A CHARACTERISATION OF TRANSIENT RANDOM WALKS ON STOCHASTIC MATRICES WITH DIRICHLET DISTRIBUTED LIMITS

S. MCKINLAY,* *University of Melbourne*

Abstract

We characterise the class of distributions of random stochastic matrices X with the property that the products $X(n)X(n - 1) \cdots X(1)$ of independent and identically distributed copies $X(k)$ of X converge almost surely as $n \to \infty$ and the limit is Dirichlet distributed. This extends a result by Chamayou and Letac (1994) and is illustrated by several examples that are of interest in applications.

Keywords: Products of random matrices; Dirichlet distribution; limit distribution; Markov chain; random exchange model; random nested simplices; service networks with polling

2010 Mathematics Subject Classification: Primary 60J05
Secondary 60B20; 60F99; 60J20

1. Introduction

Let X be a random $d \times d$ matrix with nonnegative entries, and let $\{X(n)\}_{n \geq 1}$ be an independent and identically distributed (i.i.d.) sequence of random matrices with the same distribution as X. In this paper we study the limit of the left products

$$X(n, 1) := X(n)X(n - 1) \cdots X(1) \tag{1}$$

as $n \to \infty$ for a certain class of random matrices X. Clearly, the distribution of (1) is equal to that of the right product

$$X(1, n) := X(1)X(2) \cdots X(n), \tag{2}$$

and, therefore, any assertions concerning the distribution of (1) as $n \to \infty$ apply to that of (2) as well.

The asymptotic behaviour of the left products (1) was apparently first studied by Bellman [2], who showed, under certain conditions, that $\lim_{n \to \infty} n^{-1} \mathbb{E} \log X(n, 1)_{i,j}$ exists, where $X(n, 1)_{i,j}$ is the entry in the ith row and jth column of $X(n, 1)$. This result was later strengthened to almost-sure convergence in [9], where the behaviour of the norms $\|X(n, 1)\| := \max_{i \leq d} \sum_{j=1}^{d} |X(n, 1)_{i,j}|$ was studied.

Several equivalent conditions for convergence in distribution of the right products (2) without normalisation were given in [12]. See also [3] and [18] for surveys on the properties of the products (1) and (2).

We consider the case when the random matrix X is stochastic, i.e. where all the row sums of X equal 1. The study of the products (1) in this case was apparently initiated by Rosenblatt [21], who applied existing results for compact semigroups. Recent results on the infinite product of deterministic and random stochastic matrices can be found in [23].

The class of stochastic matrices is a semigroup under matrix multiplication, and, therefore, the products (1) and (2) generate a left and a right random walk on stochastic matrices by

$$n \mapsto X(n, 1)$$

and

$$n \mapsto X(1, n), \qquad (3)$$

respectively.

The right random walk (3) is central to several related problems involving distributed averaging. These include distributed computation, distributed optimization, distributed estimation, and distributed coordination (see [23] and the references therein). They are also related to certain Markov processes called Potlatch models. These models were introduced in [11] and [16], and are described in their simplest form in [12]. We will discuss these models further in Section 4.1. A comprehensive reference for random walks on stochastic matrices and more general semigroups is [10], where the transient random walk generated by random stochastic matrices satisfying assumptions (A1)–(A3) below is presented in Appendix B.

In [6], Chamayou and Letac studied the left products (1) for random stochastic matrices X satisfying the following assumptions.

(A1) The rows of X are independent.

(A2) The rows of X are Dirichlet distributed.

(A3) Letting $(\alpha_{i,1}, \ldots, \alpha_{i,d})$ be the Dirichlet parameters of the ith row of X, we have $\sum_{j=1}^{d} \alpha_{i,j} = \sum_{j=1}^{d} \alpha_{j,i}$ for $i = 1, \ldots, d$.

It was shown in [6] that the above assumptions are sufficient to ensure that the following statements hold.

(S1) The products $X(n, 1)$ converge almost surely (a.s.) to some random matrix $\hat{X}$ as $n \to \infty$.

(S2) The limit $\hat{X}$ has identical rows a.s.

(S3) The rows of $\hat{X}$ are Dirichlet distributed.

Unfortunately, assumptions (A1)–(A3) are quite restrictive. Surprisingly, it turns out that statements (S1)–(S3) remain true under much broader conditions, and that none of the assumptions (A1)–(A3) is necessary.

Denote by $\mathcal{K}_d$ the class of all distributions of a random $d \times d$ stochastic matrix X such that (S1)–(S3) hold. In this paper we extend the result of Chamayou and Letac [6] by providing a charaterisation theorem (Theorem 4) for the class $\mathcal{K}_d$. We find that $\mathcal{K}_d$ is much broader than the class of distributions satisfying (A1)–(A3). In addition, once the Dirichlet parameters of the limiting distribution are known, it is elementary to verify that a given distribution belongs to the class $\mathcal{K}_d$.

Chamayou and Letac's result above relies on a theorem (which we present as Theorem 1 below) that admits a rather elegant proof suggested by S. Lauritzen and also presented in [6]. The authors of [6] pointed out, using Erdös's remark on the existence of God's book for the best proofs, that S. Lauritzen's proof could be taken from that book. The proof of our Theorem 3, which contains Theorem 1 as a special case, is also rather simple and concise. Similar to S. Lauritzen's proof, it is based on an insightful observation (Lemma 1 below) that, in our case, extends the well-known result by Pitman on scale independent functions of gamma-distributed random variables (see [20]).

The Dirichlet distribution appearing in the limit (S3) is widely used in statistical applications. These include the modelling of compositional data (see, e.g. [1]), Bayesian analysis, statistical genetics, and nonparametric inference (see [19] for a recent survey on the Dirichlet distribution and its applications).

The paper is structured as follows. Section 2 contains notation and the main results, while their proofs are presented in Section 3. Examples and applications to random exchange models, random nested simplices, and a service network with polling are presented in Section 4.

2. Notation and main results

For positive integers r and c, denote by $\mathcal{M}_{r,c}$ the set of $r \times c$ matrices $(p_{i,j})_{i=1}^{r}{}_{j=1}^{c}$ such that all $p_{i,j} \geq 0$, and by $\mathcal{P}_{r,c} \subset \mathcal{M}_{r,c}$ its subclass of matrices with $\sum_{j=1}^{c} p_{i,j} = 1$, $i = 1, \ldots, r$. Let $\mathcal{M}_{r,c}^{+}$ and $\mathcal{P}_{r,c}^{+}$ be the subclass of all positive matrices from $\mathcal{M}_{r,c}$ and $\mathcal{P}_{r,c}$, respectively (by a positive matrix/vector we mean a matrix/vector with all positive entries), and set $\mathbb{R}_{+} := (0, \infty)$. Clearly, $\mathcal{P}_{d} := \mathcal{P}_{d,d}$ consists of all $d \times d$ stochastic matrices, with $\mathcal{P}_{d}^{+} := \mathcal{P}_{d,d}^{+}$ its subclass of positive stochastic matrices (we similarly define $\mathcal{M}_{d} := \mathcal{M}_{d,d}$ and $\mathcal{M}_{d}^{+} := \mathcal{M}_{d,d}^{+}$). For a matrix $(Z_{i,j}) \in \mathcal{M}_{r,c}$, we denote its row and column sums by $Z_{i\bullet} := \sum_{j=1}^{c} Z_{i,j}$, $i = 1, \ldots, r$, and $Z_{\bullet j} := \sum_{i=1}^{r} Z_{i,j}$, $j = 1, \ldots, c$, respectively. Similarly, for a vector $(Y_{1}, \ldots, Y_{c})$, we denote the sum of its components by $Y_{\bullet} := \sum_{i=1}^{c} Y_{i}$.

