*The Annals of Probability*
1994, Vol. 22, No. 1, 424–430

# A TRANSIENT RANDOM WALK ON STOCHASTIC MATRICES WITH DIRICHLET DISTRIBUTIONS

BY JEAN-FRANÇOIS CHAMAYOU AND GÉRARD LETAC

*Université Paul Sabatier*

Let $X_1$ be a $(d \times d)$ random stochastic matrix such that the rows of $X_1$ are independent, with Dirichlet distributions. The rows of the $(d \times d)$ matrix $A$ are the parameters of these Dirichlet distributions, and we assume that the sums of the rows and columns of $A$ provide the same vector $r = (r_1, \dots, r_d)$. If $(X_n)_{n=1}^{\infty}$ are i.i.d., we prove that $\lim_{n\to\infty}(X_n \cdots X_1)$ almost surely has identical rows, which are Dirichlet distributed with parameter $r$. Van Assche has proved this for $d = 2$ and four identical entries for $A$.

**1. Dirichlet distributions on rectangular matrices.** If $k$ and $d$ are positive integers, $\mathscr{S}_{k,d}$ denotes the set of matrices $(p_{ij})_{i=1}^{k}{}_{j=1}^{d}$ with $k$ rows and $d$ columns such that $p_{ij} \geq 0$ and, for all $i$, $\sum_{j=1}^{d} p_{ij} = 1$. An element of $\mathscr{S}_d = \mathscr{S}_{d,d}$ is called a stochastic matrix of order $d$; $\mathscr{S}_d$ is a semigroup under matrix multiplication. An $\mathscr{S}_d$-valued random variable is called a *random stochastic matrix of order* $d$. If $\alpha = (\alpha_1, \dots, \alpha_d)$ with $\alpha_j > 0$ for all $j$, the *Dirichlet distribution* $D_\alpha$ on $\mathscr{S}_{1,d}$ is defined by its density,

$$\Gamma(\alpha_1 + \cdots + \alpha_d) \prod_{i=1}^{d} (\Gamma(\alpha_i))^{-1} x_i^{\alpha_i - 1}, \tag{1.1}$$

with respect to the Lebesgue measure on the simplex $\mathscr{S}_{1,d}$. More generally, if

$$A = (\alpha_{ij})_{i=1}^{k}{}_{j=1}^{d} \quad \text{with } \alpha_{ij} > 0,$$

we still call the *Dirichlet distribution* $D_A$ on $\mathscr{S}_{k,d}$ the law of the matrix

$$X = (X_{ij})_{i=1}^{k}{}_{j=1}^{d}$$

such that the row $X^{(i)} = (X_{i,1}, X_{i,2}, \dots, X_{i,d})$ has a distribution $D_{(\alpha_{i,1}, \dots, \alpha_{i,d})}$ and such that $X^{(1)}, \dots, X^{(k)}$ are independent.

In [3], Van Assche proves two results that we express in the above notation as follows: Let $p > 0$ and $A = \begin{bmatrix} p & p \\ p & p \end{bmatrix}$. Then the following hold:

1. $\text{Law}(Y, X) = D_{(2p, 2p)} \otimes D_A$ implies $\text{Law}(YX) = D_{(2p, 2p)}$.
2. If $(X_n)_{n=1}^{\infty}$ are i.i.d. with distribution $D_A$, then
$$Z = \lim_{n \to \infty} (X_n X_{n-1} \cdots X_1)$$
exists almost surely, $Z = \begin{pmatrix} Y \\ Y \end{pmatrix}$ and $\text{Law}(Y) = D_{(2p, 2p)}$.

The aim of this note is to extend these results as follows.

Received March 1992; July 1992.

*AMS* 1991 *subject classifications*. Primary 60J15; secondary 60E10.

*Key words and phrases*. Products of random matrices, perturbation of Markov chains, characterization of Dirichlet distributions.

