ON THE ASYMPTOTIC BEHAVIOR OF THE DIACONIS–FREEDMAN CHAIN IN A MULTI-DIMENSIONAL SIMPLEX

MARC PEIGNÉ,* *Institut Denis Poisson, Université de Tours*

TAT DAT TRAN,** *** *Max-Planck-Institut für Mathematik in den Naturwissenschaften; Mathematisches Institut, Universität Leipzig*

Abstract

We give a setting of the Diaconis–Freedman chain in a multi-dimensional simplex and consider its asymptotic behavior. By using techniques from random iterated function theory and quasi-compact operator theory, we first give some sufficient conditions which ensure the existence and uniqueness of an invariant probability measure and, in particular cases, explicit formulas for the invariant probability density. Moreover, we completely classify all behaviors of this chain in dimension two. Some other settings of the chain are also discussed.

Keywords: Iterated functions systems; quasi-compact linear operators; absorbing compact set; invariant probability measure; invariant probability density

2020 Mathematics Subject Classification: Primary 60J05
Secondary 60F05

1. Introduction

The main motivation for this article is to propose a general setting for the so-called Diaconis–Freedman chain in $\mathbb{R}^d$, $d \geq 2$. First, we give the most natural setting of this chain on a d-dimensional simplex and consider its asymptotic behavior by using techniques from random iterated function theory and quasi-compact operator theory (see [20] in dimension one). We have recently learned that this multi-dimensional setting is also considered in [23], where the authors used another approach and only considered the cases of ergodicity. We also discuss here some other possible extensions, with several invariant probability measures.

Markov chains generated by compositions of independent random iterated functions have been the object of numerous works for more than 60 years. We refer to [5, 19] for the first models designed for analyzing data in learning, and to [8, 11, 21, 22, 31] and references therein; see also [26, 27] for such processes with weak contraction assumptions on the random functions involved.

In [7], Diaconis and Freedman focused on the Markov chain $(Z_n)_{n \geq 0}$ on [0, 1] randomly generated by the two families of maps $\mathcal{H} := \{h_t \colon [0, 1] \to [0, 1], x \mapsto tx\}_{t \in [0,1]}$ and

$\mathcal{A} := \{a_t \colon [0, 1] \to [0, 1], x \mapsto tx + 1 - t\}_{t \in [0,1]}$; at each step, a map is randomly chosen with probability p in the set $\mathcal{H}$ and $q = 1 - p$ in the set $\mathcal{A}$, then uniformly with respect to the parameter $t \in [0, 1]$. They also considered the case when the weights p and $q = 1 - p$ may vary with respect to the position x. When p is constant, the random maps (see Section 3 for a detailed introduction) which control the transitions of $(Z_n)_{n \geq 0}$ are independent and identically distributed (i.i.d.), otherwise the process is no longer in the framework of compositions of independent random functions. This class of such processes has been studied for a few decades, with various assumptions put on the state space (e.g. compactness) and the regularity of the weight functions. We refer, for instance to [17] at the beginning of the 1980s, [1, 2, 3] with connections to image encoding a few years later, and [18] more recently. All these works concerned sufficient conditions for the unicity of the invariant measure and did not explore the case when there are several invariant measures. As far as we know, the coupling method does not seem to be relevant to studying this type of Markov chain when there are further invariant measures, or, equivalently, when the space of harmonic functions is not reduced to constant.

For the Diaconis–Freedman chain in dimension 1, a systematic approach was developed in [20] based on the theory of quasi-compact operators (also described in [13, 25]); the authors completely described the peripheral spectrum of the transition operator P of $(Z_n)_{n \geq 0}$ and used a precise control of the action of the family of functions generated by the sets $\mathcal{H}$ and $\mathcal{A}$. The main tool for carrying out this study is to describe the family of compact subsets of Δ_d invariant under the action of elements of the set of functions $\mathcal{H}$ and $\mathcal{A}$, according to the varying weight functions. Some compositions of maps are forbidden when these weights vanish on Δ_d; the main consequence is the emergence of several invariant sets (see Section 5), which induces the existence of several harmonic functions, then several invariant probability measures, for the chain $(Z_n)_{n \geq 0}$. This mechanism was studied in [14], with deep applications to transfer operators on the interval $[0, 1]$, corresponding to the maps $x \mapsto x/2$ and $x \mapsto (x + 1)/2$; these operators occur naturally in wavelet theory to characterize scaling filters used in multi-resolution analysis.

This model is also known in the literature as a 'stick-breaking process' or a 'stochastic give-and-take', with applications in other fields such as human genetics, robot coverage, random search, etc. We refer to [6] for more explanations in population genetics.

On a side note, instead of 'robot control' purposes, we mention a fairly rare scope of application in mathematics: the illustration of a paradox in philosophy in the conception of free will, through the sad and well-known story of Buridan's donkey:

> An equally hungry and thirsty donkey is placed just in the middle between a stack of grass and a bucket of water and hesitates to choose between the grass and the water; once this choice is done, possibly randomly, the donkey has to move towards the selected site, once again either in a deterministic or a random procedure to be defined.

The Diaconis–Freedman chain in dimension 1 is considered as a model of the behavior of Buridan's donkey: we refer to [28] for the various (deterministic or random) models describing the story and some partial results, and to [24] for generalizations of the Diaconis–Freedman chain in such a way that the stationary distribution is always of beta type. Theorem 3.1 in [20] offers a strategy for Buridan's donkey to reach either the stack of grass or the bucket of water: it is necessary and sufficient to choose weight functions p and q which are not constant, with suitable boundary conditions! By contrast, constant weights lead to erratic behavior for ever between the two final expected positions (the 'stack of grass' or the 'bucket

of water')! The reader may easily imagine some variations of this story with several stacks of food instead of the unique stack of grass, naturally modeled by the d-dimensional chain studied here.

The multi-dimensional setting for such problems has not been touched; it is our aim to introduce and analyze it here, and the Diaconis–Freedman chain appears as a rich 'toy model' for this extension.

The paper is organized as follows. In Section 2 we give our setting of the Diaconis–Freedman chain in dimension $d \geq 2$. Some properties of the transition operator and its dual operator are considered, and the uniqueness of the stationary density function is shown (Corollary 2.1). In Section 3 we give some results on uniqueness of invariant measures (Theorems 3.2 and 3.4) based on iterated function systems theory. Some special cases with explicit formulas for the unique invariant density are considered in Section 4. Section 5 contains our main result (Theorem 5.2): we classify the invariant probability measures and consider the asymptotic behavior of $(Z_n)_{n \geq 0}$. We discuss some future research directions in Section 6.

2. The Diaconis–Freedman chain in dimensions ≥ 2

In this section we consider a particular setting for the multi-dimensional problem of the Diaconis–Freedman chain; there are many ways to set it based on different application models (see, for instance, [28]). Our setting here is fit for robot control applications.

Denote by $\Delta_d := \{\mathbf{x} = (x_i)_{1 \leq i \leq d} \in \mathbb{R}_{\geq 0}^d : |\mathbf{x}| = x_1 + \cdots + x_d \leq 1\} = co\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_d\}$ a closed d-dimensional simplex with vertices $\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_d$, where $\mathbf{e}_0 = (0, \ldots, 0)$ and $\mathbf{e}_i = (0, \ldots, \underset{i\text{th}}{1}, \ldots, 0)$ for $1 \leq i \leq d$. From now on, for any $\mathbf{x} \in \Delta_d$, we set $x_0 = 1 - |\mathbf{x}|$; we have that $\mathbf{x} = x_0\mathbf{e}_0 + x_1\mathbf{e}_1 + \cdots + x_d\mathbf{e}_d$ with $x_i \geq 0$ and $x_0 + \cdots + x_d = 1$.

We consider the Markov chain $(Z_n)_{n \geq 0}$ on the simplex Δ_d corresponding to the successive positions of the robot, according to the following rules:

- The robot is put randomly at a point Z_0 in Δ_d.

- If, at time $n \geq 0$, it is located at $Z_n = \mathbf{x} \in \Delta_d$, then it chooses the vertex $\mathbf{e}_i$, $0 \leq i \leq d$, with probability $p_i(\mathbf{x})$ for the next moving direction and uniformly randomly moves to some point on the open line segment $(\mathbf{x}, \mathbf{e}_i) := \{t\mathbf{x} + (1 - t)\mathbf{e}_i \mid t \in (0, 1)\}$.

We assume that the functions p_i, $0 \leq i \leq d$, are continuous and non-negative on Δ_d and satisfy $\sum_{i=0}^n p_i(\mathbf{x}) = 1$ for any $\mathbf{x} \in \Delta_d$.

Let us make this description more rigorous. For any $i = 0, \ldots, d$ and $\mathbf{x} \in \Delta_d$, denote by $U_i(\mathbf{x}, \cdot)$ the uniform distribution on $(\mathbf{x}, \mathbf{e}_i)$; it is defined on open intervals $(\mathbf{y}_1, \mathbf{y}_2) \subset (\mathbf{x}, \mathbf{e}_i)$ as $U_i(\mathbf{x}, (\mathbf{y}_1, \mathbf{y}_2)) := |t(\mathbf{x}, \mathbf{y}_2, \mathbf{e}_i) - t(\mathbf{x}, \mathbf{y}_1, \mathbf{e}_i)|$, where the real number $t = t(\mathbf{x}, \mathbf{y}, \mathbf{e}_i) \in (0, 1)$ solves the equality $\mathbf{y} = t\mathbf{x} + (1 - t)\mathbf{e}_i$. The *one-step transition probability function P* of $(Z_n)_{n \geq 0}$ is

$$P(\mathbf{x}, d\mathbf{y}) = \sum_{i=0}^d p_i(\mathbf{x})U_i(\mathbf{x}, d\mathbf{y}), \qquad \mathbf{x} \in \Delta_d. \tag{2.1}$$

We illustrate this setting in Δ_2 in Figure 1.

We want to classify the invariant probability measures of the chain $(Z_n)_{n \geq 0}$ and to describe its behavior as $n \to +\infty$. Our approach is based on the description of the spectrum, on some suitable space (to be specified), of the operator corresponding to the one-step transition probability function P, also denoted by P. Let us first introduce this transition operator.
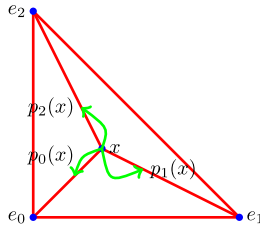
FIGURE 1. The Diaconis–Freedman chain in Δ_2.

We denote by d**x** the Lebesgue measure on Δ_d and by $\mathbb{L}^\infty(\Delta_d)$ the space of all bounded Lebesgue-measurable functions $f\colon \Delta_d \to \mathbb{C}$; the dual space of $\mathbb{L}^\infty(\Delta_d)$ with respect to the Lebesgue measure is $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$, the space of Lebesgue-measurable functions from Δ_d to $\mathbb{C}$ which are integrable with respect to d**x**. The spaces $\mathbb{L}^\infty(\Delta_d)$ and $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$ are Banach spaces, endowed respectively with the norms $\|f\|_\infty := \sup_{\mathbf{x}\in\Delta_d} |f(\mathbf{x})|$ and $\|g\|_1 := \int_{\Delta_d} |g(\mathbf{x})|\,\mathrm{d}\mathbf{x}$.

