Probab. Theory Relat. Fields 107, 219–241 (1997)

Probability Theory and Related Fields

High powers of random elements of compact Lie groups

E. M. Rains *

Harvard University, Cambridge, MA 02138, USA (e-mail: rains@ccr-p.ida.org)

Received: 15 October 1995 / In revised form: 7 March 1996

Summary. If a random unitary matrix U is raised to a sufficiently high power, its eigenvalues are exactly distributed as independent, uniform phases. We prove this result, and apply it to give exact asymptotics of the variance of the number of eigenvalues of U falling in a given arc, as the dimension of U tends to infinity. The independence result, it turns out, extends to arbitrary representations of arbitrary compact Lie groups. We state and prove this more general theorem, paying special attention to the compact classical groups and to wreath products. This paper is excerpted from the author's doctoral thesis, [9].

Mathematics Subject Classifications (1991): 60B15, 22E99

Introduction

Suppose one were given a random unitary matrix U (Haar-distributed), and wished to know how one should expect U^n to behave, for n large. In particular, how are its eigenvalues distributed? If n were much larger than the dimension of U, one might reasonably expect that the eigenvalues of U^n should be very nearly independent and uniformly distributed. It turns out, in fact, that much more can be said: for n sufficiently large (greater than or equal to the dimension), the eigenvalues are *exactly* independent and uniformly distributed. This result (Theorem 1.1 below) generalizes, with suitable modifications, to arbitrary representations of arbitrary compact Lie groups.

Theorem 1.1 was discovered as a result of the author's attempt to understand the following fact: the eigenvalues of a random unitary matrix are unusually regularly spaced in the unit circle. This is a quite visible effect; for example,

* *Present address:* Center for Communications Research, Thanet Rd., Princeton, NJ 08540, USA

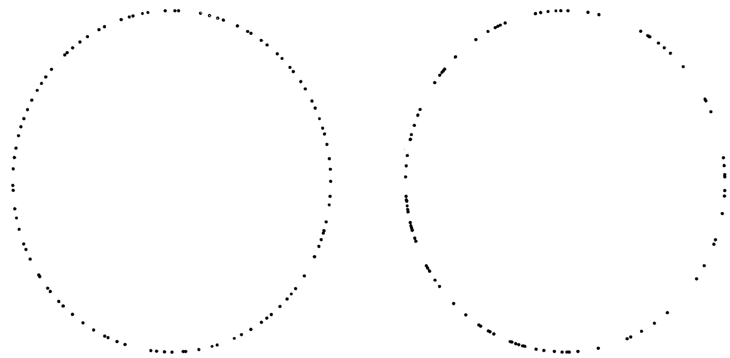

Fig. 1. a Eigenvalues of a random $U \in U(100)$. **b** 100 independent uniform points in S^1

Fig. 1a plots the eigenvalues of a random 100×100 unitary matrix, while Fig. 1b is a plot of 100 points chosen at random on the unit circle, independently and uniformly; Fig. 1a and b is noticeably more uniform.

In attempting to understand this fact, it seemed natural to investigate U^n for U a random $n \times n$ unitary matrix, in the hopes that there might be some enlightening structure to be found. This turned out to be anything but the case; in a sense, Theorem 1.1 asserts that there is as little structure as possible. The regularity of the eigenvalues is thus entirely a consequence of the structure at lower powers (together with a lack of "bad" structure at high powers). However, this lack of structure allows a number of calculations with high powers of U to be made very easily. Section 1 uses these calculations, together with some corresponding results on low powers of U from [5] to produce asymptotic quantitative results on the regularity of the eigenvalues; in a sense which will be made clear in Sect. 1, the irregularity of the eigenvalues of U is $O(\sqrt{\log n})$, while n uniform, independent, points have an $O(\sqrt{n})$ irregularity. The corresponding calculations could in principle be carried out for the other classical groups, using the appropriate versions of Theorem 1.1.

Considering the usefulness of the result on high powers of U, it was natural to look for a generalization to other compact Lie groups. Again, one has that high powers have "as little structure as possible"; the main difficulty was in figuring out how to state that formally. To give an idea as to the possible complications, consider the orthogonal group $O(2n+1)$. The eigenvalues of a typical member of this group come in n conjugate pairs, plus one eigenvalue left over, either 1 or -1. Clearly, this structure will remain unchanged for O^m, no matter how large m is taken to be. However, other than this structure, nothing else remains, for sufficiently large m; the conjugate pairs are uniformly distributed and independent, both from each other and from the ± 1 eigenvalue. Section 2 states and proves the appropriate generalization, which applies to any compact Lie group. Section 3 explores some of the consequences of Theorem 2.1; in particular, it applies the theorem to the other classical groups and to wreath products.

Section 0 is provided as a review of the relevant background material; in particular, Haar measure is defined, and algorithms for generating from Haar measure on the classical groups are given.

This paper is Sects. 1 through 3 of the author's doctoral thesis [9], plus appropriate excerpts of Sect. 0.

0 Background overview

Lie groups

We recall some facts about Lie groups.

Definition 0.1 *A* **Lie group** *is a smooth manifold G with a group structure such that the multiplication map $m : G \times G \to G$ and the inverse map $i : G \to G$ are smooth.*

Examples include $\mathbb{R}^n$, $\mathbb{C}^n$. Less trivial examples include $GL(n, \mathbb{R})$, the general linear group (invertible $n \times n$ real matrices) and $GL(n, \mathbb{C})$. Furthermore, any closed subgroup of a Lie group is also a Lie group.

One of the first constructions of Lie theory is that of the *Lie algebra* $\mathscr{L}(G)$ associated with a Lie group G. Topologically, $\mathscr{L}(G)$ is just the tangent space to G at the identity. However, there is also an induced algebraic structure, which determines much of the properties of G. First, there is an action of G on $\mathscr{L}(G)$. For any element g of G, there is a function $C_g : h \mapsto ghg^{-1}$. Taking the derivative w.r.to h at the identity, we get a linear transformation from $\mathscr{L}(G)$ to itself, denoted by $\mathrm{Ad}(g)$. Further, it is clear that $\mathrm{Ad}(g)\,\mathrm{Ad}(h) = \mathrm{Ad}(gh)$ for arbitrary g and $h \in G$; thus Ad gives a representation of G, known as the *adjoint representation*. Now, take the derivative of Ad at the identity. This gives a linear transformation from $\mathscr{L}(G)$ to $\mathscr{L}(GL(\mathscr{L}(G)))$. The Lie algebra of $GL(V)$ is fairly easily seen to be the space of linear transformations on V; thus, Ad associates to any element x of $\mathscr{L}(G)$ a linear transformation $\mathrm{Ad}(x)$ on $\mathscr{L}(G)$, defined by

$$\mathrm{Ad}(x)y = \frac{d}{dt}\left(\mathrm{Ad}(f_x(t))y\right)_{t=0},$$

where the derivative of $f_x(t)$ at 0 is x. This map satisfies

$$\mathrm{Ad}(\mathrm{Ad}(g)x) = \mathrm{Ad}(g)\,\mathrm{Ad}(x)\,\mathrm{Ad}(g)^{-1}$$

and

$$\mathrm{Ad}(x)y = -\,\mathrm{Ad}(y)x\,.$$

This motivates the definition of the operation $[x, y]$ on $\mathscr{L}(G)$ (known as the *Lie bracket*), by

$$[x, y] = \mathrm{Ad}(x)y = -[y, x].$$

This satisfies the identity

$$\left[x,[y,z]\right]+\left[y,[z,x]\right]+\left[z,[x,y]\right]=0,$$

known as the Jacobi identity. It is fairly easy to see that any smooth homomorphism between Lie groups G and H preserves this algebraic structure on their respective Lie algebras. It is a theorem that a homomorphism between connected Lie groups is determined by the induced homomorphism on their Lie algebras; see, for example, Theorem 1 in Sect. III.4 of [4].

Example If G is $GL(n,\mathbb{R})$, then the Lie algebra of G is the space of linear transformations on $\mathbb{R}^n$. The Lie bracket operation is given by $[A,B]=AB-BA$; similarly for $GL(n,\mathbb{C})$. It follows that if G is any Lie group, and R is a representation of G, then $R([A,B])=R(A)R(B)-R(B)R(A)$.

