BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 40, Number 2, Pages 155–178
S 0273-0979(03)00975-3
Article electronically published on February 12, 2003

PATTERNS IN EIGENVALUES: THE 70TH JOSIAH WILLARD GIBBS LECTURE

PERSI DIACONIS

Abstract. Typical large unitary matrices show remarkable patterns in their eigenvalue distribution. These same patterns appear in telephone encryption, the zeros of Riemann's zeta function, a variety of physics problems, and in the study of Toeplitz operators. This paper surveys these applications and what is currently known about the patterns.

Introduction

This paper surveys what we know about the distribution of the eigenvalues of typical large unitary matrices. The topic occurs naturally in problems of statistics, physics and number theory. The mathematical interconnections are also fascinating, and it is hard to escape the feeling that there is something unseen to be discovered.

To keep the paper within bounds, the following classical compact groups will be featured:

O_n the $n \times n$ real matrices M such that $MM^T = id$.
U_n the $n \times n$ complex matrices M such that $MM^* = id$.
S_n the $n \times n$ permutation matrices.

"Typical elements" are studied by using Haar measure. This is a probability measure P on a group G which is translation invariant: for any measurable set A in G and any element M in G

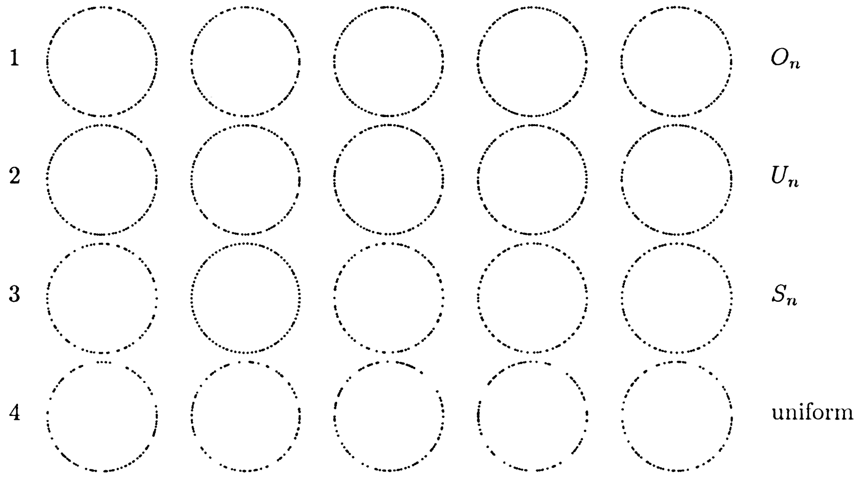
$$P(A) = P(MA).$$

For the symmetric group, we all have an intuitive feel for what it means to pick a permutation at random, at least via shuffling cards. For the other groups this is less familiar. The following method for picking a group element at random may help.

Consider the orthogonal group O_n. Here is a simple algorithm: fill out an empty $n \times n$ array with independent picks from the standard Gaussian bell-shaped curve. Then perform the Gram-Schmidt algorithm on this array: normalize the first row to have norm one, take the first row out of the second row and normalize to have norm one, and so on. This constructs an orthogonal matrix which is Haar distributed. Put more formally, put product measure on $\mathbb{R}^{n^2}$ with each factor having density $e^{-x^2/2}/\sqrt{2\pi}$. The Gram-Schmidt algorithm gives a map T from almost all of $\mathbb{R}^{n^2}$ onto O_n. The image of the product measure is Haar measure.

Received by the editors October 10, 2002.

2000 *Mathematics Subject Classification.* Primary 00-02, 60B15.



Figures 1-4: Five realizations for $n = 100$ from Haar measure on O_n, U_n, S_n, and independent uniform points.

This is easy to prove and understand. Each row of the original array has density proportional to $e^{-\frac{1}{2}\sum x_i^2} dx_1 \dots dx_n$. This measure on $\mathbb{R}^n$ is invariant under orthogonal transformations. By inspection, $T(MX) = MT(X)$ where X is the original array. Hence, $P(MT(X) \in A) = P(T(MX) \in A) = P(T(X) \in A)$. A more analytical proof can be found in Eaton [40]. Perhaps more convenient: if an $n \times n$ matrix of independent Gaussian entries is input to the QR algorithm, the resulting orthogonal piece is Haar distributed. See [32] for a survey of constructions of Haar measure.

The same construction works for the unitary group U_n using the usual complex inner product. Here the original entries of the array are chosen as independent, standard, complex Gaussian variables with density $e^{-|z|^2}/\pi$ on C.

For any of the groups O_n, U_n, S_n in its standard matrix representation, any element is diagonalizable with all eigenvalues on the unit circle. Call these $\{e^{i\theta_1}, \dots, e^{i\theta_n}\}$. The main question to be studied is: pick $M \varepsilon G$ from Haar measure; how are $\{e^{i\theta_1}, \dots, e^{i\theta_n}\}$ distributed? To begin our study, Figures 1 through 3 show eigenvalues for five independent realizations from O_n, U_n, S_n when $n = 100$. Also shown for comparison are five realizations of 100 points chosen uniformly and independently on the unit circle.

Figures 1 and 2 are similar. Each shows sets of 100 points neatly arranged around the unit circle. There are slight variations, but the points are close to 1/100 apart. A careful look shows the eigenvalues for O_n come in complex conjugate pairs. In contrast, Figure 4 shows that completely random points have much greater variability than the eigenvalues of random matrices. Figure 3 corresponds to the symmetric group. It shows neatly arranged points with varying densities. The rest of this paper presents a fairly detailed theoretical understanding of these and more subtle patterns. Before this, let us pose a basic question.

WHO CARES?

There are many questions; why study these? The next four sections of the paper offer motivation; the eigenvalues appear in applied problems of telephone encryption (Section One) and in routine statistical work (Section Two). They appear in the mathematical understanding of the zeros of Riemann's zeta function (Section Three). They also have remarkable internal properties suggesting study for their own sake (Section Four).

Section Five gives a general picture for understanding and proving things about unitary eigenvalues. This uses tools of representation theory. Section Six gives pointers to the literature on topics not covered: other ensembles, free probability, de Finetti-type theorems, largest eigenvalues and much else.

As an applied mathematician who is not a physicist, connecting my interests to Gibbs' legacy seemed like an impossible task. Despite my limitations, mathematical physics runs throughout random matrix theory. The physics of the telephone drives the analysis of Section One. Particle scattering directly connects physics and random matrices. Szegö's strong limit theorem was proved in answer to a question of Onsager on Ising phase transitions. The first rigorous proof of the equivalence of ensembles for Gibbs' measures can also be understood as a part of random matrix theory. Physics illuminates much of mathematics. We hope for the converse.

I thank my coauthors—Dan Bump, Steve Evans and Mehrdad Shahshahani—along with my students—Joe Blitzstein, Marc Coram, Jason Fulman, Eric Rains and Kelly Wieand—for their contributions to this work. Thanks, too, to my random matrix friends—Percy Deift, Kurt Johansson, Alexei Borodin, Neil O'Connell, Andre Okunkov, Craig Tracy and Harold Widom.

1. Telephone encryption

My interest in random orthogonal matrices began with an applied problem in telephone encryption. While it is well understood how to cryptographically scramble up bits, telephone encryption must make the scrambling commensurate with the physics of the telephone and be done rapidly enough to permit normal conversation. One scheme due to Aaron Wyner [85] digitized speech into 8-bit blocks and treated these as real numbers. Vectors of 256 such blocks can be encrypted by rotating with a 256 × 256 random orthogonal matrix. This scrambled vector is transmitted, and the receiver decrypts the message by multiplying by the inverse matrix. Keeping the length of the signal constant is crucial to practical encryption of speech.

All of this requires a stream of random orthogonal matrices. The Gram-Schmidt procedure previously described takes order n^3 steps to generate an $n \times n$ matrix. After all, the rows above the i^{th} have to be removed by an inner product of length n. The number of operations is thus of order

$$\sum_{i=1}^{n} in = 0(n^3).$$

When $n = 256$, putting in the constants, this algorithm takes approximately 16×10^6 operations and is simply too slow to allow natural speech on the telephone.

Neil Sloane suggested that perhaps approximately random orthogonal matrices would be practically as good. He suggested forming an approximately random element of O_n by multiplying a few random reflections: matrices of the form $I - 2uu^T$ with u chosen uniformly on the unit sphere. One can multiply a matrix by a reflection in order n^2 operations. This raised the following problem: how many reflections are required to have an approximately uniform product? Preliminary work with Shahshahani [30], brilliantly completed by Rosenthal [83] and Porod [79], shows that $\frac{1}{2}n \log n$ products are necessary and suffice to achieve approximate uniformity. The lower bound of this theorem proceeds as follows: consider a single product $I - 2uu^T$. Why isn't this random on its own? For one thing, it fixes an $n-1$ dimensional subspace. Similarly, the product of k reflections fixes an $n-k$ dimensional subspace. Thus if k is not large enough, the trace of the product will be large. This raises the question, how large will the trace of a uniformly chosen element of O_n typically be? We have finally arrived at an eigenvalue question.

Consider a uniformly chosen matrix $M \varepsilon O_n$. Its diagonal entries are small numbers (about size $\frac{1}{\sqrt{n}}$), and different rows should not be too dependent. The basic central limit theorem of probability says that if you add up a large number of approximately independent random numbers, the sum should be approximately distributed like the bell-shaped curve $e^{-x^2/2}/\sqrt{2\pi}$. Mallows and I were able to prove this:

Theorem. *Let M be chosen uniformly in O_n. Then, as n tends to ∞,*

$$\left| P\{tr\ M \leq x\} - \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| \to 0, \tag{1.1}$$

uniformly in x.

This result will be extended and refined in later sections. It implies that no matter how large n is, the trace of a random orthogonal matrix is less than three in absolute value with high probability. Using character theory, it is not hard to show that $\frac{1}{2}n \log n + cn$ random reflections are required to make the trace this small under the convolution measure.

