ALEA, Lat. Am. J. Probab. Math. Stat. 21, 1407–1416 (2024)
DOI: 10.30757/ALEA.v21-53

Markov Chains and Mappings of Distributions on Compact Spaces

David J. Aldous and Shi Feng

Department of Statistics, 367 Evans Hall # 3860, U.C. Berkeley CA 94720
E-mail address: aldous@stat.berkeley.edu

Department of Mathematics, Cornell University
E-mail address: sf599@cornell.edu

Abstract. Consider a compact metric space S and a pair (j, k) with $k \geq 2$ and $1 \leq j \leq k$. For any probability distribution $\theta \in \mathcal{P}(S)$, define a Markov chain on S by: from state s, take k i.i.d. (θ) samples, and jump to the j'th closest. Such a chain converges in distribution to a unique stationary distribution, say $\pi_{j,k}(\theta)$. So this defines a mapping $\pi_{j,k} : \mathcal{P}(S) \to \mathcal{P}(S)$. What happens when we iterate this mapping? In particular, what are the fixed points of this mapping? We present a few rigorous results, to complement our extensive simulation study elsewhere.

1. Introduction

This article discusses a rather novel topic whose motivation may seem obscure, so we start with informal background that led to the formulation of the topic. Write $S = (S, d)$ for a compact metric space.. Then the identity function $f(s) := s$ makes sense for every S. Is there any more interesting explicit function $S \to S$ whose definition makes sense for every S? For example one might try $f(s) := \arg\max_y d(s, y)$, that is the most distant point from s; this works for any space S with the property that the most distant point is always unique, but not for all S. Our introspection suggests that in fact there is no non-trivial such "general" function.

Instead let us write $\mathcal{P}(S)$ for the space of probability distributions on S, and recall that $\mathcal{P}(S)$ is a compact metric space under the usual weak topology. The observation above suggests that there may be no non-trivial function $\mathcal{P}(S) \to \mathcal{P}(S)$ whose definition makes sense for every S. But this is false! This article investigates a particular family of such functions – the reader may care to try to invent different examples.

The Markov chain. Given S and $\theta \in \mathcal{P}(S)$, consider the following discrete time Markov chain on state space S: from point s make the step to the nearer of 2 random points drawn i.i.d. from θ, breaking ties uniformly at random. This scheme naturally generalizes as follows: fix $k \geq 2$ and $1 \leq j \leq k$, and step from s to the j'th nearest of k random points drawn i.i.d. from θ, again breaking ties uniformly at random. Write the associated chain as $\mathbf{X}^{\theta,j,k} = (X^{\theta,j,k}(t), t = 0, 1, 2, \ldots)$.

Received by the editors March 29th, 2024; accepted June 22nd, 2024.
2010 *Mathematics Subject Classification.* 60J05, 37A50.
Key words and phrases. Coupling, Markov chain, compact metric space, dynamical system.

Theorem 2.1 proves that this chain always has a unique stationary distribution, which we can call $\pi_{j,k}(\theta)$. So now we have defined a mapping $\pi_{j,k} : \mathcal{P}(S) \to \mathcal{P}(S)$ for every S. Theorem 2.1 also proves that the distributions θ and $\pi_{j,k}(\theta)$ are mutually absolutely continuous, so in particular have the same support.

Invariant measures for the mappings. These mappings $\pi_{j,k}$ have apparently not been studied previously, even for special spaces S and the simplest case $k = 2$. Amongst the range of questions one could ask, we will seek to study it as a dynamical system. Given a mapping π from a space to itself, it is mathematically natural to consider iterates

$$\pi^{n+1}(\theta) = \pi(\pi^n(\theta)), n \geq 1. \tag{1.1}$$

In our setting it seems plausible that (at least for typical initial θ) the iterates should converge to some limit, that is we expect weak convergence

$$\pi^n_{j,k}(\theta) \to_w \phi \text{ as } n \to \infty \tag{1.2}$$

and then we expect[1] the limit ϕ to satisfy the *fixed point* or *invariant distribution* condition

$$\pi_{j,k}(\phi) = \phi. \tag{1.3}$$

Some comments about this set-up.
(a) The *iterative procedure* (1.1) does not have any simple stochastic process interpretation, in contrast to the mapping $\theta \to \pi_{j,k}(\theta)$ derived from the Markov chain.
(b) The equation for the stationary distribution $\pi_{j,k}(\theta)$, which for finite S is the elementary Markov chain relation $\pi = \pi\mathbf{P}$, is a linear equation, whereas the fixed point equation (1.3) is decidedly non-linear.
(c) On any S and for any (j, k), two types of measures are always invariant: we call these the *omnipresent* measures.

