SIAM J. MATH. ANAL.
Vol. 18, No. 1, January 1987

016

TIME TO REACH STATIONARITY IN THE BERNOULLI-LAPLACE DIFFUSION MODEL*

PERSI DIACONIS† AND MEHRDAD SHAHSHAHANI‡

Abstract. Consider two urns, the left containing n red balls, the right containing n black balls. At each time a ball is chosen at random in each urn and the two balls are switched. We show it takes $\frac{1}{4}n \log n + cn$ switches to mix up the urns. The argument involves lifting the urn model to a random walk on the symmetric group and using the Fourier transform (which in turn involves the dual Hahn polynomials). The methods apply to other "nearest neighbor" walks on two-point homogeneous spaces.

Key words. Markov chains, eigenvalues, Gelfand pairs, Hahn polynomials

AMS(MOS) subject classifications. Primary 60B15; secondary 60J20

1. Introduction. Daniel Bernoulli and Laplace introduced a simple model to study diffusion. Consider two urns, the left containing n red balls, the right containing n black balls. A ball is chosen at random in each urn and the two balls are switched. It is intuitively clear that after many such switches the urns will be well mixed, about half red and half black. The process is completely determined by the number of red balls in one of the urns. The stationary distribution may be described as the law of the composition of n balls drawn without replacement from n red and n black balls

$$\pi_n(j)=\binom{n}{j}\binom{n}{n-j}\Big/\binom{2n}{n}, \qquad 0\leqq j\leqq n. \tag{1.1}$$

The main question addressed here is the rate of convergence to the stationary distribution. Let P_k be the law of the process after k steps. Distance to stationarity will be measured by variation distance

$$\|P_k-\pi_n\|=\frac{1}{2}\sum_j |P_k(j)-\pi_n(j)|. \tag{1.2}$$

THEOREM 1. *Let P_k be the probability distribution of the number of red balls in one urn of the Bernoulli-Laplace diffusion model based on n of each color.*

(1.3) *Let* $k=\frac{1}{4}n\log n+cn$ *for* $c\geqq 0$. *Then for a universal constant* a, $\|P_k-\pi_n\|\leqq ae^{-2c}$.

(1.4) *With k as above, and arbitrary negative c in* $[-\frac{1}{4}\log n, 0]$ *there is a universal positive b such that* $\|P_k-\pi_n\|\geqq 1-be^{4c}$.

Remarks. Theorem 1 gives a sharp sense in which $\frac{1}{4}n\log n$ switches are needed: for somewhat fewer switches, the variation distance is essentially at its maximum value of 1. For somewhat more switches, the distance tends to zero exponentially fast. There is a fairly sharp cut-off at $\frac{1}{4}n\log n$.

---

* Received by the editors August 26, 1985, and in revised form October 30, 1985.
† Department of Statistics, Stanford University, Stanford, California 94305.
‡ Jet Propulsion Laboratories, Pasadena, California 91109.

A somewhat stronger result, starting with r red balls and $n-r$ black balls is proved in §§3 and 4 of this paper. The argument uses Fourier analysis on the symmetric group S_n and the homogeneous space $S_n/S_r \times S_{n-r}$. This last space is a "Gelfand pair" so the Fourier analysis is essentially commutative, involving spherical functions that turn out to be the dual Hahn polynomials. Section 2 develops the needed background material. Section 5 describes how essentially the same argument applies to nearest neighbor random walks on two-point homogeneous spaces. These include the Ehrenfest's model of diffusion (random walk on the "cube" Z_2^n) and random walk on the k dimensional subspaces of a vector space over a finite field.

The Bernoulli–Laplace process is discussed by Feller (1968, p. 423) who gives historical references. See also Johnson and Kotz (1977, pp. 205–207). We conclude this section by listing several "real world" problems where the model appears.

1) *r sets of an n set.* Let X be the set of r element subsets of $\{1, 2, \cdots, n\}$ so $|X| = \binom{n}{r}$. A random walk can be constructed on X as follows. Begin at $\{1, 2, \cdots, r\}$. Each time, pick an element from the present set and an element from its complement, and switch the two elements. This is a nearest neighbor walk using the metric: $d(x, y) = r - |x \cap y|$. The stationary distribution is the uniform distribution over X. Professor Laurel Smith points out that when $n = r = 2$, this becomes nearest neighbor random walk on the vertices of an octahedron. The rate of convergence to stationarity is the same as the rate for the Bernoulli–Laplace model with r red and $n-r$ black balls as shown in Lemma 1 below.