In the matrix framework, there is a function exp from the Lie algebra to the group, given by

$$\exp(A)=\sum_k \frac{A^k}{k!};$$

equivalently, it is the solution to the differential equation

$$\frac{d}{dt}\exp(tA)=A\exp(tA),$$

with boundary condition $\exp(0)=1$. This extends to the general case, giving a unique diffeomorphism exp from $\mathscr{L}(G)$ to G such that the derivative of exp at 0 is the identity transformation on $\mathscr{L}(G)$, and such that $\exp(x+y)=\exp(x)\exp(y)$ whenever $[x,y]=0$ (see Theorems 2.6 and 2.9 in [1]); this map is known as the exponential map.

Classical groups

The best known examples of Lie groups are given by the classical groups. These are compact matrix groups, defined as follows. First, we have the unitary group, $U(n)$. This is the subgroup of unitary matrices in $GL(n,\mathbb{C})$; that is, the group of $n\times n$ complex matrices such that $U^\dagger U=1$, where $A^\dagger$ is $\overline{A^t}$.

Theorem 0.2 *$U(n)$ is a closed subgroup of $GL(n,\mathbb{C})$, so is a Lie group. The Lie algebra of $U(n)$ is given by the matrices such that $A^\dagger=-A$ (anti-Hermitian matrices).*

Proof The map $A\mapsto A^\dagger A$ is polynomial, thus continuous, so $U(n)$ must be closed. That it is a subgroup follows from the facts that $1^\dagger 1=1$, that

$$(A^\dagger A)^{-1}=(A^{-1})^\dagger A^{-1},$$

and that, for $U\in U(n)$,

$$(UA)^\dagger(UA)=A^\dagger U^\dagger UA=A^\dagger A.$$

Finally, suppose $f(t)$ is a function $\mathbb{R} \to U(n)$, $f(0) = 1$. Then

$$f(t)^\dagger f(t) = 1.$$

Differentiating both sides, we get

$$f'(0)^\dagger + f'(0) = 0.$$

Thus if $A \in \mathscr{L}(U(n))$, $A^\dagger = -A$. The converse follows by taking $f(t) = \exp(tA)$, where $A^\dagger = -A$. QED

Remark The condition $U^\dagger U = 1$ states that the columns of U form an orthonormal basis of $\mathbb{C}^n$; conversely, any such (ordered) basis gives a unique element of $U(n)$.

Next, we have the orthogonal group, $O(n)$. This is defined as the group of real unitary $n \times n$ matrices.

Theorem 0.3 *$O(n)$ is a closed subgroup of $U(n)$, so is a Lie group. The Lie algebra of $O(n)$ is given by the real antisymmetric matrices.*

Proof Analogous.

Remark Similarly, $O(n)$ is bijective with the set of ordered orthonormal bases of $\mathbb{R}^n$.

Finally, we have the symplectic group. Let J be a real antisymmetric matrix satisfying $J^2 = -1$; which one is chosen is a matter of convention. Note that the eigenvalues of J must be $\pm i$, with equal multiplicity; thus such a J only exists in even dimensions. Also, by orthogonal change of basis, we can take any one such J into any other. Then the symplectic group $Sp(2n)$ is the group of unitary $2n \times 2n$ matrices such that $U^t JU = J$. Note that

$$(OUO^t)^t OJO^t (OUO^t) = O(U^t JU)O^t = OJO^t,$$

so a change in our convention for J gives an isomorphic group.

Theorem 0.4 *$Sp(2n)$ is a closed subgroup of $U(2n)$, so is a Lie group. The Lie algebra of $Sp(2n)$ is given by the anti-Hermitian matrices such that $A^t J + JA = 0$.*

Proof Analogous.

Remark $Sp(2n)$ is isomorphic to the subgroup of $GL(n, \mathbb{H})$ (where $\mathbb{H}$ is the quaternions) satisfying $S^\dagger S = 1$. The definition above corresponds to an imbedding of $GL(n, \mathbb{H})$ in $GL(2n, \mathbb{C})$ as the matrices such that $\overline{A} = -JAJ$.

Haar measure

Let G be a locally compact topological group. A left-invariant measure on G is a measure such that $\mu(gA) = \mu(A)$ for any measurable set A (we take as our σ-field the Borel σ-fnield).

Theorem 0.5 *Let G be a locally compact topological group. There exists a left-invariant measure μ on G such that $0 < \mu(A) < \infty$ for some open subset A of G. Furthermore, if μ' is any other such measure, $\mu' \propto \mu$.*

Proof See chapter 4 of [6].

This measure is known as Haar measure, or, to be precise, *left* Haar measure; there is clearly an analogous right Haar measure. If G is compact, we can say more:

Corollary 0.6 *Let G be a compact topological group. There exists a unique left-invariant probability measure on G.*

Proof Let μ be a left-invariant measure on G, and let A be an open subset of G such that $0 < \mu(A) < \infty$. G is covered by the translates of A; by compactness, there exists a finite subcover. But, then, if n is the size of the finite subcover, we have

$$\mu(A) \leq \mu(G) \leq n\mu(A),$$

by subadditivity. Then $\mu/\mu(G)$ gives a probability measure on G. Clearly, if $\mu' \propto \mu$, it gives the same probability measure on G; thus it is unique. QED

In the sequel, Haar measure will always refer to this probability measure.

Remark If G is compact Lie, the left Haar measure is equal to the right Haar measure.

For the classical groups, we can give explicit constructions of Haar measure. First, the unitary group:

Theorem 0.7 *Let $X_{\mathbb{C}}$ be a $n \times n$ matrix filled with i.i.d. complex standard normals. Then the matrix obtained by applying Gram-Schmidt to the columns of $X_{\mathbb{C}}$ is distributed according to Haar measure on $U(n)$.*

Proof First, note that applying Gram-Schmidt to the columns of M gives $M\Gamma$, for some matrix Γ. Further, Γ depends only on the inner products of the columns of M. Thus Gram-Schmidt will produce the same Γ for UM, where U is any unitary matrix. Now, the distribution of $X_{\mathbb{C}}$ is easily seen to be invariant under $X_{\mathbb{C}} \to UX_{\mathbb{C}}$ for any unitary U; by the above comments, the same applies to the output of Gram-Schmidt on $X_{\mathbb{C}}$. But then the resulting distribution is a left-invariant measure on $U(n)$, so must equal Haar measure. QED

Similarly, for the orthogonal group, we have:

Theorem 0.8 *Let $X_{\mathbb{R}}$ be a $n \times n$ matrix filled with i.i.d. real standard normals. Then the matrix obtained by applying Gram-Schmidt to the columns of $X_{\mathbb{R}}$ is distributed according to Haar measure on $O(n)$.*

Proof Analogous.

For the symplectic group, the main difficulty is extending Gram-Schmidt. It is easiest, in this case, to think of $Sp(2n)$ as $n \times n$ quaternionic matrices such that $S^\dagger S = 1$; where $S^\dagger$ is the conjugate of the transpose of S. If v and w are quaternionic vectors, define $\langle v, w\rangle$ by $v^\dagger w$, analogous to the definition of the standard Hermitian inner product. Note, in particular, that the columns of S are orthonormal with respect to this inner product. Thus, given a suitable extension of Gram-Schmidt, Theorems 0.7 and 8 will extend.

To extend Gram-Schmidt, one must simply be careful about order of multiplication. First, the norm of a vector is real:

$$\overline{v^\dagger v} = \sum_i \overline{\overline{v_i} v_i} = \sum_i \overline{v_i} v_i = v^\dagger v,$$

where the middle equality follows from the fact that conjugation is an anti-automorphism of the quaternions. Thus, we can safely divide a vector by its norm. Next, consider the vector

$$v_1 - v_0\langle v_0, v_1\rangle,$$

where $|v_0| = 1$. Then

$$\langle v_0, v_1 - v_0\langle v_0, v_1\rangle\rangle = \langle v_0, v_1\rangle - \langle v_0, v_0\rangle\langle v_0, v_1\rangle = 0.$$

And similarly for the later stages of Gram-Schmidt. The proof of Theorem 0.7 carries over directly, and we have

Theorem 0.9 *Let $X_{\mathbb{H}}$ be a $n \times n$ matrix filled with i.i.d. quaternionic standard normals. Then the matrix obtained by applying Gram-Schmidt to the columns of $X_{\mathbb{H}}$ is distributed according to Haar measure on $Sp(2n)$.*

1 Asymptotic regularity of the eigenvalue distribution on $U(n)$

Given an $n \times n$ unitary matrix U, there is an associated probability distribution on the unit circle, produced by putting a mass of $\frac{1}{n}$ at the point on the unit circle corresponding to each eigenvalue (alternatively, the distribution is that of a randomly chosen eigenvalue, if each eigenvalue is equally likely). If U is chosen with Haar measure from $U(n)$, then this associated distribution tends to be surprisingly “evenly spaced”.

