*J. Appl. Prob.* **48,** 766–777 (2011)
*Printed in England*
© *Applied Probability Trust* 2011

# REALIZATION OF AN ERGODIC MARKOV CHAIN AS A RANDOM WALK SUBJECT TO A SYNCHRONIZING ROAD COLORING

KOUJI YANO,* *Kobe University*
KENJI YASUTOMI,** *Ritsumeikan University*

**Abstract**

An ergodic Markov chain is proved to be the realization of a random walk in a directed graph subject to a synchronizing road coloring. The result ensures the existence of appropriate random mappings in Propp–Wilson's coupling from the past. The proof is based on the road coloring theorem. A necessary and sufficient condition for approximate preservation of entropies is also given.

*Keywords:* Markov chain; random walk in a directed graph; road coloring problem; Tsirelson's equation; coupling from the past

2010 Mathematics Subject Classification: Primary 60J10
Secondary 37A35; 05C81; 37H10

## 1. Introduction

Our aim is to realize an ergodic Markov chain as a suitable random walk in a directed graph, which is generated by a sequence of independent and identically distributed (i.i.d.) random variables taking values in the set of mappings of the state space.

### 1.1. Notation

Let $V$ be a set of finite symbols, say $V = \{1, \dots, m\}$. Let $Y = (Y_k)_{k\in\mathbb{Z}}$ be a (time-homogeneous) Markov chain taking values in $V$ and indexed by $\mathbb{Z}$, the set of all integers. We write $Q = (q_{x,y})_{x,y\in V}$ for the one-step transition probability matrix of $Y$, i.e.

$$q_{x,y} = \mathrm{P}(Y_1 = y \mid Y_0 = x), \qquad x, y \in V.$$

The $n$th transition probability matrix is given by the $n$th product $Q^n = (q^n_{x,y})_{x,y\in V}$. We call $Y$ *irreducible* if, for any $x, y \in V$, there exists a positive number $n = n(x, y)$ such that $q^n_{x,y} > 0$. We call $Y$ *aperiodic* if the greatest common divisor among $\{n \geq 1 : q^n_{x,x} > 0\}$ is 1 for all $x \in V$. We call $Y$ *ergodic* if $Y$ is both irreducible and aperiodic, which is equivalent to the condition that there exists a positive integer $r$ such that $q^r_{x,y} > 0$ for all $x, y \in V$.

Let $\Sigma$ denote the set of all mappings from $V$ to itself. For $\sigma_1, \sigma_2 \in \Sigma$ and $x \in V$, we simply write $\sigma_2\sigma_1 x$ for $\sigma_2(\sigma_1(x))$.

Received 14 September 2010; revision received 8 March 2011.
* Current address: Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan.
Email address: kyano@math.kyoto-u.ac.jp
Research supported by KAKENHI (20740060).
** Postal address: Department of Mathematical Sciences, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan.



https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

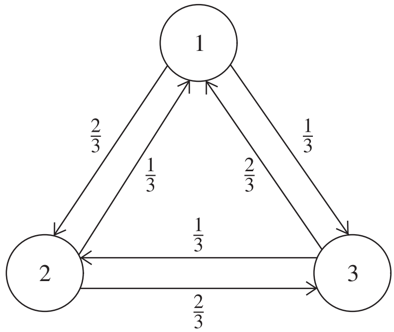


FIGURE 1: Transition probability.

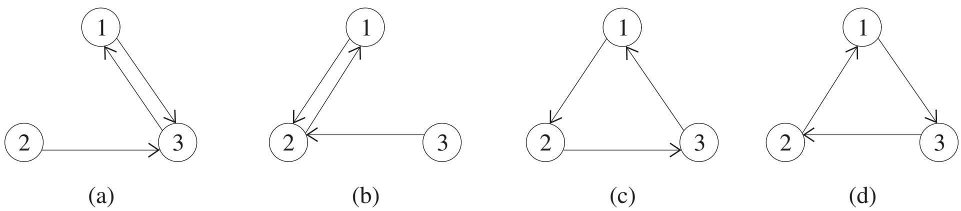


FIGURE 2: The elements (a) $\sigma^{(1)}$, (b) $\sigma^{(2)}$, (c) $\sigma^{(3)}$, and (d) $\sigma^{(4)}$ of $\Sigma$.

**Definition 1.1.** Let $Q$ be the one-step transition probability matrix of a Markov chain. A probability law $\mu$ on $\Sigma$ is called a *mapping law* for $Q$ if

$$q_{x,y} = \sum_{\{\sigma\in\Sigma:\ \sigma x=y\}} \mu(\sigma), \qquad x, y \in V. \tag{1.1}$$

**Definition 1.2.** For a probability law $\mu$ on $\Sigma$, a $\mu$*-random walk* is a Markov chain $(X, N) = (X_k, N_k)_{k\in\mathbb{Z}}$ taking values in $V \times \Sigma$ such that $N = (N_k)_{k\in\mathbb{Z}}$ is i.i.d. with common law $\mu$ such that each $N_k$ is independent of $\sigma(X_j, N_j : j \le k-1)$ and

$$X_k = N_k X_{k-1} \quad \text{almost surely for } k \in \mathbb{Z}. \tag{1.2}$$

Let $Y = (Y_k)_{k\in\mathbb{Z}}$ be an ergodic Markov chain with one-step transition probability matrix $Q$. Let $(X, N)$ be a $\mu$-random walk. Then it is obvious that $Y \overset{\mathrm{D}}{=} X$ if and only if $\mu$ is a mapping law for $Q$. For any ergodic Markov chain $Y$, we can find a mapping law $\mu$ for $Q$ (see Lemma 3.1).