Walks of this type are an essential ingredient of the currently fashionable approach to combinatorial optimization called simulated annealing. Given a function $f: X \to \mathbb{R}$ annealing algorithms perform a stochastic search for the minimum of f based on the walk. Kirkpatrick et al. (1983) or Aragon et al. (1984) give further details.

2) *Moran's model in mathematical genetics.* Moran (1958) introduced a simple process to model the stochastic behavior of gene frequencies in a finite population. In one version, there is a population of n individuals each of whom is either of type A_1 or A_2. At each time, an individual is chosen at random to reproduce. After reproduction, an individual is chosen at random to die. The model allows mutation of the newborn (from A_1 to A_2 at rate u, from A_2 to A_1 at rate v). If $u = v = 1$, the transition mechanism of Moran's model becomes precisely the transition mechanism of the Bernoulli–Laplace diffusion. A clear discussion of Moran's model is in Ewens (1979, §3.3).

Ewens gives numerous references to eigenvalue–eigenvector analysis of this Markov chain. We will use part of this literature as an ingredient of our analysis.

3) *Piaget's randomization board.* In investigating children's ability to comprehend randomness, Piaget and Inhelder (1975, pp. 1–25) worked extensively with the physical device shown in Fig. 1. The left side of the box contains 8 red balls, the right side contains 8 white balls. When the box is tipped about an axis through its center (like a child's see-saw) the balls roll across to the other side. Usually one or two balls "change sides"—a red moving into the blacks or vice versa.

Piaget asked children of varying ages questions such as "how long will we have to wait until the balls are mixed up?" Answer: 5–10 switches for 8 reds. He also asked "how long will we have to wait until the balls return to the way they started?" Piaget offered an answer to the second problem for 10 reds and 10 blacks: about 185,000 moves are needed! Naturally, children (and most adults) do not guess it takes such a long time.

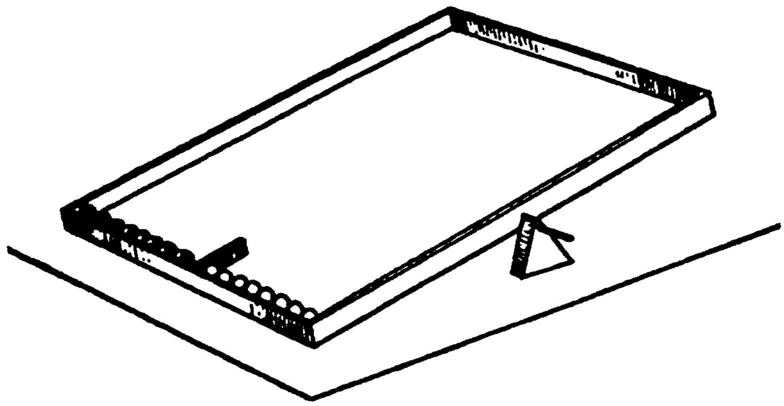

FIG. 1

One natural reaction for a mathematician is "how on earth does he know?" We began work on this paper by considering the Bernoulli–Laplace model (ignorant of its origins). Theorem 1 shows the random walk is "rapidly mixing," to use terminology of Aldous (1983). That is, the time to reach stationarity is of the order of the log of the number of states. Aldous (1983) shows that for rapidly mixing walks, the time to return to the original state has approximately the same distribution as a random walk with independent uniform steps.

If there are $|X|$ states, and W is the first time to return, then for large $|X|$,

$$P\left\{\frac{W}{|X|}>t\right\} \doteq e^{-t}.$$

For 8 red and 8 black balls, $|X| = \binom{16}{8} = 12{,}870$, so the median return time is about 9,000. For 10 red and 10 black balls $|X| = 184{,}756$; the median return time is about 128,000.

A referee points out that the associated Markov chain is doubly stochastic, irreducible, and aperiodic. Standard Markov chain theory shows that the expected number of steps to return to the starting state is $|X|$.

As explained in Diaconis and Shahshahani (1981), the analysis presented for this problem yields all the eigenvalues and eigenvectors of the associated Markov chain. Using these, it is straightforward to derive a closed form expression for the generating function of W as in Flatto, Odlyzko and Wales (1985). This can be used to get sharper asymptotic estimates for Piaget's problem.

