NON-COUPLING FROM THE PAST

GEOFFREY R. GRIMMETT[1,2] AND MARK HOLMES[2]

Abstract. The method of 'coupling from the past' permits exact sampling from the invariant distribution of a Markov chain on a finite state space. The coupling is successful whenever the stochastic dynamics are such that there is coalescence of all trajectories. The issue of the coalescence or non-coalescence of trajectories of a finite state space Markov chain is investigated in this note. The notion of the 'coalescence number' $k(\mu)$ of a Markovian coupling μ is introduced, and results are presented concerning the set $K(P)$ of coalescence numbers of couplings corresponding to a given transition matrix P.

1. Introduction

The method of 'coupling from the past' (CFTP) was introduced by Propp and Wilson [4, 5, 8] in order to sample from the invariant distribution of an irreducible Markov chain on a finite state space. It has attracted great interest amongst theoreticians and practitioners, and there is an extensive associated literature (see, for example, [7]).

The general approach of CFTP is as follows. Let X be an irreducible Markov chain on a finite state space S with transition matrix $P = (p_{i,j} : i, j \in S)$, and let π be the unique invariant distribution (see [3, Chap. 6] for a general account of the theory of Markov chains).

Let $\mathcal{F}_S$ be the set of functions from S to S, and let $\mathcal{P}_S$ be the set of all irreducible stochastic matrices on the finite set S. We write $\mathbb{N}$ for the set $\{1, 2, \dots\}$ of natural numbers, and $\mathbb{P}$ for the appropriate probability measure.

[1]Statistical Laboratory, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge CB3 0WB, UK

[2]School of Mathematics & Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

E-mail addresses: grg@statslab.cam.ac.uk, holmes.m@unimelb.edu.au.

Date: 5 July 2019.

2010 *Mathematics Subject Classification.* 60J10, 60J22.

Key words and phrases. Markov chain, coupling from the past, CFTP, perfect simulation, grand coupling, coalescence time, coalescence number.

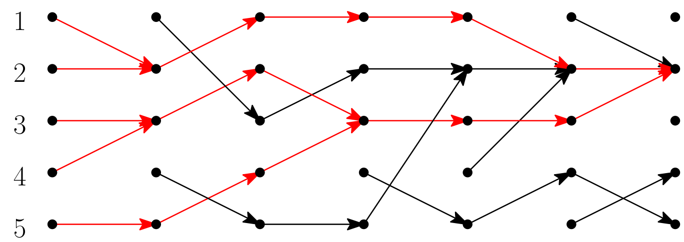

FIGURE 1.1. An illustration of coalescence of trajectories in CFTP with $|S| = 5$.

Definition 1.1. *A probability measure μ on $\mathcal{F}_S$ is* consistent *with $P \in \mathcal{P}_S$, in which case we say that the pair (P, μ) is* consistent, *if*

$$p_{i,j} = \mu\big(\{f \in \mathcal{F}_S : f(i) = j\}\big), \qquad i, j \in S. \tag{1.1}$$

Let $\mathcal{L}(P)$ denote the set of probability measures μ on $\mathcal{F}_S$ that are consistent with $P \in \mathcal{P}_S$.

Let $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$. The measure μ is called a *grand coupling* of P. Let $F = (F_s : s \in \mathbb{N})$ be a vector of independent samples from μ, let $\overleftarrow{F}_t$ denote the composition $F_1 \circ F_2 \circ \cdots \circ F_t$, and define the *backward coalescence time*

$$C = \inf\big\{t : \overleftarrow{F}_t(\cdot) \text{ is a constant function}\big\}. \tag{1.2}$$

We say that *backward coalescence occurs* if $C < \infty$. On the event $\{C < \infty\}$, $\overleftarrow{F}_C$ may be regarded as a random state.

The definition of coupling may seem confusing on first encounter. The function F_1 describes transitions during one step of the chain from time -1 to time 0, as illustrated in Figure 1. If F_1 is not a constant function, we move back one step in time to -2, and consider the composition $F_1 \circ F_2$. This process is iterated, moving one step back in time at each stage, until the earliest (random) C such that the iterated function $\overleftarrow{F}_C$ is constant. This C (if finite) is the time to backward coalescence.

Propp and Wilson proved the following fundamental theorem.

Theorem 1.2 ([4]). *Let $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$. Either $\mathbb{P}(C < \infty) = 0$ or $\mathbb{P}(C < \infty) = 1$. If it is the case that $\mathbb{P}(C < \infty) = 1$, then the random state $\overleftarrow{F}_C$ has law π.*

Here are two areas of application of CFTP. In the first, one begins with a recipe for a certain probability measure π on S, for example as the posterior distribution of a Bayesian analysis. In seeking a sample from π, one may find an aperiodic transition matrix P having π as unique invariant distribution, and then run CFTP on the associated Markov chain. In a second situation that may arise in a physical model, one begins with a Markovian dynamics with associated transition matrix $P \in \mathcal{P}_S$, and uses CFTP to sample from the invariant distribution. In the current work, we shall assume that the transition matrix P is specified, and that P is finite and irreducible.

