Chapter 1
Non-coupling from the past

Geoffrey R. Grimmett and Mark Holmes

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.

1.1 Introduction

The method of 'coupling from the past' (CFTP) was introduced by Propp and Wilson [7, 8, 11] 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, [5, 10]).

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 [4, Chap. 6] for a general account of the theory of Markov chains).

Geoffrey Grimmett
Statistical Laboratory, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge CB3 0WB, UK, e-mail: grg@statslab.cam.ac.uk
and
School of Mathematics & Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

Mark Holmes
School of Mathematics & Statistics, The University of Melbourne, Parkville, VIC 3010, Australia. e-mail: holmes.m@unimelb.edu.au

Let $\mathscr{F}_S$ be the set of functions from S to S, and let $\mathscr{P}_S$ 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 a generic probability measure.

Definition 1. A probability measure μ on $\mathscr{F}_S$ is *consistent* with $P \in \mathscr{P}_S$, in which case we say that the pair (P,μ) is *consistent*, if

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

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

Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. 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 1.1. 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.

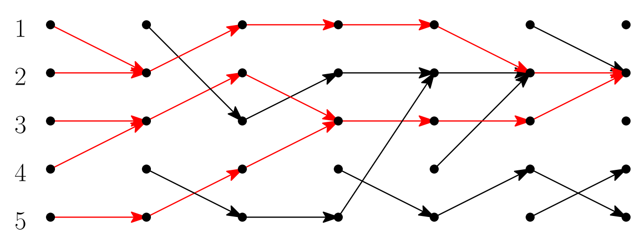

Fig. 1.1 An illustration of coalescence of trajectories in CFTP with $|S| = 5$.

Theorem 1 ([7]). *Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. 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 $P \in \mathscr{P}_S$, 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 1.2, we discuss the phenomena of backward and forward coalescence, and we define the coalescence number of a Markov coupling. Informally, the coalescence number is the (deterministic) limiting number of un-coalesced trajectories of the coupling. Theorem 3 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 1.4. These are couplings for which there is a fixed set of blocks (partitioning the state space), such that blocks of states are mapped to blocks of states, and such that coalescence occurs within but not between blocks. It is shown in Theorem 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 1.5.

1.2 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 $P \in \mathscr{P}_S$, the occurrence (or not) of coalescence depends on the choice of $\mu \in \mathscr{L}(P)$; see for example, Example 1.

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 $P \in \mathscr{P}_S$, $\mu \in \mathscr{L}(P)$, and let $F = (F_s : s \in \mathbb{N})$ be an independent sample from μ. For $i \in S$, define the Markov chain $X^i = (X_t^i : t \geq 0)$ by $X_t^i = \vec{F}_t(i)$ 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 $X_t^i = X_t^j$. 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

$$T = \inf\{t \geq 0 : X_t^i = X_t^j \text{ for all } i, j \in S\}. \tag{1.3}$$

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

Theorem 2. *Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. 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 1. 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 $\mu \in \mathscr{L}(P_n)$.

(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 $\mu \in \mathscr{L}(P_n)$ such that either (a) a.s. no pairs of states coalesce, or (b) a.s. forward coalescence occurs.

For $g \in \mathscr{F}_S$, we let $\stackrel{g}{\sim}$ be the equivalence relation on S given by $i \stackrel{g}{\sim} j$ if and only if $g(i) = g(j)$. For $f = (f_t : t \in \mathbb{N}) \subseteq \mathscr{F}_S$ 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 $\stackrel{\overleftarrow{f}_t}{\sim}$ (respectively, $\stackrel{\overrightarrow{f}_t}{\sim}$). Similarly, we define the equivalence relation $\stackrel{\overleftarrow{f}}{\sim}$ on S by $i \stackrel{\overleftarrow{f}}{\sim} j$ if and only if $i \stackrel{\overleftarrow{f}_t}{\sim} j$ for some $t \in \mathbb{N}$, and we let $k(\overleftarrow{f})$ be the number of equivalence classes of $\stackrel{\overleftarrow{f}}{\sim}$ (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 1.