Let us illustrate our notation. See Figure 1, where $V = \{1, 2, 3\}$ and

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{bmatrix} = \begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

Let $\sigma^{(1)}$, $\sigma^{(2)}$, $\sigma^{(3)}$, and $\sigma^{(4)}$ be elements of $\Sigma$, characterized by Figure 2(a), (b), (c), and (d), respectively. The transition probability $Q$ possesses several mapping laws; among others, we have $\mu^{(1)}$ and $\mu^{(2)}$ defined as

$$\mu^{(1)}(\sigma^{(1)}) = \mu^{(1)}(\sigma^{(2)}) = \mu^{(1)}(\sigma^{(3)}) = \tfrac{1}{3}, \tag{1.3}$$

$$\mu^{(2)}(\sigma^{(3)}) = \tfrac{2}{3}, \qquad \mu^{(2)}(\sigma^{(4)}) = \tfrac{1}{3}. \tag{1.4}$$

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

Identity (1.1) can be checked easily; for instance,

$$\sum_{\{\sigma\in\Sigma:\,\sigma(1)=2\}} \mu^{(1)}(\sigma) = \mu^{(1)}(\sigma^{(2)}) + \mu^{(1)}(\sigma^{(3)}) = \frac{2}{3} = q_{1,2}.$$

The two random walks $(X, N)$ corresponding to $\mu^{(1)}$ and $\mu^{(2)}$ have distinct joint laws, but have an identical marginal law of $X$, which is a Markov chain with one-step transition probability $Q$.

### 1.2. Realization of an ergodic Markov chain as a $\mu$-random walk

Our aim is to choose a mapping law $\mu$ which satisfies a nice property.

**Definition 1.3.** A subset $\Sigma_0$ of $\Sigma$ is called *synchronizing* if there exists a sequence $s = (\sigma_p, \ldots, \sigma_1)$ of elements of $\Sigma_0$ such that the composition product $\langle s\rangle := \sigma_p \cdots \sigma_1$ maps $V$ onto a singleton.

We now introduce one of our main theorems.

**Theorem 1.1.** *Suppose that $Y = (Y_k)_{k\in\mathbb{Z}}$ is ergodic. Then we can choose a mapping law $\mu$ for $Q$ such that $\mu$ has synchronizing support.*

Theorem 1.1 will be proved in Section 3.

Let us explain how our $\mu$-random walk is related to road coloring. The support of $\mu$, which we denote by $\{\sigma^{(1)}, \ldots, \sigma^{(d)}\}$, induces the adjacency matrix $A$ of a directed graph $(V, A)$ which is of constant outdegree, i.e. from every site there are $d$ roads laid. Then each element $\sigma^{(1)}, \ldots, \sigma^{(d)}$ may be regarded as a road color so that no two roads from the same site have the same color. For a $\mu$-random walk $(X, N)$, the process $X$ moves in the directed graph $(V, A)$ being driven by the randomly chosen road colors indicated by $N$ via (1.2). Thus, we may call $(X, N)$ *a random walk in a directed graph subject to a road coloring*. For an illustration of the directed graphs induced by $\mu^{(1)}$ and $\mu^{(2)}$, which are defined in (1.3) and (1.4), respectively, see Figure 3(a) and (b), respectively. Since $\sigma^{(1)}\sigma^{(2)}V = \{3\}$, we see that the support of $\mu^{(1)}$ is synchronizing, while we can easily see that the support of $\mu^{(2)}$ is nonsynchronizing.

Let us return to the general discussion. If $(X, N)$ is a $\mu$-random walk and if the support of $\mu$ is synchronizing, then the process $X$ may be represented as

$$X_k = F(N_k, N_{k-1}, \ldots), \qquad k \in \mathbb{Z}, \tag{1.5}$$

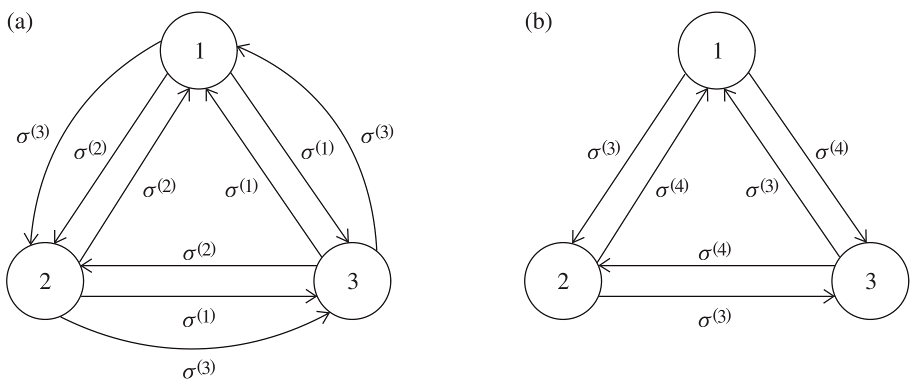


FIGURE 3: The graphs induced by (a) $\mu^{(1)}$ and (b) $\mu^{(2)}$.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