For a random element X, we denote its distribution by $\mathcal{L}(X)$, and write $X \sim F$ if $F = \mathcal{L}(X)$. If $Y \sim F_{Y}$ is independent of $X \sim F_{X}$, we write $F_{X} \otimes F_{Y}$ for the law of (X, Y).

We denote by Γ_{u} the gamma distribution with scale parameter 1 and shape parameter $u > 0$, with density

$$\frac{x^{u-1}\mathrm{e}^{-x}}{\Gamma(u)} \mathbf{1}_{(0,\infty)}(x).$$

For a positive vector $\boldsymbol{a} = (a_{1}, \ldots, a_{d})$, we denote by $D_{\boldsymbol{a}}$ the *Dirichlet distribution* on the simplex $\mathcal{S}_{d} := \mathcal{P}_{1,d}$ with density

$$\Gamma(a_{\bullet}) \prod_{i=1}^{d} \frac{x_{i}^{a_{i}-1}}{\Gamma(a_{i})}, \qquad (x_{1}, \ldots, x_{d}) \in \mathcal{S}_{d},$$

with respect to the $(d-1)$-dimensional volume measure on $\mathcal{S}_{d}$. We denote by $G_{\boldsymbol{a}}$ the distribution $\Gamma_{a_{1}} \otimes \cdots \otimes \Gamma_{a_{d}}$ on $\mathcal{M}_{1,d}$.

For $A = (\alpha_{i,j}) \in \mathcal{M}_{r,c}^{+}$, the Dirichlet distribution D_{A} on $\mathcal{P}_{r,c}$ is the law of the matrix $X = (X_{i,j}) \in \mathcal{P}_{r,c}$ such that $X^{(i)} := (X_{i,1}, \ldots, X_{i,c}) \sim D_{(\alpha_{i,1}, \ldots, \alpha_{i,c})}$ and $X^{(1)}, \ldots, X^{(r)}$ are independent. Similarly, by G_{A} we denote the law of the matrix $Z = (Z_{i,j}) \in \mathcal{M}_{r,c}$ such that $Z^{(i)} := (Z_{i,1}, \ldots, Z_{i,c}) \sim G_{(\alpha_{i,1}, \ldots, \alpha_{i,c})}$ and $Z^{(1)}, \ldots, Z^{(r)}$ are independent.

Let $A = (\alpha_{i,j}) \in \mathcal{M}_{r,c}^{+}$. The following theorems are the main results in [6]. They are extensions of earlier theorems by Van Assche [24], who proved them for the case in which $c = r = 2$ and all $\alpha_{i,j} = p > 0$ (see also [25] for an extension to finite dimensions of the result in [24]).

Theorem 1. ([6].) *If* $(Y, X) \sim D_{(\alpha_{1\bullet}, \ldots, \alpha_{r\bullet})} \otimes D_{A}$ *then* $YX \sim D_{(\alpha_{\bullet 1}, \ldots, \alpha_{\bullet c})}$.

Theorem 2. ([6].) *If* $r = c = d$, $X \sim D_{A}$, *and*

$$(\alpha_{1\bullet}, \ldots, \alpha_{d\bullet}) = (\alpha_{\bullet 1}, \ldots, \alpha_{\bullet d}), \tag{4}$$

then $\mathcal{L}(X) \in \mathcal{K}_{d}$, *and* $\hat{X}^{(1)} \sim D_{(\alpha_{1\bullet}, \ldots, \alpha_{d\bullet})}$. *Furthermore, if* Y *is a random vector taking values in* $\mathcal{S}_{d}$ *that is independent of* X, *then* $YX \overset{\mathrm{D}}{=} Y$ *if and only if* $Y \sim D_{(\alpha_{1\bullet}, \ldots, \alpha_{d\bullet})}$.

Theorem 2 is actually a simple consequence of Theorem 1. Likewise, our extension of Theorem 2 (Theorem 4 below) is based on the following result that, as we show at the end of this section, implies Theorem 1.

Theorem 3. *Let* $t = (t_1, \ldots, t_r) \in \mathbb{R}_+^r$ *and* $s = (s_1, \ldots, s_c) \in \mathbb{R}_+^c$ *with* $t_\bullet = s_\bullet$. *Suppose that* X *is a random element of* $\mathcal{M}_{r,c}$ *independent of both* $Y \sim D_t$ *and* $V \sim G_t$. *Then*

$$YX \sim D_s \quad \text{if and only if} \quad VX \sim G_s. \tag{5}$$

To state the next theorem, we need the following conditions on the random matrix $X \in \mathcal{P}_d$.

(C1) There exists a $t \in \mathbb{R}_+^d$ such that, for a random vector $V \sim G_t$ independent of X, we have
$$VX \overset{\text{D}}{=} V.$$

(C2) For an i.i.d. sequence $\{X(n)\}_{n \geq 1}$ with $X(1) \overset{\text{D}}{=} X$, there exists an $m < \infty$ such that
$$\mathbb{P}(X(m, 1) \in \mathcal{P}_d^+) > 0.$$

Remark 1. Note that, unlike the conditions of Theorem 2 (under which (C1) holds), condition (C1) does not mean that X must be positive. In Section 4 we provide examples where (C1) is met, but most of the entries of X are 0s.

Theorem 4. (i) *Relation*

$$\mathcal{L}(X) \in \mathcal{K}_d \tag{6}$$

holds if and only if (C1) and (C2) are met for X.

(ii) *If (6) holds then* $\hat{X}^{(1)} \sim D_t$, *where the vector* t *is the same as in (C1), and if* Y *is a random element of* $\mathcal{S}_d$ *independent of* X *then* $YX \overset{\text{D}}{=} Y$ *if and only if* $Y \overset{\text{D}}{=} \hat{X}^{(1)}$.

We conclude this section by showing that the assertions of Theorems 1 and 2 do follow from those of our Theorems 3 and 4. Let $\xi = (\xi_1, \ldots, \xi_d) \sim G_t$ for some $t \in \mathbb{R}_+^d$. Then it is well known (see, for instance, Formula 2.1.2 of [19]) that

$$\left(\frac{\xi_1}{\xi_\bullet}, \ldots, \frac{\xi_d}{\xi_\bullet} \right) \sim D_t. \tag{7}$$

A remarkable characteristic property of the gamma distribution is as follows (see [17]). Suppose that $\eta_1, \eta_2 > 0$ are independent, nondegenerate random variables. Then η_1 and η_2 are gamma distributed with a common scale parameter if and only if $\eta_1 + \eta_2$ is independent of $\eta_1/(\eta_1 + \eta_2)$. It follows that

$$\xi_\bullet \text{ is independent of } \left(\frac{\xi_1}{\xi_\bullet}, \ldots, \frac{\xi_d}{\xi_\bullet} \right), \tag{8}$$

and so

$$(\xi_1, \ldots, \xi_d) \overset{\text{D}}{=} \left(\frac{\xi_1}{\xi_\bullet}, \ldots, \frac{\xi_d}{\xi_\bullet} \right) \tilde{\xi}_\bullet, \tag{9}$$

where $(\tilde{\xi}_1, \ldots, \tilde{\xi}_d)$ is an independent copy of ξ.

