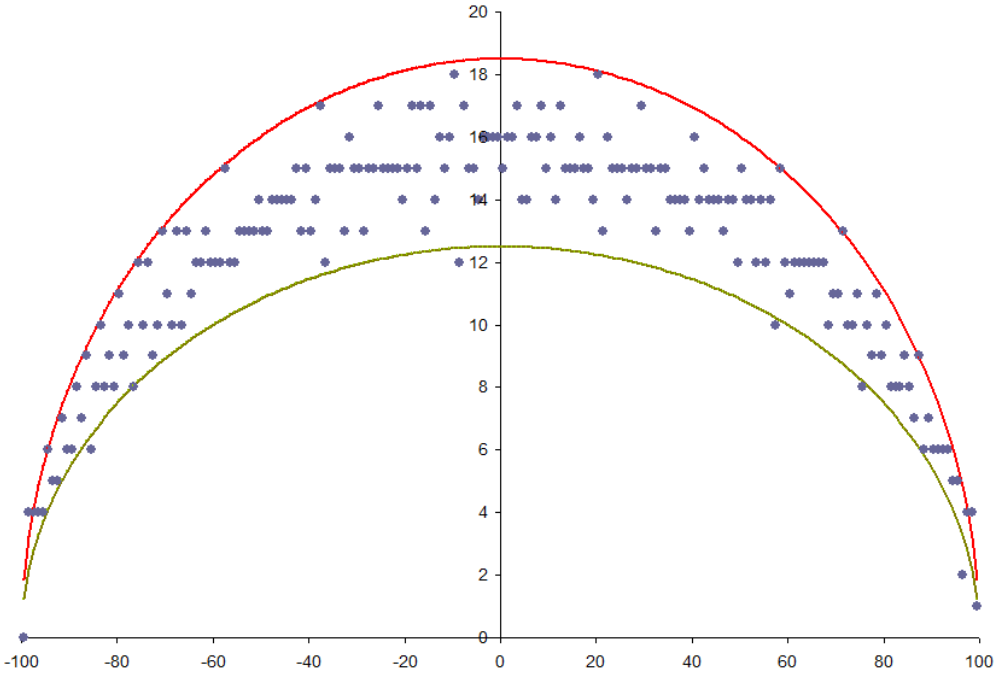

Figure 8: A frequency distribution: numbers of eigenvalues between n and $n+1$ from $n = -N$ to $+N$ for the 2435×2435 matrix in Figure 7. The count is plotted against the midpoint of each unit-width interval. Bounding curves are semi-ellipses with semi-minor axes $18 \cdot 5$ and $12 \cdot 5$.

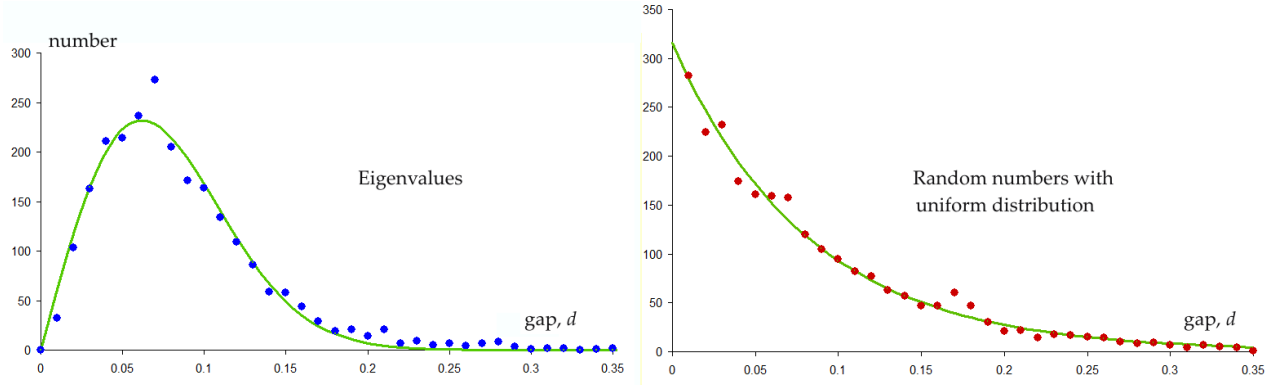

Figure 9: Frequency plots showing, left, number of gaps between eigenvalues of $N = 2435$ N[0,1] matrix in intervals of width $0{\cdot}01$, and right, comparable graph if eigenvalues were uniformly distributed.

distribution of gaps. We might speculate how a Rayleigh distribution could come about. If x and y are independent and identically distributed (i.i.d.) random numbers from the population $\mathrm{N}[0, \sigma]$, then the 'length' $r = \sqrt{x^2 + y^2}$ is Rayleigh distributed[6] with pdf $r/\sigma^2 \exp(-r^2/(2\sigma^2))$. In a two-dimensional kinetic theory of gases x and y would be components of molecular velocity. Suppose that s is a vector quantity, Gaussian in magnitude, and uniformly distributed in direction in a plane. Thus if all vectors of type s are plotted in 2 dimensions, those with magnitude between $|s|$ and $|s + \delta s|$ will lie around a circular annulus of radius $|s|$ and thickness δs. The number of planar vectors which have magnitude between $|s|$ and $|s + \delta s|$ irrespective of direction is therefore $2\pi A s \exp(-s^2/(2\sigma^2))\, \delta s$ where A is a normalisation constant. This is the Rayleigh distribution. I have no model to say how

[6] If u is uniformly distributed over (0, 1), then $r = \sigma\sqrt{-2 \ln u} = \sigma\sqrt{\ln(1/u^2)}$ is Rayleigh distributed.

this might be analogous to the gaps between eigenvalues. I can only remark that the eigenvalues are obtained by combining random matrix elements into the characteristic polynomial, and the distribution of values in these combinations will tend to be Gaussian by the central limit theorem. Some of these processes of combination may compound their components orthogonally rather than by simple summation.

3 Some non-Gaussian distributions

In the previous section we have seen something of the behaviour of real symmetric matrices whose elements come from the same normal (Gaussian) distribution N[0,1]. It is fair to ask whether the $\sqrt{N}$ rate of increase in $|\lambda^{\pm}_{max}|$ and the corresponding decrease in spacing, with the expanding apart of the larger eigenvalues, is confined to matrices with Gaussian distributed elements, or whether it is a more common feature of symmetric matrices. This will illuminate the extent to which 'universality' appertains to various classes of random matrices. I have accordingly examined matrices with three other distributions described below.

3.1 Uniform distributions

The uniform distribution over $(-\sqrt{3}, \sqrt{3})$ has mean 0 and variance 1. My first study with these has been to create a number of matrices of different sizes from this distribution to see how the largest eigenvalues, and the separation between pairs of eigenvalues, vary with N. Regarding the largest eigenvalues a sample of 11 matrices with N from 64 to 2048 showed a clear relation $|\lambda^{\pm}_{max}| = 1{\cdot}80N^{0{\cdot}52}$. This is close to the $2 \cdot 1\sqrt{N}$ of Eq 1 for Gaussian distribute matrices and can be explained by the argument of §2.2 since the mean and variance are unchanged in moving to a uniform distribution. However, I found two matrices which showed significant deviation from this expected behaviour. Of the four matrices with N = 256, three had $|\lambda^{\pm}_{max}|$ close to $31 \cdot 5$, but the fourth had values $-37 \cdot 2$ and 39.0. There was extreme departure by one of the two matrices with N = 512; whilst one had $|\lambda^{\pm}_{max}|$ at about $46 \cdot 5$, the other had these largest eigenvalues : $444 \cdot 8$, $27 \cdot 5$, $-26 \cdot 7$. It happens that the trace of this matrix is $443 \cdot 1$, which can only have come about by some chance combination of element values. This example reminds us that within any random sample there is the possibility of sporadic large departures from average behaviour.

Regarding the spacing between elements, I have found the same random scatter around a

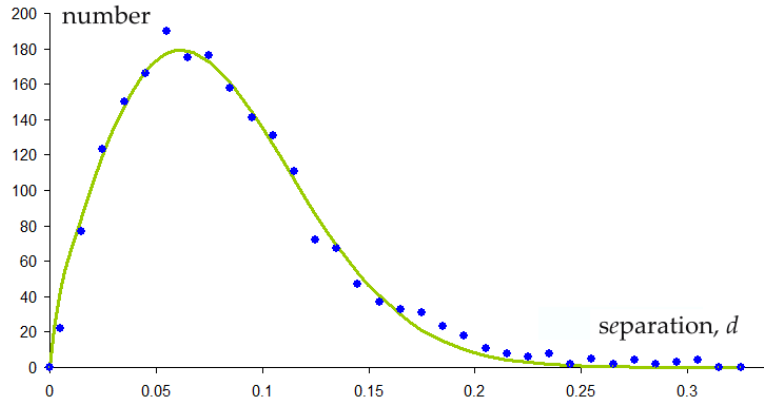

Figure 10: Frequency plot of the gaps, δ, between adjacent eigenvalues for one N = 2048 matrix with elements in U[-$\sqrt{3}$, $\sqrt{3}$].

'bath-tub' shaped curve previously seen with Gaussian matrices in Figure 7. Moreover, the Rayleigh distribution seen in Figure 9 for Gaussian matric elements can be found in matrices with uniformly distributed elements. As an example Figure 10 is a frequency plot of the gaps, δ, between adjacent eigenvalues for one $N = 2048$ matrix with elements drawn from U[$-\sqrt{3}$, $\sqrt{3}$]. The interval (bin) size is $0{\cdot}01$ and the fitted curve is again a Rayleigh distribution, already seen in Figure 9. Its parameters are $n = 4420\delta \exp(-112\delta^2)$.

3.2 A forked distribution

I will call this the M distribution on account of its shape, made from two triangles as shown in Figure 11. This can be built from the uniform distribution on $(0 < x < 2)$ by taking the positive or negative square root in equal probability: $y = \pm\sqrt{x}$. This function is almost the complement of a Gaussian pdf.

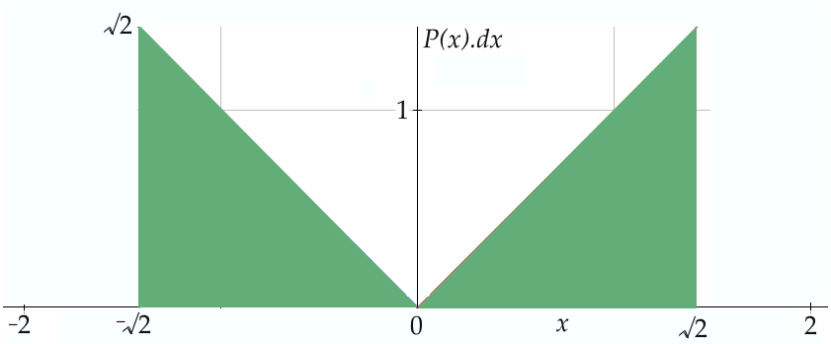

Figure 11: An M-shaped probability density function constructed from two halves of a square, side $\sqrt{2}$. $\mu = 0$, $\sigma^2 = 1$.

Figure 12 shows the probability density obtained by counting the numbers of eigenvalues in bins 1 unit wide, from five $N = 1024$ matrices with M-distributed elements. So even with such a large departure from Gaussian, the eigenvalues of the M distribution show all the signs of behaving similarly to the N[0, 1] matrices.

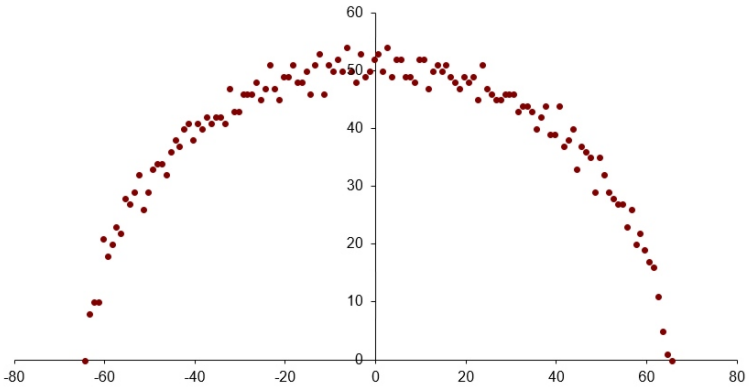

Figure 12: Frequency distribution of eigenvalues of $N = 1024$ matrices with the M-distribution. Combined results from sample of five matrices of element values.