There are many different ways in which one can quantify such a regularity condition. One can, for instance, study the size of the shortest interval between eigenvalues, and show that it tends to be close to $\frac{1}{n}$ (see, for example, [10]), or other results of this flavor. Such expressions, however, have the disadvantage of being fairly difficult to compute; some sort of $\mathscr{L}^2$ style result is easier. Thus, I will consider the following quantity:

$$R_\alpha(U) = \frac{1}{2\pi}\int_0^{2\pi}\left(\left|\{i \mid \theta_i \in [\theta-\alpha, \theta+\alpha]\}\right| - \frac{n\alpha}{\pi}\right)^2 d\theta, \tag{1.1}$$

where θ_i is the angular coordinate of the ith eigenvalue, and $\alpha \in [0, \pi]$. $\frac{n\alpha}{\pi}$ is the average number of eigenvalues falling in a randomly chosen arc of length 2α; thus, $R_\alpha(U)$ is the variance of the number of eigenvalues falling in such an arc. Note that $R_\alpha(U)$ attains its minimum value when the eigenvalues are exactly regularly spaced (this minimum value is 0 if α is an integer multiple of $\frac{\pi}{n}$); its maximum value is

$$n^2\frac{\alpha}{\pi}\left(1-\frac{\alpha}{\pi}\right),$$

attained when U is a multiple of the identity. This clearly is a measure of "clumping" of eigenvalues; moreover, it is a quadratic formula, thus much easier to compute with than, say, the length of the shortest spacing. Furthermore, since $R_\alpha(U)$ is a positive random variable, $E(R_\alpha(U))$ alone can give us fairly good bounds on tails of the distribution of $R_\alpha(U)$.

As an example, suppose the eigenvalues of U were i.i.d. uniform. In this case, the random variable $\left|\{i \mid \theta_i \in [\theta-\alpha, \theta+\alpha]\}\right|$ is the sum of n i.i.d. random variables (1 if $\theta_i \in [\theta-\alpha, \theta+\alpha]$ and 0 otherwise). As the variance of those variables is

$$\frac{\alpha}{\pi} - \left(\frac{\alpha}{\pi}\right)^2,$$

$E(R_\alpha)$, the variance of their sum, is:

$$E(R_\alpha) = n\frac{\alpha}{\pi}(1-\frac{\alpha}{\pi}). \tag{1.2}$$

Now, let us consider $E(R_\alpha(U))$ in the case of Haar measure on $U(n)$. Taking expectations allows us to throw away the integral (by phase-invariance of Haar measure); we are left with the variance of the number of eigenvalues of U falling in $[-\alpha, \alpha]$. This is an interesting quantity in its own right; although it is equal for the unitary group to R_α, this equality fails for the other classical groups (for which the average number of eigenvalues in $[-\alpha, \alpha]$ is *not* $\frac{n\alpha}{\pi}$). Furthermore, we are also interested in α of the form $\frac{\pi}{n}\beta$, as n goes to infinity, with fixed β (physicists, for instance, are interested in the limiting eigenvalue distribution scaled up by n in the limit); we can expect different asymptotic behavior in this limit.

With this in mind, let us first work out $R_\alpha(U)$ for the general case (not using translation invariance). We first note that the quantity being squared in the integrand is a sum of indicator functions of the form $\left[\theta \in [\theta_i - \alpha, \theta_i + \alpha]\right]$; we can expand this into a double sum, and take the summation out of the integral, giving us:

$$R_\alpha(U) = \sum_{1\le i,j\le n} \frac{1}{2\pi} \int_0^{2\pi} \left[\theta \in [\theta_i - \alpha, \theta_i + \alpha]\right] \left[\theta \in [\theta_j - \alpha, \theta_j + \alpha]\right] d\theta - \left(\frac{n\alpha}{\pi}\right)^2$$

We can then replace the indicator functions by their Fourier expansions, and use Parseval's theorem; after working through the algebra, we are left with

$$\begin{aligned} R_\alpha(U) &= \sum_{1\le i,j\le n} \left(\frac{1}{\pi^2} \sum_{1\le k} \frac{1}{k^2} \sin^2(k\alpha)\left(\lambda_i^k \overline{\lambda_j^k} + \overline{\lambda_i^k}\lambda_j^k\right)\right) \\ &= \frac{1}{\pi^2} \sum_{1\le k} (1 - \cos(2k\alpha)) \frac{|p_k|^2}{k^2}, \end{aligned}$$

where p_k is the sum of the kth powers of the eigenvalues (a.k.a. $\mathrm{Tr}(U^k)$).

To compute $E(R_\alpha(U))$, we thus need information about the second moments of the p_k. A result by Diaconis and Shahshahani ([5]) tells us (among other things) that for $U(n)$, $E(|p_k|^2) = k$, for $k \le n$. It thus remains to determine the moments for $k > n$. To determine this, we need only note that the density for Haar measure on $U(n)$ is given ([11]) by

$$\Delta = \frac{1}{n!} \prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^2.$$

This is a Laurent polynomial in the λ_i, of degree at most $(n-1)$ in any given λ_i. Now, consider the density of the joint eigenvalue distribution of U^k. If p is any polynomial in the kth powers of the λ_i, the expectation of p depends only on those terms in which the degree of each λ_i is a multiple of k. But then, by the method of moments (see, for example, Sect. 30 in [2]), it follows that the density of the joint eigenvalue distribution of U^k is given by taking Δ, removing all monomials in which one of the λ_i has degree not a multiple of k, then dividing the degree of each λ_i by k in every remaining monomial. But, if $k \ge n$, the only monomial remaining is the constant term, 1. It follows that every moment of the λ_i is the same as if the λ_i were i.i.d. uniform; the method of moments implies that the distributiom must then actually be i.i.d. uniform, proving:

Theorem 1.1 *If U is Haar-distributed from $U(n)$, and k is any integer $\ge n$, then the eigenvalues of U^k are independent and uniformly distributed in S^1.*

(Theorem 1.1 is a special case of Theorem 2.1, which generalizes this to any compact Lie group.)

In particular, if $k \ge n$, then $E(|p_k|^2)$ is the second moment of a sum of n i.i.d. random variables of mean 0 and variance 1; it follows that $E(|p_k|^2) = n$. This gives us the remaining information we need to complete our expression for $E(R_\alpha(U))$; we get

Theorem 1.2 *Let U be Haar-distributed on $U(n)$, and let R_α be defined by (1.1). Then*

$$E(R_\alpha(U)) = \frac{1}{\pi^2} \sum_{1\le k} (1 - \cos(2k\alpha)) \frac{\min(k, n)}{k^2}.$$

It remains now to determine the asymptotics of this formula. Before we do so, it is worth noting that $E(R_\alpha(U))$ for $U(n)$ must be $\leq E(R_\alpha(U))$ for n i.i.d. uniform points in S^1 (It is easy to see that in that case, we get the same formula, with $\min(k,n)$ replaced by n).

