

STRONG STATIONARY TIMES VIA A NEW FORM OF DUALITY

BY PERSI DIACONIS AND JAMES ALLEN FILL

Harvard University and The Johns Hopkins University

A strong stationary time for a Markov chain (X_n) is a stopping time T for which X_T is stationary and independent of T . Such times yield sharp bounds on certain measures of nonstationarity for X at fixed finite times n . We construct an absorbing dual Markov chain with absorption time a strong stationary time for X . We relate our dual to a notion of duality used in the study of interacting particle systems. For birth and death chains, our dual is again birth and death and permits a stochastic interpretation of the eigenvalues of the transition matrix for X . The duality approach unifies and extends the analysis of previous constructions and provides several new examples.

1. Overview.

1.1. *Introduction.* Strong stationary times give a probabilistic approach to bounding speed of convergence to stationarity for Markov chains. They were introduced (under the name *strong uniform times*) by Aldous and Diaconis (1986), who give a number of examples demonstrating both sharp bounds and successful analysis of problems not amenable to other techniques such as eigenvalues and coupling. Diaconis [(1988), Chapter 4] and Matthews (1987, 1988) construct strong stationary times for various random walks on groups. Aldous and Diaconis (1987) develop some basic theory, showing that an optimal strong stationary time exists for any ergodic chain, i.e., one that is irreducible, positive recurrent and aperiodic. Closely related constructions appear in Brown (1975), Athreya and Ney (1978) and Nummelin (1986). Thorisson (1988) discusses connections with coupling.

In this paper we extend the notion of strong stationary time to that of *strong stationary duality* for discrete time, finite state Markov chains. Diaconis and Fill (1990) treat problems with countably infinite state space. Fill (1990a, b) treats continuous time chains and Fill (1990c) applies strong stationary duality to diffusions. We begin with a simple example.

EXAMPLE 1.1. *Simple symmetric random walk on a d -point circle.* Let \mathbb{Z}_d be the integers modulo d , regarded as d labelled points arranged about a circle. A random walk starts at 0 and with probability $1/3$ each moves one step in either direction along the circle or remains fixed. The stationary distribution

Received November 1988; revised September 1989.

AMS 1980 subject classifications. Primary 60J10; secondary 60G40.

Key words and phrases. Markov chains, rates of convergence, stochastic monotonicity, monotone likelihood ratio, birth and death chains, eigenvalues, random walk, Ehrenfest chain, strong stationary duality, dual processes, Siegmund duality, time reversal, Doob h transform, total variation.

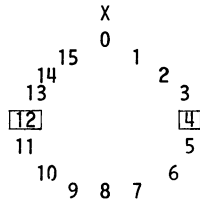


FIG. 1.

for this Markov chain is the uniform distribution U on \mathbb{Z}_d . Let π_n denote the law of the process after n steps. We measure the nonstationarity of π_n by the total variation distance

$$(1.2) \quad \|\pi_n - U\| = \max_{A \subset \mathbb{Z}_d} |\pi_n(A) - U(A)|.$$

The following construction yields a stopping time T such that the state in which the walk is stopped is uniformly distributed and is independent of T . Such times are called strong stationary times. They offer an upper bound to variation distance through the inequality

$$(1.3) \quad \|\pi_n - U\| \leq P\{T > n\}, \quad n \geq 0;$$

cf. (1.9) and (1.11).

For the construction, suppose first that $d = 2^a$, e.g., $d = 16$ (see Figure 1). The walk starts at 0. Let T_1 be the first time the walk hits 4 or 12. At time T_1 the state of the walk is equally likely to be 4 or 12 and is independent of T_1 . Let T_2 be the first time the walk hits a point at distance 2 from its position at time T_1 . At time T_2 the state of the walk is uniformly distributed in $\{2, 6, 10, 14\}$ and is independent of (T_1, T_2) . Let T_3 be the first time the walk hits a point at distance 1 from its position at time T_2 . At time T_3 the state of the walk is uniform over the odd positions and is independent of (T_1, T_2, T_3) .

To finish, let T_4 be the first time following T_3 that the process remains fixed or moves counterclockwise. At time T_4 the state of the walk is uniform over \mathbb{Z}_{16} and is independent of (T_1, T_2, T_3, T_4) and, in particular, of T_4 . Thus $T = T_4$ is a strong stationary time.

To bound the tails of T , recall (e.g., Feller (1968), Section 14.3) that a random walk on \mathbb{Z} starting at 0, that at each step changes by an amount $-1, 0$ or $+1$ with respective probabilities $\theta/2, 1 - \theta$ and $\theta/2$, has mean time to hit $\pm b$ equal to b^2/θ .

For $d = 2^a$ the random walk must successively travel amounts $\pm 2^{a-2}, \pm 2^{a-3}, \dots, \pm 1$. The final step (stay fixed or move counterclockwise) has mean $3/2$. Thus the final stopping time $T = T_a$ has (for $a \geq 2$) mean

$$\frac{3}{2}2^{2a}(2^{-4} + 2^{-6} + \dots + 2^{-2(a-1)} + 2 \times 2^{-2a}) \leq \frac{3}{16}2^{2a}.$$

Now Markov's inequality and (1.3) yield the following result.

PROPOSITION 1.4. *Let $d = 2^a$. For the simple random walk on \mathbb{Z}_d described above,*

$$\|\pi_n - U\| \leq \frac{3}{16}d^2/n.$$

REMARKS. (a) When $n = cd^2$, the variation distance is smaller than $3/(16c)$. An argument using the central limit theorem shows, conversely, that for d large it takes *at least* $n = cd^2$ steps with c large to make $\|\pi_n - U\|$ small. Indeed, it takes on the order of d^2 steps for the walk simply to have a reasonable chance of reaching positions near the antipode $d/2$ of 0.

(b) It is not hard to extend our construction to general values of d and to show again that cd^2 steps with c large (and no fewer) make the variation distance small. We sketch the extension for even d satisfying $2^{a-1} < d \leq 2^a$. As for $d = 2^a$, start the walk at a given state and wait for it to move successively $\pm 2^{a-2}, \pm 2^{a-3}, \dots, \pm 2, \pm 1$. If the total (signed) distance traveled is negative and greater than or equal to $d - 2^{a-1}$ in magnitude, repeat the entire procedure from the terminal state; otherwise, stop. Then the stopping time T' is independent of $X_{T'}$, which is uniformly distributed in $\{1, 3, 5, \dots, d - 1\}$. Now finish building a strong stationary time T as for $d = 2^a$.

(c) For this process, Fourier analysis can be used to bound the variation distance. See, e.g., Diaconis [(1988), Section 3-C].

(d) One can treat Brownian motion on the circle (or reflecting Brownian motion on an interval) in a similar fashion. The resulting stopping time is quite close to one developed by Dubins (1968) to study Skorohod embedding.

For present purposes, the point of the example is this: A strong stationary time was found by identifying sets of states increasing in size and times at which the process is uniform on each set. This transforms the original problem of analyzing convergence to stationarity into a quite different probability problem, viz., a study of first passage times. Strong stationary times serve in general to transform problems. In applications, the transformed problems can often be solved using classical techniques, such as results from the birthday problem or coupon collector's problem; see Diaconis [(1988), Chapter 4].

In this paper we give a general and often practical method for constructing strong stationary times for ergodic Markov chains. Later (in Example 3.2) we shall see how the general construction reduces to the above bisection method in the case of Example 1.1.

1.2. *The basic setup.* Let P be an irreducible aperiodic transition matrix on a discrete (finite or countably infinite) state space \mathcal{S} . When we need to restrict attention to finite state spaces we shall do so explicitly. Let π_0 (regarded as a row vector) be a distribution on \mathcal{S} . Let $X = (X_n)_{n=0,1,\dots}$ be a Markov chain on some probability space (Ω, \mathcal{F}, P) with initial distribution π_0 and one-step transition function P . We write P^n for the n -step transition function (the n th power of the matrix P) and $\pi_n = \pi_0 P^n$ for the distribution $\mathcal{L}(X_n)$ of X_n . If P is positive recurrent (in particular, if \mathcal{S} is finite), classical

theory gives a unique stationary distribution π for P , with $\pi(y) > 0$ for all y and

$$\pi_n(y) \rightarrow \pi(y) \quad \text{as } n \rightarrow \infty \text{ for all } y \in \mathcal{S},$$

regardless of the choice of π_0 .

One measure of discrepancy between π_n and π is the *total variation distance*

$$(1.5) \quad \|\pi_n - \pi\| = \max_{A \subset \mathcal{S}} |\pi_n(A) - \pi(A)|.$$

An alternative measure, exploited by Aldous and Diaconis (1986, 1987), is the *separation*

$$(1.6) \quad s(n) := \text{sep}(\pi_n, \pi) = \sup_y s(n, y),$$

where

$$(1.7) \quad s(n, y) := 1 - \pi_n(y)/\pi(y), \quad y \in \mathcal{S}.$$

Because

$$(1.8) \quad \|\pi_n - \pi\| = \sum_{y: \pi(y) \geq \pi_n(y)} [\pi(y) - \pi_n(y)],$$

it follows that

$$(1.9) \quad \|\pi_n - \pi\| \leq s(n).$$

Separation is closely allied with the stopping times of interest. A *strong stationary time* is a randomized stopping time T for X such that X_T has the stationary distribution π and is independent of T . More precisely, T takes values in $\{0, 1, 2, \dots, \infty\}$ and the law of X_T given $T = k$ is π , for every finite k . The following proposition was established by Aldous and Diaconis [(1987), Proposition 3.2].

PROPOSITION 1.10. (a) *If T is a strong stationary time, then*

$$(1.11) \quad s(n) \leq P\{T > n\}, \quad n \geq 0.$$

(b) *Conversely, there exists a strong stationary time T such that equality holds in (1.11) for each $n \geq 0$.*

We call a minimal (stochastically fastest) strong stationary time as in (b) a *time to stationarity*. The Aldous–Diaconis construction of a time to stationarity is a theoretical result requiring complete knowledge of each π_n .

On the other hand, strong stationary times leading to quite useful bounds on distance from stationarity have been constructed in an ad hoc manner for a variety of interesting examples. A chief objective of this paper is to unify these constructions by showing that each results from the building of what we shall call a dual process. This dual is absorbing and has the property that the waiting time to absorption is a strong stationary time for the given chain.

Thus a problem about time to stationarity is transformed in each case to a problem about time to absorption, and often the dual problem is tractable.

In many cases, such as Example 1.1 and the “top in at random” card shuffle described in Aldous and Diaconis (1986), the state space for the dual chain consists of subsets of the original state space. These two examples are considered further in Examples 3.1 and 3.7. In coordinate checking examples described in Diaconis [(1988), e.g., Example 2 of Chapter 4] or Matthews (1987), the dual involves, as described in Example 3.2, the positions of checked coordinates. In Section 4, birth and death chains are shown to have duals that are birth and death chains.

Our main goal is to formalize and abstract the notion of duality. Section 2 does this and shows that *any* ergodic chain has a sharp dual, i.e., one that yields a time to stationarity. Our development involves a careful study of an intertwining $\Lambda P^n = P^{*n} \Lambda$. This may be of independent interest in connection with recent work of Yor (1988). Section 2 shows how useful information not available from a strong stationary time alone can be extracted from the behavior of a dual chain *prior* to absorption. In Section 3 we investigate a special class of duals taking values in a class of subsets of the original state space and we treat a number of examples.

There is a particularly simple construction of a dual chain for a wide class of Markov chains having a certain monotonicity property described in Section 4. The dual of a birth and death chain from this class is an absorbing birth and death chain. The distribution of time to absorption can be expressed in terms of the eigenvalues of the dual chain. As it turns out, these are precisely the eigenvalues of the original chain; these eigenvalues are thereby given a stochastic interpretation. In Section 4.3 we show that if, for example, the eigenvalues are all nonnegative, then the time to stationarity can be represented as a sum of independent geometric random variables with the eigenvalues as parameters. Here eigenvalues close to 1 yield slow convergence to stationarity.

The dual chain construction of Section 4 is closely related to a notion of duality developed by Siegmund (1976) and used in the study of particle systems [e.g., Liggett (1985), Chapter 2]. In Section 5 we show that our dual is a Doob h transform of the Siegmund dual of the time-reversed chain.

2. Strong stationary duality.

2.1. Introduction. In Section 2 we describe how to build strong stationary dual processes for a given Markov chain X . These duals are absorbing processes whose time to absorption is a strong stationary time for X . The problem is closely related to the following exercise in Markov chains. We are given the marginal specification of two Markov chains X and X^* having respective state spaces \mathcal{S} and \mathcal{S}^* , along with a transition kernel, or *link*, $\Lambda = (\Lambda(x^*, x))$. We seek a bivariate chain (X^*, X) with the specified marginals so that the law of X_n given $X_n^* = x^*$ is $\Lambda(x^*, \cdot)$, in a sense we shall make precise.

Section 2.2 defines the appropriate notion of duality and establishes the connection with strong stationary times. As an example we treat the theoretical construction of a fastest strong stationary time due to Aldous and Diaconis (1987). Section 2.3 formulates and solves the bivariate Markov chain problem. Section 2.4 shows how to build one sample path of X^* from each sample path of X . The resulting X^* is dual to X as defined in Section 2.2. Section 2.5 shows how the dual process X^* constructed in Section 2.4 can be used to obtain bounds on total variation not available from strong stationary times alone. Section 3 will specialize the general construction of X^* to the case where \mathcal{S}^* is a collection of subsets of \mathcal{S} and $\Lambda(x^*, \cdot)$ is the stationary distribution of X truncated to x^* . It will also treat a number of specific applications.

Throughout the paper we write $X \sim (\pi_0, P)$ as shorthand for the statement that $X = (X_n)_{n \geq 0}$ is a (time homogeneous) Markov chain with initial distribution π_0 and transition function P . The value y is said to be a possible value of the random variable Y if $P\{Y = y\} > 0$.

2.2. Strong stationary times and duality. We begin forthwith with our definition of a dual process. Throughout Section 2.2 let $X \sim (\pi_0, P)$ with stationary distribution π be a given ergodic Markov chain on a given probability space (Ω, \mathcal{F}, P) . Write π_n for $\mathcal{L}(X_n) = \pi_0 P^n$.

DEFINITION 2.1. Let $X^* = (X_n^*)_{n \geq 0}$ be a stochastic process on (Ω, \mathcal{F}, P) taking values in a discrete state space \mathcal{S}^* . Suppose that for each $n \geq 0$,

$$(2.2a) \quad X_n^* \text{ and the chain } X \text{ are conditionally independent given } X_0, \dots, X_n.$$

Suppose also that there exists at least one state in \mathcal{S}^* , call it ∞ , for which

$$(2.2b) \quad \mathcal{L}(X_n | X_0^* = x_0^*, X_1^* = x_1^*, \dots, X_{n-1}^* = x_{n-1}^*, X_n^* = \infty) = \pi$$

for each $n \geq 0$ and each possible value of (X_0^*, \dots, X_n^*) of the form $(x_0^*, \dots, x_{n-1}^*, \infty)$. Finally, suppose that

$$(2.2c) \quad \infty \text{ is an absorbing state for } X^*,$$

i.e., that if $X_m^* = \infty$, then $X_n^* = \infty$ for all $n \geq m$. Then X^* is called a *strong stationary dual* for X .