2. Group theoretic preliminaries. One natural way to analyze the Bernoulli–Laplace process is by lifting to a random walk on the symmetric group. For integers r and b with $r + b = n$, let S_n be the symmetric group on n letters. Let $S_r \times S_b$ be the subgroup of permutations that permute the first r elements among themselves and the last b elements among themselves. Then $X = S_n/S_r \times S_b$ may be identified with the set of all $\binom{n}{r}$ r-element subsets. The random walk on X moves from x to y by choosing an element in the set x at random and an element of the complement of x at random and switching the two elements to form a new subset y. Choose a metric $d(x, y) = r - |x \cap y|$ on X. The walk is thus a nearest neighbor random walk on X. It is easy to see that the Bernoulli–Laplace process corresponds to the distance process of this walk on subsets.

It is useful to work in more generality: let G be a group and K a subgroup. Let $X = G/K$ be the associated space of cosets. Choose $x_0 = \text{id}, x_1, \cdots, x_m$ as fixed coset representatives, so $G = x_0K \cup x_1K \cdots \cup x_mK$. We will often identify X and $\{x_i\}$. Let Q be a K bi-invariant probability on G, so $Q(k_1gk_2) = Q(g)$ for all $k_1, k_2 \in K$, $g \in G$. The probability Q induces a random walk (more precisely a Markov chain) on X by the following recipe

$$Q(x, y) \overset{d}{=} Q(x^{-1}yK). \tag{2.1}$$

In (2.1), $Q(x, y)$ is the probability of going from x to y in one step. The definition comes from the following considerations: think of x_0 as the origin. Each time choose $g \in G$ from Q and move from x_0 to gx_0. This motion is then translated to motion around x via $y = xgx_0$. Thus, the chance of moving from x to y is $Q(x^{-1}yK)$.

Note that $Q(x, y)$ is well defined and satisfies

$$Q(x, y) = Q(gx, gy) \quad \text{for any} \quad g \in G. \tag{2.2}$$

Philippe Bougerol has pointed out a converse. If $Q(x, y)$ is a Markov chain on $X = G/K$ satisfying (2.2), then Q is induced by a bi-invariant probability defined by

$$Q(A) = Q(x_0, Ax_0) \quad \text{for} \quad A \subset G.$$

Alternatively, write a generic element of G as xk, then $Q(xk) = Q(x_0, x)/|K|$. This measure is K bi-invariant and $Q(x^{-1}yK) = |K|Q(x^{-1}y) = Q(x_0, x^{-1}y) = Q(x, y)$ as required. The following elementary lemma gives further connections between the random walk and Markov chain.

LEMMA 1. *Let $Q(x, y) = Q(gx, gy)$. For any $k \geqq 1$, the k step transition matrix of the Markov chain $Q(x, y)$ is induced by the kth convolution of the associated bi-invariant probability Q. The variation distance to the stationary distribution equals the variation distance to the uniform distribution.*

Because of Lemma 1, Fourier analysis on G can be used to approximate the convolution powers. We briefly review what we need from representation theory. Serre (1977) or Diaconis (1982) contain the details. Recall that a representation of G is a homomorphism $\rho: G \to GL(V)$ from G into invertible matrices on a vector space V. The dimension d_ρ of ρ is defined as the dimension of V. A representation ρ is irreducible if there are no nontrivial invariant subspaces of V. For Q a probability and ρ a representation, the Fourier transform of Q at ρ is defined by

$$\hat{Q}(\rho) = \Sigma \rho(g)Q(g).$$

The Fourier transform takes convolution into products through $P * Q(\rho) = \hat{Q}(\rho)\hat{P}(\rho)$. The uniform distribution of G: $U(g) = 1/|G|$, has $\hat{U}(\rho) = 0$ for every nontrivial irreducible representation ρ. For $X = G/K$, the set of all complex functions on X is denoted $L(X)$. The group acts on X and so $L(X)$ can be thought of as a representation as well.

The variation distance can be approximated by the following

LEMMA 2. *Let Q be a K bi-invariant probability on a finite group G*

$$\|Q - U\|^2 \leqq \tfrac{1}{4}\Sigma^*_\rho d_\rho \operatorname{Tr}(\hat{Q}(\rho)\hat{Q}(\rho)^*)$$

where the sum is over all nontrivial representations that occur in the decomposition of $L(X)$.