Returning to the original telephone encryption problem, the bounds show that $\frac{1}{2}n \log n + cn$ reflections suffice to be close to random. If all that is wanted is the image of a vector following a product of reflections, this is available at cost of order $n^2 \log n$ (it takes order n steps to multiply a vector by a reflection). This gives a substantial speedup. In summary, for this example, the eigenvalues of random orthogonal matrices came in the back door as a tool for proving lower bounds on running times in an applied problem.

2. Statistics and eigenvalues

The earliest manifestations of random matrix theory may be the fluctuation theory of correlations. Statisticians frequently analyze high dimensional data by looking at covariance matrices and their eigen-decompositions into principal components. To explain by example [67], consider the scores of 100 pupils on 5 math exams through the term. If the i^{th} students' scores are $X_i = (X_{i1}, \ldots, X_{i5})$, then the data matrix X is the 100×5 matrix with the X_i as rows. It is natural to look at linear functions of the scores, say, $\gamma \cdot X_i = \sum_{j=1}^{5} \gamma_j X_{ij}$. The norm one vector γ^* which maximizes the variance of the hundred numbers $\gamma \cdot X_1, \ldots, \gamma \cdot X_{100}$ is called the first principal component of X. The vector γ_{**} maximizing variance subject to orthogonality to γ^* is the second principal component, and so on. In the example, the first principal component is approximately the average of the five scores, while the second principal component is approximately the difference between the average of the first two tests and the last three tests. Histograms of the data on these first two principal component directions might well be used to assign final grades and assess the progress of the class. If we are to look at the patterns in γ^* and γ^{**}, it is natural to ask about their stability. If the data had come out slightly different, would the inferences change much?

It is not hard to see that the principal components are the eigenvectors of a suitably scaled version of the 5×5 covariance matrix $X^T X$. The variance of the data projected onto the maximizing eigendirections are the eigenvalues. If the data is a sample from a larger population or modelled as stochastic in other ways, understanding fluctuations of the eigenanalysis is random matrix theory.

In general, $n \times p$ data matrices are considered with rows which are independent samples from some fixed population. R. A. Fisher and J. Wishart found the sampling distribution of $X^T X$ when the population is Gaussian. In the 1930's, Wilks, Hsu, Girshick and others derived the joint distribution of the eigenvalues and eigenvectors for the Gaussian case. Anderson [5] and Muirhead [70] give a normal approximation for the eigenvalues when n is large and p is fixed for general distributions for X. The mathematical development, largely due to Alan James, is intimately linked to zonal polynomials, the spherical functions associated to the action of the orthogonal group $O(p)$ on the positive definite matrices. See [65] for details and references.

Modern statistical work, as applied in areas such as data mining or search engines, deals also with cases with p large. The empirical distribution of the bulk of eigenvalues of covariance matrices was studied by Marcenko–Pastur [66]. They showed that if $n, p \uparrow \infty$ with $n/p \to \gamma$, then

$$\frac{1}{p}\{\#e.v. \leq nt\} \to G(t) \tag{2.1}$$

with G a distribution function having density

$$g(t) = \frac{\gamma}{2\pi t}\sqrt{(b-t)(t-a)},\ a \leq t \leq b, a = (1-\gamma^{\frac{1}{2}})^2, b = (1+\gamma^{\frac{1}{2}})^2. \tag{2.2}$$

These distributions vary considerably in shape with γ.

Following work of Johansson, Johnstone [58] derived the fluctuation theory of the extreme eigenvalues for the Gaussian case. He showed that the largest eigenvalue ℓ_1 satisfies

$$\frac{\ell_1 - \mu_{np}}{\sigma_{np}} \Rightarrow F_1 \tag{2.3}$$

where $\Rightarrow$ denotes weak $*$ convergence,

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2, \sigma_{np} = (\sqrt{n-1} + \sqrt{p})\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}} \tag{2.4}$$

and F_1 is the Tracy-Widom distribution

$$F_1(s) = e^{-\frac{1}{2}\int_s^\infty q(x) + (x-s)q^2(x)dx}, \tag{2.5}$$

where q solves the Painlevé II equation

$$q'' = xq(x) + 2q^3(x), \quad q(x) \sim Ai(x) \ \text{ as } \ x \to \infty. \tag{2.6}$$

Here the scaling is non-classical – the standard deviation grows with sample size as $n^{\frac{1}{3}}$. Johnstone has shown that this approximation is useful for n as small as ten.

Much of the mathematical work on eigenvalues in statistics was done for Gaussian random variables. Because of the orthogonal invariance of Gaussian vectors, the mathematical development is closely related to the orthogonal group. Useful surveys of available results for Gaussian and more general populations appear in Bai [8] and Muirhead [70].

3. Connections with the Riemann zeta function

There is a surprising, unexplained connection between the eigenvalues of random matrices and the zeros of Riemann's zeta function. We give a brief fresh look at this from a statistical point of view, following joint work with Marc Coram [24]. Pointers to the large literature are given at the end of the section.

For complex s with $re(s) > 1$ the Riemann zeta function is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \Pi_p \left(1 - \frac{1}{p^s}\right)^{-1} \tag{3.1}$$

where the product is over all primes. The zeta function can be continued to the whole complex plane with a simple pole at $s = 1$. Riemann showed that knowledge of the zeros of $\zeta(s)$ would give information about the distribution of primes. It is known that, except for "trivial zeros" at $-2, -4, -6, \ldots,$ all the zeros are in the critical strip $0 < re(s) \leq 1$. Riemann showed that the number of zeros in the strip with imaginary parts between zero and T is

$$N(T) = \frac{T}{2\pi} \log \frac{T}{2\pi e} + 0(\log T) \quad \text{as} \quad T \to \infty. \tag{3.2}$$

This means that the zeros get denser higher up with local density $\log \frac{T}{2\pi}$ at height T. The Riemann hypothesis says that all the zeros are on the critical line $re(s) = \frac{1}{2}$.

We next connect the zeros in a neighborhood of the strip at height T to the eigenvalues of typical unitary matrices in U_n. To have the same density of eigenvalues as zeros, following an idea of Keating and Snaith [60], [61], choose $n \doteq \log \frac{T}{2\pi}$. In the data to be described below, 50,000 zeros starting at the 10^{20} zero are considered. Here $T \doteq .15 \times 10^{20}$ and $n \doteq \log \frac{T}{2\pi} \doteq 42$. To compare the zeros with eigenvalues, we wrap blocks of 42 zeros around the unit circle. More precisely, given zeros $Z_1, Z_2, \ldots, Z_N$ of the form $Z_j = \frac{1}{2} + i\tau_j$ with $\tau_1 \ldots < \tau_j \ldots < \tau_N$, form spacing $\sigma_j = \tau_{j+1} - \tau_j$. Split the spacings into disjoint groups of size $n + 1$. Each group of spacings $\sigma_1, \ldots, \sigma_n$ is mapped onto the unit circle by taking $x_j = \exp(2\pi i (\frac{\Delta_j}{\Delta_n}))$ for $1 \leq j \leq n$, where $\Delta_j = \sum_{n=1}^{j} \sigma_n$ and U is a uniform random variable on $[0, 2\pi]$ chosen independently for each group.

When $N = 50{,}000$ and $n = 42$ this gives about $50{,}000/43 \doteq 1190$ different wrapped data sets. The claim is that these data sets are distributed like the eigenvalues of matrices chosen from Haar measure on the unitary group U_{42}. This well-posed hypothesis was exhaustively tested in [24]. We present a few of the results here, but the bottom line is that the wrapped zeta data passes virtually all the tests thrown at it.

One approach to testing goodness of fit of the wrapped zeta data to unitary eigenvalues is to look at the trace. As explained in the section above

$$P_n\big(|\mathrm{Tr} M|^2 \geq t\big) \to e^{-t} \text{ uniformly in } t \text{ as } n \to \infty. \tag{3.3}$$

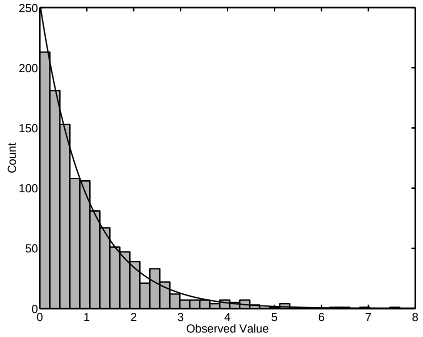


FIGURE 5. Zeta data

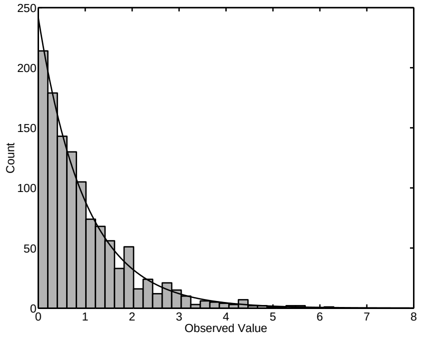


FIGURE 6. Null data

FIGURES 5-6: On the left, a histogram of 1190 zeta-function-based norm-squared "traces" with the standard exponential density function superimposed. On the right, a histogram of 1190 independent standard exponential random variables.

For application to the zeta data, $n = 42$. Work of Johansson, described in fact two of Section Four below, shows that the exponential approximation in (3.3) is remarkably good.

Figure 5 shows a histogram of the 1190 "traces" based on the wrapped zeta data with the exponential density imposed. To help the reader calibrate, Figure 6 shows a sample from a true exponential distribution. The two pictures seem interchangeable. More formal tests also show the traces match the exponential distribution remarkably closely.