We will now use (7) and relation (9) to show that Theorem 1 follows from Theorem 3. Let $A = (\alpha_{i,j}) \in \mathcal{M}_{r,c}^+$ and $(V, Z) \sim G_{(\alpha_{1\bullet}, \ldots, \alpha_{r\bullet})} \otimes G_A$. Then, from (7), we have

$$X := \begin{pmatrix} \dfrac{Z_{1,1}}{Z_{1\bullet}} & \cdots & \dfrac{Z_{1,c}}{Z_{1\bullet}} \\ \vdots & \ddots & \vdots \\ \dfrac{Z_{r,1}}{Z_{r\bullet}} & \cdots & \dfrac{Z_{r,c}}{Z_{r\bullet}} \end{pmatrix} \sim D_A \tag{10}$$

and

$$Y := \left(\frac{V_1}{V_\bullet}, \ldots, \frac{V_r}{V_\bullet} \right) \sim D_{(\alpha_1 \bullet, \ldots, \alpha_r \bullet)}.$$

Therefore,

$$(V, X) \sim G_{(\alpha_1 \bullet, \ldots, \alpha_r \bullet)} \otimes D_A \tag{11}$$

and

$$(Y, X) \sim D_{(\alpha_1 \bullet, \ldots, \alpha_r \bullet)} \otimes D_A. \tag{12}$$

Now

$$V X = \sum_{k=1}^{r} \left(\frac{Z_{k,1}}{Z_{k \bullet}}, \ldots, \frac{Z_{k,c}}{Z_{k \bullet}} \right) V_k, \tag{13}$$

and relations (9) and (11) imply that the random vector on the right-hand side of (13) is distributed as

$$\sum_{k=1}^{r} (Z_{k,1}, \ldots, Z_{k,c}) = (Z_{\bullet 1}, \ldots, Z_{\bullet c}) \sim G_{(\alpha_{\bullet 1}, \ldots, \alpha_{\bullet c})}. \tag{14}$$

It follows from Theorem 3 that, for Y satisfying (12), we have $Y X \sim D_{(\alpha_{\bullet 1}, \ldots, \alpha_{\bullet c})}$, thus establishing the claim of Theorem 1.

To see that Theorem 2 follows from Theorems 3 and 4, we suppose that $\{X(n)\}_{n \geq 1}$ are i.i.d. with law D_A, and that A satisfies (4). By Theorem 1, if $Y \sim D_{(\alpha_1 \bullet, \ldots, \alpha_d \bullet)}$ is independent of $X(1)$ then $Y X(1) \sim D_{(\alpha_{\bullet 1}, \ldots, \alpha_{\bullet d})}$, and Theorem 3 implies that $V X(1) \overset{D}{=} V$ for $V \sim G_{(\alpha_{\bullet 1}, \ldots, \alpha_{\bullet d})}$ independent of $X(1)$. Thus, (C1) holds with $t = (\alpha_{\bullet 1}, \ldots, \alpha_{\bullet d})$. Since $\mathbb{P}(X(1) \in \mathcal{P}_d^+) = 1$, (C2) is satisfied as well, and the assertion of Theorem 2 follows by applying Theorem 4.

3. Proofs

A remarkable observation made by Pitman [20] is the following extension of (8). Let, as above, $\xi \sim G_t$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be a *scale independent* measurable function, i.e. for any $a \neq 0$,

$$f(ax_1, \ldots, ax_d) \equiv f(x_1, \ldots, x_d).$$

Then the random variable $f(\xi_1, \ldots, \xi_d)$ is independent of $\xi_\bullet$.

The proof of Theorem 3 is based on the following extension of that observation to random functions.

Lemma 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(E, \mathcal{E})$ be a measurable space, and let $X : \Omega \to E$ be a random element. Suppose that $H : \mathbb{R}^r \times E \to [0, \infty)$ is jointly measurable and that, for any $a \neq 0$ and $\omega \in \Omega$,*

$$H(ay_1, \ldots, ay_r, X(\omega)) = H(y_1, \ldots, y_r, X(\omega))$$

for all $(y_1, \ldots, y_r) \in \mathbb{R}^r$. If $V = (V_1, \ldots, V_r) \sim G_t$, $t \in \mathbb{R}_+^r$, is independent of X then $V_\bullet$ is independent of $H(V, X)$.

Proof. Let $\phi(s, u)$, $(s, u) \in \mathbb{R}_+^2$, denote the joint Laplace transform of $V_\bullet$ and $H(V, X)$. Then

$$\phi(s, u) = \mathbb{E} e^{-s(V_1 + \cdots + V_r) - u H(V_1, \ldots, V_r, X)}$$

$$= C_t \int_0^\infty \cdots \int_0^\infty v_1^{t_1 - 1} \cdots v_r^{t_r - 1} e^{-(1+s) \sum v_k} \mathbb{E} e^{-u H(v_1, \ldots, v_r, X)} \, dv_1 \cdots dv_r,$$

where $C_t^{-1} = \prod_{k=1}^{r} \Gamma(t_k)$. Making the substitution $y_k = (1 + s)v_k$ and observing that $H(v_1, \ldots, v_r, X) \stackrel{D}{=} H(y_1, \ldots, y_r, X)$ by virtue of scale independence, we have

$$\phi(s, u) = \frac{C_t}{(1+s)^{\sum t_k}} \int_0^\infty \cdots \int_0^\infty y_1^{t_1-1} \cdots y_r^{t_r-1} e^{-\sum y_k} \mathbb{E} e^{-uH(y_1,\ldots,y_r,X)} \, dy_1 \cdots dy_r,$$

which is the product of two functions, one depending on s and the other on u. Therefore, $V_\bullet$ and $H(V, X)$ are independent as claimed.

The next result is an obvious consequence of Lemma 1.

Corollary 1. *Let $V = (V_1, \ldots, V_r) \sim G_t$, and let $X = (X_{i,j}) \in \mathcal{M}_{r,c}$ be random elements independent of each other. Define the function $H(v_1, \ldots, v_r, X) = (H_1(v_1, \ldots, v_r, X), \ldots, H_c(v_1, \ldots, v_r, X))$ by*

$$H_j(v_1, \ldots, v_r, X) := \sum_{i=1}^{r} \frac{v_i X_{i,j}}{v_\bullet}, \qquad 1 \le j \le c. \tag{15}$$

Then $H(V, X)$ is independent of $V_\bullet$.