We also denote by $Den(\Delta_d, \mathrm{d}\mathbf{x}) = \{g \in \mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x}) : g \geq 0 \text{ and } \int_{\Delta_d} g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1\}$ the space of all probability densities on Δ_d with respect to the reference Lebesgue measure d**x**. The set $(Den(\Delta_d, \mathrm{d}\mathbf{x}), d)$ is a complete metric space for the distance $d(f, g) := \|f - g\|_1$; furthermore, $Den(\Delta_d, \mathrm{d}\mathbf{x})$ is a non-empty closed convex subset of the Banach space $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$, and it contains the constant function $g(\mathbf{x}) \equiv d!$.

The transition operator of the chain $(Z_n)_{n\geq 0}$ is defined by

$$P\colon \mathbb{L}^\infty(\Delta_d) \to \mathbb{L}^\infty(\Delta_d), \qquad f \mapsto \left(Pf : \mathbf{x} \to \int_{\Delta_d} f(\mathbf{y})P(\mathbf{x}, \mathrm{d}\mathbf{y})\right).$$

Its dual operator $P^*\colon \mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x}) \to \mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$ with respect to the Lebesgue measure is defined by $\int_{\Delta_d} Pf(\mathbf{x})g(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\Delta_d} f(\mathbf{x})P^*g(\mathbf{x})\,\mathrm{d}\mathbf{x}$.

Let us be explicit about the form of these two operators.

Lemma 2.1. *Let P be the transition operator of $(Z_n)_{n\geq 0}$, acting on $\mathbb{L}^\infty(\Delta_d)$, and P^* its dual operator on $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$. Then, for any $f \in \mathbb{L}^\infty(\Delta_d)$ and $\mathbf{x} \in \Delta_d$,*

$$Pf(\mathbf{x}) = \sum_{i=0}^{d} p_i(\mathbf{x}) \int_0^1 f(t\mathbf{x} + (1 - t)\mathbf{e}_i)\,\mathrm{d}t, \tag{2.2}$$

and, for any $g \in \mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$ and $\mathbf{y} \in \Delta_d$,

$$P^*g(\mathbf{y}) = \sum_{i=0}^{d} \int_{1-y_i}^1 t^{-d} G_i\left(\frac{1}{t}\mathbf{y} + \left(1 - \frac{1}{t}\right)\mathbf{e}_i\right)\,\mathrm{d}t$$

$$= \sum_{i=0}^{d} \int_1^{\frac{1}{1-y_i}} s^{d-2} G_i\left(s\mathbf{y} + \left(1 - s\right)\mathbf{e}_i\right)\,\mathrm{d}s, \tag{2.3}$$

where $G_i(\mathbf{y}) = g(\mathbf{y})p_i(\mathbf{y})$.

Proof. The equality (2.1) yields

$$Pf(\mathbf{x}) = \int_{\Delta_d} f(\mathbf{y}) P(\mathbf{x}, \mathrm{d}\mathbf{y}) = \sum_{i=0}^{d} p_i(\mathbf{x}) \int_{(\mathbf{x}, \mathbf{e}_i)} f(\mathbf{y}) U_i(\mathbf{x}, \mathrm{d}\mathbf{y})$$

$$= \sum_{i=0}^{d} p_i(\mathbf{x}) \int_0^1 f(t\mathbf{x} + (1-t)\mathbf{e}_i) \, \mathrm{d}t.$$

For the computation of P^*, we assume $d = 2$; the same argument holds for any d. For all $f \in \mathbb{L}^\infty(\Delta_d)$ and $g \in \mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$,

$$\int_{\Delta_2} f(\mathbf{x}) P^* g(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \int_{\Delta_2} Pf(\mathbf{x}) g(\mathbf{x}) \, \mathrm{d}\mathbf{x}$$

$$= \sum_{i=0}^{2} \int_{\Delta_2} \left(p_i(\mathbf{x}) \int_0^1 f(t\mathbf{x} + (1-t)\mathbf{e}_i) \, \mathrm{d}t \right) g(\mathbf{x}) \, \mathrm{d}\mathbf{x}$$

$$= \sum_{i=0}^{2} \int_{\Delta_2} \left(G_i(\mathbf{x}) \int_0^1 f(t\mathbf{x} + (1-t)\mathbf{e}_i) \, \mathrm{d}t \right) \mathrm{d}\mathbf{x}.$$

Let us detail the computation of the term $\int_{\Delta_2} \left(G_0(\mathbf{x}) \int_0^1 f(t\mathbf{x}) \, \mathrm{d}t \right) \mathrm{d}\mathbf{x}$ which corresponds to $i = 0$; the same calculation holds for the other terms:

$$\int_{\Delta_2} \left(G_0(\mathbf{x}) \int_0^1 f(t\mathbf{x}) \, \mathrm{d}t \right) \mathrm{d}\mathbf{x} = \int_0^1 \left[\int_0^{1-x_1} \left(G_0(\mathbf{x}) \int_0^1 f(t\mathbf{x}) \, \mathrm{d}t \right) \mathrm{d}x_2 \right] \mathrm{d}x_1$$

$$= \int_0^1 \left[\int_0^1 \left(\int_0^{1-x_1} G_0(\mathbf{x}) f(t\mathbf{x}) \, \mathrm{d}x_2 \right) \mathrm{d}t \right] \mathrm{d}x_1$$

$$= \int_0^1 \left[\int_0^1 \left(\int_0^{1-x_1} G_0(\mathbf{x}) f(t\mathbf{x}) \, \mathrm{d}x_2 \right) \mathrm{d}x_1 \right] \mathrm{d}t$$

$$\overset{y_2 = tx_2}{=} \int_0^1 \left[\int_0^1 \left(\int_0^{(1-x_1)t} G_0\left(x_1, \frac{y_2}{t}\right) f(tx_1, y_2) \frac{\mathrm{d}y_2}{t} \right) \mathrm{d}x_1 \right] \mathrm{d}t$$

$$\overset{y_1 = tx_1}{=} \int_0^1 \left[\int_0^t \left(\int_0^{t-y_1} G_0\left(\frac{y_1}{t}, \frac{y_2}{t}\right) f(y_1, y_2) \frac{\mathrm{d}y_2}{t} \right) \frac{\mathrm{d}y_1}{t} \right] \mathrm{d}t$$

$$= \int_0^1 \left[\int_{y_1}^1 \left(\int_0^{t-y_1} \frac{1}{t^2} G_0\left(\frac{y_1}{t}, \frac{y_2}{t}\right) f(y_1, y_2) \, \mathrm{d}y_2 \right) \mathrm{d}t \right] \mathrm{d}y_1$$

$$= \int_0^1 \left[\int_0^{1-y_1} \left(\int_{y_1+y_2}^1 \frac{1}{t^2} G_0\left(\frac{y_1}{t}, \frac{y_2}{t}\right) f(y_1, y_2) \, \mathrm{d}t \right) \mathrm{d}y_2 \right] \mathrm{d}y_1$$

$$= \int_{\Delta_2} f(\mathbf{y}) \left(\int_{1-y_0}^1 t^{-2} G_0\left(\frac{1}{t}\mathbf{y}\right) \mathrm{d}t \right) \mathrm{d}\mathbf{y}.$$

Similarly,

$$\int_{\Delta_2} \left(G_i(\mathbf{x}) \int_0^1 f(t\mathbf{x} + (1-t)\mathbf{e}_i)\, dt \right) d\mathbf{x} = \int_1^{\frac{1}{1-y_i}} G_i\big(s\mathbf{y} + (1-s)\mathbf{e}_i\big)\, ds$$

for $i = 1, 2$, and (2.3) follows. $\qquad\square$

Remark 2.1. In dimension $d = 1$, Lemma 2.1 is still valid and corresponds to the expression for P^* given in [20]:

$$P^*g(y) = \int_{1-y}^1 t^{-1}G_1\left(\frac{1}{t}y + \left(1 - \frac{1}{t}\right)\right)dt + \int_y^1 t^{-1}G_0\left(\frac{1}{t}y\right)dt = \int_0^y \frac{G_1(s)}{1-s}ds + \int_y^1 \frac{G_0(s)}{s}ds.$$

Let us summarize some simple properties of P and P^*.

Proposition 2.1.

(a) *The operator P is a Markov operator on $\mathbb{L}^\infty(\Delta_d)$, i.e.*

  (i) *$Pf \geq 0$ whenever $f \in \mathbb{L}^\infty(\Delta_d)$ and $f \geq 0$;*

  (ii) *$P1 = 1$.*
  In particular, $\|Pf\|_\infty \leq \|f\|_\infty$ for any $f \in \mathbb{L}^\infty(\Delta_d)$. Furthermore, P is a Feller operator on Δ_d, i.e. $Pf \in C(\Delta_d)$ for all $f \in C(\Delta_d)$.

(c) *P^* acts on $\mathbb{L}^1(\Delta_d, d\mathbf{x})$ and, for any non-negative function $g \in \mathbb{L}^1(\Delta_d, d\mathbf{x})$, $P^*g \geq 0$ and $\|P^*g\|_1 = \|g\|_1$. Furthermore, P^* acts on $Den(\Delta_d, d\mathbf{x})$, i.e. $P^*: Den(\Delta_d, d\mathbf{x}) \to Den(\Delta_d, d\mathbf{x})$ and, for all $g_1 \neq g_2 \in Den(\Delta_d, d\mathbf{x})$,*

$$\|P^*g_1 - P^*g_2\|_1 < \|g_1 - g_2\|_1. \tag{2.4}$$

Proof. The properties of P are quite obvious; in particular, the fact that P is a Feller operator is easily checked from the representation (2.2) of P. Similarly, the first properties of P^* follow from the definition.

To establish (2.4), we fix $g_1 \neq g_2 \in Den(\Delta_d, d\mathbf{x})$ and set $h = g_1 - g_2$. We first note from (2.3) that $|P^*h| \leq P^*|h|$ on Δ_d, which implies $\|P^*g_1 - P^*g_2\|_1 = \|P^*h\|_1 \leq \|P^*|h|\|_1 = \|h\|_1 = \|g_1 - g_2\|_1$. In fact, the inequality above is strict because both functions g_1 and g_2 belong to $Den(\Delta_d, d\mathbf{x})$. Ad absurdum, let us assume $\|P^*h\|_1 = \|h\|_1$. This implies $\|P^*h\|_1 = \|P^*|h|\|_1$, and hence $|P^*h(\mathbf{x})| = P^*(|h|)(\mathbf{x})$ Lebesgue almost surely (a.s.), since $|P^*h| \leq P^*(|h|)$ on Δ_d. In other words, for Lebesgue almost all $\mathbf{x} \in \Delta_d$,

$$\left| \sum_{i=0}^d \int_1^{\frac{1}{1-x_i}} s^{d-2} p_i(s\mathbf{x} + (1-s)\mathbf{e}_i) h(s\mathbf{x} + (1-s)\mathbf{e}_i) \right|$$

$$= \sum_{i=0}^d \int_1^{\frac{1}{1-x_i}} s^{d-2} p_i(s\mathbf{x} + (1-s)\mathbf{e}_i) |h(s\mathbf{x} + (1-s)\mathbf{e}_i)|.$$

Consequently, for $i = 0, \ldots, d$, the quantities

$$\int_1^{\frac{1}{1-x_i}} s^{d-2} p_i(s\mathbf{x} + (1-s)\mathbf{e}_i) h(s\mathbf{x} + (1-s)\mathbf{e}_i)$$

all have the same sign and

$$\left| \int_1^{\frac{1}{1-x_i}} s^{d-2} p_i(s\mathbf{x} + (1-s)\mathbf{e}_i) h(s\mathbf{x} + (1-s)\mathbf{e}_i) \right|$$

$$= \int_1^{\frac{1}{1-x_i}} s^{d-2} p_i(s\mathbf{x} + (1-s)\mathbf{e}_i) |h(s\mathbf{x} + (1-s)\mathbf{e}_i)|.$$

In other words, for Lebesgue almost all $\mathbf{x} \in \Delta_d$, setting $\mathbf{x}_i := \frac{1}{1-x_i}\mathbf{x} + \left(1 - \frac{1}{1-x_i}\right)\mathbf{e}_i$ for $i = 0, \ldots, d$, the functions $p_0 h, \ldots, p_d h$, restricted respectively to $[\mathbf{x}, \mathbf{x}_0], \ldots, [\mathbf{x}, \mathbf{x}_d]$, have the same sign. Hence, the functions $hp_0, \ldots, hp_d$ have (Lebesgue a.s.) the same sign on Δ_d. Eventually, the equality $p_1 + \cdots + p_d = 1$ readily implies that the sign of h is Lebesgue a.s. constant on Δ_d. This contradicts the equality $h = g_1 - g_2$ with $g_1, g_2 \in Den(\Delta_d, \mathrm{d}\mathbf{x})$ and $g_1 \neq g_2$. □

As a direct consequence of (2.4), we may state the following corollary.

