Chapter 4-3

Aldous - Fill

October 11, 1999

xxx This is intended as a section in a Chapter near Chapter 4; maybe a new Chapter consisting of this and another section on bounding τ_2 via distinguished paths. Need some preliminary discussion, e.g. on relations between τ_1 and τ_2; observe that τ_2 is tied to reversibility whereas coupling isn't.

1 Using coupling to bound variation distance

Recall from Chapter 2 section 4.1 (yyy 9/10/99 version) several ways in which variation distance is used to measure the deviation from stationarity of the time-t distribution of a Markov chain:

$$
\begin{aligned}
d_i(t) : &= \ ||P_i(X_t = \cdot) - \pi(\cdot)|| \\
d(t) : &= \ \max_i d_i(t) \\
\bar{d}(t) : &= \ \max_{ij} ||P_i(X_t = \cdot) - P_j(X_t = \cdot)||.
\end{aligned}
$$

Recall also from yyy the definition of *variation threshold time*

$$
\tau_1 := \min\{t : \bar{d}(t) \le e^{-1}\}.
$$

Since τ_1 is affected by continuization, when using the above definition with a discrete-time chain we write τ_1^{disc} for emphasis.

Coupling provides a methodology for seeking to upper bound $\bar{d}(t)$ and hence τ_1. After giving the (very simple) theory in section 1.1 and some discussion in section 1.2, we proceed to give a variety of examples of its use. We shall say more about theory and applications of coupling in Chapter 8 (where we discuss the related idea of *coupling from the past*) and in Chapter 10 (on interacting random walks).

1

1.1 The coupling inequality

Consider a finite-state chain in discrete or continuous time. Fix states i, j. Suppose we construct a *coupling*, that is a joint process $((X_t^{(i)}, X_t^{(j)}),\ t \geq 0)$ such that

$$(X_t^{(i)}, t \geq 0) \text{ is distributed as the chain started at } i$$
$$(X_t^{(j)}, t \geq 0) \text{ is distributed as the chain started at } j. \tag{1}$$

And suppose there is a random time $T^{ij} \leq \infty$ such that

$$X_t^{(i)} = X_t^{(j)},\ T^{ij} \leq t < \infty. \tag{2}$$

Call such a T^{ij} a *coupling time*. Then the *coupling inequality* is

$$||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| \leq P(T^{ij} > t),\ 0 \leq t < \infty. \tag{3}$$

The inequality holds because

$$\begin{aligned} ||P_i(X_t \in \cdot) - P_j(X_t \in \cdot)|| &= ||P(X_t^{(i)} \in \cdot) - P(X_t^{(j)} \in \cdot)|| \\ &\leq P(X_t^{(i)} \neq X_t^{(j)}) \\ &\leq P(T^{ij} > t). \end{aligned}$$

1.2 Comments on coupling methodology

The coupling inequality provides a method of bounding the variation distance $\bar{d}(t)$, because if we can construct a coupling for an arbitrary pair (i, j) of initial states then

$$\bar{d}(t) \leq \max_{i,j} P(T^{ij} > t).$$

The reader may wish to look at a few of the examples before reading this section in detail.

In applying coupling methodology there are two issues. First we need to specify the coupling, then we need to analyze the coupling time. The most common strategy for constructing couplings is via Markov couplings, as follows. Suppose the underlying chain has state space I and (to take the continuous-time case) transition rate matrix $\mathbf{Q} = (q(i, k))$. Consider a transition rate matrix $\widetilde{\mathbf{Q}}$ on the product space $I \times I$. Write the entries of $\widetilde{\mathbf{Q}}$ as $\tilde{q}(i, j; k, l)$ instead of the logical-but-fussy $\tilde{q}((i, j), (k, l))$. Suppose that, for each pair (i, j) with $j \neq i$,

$$\tilde{q}(i, j; \cdot, \cdot) \text{ has marginals } q(i, \cdot) \text{ and } q(j, \cdot) \tag{4}$$

in other words $\sum_l \tilde{q}(i,j;k,l) = q(i,k)$ and $\sum_k \tilde{q}(i,j;k,l) = q(j,l)$. And suppose that

$$\tilde{q}(i,i;k,k) = q(i,k) \text{ for all } k$$
$$\tilde{q}(i,i;k,l) = 0 \text{ for } l \neq k.$$

Take $(X_t^{(i)}, X_t^{(j)})$ to be the chain on $I \times I$ with transition rate matrix $\widetilde{\mathbf{Q}}$ and initial position (i,j), Then (1) must hold, and $T^{ij} := \min\{t : X_t^{(i)} = X_t^{(j)}\}$ is a coupling time. This construction gives a *natural Markov coupling*, and all the examples where we use the coupling inequality will be of this form. In practice it is much more understandable to define the joint process in words, and we usually do so.

In constructing and analyzing couplings, we often exploit (explicitly or implicitly) some integer-valued metric $\rho(i,j)$ on the state space I. Then with a Markovian coupling,

$$P(T^{ij} > t) = P(\rho(X_t^i, X_t^{(j)}) \geq 1)$$

and it is enough to study the integer-valued process $Z_t := \rho(X_t^{(i)}, X_t^{(j)})$. Typically (Z_t) is *not* Markov, but one can try to compare it with some integer-valued Markov process (Z_t^*). Indeed, in defining the coupling one has in mind trying to make such a comparison possible. Often one shows that for any initial (i,j) the random time T^{ij} is stochastically smaller than the hitting time T_{a0}^* for the comparison chain (Z_t^*) to reach 0 starting from $a := \max_{i,j} \rho(i,j)$. This would imply

$$\bar{d}(t) \leq P(T_{a0}^* > t).$$

Finally, one does calculations with the integer-valued chain (Z_t^*), either bounding the tail probability $P(T_{a0}^* > t)$ directly or (what is often simpler) just bounding the expectation ET_{a0}^*, so that by Markov's inequality and the submultiplicativity property (Chapter 2 Lemma 20) (yyy 9/10/99 version) we have in continuous time

$$\tau_1 \leq eET_{a0}^*; \quad \bar{d}(t) \leq \exp\left(1 - \frac{t}{\tau_1}\right).$$

Here is perhaps the simplest comparison lemma, whose proof is left to the reader.

