On Generalizing the Cut-off Phenomenon for Random Walks on Groups

by

Jeffrey S. Rosenthal

School of Mathematics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

(November, 1992; revised June, 1993)

1. Introduction.

There has been a lot of recent work on the convergence of random walks on finite or compact groups to their stationary, uniform distribution. Particular emphasis has been placed on the *rate* of convergence, i.e. on estimates of the number of iterations until the random walk is "close" to uniformity. The best-known examples come from the card-shuffling analyses of Diaconis and co-workers; see [D] for an extensive introduction.

The most striking fact to emerge from these analyses is the existence of the "cut-off phenomenon" (see [DS], [AD], [D]) in certain examples, meaning that the variation distance to uniformity remains close to 1 for a large number of iterations, and then decreases to close to 0 in a relatively small number of further iterations. Precise definitions are given in Section 2.

The cut-off phenomenon has been observed in a number of specific examples, including Random Transpositions [DS], Top-to-Random Shuffles [AD], Riffle Shuffles [BD], Random Transvections [H], Random Rotations [R], and Random Reflections [P]. The known examples are all very specific, and it is reasonable to ask whether this phenomenon occurs more generally. On the other hand, the cut-off phenomenon does *not* occur for *all* random walks on finite and compact groups: One counter-example is simple random walk on $\mathbf{Z}/(n)$ (see [D], Section 3C, Theorem 2).

This paper presents a first step towards a more general result about the cut-off phenomenon. A large class of measures on a fairly large collection of groups (both finite and compact) are considered. The measures are still required to be conjugate-invariant, but they are defined much more generally than in the previous specific examples. For these measures, we prove the "easier half" of a cut-off phenomenon. Specifically, we provide the lower-bound part of the argument, proving that the variation distance stays close to 1 until a certain specified number of iterations, after which we *conjecture* that a cut-off occurs. Our methods involve Fourier analysis, and directly generalize previous work of others. We close by briefly discussing possible approaches to proving the corresponding upper bound, and thereby establishing the cut-off phenomenon in this generality.

Section 2 presents the necessary notation and preliminaries. Section 3 presents our main result, Section 4 provides some examples, and Section 5 presents the proof. Section 6 discusses the question of the upper bound.

2. Preliminaries.

A probability measure Q on a group G induces a random walk on the group, with the distribution after k steps given by Q^{*k}, the k-fold convolution of the measure Q with itself.

If the group G is compact [or finite], it has an associated normalized Haar measure λ [or uniform probability measure]. If Q is "nice enough" (in particular, for finite G, if the support of the measure Q is not contained in any coset of any subgroup of G), then it is well-known (see e.g. [K]) that the measures Q^{*k} will converge to λ in total variation distance. In symbols,

$$\|Q^{*k} - \lambda\|_{\mathrm{var}} \;:=\; \sup_{A \subseteq G} |Q^{*k}(A) - \lambda(A)| \;\to\; 0 \quad \text{as} \quad k \to \infty \,.$$

It is often desired to know the "rate" at which this convergence to 0 takes place, in the sense of how large k should be to make the variation distance small.

These questions can be handled by Fourier Analysis. See [DS] and [D] for background. To define things, let $\rho_0, \rho_1, \rho_2, \ldots$ be the irreducible representations of the compact (or finite) group G, with ρ_0 the trivial representation. Let d_i be the dimension of ρ_i. Recall that the Fourier Transform of the measure Q^{*k} is defined by

$$\widehat{Q^{*k}}(\rho_i) \;:=\; \int\limits_G \rho_i(s)\, Q^{*k}(ds)\ .$$

In this paper, we shall specialize to probability measures Q which are conjugate-invariant, in the sense that $Q(s^{-1}As) = Q(A)$ for all group elements s and all measurable subsets A.

For such measures, Schur's Lemma implies that $\widehat{Q}(\rho_i) = c_i I_i$ for some complex number c_i, where I_i is the $d_i \times d_i$ identity matrix. Thus

$$\widehat{Q^{*k}}(\rho_i) \;=\; (c_i)^k I_i\,.$$

Furthermore, it is seen by taking traces that $c_i = \left(\operatorname{tr} \widehat{Q}(\rho_i)\right)/d_i$; in particular, $|c_i| \le 1$.

The Upper Bound Lemma of Diaconis and Shashahani ([DS]; see also Chapter 3B of [D]) states in this case that

$$\|Q^{*k} - \lambda\|_{\mathrm{var}} \;\le\; \frac{1}{2}\sqrt{\sum_{i\ge 1}(d_i)^2 |c_i|^{2k}}\;. \tag{1}$$

In words, this states that if k is large enough to make $\sum\limits_{i\ge 1}(d_i)^2|c_i|^{2k}$ small, then the variation distance of Q^{*k} to stationarity will be small. Equation (1) has been used to prove the upper bounds in most of the previous cut-off phenomenon results; we shall discuss its possible use in our problem in Section 6.

