Strong Uniform Times and Finite Random Walks*

DAVID ALDOUS† AND PERSI DIACONIS‡

University of California, Berkeley, California 94720, and
Stanford University, Stanford, California 94305

There are several techniques for obtaining bounds on the rate of convergence to
the stationary distribution for Markov chains with strong symmetry properties, in
particular random walks on finite groups. An elementary method, strong uniform
times, is often effective. We prove such times always exist, and relate this method to
coupling and Fourier analysis.

1. INTRODUCTION

Consider a finite Markov chain, in particular one with strong “symmetry”
properties such as a random walk on a group. Let \( \pi_n \) be the distribution
after \( n \) steps. The theory of convergence of \( \pi_n \) to a stationary distribution
\( \pi \), and of the asymptotic rate of convergence, is well-understood. The
non-asymptotic behavior is harder. Two useful notions of distance between
distributions are separation distance and variation distance, defined in
Section 2. Write \( s(n) \), \( d(n) \) for these distances between \( \pi_n \) and \( \pi \). Where \( \pi_n \)
cannot be calculated explicitly, there are three available techniques for
obtaining explicit upper bounds on \( s(n) \) or \( d(n) \) for large finite \( n \):

(a) strong uniform times,
(b) coupling,
(c) Fourier analysis.

Diaconis [11] provides an introduction to this area. Examples-oriented

*Prepared under the auspices of National Science Foundation Grant MCS80-24649. Also
issued as Technical Report No. 59, Dept. of Statistics, University of California, Berkeley,
California, February 1986.
†Research supported by National Science Foundation Grant MCS80-02698.
‡Research supported by National Science Foundation Grant MCS80-24649.
accounts of the particular techniques can be found in

(a) Aldous and Diaconis [3],
(b) Aldous [1],
(c) Letac [22], Takács [34], Diaconis and Shahshahani [14].

Section 8 contains a list of all examples of random walks on groups for which good results are known. Though the authors find the examples more interesting and more deep than the theory, there are enough examples known that it seems worthwhile to set down some more detailed theory. This paper addresses two theoretical issues.

(I) How good are these techniques in principle? How do they fit with the notions of distance between distributions? How are they related to each other? In Section 3 it is shown that a minimal strong uniform time always exists and provides a sharp estimate of separation distance: this is analogous to the known relation between coupling and variation distance, reviewed in Section 4. Fourier analysis bounds are described in Section 6.

(II) What are the intrinsic properties of the sequences $d(n), s(n), n \geq 1$, and their relation? How are they related to the second-largest eigenvalue, which controls the asymptotic rate of convergence? How are they affected by symmetry conditions? Section 5 describes a hierarchy of symmetry conditions and their consequences. Section 7 describes the threshold phenomenon: one might expect the distances $d(n), s(n)$ to decrease smoothly from their initial value (near 1) to 0 as $n$ increases, but for almost all natural examples with strong symmetry one finds instead a relatively sharp transition from near 1 to near 0 around some number $a_d, a_s$ of steps. This phenomenon is related to high multiplicity of the second-longest eigenvalue.

As mentioned before, the practical application of the three techniques has been amply discussed in the literature. Here we merely use a very simple running example (random walk on the $N$-dimensional cube, Sect. 2) to illustrate the techniques. The list of examples in Section 8 indicates the effectiveness of the techniques and the sharpness of the theoretical results.

2. DEFINITIONS

Let $P = (P(i, j))$ be an irreducible aperiodic transition matrix on a finite state space $I = \{i, j, k, \ldots\}$. Fix $i_0 \in I$, and let $i_0 = X_0, X_1, X_2, \ldots$, be the associated Markov chain with initial state $i_0$. Let $\pi_n$ be the probability distribution $\pi_n(j) = P^n(i_0, j) = P(X_n = j)$. It is classical that there exists a unique stationary distribution $\pi$ for $P$ and that,

$$\pi_n(j) \to \pi(j) \quad \text{as} \quad n \to \infty; \quad \text{each} \quad j \in I.$$
A natural measure of distance between two distributions $Q_1, Q_2$ on $I$ is total variation:
\[
\|Q_1 - Q_2\| = \max_{A \subseteq I} |Q_1(A) - Q_2(A)| = \frac{1}{2} \sum_{i \in I} |Q_1(i) - Q_2(i)|.
\]

So given $P$ and $i_0$ we can define
\[
d(n) = \|\pi^n - \pi\|.
\]

Another notion of "distance", though not a metric, is separation:
\[
s(Q_1, Q_2) = \max_i \left(1 - \frac{Q_1(i)}{Q_2(i)}\right).
\]

Equivalently, $s(Q_1, Q_2)$ is the smallest $s \geq 0$ such that $Q_1 = (1 - s)Q_2 + sV$ for some distribution $V$. Given $P$ and $i_0$ define
\[
s(n) = s(\pi^n, \pi).
\]

For any $Q_1, Q_2$,
\[
0 \leq \|Q_1 - Q_2\| \leq s(Q_1, Q_2) \leq 1,
\]
the middle inequality because
\[
\|Q_1 - Q_2\| = \sum_{i: Q_2(i) \geq Q_1(i)} (Q_2(i) - Q_1(i))
\]
\[
\leq \sum_{i: Q_2(i) \geq Q_1(i)} Q_2(i) \max_j \left(1 - \frac{Q_1(j)}{Q_2(j)}\right).
\]

In particular,
\[
d(n) \leq s(n).
\]

There is no general reverse inequality: if $Q_2$ is uniform on $I$ and $Q_1$ is uniform on $I - \{i_1\}$ then $\|Q_1 - Q_2\| = 1/|I|$ while $s(Q_1, Q_2) = 1$.

The quantities $d(n), s(n)$ depend on the initial state $i_0$. Taking maxima over $i_0$ leads to maximal variation $d^*(n)$ and maximal separation $s^*(n)$:
\[
d^*(n) = \max_{i_0} \|P^n(i_0, \cdot) - \pi(\cdot)\|
\]
\[
s^*(n) = \max_{i_0} s(\pi^n(i_0, \cdot), \pi(\cdot)) = \max_{i, j} \left(1 - \frac{P^n(i, j)}{\pi(j)}\right).
\]

Alternatively, under symmetry conditions discussed in Section 5 (in particu-
lar, for random walks on groups), $d(n)$ and $s(n)$ do not depend on $i_0$, so $d^*(n) = d(n)$, $s^*(n) = s(n)$. The quantities $d^*(n), s^*(n)$ are related to classical coefficients of ergodicity (Seneta [30], Iosifescu [21]), but aside from the standard submultiplicativity property (3.7), (4.5) our concerns are different.

For our running example let $I = \{0,1\}^N$, regarded as the vertices of the unit cube in $N$ dimensions. For $i = (i_1, \ldots, i_N), \hat{i} = (\hat{i}_1, \ldots, \hat{i}_N) \in I$ let

$$P(i, \hat{i}) = \begin{cases} 1 & \text{if } \sum_s |i_s - \hat{i}_s| = 0 \text{ or } 1, \\ 0 & \text{otherwise.} \end{cases}$$

(2.6)

This is essentially the simple symmetric random walk on the cube, with a possibility of "no move" included to make the chain aperiodic. The stationary distribution is uniform on $I$. The $n$-step transition probabilities can be calculated explicitly (see (6.6): and Letac and Takács [23] discuss more general random walks on the cube), though we will only be concerned with the bounds derivable from the three techniques.

3. STRONG UNIFORM TIMES AND SEPARATION DISTRIBUTION

As in Section 2, let $(X_n: n \geq 0)$ be a Markov chain with initial state $i_0$ and stationary distribution $\pi$, and let $s(n)$ be the separation (2.2).

DEFINITION (3.1). A strong uniform time $T$ is a randomized stopping time for $(X_n: n \geq 0)$ such that

(a) $P(X_k = i | T = k) = \pi(i)$ for all $0 \leq k < \infty, i \in I$.

Note that one can reformulate (a) as

(b) $P(X_k = i | T \leq k) = \pi(i)$ for all $0 \leq k < \infty, i \in I$

or as

(c) $X_T$ has distribution $\pi$ and is independent of $T$.

PROPOSITION (3.2). (a) If $T$ is a strong uniform time for $\{X_n\}$ then

$$s(n) \leq P(T > n); \quad n \geq 0.$$  

(3.3)

(b) Conversely, there exists a strong uniform time $T$ such that (3.3) holds with equality.

