

Markov chains and mappings of distributions on compact spaces II: Numerics and Conjectures

David J. Aldous* Madelyn Cruz† Shi Feng‡

March 28, 2024

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 $\pi_{j,k}(\theta)$. This defines a mapping $\pi_{j,k} : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$. What happens when we iterate this mapping? In particular, what are the fixed points of this mapping? A few results are proved in a companion article [1]; this article, not intended for formal publication, records numerical studies and conjectures.

This document records our investigation into the novel topic described below. Numerics and heuristics suggest a variety of general conjectures, but what we can actually prove is seriously limited. The rigorous results stated here (Theorems 1, 4, 8, 9) are proved in an accompanying article [1]. This document is intended as informal discussion of a big picture. The introduction below expands on the introduction to [1].

1 Introduction

We start with informal background that led to the formulation of the questions to be studied. If I suggest a specific compact metric space (S, d) and ask you to define some “interesting” mapping $f : S \rightarrow S$ then you can certainly do so. However, if I challenge you to define a scheme which can be

*Department of Statistics, 367 Evans Hall # 3860, U.C. Berkeley CA 94720; aldous@stat.berkeley.edu; www.stat.berkeley.edu/users/aldous.

†Department of Mathematics, University of Michigan; mccruz@umich.edu

‡Department of Mathematics, Cornell University; sf599@cornell.edu

used in any “generic” such space S , without exploiting any structure of the specific space, then that seems very difficult. For instance you could define

$$f(s) := \arg \max_y d(s, y)$$

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. Are there any more interesting general schemes?

Instead let us write $\mathcal{P}(S)$ for the space of probability distributions on S . Consider the analogous challenge: define a scheme which can be used to specify a mapping $\mathcal{P}(S) \rightarrow \mathcal{P}(S)$ for a generic S . In contrast to the case of mappings $S \rightarrow S$, there are perhaps many ways to do so. In particular, there are schemes involving Markov chains. For instance, given $\theta \in \mathcal{P}(S)$ you can define a transition kernel $K(s, \cdot)$ for $s \in \text{support}(\theta)$ by conditioning θ to the ball of radius 1 around s :

$$K(s, A) = \theta(A \cap \text{ball}(s, 1)) / \theta(\text{ball}(s, 1)).$$

Under a simple connectivity condition¹ on θ , the corresponding Markov chain has a unique stationary distribution, which we can call $\pi(\theta)$. So this scheme defines a map $\pi : \mathcal{P}_0(S) \rightarrow \mathcal{P}(S)$ for every S but only for the subset $\mathcal{P}_0(S) \subseteq \mathcal{P}(S)$ satisfying the connectivity condition. One could view this as a continuous-space analog of random walk on a graph.

In this project we study a scheme which works for **all** compact metric spaces S and **all** of $\mathcal{P}(S)$. Given $\theta \in \mathcal{P}(S)$, one can consider the Markov chain that from point s makes 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. This defines a kernel $K^{\theta, j, k}$ on S . Write the associated chain as $\mathbf{X}^{\theta, j, k} = (X^{\theta, j, k}(t), t = 0, 1, 2, \dots)$. Theorem 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 $\pi_{j, k} : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$ for every S . Moreover Theorem 1 proves that the distributions θ and $\pi_{j, k}(\theta)$ are mutually absolutely continuous, so in particular have the same support.

These maps $\pi_{j, k}$ have apparently not been studied previously, even for special spaces S , so the purpose of this project is to initiate their study. There is a range of problems one might consider. The proof of Theorem 1, using the natural coupling, does not yield any helpful explicit expression for the stationary distribution. So one can ask

¹Essentially, that $\cup_{s \in \text{support}(\theta)} \text{ball}(s, 1)$ is connected.

On a given space S , is there any informative description of $\pi_{j,k}(\theta)$ in terms of θ ?

In this project our focus is different. Given a mapping π from a space to itself, it is mathematically natural to consider iterates

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

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

$$\pi_{j,k}^n(\theta) \rightarrow_w \phi \text{ as } n \rightarrow \infty \quad (1)$$

and then we expect the limit to satisfy

$$\pi_{j,k}(\phi) = \phi. \quad (2)$$

We will call such a construction $(\pi_{j,k}^n(\theta), n \geq 0)$ the *iterative procedure*. Note that this does not have any simple stochastic process interpretation, in contrast to the mapping $\theta \rightarrow \pi_{j,k}(\theta)$ derived from the Markov chain.

If ϕ satisfies (2) we call it a *fixed point* or an *invariant distribution* for $\pi_{j,k}$. One can view this property as a kind of “self-similarity under sampling” property. The original motivation for this project was the hope that a typical compact metric space might have “interesting” invariant distributions. Our first attempt at finding such distributions was via numerically implementing the iterative procedure. But in doing so, with initial distributions lacking symmetry, we observed that typically there is a limit but it is supported on only one or two points. This seemed counter-intuitive, and prompted the further study of fixed points in this article.

Let us emphasize the trivial observation:

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 $\delta_{s_1, s_2} = \frac{1}{2}(\delta_{s_1} + \delta_{s_2})$.

Are there others?

1.1 Formulating a program

There are some subtleties involved in formulating a precise program, as follows.

(i) Convergence in (1) is weak convergence. Recall that a Markov chain is a *Feller* chain if its kernel K is weakly continuous:

$$K(s_n, \cdot) \rightarrow_w K(s, \cdot) \text{ for all } s_n \rightarrow s. \quad (3)$$

It is straightforward and well known ([3] section 12.3) that a Feller chain on a compact S always has at least one stationary distribution. Our chains $\mathbf{X}^{\theta,j,k}$ are not necessarily Feller, the obstacle arising when the chain involves ties – see discussion in section 2.1. But it turns out that we do not need the Feller property for the foundational result, Theorem 1. On the other hand one might hope that the mapping $\pi_{j,k} : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$ is continuous, but we can only prove that it is continuous at (roughly speaking) distributions θ for which the Feller property holds – see Proposition 2. So convergence (1) does not automatically imply that the limit is a fixed point (2), though we do not know a counter-example.

(ii) For given θ the iterates $\pi_{j,k}^n(\theta)$, $n \geq 0$ all have the same support but we will see that the limit usually has much smaller support, so we cannot insist on limits having full support on S . On the other hand, any fixed point distribution over a space S_0 is automatically a fixed point distribution over a superspace $S_1 \supset S_0$, and to talk about “fixed points for S_1 ” we want the fixed point to have some relation with S_1 . So an appropriate precise problem formulation is

Program. (a) Given a compact metric space (S, d) and given (j, k) , find all the fixed points of $\pi_{j,k}$ with full support on S .
(b) Given a compact metric space (S, d) and given (j, k) , determine which invariant measures (not necessarily with full support on S) arise as limits (1) of the iterative procedure from some initial θ with full support on S .

1.2 Toward a big picture

This article reports on our extensive study via numerics of different various spaces S , which provides the following heuristic “big picture” of issues in the program above.

(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) converge to some δ_s , and for $j = 2, k = 2$ the iterates (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.

(c) For large k , as j increases we see 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 of the form ϕ plus a generic (not symmetry-preserving) small perturbation, the iterates converge to some δ_s or δ_{s_1, s_2} .

We have succeeded in proving only a few small parts on this picture. The shorter article [1] contains proofs of the following theorems, only stated in this document.

- Theorem 1 is the Markov chain convergence result stated above.
- Theorem 4: *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 8: *There are no $\pi_{1,2}$ or $\pi_{2,2}$ -invariant distributions on the space of finite binary tree leaves (see section 6) other than the omnipresent ones.*
- Theorem 9: *There are no $\pi_{1,2}$ or $\pi_{2,2}$ -invariant distributions on the interval $[0, 1]$ other than the omnipresent ones.*

1.3 Outline of this article

Our main focus is on simulation results, leading to formulation of conjectures. We also provide technical comments.