THEOREM 1.1. *Let* $A = (\alpha_{ij})_{i=1}^{k}{}_{j=1}^{d}$, *with* $\alpha_{ij} > 0$. *Denote the margins of* $A$ *by* $r_i = \sum_{j=1}^{d}\alpha_{ij}$ *and* $s_j = \sum_{i=1}^{k}\alpha_{ij}$. *Then* $\text{Law}(Y, X) = D_{(r_1,\dots,r_k)} \otimes D_A$ *implies*

$$\text{Law}(YX) = D_{(s_1,\dots,s_d)}.$$

For instance, if we write for $p$ and $q > 0$,

$$\beta_{p,q}(dx) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}\mathbb{1}_{(0,1)}(x)\,dx,$$

then

$$\text{Law}(C, B, A) = \beta_{p+q,p'+q'} \otimes \beta_{p,q} \otimes \beta_{p',q'}$$

implies $\text{Law}(CB + (1-C)A) = \beta_{p+p',q+q'}$, from Theorem 1.1.

THEOREM 1.2. *Let* $A = (\alpha_{ij})_{i=1}^{d}{}_{j=1}^{d}$, *with* $\alpha_{ij} > 0$, *such that*

$$r_i = \sum_{j=1}^{d} \alpha_{ij} = \sum_{j=1}^{d} \alpha_{ji} \quad \textit{for all } i = 1,\dots,d.$$

*If* $(X_n)_{n=1}^{\infty}$ *are i.i.d. with distribution* $D_A$, *then* $Z = \lim_{n\to\infty}(X_n X_{n-1} \cdots X_1)$ *exists almost surely, all rows of* $Z$ *are identical to some random* $Y$ *of* $\mathscr{S}_{1,d}$ *and* $\text{Law}(Y) = D_{(r_1,\dots,r_d)}$. *Furthermore, if* $Y_1$ *in* $\mathscr{S}_{1,d}$ *is independent of* $X_1$, *then* $\text{Law}(Y_1X_1) = \text{Law}(Y_1)$ *if and only if* $\text{Law}(Y_1) = D_{r_1,\dots,r_d}$.

For instance, if $A$, $B$ and $C$ are independent in $[0,1]$ such that $\text{Law}(B, A) = \beta_{pq} \otimes \beta_{qp}$, then Theorem 1.2 implies that

$$\text{Law}(CB + (1-C)A) = \text{Law}(C) \quad \text{if and only if } \text{Law}(C) = \beta_{p+q,p+q}. \tag{1.2}$$

An application of Theorem 1.2 is provided by considering a stochastic matrix $P = (p_{ij})_{i,j=1}^{d}$ such that $p_{ij} > 0$ for all $i, j$, with stationary distribution $(\pi_i)_{i=1}^{d}$. Assume now that a Markov chain on $\{1,\dots,d\}$, roughly governed by $P$, suffers at each transition $n-1 \to n$ small random perturbations around $P$, that is, $P$ has to be replaced by some random $X_n$ of $\mathscr{S}_d$. Taking $\lambda > 0$ and $A = (\lambda\pi_i p_{ij})_{i,j}$, then the row and column margins of $A$ are $(r_1,\dots,r_d) = \lambda(\pi_1,\dots,\pi_d)$. Assuming that $X_n$ has distribution $D_A$ is an expedient model: One has $\mathbb{E}(X_n) = P$; fluctuations around $P$ are small if $\lambda$ is big. Theorem 1.2 implies that the asymptotic distribution is random with distribution $D_{\lambda(\pi_1,\dots,\pi_d)}$, close to $(\pi_1,\dots,\pi_d)$ if $\lambda$ is big.