Proof. (a) Conditional on $\{T = m\}, X_m$ has distribution $\pi$, and the conditional distribution of each $X_n, n \geq m$, is $\pi$. So the conditional distri-
bution $X_n$ given $\{T \leq n\}$ is $\pi$. So

$$P(X_n = i) \geq P(X_n = i, T \leq n) = P(T \leq n)\pi(i) = [1 - P(T > n)]\pi(i).$$

(b) Define $a_n = \min_i \pi_n(i) / \pi(i)$. Let $k$ be the smallest integer such that $a_k > 0$. Define $T$ so that

$$P(T < k) = 0, \quad P(T = k | X_k = i) = a_k \pi(i) / P(X_k = i), \quad i \in I.$$ 

This implies $P(T = k, X_k = i) = a_k \pi(i)$. Inductively for $n > k$ make

$$P(T = n | X_n = i, T > n - 1) = \frac{a_n - a_{n-1}}{(\pi_n(i)/\pi(i)) - a_{n-1}}, \quad i \in I. \quad (3.4)$$

This makes sense because the right side is in $[0, 1]$ by definition of $a_n$. We shall show

$$P(X_n = i, T = n) = \pi(i)(a_n - a_{n-1}), \quad n \geq k, \quad i \in I. \quad (3.5)$$

This implies $T$ is a strong uniform time. Moreover, it implies $P(T = n) = (a_n - a_{n-1})$ and hence $P(T \leq n) = a_n = 1 - s(\pi_n, \pi)$, giving the desired quality in (3.3).

Equation (3.5) is proved inductively. For $n > k$,

$$P(X_n = i, T = n) = P(T = n | X_n = i, T > n - 1)P(X_n = i, T > n - 1)$$

$$= \frac{a_n - a_{n-1}}{(\pi_n(i)/\pi(i)) - a_{n-1}}$$

$$\times [P(X_n = i) - P(X_n = i, T \leq n - 1)]. \quad (3.6)$$

Now if (3.5) holds for integers less than $n$, then

$$P(X_n = i, T \leq n - 1) = \pi(i)a_{n-1}.$$ 

Using this in (3.6) give (3.5).

Remarks. (i) Part (a) explains the point of strong uniform times: by constructing one and estimating the tail of its distribution, you can bound $s(n)$ and hence $d(n)$. Aldous and Diaconis [3] treat four examples in detail. At the end of the section we illustrate the technique on our running example.

(ii) Part (b) shows that the notion of strong uniform time fits perfectly with the notion of separation distance. The construction of the minimal $T$ is purely theoretical, in that it requires complete knowledge of each $\pi_n$. 
(iii) Although explicit construction of strong uniform times is usually possible only for chains with strong symmetry, comparison lemmas like (3.8) sometimes can extend results to less symmetric chains.

(iv) Strong uniform times are related to the fundamental idea in the modern treatment of asymptotics of general state-space Markov chains. Nummelin [27, Sect. 4.4], shows that under an irreducibility condition there exists a reference distribution $\nu$ such that, for any initial state $x_0$, there is a random time $T$ such that $P(X_T \in A \mid T = n) = \nu(A)$, $n \geq 1$. This result enables standard renewal theory to be used to prove convergence of $X_n$ in total variation.

Here are two useful results involving the maximal separation $s^*(n)$ defined at (2.5). First, it is submultiplicative.

**Lemma (3.7).** $s^*(m + n) \leq s^*(m)s^*(n)$; $m, n \geq 1$.

**Proof.** From the definition we can write

$$P^n(i, j) = (1 - s^*(n))\pi(j) + s^*(n)V_n(i, j),$$

where $V_n$ is some transition matrix. A simple calculation gives

$$P^{m+n}(i, j) = [1 - s^*(m)s^*(n)]\pi(j) + s^*(m)s^*(n)\sum_k V_n(i, k)V_m(k, j).$$

$$\geq [1 - s^*(m)s^*(n)]\pi(j).$$

Second is a comparison lemma.

**Lemma (3.8).** Let $P_1, P_2$ be transition matrices on $I$, and let $s^*_1(n)$, $s^*_2(n)$ be the maximal separations (2.5). Suppose

(i) there exists $c > 0$ such that $P_2(i, j) \geq cP_1(i, j)$ for all $i, j \in I$,

(ii) $P_1P_2 = P_2P_1$.

Then $s^*_2(n) \leq \sum_{b=0}^n s^*_1(b)(\frac{n}{b})c^b(1 - c)^{n-b}$; $n \geq 1$.

**Proof.** By (i) we can write $P_2 = cP_1 + (1 - c)V$ for some transition matrix $V$. By (ii), $P_1V = VP_1$. So

$$P_2^n = \sum_{b=0}^n \binom{n}{b}c^b(1 - c)^{n-b}P_1^bV^{n-b}.$$
identical stationary distribution,  
\[ s^*(V_1 V_2) \leq s^*(V_1), \]
\[ s^*(\sum q_i V_i) \leq \sum q_i s^*(V_i) \]  for any distribution \((q_i)\).

**Remark.** Roughly, if \(n\) steps suffice to make \(s_1\) small, then \(n/c\) steps suffice to make \(s_2\) small.

Finally, here is a construction of a strong uniform time for the random walk on the cube. Let \(\theta_1, \theta_2, \ldots\) be independent, uniform on \(\{0, 1, \ldots, N\}\). Then the random walk \(X(n) = (X_i(n): 1 \leq i \leq N)\) with \(X(0) = 0\) can be defined by

\[ X_i(n) = X_i(n - 1) + 1_{(i=\theta_n)}, \quad 1 \leq i \leq N, \]

where addition is modulo 2. We now describe a scheme for “checking off coordinates.” Initially no coordinates are checked. At a general step, if coordinate \(\theta_n\) is already checked, or if \(\theta_n = 0\), then no check is made. If \(\theta_n\) is not already checked, then a fair coin is tossed: if “heads” then coordinate \(\theta_n\) is checked; if “tails” then a different unchecked coordinate (not 0) is chosen uniformly and checked. This last rule changes when there is exactly one unchecked coordinate, \(j\) say. Then wait until the next time \(n\) that \(\theta_n = j\) or 0, and at that time check coordinate \(j\).

Formally, let \(D_n \in \{H, T\}\) be the coin tosses and define \(C(n) = (C_i(n): 1 \leq i \leq N)\) as follows [\(C_i(n) = 1\) means coordinate \(i\) is checked at or before time \(n\)].

\[ C(0) = 0. \]

If \(\theta_n = 0\) or \(C_{\theta_n}(n - 1) = 1\) then \(C(n) = C(n - 1)\). If \(C_{\theta_n}(n - 1) = 0\) and \(|\{i: C_i(n - 1) = 0\}| = m \geq 2\) then:

- if \(D_n = H\) then \(C_i(n) = C_i(n - 1) + 1_{(i=\theta_n)}\)
- if \(D_n = T\) then \(C_i(n) = C_i(n - 1) + 1_{(i=\xi_n)}\), where \(\xi_n\) is uniform on \(\{i: C_i(n) = 0\} \setminus \{\theta_n\}\).

If \(|\{i: C_i(n - 1) = 0\}| = \{j\}\) then:

- if \(\theta_n \notin \{0, j\}\) then \(C(n) = C(n - 1)\),
- if \(\theta_n \in \{0, j\}\) then \(C_i(n) = C_i(n - 1) + 1_{(i=j)}\).

Having described the construction, let \(T\) be the first time that all coordinates have been checked:

\[ T = \min\{n: C(n) = (1, 1, \ldots, 1)\}. \]

Then \(T\) is a strong uniform time. To prove this, one shows that for each
$B \subseteq \{1, \ldots, N\}$ and each $n \geq 1$,

conditional on $B = \{i: C_i(n) = 1\}$, the distribution of
{$X_i(n); i \in B$} is uniform on $(0, 1)^B$.

This is a straightforward induction on $n$. Applying with $B = \{1, \ldots, N\}$, one sees that $T$ is a strong uniform time.