for some measurable function $F: \Sigma^{-\mathbb{N}} \to V$. In fact, define

$$T_k = \max\{l \in \mathbb{Z},\ l < k\colon N_k N_{k-1} \cdots N_l V \text{ is a singleton}\},$$

where we follow the convention that $\max \varnothing = -\infty$. Note that $T_k$ is measurable with respect to $(N_j\colon j \le k)$. Since the support of $\mu$ is synchronizing, it holds that $T_k$ is finite almost surely for all $k \in \mathbb{Z}$, so we may define

$$X_k = N_k N_{k-1} \cdots N_{T_k} x_0, \qquad k \in \mathbb{Z}, \tag{1.6}$$

for a fixed element $x_0 \in V$, but the resulting random walk does not depend on the choice of $x_0$. Such a representation is given in (1.5).

Letting $k = 0$ in identity (1.6), we have

$$X_0 = N_0 N_{-1} \cdots N_{T_0} x_0.$$

This shows that the stationary law of the Markov chain may be simulated exactly from an i.i.d. sequence. This method was a central idea of *Propp–Wilson's coupling from the past* (see [15] and also [11, Chapter 10]). Our Theorem 1.1 ensures theoretically that, for any ergodic Markov chain, there always exists an appropriate mapping law such that Propp–Wilson's algorithm terminates almost surely.

For the study of $\mu$-random walks in the case of nonsynchronizing supports, see [19]. Equation (1.2) is called *Tsirelson's equation in discrete time*; see [3], [12], [13] [20], [21], and [22] for the details.

The representation $Y \overset{\mathrm{D}}{=} X = F(N)$ of $Y$ by an i.i.d. sequence $N$ of the form (1.5) is called a *nonanticipating representation*. Rosenblatt [16], [17] obtained a necessary and sufficient condition for a Markov chain with countable state space to have a nonanticipating representation $Y \overset{\mathrm{D}}{=} X = F(N)$, where $N = (N_k)_{k\in\mathbb{Z}}$ is an i.i.d. sequence with uniform law on [0, 1].

### 1.3. Condition for approximate preservation of entropies

Let $Y$ and $(X, N)$ be as in Theorem 1.1. We examine the entropy information of $Y$ and $N$. See the standard textbook [4] for basic theory of entropies. Let $\lambda$ be the stationary law of $Y$, and define

$$h(Y) = -\sum_{x,y\in V} \lambda(x) q_{x,y} \log q_{x,y}$$

and

$$h(N) = -\sum_{\sigma\in\Sigma} \mu(\sigma) \log \mu(\sigma).$$

Since $Y \overset{\mathrm{D}}{=} X$ and $X$ is a measurable function of $N$ as in (1.6), we have

$$h(Y) \le h(N). \tag{1.7}$$

Note that Ornstein–Friedman's theorem (see [10] and [14]) asserts that two ergodic Markov chains which have common entropy are isomorphic. By this theorem we see that if the equality holds in (1.7) then $Y$ is isomorphic to $N$. We do not have any general criterion on $Y$ for the existence of a mapping law such that $Y$ is isomorphic to $N$. We will give an example for nonexistence in Section 5.

We are interested in a condition for the existence of mapping laws such that the corresponding $h(N)$s approximate the $h(Y)$. Following [16], we introduce the following definition.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

**Definition 1.4.** A Markov chain $Y$ is called *p-uniform* if there exist a probability law $\nu$ on $V$ and a family $\{\tau_x : x \in V\}$ of permutations of $V$ such that

$$q_{x,y} = \nu(\tau_x(y)), \qquad x, y \in V. \tag{1.8}$$

(The prefix p is the first letter of 'permutation'.)

Our second main theorem is as follows.

**Theorem 1.2.** *Let $Y$ be an ergodic Markov chain. Then the following assertions are equivalent.*

(i) *There exists a sequence $\{\mu^{(n)} : n = 1, 2, \dots\}$ of mapping laws for $Q$ with synchronizing support such that the $N^{(n)}$ corresponding to $\mu^{(n)}$ satisfy*

$$h(N^{(n)}) \to h(Y) \quad \text{as } n \to \infty. \tag{1.9}$$

(ii) *$Y$ is p-uniform.*

*In particular, if $h(N) = h(Y)$ holds for $N$ corresponding to some mapping law $\mu$ for $Q$ with synchronizing support, then $Y$ is necessarily p-uniform.*

Theorem 1.2 will be proved in Section 4.