(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 $\mathscr{F}_S$. 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 and Lemma 1, 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 $\mathscr{F}_S$, 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 (1.3), 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 2. *Let μ, μ_1, μ_2 be probability measures on $\mathscr{F}_S$.*

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

$$\boldsymbol{\mu}(k(F) \geq k^*) = 1. \tag{1.4}$$

Moreover, we 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 $F^m = (F_{mt+s} : s \geq 1)$. The event $E_{t,m} = \{k_t(F^m) = k^*\}$ depends only on $F_{mt+1}, F_{mt+2} \dots, F_{(m+1)t}$. It follows that the events $\{E_{t,m} : m \in \mathbb{N}\}$ 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 (1.4), 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 2(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 2. Let $f \in \mathscr{F}_S$ where $S = \{i_1, i_2, \dots, i_n\}$ is a finite ordered set. We write $f = (j_1 j_2 \dots j_n)$ if $f(i_r) = j_r$ for $r = 1,2,\dots,n$.

Example 2. 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 1.2.

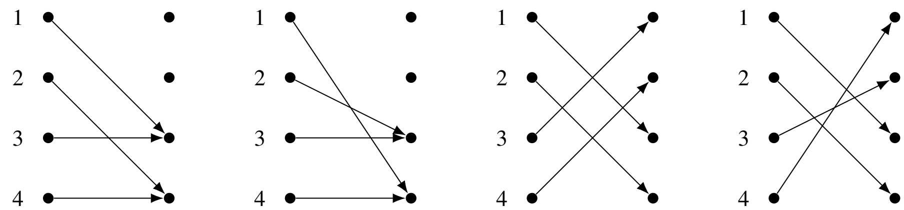

Fig. 1.2 Diagrammatic representations of the four functions f_i of Example 2. The corresponding equivalence classes are not $\boldsymbol{\mu}$-a.s. constant.

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

$$\mu = \sum_{f \in \mathrm{supp}(\mu)} \alpha_f \delta_f, \tag{1.5}$$

where α is a probability mass function on $\mathscr{F}_S$ with support $\mathrm{supp}(\mu)$, and δ_f is the Dirac delta-mass on the point $f \in \mathscr{F}_S$. Thus, $\alpha_f > 0$ if and only if $f \in \mathrm{supp}(\mu)$. If $\mu \in \mathscr{L}(P)$, by (1.1) and (1.5),

$$P = \sum_{f \in \mathrm{supp}(\mu)} \alpha_f M_f, \tag{1.6}$$

where M_f denotes the matrix

$$M_f = (1_{\{f(i)=j\}} : i, j \in S), \tag{1.7}$$

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. *Let μ have the representation (1.5), and $|S| = n$.*

(a) *We have that*

$$k(\mu) = \inf\big\{\mathrm{rank}(M_{f_t} M_{f_{t-1}} \cdots M_{f_1} : f_1, f_2, \dots, f_t \in \mathrm{supp}(\mu),\ t \geq 1\big\}. \tag{1.8}$$

(b) *There exists $T = T(n)$ such that the infimum in (1.8) 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 $k_t(F) = \text{rank}(R_t)$. Therefore, $k(\mu)$ is the decreasing limit

$$k(\mu) = \lim_{t\to\infty} \text{rank}(R_t) \qquad \text{a.s.} \tag{1.9}$$

Equation (1.8) follows since $k(\mu)$ is integer-valued and deterministic.

(b) Since the rank of a matrix is integer-valued, the infimum in (1.8) is attained. The claim follows since, for given $|S| = n$, there are boundedly many possible matrices M_f.

Let

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

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, based on ideas already in work of Doeblin [2].

Lemma 3. *We have that $1 \in K(P)$ if and only if $P \in \mathscr{P}_S$ is aperiodic.*

Proof. For $f \in \mathscr{F}_S$, 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 $i \neq j$, and i and j can never coalesce, implying $1 \notin K(P)$.

Remark 1. In a variety of cases of interest including, for example, the Ising and random-cluster models (see [3, Exer. 7.3, Sect. 8.2]), the set S has a partial order, denoted $\le$. For $P \in \mathscr{P}_S$ satisfying the so-called FKG lattice condition, it is natural to seek $\mu \in \mathscr{L}(P)$ whose transitions preserve this partial order, and such μ may be constructed via the relevant Gibbs sampler (see, for example, [4, 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 $k(\mu) = 1$.