- Section 2 serves to define the map $\pi_{j,k}$ (Theorem 1) and discusses the Feller property.
- Section 3 discusses the “symmetry is preserved” feature of $\pi_{j,k}$, implying that some invariant distributions exist “by symmetry”, but also suggesting that initial distributions with symmetry are atypical. It

also gives the counter-intuitive, though easily verified, result (Theorem 4) that the invariant distributions for $\pi_{2,2}$ coincide with the invariant distributions for $\pi_{1,2}$.

- Section 4 observes that even the 2-point case $S = \{a, b\}$ is interesting: as well as the obvious fixed points, for certain values of (j, k) there are extra fixed points, but these appear (from numerics) to be *unstable* under the mapping $\pi_{j,k}$.
- Section 5 gives examples of non-uniform invariant distributions for finite spaces S . We record the strong conjecture that, for finite S , the iterative process from *almost all* initial distributions converges to one of the omnipresent limits; this is tantamount to saying that other fixed points are unstable.
- Section 6 describes the *binary tree leaves* space, on which we can prove (Theorem 8) that for $k = 2$ the only invariant distributions are the omnipresent ones.
- The natural basic example of a continuous compact space is the unit interval, and we give a simulation study in section 7. In examples we see convergence to limit distributions supported on one or two points; we do not know whether (for any (j, k)) there exist invariant distributions with full support, though we have partial results including Theorem 9.
- Section 8 studies the case of the circle, for which the uniform distribution is invariant. Section 9 considers a high-dimensional case. Both are simulation studies.

2 Existence and uniqueness of stationary distributions

Theorem 1 ([1] Theorem 1) *Consider a compact metric space (S, d) and a probability distribution $\theta \in \mathcal{P}(S)$. For each pair $1 \leq j \leq k$, $k \geq 2$, the Markov chain $\mathbf{X}^{\theta,j,k} = (X^{\theta,j,k}(t), t = 0, 1, 2, \dots)$ 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*

$$D(2t) \leq (1 - 1/k^{k-1})^t, \quad 1 \leq t < \infty \quad (4)$$

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

$$\theta^k(A) \leq \pi(A) \leq k\theta(A), \quad A \subseteq S \quad (5)$$

and so π and θ are mutually absolutely continuous.

Remarks. (a) Note that the bound on variation distance depends only on k .

(b) The variation distance bound (4) is exponentially decreasing in time, but it is more natural to consider *mixing time* in the sense of [4]. 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 .

(c) As mentioned earlier, the proof of Theorem 1 does not say anything about $\pi_{j,k}(\theta)$ except (5). We do not know if there are informative analytic descriptions of $\pi_{j,k}(\theta)$ in terms of θ .

2.1 Concerning the Feller property

As an example to show that the Markov chains $\mathbf{X}^{\theta,j,k}$ are not necessarily Feller, consider $S = [0, 1]$ and $\theta = \frac{1}{3}(\delta_0 + \delta_1 + \lambda)$ for Lebesgue measure λ . Here the requirement (3) that $K(x_i, \cdot) \rightarrow_w K(\frac{1}{2}, \cdot)$ as $x_i \rightarrow \frac{1}{2}$ fails.

This difficulty occurs because there are ties in the distances to the points $(Y_i, 1 \leq i \leq k)$ sampled i.i.d. from θ . Suppose, for given $\theta \in \mathcal{P}(S)$ and given $s \in S$,

$$d(s, Y_i) \text{ has non-atomic distribution.} \quad (6)$$

This of course cannot hold for finite S . Property (6) implies

$$\mathbb{P}(d(s, Y_i) = d(s, Y_j)) = 0 \text{ for } j \neq i. \quad (7)$$

It is straightforward to show that, if (7) holds for all $s \in S$, then the Markov chain $\mathbf{X}^{\theta,j,k}$ is Feller. Let us consider the slightly weaker property: for given $\theta \in \mathcal{P}(S)$

$$d(s, Y_i) \text{ has non-atomic distribution for } \theta\text{-almost all } s \in S. \quad (8)$$

Note this is equivalent to

$$\mathbb{P}(d(Y_1, Y_2) = d(Y_1, Y_3)) = 0. \quad (9)$$

We can now give a partial continuity property for the maps $\pi_{j,k}$.

Proposition 2 *If $\theta \in \mathcal{P}(S)$ has property (8) then $\pi_{j,k}(\theta_n) \rightarrow_w \pi_{j,k}(\theta)$ for all $\theta_n \rightarrow_w \theta$.*

Proof. Consider $\theta_n \rightarrow_w \theta$. Write $(X(t), t = 0, 1, 2, \dots)$ for the chain $\mathbf{X}^{\theta, j, k}$ started with distribution θ and write $(X^n(t), t = 0, 1, 2, \dots)$ for the chain $\mathbf{X}^{\theta_n, j, k}$ started with distribution θ_n . It is sufficient to prove that, for fixed t , we have convergence in distribution

$$\text{dist}(X^n(t)) \rightarrow_w \text{dist}(X(t)) \text{ as } n \rightarrow \infty \quad (10)$$

because then the result will follow from (4). Proving this would make a nice exercise in a course on weak convergence. We will write out the case $(j, k) = (1, 2)$: the general case is similar.

We prove (10) by induction on t . Fix t , assume (10) for that t , and recall that by construction

$$\text{dist}(X(t)) \leq k\theta. \quad (11)$$

There exist couplings (Y_i, Y_i^n) of θ and θ_n such that

$$a(n) := \mathbb{E}d(Y_i, Y_i^n) \rightarrow 0 \text{ as } n \rightarrow \infty$$

and there exist couplings $(X(t), X^n(t))$ of ϕ and ϕ_n such that

$$b(n) := \mathbb{E}d(X(t), X^n(t)) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

We have

$$\begin{aligned} X(t+1) &= Y_1 \text{ if } d(X(t), Y_1) < d(X(t), Y_2) \\ &= Y_2 \text{ if } d(X(t), Y_1) > d(X(t), Y_2) \end{aligned}$$

and also the case of ties. Combining with the analog for $X^n(t+1)$ we have

$$\begin{aligned} d(X(t+1), X^n(t+1)) &= d(Y_1, Y_1^n) \text{ on } A_n \\ A_n &:= \{d(X(t), Y_1) < d(X(t), Y_2) \text{ and } d(X^n(t), Y_1^n) < d(X^n(t), Y_2^n)\} \\ d(X(t+1), X^n(t+1)) &= d(Y_2, Y_2^n) \text{ on } B_n \\ B_n &:= \{d(X(t), Y_1) > d(X(t), Y_2) \text{ and } d(X^n(t), Y_1^n) > d(X^n(t), Y_2^n)\} \end{aligned}$$

Now A_n contains the event $A_n(\varepsilon)$, where for $\varepsilon > 0$

$$A_n(\varepsilon) := \{d(X(t), Y_1) < d(X(t), Y_2) - 3\varepsilon; d(X(t), X^n(t)) \leq \varepsilon; d(Y_2, Y_2^n) \leq \varepsilon\}$$

and B_n contains an analogous event $B_n(\varepsilon)$. Taking account of the remaining cases, we obtain

$$\mathbb{E}d(X(t+1), X^n(t+1)) \leq 2a(n) + \Delta\mathbb{P}((A_n(\varepsilon) \cup B_n(\varepsilon))^c)$$

where $\Delta < \infty$ is the diameter of S . Using Markov's inequality to bound the probability of events like $\{d(X(t), X^n(t)) > \varepsilon\}$, we obtain

$$\mathbb{E}d(X(t+1), X^n(t+1)) \leq 2a(n) + \Delta \left(\frac{2(a(n)+b(n))}{\varepsilon} + \mathbb{P}(|d(X(t), Y_1) - d(X(t), Y_2)| \leq 3\varepsilon) \right).$$

Letting $n \rightarrow \infty$ we see that it suffices to prove

$$\mathbb{P}(|d(X(t), Y_1) - d(X(t), Y_2)| \leq 3\varepsilon) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0$$

but this follows from (7) and (10). ■

3 General remarks about fixed points of the mapping

On every compact metric space S we have an obvious “preservation of symmetry” result

Lemma 3 *If $\theta \in \mathcal{P}(S)$ is invariant under an isometry ι of S then $\pi_{j,k}(\theta)$ is also invariant under ι .*

This has several implications.