REMARK 2.3. The reason for the terminology “dual” will be explained later, in Section 5.

Theorem 2.4 shows how to use a strong stationary dual to build a strong stationary time. It also shows conversely how, in principle, every strong stationary time results from such a construction.

THEOREM 2.4. (a) *Let X^* be a given strong stationary dual of X . Let $T = T_\infty^*$ be the time to absorption in ∞ for X^* . Then T is a strong stationary time for X .*

(b) Conversely, let T be a given strong stationary time for X . Let $\mathcal{S}^* = \{0, 1, \dots\} \cup \{\infty\}$ and define

$$(2.5) \quad X_n^* = n \text{ or } \infty \text{ according as } T > n \text{ or } T \leq n.$$

Then X^* is a strong stationary dual for X and $T = T_\infty^*$.

PROOF. (a) $T = T_\infty^*$ is by (2.2a) clearly a randomized stopping time for X . Moreover, for each $n \geq 0$,

$$\mathcal{L}(X_n | T = n) = \mathcal{L}(X_n | X_0^* \neq \infty, \dots, X_{n-1}^* \neq \infty, X_n^* = \infty) = \pi,$$

as follows from (2.2b).

(b) The conditional independence condition (2.2a) is a consequence of the fact that T is a strong stationary time. That ∞ is an absorbing state for X^* is a trivial consequence of (2.5). The possible values of (X_0^*, \dots, X_n^*) are all of the form $(0, \dots, k - 1, \infty, \dots, \infty)$ with $0 \leq k \leq n + 1$. If $0 \leq k \leq n$, then

$$\mathcal{L}(X_n | X_0^* = 0, \dots, X_{k-1}^* = k - 1, X_k^* = \infty, \dots, X_n^* = \infty) = \mathcal{L}(X_n | T = k) = \pi,$$

and (2.2b) is verified.

So we have shown that X^* is a strong stationary dual for X . The identity $T = T_\infty^*$ follows immediately from (2.5). \square

The proof of the converse builds a dual X^* from a time T without regard to the distributions $\mathcal{L}(X_n | X_0^* = x_0^*, X_1^* = x_1^*, \dots, X_{n-1}^* = x_{n-1}^*, X_n^* = x_n^*)$ except [as required by (2.2b)] for the case $x_n^* = \infty$. As will be made abundantly clear later, our construction is by no means unique: There are many strong stationary duals corresponding to a given strong stationary time. Indeed, the advantage in using a dual is to exploit knowledge of $\mathcal{L}(X_n | X_0^* = x_0^*, X_1^* = x_1^*, \dots, X_{n-1}^* = x_{n-1}^*, X_n^* = x_n^*)$ even when $x_n^* \neq \infty$. See Section 2.5 for illustration.

EXAMPLE 2.6. *A stochastically fastest strong stationary time via duality.* Aldous and Diaconis (1987) show that any ergodic Markov chain X has a time to stationarity, i.e., a fastest strong stationary time; this is Proposition 1.10(b). Their construction can be described in terms of a strong stationary dual. We recall the notation $s(n, y) = 1 - \pi_n(y)/\pi(y)$, $y \in \mathcal{S}$, and $s(n) = \sup_y s(n, y)$.

Having observed a (without loss of generality possible) value x_0 of X_0 , set

$$(2.7a) \quad X_0^* = \infty \text{ with probability } (1 - s(0))\pi(x_0)/\pi_0(x_0)$$

and $X_0^* = 0$ with the complementary probability (by using randomness independent of the given chain X). Proceed inductively. Suppose that $X_0 = x_0, \dots, X_{n-1} = x_{n-1}$ have been observed and that $X_0^* = x_0^*, \dots, X_{n-1}^* = x_{n-1}^*$ have been set. If $X_{n-1}^* = \infty$, set $X_n^* = \infty$. Otherwise, observe $X_n = x_n$ and set

$$(2.7b) \quad X_n^* = \infty \text{ with probability } (s(n - 1) - s(n))/(s(n - 1) - s(n, x_n))$$

and $X_n^* = n$ with the complementary probability. Then X^* is, as Aldous and Diaconis essentially show, a strong stationary dual for X . Moreover, they show

that the corresponding strong stationary time $T = T_\infty^*$ is a time to stationarity for X .

Note that the above construction requires the separation function s as input. Typically, the reason for building strong stationary times T is to bound an intractable separation function s via Proposition 1.10 ($s(n) \leq P\{T > n\}$). Thus the construction is only of theoretical interest. Section 3 gives a practical construction of a dual process that covers a large number of cases.

REMARK 2.8. The strong stationary dual in Example 2.6 possesses a noteworthy feature which serves as motivation for Section 2.3. Both the bivariate process $(X^*, X) = (X_n^*, X_n)_{n \geq 0}$ and the marginal dual process X^* are Markov chains. The initial distribution π_0^* and the transition matrix P^* for X^* are given by

$$(2.9) \quad \pi_0^*(0) = s(0) = 1 - \pi_0^*(\infty),$$

$$(2.10a) \quad P^*(n - 1, n) = s(n)/s(n - 1) = 1 - P^*(n - 1, \infty), \quad n = 1, 2, \dots,$$

$$(2.10b) \quad P^*(\infty, \infty) = 1.$$

A link between the coordinate processes X and X^* can be defined by

$$\Lambda(x^*, x) := P\{X_n = x | X_0^* = x_0^*, X_1^* = x_1^*, \dots, X_{n-1}^* = x_{n-1}^*, X_n^* = x^*\},$$

which for Example 2.6 does not depend on x_0^*, \dots, x_{n-1}^* and is given by

$$(2.11a) \quad \Lambda(n, \cdot) = [\pi_n(\cdot) - (1 - s(n))\pi(\cdot)]/s(n), \quad n = 0, 1, \dots,$$

$$(2.11b) \quad \Lambda(\infty, \cdot) = \pi.$$

In Sections 2.3–2.4 we abstract the Markovian features of Example 2.6. That is, for given $X \sim (\pi_0, P)$, we seek to build strong stationary duals X^* that are marginally Markov and make (X^*, X) bivariate Markov. For convenience we first discuss the simultaneous construction of X and X^* and later show how to build a realization of X^* from a corresponding realization of X . In our discussion, the link between the coordinate processes will play a central role.

2.3. *The bivariate chain and the intertwining $\Lambda P^n = P^{*n} \Lambda$.* Let π_0 and π_0^* be given distributions and P and P^* be given stochastic matrices on discrete sets \mathcal{S} and \mathcal{S}^* , respectively. Let Λ be a *link*, or transition kernel, from \mathcal{S}^* to \mathcal{S} . We seek a bivariate Markov chain $(X^*, X) = (X_n^*, X_n)_{n \geq 0}$ with margins

$$(2.12) \quad X^* \sim (\pi_0^*, P^*), \quad X \sim (\pi_0, P),$$

so that X is *linked to X^** by Λ , in the sense that (for possible conditioning values)

$$(2.13) \quad \mathcal{L}(X_n | X_0^* = x_0^*, \dots, X_n^* = x_n^*) = \Lambda(x_n^*, \cdot),$$

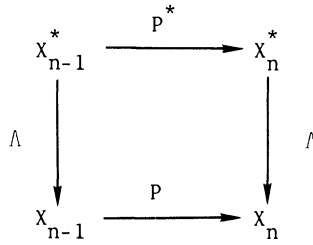


FIG. 2.

which of course implies

$$(2.14) \quad \mathcal{L}(X_n | X_n^* = x^*) = \Lambda(x^*, \cdot).$$

This problem is motivated by the desire to construct a strong stationary dual in the sense of Definition 2.1; see also Remark 2.8. Similar constructions have recently been used by Yor (1988).

If X_n^* is to be built from X_0, \dots, X_n and independent randomness—more precisely, if (2.2a) is to hold—then we necessarily have

$$(2.15) \quad X_{n-1}^* \text{ and } X_n \text{ are conditionally independent given } X_{n-1}.$$

According to (2.13) we must also have

$$(2.16) \quad X_{n-1}^* \text{ and } X_n \text{ are conditionally independent given } X_n^*.$$

The commutative diagram in Figure 2 helps to interpret (2.15) and (2.16) and Theorem 2.17. For example, according to (2.15), the three-term sequence $(X_{n-1}^*, X_{n-1}, X_n)$ is a Markov chain. Likewise, (2.16) says that (X_{n-1}^*, X_n^*, X_n) is Markov. The commutativity of the diagram is the relation $\Lambda P = P^* \Lambda$.

Theorem 2.17 shows that for an (X^*, X) satisfying (2.12), (2.14), (2.15) and (2.16) to exist, certain relations among π_0, π_0^*, P, P^* and Λ must be satisfied. Granted these relations, the proof produces such a process. To avoid trivial problems with null events, we suppose throughout that each $x^* \in \mathcal{S}^*$ is attainable, in the sense that $P\{X_n^* = x^*\} > 0$ for some $n \geq 0$.

THEOREM 2.17. *Let (π_0, P) on \mathcal{S} , (π_0^*, P^*) on \mathcal{S}^* and a transition matrix Λ from \mathcal{S}^* to \mathcal{S} be given. Then there exists a Markov chain (X^*, X) with margins $X^* \sim (\pi_0^*, P^*)$ and $X \sim (\pi_0, P)$ that has the conditional distributions (2.14) and satisfies the independence conditions (2.15) and (2.16) if and only if (π_0^*, P^*) is algebraically dual to (π_0, P) with respect to the link Λ , in the sense that*

$$(2.18) \quad \pi_0 = \pi_0^* \Lambda,$$

$$(2.19) \quad \Lambda P = P^* \Lambda.$$

PROOF. Suppose first that (X^*, X) exists. To prove (2.18), condition on X_0^* and use (2.14):

$$\begin{aligned} \pi_0(x) &= P\{X_0 = x\} = \sum_{x^*} P\{X_0^* = x^*\}P\{X_0 = x|X_0^* = x^*\} \\ &= \sum_{x^*} \pi_0^*(x^*)\Lambda(x^*, x). \end{aligned}$$

For (2.19), calculate $P\{X_n = y|X_{n-1}^* = x^*\}$ by conditioning alternatively on X_{n-1} or on X_n^* . Thus

$$\begin{aligned} P\{X_n = y|X_{n-1}^* = x^*\} &= \sum_x P\{X_{n-1} = x|X_{n-1}^* = x^*\} \\ &\quad \times P\{X_n = y|X_{n-1}^* = x^*, X_{n-1} = x\} \\ &= \sum_{y^*} P\{X_n^* = y^*|X_{n-1}^* = x^*\} \\ &\quad \times P\{X_n = y|X_{n-1}^* = x^*, X_n^* = y^*\}. \end{aligned}$$

Using (2.14) and (2.15) in the first sum gives the (x^*, y) entry of ΛP . Using (2.16) in the second sum gives the (x^*, y) entry of $P^*\Lambda$.

Conversely, if (2.18) and (2.19) are satisfied, a bivariate Markov chain can be constructed on the state space $\mathcal{S} := \{(x^*, x) : \Lambda(x^*, x) > 0\}$ having initial distribution

$$(2.20) \quad \pi_0(x^*, x) = \pi_0^*(x^*)\Lambda(x^*, x)$$

and transition function

$$(2.21) \quad \mathbf{P}((x^*, x), (y^*, y)) = \begin{cases} P(x, y)P^*(x^*, y^*)\Lambda(y^*, y)/\Delta(x^*, y), \\ \quad \text{if } \Delta(x^*, y) > 0, \\ 0, \quad \text{otherwise.} \end{cases}$$

Here $\Delta = P^*\Lambda = \Lambda P$, that is,

$$(2.22) \quad \Delta(x^*, y) := \sum_{y^*} P^*(x^*, y^*)\Lambda(y^*, y) = \sum_x \Lambda(x^*, x)P(x, y).$$

Routine calculations show that $(X^*, X) \sim (\pi_0, \mathbf{P})$ satisfies the requirements. □

REMARK 2.23. (a) Let the notations P^{*n} and $\pi_n^* = \pi_0^*P^{*n}$ for X^* mimic those for X . Then whenever (2.18) and (2.19) hold, we have the extensions

$$(2.24) \quad \Lambda P^n = P^{*n}\Lambda,$$

$$(2.25) \quad \pi_n = \pi_n^*\Lambda.$$

When (2.24) holds, one says that the transition functions $(P^n)_{n \geq 0}$ and $(P^{*n})_{n \geq 0}$ are *intertwined* by Λ .

(b) The chain (π_0, \mathbf{P}) constructed in (2.20)–(2.21) satisfies (2.13) and

$$(2.26) \quad P\{X_0^* = x_0^*|X_0 = x_0\} = \pi_0^*(x_0^*)\Lambda(x_0^*, x_0)/\pi_0(x_0)$$

and [provided $\Delta(x_{n-1}^*, x_n) > 0$]

$$\begin{aligned}
 (2.27) \quad & P\{X_n^* = x_n^* | X_0^* = x_0^*, X_0 = x_0; \dots; X_{n-1}^* = x_{n-1}^*, \\
 & X_{n-1} = x_{n-1}; X_n = x_n\} \\
 & = P^*(x_{n-1}^*, x_n^*) \Lambda(x_n^*, x_n) / \Delta(x_{n-1}^*, x_n), \quad n \geq 1.
 \end{aligned}$$

Furthermore, we have the extension

$$\begin{aligned}
 (2.28) \quad & P\{X_n = x_n | X_0^* = x_0^*, X_0 = x_0; \dots; X_{n-1}^* = x_{n-1}^*, X_{n-1} = x_{n-1}\} \\
 & = P(x_{n-1}, x_n)
 \end{aligned}$$

of (2.15).

(c) If (2.19) holds and (2.14)–(2.16) are to be satisfied, then by Bayes theorem we must have

$$(2.29) \quad P\{X_{n-1} = x | X_{n-1}^* = x^*, X_n = y\} = \Lambda(x^*, x) P(x, y) / \Delta(x^*, y)$$

and

$$(2.30) \quad P\{X_n^* = y^* | X_{n-1}^* = x^*, X_n = y\} = P^*(x^*, y^*) \Lambda(y^*, y) / \Delta(x^*, y).$$

For each (x^*, y) , let $p(x, y^* | x^*, y)$, $x \in \mathcal{S}$ and $y^* \in \mathcal{S}^*$, be any distribution on $\mathcal{S} \times \mathcal{S}^*$ having respective marginals (2.29) and (2.30). Then one can show that the declaration

$$(2.31) \quad P\{X_{n-1} = x, X_n^* = y^* | X_{n-1}^* = x^*, X_n = y\} = p(x, y^* | x^*, y)$$

leads to a (π_0, \mathbf{P}) satisfying the requirements of Theorem 2.17 via the specifications (2.20) and [for $\Lambda(x^*, x) > 0$]

$$(2.32) \quad \mathbf{P}((x^*, x), (y^*, y)) = \Delta(x^*, y) p(x, y^* | x^*, y) / \Lambda(x^*, x).$$

Our construction (2.21) simply takes the joint distribution (2.31) to be the product of the marginals (2.29) and (2.30). In terms of Figure 2, our construction makes X_{n-1} and X_n^* conditionally independent given X_{n-1}^* and X_n .