Finally, note that we consider in Theorem 1.2 the left random walk on $\mathscr{S}_d$,

$$n \mapsto X_n X_{n-1} \cdots X_1. \tag{1.3}$$

Its transient behavior contrasts with the positive recurrent behavior of the right random walk on $\mathscr{S}_d$,

$$n \mapsto X_1 \cdots X_n. \tag{1.4}$$

The fact that (1.4) has a stationary distribution is easily proved. For instance, using Proposition 1 of [1], applied to the space $E = \mathscr{S}_{1,d}$ and to the random maps $F_n$: $E \to E$ defined by $F_n(y) = yX_n$, it is easy to see that the a.s. convergence of (1.3) as indicated in Theorem 1.2 implies that for any $y$ in $\mathscr{S}_{1,d}$ the Markov chain on $\mathscr{S}_{1,d}$,

$$n \mapsto yX_1 \cdots X_n, \tag{1.5}$$

has the stationary distribution $D_{(r_1,\dots,r_n)}$.

For an arbitrary distribution in $\mathscr{S}_d$ of $X_1$ (with $X_1,\dots,X_n,\dots$ i.i.d.), (1.5) is the simplified version of the Potlatch process which is described on the first page of [2].

**2. Proofs.** We offer two proofs of Theorem 1.1. The first uses the interesting formula (2.1), which leads to (3.1). The second was suggested to us by Stephen Lauritzen (Aalborg). Paul Erdös says that God has a book for the best proofs; this one could be taken from it.

PROPOSITION 2.1. *Let $\alpha = (\alpha_1,\dots,\alpha_d)$ such that $\alpha_j > 0$ for all $j = 1,\dots,d$ and denote $\sigma = \sum_{j=1}^d \alpha_j$. Let $X = (X_1,\dots,X_d)$ be a random variable of $\mathscr{S}_{1,d}$. Then $X$ has distribution $D_\alpha$ if and only if for all $f$ in $(0,+\infty)^d$ one has*

$$\mathbb{E}\big((f_1X_1 + \cdots + f_dX_d)^{-\sigma}\big) = \prod_{i=1}^d f_i^{-\alpha_i}. \tag{2.1}$$

PROOF. ($\Rightarrow$) Clearly, from (1.1), if $(k_1,\dots,k_d) \in \mathbb{N}^d$ and $n = \sum_{j=1}^d k_j$, one has

$$\mathbb{E}\big(X_1^{k_1}\cdots X_d^{k_d}\big) = \frac{\prod_{j=1}^d \alpha_j(\alpha_j+1)\cdots(\alpha_j+k_j-1)}{\sigma(\sigma+1)\cdots(\sigma+n-1)}. \tag{2.2}$$

Let $g$ be in $\mathbb{R}^d$ and $t$ in $\mathbb{R}$ such that $|tg_j| < 1$ for all $j$:

$$\begin{aligned} &\mathbb{E}\big([1 - t(g_1X_1 + \cdots + g_dX_d)]^{-\sigma}\big)\\ &\quad= \sum_{n=0}^\infty \frac{\sigma(\sigma+1)\cdots(\sigma+n-1)}{n!}\mathbb{E}\big([g_1X_1+\cdots+g_dX_d]^n\big)t^n\\ &\quad= \sum_{k_1=0}^\infty \cdots \sum_{k_d=0}^\infty \prod_{j=1}^d \frac{\alpha_j(\alpha_j+1)\cdots(\alpha_j+k_j-1)}{k_j!} t^{k_j} g_j^{k_j}\\ &\quad= \prod_{j=1}^d (1 - tg_j)^{-\alpha_j}, \end{aligned} \tag{2.3}$$

the second equality being obtained with the multinomial expansion and (2.2).

Now to get (2.1), without loss of generality, we take $\sum_{j=1}^d f_j = 1$ and in (2.3) we take $t = 1$ and $g_j = 1 - f_j$.