3.1 Fixed points existing by symmetry

In some cases there are distributions $\phi \in \mathcal{P}(S)$ which are invariant (that is, fixed points) “by symmetry”. In particular, the omnipresent examples mentioned earlier:

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

But there are further examples:

- (iii) The Haar probability measure on a compact group S with a metric invariant under the group action.
- (iv) On a finite space S , a sufficient condition for the uniform distribution to be invariant is that S is *transitive*, that is if for each pair s, s' there is an isometry taking s to s' . This is equivalent to the finite case of Haar measure. But for finite S a weaker condition suffices, because all that matters is the *rank matrix* – see section 5.1.

Now in those cases the distribution is invariant for all $\pi_{j,k}$. So the question becomes: for a particular (j, k) , are there invariant distributions with full support, other than those “forced by symmetry” as above?

3.2 Symmetry in the limit

If $\theta \in \mathcal{P}(S)$ is invariant under an isometry ι , then by Lemma 3 each iterate $\pi_{j,k}^n(\theta)$ is invariant and so we expect (recall that we do not know that $\pi_{j,k}$ is continuous) that a limit ϕ will be invariant. This observation will be relevant for the formulation of conjectures.

3.3 Fixed points for $\pi_{1,2}$ and $\pi_{2,2}$

Perhaps the basic example of a compact space is the unit interval $[0, 1]$, and for that space we will give a detailed simulation study of the iterative process in section 7. Figure 1 shows the observed behavior of iterates of $\pi_{1,2}$ and $\pi_{2,2}$ from the initial uniform distribution. As observed above, limits of such iterates should be invariant under the reflection isometry $x \rightarrow 1 - x$. We see the former converging to $\delta_{1/2}$ and the latter to $\delta_{0,1}$. We believe this qualitative difference will hold very generally — see the related Conjectures 5 and 7. So the following general result seems very counter-intuitive.

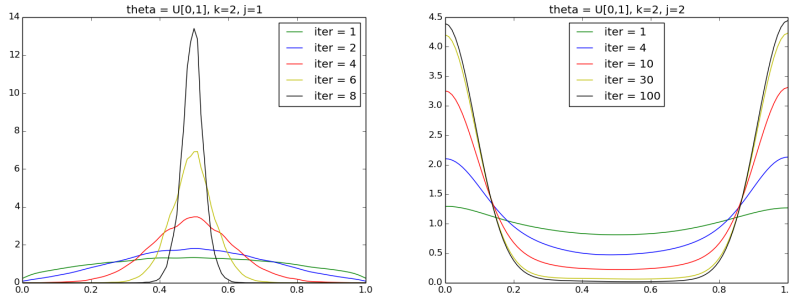


Figure 1: Iterates of $\pi_{1,2}$ (left) and $\pi_{2,2}$ (right) on the unit interval from uniform initial distribution.

Theorem 4 ([1] Theorem 3) *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 $\pi_{2,2}$.*

4 Two elements – the binomial case

One might suppose that the case of a 2-element set $S = \{a, b\}$ would be trivial, but it is not. Parametrizing a distribution θ on S by $p := \theta(a)$, we view the mapping $\pi_{j,k} : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$ as a mapping $\pi_{j,k} : [0, 1] \rightarrow [0, 1]$

defined as follows. In the associated 2-state Markov chain, the transition probabilities are

$$k_{a,b} = \mathbb{P}(\text{Bin}(k, p) < j); \quad k_{b,a} = \mathbb{P}(\text{Bin}(k, p) > k - j)$$

for Binomial random variables. From the stationary distribution we find

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

So a fixed point is a solution of the equation

$$\pi_{j,k}(p) = p. \tag{12}$$

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

We have not tried to find solutions analytically, but we will show results of numerical calculations of the iterates $\pi_{j,k}^n(p), n = 1, 2, 3, \dots$. For a given (k, j) , we observe three possible types of qualitative behavior:

- (i) $\pi_{j,k}^n(p) \rightarrow 0$ as $n \rightarrow \infty$, for all $0 < p < 1/2$.
- (ii) $\pi_{j,k}^n(p) \rightarrow 1/2$ as $n \rightarrow \infty$, for all $0 < p < 1/2$.
- (iii) There exists a critical value $p_{crit} \in (0, 1/2)$ such that
 - p_{crit} is invariant
 - and $\pi_{j,k}^n(p) \rightarrow 0$ as $n \rightarrow \infty$, for all $0 < p < p_{crit}$
 - and $\pi_{j,k}^n(p) \rightarrow 1/2$ as $n \rightarrow \infty$, for all $p_{crit} < p < 1/2$.

For us, (iii) is the interesting case. It first arises with $k = 5, j = 4$, as shown in Figure 2. We see the critical value $p_{crit} = 0.17267\dots$. We also see this is an *unstable* fixed point.

Table 1 shows the type of behavior – types (i) or (ii) or (iii) above – for all pairs (j, k) with $k \leq 9$. One take-away message is that for $S = \{a, b\}$ there exist some (j, k) for which $\pi_{j,k}$ has fixed points in addition to those existing by symmetry, but these fixed points are unstable. Another take-away message is that, except for these additional fixed points, the limit of the iterative process is either δ_a or δ_b or $\delta_{a,b}$, the latter only when j/k is somewhat close to 1. Of course the 2-point space may be very special. What properties extend to other S ?

5 Finite spaces

5.1 Non-uniform invariant distributions

In our initial investigations, via simulations of the iterative process on different spaces, we found that the iterates always converged to one- or two-point

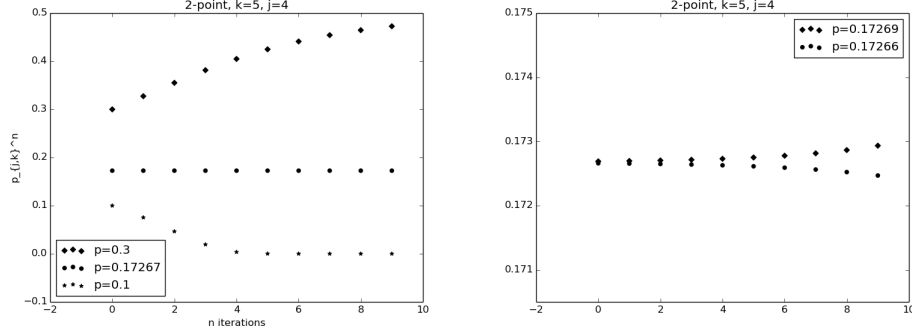


Figure 2: $S = \{a, b\}$; $k = 5, j = 4$. Iterates $n = 0, 1, 2, \dots, 10$. Left panel shows type (iii) behavior, Right panel shows the unstable fixed point at 0.17267.

support invariant distributions, δ_s or δ_{s_1, s_2} . As mentioned before, on a space with sufficient symmetry the uniform distribution will be invariant. So perhaps the existence of non-uniform invariant distributions on the 2-element space for $k \geq 5$ is an anomaly? These observations were consistent with the possibility that **all** invariant distributions supported on more than two elements are uniform on their support. But a numerical search revealed the following counter-example, in the simplest case $j = 1, k = 2$.

A finite metric space can be represented by the matrix D of distances $d(i, j)$. By taking all the non-zero distances to be between 1 and 2, the triangle inequality is automatically satisfied. Consider the example of a 5-element space with distance matrix

$$D = \begin{pmatrix} 0 & 1.714 & 1.341 & 1.656 & 1.74 \\ 1.714 & 0 & 1.298 & 1.794 & 1.03 \\ 1.341 & 1.298 & 0 & 1.715 & 1.844 \\ 1.656 & 1.794 & 1.715 & 0 & 1.524 \\ 1.74 & 1.03 & 1.844 & 1.524 & 0 \end{pmatrix}$$

What matters for our purposes, assuming as in this example that all distances are distinct², is the *rank matrix* R , where $r(i, j) = 4$ means that $d(i, j)$ is the 4'th smallest of $\{d(i, 1), d(i, 2), \dots, d(i, |S|)\}$. For the distance

²Precisely, all distances $d(i, j), 1 \leq i < j \leq n$ are distinct.