1.4 Block measures

We introduce next the concept of a block measure, which is a strong form of the *lumpability* of [6] and [4, Exer. 6.1.13].

Definition 3. Let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. For a partition $\mathscr{S} = \{S_r : r = 1, 2, \dots, l\}$ of S with $l = l(\mathscr{S}) \ge 1$, we call μ an *$\mathscr{S}$-block measure* (or just a *block measure* with l blocks) if

(a) for $f \in \mathrm{supp}(\mu)$, 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 $\mathscr{S}$-block measure μ is as follows. Since blocks are mapped a.s. to blocks, the measure μ of (1.5) induces a random permutation Π of the blocks which may be written as

$$\Pi = \sum_{f \in \mathrm{supp}(\mu)} \alpha_f \delta_{\pi_f}. \tag{1.10}$$

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

$$\text{for } r \in I \text{ and } i,j \in S_r, \text{ the pair } i,\, j \text{ coalesce a.s.,} \tag{1.11}$$

so that the equivalence classes of $\stackrel{\vec{F}}{\sim}$ 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 2 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 $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$, let $\mathscr{C} = \mathscr{C}(\mu)$ be the set of possible coalescing pairs,

$$\mathscr{C} = \big\{\{i,j\} \subseteq S : i \neq j,\, \boldsymbol{\mu}(i,j \text{ coalesce}) > 0\big\}. \tag{1.12}$$

Lemma 4. *Let $|S| = 3$ and $P \in \mathscr{P}_S$. If (P,μ) is consistent then μ is a block measure.*

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

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

Finally, if $\mathscr{C}$ contains exactly one element then we may assume, without loss of generality, that element is $\{1,2\}$. By Proposition 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 5. *For $|S| = n \geq 2$ and $1 \leq k \leq n$, there exists an aperiodic $P \in \mathscr{P}_S$ such that $k \in K(P)$.*

Proof. Let $\mathscr{S} = \{S_r : r = 1,2,\dots,l\}$ be a partition of S, and let $\mathscr{G} \subseteq \mathscr{F}_S$ 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 $\mathscr{F}_S$ with support $\mathscr{G}$ is an $\mathscr{S}$-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 $g \in \mathscr{G}$ 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. *Let $|S|=n\geq 2$ and $P\in\mathscr{P}_S$. 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 [9]).

Theorem 5 ([1, 9]). *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 2. 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 (Proof of Theorem 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 5, P is doubly stochastic if and only if it may be expressed as a convex combination

$$P=\sum_{f\in\Pi_S}\alpha_f M_f, \tag{1.13}$$

of permutation matrices M_f (recall (1.6) and (1.7)).

If P is doubly stochastic, let the α_f satisfy (1.13), and let

$$\mu=\sum_{f\in\Pi_S}\alpha_f\delta_f, \tag{1.14}$$

as in (1.5). Then $\mu\in\mathscr{L}(P)$, and $k(\mu)=n$ by part (a).

If P is not doubly stochastic and $\mu\in\mathscr{L}(P)$, then μ has no representation of the form (1.14), so that $k(\mu)<n$ by part (a).

Finally in this section, we present a necessary and sufficient condition for μ to be an $\mathscr{S}$-block measure, Theorem 6 below.

Let $P\in\mathscr{P}_S$, and let $\mathscr{S}=\{S_r:r=1,2,\dots,l\}$ 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 $\mathscr{S}$-block measure that

$$\lambda_{r,s}^{(i)}\text{ is constant for } i\in S_r. \tag{1.15}$$

When (1.15) holds, we write

$$\lambda_{r,s} = \lambda_{r,s}^{(i)}, \qquad i \in S_r. \tag{1.16}$$

Under (1.15), 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

$$\lambda_{r,s} = \mu(\Pi(r) = s), \qquad r,s \in I, \tag{1.17}$$

where Π is given by (1.10). Since $l \in K(\Lambda)$, we have by Theorem 4 that Λ is doubly stochastic, which is to say that

$$\sum_{r\in I} \lambda_{r,s} = \sum_{r\in I} \sum_{j\in S_s} p_{i_r,j} = 1, \qquad s \in I, \tag{1.18}$$

where each i_r is an arbitrarily chosen representative of the block S_r. By (1.15), equation (1.18) may be written in the form

$$\sum_{i\in S} \sum_{j\in S_s} \frac{1}{|S_{r(i)}|} p_{i,j} = 1, \qquad r,s \in I, \tag{1.19}$$

where $r(i)$ is the index r such that $i \in S_r$. The following theorem is the final result of this section.