($\Leftarrow$) Let us take $g$ in $\mathbb{R}^d$ and $t$ in $\mathbb{R}$ such that $|tg_j| < 1$ for all $j$. Applying (2.1) to $f_j = 1 - tg_j$ we get (2.3), which implies that, for all $n$,

$$\mathbb{E}\big([g_1X_1 + \cdots + g_dX_d]^n\big) = \int_{\mathscr{S}_{1,d}} [g_1x_1 + \cdots + g_dx_d]^n\, D_\alpha(dx). \tag{2.4}$$

Since polynomials are dense in the continuous functions on the compact set $\mathscr{S}_{1,d}$, (2.4) implies that $X$ has distribution $D_\alpha$. □

FIRST PROOF OF THEOREM 1.1. Let $f = (f_1, \dots, f_d)$ be in $(0, +\infty)^d$; ${}^t f$ is the transposed column. Thus $YX^t f$ is a real random variable. Write $\sigma = \sum_{i=1}^k \sum_{j=1}^d \alpha_{ij}$ and recall that the rows $X^{(1)}, \dots, X^{(k)}$ are independent. Now we have

$$\begin{aligned} \mathbb{E}\left((YX^t f)^{-\sigma}\right) &= \mathbb{E}\left(\mathbb{E}\left((YX^t f)^{-\sigma} | X\right)\right) \\ &\overset{(1)}{=} \mathbb{E}\left(\prod_{i=1}^k \left(X^{(i)t} f\right)^{-r_i}\right) \\ &\overset{(2)}{=} \prod_{i=1}^k \mathbb{E}\left(\left(X^{(i)t} f\right)^{-r_i}\right) \\ &\overset{(3)}{=} \prod_{i=1}^k \prod_{j=1}^d f_j^{-\alpha_{ij}} = \prod_{j=1}^d f_j^{-s_j}. \end{aligned}$$

Equality (1) comes from Proposition 2.1 and independence of $X$ and $Y$, (2) comes from the independence of the $X^{(i)}$ and (3) comes from Proposition 2.1. From the "if" part of Proposition 2.1, the theorem is proved. □

SECOND PROOF OF THEOREM 1.1. If $a > 0$, the gamma distribution $\gamma_a(dz)$ is

$$(\Gamma(a))^{-1} z^{a-1} \exp(-z) \mathbb{1}_{(0,+\infty)}(z)\, dz.$$

Let $(Z_{ij})_{i=1}^k{}_{j=1}^d$ be a random matrix with independent entries such that $\operatorname{Law} Z_{ij} = \gamma_{a_{ij}}$. Write

$$Z_{i\cdot} = \sum_{j=1}^d Z_{ij}, \qquad Z_{\cdot j} = \sum_{i=1}^k Z_{ij} \quad \text{and} \quad Z_{\cdot\cdot} = \sum_{i=1}^k Z_{i\cdot},$$

and define $X$ in $\mathscr{S}_{k,d}$, $Y$ in $\mathscr{S}_{1,d}$ and $Y'$ in $\mathscr{S}_{1,k}$ by

$$X_{ij} = Z_{ij}/Z_{i\cdot}, \qquad Y_i = Z_{i\cdot}/Z_{\cdot\cdot} \quad \text{and} \quad Y'_j = Z_{\cdot j}/Z_{\cdot\cdot}.$$

Clearly $Y' = YX$, but well-known properties of the Gamma distribution imply that

$$\operatorname{Law}(X) = D_A, \qquad \operatorname{Law}(Y) = D_{(r_1, \dots, r_k)}, \qquad \operatorname{Law}(Y') = D_{(s_1, \dots, s_d)}$$

and that *furthermore* $X$ *and* $Y$ are *independent*. The remainder of the proof is obvious. □

To prove Theorem 1.2, we use a proposition which, like Proposition 2.1, belongs to the folklore. We do not claim novelty, and the proposition is probably buried in the abundant literature on products of random matrices, from which it is not easy to extract the following simple statement (which is a particular case of Lemma 1(b) of [2]).