This paper is organized as follows. In Section 2 we introduce the notation needed to state the road coloring problem. Sections 3 and 4 are devoted to the proofs of Theorems 1.1 and 1.2, respectively. In Section 5 we give an example for Theorem 1.2.

## 2. Road colorings of a directed graph

Let $A = [A(y, x)]_{y,x\in V}$ be a $(V \times V)$-dimensional matrix whose entries are nonnegative integers. The pair $(V, A)$ may be called a *directed graph*, where, for $x, y \in V$, the value $A(y, x)$ is regarded as the number of directed edges from $x$ to $y$. The set $V$ is called the *set of vertices* and the matrix $A$ is called the *adjacency matrix*.

The graph $(V, A)$ is called *of constant outdegree* if there exists a constant $d$ such that

$$\sum_{y\in V} A(y, x) = d \quad \text{for all } x \in V.$$

In this case $(V, A)$ is called *d-out*. The graph $(V, A)$ is called *strongly connected* if, for any $x, y \in V$, there exists a positive integer $n = n(x, y)$ such that $A^n(y, x) \geq 1$. The graph $(V, A)$ is called *aperiodic* if the greatest common divisor among $\{n \geq 1 : A^n(x, x) \geq 1\}$ is 1 for all $x \in V$. Note that $(V, A)$ is both strongly connected and aperiodic if and only if there exists a positive integer $r$ such that $A^r(y, x) \geq 1$ for all $x, y \in V$. Following [18], we say that the graph $(V, A)$ or the adjacency matrix $A$ satisfies assumption (AGW) if $(V, A)$ is of constant outdegree, strongly connected, and aperiodic.

Recall that $\Sigma$ is the set of all mappings from $V$ to itself. For $\sigma_1, \sigma_2 \in \Sigma$ and $x \in V$, we simply write $\sigma_2\sigma_1 x$ for $\sigma_2(\sigma_1(x))$. The set $\Sigma$ acts on $V$ in the following sense:

$$(\sigma_1\sigma_2)x = \sigma_1(\sigma_2 x), \qquad \sigma_1, \sigma_2 \in \Sigma,\ x \in V.$$

The set $V = \{1, \dots, m\}$ may be identified with the set of standard basis $\{e_1, \dots, e_m\}$ of $\mathbb{R}^m$. An element $\sigma \in \Sigma$ may be identified with the 1-out adjacency matrix $\sigma = [\sigma(y, x)]_{y,x\in V}$ given by

$$\sigma = [\sigma e_1, \dots, \sigma e_m].$$

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

Under these identifications, we see that, for all $x, y \in V$,

$$\sigma(y, x) = 1 \quad \text{if and only if} \quad y = \sigma x.$$

Let $(V, A)$ be a $d$-out directed graph. A family $\{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ of elements of $\Sigma$ is called a *road coloring* of $(V, A)$ if the following identity holds:

$$A = \sigma^{(1)} + \cdots + \sigma^{(d)}. \tag{2.1}$$

Every road is assigned a color chosen from the $d$ colors $\{\sigma^{(1)}, \dots, \sigma^{(d)}\}$. Here we remark that the elements $\sigma^{(1)}, \dots, \sigma^{(d)}$ are not necessarily distinct. For an illustration, consider

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{bmatrix}$$

and see Figure 3(a) in Section 1. In this case we have $A = \sigma^{(1)} + \sigma^{(2)} + \sigma^{(3)}$ and, hence, $\{\sigma^{(1)}, \sigma^{(2)}, \sigma^{(3)}\}$ is a road coloring of $(V, A)$. Here we remark that the family $\{\sigma^{(3)}, \sigma^{(3)}, \sigma^{(4)}\}$ is another road coloring of $(V, A)$ which is different from $\{\sigma^{(1)}, \sigma^{(2)}, \sigma^{(3)}\}$.

Note that, for any graph $(V, A)$ of constant outdegree, there exists at least one road coloring of $(V, A)$. Conversely, if we are given a family $\{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ of elements of $\Sigma$, then it induces a unique $d$-out directed graph $(V, A)$, where $A$ is defined by (2.1).

Let $\Sigma_0$ be a subset of $\Sigma$. A sequence $s = (\sigma_p, \dots, \sigma_2, \sigma_1)$ of elements of $\Sigma_0$ is called a $\Sigma_0$-*word*. For a $\Sigma_0$-word $s = (\sigma_p, \dots, \sigma_2, \sigma_1)$, we write $\langle s \rangle$ for the product $\sigma_p \cdots \sigma_2 \sigma_1$. The following definition is a slight modification of Definition 1.3.

**Definition 2.1.** A road coloring $\Sigma_0 = \{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ is called *synchronizing* if $\Sigma_0$ as a subset of $\Sigma$ is synchronizing.

By this definition we see that a road coloring $\Sigma_0 = \{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ is synchronizing if and only if $\langle s \rangle V$ is a singleton for some $\Sigma_0$-word $s$. If we express

$$s = (\sigma^{(i(p))}, \dots, \sigma^{(i(2))}, \sigma^{(i(1))})$$

for some numbers $i(1), \dots, i(p) \in \{1, \dots, d\}$, the assertion '$\langle s \rangle V$ is a singleton' may be stated in other words as follows. Those who walk in the graph $(V, A)$ according to the colors $\sigma^{(i(1))}, \dots, \sigma^{(i(p))}$ in this order will lead to a common vertex, no matter where they started from.