A second test may be based on strange correlations found by Kelly Wieand [96], [97]. For $0 < a < b \leq 2\pi$ let $X_{ab}(M)$ be the number of eigenvalues of M satisfying $a < \theta_j < b$. Because of the neat distribution of eigenvalues, this random quantity has expected value $n(b-a)/2\pi$. Wieand shows that

$$Y_{ab} = \frac{X_{ab}{}^{-n(b-a)/2\pi}}{\sqrt{\log n/\pi^2}} \Rightarrow N(0,1). \tag{3.4}$$

Thus the fluctuations are at a logarithmic level, and the normalized error follows the bell-shaped normal distribution. To understand the limiting process, Wieand calculated the correlation between Y_{ab}, Y_{cd}. She found that in the large n limit

$$Corr(Y_{ab}, Y_{cd}) \to \begin{cases} 0 & \text{if} \quad \partial(a,b) \cap \partial(c,d) = 0 \\ -\frac{1}{2} & \text{if} \quad b = c \\ +\frac{1}{2} & \text{if} \quad a = c \end{cases}$$

These are strange correlations. They say that if $[a,b]$ *contains* $[c,d]$ properly, then Y_{ab}, Y_{cd} are approximately independent, while if the two intervals share a single endpoint, the limiting variables have correlation $\pm\frac{1}{2}$. Experienced probabilists found this surprising. (At first I did not think these correlations were positive definite!) In retrospect, the limiting variables make perfect sense. Suppose each point θ on the circle is assigned an independent Gaussian variable Z_θ with mean zero and

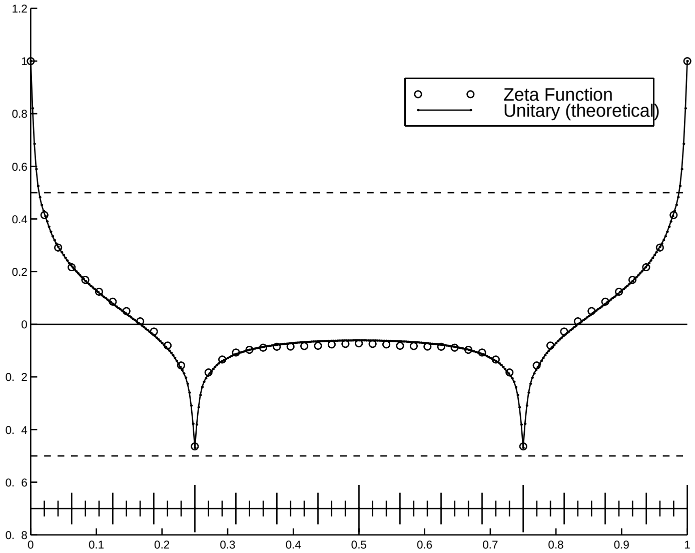


Figure 7. Correlations for $\left(0, \frac{\pi}{4}\right)$ and $\left(\theta, \theta + \frac{\pi}{4}\right)$. Solid line is the theoretical curve for Haar measure on U_n. The circles depict the empirical correlations calculated from wrapped zeta data.

variance $\frac{1}{2}$. Assign the interval $[a, b]$ to $Z_b - Z_a$. These variables have the correlations reported above. Clearly, if the intervals have distinct endpoints, the variables $Z_b - Z_a$ and $Z_d - Z_c$ are independent, while for example: $E(Z_b - Z_a)(Z_d - Z_b) = -\frac{1}{2}$.

This being observed, the obvious question is, where is the "white noise" Z_θ hidden in random matrices? A lovely, simple answer was given by Hughes et al. [53]. They showed that for each θ, with M chosen uniformly in U_n,

$$Z_\theta^n = im \text{ log DET } (e^{i\theta} - M)/\sqrt{\log\, n/\pi^2} \tag{3.5}$$

has a limiting normal distribution with the limits being independent for different θ. Thus the log characteristic polynomial is approximately normal. Using the principle of the argument, they showed that Wieand's results follow.

With all of this background, let us ask if we can find these strange correlations hidden in the zeros of the zeta function. To begin with, Wieand's results are large n limits and here $n = 42$. A finite sample correlation was calculated by Bump-Diaconis [20], with uniform approximation available through Bump-Diaconis-Keller [21]. The correlations for intervals $(0, \frac{\pi}{4})$, $(\theta, \frac{\pi}{4} + \theta)$ are shown in Figure 7. Of course, when $\theta = 0$, the correlation is one. As θ varies, the correlation drops, and when $\theta = \frac{\pi}{4}$ (so the intervals share an endpoint) it is $-\frac{1}{2}$.

Also shown are the empirical correlations based on the wrapped zeta data. We again see a striking match. The work with Coram reports extensive further testing of both specific features and an omnibus test. The upshot is a remarkable fit between the zeta data and the unitary eigenvalues. The only question to be answered now is, where is the operator?

The first connections between zeta zeros and random matrix theory were drawn by Hugh Montgomery following a conversation with Freeman Dyson. Montgomery suggested (and roughly proved) that the pair correlations – chance of finding a zero at distance x from a first zero – should match up with the pair correlations of the eigenvalues of stochastic Hermitian matrices (GUE) after suitable density adjustment to make the mean spacing one in both cases. Odlyzko [71], [72] carried out a remarkable study of data from consecutive zeros and eigenvalues. Marvelous new methods were invented for accurate computation of zeros far up. Michael Berry posited remarkable correction terms to the Dyson-Montgomery 2-point correlations, and Odlyzko was able to show an amazing match to zeta data. Much of this work is carefully reviewed in Conrey [22], Berry and Keating [12] and Katz-Sarnak [59]. In particular, the last authors suggest that random matrix theory works for general families of L-functions. Keating and Snaith [60], [61] opened a new chapter by suggesting unambiguously that the characteristic polynomial $Z(M,\theta) = \prod_{j=1}^{n}(1-e^{i\theta_j-\theta})$ was a useful surrogate for the zeta function. They showed that moments of the zeta function along the critical line matched moments of $Z(M,\theta)$ over U_n and, in joint work with several sets of coauthors (see e.g. [23]), make remarkable predictions about zeta behavior which matches data and number theoretic heuristics. This has led number theorists to ask even more detailed questions of probabilists. For example, to understand the biggest gap between zeros, one would like to understand the biggest gap between unitary eigenvalues. This is open at this writing.

4. Five surprising facts

Another reason for studying the eigenvalues is that the mathematics is surprising and beautiful. The joint probability density for the eigenvalues of a Haar-distributed random matrix in U_n is well known as the Weyl denominator formula,

$$f_2(\theta_1,\dots,\theta_n) = \frac{1}{(2\pi)^n n!}\prod_{j<k}|e^{i\theta_j}-e^{i\theta_k}|^2. \tag{4.1}$$

This is a probability density on $(0,2\pi)^n$ with respect to product Lebesgue measure. See [47] for the classical derivation. Alas, this elegant, explicit formula is not of much use in understanding the distribution of eigenvalues. All one can see is that f_2 tends to zero as θ_j and θ_k approach each other, so the eigenvalues tend to repel. Physicists write the product as

$$e^{-\beta H(\theta_1\dots\theta_n)} \text{ with } \beta=2,\ H=-\sum_{j<k}\log|e^{i\theta_j}-e^{i\theta_k}| \tag{4.2}$$

and invoke statistical physics intuition for a "Coulomb Gas" of n repelling electrical particles around a circle.

The following theorems make some of this intuition precise, but show there are many surprises hidden in the simple formula (4.1).

Fact One: *The eigenvalues are very neatly spaced* (*but slightly random*). This can be seen visually in Figure 2. To interpret the mathematical statement below, note two things: First, the sum of n complex numbers equally spaced around the unit circle is zero. Second, the sum of n complex numbers put on the unit circle at random (independently and uniformly) is of order $\sqrt{n}$. This follows from the classical central limit theorem of probability theory. In joint work with Mallows [29] we proved:

Theorem. *For M chosen from Haar measure, n large and any ball $B \subseteq \mathbf{C}$,*

$$P\{\mathrm{Tr}(M) \in B\} \sim \int_B \frac{e}{\pi}^{-|z|^2}\, dz. \tag{4.3}$$

Here the trace is *not* divided by $\sqrt{n}$, so the eigenvalues practically cancel out. If B_r is the ball in $\mathbf{C}$ centered at zero with radius r, the right side of (4.3) equals $1 - e^{-r^2}$.

I still find it mysterious, looking at (4.1) and asking how the cancellation occurs. Of course, things do not cancel perfectly. Wieand's theorem described in Section Three above shows that the number of eigenvalues in a fixed interval of the unit circle is n times the interval's length, plus fluctuations of order $\sqrt{\log n}$.

Fact Two: *The traces are amazingly close to Gaussian.* Consider the error term in the Gaussian approximation to the trace at (4.3). Johansson [56] proved:

Theorem. *There are universal constants $c, \sigma > 0$ such that*

$$\left| P\{\mathrm{Tr}\ M \ \in \ B\} - \int_B \frac{e}{\pi}^{-|z|^2} dz \right| \ \leq \ \frac{c}{n^{\sigma n}}$$

uniformly in Borel sets B.

People used to the usual error terms in probability find this result fantastic; for the classical central limit theorem for the sum of n points randomly put on the unit circle, the error is of order n. Here, the error is super-exponential.

Two closely related findings:

(a) The moments of the trace *equal* the normal moments

$$\int (\mathrm{Tr}\ M)^a\ \mathrm{Tr}\ (\bar{M})^b\ dM = \int |z|^a\ |\bar{z}|^b\ \frac{e}{\pi}^{-|z|^2}\ dz. \tag{4.4}$$

The equality (4.4) for all integers a, b smaller than n is joint work with Colin Mallows, which is discussed at length in Section Five below.

(b) An analogous result is proved for the trace of a random permutation matrix.

Theorem. *For Π, a uniformly chosen permutation matrix*

$$|P\ \{\mathrm{Tr}\,\Pi\ \varepsilon\ B\} - P\ (B)| \sim \frac{2^n}{(n+1)!}$$

$$\textit{with}\quad P(B) = \frac{1}{e}\sum_{i \in B} 1/i! \quad \textit{for} \quad B \subseteq \{1, 2, 3, \ldots\}. \tag{4.5}$$

In summary, the first thoughts suggesting a Gaussian limit for the trace were that the trace is the sum of a lot of small, approximately uncorrelated random terms. While this is true, there is some mysterious global constraint that forces the high order contact with the Gaussian law.

Fact Three: *Higher order traces go from order to chaos.* In investigating the fine structure of the eigenvalues, traces of higher powers are studied. In joint work with Shahshahani [32] discussed in Section Five below, we proved:

Theorem. *For M chosen from Haar measure on U_n, for j fixed and $n \uparrow \infty$,*

$$P\{\mathrm{Tr}(M^j)\ \in\ B\} \to P\{\sqrt{j}\ Z\ \in\ B\}$$

with Z a standard complex Gaussian variable.