The comments of Remark 2.23 (c) have established the following result.

THEOREM 2.33. *Suppose $\pi_0 = \pi_0^* \Lambda$ and $\Lambda P = P^* \Lambda$. Then the bivariate chain (π_0, \mathbf{P}) constructed in (2.20)–(2.21) is the unique chain with marginals (π_0, P) and (π_0^*, P^*) satisfying (2.14), (2.15), (2.16) and*

$$(2.34) \quad X_{n-1} \text{ and } X_n^* \text{ are conditionally independent given } X_{n-1}^* \text{ and } X_n.$$

2.4. Sample path construction of dual X^* . Let (π_0, P) on \mathcal{S} , (π_0^*, P^*) on \mathcal{S}^* and a link Λ from \mathcal{S}^* to \mathcal{S} be given as in Theorem 2.17. Suppose $\pi_0 = \pi_0^* \Lambda$ and $\Lambda P = P^* \Lambda$. In (2.20)–(2.21) we showed how to construct from scratch a bivariate chain (X^*, X) with margins $X^* \sim (\pi_0^*, P^*)$ and $X \sim (\pi_0, P)$ so that X is linked to X^* by Λ in the sense of (2.13), which for convenience we repeat here:

$$(2.35) \quad \mathcal{L}(X_n | X_0^* = x_0^*, X_1^* = x_1^*, \dots, X_n^* = x_n^*) = \Lambda(x_n^*, \cdot).$$

Now, however, we suppose that a realization of the chain $X \sim (\pi_0, P)$ is given. Thus we are required to construct one sample path for X^* out of each path for X and, perhaps, independent randomness. We shall, in fact, require that the construction of X^* be contemporaneous with the evolution of X , more precisely, that (2.2a) hold. Then, provided X is ergodic with stationary distribution π and X^* has an absorbing state ∞ with $\Lambda(\infty, \cdot) = \pi$, we shall be able to use the constructed process X^* as a strong stationary dual of X . [If such a state ∞ does not already exist, one can adjoin it to \mathcal{S}^* and set $\pi_0^*(\infty) = 0, P^*(\infty, \infty) = 1$ and $P^*(x^*, \infty) = P^*(\infty, x^*) = 0$ for $x^* \neq \infty$.] We shall call any strong stationary dual satisfying (2.35) a Λ -linked dual.

Equations (2.26) and (2.27) light the way to a solution. Explicitly, when $X_0 = x_0$ is observed, use independent randomness and set

$$(2.36a) \quad X_0^* = x_0^* \text{ with probability } \pi_0^*(x_0^*)\Lambda(x_0^*, x_0)/\pi_0(x_0).$$

Proceed inductively. Suppose that $X_0 = x_0, \dots, X_{n-1} = x_{n-1}$ have been observed and that $X_0^* = x_0^*, \dots, X_{n-1}^* = x_{n-1}^*$ have been chosen. When $X_n = x_n$ is observed, recall the notation $\Delta = P^*\Lambda$ and set

$$(2.36b) \quad X_n^* = x_n^* \text{ with probability } P^*(x_{n-1}^*, x_n^*)\Lambda(x_n^*, x_n)/\Delta(x_{n-1}^*, x_n).$$

Then $(X^*, X) \sim (\pi_0, P)$ as in (2.20)–(2.21). Starting with algebraic duality ($\pi_0 = \pi_0^*\Lambda$ and $\Lambda P = P^*\Lambda$), we have built a sample dual $X^* \sim (\pi_0^*, P^*)$ to which X is linked by Λ .

REMARK 2.37. Our construction (2.36) of a sample path dual uses a “mixture sifting” technique which one can regard as an algorithmic form of Bayes theorem. To explain, write out $\pi_0 = \pi_0^*\Lambda$ as

$$\pi_0(\cdot) = \sum_{x_0^*} \pi_0^*(x_0^*)\Lambda(x_0^*, \cdot).$$

This exhibits π_0 as a mixture of the distributions $\Lambda(x_0^*, \cdot)$. The usual operational form of this representation generates $X_0 \sim \pi_0$ by first observing the values x_0^* of $X_0^* \sim \pi_0^*$ and then conditionally generating X_0 according to $\Lambda(x_0^*, \cdot)$. We want to reverse the procedure. Given X_0 chosen from π_0 we seek to generate X_0^* so that [recall (2.14)]

$$\mathcal{L}(X_0|X_0^* = x_0^*) = \Lambda(x_0^*, \cdot).$$

Of course, Bayes theorem provides the solution (2.36a). The subsequent construction of X_1^*, X_2^*, \dots via (2.36b) is interpreted similarly.

We call a dual *sharp* if its associated strong stationary time is optimal.

DEFINITION 2.38. Let X^* be a strong stationary dual for X . Let $T = T_\infty^*$ be the corresponding strong stationary time. If T is a time to stationarity, we say that the dual X^* is *sharp*.

REMARK 2.39. When is the dual X^* constructed in (2.36) sharp? Recall from (2.25) that $\pi_n = \pi_n^* \Lambda$. From this and the definition of separation, it is easy to see that $s(n) = P\{T_\infty^* > n\}$ holds for a particular value of n if (and for finite \mathcal{S} , only if) there exists $y \in \mathcal{S}$ having the property that for each $y^* \neq \infty$ in \mathcal{S}^* , either $\pi_n^*(y^*) = 0$ or else $\Lambda(y^*, y) = 0$. Therefore if this condition holds for every $n \geq 0$, then X^* is sharp. We shall make use of this remark in Section 3.

EXAMPLE 2.40. *A sharp dual via the Aldous–Diaconis construction.* Return to Example 2.6 and recall the specific definitions (2.9)–(2.11) for that example. One checks easily that $\pi_0 = \pi_0^* \Lambda$ and that $\Lambda P = P^* \Lambda = \Delta$ with

$$(2.41a) \quad \Delta(n - 1, y) = [\pi_n(y) - (1 - s(n - 1))\pi(y)]/s(n - 1),$$

$$n = 1, 2, \dots,$$

$$(2.41b) \quad \Delta(\infty, y) = \pi(y).$$

Thus (π_0, P) and (π_0^*, P^*) are in algebraic duality with respect to the link Λ of (2.11). The sample path construction (2.7) of X^* is simply a special case of (2.36). For finite \mathcal{S} , Remark 2.39 ensures that the strong stationary dual X^* , so constructed, is sharp. The sharpness of X^* is also easy to verify when \mathcal{S} is infinite.

Remark 2.37 provides insight into the Aldous–Diaconis construction. The idea is to express one-dimensional laws of the given X as mixtures of π and other distributions so as to make the mixing coefficient for π as large as possible. To explain, first observe that π_0 is the mixture

$$(2.42) \quad \pi_0 = s(0)\lambda_0 + (1 - s(0))\pi$$

of the probability distributions

$$\lambda_0(\cdot) := \Lambda(0, \cdot) = [\pi_0(\cdot) - (1 - s(0))\pi(\cdot)]/s(0)$$

and π . To see that λ_0 is indeed a probability, note that for each $x \in \mathcal{S}$

$$\pi_0(x) = [1 - s(0, x)]\pi(x) \geq [1 - s(0)]\pi(x).$$

The infimum over x of $1 - s(0, x)$ is $1 - s(0)$, so the coefficient of π in (2.42) is as large as possible. From (2.42) we conclude that $\pi_0 = \pi_0^* \Lambda$ provided that $\pi_0^*(0) = s(0) = 1 - \pi_0^*(\infty)$, which is (2.9).

We proceed in a similar fashion. Note that

$$\begin{aligned} \lambda_0 P &= \frac{\pi_1 - (1 - s(0))\pi}{s(0)} \\ &= \frac{s(1)}{s(0)}\lambda_1 + \left(1 - \frac{s(1)}{s(0)}\right)\pi, \end{aligned}$$

where λ_1 is the probability distribution

$$\lambda_1(\cdot) := \Lambda(1, \cdot) = [\pi_1(\cdot) - (1 - s(1))\pi(\cdot)]/s(1).$$

Thus the zeroth rows of ΛP and $P^* \Lambda$ agree provided that $P^*(0, 1) = s(1)/s(0) = 1 - P^*(0, \infty)$, which is (2.10a) for $n = 1$. Continuing, we derive similar motivation for the definition of the remaining rows of P^* .

2.5. *Early stopping bounds on total variation via duality.* One advantage in building a linked dual rather than just a strong stationary time is that useful information can be gleaned by stopping the process early, that is, before absorption at ∞ . For example, each of the hitting times T_{x^*} , $x^* \in \mathcal{S}^*$, is a randomized stopping time for X . As the following simple lemma shows, stopping times can yield useful bounds on total variation distance. Throughout Section 2.5 we take $X \sim (\pi_0, P)$ to be a given ergodic Markov chain with stationary distribution π .

LEMMA 2.43. *Let T be any randomized stopping time for X . Then*

$$(2.44) \quad \|\pi_n - \pi\| \leq P\{T > n\} + \sum_{k=0}^n P\{T = k\} v_k(n), \quad n \geq 0,$$

where

$$(2.45) \quad v_k(n) := \|\mathcal{L}(X_n | T = k) - \pi\|, \quad 0 \leq k \leq n.$$

PROOF. If $A \subset \mathcal{S}$ and $n \geq 0$, then

$$\begin{aligned} \pi_n(A) - \pi(A) &\geq \sum_{k=0}^n P\{T = k\} [P\{X_n \in A | T = k\} - \pi(A)] - \pi(A) P\{T > n\} \\ &\geq - \left[P\{T > n\} + \sum_{k=0}^n P\{T = k\} v_k(n) \right]. \end{aligned}$$

Applying the same inequality to the complement of A we deduce (2.44). \square

Lemma 2.43 simplifies for random times satisfying the independence condition, but not necessarily the stationarity assumption, in the definition of a strong stationary time.

THEOREM 2.46. *Let T be a randomized stopping time for X such that T and X_T are independent. Then*

$$(2.47) \quad \|\pi_n - \pi\| \leq v + (1 - v)P\{T > n\}, \quad n \geq 0,$$

with

$$(2.48) \quad v = \|\mathcal{L}(X_T) - \pi\|.$$

PROOF. (2.47) follows simply from Lemma 2.43 and the observation

$$(2.49) \quad v_k(n) \leq v_k(k) = v, \quad 0 \leq k \leq n.$$

The inequality in (2.49) reflects the fact that total variation decreases monotonically in time for any Markov chain. The equality is an immediate consequence of the independence of T and X_T . \square

COROLLARY 2.50. *Let X^* be a Λ -linked dual for X . Let T be the hitting time of a given subset A^* of \mathcal{S}^* for X^* . Suppose that $X_T^* \sim \sigma^*$ is independent of T . Then*

$$(2.51) \quad \|\pi_n - \pi\| \leq v + (1 - v)P\{T > n\}, \quad n \geq 0,$$

with

$$(2.52) \quad v = \|\sigma^*\Lambda - \pi\|.$$

REMARK 2.53. The independence condition of Corollary 2.50 is always met when A^* is a singleton $\{x^*\}$, in which case

$$(2.54) \quad v = \|\Lambda(x^*, \cdot) - \pi(\cdot)\|.$$

A different application is given for random walk on a discrete cube in Example 3.2.

The following theorem drops the independence condition in Corollary 2.50 and so yields a cruder, though sometimes useful, bound. Its proof is transparent.

THEOREM 2.55. *Let X^* be a Λ -linked dual for X . Let T be the hitting time of a given subset A^* of \mathcal{S}^* for X^* . Then*

$$(2.56) \quad \|\pi_n - \pi\| \leq v + (1 - v)P\{T > n\}, \quad n \geq 0,$$

with

$$(2.57) \quad v = \sup_{y^* \in A^*} \|\Lambda(y^*, \cdot) - \pi\|.$$

REMARK 2.58. When $A^* = \{\infty\}$ in Corollary 2.50 or Theorem 2.55, i.e., when $T = T_\infty^*$ is the strong stationary time associated with X^* , we have $v = 0$ and the conclusion of each theorem is $\|\pi_n - \pi\| \leq P\{T > n\}$. This is otherwise a consequence of (1.9) ($\|\pi_n - \pi\| \leq s(n)$) and (1.11) ($s(n) \leq P\{T > n\}$). As Example 3.2 shows, however, judicious choice of A^* in Corollary 2.50 can yield substantial improvement.

3. Set-valued duals with truncated stationary distributions as link.

3.1. *Introduction.* Thus far the choice of \mathcal{S}^* , π_0^* , P^* and Λ in our construction (2.36) of a Λ -linked strong stationary dual chain has been left open, subject only to the restrictions $\pi_0 = \pi_0^*\Lambda$ and $\Lambda P = P^*\Lambda$. We shall often find it useful to let \mathcal{S}^* consist of subsets of the state space \mathcal{S} and to take $\Lambda(x^*, \cdot)$ to be the stationary distribution π truncated to x^* . This choice will be particularly well suited to the Markov chains considered in Section 4. Furthermore, one finds that nearly all previous examples of strong stationary times in the literature correspond naturally to such set-valued duals.

We consider three diverse applications in Section 3.2. In Section 3.3 we point out a limitation of set-valued duals: There may be none whose associated

strong stationary time is fastest among all strong stationary times. Nevertheless, a set-valued dual can be built for any (ergodic) chain, and in Section 3.4 we present a “greedy” algorithm for finite-state chains that forms the basis of the specialized dual construction in Section 4.

3.2. *Setup and examples.* Let $X \sim (\pi_0, P)$ be, as usual, an ergodic Markov chain with discrete state space \mathcal{S} and stationary distribution π . Let \mathcal{S}^* denote a collection of nonempty subsets of \mathcal{S} ; we assume $\mathcal{S} \in \mathcal{S}^*$. For each $x^* \in \mathcal{S}^*$, let $\Lambda(x^*, \cdot)$ denote the truncation of the stationary distribution π to the set x^* :

$$(3.0) \quad \Lambda(x^*, x) = I_{x^*}(x)\pi(x) / \sum_{y \in x^*} \pi(y), \quad x \in \mathcal{S}.$$

In (3.0), I_{x^*} is the indicator of the set x^* . Note that $\Lambda(x^*, \cdot) = \pi$ if and only if $x^* = \mathcal{S}$, and so \mathcal{S} will play the role of ∞ in Definition 2.1. If x^* is a singleton $\{x\}$, then $\Lambda(x^*, \cdot)$ is unit mass at x . If \mathcal{S} is finite, then so is \mathcal{S}^* .

Our setup leaves free the particular choice of \mathcal{S}^* . Here are three examples. More are given in Section 4.