To estimate the tail of the distribution of $T$, write

$$T = W_N + W_{N-1} + \cdots + W_1,$$

where $W_k$ is the number of steps during which there were $k$ unchecked coordinates. Then the $W_k$ are independent with geometric distributions

$$P(W_k = m) = \frac{k}{N+1} \left(1 - \frac{k}{N+1}\right)^{m-1}, \quad m \geq 1, \quad k \geq 2$$

$$P(W_1 = m) = \frac{2}{N+1} \left(1 - \frac{2}{N+1}\right)^{m-1}, \quad m \geq 1.$$

So

$$ET = \frac{N+1}{2} + \sum_{k=2}^N \frac{N+1}{k}$$

$$\leq (N+1)(1 + \log N),$$

and

$$\text{Var}(T) = \left(\left(1 - \frac{2}{N+1}\right)/\left(\frac{2}{N+1}\right)^2\right)$$

$$+ \sum_{k=2}^N \left(\left(1 - \frac{k}{N+1}\right)/\left(\frac{k}{N+1}\right)^2\right)$$

$$\leq (N+1)^2,$$

and Chebyshev's inequality gives

$$P(T \geq (N+1)(\log N + c)) \leq \frac{1}{(c-1)^2}, \quad c > 1.$$
Remarks.  (i) The random variable $T$ is essentially the random variable in the coupon-collector’s problem; the analysis can be sharpened to show that for large $N$, $P(T > N(\log N + c)) \approx 1 - \exp(-e^{-c})$.

(ii) Many constructions of strong uniform times involve a “checking” process, an idea originated by Andre Broder.

4. COUPLING AND VARIATION DISTANCE

Coupling is a more standard technique for obtaining explicit bounds. For completeness we give the definition and basic theorem. Keep the setting of Section 2: $P$ is a transition matrix on $I$, $\pi$ the stationary distribution, $i_0$ a prescribed state, and $d(n)$ the variation distance (2.1).

DEFINITION (4.1). A coupling consists of processes \{ $X_n; n \geq 0$\}, \{ $Y_n; n \geq 0$\} and a random coupling time $T$ such that

(a) $X_n$ is the Markov chain with transition matrix $P$ and initial state $i_0$,

(b) $Y_n$ is the Markov chain with transition matrix $P$ and initial distribution $\pi$.

(c) $X_n = Y_n, n \geq T$.

THEOREM (4.2).

(a) For any coupling,

\[ d(n) \leq P(T > n); n \geq 1. \]  \hspace{1cm} (4.3)

(b) Conversely, there exists a coupling such that (4.3) holds with equality.

Theorem 4.2 was proved by Griffeath [18]; see also Pitman [29], Goldstein [17], Thorisson [37]. Our Proposition 3.2 is closely analogous to Theorem 4.2: the strong uniform time technique fits with separation distance in exactly the same way that the coupling technique fits with the total variation distance. Aldous [1] gives several examples where couplings are constructed and the easy part (a) is used to bound variation distance. Later we illustrate the technique with our running example.

From a formal viewpoint, strong uniform times are a special case of coupling, as the next result shows.

PROPOSITION (4.4). Let $T$ be a strong uniform time for the Markov chain $(X_n)$ with starting state $i_0$. Then there exists a coupling with coupling time $T$.

Proof. Given $T$ and $(X_n)$, construct $(Y_n)$ as follows. On each nonnull \{ $T = m$ \}, define $Y_n^{(m)} = X_n, n \geq m$. Conditional on \{ $T = m$ \} the future process \{ $Y_n^{(m)}; n \geq m$ \} is distributed as the stationary Markov chain and so
can be extended to \( \{Y_n^{(m)}: n \geq 0\} \) as the stationary Markov chain. Define \( Y_n = Y_n^{(m)} \) on \( \{T = m\} \), for each \( m \). Then \( (Y_n) \) is the stationary Markov chain and \( Y_n = X_n, n \geq T \), so we have a coupling.

A second connection has been observed by Hermann Thorisson. If \( (X_n) \) is the Markov chain starting at \( i_0 \), \( (Y_n) \) the stationary chain, and \( T \) a strong uniform time for \( (X_n) \), then

\[
(T, X_T, X_{T+1}, \ldots) \overset{D}{=} (T, Y_T, Y_{T+1}, \ldots),
\]

which is called a distributional coupling in Thorisson [37]. Despite these formal connections, in concrete examples the constructions of strong uniform times and couplings are rather different.

Lemmas 3.7 and 3.8 for maximal separation have analogs for maximal variation \( d^*(n) \) defined at (2.4). Let us just observe that, although \( d^* \) is not submultiplicative, a related quantity is.

**Lemma (4.5).** Define \( d(n) = \max_{i_1, i_2} \sum_j |P^n(i_1, j) - P^n(i_2, j)| \). Then

(a) \( d(m + n) \leq d(m)d(n), m, n \geq 1, \)

(b) \( \frac{1}{2}d(n) \leq d^*(n) \leq d(n), n \geq 1. \)

The proof is straightforward.

Finally, here is a construction of a coupling for the random walk on the cube. As in Section 3 let \( X(0) = 0 \) and define the random walk \( X(n) \) by

\[
X_i(n) = X_i(n - 1) + 1_{i = \theta_n}, \quad 1 \leq i \leq N,
\]

where \( (\theta_n) \) are uniform on \( \{0, 1, \ldots, N\} \) and addition is modulo 2. Let \( Y(0) \) be uniform on \( \{0, 1\}^N \) and define \( Y(n) \) as follows.

If \( \theta_n \geq 1 \) and \( Y_i(n - 1) = X_i(n - 1) \) then \( Y_i(n) = Y_i(n - 1) + 1_{i = \theta_n}, 1 \leq i \leq N. \) Otherwise, \( Y_i(n) = Y_i(n - 1) + 1_{i = \xi}, 1 \leq i \leq N, \) where \( \xi \) is chosen uniformly from \( \{i: i = 0 \text{ or } X_i(n - 1) \neq Y_i(n - 1)\} \setminus \{\theta_n\}. \)

Roughly, \( Y \) changes in the same coordinate as \( X \) if they already agree in that coordinate, and in a different coordinate if they disagree. It is easy to verify this is a coupling, with coupling time

\[
T = \min\{n: X(n) = Y(n)\}.
\]

To estimate the tail of \( T \), observe that \( D_n = \left|\{i: X_i(n) \leq Y_i(n)\}\right| \) is the
Markov chain on states \( \{0, 1, \ldots, N\} \) with transition probabilities

\[
P(d, d) = 1 - \frac{d + 1}{N + 1} \quad d \geq 1,
\]

\[
P(d, d - 1) = \frac{2}{N + 1} \quad d \geq 1,
\]

\[
P(d, d - 2) = \frac{d - 1}{N + 1} \quad d \geq 2.
\]

Now \( T \) is stochastically dominated by the first passage time \( T^* \) for \( D_n \) from \( N \) to 0. Routine but tedious arguments show that \( T^* \) is around \( \frac{1}{2}N \log N \), for large \( N \).

5. Symmetry Conditions

So far, the only assumptions on the transition matrix \( P(i, j) \) on \( I \) are that \( P \) be irreducible and aperiodic. But all our examples have extra symmetry properties, and under symmetry properties one can prove extra theoretical results. In this section we give definitions and elementary consequences.

**Definition (5.1).** \( P \) is **doubly stochastic** if \( \Sigma_j P(i, j) = 1 \) for all \( j \).

This is precisely the condition under which the stationary distribution \( \pi \) is the uniform distribution \( U(i) = 1/|I| \).

**Definition (5.2).** \( P \) is **symmetric** if \( P(i, j) = P(j, i) \) for all \( i, j \). \( P \) is **quasi-symmetric** if there exists a bijection \( \xi: I \to I \) such that \( P(i, j) = P(\xi(j), \xi(i)) \) for all \( i, j \).

Note that quasi-symmetric implies doubly stochastic.

**Definition (5.3).** \( P \) is **transitive** if there exists a group \( G \) of bijections \( g: I \to I \) such that

(a) \( P(i, j) = P(g(i), g(j)) \); \( i, j \in I, g \in G \).

(b) \( G \) acts transitively on \( I \).

Informally, transitivity means "all initial states \( i_0 \) are equivalent." In particular, the maximal separation \( s^*(n) \) of (2.5) coincides with the separation \( s(n) \) starting from any \( i_0 \); and similarly \( d^*(n) = d(n) \).