Proof of Theorem 3. Suppose that the right-hand side relation in (5) holds, i.e. $VX \sim G_s$ for $V \sim G_t$ independent of X. Then, by Corollary 1, for the scale-independent function H defined in (15), the random vector $H(V_1, \ldots, V_r, X) \equiv (V_1/V_\bullet, \ldots, V_r/V_\bullet)X$ is independent of $V_\bullet$, and, therefore,

$$VX = \left(\frac{V_1}{V_\bullet}, \ldots, \frac{V_r}{V_\bullet}\right)XV_\bullet \stackrel{D}{=} \left(\frac{V_1}{V_\bullet}, \ldots, \frac{V_r}{V_\bullet}\right)X\tilde{V}_\bullet, \tag{16}$$

where $(\tilde{V}_1, \ldots, \tilde{V}_r) \sim G_t$ is independent of (V, X). Since $VX \sim G_s$, for $Z := (Z_1, \ldots, Z_c) \sim G_s$, we have

$$VX \stackrel{D}{=} Z \stackrel{D}{=} \left(\frac{Z_1}{Z_\bullet}, \ldots, \frac{Z_c}{Z_\bullet}\right)\tilde{Z}_\bullet, \tag{17}$$

$(\tilde{Z}_1, \ldots, \tilde{Z}_c)$ being an independent copy of Z (cf. (9)).

Equating the logarithms of the components of the vectors on the right-hand sides of (16) and (17), we obtain

$$\left(\ln\left(\frac{\sum_{i=1}^{r} V_i X_{i,1}}{V_\bullet}\right), \ldots, \ln\left(\frac{\sum_{i=1}^{r} V_i X_{i,c}}{V_\bullet}\right)\right) + \ln(\tilde{V}_\bullet)e_c$$
$$\stackrel{D}{=} \left(\ln\left(\frac{Z_1}{Z_\bullet}\right), \ldots, \ln\left(\frac{Z_c}{Z_\bullet}\right)\right) + \ln(\tilde{Z}_\bullet)e_c, \tag{18}$$

where $e_c := (1, \ldots, 1) \in \mathbb{R}^c$.

Since $t_\bullet = s_\bullet$, we have $\tilde{V}_\bullet \stackrel{D}{=} \tilde{Z}_\bullet$ as both follow $\Gamma_{t_\bullet} = \Gamma_{s_\bullet}$, and so letting ψ, φ, and χ denote the characteristic functions of the first, second (and fourth), and third terms in (18), respectively, we have

$$\psi(u_1, \ldots, u_c)\varphi(u_1, \ldots, u_c) = \chi(u_1, \ldots, u_c)\varphi(u_1, \ldots, u_c).$$

Noting that

$$\varphi(u_1, \ldots, u_c) = \mathbb{E}e^{iu_\bullet \cdot \ln(V_\bullet)}$$

$$= \frac{1}{\Gamma(t_\bullet)} \int_0^\infty e^{iu_\bullet \cdot \ln x} x^{t_\bullet - 1} e^{-x} \, dx$$

$$= \frac{1}{\Gamma(t_\bullet)} \int_0^\infty x^{iu_\bullet + t_\bullet - 1} e^{-x} \, dx$$

$$= \frac{\Gamma(t_\bullet + iu)}{\Gamma(t_\bullet)}$$

$$\neq 0$$

for $t_\bullet > 0$, we conclude that $\psi \equiv \chi$ and, therefore, the respective random vectors have a common distribution. Hence, we have

$$\left(\frac{V_1}{V_\bullet}, \ldots, \frac{V_r}{V_\bullet} \right) X = \left(\frac{\sum_{i=1}^r V_i X_{i,1}}{V_\bullet}, \ldots, \frac{\sum_{i=1}^r V_i X_{i,c}}{V_\bullet} \right) \stackrel{\mathrm{D}}{=} \left(\frac{Z_1}{Z_\bullet}, \ldots, \frac{Z_c}{Z_\bullet} \right) \sim D_s \quad (19)$$

from (7). Observing that the left-hand side of (19) has the form YX for $Y \sim D_t$ independent of X (cf. (7)), we have $YX \sim D_s$, and so the left-hand side relation in (5) holds.

Conversely, suppose that $YX \sim D_s$ for $Y \sim D_t$ independent of X. Using the same steps as above, but following them in the reverse order, we can easily conclude that $VX \sim G_s$ for $V \sim G_t$ independent of X. Theorem 3 is proved.

We will need a simple extension of Proposition 2.2 from [6].

Proposition 1. *Let $X(1)$ be a random element of $\mathcal{P}_d$ satisfying (C2), and let $\{X(n)\}_{n \geq 1}$ be an i.i.d. sequence. Then there exists a random element W of $\mathcal{S}_d$ such that*

$$X(n, 1) \stackrel{\mathrm{a.s.}}{\longrightarrow} e_d^\top W \quad (20)$$

as $n \to \infty$, where $e_d^\top$ denotes the transpose of e_d. Furthermore, if Y is a random element of $\mathcal{S}_d$ that is independent of $X(1)$ then $YX(1) \stackrel{\mathrm{D}}{=} Y$ if and only if $Y \stackrel{\mathrm{D}}{=} W$.

Proof. By Proposition 2.2 of [6], there exists a random element W of $\mathcal{S}_d$ such that

$$X(nm, 1) \stackrel{\mathrm{a.s.}}{\longrightarrow} e_d^\top W \quad (21)$$

as $n \to \infty$, where $m < \infty$ is from (C2). For $k \in (nm, (n+1)m) \cap \mathbb{N}$, we have

$$X(k, 1) = X(k, nm+1)X(nm, 1)$$
$$= X(k, nm+1)e_d^\top W + X(k, nm+1)(X(nm, 1) - e_d^\top W)$$
$$= e_d^\top W + o(1) \quad \text{a.s.}$$

as $n \to \infty$, since $Xe_d^\top = e_d^\top$ for any $X \in \mathcal{P}_d$. Hence, the limit in (20) exists and coincides with that in (21). The second part follows immediately from the proof of Proposition 2.2 of [6].

Proof of Theorem 4. (i) Suppose that (C1) and (C2) hold. Then, by Proposition 1, there exists a random element W of $\mathcal{S}_d$ such that

$$X(n, 1) \stackrel{\mathrm{a.s.}}{\longrightarrow} \hat{X} \equiv e_d^\top W \quad \text{as } n \to \infty,$$

and, therefore, (S1) and (S2) hold. From condition (C1) and Theorem 3, it follows that $YX(1) \stackrel{D}{=} Y$ for $Y \sim D_t$ independent of $X(1)$. By Proposition 1, the law of Y coincides with the distribution of the limit $\hat{X}^{(i)}$, $i = 1, 2, \ldots, d$, i.e.

$$\hat{X}^{(1)} = \hat{X}^{(2)} = \cdots = \hat{X}^{(d)} \stackrel{D}{=} Y \sim D_t, \tag{22}$$

and so (S3) holds. We have proved that $\mathcal{L}(X) \in \mathcal{K}_d$.

Conversely, let $\mathcal{L}(X) \in \mathcal{K}_d$. As $X(m, 1)$ converges a.s. to $\hat{X} \in \mathcal{P}_d^+$, we must have $X(m, 1) \in \mathcal{P}_d^+$ for all large enough m, and so (C2) holds. Furthermore, clearly,

$$\hat{X} = \lim_{n \to \infty} X(n, 1) = \lim_{n \to \infty} X(n, 2)X(1) = \hat{X}'X(1),$$

where $\hat{X}' \stackrel{D}{=} \hat{X}$ is independent of $X(1)$. In particular, we have $\hat{X}^{(1)}X(1) \stackrel{D}{=} \hat{X}^{(1)} \sim D_t$ for some $t \in \mathbb{R}_+^d$. Applying Theorem 3, we have $VX(1) \stackrel{D}{=} V$ for $V \sim G_t$, and so (C1) holds.