Proof.

$$\|Q-U\|^2=\tfrac{1}{4}\{\Sigma|Q(g)-U(g)|\}^2\leqq\tfrac{1}{4}|G|\,\Sigma|Q(g)-U(g)|^2$$
$$=\tfrac{1}{4}\Sigma_\rho^* d_\rho \operatorname{Tr}(\hat{Q}(\rho)\hat{Q}(\rho^*)).$$

Here, the Cauchy-Schwarz inequality was used and then the Plancherel theorem as in Serre (1977, p. 49) applied to $Q(g)-U(g)$. Terms corresponding to representations ρ that do not appear in the decomposition of $L(X)$ have zero Fourier transform because of Frobenius reciprocity (Serre (1977, p. 56)): this implies that a representation ρ occurs in $L(X)$ with multiplicity corresponding to the dimension of the space of K fixed vectors of ρ. Thus if ρ does not occur, then the restriction of ρ to K does not contain the trivial representation. Thus $\hat{Q}(\rho)=\Sigma_x Q(x)\rho(x)\Sigma_k\rho(k)$. The inner sum is zero because of the orthogonality of the matrix entries of the irreducible representations (Serre (1977, p. 14)). □

The Fourier transform can simplify a great deal further. Indeed, for the cases treated here the matrix $\hat{Q}(\rho)$ has only one nonzero entry in a suitable basis. The simplification in general is discussed in Volume 6 of Dieudonné (1978).

DEFINITION. The pair (G, K) is called a *Gelfand pair* if each irreducible representation of G appears in $L(X)$ with multiplicity at most 1.

Remarks. Probability theory for bi-invariant probabilities on a Gelfand pair has an extensive literature. Readable overviews appear in Letac (1981), Bougerol (1983), or Dieudonné (1978). The Bernoulli-Laplace model can be treated directly in this framework. However, the more general framework developed here is needed if one is to attack more general problems such as the natural extension to three urns where $L(X)$ has multiplicity.

For a Gelfand pair, let

$$L(X)=V_0\oplus V_1\oplus\cdots\oplus V_\lambda \tag{2.3}$$

be the decomposition into distinct irreducibles. Frobenius reciprocity implies that each V_j has a one-dimensional subspace of K invariant functions. Let $s_j(x)$ be a K invariant function in V_j normed so that $s_j(x_0)=1$. This is called the jth spherical function. The spherical functions have been explicitly computed for many Gelfand pairs.

LEMMA 3. *If (G, K) is a Gelfand pair and Q is a K bi-invariant probability, then*

$$\|Q-U\|^2\leqq\frac{1}{4}\sum_{j=0}^{\lambda} d_j|\hat{Q}(j)|^2,$$

where the sum is over the nontrivial irreducible representations occurring in (2.3) *and*

$$\hat{Q}(j)=\Sigma_g Q(g)s_j(g).$$

Proof. Fix i, and consider the vector space V_i of (2.3) as a representation ρ_i of G. Complete s_i to a basis for V_i, taking s_i as the first basis vector. With respect to this basis, the Fourier transform of any K bi-invariant function f on G becomes

$$\hat{f}(\rho)=\Sigma_g f(g)\rho_i(g)=\Sigma_x f(x)\rho_i(x)\Sigma_k\rho_i(k).$$

But the Schur orthogonality relations imply

$$\Sigma_k\rho_{ij}(k)=\begin{cases}|K| & \text{if } i=j=1,\\ 0 & \text{otherwise.}\end{cases}$$

Thus

$$\hat{f}(\rho)=\begin{pmatrix} a & & 0 \\ 0 & 0 & \\ 0 & & 0 \end{pmatrix} \quad \text{with } a=|K|\Sigma f(x)s_\rho(x).$$

Since the trace norm is invariant under unitary changes of basis, the result follows from Lemma 2. □

Remark. In the description above, the random walk associated to a bi-invariant probability Q can be represented somewhat curiously as a *right* action of G on X. Thus if the basic random elements chosen from Q on G are $g_1, g_2, g_3, \cdots$, and the walk starts at x, the successive steps have the representation $x, xg_1, xg_1g_2, xg_1g_2g_3, \cdots$, where xg denotes the coset containing xg. This is well defined for Q bi-invariant.

There is another natural way to associate a Markov chain to a probability Q, using the natural left action: $x, g_1x, g_2g_1x, g_3g_2g_1x, \cdots$. This process does not correspond to a nearest neighbor walk. It yields a Markov chain with transition matrix $\tilde{Q}(x, y) = Q(yKx^{-1})$. Chains defined in this way satisfy $\tilde{Q}(k_1x, k_2y) = \tilde{Q}(x, y)$ but we do not know a necessary and sufficient condition for a chain to lift to the left action of a bi-invariant Q. Of course, if a lifting can be found, the Fourier analysis is precisely as above, all the bounds and lemmas holding without essential change.