- The distribution δ_s degenerate at one point s;
- The uniform two-point distribution $\delta_{s_1,s_2} = \frac{1}{2}(\delta_{s_1} + \delta_{s_2})$.

(d) If an invariant distribution on S has support $S_0 \subset S$ then we can regard it as an invariant distribution on S_0. So the essential question is: given S, what are the invariant distributions *with full support*? Note that when $\pi^n_{j,k}(\theta) \to_w \phi$ the distributions $(\pi^n_{j,k}(\theta), n \geq 1)$ all have the same support (by Theorem 2.1) but ϕ may (as usually has, it turns out) have smaller support.

Motivation. There is no notion of "uniform distribution" applicable to every compact metric space S. The original motivation for this project was the hope that our invariant distributions might provide a proxy for uniform distributions on a general S. We attempted to find such distributions via numerically implementing the iterative procedure on various spaces S. What we found was that, in the absence of some special symmetry property preserved under the iterative procedure, one almost always obtained a limit supported on only one or two points, the *omnipresent* measures mentioned above. This seemed counter-intuitive, and prompted the further study of invariant measures, even though the original motivation turned out to be unsuccessful.

What numerics and simulation suggest. Our quite extensive study via numerics and simulation is described in a companion document Aldous et al. (2024), and suggests the following big picture.
(a) For $k = 2$, there are no invariant measures other than the omnipresent ones, except perhaps for "exist by symmetry" ones; with that exception, for $j = 1, k = 2$ the iterates (1.1) converge to some δ_s, and for $j = 2, k = 2$ the iterates (1.1) converge to some δ_{s_1,s_2}. The precise limits (s, s_1, s_2) may depend on the initial θ. In the case of δ_{s_1,s_2}, the pair (s_1, s_2) is a local maximum of $d(\cdot, \cdot)$.
(b) For larger k, for some types of space S there are additional *sporadic* invariant measures; we don't see a pattern.

[1] As observed in Aldous et al. (2024), the Markov chain is not always a Feller process, so (1.3) does not immediately follow from (1.2).

(c) For large k, as j increases we see (in all the examples we have studied) a transition, around $j/k = 0.7$, between convergence to some δ_s and convergence to some δ_{s_1,s_2}. However there seems no reason to believe that there is a universal value near 0.7.
(d) Except for the omnipresent ones, all invariant measures ϕ that we have encountered are *unstable*, in that from any initial distribution that is ϕ plus a generic (not symmetry-preserving) small perturbation, the iterates converge to some δ_s or δ_{s_1,s_2}.

What can we actually prove? In short: very little. Here are the results that we will derive in this article.

- Theorem 2.1 is the Markov chain convergence result.
- Results in Section 3 for $|S| = 2$ or 3 are consistent with general picture above.
- Theorem 4.1: *For every S, the set of invariant distributions for $\pi_{1,2}$ is the same as the set of invariant distributions for $\pi_{2,2}$.* This is surprising, in that apparently (as in (a) above) the iterates almost always converge to some δ_s for $\pi_{1,2}$, but to some δ_{s_1,s_2} for $\pi_{2,2}$.
- Theorem 5.1: *There are no $\pi_{1,2}$ or $\pi_{2,2}$-invariant distributions on the interval* $[0,1]$ *other than the omnipresent ones.*
- Theorem 6.1: *There are no $\pi_{1,2}$ or $\pi_{2,2}$-invariant distributions on a space of finite binary tree leaves other than the omnipresent ones.*

Of course, for any specific S, one can simply write out the *fixed point* definition (1.3) and seek some ad hoc method of finding all solutions. The results above carry this through (for $\pi_{1,2}$) for $|S| = 3$ and for the interval $[0,1]$, and for leaf-labeled binary trees. But these are essentially "proofs by contradiction" using specific features of the specific class of spaces. For general S and $\pi_{1,2}$ one feels there should be some "contraction" argument for the iterates $\pi_{1,2}^n(\theta)$ – the distributions should become more concentrated as n increases – but we are unable to formalize that general idea.

2. Existence and uniqueness of stationary distributions

Theorem 2.1. *Consider a compact metric space (S,d) and a probability distribution $\theta \in \mathcal{P}(S)$. For each pair $1 \le j \le k,\ k \ge 2$, the Markov chain $\mathbf{X}^{\theta,j,k} = (X^{\theta,j,k}(t), t = 0,1,2,\ldots)$ has a unique stationary distribution $\pi_{j,k}(\theta)$. From any initial point, the variation distance $D(t)$ between $\pi_{j,k}(\theta)$ and the distribution of $X^{\theta,j,k}(t)$ satisfies*

$$D(2t) \le (1 - 1/k^{k-1})^t, \quad 1 \le t < \infty \tag{2.1}$$

and so there is convergence to stationarity in variation distance. Moreover, for $\pi = \pi_{j,k}(\theta)$

$$(\theta(A))^k \le \pi(A) \le k\theta(A),\ \ A \subseteq S \tag{2.2}$$

and so π and θ are mutually absolutely continuous.