On the other hand, consider the character $\chi_1(s) = \operatorname{tr}\rho_1(s)$, where ρ_1 is the "first non-trivial irreducible representation", to be defined more precisely later. Under the uniform distribution λ, it is easily checked that we have $E_\lambda(\chi_1) = 0$ and $\operatorname{Var}_\lambda(\chi_1) = 1$. Under the distribution Q^{*k}, setting $m_k := E_{Q^{*k}}(\chi_1)$, we have $m_k = (c_1)^k d_1$. Now, if m_k were large for some k, then we would expect that the measures Q^{*k} and λ would be "far apart". To make this more precise, let A be the subset of G on which $\chi_1(s) \ge |m_k|/2$, and set $v_k := \operatorname{Var}_{Q^{*k}}(\chi_1) := E_{Q^{*k}}|\chi_1 - m_k|^2$. Then, using Chebychev's inequality twice (cf. p. 44 of [D]), we see that

$$\lambda(A) \;\le\; 4/|m_k|^2\,; \qquad \text{and}$$

$$Q^{*k}(A) \;\ge\; 1 \;-\; 4v_k/|m_k|^2\;.$$

Hence, it follows that

$$\|Q^{*k} - \lambda\|_{\mathrm{var}} \;\ge\; 1 \;-\; 4(1+v_k)/|m_k|^2\;. \tag{2}$$

In words, this states that if $(1+v_k)/|m_k|^2$ is small, then the variation distance of Q^{*k} to stationarity will be close to 1. We shall use equation (2) to establish our lower bound result in the present generality.

It has become standard to consider a *sequence* $G_1, G_2, \ldots$ of groups, with $|G_n| \to \infty$, and with a probability measure Q_n defined on each G_n in a "natural" way. (For example, for card shuffling, $G_n = S_n$ is the symmetric group on n letters, and Q_n is an appropriate suffling measure.) For such a sequence, it is desired to know the time to stationarity as a function of n. The sequence is said to have a *cut-off* if the time to stationarity gets more "sharply defined" with increasing n. More precisely, we have the following definition, taken from [AD].

Definition. Let $\{(G_n, Q_n)\}_{n=1}^{\infty}$ be a sequence of compact groups with probability measures. The sequence has a *cut-off at* k_n if for any $\epsilon > 0$,

$$\|Q_n^{*(1-\epsilon)k_n} - \lambda_n\|_{\text{var}} \ \to \ 1 \quad \text{as} \quad n \ \to \ \infty; \quad \text{and}$$

$$\|Q_n^{*(1+\epsilon)k_n} - \lambda_n\|_{\text{var}} \ \to \ 0 \quad \text{as} \quad n \ \to \ \infty.$$

In words, this says that k_n iterations (to first order in n) are both necessary and sufficient to approach stationarity on G_n, and that this convergence to stationarity gets relatively more and more "sudden" as n increases.

In this paper, we shall give a general proof of a *lower bound,* corresponding to the convergence to 1 above. In the final section, we briefly discuss possiblities for proving an upper bound corresponding to the convergence to 0.

3. Main Result.

We construct the measures we shall analyze as follows. Let $G_1, G_2, G_3, \ldots$ be a sequence of groups, such that G_n is a subgroup of the unitary group $U(n)$ that includes at least the even permutation matrices A_n. (In symbols, $A_n < G_n < U(n)$.) Let $M_1, M_2, \ldots, M_\ell$ be fixed matrices with $M_i \in U(r_i)$ (an $r_i \times r_i$ matrix), and let $a_1, a_2, \ldots, a_\ell$ be non-negative real numbers summing to 1. The idea is to construct a probability measure Q_n on G_n which represents a weighted (by a_i) average of measures which are uniform on conjugacy classes related to $M_1, \ldots, M_\ell$, respectively.

To this end, let $M_i^{(n)} = M_i \oplus I_{n-r_i}$ be the $n \times n$ matrix formed by extending M_i by the $(n-r_i)$-dimensional identity matrix. Thus $M_i^{(n)} \in U(n)$. Assume further that $M_i^{(n)} \in G_n$

for each n. Let $\mathcal{U}_{i,n}$ be the probability measure which is uniformly concentrated on the conjugacy class of $M_i^{(n)}$ in G_n. (To be precise, $\mathcal{U}_{i,n}$ is the measure induced from Haar measure on G_n by the map $x \to x^{-1} M_i^{(n)} x$.) Finally, let $Q_n = \sum_{i=1}^{\ell} a_i \mathcal{U}_{i,n}$. Thus, Q_n is a probability measure on G_n which is conjugate-invariant, and which represents a weighted average of the uniform distributions on the conjugacy classes of $M_1^{(n)}, \ldots, M_\ell^{(n)}$.

We observe that these measures include many of the known examples, as well as many new ones. For such measures, we prove the following.