PROPOSITION 2.2. *Let $(X_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables valued in $\mathscr{S}_d$ such that $P[\min_{ij}(X_1)_{ij} = 0] < 1$. Then there exists $Y$ in $\mathscr{S}_{1,d}$ such that, almost surely,*

$$\lim_{n\to\infty} X_n \cdots X_1 = \begin{pmatrix} Y \\ \vdots \\ Y \end{pmatrix}.$$

*Furthermore, if $Y_1$ is a random variable in $\mathscr{S}_{1,d}$ which is independent of $X_1$, then* $\mathrm{Law}(Y_1 X_1) = \mathrm{Law}(Y_1)$ *if and only if* $\mathrm{Law}(Y_1) = \mathrm{Law}(Y)$.

PROOF. We first prove that if $g$ is in $\mathbb{R}^d$, then there exists a real random variable $Y_g$ such that, almost surely,

$$\lim_{n\to\infty} X_n \cdots X_1 {}^t g = Y_g {}^t(1,\dots,1).$$

To see this, we write

$${}^t g_n = X_n \cdots X_1 {}^t g,$$

and

$$U_n = \min_i (g_n)_i, \qquad V_n = \max_i (g_n)_i.$$

Let $p$ and $q$ in $\{1,\dots,d\}$ be such that, for fixed $n \ge 1$, one has $U_{n-1} = (g_{n-1})_p$ and $V_{n-1} = (g_{n-1})_q$. We have, for all $i$,

$$\begin{aligned} U_{n-1} + (X_n)_{iq}(V_{n-1} - U_{n-1}) &= \sum_{j\neq q} (X_n)_{ij} U_{n-1} + (X_n)_{iq} V_{n-1} \\ &\le \sum_{j=1}^{d} (X_n)_{ij} (g_{n-1})_j = (g_n)_i. \end{aligned}$$

Similarly,

$$(g_n)_i \le V_{n-1} - (X_n)_{ip}(V_{n-1} - U_{n-1}).$$

Thus if $M_n = \min_{ij}(X_n)_{ij}$, we get

$$\begin{aligned} U_{n-1} \le U_{n-1} + M_n(V_{n-1} - U_{n-1}) &\le U_n \le V_n \\ &\le V_{n-1} - M_n(V_{n-1} - U_{n-1}) \le V_{n-1}, \end{aligned}$$

which implies

$$V_n - U_n \le (1 - 2M_n)(V_{n-1} - U_{n-1}) \le (V_0 - U_0) \prod_{j=1}^{n} (1 - 2M_j).$$

Since $P[M_1 = 0] < 1$, clearly almost surely $\sum_{n=1}^{\infty} M_n = +\infty$, $(U_n)_{n=0}^{\infty}$ and $(V_n)_{n=0}^{\infty}$ are adjacent sequences and their common limit $Y_g$ has the required property.

The first part of the proposition is now easily obtained by applying this to each $g$ of the canonical basis of $\mathbb{R}^d$.

To prove the "if" in the second part, one has just to observe that $Y_1$ defined by

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_1 \end{pmatrix} = \lim_{n\to\infty} (X_n \cdots X_2)$$

has the same distribution as $Y$, and that

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_1 \end{pmatrix} X_1 = \begin{pmatrix} Y \\ \vdots \\ Y \end{pmatrix}.$$

To prove the "only if" in the second part, one may assume without loss of generality that $Y_1$ is independent of $(X_n)_{n=1}^{\infty}$. Thus,

$$\mathrm{Law}(Y_1) = \mathrm{Law}(Y_1 X_n) = \mathrm{Law}(Y_1 X_n \cdots X_1) = \mathrm{Law}\, Y_1 \begin{pmatrix} Y \\ \vdots \\ Y \end{pmatrix}.$$

Since $Y_1 \begin{pmatrix} Y \\ \vdots \\ Y \end{pmatrix} = Y$, the proof is done. □