EXAMPLE 3.1. *Random walk on the d -point circle, revisited.* We return to the setting of Example 1.1 and suppose again that $d = 2^a$, for simplicity that $d = 16$, and that the walk starts at $2^{a-1} = 8$ rather than 0. Here $\mathcal{S}^* = A_0^* \cup \dots \cup A_4^*$, with $A_0^* = \{\{8\}, \{7, 9\}, \{6, 10\}, \{5, 11\}\}$, $A_1^* = \{\{4, 12\}, \{3, 5, 11, 13\}\}$, $A_2^* = \{\{2, 6, 10, 14\}\}$, $A_3^* = \{\{\text{odds}\}, \{\text{evens}\}\}$ and $A_4^* = \{\mathcal{S}\}$. The dual chain X^* advances progressively from A_0^* to A_4^* , in each case entering A_j^* through its first listed element. According to Remark 2.39 the dual is *not* sharp; the problem is that for $n \geq 8$ we have $\pi_n^*(y^*) > 0$ for every $y^* \in \mathcal{S}^*$, while for any $y \in \mathcal{S}$ we have either $\Lambda(\{\text{odds}\}, y) > 0$ or $\Lambda(\{\text{evens}\}, y) > 0$.

Suppose, however, that the holding probability at each step is changed from $1/3$ to $1/2$. In this case, once the dual X^* has reached the set $\{\text{odds}\}$ (which happens at time $T_{a-1} = T_3$ in the terminology of Example 1.1), we need only stop the walk after one more step. Thus $\{\text{evens}\}$ can be removed from \mathcal{S}^* , and by Remark 2.39 with (say) $y = 2^a - 1 = 15$, the resulting dual is sharp.

EXAMPLE 3.2. *Simple symmetric random walk on d -dimensional cube.* Consider a random walk X on the discrete d -dimensional cube \mathbb{Z}_2^d that at each step stays fixed or moves to a uniformly chosen nearest neighbor. For simplicity we take the probability of staying fixed to be $1/2$. The random walk starts at a fixed point, say, $\mathbf{0} = (0, \dots, 0)$.

It is known [see, e.g., Chapter 4 in Diaconis (1988)] that when d is large it takes $\frac{1}{2}d \log d + cd$ steps with c large to make variation distance small, but $d \log d + cd$ steps with c large to make separation small. Thus the bound $\|\pi_n - \pi\| \leq s(n)$ is inadequate for asymptotic evaluation of total variation as $d \rightarrow \infty$. However, we show here how a construction of a strong stationary time for this walk discovered by Mark Brown and Andrei Broder readily yields a

strong stationary dual and we exploit this dual via Corollary 2.50 to recapture the factor of $\frac{1}{2}$ in the leading $d \log d$ term for the number of steps.

The construction is simple to describe in terms of checked coordinates of the vertices. Initially none of the d coordinates is checked. When the walk takes a step along the c th coordinate, check that coordinate (if indeed it is not already checked). When the walk stays fixed, select a coordinate at random (uniformly) and, if unchecked, check it. Thus coordinates can be checked even when the walk stays fixed. The claim is that the first time T that all coordinates are checked is a time to stationarity for the walk.

To see this, let X_n^* record the positions of the checked coordinates at time n . If we identify $x^* = \{c_1, \dots, c_j\} \in \mathcal{S}^*$ with the set of all points in $\mathcal{S} = \mathbb{Z}_2^d$ with zeros in the unchecked positions, then X^* is the strong stationary dual (2.36) having (3.0) as link. The dual is sharp: Use Remark 2.39 with $y = (1, 1, \dots, 1)$.

To apply the early stopping Corollary 2.50, fix j , $0 \leq j \leq d$, and let A^* consist of all j -element members of \mathcal{S}^* . Thus T is the first time j coordinates have been checked. Clearly T and X_T^* are independent and $\sigma^* = \mathcal{L}(X_T^*)$ is uniform probability on A^* . We need to calculate $v = \|\sigma^* \Lambda - \pi\|$.

Since the respective sizes of A^* and any $x^* \in A^*$ are $\binom{d}{j}$ and 2^j , it follows that

$$(3.3) \quad (\sigma^* \Lambda)(x) = \left[\binom{d}{j} 2^j \right]^{-1} \times \#\{x^*: x^* \in A^* \text{ and } x_c = 0 \text{ for all } c \notin x^*\}$$

for $x = (x_1, \dots, x_d) \in \mathcal{S}$. Let $|x| = k$ be the number of ones in a given x . If $(\sigma^* \Lambda)(x) > 0$, then $k \leq j$. Conversely, if $k \leq j$, then $x^* \in \mathcal{S}^*$ belongs to the set in (3.3) if and only if it consists of the k positions of ones in x along with any other $j - k$ positions. There are $\binom{d-k}{j-k}$ such x^* . Hence

$$(3.4) \quad (\sigma^* \Lambda)(x) = \begin{cases} \left[\binom{d}{j} 2^j \right]^{-1} \binom{d-k}{j-k} = 2^{-j} \binom{j}{|x|} / \binom{d}{|x|}, & \text{if } |x| \leq j, \\ 0, & \text{otherwise.} \end{cases}$$

Now a straightforward computation shows that $v = \|\sigma^* \Lambda - \pi\|$ equals the total variation distance between the binomial $(j, \frac{1}{2})$ and binomial $(d, \frac{1}{2})$ distributions. This result is not surprising, as the number of ones in X_T has the binomial $(j, \frac{1}{2})$ distribution, while the number of ones for the chain in stationarity has the binomial $(d, \frac{1}{2})$ distribution.

To proceed we turn to asymptotics in the dimension d . A simple central limit argument shows that, for large d , v is small when j is chosen to be $d - bd^{1/2}$ with b small. More precisely, let $b > 0$ be fixed and let $j = j(d)$ be the smallest integer at least $d - bd^{1/2}$. Then $v = v(d)$ converges as $d \rightarrow \infty$ to the variation distance between unit normals with means differing by b , namely, $P\{|Z| \leq b/2\}$ with Z a standard normal. Furthermore (recall the coupon collector's problem), $T = T(d)$ then has mean $d \sum l^{-1}$ and variance $d^2 \sum l^{-2} - d \sum l^{-1}$, where the sums are over $bd^{1/2} < l \leq d$. Fix $c \geq 0$ and let

$n = n(d)$ equal $\frac{1}{2}d \log d + cd$ (rounded to an integer). Straightforward calculations involving Chebychev's inequality then show that if $c > \log(1/b)$, then $P\{T > n\} \rightarrow 0$ as $d \rightarrow \infty$. Hence from (2.51) it follows that

$$(3.5) \quad \limsup_{d \rightarrow \infty} \|\pi_n - \pi\| \leq P\{|Z| \leq e^{-c}/2\}.$$

That the left side of (3.5) tends to 0 as $c \rightarrow \infty$ is precisely what is meant by saying that $\frac{1}{2}d \log d + cd$ steps with c large are sufficient to make total variation small for the walk. For this example Corollary 2.50 is remarkably sharp: Diaconis, Graham and Morrison (1988) show

$$\lim_{d \rightarrow \infty} \|\pi_n - \pi\| = P\{|Z| \leq e^{-c}/2\}.$$

REMARK 3.6. The factor $\frac{1}{2}$ differentiating the respective numbers of steps needed to make total variation and separation small in the previous example is typical of many problems. For random walks on finite groups Aldous and Diaconis (1987) showed that $\phi(s(2n)) \leq \|\pi_n - \pi\| \leq s(n)$ for a universal continuous function ϕ satisfying $\phi(x) \sim x^2/32$ as $x \rightarrow 0$. Thus if separation is small, then variation is small and if variation is small, then separation is small in at most twice as many steps.

EXAMPLE 3.7. *Top in at random card shuffle.* Let X_n record the order of a deck of d cards started in a given order, say, d (on top), \dots , 1 (on bottom), after n "top in at random" shuffles as described in Aldous and Diaconis (1986). Let $x^*(a_1, \dots, a_l)$, $0 \leq l \leq d - 2$, denote the set of permutations of the deck that begin with the cards (in order from top to bottom) $a_1, \dots, a_l, 2$, where $a_i \neq 1$ for every i . Let \mathcal{S}^* consist of all these $x^*(a_1, \dots, a_l)$ and \mathcal{S} . Then, for suitable π_0^* and P^* , the (π_0, P) governing X and (π_0^*, P^*) are in algebraic duality with respect to the link (3.0). Moreover, the strong stationary dual constructed by (2.36) is sharp, as follows from Remark 2.39, and the associated strong stationary time T has the simple description "wait until the card labelled 2 rises to the top and is shuffled in at random." This T has a simple analysis in terms of the standard coupon collector's problem; see Diaconis (1988) for details.

For this problem, sharp asymptotics for total variation can be derived. If $n = d \log d + cd$, with $c \geq 0$, then

$$\lim_{d \rightarrow \infty} \|\pi_n - \pi\| = \frac{1}{2} [1 - (1 + e^{-c})e^{-e^{-c}}].$$

3.3. *A sharp set-valued dual need not exist.* In Example 2.40 we showed how to construct a sharp dual for any ergodic Markov chain, but the dual is not in general set-valued. On the other hand, in each of Examples 3.1 (with holding probability $1/2$), 3.2 and 3.7 we constructed a sharp set-valued dual.

The following question therefore arises naturally. Can a sharp dual always be built using the truncated stationary distributions as link? The answer is no.

EXAMPLE 3.8. Let $\mathcal{S} = \{0, 1, 2\}$, $\pi_0 = (1, 0, 0)$ and

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{pmatrix} 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0.9 \\ 0.9 & 0.1 & 0 \end{pmatrix} \end{matrix}.$$

This chain is irreducible and aperiodic with stationary distribution $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Also,

$$\begin{aligned} \pi_1 &= (0, 0.9, 0.1), \\ \pi_2 &= (0.18, 0.01, 0.81), \\ \pi_3 &= (0.73, 0.243, 0.027). \end{aligned}$$

Clearly π_0^* must put unit mass at $\{0\}$, while, for some $0 \leq \alpha \leq 1$, $\pi_1^* = P^*({0}, \cdot)$ puts masses $0.8 + 0.1\alpha$, 0.1α and $0.2(1 - \alpha)$ at $\{1\}$, $\{2\}$ and $\{1, 2\}$, respectively. But $P^*({1}, \mathcal{S}) = 0 = P^*({2}, \mathcal{S})$ and $P^*({1, 2}, \mathcal{S}) \leq 0.15$, so π_2^* puts mass $\leq 0.03(1 - \alpha)$ at \mathcal{S} . If (π_0^*, P^*) is sharp, then $\pi_2^*(\mathcal{S}) = 1 - s(2) = \pi_2(1)/\pi(1) = 0.03$, and so $\alpha = 0$ and $P^*({1, 2}, \mathcal{S}) = 0.15$. Then $P^*({1, 2}, \{0, 2\}) \leq 0.8$, whence

$$\begin{aligned} \pi_2^*({0, 2}) &= \pi_1^*({1})P^*({1}, \{0, 2\}) + \pi_1^*({1, 2})P^*({1, 2}, \{0, 2\}) \\ &\leq (0.8)(0.2) + (0.2)(0.8) = 0.32. \end{aligned}$$

But then

$$\begin{aligned} \pi_3^*(\mathcal{S}) &= \pi_2^*({0, 2})P^*({0, 2}, \mathcal{S}) + \pi_2^*(\mathcal{S})P^*(\mathcal{S}, \mathcal{S}) \\ &\leq (0.32)(0.15) + (0.03)(1) = 0.078, \end{aligned}$$

while if (π_0^*, P^*) is sharp, then $\pi_3^*(\mathcal{S}) = 1 - s(3) = \pi_3(2)/\pi(2) = 0.081$. Thus no sharp set-valued dual exists.

3.4. *Greedy construction of a set-valued dual.* Although no sharp set-valued dual need exist (consult the previous subsection), there is for finite-state chains a natural “greedy” construction of a set-valued dual which we describe here. As evidence that such a dual might generally work well, we note that the greedy dual of any chain in the class considered in Section 4—which includes most finite-state birth and death chains—is sharp.

Let $X \sim (\pi_0, P)$ be a given ergodic Markov chain with finite state space \mathcal{S} and stationary distribution π . We show how to find an algebraic dual (π_0^*, P^*) on the full space $\mathcal{S}^* = \{x^* \neq \emptyset: x^* \subset \mathcal{S}\}$ with link Λ . Then a strong stationary dual chain $X^* \sim (\pi_0^*, P^*)$ can be built using the general construction (2.36).

First we construct π_0^* . Although the details (3.9)–(3.13) look complicated, the idea is simple. To establish the algebraic duality $\pi_0 = \pi_0^*\Lambda$, we need to express π_0 as a mixture of truncated stationary distributions $\Lambda(x^*, \cdot)$; the

mixing coefficients will be the numbers $\pi_0^*(x^*)$. Adopting a greedy demeanor, we seek to make $\pi_0^*(\mathcal{S})$ as large as possible; after all, this is the probability that the strong stationary time associated with the dual we are going to construct stops at time 0. It is clear that the largest possible value for $\pi_0^*(\mathcal{S})$ is $\min\{\pi_0(z)/\pi(z) : z \in \mathcal{S}\} = 1 - s(0)$. Now, going beyond the dual construction in Example 2.40, we need to resolve further the remainder subprobability distribution $\sigma = \pi_0 - (1 - s(0))\pi$, specifically as a mixture of truncated stationary distributions. The largest possible set z^* which can appear with positive coefficient in such a mixture (remember, we're greedy) is clearly the collection of $z \in \mathcal{S}$ with $\sigma(z) > 0$, and the largest possible coefficient for this $\Lambda(z^*, \cdot)$ is $\min\{\sigma(z)/\pi(z) : z \in \mathcal{S}\} = c$ (say). The remainder $\sigma - c\pi$ is a subprobability on z^* which we resolve, etc., continuing until π_0 is completely expressed as a mixture of truncated stationary distributions.

We now proceed to a formal description. We shall show recursively how to define a strictly decreasing sequence $z_1^* \supset z_2^* \supset \dots \supset z_\nu^* (1 \leq \nu < \infty)$ of nonempty subsets of \mathcal{S} and strictly positive numbers c_1, \dots, c_ν such that

$$(3.9) \quad \pi_0 = \sum_{r=1}^{\nu} c_r H(z_r^*) \Lambda(z_r^*, \cdot),$$

where here $H(z^*) := \sum_{z \in z^*} \pi(z)$. One may then define

$$(3.10) \quad \pi_0^*(z_r^*) := c_r H(z_r^*), \quad r = 1, \dots, \nu,$$

and $\pi_0^*(z^*) := 0$, otherwise, in order to satisfy the duality condition $\pi_0 = \pi_0^* \Lambda$ for initial distributions. In performing the recursion we shall produce auxiliary subprobability measures $\pi_0^{(r)}$ on z_r^* . We begin the recursion by defining $z_0^* = \mathcal{S}$ and $\pi_0^{(0)} = \pi_0$.

Suppose that z_{r-1}^* and $\pi_0^{(r-1)}$ have been constructed. Set

$$(3.11) \quad c_r := \min \left\{ \frac{\pi_0^{(r-1)}(z)}{\pi(z)} : z \in z_{r-1}^*, \pi_0^{(r-1)}(z) > 0 \right\}.$$

If the set on the right is empty, i.e., if $\pi_0^{(r-1)}$ is the zero measure on z_{r-1}^* , stop and set $\nu = r - 1$. Otherwise, $c_r > 0$ and we continue by setting

$$(3.12) \quad z_r^* := \left\{ z \in z_{r-1}^* : \frac{\pi_0^{(r-1)}(z)}{\pi(z)} \geq c_r \right\} \neq \emptyset,$$

$$(3.13) \quad \pi_0^{(r)} := \pi_0^{(r-1)} - c_r \pi \quad \text{on } z_r^*.$$

Upon completion of the algorithm, (3.9) clearly holds.