Theorem 1. *Let G_n and Q_n be as above. Then there are positive numbers A, B, C, depending only on $M_1, \ldots, M_\ell$ and $a_1, \ldots, a_\ell$, but independent of n and G_n, such that if $k = \Re e\,(\alpha^{-1}) n \log n + tn$ for any $t < 0$, then*

$$\|Q_n^{*k} - \lambda_n\|_{\text{var}} \;\geq\; 1 - Ae^{Bt} - C\left(\frac{\log n}{n}\right).$$

Here Q_n^{*k} is the k-fold convolution of Q_n with itself, λ_n is normalized Haar measure on G_n, $\|\cdot\|_{\text{var}}$ is total variation distance on G_n, and $\alpha = \sum_{i=1}^{\ell} a_i (r_i - \operatorname{tr} M_i)$.

If $t << 0$ and n is large, the right-hand side of this inequality is close to 1. The theorem thus says that the variation distance to Haar measure is close to 1 if we do significantly less than $\Re e\,(\alpha^{-1}) n \log n$ iterations of the random walk.

The real significance of this assertion lies in the following conjecture, which asserts that the result of Theorem 1 corresponds to a cut-off phenomenon for the measures Q_n.

Conjecture 2. *Let G_n, Q_n, and α be as above. Assume further that the support of Q_n is not contained in any proper coset of G_n for sufficiently large n. Then for many* such Q_n, there are constants D and E, depending only on $M_1, \ldots, M_\ell$ and $a_1, \ldots, a_\ell$, such that for sufficiently large n, if $k = \Re e\,(\alpha^{-1}) n \log n + tn$ for any $t > 0$, then*

$$\|Q_n^{*k} - \lambda_n\|_{\text{var}} \;\leq\; De^{-Et}.$$

* *We originally made our conjecture for all such Q_n. However, Ursula Porod, at Johns Hopkins University, has recently constructed a counter-example on $U(n)$, for which Q_n^{*k} is not even in L^2 for $k < O(n^2)$. The precise conditions on G_n and Q_n required for the conjecture to hold remains an open problem.*

If $t >> 0$, the right-hand side of this inequality is very close to 0. The conjecture thus asserts that the variation distance to Haar measure is close to 0 if we do significantly more than $\Re e\,(\alpha^{-1}) n \log n$ iterations of the random walk.

To make the conjecture plausible, it should be noted that since the support of Q_n is not contained in any coset of G_n, the measure Q_n^{*k} will usually be absolutely continuous with respect to Haar measure for $k \geq O(n)$. Furthermore, there will be no periodicity problems. Evidence for the conjecture comes from examining the computations for the Upper Bound Lemma (equation (1)) in previously known examples, as well as the way these computations appear to work in the present generality. Further comments on possibilities for proving this conjecture will be given in Section 6.

Remark. Theorem 1 remains true if the numbers a_i are replaced by numbers $a_{i,n}$ depending on n, provided these numbers approach limits r_i in such a way that $|a_{i,n} - r_i| \leq$ (const)$/n$. While for clarity we shall not consider this extension further, we note that our proof goes through with only minor modifications. Also, Conjecture 2 remains unchanged as well, except that we must now require that Q_n assigns probability at least (const)$/n$ to the complement of each proper coset of G_n.

4. Examples.

In this section, we present two straightforward examples of Theorem 1 and Conjecture 2.

Example 1. On S_n, Theorem 1 says that it takes the product of at least $(1/j) n \log n$ randomly chosen j-cyles to get random, for j fixed and n large. (The case $j = 2$ corresponds to the Random Transpositions studied in [DS].) More generally, if we define Q_n on S_n to share weight a_i uniformly over all elements with non-trivial cycles of sizes $s_{i1}, \ldots, s_{iN_i}$ (where $\sum_i a_i = 1$), then Theorem 1 says it takes at least $\alpha^{-1} n \log n$ iterations to get random, where

$$\alpha = \sum_i \left(a_i \sum_j s_{ij} \right) .$$

If for different i the sum $\sum\limits_j s_{ij}$ is both even and odd, or if we restrict attention to A_n instead of S_n, then Conjecture 2 asserts that $\alpha^{-1} n \log n$ iterations also *suffices* to get random.

Example 2. On $SO(n)$, if we define Q_n to share weight a_i uniformly over all elements corresponding to rotations in orthogonal planes of angles $s_{i1}, \ldots, s_{iN_i}$, then Theorem 1 says it takes at least $\alpha^{-1} n \log n$ iterations to get random, with

$$\alpha = \sum_i a_i \sum_j (2 - 2\cos s_{ij}) \ ,$$

while Conjecture 2 asserts that this number of iterations also suffices. The case $a_1 = 1$, $N_1 = 1$ and $s_{11} = \pi$ was studied in detail in [R].

5. Proof of Main Result.

In this section, we provide a proof of Theorem 1 above, making use of the method implied by equation (2). We emphasize that this method has been used previously (see p. 44 of [D]) in specific cases. Our contribution consists of proving a much more general result using similar methods. In addition to the much greater generality, our approach may help to "explain" why the computations work out as they do in the previous specific cases.

We proceed by fixing n, and letting ρ_1 be the "first non-trivial" irreducible representation of G_n given as follows. Think of G_n as acting on $\mathbf{R}^n$. Let $e_1, \ldots, e_n$ be the standard basis for $\mathbf{R}^n$, and let $e_+ = e_1 + \ldots + e_n$. We distinguish two cases. Case (A) below includes A_n and S_n, while case (B) below includes any G_n with $SO(n) < G_n < U(n)$.