Note that the bound on variation distance depends only on k.

Proof: First note that for any partition $(B_i, 1 \le i \le k)$ of S we have

$$\sum_i (\theta(B_i))^k \ge 1/k^{k-1} \tag{2.3}$$

because by convexity the sum is minimized when $\theta(B_i) \equiv 1/k$.

We construct the process $X(t) = X^{\theta,j,k}(t)$ in the natural way, by creating i.i.d. θ-distributed $(\mathbf{Y}(t) = (Y_i(t), 1 \le i \le k), t \ge 1)$ and defining for $t \ge 1$

$X(t)$ is the element of $(Y_i(t), 1 \le i \le k)$ attaining the j'th smallest value of $(d(X(t-1), Y_i(t)), 1 \le i \le k)$.

In defining the re-ordering to determine "j'th smallest", we break ties in accordance with the original i – that is, if $d(X(t-1), Y_{i_1}(t)) = d(X(t-1), Y_{i_2}(t))$ for $i_1 < i_2$ then we put the i_1 term before the i_2 term in the reordering. Because the Y_i are i.i.d. this has the same effect as breaking the tie randomly.

We define the natural coupling $(X(t), X'(t))$ of two chains started from arbitrary different states by using the same realizations of $Y_i(t)$ for each chain. We first seek to upper bound the coupling time $T := \min\{t : X(t) = X'(t)\}$. Consider a realization $\mathbf{y} = (y_i, 1 \leq i \leq k)$ of $\mathbf{Y}(t+2)$. This $\mathbf{y}$ induces a partition of S, say $(B_i(\mathbf{y}), 1 \leq i \leq k)$, where $B_i(\mathbf{y})$ is the set of $s \in S$ such that $d(s, y_i)$ is the j'th smallest of $(d(s, y_u), 1 \leq u \leq k)$, breaking ties as above. The central part of the proof is the observation that the event $\{T \leq t+2\}$ includes the event

$$\text{each component of } \mathbf{Y}(t+1) \text{ is in the same set of the partition } (B_i(\mathbf{Y}(t+2)), 1 \leq i \leq k). \quad (2.4)$$

Now $\mathbf{Y}(t+1)$ is independent of $\mathbf{Y}(t+2)$, so we can apply (2.3) to show that event (2.4) has probability $\geq 1/k^{k-1}$. This remains true conditional on $(X(t), X'(t))$, and hence conditional on $\{X(t) \neq X'(t)\}$, implying that

$$\mathbb{P}(T \leq t+2 | T > t) \geq 1/k^{k-1}.$$

So inductively

$$\mathbb{P}(T > 2t) \leq (1 - 1/k^{k-1})^t, \quad 1 \leq t < \infty. \quad (2.5)$$

This is true for arbitrary initial distributions θ and $\theta' \in \mathcal{P}(S)$, and so in particular for θ and $\theta^{(2)}$, where $\theta^{(t)}$ denotes the distribution of $X^{\theta,j,k}(t)$. So (2.5) bounds the variation distance

$$||\theta^{(2t+2)} - \theta^{(2t)}||_{VD} \leq (1 - 1/k^{k-1})^t, \quad 1 \leq t < \infty$$

and similarly

$$||\theta^{(2t+1)} - \theta^{(2t)}||_{VD} \leq (1 - 1/k^{k-1})^t, \quad 1 \leq t < \infty.$$

Now variation distance is a complete metric on $\mathcal{P}(S)$, so $\theta^{(t)}$ converges in variation distance to a limit π, and π is a stationary distribution for the kernel $K^{\theta,j,k}$. Then applying (2.5) to π and an arbitrary other initial distribution establishes (2.1) and shows that π is the *unique* stationary distribution. Then (2.2) follows by considering the first step $(X(0), X(1))$ of the stationary chain, because for $A \subset S$

$$\cap_i \{Y_i(1) \in A\} \subseteq \{X(1) \in A\} \subseteq \cup_i \{Y_i(1) \in A\}.$$

□

Remarks. The variation distance bound (2.1) is exponentially decreasing in time, but it is more natural to consider *mixing time* in the sense of Levin et al. (2009). The example of the uniform distribution θ on a 2-point space with $j = 1$ shows that the mixing time as a function of k can be order 2^k.

The proof of Theorem 2.1 does not say anything about $\pi_{j,k}(\theta)$ except (2.2). We do not know if there are informative analytic descriptions of $\pi_{j,k}(\theta)$ in terms of θ.