3.3 A distribution from continued fractions

Several years ago I wrote a monograph on continued fractions, available on *ww.mathstudio.co.uk*. A continued fraction such as

$$a + \cfrac{1}{b + \cfrac{1}{c + \cfrac{1}{d + \cfrac{1}{\cdots}}}}, \qquad a, b, c, d, \ldots \text{ integers},$$

is written as $\{a : b, c, d, \ldots\}$ where the integers in the sequence are called the partial quotients of the fraction. For rational numbers this is a finite sequence, and for square roots it is a recurring palindromic sequence; for example

$$\sqrt{31} \;=\; \{5 : \underline{1, 1, 3, 5, 3, 1, 1, 10}\}$$

where the underlined sequence recurs *ad infinitum*. For other real numbers in $\mathbb{R}$ the sequence is unending and random, with a preponderance of 1s and 2s, fewer 3s and 4s, and a sparse smattering of higher integers, though there is no upper limit. As an example

$$\pi \;=\; \{\, 3 : 7,\ 15,\ 1,\ 292,\ 1,\ 1,\ 1,\ 2,\ 1,\ 3,\ 1,\ 14,\ 2,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 1,\ 84,\ 2,\ 1,\ 1,\ 15,\ 3,\ 13,$$
$$1,\ 4,\ 2,\ 6,\ 6,\ 99,\ 1,\ 2,\ 2,\ 6,\ 3,\ 5,\ 1,\ 1,\ 6,\ 8,\ 1,\ 7,\ 1,\ 2,\ 3,\ 7,\ 1,\ 2,\ 1,\ 1,\ 12,\ 1,\ 1,\ 1,\ 3,\ 1,\ 1,\ 8,\ 1,\ 1,$$
$$2,\ 1,\ 6,\ 1,\ 1,\ 5,\ 2,\ 2,\ 3,\ 1,\ 2,\ 4,\ 4,\ 16,\ 1,\ 161,\ 45,\ 1,\ 22,\ 1,\ 2,\ 2,\ 1,\ 4,\ 1,\ 2,\ 24,\ 1,\ 2,\ 1,\ 3,\ 1,\ 2,\ 1,\ 1,$$
$$10,\ 2,\ 5,\ 4,\ 1,\ 2,\ 2,\ 8,\ 1,\ 5,\ 2,\ 2,\ 26,\ 1,\ 4,\ 1,\ 1,\ 8,\ 2,\ 42,\ 2,\ 1,\ 7,\ 3,\ 3,\ 1,\ 1,\ 7,\ 2,\ 4,\ 9,\ 7,\ 2,\ 3,\ 1,\ 57,\ 1,\ \ldots\ \}\ .$$

The distribution function of these random partial quotients is given in section §16 of my article. It was derived first by the great Friedrich Gauss and elaborated by the Russian mathematician Alexander Khinchin who did much to develop probability theory in the 1920s. Gauss found that the probability that any partial quotient a_k in the unending tail of the sequence has the value v tends asymptotically to

$$\mathcal{P}(a_k = v) = \log_2\left(1 + \frac{1}{v(v+2)}\right) \;=\; \frac{1}{\ln 2}\,\ln\left(1 + \frac{1}{v(v+2)}\right) \qquad k \to \infty. \tag{8}$$

Clearly this distribution is far removed from a normal one. All the possible values are positive integers. Moreover, it does not have an arithmetic mean nor a finite variance nor higher moments because the sum

$$\sum_1^\infty v^k \ln\left(1 + \frac{1}{v(v+2)}\right)$$

does not converge for any $k > 1$. It is possible, to calculate other averages such as the geometric and harmonic means. The geometric mean $\mathcal{G}$ of a set of N numbers $\{v_1, v_1, \ldots v_1, v_2, v_2, \ldots v_2, v_3, \ldots\ v_r\}$, in which v_1 occurs p_1 times, v_2 occurs p_2 times, *etc.* is

$$\left(v_1^{p_1}.v_2^{p_2}.v_3^{p_3}.\ldots.v_r^{p_r}\right)^{\frac{1}{N}}, \qquad p_1 + p_2 + \cdots + p_r = N\ .$$

In the limit $N \to \infty$

$$\ln \mathcal{G} \;=\; \frac{1}{\ln 2}\sum_{v=1}^{\infty} \ln v\,\ln\left(1 + \frac{1}{v(v+2)}\right) \;=\; 0\cdot 987849$$

from which $\mathcal{G} = 2\cdot 6854520\ldots$, known as Khinchin's constant. It states the remarkable tendency of the geometric mean of almost all the partial quotients of 'almost all' real numbers to converge on this value. The harmonic mean $\mathcal{H}$ is similarly obtained from the formula

$$\frac{1}{\mathcal{H}} \;=\; \frac{1}{\ln 2}\sum_{v=1}^{\infty} \frac{1}{v}\ln\left(1 + \frac{1}{v(v+2)}\right) \;=\; \frac{0\cdot 39713}{\ln 2} \;=\; 0\cdot 5729$$

from which $\mathcal{H} = 1 \cdot 7454$. $\mathcal{H}$ is less than the geometric mean.

I have generated a number of symmetric matrices with elements drawn from the Gauss-Khinchin distribution, independent and identically distributed. Their eigenvalues vary wildly because they are dominated by sporadic large values of some matrix elements.

Since 1 is the most common element value, one limiting form for these matrices is that in which all elements equal 1. Such an $N \times N$ matrix has characteristic polynomial $\lambda^{N-1}(\lambda - N)$ and hence one eigenvalue equal to N and the rest all zero. For each diagonal element which is changed from 1, the multiplicity of the $\lambda = 0$ eigenvalue decreases by 1, so if $N - 1$ diagonal elements are changed, no eigenvalue will be zero and all will in general be different. The off-diagonal elements have a less strong action towards decreasing the number of zero eigenvalues, but as a guide between $2N$ and $3N$ off diagonal elements must not equal 1 for a zero eigenvalue to be avoided, the diagonal ones all remaining at 1.

Another limiting form is that in which all off-diagonal elements are very small relative to those on the diagonal. The matrix then looks almost like a diagonalised matrix, whose eigenvalues are necessarily close to the diagonal elements themselves. The behaviour is not so simple when the large elements are an off-diagonal pair (the matrix being symmetric).

I conclude that when the matrix elements come from the Gauss-Khinchin distribution, the eigenvalues do not follow the pattern shown by the other matrices studies, which have an arithmetic mean and variance 1. The conclusion of this whole section is that the eigenvalues of a real symmetric random matrix will be statistically similar to one drawn from N[0, 1] only provided i) the mean of the elements is 0, ii) their variance is 1, ii) all higher moments are finite. The Gauss-Khinchin distribution fails these criteria.

4 Wigner's semicircle theorem

This theorem was first proved by Wigner in 1956. It states essentially that the limiting form, $N \to \infty$, of the probability density function (pdf) of eigenvalues from an i.i.d. random matrix with mean zero, variance 1 and finite higher moments, suitably normalised, is

$$P(x)\,dx \;=\; \frac{1}{2\pi}\sqrt{4 - x^2}\;dx\,, \qquad |\lambda| \le 2. \tag{9a}$$

He chose his semi-circle to have radius 2 units. The units are normalised by dividing the eigenvalues by $\sqrt{N}$ in recognition of Eq 2, that the largest eigenvalues grow as $\sqrt{N}$: $x = \lambda/\sqrt{N}$. In terms of λ

$$P(\lambda) \;=\; \frac{1}{2\pi N}\sqrt{4N - \lambda^2}\,, \qquad |\lambda| \le 2\sqrt{N}. \tag{9b}$$

You may care to look back at Figures 7 and 8 for one $N = 2435$ N[0,1] matrix and compare the empirically fitted curves with these formulae. $2\sqrt{N} = 98 \cdot 7$ is to be compared with 100 for the half width, and $\sqrt{N}/\pi = 15 \cdot 7$ with $15 \cdot 5$ in Figure 7 for the central level.

Wigner proved his law using the moments of this distribution. Recall that the moments of a variable x which has a probability density $P(x)\,dx$ are defined by

$$M_n \;=\; \int_a^b x^n P(x)\,dx$$

where a and b are the extreme values that x can take. The moments are the expectation values of the respective powers of x and give increasing detail about the position, size and shape of the distribution. The full set of moments in almost all cases defines the distribution uniquely. In the derivation below moments are approached from two opposite ends of the problem and meet in the middle; I show that the moments arising from the distribution of matrix elements values approach the moments of the semi-circle distribution as $N \to \infty$, and so the two must be asymptotically identical. Technically they are said to converge to each other 'in expectation' which is a rather weak form of converging 'in probability'.

4.1 Catalan numbers

Coming in one direction at the problem, we assume for the time being that the semi-circle distribution is correct and calculate the sequence of its moments. Since the semi-circle is symmetric about $x = 0$, it is clear that the mean, M_1, and all higher odd moments must be zero. The even moments are therefore central moments, of which M_2 is the variance. Higher even moments are given by

$$M_{2k} \;=\; \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}\;dx\,, \qquad k = 1, 2, 3, ...$$

You can use a computer integration program like Mathematica to evaluate the first few of these to find

$$M_2 = 1, \quad M_4 = 2, \quad M_6 = 5, \quad M_8 = 14, \quad M_{10} = 42, \quad M_{12} = 132 \tag{10a}$$

and then use the On-Line Encyclopaedia of Integer Sequences (*oeis.org*) to see that these are the Catalan numbers defined by

$$C_k \;= \frac{(2k)!}{k!\,(k+1)!}\,, \qquad \text{so } M_{2k} = C_k\,. \tag{10b}$$

If you wish to press on with evaluating the integrals by hand, first make the substitution $x = 2y$ to obtain

$$\frac{2}{\pi}\int_{-1}^{1} (2y)^{2k}\sqrt{1-y^2}\;dy\,,$$

then make the further change of variable, $y = \sin\theta$, to obtain

$$M_{2k} \;= \frac{2^{2k+1}}{\pi}\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta\,\cos^2\theta\;d\theta\,.$$

From here we integrate by parts

$$\int_a^b u.dv \;=\; u.v\big|_a^b \;-\; \int_a^b v.du\,, \qquad u = u(\theta), \quad v = v(\theta)\,,$$

using the particular division of the integrand

$$u \;=\; \sin^{2k-1}\theta\,\cos^2\theta, \qquad dv \;=\; \sin\theta\,d\theta$$

$$du \;=\; (2k-1)\sin^{2k-2}\theta\,\cos^3\theta\,d\theta \;-\; 2\sin^{2k}\theta\,\cos\theta\,d\theta, \qquad v = -\cos\theta\,.$$

The product uv is zero at both limits so contributes nothing, while

$$-v.du \;=\; (2k-1)\sin^{2k-2}\cos^2\theta\,(1-\sin^2\theta)\,d\theta \;-\; 2\sin^{2k}\theta\,\cos^2\theta\,d\theta\,.$$

There are signs of a recursion relation here so we note that

$$M_{2k-2} \;= \frac{2^{2k-1}}{\pi}\int_{-\pi/2}^{\pi/2} \sin^{2k-2}\theta\,\cos^2\theta\,d\theta\,.$$

$$\text{Then} \qquad M_{2k} \;=\; 4(2k-1)M_{2k-2} - (2k+1)M_{2k}\,.$$

$$M_{2k} \;=\; \frac{2(2k-1)}{k+1}M_{2k-2} \tag{10c}$$

which generates the Catalan numbers at Eq 9 above.

The Catalan numbers, named after the 19th century Belgian mathematician Eugène Catalan, feature widely in combinatorics. Essentially C_k counts the number of matched pairs which can be made from a two equal set of objects, k objects in each set, when order is important. Some instances are illustrated in Figure 13. Others include

1. the number of ways k left brackets and k right brackets can be placed so that valid pairs of open and closed brackets result. For $k = 2$ there are the two pairs {} {} and {{ }}. For $k = 3$ there are five:
   { }{ }{ }, {{ }}{ }, { }{{ }} , {{ }{ }}, {{{ }}} .
2. the brackets above can be replaced by letters or numerals to form a Dyck word (after Walther von Dyck) in which, at any position along the word, the number of 'b's does not exceed the numbers of 'a's. Thus the above list for $k = 3$ translates to
   ababab, aabbab, abaabb , aababb, aaabbb .
3. the number of ways to draw a peaked 'mountain range' from k diagonal up strokes / and k down strokes \ so that no line dips below the starting point at sea level (Figures 13 and 14),
4. the number of ways to dissect a convex polygon with k sides into triangles by non-crossing cuts between vertices,
5. the number of ways $2k$ people sitting in a circle can shake hands with each other pairwise,
6. in graph theory, the number of ordered rooted trees which have k edges, $k+1$ vertices (Figure 13, right). A tree is a graph with no cycles. Ordered means that the vertices are labelled 1, 2, 3, ... ; this labelling means that trees that would otherwise be equivalent under some symmetry operation remain distinct. Rooted means that one external vertex is chosen as being 'in the ground' and acts as a starting point for growing the tree. The pairing is of the paths one way then back along each branch, *i.e.* edge of the graph.
7. the number of ways k applications of a binary operator, such as +, can be associated. This is illustrated by using k pairs of brackets to show how $k+1$ quantities can be added in different stages. For $k = 3$ the quantity $a+b+c+d$ can be built up in 5 ways using three + operations:
   $$(((a+b)+c)+d),\quad ((a+b)+(c+d)),\quad ((a+(b+c))+d),\quad (a+((b+c)+d)),\quad (a+(b+)c+d))).$$
8. in graph theory, the number of 'full binary trees' with k internal vertices. A full binary tree is a rooted tree in which every internal vertex has exactly 2 children. It has $k+1$ leaves (out to the end vertices), $2k$ edges and $k+1$ external vertices, making $2k+1$ vertices in all.

There are other interesting and well illustrated examples on the internet. Figure 14 illustrates the 2, 5, and 14 matched pairing of 2, 3 and 4 pairs of up and down pen strokes drawing a mountain range. Notice how the shapes from the smaller pairings form building blocks within the larger shapes towards the bottom of the diagram. This is the geometric basis of recursion.

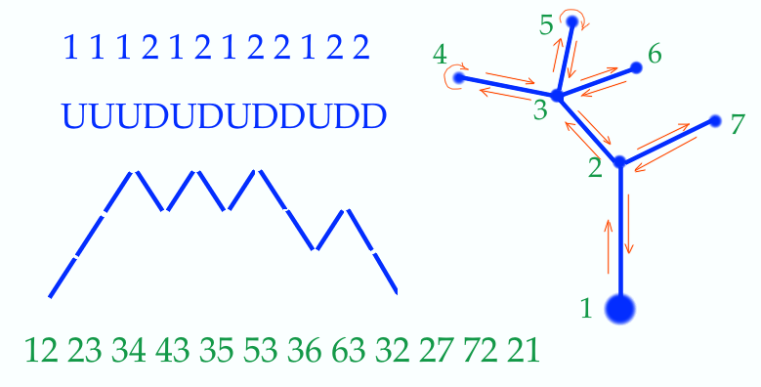

Figure 13: Five ways to represent a set of 6 pairings. U stands for Up, D for Down. There are $C_6 = 132$ possible matched pairings of 6+6 items. In the tree graph each edge defines the there-and-back in a walk around the tree.

Out of passing interest, I will also point out an algorithm to translate the bracketing at item 7 with the full binary trees at item 8. Place the quantities x, y which are at the innermost pairing(s) at two vertices (labelled x, y) at a level furthest from the root. Their combination under $+$ is a vertex at the next level near the root, such that they form two leaves from that vertex. Label this parent vertex $(x+y)$. Continue in this way to the root vertex, with corresponds with the total expression. As examples the binary trees representing $(((a+b)+c)+d)$ and $((a+b)+(c+d))$ are shown in Figure 15.