Case (A): $g(e_+)$ is a scalar multiple of e_+ for each $g \in G_n$. This means that G_n leaves the perpendicular subspace $(e_+)^\perp$ invariant. We define the representation ρ_1 by $\rho_1(g) = g|_{(e_+)^\perp}$. In words, ρ_1 takes each group element g to the $(n-1)$-dimensional operator given by the restriction of g to $(e_+)^\perp$.

Case (B): G_n does *not* leave the subspace $(e_+)^\perp$ invariant. In this case we simply define ρ_1 by $\rho_1(g) = g$. In words, ρ_1 takes each group element g to *itself.*

Thus, ρ_1 is essentially the "natural" representation of G_n, with the proviso that we mod out the span of e_+ in case (A) above. It is well known that this representation ρ_1, so constructed, is irreducible (cf. [JG], p. 97).

Recall that d_1 is the dimension of ρ_1, and $\chi_1 = \operatorname{tr} \rho_1$ is the corresponding character. In case (A) we have $d_1 = n - 1$, and $\chi_1(g) = \operatorname{tr}(g) - 1$, while in case (B) we have $d_1 = n$

and $\chi_1(g) = \mathrm{tr}\,(g)$. The two cases are not identical, but our computation will be similar in each of them.

Following the method implied by equation (2), we let $m_{n,k} = E_{Q_n^{*k}}(\chi_1)$ and $v_{n,k} = \mathrm{Var}\,_{Q_n^{*k}}(\chi_1) := E_{Q_n^{*k}}|\chi_1 - m_{n,k}|^2$. Our goal is to bound $(1+v_{n,k})/|m_{n,k}|^2$ and show it is small (for large n) if $k = \Re e\,(\alpha^{-1}) n \log n + tn$ with $t << 0$.

We proceed by considering the representation $\rho_1 \otimes \overline{\rho_1}$ of G_n given by taking the tensor product of ρ_1 with its (entry-wise) complex conjugate. This new representation will not be irreducible but will split into a (finite) direct sum of irreducible representations:

$$\rho_1 \otimes \overline{\rho_1} \sim \bigoplus_j \rho_j$$

for some irreducible representations ρ_j. Once we understand this splitting, then with d_j and χ_j the dimension and character of ρ_j, we can write

$$E_{Q_n^{*k}}(|\chi_1|^2) \;=\; E_{Q_n^{*k}}\,(\mathrm{tr}\,(\rho_1 \otimes \overline{\rho_1})) \;=\; E_{Q_n^{*k}}\left(\mathrm{tr} \bigoplus_j \rho_j\right) \;=\; E_{Q_n^{*k}} \sum_j \chi_j \;=\; \sum_j (c_j)^k d_j\ ,$$

with c_j as in Section 2. Then, recalling that $v_{n,k} = E_{Q_n^{*k}}\left(|\chi_1 - m_{n,k}|^2\right) = E_{Q_n^{*k}}(|\chi_1|^2) - |m_{n,k}|^2$, that $(d_1)^2 = \sum_j d_j$, and that $m_{n,k} = (c_1)^k d_1$, we have that

$$v_{n,k} = \sum_j \left((c_j)^k - |c_1|^{2k}\right) d_j\ . \tag{3}$$

We shall use this equation to obtain bounds on $v_{n,k}$ in Lemma 3 below. The proof is somewhat computational, and involves considering the details of how the representation $\rho_1 \otimes \overline{\rho_1}$ splits into irreducibles.

Lemma 3. *Let G_n, Q_n be as in Section 2, let α be as in Theorem 1, let χ_1 be as above, and let $k = \Re e\,(\alpha^{-1}) n \log n + tn$ for any $t < 0$. Then there are numbers A and B, bounded independently of n and t, such that*

$$m_{n,k} := E_{Q_n^{*k}}(\chi_1) = e^{-\alpha t} + A\left(\frac{\log n}{n}\right)\ ; \quad \textit{and}$$

$$v_{n,k} := \mathrm{Var}\,_{Q_n^{*k}}(\chi_1) \le 1 + B\left(\frac{\log n}{n}\right) e^{-2\Re e\,(\alpha) t} + |m_{n,k}|\ .$$

Proof. We begin with the computation for $m_{n,k}$. Recall that $m_{n,k} = (c_1)^k d_1$. If we are in Case (B) above, then $d_1 = n$ and

$$c_1 = \left(\operatorname{tr}\widehat{Q}(\rho_1)\right)/d_1 = \left(\sum_i a_i\,\chi_1(M_i^{(n)})\right)/n$$
$$= \left(\sum_i a_i(n - r_i + \operatorname{tr} M_i)\right)/n = (n-\alpha)/n = \left(1 - \frac{\alpha}{n}\right) .$$

In Case (A) above, both the numerator and denominator of $c_1 = \frac{\operatorname{tr}\widehat{Q}(\rho_1)}{d_1}$ will be decreased by 1, so we instead find that $c_1 = \left(1 - \frac{\alpha}{n-1}\right)$. Hence, in either case (writing $n \pm 1$ for n or $n-1$) we have

$$\begin{aligned} m_{n,k} &= \left(1 - \frac{\alpha}{n \pm 1}\right)^{(\alpha^{-1} n \log n + tn)} (n \pm 1) \\ &= \left((1/n) + O(\log n/n^2)\right)\left(e^{-\alpha t} + O(1/n)\right)(n \pm 1) \;=\; e^{-\alpha t} + O(\log n/n)\,, \end{aligned}$$

as asserted.