In just the same way that π_0^* is constructed from π_0 , so each $P^*(x^*, \cdot)$, $x^* \subset \mathcal{S}$, is constructed from the x^* row of ΛP . Note that if $x^* = \mathcal{S}$, then $\nu = 1$, $c_1 = 1$ and $z_1^* = \mathcal{S}$: $P^*(\mathcal{S}, \mathcal{S}) = 1$. Thus the greedy algorithm makes \mathcal{S} an absorbing state for P^* , in accordance with (2.2c) in the definition (Definition 2.1) of a strong stationary dual.

REMARK 3.14. Application of the greedy algorithm to the (π_0, P) of Example 3.8 leads to $\pi_3^* = \pi_0^* P^{*3}$ putting masses 0.512, 0.024, 0.384, 0.002 and 0.078, respectively, at $\{0\}$, $\{1\}$, $\{0, 1\}$, $\{1, 2\}$ and $\mathcal{S} = \{0, 1, 2\}$. Since $1 - s(3) = 0.081 > 0.078$, we see directly that the greedy (π_0^*, P^*) is not sharp.

4. Chains with monotone likelihood ratio.

4.1. *Introduction.* The construction and analysis of dual chains simplifies considerably if the observed chain has a certain monotonicity property. To be specific, suppose (throughout Section 4) that $X \sim (\pi_0, P)$ is an ergodic Markov chain on the finite, linearly ordered state space $\mathcal{S} = \{0, \dots, d\}$ with stationary distribution π . Suppose that $\pi_0(x)/\pi(x)$ is monotone (say, decreasing) in x . Then the initial distribution π_0^* for the set-valued dual constructed according to the greedy algorithm of Section 3.4 is concentrated on intervals of \mathcal{S} of the form $\{0, \dots, z^*\}$. We obtain an analogous result for the dual transitions P^* under a suitable monotonicity assumption for P . Then, by identifying initial segments of \mathcal{S} with their right endpoints, the dual chain can be taken to have values on the original state space \mathcal{S} .

Section 4.2 identifies exactly when such a construction is possible and carries out the construction in that case. It shows that the resulting dual is sharp. Section 4.3 tailors the results to birth and death chains. This gives the promised stochastic interpretation of the eigenvalues of P . Section 4.4 works through some examples, including the two-state chain, the Ehrenfest chain, simple (asymmetric) random walk on $\{0, \dots, d\}$ and simple symmetric random walk on a finite tree.

4.2. *Duality for chains with monotone likelihood ratio.* In this section we derive a necessary and sufficient condition for the given (π_0, P) to possess, with respect to the link Λ , an algebraic dual (π_0^*, P^*) whose state space \mathcal{S}^* consists of precisely those subsets of \mathcal{S} of the form $\{0, \dots, x^*\}$ with $x^* \in \mathcal{S}$. Then \mathcal{S}^* can be identified with \mathcal{S} in an obvious fashion. Here we take the link Λ to be the family

$$(4.1) \quad \Lambda(x^*, x) = I_{\{0, \dots, x^*\}}(x) \pi(x) / H(x^*), \quad x^* \in \mathcal{S}, x \in \mathcal{S},$$

of truncated stationary distributions, where

$$(4.2) \quad H(x^*) := \sum_{x \in \mathcal{S}: x \leq x^*} \pi(x), \quad x^* \in \mathcal{S},$$

is the cumulative distribution function for the stationary distribution π .

Once we have an algebraic dual, we shall use the general sample-path construction of Section 2.4 to produce a Λ -linked strong stationary dual X^* for X . We shall show that this dual is sharp (in the sense of Definition 2.38).

To construct an algebraic dual, π_0^* and P^* must be found to satisfy

$$\pi_0 = \pi_0^* \Lambda \quad \text{and} \quad \Lambda P = P^* \Lambda.$$

The first of these relations says

$$\pi_0(x) = \sum_{x^* \geq x} \pi_0^*(x^*)\pi(x)/H(x^*), \quad x \in \mathcal{S},$$

or, equivalently,

$$(4.3) \quad \pi_0^*(x^*) = H(x^*) \left[\frac{\pi_0(x^*)}{\pi(x^*)} - \frac{\pi_0(x^* + 1)}{\pi(x^* + 1)} \right], \quad x^* \in \mathcal{S},$$

with the convention $\pi_0(d + 1)/\pi(d + 1) = 0$. The solution (4.3) to $\pi_0 = \pi_0^*\Lambda$ is nonnegative if and only if the likelihood ratio π_0/π is decreasing (here and elsewhere we use decreasing for nonincreasing). This solution sums to unity. Note that if $\pi_0 = \delta_0$, then $\pi_0^* = \delta_0$ also.

The relation $\Lambda P = P^*\Lambda$ says

$$\left[\sum_{x \leq x^*} \pi(x)P(x, y)/\pi(y) \right] / H(x^*) = \sum_{y^* \geq y} P^*(x^*, y^*)/H(y^*), \quad x^*, y^* \in \mathcal{S}.$$

To interpret this, let the time reversal of P be denoted

$$(4.4) \quad \tilde{P}(y, x) = \pi(x)P(x, y)/\pi(y), \quad x, y \in \mathcal{S}.$$

Let $\tilde{X} = (\tilde{X}_n)_{n \geq 0}$ be a Markov chain with transition function \tilde{P} . When \tilde{X} is started deterministically in state y , write P_y for the probability measure P . The relation $\Lambda P = P^*\Lambda$ then says

$$P_y\{\tilde{X}_1 \leq x^*\}/H(x^*) = \sum_{y^* \geq y} P^*(x^*, y^*)/H(y^*), \quad x^*, y \in \mathcal{S},$$

or, equivalently,

$$(4.5) \quad P^*(x^*, y^*) = \frac{H(y^*)}{H(x^*)} [P_{y^*}\{\tilde{X}_1 \leq x^*\} - P_{y^*+1}\{\tilde{X}_1 \leq x^*\}], \quad x^*, y^* \in \mathcal{S}.$$

In (4.5) we have set $P_{d+1}\{\tilde{X}_1 \leq x^*\} = 0$ for all $x^* \in \mathcal{S}$.

The upshot is that $\Lambda P = P^*\Lambda$ has a nonnegative solution if and only if $P_{y^*}\{\tilde{X}_1 \leq x^*\}$ decreases in y^* for each fixed x^* , in which case the unique P^* is given by (4.5). From $\Lambda P = P^*\Lambda$ it is clear that each row of P^* sums to unity.

The condition that $P_y\{\tilde{X}_1 \leq x\}$ be decreasing in y for each x has been called (*stochastic monotonicity*) by Daley (1968). We discuss this and its relation to Siegmund duality in Section 5.

These observations are summarized in the following theorem.

THEOREM 4.6. *Let X be an irreducible aperiodic Markov chain on $\mathcal{S} = \{0, \dots, d\}$ with initial distribution π_0 and transition function P . Let π denote the stationary distribution and H the cumulative of π . Let the time reversal of*

P be denoted $\tilde{P}(x, y) = \pi(y)P(y, x)/\pi(x)$. The (π_0, P) has an algebraic dual (π_0^*, P^*) on $\mathcal{S}^* = \mathcal{S}$ with respect to the link $\Lambda(x^*, x) = I_{\{0, \dots, x^*\}}(x)\pi(x)/H(x^*)$ if and only if

$$(4.7) \quad \frac{\pi_0(x)}{\pi(x)} \text{ decreases in } x$$

and

$$(4.8) \quad \tilde{P} \text{ is monotone.}$$

In this case, the dual (π_0^*, P^*) is uniquely determined as

$$(4.9) \quad \pi_0^*(x^*) = H(x^*) \left[\frac{\pi_0(x^*)}{\pi(x^*)} - \frac{\pi_0(x^* + 1)}{\pi(x^* + 1)} \right], \quad x^* \in \mathcal{S},$$

$$(4.10) \quad P^*(x^*, y^*) = \frac{H(y^*)}{H(x^*)} [P_{y^*}\{\tilde{X}_1 \leq x^*\} - P_{y^*+1}\{\tilde{X}_1 \leq x^*\}],$$

$x^*, y^* \in \mathcal{S}.$

REMARK 4.11. The monotonicity conditions (4.7)–(4.8) can be expressed another way. Say that (π_0, P) is $\text{MLR}(n)$ [has monotone (decreasing) likelihood ratio at time n] if $\pi_n(y)/\pi(y)$ decreases as y increases. Call the transition matrix P MLR preserving if (π_0, P) is $\text{MLR}(1)$ whenever (π_0, P) is $\text{MLR}(0)$. If (π_0, P) is $\text{MLR}(0)$ and P is MLR preserving, say that (π_0, P) is MLR . Of course, an MLR chain is $\text{MLR}(n)$ for every n . We have not investigated the converse.

The connection with duality comes from the following obvious fact: A distribution σ on $\{0, 1, \dots, d\}$ has monotone decreasing likelihood ratio relative to the stationary distribution π if and only if σ is a mixture of the truncated stationary distributions $\Lambda(x^*, \cdot)$ of (4.1). Thus an algebraic dual as in Theorem 4.6 exists if and only if X is MLR . For example, we can argue that $\Lambda P = P^* \Lambda$ has a stochastic solution P^* if and only if P preserves MLR . Indeed, if P preserves MLR , fix $x^* \in \mathcal{S}$. Since $(\Lambda(x^*, \cdot), P)$ is $\text{MLR}(0)$, it must be that $(\Lambda(x^*, \cdot), P)$ is $\text{MLR}(1)$, i.e., that $\sum_x \Lambda(x^*, x)P(x, \cdot)$ is a mixture of truncated stationary distributions. So $\Lambda P = P^* \Lambda$ for some P^* . Conversely, if $\Lambda P = P^* \Lambda$, then $(\Lambda(x^*, \cdot), P)$ is $\text{MLR}(1)$ for every $x^* \in \mathcal{S}$; taking mixtures, one concludes that P preserves MLR .

REMARK 4.12. If (π_0, P) is an MLR chain, then the greedy construction of an algebraic dual described in Section 3.4 produces the (π_0^*, P^*) of Theorem 4.6.

We now use (2.36) in Section 2.4 to build, pathwise, a Λ -linked strong stationary dual X^* for an MLR chain $X \sim (\pi_0, P)$. First, consider the inductive step (2.36b). Using the respective definitions (4.10) and (4.1) for P^* and Λ ,

we calculate

$$\begin{aligned}
 &P^*(x_{n-1}^*, x_n^*) \Lambda(x_n^*, x_n) \\
 &= \frac{\pi(x_n)}{H(x_{n-1}^*)} [P_{x_n^*}\{\tilde{X}_1 \leq x_{n-1}^*\} - P_{x_n^*+1}\{\tilde{X}_1 \leq x_{n-1}^*\}] \\
 &\quad \times I_{\{0, \dots, x_n^*\}}(x_n)
 \end{aligned}$$

for the numerator in (2.36b) and hence

$$\Delta(x_{n-1}^*, x_n) = \frac{\pi(x_n)}{H(x_{n-1}^*)} P_{x_n}\{\tilde{X}_1 \leq x_{n-1}^*\}$$

for the denominator. Thus the construction sets $X_n^* = x_n^*$ with probability

$$\begin{aligned}
 &P\{X_n^* = x_n^* | X_{n-1}^* = x_{n-1}^*, X_n = x_n\} \\
 (4.13) \quad &= \frac{[P_{x_n^*}\{\tilde{X}_1 \leq x_{n-1}^*\} - P_{x_n^*+1}\{\tilde{X}_1 \leq x_{n-1}^*\}]}{P_{x_n}\{\tilde{X}_1 \leq x_{n-1}^*\}} I_{\{0, \dots, x_n^*\}}(x_n).
 \end{aligned}$$

Notice that the cumulative H appears nowhere in (4.13). This is important because in many examples the only quantity difficult to compute is H .

A similar calculation shows that

$$\begin{aligned}
 &P\{X_0^* = x_0^* | X_0 = x_0\} \\
 (4.14) \quad &= \frac{[\pi_0(x_0^*)/\pi(x_0^*) - \pi_0(x_0^* + 1)/\pi(x_0^* + 1)] I_{\{0, \dots, x_0^*\}}(x_0)}{\pi_0(x_0)/\pi(x_0)},
 \end{aligned}$$

which is again free of H .

REMARK 4.15. According to Remark 2.39 the dual X^* constructed by (4.13)–(4.14) is sharp. Indeed, the only dual state as large as $y = d \in \mathcal{S}$ is $y^* = d$ [which, since $\Lambda(x^*, \cdot) = \pi$ if and only if $x^* = d$, plays the role of ∞ in the general definition of a strong stationary dual].

4.3. *Birth and death chains.* Let X be an irreducible birth and death chain on $\mathcal{S} = \{0, \dots, d\}$ with initial distribution π_0 and transition matrix P . Write q_x for $P(x, x - 1)$, r_x for $P(x, x)$ and p_x for $P(x, x + 1)$. Irreducibility is the requirement that $q_x > 0$ for $x > 0$ and $p_x > 0$ for $x < d$. By natural convention $q_0 = 0 = p_d$. The stationary distribution is given by

$$\pi(x) = c \prod_{y=1}^x \frac{p_{y-1}}{q_y},$$

with $c = \pi(0)$ a normalizing constant.

As is well known, X is time reversible; that is, $\tilde{P} = P$. Thus if P is aperiodic, then (π_0, P) has an \mathcal{S} -valued dual (π_0^*, P^*) as in Theorem 4.6 if and

only if

$$(4.16a) \quad \frac{\pi_0(x)}{\pi(x)} \text{ decreases in } x, \text{ i.e., } q_{x+1}\pi_0(x+1) \leq p_x\pi_0(x) \text{ for } x < d;$$

$$(4.16b) \quad P \text{ is monotone, i.e., } p_x + q_{x+1} \leq 1, \quad x < d.$$

The equivalence in (4.16b) has been pointed out by Cox and Rösler (1983) and is immediate from the definition. Observe that if (4.16b) holds, then P is automatically aperiodic; e.g., $r_0 = 1 - p_0 > 1 - (p_0 + q_1) \geq 0$.