Corollary 2.1. (Uniqueness of the stationary density function) *If a stationary density function for the Markov chain $(Z_n)_{n\geq 0}$ exists then it is unique.*

Proof. Assume that there are two different stationary density functions $f \neq g \in Den(\Delta_d, \mathrm{d}\mathbf{x})$, i.e. $P^* f = f$ and $P^* g = g$. This implies that $d(P^* f, P^* g) = d(f, g)$, in contradiction with (2.4). □

Remark 2.2.

(a) Although $(Den(\Delta_d, \mathrm{d}\mathbf{x}), d)$ is a complete metric space, the operator P^* is not uniformly contractive, i.e. there is no $\rho \in [0, 1)$ such that, for any $f, g \in Den(\Delta_d, \mathrm{d}\mathbf{x})$, $d(P^* f, P^* g) \leq \rho d(f, g)$. Therefore, we cannot apply the Banach fixed point theorem. In [29, Proposition 2, pp.988–989], the authors applied the Banach fixed point theorem to prove the existence of the stationary density function, but their argument does not work. A precise proof can be found in [20, Theorem 3.1], which covered all cases of $p_i(\mathbf{x})$ in dimension 1.

(b) Although $Den(\Delta_d, \mathrm{d}\mathbf{x})$ is a non-empty closed convex subset in a Banach space $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$, we cannot apply the Browder fixed point theorem, because $\mathbb{L}^1(\Delta_d, \mathrm{d}\mathbf{x})$ is not uniformly convex.

(c) There are many cases of $p_i(\mathbf{x})$ such that there is no stationary density function for the $(Z_n)_{n\geq 0}$ even in dimension 1: see cases 2 and 3 in [20, Theorem 3.1], where the set of invariant probability measures consists of convex combinations of Dirac measures δ_0 and δ_1. It is of interest to find a criteria on the weight functions $p_i(\mathbf{x})$ which ensures the existence (then the unicity) of a stationary density function. This is still an open question (see the last section of the present paper).

3. General framework for iterated function systems with a unique invariant probability measure

In this section we recall some well-known results for iterated function systems and apply them to our model.

The approach we use here is based on the description of the spectrum of the transition operator associated with $(X_n)_{n\geq 0}$; it is totally different from the one used in [23], based on

criteria of ergodicity and stability of stochastic processes due to A. A. Borovkov [4]. We shall use the assumption that our weight functions p_i are Hölder, and there is at least one of these functions with full support. This assumption and Assumption 2.2(ii) in [23] cannot be compared. Therefore, our Theorem 3.4 and their Theorem 2.4 cannot be compared, although their Assumptions 2.2(i) and (iii) for ξ cover ours for $\xi = U(0, 1)$. Moreover, the main interest is its flexibility: indeed, it allows us to study the case when several invariant probability measures exist (see Section 5).

3.1. Iterated function systems with place-independent probabilities

Let (E, d) be a compact metric space and denote by $\mathbb{Lip}(E, E)$ the space of Lipschitz-continuous functions from E to E, i.e. functions $T : E \to E$ such that

$$[T] := \sup_{\substack{x,y \in E \\ x \neq y}} \frac{d(T(x), T(y))}{d(x, y)} < \infty.$$

Let $(T_n)_{n \geq 1}$ be a sequence of i.i.d. random functions defined on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ with values in $\mathbb{Lip}(E, E)$ and common distribution μ. We consider the Markov chain $(X_n)_{n \geq 0}$ on E defined, for any $n \geq 0$, by

$$X_{n+1} := T_{n+1}(X_n), \tag{3.1}$$

where X_0 is a fixed random variable with values in E. We say that the chain $(X_n)_{n \geq 0}$ is generated by the *iterated function system* $(T_n)_{n \geq 1}$. Notice that $X_n = T_n \circ \cdots \circ T_1(X_0)$; it thus corresponds to the forward iterations of the function system $(T_n)_{n \geq 1}$. The stochastic behavior of $(X_n)_{n \geq 0}$ is related to the contraction properties of the backward iterations $T_1 \circ \cdots \circ T_n$ of the function system $(T_n)_{n \geq 1}$, which usually behave more nicely.

The transition operator Q of $(X_n)_{n \geq 0}$ is defined, for any bounded Borel function $\varphi : E \to \mathbb{C}$ and any $x \in E$, by $Q\varphi(x) := \int_{\mathbb{Lip}(E,E)} \varphi(T(x))\mu(\mathrm{d}T)$. The chain $(X_n)_{n \geq 0}$ has the 'Feller property', i.e. the operator Q acts on the space $C(E, \mathbb{C})$ of continuous functions from E to $\mathbb{C}$. The maps T_n being Lipschitz continuous on E, the operator Q also acts on the space $\mathbb{Lip}(E, \mathbb{C})$ of Lipschitz-continuous functions from E to $\mathbb{C}$, and more generally on the space $\mathcal{H}_\alpha(E, \mathbb{C})$, $0 < \alpha \leq 1$, of α-Hölder-continuous functions from E to $\mathbb{C}$ defined by $\mathcal{H}_\alpha(E, \mathbb{C}) := \{f \in C(E, \mathbb{C}) \mid \|f\|_\alpha := \|f\|_\infty + m_\alpha(f) < +\infty\}$, where

$$m_\alpha(f) := \sup_{\substack{x,y \in E \\ x \neq y}} \frac{|f(x) - f(y)|}{d(x, y)^\alpha} < \infty.$$

Endowed with the norm $\| \cdot \|_\alpha$, the space $\mathcal{H}_\alpha(E, \mathbb{C})$ is a Banach space.

The behavior of the chain $(X_n)_{n \geq 0}$ is closely related to the spectrum of the restriction of Q to these spaces. Under some 'contraction in mean' assumption on the T_n, the restriction of Q to $\mathcal{H}_\alpha(E, \mathbb{C})$ satisfies some spectral gap property. We first cite the following result from [20, Proposition 2.1].

Theorem 3.1. ([20].) *Assume that there exists $\alpha \in (0, 1]$ such that*

$$r := \sup_{\substack{x,y \in E \\ x \neq y}} \int_{\mathbb{Lip}(E,E)} \left(\frac{d(T(x), T(y))}{d(x, y)} \right)^\alpha \mu(\mathrm{d}T) < 1. \tag{3.2}$$

Then, there exists on E a unique Q-invariant probability measure ν. Furthermore, there exist constants $\kappa > 0$ and $\rho \in (0, 1)$ such that, for all $\varphi \in \mathcal{H}_\alpha(E, \mathbb{C})$, $\|Q^n\varphi - \nu(\varphi)\|_\alpha \leq \kappa \rho^n \|\varphi\|_\alpha$.

Remark 3.1. In the theory of random iterative function systems, it is common to assume contraction in mean in the form

$$\mathbb{E}\left(\log \sup_{\substack{x,y \in E \\ x \neq y}} \frac{d(T(x), T(y))}{d(x, y)}\right) < 0,$$

which is secured, for instance, by

$$\mathbb{E}\left(\sup_{\substack{x,y \in E \\ x \neq y}} \frac{d(T(x), T(y))}{d(x, y)}\right)^{\alpha} < 1.$$

The log-contraction condition instead of this last, more restrictive, contraction condition is in fact an improvement of a cosmetic nature, since cases when log-contraction conditions are fulfilled but average contraction conditions are not typically indicate that a change of metric is appropriate (see [31] for more details). Notice also that the condition (3.2) (see also condition H1 in Theorem 3.3) is less restrictive, with the supremum put out of the expectation; it already appeared as the 'global average condition' in [31] and references therein.

3.1.1. *Application to the Diaconis–Freedman chain for p fixed in Δ_d* We assume $p_i(\mathbf{x}) = p_i$ for all $i = 0, \ldots, d$. This puts the Diaconis–Freedman chain into the framework of iterated random functions as follows. For each $i = 0, \ldots, d$ and $t \in [0, 1]$, we set $H_i(t, \cdot) \colon \Delta_d \to \Delta_d$, $\mathbf{x} \mapsto t\mathbf{x} + (1 - t)\mathbf{e}_i$ as an affine transformation; these functions $H_i(t, \cdot)$ belong to the space $\mathbb{L}\mathrm{ip}(\Delta_d, \Delta_d)$ of Lipschitz-continuous functions from Δ_d to Δ_d, with Lipschitz coefficient $m(H_i(t, \cdot)) = t$. Then, we consider the probability measure μ on $\mathbb{L}\mathrm{ip}(\Delta_d, \Delta_d)$ defined by

$$\mu(\mathrm{d}T) := \sum_{i=0}^{d} p_i \int_0^1 \delta_{H_i(t, \cdot)}(\mathrm{d}T)\,\mathrm{d}t, \tag{3.3}$$

where δ_T is the Dirac mass at T. Equation (2.2) may be rewritten as: for all $f \in \mathbb{L}^{\infty}(\Delta_d)$ and $\mathbf{x} \in \Delta_d$, $Qf(\mathbf{x}) = \int_{\mathbb{L}\mathrm{ip}(\Delta_d, \Delta_d)} f(T(\mathbf{x}))\mu(\mathrm{d}T)$. Hence, the Diaconis–Freedman chain $(Z_n)_{n \geq 0}$ on Δ_d is generated by the *iterated function system* $(T_n)_{n \geq 1}$ in the sense of (3.1), where $(T_n)_{n \geq 1}$ is a sequence of i.i.d. random functions with common distribution μ defined by (3.3).

Theorem 3.2. *If $p_i(\mathbf{x}) = p_i$ for all $i = 0, \ldots, d$, then the Diaconis–Freedman chain in Δ_d admits a unique P-invariant probability measure ν on Δ_d. Furthermore, for any $\alpha \in (0, 1]$, there exist constants $\kappa > 0$ and $\rho \in (0, 1)$ depending on α such that, for any $\varphi \in \mathcal{H}_{\alpha}(\Delta_d, \mathbb{C})$ and $\mathbf{x} \in \Delta_d$, $\left|Q^n\varphi(\mathbf{x}) - \nu(\varphi)\right| \leq \kappa\rho^n\|\varphi\|_{\alpha}$.*

Proof. This is a direct consequence of Theorem 3.1, because, for all $\alpha \in (0, 1]$, we have

$$r = \sup_{\substack{\mathbf{x}, \mathbf{y} \in \Delta_d \\ \mathbf{x} \neq \mathbf{y}}} \int_{\mathbb{L}\mathrm{ip}(\Delta_d, \Delta_d)} \left(\frac{|T(\mathbf{x}) - T(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|}\right)^{\alpha} \mu(\mathrm{d}T) \leq \sum_{i=0}^{d} p_i \int_0^1 m(H_i(t, \cdot))^{\alpha}\,\mathrm{d}t$$

$$= \int_0^1 t^{\alpha}\,\mathrm{d}t = \frac{1}{1 + \alpha} < 1. \qquad \square$$

Remark 3.2. The unique *P*-invariant probability measure ν is usually nothing but the Dirichlet distribution $Dir[\theta_0, \ldots, \theta_k]$, as will be shown later. If $\theta_i > 0$ for all $i = 0, \ldots, k$, we have a

unique invariant probability density which is the Dirichlet density. Otherwise, the Dirichlet distribution is singular and can be understood in the sense of [9, p. 211], [10, Definition 3.1.1, p. 89], or [16, Definition 4.2].