As j increases, $\sqrt{j}Z$ becomes more spread out. Eric Rains [80] proved that a kind of phase transition occurs for $j \geq n$.

Theorem. *Let M be chosen from Haar measure on U_n. For any n, and any $j \geq n$, the eigenvalues of M^j are exactly distributed as n points chosen independently and uniformly on the unit circle.*

Thus high powers of M have *no* structure in their eigenvalues. The trace of M^n is approximately Gaussian, but with error $\frac{1}{n}$. I find the contrast between facts one, two and three unsettling. I still have no mental picture that explains how these can all hold. I have tried to think about generating Haar-distributed eigenvalues by putting down n uniform points independently and taking appropriate n^{th} roots. Rains' result can be demystified slightly by computation:

The joint mixed moments of the eigenvalues of M^j are

$$\int \prod_{k=1}^{n} e^{ij\theta_k(b_k-a_k)}\; f_2(\theta_1,\ldots,\theta_k)\; d\theta_1,\ldots,d\theta_n.$$

Expanding f_2 from (4.1) as a polynomial, we see that for $j \geq n$ we get zero unless $b_k = a_k$ for $1 \leq k \leq n$. These are just the joint mixed moments for independent uniform points.

Rains [80] proved similar results for all the classical compact Lie groups. There is a subtlety here. Take the orthogonal group O_{2n}. The eigenvalues come in conjugate pairs, and this is preserved for powers. Rains shows that for suitably large j the eigenvalues of M^j are exactly distributed as n random points and their conjugates.

Some light on these strange doings follows from work of Forrester-Rains [42], Haake [49] and Rains [81]. For simplicity, consider M chosen from Haar measure on U_{2n}. They prove that for all n, the eigenvalues of M^2 are exactly distributed as the union of two independent sets of eigenvalues from M_1, M_2 chosen from Haar measure on U_n. Similarly, the eigenvalues of $M^3 \in U_{3n}$ are exactly distributed as the union of eigenvalues of M_1, M_2, M_3 independently chosen in U_n. Their final result has no divisibility requirements and applies for all powers j. It shows that the eigenvalues of $M^n \in U_n$ are exactly distributed as independent points from U_1.

These extra facts compound the mystery.

Fact Four: *Neat marginals.*

The foundations of random matrix theory were laid out in a great series of papers by Dyson [38], [39], who labeled the unitary ensemble CUE (Compact Unitary Ensemble). Dyson showed that the marginal distribution of n eigenvalues has a neat form. Physicists call these "n-point correlations" for mysterious historical reasons (they are *not* correlations!). Thus f_2 from (4.1), which could be written f_2^n, is the n-point distribution. Informally, $f_2^k(\theta_1,\ldots,\theta_k)$ is the probability density for an eigenvalue in $d\theta_1, d\theta_2,\ldots,d\theta_k$ from a Haar-distributed matrix in U_n. The elegant fact (due to Dyson) is a simple, closed-form formula:

$$f_2^k(\theta_1\ldots\theta_k) = \mathrm{DET}\left(\frac{\sin(n(\theta_a-\theta_b))}{\sin(\theta_a-\theta_b)}\right) 1 \leq a,b \leq k.$$

For example, $f^1 = \text{Constant}$, $f^2 = 1 - \left(\frac{\sin n(\theta_1-\theta_2)}{\sin(\theta_1-\theta_2)}\right)^2$.

Maachi [63], [64] created a general theory for point processes with k-point distributions having determinental form. See Daley-Verre-Jones [25] for a concise and very readable treatment of point processes and Macchi's work. Soshnikov [88] gives a marvelous survey showing how surprisingly many ensembles admit neat determinental forms, and how a wide variety of limit theorems can be proved from these forms. Following work of Macchi, Diaconis-Evans [34] survey developments where determinants are replaced by permanents or immanents.

Fact Five: *The Toeplitz connection.*

I find the following connection surprising. Shahshahani and I proved that if M is chosen from Haar measure on U_n, the trace of successive powers has limiting Gaussian distributions. As $n \to \infty$, for any fixed k and Borel sets $B_1, \dots, B_k$

$$P(\mathrm{Tr}M \in B_1, \dots, \mathrm{Tr}M^k \in B_k) \to \prod_{j=1}^{k} P(\sqrt{j}\, Z \in B_j) \tag{4.6}$$

with Z a standard complex Gaussian variable.

In a seemingly very different sphere, G. Szegö derived the limiting asymptotics for the eigenvalues of Toeplitz matrices. This is a rich subject, and I will not try to develop it in detail. Böttcher and Silbermann [17] is a remarkably good introduction. Briefly, a Toeplitz matrix is an $n \times n$ matrix with complex entries which are constant down the main diagonals, such as

$$\begin{bmatrix} a & b & c & d \\ e & a & b & c \\ f & e & a & b \\ g & f & e & a \end{bmatrix}$$

Szegö determined the asymptotics of the determinants of a sequence of Toeplitz matrices $T_n(g)$ constructed with (j,k) entry $\hat{g}(j-k)$ where $\hat{g}(m) = \frac{1}{2\pi}\int_0^{2\pi} g(e^{im\theta})\, d\theta$ is the Fourier transform of a function on the unit circle.

Theorem (Strong-Szegö)**.**

$$\textit{Let} \quad g(e^{i\theta}) = e^{f(e^{i\theta})} \quad \textit{with} \quad \sum_{-\infty}^{\infty} |\hat{f}\ (k)|^2|k| < \infty. \tag{4.7}$$

Then

$$DET\, T_n(g) = e^{n\hat{f}(0) + \sum_{k=1}^{\infty} f(k)f(-k)k + o(1)}.$$

Szegö's theorem has dozens of variations and applications from time series and electrical engineering to the first proof of phase transitions in the Ising model. Grenander-Szegö [48] is a classical, readable overview of this material, as is Chapter Five in Böttcher-Silbermann [17].

The point of the present presentation is that (4.6) and (4.7) seem completely unconnected, and yet they are easily equivalent. The key connection is a classical determinant identity of Heine and Szegö.

Proposition. *For f as above and $g = e^{f(e^{i\theta})}$,*

$$\frac{1}{(2\pi)^n}\int_0^{2\pi} \cdots \int_0^{2\pi} \prod_{j=1}^{n} e^{f(e^{i\theta_j})} \prod_{1\le\, a\le\, b\le\, n} |e^{i\theta_a} - e^{i\theta_b}|^2\, d\theta_1 \dots d\theta_n = T_n(g). \tag{4.8}$$

It is not hard to see this directly; expand the determinant on the right side as a sum of products of n terms. Each term is a Fourier transform, and so the product is an n-fold integral of the same form as the left side. Recognizing a Vandermonde and elementary manipulations complete the proof. Alternatively, in joint work with Bump [20], we have shown that (4.8) follows from the classical Jacobi-Trudi identity of symmetric function theory.

With (4.8) established we can now see the equivalence of (4.6) and (4.7). Begin with (4.6). This is a limit theorem for the traces. Passing to Fourier transforms, let

$$f(e^{i\theta}) = \sum_{j=1}^{k} a_j \ \cos(j\theta) + b_j \ \sin(j\theta).$$

Then, for $M \in U_n$ with eigenvalues $e^{i\theta_1}, \ldots, e^{i\theta_n}$

$$a_1 re \ \mathrm{Tr}(M) + b_1 im \ \mathrm{Tr}(M) + \ldots + a_k \ re \ \mathrm{Tr}(M^k) + b_k im \ \mathrm{Tr}(M^k) = \sum_{j=1}^{n} f(e^{i\theta_j}).$$

We will use a_j, b_j as transform variables. Noting next that the transform of a Gaussian variable $\sqrt{j}Z = \sqrt{j}(X+iY)$ is

$$\int e^{\sqrt{j}(ax+by)-x^2-y^2} \frac{dxdy}{\pi} = e^{\frac{j}{4}(a^2+b^2)},$$

we see that the limiting normality of (4.6) is equivalent to convergence of the transforms

$$\frac{1}{(2\pi)^n} \int \cdots \int \prod_{j=1}^{n} e^{f(e^{i\theta_1})} \prod_{1 \le j < k \le n} |e^{i\theta_j} - e^{i\theta_k}|^2 \ d\theta_1 \ldots d\theta_n \to \exp\left(\sum_{j=1}^{k} \frac{j}{4}(a_j^2 + b_j^2) \right).$$

This is just Szegö's theorem for the trigonometric polynomial f. Thus Szegö's theorem implies (4.6). Conversely, (4.6) shows the Fourier transforms converge, and so Szegö's theorem holds for trigonometric polynomials. It is not hard to pass to the limit and derive the result for general f.

There are *many* proofs of Szegö's theorem in the references above. Johansson [55] gives a careful development of the asymptotic analysis required to go from polynomials to more general functions.

The present approach leads to straightforward generalizations in two directions. First, the limiting normality of traces of powers for other groups such as O_n or SP_{2n} gives Szegö-type theorems for determinants with entries the coefficients of expansions in Jacobi polynomials. To be fair, these are classically known [52]. A second generalization was determined in joint work with Bump [20]. This gives asymptotics for the determinants of Toeplitz minors. These seem new, but closely related work has been done by Tracy-Widom [94].

Here is a simple example. Let λ be a partition of m. Define the Toeplitz minor

$$T_n^{\lambda}(g) = det \ (\hat{g}(\lambda_i - i + j)_{1 \le i, j \le n}.$$

Then, for λ fixed and $n \to \infty$, and $g = e^f$ as in (4.7)

$$T_n^{\lambda}(g)/T_n(g) \sim \frac{1}{m!} \sum_{\sigma \in S_m} x^{\lambda}(\sigma) \prod_{k=1}^{\infty} (k\hat{f}(k))^{\gamma_n(\sigma)}.$$

Here the sum is over the symmetric group S_m, $x^\lambda(\sigma)$ is the irreducible character associated to λ, and $\gamma_n(\sigma)$ is the number of cycles of length k in σ. The minor $T_n^\lambda(g)$ is obtained from the original Toeplitz matrix by striking the first λ_1 columns, keeping the first row but striking the next $\lambda_1 - \lambda_2$ rows, keeping the next row and so on. For example, when $\lambda = (1)$ the minor is obtained by striking the first column and the second row, and the right side is $\hat{f}(1)$. It seems strange that characters appear. See [20] for extensions.