Now we state the *road coloring theorem.*

**Theorem 2.1.** ([18].) *Suppose that the directed graph* $(V, A)$ *satisfies assumption (AGW). Then there exists a synchronizing road coloring of* $(V, A)$.

This was first conjectured in the case of no multiple directed edges by Adler *et al.* [1] (see also [2, Section 11]) in the context of the isomorphism problem of symbolic dynamics with common topological entropy. For related studies published prior to that of Trahtman [18], see [8] and [9]; see also [5], [6], and [7].

## 3. Construction of a mapping law on a synchronizing road coloring

We need the following lemma.

**Lemma 3.1.** *Let* $Y$ *be a Markov chain with one-step transition probability matrix* $Q$. *Then there exists a mapping law* $\mu$ *for* $Q$.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

*Proof.* First, we suppose that $q_{x,y}$ is a rational number for all $x, y \in V$. Then we may take an integer $d$ sufficiently large so that $A(y, x) := q_{x,y}d$ is an integer for all $x, y \in V$. Then $A := [A(y, x)]_{x,y \in V}$ is the adjacency matrix of a $d$-out directed graph $(V, A)$; in fact,

$$\sum_{y \in V} A(y, x) = d \sum_{y \in V} q_{x,y} = d.$$

Let $\{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ be a road coloring of $(V, A)$, and define

$$\mu(\sigma) = \frac{1}{d} \#(\{i = 1, \dots, d : \sigma^{(i)} = \sigma\}),$$

where $\#(\cdot)$ denotes the number of elements of the set indicated. Thus, for any $x, y \in V$, we see that

$$\sum_{\{\sigma \in \Sigma : y = \sigma x\}} \mu(\sigma) = \frac{1}{d} \#(\{i = 1, \dots, d : \sigma^{(i)}(y, x) = 1\}) = \frac{1}{d} A(y, x) = q_{x,y},$$

which shows that $\mu$ is a mapping law for $Q$.

Second, we consider the general case. Let us take a sequence $\{Q^{(n)} : n = 1, 2, \dots\}$ of one-step transition probability matrices such that $q^{(n)}_{x,y}$ is a rational number for all $n$ and $x, y \in V$ and that $q^{(n)}_{x,y} \to q_{x,y}$ as $n \to \infty$ for all $x, y \in V$. Then, for any $n$, there exists a mapping law $\mu^{(n)}$ for $Q^{(n)}$. Since $\Sigma$ is a finite set, we can choose some subsequence $\{\mu^{(n(k))} : k = 1, 2, \dots\}$ and some probability law $\mu$ on $\Sigma$ such that $\mu^{(n(k))}(\sigma) \to \mu(\sigma)$ as $k \to \infty$. This shows that $\mu$ is a mapping law for $Q$. This completes the proof.

Now we proceed to prove Theorem 1.1.

*Proof of Theorem 1.1.* Let $Q = (q_{x,y})_{x,y \in V}$ be the one-step transition probability matrix for an ergodic Markov chain $Y$.

First, we take an adjacency matrix $A$ which is of constant outdegree and satisfies

$$A(y, x) \begin{cases} \geq 1 & \text{if } q_{x,y} > 0, \\ = 0 & \text{if } q_{x,y} = 0. \end{cases} \tag{3.1}$$

For this, we introduce a subset $V \times V$ defined by

$$E = \{(x, y) \in V \times V : q_{x,y} > 0\}.$$

For each $x \in V$, we define the outdegree of $E$ at $x$ by

$$d(x) = \#\{(x, y) \in E : y \in V\},$$

and write $d = \max_{x \in V} d(x)$ for the maximum outdegree of $E$. For each $x \in V$, we choose a site $\sigma(x) \in V$ so that $(x, \sigma(x)) \in E$. Now we set

$$A(y, x) = \begin{cases} d - d(x) + 1 & \text{if } y = \sigma(x), \\ 1 & \text{if } y \neq \sigma(x) \text{ and } (x, y) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Then this $(A(y, x))_{x,y \in V}$ is of constant outdegree and satisfies (3.1).