3.2. Iterated function systems with place-dependent probabilities

In this subsection we extend the measure μ to a collection $(\mu_x)_{x \in E}$ of probability measures on $\mathbb{L}\mathrm{ip}(E, E)$, depending continuously on x. We consider the Markov chain $(X_n)_{n \geq 0}$ on E whose transition operator Q is given, for any bounded Borel function $\varphi : E \to \mathbb{C}$ and any $x \in E$, by $Q\varphi(x) = \int_{\mathbb{L}\mathrm{ip}(E,E)} \varphi(T(x))\mu_x(\mathrm{d}T)$. First, we introduce the following definition.

Definition 3.1. A sequence $(\xi_n)_{n \geq 0}$ of continuous functions from E to E is a contracting sequence if there exists $x_0 \in E$ such that, for all $x \in E$, $\lim_{n \to +\infty} \xi_n(x) = x_0$.

We cite the following result from [20, Proposition 2.2].

Theorem 3.3. ([20].) *Assume that there exists $\alpha \in (0, 1]$ such that*

H1 $\quad r := \sup\limits_{\substack{x,y \in E \\ x \neq y}} \int_{\mathbb{L}\mathrm{ip}(E,E)} \left(\dfrac{d(T(x), T(y))}{d(x, y)} \right)^{\alpha} \mu_x(\mathrm{d}T) < 1;$

H2 $\quad R_\alpha := \sup\limits_{\substack{x,y \in E \\ x \neq y}} \dfrac{|\mu_x - \mu_y|_{\mathrm{TV}}}{d(x, y)^{\alpha}} < +\infty,$ *where $|\mu_x - \mu_y|_{\mathrm{TV}}$ is the total variation distance between μ_x and μ_y;*

H3 $\quad$ *there exist $\delta > 0$ and a probability measure μ on E such that*

    (i) *for all $x \in E$, $\mu_x \geq \delta\mu$;*

    (ii) *the closed semi-group T_μ generated by the support S_μ of μ possesses a contracting sequence.*

Then, there exists on E a unique Q-invariant probability measure ν. Furthermore, for some constants $\kappa > 0$ and $\rho \in (0, 1)$, for any $\varphi \in \mathcal{H}_\alpha(E)$ and $x \in E$, $|Q^n\varphi(x) - \nu(\varphi)| \leq \kappa\rho^n \|\varphi\|_\alpha$.

Let us now apply this statement to the Diaconis–Freedman chain on Δ_d. For each $\mathbf{x} \in \Delta_d$, we define a space-dependent probability measure $\mu_{\mathbf{x}}$ on $\mathbb{L}\mathrm{ip}(\Delta_d, \Delta_d)$ by

$$\mu_{\mathbf{x}}(\mathrm{d}T) := \sum_{i=0}^{d} p_i(\mathbf{x}) \int_0^1 \delta_{H_i(t,\cdot)}(\mathrm{d}T) \, \mathrm{d}t,$$

where δ_T is the Dirac mass at T. With this collection $(\mu_{\mathbf{x}})_{\mathbf{x} \in \Delta_d}$ of probability measures, the Diaconis–Freedman chain falls within the scope of Theorem 3.3.

Theorem 3.4. *Assume that*

    (i) *for all $j = 0, \ldots, d$, the functions p_j belong to $\mathcal{H}_\alpha(\Delta_d, \mathbb{C})$;*

    (ii) *there exists $i \in \{0, \ldots, d\}$ such that $\delta_i := \min_{\mathbf{x} \in \Delta_d} p_i(\mathbf{x}) > 0$.*

Then, the Diaconis–Freedman chain in Δ_d has a unique P-invariant probability measure ν on Δ_d. Furthermore, there exist constants $\kappa > 0$ and $\rho \in (0, 1)$ such that, for any $\varphi \in \mathcal{H}_\alpha(\Delta_d, \mathbb{C})$ and $\mathbf{x} \in \Delta_d$, $|P^n\varphi(\mathbf{x}) - \nu(\varphi)| \leq \kappa\rho^n \|\varphi\|_\alpha$.

Proof. This is a direct consequence of Theorem 3.3 since conditions H1–H3 hold in this context:

H1 For any $\mathbf{x} \neq \mathbf{y} \in \Delta_d$,

$$\int_{\mathbb{Lip}(\Delta_d, \Delta_d)} \left(\frac{|T(\mathbf{x}) - T(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|} \right)^\alpha \mu_{\mathbf{x}}(\mathrm{d}T) = \sum_{i=0}^d p_i(\mathbf{x}) \int_0^1 \left(\frac{|H_i(t, \mathbf{x}) - H_i(t, \mathbf{y})|}{|\mathbf{x} - \mathbf{y}|} \right)^\alpha \mathrm{d}t$$

$$= \frac{1}{1+\alpha} < 1.$$

H2 For any $\mathbf{x} \neq \mathbf{y} \in \Delta_d$ and any Borel set $A \subseteq \mathbb{Lip}(\Delta_d, \Delta_d)$,

$$\frac{|\mu_{\mathbf{x}}(A) - \mu_{\mathbf{y}}(A)|}{|\mathbf{x} - \mathbf{y}|^\alpha} \leq \sum_{i=0}^d \frac{|p_i(\mathbf{x}) - p_i(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^\alpha} \int_0^1 \delta_{H_i(t, \cdot)}(A) \, \mathrm{d}t$$

$$\leq \sum_{i=0}^d m_\alpha(p_i) < \infty.$$

H3 Set $\mu(\mathrm{d}T) := \int_0^1 \delta_{H_i(t, \cdot)}(\mathrm{d}T) \, \mathrm{d}t$; then, for all $\mathbf{x} \in \Delta_d$,

$$\mu_{\mathbf{x}}(\mathrm{d}T) \geq p_i(\mathbf{x}) \int_0^1 \delta_{H_i(t, \cdot)}(\mathrm{d}T) \, \mathrm{d}t \geq \delta \mu(\mathrm{d}T).$$

Moreover, the constant function $\mathbf{x} \mapsto 0$ belongs to the support of μ so that the semigroup T_μ contains a contracting sequence with limit point 0.

We emphasize that under conditions H1–H3, the eigenvalue 1 is simple and is the unique eigenvalue of Q with modulus 1; the proof requires precise control of the peripherical spectrum of Q; see the proof of [20, Proposition 2.2]. □

4. Some explicit invariant probability densities

In this section we consider some special cases of weights for which it is possible to compute explicitly the unique invariant probability density. When $d = 1$, it is known that, when both conditions $p_1(0) > 0$ and $p_0(1) > 0$ hold, there exists a unique invariant probability density of $(Z_n)_{n \geq 0}$ given by

$$g_\infty(y) = C \exp \left(\int_{1/2}^y \frac{p_1(t)}{1-t} \, \mathrm{d}t - \int_{1/2}^y \frac{p_0(t)}{t} \, \mathrm{d}t \right);$$

see, for instance, [29] or [20]. We do not get such a general result when $d \geq 2$, we can do it only in some specific cases. We would also like to emphasize that in [23], based on Sethuraman's construction of the Dirichlet distributions [30], the authors also gave general results for the explicit formula of the stationary density in these special cases. Our approach, however, is very natural and worth taking into account.

4.1. The case of constant weights

We first consider the case of constant weights, i.e. $p_i(\mathbf{x}) = p_i > 0$ for all $i = 0, \ldots, d$.

Theorem 4.1. *If $p_i(\mathbf{x}) = p_i > 0$ for all $i = 0, \ldots, d$, then the unique invariant probability density g_∞ of $(Z_n)_{n \geq 0}$ is the density of the Dirichlet distribution $Dir[p_0, \ldots, p_d]$, i.e.*

$$g_\infty(\mathbf{y}) = \frac{1}{\prod_{i=0}^{d} \Gamma(p_i)} \prod_{i=0}^{d} y_i^{p_i - 1} \mathbf{1}_{\Delta_d}(\mathbf{y})$$

for any $\mathbf{y} = (y_1, \ldots, y_d)$ in Δ_d, where $y_0 = 1 - y_1 - \cdots - y_d$.

Proof. It suffices to prove that $g_\infty(\mathbf{y}) = \prod_{i=0}^{d} y_i^{\alpha_i}$ with $\alpha_i = p_i - 1$ is the unique solution of the equation $P^* g(\mathbf{y}) = g(\mathbf{y})$. Indeed, according to (2.3),

$$P^* g_\infty(\mathbf{y}) = \sum_{i=0}^{d} p_i \int_1^{\frac{1}{1-y_i}} s^{d-2} g_\infty\left(s\mathbf{y} + (1-s)\mathbf{e}_i\right) ds$$

$$= \sum_{i=0}^{d} p_i \int_1^{\frac{1}{1-y_i}} s^{d-2} (sy_i + 1 - s)^{\alpha_i} \prod_{j \neq i} (sy_j)^{\alpha_j} ds$$

$$= \sum_{i=0}^{d} p_i \int_1^{\frac{1}{1-y_i}} s^{d-2+\sum_{j \neq i} \alpha_j} (sy_i + 1 - s)^{\alpha_i} \frac{g_\infty(\mathbf{y})}{y_i^{\alpha_i}} ds$$

$$= \sum_{i=0}^{d} \frac{p_i}{(1-y_i) y_i^{\alpha_i} (1-y_i)^{-2-\alpha_i}} \left(\int_0^{y_i} t^{\alpha_i} (1-t)^{-2-\alpha_i} dt\right) g_\infty(\mathbf{y}), \qquad (4.1)$$

where the last equality follows by using the change of variables $t = sy_i + 1 - s$ and the equality $\sum_{j=0}^{d} \alpha_j = -d$. Notice that, for $i = 0, \ldots, d$ and $\alpha_i > -1$,

$$(1 + \alpha_i) \int_0^{y_i} t^{\alpha_i} (1-t)^{-2-\alpha_i} dt = y_i^{1+\alpha_i} (1 - y_i)^{-1-\alpha_i}.$$

As a matter of fact, the function $F \colon y_i \mapsto (1 + \alpha_i) \int_0^{y_i} t^{\alpha_i} (1-t)^{-2-\alpha_i} dt - y_i^{1+\alpha_i} (1 - y_i)^{-1-\alpha_i}$ satisfies $F(0) = 0$ and $F'(y_i) = 0$; therefore, $F(y_i) \equiv 0$. Hence, the equality (4.1) yields

$$P^* g_\infty(\mathbf{y}) = \sum_{i=0}^{d} y_i g_\infty(\mathbf{y}) = g_\infty(\mathbf{y}).$$

The uniqueness stems from Corollary 2.1. □

Example 4.1. When $d = 2$, $p_1 = p_2 = p_0 = 1/3$, the unique invariant probability density is

$$g_\infty(\mathbf{y}) = \frac{1}{\Gamma(1/3)^3} y_1^{-\frac{2}{3}} y_2^{-\frac{2}{3}} (1 - y_1 - y_2)^{-\frac{2}{3}} \mathbf{1}_{\Delta_2}(\mathbf{y})$$
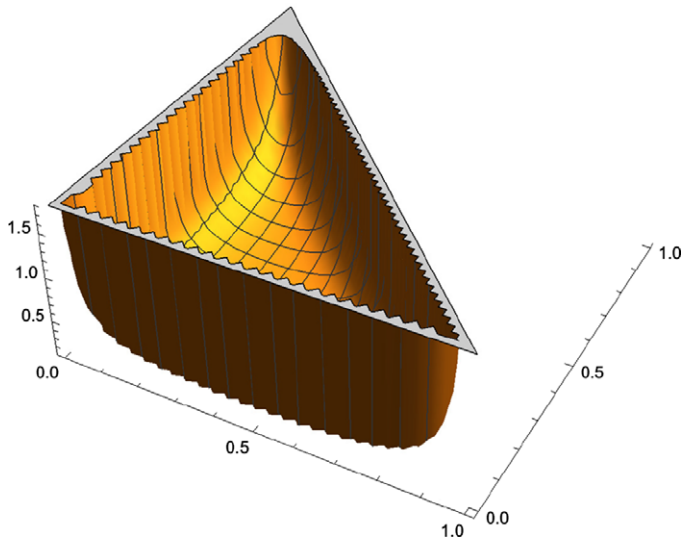
(see Figure 2).