We now consider $v_{n,k}$. We shall make use of formula (3). We begin by restricting our attention to the case where G_n satisfies $SO(n) < G_n < O(n)$. (We are thus in case (B) above.) In this case, it is well-known (see e.g. [R]) that

$$\rho_1 \otimes \overline{\rho_1} \sim \rho_0 \oplus \rho_2 \oplus \rho_3 \tag{4}$$

where ρ_0 is the trivial representation given by the restriction of $\rho_1 \otimes \overline{\rho_1}$ to the 1-dimensional subspace spanned by the identity matrix; ρ_2 is given by the restriction of $\rho_1 \otimes \overline{\rho_1}$ to the skew-symmetric matrices; and ρ_3 is given by the restriction of $\rho_1 \otimes \overline{\rho_1}$ to the symmetric, traceless matrices. Thus $d_2 = n(n-1)/2$ and $d_3 = (n(n+1)/2) - 1$. We begin with the computation for this case, and then derive the general case afterwards.

The sum in (3) thus consists of three terms. The term $j = 0$ corresponding to the trivial representation ρ_0 (where $d_0 = 1$) is clearly bounded by 1:

$$\left((c_0^k - |c_1|^{2k}\right) d_0 \leq 1 \ . \tag{5}$$

For the term $j = 2$ corresponding to ρ_2, we shall show that $(c_2)^k - |c_1|^{2k}$ is small, which we shall do by showing that $c_2 - |c_1|^2$ is small. The reason this is true, roughly, is that while $\rho_1(M_i^{(n)})$ leaves about $n - r_i$ out of n basis elements invariant, $\rho_2(M_i^{(n)})$ will leave about $(n-r_i)^2/2$ out of $n^2/2$ basis elements invariant, so that $\Big(\operatorname{tr}\rho_2(M_i^{(n)})/d_2\Big) \approx \Big(\operatorname{tr}\rho_1(M_i^{(n)})/d_1\Big)^2$.

We now proceed to the details of the calculation. Recall that

$$c_2 = \frac{\operatorname{tr}\widehat{Q}(\rho_2)}{d_2} = \sum_i a_i \, \frac{\operatorname{tr}\rho_2(M_i^{(n)})}{d_2} \; .$$

Since a basis for ρ_2 (thought of as a sub-representation of $\rho_1 \otimes \overline{\rho_1}$) is given by $\{e_r \otimes e_s - e_s \otimes e_r \,|\, r < s\}$, we have

$$\operatorname{tr}\rho_2\Big(M_i^{(n)}\Big) = \frac{1}{2}\sum_{r<s} \langle\ (\rho_1 \otimes \overline{\rho_1})(M_i^{(n)})(e_r \otimes e_s - e_s \otimes e_r),\ e_r \otimes e_s - e_s \otimes e_r\ \rangle\ .$$

But $\rho_1(M_i^{(n)})$ leaves e_r fixed for $r > r_i$, so $(\rho_1 \otimes \overline{\rho_1})(M_i^{(n)})(e_r \otimes e_s - e_s \otimes e_r) = e_r \otimes e_s - e_s \otimes e_r$ for $r, s > j$. Also for $r \le r_i < s$, we have

$$\begin{aligned}
&\langle\ (\rho_1 \otimes \overline{\rho_1})(M_i^{(n)})(e_r \otimes e_s - e_s \otimes e_r),\quad e_r \otimes e_s - e_s \otimes e_r\ \rangle \\
&= \langle\ \rho_1(M_i^{(n)})(e_r) \otimes e_s - e_s \otimes \overline{\rho_1(M_i^{(n)})(e_r)},\quad e_r \otimes e_s - e_s \otimes e_r\ \rangle \\
&= 2\,\Re e\,\langle\ \rho_1(M_i^{(n)})(e_r),\quad e_r\ \rangle
\end{aligned}$$