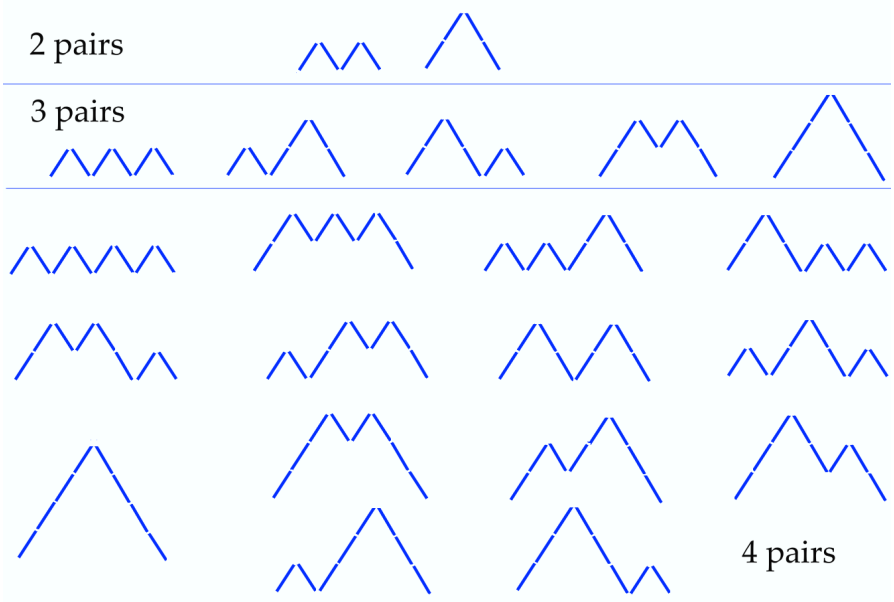

Figure 14: Mountain-style representations of the matched pairings of 2+2, 3+3 and 4+4 items, illustrating Catalan numbers $C_2 = 2$, $C_3 = 5$, $C_4 = 14$.

I remarked above that the eigenvalue parameter used in the semi-circle law is normalised by dividing λ by $\sqrt{N}$. In terms of the actual eigenvalues the moments are

$$M_2 = N, \qquad M_4 = 2N^2, \qquad M_6 = 5N^3, \qquad M_8 = 14N^4, \qquad M_{10} = 42N^5\,, \quad M_{2k} = C_k N^k\,. \tag{11}$$

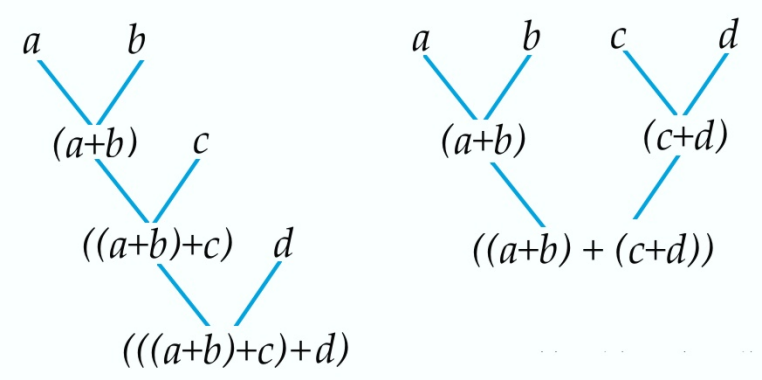

Figure 15: Binary trees representing two associations of a, b, c, d under $+$.

4.2 Moments of an equally-spaced distribution

In the frequency plots of Figures 8 and 12 we have pictured the eigenvalues as being points along the number line from λ^-_{max} to λ^+_{max} and counted the number within each short interval of length $\delta\lambda$. Take λ to be a variable position on the number line and let $\nu(\lambda)\delta\lambda$ be the number of eigenvalues in length $\delta\lambda$ found in a numerical experiment with a large random matrix. The total number of eigenvalues is N so $\nu(\lambda).\delta\lambda/N$ is their probability density for this matrix, usually called the 'empirical spectral density' or ESD of the eigenvalues. Its k^{th} moment is

$$M_k \;=\; \int_{\lambda^-_{max}}^{\lambda^+_{max}} \lambda^k \,\frac{\nu(\lambda)}{N}\, d\lambda\,.$$

This can be written in a better way for our purposes by seeing that as $\delta\lambda$ becomes very short, less than the minimum distance between adjacent eigenvalues, it will contain either no eigenvalue or just the one, if there happens to be one close to position λ_j. (I assume that there are no multiple eigenvalues.) The integral over λ degenerates to a sum over the λ_j, for each of which $\nu(\lambda_j) = 1$. The k^{th} moment is therefore

$$M_k \;=\; \frac{1}{N}\sum_{j=1}^{N} \lambda_j^k\,. \tag{12}$$

Before attempting to calculate the expectation value of these for a random matrix, let us pause briefly to use Eq 12 to calculate the moments of the artificial approximate distribution of eigenvalues postulated in §2.2. In this the eigenvalues are equally spaced with separation distance $2q$ where q is $\beta/\sqrt{N+1}$, and β has been identified as somewhere between $2\cdot 1$ and $\sqrt{3}$ (see Eqs 5 and 6). We do not expect the moments to equal those of the semi-circle distribution since the uniform, comb-like distribution does not have the increasingly wider spacings of the outermost eigenvalues. However, there may be similarities which could prove helpful guides to random matrices. In the uniform distribution the N eigenvalues are at λ values of $-(N-1)q$, $-(N-3)q$, ... $-q$, q, $3q$, $5q$, ... $(N-1)q$. The moments are

$$M_{2k} \;=\; \frac{2}{N}\sum_{j=1}^{N/2} \lambda_j^{2k}, \qquad \lambda_j = (2j-1)q, \qquad q = \frac{\beta}{\sqrt{N+1}}\,.$$

As an example,

$$M_4 \;=\; \frac{\beta^2}{15(N+1)}\,(3N^3 - 3N^2 - 7N + 7) \quad \text{which tends to} \quad \frac{\beta^4}{5}(N^2 - 2N - ...) \text{ as } N \to \infty.$$

In general M_{2k} tends to a polynomial of degree k as

$$M_{2k} \to \frac{\beta^{2k}}{2k+1} N^k \; - \; \mathcal{O}(N^{k-1}) \,. \tag{13}$$

The limiting values of the first few moments are :

$$M_2 \to \frac{\beta^2}{3} N, \qquad M_4 \to \frac{\beta^4}{5} N^2, \qquad M_6 \to \frac{\beta^6}{7} N^3, \qquad M_8 \to \frac{\beta^8}{9} N^4.$$

This sequence has the N^k dependence seen in Eq 11. A log-log plot of $J_{2k} = \beta^{2k}/(2k+1)$ against the Catalan numbers C_k gives a fairly straight line indicating a relationship approximately of the form $J_{2k} = A {C_k}^b$. I find that $\beta = 1 \cdot 9186$ makes $b = 1$ and so gives J_{2k} directly proportional to C_k: $J_{2k} \approx 1 \cdot 3 \, C_k$. It is surprising that this equally-spaced distribution and the semi-circle distributions have moments sufficiently similar that those of one are roughly equal to those of the other scaled up by about $1 \cdot 3$.

4.3 Trace of matrix powers

We now return to the main task of calculating the moments of a symmetric random matrix with mean zero, unit variance and finite higher moments, and approach the problem from the opposite direction from §4.2. We need expressions for the sequence of expectation values of moments of the eigenvalues to compare with the moments of the semi-circle at Eq 11. The first step is to use the fact that for any $N \times N$ matrix A raised to a power k

$$Tr(A^k) \; = \; \sum_{j=1}^{N} \lambda_j^k \,. \tag{14}$$

We therefore replace the sum over eigenvalues in Eq 12 by the trace of a power of the matrix. The proof of Eq 14 is by induction on k and runs as follows. For any matrix $A\phi = \lambda\phi$ by definition of an eigenvalue and its eigenvectors. Suppose that for $k \geq 1$ that $A^k\phi = \lambda^k\phi$. Then

$$A^{k+1}\phi \; = \; A^k A\phi \; = \; A^k \lambda\phi \; = \; \lambda A^k \phi \; = \; \lambda^{k+1}\phi \,.$$

So A and A^k share the same eigenvectors, and the eigenvalues of A^k are those of A raised to the k^{th} power. Now use the fact that the trace of a matrix is the sum of its eigenvalues to obtain Eq 14.

The challenge, therefore, has now moved to finding the expectation value of $Tr(A^k)$. The argument here has similarities with that of §2.3 since if k is odd, all the terms which contribute to the trace have an odd number of factors such as $....a_{11}a_{13}^2 + a_{12}^2 a_{22}....$ Since all elements of A are assumed independent, with no correlation between any two, $\mathrm{E}(x\,y) = \mathrm{E}(x)\,\mathrm{E}(y)$ where x, y represent any distinct two elements a_{ij} or their powers. When the expectation value of these products is taken, every term which has a factor a_{ij} to a single power will average to zero. What about the higher odd powers such as a_{ij}^3 which occur in all odd moments? Clearly these will also tend to be distributed symmetrically about zero so will all average to zero. We may conclude that all odd moments tend to zero on average as $N \to \infty$.

So we now tackle the problem of determining the expectation values of $Tr(A^{2k})$, $k = 1, 2, 3, ...$ as $N \to \infty$. In the accounts of this proof which I have read in the literature, mathematicians move forwards quickly to equate the limiting expectation value with the number of trees graphs which have k edges, and hence with the Catalan numbers and the moments of the semi-circle distribution.

Personally I do no find this at all obvious, so I have examined the matter in small stages. To start, I examined the types of term which are summed in the trace and how many of each there are.

If we start with a general $N \times N$ matrix A and square it, each element of A^2 is a sum of N terms, each of which is a product of 2 factors. Denote this by ${}^{+}N(2)$. The diagonal has N such elements so $Tr(A^2) =^{+} N^2(2)$, *i.e.* the sum of N^2 terms each of which is the product of two elements of A. Now square A^2. Each element of A^4 is a sum of N terms, each of which has N^2 sub-terms, each sub-term being the product of 4 factors. So a representative element of A^4 has structure ${}^{+}N^3(4)$. The diagonal has N such elements so $Tr(A^4) =^{+} N^4(4)$. Continuing in this shows see that $Tr(A^{2k})$ is the sum of N^{2k} terms, each of which is the product of $2k$ factors; that is, the degree of each term is $2k$. Its structure is ${}^{+}N^{2k}(2k)$.

As an example, here is a selection of terms from $Tr(A^8)$ where A is the general algebraic 6×6 symmetric matrix, so $N = 6,\ 2k = 8$:

$$a_{11}^8 + 8\,a_{11}^6\,a_{12}^2 + 20\,a_{11}^4\,a_{12}^4 + 16\,a_{11}^2\,a_{12}^6 + 2\,a_{12}^8 + 8\,a_{11}^6\,a_{13}^2$$
$$+40\,a_{11}^4\,a_{12}^2\,a_{13}^2 + 48\,a_{11}^2\,a_{12}^4\,a_{13}^2 + 8\,a_{12}^6\,a_{13}^2 + 20\,a_{11}^4\,a_{13}^4 + \ ....$$
$$..... \quad + 96\,a_{12}\,a_{13}\,a_{14}^2\,a_{16}^2\,a_{24}\,a_{34} + 32\,a_{12}\,a_{13}\,a_{15}^2\,a_{16}^2\,a_{24}\,a_{34} + 16\,a_{12}\,a_{13}\,a_{16}^4\,a_{24}\,a_{34}$$
$$+16\,a_{11}^3\,a_{12}\,a_{13}\,a_{22}\,a_{24}\,a_{34} + 64\,a_{11}\,a_{12}^3\,a_{13}\,a_{22}\,a_{24}\,a_{34}$$
$$+96\,a_{11}\,a_{12}\,a_{13}\,a_{14}^2\,a_{22}\,a_{24}\,a_{34} + 16\,a_{11}^2\,a_{12}\,a_{16}\,a_{23}\,a_{34}^2\,a_{36}.....$$

In all there are $6^8 = 1,679,616$ underlying terms, each the product of 8 factors. Even though the symmetry means that equal terms a_{ij}, a_{ji} will be collected together, the sum of the coefficients remains at 1,679,616. The expectation value E( ) of the trace is the sum of the expectation values of its N^{2k} terms. Again, since the elements a_{ij} are drawn from a population (not necessarily Gaussian) symmetrical about the mean 0, $\mathrm{E}(a_{ij}) = \mathrm{E}(a_{ij}^3) = \mathrm{E}(a_{ij}^5) = ... = 0$. This eliminates a vast numbers of terms from $\mathrm{E}(Tr(A^{2k}))$, leaving only those which are products of even powers to contribute.

That is perhaps all I can say at present about the general large random symmetric matrix. I will now look at small matrices for which explicit values can be calculated numerically and see if I can discern patterns which point to the behaviour as N becomes large. I have used software to calculate algebraically the trace of powers of symmetric matrices up to $N = 7$ and examined the number of terms of each type. We can expect the types of term in the trace of A^{2k} to be related to the partitions of the index $2k$. Thus, in terms of powers of elements, $Tr(A^4)$ for $N = 3$ has the partitions

$$4, \quad 2{+}2, \quad 2{+}1{+}1, \quad 1{+}1{+}1{+}1, \quad \text{but not } 3{+}1.$$

To be clear, '2+1+1' means terms of the form $a_{ij}^2\,a_{kl}\,a_{mn}$. The index 6 in $Tr(A^6)$ for $N = 3$ has 11 partitions and 9 of these occur:

$$6, \quad 4{+}2, \quad 4{+}1{+}1, \quad 3{+}2{+}1, \quad 3{+}1{+}1{+}1, \quad 2{+}2{+}2, \quad 2{+}2{+}1{+}1, \quad 2{+}1{+}1{+}1{+}1, \quad 1{+}1{+}1{+}1{+}1{+}1\,.$$

The missing ones are 5+1, 3+3, meaning there are no terms of the forms $a_{ij}^5\,a_{kl}$ or $a_{ij}^3\,a_{kl}^3$. The only partitions which can contribute to the expectation value are ones entirely with even numbers, 6, 4+2 and 2+2+2, and there are 9, 126 and 132 of these respectively. Table 2 lists the total numbers of terms which have no factors with odd index; these are the ones which contribute to $\mathrm{E}(Tr(A^{2k}))$. The upper panel gives the absolute number and the lower panel the percentage.