Lemma 1 (Decreasing functional lemma) *Let (Y_t) be a Markov chain on S and $f : S \to \{0, 1, 2, \ldots, \Delta\}$ a function. Suppose that for each $1 \leq i \leq$*

3

Δ *and each initial state* y *with* $f(y) = i$,

(i) $f(Y_1) \leq i$;

(ii) $P(f(Y_1) \leq i - 1) \geq a_i > 0$.

Then

$$\max_{y \in S} E_y T_A \leq \sum_{i=1}^{\Delta} 1/a_i$$

where $A := \{y : f(y) = 0\}$.

We now start a series of examples. Note that when presenting a coupling proof we don't need to explicitly check irreducibility, because the conclusion of a bound on coupling time obviously implies irreducibility.

1.3 Random walk on a dense regular graph

(Chapter 5 Example 16).

Consider an r-regular n-vertex graph. Write $\mathcal{N}(v)$ for the set of neighbors of v. For any pair v, w we can define a $1 - 1$ map $\theta_{v,w} : \mathcal{N}(v) \to \mathcal{N}(w)$ such that $\theta_{v,w}(x) = x$ for $x \in \mathcal{N}(v) \cap \mathcal{N}(w)$. Consider discrete-time random walk on the graph. We define a "greedy coupling" by specifying the joint transition matrix

$$\tilde{p}(v, w; x, \theta_{v,w}(x)) = 1/r, \quad x \in \mathcal{N}(v).$$

That is, from vertices v and w, if the first chain jumps to x then the second chain jumps to $\theta_{v,w}(x)$, and we maximize the chance of the two chains meeting after a single step. In general one cannot get useful bounds on the coupling time. But consider the dense case, where $r > n/2$. As observed in Chapter 5 Example 16, here $|\mathcal{N}(v) \cap \mathcal{N}(w)| \geq 2r - n$ and so the coupled process (X_t, Y_t) has the property that for $w \neq v$

$$P(X_{t+1} = Y_{t+1} | X_t = v, Y_t = w) = \frac{|\mathcal{N}(v) \cap \mathcal{N}(w)|}{r} \geq \frac{2r - n}{r}$$

implying that the coupling time T (for any initial pair of states) satisfies

$$P(T > t) \leq \left(\frac{n - r}{r} \right)^t.$$

So the coupling inequality implies $\bar{d}(t) \leq (\frac{n-r}{r})^t$. In particular the variation threshold satisfies

$$\tau_1^{\text{disc}} = O(1) \text{ as } n \to \infty, \ r/n \to \alpha > 1/2.$$

4

1.4   Continuous-time random walk on the d-cube

(Chapter 5 Example 15).

For $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$ in $I = \{0, 1\}^d$, let $\mathcal{D}(\mathbf{i}, \mathbf{j})$ be the set of coordinates u where $\mathbf{i}$ and $\mathbf{j}$ differ. Write $\mathbf{i}^u$ for the state obtained by changing the u'th coordinate of $\mathbf{i}$. Recall that in continuous time the components move independently as 2-state chains with transition rates $1/d$. Define a coupling in words as "run unmatched coordinates independently until they match, and then run them together". Formally, the non-zero transitions of the joint process are

$$\begin{aligned}
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}^u, \mathbf{j}^u) &= 1/d \text{ if } u \notin \mathcal{D}(\mathbf{i}, \mathbf{j}) \\
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}^u, \mathbf{j}) &= 1/d \text{ if } u \in \mathcal{D}(\mathbf{i}, \mathbf{j}) \\
\tilde{q}(\mathbf{i}, \mathbf{j}; \mathbf{i}, \mathbf{j}^u) &= 1/d \text{ if } u \in \mathcal{D}(\mathbf{i}, \mathbf{j}).
\end{aligned}$$

For each coordinate which is initially unmatched, it takes exponential (rate $2/d$) time until it is matched, and so the coupling time $T = T^{\mathbf{ij}}$ satisfies

$$T \stackrel{d}{=} \max(\xi_1, \ldots, \xi_{d_0})$$

where the (ξ_u) are independent exponential (rate $2/d$) and $d_0 = |\mathcal{D}(\mathbf{i}, \mathbf{j})|$ is the initial number of unmatched coordinates. So

$$P(T \leq t) = (1 - \exp(-2t/d))^{d_0}$$

and the coupling inequality bounds variation distance as

$$\bar{d}(t) \leq 1 - (1 - \exp(-2t/d))^d.$$

This leads to an upper bound on the variation threshold time

$$\tau_1 \leq (\tfrac{1}{2} + o(1)) d \log d \text{ as } d \to \infty.$$

This example is discussed in more detail in Chapter 5 Example 15 (yyy 4/23/96 version) where it is shown that

$$\tau_1 \sim \tfrac{1}{4} d \log d \text{ as } d \to \infty$$

so the coupling bound is off by a factor of 2.