3. The upper bound. Let r and $n - r$ be positive integers. The stationary distribution for the Bernoulli-Laplace model based on two urns, one containing r red balls initially, the second containing $n - r$ black balls may be described as the distribution of the number of red balls in a random sample of size r from the total population of n

$$\pi_r^n(j)=\binom{r}{j}\binom{n-r}{r-j}\bigg/\binom{n}{r}, \qquad 0 \leqq j \leqq r.$$

Let P_k be the probability distribution of the number of red balls in the urn containing r balls after k switches have been made.

THEOREM 2. *If*

$$k=\frac{r}{2}\left(1-\frac{r}{n}\right)[\log n+c] \quad \textit{for } c \geqq 0,$$

then, for a universal constant a,

$$\|P_k-\pi_r^n\| \leqq ae^{-c/2}.$$

Proof. Without loss, take $r \leqq n/2$. The decomposition of the space $L(X)$ is a standard result in the representation theory of the symmetric group. James (1978, p. 52) proves that $L(X) = V_0 \oplus V_1 \oplus \cdots \oplus V_r$ where V_i are distinct irreducible representations of the symmetric group corresponding to the partition $(n - i, i)$. In particular, the pair $S_n, S_r \times S_{n-r}$ is a Gelfand pair. Since this result holds for all $r \leqq n/2$, induction gives $\dim (V_i) = \binom{n}{i} - \binom{n}{i-1}$.

The spherical functions have essentially been determined by Karlin and McGregor (1961). Stanton (1984) contains this result in modern language. The spherical functions turn out to be classically studied orthogonal functions called the dual Hahn polynomials. As a function on X, the function $s_i(x)$ only depends on the distance $d(x, x_0)$ and is a polynomial in d given by

$$s_i(d) = \sum_{m=0}^{i} \frac{(-i)_m(i-n-1)_m(-d)_m}{(r-n)_m(-r)_m m!}, \qquad 0 \leqq i \leqq r, \tag{3.1}$$

where $(j)_m = j(j+1)\cdots(j+m-1)$. Thus

$$\begin{gathered} s_0(d) = 1, \qquad s_1(d) = 1 - \frac{dn}{r(n-r)}, \\ s_2(d) = 1 - \frac{2d(n-1)}{r(n-r)} + \frac{(n-1)(n-2)d(d-1)}{(n-r)(n-r-1)r(r-1)}. \end{gathered} \tag{3.2}$$

The basic probability Q for this problem may be regarded as the uniform distribution on the $r(n-r)$ sets of distance one from the set $\{1, \cdots, r\}$. Thus, the Fourier transform of Q at the ith spherical function is

$$\hat{Q}(i) = s_i(1) = 1 - \frac{i(n-i+1)}{r(n-r)}, \qquad 0 \leqq i \leqq r.$$

Using this information in the upper bound lemma (Lemma 3) yields

$$\|P_k - \pi_r^n\|^2 \leqq \frac{1}{4} \sum_{i=1}^{r} \left\{ \binom{n}{i} - \binom{n}{i-1} \right\} \left(1 - \frac{i(n-i+1)}{r(n-r)}\right)^{2k}. \tag{3.3}$$

To bound the sum, consider first the term for $i = 1$

$$(n-1)\left(1 - \frac{n}{r(n-r)}\right)^{2k}.$$

This is smaller than

$$\exp\left(-\frac{2kn}{r(n-r)} + \log n\right).$$

Thus k must be at least

$$\frac{r}{2}\left(1 - \frac{r}{n}\right)[\log n + c]$$

to drive this term to zero. With k of this form, the problem is reduced to bounding

$$\sum_{i=1}^{r} e^{a(i)+b(i)}$$

where

$$a(i) = ci\left(\frac{(i-1)}{n} - 1\right), \qquad b(i) = \frac{i(i-1)\log n}{n} - \log(i!).$$

Calculus shows that $a(i) \leqq a(1) = -c$ for all $i = [2, n/2]$. Thus, to prove Theorem 1 it suffices to prove

$$\sum_{i=1}^{n/2} e^{b(i)} \leqq B \quad \text{independently of } n.$$

Clearly, the sum of $e^{b(i)}$ over $1 \leqq i \leqq 21$ is uniformly bounded. For the remaining range, upper bound $b(i)$ by $i^2(\log n/n) - i \log i + i$. It will be argued that

$i^2(\log n/n) - i\log i + i < -i$ for $21 \leqq i \leqq n/2$, equivalently, $\log n/n < (\log i - 2)/i$. Now if $f(x) = (\log x - 2)/x$, $f'(x) = (3 - \log x)/x^2$. This is negative for $x > e^3$, so for $i > 21 > e^3$,

$$\frac{\log i - 2}{i} > \frac{\log (n/2) - 2}{n/2}.$$

This last is greater than $\log n/n$ for $n \geqq e^{4+2\log 2}$. Thus,

$$\sum_{i=22}^{n/2} e^{b(i)} \leqq \sum_{i=22}^{\infty} e^{-i} < B \quad \text{uniformly in } n. \qquad \square$$