3. Two or 3 points

3.1. *Two points – the binomial case.* The case of a 2-element space $S = \{a, b\}$ and general (j, k) is not completely trivial. Here is an outline – for more details see Aldous et al. (2024).

Parametrizing a distribution θ on S by $p := \theta(a)$, we view the mapping $\pi_{j,k} : \mathcal{P}(S) \to \mathcal{P}(S)$ as a mapping $\pi_{j,k} : [0,1] \to [0,1]$. From the stationary distribution we find, in terms of binomial variables,

$$\pi_{j,k}(p) = \frac{\mathbb{P}(\text{Bin}(k,p) > k-j)}{\mathbb{P}(\text{Bin}(k,p) > k-j) + \mathbb{P}(\text{Bin}(k,p) < j)}.$$

So a fixed point is a solution of the equation

$$\pi_{j,k}(p) = p. \quad (3.1)$$

The omnipresent fixed points are $p = 0, p = 1/2, p = 1$; are there others? By symmetry it is enough to consider $0 < p < 1/2$.

For given (j, k), we observe in Aldous et al. (2024) three possible types of qualitative behavior:
(i) $\pi^n_{j,k}(p) \to 0$ as $n \to \infty$, for all $0 < p < 1/2$.
(ii) $\pi^n_{j,k}(p) \to 1/2$ as $n \to \infty$, for all $0 < p < 1/2$.
(iii) There exists a critical value $p_{crit} \in (0, 1/2)$ which is unstable: that is,
p_{crit} is invariant
and $\pi^n_{j,k}(p) \to 0$ as $n \to \infty$, for all $0 < p < p_{crit}$
and $\pi^n_{j,k}(p) \to 1/2$ as $n \to \infty$, for all $p_{crit} < p < 1/2$.
Case (iii) first arises with $k = 5, j = 4$; then for larger values of k we see one or more values of j (depending on k) which fit case (iii). For instance, with $k = 8$ we observe case (i) for $1 \leq j \leq 5$, case (iii) for $j = 6$ with $p_{crit} = 0.26405$, and case (ii) for $j = 7, 8$.

Of course the 2-point space is very special. The occurrence of these "sporadic" case (iii) fixed points seems much rarer in other spaces.

3.2. *Three elements.* Here we consider $S = \{a, b, c\}$ where the three distances are distinct, say

$$d(a, b) < d(a, c) < d(b, c). \tag{3.2}$$

Theorem 3.1. *If S satisfies (3.2) then there is no $\pi_{1,2}$-invariant distribution except the omnipresent ones.*

Proof: It is enough to prove that there is no invariant distribution $\theta = (\theta_a, \theta_b, \theta_c)$ with each term strictly positive. So suppose, to get a contradiction, such θ exists.

Take Y, Y_1, Y_2 independent with distribution θ. Invariance says that the random variable X defined as

$$\begin{aligned} X =& Y_1 \text{ if } d(Y, Y_1) < d(Y, Y_2) \\ =& Y_2 \text{ if } d(Y, Y_2) < d(Y, Y_1) \end{aligned}$$

will also have distribution θ. Writing out the ways that X can be c or b or a gives the equations

$$\begin{aligned} \theta_c =& \theta_c(1 - (1 - \theta_c)^2) + \theta_b \theta_c^2 + \theta_a \theta_c^2 \\ \theta_b =& \theta_c \theta_b^2 + \theta_b(1 - (1 - \theta_b)^2) + \theta_a(\theta_b^2 + 2\theta_b\theta_c) \\ \theta_a =& \theta_c(\theta_a^2 + 2\theta_a\theta_b) + \theta_b(\theta_a^2 + 2\theta_a\theta_c) + \theta_a(1 - (1 - \theta_a)^2). \end{aligned}$$

Because each term of θ is strictly positive, we can cancel the common terms to get

$$1 = 1 - (1 - \theta_c)^2 + (1 - \theta_c)\theta_c \tag{3.3}$$

$$1 = \theta_c\theta_b + (1 - (1 - \theta_b)^2) + \theta_a(\theta_b + 2\theta_c) \tag{3.4}$$

$$1 = \theta_c(\theta_a + 2\theta_b) + \theta_b(\theta_a + 2\theta_c) + 1 - (1 - \theta_a)^2).$$

Equation (3.3) reduces to $2\theta_c^2 - 3\theta_c + 1 = 0$ with solutions $\theta_c = 1$ or $1/2$. The solution with $\theta_c = 1$ is excluded by supposition, so we must have $\theta_c = 1/2$. Now we have $\theta_a = 1/2 - \theta_b$; inserting into (3.4), the equation reduces to $2\theta_b^2 - 2\theta_b + \frac{1}{2} = 0$ with solution $\theta_b = \frac{1}{2}$. So $\theta_a = 0$, contradicting the supposition. □ □

Theorem 6.1 establishes a more general result, but we have given the simpler proof above to demonstrate the style of "proof by contradiction" to be used later.