A natural class of examples are random walks on groups. Let \( Q \) be a probability distribution on a finite group \( G \). The random walk on \( G \) with
step distribution $Q$ is
\[ X_n = \xi_n \ast \xi_{n-1} \ast \cdots \ast \xi_1 \quad (X_0 = \text{identity}), \]
(5.4)
where $(\xi_i)$ are independent with distribution $Q$. So $\{X_n\}$ is the Markov chain on $I = G$ with
\[ P(g_1, g_2) = Q(g_2g_1^{-1}). \]
(5.5)

Our background assumptions of irreducibility and aperiodicity correspond to the assumption that $Q$ is not supported on any coset of any proper subgroup. Clearly $P$ is automatically transitive (under the action of $G$ on itself) and quasisymmetric (under $g \to g^{-1}$); $P$ is symmetric iff $Q(g) = Q(g^{-1})$ for all $g \in G$.

Another natural class of examples are random walks on graphs. For a finite undirected graph $\Gamma$ with vertex-set $V$ and edge-set $E$, the natural random walk on $\Gamma$ is the Markov chain with
\[ P(v_1, v_2) = \frac{1}{\gamma_{v_1}} \quad \text{if } (v_1, v_2) \in E 
= 0 \quad \text{if not} \]
(5.6)
where $\gamma_v$ is the degree of $v$. Here $P$ is irreducible and aperiodic iff $\Gamma$ is connected and not 2-colorable. And $P$ is doubly stochastic iff $\Gamma$ is regular, that is $\gamma_v = \gamma$ for all $v$. Given these conditions, $P$ is automatically symmetric; it is transitive iff $\Gamma$ is vertex-transitive in the sense of Biggs [6].

Our two final symmetry conditions concern the case of a random walk on a group $G$ with step distribution $Q$.

**Definition** (5.7). $Q$ is constant on conjugacy classes if $Q(g) = Q(hgh^{-1})$; all $g, h \in G$.

**Definition** (5.8). Let $K$ be a subgroup of $G$. Call $Q$ $K$-biinvariant if $Q(g) = Q(k_1gk_2)$; all $g \in G$, $k_1, k_2 \in K$. Call $(G, K)$ a Gelfand pair if convolution of $K$-biinvariant distributions is commutative. Write $I = G/K$ for the space of cosets. Given a $K$ biinvariant distribution $Q$ on a Gelfand pair $(G, K)$, the associated random walk is the Markov chain on $I$ with transitions
\[ P(g_1K, g_2K) = Q(g_1^{-1}g_2K). \]

See Letac [22], Diaconis and Shahshahani [15] for discussion and examples of random walks on Gelfand pairs.

Our running example, random walk on the $N$-cube, satisfies all these definitions. It is the random walk on the group $G = \{0, 1\}^N$ associated with the step distribution $Q(0, 0, \ldots, 0) = Q(1, 0, 0, \ldots, 0) = \cdots = \cdots = \cdots$
Q(0, 0, \ldots, 1) = 1/(N + 1). Since $G$ is Abelian, (5.7) is automatically satisfied. Next, one can consider $\{0, 1\}^N$ as a graph in the natural way. Then the process is the random walk on the graph in the sense (5.6). Finally, let $\hat{G}$ be the group of automorphisms of the graph, and $K$ the subgroup of automorphisms which fix $0$. Then one can identify $I = \{0, 1\}^N$ with $\hat{G}/K$ and hence identify the process with a random walk on the Gelfand pair ($\hat{G}, K$).

We now start on the elementary consequences of these symmetry conditions. In the case where the matrix $P$ is symmetric (5.2), it has real eigenvalues which (by irreducibility and aperiodicity) satisfy

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|I|} > -1.$$  

(5.9)

One can bound the maximal total variation $d^*(n)$ in terms of these eigenvalues.

PROPOSITION (5.10). If $P$ is symmetric then

$$d^*(n) \leq \frac{1}{2} |I|^{1/2} \left\{ \sum_{k \geq 2} \lambda_k^{4n} \right\}^{1/4}.$$  

Proof. Matrix theory says $P^n(i, i) = \sum_k a_{i, k} \lambda_k^n$, where $\sum_k a_{i, k}^2 = 1$ and $a_{i, 1} = 1/|I|$. Hence

$$P^{2n}(i, i) - \frac{1}{|I|} = \sum_{k \geq 2} a_{i, k} \lambda_k^{2n} \leq \left\{ \sum_{k \geq 2} \lambda_k^{4n} \right\}^{1/2}$$  

by Cauchy–Schwarz. (5.11)

Now

$$\left( \sum_j \left| P^n(i, j) - \frac{1}{|I|} \right| \right)^2 \leq |I| \sum_j \left( P^n(i, j) - \frac{1}{|I|} \right)^2$$  

by Cauchy–Schwarz

$$= |I| \sum_j (P^n(i, j))^2 - 1$$

$$= |I| P^{2n}(i, i) - 1$$  

by symmetry

$$\leq |I| \left\{ \sum_{k \geq 2} \lambda_k^{4n} \right\}^{1/2}$$  

by (5.11),

and the result follows from the definition of $d^*(n)$.

Our next result concerns the transitive case. Recall $\pi_n$ is the distribution of $X_n$ resulting from some initial $i_0$; and $s(n)$ and $d(n)$ are the separation and variation distances between $\pi_n$ and the uniform distribution $U$, which by transitivity do not depend on $i_0$. By (2.3), $d(n) \leq s(n)$. Here is a kind of
converse. For $0 < \varepsilon < \frac{1}{4}$ define
\[ \phi(\varepsilon) = 1 - (1 - 2\varepsilon^{1/2})(1 - \varepsilon^{1/2})^2 \quad (5.12) \]
and observe that $\phi(\varepsilon)$ decreases as $\varepsilon$ decreases, and $\phi(\varepsilon) \sim 4\varepsilon^{1/2}$ as $\varepsilon \to 0$.

**PROPOSITION (5.13).** If $P$ is transitive and quasisymmetric on a set $G$, (in particular, for any random walk on a group $G$),
\[ d(2n) \leq s(2n) \leq \phi(2d(n)), \quad n \geq 1 \quad (5.14) \]
provided $d(n) < \frac{1}{8}$. The proof requires a lemma.

**LEMMA (5.15).** Let $(q_1, \ldots, q_J)$ be a probability distribution, and suppose $\sum_i |q_i - 1/J| \leq \varepsilon < \frac{1}{4}$. Then for any permutation $\pi \in S_J$, $\sum_i q_i q_{\pi(i)} \geq (1 - \phi(\varepsilon))/J$.

**Proof of Lemma 5.15.** Fix $\varepsilon^{-1} > \alpha > 2$. Let $I$ be the set of indices $i$ such that $|q_i - 1/J| > \alpha \varepsilon/J$. Then
\[ \sum_{j \in I} |q_i - 1/J| \geq \frac{\alpha |I|}{J}, \]
and so using the hypothesis we must have $|I| \leq J/\alpha$. Now consider the sum $\sum_i q_i q_{\pi(i)}$. For at least $J - 2J/\alpha$ terms, both $i$ and $\pi(i)$ are outside $I$, and so both $q_i$ and $q_{\pi(i)}$ are $> (1/J) - (\alpha \varepsilon/J)$. Thus
\[ \sum_i q_i q_{\pi(i)} \geq \left( J - \frac{2J}{\alpha} \right) \left( \frac{1}{J} - \frac{\alpha \varepsilon}{J} \right)^2 \]
\[ = \frac{1}{J} \cdot \left( 1 - \frac{2}{\alpha} \right) (1 - \alpha \varepsilon)^2. \]
Setting $\alpha = \varepsilon^{-1/2}$ yields the results.

**Proof of Proposition 5.13.** We need only prove the second inequality. Suppose $d(n) < \frac{1}{8}$. Then
\[ \pi^{2n}(j) = \sum_i P^n(i_0, i) P^n(i, j) \]
\[ = \sum_i P^n(i_0, i) P^n(\xi(j), \xi(i)) \quad \text{for some bijection } \xi: \ G \to G \]
\[ = \sum_i P^n(i_0, i) P^n(i_0, h_j \xi(i)) \quad \text{for some bijection } h_j: \ G \to G \]
\[ \geq (1 - \phi(2d(n))/|G|) \quad \text{by Lemma 5.15}, \]
and the result follows from the definition of $s(2n)$. 
Our next result is a comparison lemma for the case of random walks on groups.

**Lemma (5.16).** Let $Q_1, Q_2$ be distributions on a group $G$ satisfying
\[ Q_1 * Q_2 = Q_2 * Q_1. \] (5.17)
Let $s_1(n), s_2(n)$ be the separation distances for the associated random walks.