Remarks. Change n to $2n$ and take $r = n$. The result becomes $(n/4)\log n + (c/2)n$ which gives (1.3) of Theorem 1. If $r = o(n)$, the result becomes $(r/2)\log n + (c/2)r$. As usual with approximations, some precision has been lost to get a clean statement. When $r = 1$, for example, there is only one term: $(n-1)(1/(n-1))^{2k}$. For $k = 1$ this gives $\frac{1}{2}(1/\sqrt{n-1})$ as an upper bound for the variation distance. Elementary considerations show that for this case the correct answer is $1/n$. Thus, the upper bound lemma gives the right answer for the number of steps required (namely, 1) but an over estimate for the distance.

4. The lower bound. A lower bound for the variation distance will be found by using the easily derived relation

$$\|P - Q\| = \sup_A |P(A) - Q(A)|. \tag{4.1}$$

Any specific set A thus provides a lower bound. Intuitively, if the number of steps k is too small, there will tend to be too many of the original color in the urn. The argument below gives a sharp form of this. For ease of exposition, we only prove the result for $r = n - r$ (for example (1.4)) but the proof works in the general case.

The idea of the proof is to again use spherical functions, but this time as random variables, not transforms. Thus for any Gelfand pair (G, K) with $X = G/K$, consider $s_j\colon X \to \mathbb{R}$ as a random variable. If Z is a point chosen uniformly in X, the orthogonality relations (as in Stanton (1984, eq. (2.9))) give

$$E\{s_j(Z)\} = \delta_{0j}, \qquad \operatorname{Var}(s_j(Z)) = \frac{1}{\dim (V_j)}. \tag{4.2}$$

If Z_k denotes an X valued random variable with distribution P^{*k} for P a bi-invariant probability on X the basic convolution property of spherical functions becomes

$$E\{s_j(Z_k)\} = E\{s_j(Z_1)\}^k. \tag{4.3}$$

On $S_{2n}/S_n \times S_n$ the first three spherical functions, as functions of the distance d are given (from (3.1)) as

$$s_0(d) = 1, \qquad s_1(d) = 1 - \frac{2d}{n},$$

$$s_2(d) = 1 - \frac{2(2n-1)d}{n^2} + \frac{(2n-1)(2n-2)d(d-1)}{[n(n-1)]^2}.$$

Since these are polynomials in d, it follows that for some a, b, c, $s_1^2 = a + bs_1 + cs_2$. After a computation

$$s_1^2 = \frac{1}{2n-1} + \frac{2n-2}{2n-1} s_2. \tag{4.4}$$

Remark. When working with general r, $n-r$ values the term s_1 appears in the expression for s_1^2.

To lower bound the variation distance, consider the normalized spherical function $f(x) \stackrel{d}{=} \sqrt{n-1}\, s_1(x)$. Now (4.2) implies for Z uniform on x,

$$E\{f(Z)\} = 0, \qquad \mathrm{Var}\{f(Z)\} = 1.$$

Under the convolution measure

$$E\{f(Z_k)\} = \sqrt{n-1}\left(1-\frac{2}{n}\right)^k, \tag{4.5}$$

$$\mathrm{Var}\{f(Z_k)\} = \frac{n-1}{2n-1} + \frac{(n-1)(2n-2)}{(2n-1)}\left(1-\frac{2(2n-1)}{n^2}\right)^k - (n-1)\left(1-\frac{2}{n}\right)^{2k}. \tag{4.6}$$

For k of the form $\frac{1}{4}n\log n - cn$, the mean becomes

$$E\{f(Z_k)\} = \exp\left(2c + O\left(\frac{\log n}{n}\right) + O\left(\frac{c}{n}\right)\right), \tag{4.7}$$

where $c > 0$, and all error terms are uniform in both n and c. Thus, for c large, the mean is large. Similarly

$$\begin{aligned}\mathrm{Var}\{f(Z_k)\} &= \frac{1}{2} + O\left(\frac{1}{n}\right) + \exp\left(4c + O\left(\frac{\log n}{n}\right) + O\left(\frac{c}{n}\right)\right)\\ &\quad - \exp\left(4c + O\left(\frac{\log n}{n}\right) + O\left(\frac{c}{n}\right)\right)\\ &= \frac{1}{2} + e^{4c}\left\{O\left(\frac{\log n}{n}\right) + O\left(\frac{c}{n}\right)\right\}.\end{aligned}$$