5

1.5 The graph-coloring chain

Fix a n-vertex graph with maximal degree r. Fix an integer $c \geq r + 2$ and consider $[c] := \{1, 2, \ldots, c\}$ as a set of c colors. Let $\mathrm{col}(G, c)$ be the set of c-colorings of G, where a *c-coloring* is an assignment of a color to each vertex, in such a way that no two adjacent vertices have the same color. One can put a natural "product graph" structure on $\mathrm{col}(G, c)$, in which two colorings are adjacent if they differ at only one vertex. It is not hard to check that the condition $c \geq r + 2$ ensures that $\mathrm{col}(G, c)$ is non-empty and the associated graph is connected. There is a natural discrete-time Markov chain on $\mathrm{col}(G, c)$:

> Pich a vertex v of G uniformly at random, pick a color γ uniformly at random, assign color γ to vertex v if feasible (i.e. if no neighbor of v has color γ), else retain existing color of v.

Under certain conditions a simple coupling analysis succeeds in bounding the mixing time. (The bound is far from sharp – see Notes).

Proposition 2 *If $c > 4r$ then $\bar{d}(t) \leq n \exp\left(-\frac{(c-4r)t}{cn}\right)$ and so $\tau_1^{\mathrm{disc}} \leq 1 + \frac{cn}{c-4r}(1 + \log n)$.*

Proof. We couple two versions of the chain by simply using the same v and γ in both chains at each step. Write D_t for the number of vertices at which the colors in the two chains differ. Then $D_{t+1} - D_t \in \{-1, 0, 1\}$ and the key estimate is the following.

Lemma 3 *Conditional on the state of the coupled process at time t,*

$$P(D_{t+1} = D_t + 1) \quad \leq \quad \frac{2rD_t}{cn} \tag{5}$$

$$P(D_{t+1} = D_t - 1) \quad \geq \quad \frac{(c - 2r)D_t}{cn} \tag{6}$$

Proof. In order that $D_{t+1} = D_t + 1$ it is necessary that the chosen pair (v, γ) is such that

> (*) there exists a neighbor (w, say) of v such that w has color γ in one chain but not in the other chain.

But the total number of pairs (v, γ) equals nc while the number of pairs satisfying (*) is at most $D_t \cdot 2r$. This establishes (5). Similarly, for $D_{t+1} = D_t - 1$ it is sufficient that v is currently unmatched and that no neighbor

6

of v in either chain has color γ; the number of such pairs (v, γ) is at least $D_t \cdot (c - 2r)$. $\square$

Lemma 3 implies $E(D_{t+1} - D_t | D_t) \leq -(c - 4r) D_t / (cn)$ and so

$$E D_{t+1} \leq \kappa E D_t; \quad \kappa := 1 - \frac{c-4r}{cn}.$$

Since $D_0 \leq n$ we have, for any initial pair of states,

$$P(D_t \geq 1) \leq E D_t \leq \kappa^t n \leq n \exp\left(-\frac{(c-4r)t}{cn}\right)$$

and the coupling lemma establishes the Proposition.

1.6 Permutations and words

The examples in sections 1.4 and 1.5 were simple prototypes of *interacting particle systems*, more examples of which appear in Chapter 10, whose characteristic property is that a step of the chain involves only "local" change. Chains making "global" changes are often hard to analyze, but here is a simple example.

Fix a finite alphabet A of size $|A|$. Fix m, and consider the set A^m of "words" $\mathbf{x} = (x_1, \ldots, x_m)$ with each $x_i \in A$. Consider the discrete-time Markov chain on A^m in which a step $\mathbf{x} \to \mathbf{y}$ is specified by the following two-stage procedure.

Stage 1. Pick a permutation σ of $\{1, 2, \ldots, m\}$ uniformly at random from the set of permutations σ satisfying $x_{\sigma(i)} = x_i \forall i$.

Stage 2. Let $(c_j(\sigma); j \geq 1)$ be the cycles of σ. For each j, and independently as j varies, pick uniformly an element α_j of A, and define $y_i = \alpha_j$ for every $i \in c_j(\sigma)$.

Here is an alternative description. Write Π for the set of permutations of $\{1, \ldots, m\}$. Consider the bipartite graph on vertices $A^m \cup \Pi$ with edge-set $\{(\mathbf{x}, \sigma) : x_{\sigma(i)} = x_i \forall i\}$. Then the chain is random walk on this bipartite graph, watched every second step (that is, when it is in A^m).

From the second description, it is clear that the stationary probabilities $\pi(\mathbf{x})$ are proportional to the degree of $\mathbf{x}$ in the bipartite graph, giving

$$\pi(\mathbf{x}) \propto \prod_a n_a(\mathbf{x})!$$

where $n_a(\mathbf{x}) = |\{i : x_i = a\}|$. We shall use a coupling argument to establish the following bound on variation distance:

$$\bar{d}(t) \leq m \left(1 - \frac{1}{|A|}\right)^t \tag{7}$$

implying that the variation threshold satisfies

$$\tau_1^{\mathrm{disc}} \le 1 + \frac{1 + \log m}{-\log(1 - \frac{1}{|A|})} \le 1 + (1 + \log m)|A|.$$

The construction of the coupling depends on the following lemma.

Lemma 4 *Given finite sets F^1, F^2 we can construct (for $u = 1, 2$) a uniform random permutation σ^u of F^u with cycles $(C_j^u; j \ge 1)$, where the cycles are labeled such that*

$$C_j^1 \cap F^1 \cap F^2 = C_j^2 \cap F^1 \cap F^2 \ \text{for all } j.$$

In the equality we interpret the C_j^u as sets.