A third benefit of the present approach: it offers some explanation for the form $\sum_{k=1}^{\infty} k\hat{f}(k)\hat{f}(-k)$ in the Szegö corrections. The k appears because var $(M^k) = k$.

I do not want to leave this area without pointing to a beautiful related development due to Estelle Basor. She has derived the limit theory for the spectrum of a variety of operators of Hankel, Toeplitz and mixed-type in sweeping generality. Her results are paired with Gaussian limit theorems of the same flavor as those of this section. A readable survey with extensive references is in [10], [11].

I do not feel we understand the parallel between (4.6) and (4.7). The determinant identity seems like a magic trick!

5. A general approach

A general approach to studying unitary eigenvalues has gradually been developed. This begins in joint work with Mallows [29] and Shahshahani [32]. The present refined account was developed with Steve Evans [33].

For $M_n \in U_n$ with eigenvalues $\{e^{i\theta_j}\}$, let Ξ_n be the measure on the unit circle T which puts mass one at each eigenvalue. We may study Ξ_n via linear, quadratic, $\ldots$, functionals. Thus if $f : T \to C$ has Fourier expansion $f(e^{i\theta}) = \sum_{j\in\mathbf{Z}} \hat{f}_j e^{ij\theta}$

$$\Xi_n(f) = \sum_{j=1}^{n} f(e^{i\theta_j}) = n\hat{f}_0 + \sum_{j=1}^{\infty} \hat{f}_j \mathrm{Tr}(M_n^j) + \sum_{j=1}^{\infty} \hat{f}_{-j}\, \overline{\mathrm{Tr}(M_n^j)}. \tag{5.1}$$

The key to studying such functionals is an explicit formula for the joint mixed moments of the traces. It was proved in joint work with Evans and Shahshahani.

Proposition. *(a) Consider* $a = (a_1, \ldots, a_k)$ *and* $b = (b_1, \ldots, b_k)$ *with* $a_j, b_j \in \{0, 1, 2, \ldots\}$*. Let* $Z_1, Z_2, \ldots, Z_n$ *be independent standard complex normal random variables.*

Then, for $n \geq$ *max* $(\sum_1^k ja_j, \sum_1^k jb_j)$

$$\int_{U_n} \prod_{j=1}^{k} (trM_n^j)^{a_j} \overline{(\mathrm{Tr}M_n^j)}^{b_j} dM_n = \delta_{ab} \prod_{j=1}^{k} j^{a_j} a_j!$$

$$= E\left(\prod_{j=1}^{k} (\sqrt{j}Z_j)^{a_j} (\sqrt{j}\bar{Z}_j)^{b_j}\right).$$

(b) For any j, k, $\int \mathrm{Tr}(M_n^j)\, \overline{\mathrm{Tr}(M_n^k)} dM_n = \delta_{jk}$ *min* (j, k).

The proof of this proposition is basic Schur-Weyl duality. First, introduce the power sum symmetric functions $P_j(x_1, \ldots, x_n) = x_1^j + \ldots + x_n^j$ and for μ a partition of some integer K with a_i parts equal to i, let $P_\mu = \prod_j P_j^{a_j}$. Since $\mathrm{Tr}M_n^j = P_j(e^{i\theta_1}, \ldots, e^{i\theta_n})$, the left side in (a) is $< P_\mu, P_\nu >$ for the inner product given by integration over U_n. The power sums form a basis for the homogeneous symmetric polynomials of degree K in $x_1, \dots, x_n$. A second basis is given by the Schur functions s_λ. All needed properties are in MacDonald [65] or Stanley [89]. The two key properties for present purposes are:

(5.2) Orthogonality: $\langle s_\lambda, s_\tau \rangle = \delta_{\lambda\tau}\delta_{\ell(\lambda)\le n}$ for any partitions λ, τ with $\ell(\lambda)$ the number of parts in λ.

(5.3) Schur-Weyl Duality: For any partition μ of K,

$$P_\mu = \sum_{\lambda \vdash K} \chi^\lambda_\mu s_\lambda$$

where the sum is over partitions of K, and χ^λ_μ is the irreducible character of the symmetric group S_K associated with λ on the μ^{th} conjugacy class.

Using these formulae we simply compute: If μ is a partition of K and ν is a partition of L,

$$\begin{aligned}\langle P_\mu, P_\nu \rangle &= \left\langle \sum_{\lambda \vdash K} \chi^\lambda_\mu s_\lambda, \sum_{\tau \vdash L} \chi^\tau_r s_\tau \right\rangle \\ &= \delta_{KL} \sum_{\lambda \vdash K} \chi^\lambda_\mu \chi^\lambda_\tau \delta(\ell(\lambda) \le n).\end{aligned}$$

When $K \le n$, all partitions of K appear and the second orthogonality relation for characters shows our expression equals

$$\delta_{KL}\delta_{\mu\nu} \prod_{j=1}^{K} j^{a_j} a_j! = \delta_{ab} \prod j^{a_j} a_j!$$

This last equals the joint mixed moments of $\sqrt{j}Z_j$ by an easy calculation. This proves (a). To prove (b), observe that the calculation above gives, for any j, k,

$$\int \mathrm{Tr}(M_n^j)\mathrm{Tr}(M_n^k) dM_n = \delta_{jn} \sum_{\lambda \vdash j} |\chi^\lambda_{(j)}|^2 \delta(\ell(\lambda) \le n).$$

Here $\chi^\lambda_{(j)}$ is the character of a j-cycle. These are explicitly known. They are zero unless λ is a hook shape and $(-1)^{\ell(\lambda)-1}$ if λ is a hook shape. Now (b) follows.

The proposition shows that $\mathrm{Tr}(M_n^j)$ behaves asymptotically like $\sqrt{j}Z_j$, and one can then plug into Fourier series expansions. My work with Evans exploits this carefully for a variety of functionals. Of course, care is needed in bounding the tail of these sums and that is where (b) comes in.

Here is one carefully stated example:
Let $H_2^{\frac{1}{2}}$ denote the space of functions f in $L^2(T)$ such that $\| f \|^2_{\frac{1}{2}} = \sum_{j \in \mathbf{Z}} |\hat{f}_j|^2 |j| < \infty$. For example, $f(z) = z^j \in H_2^{\frac{1}{2}}$; $H_2^{\frac{1}{2}}$ is precisely the functions such that $\sum \hat{f}(n)\sqrt{n}Z_n$ converges almost surely, where Z_n are independent, standard, complex Gaussian variables.

Proposition. *If $f_1, f_2, \dots, f_k \in H_2^{\frac{1}{2}}$ with $\hat{f}_i(0) = 0, 1 \le i \le k$, then the random vector $(\Xi_n(f_1), \dots, \Xi_n(f_k))$ converges in distribution to a jointly normal, centered random vector $(V_1, V_2, \dots, V_k)$ with $E(V_a V_b) =< f_a, f_b >_{\frac{1}{2}}$.*

This proposition shows there is a limiting Gaussian field indexed by $H_2^{\frac{1}{2}}$ naturally associated with unitary eigenvalues. As discussed in Diaconis-Evans [33], [35], the Hilbert space $H_2^{\frac{1}{2}}$ of "$\frac{1}{2}$ differentiable functions" or functions of "finite energy" appears in many contexts, and it is natural to seek an explanation for its appearance here. This is lacking at present.

Let $P_{M_n}(Z) = det(M_n - Iz)$ be the characteristic polynomial of $M_n \in U_n$. Then $P'_{M_n}/P_{M_n} = \sum\limits_{j=1}^{n}(z - e^{i\theta_j})^{-1} = \sum\limits_{j=1}^{\infty} \overline{\mathrm{Tr}(M_n^j)} z^j$. From the proposition, this random power series converges in distribution to the random analytic function

$$G(z) = \sum_{j=1}^{\infty} \sqrt{j} Z_j z^j \quad |z| < 1$$

where the Z_j are independent, standard Gaussian random variables. With a bit more work, one can check that this convergence occurs in the space of continuous $\mathbf{C}$-valued functions on $\{z \in \mathbf{C} : |z| < 1\}$ in the topology of uniform convergence on compacts.

Random analytic functions like G have been intensely studied, and one can show that G takes all values in $\mathbf{C}$ infinitely, often with probability one. More precise statements are in [33]. Figure 8 shows a plot of the size of P'_{M_n}/P_{M_n} for a single random choice of M_n when $n = 100$. The zeros show up as the tree-like shape, and the original eigenvalues show up as the crosses on the unit circle. Figure 9 shows a similar plot of $G(z)$. Similar tree-like shapes appear when the plots are based on the zeta zeros as explained in Section Three.

This section shows how functions of the eigenvalues can be studied using traces. The method works for non-smooth functions such as the number of eigenvalues in an interval and for functions of several eigenvalues.

There are other approaches available as well. Soshnikov [86] gives a determinental expression for the Laplace transform of a linear statistic and uses it to derive Gaussian limits such as the proposition above. Hughes and Rudnick [54] use Soshnikov's method to derive non-Gaussian limit theorems for the number of eigenvalues in intervals of length $1/n$.

Adler and Van Moerbeke [1], [2], [3] have studied the eigenvalues by studying the joint Laplace transform of the traces of all powers. They show the transform satisfies a hierarchy of non-linear equations with Virasoro constraints, similar to the well-studied Toda lattice. This puts them into the well-studied territory of integrable systems, and some of the remarkable tools developed there can be used to get novel asymptotics and even simple recursions for quantities, like the distribution of the length of the longest increasing sequence in a random permutation.

The eigenvalues can also be studied by looking at the joint distribution of the coefficients of the characteristic polynomial. This is instituted in Haake [49]. Of course, the coefficients are just the elementary symmetric functions in the eigenvalues, and the traces are the power sum symmetric functions. Change of basis formulae between these two sets of symmetric functions give some information. For example, Alex Gamburd and I have shown that the k^{th} moment of the j^{th} coefficient equals the number of $k \times k$ "magic squares" with all row and column sums equal to j. Nonetheless, the coefficients of the characteristic polynomial are less tractable than the traces – the limiting distribution of e.g., the middle coefficient, is not known at this writing.