(ii) If (6) holds then, by (i), (C1) and (C2) are satisfied, in which case we have (22), where t is the same as in (C1). Furthermore, for Y a random element of $\mathcal{S}_d$ that is independent of X, by Proposition 1, $YX \stackrel{D}{=} Y$ if and only if $Y \stackrel{D}{=} \hat{X}^{(1)}$, where $\hat{X}^{(1)} \sim D_t$ by (22). $\square$

4. Examples and applications

4.1. Random exchange models

Consider the following random exchange model, which is a discrete-time analogue of certain continuous-time Markov processes, called Potlatch models (see, for instance, [11] and [16]). Suppose that we have $d < \infty$ bins labelled by the numbers $1, 2, \ldots, d$ that hold amounts $q_k(n)$, $k = 1, \ldots, d$, of a homogeneous commodity at times $n = 0, 1, 2, \ldots$, respectively. The dynamics of the model are as follows: at time $n \geq 1$, the vector $\boldsymbol{q}(n-1) := (q_1(n-1), \ldots, q_d(n-1))$ changes to $\boldsymbol{q}(n) := \boldsymbol{q}(n-1)X(n)$, where $\{X(n)\}_{n \geq 1}$ are i.i.d. random elements of $\mathcal{P}_d$ with distribution $\mathcal{L}(X)$, which generates the Markov chain

$$\boldsymbol{q}(n) = \boldsymbol{q}(0)X(1, n), \qquad n \geq 1. \tag{23}$$

Clearly, $\sum_{k=1}^d \boldsymbol{q}_k(n) = \sum_{k=1}^d \boldsymbol{q}_k(0)$ for all n, and without loss of generality, we assume that this quantity is equal to 1.

As noted in [6], it is easy to prove that Markov chain (23) has a stationary distribution when $\mathcal{L}(X) \in \mathcal{K}_d$. Indeed, defining the random maps $K_n : \mathcal{S}_d \to \mathcal{S}_d$ by $K_n(\boldsymbol{y}) := \boldsymbol{y}X(n)$, we have, for any $\boldsymbol{y} \in \mathcal{S}_d$,

$$K_1 \circ \cdots \circ K_n(\boldsymbol{y}) = \boldsymbol{y}X(n, 1) \xrightarrow{\text{a.s.}} \hat{X}^{(1)} \quad \text{as } n \to \infty.$$

Applying Proposition 1 of [5], we find that $\mathcal{L}(\hat{X}^{(1)})$ is the stationary distribution of Markov chain (23).

The random exchange model (23) is a higher dimension analogue of the stochastic give-and-take model studied in [7] and introduced in its deterministic form in the context of human genetics in [15]. This model is applicable to the study of dynamical systems (see, for instance, [4]), e.g. it provides a model of a closed economy, where agents exchange real wealth at each step in a random manner.

Now we consider two special cases of random stochastic matrices X with $\mathcal{L}(X) \in \mathcal{K}_d$. The purpose of these examples is to demonstrate that none of the conditions (S1)–(S3) from [6] is necessary. We will also discuss the respective special cases of random exchange model (23). Our first example leads to a generalisation of the result in [6] to the case of 'extended' Dirichlet distributions, where the parameters of the distribution are permitted to be 0. Our second example demonstrates that the rows of X need not be independent.

Example 1. For a vector $\boldsymbol{a} \in \mathcal{M}_{1,d}$, we set $D_{\boldsymbol{a}}$ to be the weak limit of the distribution $D_{\boldsymbol{a}+\varepsilon}$ as $\varepsilon \downarrow 0$, with the usual interpretation of the sum $\boldsymbol{a} + \varepsilon$ with $\boldsymbol{a} \in \mathbb{R}^d$ and $\varepsilon \in \mathbb{R}$. In other words, the components of $\boldsymbol{Y} \sim D_{\boldsymbol{a}}$ that correspond to zero components of $\boldsymbol{a}$ are identically 0, whereas the subvector of $\boldsymbol{Y}$ consisting of the components Y_j of that random vector that correspond to $a_j > 0$ form a usual Dirichlet distributed vector. Likewise, for a matrix $A \in \mathcal{M}_{r,c}$, we set D_A to be the weak limit of the distribution of $D_{A+\varepsilon}$ as $\varepsilon \downarrow 0$. We define the distributions $G_{\boldsymbol{a}}$ and G_A in a similar way.

Let $A = (\alpha_{i,j}) \in \mathcal{M}_d$ with $\alpha_{i\bullet} = \alpha_{\bullet i} > 0$ for $i = 1, \dots, d$. If $X \sim D_A$ and X satisfies (C2), then it turns out that $\mathcal{L}(X) \in \mathcal{K}_d$ and $\hat{X}^{(1)} \sim D_{(\alpha_{1\bullet}, \dots, \alpha_{d\bullet})}$, and so $D_{(\alpha_{1\bullet}, \dots, \alpha_{d\bullet})}$ is the stationary distribution of Markov chain (23).

Indeed, by Theorem 4, it suffices to show that $\boldsymbol{V}X \overset{\mathrm{D}}{=} \boldsymbol{V}$ for $\boldsymbol{V} \sim G_{(\alpha_{1\bullet}, \dots, \alpha_{d\bullet})}$ independent of X. To this end, let $(\boldsymbol{V}, Z) \sim G_{(\alpha_{1\bullet}, \dots, \alpha_{d\bullet})} \otimes G_A$, and define $X \in \mathcal{P}_d$ by (10). Then $\boldsymbol{V}$ is independent of X, and $\boldsymbol{V}X \sim G_{(\alpha_{1\bullet}, \dots, \alpha_{d\bullet})}$ by (13) and (14), as required.

In particular, we have obtained the stationary distribution for the following simple model, which, to the best of the author's knowledge, has not previously been studied: at time $n \geq 1$, a uniform proportion of the commodity previously held in bin k, $k = 1, 2, \dots, d$, is shifted to the (neighbouring) bin $k + 1 \pmod d$. In this case vector $\boldsymbol{q}(n)$ is defined by (23) with

$$
X := \begin{pmatrix}
U_1 & 1 - U_1 & 0 & \cdots & & 0 \\
0 & U_2 & 1 - U_2 & \ddots & & \vdots \\
\vdots & \ddots & \ddots & \ddots & & 0 \\
0 & \cdots & 0 & U_{d-1} & 1 - U_{d-1} \\
1 - U_d & 0 & \cdots & 0 & & U_d
\end{pmatrix}, \tag{24}
$$

where the U_k, $k = 1, \dots, d$, are i.i.d. uniformly distributed random variables on $(0, 1)$. Observing that X defined by (24) satisfies (C2) for $m = d - 1$, it follows from above that $\hat{X}^{(1)} \sim D_{(2, \dots, 2)}$, and so $D_{(2, \dots, 2)}$ is the stationary distribution of Markov chain (23) with $X(n)$ distributed as (24).