4. The $k = 2$ case.

Theorem 4.1. *For every compact metric space S, the set of invariant distributions for $\pi_{1,2}$ is the same as the set of invariant distributions for $\pi_{2,2}$.*

Proof: Given $\theta \in \mathcal{P}(S)$, the transition kernel $K = K^{\theta,1,2}$ for $\pi_{1,2}$ can be written as a Radon-Nikodym density w.r.t. θ as follows.[2]

$$\begin{aligned}\frac{dK(x,\cdot)}{d\theta(\cdot)}(y) =& 2\theta\{z : d(x,z) > d(x,y)\} + \theta\{z : d(x,z) = d(x,y)\}\\ =& \int (2 \cdot 1_{\{z:d(x,z)>d(x,y)\}} + 1_{\{z:d(x,z)=d(x,y)\}})\theta(dz).\end{aligned}$$

So the identity $\theta = \theta K$ characterizing a $\pi_{1,2}$-invariant distribution θ can be written in density form as

$$\begin{aligned}1 &= \int \theta(dx)\ \frac{dK(x,\cdot)}{d\theta(\cdot)}(y)\\ &= \int\int (2 \cdot 1_{\{z:d(x,z)>d(x,y)\}} + 1_{\{z:d(x,z)=d(x,y)\}})\theta(dz)\theta(dx) \qquad (4.1)\end{aligned}$$

where the equality holds for θ-a.a. y. Because

$$\begin{aligned}1 &= \int\int 1\ \ \theta(dz)\theta(dx)\\ &= \int\int (1_{\{z:d(x,z)>d(x,y)\}} + 1_{\{z:d(x,z)=d(x,y)\}} + 1_{\{z:d(x,z)<d(x,y)\}})\ \ \theta(dz)\theta(dx)\end{aligned}$$

we have from (4.1) that

$$\int\int 1_{\{z:d(x,z)>d(x,y)\}}\ \ \theta(dz)\theta(dx) = \int\int 1_{\{z:d(x,z)<d(x,y)\}}\ \ \theta(dz)\theta(dx). \qquad (4.2)$$

Analogous to (4.1), the identity characterizing a $\pi_{2,2}$-invariant distribution ϕ can be written as

$$1 = \int\int (2 \cdot 1_{\{z:d(x,z)<d(x,y)\}} + 1_{\{z:d(x,z)=d(x,y)\}})\phi(dz)\phi(dx). \qquad (4.3)$$

By (4.1) and (4.2), any $\pi_{1,2}$-invariant distribution θ satisfies (4.3) and is therefore a $\pi_{2,2}$-invariant distribution. The converse holds via the analog of (4.2) for ϕ. □

5. The case $S = [0, 1]$

Numerical study in Aldous et al. (2024) suggests that there are no invariant distributions on $[0, 1]$ with full support, for any (j, k). Theorem 5.1 proves a slightly stronger result in the case $k = 2$ (recall that by Theorem 4.1 the cases $j = 1$ and $j = 2$ here are identical). The stronger form is not true for general (j, k), for instance the uniform distribution on the 4 points $\{0, 0.4, 0.6, 1\}$ is invariant for $\pi_{3,4}$.

Theorem 5.1. *There are no $\pi_{2,2}$-invariant distributions on $[0, 1]$ other than those of the form δ_s or δ_{s_1,s_2}.*

By considering the endpoints of the support of an invariant distribution, and scaling, this reduces to proving

> **equivalent assertion:** The only $\pi_{2,2}$-invariant distribution on $[0, 1]$ whose support contains both 0 and 1 is the distribution $\delta_{0,1}$.

[2]In a step from x with sampled Y_1, Y_2, to jump to y we need $(Y_1 = y, d(Y_2, x) > d(x, y))$ or $(Y_2 = y, d(Y_1, x) > d(x, y))$ or $Y_1 = Y_2 = y$.

We will prove this in two steps.

Lemma 5.2. *There is no $\pi_{2,2}$-invariant distribution whose support contains* 0 *and which assigns zero weight to the point* 0.