Hence,

$$\begin{aligned}
&\frac{1}{2}\sum_{s<r} \langle\ (\rho_1 \otimes \overline{\rho_1})(M_i^{(n)})(e_r \otimes e_s - e_s \otimes e_r),\quad e_r \otimes e_s - e_s \otimes e_r\ \rangle \\
&= \frac{(n-r_i-1)(n-r_i-2)}{2} + (n-r_i)\,\Re e \sum_{r=1}^{r_i} \langle \rho_1(M_i^{(n)})(e_r), e_r\rangle + \Delta\,, \\
&= \frac{(n-r_i-1)(n-r_i-2)}{2} + (n-r_i)\,\Re e \operatorname{tr} M_i + \Delta\,,
\end{aligned}$$

where $\Delta = \frac{1}{2}\sum_{r<s\le r_i} \langle\ (\rho_1 \otimes \overline{\rho_1})(M_i^{(n)})(e_r \otimes e_s - e_s \otimes e_r),\quad e_r \otimes e_s - e_s \otimes e_r\ \rangle = O(1)$ as $n \to \infty$. Recalling that $d_2 = (n-1)(n-2)/2$, we see that

$$\begin{aligned}
c_2 &= \sum_i a_i \left(\frac{(n-r_i-1)(n-r_i-2) + 2(n-r_i)\Re e \operatorname{tr} M_i}{(n-1)(n-2)}\right) + O(1/n^2) \\
&= \sum_i a_i \left(1 - \frac{2\,(r_i - \Re e \operatorname{tr} M_i)}{n}\right) + O(1/n^2) \\
&= 1 - \frac{2\Re e\,(\alpha)}{n} + O(1/n^2)\ .
\end{aligned}$$

But

$$c_1 = \sum_i a_i \chi_1(M_i^{(n)})/n = \sum_i a_i \left(1 - \frac{(r_i - \operatorname{tr} M_i)}{n}\right) = 1 - \frac{\alpha}{n}\,,$$

so that

$$|c_1|^2 = 1 - \frac{2\Re e\,(\alpha)}{n} + O(1/n^2)\,,$$

and hence

$$\left|c_2 - |c_1|^2\right| = O(1/n^2)\ . \tag{6}$$

Using the basis

$$\{e_r \otimes e_s + e_s \otimes e_r \mid r < s\} \cup \{e_r \otimes e_r - e_n \otimes e_n \mid 1 \le r \le n\}$$

for the term corresponding to $j = 3$, it is similarly computed that

$$c_3 = \sum_i a_i \left(1 - (2/n)\left(r_i - \Re e \operatorname{tr} M_i\right)\right) + O(1/n^2)\ ,$$

so that also

$$\left|c_3 - |c_1|^2\right| = O(1/n^2)\ . \tag{7}$$

To make good use of these bounds, we bound a typical term in equation (3) by noting that

$$|x^k - y^k| = |(x-y)(x^{k-1} + x^{k-2}y + \ldots + y^{k-1})| \le |x-y|\, k\, \max(|x|,|y|)^k\,.$$

Thus, using equation (6), we have that for the term $j = 2$,
$\left|(c_2)^k - |c_1|^{2k}\right| d_2 \le \left|c_2 - |c_1|^2\right|\ k\ \max(|c_2|, |c_1|^2)^k\, d_2$

$$= O(1/n^2)(\Re e\,(\alpha^{-1}) n \log n + tn)\left(1 - \frac{2\Re e\,(\alpha)}{n} + O(1/n^2)\right)^{\Re e\,(\alpha^{-1}) n \log n + tn} O(n^2)$$

$$\le O\left(\frac{\log n}{n}\right) e^{-\Re e\,(\alpha) t} \qquad \text{(for } t < 0\text{)}$$

(and similarly for the term $j = 3$). The asserted bound on $v_{n,k}$ now follows, for the case $SO(n) < G_n < U(n)$, using equations (3), (5), (6), and (7) (and with the $|m_{n,k}|$ term unnecessary).

We now proceed to the more general case $A_n < G_n < O(n)$. The splitting given by equation (4) will be slightly different in this case, but the changes do not matter significantly as we shall see. There are several ways in which our computation for $v_{n,k}$ could differ from the previous case. We list them all, and then argue point by point that the conclusion about $v_{n,k}$ does not change.

The key point is that the splitting (4) cannot change by very much. Indeed, we have (see [JG], p. 97 for S_n, and [FOW] for the extension to A_n) that for the case $G_n = A_n$,

$$\rho_1 \otimes \overline{\rho_1} \sim \rho_0 \otimes \rho_1 \otimes \rho_2 \otimes \rho_3 \tag{8}$$

where the representations are as follows. Recall that A_n falls in case (A) above, so that ρ_1 is $(n-1)$-dimensional, and in fact corresponds to restricting each group element's action to those elements of $\mathbf{R}^n$ whose coordinates sum to zero. Thus $\rho_1 \otimes \overline{\rho_1}$ acts on those elements of the vector space $\mathbf{R}^n \otimes \mathbf{R}^n$ each of whose rows and columns sum to zero. The splitting of this vector space is as follows: ρ_0 is again the trivial representation, and corresponds to restricting $\rho_1 \otimes \overline{\rho_1}$ to the 1-dimensional subspace given by

$$V_0 = \text{span}\left\{ (n-1)\sum_{i=1}^{n} e_i \otimes e_i - \sum_{i \neq j} (e_i \otimes e_j) \right\} \; ;$$