Example 2. In this example we consider a random stochastic matrix X with all rows dependent, which generates Markov chain (23) with the same stationary distribution as that corresponding to the random stochastic matrix (24). The behaviour of this model is controlled by the decisions of a 'leader' as follows. At time $n \geq 1$, the 'leader' shifts a uniform proportion of the commodity held in bin 1 to bin 2. If the proportion shifted is greater than $\frac{1}{2}$, then no other shifts occur in the system at time n. However, if the proportion shifted is less than or equal to $\frac{1}{2}$ then the commodity previously held in bin k, $k = 2, 3, \dots, d$, $d \geq 2$, is shifted to the (neighbouring) bin $k + 1 \pmod d$. In this model, the random vector $\boldsymbol{q}(n)$ is given by (23), with the random element $X \in \mathcal{P}_d$ defined by

$$
X := \begin{pmatrix}
U & 1 - U & 0 & \cdots & & 0 \\
0 & I & 1 - I & \ddots & & \vdots \\
\vdots & \ddots & \ddots & \ddots & & 0 \\
0 & \cdots & 0 & I & 1 - I \\
1 - I & 0 & \cdots & 0 & & I
\end{pmatrix}, \tag{25}
$$

where U is a uniform random variable on $(0, 1)$, and $I := \mathbf{1}_{\{U > 1/2\}}$, $\mathbf{1}_A$ being the indicator function for event A.

We will show that $\mathcal{L}(X) \in \mathcal{K}_d$ and $\hat{X}^{(1)} \sim D_{(2,\ldots,2)}$. By Theorem 4, it suffices to show that (C1) and (C2) hold for X, $\boldsymbol{t} = (2, \ldots, 2)$ being the vector from (C1). It is not hard to directly verify that X defined by (25) satisfies (C2) for $m = 2d - 2$. Now let $V \sim G_{(2,\ldots,2)}$ be independent of X, and denote by $\varphi(u_1, \ldots, u_d)$ the characteristic function of VX. Then, setting $u_{d+1} := u_1$, we have

$$
\varphi(u_1, \ldots, u_d) = \mathbb{E} \exp \left\{ i(u_1 U + u_2(1 - U))V_1 + i \sum_{j=2}^{d} (u_j I + u_{j+1}(1 - I))V_j \right\}
$$

$$
= \mathbb{E} \exp \left\{ i \sum_{j=2}^{d} u_{j+1} V_j \right\} \int_0^{1/2} \frac{ds}{[1 - i(u_1 s + u_2(1 - s))]^2}
$$

$$
+ \mathbb{E} \exp \left\{ i \sum_{j=2}^{d} u_j V_j \right\} \int_{1/2}^1 \frac{ds}{[1 - i(u_1 s + u_2(1 - s))]^2}
$$

$$
= \left(\frac{(1 - iu_2)^2}{\prod_{k=1}^d (1 - iu_k)^2} \right) \left(\frac{i}{(1 - iu_2)(2i + u_1 + u_2)} \right)
$$

$$
+ \left(\frac{(1 - iu_1)^2}{\prod_{k=1}^d (1 - iu_k)^2} \right) \left(\frac{i}{(1 - iu_1)(2i + u_1 + u_2)} \right)
$$

$$
= \prod_{k=1}^d (1 - iu_k)^{-2},
$$

which is the characteristic function of V as well, and so we conclude that $VX \overset{\mathrm{D}}{=} V$. Therefore, (C1) holds with $\boldsymbol{t} = (2, \ldots, 2)$, as required.

4.2. Random nested simplices

The study of random triangles and, in particular, nested sequences of random triangles, has been extensive in the probabilistic literature (see, for instance, [14] and the references therein). In one such study [14], the authors used Theorem 1.2 of [6] (see Theorem 2 above) to prove that the limiting distribution of certain random nested simplices is Dirichlet. More precisely, they considered a real $(d-1)$-dimensional affine space E and a convex hull of d points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d$ in E which forms a simplex S. Choosing d points $\boldsymbol{p}_1(1), \ldots, \boldsymbol{p}_d(1)$ in S and taking their convex hull yields a new simplex $S(1) \subset S$. Similarly, choosing points $\boldsymbol{p}_1(2), \ldots, \boldsymbol{p}_d(2)$ in $S(1)$ to obtain $S(2)$ spanned by them, and continuing in this fashion, we obtain a sequence $\{S(n)\}_{n \geq 1}$ of nested simplices.

Suppose that, in the above procedure, the points are chosen at random using the following mechanism. Let $\{X(n)\}_{n \geq 1}$ be a sequence of i.i.d. elements of $\mathcal{P}_d$ with distribution D_A, $A \in \mathcal{M}_d$, satisfying (4). Then, for each $n = 1, 2, \ldots$ and $k = 1, \ldots, d$, the point $\boldsymbol{p}_k(n)$ is specified by its barycentric coordinates given by $X^{(k)}(n)$ with respect to $(\boldsymbol{p}_1(n-1), \ldots, \boldsymbol{p}_d(n-1))$. As shown in [14], the so-defined simplices $S(n)$ shrink to a random point $Y \in E$ as $n \to \infty$, i.e.

$$
\boldsymbol{p}_k(n) \overset{\text{a.s.}}{\longrightarrow} Y, \qquad k = 1, \ldots, d, \tag{26}
$$

where the barycentric coordinates of Y with respect to $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$ are Dirichlet distributed with parameters $(\alpha_{1\bullet}, \ldots, \alpha_{d\bullet})$. We extend this result as follows (see Section 2 in [14] for definitions related to affine spaces, affine frames, and barycentric coordinates).

Theorem 5. *If $X(1)$ satisfies (C1) and (C2), then (26) holds. Furthermore, the barycentric coordinates of Y with respect to $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$ have distribution D_t, t being the vector from (C2).*

Proof. The proof is basically a repetition of that of Theorem 2.2 from [14], except we use Theorem 4 in place of Theorem 2. For $\boldsymbol{B}(0) \in \mathcal{S}_d$, define $\boldsymbol{Z}(n) \in S(n)$ as the point with barycentric coordinates $\boldsymbol{B}(0)$ with respect to the affine frame $(\boldsymbol{p}_1(n), \ldots, \boldsymbol{p}_d(n))$, and denote by $\boldsymbol{B}(n)$ the barycentric coordinates of $\boldsymbol{Z}(n)$ with respect to the affine frame $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$. Then, as shown in the proof of Theorem 2.2 from [14], we have

$$\boldsymbol{B}(n) = \boldsymbol{B}(0)X(n, 1), \qquad n = 0, 1, \ldots.$$

Now suppose that $X(1)$ satisfies (C1) and (C2). By Theorem 4, $\mathcal{L}(X(1)) \in \mathcal{K}_d$, with $\hat{X}^{(1)} = (Y_1, \ldots, Y_d) \sim D_t$, t being the vector from (C2). Then, clearly, $\boldsymbol{B}(n) \xrightarrow{\text{a.s.}} \hat{X}^{(1)}$ as $n \to \infty$ for any $\boldsymbol{B}(0)$. In particular, for any $k = 1, \ldots, d$, we can set $\boldsymbol{B}(0) := (0, \ldots, 0, 1, 0, \ldots, 0)$, where 1 is at the kth place, to obtain $\boldsymbol{p}_k(n) \xrightarrow{\text{a.s.}} Y$ as $n \to \infty$, where Y has barycentric coordinates $(Y_1, \ldots, Y_d)$ with respect to $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$, as required.