Here is a summary of the work presented here. In Section 2, we discuss the phenomena of backward and forward coalescence, and we define the coalescence number of a Markov coupling. Theorem 3.4 explains the relationship between the coalescence number and the ranks of products of extremal elements in a convex representation of the stochastic matrix P. The question is posed of determining the set $K(P)$ of coalescence numbers of couplings consistent with a given P. A sub-family of couplings, termed 'block measures', is studied in Section 4. It is shown in Theorem 4.4, via Birkhoff's convex representation theorem for doubly stochastic matrices, that $|S| \in K(P)$ if and only if P is doubly stochastic. Some further results about $K(P)$ are presented in Section 5.

2. Coalescence of trajectories

CFTP relies upon almost-sure backward coalescence, which is to say that $\mathbb{P}(C < \infty) = 1$, where C is given in (1.2). For given $P \in \mathcal{P}_S$, the occurrence (or not) of coalescence depends on the choice of $\mu \in \mathcal{L}(P)$; see for example, Example 2.2.

We next introduce the notion of 'forward coalescence', which is to be considered as '*coalescence*' but with the difference that time runs forwards rather than backwards. As before, let $P \in \mathcal{P}_S$, $\mu \in \mathcal{L}(P)$, and let $F = (F_s : s \in \mathbb{N})$ be an independent sample from μ. For $i \in S$, define the Markov chain $X^i = (X^i_t : t \geq 0)$ by $X^i_t = \vec{F}_t(i)$ where $\vec{F}_t = F_t \circ F_{t-1} \circ \cdots \circ F_1$. Then $(X^i : i \in S)$ is a family of coupled Markov chains, running forwards in time, each having transition matrix P, and such that X^i starts in state i.

The superscript $\rightarrow$ (respectively, $\leftarrow$) is used to indicate that time is running forwards (respectively, backwards). For $i, j \in S$, we say that i and j *coalesce* if there exists t such that $X^i_t = X^j_t$. We say that *forward coalescence occurs* if, for all pairs $i, j \in S$, i and j coalesce. The *forward coalescence time* is given by

$$T = \inf\{t \geq 0 : X^i_t = X^j_t \text{ for all } i, j \in S\}. \tag{2.1}$$

Clearly, if P is periodic then $T = \infty$ a.s. for any $\mu \in \mathcal{L}(P)$. A simple but important observation is that C and T have the same distribution.

Theorem 2.1. *Let $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$. The backward coalescence time C and the forward coalescence time T have the same distribution.*

Proof. Let $(F_i : i \in \mathbb{N})$ be an independent sample from μ. For $t \geq 0$, we have

$$\mathbb{P}(C \leq t) = \mathbb{P}\big(\overleftarrow{F}_t(\cdot) \text{ is a constant function}\big).$$

By reversing the order of the functions $F_1, F_2, \dots, F_t$, we see that this equals $\mathbb{P}(T \leq t) = \mathbb{P}(\overrightarrow{F}_t(\cdot) \text{ is a constant function})$. ■

Example 2.2. *Let $S = \{1, 2, \dots, n\}$ where $n \geq 2$, and let $P_n = (p_{i,j})$ be the constant matrix with entries $p_{i,j} = 1/n$ for $i, j \in S$. Let $F = (F_i : i \in \mathbb{N})$ be an independent sample from $\mu \in \mathcal{L}(P_n)$.*

(a) *If each F_i is a uniform random permutation of S, then $T \equiv \infty$ and $\overrightarrow{F}_t(i) \neq \overrightarrow{F}_t(j)$ for all $i \neq j$ and $t \geq 1$.*

(b) *If $(F_1(i) : i \in S)$ are independent and uniformly distributed on S, then $\mathbb{P}(T < \infty) = 1$.*

In this example, there exist measures $\mu \in \mathcal{L}(P_n)$ such that either (a) *a.s. no pairs of states coalesce, or* (b) *a.s. forward coalescence occurs.*

For $g \in \mathcal{F}_S$, we let $\overset{g}{\sim}$ be the equivalence relation on S given by $i \overset{g}{\sim} j$ if and only if $g(i) = g(j)$. For $f = (f_t : t \in \mathbb{N}) \subseteq \mathcal{F}_S$ and $t \geq 1$, we write

$$\overleftarrow{f}_t = f_1 \circ f_2 \circ \cdots \circ f_t, \quad \overrightarrow{f}_t = f_t \circ f_{t-1} \circ \cdots \circ f_1.$$

Let $k_t(\overleftarrow{f})$ (respectively, $k_t(\overrightarrow{f})$) denote the number of equivalence classes of the relation $\overset{\overleftarrow{f}_t}{\sim}$ (respectively, $\overset{\overrightarrow{f}_t}{\sim}$). Similarly, we define the equivalence relation $\overset{\overleftarrow{f}}{\sim}$ on S by $i \overset{\overleftarrow{f}}{\sim} j$ if and only if $i \overset{\overleftarrow{f}_t}{\sim} j$ for some $t \in \mathbb{N}$, and we let $k(\overleftarrow{f})$ be the number of equivalence classes of $\overset{\overleftarrow{f}}{\sim}$ (and similarly for $\overrightarrow{f}$). We call $k(\overleftarrow{f})$ the *backward coalescence number* of $\overleftarrow{f}$, and likewise $k(\overrightarrow{f})$ the *forward coalescence number* of $\overrightarrow{f}$. The following lemma is elementary.