Suppose that (4.16) holds. Then Theorem 4.6 and the sample path construction (4.13)–(4.14) yield a strong stationary dual X^* that is also a birth and death chain on $\{0, \dots, d\}$. It has initial distribution

$$(4.17) \quad \pi_0^*(x^*) = \begin{cases} \frac{H(x^*)}{\pi(x^*)p_{x^*}} [p_{x^*}\pi_0(x^*) - q_{x^*+1}\pi_0(x^* + 1)], & x^* < d, \\ \frac{\pi_0(d)}{\pi(d)}, & x^* = d \end{cases}$$

and transition parameters

$$(4.18) \quad \begin{aligned} q_{x^*}^* &= \frac{H(x^* - 1)}{H(x^*)} p_{x^*}, & r_{x^*}^* &= 1 - (p_{x^*} + q_{x^*+1}), \\ p_{x^*}^* &= \frac{H(x^* + 1)}{H(x^*)} q_{x^*+1}, & x^* &\in \mathcal{S}. \end{aligned}$$

Here H is the cumulative of the stationary distribution π . For a reversible process, the sample path construction sets $X_n^* = x_n^*$ with probability

$$(4.19) \quad \alpha [P_{x_n^*}\{X_1 \leq x_{n-1}^*\} - P_{x_{n+1}^*}\{X_1 \leq x_{n-1}^*\}] I_{\{0, \dots, x_n^*\}}(x_n)$$

given $X_{n-1}^* = x_{n-1}^*$ and $X_n = x_n$; α is a normalizing constant. In the present birth and death context the right side simplifies. If $x_{n-1}^* \geq x_n + 1$, then (4.19) is $p_{x_{n-1}^*}$, $1 - (p_{x_{n-1}^*} + q_{x_{n-1}^*+1})$, $q_{x_{n-1}^*+1}$ or 0, according as $x_n^* = x_{n-1}^* - 1$, $x_n^* = x_{n-1}^*$, $x_n^* = x_{n-1}^* + 1$ or $|x_n^* - x_{n-1}^*| > 1$. If $x_{n-1}^* = x_n$, then (4.19) is $[1 - (p_{x_{n-1}^*} + q_{x_{n-1}^*+1})]/[1 - p_{x_{n-1}^*}]$, $q_{x_{n-1}^*+1}/[1 - p_{x_{n-1}^*}]$ or 0, according as $x_n^* = x_{n-1}^*$, $x_n^* = x_{n-1}^* + 1$ or neither. If $x_{n-1}^* = x_n - 1$, then (4.19) is 1 or 0, according as $x_n^* = x_{n-1}^* + 1$ or not.

For the remainder of Section 4.3 suppose for simplicity that $\pi_0 = \delta_0$, so that also $\pi_0^* = \delta_0$. By Theorem 2.4(a), the first time T_d^* that the dual chain X^* hits d is a strong stationary time for X . In fact, since X^* is sharp (Remark 4.15), T_d^* is a time to stationarity. In continuous time Keilson (1979) shows that the hitting time of state d for a birth and death chain on $\{0, \dots, d\}$ can be represented as a sum of independent exponential random variables with parameters related to the eigenvalues of the chain. An analogous result holds in discrete time and will lead to a stochastic interpretation of the eigenvalues.

THEOREM 4.20. *Let X be an irreducible monotone birth and death chain on $\mathcal{S} = \{0, \dots, d\}$ with transition matrix P , started in state 0. Then P has $d + 1$ distinct eigenvalues $1, \theta_1, \dots, \theta_d$, with $-1 < \theta_j < 1$ for $j = 1, \dots, d$. The time to stationarity has probability generating function*

$$(4.21) \quad u \rightarrow \prod_{j=1}^d \left[\frac{(1 - \theta_j)u}{1 - \theta_j u} \right].$$

REMARK 4.22. (a) If $\theta_1, \dots, \theta_d$ are all nonnegative, then the time T_d^* to stationarity is distributed as the sum of independent geometric random variables with success probabilities $1 - \theta_1, \dots, 1 - \theta_d$. These geometrics will tend to be large if the θ_j 's are close to 1. Since

$$s(n) = P\{T_d^* > n\},$$

Theorem 4.20 establishes a nonasymptotic stochastic relationship between approach to stationarity and spectral gap.

(b) It need not be that every θ_j is nonnegative: see Example 4.38 [random walk on the d -cube, with holding probability $r \in [1/(d + 1), 1/2]$] for a counterexample.

(c) Suppose some of the eigenvalues, say $\theta_{l+1}, \dots, \theta_d$, are negative. Then Theorem 4.20 yields the curious distributional identity

$$(4.23) \quad T_d^* - \sum_{j=l+1}^d B_j =_{\mathcal{L}} \sum_{j=1}^l G_j,$$

wherein on the left the time to stationarity T_d^* and B_{l+1}, \dots, B_d are independent random variables with $B_j \sim \text{Bernoulli}(1/(1 + |\theta_j|))$ and on the right G_1, \dots, G_l are independent geometric $(1 - \theta_j)$ variables as before.

(d) In any case, the time to stationarity has mean $\sum_{j=1}^d (1 - \theta_j)^{-1}$ and variance $\sum_{j=1}^d \theta_j (1 - \theta_j)^{-2} = \sum_{j=1}^d (1 - \theta_j)^{-2} - \sum_{j=1}^d (1 - \theta_j)^{-1}$.

(c) To make use of Theorem 4.20, one needs to compute or at least bound the eigenvalues. Of course, with these at hand, techniques like Fourier analysis can also be used to bound total variation or separation.

REMARK 4.24. Arguing as at (1.2) of Brown and Shao (1987), the tails of T_d^* have the explicit form

$$(4.25) \quad s(n) = P\{T_d^* > n\} = \sum_{j=1}^d \left(\prod_{k \neq j} \frac{1 - \theta_k}{\theta_j - \theta_k} \right) \theta_j^n.$$

Because of monotonicity, the separation is achieved at the state d . The right side of (4.25) is thus $1 - P\{X_n = d\}/\pi(d)$.

PROOF OF THEOREM 4.20. Let $P_{(d)}^*$ be the $d \times d$ matrix obtained by deleting the final row and column from P^* . Arguing as in Brown and Shao (1987), the eigenvalues $\theta_1, \dots, \theta_d$ of $P_{(d)}^*$ satisfy $|\theta_j| < 1$ for every j , and the

first time T_d^* that X^* hits d has generating function $\prod_{j=1}^d [(1 - \theta_j)u / (1 - \theta_j u)]$.

Because d is an absorbing state for X^* , it is a simple matter to relate the eigenstructures of P^* and $P_{(d)}^*$. In particular, the eigenvalues of P^* are precisely $1, \theta_1, \dots, \theta_d$. Finally, because $\Lambda P = P^* \Lambda$ and Λ is invertible (it is lower triangular with strictly positive diagonal entries), P^* and P are similar matrices and so have the same eigenvalues. $P - I$ is the generator matrix of a continuous-time birth and death chain, and its eigenvalues are $0, \theta_1 - 1, \dots, \theta_d - 1$. From Keilson [(1979), Sections 3.2 and 5.1], these $d + 1$ numbers are real and distinct.

REMARK 4.26. We consider early stopping to obtain bounds on total variation that improve $\|\pi_n - \pi\| \leq s(n)$.

(a) In general, if in Theorem 2.55 we take Λ to be the link of truncated stationary distributions and A^* to be $\{y^* \in \mathcal{S}^*: y^* \supset x^*\}$ for a fixed $x^* \in \mathcal{S}^*$, then v of (2.57) equals $\|\Lambda(x^*, \cdot) - \pi\| = \sum_{x \notin x^*} \pi(x)$.

(b) Specializing (a) for X an MLR chain on $\mathcal{S} = \{0, \dots, d\}$, we have

$$(4.27) \quad \|\pi_n - \pi\| \leq (1 - H(x^*)) + H(x^*)P\{T > n\},$$

where T is the hitting time of $\{x^*, x^* + 1, \dots, d\}$ for X^* . If X is a birth and death chain and (say) $\pi_0 = \delta_0$, then T is the hitting time of state x^* for X^* .

4.4. *Examples.* In this section we present the detailed analysis of four birth and death chains.

EXAMPLE 4.28. *Two-state Markov chains.* Let $d = 1$ and introduce the abbreviations

$$(4.29) \quad p = p_0 > 0, \quad q = q_1 > 0.$$

Here

$$(4.30) \quad \pi(0) = \frac{q}{p + q}, \quad \pi(1) = \frac{p}{p + q}.$$

According to (4.16)–(4.18), our dual construction will apply precisely when

$$(4.31a) \quad \pi_0 \leq \pi \text{ stochastically } \left(\text{i.e., } \pi_0(0) > \pi(0) = \frac{q}{p + q} \right)$$

and

$$(4.31b) \quad p + q \leq 1,$$

in which case X^* is a two-state Markov chain with

$$(4.32) \quad \pi_0^*(1) = 1 - \pi_0^*(0) = \frac{\pi_0(1)}{\pi(1)} = \frac{p + q}{p} \pi_0(1)$$

and

$$(4.33) \quad p^* = p + q, \quad q^* = 0.$$

It is particularly simple to describe the sample path construction of X^* from X for this example. If $X_0 = 1$, then (with certainty) set $X_0^* = 1$. If $X_0 = 0$, then set $X_0^* = 1$ with probability $(\pi_0(1)/\pi_0(0))/(\pi(1)/\pi(0)) = (q/p) \times (\pi_0(1)/\pi_0(0))$ and $X_0^* = 0$ with the complementary probability. Let $n \geq 1$. If $X_{n-1}^* = 1$, or if $X_{n-1}^* = 0$ and $X_n = 1$, then (with certainty) set $X_n^* = 1$. If $X_{n-1}^* = 0$ and $X_n = 0$, then set $X_n^* = 0$ with probability $q/(1-p)$ and $X_n^* = 1$ with the complementary probability.

From (4.32)–(4.33) we see that the time T_1^* to stationarity for X has probability mass function

$$(4.34) \quad P\{T_1^* = 0\} = \pi_0^*(1) = \frac{\pi_0(1)}{\pi(1)} = \frac{p+q}{p}\pi_0(1),$$

$$P\{T_1^* = n\} = \pi_0^*(0)[1 - (p+q)]^{n-1}(p+q), \quad n \geq 1.$$

That is, $\mathcal{L}(T_1^*)$ is the $(\pi_0^*(0), \pi_0^*(1))$ mixture of the geometric $(p+q)$ distribution on $\{1, 2, \dots\}$ and point mass at 0. Thus

$$(4.35) \quad s(n) = P\{T_1^* > n\} = \pi_0^*(0)[1 - (p+q)]^n = \pi_n^*(0), \quad n \geq 0.$$

When $\pi_0 = \delta_0$, these results agree with Theorem 4.20 and Remark 4.22(a). For comparison, the total variation distance is

$$(4.36) \quad \|\pi_n - \pi\| = \frac{p}{p+q}s(n), \quad n \geq 0.$$

REMARK 4.37. As the preceding example demonstrates, the mapping $P \rightarrow P^*$ is many-to-one: Every two-state P with $p > 0$, $q > 0$ and the same value of $p+q \leq 1$ has by (4.33) the same dual transitions P^* .

EXAMPLE 4.38. *Simple symmetric random walk on d -dimensional cube, revisited: Ehrenfest chain.* Example 3.2 presented an analysis of simple symmetric random walk on the d -cube \mathbb{Z}_2^d . However, holding probability of $1/2$ was used in a rather special way and the construction of a dual required the clever coordinate checking argument of Broder. We return to this example here, showing how the dual construction of Section 4.3 effortlessly yields an analyzable time to stationarity whenever the holding probability at each vertex is a constant $r \in [1/(d+1), 1)$.

To draw the connection with birth and death chains, let $Y_n = (Y_{n1}, \dots, Y_{nd}) \in \mathbb{Z}_2^d$ record the position of the walk at time $n \geq 0$. Consider the number

$$X_n = |Y_n| = \sum_{j=1}^d Y_{nj}$$

of ones among Y_{n1}, \dots, Y_{nd} . Then $(X_n)_{n \geq 0}$, an aperiodic version of the Ehrenfest chain, is a birth and death chain on $\mathcal{S} = \{0, \dots, d\}$ with transition

parameters

$$(4.39) \quad q_x = (1 - r) \frac{x}{d}, \quad r_x = r, \quad p_x = (1 - r) \frac{d - x}{d}$$

for $x \in \mathcal{S}$. Observe that $p_x + q_{x+1} = (1 - r)(d + 1)/d \leq 1$ by the assumption that $r \geq 1/(d + 1)$; thus X is irreducible (since $r < 1$), aperiodic and monotone. The stationary distribution π is binomial $(d, \frac{1}{2})$:

$$(4.40) \quad \pi(x) = 2^{-d} \binom{d}{x}, \quad x \in \mathcal{S}.$$

If X is started in a distribution π_0 satisfying (4.16a), then Section 4.3 shows how to build a strong stationary dual X^* . Moreover, if Y_0 is conditionally uniform given X_0 , i.e., if for each $y = (y_1, \dots, y_d) \in \mathbb{Z}_2^d$,

$$(4.41) \quad P\{Y_0 = y\} = \pi_0(|y|) \binom{d}{|y|},$$

($|y| = \sum_{j=1}^d y_j$), then it is easy to see that X^* is a set-valued dual for Y with respect to the link

$$(4.42) \quad \Lambda(x^*, y) = I_{x^*}(y) / \#(x^*), \quad x^* \subset \mathbb{Z}_2^d, \quad y \in \mathbb{Z}_2^d,$$

of truncated stationary distributions. Here $\#(x^*)$ is the size of the set x^* and we have identified $x^* \in \{0, \dots, d\}$ with $\{x \in \mathbb{Z}_2^d: |x| \leq x^*\}$.

For simplicity suppose for the remainder of this example that $\pi_0 = \delta_0$, that is, that the random walk Y begins at the vertex $(0, \dots, 0)$. Theorem 4.20 then applies to X . Following Kac (1947) we find for the *unmodified* Ehrenfest chain with transition probabilities $q_x = x/d, r_x = 0, p_x = (d - x)/d, x \in \mathcal{S}$, that the eigenvalues are $1 - 2j/d, j = 0, \dots, d$. Hence our P has the $d + 1$ distinct eigenvalues $\theta_j = r + (1 - r)(1 - 2j/d) = 1 - (1 - r)2j/d, j = 0, \dots, d$. According to Theorem 4.20, the time to stationarity has generating function $\prod_{j=1}^d [(1 - \theta_j)u / (1 - \theta_j)]$. From (4.25),

$$(4.43) \quad s(n) = \sum_{j=1}^d (-1)^{j-1} \binom{d}{j} \theta_j^n, \quad n \geq 0.$$

Standard asymptotics shows that it takes $\frac{1}{2}(1 - r)^{-1}d \log d + cd$ steps with c large to make separation small. It takes half as long to make total variation small.

It is worthwhile to consider two special values of the holding probability r . First suppose $r = \frac{1}{2}$, as in our earlier treatment of the walk (Example 3.2). Then the eigenvalues $\theta_j = 1 - j/d, j = 0, \dots, d$, are all nonnegative, and so by Remark 4.22(a) the time to stationarity is distributed as the independent sum of geometric (j/d) random variables, $j = 1, \dots, d$. Example 3.2 gave a direct interpretation to each geometric term: The time required to advance from j to $j - 1$ unchecked coordinates is geometric (j/d) .