Proof. Given a permutation σ of a finite set G, there is an induced permutation on a subset G' obtained by deleting from the cycle representation of σ those elements not in G'. It is easy to check that, for a uniform random permutation of G, the induced random permutation of G' is also uniform. In the setting of the lemma, take a uniform random permutation σ of $F^1 \cup F^2$, and let σ^u be the induced random permutations of F^u. Then the equality holds because each side is representing the cycles of the induced permutation on $F^1 \cap F^2$. $\square$

We construct a step $(\mathbf{x}^1, \mathbf{x}^2) \to (\mathbf{Y}^1, \mathbf{Y}^2)$ of the coupled processes as follows. For each $a \in A$, set $F^{1,a} = \{i : x_i^1 = a\}$, $F^{2,a} = \{i : x_i^2 = a\}$. Take random permutations $\sigma^{1,a}$, $\sigma^{2,a}$ as in the lemma, with cycles $C_j^{1,a}, C_j^{2,a}$. Then $(\sigma^{1,a}, a \in A)$ define a uniform random permutation σ^1 of $\{1, \ldots, m\}$, and similarly for σ^2. This completes stage 1. For stage 2, for each pair (a, j) pick a uniform random element α_j^a of A and set

$$Y_i^1 = \alpha_j^a \ \text{for every } i \in C_j^{1,a}$$

$$Y_i^2 = \alpha_j^a \ \text{for every } i \in C_j^{2,a}.$$

This specifies a Markov coupling. By construction

$$\text{if } x_i^1 = x_i^2 \qquad \text{then } Y_i^1 = Y_i^2$$
$$\text{if } x_i^1 \ne x_i^2 \quad \text{then } P(Y_i^1 = Y_i^2) = 1/|A|$$

because Y_i^1 and Y_i^2 are independent uniform choices from A. So the coupled processes $(\mathbf{X}^1(t), \mathbf{X}^2(t))$ satisfy

$$P(X_i^1(t) \ne X_i^2(t)) = \left(1 - \frac{1}{|A|}\right)^t P(X_i^1(0) \ne X_i^2(0)).$$

In particular $P(\mathbf{X}^1(t) \ne \mathbf{X}^2(t)) \le m(1 - 1/|A|)^t$ and the coupling inequality (3) gives (7).

8

1.7 Card-shuffling by random transpositions

We mentioned in Chapter 1 section 1.4 (yyy 7/20/99 version) that card-shuffling questions provided a natural extrinsic motivation for the study of mixing times. The example here and in section 1.9 give a first study of mathematically (if not physically) simple random shuffles, and these discrete-time chains are prototypes for more complex chains arising in other contexts.

Consider a d-card deck. The *random transpositions* shuffle is:

> Make two independent uniform choices of cards, and interchange them.

With chance $1/d$ the two choices are the same card, so no change results. To make a coupling analysis, we first give an equivalent reformulation.

> Pick a label a and a position i uniformly at random; interchange the label-a card with the card in position i.

This reformulation suggests the coupling in which the same choice of (a, i) is used for each chain. In the coupled process (with two arbitrary starting states) let D_t be the number of unmatched cards (that is, cards whose positions in the two decks are different) after t steps. Then

(i) $D_{t+1} \leq D_t$.

(ii) $P(D_{t+1} \leq j - 1 \mid D_t = j) \geq j^2/d^2$.

Here (i) is clear, and (ii) holds because whenever the card labeled a and the card in position i are both unmatched, the step of the coupled chain creates at least one new match (of the card labeled a).

Noting that D_t cannot take value 1, we can use the decreasing functional lemma (Lemma 1) to show that the coupling time $T := \min\{t : D_t = 0\}$ satisfies

$$ET \leq \sum_{j=2}^{d} d^2/j^2 \leq d^2\left(\tfrac{\pi^2}{6} - 1\right).$$

In particular, the coupling inequality implies $\tau_1^{\text{disc}} = O(d^2)$.

We revisit this example in Chapter 7 Example 18 (yyy 1/31/94 version) where it is observed that in fact

$$\tau \sim \tfrac{1}{2} d \log d \tag{8}$$

An analogous continuous-space chain on the simplex is studied in Chapter 13-4 Example 3 (yyy 7/29/99 version)

9

1.8 Reflection coupling on the n-cycle

Consider continuous-time random walk on the n-cycle $I = \{0, 1, 2, \ldots, n-1\}$. That is, the transition rates are

$$i \xrightarrow{1/2} i + 1; \quad i \xrightarrow{1/2} i - 1$$

where here and below ± 1 is interpreted modulo n. One can define a coupling by specifying the following transition rates for the bivariate process.

$$(i, i) \xrightarrow{1/2} (i+1, i+1); \quad (i, i) \xrightarrow{1/2} (i-1, i-1)$$

$$\text{(if } |j - i| > 1) \quad (i, j) \xrightarrow{1/2} (i+1, j-1); \quad (i, j) \xrightarrow{1/2} (i-1, j+1)$$

$$(i, i+1) \xrightarrow{1/2} (i, i); \quad (i, i+1) \xrightarrow{1/2} (i+1, i+1); \quad (i, i+1) \xrightarrow{1/2} (i-1, i+2) \quad (9)$$

and symmetrically for $(i+1, i)$. The joint process $((X_t^{(0)}, X_t^{(k)}), t \geq 0)$ started at $(0, k)$ can be visualized as follows. Let $\phi(i) := k - i \bmod n$. Picture the operation of ϕ as reflection in a mirror which passes through the points $\{x_1, x_2\} = \{k/2, k/2 + n/2 \bmod n\}$ each of which is either a vertex or the middle of an edge. In the simplest case, where x_1 and x_2 are vertices, let (X_t^0) be the chain started at vertex 0, let $T^{0k} = \min\{t : X_t \in \{x_1, x_2\}\}$ and define

$$
\begin{aligned}
X_t^{(k)} &= \phi(X_t^{(0)}), \ t \leq T^{0k} \\
&= X_t^{(0)}, \ t > T^{0k}.
\end{aligned}
$$

This constructs a bivariate process with the transition rates specified above, with coupling time T^{0k}, and the pre-T^{0k} path of $X^{(k)}$ is just the reflection of the pre-T^{0k} path of $X^{(0)}$. In the case where a mirror point is the middle of an edge $(j, j + 1)$ and the two moving particles are at j and $j + 1$, we don't want simultaneous jumps across that edge; instead (9) specifies that attempted jumps occur at *independent* times, and the process is coupled at the time of the first such jump.