Proof: For a proof by contradiction, suppose such an invariant distribution θ exists. Take Y, Y_1, Y_2 independent with distribution θ. Invariance says that the random variable X defined as

$$\begin{aligned} X =& Y_2 \text{ if } |Y - Y_1| < |Y - Y_2| \\ =& Y_1 \text{ if } |Y - Y_2| < |Y - Y_1| \end{aligned}$$

(with our usual convention about ties) will also have distribution θ. Fix $0 < x < 1$. From the definition we have the inclusion of events

$$\{X \leq x\} \subseteq A_1 \cup A_2 \cup A_3$$

where

$$\begin{aligned} A_1 :=& \{Y_1 \leq x,\ Y_2 \leq x\} \\ A_2 :=& \{Y_1 \leq x,\ Y_2 > x,\ Y \geq \tfrac{1}{2}(Y_1 + Y_2)\} \\ A_3 :=& \{Y_2 \leq x,\ Y_1 > x,\ Y \geq \tfrac{1}{2}(Y_1 + Y_2)\} \end{aligned}$$

and the (A_i) are disjoint. Now note that

$$A_2 \subseteq \{Y_1 \leq x,\ Y \leq \tfrac{1}{2}(x + Y_2)\}$$

and similarly for A_3. So by independence, the distribution function F of θ satisfies

$$F(x) \leq F^2(x) + 2F(x)\mathbb{P}(Y \leq \tfrac{1}{2}(x + Y_2)). \tag{5.1}$$

By hypothesis, $F(x) > 0$ for small $x > 0$ and $F(x) \downarrow 0$ as $x \downarrow 0$. So we can divide both sides of (5.1) by $F(x)$ and take limits as $x \downarrow 0$ and deduce

$$\mathbb{P}(Y \leq \tfrac{1}{2}Y_2) \geq \tfrac{1}{2}.$$

By symmetry we also have $\mathbb{P}(Y_2 \leq \frac{1}{2}Y) \geq \frac{1}{2}$, and so

$$\mathbb{P}(\tfrac{1}{2}Y < Y_2 < 2Y) = 0.$$

But this is impossible for i.i.d. samples from a distribution θ on $(0, 1]$, because it would remain true for θ conditioned on an interval of the form $[y, 3y/2]$. □

Using Lemma 5.2 and reflection-symmetry of $[0, 1]$, to prove the *equivalent assertion* and hence Theorem 5.1 it will be sufficient to prove

Lemma 5.3. *If θ is a $\pi_{2,2}$-invariant distribution and $\theta_0 > 0, \theta_1 > 0$ then $\theta = \delta_{0,1}$.*

Here we write θ_s for $\theta(\{s\})$.

Proof: First note an elementary fact:

$$\text{if } 0 < x < 1 \text{ and } \beta \geq 0 \text{ and } x \geq x^2 + 2x(1 - x)\beta \text{ then } \beta \leq 1/2. \tag{5.2}$$

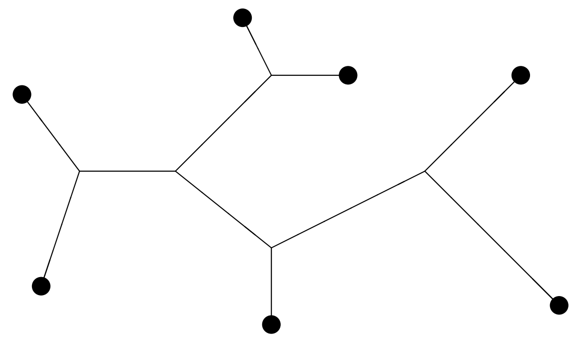

Figure 6.1. A BTL space S with $|S| = 7$.

From the construction with (Y, Y_1, Y_2, X) we have

$$\begin{aligned}
\theta_0 =&\theta_0^2 + 2\mathbb{P}(Y_1 = 0, Y_2 > 0, Y > Y_2/2) + \mathbb{P}(Y_1 = 0, Y_2 > 0, Y = Y_2/2)\\
=&\theta_0^2 + 2\theta_0(\mathbb{P}(2Y > Y_2, Y_2 > 0) + \tfrac{1}{2}\mathbb{P}(2Y = Y_2, Y_2 > 0))\\
=&\theta_0^2 + 2\theta_0(1-\theta_0)(\mathbb{P}(2Y > Y_2|Y_2 > 0) + \tfrac{1}{2}\mathbb{P}(2Y = Y_2|Y_2 > 0))\\
\geq&\theta_0^2 + 2\theta_0(1-\theta_0)(\mathbb{P}(Y > 1/2) + \mathbb{P}(Y = 1/2, Y_2 < 1|Y_2 > 0) + \tfrac{1}{2}\mathbb{P}(Y = 1/2, Y_2 = 1|Y_2 > 0)) \qquad (5.3)\\
=&\theta_0^2 + 2\theta_0(1-\theta_0)(\mathbb{P}(Y > 1/2) + \theta_{1/2}\ [\mathbb{P}(Y_2 < 1|Y_2 > 0) + \tfrac{1}{2}\mathbb{P}(Y_2 = 1|Y_2 > 0)])\\
=&\theta_0^2 + 2\theta_0(1-\theta_0)(\mathbb{P}(Y > 1/2) + \theta_{1/2}\ [\tfrac{1}{2} + \tfrac{1}{2}\mathbb{P}(Y_2 < 1|Y_2 > 0)])\\
\geq&\theta_0^2 + 2\theta_0(1-\theta_0)(\mathbb{P}(Y > 1/2) + \tfrac{1}{2}\mathbb{P}(Y = 1/2)). \qquad (5.4)
\end{aligned}$$