Table 1: $S = \{a, b\}$ and $2 \leq k \leq 9$. The values of j with each type of behavior, and (critical values) in type (iii) behavior.

k	(i)	(iii)	(ii)
2	1		2
3	[1, 2]		3
4	[1 – 3]		4
5	[1 – 3]	4 (0.17267)	5
6	[1 – 4]	5 (0.09558)	6
7	[1 – 5]	6 (0.06276)	7
8	[1 – 5]	6 (0.26405)	[7, 8]
9	[1 – 6]	7 (0.18884); 8 (0.03364)	9

matrix D above, the rank matrix is

$$R = \begin{pmatrix} 1 & 4 & 2 & 3 & 5 \\ 4 & 1 & 3 & 5 & 2 \\ 3 & 2 & 1 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \\ 4 & 2 & 5 & 3 & 1 \end{pmatrix}$$

By numerical calculation, for $\pi_{1,2}$ on this space there is an invariant distribution

$$\theta \approx (0.149 \ 0.188 \ 0.203 \ 0.298 \ 0.162)$$

for which the transition matrix is

$$K \approx \begin{pmatrix} 0.276 & 0.097 & 0.304 & 0.297 & 0.026 \\ 0.111 & 0.341 & 0.222 & 0.089 & 0.237 \\ 0.159 & 0.265 & 0.365 & 0.185 & 0.026 \\ 0.139 & 0.036 & 0.118 & 0.507 & 0.201 \\ 0.083 & 0.28 & 0.041 & 0.298 & 0.298 \end{pmatrix}$$

This example was found by simulating random distance matrices D , obtaining the rank matrix R , and then numerically solving for invariant distributions θ until finding a solution with full support. Note this involved non-linear equations: we need to solve $\theta K = \theta$ but here K depends on θ , for instance for $\pi_{1,2}$

$$k(i, i) = 1 - (1 - \theta(i))^2$$

$$\text{if } r(i, j) = 5 \text{ then } k(i, j) = \theta^2(j).$$

Note also that for $|S| = 5$ there are only a finite number of possible rank matrices R , so this counter-example is not like a counter-example depending on a real parameter taking a specific value.

To be precise, let us define a *rank matrix* R to be a matrix where each row consists of $\{1, 2, \dots, |S|\}$ in some order and each diagonal element $r_{ii} = 1$. We are concerned with rank matrices which are *feasible* in that they arise from some distance matrix D with distinct entries. We do not know precise conditions to determine whether a given rank matrix is feasible. Also, we do not have any general conjecture about conditions under which non-uniform invariant distributions exist.

We observed that the invariant distribution in Figure 2 was unstable. Here is another example. Consider again $\pi_{1,2}$ and the rank matrix (easily checked to be feasible)

$$R = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 2 & 1 & 3 & 4 \\ 4 & 2 & 1 & 3 \\ 2 & 4 & 3 & 1 \end{pmatrix} \quad (13)$$

for which the invariant distribution is

$$\theta = (\frac{1}{6} \ \frac{1}{6} \ \frac{2}{6} \ \frac{2}{6})$$

and the transition matrix is

$$K = \begin{pmatrix} 11/36 & 1/4 & 1/9 & 1/3 \\ 1/4 & 11/36 & 1/3 & 1/9 \\ 1/36 & 7/36 & 5/9 & 2/9 \\ 7/36 & 1/36 & 2/9 & 5/9 \end{pmatrix}$$

This is unstable under a generic small perturbation of the invariant distribution, as illustrated in Figure 3.

However if instead of $\pi_{1,2}$ we consider $\pi_{2,2}$ for the same rank matrix (13), then by Theoremj 4 we have same invariant distribution $\theta = (\frac{1}{6} \ \frac{1}{6} \ \frac{2}{6} \ \frac{2}{6})$, but if we take an initial distribution of the form $(\frac{1}{6} + \varepsilon \ \frac{1}{6} + \varepsilon \ \frac{2}{6} - \varepsilon \ \frac{2}{6} - \varepsilon)$ then iterates do converge to the fixed point. This example seems to depend on a certain “partial symmetry” property of the rank matrix which is copied to the perturbed initial distribution. We do not know a precise formulation of the “partial symmetry” idea here. We avoid that issue as follows, formulating the strongest conjecture to which we do not have a counter-example.

Conjecture 5 *On any finite space S with distinct distances, for almost all initial distributions θ the iterative process $(\pi_{j,k}^n(\theta), n \geq 1)$ converges to a limit of the form δ_s or δ_{s_1, s_2} .*

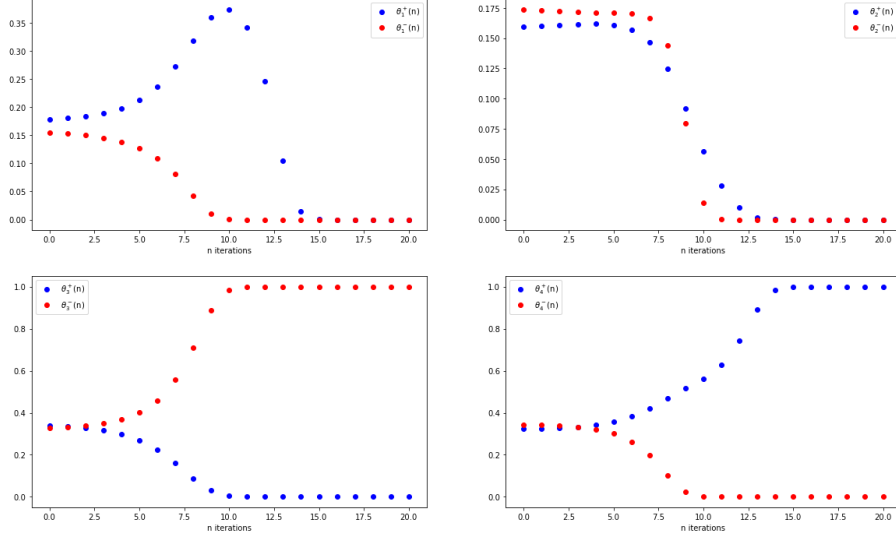


Figure 3: $|S| = 4$, rank matrix R at (13), $\pi_{1,2}$. Unstable behavior of iterates $\theta_i(n)$, $n = 0, 1, 2, \dots, 20$ in panels $i = 1, 2, 3, 4$, starting from two different initial distributions θ^+, θ^- near the fixed point.

Here *almost all* is with respect to the natural uniform measure on the simplex of probability distributions on S . This would imply that our original plan for finding invariant distributions, by iterating from some haphazardly-chosen initial distribution, would not be effective in finding **all** invariant distributions.

5.2 Uniform invariant distributions

We noted earlier that strong symmetry conditions on a compact metric space S would be sufficient to imply that the uniform distribution was invariant. For finite S , it is clearly sufficient that the rank matrix R is a Latin square. To fit our context we need the rank matrix to be feasible, so we mention the following result: we do not know if it is new.

Lemma 6 *If a rank matrix R is a Latin square, then R is feasible if and only if R is a symmetric matrix.*

5.3 Finite sets in Euclidean space

As another exploratory example of finite S , Figure 4 shows results of simulations for a 9-point set. The points are located as a perturbed 3×3 pattern,

marked as $\bullet$ in the Figure, with Euclidean distance in the plane. We fix $k = 10$ and consider each $j = 1, 2, \dots, 10$. Results for three different initial distributions are shown, indicated by the three different colors. The colored vertical lines in each square indicate the initial probabilities at the associated point $\bullet$. The results show behavior analogous to the previous settings, in that the “fixed point” limit of the iterative procedure is either a unit mass δ_s or $\delta_{s_1, s_2} = \frac{1}{2}(\delta_{s_1} + \delta_{s_2})$. In Figure 4, a δ_s limit for the j 'th nearest process is represented by a colored j in square s , or a limit δ_{s_1, s_2} is represented by a colored (j) in both square s_1 and square s_2 . For instance, for $j = 8$ the limit is (for each initial distribution color) uniform on some two points: top left and top right for blue, bottom left and middle right for red, and top center and bottom left for brown.