Now suppose $r = 1/(d + 1)$, so that at each step the walk moves at random to a vertex within distance 1 of the current vertex. Then the eigenvalues are $\theta_j = 1 - 2j/(d + 1), j = 0, \dots, d$, about half of which are negative. However,

rather than employ Remark 4.22(c), we note that the eigenvalues occur in \pm pairs. For definiteness, suppose d is even. Then the time to stationarity has generating function $\prod_{j=1}^{d/2} [(1 - \theta_j^2)u^2 / (1 - \theta_j^2 u^2)]$ and so is distributed as twice the sum of independent geometric $(1 - \theta_j^2 = 4 \times j / (d + 1) \times (1 - j / (d + 1)))$ random variables, $j = 1, \dots, d/2$. The exact expression (4.43) for the separation simplifies in this case, too. For example, if n is odd, then

$$(4.44) \quad s(n) = \sum_{j=1}^{d/2} (-1)^{j-1} \binom{d+1}{j} \left(1 - \frac{2j}{d+1}\right)^n, \quad n \geq 0.$$

REMARK 4.45. A very similar analysis can be carried out for nearest neighbor random walk on any distance regular graph. These graphs include the Bernoulli-Laplace model of diffusion discussed by Diaconis and Shahshahani (1987), who give further references. For these graphs, all of the eigenvalues are available in closed form. Bannai and Itô (1987) give references to extensive lists of distance regular graphs and the active attempts at the classification of all such graphs.

EXAMPLE 4.46. *Simple random walk on $\{0, \dots, d\}$.* Consider random walk X on $\mathcal{S} = \{0, \dots, d\}$ with probability $0 < \lambda < 1$ of moving right and $\mu = 1 - \lambda$ of moving left from each interior state and with complete reflection at the endpoints. In order to avoid periodicity we shall put holding probability $[1 + (\max(\lambda, \mu))^{-1}]^{-1} \leq r < 1$ at each point in \mathcal{S} leading to the parameters

$$(4.47a) \quad q_x = (1 - r)\mu, \quad r_x = r, \quad p_x = (1 - r)\lambda, \quad 0 < x < d;$$

$$(4.47b) \quad q_0 = 0, \quad r_0 = r, \quad p_0 = 1 - r;$$

$$(4.47c) \quad q_d = 1 - r, \quad r_d = r, \quad p_d = 0.$$

Observe that $p_x + q_{x+1} = 1 - r \leq 1$, $0 < x < d$, and $p_0 + q_1 = (1 - r)(1 + \mu) \leq 1$ and $p_{d-1} + q_d = (1 - r)(1 + \lambda) \leq 1$. Thus X is irreducible, aperiodic and monotone. The stationary distribution is

$$(4.48) \quad \pi(x) = \begin{cases} \gamma, & \text{if } x = 0, \\ \gamma \times \frac{1}{\mu} \left(\frac{\lambda}{\mu}\right)^{x-1}, & \text{if } 0 < x < d, \\ \gamma \times \left(\frac{\lambda}{\mu}\right)^{d-1}, & \text{if } x = d, \end{cases}$$

where the normalizing constant

$$(4.49) \quad \gamma = \begin{cases} \frac{1}{2} \left(\frac{\lambda}{\mu} - 1\right) / \left[\left(\frac{\lambda}{\mu}\right)^d - 1\right], & \text{if } \lambda \neq \frac{1}{2}, \\ \frac{1}{2d}, & \text{if } \lambda = \frac{1}{2}. \end{cases}$$

Arguing as in Chapter 10 of Karlin and Taylor (1981) we find that the eigenvalues of the transition matrix are the $d + 1$ distinct values

$$(4.50) \quad \begin{aligned} 1; \quad \theta_j &= r + (1 - r)2(\lambda\mu)^{1/2} \cos(j\pi/d), \quad j = 1, \dots, d - 1; \\ \theta_d &= -1 + 2r. \end{aligned}$$

Suppose from now on that the walk starts in state 0. The time to stationarity then has mean

$$(4.51) \quad ET_d^* = (1 - r)^{-1} \left\{ \sum_{j=1}^{d-1} [1 - 2(\lambda\mu)^{1/2} \cos(j\pi/d)]^{-1} + \frac{1}{2} \right\}$$

and variance

$$(4.52) \quad \text{Var } T_d^* = (1 - r)^{-2} \left\{ \sum_{j=1}^{d-1} [1 - 2(\lambda\mu)^{1/2} \cos(j\pi/d)]^{-2} + \frac{1}{4} \right\} - ET_d^*.$$

(a) Consider first the asymmetric case $\lambda \neq \frac{1}{2}$. Then as $d \rightarrow \infty$,

$$(4.53) \quad ET_d^* = \alpha_1 d + O(1),$$

$$(4.54) \quad \text{Var } T_d^* = \alpha_2^2 d + O(1),$$

where

$$(4.55) \quad \begin{aligned} \alpha_1 &= (1 - r)^{-1} \int_0^1 [1 - 2(\lambda\mu)^{1/2} \cos(\pi u)]^{-1} du \\ &= (1 - r)^{-1} \mu^{-1} \left| 1 - \frac{\lambda}{\mu} \right|^{-1} \end{aligned}$$

and

$$(4.56) \quad \alpha_2^2 = (1 - r)^{-2} \int_0^1 [1 - 2(\lambda\mu)^{1/2} \cos(\pi u)]^{-2} du - \alpha_1 > 0$$

can also be calculated explicitly. Moreover, one can show that T_d^* is asymptotically normal. Thus for any real c , as $d \rightarrow \infty$ the separation after $\alpha_1 d + c\alpha_2 d^{1/2}$ steps converges to $P\{Z > c\}$, where Z is standard normal.

If $\lambda > \frac{1}{2}$, then from the inequality

$$(4.57) \quad s(n) \geq \|\pi_n - \pi\| \geq |\pi_n(d) - \pi(d)| = \pi(d)s(n),$$

it follows that $\alpha_1 d + c\alpha_2 d^{1/2}$ steps, with c large, are required to make total variation small. If $\lambda < \frac{1}{2}$, then from the early stopping bound (4.27) with $x^* \rightarrow \infty$ as $d \rightarrow \infty$ it follows that c steps, with c large, are enough to make total variation small.

Results asymptotic in λ are also available. For example, if d is held fixed and $\mu \rightarrow 0$ and we assume for definiteness that $r \geq \frac{1}{2}$, then the time to stationarity converges in distribution to the independent sum of negative binomial $(d - 1, 1 - r)$ and geometric $(2(1 - r))$ random variables. This result

is easy to interpret. If we set $\mu = 0$ in our definition of the random walk, then the chain moves from 0 to $d - 1$ in negative binomial $(d - 1, 1 - r)$ steps and thereafter behaves like a two-state chain with $p = q = 1 - r$.

(b) Now consider the symmetric case $\lambda = \frac{1}{2}$, for which $\frac{1}{3} \leq r < 1$ is required. In this case the separation has the simple form

$$(4.58) \quad s(n) = 2 \sum_{j=0}^d (-1)^{j-1} \cos\left(\frac{j\pi}{2d}\right) \left[r + (1 - r) \cos\left(\frac{j\pi}{d}\right) \right]^n.$$

It takes cd^2 steps with c large to make separation small.

There is a simple connection between the present symmetric random walk on $\{0, \dots, d\}$ and symmetric random walk Y on the discrete circle of $2d$ points (cf. Examples 1.1 and 3.1). Indeed, if the holding probability at each point on the circle is r , let

$$X_n = \min(Y_n, d - Y_n)$$

record the distance of the circle-walk from its initial position of 0. Then X is our segment-random walk. Moreover, X^* is (with the obvious identifications) a set-valued dual for Y with respect to the link of uniform distributions on the respective sets $\{0\}$, $\{d\}$ and $\{j, d + j\}$, $j = 1, \dots, d - 1$. Thus both separation and total variation are the same for the two walks. This has two interesting consequences.

First, one can argue by use of the central limit theorem that, as $d \rightarrow \infty$, the total variation after cd^2 steps converges to the total variation distance between the fractional part of a normal $(0, (1 - r)^2c)$ random variable and a uniform $(0, 1)$ random variable. Second, suppose $r = \frac{1}{2}$ and $d = 2^{a-1}$. Then by the analysis of Examples 1.1 and 3.1, the time to stationarity has mean equal to $2(\sum_{j=1}^{a-1} \mu_j + 1)$, where μ_j is the mean duration of the game for a gambler's ruin chain on $\{0, \dots, (2d)/2^j\}$ started at the middle point $(2d)/2^{j+1}$. As is well known, $\mu_j = ((2d)/2^{j+1})^2$. Hence $ET_d^* = (2d^2 + 1)/3$ and (4.51) gives the curious identity

$$(4.59) \quad \sum_{j=1}^d [1 - \cos(j\pi/d)]^{-1} = (2d^2 + 1)/6,$$

where d is a power of 2.

EXAMPLE 4.60. Random walk on a finite tree. Consider a rooted m -ary tree of depth d . For example, Figure 3 is a rooted 4-ary tree of depth 2. Suppose we start a random walk Y at the root. Thereafter, with probability r the walk remains at its present state and with probability $1 - r$ one of the nearest neighbors of the present node is chosen at random as the next node. Let X_n denote the level of Y_n , where the root is on level 0, its offspring are on level 1, ... and the leaves are on level d . Then X is a simple random walk on $\{0, \dots, d\}$ with transition probabilities as in Example 4.46 [recall (4.47)] with $\lambda = m/(m + 1)$. If $m = 1$, then $Y = X$ was treated in Example 4.46(b); otherwise, as we henceforth assume, Example 4.46(a) applies. The stationary

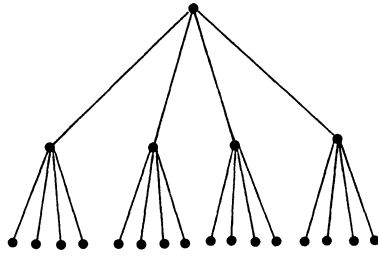


FIG. 3.

distribution for Y puts mass γ at the root, $\gamma(1 + 1/m)$ at each internal node and γ/m at each leaf, where $\gamma = \frac{1}{2}(m - 1)/(m^d - 1)$. With m fixed, it takes $\alpha_1 d + c\alpha_2 d^{1/2}$ steps, with c large, to make either separation or total variation small, where α_1 and α_2 are given by (4.55) and (4.56), respectively, with $\lambda = m/(m + 1) = 1 - \mu$. The result of Example 4.46 for d fixed and $\mu \rightarrow 0$ likewise applies to the case $m \rightarrow \infty$ for Y .

Now suppose instead that Y is started at a random leaf. Again let X_n denote the level of Y_n , but now assign level 0 to the leaves, 1 to their parents, ... and d to the root. Then X is as in Example 4.46 with $\lambda = 1/(m + 1)$. The time to stationarity has the same distribution as in the root-started walk. However, as $d \rightarrow \infty$ it takes only c steps, with c large, to make total variation small.

REMARK 4.61. The argument works for variations of the tree with the same conclusion. First, the number of offspring of the root can be changed to any finite number. Second, there can be an arbitrary (constant) number of terminal leaves at depth d . The distribution of the level-walk, and hence the separation and total variation for the tree-walk, remain unchanged. In particular, the analyses of Example 4.60 apply without change to the symmetric tree. For example, the symmetric 4-ary tree of depth 2 can be drawn as shown in Figure 4.

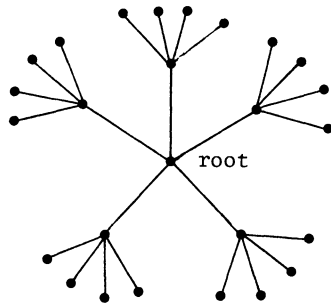


FIG. 4.

Passing to the limit in this example as the number of levels tends to ∞ we recapture the infinite d -ary tree of Sawyer (1978). We hope to explore the relation between the eigenanalysis above and the spherical functions of Sawyer's paper.

5. Duality.

5.1. *Strong stationary duality and Siegmund duality.* The purpose of Section 5 is to relate the strong stationary duality of Section 4.2 to other notions of duality. General work on duality began with the paper of Siegmund (1976). To describe this work in the context of the present paper, let Y and Z be Markov chains on the state space $\{0, 1, 2, \dots\}$. Call Z the dual of Y if

$$(5.1) \quad P_y\{Y_n \leq z\} = P_z\{y \leq Z_n\} \text{ for all } y \text{ and } z \text{ and } n \geq 0.$$

Siegmund noted that any random walk reflecting at 0 is the dual of the negative of the same walk absorbed at 0. This relation had allowed Lindley (1952) to transform solutions to problems for the absorbing walk into solutions to problems for the reflecting walk. Lévy (1948) had used the analogue for Brownian motion to prove theorems about reflecting Brownian motion.

The term "dual" was perhaps first employed in this context by Karlin and McGregor [(1957), Section 6], who discussed the notion for birth and death processes. Siegmund's work shows that Y has a dual Z if and only if Y has an absorbing barrier at 0 and for $n = 1$, $P_y\{Y_n \leq z\}$ is decreasing in y for each fixed z , in which case the same is true for every n . This condition on Y has been termed (*stochastic*) *monotonicity* by Daley (1968). Siegmund gave necessary and sufficient conditions for the existence of a dual in a much broader context and exhibited several pairs of processes in duality. Van Doorn (1980), Cox and Rösler (1983) and Clifford and Sudbury (1985) contain further development.

Our aim is to show how Siegmund duality can facilitate the construction of a strong stationary dual. To set the stage, let $X \sim (\pi_0, P)$ be an ergodic Markov chain on $\mathcal{S} = \{0, \dots, d\}$ with stationary distribution π . Let $\check{P}(x, y) = \pi(y)P(y, x)/\pi(x)$ denote the time reversal of P and write \check{X} for any chain with transition matrix \check{P} .

The construction of a Siegmund dual to \check{X} clearly requires that \check{X} have an absorbing barrier at its smallest state. Accordingly, append an absorbing state -1 to the \check{X} state space and define $\check{P}(x, -1) = 0$ for $x = 0, \dots, d$. Then a Siegmund dual, call it $\check{\check{X}}$, exists if and only if \check{X} (original or extended) is monotone, in which case the transition function for $\check{\check{X}}$ is given by

$$(5.2) \quad \check{\check{P}}(x, y) = P_y\{\check{X}_1 \leq x\} - P_{y+1}\{\check{X}_1 \leq x\}, \quad x, y \in \{-1, 0, \dots, d\}.$$

Note that -1 is an absorbing state for $\check{\check{P}}$ but that $\check{\check{P}}(x, -1)$ is strictly positive for some values of $x \geq 0$. Note too that d is an absorbing state for $\check{\check{P}}$.

In Section 4.2 (cf. Theorem 4.6) it was shown that P^* satisfying $\Lambda P = P^* \Lambda$ exists if and only if \check{P} is monotone, in which case

$$(5.3) \quad P^*(x^*, y^*) = \frac{H(y^*)}{H(x^*)} \check{P}(x^*, y^*), \quad x^*, y^* \in \mathcal{S},$$

where H is the cumulative of π . Observe that

$$(5.4) \quad H \text{ is a harmonic function for } \check{P} \text{ on } \{-1, 0, \dots, d\},$$

i.e.,

$$\begin{aligned} \sum_{y^*=-1}^d \check{P}(x^*, y^*) H(y^*) &= \sum_{y^*=0}^d \check{P}(x^*, y^*) H(y^*) \\ &= \begin{cases} H(x^*) \sum_{y^*=0}^d P^*(x^*, y^*), & \text{if } x^* \in \{0, \dots, d\}, \\ 0, & \text{if } x^* = -1, \end{cases} \\ &= H(x^*). \end{aligned}$$

Recall that if Q is a transition function and h is harmonic for Q , i.e., h is nonnegative and $Qh = h$, then $Q_h(x, y) = Q(x, y)h(y)/h(x)$ defines a transition function on $\{x: h(x) > 0\}$. Q_h is called the *Doob h transform* of Q . See, e.g., Kemeny, Snell and Knapp (1976). Thus P^* is the Doob H transform of the Siegmund dual of the time reversal of P . We record these results formally.