Lemma 2.3.

(a) *We have that $k_t(\overleftarrow{f})$ and $k_t(\overrightarrow{f})$ are monotone non-increasing in t. Furthermore, $k_t(\overleftarrow{f}) = k(\overleftarrow{f})$ and $k_t(\overrightarrow{f}) = k(\overrightarrow{f})$ for all large t.*

(b) *Let $F = (F_s : s \in \mathbb{N})$ be independent and identically distributed elements in $\mathcal{F}_S$. Then $k_t(\overleftarrow{F})$ and $k_t(\overrightarrow{F})$ are equidistributed, and similarly $k(\overleftarrow{F})$ and $k(\overrightarrow{F})$ are equidistributed.*

Proof. (a) The first statement holds by consideration of the definition, and the second since $k(\overleftarrow{F})$ and $k(\overrightarrow{F})$ are integer-valued.

(b) This holds as in the proof of Theorem 2.1. ■

3. Coalescence numbers

In light of Theorem 2.1 and Lemma 2.3, we henceforth consider only Markov chains running in *increasing positive time. Henceforth, expressions involving the word 'coalescence' shall refer to* forward *coalescence.* Let μ be a probability measure on $\mathcal{F}_S$, and let $\mathrm{supp}(\mu)$ denote the support of μ. Let $F = (F_s : s \in \mathbb{N})$ be a vector of independent and identically distributed random functions, each with law μ. The law of F is the product measure $\boldsymbol{\mu} = \prod_{i \in \mathbb{N}} \mu$. The coalescence time T is given by (2.1), and the term *coalescence number* refers to the quantities $k_t(\vec{F})$ and $k(\vec{F})$, which we denote henceforth by $k_t(F)$ and $k(F)$, respectively.

Lemma 3.1. *Let μ, μ_1, μ_2 be probability measures on $\mathcal{F}_S$.*

(a) *Let $F = (F_s : s \in \mathbb{N})$ be a sequence of independent and identically distributed functions each with law μ. We have that $k(F)$ is $\boldsymbol{\mu}$-a.s. constant, and we write $k(\mu)$ for the almost surely constant value of $k(F)$.*

(b) *If $\mathrm{supp}(\mu_1) \subseteq \mathrm{supp}(\mu_2)$, then $k(\mu_1) \geq k(\mu_2)$.*

(c) *If $\mathrm{supp}(\mu_1) = \mathrm{supp}(\mu_2)$, then $k(\mu_1) = k(\mu_2)$.*

We call $k(\mu)$ the *coalescence number* of μ.

Proof. (a) For $j \in \{1, 2, \ldots, n\}$, let $q_j = \boldsymbol{\mu}(k(F) = j)$, and $k^* = \min\{j : q_j > 0\}$. Then

$$\boldsymbol{\mu}(k(F) \geq k^*) = 1. \tag{3.1}$$

Moreover, we may choose $t \in \mathbb{N}$ such that

$$\kappa := \boldsymbol{\mu}(k_t(F) = k^*) \quad \text{satisfies} \quad \kappa > 0.$$

For $m \in \mathbb{N}$, write $F^m = (F_{mt+s} : s \geq 0)$. The events $\{k_t(F^m) = k^*\}$, $m \in \mathbb{N}$, are independent, and each occurs with probability κ. Therefore, almost surely at least one of these events occurs, and hence $\boldsymbol{\mu}(k(F) \leq k^*) = 1$. By (3.1), this proves the first claim.

(b) Assume $\mathrm{supp}(\mu_1) \subseteq \mathrm{supp}(\mu_2)$, and let k_i^* be the bottom of the $\boldsymbol{\mu}_i$-support of $k(F)$. Since, for large t, $\boldsymbol{\mu}_1(k_t(F) = k_1^*) > 0$, we have also that $\boldsymbol{\mu}_2(k_t(F) = k_1^*) > 0$, whence $k_1^* \geq k_2^*$. Part (c) is immediate. ∎

Whereas $k(F)$ is a.s. constant (as in Lemma 3.1(a)), the equivalence classes of $\overset{\vec{F}}{\sim}$ need not themselves be a.s. constant. Here is an example of this, preceded by some notation.

Definition 3.2. *Let $f \in \mathcal{F}_S$ where $S = \{i_1, i_2, \ldots, i_n\}$ is a finite ordered set. We write $f = (j_1 j_2 \ldots j_n)$ if $f(i_r) = j_r$ for $r = 1, 2, \ldots, n$.*

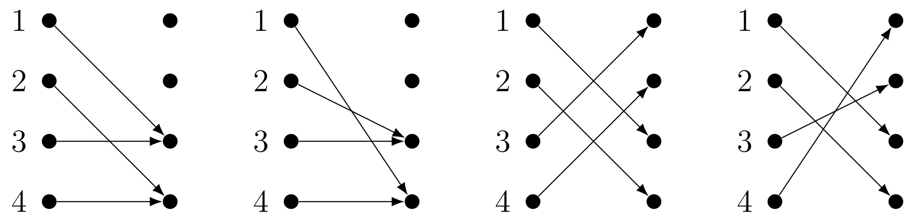

FIGURE 3.1. Diagrammatic representations of the four functions f_i of Example 3.3. The corresponding equivalence classes are not $\boldsymbol{\mu}$-a.s. constant.