It would be fortunate to extrapolate a pattern in Table 2 to indefinitely large N. The following patterns do occur:

No. of contributing (even index) terms

$N \setminus 2k$	2	4	6	8	10	12	14
2	4	12	40	144	544	2112	8320
3	9	45	267	1785	12999		
4	16	112	952	9184			
5	25	225	2485	31185			
6	36	396	5376	82896			
7	49	637	10255				

Percentage contributing

2	100	75	63	56	53	52	51
3	100	56	37	27	22		
4	100	44	23	14			
5	100	36	16	8			
6	100	31	12	5			
7	100	27	9				

Table 2: Number (top panel) and percentage (lower) of terms in the trace of A^{2k}, an $N \times N$ symmetric $[0, 1]$ random matrix, which contribute to its expectation value.

- the total number of terms is N^{2k}, and the sum of indices in each term is $2k$,
- for $N = 2$ the number of contributing terms is $2^k(2^{k-1} + 1)$. The fraction of contributing terms is $\frac{1}{2} + \frac{1}{2^k}$.
- for $2k = 2$ the number of contributing terms is N^2,
- for $2k = 4$ the number of contributing terms is $2N^3 - N^2$. This is a polynomial of the form found for the equally spaced distribution at Eq 13. As a fraction of the total this is $\frac{2N-1}{N^2}$ which tends to zero like $2/N$ as $N \to \infty$.

Prompted by Eq 13 we may suspect that the other columns also fit to polynomials of degree k. It has been easy to find that for $2k = 6$ the polynomial

$$5N^4 - 5N^3 - N^2 + 2N \tag{15a}$$

is an exact fit to the numbers of contributing terms. There is one integer too few in the $2k = 8$ column to fit a unique degree 5 polynomial, but, suspecting that the coefficient of N^5 is the Catalan number 14, I find a 4th order polynomial can be fitted exactly to the remainders. Thus the number of contributing terms is

$$14N^5 - 19N^4 - 10N^3 + 24N^2 - 8N\,. \tag{15b}$$

These values, of course, are not the $2k^{th}$ moments but just the number of contributing terms. Nevertheless it is encouraging to see that the leading terms for $2k = 2$, 4, 6, 8 are respectively $1N^2$, $2N^3$, $5N^4$ and $14N^5$. When divided by N as required by Eq 12, these would be the limiting forms of the moments on the semi-circle distribution. This fact suggests both that the semi-circle distribution is indeed the limiting form and that it is the number of contributing terms which has the largest influence on the moments.

To quantify the actual moments we need the expectation values of powers of a_{ij} and for this we need to specify their distribution. Here are the calculations for a Gaussian distribution, and a

uniform distribution, both with variance 1.

Gaussian N[0, 1]: $\mathrm{E}(x^2) = 1$, by definition of the variance.

$$k = 2: \qquad \mathrm{E}(x^4) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2}\, dx \;=\; 3\,, \tag{16}$$

$$k = 3: \qquad \mathrm{E}(x^6) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^6 e^{-x^2/2}\, dx \;=\; 15\,,$$

$$k = 4: \qquad \mathrm{E}(x^8) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^8 e^{-x^2/2}\, dx \;=\; 105\,.$$

The sequence continues with rapidly increasing values: 1, 3, 15, 105, 945, 10395, 135135, 2027025, ... $\frac{(2k-1)!}{2^{k-1}\,(k-1)!}$. These are the number of ways to choose k disjoint pairs of items from $2k$ items.

Uniform U[0, 1]:

$$k = 2: \qquad \mathrm{E}(x^4) \;=\; \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} x^4\, dx \;=\; \frac{9}{5}\,, \tag{17}$$

$$k = 3: \qquad \mathrm{E}(x^6) \;=\; \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} x^6\, dx \;=\; \frac{27}{7}.$$

This sequence does not grow so quickly because the maximum value of $|x|$ is capped at $\sqrt{3}$: 1, $\frac{9}{5}$, $\frac{27}{7}$, 9, $\frac{243}{11}$, $\frac{729}{13}$, ..., $\frac{3^k}{2k+1}$.

I have examined the number of terms of each partition type for the matrices referenced in Table 2 and, by weighting them according to the moments just calculated, arrived at values for $\mathrm{E}(Tr(A^{2k}))$. For example, the theoretical value of $\mathrm{E}(Tr(A^4))$ is $9 \times 3 + 36 = 63$. I attempted comparison with numerical samples by using Mathematica to calculate algebraically $Tr(A^4)$ for A a 3×3 symmetric matrix, then substituting 100 sets of random numbers from N[0, 1] for the six independent coefficients a_{11}, $a_{12} = a_{21}$, etc. I hence obtained mean values over the 100 samples of each of the different types of term in the trace. There are four types of term as listed in Table 3. Here a, b, c, d represent any of the elements $a_{ij}, 1 \le i \le 4,\ 1 \le j \le 4$. Observe that there are none of type $a^3 b$. The agreement between theory and experiment is not particularly good because the variance of powers of the elements is high. For instance a_{11}^4 varied from almost 0 to 52, the latter due to a_{11} being $2 \cdot 68$ – rare but possible in a normal distribution.

Type	No. of terms	theory mean	exptl. mean	sum
a^4	9	3	3.67	33.04
a^2b^2	36	1	1.46	52.46
a^2bc	12	0	–0.03	–0.38
$abcd$	24	0	0.08	1.85
Total	81			86.97

Table 3: Summary of types of term in trace of 4th moment of a 3×3 symmetric matrix with elements in N[0, 1]. The theoretical value of $\mathrm{E}(Tr(A^4))$ is $9 \times 3 + 36 = 63$.

Another example is the theoretical $\mathrm{E}(Tr(A^6))$ for elements from N[0, 1]:

$$9 \times 15 + 126 \times 3 \times 1 + 132 \times 1 \times 1 \times 1 \;=\; 135 + 378 + 132 \;=\; 645\,.$$

If this had been the uniform distribution U[0, 1] of aij, the result would have been

$$9 \times \tfrac{27}{7} + 126 \times \tfrac{9}{5} + 132 \;=\; 34{\cdot}7 + 226{\cdot}8 + 132 \;=\; 393{\cdot}5\,.$$

The weightings given to the terms from the higher powers of a_{ij} are large and mean that there is no simple relation between the number of even-index terms in the trace and the value of the trace. Take the case of a 4×4 matrix raised to its 8th power. Table 4 lists the number of each type of term (according to the indices of its factors) and the contribution each makes to the trace.

Partition	Number	Weight	Contribution
8	16	105	1680
6+2	480	15	7200
4+4	384	9	3456
4+2+2	4248	3	12744
2+2+2+2	4096	1	4096
some odd powers	56312	0	0
Total	65536		29176

Table 4: The number of various type of term in $Tr(A^8)$ for A a 4×4 symmetric matrix, and their contributions to the expectation value of the trace.

The remaining question is: in the general case of large N, how many are there of these various contributing types of term? Knowing the weighting of each, we would ideally like to decide on the dominant types and hence estimate the limiting values of $\mathrm{E}(Tr(A^{2k}))$ for all k as $N \to \infty$. Terms which grow as powers of N will dominate over terms which depend just on k.

Define ratios a) r_t of mean value of trace to number of contributing terms, and b) r_f of mean value of trace to the first term in the polynomial, $C_k N^{k+1}$. Values are listed together with the total value of the trace in Table 5. I find that for each value of $2k$ the ratio r_t falls closely as

$$r_t \;=\; \exp\left(\frac{\alpha}{N^{1+\varepsilon}}\right) \tag{18}$$

where α and ε are as given below. The values in italics are extrapolated and hence approximate. Clearly all these formulae tend to 1 as $N \to \infty$, meaning that the larger weightings of the higher powers of a_{ij} have a diminishing effect relative to the increasing number of terms. In this sense the value of $Tr(A^{2k})$ tends to the number of terms in $Tr(A^{2k})$ as $N \to \infty$. By retaining only the leading term in the polynomial for the number of terns (meaning that we do not subtract a term in N^k), we compensate in a loose way for the higher weightings of the factors in a_{ij}^4 and higher powers. For this reason the ratios r_f =$\mathrm{E}(Tr(A^{2k}))/(C_k N^{k+1})$ in the lower panel of Table 5 converge more quickly than those in the middle one. For matrix elements drawn from a uniform distribution, for which the weightings are all smaller, this convergence can be expected to be quicker.

Let us take stock of where we are. We have reached a position at which the numerical evidence from small matrices points to the number of contributing terms in the trace $Tr(A^{2k})$ being given by a polynomial whose leading term is $C_k N^{k+1}$ where C_k is a Catalan number. The contributing terms are those which do not statistically average to zero, and they all are composed only of factors which have an even power. Each factor of the form a_{ij}^2 is weighted 1, and higher powers have a higher weighting

N\ 2k	2	4	6	8	10	12	14
Trace value							
2	4	20	156	1656	22320	365760	7071120
3	9	63	645	8601	141975		
4	16	144	1824	29136			
5	25	275	4155	77385			
6	36	468	8220	175080			
7	49	735	14721				
ratio r_t							
2	1	1.667	3.900	11.500	41.03	173.18	849.89
3	1	1.400	2.416	4.818	10.92		
4	1	1.286	1.916	3.172			
5	1	1.222	1.672	2.481			
6	1	1.182	1.529	2.112			
7	1	1.154					
ratio r_f							
2	1	1.250	1.950	3.696	8.30	21.65	64.39
3	1	1.167	1.593	2.528	4.64		
4	1	1.125	1.425	2.032			
5	1	1.100	1.330	1.769			
6	1	1.083	1.269	1.608			
7	1	1.071	1.226				

Table 5: Upper panel: mean values of traces of random symmetric matrices from N[0, 1], including weighting of terms. Middle panel: ratios r_t of mean value of trace to number of contributing terms. Lower panel: ratio r_f of mean value of trace to the first term in the polynomial, $C_k N^{k+1}$.

$2k$	2	4	6	8	10	12	14
α	0	1.029	2.832	5.1451	*7.98*	*11.6*	*16*
ε		0.015	0.0603	0.0772	*0.0866*		

Table 6: Parameters in the ratio r_t in Eq 18.

depending on the precise statistics of the distribution from which the a_{ij} are drawn. We found that as N increases, the number of contributing terms in the trace becomes an increasingly closer approximation to the value of the expectation value of that trace, as Eq 15a, b. This approximation is equivalent to replacing the true weighting of the higher powers by 1. From Eq 12, dividing the trace by N gives the k^{th} moment in terms of the actual eigenvalues, and dividing by N^k gives the moments in terms of the normalised eigenvectors $\lambda/\sqrt{N}$ as used by Wigner in his semi-circle law.

The outstanding point we have yet to explain is why the Catalan numbers appear as coefficients of the leading terms in the polynomials, as at Eq 15. A subsidiary point is why these leading terms over-estimate the number of contributing terms – in other words, why the second terms in the polynomials appear to be $-\mathcal{O}(N^{k-1})$. The Catalan numbers essentially count the number of ways two equal sets of objects can be paired. The pairings in the trace are the pairing of a_{ij} with a_{ji} (which are equal by symmetry of the matrix) to give the even power a_{ij}^2.

It is not difficult to see why the numbers of contributing terms for $2k$ = 2 and $2k$ = 4 are

N^2 and $2N^3 - N^2$ respectively. I illustrate this for $N = 4$ and record only the indicies, so that ij represents the element a_{ij}. The square of A_4 is

$$ {A_4}^2 = \begin{pmatrix} \Sigma_j 1j.j1 & \Sigma_j 1j.j2 & \Sigma_j 1j.j3 & \Sigma_j 1j.j4 \\ \Sigma_j 2j.j1 & \Sigma_j 2j.j2 & \Sigma_j 2j.j3 & \Sigma_j 2j.j4 \\ \Sigma_j 3j.j1 & \Sigma_j 3j.j2 & \Sigma_j 3j.j3 & \Sigma_j 3j.j4 \\ \Sigma_j 4j.j1 & \Sigma_j 4j.j2 & \Sigma_j 4j.j3 & \Sigma_j 4j.j4 \end{pmatrix} $$

When the matrix is symmetric, $ij = ji$ and this pairing of elements across the matrix diagonal produces elements raised to even powers. Clearly in the above matrix this can occur only for elements down its diagonal. In the top row $\Sigma_j 1j.j1$ contributes a sum of four squares, $11^2 + 12^2 + 13^2 + 14^2$ to the trace. There are thus 4 rows each contributing 4 terms, giving 4^2, or N^2 contributing terms. The argument readily generalises to higher N.