THEOREM 5.5. *Let X be an irreducible aperiodic Markov chain on $\mathcal{S} = \{0, \dots, d\}$ with transition function P , initial distribution π_0 and stationary distribution π . Let H be the cumulative of π . Let $\check{P}(x, y) = \pi(y)P(y, x)/\pi(x)$ be the time reversal of P . Suppose that $\pi_0(x)/\pi(x)$ decreases in x and \check{P} is monotone, so that a dual P^* for P exists as in Theorem 4.6. Then P^* can be computed from P in three steps:*

- (i) Calculate the time reversal \check{P} of P .
- (ii) Calculate the Siegmund dual $\check{\check{P}}$ of \check{P} .
- (iii) Calculate the Doob H transform $P^* = (\check{\check{P}})_H$ of $\check{\check{P}}$.

Using results from Cox and Rösler (1983) and similar results, one can argue that the entrance law $(\pi_n)_{n \geq 0}$, with $\pi_n = \pi_0 P^n$, for P is in a certain natural sense transformed by the successive steps (1)–(3) of Theorem 5.5 first to the exit law $(\pi_n(\cdot)/\pi(\cdot))_{n \geq 0}$ for \check{P} , then to the entrance law $(\pi_n(\cdot)/\pi(\cdot) - \pi_n(\cdot + 1)/\pi(\cdot + 1))_{n \geq 0}$ for $\check{\check{P}}$ and finally to the entrance law $(\pi_n^*)_{n \geq 0}$ for P^* . We omit the details. (Our use of exit law is slightly nonstandard; we have negated the time index.) Likewise, the harmonicity of H for \check{P} is a natural reflection of the fact that π is stationary, i.e., a constant nonnegative entrance law, for \check{P} .

Turning to the sample path construction of a strong stationary dual X^* , we note that Siegmund duality helps us to interpret the probabilities (4.13)–(4.14).

Indeed, the inductive construction sets $X_n^* = x_n^*$ with probability

$$(5.6) \quad \begin{aligned} P\{X_n^* = x_n^* | X_{n-1}^* = x_{n-1}^*, X_n = x_n\} \\ = P\{\check{X}_n = x_n^* | \check{X}_{n-1} = x_{n-1}^*, \check{X}_n \geq x_n\}. \end{aligned}$$

At the start the construction sets $X_0^* = x_0^*$ with probability

$$(5.7) \quad P\{X_0^* = x_0^* | X_0 = x_0\} = P\{\check{X}_0 = x_0^* | \check{X}_0 \geq x_0\},$$

where on the right \check{X}_0 has the (unconditional) distribution

$$(5.8) \quad P\{\check{X}_0 = x_0^*\} = \frac{\pi(0)}{\pi_0(0)} \left[\frac{\pi_0(x_0^*)}{\pi(x_0^*)} - \frac{\pi_0(x_0^* + 1)}{\pi(x_0^* + 1)} \right], \quad x_0^* = 0, \dots, d,$$

with the convention $\pi_0(d + 1)/\pi(d + 1) := 0$. [Note $\pi_0(0) > 0$.]

To complete Section 5.1 we specialize to birth and death chains. Let P be the transition matrix of a monotone birth and death chain. Recall that P is reversible. Once (5.2) is used to obtain the Siegmund dual birth and death chain transition probabilities

$$(5.9) \quad \check{q}_x = p_x, \quad \check{r}_x = 1 - (p_x + q_{x+1}), \quad \check{p}_x = q_{x+1},$$

it is clear that the definition (4.18) of $q_{x^*}^*, r_{x^*}^*, p_{x^*}^*$ is just a special case of Theorem 5.5.

5.2. Strong stationary duality as duality with respect to a function. A definition of duality generalizing Siegmund's was used by Holley and Stroock (1979) in studying interacting particle systems. Accessible treatments are in Cox and Rösler [(1983), Section 1] and Liggett [(1985), Definition 2.3.1]. A related reference is Vervaat (1987).

DEFINITION 5.10. If Y and Z are Markov chains with countable state spaces \mathcal{S}_Y and \mathcal{S}_Z and f is a bounded nonnegative function on $\mathcal{S}_Y \times \mathcal{S}_Z$, then Z is said to be a *dual of Y with respect to f* if

$$(5.11) \quad E_y f(Y_n, z) = E_z f(y, Z_n), \quad y \in \mathcal{S}_Y, z \in \mathcal{S}_Z, n \geq 0.$$

In this generality, duality is a nearly symmetric notion: Z is a dual of Y with respect to f if and only if Y is a dual of Z with respect to the transpose

$$f^T(z, y) = f(y, z)$$

of f . Siegmund duality is recaptured by taking $f(y, z) = 1$ or 0 according as $y \leq z$ or $y > z$.

Adopt the setup of Theorem 5.5 and recall

$$\Lambda(x^*, x) = I_{\{0, \dots, x^*\}}(x) \pi(x) / H(x^*).$$

We present two reasons why the term dual deserves to be applied to P^* . The first reason is Theorem 5.12.

THEOREM 5.12. *The reversed chain \check{P} and the strong stationary dual P^* are dual with respect to the function*

$$(5.13) \quad f_H(y, z) := \begin{cases} 1/H(z), & \text{if } y \leq z, \\ 0, & \text{otherwise,} \end{cases}$$

on $\mathcal{S} \times \mathcal{S}$.

PROOF.

$$\begin{aligned} E_x f_H(\check{X}_n, x^*) &= \sum_y \check{P}^n(x, y) f_H(y, x^*) \\ &= \frac{1}{H(x^*)} \sum_{y \leq x^*} \check{P}^n(x, y) = \frac{1}{\pi(x)} \sum_y \Lambda(x^*, y) P^n(y, x) \\ &= \frac{1}{\pi(x)} \sum_{y^*} P^{*n}(x^*, y^*) \Lambda(y^*, x) \\ &= \sum_{y^* \geq x} \frac{P^{*n}(x^*, y^*)}{H(y^*)} = E_{x^*} f_H(x, X_n^*). \quad \square \end{aligned}$$

An alternative proof of Theorem 5.12 can be based on the observations $P_y\{\check{X}_n \leq x\} = P_x\{\check{X}_n \geq y\}$ and $P^{*n}(x^*, y^*) = (H(y^*)/H(x^*))\check{P}^n(x, y)$. Yet another proof can be built by modifying the following lemma to account for the fact that the harmonic function H for \check{P} vanishes at -1 . Lemma 5.14 describes how the Doob h transform affects duality.

LEMMA 5.14. *If Y and Z are Markov chains dual with respect to some function f and h is an everywhere strictly positive harmonic function for Z , then Y and the Doob h transform Z^h of Z are dual with respect to*

$$(5.15) \quad f_h(y, z) := f(y, z)/h(z).$$

PROOF. Let Z and Z^h have respective transition functions Q and Q_h . Then

$$\begin{aligned} E_y f_h(Y_n, z) &= \frac{1}{h(z)} E_y f(Y_n, z) \\ &= \frac{1}{h(z)} E_z f(y, Z_n) = \frac{1}{h(z)} \sum_{z'} Q^n(z, z') f(y, z') \\ &= \sum_{z'} Q_h^n(z, z') \frac{f(y, z')}{h(z')} = E_z f_h(y, Z_n^h). \quad \square \end{aligned}$$

Our other reason for calling P^* dual to P is explained by Theorem 5.19. First we extend the definition of duality (Definition 5.10). As we shall see, the extended definition unites the various notions of duality we have discussed.

DEFINITION 5.16. Let Q_Y and Q_Z be transition matrices on countable state spaces \mathcal{S}_Y and \mathcal{S}_Z . Let μ be a distribution on $\mathcal{S}_Y \times \mathcal{S}_Z$ and f a bounded nonnegative function on $\mathcal{S}_Y \times \mathcal{S}_Z$. Let Y and Z be Markov chains defined on a common probability space (Ω, \mathcal{F}, P) having transition functions Q_Y and Q_Z , respectively. We suppose that $(Y_0, Z_0) \sim \mu$, that the process Y and Z_0 are conditionally independent given Y_0 and that the process Z and Y_0 are conditionally independent given Z_0 . If

$$(5.17) \quad Ef(Y_n, Z_0) = Ef(Y_0, Z_n), \quad n \geq 0,$$

we say that Y and Z are *dual with respect to* (f, μ) .

REMARK 5.18. (a) Since the left side of (5.17) equals $\sum_{(y,z)} \mu(y, z) E_y f(Y_n, z)$ and the right side equals $\sum_{(y,z)} \mu(y, z) E_z f(y, Z_n)$, duality is in reality an assertion about transition functions.

(b) Chains Y and Z are dual in the sense of Definition 5.10 if and only if they are dual with respect to the given f and every point-mass distribution μ on $\mathcal{S}_Y \times \mathcal{S}_Z$, in which case they are also dual with respect to (f, μ) for every distribution μ .

(c) The more general Definition 5.16 makes time reversal a form of duality. Let

$$\mu(x, \tilde{x}) = \begin{cases} \pi(x), & \text{if } x = \tilde{x}, \\ 0, & \text{otherwise.} \end{cases}$$

Then Q is the time reversal \tilde{P} of P if and only if P and Q are dual with respect to every point-mass f and our specified μ , in which case P and $Q = \tilde{P}$ are dual with respect to (f, μ) for every bounded nonnegative f on $\mathcal{S} \times \mathcal{S}$.

THEOREM 5.19. *With assumptions as in Theorem 5.5, P^* and P are dual with respect to Λ and every μ of the form*

$$(5.20) \quad \mu(x^*, x) = \pi_0^*(x^*) \Lambda(x^*, x),$$

with π_0^* a distribution on \mathcal{S} [compare (2.20)].

PROOF. If μ is of the form (5.20), then

$$\begin{aligned} E\Lambda(X_n^*, X_0) &= \sum_{x^*} \sum_y \pi_0^*(x^*) \Lambda(x^*, y) \sum_{y^*} P^{*n}(x^*, y^*) \Lambda(y^*, y) \\ &= \sum_{x^*} \sum_y \pi_0^*(x^*) \Lambda(x^*, y) \sum_x \Lambda(x^*, x) P^n(x, y) \\ &= E\Lambda(X_0^*, X_n). \end{aligned} \quad \square$$

Acknowledgments. This material was presented as part of The 1987 Wald Memorial Lectures. We thank David Aldous for first suggesting that there must be a connection between strong uniform times and duality. Mark Brown has been a particularly valuable ally and Jim Pitman, David Siegmund and Mark Yor have all helped a great deal!

REFERENCES

- ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348.
- ALDOUS, D. and DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. in Appl. Math.* **8** 69–97.
- ATHREYA, K. B. and NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.
- BANNAI, E. and ITÔ, T. (1987). The study of distance-regular graphs from the algebraic (i.e., character theoretical) viewpoint. In *Proc. Symp. Pure Math.* (P. Fong, ed.) **47** (2) 343–350. Amer. Math. Soc., Providence, R.I.
- BROWN, M. (1975). The first passage time distribution for a parallel exponential system with repair. In *Reliability and Fault Tree Analysis* (R. E. Barlow, J. Fussell and N. Singpurwalla, eds.) SIAM, Philadelphia.
- BROWN, M. and SHAO, Y. S. (1987). Identifying coefficients in the spectral representation for first passage time distributions. *Probab. Eng. Inform. Sci.* **1** 69–74.
- CLIFFORD, P. and SUDBURY, A. (1985). A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.* **13** 558–565.
- COX, J. T. and RÖSLER, U. (1983). A duality relation for entrance and exit laws for Markov processes. *Stochastic Process. Appl.* **16** 141–156.
- DALEY, D. J. (1968). Stochastically monotone Markov chains. *Z. Wahrsch. Verw. Gebiete* **10** 305–317.
- DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, Calif.
- DIACONIS, P. and FILL, J. A. (1990). Examples for the theory of strong stationary duality with countable state spaces. *Probab. Eng. Inform. Sci.* **4** 157–180.
- DIACONIS, P., GRAHAM, R. L. and MORRISON, J. A. (1988). Asymptotic analysis of a random walk on a hypercube with many dimensions. Technical Report, AT & T Bell Labs, Murray Hill, N.J.
- DIACONIS, P. and SHAHSHAHANI, M. (1987). Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM J. Math. Anal.* **18** 208–218.
- DUBINS, L. (1968). On a theorem of Skorohod. *Ann. Math. Statist.* **39** 2094–2097.
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications*. 1, 3rd ed. Wiley, New York.
- FILL, J. A. (1990a). Time to stationarity for a continuous-time Markov chain. *Probab. Eng. Inform. Sci.* To appear.
- FILL, J. A. (1990b). Strong stationary duality for continuous-time Markov chains. Unpublished.
- FILL, J. A. (1990c). Strong stationary duality for diffusion. Unpublished.
- HOLLEY, R. and STROCK, D. (1979). Dual processes and their application to infinite interacting systems. *Adv. in Math.* **32** 149–174.
- KAC, M. (1947). Random walk and the theory of Brownian motion. *Amer. Math. Monthly* **54** 369–391.
- KARLIN, S. and MCGREGOR, J. (1957). The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86** 366–400.
- KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*, 2nd ed. Academic, New York.
- KARLIN, S. and TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic, New York.
- KEILSON, J. (1979). *Markov Chain Models—Rarity and Exponentiality*. Springer, New York.
- KEMENY, J., SNELL, L. and KNAPP, A. (1976). *Denumerable Markov Chains*. Springer, New York.
- LÉVY, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier–Villars, Paris.
- LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.
- LINDLEY, D. (1952). The theory of queues with a single server. *Proc. Cambridge Philos. Soc.* **48** 277–289.
- MATTHEWS, P. (1987). Mixing rates for a random walk on the cube. *SIAM J. Algebraic Discrete Methods* **8** 746–752.

- MATTHEWS, P. (1988). Some sample path properties of a random walk on the cube. *J. Theoret. Probab.* **2** 313–324.
- NUMMELIN, E. (1986). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge Univ. Press, Cambridge, Mass.
- SAWYER, S. (1978). Isotropic random walks in a tree. *Z. Wahrsch. Verw. Gebiete* **42** 279–292.
- SIEGMUND, D. (1976). The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4** 914–924.
- THORISSON, H. (1988). Future independent times and Markov chains. *Probab. Theory Related Fields* **78** 143–148.
- VAN DOORN, E. (1980). Stochastic monotonicity of birth–death processes. *Adv. in Appl. Probab.* **12** 59–80.
- VERVAAT, W. (1987). Algebraic duality of Markov processes. Technical Report, Catholic Univ., Nijmegen.
- YOR, M. (1988). Intertwinings of Bessel processes. Technical Report, Univ. California, Berkeley.

DEPARTMENT OF MATHEMATICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138

DEPARTMENT OF MATHEMATICAL STATISTICS
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218