Example 3.3. *Take $S = \{1, 2, 3, 4\}$ and any consistent pair (P, μ) with* $\mathrm{supp}(\mu) = \{f_1, f_2, f_3, f_4\}$*, where*

$$f_1 = (3434), \quad f_2 = (4334), \quad f_3 = (3412), \quad f_4 = (3421).$$

Then $k(\mu) = 2$ but the equivalence classes of $\overset{\vec{F}}{\sim}$ may be either $\{1, 3\}$, $\{2, 4\}$ or $\{1, 4\}$, $\{2, 3\}$, each having a strictly positive probability. The four functions f_i are illustrated diagrammatically in Figure 3.1.

A probability measure μ on $\mathcal{F}_S$ may be written in the form

$$\mu = \sum_{f \in \mathrm{supp}(\mu)} \alpha_f \delta_f, \tag{3.2}$$

where α is a probability mass function on $\mathcal{F}_S$ with support $\mathrm{supp}(\mu)$, and δ_f is the Dirac delta-mass on the point $f \in \mathcal{F}_S$. Thus, $\alpha_f > 0$ if and only if $f \in \mathrm{supp}(\mu)$. If $\mu \in \mathcal{L}(P)$, by (1.1) and (3.2),

$$P = \sum_{f \in \mathrm{supp}(\mu)} \alpha_f M_f, \tag{3.3}$$

where M_f denotes the matrix

$$M_f = (1_{\{f(i)=j\}} : i, j \in S), \tag{3.4}$$

and 1_A is the indicator function of A.

Let Π_S be the set of permutations of S. We denote also by Π_S the set of matrices M_f as f ranges over the permutations of S.

Theorem 3.4. *Let μ have the representation (3.2), and $|S| = n$.*

(a) *We have that*

(3.5)
$$k(\mu) = \inf\bigl\{\mathrm{rank}(M_{f_t} M_{f_{t-1}} \cdots M_{f_1} : f_1, f_2, \dots, f_t \in \mathrm{supp}(\mu),\ t \geq 1\bigr\}.$$

(b) *There exists $T = T(n)$ such that the infimum in (3.5) is achieved for some t satisfying $t \leq T$.*

Proof. (a) Let $F = (F_s : s \in \mathbb{N})$ be drawn independently from μ. Then

$$R_t := M_{F_t} M_{F_{t-1}} \cdots M_{F_1}$$

is the matrix with (i, j)th entry $1_{\{\vec{F}_t(i)=j\}}$. Therefore, $k_t(F)$ equals the number of non-zero columns of R_t. Since each row of R_t contains a unique 1, we have that $k_t(F) = \operatorname{rank}(R_t)$. Therefore, $k(\mu)$ is the decreasing limit

$$k(\mu) = \lim_{t\to\infty} \operatorname{rank}(R_t) \qquad \text{a.s.} \tag{3.6}$$

Equation (3.5) follows since $k(\mu)$ is integer-valued and deterministic.

(b) Since the rank of a matrix is integer-valued, the infimum in (3.5) is attained. The claim follows since, for given $|S| = n$, there are boundedly many possible matrices M_f. ■

Let

$$K(P) = \big\{k : \text{ there exists } \mu \in \mathcal{L}(P) \text{ with } k(\mu) = k\big\}.$$

It is a basic question to ask: what can be said about K as a function of P? We first state a well-known result.

Lemma 3.5. *We have that $1 \in K(P)$ if and only if $P \in \mathcal{P}_S$ is aperiodic.*

Proof. For $f \in \mathcal{F}_S$, let $\mu(\{f\}) = \prod_{i\in S} p_{i,f(i)}$. This gives rise to $|S|$ chains with transition matrix P, starting from $1, 2, \dots, n$, respectively, that evolve independently until they meet. If P is aperiodic (and irreducible) then all n chains meet a.s. in finite time.

Conversely, if P is periodic and $p_{i,j} > 0$ then i and j can never coalesce, so $1 \notin K(P)$. ■

Remark 3.6. *In a variety of cases of interest including, for example, the Ising and random-cluster models (see* [2, Exer. 7.3, Sect. 8.2]*), the set S has a partial order, denoted $\le$. For $P \in \mathcal{P}_S$ satisfying the so-called FKG lattice condition, it is natural to seek $\mu \in \mathcal{L}(P)$ whose transitions preserve this partial order, and such μ may be constructed via the relevant Gibbs sampler (see, for example,* [3, Sect. 6.14]*). By the irreducibility of P, the trajectory starting at the least state of S passes a.s. through the greatest state of S. This implies that coalescence occurs, so that $k(\mu) = 1$.*

4. Block measures

We introduce next the concept of a block measure.