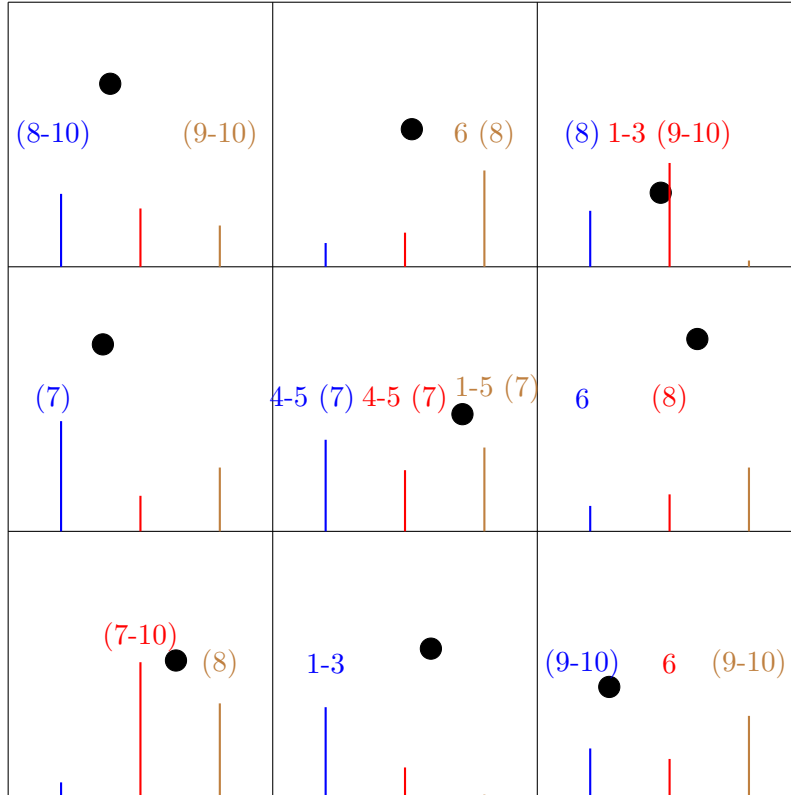


Figure 4: A 9-point set in the plane. See text for explanation.

What we observe in these results (in each of 11 trials from different initial

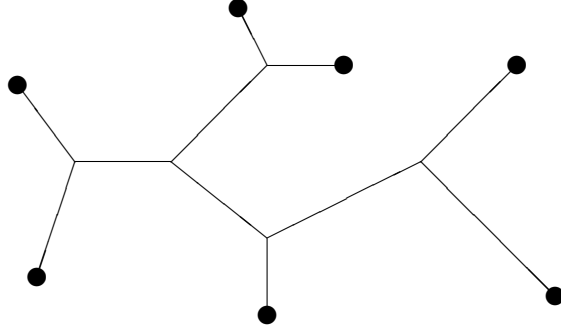


Figure 5: A BTL space S with $|S| = 7$.

distributions, extending the 3 trials shown in the figure) is that

(*) for $j \leq 6$ the limit is always some δ_s whereas for $j \geq 7$ the limit is always some δ_{s_1, s_2} . But the precise limit – which s or s_1, s_2 – depends both on j and the initial distribution.

Remarkably, simulations for an analogous perturbed 5×5 pattern S show exactly the same behavior described in (*), and so do simulations for an analogous perturbed $3 \times 3 \times 3$ pattern in three dimensions.

Here is one conjecture.

Conjecture 7 *For finite S embedded in the plane, with Euclidean distance, consider $\pi_{1,2}$. Then for almost all initial distributions, the limit of the iterative process is some δ_s .*

Indeed this might be true on every compact metric space S , for some appropriate notion of “almost all”.

6 A discrete tree space

In [1] we study binary tree leaves (BTL), illustrated in Figure 5, as a class of finite spaces. Here S is the finite set of leaves; the edges have lengths which serve to determine the 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

$$\text{all distances } (d(s_i, s_j), j \neq i) \text{ are distinct.} \quad (14)$$

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 8 ([1] Theorem 7) : *On a BTL space S with $|S| \geq 3$ and satisfying (14), and for $k = 2$, there are no invariant measures with full support.*

As with the analogous result below on the unit interval, the proof is a “proof by contradiction” depending on the specific space S , which does not suggest possible general proofs.

7 The unit interval

Here we take S to be the unit interval $[0, 1]$. We will show simulations and discuss what they suggest regarding convergence to fixed points. These are Monte Carlo simulations: a distribution is represented as a population of (typically 500,000) sample points. In the Figures, densities are drawn via Gaussian KDE, which becomes inaccurate for sharply peaked distributions, and also tends to be inaccurate at endpoints.

7.1 Initial distribution $U[0, 1]$.

We first study, by simulation, iterates starting from the uniform distribution $U[0, 1]$. Because $U[0, 1]$ is symmetric about $1/2$, all iterates must be symmetric about $1/2$; in section 7.2 we will study non-symmetric initial distributions, to investigate whether symmetry might be forcing some non-generic behavior.

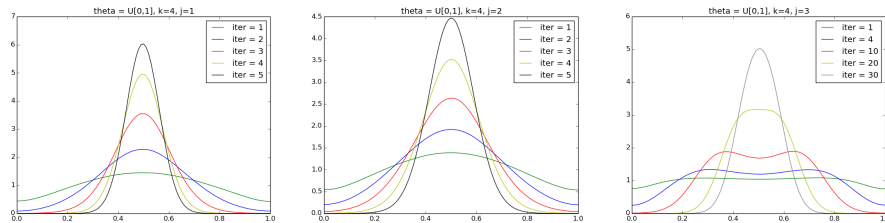


Figure 6: Iterates from $U[0, 1]$; $k = 4, j = 1, 2, 3$.

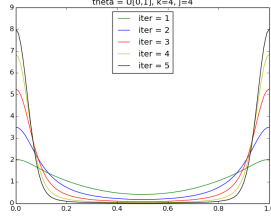


Figure 7: Iterates from $U[0, 1]$; $k = 4, j = 4$

First let us show the case $k = 4$. Figure 6 shows the first few iterates of $U[0, 1]$ for $j = 1, 2, 3$. Note that, here and throughout, the vertical scale and the numbers of iterations shown may not be the same from one panel to the next. What we see strongly suggests that the iterates are converging, quickly for $j = 1$ but rather slowly for $j = 3$, toward the degenerate distribution $\delta_{1/2}$. This is strongly supported by examining the standard deviations of the iterates, shown on log scale in Figure 10, which will be discussed below. In contrast, Figure 7 for $j = 4$ strongly suggests that the iterates are converging quickly toward the mixture $\delta_{0,1}$. These two “extreme” behaviors – convergence to $\delta_{1/2}$ for smaller j or to $\delta_{0,1}$ for larger j – appear to hold for all k . Figure 8 shows simulations for $k = 6$, and Table 2 shows which behavior appears to hold in simulations for each pair (j, k) with $k \leq 9$.

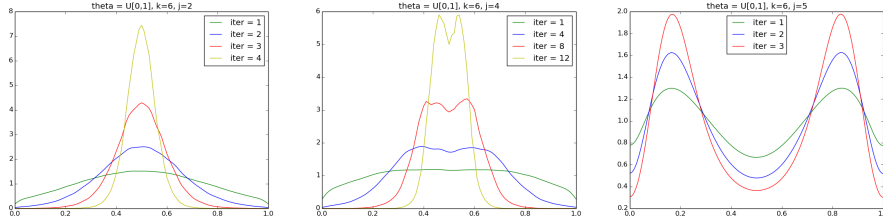


Figure 8: Iterates from $U[0, 1]$; $k = 6, j = 2, 4, 5$.