PROOF OF THEOREM 1.2. Clearly, Proposition 2.1 is applicable and there exists a random variable $Y$ in $\mathscr{S}_{1,d}$ such that, almost surely,

$$\lim_{n\to\infty} X_n X_{n-1} \cdots X_1 = \begin{pmatrix} Y \\ \vdots \\ Y \end{pmatrix}.$$

Furthermore, if $Y_1$ is independent of $X_1$ and has distribution $D_{(r_1,\dots,r_d)}$, then from Theorem 1.1 $\mathrm{Law}(Y_1 X_1) = \mathrm{Law}(Y_1)$. The second part of Proposition 2.1 implies that $\mathrm{Law}(Y) = D_{r_1,\dots,r_d}$. □

**3. Comments.** One can see two possible lines of generalization. The first is deceptive; it consists of trying to extend Theorem 1.1 to random Dirichlet measures on an abstract measurable space $(E, \mathscr{E})$. If $\alpha$ is a bounded measure on $(E, \mathscr{E})$ with total mass $\sigma$, a random Dirichlet distribution $X$ with parameter $\alpha$ is a map of some probability space $\Omega$ on the set of probability measures on $(E, \mathscr{E})$ such that, for any finite measurable partition $(A_1, \dots, A_d)$ of $E$ with $\alpha_j = \alpha(A_j)$ positive for all $j = 1 \dots d$, the following hold:

1. the map $\omega \mapsto (X(A_1), \dots, X(A_d))$, $\Omega \to \mathbb{R}^d$, is measurable;
2. the law of $(X(A_1), \dots, X(A_d))$ in $\mathscr{S}_{1,d}$ is $D_{\alpha_1,\dots,\alpha_d}$.

This is standard for generalizing Proposition 2.1 to this framework: A random probability $X$ on $(E, \mathscr{E})$ is a Dirichlet random measure with parameter $\alpha$ if and only if, for any measurable $f\colon E \to (0, +\infty)$ such that $\log f$ is bounded, one has

$$\mathbb{E}\left(\left(\int_E f\,dX\right)^{-\sigma}\right) = \exp\left(-\int_E (\log f)\,d\alpha\right). \tag{3.1}$$

The tool for this is the extension of (2.3); if $X$ is a Dirichlet random probability with parameter $\alpha$, if $g\colon E \to \mathbb{R}$ is measurable and bounded and if $|t|\max|g| < 1$, one has

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\sigma(\sigma+1)\cdots(\sigma+n-1)}{n!}\mathbb{E}\left(\left(\int_E g(x)X(dx)\right)^n\right)t^n \\ = \exp \sum_{n=1}^{\infty} \frac{t^n}{n}\int_E (g(x))^n \alpha(dx).\end{aligned} \tag{3.2}$$

Although (3.1) and (3.2) are attractive, we have not been able to extend Theorem 1.1, which corresponds to the case where $\alpha$ is concentrated on a finite number of atoms, beyond the trite case where $\alpha$ is concentrated on a countable number of atoms.

The second extension is more promising. It consists of using the definition of beta distributions on the cone of positive definite matrices, as is well known in multivariate analysis, and even on a symmetric cone. For this, we must replace the gamma distributions of the second proof of Theorem 1.1 by Wishart distributions on symmetric cones.

Finally, we would be pleased to extend Theorem 1.2 at least to the case where $A$ does not satisfy this rather artificial symmetry condition on margins. However, we do not know how to generalize the Dirichlet distributions, or even the Beta distribution, even in a simple case like $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$.

## REFERENCES

[1] Chamayou, J. F. and Letac, G. (1991). Explicit stationary distributions for compositions of random functions and products. *J. Theoret. Probab.* **4** 3–36.

[2] Kesten, H. and Spitzer, F. (1984). Convergence in distribution of random matrices. *Z. Wahrsch. Verw. Gebiete* **67** 363–386.

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

Laboratoire de Statistique
et Probabilités
U.R.A. C.N.R.S. D0745
118, route de Narbonne
31062 Toulouse Cedex
France