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

Since $Y$ is an ergodic Markov chain, there exists a positive integer $r$ such that $q^r_{x,y} > 0$ for all $x, y \in V$. Hence, we have $A^r(y, x) \geq 1$ for all $x, y \in V$; in fact, there exists a path $x = x_0, x_1, \dots, x_n = y$ such that $q_{x_{k-1},x_k} > 0$ for $k = 1, 2, \dots, n$, which implies that $A(x_k, x_{k-1}) \geq 1$ for $k = 1, 2, \dots, n$. Thus, we see that $(V, A)$ satisfies assumption (AGW). This means that we can apply Theorem 2.1 to obtain a synchronizing road coloring $\{\sigma^{(1)}, \dots, \sigma^{(d)}\}$ of $(V, A)$. Define

$$\hat{\mu}(\sigma) = \frac{1}{d}\#(\{i = 1, \dots, d : \sigma^{(i)} = \sigma\}), \qquad \sigma \in \Sigma,$$

and define

$$\hat{q}_{x,y} = \sum_{\{\sigma \in \Sigma :\, y = \sigma x\}} \hat{\mu}(\sigma), \qquad x, y \in V.$$

Then $\hat{\mu}$ is a mapping law for $\hat{Q}$ and has synchronizing support. We also note that

$$\hat{q}_{x,y} = 0 \quad \text{if } (x, y) \notin E.$$

Let

$$\varepsilon = \min\{q_{x,y} : (x, y) \in E\} > 0.$$

If $\varepsilon = 1$ then we have $Q = \hat{Q}$, so that $\hat{\mu}$ is as desired. Let us assume that $\varepsilon < 1$. Define

$$Q^{(\varepsilon)} = \frac{1}{1 - \varepsilon}(Q - \varepsilon \hat{Q}).$$

Then $Q^{(\varepsilon)} = (q^{(\varepsilon)}_{x,y})_{x,y \in V}$ is a one-step transition probability matrix of a Markov chain. In fact, we see that

$$(1 - \varepsilon) q^{(\varepsilon)}_{x,y} = q_{x,y} - \varepsilon \hat{q}_{x,y} \geq q_{x,y} - \varepsilon\, \mathbf{1}_{\{(x,y) \in E\}} \geq 0, \qquad x, y \in V,$$

and that

$$\sum_{y \in V} q^{(\varepsilon)}_{x,y} = \frac{1}{1 - \varepsilon}\left(\sum_{y \in V} q_{x,y} - \varepsilon \sum_{y \in V} \hat{q}_{x,y}\right) = 1.$$

Now we apply Lemma 3.1 to obtain a mapping law $\mu^{(\varepsilon)}$ for $Q^{(\varepsilon)}$. Define

$$\mu = (1 - \varepsilon)\mu^{(\varepsilon)} + \varepsilon \hat{\mu}.$$

Since $\mu^{(\varepsilon)}$ has synchronizing support, so does $\mu$. For $x, y \in V$, we have

$$\begin{aligned}
\sum_{\{\sigma \in \Sigma :\, y = \sigma x\}} \mu(\sigma) &= (1 - \varepsilon) \sum_{\{\sigma \in \Sigma :\, y = \sigma x\}} \mu^{(\varepsilon)}(\sigma) + \varepsilon \sum_{\{\sigma \in \Sigma :\, y = \sigma x\}} \hat{\mu}(\sigma) \\
&= (1 - \varepsilon) q^{(\varepsilon)}_{x,y} + \varepsilon \hat{q}_{x,y} \\
&= q_{x,y},
\end{aligned}$$

which shows that $\mu$ is a mapping law for $Q$. This completes the proof.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

## 4. Approximate preservation of entropies

Let us prove Theorem 1.2.

*Proof of Theorem 1.2.* We prove that (i) implies (ii). Note that

$$h(Y) = -\sum_{x,y\in V} \lambda(x) q_{x,y} \log q_{x,y}, \tag{4.1}$$

$$h(N^{(n)}) = -\sum_{\sigma\in\Sigma} \mu^{(n)}(\sigma) \log \mu^{(n)}(\sigma).$$

Taking a subsequence if necessary, we assume that there exists a probability law $\mu$ on $\Sigma$ such that $\mu^{(n)}(\sigma) \to \mu(\sigma)$ for all $\sigma \in \Sigma$. Note that $\mu$ is a mapping law for $Q$ but does not necessarily have synchronizing support. By assumption (1.9) we see that

$$h(Y) = \lim_{n\to\infty} h(N^{(n)}) = -\sum_{\sigma\in\Sigma} \mu(\sigma) \log \mu(\sigma).$$

For $x, y \in V$, we set

$$\Sigma(y, x) = \{\sigma \in \Sigma : y = \sigma x\},$$

so that we have

$$q_{x,y} = \sum_{\sigma\in\Sigma(y,x)} \mu(\sigma).$$

Hence, we have

$$\mu(\sigma) \le q_{x,y} \quad \text{whenever } \sigma \in \Sigma(y, x). \tag{4.2}$$

Since $t \mapsto \log t$ is increasing, we have

$$-\sum_{\sigma\in\Sigma(y,x)} \mu(\sigma) \log \mu(\sigma) \ge -\sum_{\sigma\in\Sigma(y,x)} \mu(\sigma) \log q_{x,y} = -q_{x,y} \log q_{x,y}. \tag{4.3}$$

Since $\bigcup_{y\in V} \Sigma(y, x) = \Sigma$, we have

$$h(Y) = -\sum_{y\in V} \sum_{\sigma\in\Sigma(y,x)} \mu(\sigma) \log \mu(\sigma) \ge q(x) \quad \text{for all } x \in V, \tag{4.4}$$

where we set

$$q(x) = -\sum_{y\in V} q_{x,y} \log q_{x,y}, \qquad x \in V.$$

We take $\hat{x} \in V$ such that

$$q(\hat{x}) = \max_{x\in V} q(x).$$

Using (4.4) and (4.1), we have

$$q(\hat{x}) \le h(Y) = \sum_{x\in V} \lambda(x) q(x) \le q(\hat{x}). \tag{4.5}$$