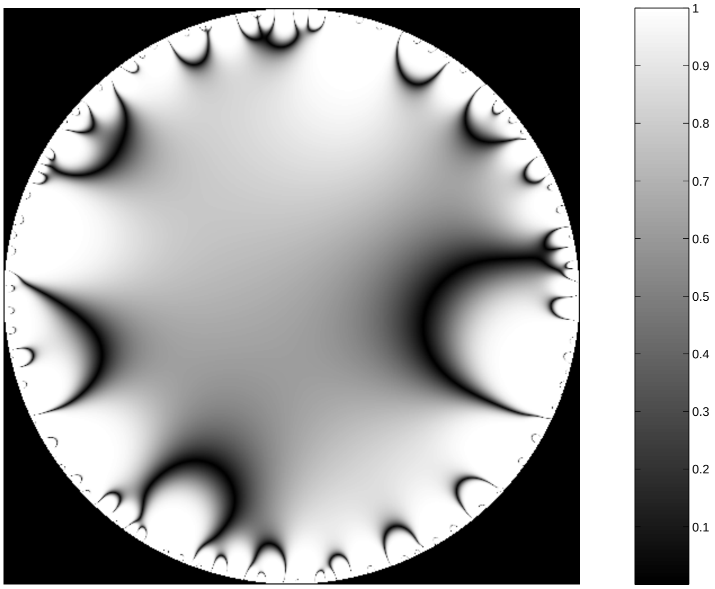


FIGURE 8. A realization of P'_{M_n}/P_{M_n} for M_n drawn from Haar on U_n and $n = 100$ is depicted here. The grayscale indicates the tanh of the absolute value of real part of this function.

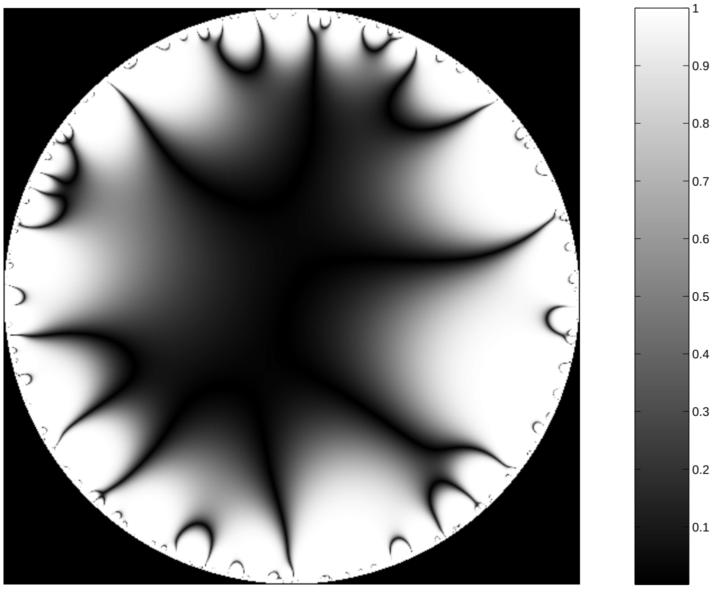


FIGURE 9. A realization of the random analytic function $G(z)$ (truncated to 1000 terms) is depicted here. The grayscale indicates the tanh of the absolute value of real part of this function.

6. Some topics not covered

I want to point readers to three rich, related areas. The first "enumeration" sets the present topic in a much broader context: study a group through understanding what "typical elements" look like. The second "free probability" is a growing set of tools to answer questions like: "Suppose you know the eigenvalues of each of two symmetric matrices. For typical matrices, what can you say about the eigenvalues of their sum?" The third topic describes a different application of random matrix theory to de Finetti's theorem in statistics and the equivalence of ensembles. To keep with the title of this section, the treatment is brief.

6.1. **Enumeration.** One way of understanding a group is to ask about the properties of typical elements. For the symmetric group S_n this is actively developed as the subject of permutation enumeration. Thus consider the following questions:

Pick $w \in S_n$ at random:

- How many fixed points does w have?
- How many cycles?
- What is the length of the longest cycle?
- What is the order of w (smallest k so $w^k = 1$)?

All of these questions only depend on w through its conjugacy class – they are invariant under irrelevant relabeling. The results are given in terms of the proportions of permutations:

- $P\{FP(w) = j\} = \frac{1}{e}\frac{1}{j!} + O\left(\frac{2^n}{n!}\right)$ Monmort (1708)
- $P\left\{\frac{\#\text{ cycles}-(\log n)}{\sqrt{\log n}} \leq x\right\} \to \int_{-\infty}^{x} e^{-t^2/2}\,\frac{dt}{\sqrt{2\pi}}$ Goncharov (1942)
- AV length of longest cycle is cn with $c \doteq .624...$ Shepp-Lloyd (1966)
- $P\left\{\frac{\log \text{order}(w)-(\log n)^2/2}{((\log n)/3)^{\frac{3}{2}}} \leq x\right\} \to \int_{-\infty}^{x} e^{-t^2/2}\,\frac{dt}{\sqrt{2\pi}}$ Erdös-Turan (1965)

These theorems give a good feel for the behavior of typical permutations. Related questions also arise in practical statistical problems [28] and in the analysis of the running time of computer algorithms. The results on the length of the longest cycle explain the fluctuations in the density of Figure 3 above.

As an abstraction, let G be a finite or compact group. Pick $g \in G$ from the uniform distribution. One may ask for the limiting distribution of the conjugacy class containing g. All of the problems discussed in the bulk of this paper are of this type. Indeed, two unitary matrices are conjugate if and only if they have the same eigenvalues. Of course, put this bald way, the question seems strange. It is an empirical fact that the general question seems to lead to elegant mathematics which has surprisingly useful consequences. A marvelous survey, making just these points and verifying them on finite groups of Lie type, is given in Fulman [43].

With all this evidence, what are you waiting for? Go get a group you'd like to learn about and try a simple case.

6.2. **Eigenvalues of typical conjugates.** How are the eigenvalues of a sum of symmetric matrices related to the eigenvalues of the summands? Of course, the traces add up, and an amazing host of further inequalities are satisfied. The exact determination of these inequalities is a great achievement of modern mathematics. Fulton [44] gives an inspiring account. These theorems give extreme or worst case bounds.

The calculus of free probability may be presented as the typical case answer. Thus let Σ_1, Σ_2 be real symmetric matrices with eigenvalues $\{\lambda_1', \ldots, \lambda_n'\}$, $\{\lambda_1'', \ldots, \lambda_n''\}$. Conjugate Σ_1 and Σ_2 by randomly chosen orthogonal matrices. The eigenvalues of the sum will now be random, and their distribution can be described via the free convolution of λ' and λ''. Here is a specific example drawn from the splendid introduction of Biane [13]. Let $n = 2m$ be even. Choose an m-dimensional subspace at random (uniformly) and let Σ_1 be the matrix for projection onto this subspace. Thus, Σ_1 has m zero eigenvalues and m eigenvalues equal to one. Let Σ_2 be independently formed in a similar way. How are the eigenvalues of $\Sigma_1 + \Sigma_2$ distributed? Free probability shows that for large n,

$$P\left\{\frac{\#\, e.v.(\Sigma_1 + \Sigma_2)}{n} \leq x\right\} \to \int_0^x \frac{1}{\pi\sqrt{t(1-t)}}dt.$$

This striking result is the tip of a remarkable set of results and tools which are rapidly becoming a major area of probability and functional analysis.

Free probability was created by Dan Voiculescu to answer questions in Von Neumann algebras. It is not known if the Von Neumann algebras associated to the free groups on 2-generators and on 3-generators are isomorphic. Voiculescu hoped to introduce an entropy-like invariant to distinguish these cases. Voiculescu's summary [95] is replete with many pointers to the huge emerging literature.

6.3. **Beyond eigenvalues.** The eigenvalues capture the coordinate-free aspects of a matrix. A different set of properties is captured by the actual matrix entries. Consider the following result:

Theorem (E. Borel)**.** *Pick Γ from the uniform distribution on the orthogonal group O_n. Then*

$$P\{\sqrt{n}\,\Gamma_{11} \leq x\} \to \int_{-\infty}^x e^{-t^2/2}dt.$$

Borel [15] proved the result in studying "Equivalence of Ensembles" in statistical mechanics. There, the "microcanonical distribution" is a suitable uniform distribution $U(dx)$ on the constant energy surface $\{x \in \mathbb{R}^N : H(x) = h^*\}$. One can predict properties by calculating averages as $\int f(x)U(dx)$. Maxwell, Boltzmann and Gibbs also considered a canonical measure $U_\beta(dx) = Z^{-1}e^{-\beta H(x)}dx$ on $\mathbb{R}^N$ with β chosen so $\int H(x)U_\beta(dx) = h^*$. The equivalence of ensembles asserts that (under conditions)

$$\int f(x)\, U(dx) \simeq \int f(x)\, U_\beta(dx),$$

when N is large. Borel took the simplest case: $H(x) = x_1^2 + \cdots + x_N^2$. Then the microcanonical measure becomes uniform on the sphere, and the canonical measure becomes product Gauss measure. Taking f to be a continuous function depending only on x_1 and using the fact that the first row of a random orthogonal matrix is Haar-distributed show that the stated theorem on matrices gives a version of equivalence of ensembles.

Borel himself, followed by Paul Levy and others, extended the result to functions depending on many coordinates. In joint work with D'Aristotle, Eaton, Freedman, Lauritzan and Newman, this result has been extended and applied to give a variety of results in mathematical statistics [7], [36], [37]. The results show that in a suitable sense, the entries $\{\sqrt{n}\,\Gamma_{i,j}\}$ are jointly distributed as independent standard Gaussian variables. While this is true in a suitable sense (arbitrary linear combinations), it is not true for the eigenvalues. The eigenvalues of $\Gamma \in \ O_n$ lie on the unit circle. The eigenvalues of a matrix of independent Gaussian variables fill out the disc with radius $\sqrt{n}$ uniformly, with order $\sqrt{n}$ on the real axis [8], [41]. Determining the right class of functions for equivalence of ensembles is still an open problem.