By hypothesis $\theta_0 > 0$, so by (5.2) we have $\mathbb{P}(Y > 1/2) + \frac{1}{2}\mathbb{P}(Y = 1/2) \leq 1/2$. However we have the analogous sequence of equalities and inequalities for θ_1, which imply $\mathbb{P}(Y < 1/2) + \frac{1}{2}\mathbb{P}(Y = 1/2) \leq 1/2$, and so we must have

$$\mathbb{P}(Y > 1/2) + \tfrac{1}{2}\mathbb{P}(Y = 1/2) = 1/2 = \mathbb{P}(Y < 1/2) + \tfrac{1}{2}\mathbb{P}(Y = 1/2).$$

The quantity (5.4) now *equals* θ_0, so the inequalities at (5.3) and (5.4) must in fact be equalities. In order for the inequality leading to (5.4) to be an equality, we must have either $\theta_{1/2} = 0$ or $\mathbb{P}(Y_2 < 1|Y_2 > 0) = 0$. In the latter case, θ is supported on $\{0, 1\}$ and so $\theta = \delta_{0,1}$, as desired. So the remaining case is $\theta_{1/2} = 0$. In this case, for the inequality leading to (5.3) to be an equality, we must have $\mathbb{P}(Y_2 < 2Y, Y < 1/2|Y_2 > 0) = 0$. But, as at the end of the proof of Lemma 5.2, this can only happen if $\mathbb{P}(0 < Y < 1/2) = 0$. By the analogous argument for θ_1 we have $\mathbb{P}(1/2 < Y < 1) = 0$, and so the distribution is supported on $\{0, 1\}$ and must be $\delta_{0,1}$, as desired. □

This line of argument can be extended to some other values of (j, k) – see Aldous et al. (2024).

6. A class of tree spaces

In this section we consider binary[3] tree leaves (BTL), illustrated in Figure 6.1, as a class of finite spaces. Here S is the finite set of leaves; the edges have lengths which serve to determine the

[3]Essentially the same argument works without the *binary* assumption.

distance between two leaves as the length of the unique path between them; the edges also define $|S|-2$ branchpoints. To "break symmetry" we assume

$$\text{all distances } (d(s_i,s_j), j \neq i) \text{ are distinct.} \tag{6.1}$$

We claim that, as suggested by the general picture from numerics, for $k=2$ there are no invariant measures other than the omnipresent ones. An invariant measure supported on a subset of leaves is an invariant measure on the induced spanning tree of that subset, so to prove that claim it suffices to prove

Theorem 6.1. *On a BTL space S with $|S| \geq 3$ and satisfying (6.1), and for $k=2$, there are no invariant measures with full support.*

Proof: As in previous proofs, consider a $\pi_{1,2}$ - invariant distribution θ with full support on S, where $|S| \geq 3$. Take Y_0, Y_1, Y_2 independent with distribution θ. Invariance says that the random variable X defined as

$$\begin{aligned} X =& Y_1 \text{ if } d(Y_0,Y_1) < d(Y_0,Y_2) \\ =& Y_2 \text{ if } d(Y_0,Y_2) < d(Y_0,Y_1) \end{aligned} \tag{6.2}$$

will also have distribution θ. We proceed to a proof by contradiction.

We quote an elementary fact.

Lemma 6.2. *For any probability distribution θ on a BTL space S, either*
(i) $\theta(s_0) > \frac{1}{2}$ for some $s_0 \in S$
or (ii) there exists a centroid, that is a branchpoint such that the associated partition $S = \cup_{i=1}^3 A_i$ of leaves satisfies $0 < \theta(A_i) \leq \frac{1}{2}$ for all i.

Consider case (i). That is, suppose θ is invariant and $\theta(s_0) \in (\frac{1}{2}, 1)$. From the invariance relation (6.2), in order that $X = s_0$ it is sufficient that

$$(Y_0 \neq s_0, Y_1 = s_0, Y_2 = s_0) \text{ or } (Y_0 = s_0, Y_1 \text{ or } Y_2 = s_0).$$

So, setting $\theta(s_0) = x \in (\frac{1}{2}, 1)$,

$$x \geq (1-x)x^2 + x(2x - x^2).$$

Cancelling x, this reduces to $2x^2 - 3x + 1 \geq 0$, but this inequality is false for $x \in (\frac{1}{2}, 1)$.