To figure out an asymptotic expansion of this sum, we first split the sum into two sums:

$$\frac{1}{\pi^2}\sum_{1\leq k}(1-\cos(2k\alpha))\frac{\min(k,n)}{k^2} = \frac{1}{\pi^2}\sum_{1\leq k\leq n}(1-\cos(2k\alpha))\frac{k-n}{k^2} + \frac{1}{\pi^2}\sum_{1\leq k}(1-\cos(2k\alpha))\frac{n}{k^2}.$$

We can then rewrite the finite sum as a telescoping infinite sum, then combine the sums:

$$\frac{1}{\pi^2}\sum_{1\leq k}\Big((1-\cos(2k\alpha))\frac{1}{k} - (1-\cos(2(k+n)\alpha))\frac{1}{k+n} + (1-\cos(2(k+n)\alpha))\frac{n}{(k+n)^2}\Big).$$

Using trig identities and rearranging, we get

$$\begin{aligned}\frac{1}{\pi^2}\sum_{1\leq k}\Bigg\{ &-\frac{\cos(2k\alpha)-\cos(2(k+n)\alpha)}{k}\\ &+\left[\left(\frac{1}{k}-\frac{1}{k+n}\right)+\frac{n}{(k+n)^2}\right]\\ &-\cos(2n\alpha)\left[\left(\cos(2k\alpha)\left(\frac{1}{k}-\frac{1}{k+n}\right)\right)+\frac{n\cos(2k\alpha)}{(k+n)^2}\right]\\ &+\sin(2n\alpha)\left[\left(\sin(2k\alpha)\left(\frac{1}{k}-\frac{1}{k+n}\right)\right)+\frac{n\sin(2k\alpha)}{(k+n)^2}\right]\Bigg\}.\end{aligned}$$

Note that each expression in brackets is of the form $f(n)+nf'(n)$; once we have an asymptotic expansion for the sum of the first term in the bracket, the asymptotic expansion for the sum of the second term follows easily. Thus, we need to determine the asymptotic behavior of the following formulae:

$$\begin{aligned}A_1 &= \sum_{1\leq k}\left(\frac{1}{k}-\frac{1}{k+n}\right)\\ A_2 &= \sum_{1\leq k}\frac{\cos(2k\alpha)-\cos(2(k+n)\alpha)}{k}\\ A_3 &= \sum_{1\leq k}e^{2ik\alpha}\left(\frac{1}{k}-\frac{1}{k+n}\right).\end{aligned}$$

A_1 is simply H_{n-1}, the $(n-1)$st harmonic number, thus has asymptotic expansion

$$\log(n) + \gamma + \frac{B_1}{n} - \sum_{2\le k<N} \frac{B_k}{kn^k} + O(n^{-N}),$$

where γ is Euler's constant (0.57721 ...), and B_k is the kth Bernoulli number ([7], Eq. (16) in Sect. 1.2.11.2). A_2 can be simplified as follows:

$$\begin{aligned} A_2 &= \sum_{1\le k} \frac{\cos(2k\alpha) - \cos(2(k+n)\alpha)}{k} \\ &= \sum_{1\le k} \frac{e^{2ik\alpha} + e^{-2ik\alpha} - e^{2i(k+n)\alpha} - e^{-2i(k+n)\alpha}}{2k} \\ &= \frac{1}{2}\sum_{1\le k} \frac{e^{2ik\alpha} - e^{2i(k+n)\alpha}}{k} + \frac{1}{2}\sum_{1\le k} \frac{e^{-2ik\alpha} - e^{-2i(k+n)\alpha}}{k} \end{aligned}$$

These two sums are complex conjugates, so this

$$= \Re\Big((1 - e^{2in\alpha}) \sum_{1\le k} \frac{e^{2ik\alpha}}{k}\Big).$$

Now, this sum does not converge absolutely, but by replacing $e^{2ik\alpha}$ with z^k, we get a limit of absolutely converging sums:

$$\begin{aligned} A_2 &= \lim_{z\to e^{2i\alpha}} \Re\Big((1 - e^{2in\alpha}) \sum_{1\le k} \frac{z^k}{k}\Big). \\ &= -\Re\big((1 - e^{2in\alpha}) \log(1 - e^{2i\alpha})\big) \\ &= -\log\big|2\sin(\alpha)\big| + \Re\big(e^{2in\alpha} \log(1 - e^{2i\alpha})\big). \end{aligned}$$

Finally, we have A_3. Again, one can replace $e^{2ik\alpha}$ with z^k, then take the limit as $z \to e^{2i\alpha}$ (allowable by dominated convergence). It is easy to verify that under that substitution, A_3 simplifies to

$$-\log(1-z) - \int_0^\infty \frac{e^{-nt}}{1 - ze^{-t}}\, dt$$

(Compare the Taylor series around z=0). Since the second integral is a Laplace transform, Watson's lemma (stated without proof in [3], p. 253) gives an asymptotic series for the limit as $z \to e^{2i\alpha}$, in terms of the Taylor series of $\frac{1}{1-ze^{-t}}$ around $t = 0$:

$$-\log(1 - e^{2i\alpha}) - \frac{1}{2n} - \sum_{1\le k<N} i^k \cot^{(k-1)}(\alpha)(2n)^{-k} + O(n^{-N});$$

$\cot^{(k-1)}(n)$ is the $(k-1)$st derivative of $\cot(n)$.

Now, we have

$$E(R_\alpha(U)) = \frac{1}{\pi^2}\left(-A_2 + (A_1 + n\frac{d}{dn}A_1) - \Re\left(e^{2in\alpha}(A_3 + n\frac{d}{dn}A_3)\right)\right).$$

Simplifying

$$
\begin{aligned}
A_1 + n\frac{d}{dn}A_1 &= \log(n) + \gamma + \frac{B_1}{n} - \sum_{2\le k<N} \frac{B_k}{kn^k} \\
&\quad + 1 - \frac{B_1}{n} + \sum_{2\le k<N} \frac{B_k}{n^k} + O(n^{-N}) \\
&= \log(n) + \gamma + 1 + \sum_{2\le k<N} \frac{(k-1)B_k}{kn^k} + O(n^{-N}),
\end{aligned}
$$

and

$$
\begin{aligned}
A_3 + n\frac{d}{dn}A_3 &= -\log(1-e^{2i\alpha}) - \frac{1}{2n} - \sum_{1\le k<N} i^k \cot^{(k-1)}(\alpha)(2n)^{-k} \\
&\quad + \frac{1}{2n} + \sum_{1\le k<N} i^k k \cot^{(k-1)}(\alpha)(2n)^{-k} + O(n^{-N}) \\
&= -\log(1-e^{2i\alpha}) + \sum_{2\le k<N} i^k (k-1)\cot^{(k-1)}(\alpha)(2n)^{-k} \\
&\quad + O(n^{-N}),
\end{aligned}
$$

then plugging in and combining terms, we get the following result:

Theorem 1.3 *Let U be uniformly distributed from the unitary group $U(n)$, and let R_α be defined by (1.1). Then, for any fixed N, as $n \to \infty$,*

$$
\begin{aligned}
E(R_\alpha(U)) = \frac{1}{\pi^2}\Big(\log(n) \\
&+ \big(\gamma + 1 + \log\big|2\sin(\alpha)\big|\big) \\
&+ \sum_{2\le k<N} \Big(\frac{(k-1)B_k}{kn^k} \\
&\qquad + \Re(i^k e^{2in\alpha})(1-k)\cot^{(k-1)}(\alpha)(2n)^{-k}\Big) \\
&+ O(n^{-N})\Big).
\end{aligned}
$$

In particular, for $N = 5$, we have:

$$
\begin{aligned}
E(R_\alpha(U)) = \frac{1}{\pi^2}\Big(\log(n) \\
&+ (\gamma + 1 + \log\big|2\sin(\alpha)\big|) \\
&+ (\frac{B_2}{2} - \frac{1}{4}\csc^2(\alpha)\cos(2n\alpha))n^{-2} \\
&- (\frac{1}{2}\cot(\alpha)\csc^2(\alpha)\sin(2n\alpha))n^{-3} \\
&+ (\frac{3B_4}{4} - (\frac{3}{4}\csc^2(\alpha) - \frac{9}{8}\csc^4(\alpha))\cos(2n\alpha))n^{-4} \\
&+ O(n^{-5})\Big).
\end{aligned}
$$

In contrast, R_α for the uniform independent distribution is of order n (as is immediately apparent from Eq. (1.2)). Thus, the eigenvalues of a Haar-distributed random matrix are significantly more regularly distributed than a similar number of independent, uniform random eigenvalues.