For the fourth power, ${A_4}^4$, the top row of the matrix makes the following contribution to the trace:

$$ (\Sigma_h 1h.h1)(\Sigma_j 1j.j1) + (\Sigma_h 1h.h2)(\Sigma_j 2j.j1) + (\Sigma_h 1h.h3)(\Sigma_j 3j.j1) + (\Sigma_h 1h.h4)(\Sigma_j 4j.j1). $$

with the other three rows contributing similar terms. The first term above, when expanded, has 16 sub-terms involving 2^{nd} powers, while the other three terms have only 4 each. For example, the third term above is

$$ (11.13 + 12.23 + 13.33 + 14.43)(31.11 + 32.21 + 33.31 + 34.41) $$

of which only $11^2.13^2$, $12^2.23^2$, $13^2.33^2$ and $14^2.34^2$ – all products between corresponding terms – produce square factors. The number of contributing terms in $Tr({A_4}^4)$ is therefore 16+4+4+4 from row 1, 4+16+4+4 from row 2, 4+4+16+4 from row 3 and 4+4+4+16 from the bottom row. Generalising this to other N, the total number of contributing terms is $(N^2 + [N-1]N)N = 2N^3 - N^2$. We are interested in how the Catalan coefficient 2 arises, and note that there are two types of contributing terms, one arising from the diagonal element of A_N^{2k} and the other from the off-diagonal elements. Unfortunately the sixth moment, $2k = 6$ does not furnish so easy a rationale for the Catalan number $C_3 = 5$. We would be cheerful if we could find 5 types of combination each producing $N^3 - N^2$ contributing terms in each of the N rows to explain the polynomial $5N^4 - 5N^3 - N^2 + 2N$. I examined the matrix with $N = 3$ in detail but found no obvious partition of the 89 contributing terms into 5 sets; perhaps 3 is too small a matrix for this partition to show up.

4.4 Tree graphs and paired matrix elements

Each of the terms in the trace, such as $a_{11}\, a_{12}\, a_{13}\, a_{14}^2\, a_{22}\, a_{24}\, a_{34}$ in the example of $Tr(A^8)$ listed in §4.3, can be represented by a graph. The procedure is that used to draw the tree graph in Figure 13, by letting each i or j in the index of a_{ij} label a vertex, and letting ij be the edge between vertices i and j. By symmetry of the matrix, a squared factor such as $a_{14}^2 = a_{14}a_{41}$ is a path there and back between vertices 1 and 4. Elements a_{ii} on the diagonal correspond to a loop at vertex i, and ${a_{ii}}^{2m}$ is represented by m loops at vertex i. The graphs for $a_{11}\, a_{12}\, a_{13}\, a_{14}^2\, a_{22}\, a_{24}\, a_{34}$, $a_{11}^2\, a_{12}\, a_{16}\, a_{23}\, a_{34}^2\, a_{36}$ and $a_{11}^4\, a_{12}^2\, a_{13}^2$ are shown in Figure 16. The odd index terms create cycles in the graph, all of which contribute zero to the expectation value of the trace. The terms which do contribute have double edges and/or loops. What is a remarkable property of all terms is that there seem to be no disjoint graphs, only connected ones. This happens because the numbers of indices i and j in each term are such that some of the i and/or j recur and overlap between factors. For instance, we do have 11 12 24 but not 11 22 24 or 11 23 45. I have not sought an analytic reason for why this is the case, but have verified it for the trace of A^8 where A is the general symmetric real 6×6 matrix. I wrote

a computer program to examine the indices of all factors in all 48,966 different varieties of terms, there being $6^8 = 1,679,616$ terms in all, though many share the same factors. There were no disjoint sets of indices.

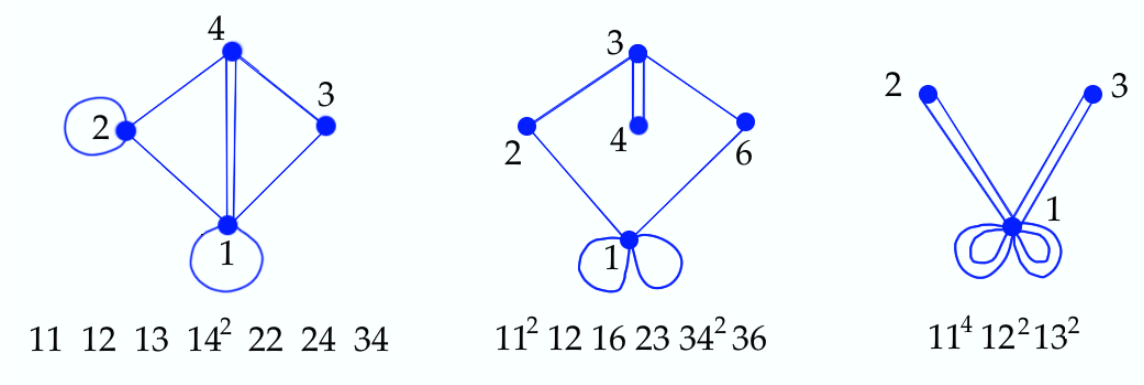

Figure 16: Graphs representing three typical terms in the trace of A^8 for A a 4×4 symmetric matrix.

Of the three graphs in Figure 16 only the third one contributes to the trace; the other two, with cycles, have zero expectation value. The rightmost graph in Figure 16 could also be seen as describing a path around a V-shaped tree – compare it with the path in red arrows round the tree on the right of Figure 13 – were it not for the four loops at the tree root. Recall that a vertex with multiple loops makes a heavily weighted contribution to the expectation value. In the approximation of the trace by the number of contributing terms, in §4.3, we weighted each square factor with 1 (instead of the true values, 3, 15, 105, etc., for higher powers). In a representation of this as a tree graph, this approximation is equivalent to converting each loop at a vertex into a leaf. This is done by adding a fictitious vertex half way round the loop, as shown in Figure 17 where four loops have been replaced by two there-and-back paths to fictitious vertices $1', 1''$. In this way each term becomes represented by a tree with k edges, and a path $2k$ long around these edges.

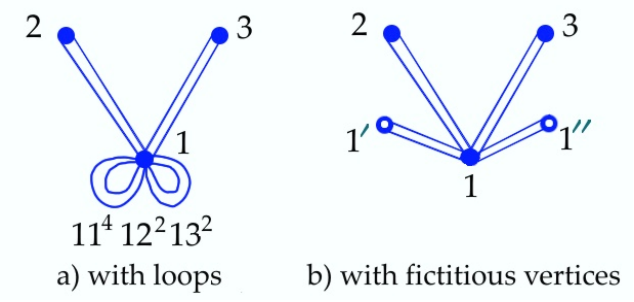

Figure 17: Removing loops by introducing fictitious vertices to produce a tree graph.

So every square factor a_{ij}^2 in the expectation value of $Tr(A_N^{2k})$ becomes represented by the there-and-back path round a single edge in a tree graph. There are C_N edges, $C_N + 1$ vertices in the tree representing each contributing term in the trace, thus making the total degree of the term $2k$. The approximation being made in the limit of large N is that the total number of terms is given by saying that each vertex can take any of N values independent of what happens at the other vertices. As an example, for $N = 3$, $2k = 4$ there are 2 edges, 3 vertices which can be arranged in $C_2 = 2$ configurations (see Figure 14, top panel). If each vertex can be assigned the index 1, 2 or 3 independently of the others, there will be $2 \times 3^3 = 54$ contributing terms. The true number is 45

because there are overlaps which cause some configurations such as a_{11}^4 to be counted more than once. As a further example, take the case of A_6^8 where $N = 6$, $k = 4$. There are now 5 vertices in the suite of $C_4 = 14$ tree graphs. These graphs can readily be drawn from the mountain range diagrams in Figure 14. The top left /VV\ shape in Figure 14 translates to the right tree in Figure 17. In this if each vertex is allowed to take any of the 6 indices, 1 to 6, there will be $5^6 = 15625$ terms contributing to the expectation values of the trace. Multiply this by the 14 configurations and the approximation predicts 218,750 contributing terms. Table 2 shows that this is a large over-estimate; the correct numbers is 82,896. However, we also understand that this over-counting partly compensates for neglecting the weighting of the higher powers such as a_{ij}^4. When the weightings for a Gaussian distribution are including, the expectation value of the trace is 175,080, from the list in Table 5. The 218,750 is 25% over this. Generalising, the approximation is $C_k N^{k+1}$ as found in §4.3, Table 5, as the limiting value of r_f.

The two approaches to proving Wigner's semicircle theorem, discussed in §4.1 and and §4.3 have now met in the middle so we can consider the theorem proved. In fact, the analysis in §4.3 gives insight into the trace of the matrix powers so that we see that, for finite N, approximating the moments by $C_k N^k$ is an over estimate for matrices whose elements come from both an uniform distribution and a Gaussian one. We might ask whether there exists a distribution of a_{ij} which fits closely to the $C_k N^k$ formula. It probably will be more peaked and narrow than the N[0, 1] Gaussian.

5 Electric charges spaced under Coulomb repulsion

The spacing of eigenvalues along the number line has been likened to the equilibrium spacing of negatively charged particles against a restraining force. The negative particles repel one another and would fly apart if they were not held by some force of attraction. The models used are 2-dimensional, and the general concept is illustrated in the left panel of Figure 18. It shows 13 lines of negative electric charge in a planar array, at positions from u_1 on the right to u_{13} on the left, with the central charge, number 7, at the origin. Several forms of constraining force can be postulated and I have examined two:

1. a positive electrostatic charge applied on parallel plates either side of negative array, as shown by the red strips in the right panel of Figure 18,

2. mechanical constraint through a set of springs, one spring attached to each line charge at one end and fixed to the origin at its other end.

The purely electrostatic model, 1, is easy to picture. Within the region occupied by the positive charge, the negative line charges space themselves almost equally, but outside this region the forces of constraint are less so the charges move further apart, just as do the eigenvalues in the semi-circle law limit.

5.1 The electrostatic analogue

The starting point for both these models is Coulomb's inverse square law for the force between two point charges. We will first derive the repulsive forces in the 2D model illustrated in the left panel of Figure 18. Here are a finite number of lines of negative charge in the $x-y$ plane, extending infinitely along the z axis and spaced apart along the x axis at positions u_j, $j = 1, ..., N$. The constraints we apply will ensure symmetry about the origin, making $u_N = -u_1$, $u_{N-1} = -u_2$, etc. The calculation is in four steps:

1. calculate the repulsive force between two isolated lines of negative charge,

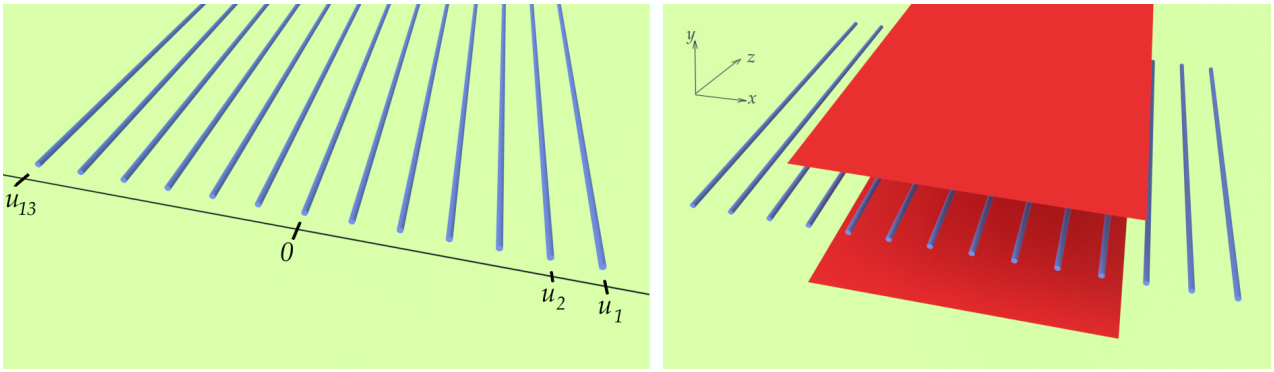

Figure 18: Left: Lines of negative change in the $x-z$ plane at x positions u_1 to $u_N, N = 13$. Right: the array lying between two sheets of positive charge at equal separation in y. All components extend to $z = \pm\infty$.

2. propose a restraining force. In the electrostatic model this is between one line of negative charge and the parallel strips of positive charge,

3. specify a number of line charges, add all components of force on each, and look for positions of the lines at which the net force on each is zero. This involves the solution of simultaneous non-linear equations.

4. increase the number N of line charges indefinitely, calculate their separations, look for trends, and compare with the statistics of samples of random matris.

Step 1: Inter-charge repulsive force The left panel of Figure 19 shows a 'point' negative charge q in three dimensions a distance h from a line of negative charge with charge density σ_1 per unit length. In the usual way of these calculations, consider the Coulomb repulsive force between q and the element of line charge at position z which is subtended by angle $\delta\alpha$ at q. If z is measured from the foot of the perpendicular from q, we have $\cos\alpha = h/r$, $\tan\alpha = z/h$, $\tan(\alpha+\delta\alpha) = (z+\delta z)/h$. From this

$$\delta z \;=\; \frac{h}{\cos^2\alpha}\,\delta\alpha\,.$$

Taking the electric permittivity constant to be 1, by Coulomb's law the force at separation r is

$$\delta f \;=\; q\,(\sigma_1\,\delta z)\,\frac{1}{r^2} \;=\; q\,\sigma_1\,\frac{h}{\cos^2\alpha}\,\frac{\cos^2\alpha}{h^2}\,\delta\alpha \;=\; \frac{q\,\sigma_1}{h}\,\delta\alpha$$

and acts along the direction of r. Since the line charge extends to infinity, the total force is the sum over all such elementary forces between $\alpha = \pm\frac{\pi}{2}$. The z components sum to zero by symmetry while the x components produce a net repulsion of

$$f_x \;=\; \frac{q\,\sigma_1}{h}\int_{-\pi/2}^{+\pi/2}\cos\alpha\,d\alpha \;=\; 2\,\frac{q\,\sigma_1}{h} \tag{19}$$

So the force falls off inversely proportional to the separation distance. We now want the force between two lines of charge, so consider q to be an element δz of a line of charge with density also σ_1 per unit length: $q = \sigma_1\delta z$. The magnitude of the force per unit length of one line upon the other is therefore $f_x \;=\; 2\,\sigma_1^2/h$. I use the sign convention that if a force pushes a line charge to the right it is reckoned as positive, and to the left, negative.