(a) Suppose there exists $c > 0$ such that $Q_2(g) \geq c Q_1(g)$ for all $g \in G$. Then
\[ s_2(n) \leq \sum_{b=0}^{n} s_1(b) \binom{n}{b} c^b (1 - c)^{n-b}, \quad n \geq 1. \]

(b) Let $Q = Q_1 * Q_2$ and let $s(n)$ be the associated separation distance. Then
\[ s(n) \leq s_1(n), \quad n \geq 1. \]

**Proof.** Part (a) is a specialization of Lemma 3.8, and part (b) is straightforward.

**Remarks.** The reason for mentioning this result here is that symmetry conditions (5.7), (5.8) are helpful in checking the commutativity condition (5.17). If $Q_1, Q_2$ are $K$-biinvariant distributions on a Gelfand pair then (5.17) holds by definition. If $Q$ is constant on conjugacy classes then
\[ Q * P = P * Q \quad \text{for all distributions $P$ on $G$.} \] (5.18)
Intuitively, one might expect (b) to remain true without restriction on $Q_1, Q_2$, since $Q$ should be “more random” than $Q_1$. But this is false. Diaconis and Shahshahani [13] give examples on the symmetric group $S_n$ ($n \geq 3$) of $Q_1, Q_2$ such that $Q_1 * Q_2$ is uniform while $(Q_1 * Q_2)^*k$ is not uniform for any $k$.

Our final results concern the following setting.

$G$ is a group, $S$ is a subset of $G$ which is not contained in any coset of any subgroup, $Q$ is the uniform distribution on $S$: $Q(g) = 1/|S|$, $g \in S$. In this setting, we show there is a construction of strong uniform times which can be related to the Fourier analytic bounds of the next section.

(5.19)

For $k \geq 1$ and $g \in G$ define
\[ B_g^k = \{(s_1, \ldots, s_k) \in S^k : s_1 s_2 \cdots s_k = g\}, \]
\[ b_k = \min_{g \in G} |B_g^k|. \]
Since $Q^*k(g) = |B^k_g|/|S|^k$, the definition of separation distance $s(k)$ gives
\[ s(k) = 1 - \frac{|G| \cdot b_k}{|S|^k}. \quad (5.20) \]

Pick subsets $\hat{B}^k_g \subset B^k_g$ such that $|\hat{B}^k_g| = b_k$. Write the random walk associated with $Q$ as
\[ X_n = \xi_n \xi_{n-1} \cdots \xi_1. \]

Provided $b_k > 0$ we can define a random time $T$ taking values in \{k, 2k, 3k, \ldots\} by
\[ T = k \min \{ j \geq 1 : (\xi_{jk}, \xi_{jk-1}, \ldots, \xi_{(j-1)k+1}) \in \hat{B}^k_g \text{ for some } g \in G \}. \quad (5.21) \]

Then
\[ P(T = k, X_k = g) = \frac{|\hat{B}^k_g|/|S|^k}{(1 - s(k))/|G|}, \]
and so $P(T = k) = 1 - s(k)$. It is straightforward to show $P(T = jk, X_{jk} = g) = s^{-1}(k)(1 - s(k))/(1 - s(k))/|G|, j \geq 1$, establishing the following result.

**Proposition (5.22).** In the setting of (5.19), define $T$ by (5.21) for some $k \geq 1$ such that $b_k > 0$ (equivalently, $s(k) < 1$). Then $T$ is a strong uniform time for which
\[ P(T > jk) = s^j(k); \quad j \geq 1. \]

**Remarks.** For any strong uniform time $T$,
\[ s(n) \leq P(T > n); \quad n \geq 1, \]
and Proposition 3.2 gave a rather non-explicit construction of $T$ attaining equality for all $n$. The point of construction (5.21) is to obtain a more explicit strong uniform time attaining equality at a specified time $k$. Note also a technical difference: Proposition 3.2 involved a randomized stopping time $T$ whereas (5.21) defines a natural stopping time $T$.

Specializing further, suppose that subset $S$ of (5.19) satisfies
\[ S = S^{-1}; \quad S \text{ contains the identity}. \quad (5.23) \]
Define the "diameter"
\[ \Delta = \min \{ n : \text{each } g \in G \text{ is of the form } g = h_1 h_2 \cdots h_n \text{ for some } h_i \in S \} \].

**Proposition (5.24).** Under conditions (5.19), (5.23),
\[ d(n) \leq \frac{1}{2} |G|^{3/4} \left( 1 - \frac{1}{20 \Delta^2 |S|} \right)^n, \quad n \geq 1. \]

*Remark.* Roughly, \( d(n) \) becomes small for \( n = O(\Delta^2 |S| \log |G|) \). This is the only reasonably general and reasonably explicit bound for \( d(n) \) known.

*Proof.* Write \( S_0 = S \setminus \{ \text{identity} \} \). Let \( Q_0 \) be the uniform distribution on \( S_0 \), and let \( P \) and \( P_0 \) be the transition matrices arising from \( Q \) and \( Q_0 \). So
\[ P = \frac{1}{|S|} I + \frac{|S| - 1}{|S|} P_0. \quad (5.25) \]

Using a result of Alon [5], it is observed by Aldous [2] that the second largest eigenvalue \( \lambda_2 \) of \( P_0 \) satisfies
\[ \lambda_2 \leq 1 - \frac{1}{20 \Delta^2 |S_0|}. \]

It easily follows from (5.25) that all the eigenvalues \( \lambda_k, k \geq 2, \) of \( P \) satisfy
\[ |\lambda_k| \leq 1 - \frac{1}{20 \Delta^2 |S|}, \]
and the proposition follows from Proposition 5.10.

6. **Fourier Analysis**

Group representation theory provides a powerful tool for obtaining bounds on separation and variation distance. We review the basic definitions below. Serre [31] is a good reference for representation theory, and Diaconis [11] details its application to the problem at hand.

A representation \( \rho \) of a finite group \( G \) is a homomorphism from \( G \) to \( GL(V) \). Thus \( \rho \) assigned matrices to group elements is such a way that \( \rho(gh) = \rho(g)\rho(h) \). The dimension of the vector space \( V \) is denoted by \( d_\rho \). The representation is irreducible if there is no non-trivial subspace \( W \) such that \( \rho(g)W \subseteq W \) for all \( g \in G \). The Fourier transform of a distribution \( Q \) at a representation \( \rho \) is defined by
\[ \hat{Q}(\rho) = \sum_{g \in G} Q(g)\rho(g). \]
Then \( \hat{Q}_1 \star \hat{Q}_2 = \hat{Q}_3 \hat{Q}_2 \). For the uniform distribution \( U \) on \( G \), \( \hat{U}(\rho) = 0 \) for each nontrivial irreducible representation. In what follows, \( \sum_{\rho} \) denotes summation over all nontrivial irreducible representations. For a matrix \( A \) write \( \|A\|^2 = \text{trace}(AA^*) \).

A useful upper bound for variation distance to the uniform distribution is given in Diaconis and Shahshahani [12].

**Lemma (6.1).** For a distribution \( Q \) on the finite group \( G \),

\[
\|Q^{**} - U\|^2 \leq \frac{1}{4} \sum_{\rho} d_{\rho} \|\hat{Q}^n(\rho)\|^2.
\]

So the problem of bounding variation distance \( d(n) \) for the random walk on \( G \) associated with \( Q \) reduces to the problem of estimating eigenvectors and eigenvalues of \( \hat{Q}(\rho) \). This technique is particularly useful under the symmetry conditions (5.7) or (5.8). If \( Q \) is constant on conjugacy classes then each \( \hat{Q}(\rho) \) is of the form \( cI \) for some constant \( c \) (\( I \) is the identity matrix). If \( Q \) is a random walk on a Gelfand pair then each \( \hat{Q}(\rho) \) is of the form \( (c' \circ 0) \). In either case, the matrix products reduce to products of numbers. Takács [34] discusses these cases and suggests some others. The following propositions show what can be proved by easy arguments from the Fourier inversion formula

\[
Q(g) = \frac{1}{|G|} \sum_{\rho} d_{\rho} \text{trace}(\hat{Q}(\rho)\rho(g^{-1})). \tag{6.2}
\]

**Proposition (6.3).** Suppose \( Q \) is constant on conjugacy classes of \( G \). For an irreducible representation \( \rho \) let \( \hat{Q}(\rho) = (1/d_{\rho}) \sum_{g} Q(g) \chi_{\rho}(g) \), where \( \chi_{\rho}(g) = \text{trace}(\rho(g)) \). Then