The behavior of the matrix entries under conjugation by a random unitary matrix has been studied by Pickrell [78], Olshansky-Vershik [77], and Borodin-Olshansky [16]. Their interest is in the representation theory of the injective limit $U(\infty)$. Results are often proved by passage to the limit from $U(n)$. Their elegant results are too rich to state completely, but for a broad class of examples, the resulting conjugation is approximately a constant times the identity when the dimension is large.

6.4. **Topics really not covered.** The present paper is based on my Gibbs Lecture but incorporates a few recent developments. The field of random matrix theory has had an explosive growth. Much of this has been on the distribution of the largest eigenvalue of random symmetric or Hermitian matrices. There have been *many* fine surveys. The work of Baik-Deift-Johansson on an integrable systems approach to largest eigenvalues and the longest increasing subsequence of a random permutation is surveyed in [27]. The work of Tracy-Widom on a wide variety of random matrix results and applications using Painlevé transcendents is surveyed in [90], [91] and [92], [93]. The work of Adler-Van Moerbeke linking random matrices to Virasoro algebras and much else is surveyed in [3]. For work of Okounkov, connecting random matrices and random permutations through Riemann surfaces, see [75]. A wonderful connection between classical queuing theory, tableaux combinatorics and random matrix theory has been developed by Bougerol-Jeulin [18] and by O'Connell-Yor [74]. I have not really touched on the physical applications of random matrix theory, though Mehta [68] and Bohigas et al. [14] give extensive pointers. Similarly, I lament not describing the many interactions with algebraic combinatorics. See [4], [43], [51]. Many of the results stated here for unitary matrices are "universal", applying to many other matrix ensembles (just as the central limit theorem): see Tracy-Widom [93] for an overview of the many great results. I have focused on eigenvalues of unitary or Hermitian matrices. There are remarkable probabilistic theorems for non-Hermitian matrices. See [8] and [45], [46] for pointers.

I hope I have given a picture of a thriving zoo with a wealth of novel findings that touch many areas of pure and applied mathematics.

References

[1] Adler, M.; Van Moerbeke, P., Integrals over Classical Groups, Random Permutations, Toda and Toeplitz Lattices. Comm. Pure Appl. Math. **2001**, *54*, 153–205.

[2] Adler, M.; Van Moerbeke, P., Recursion Relations for Unitary Integrals, Combinatorics and the Toeplitz Lattice. Technical Report, Dept. of Mathematics, Brandeis, 2002.

[3] Adler, M.; Shiota, T.; Van Moerbeke, P., Random Matrices, Virasoro Algebras, and Non-Commutative KP. Duke Math. J. **1998**, *94*, 379–431. MR **99e:**58088

[4] Aldous, D.; Diaconis, P., Longest Increasing Subsequences: From Patience Sorting to the Baik-Deift-Johansson Theorem. Bull. Amer. Math. Soc. **1999**, *36*, 413–432. MR **2000g:**60013

[5] Anderson, T., Asymptotic Theory for Principal Component Analysis. Ann. Math. Statist. **1963**, *34*, 122–148. MR **26:**3149

[6] d'Aristotle, A., An Invariance Principle for Triangular Arrays. Jour. Theoret. Probab. **2000**, *13*, 327–342.

[7] d'Aristotle, A.; Diaconis, P; Newman, C., Brownian Motion and the Classical Groups. Technical Report, Stanford University, 2002.

[8] Bai, Z., Methodologies in Spectral Analysis of Large Dimensional Random Matrices: A Review. Statistica Sinica **1999**, *9*, 611-677. MR **2000e:**60044

[9] Baik, J.; Deift, P.; Johansson, K., On the Distribution of the Length of the Longest Increasing Subsequence of Random Permutations. Jour. Amer. Math. Soc. **1999**, *12*, 1119–1178. MR **2000e:**05006

[10] Basor, E., Distribution Functions for Random Variables for Ensembles of Positive Hermitian Matrices. Comm. Math. Phys. **1997**, *188*, 327–350. MR **99b:**82046

[11] Basor, E., Connections Between Random Matrices and Szegö Limit Theorem. In *Spectral Problems in Geometry and Arithmetic*; Branson, T., Ed.; Amer. Math. Soc.: Providence, RI, 1999; 1–7. MR **2000e:**47049

[12] Berry, M.; Keating, J., The Riemann Zeros and Eigenvalue Asymptotics. Siam Review **1999**, *41*, 236–266. MR **2000f:**11107

[13] Biane, P., Free Probability for Probabilists, 2000, preprint.

[14] Bohigas, O.; Giannoni, M., Chaotic Motion and Random Matrix Theories. In *Mathematical and Computational Methods in Nuclear Physics*; Dehesa, J.L., Ed.; Springer Lecture Notes in Physics, 1984, *209*, 1–99. MR **86c:**58129

[15] Borel, E., Sur les Principes de la Theorie Cinetique des Gaz. Annales, L'Ecole Normal Sup. **1906**, *23*, 9–32.

[16] Borodin, A.; Olshansky, G., Correlation Kernels Arising from the Infinite-Dimensional Unitary Group and Its Representations. University of Pennsylvania, Department of Mathematics, 2001, preprint.

[17] Böttcher, A. and Silbermann, B., *Introduction to Large Truncated Toeplitz Matrices.* Springer-Verlag: New York, 1999. MR **2001b:**47043

[18] Bougerol, Ph. and Jeulin, Th., Paths in Weyl Chambers and Random Matrices. Laboratoire de Probabilities: Paris **2001**, preprint.

[19] Boutet de Monvel, A.; Pastur, L.; Shcherbina, M., On the Statistical Mechanics Approach in the Random Matrix Theory: Integrated Density of States. Jour. Statist. Phys. **1995**, *79*, 585–611. MR **96d:**82033

[20] Bump, D. and Diaconis, P., Toeplitz Minors. Jour. Combin. Th. A. **2002**, *97*, 252–271. MR **2002j:**47052

[21] Bump, D.; Diaconis, P.; Keller, J., Unitary Correlations and the Fejer Kernel. Mathematical Phys., Analysis, Geometry **2002**, *5*, 101–123.

[22] Conrey, B., *L*-Functions and Random Matrices. In *Mathematics Unlimited 2001 and Beyond*; Enquist, B., Schmid, W. Eds.; Springer-Verlag: Berlin, 2001; 331–352.

[23] Conrey, B.; Farmer, D.; Keating, J.; Rubinstein, M.; Snaith, W., Correlation of Random Matrix Polynomials. Technical Report, American Institute of Mathematics, 2002.

[24] Coram, M.; Diaconis, P., New Tests of the Correspondence Between Unitary Eigenvalues and the Zeros of Riemann's Zeta Function. Jour. Phys. A. **2002**, to appear.

[25] Daley, D.; Verre-Jones, D., *An Introduction to the Theory of Point Processes.* Springer-Verlag: New York, 1988. MR **90e:**60060

[26] Deift, P., Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach. Courant Lecture Notes #3, NYU/Courant Institute: New York, and Amer. Math. Soc.: Providence, RI, 1999. MR **2000g:**47048

[27] Deift, P., Integrable Systems and Combinatorial Theory. Notices, Amer. Math. Soc. **2000**, 47, 631–640. MR **2001g:**05012

[28] Diaconis, P., *Group Representations in Probability and Statistics.* Ins. Math. Statist., Hayward, CA, 1986. MR **90a:**60001

[29] Diaconis, P., Applications of the Method of Moments in Probability and Statistics. In *Moments in Mathematics*; Landau, H., Ed.; Amer. Math. Soc.: Providence, RI, 1987; 125–142. MR **89m:**60006

[30] Diaconis, P.; Shahshahani, M., Products of Random Matrices as They Arise in the Study of Random Walks on Groups. Contemp. Math. **1986**, *50*, 183–195. MR **87k:**60025

[31] Diaconis, P.; Shahshahani, M., The Subgroup Algorithm for Generating Uniform Random Variables. Prob. Eng. and Info. Sci. **1987**, *1*, 15–32.

[32] Diaconis, P.; Shahshahani, M., On the Eigenvalues of Random Matrices. In *Studies in Applied Probablility*; Gani, J., Ed.; Jour. Appl. Probab.: Special Vol. 31A, 1994; 49–62. MR **95m:**60011

[33] Diaconis, P.; Evans, S., Linear Functionals of Eigenvalues of Random Matrices. Transactions Amer. Math. Soc. **2001**, *353*, 2615–2633. MR **2002d:**60003

[34] Diaconis, P.; Evans, S., Immanants and Finite Point Processes. Jour. Combin. Th. A. **2000**, *91*, 305–321. MR **2001m:**15018

[35] Diaconis, P.; Evans, S., A Different Construction of Gaussian Fields from Markov Chains: Dirichlet Covariances. Ann. Inst. Henri Poincaré, **2002**, to appear.