Definition 4.1. *Let $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$. For a partition $\mathcal{S} = \{S_r : r = 1, 2, \dots, l\}$ of S with $l = l(\mathcal{S}) \geq 1$, we call μ an* $\mathcal{S}$-block measure *(or just a* block measure *with l blocks) if*

(a) *for $f \in \operatorname{supp}(\mu)$, there exists a unique permutation $\pi = \pi_f$ of $I := \{1, 2, \dots, l\}$ such that, for $r \in I$, $fS_r \subseteq S_{\pi(r)}$, and*
(b) *$k(\mu) = l$.*

The action of an $\mathcal{S}$-block measure μ is as follows. Since blocks are mapped a.s. to blocks, the measure μ of (3.2) induces a random permutation Π of the blocks which may be written as

$$\Pi = \sum_{f \in \operatorname{supp}(\mu)} \alpha_f \delta_{\pi_f}. \tag{4.1}$$

The condition $k(\mu) = l$ implies that

$$\text{for } r \in I \text{ and } i, j \in S_r, \text{ the pair } i,\ j \text{ coalesce a.s.,} \tag{4.2}$$

so that the equivalence classes of $\overset{\vec{F}}{\sim}$ are a.s. the blocks $S_1, S_2, \dots, S_l$. If, as the chain evolves, we observe only the evolution of the blocks, we see a Markov chain on I with transition probabilities $\lambda_{r,s} = \mathbb{P}(\Pi(r) = s)$ which, since P is irreducible, is itself irreducible.

Example 3.3 illustrates the existence of measures μ that are not block measures, when $|S| = 4$. On the other hand, we have the following lemma when $|S| = 3$. For $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$, let $\mathcal{C} = \mathcal{C}(\mu)$ be the set of possible coalescing pairs,

$$\mathcal{C} = \big\{\{i, j\} \subseteq S : i \neq j,\ \boldsymbol{\mu}(i, j \text{ coalesce}) > 0\big\}. \tag{4.3}$$

Lemma 4.2. *Let $|S| = 3$ and $P \in \mathcal{P}_S$. If (P, μ) is consistent then μ is a block measure.*

Proof. Let S, (P, μ) be as given. If $\mathcal{C}$ is empty then $k(\mu) = 3$ and μ is a block measure with 3 blocks.

If $|\mathcal{C}| \geq 2$, we have by the forthcoming Proposition 5.1(a, b) that $k(\mu) \leq 1$, so that μ is a block measure with 1 block.

Finally, if $\mathcal{C}$ contains exactly one element then we may assume, without loss of generality, that element is $\{1, 2\}$. By Proposition 5.1(b), we have $k(\mu) = 1$, whence a.s. some pair coalesces. By assumption only $\{1, 2\}$ can coalesce, so in fact a.s. we have that 1 and 2 coalesce, and they do not coalesce with 3. Therefore, μ is a block measure with the two blocks $\{1, 2\}$ and $\{3\}$. ■

We show next that, for $1 \leq k \leq |S|$, there exists a consistent pair (P, μ) such that μ is a block measure with $k(\mu) = k$.

Lemma 4.3. *For $|S| = n \geq 2$ and $1 \leq k \leq n$, there exists an aperiodic $P \in \mathcal{P}_S$ such that $k \in K(P)$.*

Proof. Let $\mathcal{S} = \{S_r : r = 1, 2, \dots, l\}$ be a partition of S, and let $\mathcal{G} \subseteq \mathcal{F}_S$ be the set of all functions g satisfying: there exists a permutation π of $\{1, 2, \dots, l\}$ such that, for $r = 1, 2, \dots, l$, we have $gS_r \subseteq S_{\pi(r)}$. Any probability measure μ on $\mathcal{F}_S$ with support $\mathcal{G}$ is an $\mathcal{S}$-block measure.

Let μ be such a measure and let P be the associated stochastic matrix on S, given in (1.1). For $i, j \in S$, there exists $g \in \mathcal{G}$ such that $g(i) = j$. Therefore, P is irreducible and aperiodic. ■

We identify next the consistent pairs (P, μ) for which either $k(\mu) = |S|$ or $|S| \in K(P)$.

Theorem 4.4. *Let $|S| = n \geq 2$ and $P \in \mathcal{P}_S$. We have that*

(a) *$k(\mu) = n$ if and only if $\operatorname{supp}(\mu)$ contains only permutations of S,*
(b) *$n \in K(P)$ if and only if P is doubly stochastic.*

Before proving this, we remind the reader of Birkhoff's theorem [1] (sometimes attributed also to von Neumann [6]).

Theorem 4.5 ([1, 6])**.** *A stochastic matrix P on the finite state space S is doubly stochastic if and only if it lies in the convex hull of the set Π_S of permutation matrices.*

Remark 4.6. *We note that the simulation problem confronted by CFTP is trivial when P is irreducible and doubly stochastic, since such P are characterized as those transition matrices with the uniform invariant distribution $\pi = (\pi_i = n^{-1} : i \in S)$.*

Proof of Theorem 4.4. (a) If $\operatorname{supp}(\mu)$ contains only permutations, then a.s. $k_t(F) = n$ for every $t \in \mathbb{N}$. Hence $n \in K(P)$. If $\operatorname{supp}(\mu)$ contains a non-permutation, then with positive probability $k_1(F) < n$ and hence $k(\mu) < n$.

(b) By Theorem 4.5, P is doubly stochastic if and only if it may be expressed as a convex combination

$$P = \sum_{f \in \Pi_S} \alpha_f M_f, \tag{4.4}$$

of permutation matrices M_f (recall (3.3) and (3.4)).