(a) \( 1 - |G|Q(g) \leq \sum_{\rho} d_{\rho} |\chi_{\rho}(g)\hat{Q}(\rho)| \leq \sum_{\rho} d_{\rho}^2 |\hat{Q}(\rho)|, \ g \in G, \)

(b) \( \|Q - U\|^2 \leq \frac{1}{4} \sum_{\rho} d_{\rho}^2 |\hat{Q}(\rho)|^2. \)

**Proposition (6.4).** Let \( Q \) be a \( K \)-biinvariant distribution on a Gelfand pair \( (G, K) \). For an irreducible representation \( \rho \) of \( G \), let \( s_{\rho}(g) = (1/|K|) \sum_{k \in K} \chi_{\rho}(kg) \), where \( \chi_{\rho}(g) = \text{trace}(\rho(g)) \). Define \( \overline{Q}(\rho) = \sum_{g \in G} Q(g) s_{\rho}(g) \). Then

(a) \( 1 - |G|Q(g) \leq \sum_{\rho} d_{\rho} |\overline{Q}(\rho)\overline{s}_{\rho}(g)| \leq \sum_{\rho} d_{\rho} |\overline{Q}(\rho)|, \)

(b) \( \|Q - U\|^2 \leq \frac{1}{4} \sum_{\rho} d_{\rho} |\overline{Q}(\rho)|^2. \)

For examples see Diaconis and Shahshahani [12, 15], Letac [22], Takács [34]. We treat our running example later.
The final bound in each (a) does not involve \( g \), and so is a bound on the separation \( s(Q, U) \). Applying the bounds to \( Q^n \) gives bounds on \( s(n) \) and \( d(n) \).

A curious fact which emerges from the Fourier analysis is that, under the hypotheses of either Proposition above,

\[
Q^k(\text{identity}) + Q^k(g) \geq \frac{2}{|G|}, \quad g \in G, \quad k \text{ even.} \tag{6.5}
\]

It is not clear if this remains true if \( Q \) is merely assumed to be symmetric \((Q(g) = Q(g^{-1}))\), although symmetry implies the weaker condition

\[
Q^k(\text{identity}) \geq \frac{1}{|G|}, \quad k \text{ even.}
\]

For our running example, \( G = \{0, 1\}^N \). For \( x = \{x_1, \ldots, x_N\} \), \( y = \{y_1, \ldots, y_N\} \in G \) write \(|x| = \sum x_u \), \( x \cdot y = \sum_{u \neq 0} x_u y_u \). For each \( y \) the map \( x \to y(x) = (-1)^{x \cdot y} \) is a representation, and as \( y \) varies we obtain all representations, which are all 1-dimensional. The Fourier transform of \( Q \) is \( \hat{Q}(y) = (1 - (2|y|/(N + 1))) \). Applying the inversion formula (6.2) for \( Q^k \) gives an explicit formula

\[
Q^k(x) = 2^{-N} \sum_{y} (-1)^{x \cdot y} \left( 1 - \frac{2|y|}{N + 1} \right)^k. \tag{6.6}
\]

For \( (\hat{Q}^k) \) defined in Proposition (6.3) we get

\[
(\hat{Q}^k(y)) = \left( 1 - \frac{2|y|}{N + 1} \right)^k.
\]

Then part (b) of Proposition 6.3 gives

\[
\|Q^k - U\|^2 \leq \frac{1}{4} \sum_{y \neq 0} \left( 1 - \frac{2|y|}{N + 1} \right)^{2k}
\]

\[
= \frac{1}{4} \sum_{j=1}^{N} \left( \frac{N}{j} \right) \left( 1 - \frac{2j}{N + 1} \right)^{2k}
\]

\[
\leq \frac{1}{4} \sum_{j=1}^{\infty} \frac{N^j}{j!} \exp \left( - \frac{4kj}{N + 1} \right)
\]

\[
\leq \frac{1}{4} \exp \left\{ Ne^{-4k/(N+1)} \right\}.
\]
So \( d(k) \leq \frac{1}{2} \exp\left(\frac{1}{2} Ne^{-4k/N+1}\right) \), or equivalently
\[
d\left[\frac{1}{4}(N + 1)(\log N + c)\right] \leq \frac{1}{2} \exp(-\frac{1}{2}e^{-c}).
\]

7. **The Threshold Phenomenon**

We have been concerned with the difference between the distribution after \( n \) steps and the limit distribution, and have described techniques which lead to explicit bounds for finite \( n \). The asymptotic behavior as \( n \to \infty \) is much simpler. In the finite Markov chain setting of Section 2, one typically finds
\[
P^n(i, j) = \pi(j) - 4)Kw,1^n \text{ as } n \to \infty, \quad (7.1)
\]
where \( \lambda_2 \) is the second largest (in absolute value) eigenvalue of \( P \) (the largest being \( \lambda_1 = 1 \)). Thus the variation distance \( d(n) \) for the chain started at \( i \) satisfies
\[
d(n) \sim c_i \exp(-n/\tau_e) \text{ as } n \to \infty; \text{ where } \tau_e = -1/\log|\lambda_2|. \quad (7.2)
\]
A similar result holds for separation, or any other reasonable notion of distance between distributions. In application of Markov processes to the physical sciences, \( \tau_e \) is called the *relaxation time* of the process.

But there is a different notion of asymptotics available. Most concrete examples involve a parameter \( N \), which can loosely be interpreted as "dimension," and one can study the behavior of the processes as \( N \to \infty \). Formally, consider a sequence \( G_N \) of groups with \( |G_N| \to \infty \), and consider the random walks associated with distributions \( Q_{\theta} \) on \( G_N \). Write \( d_N(n), s_N(n) \) for variation distance and separation distance. Say a sequence \( \tau_d(N) \) is a *variation threshold* for \((G_N, Q)\) if
\[
 d_N((1 - \epsilon)\tau_d(N)) \to 1 \quad \text{as } N \to \infty; \epsilon > 0 \text{ fixed.} \quad (7.3)
\]
The *separation threshold* \( \tau_s(N) \) is defined similarly. It turns out that most natural examples do have such thresholds. In this case, the behavior (for large \( N \)) of \( d_N(n), n \geq 1 \), is as follows: it stays near its maximum value 1 for almost \( \tau_d(N) \) steps, then cuts down to almost 0, ultimately converging to 0 at the exponential rate determined by the \( \tau_d(N) \) of (7.2). Thus for large \( N \) the threshold provides a sharp formalization of the notion "number of steps of the random walk required to approach the uniform distribution."
A table of examples where the thresholds are known is given in the next section. We do not know any useful general theorems which guarantee existence of thresholds. Informally, they are a consequence of a high degree of symmetry of the walks \((G_N, Q_N)\). This causes the dominant eigenvalue \(\lambda_2^{(N)}\) of (7.1) to have high multiplicity, \(\partial_N\) say, which can be interpreted as the “degree of symmetry” of \(Q_N\). Then, informally, the contribution of these eigenvalues to \(Q^{(N)}_* - U\) is of order
\[
\partial_N (\lambda_2^{(N)})^n \approx \partial_N \exp(-n/\tau_e(N)), \tag{7.4}
\]
which as a function of \(n\) becomes small around
\[
\hat{\tau}_N = \tau_e(N) \log \partial_N. \tag{7.5}
\]

This heuristic estimate \(\hat{\tau}_N\) gives the correct order of magnitude of the threshold in the known examples. As a simple instance, consider \((G, Q)\) satisfying (7.1), (7.2); that is
\[
Q^\ast^n(g) \sim u(g) - q(\gamma) as \quad n \to \infty. \tag{7.6}
\]
Now consider the “product” random walk \(X_N^{(N)}\) on the product group \(G^N = G \times \cdots \times G\) associated with the product distribution \(Q \times \cdots \times Q\). For this product random walk, the dominant eigenvalue is unaffected by \(N\), and there are \(N\) “degrees of symmetry,” so the heuristic estimate (7.5) is consistent with the result below.

**Proposition (7.7).** Suppose \((G, Q)\) satisfies (7.6). Then the product random walks \((G_N, Q_N)\) have \(\tau_e(N) = \tau_e\) and have thresholds
\[
\tau_s(N) = \frac{1}{2} \tau_e \log N, \quad \tau_d(N) = \tau_e \log N.
\]

Before proving this, let us state and prove what general results we have concerning thresholds.