The asymptotics of $E(R_\alpha(U))$ for $\alpha = \frac{\beta\pi}{n}$, β constant are relatively easy to compute; the Euler-Maclaurin summation formula applies in this case, to give

$$E(R_{\frac{\beta\pi}{n}}(U)) = \pi^2|\beta| + 1 - cos(2\pi\beta) - 2\pi\beta \int_0^{2\pi\beta} \frac{\sin(t)}{t} dt + \int_0^{2\pi\beta} \frac{1-\cos(t)}{t} dt \\ - \frac{1-\cos(2\pi\beta) - 2\pi^2\beta^2}{12n^2} + O(n^{-4}).$$

The $O(1)$ terms of this are given in [8], A.38; however, the method used there does not appear to extend to give the later terms in the asymptotic expansion. For i.i.d. uniform, we get

$$E(R_{\frac{\beta\pi}{n}}) = n\frac{\beta}{n}(1-\frac{\beta}{n}) = \beta - \beta^2 n^{-1}.$$

2 Uniformity of the eigenvalue distribution of U^n on general compact groups

A key to the results in Sect. 1 was the observation that, because the formula for the density of Haar measure for $U(N)$ is a Laurent polynomial of degree $N-1$ in the eigenvalues, for $n \geq N$, the eigenvalues of U^n are i.i.d. uniform. While it is certainly to be expected that the eigenvalues of U^n should tend to become independent and uniform as $n \to \infty$, it is surprising that they attain exact independence at some point. This phenomenon is in fact not unique to the unitary group, but is true (with some important caveats) for an arbitrary compact Lie group.

For example, consider the special orthogonal group $SO(2N+1)$. The eigenvalues of the generic matrix from this group split into a number of conjugate pairs, with the remaining eigenvalue forced to be 1. Clearly, no matter how high m is, the eigenvalues of O^m (O Haar-distributed from $SO(2N+1)$ can never be i.i.d. uniform. However, if one chooses a representative from each conjugate pair (getting $\mu_1 \dots \mu_n$), one can write the density for Haar measure as a Laurent polynomial of degree $2N-1$ in the μ_i. As a consequence, if one raises a random special orthogonal matrix U to the nth power, where $n > 2N-1$, the μ_i are i.i.d. uniform. Said another way, the law of the eigenvalues of U^n is the same as the law of a particular set of $2N+1$ Laurent monomials in N i.i.d. uniform random phases. ($\{1, \mu_i, \mu_i^{-1}\}$, to be precise)

The other caveat involves non-connected Lie groups. In such a group, the restrictions on the eigenvalues will in general vary from component to component; the number of degrees of freedom can even vary. For instance, consider the orthogonal group $O(2N)$. In the determinant 1 component, the eigenvalues form N conjugate pairs, while in the determinant -1 component, the eigenvalues form only $N-1$ conjugate pairs, with the remaining two eigenvalues 1 and -1. (This can be seen as follows: Since orthogonal matrices are real and unitary, the set of eigenvalues must consist of some number of conjugate pairs (norm 1), some number of 1s, and some number of -1s. Now, since the number of eigenvalues is even ($2N$), either there are an even number of both 1s and -1s (determinant 1), or an odd number of both (determinant -1). Since a pair of 1s is a conjugate pair, and similarly for a pair of -1s, the statement clearly holds.) Thus, the determinant -1 component has one fewer continuous degree of freedom in its eigenvalues. Clearly, then, each component must be considered somewhat separately.

With these caveats in mind, we can now state the following theorem:

Theorem 2.1 *Let L be a compact Lie group, φ a continuous (unitary) representation of L. Then for every connected component C of L, there is a set of Laurent monomials on a finite set of random variables μ_i (where the μ_i are i.i.d. uniform on S^1), and an integer d such that for any integer $n > d$, and any Haar-distributed random variable U on L, the conditional distribution of the set of eigenvalues of $\varphi(U^n)$ given that $U \in C$ is the same as the distribution of the set of monomials. Furthermore, d may be chosen independently of C and of the representation φ.*

It should be noted that this is the largest class of Lie groups on which this could be expected to hold; the compactness condition is necessary for Haar measure to be a probability. If one leaves the probability setting, using the invariant measure, despite the fact that it is infinite on the whole group, then the result appears to be valid to a very limited extent. However, the result hinges on the fact that on a compact group, raising a matrix to some power throws away information about the eigenvalues; if the eigenvalues of the matrix can be determined from the eigenvalues of a power of the matrix, a non-uniform measure can never become uniform. It is unclear to what extent Theorem 2.1 can be extended to general compact groups. (It does hold, in general, in the case of general finite topological groups, if vacuously so: any continuous representation of a finite topological group has discrete image, so must be constant on each component.)

Theorem 2.1 is proved by finding a set of phase variables μ_i that generate the eigenvalues through a set of monomials, then showing that the density of Haar measure can be written as a Laurent polynomial in the μ_i. Then the method of moments easily gives the result.

It should be noted that the proofs of Lemmas 2.3, 2.4, and 2.5 below are generalized from the proofs of Theorems 4.21 and 6.1 in [1].

Suppose we are given a compact Lie group L, and a component C thereof. Further, let T be a maximal torus of L (a torus of L is an abelian subgroup homeomorphic to a torus of some dimension; a maximal torus is a torus not properly contained in any torus). We have the following lemma:

Lemma 2.2 *There is an element $a \in C$ of finite order, such that $aTa^{-1} = T$.*

Proof Let x be an arbitrary element of C, and consider the image of T under conjugation by x, $S = xTx^{-1}$. It is a well known Theorem ([1], Corollary 4.23), that for any two maximal tori S and T of L, there is an element g of the identity component L_e of L such that $gSg^{-1} = T$. Let $a = gx$. By continuity of multiplication, $a \in C$; furthermore, $aTa^{-1} = T$. Now, suppose a is not of finite order, and consider the cyclic subgroup $\langle a \rangle$ generated by a. In particular, first consider its image under the quotient map $L \to L/L_e$. By compactness of L, this image must be finite; therefore, $\langle a \rangle$ intersects L_e in an infinite cyclic subgroup. Consider the subgroup generated by $(\langle a \rangle \cap L_e) \cup T$; this is clearly abelian, and contains T. But then by proposition 4.26 of [1], it must equal T. It follows then that for some n, $a^n \in T$. Now, it is easy to see that there exists $t \in T$ such that $t^n = a^n$. Then $(t^{-1}a)^n = t^{-n}a^n = e$, so $t^{-1}a$ is of finite order. QED

Example Consider, for example, the complement of $SO(2N)$ in $O(2N)$. In this case, T can be chosen to be the subgroup of block-diagonal matrices with 2×2 rotations down the diagonal. In this case, reflection through a coordinate hyperplane gives a suitable a, although this is hardly exhaustive; it could, for example, be composed with an arbitrary orthogonal transformation of finite order in T that fixes the hyperplane.

Now, since the components of L are also the cosets of L_e (for any coset of L_e, there is a homeomorphism of L carrying it to L_e; it follows that it must be a component of L; the converse follows from the fact that the cosets of L_e exhaust L), we can write every element of C in the form xa, where $x \in L_e$. Conjugation by g gives

$$gxag^{-1} = gxg^{-a}a,$$

where a as an exponent stands for the automorphism of L_e induced by conjugation by a. (In the sequel, if an element of L is used in place of an automorphism, the corresponding inner automorphism will be understood). Since the eigenvalues of $\varphi(x)$ are preserved by conjugation, we need to understand the action of $x \mapsto gxg^{-a}$ on L_e, for a as in Lemma 2.2.

Definition 2.3 *Let a be an automorphism of L_e. A maximal a-torus of L_e is a torus in L_e maximal among all tori fixed by a.*

Example Again, consider $O(2N) - SO(2N)$, with a reflection through a coordinate hyperplane. Then a maximal a-torus is given by block-diagonal matrices consisting of $(N - 1)$ 2×2 rotations down the diagonal, followed by two 1s. Note that in the case $N = 1$, the maximal a-torus is 0-dimensional (that is, a point).

Lemma 2.4 *Let a be an automorphism of L of finite order, and let T_a be any maximal a-torus of L_e. Then for any $x \in L_e$, there is some $g \in L_e$ such that $gxg^{-a} \in T_a$.*

Proof The proof follows the proof of Theorem 4.21 in [1]. Given $x \in L_e$, consider the function $f_x : L_e/T_a \to L_e/T_a$ that takes gT_a to xg^aT_a. Since a fixes T_a, this map is clearly well-defined. Now, if $g^{-1}T_a$ is a fixed point of f_x, then $gxg^{-a} \in T_a$, and we are done. We now show that any function in the homotopy class of f_x must have a fixed point, using the Lefschetz fixed point theorem.