Thus, we see that the equalities hold in (4.5) and that $q(x) = q(\hat{x})$ for all $x \in V$. For any $x \in V$, we combine $h(N) = q(x)$ together with (4.3) to obtain

$$-\sum_{\sigma\in\Sigma(y,x)} \mu(\sigma) \log \mu(\sigma) = -q_{x,y} \log q_{x,y}, \qquad x, y \in V.$$

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

Combining this with (4.2), we obtain

$$\mu(\sigma) = q_{x,y} \quad \text{whenever } \sigma \in \Sigma(y, x).$$

Let $x_0 \in V$ be fixed, and let $x \in V$. Since $\{\Sigma(y, x) \colon y \in V\}$ is a partition of $\Sigma$, we may choose a permutation $\tau_x$ of $V$ so that

$$\Sigma(\tau_x(y), x) \cap \Sigma(y, x_0) \neq \varnothing, \qquad y \in V.$$

This shows that

$$q_{x,\tau_x(y)} = q_{x_0,y}, \qquad x, y \in V,$$

which implies p-uniformity of $Y$. This completes the proof of the implication (i) $\Rightarrow$ (ii).

We now prove that (ii) implies (i). Let $\{x_1, \ldots, x_d\}$ be an enumeration of the support of the law $\nu$ in (1.8). For $i = 1, \ldots, d$, we define

$$\sigma^{(i)}(y, x) = \mathbf{1}_{\{\tau_x(y)=x_i\}}.$$

For each $x \in V$, there exists a unique $y \in V$ such that $\sigma^{(i)}(y, x) = 1$, so that we have $\sigma^{(i)} \in \Sigma$. By (1.8) we obtain

$$q_{x,y} = \sum_{i=1}^{d} \sigma^{(i)}(y, x)\nu(x_i), \qquad x, y \in V.$$

Let $A$ be as in (3.1), and let $\Sigma_1$ be a synchronizing subset corresponding to some synchronizing road coloring of $(V, A)$. For a sufficiently large integer $n$, we define a probability law $\mu^{(n)}$ on $\Sigma$ by

$$\mu^{(n)}(\sigma) = \sum_{\{i \colon \sigma^{(i)}=\sigma\}} \left\{ \nu(x_i) - \frac{1}{nd} \right\} + \frac{1}{n|\Sigma_1|} \mathbf{1}_{\{\sigma\in\Sigma_1\}}.$$

Then it is obvious that $\mu^{(n)}$ is a mapping law for $Q$ and has synchronizing support.

Let us verify condition (1.9). On the one hand, we have

$$h(N^{(n)}) \to -\sum_{i=1}^{d} \nu(x_i) \log \nu(x_i) \quad \text{as } n \to \infty.$$

On the other hand, we have

$$\begin{aligned} h(Y) &= -\sum_{x,y\in V} \lambda(x) q_{x,y} \log q_{x,y} \\ &= -\sum_{x,y\in V} \lambda(x) \sum_{i=1}^{d} \sigma^{(i)}(y, x)\nu(x_i) \log \nu(x_i) \\ &= -\sum_{i=1}^{d} \left\{ \sum_{x,y\in V} \lambda(x)\sigma^{(i)}(y, x) \right\} \nu(x_i) \log \nu(x_i) \\ &= -\sum_{i=1}^{d} \nu(x_i) \log \nu(x_i). \end{aligned}$$

This shows (1.9), completing the proof.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

## 5. An example

Let $V = \{1, 2\}$. Then $\Sigma = \{(12), (21), (11), (22)\}$, where

$$(ij) = \begin{bmatrix} 1 \mapsto i \\ 2 \mapsto j \end{bmatrix}, \qquad i, j = 1, 2.$$

Let $0 < p < 1$, and consider a Markov chain $Y$ with one-step transition probability given by

$$\begin{bmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{bmatrix} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Then it is obvious that $Y$ is an ergodic Markov chain. Since

$$\begin{bmatrix} q_{1,1} \\ q_{2,1} \end{bmatrix} = \begin{bmatrix} q_{2,2} \\ q_{1,2} \end{bmatrix} = \begin{bmatrix} p \\ 1-p \end{bmatrix},$$

we see that $Y$ is p-uniform.

It is obvious that the stationary law is given as

$$\lambda(1) = \lambda(2) = \tfrac{1}{2}.$$

We now see that

$$h(Y) = \varphi(p) + \varphi(1-p),$$

where $\varphi(t) = -t \log t$.