ρ_1 appears in the splitting as the restriction of $\rho_1 \otimes \overline{\rho_1}$ to the $(n-1)$-dimensional subspace

$$\begin{aligned} V_1 = (V_0)^{\perp} \;\cap\; \text{span}\; \Big\{ \; & (n-1)^2 \sum_{i=1}^{n} e_i \otimes e_i - (n-1)\sum_{j \neq i} (e_i \otimes e_j + e_j \otimes e_i) \\ & + \sum_{j,k \neq i} e_j \otimes e_k \mid 1 \leq i \leq n \Big\} \; ; \end{aligned}$$

ρ_2 corresponds to the restriction to the $\left(\frac{(n-1)(n-2)}{2}\right)$-dimensional subspace

$$V_2 = \text{span}\left\{ (e_i \otimes e_j - e_j \otimes e_i) + (e_k \otimes e_i - e_i \otimes e_k) + (e_j \otimes e_k - e_k \otimes e_j) \mid 1 \leq i < j < k \leq n \right\} \; ;$$

and ρ_3 corresponds to a restriction to the $\left(\frac{n(n-3)}{2}\right)$-dimensional subspace

$$\begin{aligned} V_3 = (V_0)^{\perp} \;\cap\; (V_1)^{\perp} \;\cap\; \text{span}\; \Big\{ \; & e_i \otimes e_i + e_j \otimes e_j - 2e_k \otimes e_k \\ & - 2(e_i \otimes e_j + e_j \otimes e_i) + (e_i \otimes e_k + e_k \otimes e_i) + (e_j \otimes e_k + e_k \otimes e_j) \mid 1 \leq i < j < k \leq n \Big\} \; . \end{aligned}$$

Thus, ρ_2 and ρ_3 are roughly as they were in equation (4) above, except for some minor modifications because they are restricted to $(e_+)^\perp \otimes (e_+)^\perp$. The only other new feature is the appearance of ρ_1 in the right-hand side of this splitting, where it didn't appear before.

Since the splittings (4) and (8) for the extremes $O(n)$ and A_n are so similar, this allows us to control the splittings for any G_n sandwiched between them. Indeed, for any such G_n, the splitting will be something "in between" equations (4) and (8), in the sense that each of ρ_1, ρ_2, and ρ_3 could "shrink" a little bit from its value in (4), and also ρ_1 could appear in the right-hand side of the splitting, but these changes will always be controlled by being extensions of the corresponding representations appearing in (8).

To be more precise, the possible differences between the computation for $v_{n,k}$ in the case $SO(n) < G_n < O(n)$, and in the current more general case, are as follows.

- (i) The value of $c_1 = \frac{\operatorname{tr} \widehat{Q}(\rho_1)}{d_1}$ could be changed (both denominator and numerator could decrease by 1).
- (ii) The value of $c_2 = \frac{\operatorname{tr} \widehat{Q}(\rho_2)}{d_2}$ could be changed (both denominator and numerator could decrease because we're restricting to $(e_+)^\perp$).
- (iii) The value of $c_3 = \frac{\operatorname{tr} \widehat{Q}(\rho_3)}{d_3}$ could be similarly changed.
- (iv) The extra representation ρ_1 may appear in the splitting.

We examine each of these issues in turn. For (i), we have already seen that the value of c_1 will be changed by at most $((\alpha/(n-1)) - (\alpha/n)) = O(1/n^2)$. For (ii) and (iii), it is seen that the denominator (d_2 or d_3) will be decreased by an amount $\delta = O(n)$, and the numerator ($\operatorname{tr} \widehat{Q}(\rho_2)$) will be decreased by an amount $\delta \pm O(1)$. Since the denominator is $O(n^2)$ and the numerator is within $O(n)$ of the denominator in any case, it is seen that the value of c_2 or c_3 will change by only $O(1/n^2)$. Specifically, writing the denominator as xn^2, and the numerator as $xn^2 - yn$, the change in c_2 or c_3 will be of the form

$$\left(\frac{xn^2 - yn}{xn^2} \right) \; - \; \left(\frac{xn^2 - yn - \delta \pm O(1)}{xn^2 - \delta} \right) \; = \; O(1/n^2) \, .$$

Thus, differences (i), (ii), and (iii) will only affect the values of $|c_j - |c_1|^2|$ in equations (6) and (7) by $O(1/n^2)$ which will not affect the conclusion. Also, difference (iv) will merely add $m_{n,k}$ to $v_{n,k}$, which accounts for the extra term $|m_{n,k}|$ in the lemma.

We conclude that the previous computation for $v_{n,k}$ will not be changed in a way that affects the conclusion. Thus, the present statement about $v_{n,k}$ in the more general situation is verified.