It's noteworthy that in this example the coupling inequality

$$||P_0(X_t \in \cdot) - P_k(X_t \in \cdot)|| \leq P(X_t^{(0)} \neq X_t^{(k)})$$

is in fact an equality. Indeed this assertion, at a given time t, is equivalent to the assertion

$$P(X_t^{(0)} \in \cdot, T > t) \text{ and } P(X_t^{(k)} \in \cdot, T > t) \text{ have disjoint support.}$$

But the support A^0 of the first measure is the set of vertices which can be reached from 0 without meeting or crossing any mirror point (and similarly for A^k); and A^0 and A^k are indeed disjoint.

It is intuitively clear that the minimum over k of T^{0k} is attained by $k = \lfloor n/2 \rfloor$: we leave the reader to find the simple non-computational proof. It follows, taking e.g. the simplest case where n is multiple of 4, that we can write

$$\bar{d}(t) = P(T_{\{-n/4,n/4\}} > t) \tag{10}$$

where $T_{\{-n/4,n/4\}}$ is the hitting time for continuous-time random walk on the integers.

Parallel results hold in discrete time but only when the chains are suitably *lazy*. The point is that (9) isn't allowable as transition *probabilities*. However, if we fix $0 < a \leq 1/3$ then the chain with transition probabilities

$$i \overset{a}{\longrightarrow} i + 1; \quad i \overset{a}{\longrightarrow} i - 1$$

(and which holds with the remaining probability) permits a coupling of the form (9) with all transition probabilities being a instead of $1/2$. The analysis goes through as above, leading to (10) where T refers to the discrete-time lazy walk on the integers.

Similar results hold for random walk on the n-path (Chapter 5 Example 8) (yyy 4/23/96 version). and we call couplings of this form *reflection couplings*. They are simpler in the context of continuous-path Brownian motion – see Chapter 13-4 section 1 (yyy 7/29/99 version).

1.9   Card-shuffling by random adjacent transpositions

As in section 1.7 we take a d-card deck; here we define a (lazy) shuffle by

> With probability $1/2$ make no change; else pick a uniform random position $i \in \{1, 2, \ldots, d\}$ and interchange the cards in positions i and $i + 1$ (interpret $d + 1$ as 1).

To study this by coupling, consider two decks. In some positions i the decks match (the label on the card in position i is the same in both decks). Write $\mathcal{D}$ for the set of i such that either position i or position $i + 1$ or both match. Specify a step of the coupled chain by:

$$
\begin{aligned}
P(\text{interchange } i \text{ and } i + 1 \text{ in each deck}) &= \tfrac{1}{2d}, & i \in \mathcal{D} \\
P(\text{interchange } i \text{ and } i + 1 \text{ in first deck, no change in second deck}) &= \tfrac{1}{2d}, & i \notin \mathcal{D}
\end{aligned}
$$

$$P(\text{interchange } i \text{ and } i+1 \text{ in second deck, no change in first deck}) \;=\; \tfrac{1}{2d}, \quad i \notin \mathcal{D}$$
$$P(\text{no change in either deck}) \;=\; \tfrac{|\mathcal{D}|}{2d}.$$

Consider a particular card a. From the coupling description we see
(a) if the card gets matched then it stays matched;
(b) while unmatched, at each step the card can move in at most one of the decks.
It follows that the "clockwise" distance $D(t) := X_a^1(t) - X_a^2(t)$ mod d between the positions of card a in the two decks behaves exactly as a lazy random walk on the d-cycle:

$$p_{j,j+1} = p_{j,j-1} = 1/d, \quad 1 \le j \le d$$

until $D(t)$ hits 0. By the elementary formula for mean hitting times on the cycle (Chapter 5 eq. (24)) (yyy 4/23/96 version), the mean time $T^{(a)}$ until card a becomes matched satisfies

$$ET^{(a)} \le \tfrac{d}{2} \, \tfrac{d^2}{4}$$

uniformly over initial configurations. By submultiplicativity (Chapter 2 section 4.3) (yyy 9/10/99 version)

$$P(T^{(a)} > md^3/4) \le 2^{-m}, \; m = 1, 2, \ldots.$$

The chains couple at time $T := \max_a T^{(a)}$ and so

$$\bar{d}(md^3/4) \le P(T > md^3/4) \le d2^{-m}.$$

In particular
$$\tau_1^{\text{disc}} = O(d^3 \log d).$$

In this example it turns out that coupling does give the correct order of magnitude; the corresponding lower bound

$$\tau_1^{\text{disc}} = \Omega(d^3 \log d)$$

was proved by Wilson [18]. Different generalizations of this example appear in section 1.13 and in Chapter 14 section 5 (yyy 3/10/94 version), where we discuss relaxation times.

A generalization of this example, the *interchange process*, is studied in Chapter 14 section 5 (yyy 3/10/94 version).