Let x_0 be an element of T_a such that x_0^n (where n is the order of a) generates a dense subgroup of T_a. This can be done by choosing x_0 with irrational, incommensurable coordinates. Now, the map f_{x_0} is clearly homotopic to f_x, since L_e is path-connected. f_{x_0} clearly has a fixed point (namely T_a); we now wish to show that it has only a finite number of fixed points. If gT_a is a fixed point of f_{x_0}, then

$$f_{x_0}^n(gT_a) = x_0^n gT_a = gT_a;$$

since x_0 generates a dense subgroup of T_a, we can conclude that $T_ag = gT_a$, so g normalizes T_a. From this, we can conclude that $T_ag = g^aT_a$.

Now, consider $N_a(T_a)$, the group of all $g \in L_e$ such that $T_ag = g^aT_a$. Conjugating both sides by a, we deduce that $N_a(T_a)$ is preserved by conjugation by a. We can also deduce that $N_a(T_a) \subset N(T_a)$, the normalizer of T_a. Now, for $g \in N_a(T_a)$, consider $u = gg^{-a}$. This is clearly in T_a, so is fixed by a. Then

$$\begin{aligned} u^n &= (gg^{-a})^n \\ &= (gg^{-a})(gg^{-a})^a(gg^{-a})^{a^2} \dots (gg^{-a})^{a^{n-1}} \\ &= gg^{-a}g^ag^{-a^2}g^{a^2} \dots g^{-1} \\ &= 1 \end{aligned}$$

Now, elements of order n are discrete in a torus (in fact, the set of elements of order n is finite; there are precisely $\varphi(n)^d$ such elements, where φ is Euler's totient function, and d is the dimension of the torus). Therefore, if g is in the identity component of $N_a(T_a)$, $gg^{-a} = 1$, so $g = g^a$. Furthermore, such a g must induce the trivial automorphism on T_a. If $g \notin T_a$, this would contradict the maximality of T_a. Therefore, the identity component of $N_a(T_a)$ is T_a, and furthermore, the number of cosets of T_a in $N_a(T_a)$ must be finite. Since every fixed point of f_{x_0} is a coset of T_a in $N_a(T_a)$, f_{x_0} has only a finite number of fixed points.

For any hT_a a fixed point of f_0, the map r_h, which takes gT_a to gT_ah, commutes with f_0:

$$\begin{aligned} r_hf_{x_0}r_h^{-1}(gT_a) &= r_hf_0(gT_ah^{-1}) \\ &= r_h(x_0g^aT_ah^{-a}) \\ &= x_0g^aT_ahh^{-a} \\ &= f_{x_0}(gT_a). \end{aligned}$$

Now, r_h takes T_a to hT_a; therefore, the multiplicity of f_{x_0} at hT_a is the same as that at T_a. It thus suffices to show that the multiplicity at T_a is nonzero; this is easily verified. Thus, the Lefschetz number of f_{x_0} is nonzero, so the Lefschetz number of f_x is nonzero, and f_x has a fixed point. QED

Example Consider $O(2N)-SO(2N)$, with a and T_a as in the previous examples. Lemma 2.4 is equivalent to the fact that any $2N$-dimensional orthogonal matrix can be conjugated into block-diagonal form, with the diagonal consisting of $N-1$ 2×2 rotations, followed by a 1 and a -1; this easily follows from the fact that the eigenvalues fall into $N-1$ conjugate pairs, plus a 1 and a -1.

Note that it easily follows from this that for any $x \in C$, there exists a $g \in L_e$ such that $gxg^{-1} \in T_a a$; thus, we can restrict our attention to the eigenvalues of elements of $T_a a$, by conjugating the random x into $T_a a$. This, plus the fact that $[N_a(T_a) : T_a]$ is finite, allows us to make the following definition:

Definition 2.5 *Let $a \in C$ be of finite order, and let T_a be a maximal a-torus. The induced distribution on T_a is the distribution of the following random variable: Pick x at random (uniformly) from C, choose an element of $T_a a$ conjugate to x, then conjugate that element by a (uniform) random element of $N_a(T_a)$.*

The significance of this definition is that the eigenvalues of ta, with t chosen from the induced distribution on T_a, clearly have the same distribution as the eigenvalues of a Haar-distributed element of C.

Now, consider the subalgebra of the Lie algebra of L corresponding to T_a. There is a lattice in this subalgebra given by the inverse image of the identity under the exponential map. (When a is the identity, this lattice is called the "integer lattice" of L_e.) Now, if we choose a set H_i of generators of this lattice, we can express any point in the subalgebra as a linear combination of the generators; this induces a coordinatization of T_a, assigning to each point in T_a an m-tuple of angles θ_i. By exponentiating these angles, we get the desired μ_i. It remains to show that the induced density on T_a is a Laurent polynomial in the μ_i, and that the eigenvalues of a group element in any representation are monomials in the μ_i. Note that the μ_i are not unique; any isomorphism of the lattice will give valid μ_i; this gives a freedom of $SL(r,\mathbb{Z})$, where r is the dimension of T_a.

Example Again we consider $O(2N)-SO(2N)$. The above lattice is generated by matrices H_i $(0 \le i < N-1)$, where $H_i(e_{2j}) = 2\pi\delta_{ij}e_{2j+1}$, and $H_i(e_{2j+1}) = -2\pi\delta_{ij}e_{2j}$ on the standard basis e_j $(0 \le j < 2N)$. This gives $\mu_j = e^{i\theta_j}$, where the jth rotation matrix rotates by θ_j.

It is convenient first to show

Lemma 2.6 *Let φ be a continuous unitary representation of L, let a be an element of C of finite order, and let T_a be a maximal a-torus; let the μ_i be as above. Then there is some set S of Laurent monomials in the μ_i such that for any $x \in C$, the set of eigenvalues of x is the same as S evaluated at any representative of x in $T_a a$.*

Proof By Lemma 2.4, and the invariance of eigenvalues under conjugation, we need only show this for $x \in T_a a$. By definition of T_a, xa^{-1} and a commute, so their representations can be simultaneously diagonalized. Thus, we need only show that the eigenvalues of xa^{-1} can be written as monomials in the μ_i, since multiplication by a will simply change the coefficient of the monomials. Now, the μ_i give an isomorphism between T_a and a product of $\dim(T_a)$ copies of S^1. Factoring the representation that φ induces on T_a through the μ_i, we get a representation of $(S^1)^{\dim(T_a)}$. This representation, then, is a sum of 1-dimensional representations, each of which clearly corresponds to a Laurent monomial in the μ_i; the lemma follows immediately. QED

Now we can show (by a proof analogous to that of Theorem 6.1 in [1]):

Lemma 2.7 *The induced distribution on T_a has density given by a Laurent polynomial in the μ_i.*

Proof Consider the function $f : L_e/T_a \times T_a \to L_e$, given by

$$(g, t) \mapsto g t g^{-a}.$$

By Lemma 2.4, this map is surjective. Furthermore, with probability 1, a random element of L_e has exactly $[N_a(T_a) : T_a]$ inverse images: It suffices to show that the subset of elements of L_e not satisfying this condition is a countable union of lower-dimensional submanifolds. Since the dimension of $L_e/T_a \times T_a$ is equal to that of L_e, it suffices to show that the corresponding inverse image is a countable union of lower-dimensional submanifolds. Clearly, the size of $f^{-1}(f(g, t))$ is independent of g. Now, it is fairly easy to see that if the closure of the cyclic subgroup generated by t is T_a, then $f^{-1}(f(e, t)) = N_a(T_a)/T_a$. But the closure of $\langle t\rangle \neq T_a$ only if t is in a lower-dimensional subtorus of T_a, and there are only countably many such subtori.

Now, if we lift Haar distribution on L through f to $L/T_a \times T_a$ (pick an element of the inverse image at random), then integrate over L/T_a, we clearly get the induced distribution on T_a. With this in mind, we wish to compute the Jacobian of f. The derivative of f can be computed as follows:

$$\begin{aligned} f(g + g d_g, t + t d_t) - f(g, t) &= g d_g t g^{-a} + g t d_t g^{-a} - g t a d_g a^{-1} g^{-a} \\ &= f(g, t)\left[g^a d_t g^{-a} + g^a t^{-1} d_g t g^{-a} - g^a a d_g a^{-1} g^{-a}\right] \\ &= f(g, t) g^a \left[d_t + \mathrm{Ad}(t)(d_g) - \mathrm{Ad}(a)(d_g)\right], \end{aligned}$$

where Ad is the adjoint representation of L (L acting by conjugation on its Lie algebra). Thus

$$|\det(f')| = \left|\det_{\mathscr{L}(L_e/T_a)}(\mathrm{Ad}(t) - \mathrm{Ad}(a))\right|;$$

$\mathscr{L}(L_e/T_a)$ is the subalgebra of the Lie algebra of L corresponding to L_e/T_a. By Lemma 2.6, this determinant (without the absolute value) is a product of Laurent polynomials in the μ_i. Thus, we need only show that the determinant is, in fact, nonnegative real, not identically zero, and the result follows.