[36] Diaconis, P.; Freedman, D., A Dozen deFinetti-Style Results in Search of a Theory. Ann. Inst. Henri Poincaré, **1987**, *23*, 397–423. MR **88f:**60072

[37] Diaconis, P.; Eaton, M.; Lauritzan, S., Finite deFinetti Theorems in Linear Models and Multivariate Analysis. Scand. Jour. Statist. **1992**, *19*, 289–315. MR **94g:**60065

[38] Dyson, F., Statistical Theory of the Energy Levels of Complex Systems, I, II, III. J. Math. Phys. **1962**, *3*, 140–156, 157–165, 166–175. MR **26:**1111, MR **26:**1112, MR **26:**1113

[39] Dyson, F., Correlations Between Eigenvalues of a Random Matrix. Comm. Math. Phys. **1970**, *19*, 235–250. MR **43:**4398

[40] Eaton, M., *Multivariate Statistics*; Wiley: New York, 1983. MR **86i:**62086

[41] Edelman, A.; Kostlan, E.; Shub, M., How Many Eigenvalues of a Random Matrix Are Real? Jour. Amer. Math. Soc. **1994**, *7*, 297–267. MR **94f:**60053

[42] Forrester, P.; Rains, E., Inter-Relationships Between Orthogonal, Unitary and Symplectic Matrix Ensembles. MSRI Publications **2001**, *40*, 171–207. MR **2002h:**82008

[43] Fulman, J., Random Matrix Theory Over Finite Fields. Bull. Amer. Math. Soc. **2002**, *39*, 51–86. MR **2002i:**60012

[44] Fulton, W., Eigenvalues, Invariant Factors, Highest Weights and Schubert Calculus. Bull. Amer. Math. Soc. **2000**, *37*, 209–249. MR **2001g:**15023

[45] Fyodorov, Y.; Khoruzhenko, B.; Sommers, H., Universality in the Random Matrix Spectra in the Regime of Weak Non-Hermiticity. Ann. Inst. Henri Poincaré: Physique Théorique **1998**, *68*, 440–489. MR **99i:**60080

[46] Goldsheid, I.; Khoruzhenko, B., Eigenvalue Curves of Asymmetric Tri-Diagonal Random Matrices. Electronic Jour. Probab. **2000**, *5*, Paper 16. MR **2002j:**82061

[47] Goodman, R.; Wallach, W., *Representations and Invariants of the Classical Groups*. Cambridge Press: Cambridge, 1998. MR **99b:**20073

[48] Grenander, U.; Szegö, G., *Toeplitz Forms and Their Applications*. University of California Press: Berkeley, 1958. MR **20:**1349

[49] Haake, F., Secular Determinants of Random Unitary Matrices. Jour. Pys. A. **1996**, *29*, 3641–3658. MR **97g:**82002

[50] Haake, F., *Quantum Signatures of Chaos*, 2nd Ed.; Springer-Verlag: Berlin, 2001.

[51] Hanlon, P.; Stanley, R.; Stembridge, J., Some Combinatorial Aspects of the Spectra of Normally Distributed Random Matrices. Contemp. Math. **1992**, *138*, 151–174. MR **93j:**05164

[52] Hirschman, I., The Strong Szegö Limit Theorem for Toeplitz Determinants. Amer. Jour. Math. **1966**, *88*, 577–614. MR **35:**2064

[53] Hughes, C.; Keating, J.; O'Connell, W., On the Characteristic Polynomial of a Random Unitary Matrix. Comm. Math. Phys. **2001**, *220*, 429–451. MR **2002m:**82028

[54] Hughes, C.; Rudnick, Z., Mock-Gaussian Behavior for Linear Statistics of Classical Compact Groups. Department of Mathematics, Tel Aviv University, 2002, preprint.

[55] Johansson, K., On Szegö's Asymptotic Formula for Toeplitz Determinants and Generalizations. Bull. Sc. Math. **1988**, *112*, 257–304. MR **89m:**47021

[56] Johansson, K., On Random Matrices from the Compact Classical Groups. Ann. Math. **1997**, *145*, 519–545. MR **98e:**60016

[57] Johansson, K., The Longest Increasing Subsequence in a Random Permutation and a Unitary Random Matrix Model. Math. Res. Lett. **1998**, *5*, 63–82. MR **99e:**60033

[58] Johnstone, I., On the Distribution of the Largest Eigenvalue in Principal Component Analysis. Ann. Statist. **2001**, *29*, 295–327. MR **2002i:**62115

[59] Katz, N.; Sarnak, P., *Random Matrices, Frobenius Eigenvalues, and Monodromy.* Amer. Math. Soc.: Providence, RI, 1999. MR **2000b:**11070

[60] Keating, J.; Snaith, N., Random Matrix Theory and $\xi(\frac{1}{2}+it)$. Commun. Math. Phys. **2000**, *214*, 57–89. MR **2002c:**11107

[61] Keating, J.; Snaith, N., Random Matrix Theory and L-Functions at $s=\frac{1}{2}$. Commun. Math. Phys. **2000**, *214*, 91–110. MR **2002c:**11108

[62] Kiessling, M.; Spohn, H., A Note on the Eigenvalue Density of Random Matrices. Comm. Math. Phys. **1999**, *199*, 638–695. MR **2000a:**82031

[63] Macchi, O., Stochastic Processes and Multicoincidences. IEEE Transactions **1971**, *17*, 1–7.

[64] Macchi, O., The Coincidence Approach to Stochastic Point Processes. Adv. Appl. Probab. **1975**, *7*, 83–122. MR **52:**1876

[65] MacDonald, I., *Symmetric Functions and Hall Polynomials*, 2nd Ed.; Clarendon Press: Oxford, 1995. MR **96h:**05207

[66] Marchenko, V.; Pastur, L., Distribution of Some Sets of Random Matrices. Mat. Sb. **1967**, *1*, 507–536.

[67] Mardia, K.; Kent, J.; Bibby, J., *Multivariate Analysis.* Academic Press: New York, 1979. MR **81h:**62003

[68] Mehta, M., *Random Matrices*, 2nd Ed.; Acad. Press: New York, 1991. MR **92f:**82002

[69] Mezzadri, F., Random Matrix Theory and the Zeros of $\xi'(s)$. Dept. of Mathematics, University of Bristol, **2002**, preprint.

[70] Muirhead, R., Latent Roots and Matrix Variates: A Review of Some Aymptotic Results. Ann. Statist. **1978**, *6*, 5–33. MR **56:**16919

[71] Odlyzko, A., On the Distribution of Spacings Between Zeros of the Zeta Function. Math. Comp. **1987**, *48*, 273–308. MR **88d:**11082

[72] Odlyzko, A., The 10^{20}-th Zero of the Riemann Zeta Function and 175 Million of Its Neighbors. ATT Laboratories, 1992, preprint.

[73] O'Connell, N., Random Matrices, Non-Colliding Processes and Queues. Laboratoire de Probabilites, Paris 6, 2002, preprint.

[74] O'Connell, N.; Yor, M., Brownian Analogues of Burke's Theorem. Stoch. Proc. Appl. **2001**, *96*, 285–304. MR **2002h:**60175

[75] Okounkov, A., Random Matrices and Random Permutations. Math. Res. Notices **2000**, *20*, 1043–1095. MR **2002c:**15045

[76] Olshansky, G., An Introduction to Harmonic Analysis on the Infinite-Dimensional Unitary Group. University of Pennsylvania, Dept. of Mathematics, 2001, preprint.

[77] Olshanski, G.; Vershik, A., Ergodic Unitarily Invariant Measures on the Space of Infinite Hermitian Matrices. In *Contemporary Mathematical Physics*; Amer. Soc. Transl. Ser. 2, 1996, *175*, 137–175. MR **98e:**28015

[78] Pickrell, D., Mackey Analysis of Infinite Classical Motion Groups. Pacific Jour. **1991**, *150*, 139–166. MR **92g:**22041

[79] Porod, U., The Cut-Off Phenomenon for Random Reflections. Ann. Probab. **1996**, *24*, 74–96. MR **97e:**60012

[80] Rains, E., High Powers of Random Elements of Compact Lie Groups. Probab. Th. Related Fields *107*, 219–241. MR **98b:**15026

[81] Rains, E., Images of Eigenvalue Distributions Under Power Maps. ATT Laboratories, 1999, preprint.

[82] Rains, E., *Probability Theory on Compact Classical Groups.*, Harvard University: Department of Mathematics, 1991, Ph.D. thesis.

[83] Rosenthal, J., Random Rotations, Characters and Random Walks on SO(N). Ann Probab. **1994**, *22*, 398–423. MR **95c:**60008

[84] Sinai, Y.; Soshnikov, A., Central Limit Theorem for Traces of Large Random Symmetric Matrices with Independent Matrix Elements. Bol. Soc. Brasil. Mat. (N.S.) **1998**, *29*, 1–24. MR **99f:**60053

[85] Sloane, N., Encrypting by Random Rotations. Technical Memorandum, Bell Laboratories, 1983.

[86] Soshnikov, A., The Central Limit Theorem for Local Linear Statistics in Classical Compact Groups and Related Combinatorial Identities. Ann. Probab. **2000**, *28*, 1353–1370. MR **2002f:**15035

[87] Soshnikov, A., Level Spacings Distribution for Large Random Matrices: Gaussian Fluctuations. Ann. Math. **1998**, *148*, 573–617. MR **2000f:**15014

[88] Soshnikov, A., Determinantal Random Point Fields. Russian Math. Surveys **2000**, *55*, 923–975. MR **2002f:**60097

[89] Stanley, R., *Enumerative Combinatorics. Vol. 2*; Cambridge University Press: Cambridge, 1999. MR **2000k:**05026

[90] Tracy, C.; Widom, H., Introduction to Random Matrices. In *Geometric and Quantum Aspects of Integrable Systems*; Springer-Verlag: Berlin, 1993, 103–130. MR **95a:**82050

[91] Tracy, C.; Widom, H., Random Unitary Matrices, Permutations and Painlevé. Comm. Math. Physics **1999**, *207*, 665–685. MR **2001h:**15019

[92] Tracy, C.; Widom, H., On the Relations Between Orthogonal, Symplectic and Unitary Ensembles. Jour. Statist. Phys. **1999**, *94*, 347–363.

[93] Tracy, C.; Widom, H., Universality of the Distribution Functions of Random Matrix Theory. CRM Proceedings **2000**, *26*, 251–264. MR **2002f:**15036

[94] Tracy, C.; Widom, H., On the Limit of Some Toeplitz-Like Determinants. SIAM J. Matrix Anal. Appl. **2002**, *23*, 1194–1196.

[95] Voiculescu, D., Lectures on Free Probability Theory. Springer Lecture Notes in Mathematics **2000**, *1738*, 279–349. MR **2001g:**46121

[96] Wieand, K., *Eigenvalue Distributions of Random Matrices in the Permutation Group and Compact Lie Groups*, Harvard University: Department of Mathematics, 1998, Ph.D. thesis.

[97] Wieand, K., Eigenvalue Distributions of Random Permutation Matrices. Ann. Probab. **2000**, *28*, 1563–1587. MR **2002d:**15027

Department of Mathematics and Statistics, Stanford University, Stanford, CA 94305
E-mail address: diaconis@math.stanford.edu