1.10   Independent sets

Fix a graph G on n vertices with maximal degree r, An *independent set* is a set of vertices which does not contain any adjacent vertices. Fix m and consider the space of all independent sets of size m in G. Picture an independent set $\mathbf{x}$ as a configuration of m particles at distinct vertices, with no two particles at adjacent vertices. A natural discrete-time chain $(\mathbf{X}_t)$ on I is

> pick a uniform random particle a and a uniform random vertex v; move particle a to vertex v if feasible, else make no move.

To study mixing times, we can define a coupling $(\mathbf{X}_t, \mathbf{Y}_t)$ by simply making the same choice of (a, v) in each of the two coupled chains, where at each time we invoke a matching of particles in the two realizations which is arbitrary except for matching particles at the same vertex. To analyze the coupling, let ρ be the natural metric on I: $\rho(\mathbf{x}, \mathbf{y}) = $ number of vertices occupied by particles of $\mathbf{x}$ but not by particles of $\mathbf{y}$. Clearly $D_t := \rho(\mathbf{X}_t, \mathbf{Y}_t)$ can change by at most 1 on each step. Let us show that, for initial states with $\rho(\mathbf{x}, \mathbf{y}) = d > 0$,

$$P_{(\mathbf{x},\mathbf{y})}(D_1 = d + 1) \quad \leq \quad \frac{m - d}{m} \frac{2d(r + 1)}{n} \tag{11}$$

$$P_{(\mathbf{x},\mathbf{y})}(D_1 = d - 1) \quad \geq \quad \frac{d}{m} \frac{n - (m + d - 2)(r + 1)}{n}. \tag{12}$$

For in order that $D_1 = d + 1$ we must first choose a matched particle a (chance $(m - d)/d$) and then choose a vertex v which is a neighbor of (or the same as) some vertex v' which is in exactly one of $\{\mathbf{x}, \mathbf{y}\}$: there are $2d$ such vertices v' and hence at most $2d(r + 1)$ possibilities for v. This establishes (11). Similarly, in order that $D_t = d - 1$ it is sufficient that we pick an unmatched particle a (chance d/m) and then choose a vertex v which is not a neighbor of (or the same as) any vertex v' which is occupied in one or both realizations by some particle other than a: there are $m + d - 2$ such forbidden vertices v' and hence at most $(m + d - 2)(r + 1)$ forbidden positions for v. This establishes (12).

From (11,12) a brief calculation gives

$$\begin{aligned}
E_{(\mathbf{x},\mathbf{y})}(D_1 - d) \quad &\leq \quad \tfrac{-d}{mn}(n - (3m - d - 2)(r + 1)) \\
&\leq \quad \tfrac{-d}{mn}(n - 3(m - 1)(r + 1)).
\end{aligned}$$

In other words

$$E_{(\mathbf{x},\mathbf{y})}D_1 \leq \kappa d; \quad \kappa := 1 - \frac{n - 3(m-1)(r+1)}{mn}.$$

If $m < 1 + \frac{n}{3(r+1)}$ then $\kappa < 1$. In this case, by copying the end of the analysis of the graph-coloring chain (section 1.5)

$$\bar{d}(t) \leq m\kappa^t; \quad \tau_1 = O\left(\frac{\log m}{1-\kappa}\right).$$

To clarify the size-asymptotics, suppose $m, n \to \infty$ with $m/n \to \rho < \frac{1}{3(r+1)}$. Then for fixed ρ

$$\tau_1 = O(n \log n).$$

1.11 Two base chains for genetic algorithms

One way of motivating study of Markov chains on combinatorial sets with uniform stationary distributions is as "base chains" on which to base Markov chain Monte Carlo, that is to create other chains designed to have some specified distribution as their stationary distributions. Here is a typical base chain underlying genetic algorithms.

Fix integers $K, L \geq 1$ with K even. A state of the chain is a family of *words* $(\mathbf{x}^k, 1 \leq k \leq K)$, where each word is a binary L-tuple $\mathbf{x}^k = (x_l^k, 1 \leq l \leq L)$. A step of the chain is defined as follows.

Use a uniform random permutation π of $\{1, 2, \ldots, K\}$ to partition the words into $K/2$ pairs $\{\mathbf{x}^{\pi(1)}, \mathbf{x}^{\pi(2)}\}$, $\{\mathbf{x}^{\pi(3)}, \mathbf{x}^{\pi(4)}\}$ . . .. Create a new pair $\{\mathbf{y}^1, \mathbf{y}^2\}$ from $\{\mathbf{x}^{\pi(1)}, \mathbf{x}^{\pi(2)}\}$ by setting, independently for each $1 \leq l \leq L$

$$P((y_l^1, y_l^2) = (x_l^{\pi(1)}, x_l^{\pi(2)})) = P((y_l^1, y_l^2) = (x_l^{\pi(2)}, x_l^{\pi(1)})) = 1/2. \tag{13}$$

Repeat independently for $1 \leq i \leq K/2$ to create new pairs $\{\mathbf{y}^{2i-1}, \mathbf{y}^{2i}\}$ from $\{\mathbf{x}^{\pi(2i-1)}, \mathbf{x}^{\pi(2i)}\}$. The new state is the family of words $\mathbf{y}^k$.

Associated with an initial state $(\mathbf{x}^k)$ is a vector of column sums $\mathbf{m} = (m_l, 1 \leq l \leq L)$ where $m_l = \sum_k x_l^k$. These sums are preserved by the chain, so the proper state space is the space $I_\mathbf{m}$ of families with column-sums $\mathbf{m}$. The transition matrix is symmetric and so the chain is reversible with uniform stationary distribution on $I_\mathbf{m}$.

14

To describe the coupling, first rephrase (13) in words as "(y_l^1, y_l^2) is $\{x_l^{\pi(1)}, x_l^{\pi(2)}\}$ in random order, either forwards or backwards". Now specify the coupling as follows.