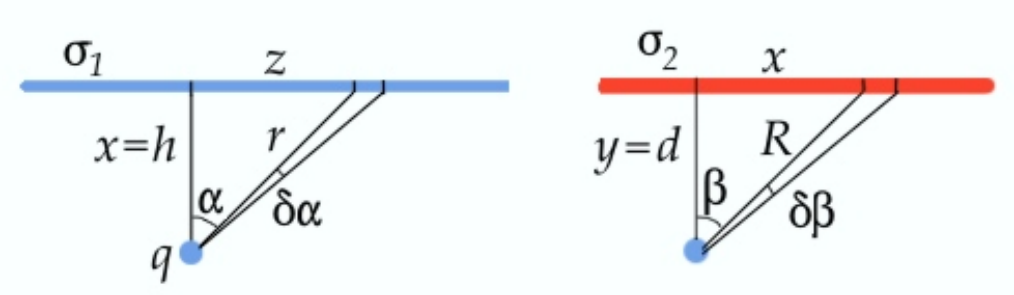

Figure 19: Left: negative point charge q at distance h from a line of negative charge along the z axis. Right: End view of line of negative charge distance d from a strip of positive charge (red).

When there are N line charges, the total repulsive force on the charge at u_j is

$$2\,\sigma_1^2 \sum_{k,k\neq j}^{N} \frac{1}{u_j - u_k}\,. \tag{20}$$

This takes account of the signs + and −; charges to the right of u_j push it to the left, while those to the left ($u_j > u_k$) push it to the right.

Note that this Coulomb model of line charges is referred to in the literature as the 'log-gas' model. This comes from the fact that the potential energy stored in a collection of line charges is equal to the energy required to bring each from infinity to be a distance h from a reference line charge at the origin. This is the integral of $2\,\sigma_1^2/h$ over h and equals $2\,\sigma_1^2 \ln h$. Thus the spacing of eigenvalues is being seen as determined by the energy stored in an array of charges, which is a sum over terms involving the logarithm of their separation.

Step 2: Constraining force. In the model where the constraining force is due to a spring attached at the origin, this force is simply $-S(u_j - L_j)$ where S is the spring constant and L_j the natural length. I discuss the choices of S and L_j in §5.2 below.

In the electrostatic model with the positively charged parallel plates, the formula is more complicated. The right panel of Figure 19 is viewed at a right angle to the left panel, with the z axis into the paper. We are looking end on at one infinite line of negative charges positioned d from a sheet of positive charge, which is also infinite into the paper. The charge density on the line is σ_1 per unit length and that on the positive strip is σ_2 per unit area. There is a similar sheet of positive charge at $y = -d$ to constrain the line charge to the $x - z$ plane. Similar to Step 1, $\delta x = d\,\delta\beta/\cos^2\beta$ and the x component of force on the line charge from the elements at x on the two sheets is

$$\delta F_x \;=\; 4\sigma_1\,(\sigma_2\,\delta x)\,\frac{1}{R}\,\sin\beta \;=\; 4\sigma_1\,\sigma_2\,\tan\beta\,\delta\beta\,.$$

The positive strips have finite width, from β_2 on the left to β_1 on the right, so the total x component of force is

$$F_x \;=\; 4\sigma_1\,\sigma_2[\ln(\cos\beta_2) - \ln(\cos\beta_1)] \;=\; 4\sigma_1\,\sigma_2 \ln\left(\frac{\cos\beta_2}{\cos\beta_1}\right).$$

The force attracts the line charge towards the mid-line of the two parallel strips. Note that if $|\beta_1| = |\beta_2|$ the force is zero; if $\beta_1 = \beta_2$ the strip has no width so cannot exert a force, and if $\beta_1 = -\beta_2$ the net force is zero because of symmetry. Only the excess length to one edge of the strip compared

with the other edge contributes to the x component of force. If x is measured from the foot of the perpendicular, and the left and right edges of each strip are at x_2 and x_1 respectively, this x–directed force can be written also as

$$F_x \;=\; 2\sigma_1\,\sigma_2 \ln\left(\frac{d^2+{x_1}^2}{d^2+{x_2}^2}\right). \qquad (21a)$$

In the right panel of Figure 19 $x = u = 0$ at the mid-line of the plates, the right hand end line charge is at $u_1 > 0$, the left at $-u_1 < 0$, and the edges of the plates at $u = \pm a$. The attractive force in the x-direction on the charge at u_j due to the two plates, per unit depth in z, is

$$2\sigma_1\,\sigma_2 \ln\left(\frac{d^2+(a-u_j)^{\,2}}{d^2+(a+u_j)^{\,2}}\right), \qquad u_j > 0\,.$$

This term, being attractive, should be negative in the sense that it moves the line charge towards the origin. For $u_j > 0$ this is indeed less than zero, hence the + sign. Since we require the whole system to have no net charge, $4a\,\sigma_2 = 2\sigma_1$. We can also normalise distances by taking $d = 1$, and so arrive at the attractive force of constraint

$$\frac{2N}{a}\,\ln\left(\frac{1+(a-u_j)^{\,2}}{1+(a+u_j)^{\,2}}\right), \qquad u_j > 0\,. \qquad (21b)$$

5.2 Equilibrium positions.

This is Step 3. The equilibrium positions are found by solving the $N\,div\,2$ simultaneous equations

$$2\,\sigma_1^2 \sum_{k,k\neq j}^{N} \frac{1}{u_j-u_k} \;-\; C(u_j) \;=\; 0, \qquad u_{N+1-k} = -u_k \qquad (22)$$

where $C(u_j)$ is the constraint force on the j^{th} line charge. The purpose of this analogy is to have the equilibrium positions correspond to the expectation values of the eigenvalues of an $N \times N$ symmetric random matrix with elements from N[0, 1], U[0, 1], etc. so the parameters of the constraint need to be chosen appropriately.

Regarding the spring model, I tried various values of S and L_J and quickly found that $S = 1$, $L_j = 0$ gives the best fit – indeed, in some cases it can be shown to give an exact fit. $L_j = 0$ means that each spring has zero natural length, so the displacement of a line charge from the origin to u_j means an extension of u_j. I will soon show that this spring model predicts *exactly* the expectation values of the characteristic polynomials of random matrices. There then seems little point exploring further the other model of constraint using positively charged plates. The latter, however, was the first model I investigated, and it is not without interest, so I give some results from it in the Appendix, §8.

Focusing therefore on the spring model, consider as an example the case of 5 charges, the central one at O. The two simultaneous equations for zero net force are

$$\frac{3}{u_1}+\frac{2}{u_1-u_2}+\frac{2}{u_1+u_2}-u_1 \;=\; 0\,,$$
$$\frac{3}{u_2}+\frac{2}{u_2-u_1}+\frac{2}{u_1+u_2}-u_2 \;=\; 0\,.$$

These are symmetric in u_1 and u_2. By multiplying out each we obtain from the numerators two quadratics in $p = u_1^2$, $q = u_2^2$:

$$u_1^4-7u_1^2-u_1^2u_2^2+3u_2^2 \;=\; 0, \qquad u_2^4-7u_2^2-u_1^2u_2^2+3u_1^2 \;=\; 0\,.$$

$$p^2 - 7p - pq + 3q \;=\; 0, \qquad\qquad q^2 - 7q - pq + 3p \;=\; 0\,.$$

From the first $q = (p^2 - 7p)/(p - 3)$ which, when substituted into the second, gives

$$\frac{8p(p^2 - 10p + 15)}{p^2 - 6p + 9} \;=\; 0\,.$$

The numerator contains the defining polynomial $u_1^5 - 10u_1^3 + 15u_1$ which is exactly the characteristic polynomial for a $N = 5$ random symmetric matrix from N[0, 1] as determined in §2.2, Eq 4. In other words, for $N = 5$ at least, this spring model is an exact analogy.

For $N = 11$, the largest random matrix for which I could directly compute the mean characteristic equation (§2.2), the spring analogue gives roots from which the characteristic equation can be calculated as

$$(\lambda^2 - 5.188^2)(\lambda^2 - 3.9362^2)(\lambda^2 - 2.8651^2)(\lambda^2 - 1.8760^2)(\lambda^2 - 0.9289^2)\lambda \;=\; 0\,,$$

which evaluates to machine accuracy to

$$N = 11: \quad \lambda^{11} - 55\lambda^9 + 990\lambda^7 - 6930\lambda^5 + 17325\lambda^3 - 10395\lambda\,,$$

precisely as determined directly from the matrix by averaging products of elements. This holds for all the other characteristic polynomials at Eq 4, §2.2 and gives strong evidence that this model will hold for all N. The next section reveals the true nature of these polynomials.

6 Matrix mean characteristic polynomials revisited

The process of solving the simultaneous equations and creating the polynomial of which they are the roots allows us to extend the sequence of mean characteristic polynomials above 11. I find

$$N = 12: \qquad \lambda^{12} - 66\lambda^{10} + 1485\lambda^8 - 13860\lambda^6 + 51975\lambda^4 - 62370\lambda^2 + 10395\,,$$

$$N = 13: \quad \lambda^{13} - 78\lambda^{11} + 2145\lambda^9 - 25740\lambda^7 + 135135\lambda^5 - 270270\lambda^3 + 135135\lambda\,.$$

Reference to the On-Line Encyclopaedia of Integer Sequences (*www.oeis.org*, sequences A001498, A001498) shows that the coefficients are the so-called Bessel numbers. These are less famous than the Bessel functions, the renowned functions associated with cylindrical symmetry. The general formula for the mean characteristic polynomials is

$$\boxed{P_N(\lambda) \;=\; \sum_{k=0}^{N} C_{k,N}\,\lambda^{N-2k}\,, \qquad C_{k,N} \;=\; \frac{N!}{(-2)^k\,k!\,(N-2k)!}\,, \qquad 2k \le N} \tag{23}$$

This is one of the main results of this article. It summarises all the polynomials in §2.2 and above. Using it we can in principle solve numerically for the mean characteristic polynomial for any N, and so determine the expectation values of all eigenvalues of any symmetric random matrix with elements drawn from a population having mean 0, variance 1.

Naturally I wondered whether these polynomials $P_N(\lambda)$ satisfy a differential equation. With some algebraic over-kill, and setting coefficients to zero, I find that they do satisfy the rather general fourth order equation

$$A\frac{d^4P_N}{d\lambda^4} \;+\; B\frac{d^3P_N}{d\lambda^3} \;+\; C\frac{d^2P_N}{d\lambda^2} \;+\; D\frac{dP_N}{d\lambda} \;+\; EP_N \;=\; 0$$

where

$$\begin{aligned} A &= c\lambda + d \\ B &= c(N-2-\lambda^2) + g\lambda - m - r \\ C &= -(d+g)\lambda^2 + m\lambda + (2N-3)d + (N-1)g - s \\ D &= r\lambda^2 + s\lambda + (N-2)(N-1)c - (N-1)(m+r) \\ E &= -Nr\lambda + N(N-1)(d+g) - Ns \end{aligned}$$

where c, d, ... s are arbitrary parameters. Setting $c = d = g = 0$, $r = -m$ gives the second order equation

$$\frac{d^2P_N}{d\lambda^2} - \lambda\frac{dP_N}{d\lambda} + N\,P_N = 0 \tag{24}$$

which I take to be the lowest order differential equation which these polynomials satisfy. If λ were to be replaced in Eq ?? by 2λ and N by $2N$, this would be the standard form for Hermite's equation. So the characteristic polynomials of §2.2 are cousins of the familiar Hermite polynomials. Indeed, for all N

$$\boxed{P_N(\lambda) = \frac{1}{2^{N/2}}\mathcal{H}_N\left(\frac{\lambda}{\sqrt{2}}\right)} \tag{25}$$

where $\mathcal{H}_N(x)$ are the Hermite polynomials. This is a remarkably simple result. Since the Hermite polynomials are readily available in computer software packages, their zeros can be found and the development of the mean eigenvalues explored as N become large. Recall that Hermite polynomials are the eigenvalues of the quantum mechanical treatment of the harmonic oscillator. Perhaps therefore they are not unexpected in our context because a linear spring force is equivalent to a quadratic potential, as assumed in the harmonic oscillator.

6.1 Statistical comparison with random matrices

It is important to bear in mind that the Hermite-like polynomials derived from the Coulomb charge-linear spring analogy strictly only give the equilibrium positions of the line charges in this model. The only evidence that the zeros of these $P_N(\lambda)$ also give the mean eigenvalues of random matrices with mean 0, variance 1 is the fact, presented in §2.2, that they do agree with the mean characteristic polynomials for random matrices up to $N \leq 11$, 11 being the largest square matrix for which my computer could calculate the characteristic polynomial. I have not proved this correspondence in any deeper sense. For this reason I consider it necessary to make at least some limited direct comparison between the charge equilibrium positions and the mean values of eigenvalues from several samples of matrices. Such is the spread in eigenvalues that one would need hundreds of samples to obtain close statistics. I offer only a limited comparison for $N = 21$ and 45.

I constructed 100 $N = 21$ symmetric matrices with all elements (including the diagonal ones) drawn from the Gaussian N[0, 1] population, and sorted their eigenvalues into ascending order. The negative ones were paired with the corresponding positive ones by order since these should have the same magnitude on average, and hence I calculated their aggregated mean absolute values and standard deviations. Thus 100 matrices furnish 200 instances of each eigenvalue. In Table 7 the results are compared i) with the roots of the characteristics polynomial and ii) the equilibrium positions u in the fully Coulomb electrostatic analogy (see Appendix). The roots of the polynomial from the spring model compare quite well with the experimental means, though the standard deviations are relatively large. The alternative model with electrostatic retaining force consistently under-estimates the average eigenvalues. It became clear at this stage that this alternative model fails by increasing amounts as N increases.