FIGURE 2. The density function of the Dirichlet distribution $Dir\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$.
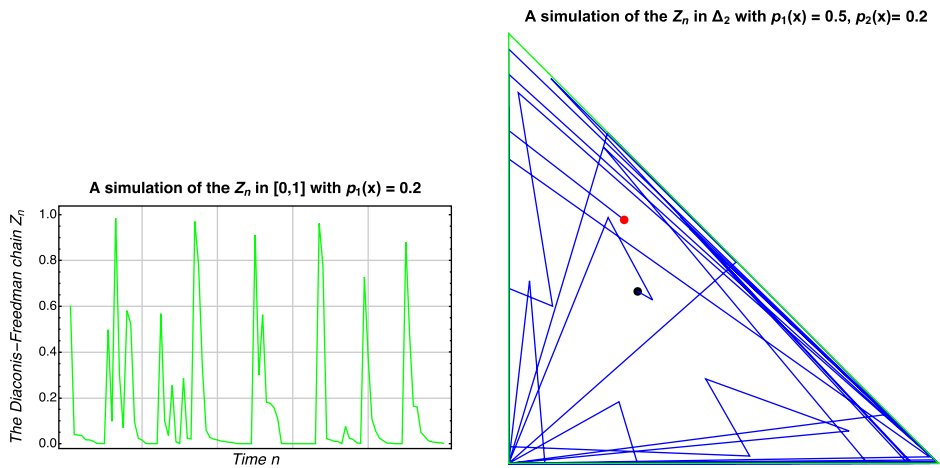

FIGURE 3. Left: Random trajectory of Z_n in [0,1] starting at $x_0 = 0.6$. Right: Random trajectory of Z_n in Δ_2 starting at $\mathbf{x}_0 = (0.3, 0.4)$.

Figure 3 represents random trajectories of $(Z_n)_{n \geq 0}$:

- in [0, 1], starting at $x_0 = 0.6$ with $p_1(x) = 0.2$, $p_0(x) = 0.8$, in the left panel;

- in Δ_2, starting at $\mathbf{x}_0 = (0.3, 0.4)$ with $p_1(\mathbf{x}) = 0.5$, $p_2(\mathbf{x}) = 0.2$, $p_0(\mathbf{x}) = 0.3$ in the right panel.

They both illustrate the ergodic behavior of the chain.

4.2. The case of affine weights

In this section we present a special case of non-constant weight functions $p_i(\mathbf{y})$ which yield an explicit form of the unique invariant density function.

Theorem 4.2. *Fix positive constants $\theta_0, \ldots, \theta_d > 0$ with $|\boldsymbol{\theta}| = \theta_0 + \cdots + \theta_d \leq 1$, and assume that, for any $i = 1, \ldots, d$ and $\mathbf{x} = (x_1, \ldots, x_d)$ in Δ_d, $p_i(\mathbf{x}) = \tilde{p}_i(x_i) := \theta_i + (1 - |\boldsymbol{\theta}|)x_i$ (which implicitly implies $p_0(\mathbf{x}) = \tilde{p}_0(x_0) := \theta_0 + (1 - |\boldsymbol{\theta}|)x_0$). Then, the unique invariant probability density g_∞ of $(Z_n)_{n \geq 0}$ is the Dirichlet distribution $Dir[\theta_0, \ldots, \theta_d]$ given by*

$$g_\infty(\mathbf{y}) = Dir[\theta_0, \ldots, \theta_d](\mathbf{y}) = \frac{\Gamma(|\boldsymbol{\theta}|)}{\prod_{i=0}^d \Gamma(\theta_i)} \prod_{i=0}^d y_i^{\theta_i - 1} \mathbf{1}_{\Delta_d}(\mathbf{y})$$

for any $\mathbf{y} = (y_1, \ldots, y_d)$ in Δ_d, where $y_0 = 1 - y_1 - \cdots - y_d$.

Proof. Using the same techniques as in the proof of Theorem 4.1, we only need to check that

$$\frac{\int_0^{y_i} p_i(t) t^{\alpha_i} (1 - t)^{|\boldsymbol{\alpha}| - \alpha_i + d - 2} \, \mathrm{d}t}{y_i^{\alpha_i} (1 - y_i)^{|\boldsymbol{\alpha}| - \alpha_i + d - 1}} = y_i,$$

which can be easily done by a direct calculation.

Remark 4.1.

(i) The case when $|\boldsymbol{\theta}| = 1$ corresponds to constant weights.

(ii) This result has a very close connection to results studied in Wright–Fisher models with mutations (see, for instance, [15, 32, 33]).

In Figure 4 we simulate random trajectories of $(Z_n)_{n \geq 0}$:

- in $[0, 1]$, starting at $x_0 = 0.6$ with $p_1(x) = x$, $p_0(x) = 1 - x$ in the left panel;

- in Δ_2 starting at $\mathbf{x}_0 = (0.3, 0.4)$ with $p_1(\mathbf{x}) = x_1$, $p_2(\mathbf{x}) = x_2$, $p_0(\mathbf{x}) = 1 - x_1 - x_2$ in the right panel. They both illustrate the absorbing behavior of the chain.

5. Asymptotic behavior of $(Z_n)_{n \geq 0}$

In this section we describe the asymptotic behavior of $(Z_n)_{n \geq 0}$ using the notion of *minimal P-absorbing compact subsets*. First, we establish some properties of the family $\mathcal{K}_m$ of these subsets and propose their classification. By a general results of [14], this yields the classification of the set of P-invariant probability measures as well as the description of the asymptotic behavior of $(Z_n)_{n \geq 0}$. The classification is complete in Δ_2 but partial in Δ_d, $d > 2$.

5.1. The set $\mathcal{K}_m$ of minimal *P*-absorbing compact subsets

Definition 5.1. A non-empty compact subset $K \subseteq \Delta_d$ is said to be *P-absorbing* if, for all $\mathbf{x} \in K$,

$$P(\mathbf{x}, K^c) := P\mathbf{1}_{K^c}(\mathbf{x}) = \sum_{i=0}^d p_i(\mathbf{x}) \int_0^1 \mathbf{1}_{K^c}(t\mathbf{x} + (1 - t)e_i) \, \mathrm{d}t = 0,$$

where $K^c = \Delta_d \setminus K$. It is *minimal* when it does not contain any proper P-absorbing compact subset.

Four simulations of the Z_n in Δ_2 with $p_1(x) = x_1$, $p_2(x)=x$

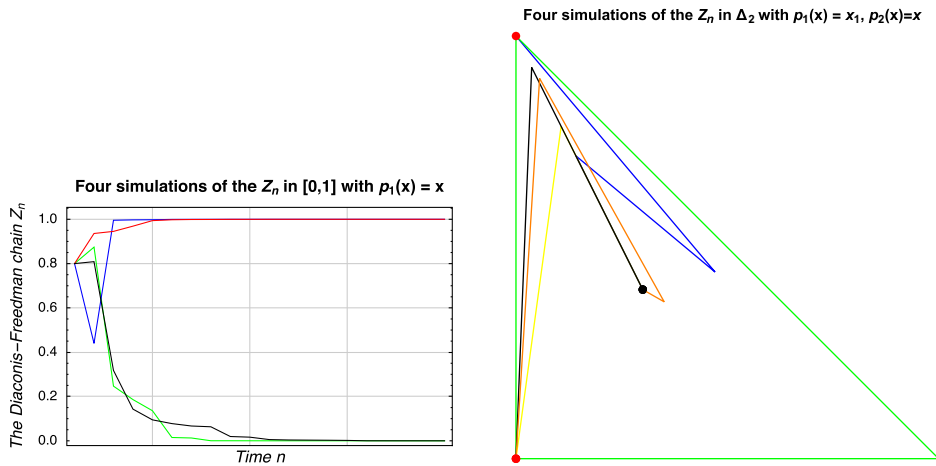**Four simulations of the Z_n in [0,1] with $p_1(x) = x$**

FIGURE 4. Left: Four random trajectories of Z_n in $[0,1]$ starting at $x_0 = 0.8$; they will absorb in $\{0, 1\}$. Right: Four random trajectories of Z_n in Δ_2 starting at $\mathbf{x}_0 = (0.3, 0.4)$; they will absorb in $\{\mathbf{e}_0, \mathbf{e}_2\}$.
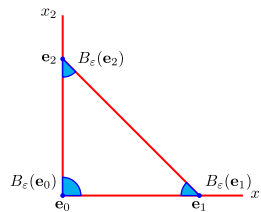
FIGURE 5. The domain $B_\varepsilon(\mathbf{e}_i)$.

We denote by $\mathcal{K}_m$ the set of all minimal P-absorbing compact subsets. For any $\mathbf{x}_0 \in \Delta_d$ and $\varepsilon > 0$, we set $B_\varepsilon(\mathbf{x}_0) = \{\mathbf{x} \in \Delta_d : |\mathbf{x} - \mathbf{x}_0| < \varepsilon\}$ and $B_\varepsilon = \cup_{i=0}^d B_\varepsilon(\mathbf{e}_i)$; Figure 5 shows the domain $B_\varepsilon(\mathbf{e}_i)$.

The following rules are useful for describing the minimal P-absorbing sets K.

Proposition 5.1. *Given continuous weight functions $p_i(\mathbf{x})$ on Δ_d, i.e. $p_i(\mathbf{x}) \geq 0$ for every $i = 0, \ldots, d$ and $\sum_{i=0}^d p_i(\mathbf{x}) = 1$:*

(i) *If $K \in \mathcal{K}_m$ then K contains at least one vertex.*

(ii) *If $K \in \mathcal{K}_m$, $\mathbf{e}_i \in K$, and $p_i(\mathbf{e}_i) = 1$ then $K = \{\mathbf{e}_i\}$.*

(iii) *If $K \in \mathcal{K}_m$, $\mathbf{e}_i \in K$, and $p_j(\mathbf{e}_i) > 0$ for some $j \neq i$ then $[\mathbf{e}_i, \mathbf{e}_j] \subseteq K$.*

(iv) *If $K \in \mathcal{K}_m$ and $p_i(\mathbf{x}) > 0$ for some $\mathbf{x} \in K \setminus \{\mathbf{e}_i\}$ then $[\mathbf{e}_i, \mathbf{x}] \subseteq K$.*

Proof.