As noted in the proof of Lemma 2.6, we can simultaneously diagonalize the images of a and T_a in the adjoint representation; since the adjoint representation is an orthogonal representation, it splits on a and T_a into a direct sum of 1- and 2-dimensional irreducible real representations. Clearly, the determinant we need

to compute is the product of the corresponding determinants restricted to each representation.
Case 1: $v \in \mathscr{L}(L_e/T_a)$ is a basis vector for a 1-dimensional representation. Then $av = \pm v$ and $tv = \pm v$, for all $t \in T_a$. By continuity, then, $tv = v$, for all $t \in T_a$. Now, if $av = v$, then we could add the one-parameter subgroup corresponding to v to T_a, thus contradicting the maximality of T_a. Thus, $av = -v$, and $\det_v(\mathrm{Ad}(t) - \mathrm{Ad}(a)) = 2$.
Case 2: $v, w \in \mathscr{L}(L_e/T_a)$ are basis vectors for a 2-dimensional irreducible representation. Then $\det_{vw}(\mathrm{Ad}(t) - \mathrm{Ad}(a))$ is the product of two numbers complex conjugate to each other, thus is nonnegative. We thus need only show that it is non-zero on every element of T_a whose nth powers (again, n is the order of a) generate a dense subgroup of T_a Suppose x is such an element. If $\det_{vw}(\mathrm{Ad}(x) - \mathrm{Ad}(a)) = 0$, then $\mathrm{Ad}(x) = \mathrm{Ad}(a)$ on v, w, so $\mathrm{Ad}(x^n) = \mathrm{Ad}(a^n) = 1$ on v, w. This implies that $\mathrm{Ad}(t) = 1$ on the subspace, for all $t \in T_a$, and further that $\mathrm{Ad}(a) = 1$ on the subspace. This contradicts the irreducibility of the representation. QED

This completes the proof of Theorem 2.1. In the next section, we will give refinements of the lemmas which will allow us to give more precise results for determining the independence threshold d, and also give examples in some special cases. Note that it follows from the proof of Lemma 2.7 that

$$\varDelta = |\det(f')|/[N_a(T_a) : T_a] = \frac{1}{[N_a(T_a) : T_a]}\left|\det_{\mathscr{L}(L_e/T_a)}(\mathrm{Ad}(t) - \mathrm{Ad}(a))\right|, \quad (2.1)$$

where $\varDelta$ gives the density of the joint distribution of the μ_i; this equation will be simplified in the next section.

Example Again, consider the complement of $SO(2N)$ in $O(2N)$. We have

$$\varDelta \propto \left|\prod_{1\le i\le(n-1)}(\lambda_i - \lambda_i^{-1})\prod_{1\le i<j\le(n-1)}(\lambda_i - \lambda_j)(\lambda_i - \lambda_j^{-1})\right|^2;$$

see [11], or Sect. 3. This is degree $(2n-2)$ in each λ_i; thus, the threshold degree is $d = (2n-2)$.

3 Refinements and examples

The formula (2.1) is not especially convenient for most purposes; it can be simplified to a significant extent, however. As before, we have L a compact Lie group, and C a component thereof; choose $a \in C$ of finite order, and choose a maximal a-torus T_a. First, we can extend T_a to a maximal torus T such that $aTa^{-1} = T$: if T' is a torus such that $aT'a^{-1} = T'$, then the set $\mathscr{S}$ of $x \in \mathscr{L}$ such that $[\mathscr{L}(T'), x] = 0$ is preserved by the automorphism a:

$$[\mathscr{L}(T'), x^a] = [\mathscr{L}(T'^a), x^a] = [\mathscr{L}(T'), x]^a = 0.$$

If $\mathscr{L}(T') = \mathscr{S}$, then $\mathscr{L}(T')$ is a maximal torus; otherwise, let $v \in \mathscr{S}$ be an eigenvector of $\mathrm{Ad}(a)$ (not in $\mathscr{L}(T')$). If v is real, then v can be added to $\mathscr{L}(T')$; else, the compactness of L implies that $[v, \overline{v}] = 0$; v and $\overline{v}$ can thus both be added to T'. Once we have extended T_a to a maximal torus T' preserved by a, we can then conjugate T' to a chosen maximal torus; as a result, for any maximal torus T, we can choose a finite-order $a \in C$ and a maximal a-torus T_a so that $T_a \in T$.

Since L is compact, we can choose a basis of $\mathscr{L}$ of the form $\{H_i\} \cup \{E_\alpha\}$, where the H_i are in T, and the E_α are simultaneous eigenvectors of T (in the adjoint representation, $x \mapsto gxg^{-1}$), and where $[E_\alpha, E_\beta] = N_{\alpha\beta} E_{\alpha+\beta}$, with $N_{\alpha\beta}$ real; α ranges over elements of the “root system” of $\mathscr{L}$. Now, E_α^a must also be a simultaneous eigenvector of T; therefore, $E_\alpha^a = s(\alpha) E_{\alpha^a}$, where $|s(\alpha)|^2 = 1$. The reality condition on $N_{\alpha\beta}$ means that if $\alpha + \beta$ is in the root system, then $s(\alpha+\beta) = \pm s(\alpha)s(\beta)$. Thus, modulo sign, we can replace a by $t_1 t_2 a t_2^{-1}$, where $t_1 \in T_a$, $t_2 \in T$, and $t_1^n = 1$, and make every $s(\alpha) = \pm 1$. Now, we can write the density formula (2.1) as a product over orbits of the E_α under the action of a; for $t \in T_a$ and an orbit O, $\det_O(\mathrm{Ad}(t) - \mathrm{Ad}(a))$ is $l_O(t)^n - \prod_{\alpha \in O} s(\alpha)$, where $\mathrm{Ad}(t) E_\alpha = l_O(t) E_\alpha$ for every $\alpha \in O$ (l_O is independent of α, since a fixes T_a). Thus, we have proved:

Theorem 3.1 *Let T be a maximal torus of a compact Lie group L. Let C be a connected component of L. Then one can choose $a \in C$ of finite order and a maximal a-torus T_a such that the induced density on T_a can be written*

$$\Delta = \frac{1}{[N_a(T_a) : T_a]} \Big(\det_{\mathscr{L}(T/T_a)} (1 - a) \Big) \prod_O \Big(l_O(t)^n - \prod_{\alpha \in O} s(\alpha) \Big). \tag{3.1}$$

Note that this is a density relative to the uniform measure on T_a; to transform it into an integral on $[0, 2\pi]^r$, a factor of $(\frac{1}{2\pi})^r$ must be added, as well as a rational factor (the reciprocal of the number of points in $[0, 2\pi]^r$ that correspond to the identity); since the integral must be 1, the constant term will be $(\frac{1}{2\pi})^r$.