4.3. Service networks with polling

The well-known story of Buridan's donkey motivated the authors of [22] to consider the following model. Let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d$ be $d \geq 2$ points in the plane. Starting at a point $\boldsymbol{R}(0)$ inside the convex hull of $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$ at time 0, at time $td + r$, $t \geq 0$, $r \in \{1, \ldots, d\}$, the donkey moves from the point $\boldsymbol{R}(td + r - 1)$ to a point $\boldsymbol{R}(td + r)$ which is uniformly distributed on the straight line segment connecting the points $\boldsymbol{R}(td + r - 1)$ and $\boldsymbol{p}_r$. As noted in [13], although it is easy to prove existence of the stationary distributions for the d homogeneous Markov chains $\{\boldsymbol{R}(td + r)\}_{t \geq 0}$, $r = 1, \ldots, d$, their computation for $d > 2$ is a difficult problem (the authors of [22] focused on the case where $d = 2$).

A modification of that scheme in which the donkey travels inside a simplex S with vertices $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d$ in $\mathbb{R}^{d-1}$ by choosing on step $td + r$ its new position at random inside the convex hull spanned by its current location and vectors from the set $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d\} \setminus \{\boldsymbol{p}_r\}$ was considered in [13]. The main result of [13] establishing the form of the stationary distribution of the thus modified donkey walk was actually proved for yet another version of the model. Namely, rather than the current position of the donkey determining the range for the next step, one instead takes a point with barycentric coordinates (with respect to the affine frame $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$) taken from rows of i.i.d. random stochastic matrices.

Specifically, let $X := (X_{i,j})$ be a random element of $\mathcal{P}_d$, and let $\{X(t)\}_{t \geq 0}$ be a sequence of i.i.d. random matrices with distribution $\mathcal{L}(X)$. The donkey's position at time n is specified by the vector $\boldsymbol{B}(n) = (B_1(n), \ldots, B_d(n)) \in \mathcal{S}_d$ of its barycentric coordinates with respect to the affine frame $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$. Starting at a nonrandom point $\boldsymbol{B}(0)$, for $r = 1, \ldots, d$ and $t = 0, 1, \ldots$, given $\boldsymbol{B}(td + r - 1) = (x_1, \ldots, x_d)$, we have

$$\boldsymbol{B}(td + r) = (x_1, \ldots, x_{r-1}, 0, x_{r+1}, \ldots, x_d) + (x_r X_{r,1}(t), \ldots, x_r X_{r,d}(t)). \qquad (27)$$

In other words, setting $\tilde{\boldsymbol{p}}_r := \sum_{j=1}^{d} X_{r,j}(t) \boldsymbol{p}_j$, the location of the donkey at time $td + r$ will have the same barycentric coordinates in the frame $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{r-1}, \tilde{\boldsymbol{p}}_r, \boldsymbol{p}_{r+1}, \boldsymbol{p}_d)$ as those for the donkey's location at time $td + r - 1$ in the original frame $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d)$. It is clear that each of the sequences $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$, $r = 1, \ldots, d$, forms a homogeneous Markov chain.

This model actually describes the progression of a customer in the following discrete-time closed network with polling. Suppose that there are d buffer nodes in the network, accessed

in cyclic order by a single server, with customers never leaving the network. At time $n = 1$, the server accesses node 1 and moves customers from that node to other nodes in the network randomly, with transition probabilities taken from the first row of a random $X(1) \in \mathcal{P}_d$. At time $n \leq d$, the server accesses node n, and similarly moves customers from that node to other nodes with transition probabilities taken from the nth row of $X(1)$. The same procedure is repeated in a cyclical fashion, with the server accessing node 1 at time $n = d + 1$, and so on, with transition probabilities taken from the rows of random elements $X(j) \in \mathcal{P}_d$ for the jth cycle. In this formulation, the vector $\boldsymbol{B}(n)$ represents the distribution of customers in the network after n steps, with initial distribution $\boldsymbol{B}(0)$.

Markov chains with the cyclic property that each of the sequences $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$, $r = 1, \ldots, d$, forms a homogeneous Markov chain, were also considered in [8], where several examples of such Markov chains were provided.

The following assertion was proved in [13] (instead of the stated Theorem 3 concerning the 'true donkey walk' discussed at the beginning of this section).

Theorem 6. ([13].) *Let $A = (\alpha_{i,j}) \in \mathcal{M}_d^+$ satisfy (4), and let $X \sim D_A$. Then, for $r = 1, \ldots, d$, the stationary distribution of the rth homogeneous Markov chain $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$ given by (27) is Dirichlet with parameters $(\beta_{r,1}, \ldots, \beta_{r,d})$ defined by*

$$
\beta_{r,j} := \begin{cases} \displaystyle\sum_{i=j}^{r} \alpha_{i,j} & \text{for } j \leq r, \\ \displaystyle\sum_{i=j}^{d+r} \alpha_{i,j} & \text{for } j > r, \end{cases}
\tag{28}
$$

with the convention that $\alpha_{i,j} = \alpha_{i,j'}$ when $j = j' \pmod{d}$.

For a matrix $X = (X_{i,j}) \in \mathcal{M}_d$ and $r = 1, 2, \ldots, d$, set

$$
(T_r(X))_{ij} := (\delta_{ij} + X_{r,j}\delta_{ri} - \delta_{rj}\delta_{ij}), \qquad 1 \leq i, j \leq d,
$$

δ_{ij} being Kroneker's delta (i.e. $T_r(X)$ is the $d \times d$ identity matrix with the rth row replaced by the rth row of X).

An extension of Theorem 6 is given by the following theorem.

Theorem 7. *Let X be a random element of $\mathcal{P}_d$ such that $T_1(X) \cdots T_d(X)$ satisfies (C2). Let $V = (V_{i,j}) \sim G_B$ be a random matrix independent of X, where $B \in \mathcal{M}_d^+$ has entries $\beta_{i,j}$ given by (28). Furthermore, suppose that*

$$
V^{(r)} \stackrel{\mathrm{D}}{=} V^{(r-1)}T_r(X), \qquad 1 \leq r \leq d,
\tag{29}
$$

where $V^{(0)} := V^{(d)}$.

Then, for $r = 1, \ldots, d$, the stationary distribution of the rth homogeneous Markov chain $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$ is Dirichlet with parameters $(\beta_{r,1}, \ldots, \beta_{r,d})$.