(i) Assume that $\mathbf{e}_i \notin K$ for all $i = 0, \ldots, d$. Since K^c is open, there exists $\varepsilon > 0$ such that $B_\varepsilon \subseteq K^c$. Therefore, for all $\mathbf{x} \in K$,

$$P(\mathbf{x}, K^c) = \sum_{i=0}^{d} p_i(\mathbf{x}) \int_0^1 \mathbf{1}_{K^c}(t\mathbf{x} + (1-t)e_i)\,\mathrm{d}t$$

$$\geq \sum_{i=0}^{d} p_i(\mathbf{x}) \int_0^1 \mathbf{1}_{B_\varepsilon}(t\mathbf{x} + (1-t)e_i)\,\mathrm{d}t = \sum_{i=0}^{d} p_i(\mathbf{x})\varepsilon = \varepsilon > 0.$$

This contradicts the fact that K is P-absorbing.

(ii) It suffices to prove that $\{\mathbf{e}_i\} \in \mathcal{K}_m$. This is true because

$$P(\mathbf{e}_i, \{\mathbf{e}_i\}^c) = \sum_{j \neq i} p_j(\mathbf{e}_i) = 0.$$

(iii) If $[\mathbf{e}_i, \mathbf{e}_j] \cap K^c \neq \emptyset$, then there exist $\mathbf{x}_0 \in [\mathbf{e}_i, \mathbf{e}_j] \cap K^c$ and $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{x}_0) \subseteq K^c$. Therefore, $P(\mathbf{e}_i, K^c) \geq p_j(\mathbf{e}_i)\varepsilon > 0$, a contradiction.

(iv) Again, if $[\mathbf{e}_i, \mathbf{e}_j] \cap K^c \neq \emptyset$, then there exist $\mathbf{x}_0 \in [\mathbf{e}_i, \mathbf{e}_j] \cap K^c$ and $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{x}_0) \subseteq K^c$; hence, $P(\mathbf{x}, K^c) \geq p_i(\mathbf{x})\varepsilon > 0$, which is a contradiction. $\square$

This proposition easily yields the classification of $\mathcal{K}_m$ when $d = 1$ (see [20]):

(i) If $p_1(1) < 1$ and $p_0(0) < 1$ then $\mathcal{K}_m = \{[0, 1]\}$.

(ii) If $p_1(1) < 1$ and $p_0(0) = 1$ then $\mathcal{K}_m = \{\{1\}\}$.

(iii) If $p_1(1) = 1$ and $p_0(0) < 1$ then $\mathcal{K}_m = \{\{0\}\}$.

(iv) If $p_1(1) = 1$ and $p_0(0) = 1$ then $\mathcal{K}_m = \{\{0\}, \{1\}\}$.

Proposition 5.1 is also sufficient to get the full classification of the set $\mathcal{K}_m$ in Δ_2. In the next two subsections we focus on this classification and the asymptotic behavior of $(Z_n)_{n \geq 0}$ in Δ_2. The reader may be easily convinced that similar statements hold in higher dimensions, with technical difficulties we do not want to contend with here; an expanded version of Proposition 5.1 is required when $d \geq 3$.

5.2. Classification of $\mathcal{K}_m$ in Δ_2

We assume $d = 2$ in this subsection. Unlike the case $d = 1$, for the case of $d = 2$ we need to classify the values of p_i not only on the vertices but also on the edges. We denote $L_0 = \{\mathbf{x} \in [\mathbf{e}_1, \mathbf{e}_2] : p_0(\mathbf{x}) > 0\}$, $L_1 = \{\mathbf{x} \in [\mathbf{e}_0, \mathbf{e}_2] : p_1(\mathbf{x}) > 0\}$, and $L_2 = \{\mathbf{x} \in [\mathbf{e}_0, \mathbf{e}_1] : p_2(\mathbf{x}) > 0\}$. Let L_0^c (respectively L_1^c and L_2^c) be the complement of L_0 (resp. L_1 and L_2) in $[\mathbf{e}_1, \mathbf{e}_2]$ (resp. $[\mathbf{e}_0, \mathbf{e}_2]$ and $[\mathbf{e}_0, \mathbf{e}_1]$).

Let us fix $K \in \mathcal{K}_m$. There are several cases to consider.

5.2.1 $p_0(e_0) = p_1(e_1) = p_2(e_2) = 1$ By Proposition 5.1(i), either $\mathbf{e}_0 \in K$ or $\mathbf{e}_1 \in K$ or $\mathbf{e}_2 \in K$. When $\mathbf{e}_0 \in K$, Proposition 5.1(ii) implies $K = \{\mathbf{e}_0\}$; similarly for $\mathbf{e}_1$ or $\mathbf{e}_2$. Finally, $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}, \{\mathbf{e}_2\}\}$.

5.2.2 $p_0(e_0) = p_1(e_1) = 1$ *but* $p_2(e_2) < 1$ As above, if $\mathbf{e}_0 \in K$ (resp. $\mathbf{e}_1 \in K$) then $K = \{\mathbf{e}_0\}$ (resp. $\mathbf{e}_1 \in K$). Assume now that $\mathbf{e}_2 \in K$. Proposition 5.1(iii) implies $[\mathbf{e}_0, \mathbf{e}_2] \subseteq K$ when $p_2(\mathbf{e}_0) > 0$ and $[\mathbf{e}_1, \mathbf{e}_2] \subseteq K$ when $p_2(\mathbf{e}_1) > 0$. Therefore, K contains $\mathbf{e}_1$ or $\mathbf{e}_2$, in contradiction with the minimality of K. Finally, $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}\}$.
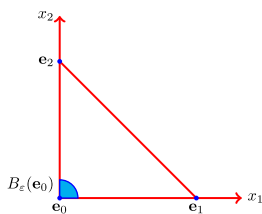
FIGURE 6. Domain $B_\varepsilon(\mathbf{e}_0)$.

Similar statements hold when $p_0(\mathbf{e}_0) = p_2(\mathbf{e}_2) = 1$ but $p_1(\mathbf{e}_1) < 1$ or $p_1(\mathbf{e}_1) = p_2(\mathbf{e}_2) = 1$ but $p_0(\mathbf{e}_0) < 1$.

5.2.3 $p_0(\mathbf{e}_0) = 1$ *but* $p_1(\mathbf{e}_1), p_2(\mathbf{e}_2) < 1$ Firstly, $\{\mathbf{e}_0\} \in \mathcal{K}_m$ and $K = \{\mathbf{e}_0\}$ as soon as $\mathbf{e}_0 \in K$.

Assume now $\mathbf{e}_0 \notin K$ (thus $B_\varepsilon(\mathbf{e}_0) \subseteq K^c$ for some $\varepsilon > 0$; see Figure 6) and suppose, for instance, $\mathbf{e}_1 \in K$ (the same argument holds with $\mathbf{e}_2$). Hence, $p_0(\mathbf{e}_1) = 0$; indeed, the condition $p_0(\mathbf{e}_1) > 0$ implies $[\mathbf{e}_0, \mathbf{e}_1] \subseteq K$, a contradiction. Consequently, $p_2(\mathbf{e}_1) > 0$, which implies $[\mathbf{e}_1, \mathbf{e}_2] \subseteq K$, so $p_0(\mathbf{e}_2) = 0$ and $p_1(\mathbf{e}_2) > 0$. This readily implies that $L_0 = \emptyset$; otherwise, $p_0(\mathbf{x}_0) > 0$ for some $\mathbf{x}_0 \in [\mathbf{e}_1, \mathbf{e}_2]$, and therefore $P(\mathbf{x}_0, K^c) \geq p_0(\mathbf{x}_0)\varepsilon > 0$, contradicting the fact that $\mathbf{x}_0 \in K$ and K is invariant. The equality $L_0 = \emptyset$ yields $K = [\mathbf{e}_1, \mathbf{e}_2]$ and $\mathcal{K}_m = \{\{\mathbf{e}_0\}, [\mathbf{e}_1, \mathbf{e}_2]\}$. Finally,

$$\mathcal{K}_m = \begin{cases} \{\{\mathbf{e}_0\}, [\mathbf{e}_1, \mathbf{e}_2]\} & \text{if } L_0 = \emptyset, \\[2mm] \{\{\mathbf{e}_0\}\} & \text{otherwise.} \end{cases}$$

Similar statements hold when $p_1(\mathbf{e}_1) = 1$ but $p_0(\mathbf{e}_0), p_2(\mathbf{e}_2) < 1$, or $p_2(\mathbf{e}_2) = 1$ but $p_0(\mathbf{e}_0), p_1(\mathbf{e}_1) < 1$.

5.2.4 $p_0(\mathbf{e}_0), p_1(\mathbf{e}_1), p_2(\mathbf{e}_2) < 1$ *and* $L_0 = \emptyset$ By Proposition 5.1(i), the set K contains at least one of the vertices. Assume, for instance, $\mathbf{e}_1 \in K$; thus, $p_2(\mathbf{e}_1) > 0$ since $L_0 = \emptyset$, which implies $[\mathbf{e}_1, \mathbf{e}_2] \subseteq K$. The condition $L_0 = \emptyset$ also implies $P(\mathbf{x}, [\mathbf{e}_1, \mathbf{e}_2]^c) = p_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in [\mathbf{e}_1, \mathbf{e}_2]$, and finally $K = [\mathbf{e}_1, \mathbf{e}_2]$. The same conclusion holds when $\mathbf{e}_2 \in K$.

Now, the set K cannot contain $\mathbf{e}_0$. Otherwise, the condition $p_0(\mathbf{e}_0) < 1$ implies either $[\mathbf{e}_0, \mathbf{e}_1] \subseteq K$ or $[\mathbf{e}_0, \mathbf{e}_2] \subseteq K$; thus, $\mathbf{e}_1 \in K$ or $\mathbf{e}_2 \in K$. This yields $K = [\mathbf{e}_1, \mathbf{e}_2]$, a contradiction. Finally, $\mathcal{K}_m = \{[\mathbf{e}_1, \mathbf{e}_2]\}$.

Similar statements hold when $L_1 = \emptyset$ or $L_2 = \emptyset$.

5.2.5 $p_0(\mathbf{e}_0), p_1(\mathbf{e}_1), p_2(\mathbf{e}_2) < 1$ *and* L_0, L_1, L_2 *non-empty* In this case we always have $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\} \subseteq K$. Indeed, by Proposition 5.1(i), the set K contains at least one vertex, say $\mathbf{e}_0 \in K$; since $p_0(\mathbf{e}_0) < 1$, it contains one of the two sides $[\mathbf{e}_0, \mathbf{e}_1]$ or $[\mathbf{e}_0, \mathbf{e}_2]$. Assume $[\mathbf{e}_0, \mathbf{e}_1] \subset K$ (thus $\mathbf{e}_1 \in K$) and let us check that $\mathbf{e}_2 \in K$. Otherwise, $B_\varepsilon(\mathbf{e}_2) \subseteq K^c$ for some $\varepsilon > 0$; since L_2 is a proper subset of $[\mathbf{e}_0, \mathbf{e}_1]$, there exists $\mathbf{x} \in [\mathbf{e}_0, \mathbf{e}_1] \subseteq K$ such that $P(\mathbf{x}, K^c) \geq p_2(\mathbf{x})\varepsilon > 0$, a contradiction.