(i) Use the same random permutation π for both chains.

(ii) For each i and each l, in creating the new words (y_l^{2i-1}, y_l^{2i}) from the old words $\{x_l^{\pi(2i-1)}, x_l^{\pi(2i)})\}$ use the same choice (forwards or backwards) in both chains, except when $(x_l^{\pi(2i-1)}, x_l^{\pi(2i)}) = (1, 0)$ for one chain and $= (0, 1)$ for the other chain, in which case use opposite choices of (forwards, backwards) in the two chains.

To study the coupled processes $(\mathbf{X}(t), \hat{\mathbf{X}}(t))$, fix l and consider the number $W(t) := \sum_{k=1}^{K} |X_l^k(t) - \hat{X}_l^k(t)|$ of words in which the l'th letter is not matched in the two realizations. Suppose $W(0) = w$. Consider the creation of the first two new words in each chain. The only way that the number of matches changes is when we use opposite choices of (forwards, backwards) in the two chains, in which case two new matches are created. The chance that the l'th letter in the two chains is 1 and 0 in the $\pi(1)$'th word and is 0 and 1 in the $\pi(2)$'th word equals $\frac{w/2}{K} \times \frac{w/2}{K-1}$, and so (taking into account the symmetric case) the mean number of new matches at l in these two words equals $\frac{w^2}{K(K-1)}$. Summing over the $K/2$ pairs,

$$E(W(1)|W(0) = w) = w - \frac{w^2}{2(K-1)}.$$

We can now apply a comparison lemma (Chapter 2 Lemma 32) (yyy 9/10/99 version) which concludes that the hitting time T^l of $W(t)$ to 0 satisfies

$$ET^l \leq \sum_{w=2}^{K} \frac{2(K-1)}{w^2} \leq 2K.$$

Since $T := \max_l T^l$ is a coupling time, a now-familiar argument shows that for $u = 1, 2, \ldots$

$$\bar{d}(4uK) \leq P(T > 4uK) \leq LP(T^l > u \cdot 4K) \leq L \, 2^{-u}$$

and so

$$\tau_1^{\text{disc}} = O(K \log L).$$

Open Problem 5 *Show* $\tau_1^{\text{disc}} = O(\log K \times \log L)$.

15

We expect this bound by analogy with the "random transpositions" shuffle (section 1.7). Loosely speaking, the action of the chain on a single position in words is like the random transpositions chain speeded up by a factor $K/2$, so from (8) we expect its mixing time to be $\Theta(\log K)$. It would be interesting to study this example via the group representation or strong stationary time techniques which have proved successful for the random transpositions chain.

To make a metaphor involving biological genetics, the letters represent chromosomes and the words represent the chromosomal structure of a gamete; the process is "sexual reproduction from the viewpoint of gametes". If instead we want a word to represent a particular chromosome and the letters to represent genes within that chromosome, then instead of flipping bits independently it is more natural to model *crossover*. That is, consider a chain in which the rule for creating a new pair $\{\mathbf{y}^{2i-1}, \mathbf{y}^{2i}\}$ from $\{\mathbf{x}^{\pi(2i-1)}, \mathbf{x}^{\pi(2i)}\}$ becomes

Take U_i uniform on $\{1, 2, \ldots, L, L+1\}$. Define

$$
\begin{aligned}
(y_l^{2i-1}, y_l^{2i}) &= (x_l^{\pi(2i-1)}, x_l^{\pi(2i)}), \ l < U \\
(y_l^{2i-1}, y_l^{2i}) &= (x_l^{\pi(2i)}, x_l^{\pi(2i-1)}), \ l \geq U.
\end{aligned}
$$

As an exercise (hint in Notes), find a coupling argument to show that for this chain

$$
\tau_1^{\text{disc}} = O(KL^2). \tag{14}
$$

1.12   Path coupling

In certain complicated settings it is useful to know that it is enough to couple versions of the chain which start in "nearby" states. To say this carefully, let I be finite and consider a $\{0, 1, 2, \ldots\}$-valued function $\rho(i, j)$ defined on some symmetric subset $\mathcal{E} \subset I \times I$. Call ρ a *pre-metric* if
(i) $\rho(i, j) = 0$ iff $i = j$.
(ii) $\rho(i, j) = \rho(j, i)$.
(iii) $\rho(i_0, i_k) \leq \sum_{u=0}^{k-1} \rho(i_u, i_{u+1})$, whenever (i_0, i_k) and each (i_u, i_{u+1}) are in $\mathcal{E}$.

Clearly a pre-metric extends to a metric $\bar{\rho}$ by defining

$$
\bar{\rho}(i, j) := \min \left\{ \sum_u \rho(i_u, i_{u+1}) \right\} \tag{15}
$$

the minimum over all paths $i = i_0, i_1, \ldots, i_k = j$ with each $(i_u, i_{u+1}) \in \mathcal{E}$. Note $\bar{\rho}(i, j) \leq \rho(i, j)$ for $(i, j) \in \mathcal{E}$.

Lemma 6 *Let S be a state space. Let $(\mu_{i,i+1}, \ 0 \le i \le d-1)$ be probability distributions on $S \times S$ such that the second marginal of $\mu_{i,i+1}$ coincides with the first marginal of $\mu_{i+1,i+2}$ for $0 \le i \le d-2$. Then there exists a S-valued random sequence $(V_i, 0 \le i \le d)$ such that $\mu_{i,i+1} = \text{dist}(V_i, V_{i+1})$ for $0 \le i \le d-1$.*

Proof. Just take (V_i) to be the non-homogeneous Markov chain whose transition probabilities $P(V_{i+1} \in \cdot | V_i = v)$ are the conditional probabilities determined by the specified joint distribution $\mu_{i,i+1}$.