Theorem 6. *Let S be a non-empty, finite set, and let $\mathscr{S} = \{S_r : r = 1,2,\dots,l\}$ be a partition of S. For $P \in \mathscr{P}_S$, a measure $\mu \in \mathscr{L}(P)$ is an $\mathscr{S}$-block measure if and only if* (1.15), (1.19) *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 (1.15)–(1.16). By (1.16) and (1.19), Λ is doubly stochastic. By Theorem 4, we may find a measure $\rho \in \mathscr{L}(\Lambda)$ 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 $\mathscr{S}$-block measure that is consistent with P.

1.5 The set $K(P)$

We begin with a triplet of conditions.

Proposition 1. *Let $S = \{1,2,\dots,n\}$ where $n \geq 3$, and let $P \in \mathscr{P}_S$ and $\mu \in \mathscr{L}(P)$. Let $\mathscr{C} = \mathscr{C}(\mu)$ be the set of possible coalescing pairs, as in (1.12).*

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

$$\sum_{j=3}^{n} p_{1,j} = \sum_{j=3}^{n} p_{2,j} = \sum_{i=3}^{n} (p_{i,1} + p_{i,2}). \tag{1.20}$$

Proof. (a) See Theorem 4(a).

(b) By part (a), $k(\mu) \leq n-1$ when $|\mathscr{C}| = 1$. It is trivial by definition of k and $\mathscr{C}$ that, if $k(\mu) \leq n-2$, then $|\mathscr{C}| \geq 2$. It suffices, therefore, to show that $k(\mu) \leq n-2$ when $|\mathscr{C}| \geq 2$. Suppose that $|\mathscr{C}| \geq 2$. Without loss of generality we may assume that $\{1,2\} \in \mathscr{C}$ and either that $\{1,3\} \in \mathscr{C}$ or (in the case $n \geq 4$) that $\{3,4\} \in \mathscr{C}$. 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 $k(\mu) \leq n-1$. Assume that

$$k(\mu) = n-1. \tag{1.21}$$

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

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

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

Thus $\mu(B \cap J) = 0$ by (1.21). On $\overline{B} \cap J$, the $\vec{F}_M(i)$, $i \geq 3$, are by (1.21) a.s. distinct, and in addition take values in $S \setminus \{1,2,3\}$. Thus there exist $n-2$ distinct values of $\vec{F}_M(i)$, $i \geq 3$, but at most $n-3$ values that they can take, which is impossible, whence $\mu(\overline{B} \cap J) = 0$. It follows that

$$0 < \boldsymbol{\mu}(J) = \boldsymbol{\mu}(B \cap J) + \boldsymbol{\mu}(\overline{B} \cap J) = 0, \tag{1.22}$$

a contradiction.

Suppose secondly that $\{1,2\},\{3,4\} \in \mathscr{C}$. Let C be the event that either (i) there exists $i \geq 3$ such that $\vec{F}_M(i) \in \{1,2\}$, or (ii) $\{\vec{F}_M(i) : i \geq 3\} \supseteq \{3,4\}$. On $C \cap J$, we have $k(F) \leq n-2$ a.s. On $\overline{C} \cap J$, by (1.21) the $\vec{F}_M(i)$, $i \geq 3$, 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 (1.22).