Now consider case (ii). There is a centroid branchpoint defining a partition $S = \cup_{i=1}^3 A_i$. Consider the leaf s_1 which is closest to the centroid. We may assume $s_1 \in A_1$. From the invariance relation (6.2), in order that $X = s_1$ it is sufficient that the following condition (*) holds:

exactly one of (Y_1, Y_2) equals s_1
and
Y_0 and the other[4] Y are in different components of $\cup_{i=1}^3 A_i$.

For instance, if $Y_0 \in A_2$ and $Y_1 = s_1$ and $Y_2 \in A_3$, then Y_2 is some leaf in A_3 which is farther from the centroid than is s_1, so $d(Y_0, s_1) < d(Y_0, Y_2)$. The other possibilities are similar.

By considering the three possibilities for " different components of $\cup_{i=1}^3 A_i$" we see that the probability of (*) equals $\theta(s_1)$ times

$$\begin{gathered} 2\theta(A_1)\theta(A_2) + 2(\theta(A_1) - \theta(s_1))\theta(A_2) \\ +2\theta(A_1)\theta(A_3) + 2(\theta(A_1) - \theta(s_1))\theta(A_3) \\ +4\theta(A_2)\theta(A_3) \end{gathered}$$

which rearranges to

$$\theta(s_1)[4(\theta(A_1)\theta(A_2) + \theta(A_1)\theta(A_3) + \theta(A_2)\theta(A_3)) - 2\theta(s_1)(\theta(A_2) + \theta(A_3))]$$

[4]The leaf from (Y_1, Y_2) that is not s_1.

$$= \theta(s_1) \cdot B, \text{ say.} \tag{6.3}$$

A disjoint sufficient condition for $X = s_1$ is that $Y_1 = Y_2 = s_1$, which has probability $\theta^2(s_1)$. So

$$\theta(s_1) = \mathbb{P}(X = s_1) \geq \theta(s_1)(B + \theta(s_1)).$$

Cancelling the $\theta(s_1)$ term,

$$1 \geq 4(\theta(A_1)\theta(A_2) + \theta(A_1)\theta(A_3) + \theta(A_2)\theta(A_3)) - 2\theta(s_1)(\theta(A_2) + \theta(A_3) - \tfrac{1}{2}).$$

Because $\sum_i \theta(A_i) = 1$ we have $\theta(A_2) + \theta(A_3) - \frac{1}{2} = \frac{1}{2} - \theta(A_1)$ and

$$2(\theta(A_1)\theta(A_2) + \theta(A_1)\theta(A_3) + \theta(A_2)\theta(A_3)) = 1 - \sum_i \theta^2(A_i)$$

and the inequality above reduces to

$$1 \geq 2 - 2\sum_i \theta^2(A_i) - 2\theta(s_1)(\tfrac{1}{2} - \theta(A_1)).$$

Because $\theta(s_1) \leq \theta(A_1)$, this implies

$$C := \sum_i \theta^2(A_i) + \theta(A_1)(\tfrac{1}{2} - \theta(A_1)) \geq \tfrac{1}{2}. \tag{6.4}$$

We need to show that $C \geq \frac{1}{2}$ cannot in fact occur under the constraints $0 < P(A_i) \leq \frac{1}{2}$ and $\sum_i \theta(A_i) = 1$. Given $\theta(A_1) = x$, the quantity C is maximized when $(\theta(A_2), \theta(A_3)) = (\frac{1}{2}, \frac{1}{2} - x)$ and so

$$C \leq x^2 + \tfrac{1}{4} + (\tfrac{1}{2} - x)^2 + x(\tfrac{1}{2} - x) = x^2 - \tfrac{1}{2}x + \tfrac{1}{2}.$$

This implies that $C < \frac{1}{2}$ on the open interval $x \in (0, \frac{1}{2})$, and we cannot have $x = 0$ or $\frac{1}{2}$ by the $\theta(A_i) > 0$ constraint. □

Acknowledgements

The authors thank an anonymous referee for careful reading of the submission.

References

Aldous, D. J., Cruz, M., and Feng, S. Markov chains and mappings of distributions on compact spaces II: Numerics and Conjectures. *ArXiv Mathematics e-prints* (2024). arXiv: 2403.18153.

Levin, D. A., Peres, Y., and Wilmer, E. L. *Markov chains and mixing times.* American Mathematical Society, Providence, RI (2009). ISBN 978-0-8218-4739-8. MR2466937.