If P is doubly stochastic, let the α_f satisfy (4.4), and let

$$\mu = \sum_{f \in \Pi_S} \alpha_f \delta_f, \tag{4.5}$$

as in (3.2). Then $\mu \in \mathcal{L}(P)$, and $k(\mu) = n$ by part (a).

If P is not doubly stochastic and $\mu \in \mathcal{L}(P)$, then μ has no representation of the form (4.5), so that $k(\mu) < n$ by part (a). ■

Finally in this section, we present a necessary and sufficient condition for μ to be an $\mathcal{S}$-block measure. Let $P \in \mathcal{P}_S$, and let $\mathcal{S} = \{S_r : r = 1, 2, \dots, l\}$ be a partition of S with $l \geq 1$. For $r, s \in I := \{1, 2, \dots, l\}$ and $i \in S_r$, let

$$\lambda_{r,s}^{(i)} = \sum_{j \in S_s} p_{i,j}.$$

Since a block measure comprises a transition operator on blocks, combined with a shuffling of states within blocks, it is necessary in order that μ be an $\mathcal{S}$-block measure that

$$\lambda_{r,s}^{(i)} \text{ is constant for } i \in S_r. \tag{4.6}$$

When (4.6) holds, we write

$$\lambda_{r,s} = \lambda_{r,s}^{(i)}, \qquad i \in S_r. \tag{4.7}$$

Under (4.6), the matrix $\Lambda = (\lambda_{r,s} : r, s \in I)$ is the irreducible transition matrix of the Markov chain derived from P by observing the evolution of blocks, which is to say that

$$\lambda_{r,s} = \mu(\Pi(r) = s), \qquad r, s \in I, \tag{4.8}$$

where Π is given by (4.1). Since $l \in K(\Lambda)$, we have by Theorem 4.4 that Λ is doubly stochastic, which is to say that

$$\sum_{r \in I} \lambda_{r,s} = \sum_{r \in I} \sum_{j \in S_s} p_{i_r,j} = 1, \qquad s \in I, \tag{4.9}$$

where each i_r is an arbitrarily chosen representative of the block S_r. By (4.6), equation (4.9) may be written in the form

$$\sum_{i \in S} \sum_{j \in S_s} \frac{1}{|S_{r(i)}|} p_{i,j} = 1, \qquad r, s \in I, \tag{4.10}$$

where $r(i)$ is the index r such that $i \in S_r$. We summarise this in a theorem.

Theorem 4.7. *Let S be a non-empty, finite set, and let $\mathcal{S} = \{S_r : r = 1, 2, \dots, l\}$ be a partition of S. For $P \in \mathcal{P}_S$, a measure $\mu \in \mathcal{L}(P)$ is an $\mathcal{S}$-block measure if and only if (4.6), (4.10) hold, and also $k(\mu) = l$.*

Proof. The necessity of the conditions holds by the definition of block measure and the above discussion.

Suppose conversely that the stated conditions hold. Let $\Lambda = (\lambda_{r,s})$ be given by (4.6)–(4.7). By (4.7) and (4.10), Λ is doubly stochastic. By Theorem 4.4, we may find a measure $\rho \in \mathcal{L}(\Lambda)$ supported on a subset of the set Π_I of permutations of I, and we let Π have law ρ. Conditional on Π, let $Z = (Z_i : i \in S)$ be independent random variables such that

$$\mathbb{P}(Z_i = j \mid \Pi) = \begin{cases} p_{i,j}/\lambda_{r,s} & \text{if } S_r \ni i,\ S_s \ni j,\ \Pi(r) = s, \\ 0 & \text{otherwise.} \end{cases}$$

The law μ of Z is an $\mathcal{S}$-block measure that is consistent with P. ■

5. The set $K(P)$

We begin with a triplet of conditions.

Proposition 5.1. *Let $S = \{1, 2, \dots, n\}$ where $n \ge 3$, and let $P \in \mathcal{P}_S$ and $\mu \in \mathcal{L}(P)$. Let $\mathcal{C} = \mathcal{C}(\mu)$ be the set of possible coalescing pairs, as in (4.3).*

(a) *$k(\mu) = n$ if and only if $|\mathcal{C}| = 0$.*
(b) *$k(\mu) = n - 1$ if and only if $|\mathcal{C}| = 1$.*
(c) *If $|\mathcal{C}|$ comprises the single pair $\{1, 2\}$, then P satisfies*

$$\sum_{j=3}^{n} p_{1,j} = \sum_{j=3}^{n} p_{2,j} = \sum_{i=3}^{n} (p_{i,1} + p_{i,2}). \tag{5.1}$$

Proof. (a) See Theorem 4.4(a).