Thus, the variance is uniformly bounded for $O \leqq c \leqq \frac{1}{4}\log n$. Now use Chebyshev's inequality: if $A_\alpha = \{x: |f(x)| \leqq \alpha\}$, $\pi_n(A_\alpha) \geqq 1 - 1/\alpha^2$ while $P_k(A_\alpha) < B/(e^{2c}-\alpha)^2$ where B is uniformly bounded for $c \leqq \frac{1}{4}\log n$. Thus, for any fixed α and c, for all sufficiently large n,

$$\|P_k - \pi_n\| \geqq 1 - \frac{1}{\alpha^2} - \frac{B}{(e^{2c}-\alpha)^2}.$$

This completes the proof of (1.4), choosing $\alpha = e^{2c}/2$, for example. □

Remark. From the definition of s_1, the set A_α can be interpreted as the event

$$\left|\#\text{reds} - \frac{n}{2}\right| \Big/ \sqrt{n} \geqq \alpha.$$

5. Other nearest neighbor walks. A class of problems that can be treated by following the steps above involves a connected graph with vertex set X and an edge set E. Define a metric on X as

$$d(x, y) = \text{length of shortest path from } x \text{ to } y.$$

We want to analyze nearest neighbor walks on this graph. An automorphism of the graph is a 1–1 mapping from X to X which preserves the edge set. Let G be the group of automorphisms of X. Call the graph *2-point homogeneous* if $d(x, y) = d(x', y')$ implies there is a $g \in G$ such that $gx = x'$, $gy = y'$. Taking $x = y$, $x' = y'$ shows G operates transitively on X, so $X \cong G/K$ where $K = \{g \in G: gx_0 = x_0\}$ for some fixed point x_0. Stanton (1984) shows

THEOREM. *For a 2-point homogeneous graph, (G, K) form a Gelfand pair and the spherical functions are orthogonal polynomials.*

This means that in principle the analysis above can be carried out for such examples. Here is a list of some of the examples in Stanton:

Example 1. $X = Z_2^n$, $d(x, y) =$ number of coordinates where x and y differ. Here the random walk becomes nearest neighbor walk on the n cube. This is a well studied problem equivalent to the well-known model of diffusion known as the Ehrenfest urn model. A wonderful discussion of this model is in Kac (1945). Further references are in Letac and Takacs (1979). The straightforward random walk never converges because of parity—after an even number of steps the walk is at a point at an even distance from 0. One simple way to get convergence is to stay fixed with probability $1/(n + 1)$ and move to a vertex 1 away with probability $1/(n + 1)$. For this process, the analysis can be carried out just as in §§3 and 4 to show $\frac{1}{4}n \log n + cn$ steps suffice and that this many steps are needed.

THEOREM 3. *Let $X = Z_2^n$. Let $P(00 \cdots 0) = P(10 \cdots 0) = \cdots = P(00 \cdots 1) = 1/(n + 1)$. Let U be the uniform distribution on X. Suppose $k = \frac{1}{4}(n + 1) \log n + c(n + 1)$ for $c > 0$. Then*

$$\|P^{*k} - U\|^2 \leqq \tfrac{1}{2}(e^{e^{-4c}} - 1).$$

Conversely, for $k = \frac{1}{4}(n + 1) \log n - c(n + 1)$, for $c > 0$, the variation distance does not tend to zero as n tends to infinity: $\underline{\lim}\ \|P^{*k} - U\| \geqq (1 - 8e^{-c})$.

Remark. It is curious that the critical rate is precisely the same $\frac{1}{4}n \log n$, for the cube and n sets of a $2n$ set.

Example 2. Let F_q be a finite field with q elements. Let V be a vector space of dimension n over F_q. Let X be the set of k-dimensional subspaces of V, with metric $d(x, y) = k - \dim (x \cap y)$. Here, $G = GL_n(q)$ operates transitively on X. Stanton (1984) gives all the ingredients needed to carry out the analysis.

Example 3. Let X be the set of $(n - r) \times r$ matrices over F_q with metric $d(x, y) = \operatorname{rank}(x - y)$. Here, $GL_{n-r} \times GL_r$ operates transitively on X. Again a complete analysis seems in reach using results given by Stanton.

Example 4. For q odd, let X be the set of skew-symmetric matrices over F_{q^2} with metric $\frac{1}{2} \operatorname{rank}(x - y)$. Here $G = GL_n$ acts on X by $x \to A^T x A$. Again Stanton gives enough information about spherical functions and dimensions to allow a complete analysis.