The pattern in Table 2 is reminiscent of the pattern seen earlier for $S = \{a, b\}$, but one must be careful about the comparison. In Figure 2 for $S = \{a, b\}$ each $\bullet$ represents a probability distribution, whereas in Figures 6 - 8 for $S = [0, 1]$ a probability distribution is represented by the estimated density curve. Figure 2 shows iterates of the same $\pi_{j,k}$ from different initial distributions, all on the same graphic. But for $S = [0, 1]$ we will show

Table 2: Conjectured limits of iterates from $U[0, 1]$; the values of j with each type of behavior.

k	$\rightarrow \delta_{1/2}$	$\rightarrow \delta_{0,1}$
2	1	2
3	[1, 2]	3
4	[1 - 3]	4
5	[1 - 4]	5
6	[1 - 4]	[5, 6]
7	[1 - 5]	[6, 7]
8	[1 - 6]	[7, 8]
9	[1 - 7]	[8, 9]

different graphics for the different initial distributions – that is, we should compare Figures 6 - 7 with Figures 12 - 13 below.

Are there any cases on $[0, 1]$ where the iterations of $\pi_{j,k}$ from the uniform initial distribution converge to a limit other than $\delta_{1/2}$ or $\delta_{0,1}$? Consider the case $k = 9, j = 7$ in Figure 9 (the case $k = 6, j = 5$ in Figure 8 is similar). Once two tight peaks are formed, one can see heuristically how the Markov chain behaves. Suppose s is near the left peak. Amongst the 9 samples, 4 or 5 will typically be near the right peak, in which case the 7'th closest to s will be the 2nd of 4 or the 3rd of 5, So on average the chain jumps to slightly to the left of center of the right peak. This is consistent with what we observe in Figure 9: the peaks grow in height and move slowly in the direction of $1/2$. It may be possible, in a case like this, that the iterates converge to a limit of the form $\delta_{s,1-s}$ for $0 < s < 1/2$ rather than to $\delta_{1/2}$. We cannot decide convincingly from our simulations. Thus heuristic may be more widely applicable – see section 10.

Simulations also suggest a scaling limit. Figure 10 shows, for some (j, k) for which the iterates converge to $\delta_{1/2}$ (others are similar), a remarkably precise geometric decrease in the s.d. as a function of the number of iterations, after the first few iterations. This strongly suggests a certain form of asymptotic self-similarity under scaling: that there exists a mean-zero distribution, say $\text{dist}(\xi_{j,k})$, on $\mathbb{R}$ and a constant $0 < c_{j,k} < 1$ such that $\pi_{j,k}[\text{dist}(\xi_{j,k})] = \text{dist}(c_{j,k}\xi_{j,k})$. And that $\text{dist}(\xi_{j,k})$ is the scaling limit of the iterates, and $c_{j,k}$ the geometric rate constant. Figure 11 shows renormalized (by mean and s.d.) iterates approximating $\text{dist}(\xi_{2,4})$.

One can imagine analogous scaling limits near 0 and 1 for pairs (j, k) for

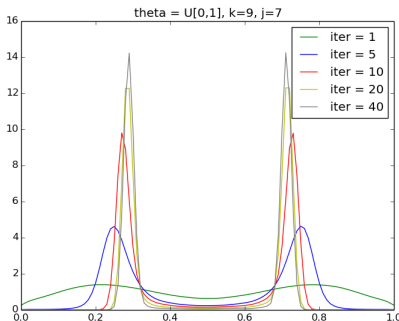


Figure 9: Iterates from $U[0, 1]$; $k = 9, j = 7$.

which the iterates converge to $\delta_{0,1}$.

7.2 Non-uniform initial distributions.

Some aspects of the observed behavior above might be consequences of symmetry (about $1/2$) of the initial distribution $U[0, 1]$. So here we study non-symmetric initial distributions.

First consider the slightly “tilted” distribution $T[0, 1]$ with density $f(u) = \frac{1}{2} + u$. We find that the behavior remains qualitatively unchanged; compare Figure 12 below with Figures 6 and 7 above. Convergence to $\delta_{1/2}$ is replaced by convergence to some δ_s where s depends on (j, k) . Convergence to $\delta_{0,1}$ is unchanged: Figure 12 shows that asymmetry persists through at least 5 iterations, but the probability masses near each endpoint tend to $1/2$ in the limit, consistent with the result that (for this case $k = 4$) the only invariant distribution supported on $\{0, 1\}$ is the uniform distribution. Figure 13 shows similar behavior for the “more tilted” distribution $TT[0, 1]$ with density $f(u) = 2u$.

7.3 Rigorous results for $S = [0, 1]$

The previous sections stated “results” about $\pi_{j,k}$ which we confidently believe via simulations, but without proofs. At a rigorous level, the key open questions are

- Does there exist (for any (j, k)) any invariant distribution with full support?

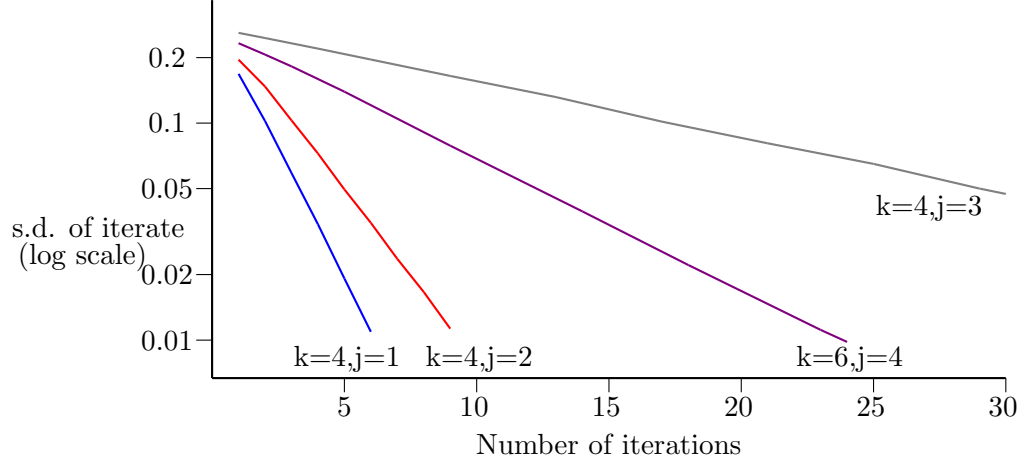


Figure 10: The geometric decrease in s.d. for some $U[0, 1]$ models. The lines pass through the actual simulated values, without being fitted.

- Does there exist (for any (j, k)) any distribution other than δ_s or $\delta_{0,1}$ that occurs as a limit of iterates from some initial distribution with full support?

We suspect the answer to each is “no”. Of course, “no” to the second question would imply “no” to the first question.

A natural way to start a study is to write down the identity for an invariant density $f_{j,k}(x)$. But because we are seeking to prove non-existence, we seek to prove there is no solution. Our partial results involve considering the density at endpoints; the proof in [1] of Theorem 9 was constructed by first considering the density case and then re-writing the argument for general distributions.

Let us discuss in detail only the simplest case, that is $k = 2$. Recall from Theorem 4 that the invariant distributions for $\pi_{2,2}$ and $\pi_{1,2}$ are the same.

Theorem 9 ([1] Theorem 4) : There are no $\pi_{1,2}$ or $\pi_{2,2}$ -invariant distributions on the interval $[0, 1]$ other than the omnipresent ones.

The proof consists of analyzing the equation for an invariant density $f(x)$ to obtain a contradiction at $f(0)$. A more elaborate argument (omitted) extends the result to some other values of (j, k) , as follows.

There is no $\pi_{j,k}$ -invariant density on $[0, 1]$ if $j = k$ or if

$$\binom{k}{j-1 \quad 1 \quad k-j} \frac{(j-1)^{j-1}(k-j)^{k-j}}{(k-1)^{k-1}} \leq 2.$$

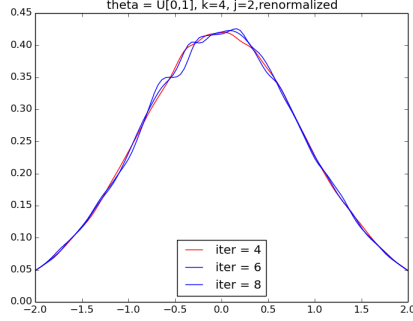


Figure 11: Renormalized iterates from $U[0, 1]$; $k = 4, j = 2$ strongly suggest a scaling limit.