As an example, consider $O(2n) - SO(2n)$; T can be taken to be the subgroup of block-diagonal matrices with each block a 2×2 rotation matrix; a can be taken to be reflection through a coordinate hyperplane, and T_a can be taken to be the subtorus of T fixed by a. The E_α that appear have α of the form $e_i - e_j$, $e_i + e_j$, or $-e_i - e_j$, $(1 \le i \ne j \le n)$, where e_i correspond to an integral basis of T (each corresponds to a μ_i); a takes e_n to $-e_n$, and fixes the remaining e_i; the $s(\alpha)$ are all 1. The resulting root orbits are: $\{e_i - e_j\}$, $\{e_i + e_j\}$, $\{-e_i - e_j\}$, $\{e_i - e_n, e_i + e_n\}$, and $\{e_n - e_i, -e_n - e_i\}$, for $1 \le i \ne j \le (n-1)$. This gives a density formula of:

$$\Delta = K \prod_{1 \le i < j \le (n-1)} (\lambda_i \lambda_j - 1)(\lambda_i \lambda_j^{-1} - 1)(\lambda_i^{-1} \lambda_j - 1)(\lambda_i^{-1} \lambda_j^{-1} - 1) \prod_{1 \le i \le (n-1)} (\lambda_i^2 - 1)(\lambda_i^{-2} - 1),$$

where K is a constant scale factor (irrelevant for our purposes); this formula can be simplified to

$$\Delta = K \left| \prod_{1 \le i \le (n-1)} (\lambda_i - \lambda_i^{-1}) \prod_{1 \le i < j \le (n-1)} (\lambda_i - \lambda_j)(\lambda_i - \lambda_j^{-1}) \right|^2 .$$

This agrees with the formula given in [11]. For our purposes, it suffices to notice that this is of degree $(2n - 2)$ in each λ_i; therefore, the threshold degree for independence here is $d = (2n - 2)$. It is fairly straightforward to verify similar formulae for $SO(2n)$, $SO(2n+1)$, and $O(2n+1) - SO(2n+1)$; we can conclude that the independence threshold for $O(n)$ is $d = (n - 2)$.

Despite the fact that Theorem 2.1 refers to a threshold degree, there can in general be degrees below the threshold that give independence. The easiest examples of this phenomenon are wreath products of a finite permutation group H (acting on a finite set S) and a (connected) compact Lie group G. In the case of a wreath product, we can take the torus T to be the same for each component: let T_0 be a maximal torus of G; then $T = T_0^S$. Now, pick a component of $G \wr_S H$. The components are parametrized by elements of H; we can thus take a to be the element of H corresponding to the chosen component. a clearly preserves T. Now, S breaks up into orbits of $\langle a \rangle$; T_a is the subtorus of T constant on orbits of $\langle a \rangle$. Let λ_i be eigenvalue generators of G; then in our component, the eigenvalue generators are given by a copy of λ_i for each orbit of $\langle a \rangle$. Noting finally that $s(\alpha) = 1$ for each root α (a simply permutes the factors), (3.1) becomes:

$$K \prod_{O} \left(\Delta_G(\lambda_{O1}^{|O|}, \lambda_{O2}^{|O|}, \ldots) \right), \tag{3.2}$$

where O ranges over orbits of $\langle a \rangle$ in S, λ_{Oi} is the ith eigenvalue generator corresponding to O, and Δ_G is the density formula for G. Now, suppose Δ_G has degree at most d in each λ_i. Then, for every term in (3.2), the degree of λ_{Oi} in that term can be written $|O|\delta$, where $1 \le \delta \le d$. Thus, for every component of $G \wr_S H$, every eigenvalue generator appears with degree of the form $\sigma\delta$, where $1 \le \sigma \le |S|$ and $1 \le \delta \le d$. It thus follows that for any m that cannot be expressed in this form, the eigenvalues of M^m are independent (in the sense of Theorem 2.1), for M Haar distributed from $G \wr_S H$. If we consider the special case $H = S = Z_n$, two things become quite apparent. Firstly, the threshold is nd, by inspection, whereas there are clearly $m < nd$ that cannot be written as $\sigma\delta$. Secondly, the set of m which give independence is relatively complicated, even in such a simple case (to be precise, it is the set of m that cannot be written in the form $\nu\delta$, where $\nu | n$ and $1 \le \delta d$); if H is a more complicated group, the situation becomes quite a bit more complicated. However, it is easy to give a threshold for general H and S: $d' = |S| d$. Thus, although stating things in terms of a threshold can lose information, the added ease of calculation more than makes up for it.

The thresholds of greatest interest in the sequel are the following: for $U(n)$, $d = n-1$, for $O(n)$, $SO(n)$, and $O(n)-SO(n)$, $d = n-2$, and for $Sp(2n)$, $d = 2n$. With care, this, combined with the theorems in [5] (given here as Theorems 6.1 and 6.2 for $O(n)$ and $Sp(2n)$), can give us some formulae for the means and covariances of the $\mathrm{Tr}(M^i)$.

For $U(n)$, a simple rotational symmetry argument gives $E(\mathrm{Tr}(U^i)) = 0$ and $E(\mathrm{Tr}(U^i)\overline{\mathrm{Tr}(U^j)}) = 0$, unless $i = j$. In that case, the formula in [5] gives $E(|\mathrm{Tr}(U^i)|^2) = i$ for $i \leq n$. For $i > n$, Theorem 2.1 kicks in, giving $E(|\mathrm{Tr}(U^i)|^2) = n$ for $i > n$, as shown in Sect. 1.

For $O(n)$, Theorem 6.1 and Theorem 2.1 (plus a slight refinement thereof, to the effect that

$$E\left(\sum_{i \neq j}(\lambda_i^k \lambda_j^l)\right) = 0$$

if either k or l is greater than $n-2$) give the following formulae (where the notation is used that $[i \text{ even}]$ is 1 if i is even, and 0 otherwise, and similarly for other predicates):

$$E(\mathrm{Tr}(O^i)) = [i \text{ even}], \tag{3.3}$$

$$\mathrm{Cov}(\mathrm{Tr}(O^i), \mathrm{Tr}(O^j)) = \min(i, n-1)\delta_{ij} + [i-n \text{ even}, i \geq n][j-n \text{ even}, j \geq n]. \tag{3.4}$$

For $SO(n)$, we have

$$E(\mathrm{Tr}(O^i)) = [i \text{ even}] + (-1)^n[i-n \text{ even}, i \geq n],$$

while for $O(n) - SO(n)$ we have

$$E(\mathrm{Tr}(O^i)) = [i \text{ even}] - (-1)^n[i-n \text{ even}, i \geq n].$$

To compute the covariances for those cases, as well as for $Sp(2n)$ would require stronger results than those given in Sect. 6.

Finally, for $Sp(2n)$, we get

$$E(\mathrm{Tr}(S^i)) = -[i \text{ even}, i \leq 2n].$$

Acknowledgements. The greatest thanks are due to Persi Diaconis (the author's thesis advisor), for his many suggestions of problems (a small sampling of which appear in the present work), for his helpful advice when it came time to revise what had been written, and for his encouragement of the author's procrastination.

Thanks are also due to the following people and institutions, in no particular order: Andrew Odlyzko, for many helpful comments on Sect. 5 of [9]; AT&T Bell Laboratories (Murray Hill) and the Center for Communications Research (Princeton) for generous summer support; the Harvard University Mathematics Department and the National Science Foundation, for generous support for the rest of the year. Also, thanks are due, for helpful comments, to Nantel Bergeron, Maurice Rojas, Richard Stanley, and Dan Stroock. Last, but not least, the author owes thanks to Wojbor Woyczynski, both for introducing him to probability theory and for introducing him to Prof. Diaconis; the thesis excerpted here would be very different, had either introduction not been made.

References

1. Adams, J.F.: Lectures on Lie Groups W.A. Benjamin, New York (1969)
2. Billingsley, P.: Probability and Measure, 2nd edition. Wiley, New York (1986)
3. Boas, R.P.: Invitation to Complex Analysis Random House, New York (1987)

4. Bourbaki, N.: Elements of Mathematics: Lie groups and Lie algebras, part I, Addison-Wesley, Reading, Mass. (1975)
5. Diaconis, P., Shahshahani, M.: *On the Eigenvalues of Random Matrices* J. Appl. Prob. **31** (1994), 49–61
6. Hewitt, E., Ross, K.A.: Abstract Harmonic Analysis, vol. 1 Springer-Verlag, Berlin (1963)
7. Knuth, D.E.: The Art of Computer Programming, vol. 1 Fundamental Algorithms 2nd edition. Addison Wesley, Reading, Mass. (1973)
8. Mehta, M.L.: Random Matrices, 2nd edition. Acadmic Press, Boston (1991)
9. Rains, E.M.: Topics in Probability on Compact Lie Groups, Ph.D. thesis, Harvard University (1995).
10. Tracy, C., Widom, H.: *Introduction to Random Matrices* Proc. 8th Scheveningen Conf. Springer-Verlag, to appear?
11. Weyl, H.: Classical Groups, Princeton University Press, Princeton (1942)

This article was processed by the author
using the Springer-Verlag TEX QPMZGHB macro package 1991.