Proof. Let $\{X(t)\}_{t \geq 0}$ be i.i.d. with law $\mathcal{L}(X)$, and fix $r \in \{1, \ldots, d\}$. Then

$$
R(t) := \begin{cases} T_1(X(t)) \cdots T_d(X(t)) & \text{for } r = d, \\ T_{r+1}(X(t)) \cdots T_d(X(t))T_1(X(t)) \cdots T_r(X(t)) & \text{for } r < d, \end{cases} \qquad t = 0, 1, \ldots,
$$

are i.i.d. random elements in $\mathcal{P}_d$, and

$$\boldsymbol{B}(td + r) = \boldsymbol{B}(r)R(1, t),$$

where $R(1, t)$ is given by (2). Clearly, $R(1)$ satisfies (C2) since $T_1(X) \cdots T_d(X)$ satisfies (C2), and so, by Proposition 1, there exists a random vector $\boldsymbol{W} \in \mathcal{S}_d$ such that

$$R(t, 1) \xrightarrow{\text{a.s.}} \boldsymbol{e}_d^{\top} \boldsymbol{W}$$

as $t \to \infty$. Using Proposition 1 of [5], applied to the space $\mathcal{S}_d$ and to the random mappings from $\mathcal{S}_d$ to $\mathcal{S}_d$ defined by $\boldsymbol{x} \mapsto \boldsymbol{x} Z(t)$, $t \geq 0$, it follows that $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$ has a unique stationary distribution.

Now $V^{(r)} \stackrel{\text{D}}{=} V^{(r)} R(1)$ by (29), and, therefore, Theorem 4 implies that $\hat{R}^{(1)} \sim D_{(\beta_{r,1}, \ldots, \beta_{r,d})}$. Hence, the stationary distribution of $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$ is $D_{(\beta_{r,1}, \ldots, \beta_{r,d})}$, as required.

The following remark shows that our Theorem 7 is indeed an extension of Theorem 6.

Remark 2. Let $A = (\alpha_{i,j}) \in \mathcal{M}_d$ be such that $\alpha_{i\bullet} = \alpha_{\bullet i} > 0$ for $i = 1, \ldots, d$, and $B = (\beta_{i,j}) \in \mathcal{M}_d^+$, where the numbers $\beta_{i,j}$ are given by (28). Let $(V, Z) \sim G_B \otimes G_A$, and $X \sim D_A$ defined by (10) be such that $T_1(X^{(1)}) \cdots T_d(X^{(d)})$ satisfies (C2). Then, for $r = 1, \ldots, d$ and using the convention that $r - 1 = d$ for $r = 1$, we have

$$V^{(r-1)} T_r(X) = (V_{r-1,1}, \ldots, V_{r-1,r-1}, 0, V_{r-1,r+1}, \ldots, V_{r-1,d}) + V_{r-1,r} \left(\frac{Z_{r,1}}{Z_{r\bullet}}, \ldots, \frac{Z_{r,d}}{Z_{r\bullet}} \right)$$

$$\stackrel{\text{D}}{=} (V_{r-1,1}, \ldots, V_{r-1,r-1}, 0, V_{r-1,r+1}, \ldots, V_{r-1,d}) + (Z_{r,1}, \ldots, Z_{r,d}),$$

where the equality in distribution holds by (9) and the fact that $V_{r-1,r} \stackrel{\text{D}}{=} Z_{r\bullet}$, as both follow $\Gamma_{\beta_{r-1,r}} = \Gamma_{\alpha_{r\bullet}}$. It remains to observe that $\beta_{r,r} = \alpha_{r,r}$ and, for $j \neq r$, $\beta_{r,j} = \beta_{r-1,j} + \alpha_{r,j}$; hence, the vector in the last line above is distributed as $V^{(r)}$.

The conditions of Theorem 7 are met, and we conclude that, for all $r = 1, \ldots, d$, the homogeneous Markov chain $\{\boldsymbol{B}(td + r)\}_{t \geq 0}$ has stationary distribution $D_{(\beta_{r,1}, \ldots, \beta_{r,d})}$. In particular, for X defined by (24), we have $\beta_{r,j} = 1$ for $j \neq r + 1 \pmod d$, and $\beta_{r,j} = 2$ for $j = r + 1 \pmod d$.

References

[1] AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data.* Chapman and Hall, London.

[2] BELLMAN, R. (1954). Limit theorems for non-commutative operators. I. *Duke Math. J.* **21,** 491–500.

[3] BOUGEROL, P. AND LACROIX, J. (1985). *Products of Random Matrices with Applications to Schrödinger Operators.* Birkhäuser, Boston, MA.

[4] BRUNEAU, L., JOYE, A. AND MERKLI, M. (2010). Infinite products of random matrices and repeated interaction dynamics. *Ann. Inst. H. Poincaré Prob. Statist.* **46,** 442–464.

[5] CHAMAYOU, J.-F. AND LETAC, G. (1991). Explicit stationary distributions for compositions of random functions and products of random matrices. *J. Theoret. Prob.* **4,** 3–36.

[6] CHAMAYOU, J.-F. AND LETAC, G. (1994). A transient random walk on stochastic matrices with Dirichlet distributions. *Ann. Prob.* **22,** 424–430.

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

[8] Dufresne, D. (1998). Algebraic properties of beta and gamma distributions, and applications. *Adv. Appl. Math.* **20,** 285–299.

[9] Furstenberg, H. and Kesten, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31,** 457–469.

[10] Högnäs, G. and Mukherjea, A. (2011). *Probability Measures on Semigroups,* 2nd edn. Springer, New York.

[11] Holley, R. and Liggett, T. M. (1981). Generalized potlatch and smoothing processes. *Z. Wahrscheinlichkeitsth.* **55,** 165–196.

[12] Kesten, H. and Spitzer, F. (1984). Convergence in distribution of products of random matrices. *Z. Wahrscheinlichkeitsth.* **67,** 363–386.

[13] Letac, G. (2002). Donkey walk and Dirichlet distributions. *Statist. Prob. Lett.* **57,** 17–22.

[14] Letac, G. and Scarsini, M. (1998). Random nested tetrahedra. *Adv. Appl. Prob.* **30,** 619–627.

[15] Li, J. (1961). *Human Genetics.* McGraw-Hill, New York.

[16] Liggett, T. M. and Spitzer, F. (1981). Ergodic theorems for coupled random walks and other systems with locally interacting components. *Z. Wahrscheinlichkeitsth.* **56,** 443–468.

[17] Lukacs, E. (1955). A characterisation of the gamma distribution. *Ann. Math. Statist.* **26,** 319–324.

[18] Mukherjea, A. (2000). *Topics in Products of Random Matrices.* Narosa, New Delhi.

[19] Ng, K. W., Tian, G.-L. and Tang, M.-L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications.* John Wiley, Chichester.

[20] Pitman, E. J. G. (1937). The "closest" estimates of statistical parameters. *Proc. Camb. Phil. Soc.* **33,** 212–222.

[21] Rosenblatt, M. (1965). Products of independent and identically distributed stochastic matrices *J. Math. Anal. Appl.* **11,** 1–10.

[22] Stoyanov, J. and Pirinsky, C (2000). Random motions, classes of ergodic Markov chains and beta distributions. *Statist. Prob. Lett.* **50,** 293–304.

[23] Touri, B. (2012). *Product of Random Stochastic Matrices and Distributed Averaging.* Springer, Berlin.

[24] Van Assche, W. (1986). Products of 2×2 stochastic matrices with random entries. *J. Appl. Prob.* **23,** 1019–1024.

[25] Volodin, N. A., Kotz, S. and Johnson. N. L. (1993). Use of moments in distribution theory: a multivariate case. *J. Multivariate Anal.* **46,** 112–119.