Stanton also gives results for orthogonal, hermitian, and symplectic matrices over finite fields. He also gives results for a variety of less familiar combinatorial objects. Combinatorialists have also studied such objects: see Biggs (1974, Chaps. 20, 21). Further surveys and examples of Gelfand pairs are given by Heyer (1983) and Sloane (1982).

Finally, there are Gelfand pairs that do not arise from two point homogeneous graphs. An example is the set X of all partitions of $\{1, 2, \cdots, 2n\}$ into n two-element subsets. In graph theoretic language this is the set of all "matchings" of a $2n$ set. The symmetric group S_{2n} acts transitively on X and yields a Gelfand pair. A natural random walk involves picking two elements at random and switching them to form a new partition. This gives an algorithm that converges to a random matching. The spherical functions are "zonal polynomials." It can be shown that $\frac{1}{2}n \log n$ switches suffice.

REFERENCES

D. ALDOUS (1983), *On the time taken by random walks on finite groups to visit every state*, Z. Wahrsch. Verw. Geb., 62, pp. 361–374.

C. ARAGON, D. JOHNSON, L. MEGEOCH AND C. SCHEVON (1984), *Optimization by simulated annealing: An experimental evaluation*, Technical memorandum, Bell Laboratories, Murray Hill, NJ.

N. BIGGS (1974), *Algebraic Graph Theory*, Cambridge Univ. Press, Cambridge.

P. BOUGEROL (1983), *Un mini-cours sur les couples de Gelfand*, Publication Laboratoire de Statistique et Probabilities, Université Paul Sabatier, Toulouse, France, No. 1-83.

P. DIACONIS (1982), *Lecture notes on the use of group representations in probability and statistics*, IMS Lecture Note Series, to appear.

P. DIACONIS AND M. SHAHSHAHANI (1981), *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Geb., 57, pp. 159–179.

J. DIEUDONNÉ (1978), *Treatise on Analysis* VI, Academic Press, New York.

W.J. EWENS (1979), *Mathematical Population Genetics*, Springer-Verlag, Berlin.

W. FELLER, (1968), *An Introduction to Probability and Its Applications*, Vol. 1, 3rd ed., John Wiley, New York.

L. FLATTO, A. ODLYZKO AND D. WALES (1985), *Random shuffles and group representations*, Ann. Prob., 13, pp. 154–178.

H. HEYER (1983), *Convolution semi-groups of probability measures on Gelfand pairs.*, Expos. Math., 1, pp. 3–45.

G.D. JAMES (1978), *Representation of the Symmetric Groups*, Springer Lecture Notes in Mathematics 682, Springer-Verlag, Berlin.

N.L. JOHNSON AND S. KOTZ (1977), *Urn Models and Their Application*, John Wiley, New York.

M. KAC (1945), *Random walk and the theory of Brownian motion*, Amer. Math. Monthly, 54, pp. 369–391.

S. KARLIN AND J. MCGREGOR (1961), *The Hahn polynomials, formulas and an application*, Scripta Math., 26, pp. 33–46.

——— (1975). *Linear growth models with many types and multidimensional Hahn polynomials*, in Theory and Applications of Special Functions, R. Askey, ed., Academic Press, New York, pp. 261–288.

S. KIRKPATRICK, C. GELATT AND M. VECCHI (1983), *Optimization by simulated annealing*, Science, 220, pp. 671–680.

G. LETAC (1981), *Problèmes Classiques de Probabilite sur un Couple de Gelfand*, in Springer Lecture Notes in Mathematics 861, Springer-Verlag, Berlin.

G. LETAC AND L. TAKACS (1979), *Random walks on the m-dimensional cube*, J. Reine Angew. Math., 310, pp. 187–195.

P.A.P. MORAN (1958), *Random Processes in Genetics*, Proc. Cambridge Phil. Soc., 54, pp. 60–72.

J. PIAGET AND B. INHELDER (1975), *The Origin of the Idea of Chance in Children*, Norton, New York.

J.P. SERRE (1977), *Linear Representations of Finite Groups*, Springer-Verlag, New York.

N.J.A. SLOANE (1982), *Recent bounds for codes, sphere packings, and related problems obtained by linear programming and other methods*, Contemp. Math, 9, pp. 153–185.

D. STANTON (1984), *Orthogonal polynomials and Chevalley groups*, in Special Functions: Group Theoretical Aspects and Applications, R. Askey, T.H. Koornwinder and W. Schempp, eds., Reidel, Dordrecht, pp. 87–128.