Now, the inclusion $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\} \subseteq K$ combined with Proposition 5.1(iv) yields $co\{\mathbf{e}_0, L_0\} \subseteq K$, $co\{\mathbf{e}_1, L_1\} \subseteq K$, and $co\{\mathbf{e}_2, L_2\} \subseteq K$, so that, by the compactness of K,

$$K_0 := \overline{co\{\mathbf{e}_0, L_0\} \cup co\{\mathbf{e}_1, L_1\} \cup co\{\mathbf{e}_2, L_2\}} \subseteq K. \tag{5.1}$$
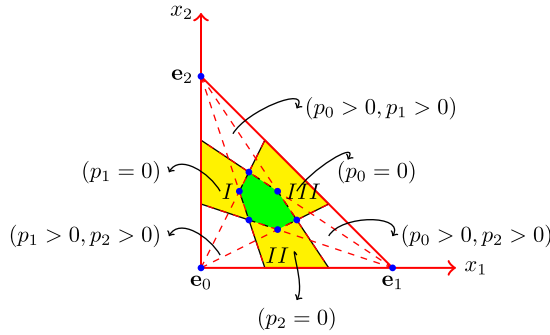
FIGURE 7. $K_0 =$ white domains $\cup$ yellow domains $= \Delta_2 \setminus$ green domain; K_0^c is the green domain; Ω_1, Ω_2, and Ω_0 are the yellow domains *I*, *II*, and *III* respectively.

Next, for all $i \in \{0, 1, 2\}$ we denote $L_i(1) := \{\mathbf{x} \in K_0 : p_i(\mathbf{x}) > 0\}$. It is easy to see that $L_i(1) \supseteq L_i$ for all $i \in \{0, 1, 2\}$. Then, Proposition 5.1(iv) and the compactness of K yield

$$K_1 := \overline{co\{\mathbf{e}_0, L_0(1)\} \cup co\{\mathbf{e}_1, L_1(1)\} \cup co\{\mathbf{e}_2, L_2(1)\}} \subseteq K. \tag{5.2}$$

We construct by induction a sequence of increasing compact subsets $\{K_n\}_{n \geq 0} \subseteq K$ and consider two possibilities: if there is a finite n such that $K_n = \Delta_2$ then $K = \Delta_2$; otherwise, $K = K_\infty := \overline{\{\cup_{n \geq 0} K_n\}} \subset \Delta_2$. In this case $\mathcal{K}_m$ consists of a unique minimal P-absorbing compact set,

$$\mathcal{K}_m = \begin{cases} \{\Delta_2\} & \text{if there exists a finite } n \text{ such that } K_n = \Delta_2, \\ \{K_\infty\} & \text{otherwise.} \end{cases}$$

We illustrate here two cases when $\mathcal{K}_m = \{K_0\}$ and $\mathcal{K}_m = \{K_1\}$. If we assume that, for $i = 0, 1, 2$,

$$p_i(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \Omega_i := \{\mathbf{y} \in K_0 : [\mathbf{y}, \mathbf{e}_i] \cap K_0^c \neq \emptyset\}, \tag{5.3}$$

then $K_1 = K_0$ and, by induction, $K_\infty = K_0$. Moreover, from (5.3), $P(\mathbf{x}, K_0^c) = 0$ for all $\mathbf{x} \in K_0$, and therefore $K = K_0$ by the minimality of K and (5.1) (see Figure 7).

If we assume now, for $i = 0, 1, 2$,

$$p_i(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \Omega_i(1) := \{i\text{th yellow region}\}, \tag{5.4}$$

then $K_1 \neq K_0$, $K_2 = K_1$, and iteratively $K_\infty = K_1$. Moreover, by (5.4), $P(\mathbf{x}, K_1^c) = 0$ for all $\mathbf{x} \in K_1$, and therefore $K = K_1$ by the minimality of K and (5.2) (see Figure 8).

5.2.6. *Summary* The complete classification of $\mathcal{K}_m$ in Δ_2 is as follows:

Theorem 5.1. *In Δ_2, we have the following options, which are mutually exclusive:*

 (i) *$\mathcal{K}_m$ consists of 3 vertices.*

 (ii) *$\mathcal{K}_m$ consists of 2 vertices.*

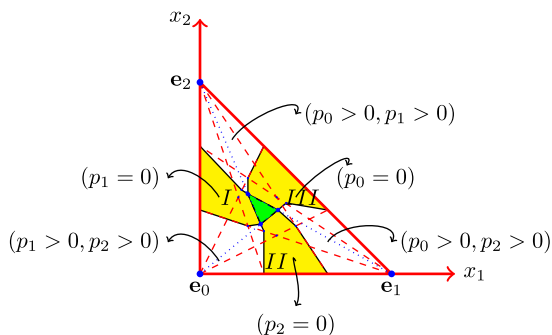 (iii) *$\mathcal{K}_m$ consists of 1 vertex.*

FIGURE 8. $K_1 =$ white domains + yellow domains $= \Delta_2$ minus; green domain; K_1^c is the green domain; $\Omega_1(1)$, $\Omega_2(1)$, and $\Omega_0(1)$ are the yellow domains *I*, *II*, and *III* respectively.

(iv) $\mathcal{K}_m$ consists of 1 edge.

(v) $\mathcal{K}_m$ consists of 1 vertex and 1 opposite edge.

(vi) $\mathcal{K}_m$ consists of a compact subset $K_\infty \subseteq \Delta_2$ such that $K_\infty \cap \Delta_2 \neq \emptyset$. This set K_∞ may equal the whole set Δ_2.

5.3. Asymptotic behavior of $(Z_n)_{n \geq 0}$ in Δ_2

Using Theorem 5.1, we may state the following theorem in Δ_2.

Theorem 5.2. *Let $(Z_n)_{n\geq0}$ be the Diaconis–Freedman chain in Δ_2 with weight functions $p_i(\mathbf{x}) \in \mathcal{H}_\alpha(\Delta_2)$ for some $\alpha \in (0, 1]$. Denote by $\mathcal{P}_{\mathcal{I}nv}(\Delta_2)$ the set of invariant probability measures of $(Z_n)_{n\geq0}$ in Δ_2. Then, we have the following options, which are mutually exclusive:*

(i) *If $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}, \{\mathbf{e}_2\}\}$ then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = co\{\delta_{\mathbf{e}_0}, \delta_{\mathbf{e}_1}, \delta_{\mathbf{e}_2}\}$ and, for all $\mathbf{x} \in \Delta_2$, the chain $(Z_n)_{n\geq0}$ converges $\mathbb{P}_\mathbf{x}$-a.s. to a random variable Z_∞ with values in $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2\}$ and distribution $\mathbb{P}_\mathbf{x}(Z_\infty = \mathbf{e}_i) = h_i(\mathbf{x})$, $i \in \{0, 1, 2\}$, where h_i is a non-negative function in $\mathcal{H}_\alpha(\Delta_2)$ such that $Ph_i = h_i$, $h_0 + h_1 + h_2 \equiv 1$, and $h_i(\mathbf{e}_j) = 0$ for all $i \neq j \in \{0, 1, 2\}$. Moreover, there exist $\kappa > 0$ and $\rho \in [0, 1)$ such that, for all $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ and $\mathbf{x} \in \Delta_2$, $\left|P^n\varphi(\mathbf{x}) - h_0(\mathbf{x})\varphi(\mathbf{e}_0) - h_1(\mathbf{x})\varphi(\mathbf{e}_1) - h_2(\mathbf{x})\varphi(\mathbf{e}_2)\right| \leq \kappa\rho^n\|\varphi\|_\alpha$.*

(ii) *If $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}\}$ then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = co\{\delta_{\mathbf{e}_0}, \delta_{\mathbf{e}_1}\}$ and, for any $\mathbf{x} \in \Delta_2$, the chain $(Z_n)_{n\geq0}$ converges $\mathbb{P}_\mathbf{x}$-a.s. to a random variable Z_∞ with values in $\{\mathbf{e}_0, \mathbf{e}_1\}$ and distribution $\mathbb{P}_\mathbf{x}(Z_\infty = \mathbf{e}_i) = h_i(\mathbf{x})$, $i \in \{0, 1\}$, where h_i is the unique function in $\mathcal{H}_\alpha(\Delta_2)$ such that $Ph_i = h_i$, $h_0 + h_1 \equiv 1$, and $h_i(\mathbf{e}_j) = \delta_{ij}$ for all $i, j \in \{0, 1\}$. Moreover, there exist $\kappa > 0$ and $\rho \in [0, 1)$ such that, for all $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ and $\mathbf{x} \in \Delta_2$, $\left|P^n\varphi(\mathbf{x}) - h_0(\mathbf{x})\varphi(\mathbf{e}_0) - h_1(\mathbf{x})\varphi(\mathbf{e}_1)\right| \leq \kappa\rho^n\|\varphi\|_\alpha$.*
Similar statements hold when $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_2\}\}$ or $\mathcal{K}_m = \{\{\mathbf{e}_1\}, \{\mathbf{e}_2\}\}$.

(iii) *If $\mathcal{K}_m = \{\{\mathbf{e}_0\}\}$ then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = \{\delta_{\mathbf{e}_0}\}$ and, for any $\mathbf{x} \in \Delta_2$, the chain $(Z_n)_{n\geq0}$ converges $\mathbb{P}_\mathbf{x}$-a.s. to $\mathbf{e}_0$. Moreover, there exist $\kappa > 0$ and $\rho \in [0, 1)$ such that, for any $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ and $\mathbf{x} \in \Delta_2$, $\left|P^n\varphi(\mathbf{x}) - \varphi(\mathbf{e}_0)\right| \leq \kappa\rho^n\|\varphi\|_\alpha$.*
Similar statements hold when $\mathcal{K}_m = \{\{\mathbf{e}_1\}\}$ or $\mathcal{K}_m = \{\{\mathbf{e}_2\}\}$.

(iv) *If $\mathcal{K}_m = \{[\mathbf{e}_1, \mathbf{e}_2]\}$ then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = \{\mu_\infty^{12}(d\mathbf{x} \cap [\mathbf{e}_1, \mathbf{e}_2])\}$, where μ_∞^{12} is the probability measure on $[\mathbf{e}_1, \mathbf{e}_2]$ with density*

$$g_\infty^{12}((t, 1-t), (s, 1-s))$$
$$:= C \exp\left(\int_s^t \frac{p_1(u, 1-u)}{1-u} \, du + \int_t^s \frac{p_2(u, 1-u)}{u} \, du \right). \quad (5.5)$$

For any $\mathbf{x} \in \Delta_2$, the chain $(Z_n)_{n \geq 0}$ converges $\mathbb{P}_\mathbf{x}$-a.s. to a random variable Z_∞ with values on $[\mathbf{e}_1, \mathbf{e}_2]$ and distribution $\mu_\infty^{12}(d\mathbf{x} \cap [\mathbf{e}_1, \mathbf{e}_2])$. Moreover, there exist $\kappa > 0$ and $\rho \in [0, 1)$ such that, for any $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ and $\mathbf{x} \in \Delta_2$, $\left| P^n \varphi(\mathbf{x}) - \mu_\infty^{12}(\varphi) \right| \leq \kappa \rho^n \|\varphi\|_\alpha$. Similar statements hold when $\mathcal{K}_m = \{[\mathbf{e}_0, \mathbf{e}_1]\}$ or $\mathcal{K}_m = \{[\mathbf{e}_0, \mathbf{e}_2]\}$.