**Proposition (7.8).** Consider the sequence of random walks associated with \((G_N, Q_N)\), where \(|G_N| \to \infty\).

(a) If the separation threshold and the variation threshold both exist, then we may take \(\tau_d(N) \leq \tau_s(N) \leq 2\tau_d(N)\).

(b) If the separation threshold [resp. variation threshold] exists then
\[
\frac{\tau_s(N)}{\tau_e(N)} \to \infty \quad \text{[resp.} \frac{\tau_d(N)}{\tau_e(N)} \to \infty].
\]
(c) If each $Q_N$ is symmetric and if the variation threshold exists then we may take $\tau_d(N) \leq \frac{3}{4} \tau_v(N) \log |G_N|$. 

Proof. Part (a) follows from Proposition 5.13, and part (c) from Proposition 5.10, since $\sum \lambda_k^n \leq |G_N| \lambda$. For part (b), fix $\varepsilon > 0$. The threshold for separation implies that, for $N$ sufficiently large,

$$s_N(2\tau_s(N)) \leq \varepsilon.$$ 

Then by Lemma 3.7,

$$s_N(2j\tau_s(N)) \leq \varepsilon^j, \quad j \geq 1.$$ 

But $s_N(2j\tau_s(n)) \sim a_N \exp(-2j\tau_s(N)/\tau_v(N))$ as $j \to \infty$; with $N$ fixed, by (7.1), and so

$$\exp(-2\tau_s(N)/\tau_v(N)) \leq \varepsilon.$$ 

Rearranging, $\tau_s(N)/\tau_v(N) \geq \frac{1}{2} \log(1/\varepsilon)$. Since $\varepsilon$ is arbitrary, $\tau_s(N)/\tau_v(N) \to \infty$ as claimed. The argument for variation is similar.

Proof of Proposition 7.7. Let $(a_g), \tau_s$ be as at (7.6). For fixed $N$ and $(g_1, \ldots, g_N)$,

$$P(X_n^N = (g_1, \ldots, g_N)) - |G|^{-N} \sim |G|^{-N} \sum_{i=1}^N a_{g_i}|\lambda|^n \text{ as } n \to \infty,$$

and so $\tau_s(N) = \tau_v$. Next, from the definition of separation,

$$s_N(n) = \max_{g_1, \ldots, g_N} \left\{ 1 - |G|^{-N} P(X_n^N = (g_1, \ldots, g_N)) \right\}$$

$$= 1 - \left( |G| \min_g P(X_n = g) \right)^N$$

$$= 1 - (1 - s(n))^N.$$

But from (7.6),

$$s(n) \sim \hat{a} \exp(-n/\tau_v) \quad \text{as } n \to \infty; \quad \hat{a} = \max(-a_g).$$

Calculus shows

$$s_N(c\tau_v \log N) \to 1 \quad \text{as } N \to \infty \quad (c < 1 \text{ fixed})$$

$$\to 0 \quad \text{as } N \to \infty \quad (c > 1 \text{ fixed}).$$

So $\tau_s(N) = \tau_v \log N$ is indeed the threshold for separation distance.
To prove the final assertion of the proposition, we first quote an easy lemma.

**Lemma (7.9).** Let $U^N$ be the uniform distributions on groups $G^N$, and let $Q^N$ be some distributions on $G^N$.

(a) If $U^N\{g: |\log(Q^N(g)/U^N(g))| \leq \epsilon\} \to 1$ as $N \to \infty$; each $\epsilon > 0$, then $||Q^N - U^N|| \to 0$.

(b) If $U^N\{g: \log(Q^N(g)/U^N(g)) \leq -K\} \to 1$ as $N \to \infty$; each $K < \infty$ then $||Q^N - U^N|| \to 1$.

Fix $c > 0$ and let $n(N) \sim c\log N$. Let $(\xi_i)$ be i.i.d. uniform on $G$ and define

$$Z_i^N = \log(|G|P(X_{n(N)} = \xi_i|\xi_i)), \quad 1 \leq i \leq N,$$

$$S_N = \sum_{i=1}^{N} Z_i^N.$$ 

Then the distribution of $S_N$ equals the distribution of $\log(|G|^NP(X_{n(N)} = g))$ where $g = (g_1, \ldots, g_N)$ is uniformly distributed on $G^N$. So by Lemma 7.9 it suffices to prove

$$S_N \to 0 \quad \text{as } N \to \infty \quad (c > \frac{1}{2})$$

$$\to -\infty \quad \text{as } N \to \infty \quad (c < \frac{1}{2}). \quad (7.10)$$

Let $Y_i^N = |G|P(X_{n(N)} = \xi_i|\xi_i) - 1$. Then $EY_i^N = 0$ and, using (7.6),

$$Y_i^N \sim |G|N^{-c}a_{\xi_i} \quad \text{as } N \to \infty$$

$$\text{Var}(Y_i^N) \sim |G|^2N^{-2c}\sigma^2, \quad \text{where } \sigma^2 = \sum_{g} a_g^2.$$ 

Since the $Y_i^N$ take at most $|G|$ values, it is easy to justify the central limit theorem and weak law of large numbers:

$$|G|^{-1}N^{c-1/2} \sum_{i=1}^{N} Y_i^N \overset{D}{\to} \text{Normal } (0, \sigma^2), \quad (7.11a)$$

$$|G|^{-2}N^{2c-1} \sum_{i=1}^{N} (Y_i^N)^2 \overset{D}{\to} \sigma^2. \quad (7.11b)$$

For $|x|$ sufficiently small we have bounds

(i) $\log(1 + x) \leq x - \frac{1}{4}x^2$.

(ii) $|\log(1 + x) - x| \leq x^2$. 
Now \( Z_i^N = \log(1 + Y_i^N) \), and \( \|Y_i^N\|_\infty \rightarrow 0 \), so for \( N \) sufficiently large we may use the bounds above. For \( c > \frac{1}{2} \),

\[
\left| S_N - \sum_{i=1}^{N} Y_i^N \right| \leq \sum_{i=1}^{N} (Y_i^N)^2 \quad \text{by (ii)},
\]

and then (7.11) implies that each sum \( \rightarrow_p 0 \), hence \( S_N \rightarrow_p 0 \). For \( c < \frac{1}{2} \), (i) shows

\[
S_N \leq \sum_{i=1}^{N} Y_i^N - \frac{1}{4} \sum_{i=1}^{N} (Y_i^N)^2.
\]

Take \( \frac{1}{2} - c < b < 1 - 2c \). Then

\[
N^{-b} S_N \leq N^{-b} \sum_{i=1}^{N} Y_i^N - \frac{1}{4} N^{-b} \sum_{i=1}^{N} (Y_i^N)^2.
\]

By (7.11) the first sum \( \rightarrow_p 0 \) and the second sum \( \rightarrow_p \infty \), so \( S_N \rightarrow_p -\infty \) as required.

Remark. We use the term “threshold” by analogy with similar effects in the study of random graphs: see Bollobas [7].

8. Examples

Here we give a complete (to the authors’ knowledge) list of examples of random walks on groups which have been studied in the present context. We shall quote only “threshold” results; the references often give more precise bounds.