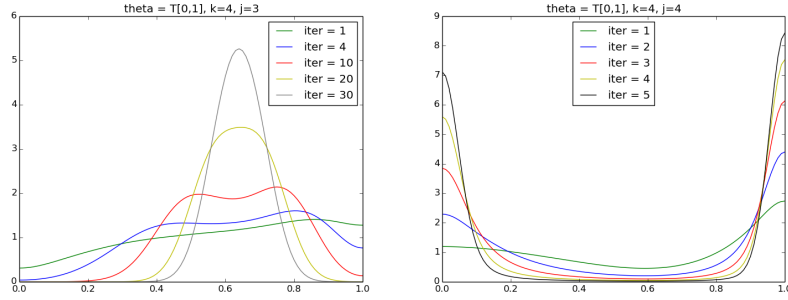


Figure 12: Iterates on the unit interval from a slightly tilted initial distribution, $k = 4, j = 3, 4$.

7.4 Discrete points within $[0, 1]$

Somewhat relevant to the open problems at the start of section 7.3 is the following discrete example.

Proposition 10 *The uniform distribution on the 4 points $0, 0.4, 0.6, 1$ is invariant for $\pi_{3,4}$.*

This is true by calculation: the transition matrix is

$$\frac{1}{256} \begin{pmatrix} 13 & 67 & 109 & 67 \\ 109 & 13 & 67 & 67 \\ 67 & 67 & 13 & 109 \\ 67 & 109 & 67 & 13 \end{pmatrix}$$

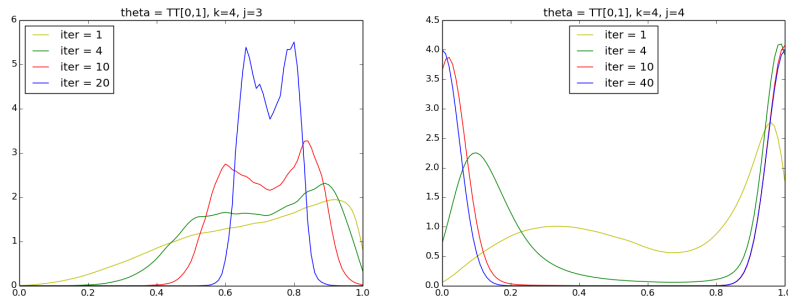


Figure 13: Iterates on the unit interval from a more tilted initial distribution, $k = 4, j = 3, 4$.

which is doubly-stochastic (like a variant of a Latin square). But it is not clear if any analog holds for larger point sets.

8 The circle

Here we take S to be the unit circle $\mathbb{C} = [0, 1)$.

8.1 The uniform distribution

On the circle, the uniform distribution $U[\mathbb{C}]$ is invariant, by symmetry. However, when we do Monte Carlo simulations of iterates, the symmetry breaks, as shown in Figure 14. We interpret this as implying that the uniform distribution is an *unstable* fixed point.

In detail, the figure shows two realizations of each of the four cases ($j = 1, 2, 3, 4$) with $k = 4$. For each j , different realizations are qualitatively similar. The symmetry is broken by the emergence of “waves” with some number $\nu = \nu(j, k)$ of peaks evenly spread around the circle. The time (number of iterations) of emergence is similar in different realizations. In simulations one wave eventually one takes over, as shown in Figure 15.

Implications for the true iterative process are elusive. Recall that by symmetry (section 3.1) the uniform distribution on any number ν of points evenly spread around the circle (say θ_ν) is invariant. If a non-uniform initial distribution has some symmetry property, such as invariant under a $1/\nu$ ’th fraction of one rotation³, then by the general “preservation of symmetry” result (Lemma 3) that symmetry property persists in the iterations and hence

³For the maximum such ν .

in any limit. So it is natural to conjecture that the iterative process started from such a distribution will converge to θ_ν . However, the significance for the iterative process of the numbers $\nu(j, k)$ observed in initial iterates in simulations is elusive.

8.2 Non-uniform initial distributions

We have only done brief investigations of iterates on $\mathbb{C}$ from non-symmetric starting distributions. It appears that behavior here corresponds to the case of the unit interval, in that limits are either δ_s or δ_{s_1, s_2} where s_1 and s_2 are opposite points of $\mathbb{C}$. Figure 16 shows the case of the discontinuous initial density

$$f(t) = t + 0.5, 0 < t < 1. \quad (15)$$

9 A high-dimensional example

We wish to study a high-dimensional example without any special symmetry. Here is our rather arbitrary choice.

Take $S = [0, 1]^{10}$ with elements $\mathbf{x} = (x_i)$ and with non-Euclidean metric

$$d(\mathbf{x}, \mathbf{y}) = \sum_i \beta^i |x_i - y_i| \quad (16)$$

for a parameter $0 < \beta < 1$. And take initial distributions with density of the form

$$f(\mathbf{x}) \propto \exp(-\alpha \sum_i x_i^2)$$

for a parameter $\alpha > 0$.

Figure 17 shows simulations with $k = 2$. What is shown is the distribution of “distance to $\mathbf{0}$ ” after varying numbers of iterations. The 8 panels show the 8 combinations of

$$\alpha = 0.1 \text{ or } 1.0; \beta = 0.7 \text{ or } 0.9; j = 1 \text{ or } 2.$$

The qualitative behavior is similar to previous examples: for $j = 1$ there is apparent convergence to some $\delta_{\mathbf{x}}$ and for $j = 2$ we suspect there is convergence to $\delta_{\mathbf{0}, \mathbf{1}}$.