(b) By part (a), $k(\mu) \le n - 1$ when $|\mathcal{C}| = 1$, It suffices, therefore, to show that $k(\mu) \le n - 2$ when $|\mathcal{C}| \ge 2$. Suppose that $|\mathcal{C}| \ge 2$. Without loss of generality we may assume that $\{1, 2\} \in \mathcal{C}$ and either that $\{1, 3\} \in \mathcal{C}$ or (in the case $n \ge 4$) that $\{3, 4\} \in \mathcal{C}$. Let $F = (F_s : s \in \mathbb{N})$ be an independent sample from μ. Let M be the Markov time $M = \inf\{t > 0 : \vec{F}_t(1) = \vec{F}_t(2) = 1\}$, and write $J = \{M < \infty\}$. By irreducibility, $\boldsymbol{\mu}(J) > 0$, implying that $k(\mu) \le n - 1$. Assume that

$$k(\mu) = n - 1. \tag{5.2}$$

We shall obtain a contradiction, and the conclusion of the lemma will follow.

Suppose first that $\{1, 2\}, \{1, 3\} \in \mathcal{C}$. Let B be the event that there exists $i \ge 3$ such that $\vec{F}_M(i) \in \{1, 2, 3\}$. On $B \cap J$, we have $k(F) \le n-2$ a.s., since

$$\boldsymbol{\mu}(\text{at least 3 states belong to coalescing pairs}) > 0.$$

Thus $\mu(B \cap J) = 0$ by (5.2). On $\overline{B} \cap J$, the $\vec{F}_M(i)$, $i \ge 3$, are by (5.2) a.s. distinct, and in addition take values in $S \setminus \{1, 2, 3\}$. Thus there exist $n - 2$ distinct values of $\vec{F}_M(i)$, $i \ge 3$, but at most $n - 3$ values that they can take, which is impossible, whence $\mu(\overline{B} \cap J) = 0$. It follows that

$$0 < \boldsymbol{\mu}(J) = \boldsymbol{\mu}(B \cap J) + \boldsymbol{\mu}(\overline{B} \cap J) = 0, \tag{5.3}$$

a contradiction.

Suppose secondly that $\{1,2\}, \{3,4\} \in \mathcal{C}$. Let C be the event that either (i) there exists $i \ge 3$ such that $\vec{F}_M(i) \in \{1,2\}$, or (ii) $\{\vec{F}_M(i) : i \ge 3\} \supseteq \{3,4\}$. On $C \cap J$, we have $k(F) \le n-2$ a.s. On $\overline{C} \cap J$, by (5.2) the $\vec{F}_M(i)$, $i \ge 3$, are a.s. distinct, and in addition take values in $S \setminus \{1,2\}$ and no pair of them equals $\{3,4\}$. This provides a contradiction as in (5.3).

(c) Let F_1 have law μ. Write $A_i = \{F_1(i) \in \{1,2\}\}$, and

$$M = |\{i \le 2 : A_i \text{ occurs}\}|, \qquad N = |\{i \ge 3 : A_i \text{ occurs}\}|.$$

If $\mu(A_i \cap A_j) > 0$ for some $i \ge 3$ and $j \ne i$, then $\{i,j\} \in \mathcal{C}$, in contradiction of the assumption that $\mathcal{C}$ comprises the singleton $\{1,2\}$. Therefore, $\mu(A_i \cap A_j) = 0$ for all $i \ge 3$ and $j \ne i$, and hence

$$\mu(N \ge 2) = 0, \tag{5.4}$$

$$\mu(M \ge 1, N = 1) = 0. \tag{5.5}$$

By similar arguments,

$$\mu(M < 2, N = 0) = 0, \tag{5.6}$$

$$\mu(M = 1) = 0. \tag{5.7}$$

It follows that

$$\begin{aligned} \mu(N=1) &= \mu(N=1, M=0) && \text{by (5.5)} \\ &= \mu(M=0) && \text{by (5.6) and (5.4)} \\ &= \mu(\overline{A}_1 \cap \overline{A}_2) \\ &= \mu(\overline{A}_r), \quad r = 1,2, && \text{by (5.7).} \end{aligned}$$

Therefore,

$$\mu(N=1) = \mu(\overline{A}_r) = \mu(F_1(r) \ge 3) = \sum_{j=3}^{n} p_{r,j}, \qquad r = 1,2.$$

By (5.4),

$$\mu(N=1) = \mu(N) = \sum_{i=3}^{n} \mu(A_i) = \sum_{i=3}^{n} (p_{i,1} + p_{i,2}),$$

where $\mu(N)$ is the mean value of N. This yields (5.1). ■

The set $K(P)$ can be fairly sporadic, as illustrated in the next two examples.

Example 5.2. *Consider the matrix*

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{5.8}$$

Since P is doubly stochastic, by Theorem 4.4(a), there exists $\mu \in \mathcal{L}(P)$ such that $k(\mu) = 3$ (one may take $\mu(123) = \mu(231) = \frac{1}{2}$). By Lemma 3.5, we have that $1 \in K(P)$, and thus $\{1, 3\} \subseteq K(P)$. We claim that $2 \notin K(P)$, and we show this as follows.

Let $\mu \in \mathcal{L}(P)$, with $k(\mu) < 3$, so that $|\mathcal{C}| \geq 1$. There exists no permutation of S for which the matrix P satisfies (5.1), whence $|\mathcal{C}| \geq 2$ by Proposition 5.1(c). By parts (a, b) of that proposition, $k(\mu) \leq 1$. In conclusion, $K(P) = \{1, 3\}$.