(v) *If $\mathcal{K}_m = \{\{\mathbf{e}_0\}, [\mathbf{e}_1, \mathbf{e}_2]\}$ then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = co\{\delta_{\mathbf{e}_0}, \mu_\infty^{12}(d\mathbf{x} \cap [\mathbf{e}_1, \mathbf{e}_2])\}$, where μ_∞^{12} is the probability measure on $[\mathbf{e}_1, \mathbf{e}_2]$ with density given by (5.5). Furthermore, for any $\mathbf{x} \in \Delta_2$, the chain $(Z_n)_{n \geq 0}$ converges to $\mathbf{e}_0$ with probability $h_0(\mathbf{x})$ and to $[\mathbf{e}_1, \mathbf{e}_2]$ with probability $h_{12}(\mathbf{x})$, where h_0 and h_{12} are non-negative functions in $\mathcal{H}_\alpha(\Delta_2)$ such that $Ph_0 = h_0$, $Ph_{12} = h_{12}$, $h_0 + h_{12} \equiv 1$, and $h_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in [\mathbf{e}_1, \mathbf{e}_2]$ and $h_{12}(\mathbf{e}_0) = 0$. Moreover, there exist $\kappa > 0$ and $\rho \in [0, 1)$ such that, for all $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ and all $\mathbf{x} \in \Delta_2$, $\left| P^n \varphi(\mathbf{x}) - h_0(\mathbf{x})\varphi(\mathbf{e}_0) - h_{12}(\mathbf{x})\mu_\infty^{12}(\varphi) \right| \leq \kappa \rho^n \|\varphi\|_\alpha$.*

(vi) *If $\mathcal{K}_m = \{K_\infty\}$ (possibly equal to Δ_2) then $\mathcal{P}_{\mathcal{I}nv}(\Delta_2) = \{\mu_\infty\}$, where μ_∞ is a probability measure on Δ_2 with support K_∞ (which possibly equals Δ_2).*

Proof. The operator P is non-negative, and bounded on $\mathcal{H}_\alpha(\Delta_2)$ with spectral radius 1. Moreover, for all $\varphi \in \mathcal{H}_\alpha(\Delta_2)$, we have

$$\|P\varphi\|_\alpha \leq \frac{1}{1+\alpha} m_\alpha(\varphi) + \left(1 + \sum_{i=0}^2 m_\alpha(p_i) \right) \|\varphi\|_\infty.$$

Hence, by [12], the operator P is quasi-compact on $\mathcal{H}_\alpha(\Delta_2)$. Note that $\mathbf{1} \in \mathcal{H}_\alpha(\Delta_2)$ satisfies $P\mathbf{1} = \mathbf{1}$. Therefore, by using [14, Theorem 2.2], the eigenspace corresponding to eigenvalue 1 is nothing but $\ker(P - Id)$. The decomposition of $\ker(P - Id)$ in all six of the above options is then done using [14, Theorem 2.3] (i.e. if $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}, \{\mathbf{e}_2\}\}$ then $\ker(P - Id) = \mathbb{C}\mathbf{1} \oplus \mathbb{C}h_1 \oplus \mathbb{C}h_2$; if $\mathcal{K}_m = \{\{\mathbf{e}_0\}, \{\mathbf{e}_1\}\}$ then $\ker(P - Id) = \mathbb{C}\mathbf{1} \oplus \mathbb{C}h_1$; and so on).

Furthermore, by following [20, Theorem 3.1, Step 3], we easily prove that the peripheral spectrum of P is reduced to $\{1\}$; the argument relies on the fact that, if $\varphi \in \mathcal{H}_\alpha(\Delta_2)$ satisfies $P\varphi = \lambda\varphi$ with $|\lambda| = 1$, then the sequence $(\lambda^{-n}\varphi(Z_n))_{n \geq 0}$ is a bounded martingale, which thus converges $\mathbb{P}$-a.s. (see [20, Lemma 2.4]). This implies the proof in all six of the above options. $\qquad \square$

Remark 5.1. The cases considered in Section 4 where there is a unique invariant probability density all satisfy case (vi) where $\mathcal{K}_m = \{\Delta_2\}$, i.e. when $p_i(\mathbf{e}_i) < 1$ for all $i = 0, 1, 2$. The question of the existence (hence unicity) of an invariant probability density when $p_i(\mathbf{e}_i) < 1$ for all $i = 0, 1, \ldots, d$ is still open for $d \geq 2$ (it has been solved for $d = 1$ in [20]).

6. Discussion

We would like to briefly present here another interesting setting for the Diaconis–Freedman chain in Δ_d. For any $i = 0, \ldots, d$ and $\mathbf{x} \in \Delta_d$, let $S_i(\mathbf{x})$ be the strict convex combination of $\mathbf{x}$ and all vertices $\mathbf{e}_j$ except $\mathbf{e}_i$, i.e. $S_i(\mathbf{x}) := \left(co\{\mathbf{x}, \{\mathbf{e}_j\}_{j \neq i}\} \right)$.
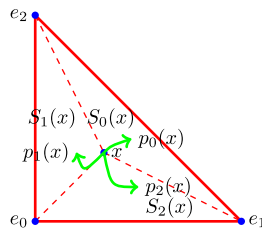
FIGURE 9. An alternative model.

Assume that at time n, a walker $\mathbf{Z}$ is located at site $Z_n = \mathbf{x} \in \Delta_d$ and has probability $p_i(\mathbf{x})$ to move to the domain $S_i(\mathbf{x})$, the arrival point being chosen according to the uniform distribution $U_{S_i}(\mathrm{d}\mathbf{x}))$ on this domain. In other words, the one-step transition probability function of the Markov chain generated by this walker is

$$P(\mathbf{x}, \mathrm{d}\mathbf{y}) = \sum_{i=0}^{d} p_i(\mathbf{x}) \frac{1}{|S_i(\mathbf{x})|} \mathbf{1}_{S_i(\mathbf{x})}(\mathbf{y}) \, \mathrm{d}\mathbf{y}, \qquad \mathbf{x} \in \Delta_d.$$

We illustrate this setting in Δ_2 in Figure 9; such a setting and its applications will be considered in detail elsewhere.

Acknowledgements

Funding Information

Competing Interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BARNSLEY, M. F., DEMKO, S. G., ELTON, J. H. AND GERONIMO, J. S. (1988). Invariant measures for Markov processes arising from iterated functions systems with place-dependent probabilities. *Ann. Inst. H. Poincaré Prob. Statist.* **24**, 367–394.

[2] BARNSLEY, M. F. AND ELTON, J. H. (1988). A new class of Markov processes for image encoding. *Adv. Appl. Prob.* **20**, 14–32.

[3] BARNSLEY, M. F., ELTON, J. H. AND HARDIN, D. P. (1989). Recurrent iterated function systems. *Constr. Approx.* **5**, 3–31.

[4] BOROVKOV A. A. (1998). *Ergodicity and Stability of Stochastic Processes*. Wiley, New York.

[5] BUSH, R. R. AND MOSTELLER, F. (1953). A stochastic model with applications to learning. *Ann. Math. Statist.* **24**, 559–585.

[6] DEGROOT M. H.and RAO M. M. (1963). Stochastic give-and-take. *J. Math. Anal. Appl.* **7**, 489–498.

[7] DIACONIS, P. AND FREEDMAN, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 45–76.

[8] DUBINS, L. E. AND FREEDMAN, D. A. (1966). Invariant probabilities for certain Markov processes. *Ann. Math. Statist.* **37**, 837–848.

[9] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

[10] GHOSH, J. K. AND RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. Springer, New York.

[11] GUIVARC'H, Y. AND RAUGI, A. (1986). Products of random matrices: convergence theorems. In *Random Matrices and their Applications* (Contemp. Math. **50**), eds. J. E. COHEN, H. KESTEN AND C. M. NEWMAN. American Mathematical Society, Providence, RI, pp. 31–54.

[12] HENNION, H. (1993). Sur un théorème spectral et son application aux noyaux lipchitziens. *Proc. Amer. Math. Soc.* 118, 627–634.

[13] HENNION, H. AND HERVÉ, L. c. (2001). *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness* (Lect. Notes Math. 1766). Springer, Berlin.

[14] HERVÉ, L. c. (1994). Étude d'opérateurs quasi-compacts positifs. Applications aux opérateurs de transfert. *Ann. Inst. H. Poincaré Prob. Statist.* 30, 437–466.

[15] HOFRICHTER, J., JOST, J., and TRAN, T. D. (2017). *Information Geometry and Population Genetics*. Springer, Cham.

[16] JOST, J., LE, H., and TRAN, T. (2021). Probabilistic morphisms and Bayesian nonparametrics. *Eur. Phys. J. Plus* **136**, 441.

[17] KAIJSER, T. (1981). On a new contraction condition for random systems with complete connections. *Rev. Roumaine Math. Pure Appl.* **26**, 1075–1117.

[18] KAPICA, R. AND SLECZKA, M. (2017). Random iteration with place-dependent probabilities. Preprint, arXiv:1107.0707.

[19] KARLIN, S. (1953). Some random walks arising in learning models. I. *Pacific J. Math.* **3**, 725–756.

[20] LADJIMI, F. AND PEIGNÉ, M. (2019). On the asymptotic behavior of the Diaconis and Freedman chain on [0, 1]. *Statist. Prob. Lett.* **145**, 1–11.

[21] LETAC, G. (1986). A contraction principle for certain Markov chains and its applications. In *Random Matrices and their Applications* (Contemp. Math. **50**), eds. J. E. Cohen, H. Kesten and C. M. Newman. American Mathematical Society, Providence, RI, pp. 263–273.

[22] MIREK, M. (2011). Heavy tail phenomenon and convergence to stable laws for iterated Lipschitz maps. *Prob. Theory Relat. Fields* **151**, 705–734.

[23] NGUYEN, T.-M. AND VOLKOV, S. (2020). On a class of random walks in simplexes. *J. Appl. Prob.* **57**, 409–428.

[24] PACHECO-GONZALEZ, C. G. AND STOYANOV, J. (2008) A class of Markov chains with beta ergodic distribution. *Math. Scientist* **33**, 110–119.

[25] PEIGNÉ, M. (1993). Iterated function systems and spectral decomposition of the associated Markov operator. *Publications mathématiques et informatique de Rennes* **2**, 1–28.

[26] PEIGNÉ, M. AND WOESS, W. (2011a). Stochastic dynamical systems with weak contractivity properties I. Strong and local contractivity. *Colloq. Math.* **125**, 31–54.

[27] PEIGNÉ, M. AND WOESS, W. (2011b). Stochastic dynamical systems with weak contractivity properties II. Iteration of Lipschitz mappings. *Colloq. Math.* **125**, 55–81.

[28] PIRINSKYB, C. AND STOYANOV, J. (2000) Random motions, classes of ergodic Markov chains and beta distributions. *Statist. Prob. Lett.* **50**, 293–304.

[29] RAMLI, M. A. AND LENG, G. (2010). The stationary probability density of a class of bounded Markov processes. *Adv. Appl. Prob.* **42**, 986–993.

[30] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

[31] STENFLO, O. (2012). A survey of average contractive iterated function systems. *J. Difference Equat. Appl.* **18**, 1355–1380.

[32] TRAN, T., HOFRICHTER, J. AND JOST, J. (2015a). The free energy method and the Wright–Fisher model with 2 alleles. *Theory Biosci.* **134**, 83–92.

[33] TRAN, T., HOFRICHTER, J. AND JOST, J. (2015b). The free energy method for the Fokker–Planck equation of the Wright–Fisher model. MIS-Preprint 29/2015.