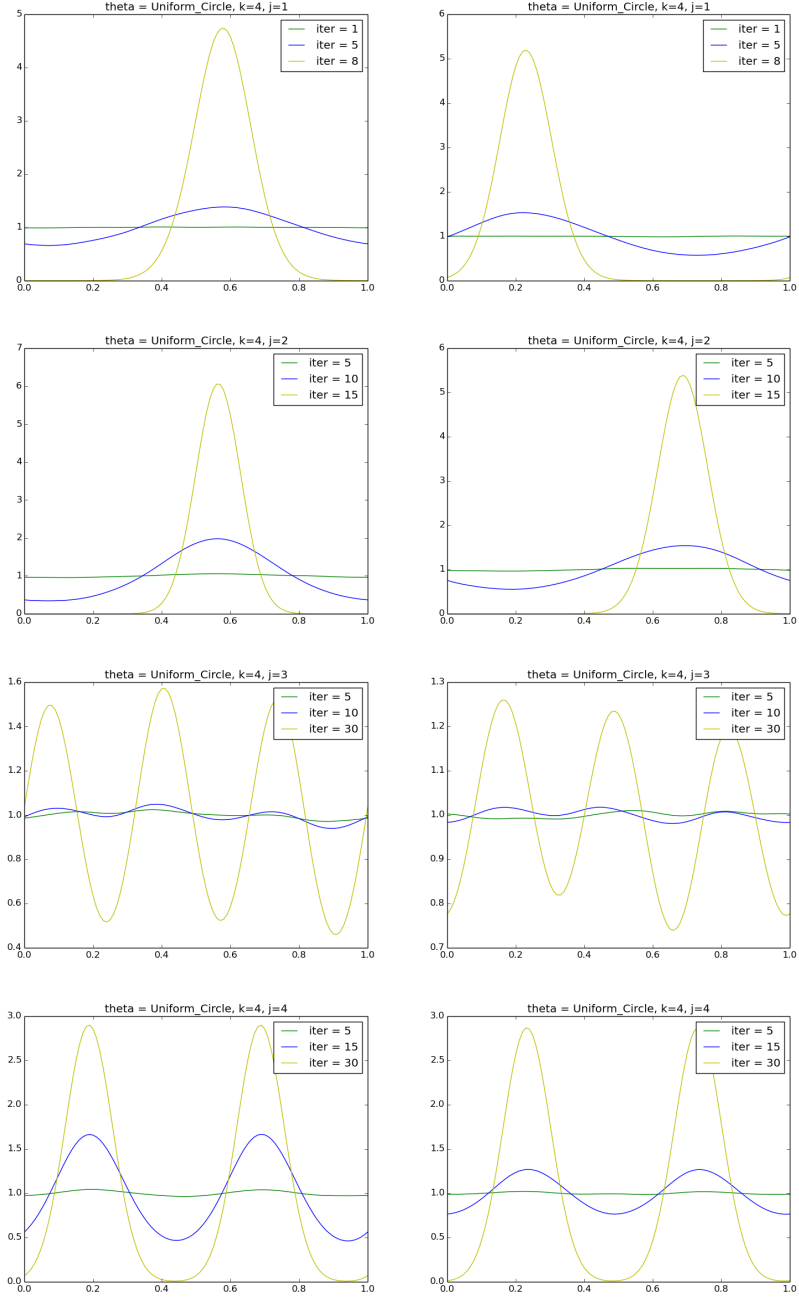


Figure 14: Iterates for $U[\mathbb{C}]$; $k=4, j=1, 2, 3, 4$

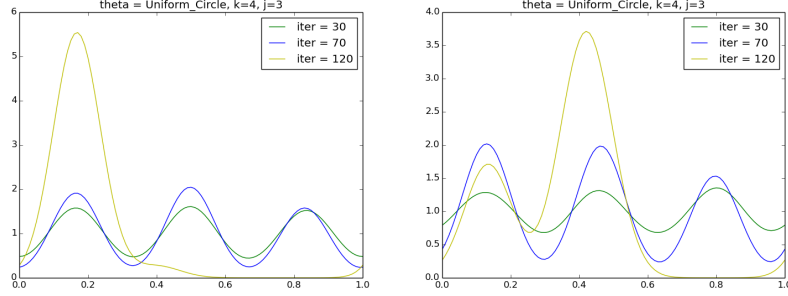


Figure 15: Iterates for $U[\mathbb{C}]$, long-term: $k = 4, j = 3$.

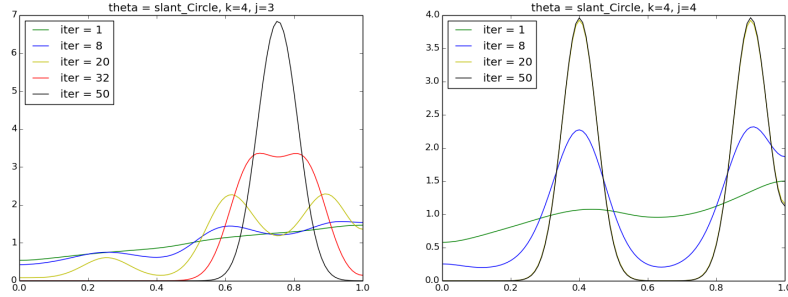


Figure 16: Iterates on $\mathbb{C}$ from initial density (15), $k = 4, j = 3, 4$.

10 Final remarks

1: The $j \approx 3k/4$ heuristic. The heuristics in section 7.1 in the context of the unit interval can be repeated more generally. Suppose k is large, and suppose S and (j, k) are such that there is an initial tendency for the iterate distributions to concentrate near two distinct regions. Then about half of the k samples will be in each region, so for $j > k/2$ the chain will switch back and forth between the regions. And the average distance between regions in successive iterations will tend to increase or decrease according to whether the chosen point at each step of the chain is farther (from the other region) or nearer than the average, that is according to whether $j > 3k/4$ or $j < 3k/4$.

2. The only rigorous analyses of continuous S that we have completed involve explicit calculations. To give rigorous proofs of, for instance, convergence of iterates to some δ_s , it seems natural to investigate some kind of monotonicity property for concentration of distributions, but we have been

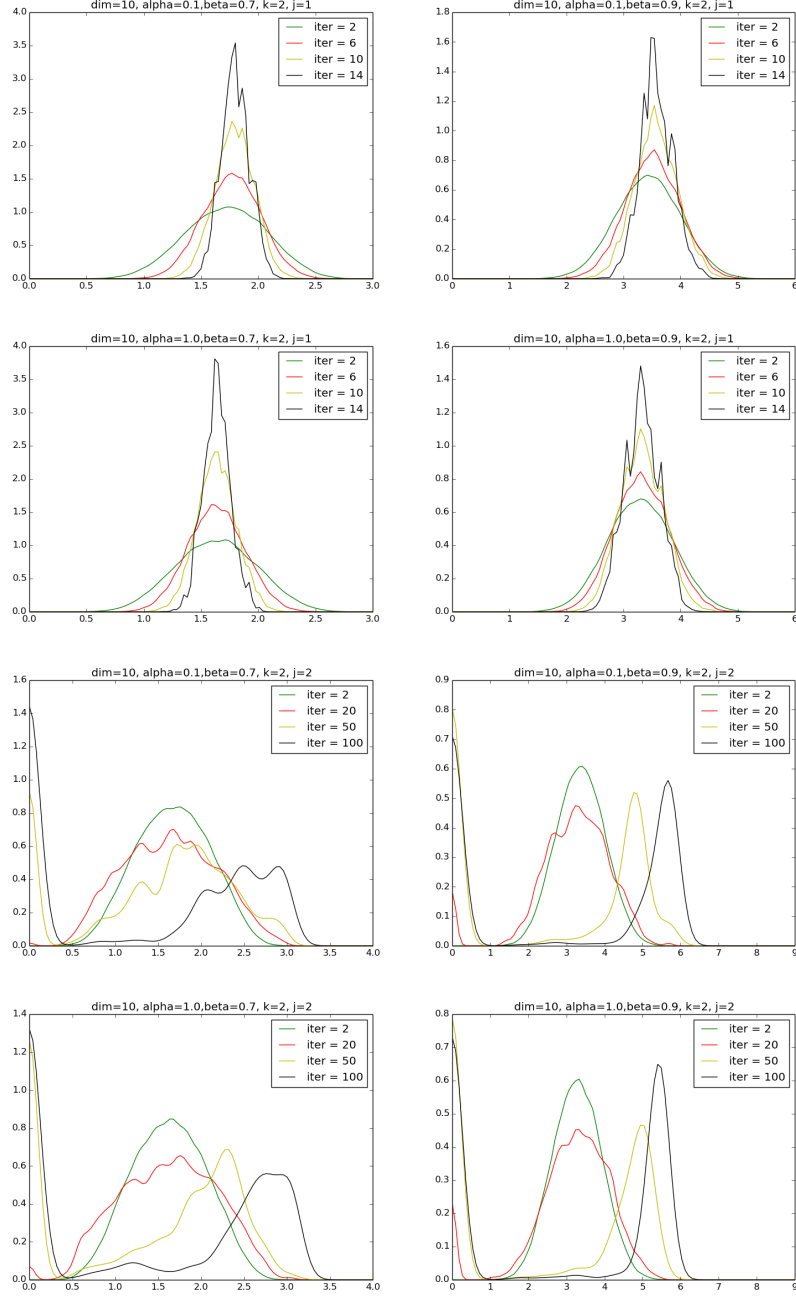


Figure 17: The 10-dimensional example (16); $k = 2$.

unable to implement that idea.

3. This article studies a certain process which can be defined on any compact metric space S , parametrized by any $\theta \in \mathcal{P}(S)$. Are there other interesting such processes? As one example, a certain such *coverage process* is studied in [2]: “seeds” arrive as a Poisson process in time at i.i.d. (θ) locations and form the center of growing balls.

References

- [1] David J. Aldous and Shi Feng. Markov chains and mappings of distributions on compact spaces, 2024. Unpublished.
- [2] D.J. Aldous. Covering a compact space by fixed-radius or growing random balls. *ALEA Lat. Am. J. Probab. Math. Stat.*, 19:755–767, 2022.
- [3] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. *Markov chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018.
- [4] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.