Expectation eigenvalue	mean of 200 eigenvalues	standard deviation	u	% difference
$7 \cdot 85$	$8 \cdot 17$	$0 \cdot 77$	$7 \cdot 85$	4%
$6 \cdot 75$	$6 \cdot 87$	$0 \cdot 59$	$6 \cdot 02$	14%
$5 \cdot 83$	$5 \cdot 90$	$0 \cdot 53$	$5 \cdot 19$	14%
$4 \cdot 99$	$4 \cdot 98$	$0 \cdot 48$	$4 \cdot 47$	11%
$4 \cdot 21$	$4 \cdot 20$	$0 \cdot 46$	$3 \cdot 81$	10%
$3 \cdot 47$	$3 \cdot 42$	$0 \cdot 42$	$3 \cdot 15$	8%
$2 \cdot 75$	$2 \cdot 69$	$0 \cdot 42$	$2 \cdot 52$	7%
$2 \cdot 05$	$2 \cdot 01$	$0 \cdot 45$	$1 \cdot 88$	7%
$1 \cdot 36$	$1 \cdot 34$	$0 \cdot 43$	$1 \cdot 25$	7%
$0 \cdot 68$	$0 \cdot 65$	$0 \cdot 39$	$0 \cdot 63$	4%
0	$0 \cdot 03$	$0 \cdot 42$	0	0

Table 7: N = 21. Comparison of expectation values of eigenvalue (column 1: from roots of characteristic polynomial) with mean experimental values (columns 2, 3) from sample of 100 N = 21 symmetric matrices from N[0, 1]. Columns 4, 5 shows the equivalent results from the alternative electrostatic model and its percentage difference from the mean – see Appendix.

Table 8 presents an equivalent comparison between the roots of the spring model polynomial and the average of 60 samples of each eigenvalue obtained from 30 N = 45 random matrices. The agreement again gives confidence in the spring model, though the standard deviations are still typically 5 to 10 percent of the mean. Clearly one could go on making such statistical comparisons, but I accept at this stage that the Coulomb gas model with springs providing the constraint forces does correctly predict the expectation values of the eigenvalues.

eigenvalue	mean	std. devn.	eigenvalue	mean	std. devn.
$12 \cdot 24$	$12 \cdot 58$	$0 \cdot 60$	$4 \cdot 76$	$4 \cdot 76$	$0 \cdot 28$
$11 \cdot 29$	$11 \cdot 42$	$0 \cdot 57$	$4 \cdot 26$	$4 \cdot 27$	$0 \cdot 33$
$10 \cdot 50$	$10 \cdot 54$	$0 \cdot 47$	$3 \cdot 78$	$3 \cdot 78$	$0 \cdot 35$
$9 \cdot 79$	$9 \cdot 89$	$0 \cdot 42$	$3 \cdot 29$	$3 \cdot 32$	$0 \cdot 29$
$9 \cdot 14$	$9 \cdot 21$	$0 \cdot 43$	$2 \cdot 81$	$2 \cdot 87$	$0 \cdot 32$
$8 \cdot 52$	$8 \cdot 60$	$0 \cdot 36$	$2 \cdot 34$	$2 \cdot 33$	$0 \cdot 36$
$7 \cdot 94$	$7 \cdot 99$	$0 \cdot 39$	$1 \cdot 87$	$1 \cdot 84$	$0 \cdot 33$
$7 \cdot 37$	$7 \cdot 44$	$0 \cdot 36$	$1 \cdot 40$	$1 \cdot 42$	$0 \cdot 33$
$6 \cdot 82$	$6 \cdot 91$	$0 \cdot 36$	$0 \cdot 93$	$0 \cdot 92$	$0 \cdot 35$
$6 \cdot 29$	$6 \cdot 37$	$0 \cdot 33$	$0 \cdot 47$	$0 \cdot 53$	$0 \cdot 34$
$5 \cdot 77$	$5 \cdot 82$	$0 \cdot 36$	$0 \cdot 00$	$0 \cdot 00$	$0 \cdot 35$
$5 \cdot 26$	$5 \cdot 32$	$0 \cdot 34$			

Table 8: N = 45 : Comparison of roots of characteristic polynomial with the mean of 60 sets of all eigenvalues from 30 random matrices.

In the 1990s Craig Tracy and Harold Widom from the University of California published an analysis of the distribution function of the largest eigenvalues[7]. The so-called Tracy-Widom probability distribution is a skewed bell-shaped function with a negative mean and a longer tail on

[7]Commun. Math. Phys. 177, 727-754 (1996), ICM 2002 Vol III-1-3 ,

the high side[8]. Figure 20 is a plot of the Tracy-Widom probability density for $\beta = 1$ appropriate to real values random matrices, obtained from data in a paper by Andrei Bejan (on the internet). Note that the mean is at $-1 \cdot 27$ and the mode at $-1 \cdot 35$. The derivation of this distribution involves Airy functions, these being Bessel functions of order 1/3, and the exponent 1/3 or 1/6 features in its properties. For instance, the width across the central body of the probability density curve has been shown to scale as $1/N^{1/6}$, which evaluates to $0 \cdot 53$ for $N = 21$ and $0 \cdot 60$ for $N = 45$. The distribution has achieved a degree of fame because it has been shown to be another statistical phenomenon which has universality; it applies to many situations where the details of the underlying sub-processes do not have a strong affect on the combined outcome.

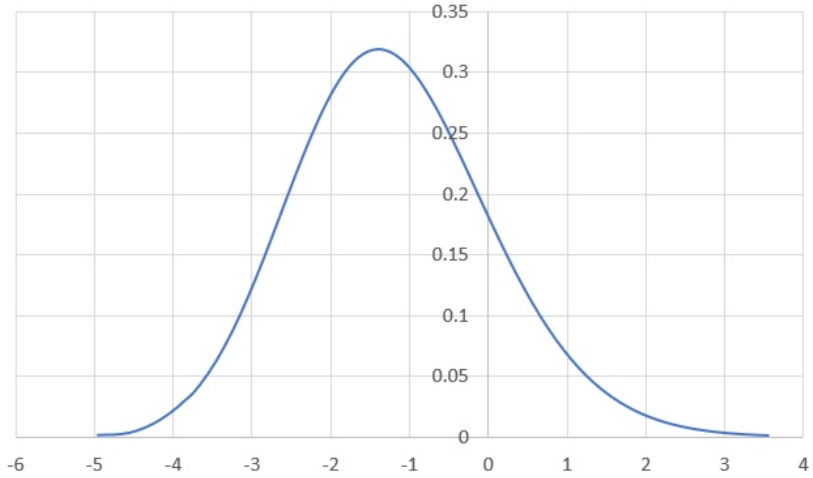

Figure 20: Probability distribution of the Tracy-Wisdom distribution for $\beta = 1$, for the largest eigenvalues of real symmetric random matrices.

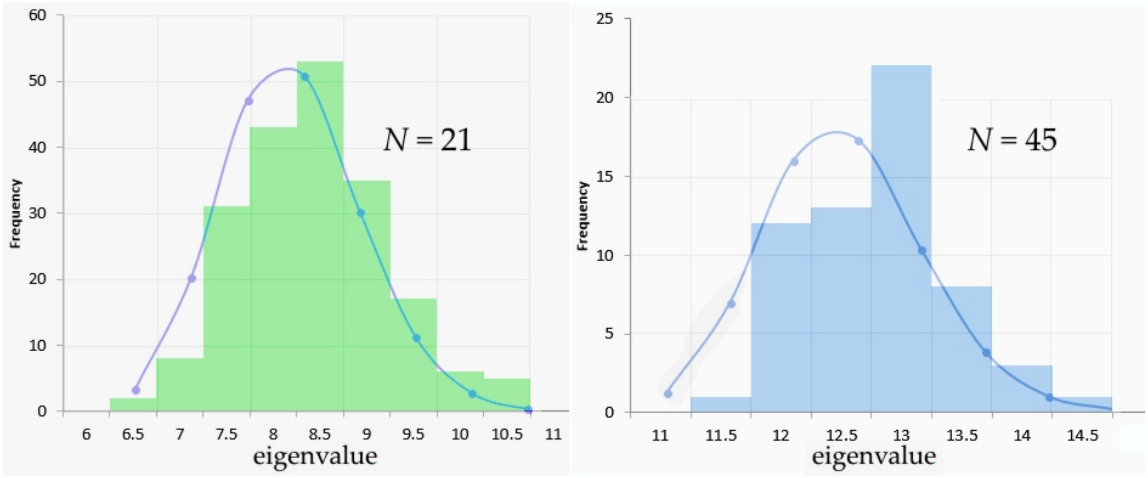

Figure 21: Histograms of $|\lambda_{max}|$ for 200 eigenvalues for $N = 21$, and 60 for $N = 45$. The over-laid curves are the Tracy-Widom distribution for that value of N.

[8] The T-W distribution has three versions which strictly apply to the three types of symmetry identified by Dyson and labelled $\beta = 1$, 2 or 4. $\beta = 1$ corresponds to real symmetric matrices, but ones which have twice the variance of diagonal elements than the matrices which I have used.

The two panels of Figure 21 present histograms of $|\lambda_{max}|$ for the 200 samples at $N = 21$ and 60 at $N = 45$ respectively. Over-laid on these are the corresponding Tracy-Widom probability density curves, suitably scaled. The agreement is fair.

6.2 Trends in mean eigenvalues as $N \to \infty$

I have solved 20 polynomials, Eq 25, for N up to 500 to obtain the line charge equilibrium positions and hence mean eigenvalues. Wigner's semi-circle law is well illustrated in Figure 22 for $N = 500$. This plots the mean gaps δ between adjacent eigenvalues as a function of their mid-value. Also shown is the reciprocal, $1/\delta$, which is proportional to the density of equilibrium positions (*i.e.* eigenvalues) along the number line. Compare with Figures 7 and 8 for one $N = 2435$ matrix.

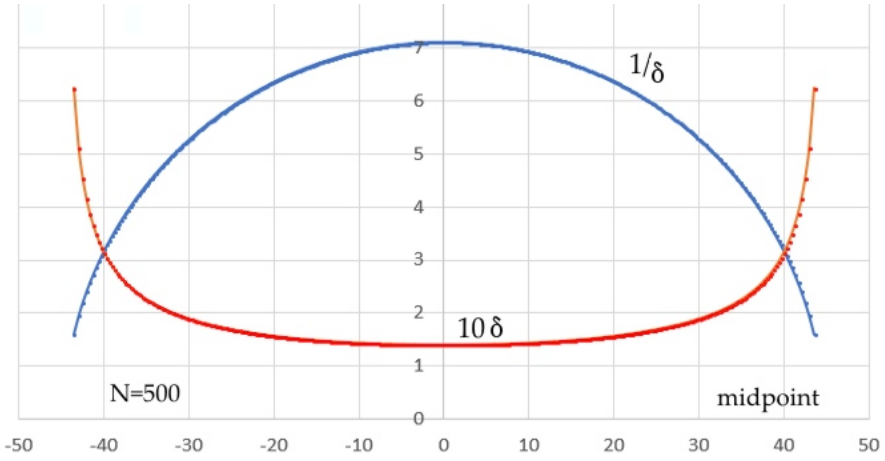

Figure 22: Wigner's semi-circle law as expressed by $1/\delta$ for $N = 500$ (blue curve). The red curve plots the gap, δ, between adjacent eigenvalues (scaled up by 10) against mid-point between adjacent eigenvalues.

The exact mean eigenvalues means that an improved approximation to $|\lambda_{max}|$ as a function of N can be constructed. A modified version of Eq 6, §2.2., is

$$|\lambda_{max}| \approx \beta \frac{N-1}{\sqrt{N+2}}, \qquad \beta = 1 \cdot 81952\, N^{0 \cdot 0131495} \tag{26}$$

which has error typically less than $0 \cdot 05$ for $N < 500$. Note the $N + 2$ in the denominator, compared with $N + 1$ before. For the $N = 2435$ matrix presented in §2.2 and 2.3 this predicts an expectation value of $99 \cdot 40$, very close to the observed values of $99 \cdot 378$ and -98.996. There is another form of approximation in the literature which maintains a close link with the limit of the Wigner semi-circle law by approximating the deficit from the limiting value $2\sqrt{N}$. The form is

$$2\sqrt{N} - |\lambda_{max}| \approx \frac{\mu}{N^{1/\nu}} \tag{27}$$

where $|\lambda_{max}|$ means the expectation value. Fitting this to my 20 data points up to $N = 500$, $\mu = 2 \cdot 07$, $\nu = 6 \cdot 66$. Indeed, this is a marginally better fit than the revised expression involving β above. The power of 1/6 has resonances with the Tracy-Widom law, where a cube power is involved in the derivation. Also I point out below that the ratio of largest to smallest gap seems to be a cubic function. If ν is constrained to be precisely 6, the best fit value of μ is $\mu = 2 \cdot 20$. This should be compared with the mean of the Tracy-Widom distribution, which is $1 \cdot 21$. There is a small possibility that the largest experimental value of $2 \cdot 2$ may be closer to the correct value for the matrices I have

used, which have a smaller variance of the diagonal elements than the GOE matrices for which the Wigner semi-circle and Tracy-Widom distributions were derived (1 compared with $\sqrt{2}$).

Another set of interesting quantities is the gaps between the largest pair and smallest pair of eigenvalues for any given N. These decrease as N increases, rapidly for N small, much more gradually for N large. Call the gap $\lambda_{max} - \lambda_2 = \delta_1$, the gap $\lambda_2 - \lambda_3 = \delta_2$, and call the smallest gap, at $\lambda = 0$, λ_s. I find that all these gaps and the ratios between them can be fairly well represented by functions of the form bN^p. Table 9 below gives the values of b and p. The ratio in the bottom row means that the largest gap grows relative to the smallest roughly as $0 \cdot 6 \sqrt[3]{N}$ – another appearance of 1/3 power. These growing gaps between adjacent eigenvalues at the two extremes is the mechanism by which the semi-circle distribution develops.

quantity	b	p
δ_1	$1 \cdot 97636$	$-1/5 \cdot 3$
δ_2	$1 \cdot 77227$	$-0 \cdot 20588$
δ_s	$2 \cdot 98029$	$-0 \cdot 48864$
δ_1/δ_s	$0 \cdot 62255$	$+0 \cdot 31498$

Table 9: Parameters b and p in the functions bN^p which approximate the two largest gaps, δ_1 and δ_2, the smallest, δ_s and also the ratio δ_1/δ_s.