Example 5.3. *Consider the matrix*

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{5.9}$$

We have, as in Example 5.2, that $\{1, 4\} \subseteq K(P)$. Taking

$$\mu(1234) = \mu(2244) = \mu(1331) = \mu(2341) = \tfrac{1}{4}$$

reveals that $2 \in K(P)$, and indeed μ is a block measure with blocks $\{1, 2\}$, $\{3, 4\}$. As in Example 5.2, we have that $3 \notin K(P)$, so that $K(P) = \{1, 2, 4\}$.

We investigate in greater depth the transition matrix on S with equal entries. Let $|S| = n \geq 2$ and let $P_n = (p_{i,j})$ satisfy $p_{i,j} = n^{-1}$ for $i, j \in S = \{1, 2, \dots, n\}$.

Theorem 5.4. *For $n \geq 2$ there exists a block measure $\mu \in \mathcal{L}(P_n)$ with $k(\mu) = l$ if and only if $l \mid n$. In particular, $K(P_n) \supseteq \{l : l \mid n\}$. For $n \geq 3$, we have $n - 1 \notin K(P_n)$.*

We do not know whether $K(P_n) = \{l : l \mid n\}$, and neither do we know if there exists $\mu \in \mathcal{L}(P_n)$ that is not a block measure.

Proof. Let $n \geq 2$. By Lemma 3.5 and Theorem 4.4, we have that $1, n \in K(P_n)$. It is easily seen as follows that $l \in K(P_n)$ whenever $l \mid n$. Suppose $l \mid n$ and $l \neq 1, n$. Let

$$S_r = (r - 1)n/l + \{1, 2, \dots, n/l\}, \qquad r = 1, 2, \dots, l.$$

We describe next a measure $\mu \in \mathcal{L}(P_n)$. Let Π be a uniformly chosen permutation of $\{1, 2, \dots, l\}$. For $i \in S$, let Z_i be chosen uniformly at random from $S_{\Pi(i)}$, where the Z_i are conditionally independent given Π. Let μ be the block measure governing the vector $Z = (Z_i : i \in S)$. By symmetry,

$$q_{i,j} := \mu\big(\{f \in \mathcal{F}_S : f(i) = j\}\big), \qquad i, j \in S,$$

is constant for all pairs $i, j \in S$. Since μ is a probability measure, $Q = (q_{i,j})$ has row sums 1, whence $q_{i,j} = n^{-1} = p_{i,j}$, and therefore $\mu \in \mathcal{L}(P_n)$. By examination of μ, μ is an $\mathcal{S}$-block measure.

Conversely, suppose there exists an $\mathcal{S}$-block measure $\mu \in \mathcal{L}(P_n)$ with corresponding partition $\mathcal{S} = \{S_1, S_2, \dots, S_l\}$ with index set $I = \{1, 2, \dots, l\}$. By Theorem 4.7, equations (4.6) and (4.10) hold. By (4.6), the matrix $\Lambda = (\lambda_{r,s} : r, s \in I)$ satisfies

$$\lambda_{r,s} = \frac{|S_s|}{n}, \qquad r, s \in I. \tag{5.10}$$

By (4.10),

$$\frac{|S_s|}{|S_r|} = 1, \qquad s, r \in I,$$

whence $|S_s| = n/l$ for all $s \in I$, and in particular $l \mid n$.

Let $n \geq 3$. We prove next that $k(\mu) \neq n - 1$ for $\mu \in \mathcal{L}(P_n)$. Let $\mathcal{C} = \mathcal{C}(\mu)$ be given as in (4.3). By Proposition 5.1(b), it suffices to prove that $|\mathcal{C}| \neq 1$. Assume on the contrary that $|\mathcal{C}| = 1$, and suppose without loss of generality that $\mathcal{C}$ contains the singleton pair $\{1, 2\}$. With $P = P_n$, the necessary condition (5.1) becomes

$$(n-2)\frac{1}{n} = (n-2)\frac{2}{n},$$

which is false when $n \geq 3$. Therefore, $|\mathcal{C}| \neq 1$, and the proof is complete. ■

Acknowledgements

The authors thank Tim Garoni and Wilfrid Kendall for their comments on this work. MH is supported by Future Fellowship FT160100166 from the Australian Research Council.

References

[1] G. Birkhoff, *Three observations on linear algebra*, Univ. Nac. Tucumán. Revista A. **5** (1946), 147–151.

[2] G. R. Grimmett, *Probability on Graphs*, 2nd ed., Cambridge University Press, Cambridge, 2018.

[3] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 4th ed., Oxford University Press, Oxford, 2020.

[4] J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures Algorithms **9** (1996), 223–252, Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995).

[5] ______, *How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph*, J. Algorithms **27** (1998), 170–217, 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996).

[6] J. von Neumann, *A certain zero-sum two-person game equivalent to the optimal assignment problem*, Contributions to the theory of games, vol. 2, Annals of Mathematics Studies, no. 28, Princeton University Press, Princeton, NJ, 1953, pp. 5–12.

[7] D. B. Wilson, *Perfectly random sampling with Markov chains*, http://www.dbwilson.com/exact/.

[8] D. B. Wilson and J. G. Propp, *How to get an exact sample from a generic Markov chain and sample a random spanning tree from a directed graph, both within the cover time*, Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996), ACM, New York, 1996, pp. 448–457.