Finally, we turn our attention to the general (possibly complex) case $A_n < G_n < U(n)$. Here, in addition to the issues mentioned above, it is possible (because of the complex conjugate of the second ρ_1) that the two large representations in the splitting of $(\rho_1 \otimes \overline{\rho_1})$ (namely ρ_2 and ρ_3 corresponding to skew-symmetric and symmetric matrices, respectively) will themselves not be separately invariant. If so, the representations ρ_2 and ρ_3 would be replaced by a representation ρ_4 corresponding to their direct sum (modulo the possible small differences mentioned above). Setting $d_j = \dim \rho_j$ and $N_j = E_{Q^{*k}} \chi_j$ (so that $c_j = N_j / d_j$), we see that the corresponding dimension and character values for ρ_4 would be $d_4 = d_2 + d_3$ and $N_4 = N_2 + N_3$. Thus, writing

$$\frac{N_2 + N_3}{d_2 + d_3} \;=\; \frac{d_2}{d_2 + d_3} \left(\frac{N_2}{d_2} \right) \;+\; \frac{d_3}{d_2 + d_3} \left(\frac{N_3}{d_3} \right) ,$$

we see that we will similarly have

$$\left| c_4 \;-\; |c_1|^2 \right| \;=\; O(1/n^2) \, .$$

The conclusion about $v_{n,k}$ is thus valid for any $A_n < G_n < U(n)$. ■

Lemma 3 allows us to complete the proof of Theorem 1. Indeed, using equation (2), we have that

$$\|Q_n^{*k} - \lambda_n\|_{\mathrm{var}} \geq 1 \;-\; \frac{4 \left(1 + B \left(\frac{\log n}{n} \right) e^{-2\Re e\,(\alpha) t} + (e^{-\Re e\,(\alpha) t} + A \left(\frac{\log n}{n} \right)) \right)}{\left(e^{-\Re e\,(\alpha) t} + A \left(\frac{\log n}{n} \right) \right)^2}$$

$$\geq 1 - (\mathrm{const}) e^{\Re e\,(\alpha) t} - (\mathrm{const}) \left(\frac{\log n}{n} \right) ,$$

establishing Theorem 1.

6. On Proving the Conjecture.

Theorem 1 is unsatisfying in that it only provides a lower bound on the time to uniformity for the random walks being considered, and does not establish the existence of a cut-off phenomenon. It is reasonable to ask about the possibility of proving Conjecture 2, and thus establishing the full result.

We believe that the methods used in this paper could be used to prove Conjecture 2, though we have been unable to do so. The idea is to use the Upper Bound Lemma (1), and to show that the sum there is small if $k = \alpha^{-1} n \log n + tn$ with $t >> 0$. Now, we already know that in this case the first term $(d_1)^2 (c_1)^{2k}$ will be small. But Lemmas 3 and 4 assert that also $v_{n,k}$ is small, which amounts to saying that $d_2(c_2)^k \approx (d_1 |c_1|^k)^2/2$, and similarly $d_3(c_3)^k \approx (d_1|c_1|^k)^2/2$. This means that the next two terms in the Upper Bound Lemma sum are correspondingly smaller. It *appears* that this pattern continues through all of the "small" irreducible representations. If so, then the Upper Bound Lemma sum can be easily summed (by comparison to a geometric sum) and shown to be small, establishing Conjecture 2. On the other hand, to carry out this program it is insufficient to consider $v_{n,k} = Var_{Q_n^{*k}}(\chi_1)$. Rather, variances of other characters have to be computed, and this involves considering how representations $\rho_i \otimes \overline{\rho_i}$ split for $i > 1$. This program appears to be somewhat involved, but still promising.

Acknowledgement: I am very grateful to Persi Diaconis for many valuable discussions. I sincerely thank Ursula Porod for very insightful comments about an earlier draft of this manuscript.

REFERENCES

[AD] D. Aldous and P. Diaconis (1987), *Strong Stopping Times and Finite Random Walks,* Adv. Appl. Math. **8**, 69-97.

[BD] D. Bayer and P. Diaconis (1992), *Trailing the Dovetail Shuffle to its Lair,* Ann. Prob. **2**, 294-313.

[D] P. Diaconis (1988), *Group Representations in Probability and Statistics,* IMS Lecture Series volume **11**, Institute of Mathematical Statistics, Hayward, California.

[DS] P. Diaconis and M. Shashahani (1981), *Generating a Random Permutation with Random Transpositions,* Z. Wahrscheinlichkeitstheorie Verw. Gebiete **57**, 159-179.

[FOW] L. Flatto, A.M. Odlyzko, and D.B. Wales (1985), *Random shuffles and group representations,* Ann. Prob. **13**, 154-178.

[H] M. Hildebrand (1992), *Generating random elements in $SL_n(F_q)$ by random transvections,* J. Alg. Comb. **1**, 133-150.

[JG] G. James and A. Kerber (1981), *The Representation Theory of the Symmetric Group,* Addison-Wesley, Reading, Massachusetts.

[K] B.M. Kloss (1959), *Limiting distributions on bicompact topological groups,* Th. Prob. Appl. **4**, 237-270.

[P] U. Porod (1992), *Random Reflections. Analysis of a Random Walk on $O(N)$,* Tech. Rep., Department of Mathematical Sciences, Johns Hopkins University.

[R] J.S. Rosenthal (1991), *Random Rotations: Characters and Random Walks on $SO(N)$,* Ann. Prob., to appear.