7 Concluding remarks

This has been a amateur's attempt to explore some basic aspects of random matrices. The literature shows a vast mountain of sophisticated mathematics and applications built on foundations like these. It has been a surprise to find that the average values of eigenvalues of a wide class of symmetric matrices are numerically identical with the equilibrium positions of negative electric line charges constrained by springs about a central position. The matrices for which this is true all have elements drawn from populations with a) mean = 0, b) variance = 1, c) all higher moments being finite. Recall that where the statistics do not meet these requirements, as with the continued fraction distribution of §3.3, the eigenvalues do not match the equilibrium positions.

In the physical analogues the equilibrium positions of line charges are determined by the balance of mutual repulsion according to Coulomb's law, whose force varies as 1/distance (in 2 dimensions), and the restraining spring force which varies directly as distance. What are the corresponding forces on the eigenvalues of random matrices? For instance, ould these be equivalent to the mean and the variance which are the only parameters of the matrix elements. Consider that the variance describes how separated are the matrix elements – it measures what is 'pushing' the elements apart. The mean, in contrast, is a anchor holding the elements to a central position. Moreover the variance is a quadratic quantity, similar to the retaining electric potential in the Coulomb gas model. Of course, the link between matrix elements and eigenvalues is a convoluted one, so 'forces' between elements cannot map readily to 'forces' between eigenvalues, though the mean value of 0 is the same for each. I suspect – but have not examined the literature on this – that the spread in values of the eigenvalues about their respective mean positions can be likened to the spread in displacements about their equilibrium positions of line charges under thermal agitation.

Can this be pushed to speculation as to whether there are 'forces' between the complex zeroes of the Riemann zeta functions, against a background of constraint? One obvious difference

from matrices is that there is a sequence of matrices of increasing size N, from which the study of smaller matrices can enlighten ideas about how N tends to infinity. In contrast I know of no N-indexed sequence of entities whose limit is the zeta function, which has an infinity of zeros. Perhaps zeta-like functions over finite fields of increasing order have a role here? I must leave this to the professionals.

John Coffey, Cheshire, England, 2018

8 Appendix: Electrostatic constraint in Coulomb line charge model

I explained in the main text that the spring model of constraint furnishes an exact analogy with the mean eigenvalues of a random matrix, and that the electrostatic model of constraint does not hold as N increases. It does, however, give decent approximations to the eigenvalues for small N, and since my approach has been to start with N small and build on that, this model initially seemed promising. This Appendix therefore examines what would be Steps 3 and 4 of the equilibrium of Coulomb forces model analysed in §5.1.

8.1 Two line charges in equilibrium

The formula for the x component of force between a line charge and the pair parallel plates is given in §5.1, Eq 21a, b. Taking things in small steps, I first introduce only 2 line charges between the two positive sheets of charge and look for their equilibrium positions. Clearly these will be symmetrically placed with respect to the plates, which each have half-width a. The x-position must now be reckoned from the mid-line of the plates. The only variable here is u_1 so the equilibrium positions are at $\pm u$ where the attractive and repulsive forces have equal magnitudes. So we look for roots u_1 of

$$\frac{a}{u_1} + \ln\left(\frac{1+(a-u_1)^2}{1+(a+u_1)^2}\right) \;=\; 0\,.$$

This is intractable analytically, so I have found numerical solutions for a range of the parameter a. The two panels in Figure 23 plot the equilibrium position of the right hand line charge as a function of a. The left panel is where the positive plates are narrow and the negative charges may protrude beyond the ends of the positive strips. When $a = \frac{1}{2}\sqrt{e-1} = 0\cdot655$, $u_1 = a$ meaning that the line charges at precisely at the edges of the strips. When the strips are reduced to a line's width, $u_1 = 1/\sqrt{3} = 0\cdot577$. In this limit the positive and negative charges are arranged at the vertices of two touching equilateral triangles, which is a close-packed configuration as one might expect. In this left graph the fitted curve is a fourth order polynomial but the coefficients of a^4 and a^3 are small, so the curve is close to being a parabola. The right hand graph is for where the parallel strips are wider and the line charges are always well inside them. Here the equilibrium positions are a linear function of strip width.

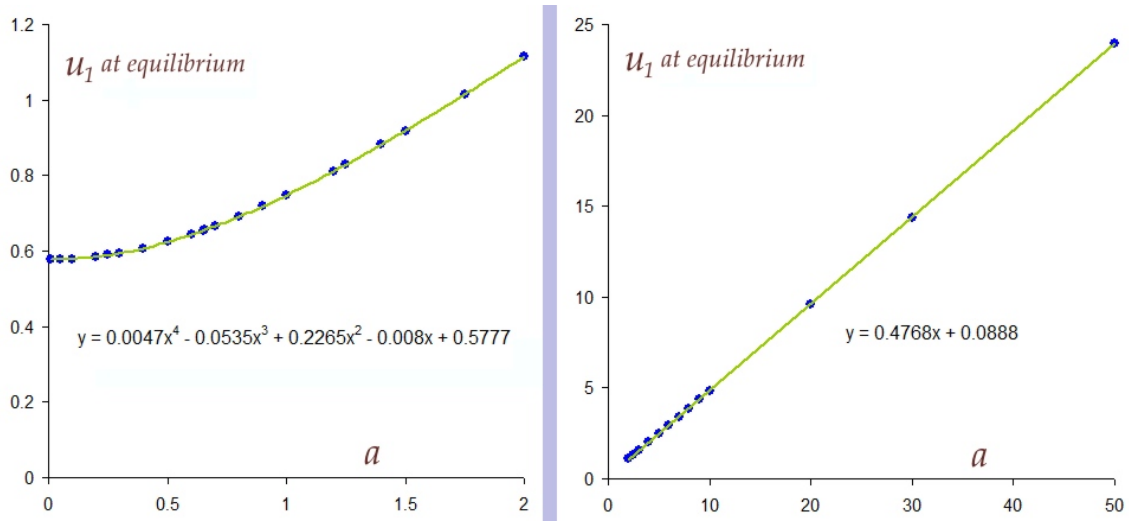

Figure 23: Equilibrium positions $\pm u_1$ of two symmetrically placed lines of negative charge as a function of the half-width a of the two parallel strips of positive charge, 2 units apart.

As an aside, I checked the above theory against a finite element model to gain confidence that it was consistent. I used the free Ansoft Corporation program MaxwellSV, 2002, and set up a model in the $x-y$ plane with two symmetrically placed line charges each of 1×10^{-5} Coulombs per unit depth in z, with the plate semi-spacing d set at 1 mm. The force in newtons on one of the lines charges was calculated, and Figure 24 is a plot of the x components of force for $a = 1\cdot25$ mm. In the FEM calculation the y component of force proved a good way to monitor the quality of the convergence since theoretically it should be zero. No arbitrary scaling factors have been introduced in plotting these results – the scales are absolute. As Figure 24 makes plain, the agreement is very good and shows that the net x component of force is zero when $u_1 = 0\cdot930$ mm.

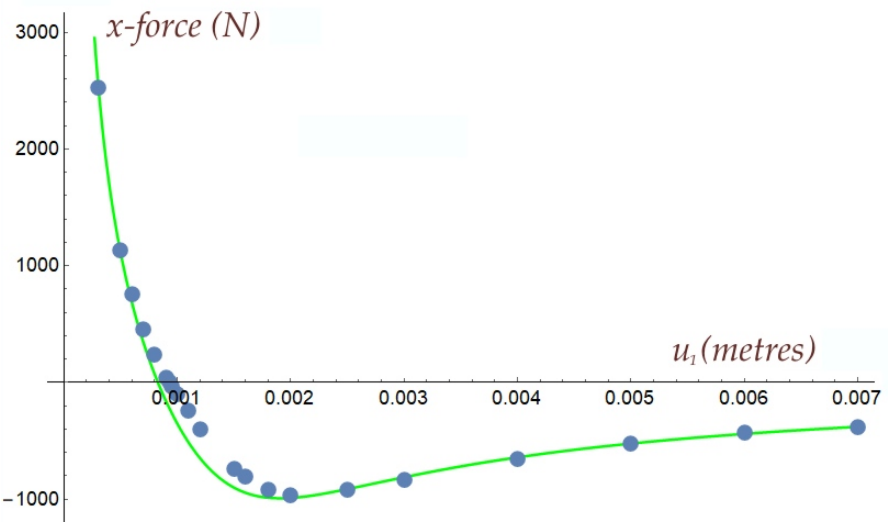

Figure 24: The net force on one line charge from the combined effect of a second line charge and the two parallel strips. Comparison of theory (green line) with values calculated by a finite element model for the case $a = 1\cdot25$ mm $= 0\cdot00125$ m.

8.2 Equilibrium spacings of several line charges

More line charges are now added and we seek a method to determine the equilibrium positions as N, the number of lines, increases to infinity. When N is odd, the middle line charge will position itself centrally at $x = 0$ with the others symmetrically positioned to either side at $\pm u_1, \pm u_2$, *etc.*, $u_1 > u_2 > u_3 > ...$ For $N = 2n$ or $2n+1$ there are n simultaneous equations to solve, and for small n this can be done with a multi-dimensional version of Newton's method. I have found that Mathematica 10 copes well. Overall charge neutrality requires that $N\sigma_1 = 4a\sigma_2$. If we take $\sigma_1 = 1$, then $\sigma_2 = N/(4a)$. As an example, here are the two simultaneous equations for $N = 5$:

$$\frac{2}{u_1-u_2}+\frac{2}{u_1}+\frac{2}{u_1+u_2}+\frac{2}{2u_1}+\frac{5}{4a}.2\sigma_1\sigma_2\ln\left(\frac{d^2+(a-u_1)^2}{d^2+(a+u_1)^2}\right) = 0,$$

$$-\frac{2}{u_1-u_2}+\frac{2}{u_2}+\frac{2}{u_1+u_2}+\frac{2}{2u_2}+\frac{5}{4a}.2\sigma_1\sigma_2\ln\left(\frac{d^2+(a-u_2)^2}{d^2+(a+u_2)^2}\right) = 0.$$

I have set up and solved numerically similar equations for the number of line charges ranging from 2 to 21 for various values of a. The positions of the line charges vary widely with a, obliging us to choose one value of a at which to make the comparison. The need to make this selection is a

distinct weakness of this model. I choose $a = a_{ev}$, the value of a at which $u_1 = \lambda_{max}$. This has found for each N by numerical searching. Table 10 lists the value of a_{ev} and the corresponding equilibrium positions of the line charges for N up to 11. These compare quite well with the mean eigenvalues.

N	a_{ev}		1	2	3	4	5	6
2	1.71	eigenvalues	1					
		positions	1					
3	2.24	eigenvalues	1.732	0				
		positions	1.732	0				
4	2.69	eigenvalues	2.334	0.742				
		positions	2.332	0.745				
5	3.09	eigenvalues	2.857	1.356	0			
		positions	2.856	1.362	0			
6	3.45	eigenvalues	3.324	1.889	0.617			
		positions	3.324	1.896	0.625			
7	3.78	eigenvalues	3.750	2.367	1.154	0		
		positions	3.752	2.371	1.170	0		
8	4.08	eigenvalues	4.145	2.802	1.637	0.539		
		positions	4.146	2.798	1.654	0.549		
9	4.37	eigenvalues	4.513	3.205	2.077	1.023	0	
		positions	4.523	3.195	2.097	1.042	0	
10	4.60	eigenvalues	4.859	3.582	2.484	1.466	0.485	
		positions	4.851	3.538	2.486	1.481	0.492	
11	4.85	eigenvalues	5.188	3.936	2.865	1.876	0.929	0
		positions	5.188	3.878	2.858	1.891	0.942	0

Table 10: Expectation values of eigenvalues of $N \times N$ symmetric random matrices with mean 0, variance 1 compared with the equilibrium positions of N negative line charges between two positively charged plates of half-width a_{ev} and separation 2 units. The ± signs are implied.

The agreement, of course, does hang on our seeding the value of $|\lambda_{max}|$, and the only guide to this is an approximation such as Eq 26 or 27 obtained by extrapolating from small values of N. As an illustration, consider $N = 21$. I used Mathematica 10 to solve the 10 simultaneous equations for the charges. To get started take λ_{max} from the formula concocted for β at Eqs 6, 26 which is $7 \cdot 847$. We do not know the value of a_{ev} which will make $u_1 = 7 \cdot 847$, so an initial guess is a value somewhat less than this. After a little iteration I have found that $a_{ev} = 6 \cdot 147$ and the equilibrium positions are

$$u_1 = 7 \cdot 847, \quad u_2 = 6 \cdot 023, \quad u_3 = 5 \cdot 185, \quad u_4 = 4 \cdot 475, \quad u_5 = 3 \cdot 805,$$

$$u_6 = 3 \cdot 155, \quad u_7 = 2 \cdot 515, \quad u_8 = 1 \cdot 882, \quad u_9 = 1 \cdot 253, \quad u_{10} = 0 \cdot 626, \quad u_{11} = 0.$$

The results are compared with the equilibrium positions u in the Coulomb electrostatic analogy in the rightmost columns of Table 7 in §6.1 of the main article. It is already clear that for all eigenvalues this fully electrostatic model under-estimates the value.

Having understood the success of the spring model of constraining force, it is easy to see why the electrostatic constrain model works for N small. By expanding the logarithmic force term as a Taylor series about $u = 0$ we find that it is approximately

$$-\frac{2N}{a^2+1}\ u\ -\ \frac{2N(a^2-3)}{3(a^2+1)^3}\ u^3\,.$$

Where the second term contributes little there is an almost linear relation between force and position u, just as in the spring model. Indeed, equating the first term with the spring force $-u$ gives the critical value of a as

$$a_{ev}\ \approx\ \sqrt{2N-1}\,. \tag{28}$$

This agrees well with the values found in the numerical experiments. For large N, the ratio $a_{ev}/|\lambda_{max}|$ would tend to $1/\sqrt{2}$, but the model has broken down long before then.