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

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

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

$i \geq 3$ and $j \neq i$, and hence

$$\mu(N \geq 2) = 0, \tag{1.23}$$
$$\mu(M \geq 1, N = 1) = 0. \tag{1.24}$$

By similar arguments,

$$\mu(M < 2, N = 0) = 0, \tag{1.25}$$
$$\mu(M = 1) = 0. \tag{1.26}$$

It follows that

$$\begin{aligned}
\mu(N = 1) &= \mu(N = 1, M = 0) && \text{by (1.24)} \\
&= \mu(M = 0) && \text{by (1.25) and (1.23)} \\
&= \mu(\overline{A}_1 \cap \overline{A}_2) \\
&= \mu(\overline{A}_r), \quad r = 1, 2, && \text{by (1.26)}.
\end{aligned}$$

Therefore,

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

By (1.23),

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

where $\mu(N)$ is the mean value of N. This yields (1.20).

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

Example 3. Consider the matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \tag{1.27}$$

Since P is doubly stochastic, by Theorem 4(a), there exists $\mu \in \mathscr{L}(P)$ such that $k(\mu) = 3$ (one may take $\mu(123) = \mu(231) = \frac{1}{2}$). By Lemma 3, 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 $\mu \in \mathscr{L}(P)$, with $k(\mu) < 3$, so that $|\mathscr{C}| \geq 1$. There exists no permutation of S for which the matrix P satisfies (1.20), whence $|\mathscr{C}| \geq 2$ by Proposition 1(c). By parts (a, b) of that proposition, $k(\mu) \leq 1$. In conclusion, $K(P) = \{1, 3\}$.

Example 4. Consider the matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{1.28}$$

We have, as in Example 3, 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 3, 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 7. *For $n \geq 2$ there exists a block measure $\mu \in \mathscr{L}(P_n)$ 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 $\mu \in \mathscr{L}(P_n)$ that is not a block measure.

Proof. Let $n \geq 2$. By Lemma 3 and Theorem 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

$$\begin{aligned} S_r &= (r-1)n/l + \{1,2,\dots,n/l\} \\ &= \{(r-1)n/l + 1, (r-1)n/l + 2, \dots, rn/l\}, \qquad r = 1,2,\dots,l. \end{aligned} \tag{1.29}$$

We describe next a measure $\mu \in \mathscr{L}(P_n)$. 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,

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

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 $\mu \in \mathscr{L}(P_n)$. By examination of μ, μ is an $\mathscr{S}$-block measure.

Conversely, suppose there exists an $\mathscr{S}$-block measure $\mu \in \mathscr{L}(P_n)$ with corresponding partition $\mathscr{S} = \{S_1, S_2, \dots, S_l\}$ with index set $I = \{1,2,\dots,l\}$. By Theorem 6, equations (1.15) and (1.19) hold. By (1.15), the matrix $\Lambda = (\lambda_{r,s} : r, s \in I)$ satisfies

$$\lambda_{r,s} = \frac{|S_s|}{n}, \qquad r, s \in I. \tag{1.30}$$

By (1.19),

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

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 $\mu \in \mathscr{L}(P_n)$. Let $\mathscr{C} = \mathscr{C}(\mu)$ be given as in (1.12). By Proposition 1(b), it suffices to prove that $|\mathscr{C}| \neq 1$. Assume on the contrary that $|\mathscr{C}| = 1$, and suppose without loss of generality that $\mathscr{C}$ contains the singleton pair $\{1,2\}$. With $P = P_n$, the necessary condition (1.20) becomes

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

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

Acknowledgements

The authors thank Tim Garoni, Wilfrid Kendall, and an anonymous referee for their comments. MH was 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. W. Doeblin, *Exposé de la théorie des chaînes simples constants de Markoff á un nombre fini d'états.* Revue Math. de l'Union Interbalkanique **2** (1938), 77–105.
3. G. R. Grimmett, *Probability on Graphs*, 2nd ed., Cambridge University Press, Cambridge, 2018.
4. G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 4th ed., Oxford University Press, Oxford, 2020.
5. M. L. Huber, *Perfect Simulation*. Boca Raton: Chapman and Hall/CRC, 2015.
6. J. G. Kemeny, J. L. Snell, *Finite Markov Chains*. Van Nostrand, 1963.
7. 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).
8. J. G. Propp and D. B. Wilson, *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).
9. 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.
10. D. B. Wilson, *Perfectly random sampling with Markov chains*, `http://www.dbwilson.com/exact/`.
11. 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.