a. Known Thresholds

The table lists the examples of families \((G_N, Q_N)\) where the thresholds are known. \( \tau_e \) is the relaxation time (7.2), \( \tau_d \) and \( \tau_s \) the thresholds (7.3) for variation and for separation, and we give the best known upper bounds on the thresholds obtainable by the coupling and strong uniform time techniques. Here are brief descriptions of the examples. When \( G_N \) is the permutation group \( S_N \), a distribution \( Q \) can be interpreted as “one shuffle” of a deck of \( N \) cards, and the associated random walk represents the process of repeated shuffling. Example 2 is “take the top card off the deck and insert it at random.” Example 3 is “pick a card at random and switch it with the top card.” Example 4 is “pick two cards at random and switch them.” Example 5 is a model of riffle shuffling, the usual practical method of cutting a deck into two piles and interleafing the two piles. In this model, all possible such riffles are assumed equally likely. Examples 6, 7
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Coupling bound</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Random walk on $N$-cube</td>
<td>$\frac{1}{2}N$</td>
<td>$\frac{1}{4}N \log N$</td>
<td>$\frac{1}{4}N \log N$</td>
<td>$\frac{1}{2}N \log N$</td>
<td>$\frac{1}{2}N \log N$</td>
</tr>
<tr>
<td>2) Card-</td>
<td>?</td>
<td>$N \log N$</td>
<td>$N \log N$</td>
<td>$N \log N$</td>
<td>$N \log N$</td>
</tr>
<tr>
<td>3) Card-</td>
<td>transpose top/random</td>
<td>?</td>
<td>$N \log N$</td>
<td>$N \log N$</td>
<td>?</td>
</tr>
<tr>
<td>4) Shuffling</td>
<td>random transposition</td>
<td>$\frac{1}{2}N$</td>
<td>$\frac{1}{2}N \log N$</td>
<td>$O(N^2)$</td>
<td>$\frac{1}{2}N \log N$</td>
</tr>
<tr>
<td>5) riffle</td>
<td>$\log_2 e N^2 N$</td>
<td>$\log_2 N$</td>
<td>$\log_2 N$</td>
<td>$\log_2 N$</td>
<td>$\log_2 N$</td>
</tr>
<tr>
<td>6) $[N, M]$-urn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8) $2N$ Greenwich Village</td>
<td>$N$</td>
<td>$\frac{1}{2}N \log N$</td>
<td>$O(N^2)$</td>
<td>$\frac{1}{2}N \log N$</td>
<td>$O(N \log N)$</td>
</tr>
</tbody>
</table>

References: [1, 3, 11]
concerns 2 urns with \( N, M \) balls, respectively; at each step, one ball is picked from each urn and the balls switched. Example 8 concerns partitions of \( 2N \) elements into \( N \) unordered pairs; at each step, two elements are picked and change partners.

**Remarks.** (i) Proposition 7.8 showed that \( r_\tau/r_d \) was bounded between 1 and 2. Typically one extreme is observed, but Example 7 shows that all intermediate values are possible.

(ii) In all these examples \( r_d/r_\tau \) is of order log\( (N) \), consistent with the heuristic (7.5).

(iii) The techniques described are aimed at obtaining upper bounds on \( d(n) \) or \( s(n) \), and hence on \( r_d, r_\tau \). Lower bounds are usually obtained directly from the definitions, by estimating \( P(X_n \in A) \) or \( P(X_n = i) \) for some "bad" subset \( A \) or state \( i \).

(iv) Apart from 2, 3, 5, these examples are Gelfand pairs, and the Fourier techniques of Section 6 lead to sharp bounds. Other examples of Gelfand pairs (involving vector spaces and matrices over finite fields) where similar analyses can be carried out are mentioned in [15, 32].

b. Suspected Thresholds

Here are examples of families \((G_N, Q_N)\) where thresholds are conjectured to exist but only imprecise bounds are known. Say a family has threshold \( O(a_N) \) if

\[
d_N(a_N t_N) \to 0 \quad \text{all } t_N \to \infty.
\]

Say a family has threshold \( \Omega(a_N) \) if

\[
d_N(\varepsilon_N a_N) \to 1 \quad \text{all } \varepsilon_N \to 0.
\]

The next 4 examples are card-shuffling schemes.

(9) **Example.** Adjacent Transpositions [1]. Two adjacent cards in a \( N \)-card deck are picked at random, and switched. Here the threshold is \( \Omega(N^3) \) and \( O(N^3 \log N) \), by coupling.

(10) **Example.** \( K' \)th to random [26]. The \( K' \)th card (\( K \) fixed) is removed and replaced at random. The threshold is \( \Omega(N \log N) \) and \( O(N \log N \log N) \), by strong uniform times.

(11) **Example.** Overhand shuffle [28]. The deck is divided into packets with geometric (\( \theta \)) distributed lengths, and the order of the packets is reversed. For fixed \( \theta \) the threshold is \( \Omega(N^2) \) and \( O(N^2 \log N) \), by coupling.
(12) **Example.** Another riffle shuffle [38]. Divide the deck into two equal stacks. Merge by putting the j’th cards in each stack in adjacent positions in random order (for each j). An informal argument [4] suggests the threshold of $O(\log^2 N)$.

(13) **Example.** $X \rightarrow 2X + \varepsilon \mod(N)$ [3, 10]. For odd $N$ let $X_{n+1} = 2X_n + \varepsilon_n$ modulo $(N)$, where the $\varepsilon_n$ are independent and

$$P(\varepsilon_n = -1) = P(\varepsilon_n = 0) = P(\varepsilon_n = 1) = \frac{1}{3}.$$ 

For $N$ of the form $2^l - 1$, a strong uniform time argument shows the threshold is $O(\log N)$. For general odd $N$, a Fourier analysis argument shows the thresholds is $\Omega(\log N)$ and $O(\log N \log \log N)$.

**Remark.** More generally, for prime $N$ one can consider $X_{n+1} = A_nX_n + B_n$ modulo $(N)$, where $(A_n, B_n)$ are i.i.d., $X_n$ and $B_n$ are vectors, and $A_n$ is a matrix. It seems reasonable that $\log N$ steps should suffice, under fairly general conditions, for $X_n$ to approach the uniform distribution. See [11] for $O(N^2)$ bounds.

c. **Other Families**

For the natural random walks on the complete graph (or complete bipartite graph) on $N$ vertices, one can write down explicitly the $n$-step distributions, and hence $d(n)$ and $s(n)$, but there is nothing interesting to say. For aperiodic simple symmetric random walk on the $q$-dimensional integers modulo $N$, as $N \to \infty$ ($q$ fixed) the random walks can be approximated by $q$-dimensional Brownian motion processes, and the distance functions $d_{N,q}(n), S_{N,q}(n)$ rescale as

$$d_{N,q}(tN^2) \to \hat{d}_q(t), \quad s_{N,q}(tN^2) \to \hat{s}_q(t)$$

where $\hat{d}_q(t), \hat{s}_q(t)$ are the variation and separation distances associated with the limiting Brownian motion density. Here there is no threshold, since the limit functions are continuous for each $q$; but as $q \to \infty$ the functions $\hat{d}_q(t), \hat{s}_q(t)$ approach step functions. This is consistent with the heuristic explanation of the threshold phenomenon as indicating increasing degrees of symmetry.

d. **Sporadic Examples**

Random walks on the regular polytopes have been studied in detail by Takács and Letac [23–25, 33–36]. A harder problem concerns Rubic’s cube, where one picks uniformly from the 27 possible twists: presumably around 20 moves are required to approach the uniform distribution, but the authors do not know any explicit results.
e. Random Distributions

Imagine first picking a transition matrix \( P \) on \( \{1, \ldots, N\} \) at random in some way, then running the Markov chain associated with \( P \). Or picking a distribution \( Q \) on a group \( G \) at random in some way, and then running the random walk associated with \( Q \). The variation distance after \( n \) steps is now a random variable \( D(n) \), and one can ask how many steps are required until \( ED(n) \) becomes small. In the group setting, a tractable choice of distribution is to take \( Q = (Q(g): g \in G) \) to be uniform on the \( |G| \)-simplex. A simple calculation \[4\] then shows \( ED(2) \to 0 \) as \( |G| \to \infty \). In other words, \( Q \ast Q \) is close to \( U \) for “typical” distributions \( Q \) on a large group \( G \).

In the Markov chain setting, choose \( P(i, \cdot) \) uniform on the simplex for each \( i \), independently as \( i \) varies. It seems plausible that \( n = O(\log N) \), or maybe even \( n = O(1) \) again, suffices to make \( ED(n) \) small.

Similarly, one can consider random \( r \)-regular graphs on \( N \) vertices, as in \[7\]. For random walks on such graphs, it can be shown that \( n = O(\log N) \) steps suffice to make \( ED(n) \) small.

f. Infinite and Continuous Groups

For random walks on infinite discrete groups questions of convergence to the uniform distribution do not arise; random walks on such groups are studied in connection with their recurrence properties (see, e.g., Heyer \[19, 20\]). For continuous compact groups, one in general obtains weak convergence (rather than variation convergence) to the uniform (Haar) distribution. Nothing is known about quantitative bounds for the weak convergence case. Where variation convergence occurs, the techniques of this paper can be used to obtain bounds. For instance, Brownian motion on the surface of the \( N \)-sphere can be handled by coupling. More interesting questions concern various random walks on the orthogonal group of \( N \times N \) matrices; some results are given in Diaconis and Shahshahani \[14\], but many open problems remain.

REFERENCES

26 P. Matthews, Mixing rates for a random walk on the cube, technical report, University of Maryland, Baltimore, Md.