If $\mu$ is a mapping law for $Q$ then we have

$$\mu(12) + \mu(11) = p, \qquad \mu(21) + \mu(11) = 1 - p.$$

From this, we see that there exists some $\varepsilon$ with $0 \le \varepsilon \le \min\{p, 1-p\}$ such that

$$\varepsilon = \mu(11) = \mu(22), \qquad \mu(12) = p - \varepsilon, \qquad \mu(21) = 1 - p - \varepsilon. \tag{5.1}$$

Conversely, for any $\varepsilon$ with $0 \le \varepsilon \le \min\{p, 1-p\}$, we may define $\mu = \mu^{(\varepsilon)}$ by (5.1) so that $\mu^{(\varepsilon)}$ is a mapping law for $Q$.

If $\mu^{(\varepsilon)}$ has synchronizing support, $\varepsilon$ should be positive. Let $\{X^{(\varepsilon)}, N^{(\varepsilon)}\}$ be the $\mu^{(\varepsilon)}$-random walk. We then see that

$$h(N^{(\varepsilon)}) = 2\varphi(\varepsilon) + \varphi(p - \varepsilon) + \varphi(1 - p - \varepsilon).$$

If $p = \frac{1}{2}$, we see that $h(Y) = h(N^{(1/2)})$.

Suppose that $p \ne \frac{1}{2}$. Then, by an easy computation we see that

$$h(Y) < h(N^{(\varepsilon)})$$

for all $\varepsilon$ with $0 < \varepsilon \le \min\{p, 1-p\}$. However, it holds that $h(N^{(\varepsilon)}) \to h(Y)$ as $\varepsilon \to 0+$.

https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press

## References


[1] Adler, R. L., Goodwyn, L. W. and Weiss, B. (1977). Equivalence of topological Markov shifts. *Israel J. Math.*, **27,** 48–63.

[2] Adler, R. L. and Weiss, B. (1970). *Similarity of Automorphisms of the Torus* (Memoirs Amer. Math. Soc. **98**). American Mathematical Society, Providence, RI.

[3] Akahori, J., Uenishi, C. and Yano, K. (2008). Stochastic equations on compact groups in discrete negative time. *Prob. Theory Relat. Fields* **140,** 569–593.

[4] Billingsley, P. (1978). *Ergodic Theory and Information*. Robert E. Krieger Publishing, Huntington, NY.

[5] Budzban, G. (2004). Semigroups and the generalized road coloring problem. *Semigroup Forum* **69,** 201–208.

[6] Budzban, G. and Feinsilver, P. (2007). Completely simple semigroups, Lie algebras, and the road coloring problem. *Semigroup Forum* **74,** 206–226.

[7] Budzban, G. and Mukherjea, A. (2000). A semigroup approach to the road coloring problem. In *Probability on Algebraic Structures* (Gainesville, FL, 1999; Contemp. Math. **261**), American Mathematical Society, Providence, RI, pp. 195–207.

[8] Culik, K., II, Karhumäki, J. and Kari, J. (2002). A note on synchronized automata and road coloring problem. In *Developments in Language Theory* (Vienna, 2001; Lecture Notes Comput. Sci. **2295**), Springer, Berlin, pp. 175–185.

[9] Friedman, J. (1990). On the road coloring problem. *Proc. Amer. Math. Soc.* **110,** 1133–1135.

[10] Friedman, N. A. and Ornstein, D. S. (1970). On isomorphism of weak Bernoulli transformations. *Adv. Math.* **5,** 365–394.

[11] Häggström, O. (2002). *Finite Markov Chains and Algorithmic Applications* (London Math. Soc. Student Texts **52**). Cambridge University Press.

[12] Hirayama, T. and Yano, K. (2010). Extremal solutions for stochastic equations indexed by negative integers and taking values in compact groups. *Stoch. Process. Appl.* **120,** 1404–1423.

[13] Hirayama, T. and Yano, K. (2010). Strong solutions of Tsirelson's equation in discrete time taking values in compact spaces with semigroup action. Preprint. Available at http://arxiv.org/abs/1005.0038v1.

[14] Ornstein, D. (1970). Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.* **4,** 337–352.

[15] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Internat. Random Structures Algorithms* **9,** 223–252.

[16] Rosenblatt, M. (1959). Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* **8,** 665–681.

[17] Rosenblatt, M. (1962). Stationary Markov chains and independent random variables. *J. Math. Mech.* **9,** 945–949. (Addendum: **11** (1962), 317.)

[18] Trahtman, A. N. (2009). The road coloring problem. *Israel J. Math.* **172,** 51–60.

[19] Yano, K. (2010). Random walk in a finite directed graph subject to a road coloring. Preprint. Available at http://arxiv.org/abs/1005.0079v3.

[20] Yano, K. and Takahashi, Y. (2007). Time evolution with and without remote past. *Sūrikaisekikenkyūsho Kōkyūroku* **1552,** 164–171.

[21] Yano, K. and Yor, M. (2010). Around Tsirelson's equation, or: the evolution process may not explain everything. Preprint. Available at http://arxiv.org/abs/0906.3442v2.

[22] Yor, M. (1992). Tsirel′son's equation in discrete time. *Probab. Theory Relat. Fields* **91,** 135–152.


https://doi.org/10.1239/jap/1316796913 Published online by Cambridge University Press