Lemma 7 (Path-coupling lemma) *Take a discrete-time Markov chain (X_t) with finite state space I. Write $X_1^{(i)}$ for the time-1 value of the chain started at state i. Let ρ be a pre-metric defined on some subset $\mathcal{E} \subset I \times I$. Suppose that for each pair (i,j) in $\mathcal{E}$ we can construct a joint law $(X_1^{(i)}, X_1^{(j)})$ such that*

$$E\bar{\rho}(X_1^{(i)}, X_1^{(j)}) \le \kappa \rho(i,j) \tag{16}$$

for some constant $0 < \kappa < 1$. Then

$$\bar{d}(t) \le \Delta_\rho \kappa^t \quad (\kappa < 1) \tag{17}$$

where $\Delta_\rho := \max_{i,j \in I} \bar{\rho}(i,j)$.

See the Notes for comments on the case $\kappa = 1$.

Proof. Fix states i, j and consider a path (i_u) attaining the minimum in (15). For each u let $(X_1^{(i_u)}, X_1^{(i_u+1)})$ have a joint distribution satisfying (16). By Lemma 6 there exists a random sequence $(X_1^{(i)} = X_1^{(i_0)}, X_1^{(i_1)}, \ldots, X_1^{(j)})$ consistent with these bivariate distributions. In particular, there is a joint distribution $(X_1^{(i)}, X_1^{(j)})$ such that

$$E\bar{\rho}(X_1^{(i)}, X_1^{(j)}) \le E \sum_u \bar{\rho}(X_1^{(i_u)}, X_1^{(i_u+1)}) \le \kappa \sum_u \rho(i_u, i_{u+1}) = \kappa \bar{\rho}(i,j).$$

This construction gives one step of a coupling of two copies of the chain started at arbitrary states, and so extends to a coupling $((X_t^{(i)}, X_t^{(j)}), \ t = 0, 1, 2, \ldots)$ of two copies of the entire processes. The inequality above implies

$$E(\bar{\rho}(X_{t+1}^{(i)}, X_{t+1}^{(j)}) | X_t^{(i)}, X_t^{(j)}) \le \kappa \bar{\rho}(X_t^{(i)}, X_t^{(j)})$$

and hence

$$P(X_t^{(i)} \ne X_t^{(j)}) \le E\bar{\rho}(X_t^{(i)}, X_t^{(j)}) \le \kappa^t \bar{\rho}(i,j) \le \kappa^t \Delta_\rho$$

17

establishing (17). □

Bubley and Dyer [2] introduced Lemma 7 and the name *path-coupling*. It has proved useful in extending the range of applicability of coupling methods in settings such as graph-coloring (Bubley et al [4] Vigoda [17]) and independent sets (Luby and Vigoda [13]). These are too intricate for presentation here, but the following example will serve to illustrate the use of path-coupling.

1.13   Extensions of a partial order

Fix a partial order $\preceq$ on an n-element set, and let I_m be the set of *linear extensions* of $\preceq$, that is to say total orders consistent with the given partial order. We can define a discrete-time Markov chain on I_m by re-using the idea in the "random adjacent transpositions" example (section 1.9). Let $w(\cdot)$ be a probability distribution on $\{1, 2, \ldots, n-1\}$. Define a step of the chain as follows.

> Pick position i with probability $w(i)$, and independently pick one of { stay, move } with probability $1/2$ each. If pick "move" then interchange the elements in positions i and $i + 1$ if feasible (i.e. if consistent with the partial order); else make no change.

The transition matrix is symmetric, so the stationary distribution is uniform on I_m.

To analyze by coupling, define one step of a bivariate coupled process as follows.

> Make the same choice of i in both chains. Also make the same choice of { move, stay }, except in the case where the elements in positions i and $i + 1$ are the same elements in opposite order in the two realizations, in which case use the opposite choices of { stay, move }.

The coupling is similar (but not identical) to that in section 1.9, where the underlying chain is that corresponding to the "null" partial order. For a general partial order, the coupling started from an arbitrary pair of states seems hard to analyze directly. For instance, an element in the same position in both realizations at time t may not remain so at time $t + 1$. Instead we use path-coupling, following an argument of Bubley and Dyer [3]. Call two states $\mathbf{x}$ and $\mathbf{y}$ *adjacent* if they differ by only one (not necessarily adjacent) transposition; if the transposed cards are in positions $i < j$ then let $\rho(\mathbf{x}, \mathbf{y}) =$

$j - i$. We want to study the increment $\Phi := \rho(X_1, Y_1) - \rho(\mathbf{x}, \mathbf{y})$ where (X_1, Y_1) is the coupled chain after one step from $(\mathbf{x}, \mathbf{y})$. The diagram shows a typical pair of adjacent states.

a	b	c	α	d	e	f	β	g	h
a	b	c	β	d	e	f	α	g	h
position	$\cdot$	$\cdot$	i	$\cdot$	$\cdot$	$\cdot$	j	$\cdot$	$\cdot$

Observe first that any choice of position other than $i - 1, i, j - 1, j$ will have no effect on Φ. If position i and "move" are chosen, then $\{\alpha, d\}$ are interchanged in the first chain and $\{\beta, d\}$ in the second; both lead to feasible configurations by examining the relative orders in the other chain's previous configuration. This has chance $w(i)/2$ and leads to $\Phi = -1$. If position $i - 1$ and "move" are chosen (chance $w(i-1)/2$), then if either or both moves are feasible $\Phi = 1$, while if neither are feasible then $\Phi = 0$. Arguing similarly for choices $j - 1, j$ leads to

$$E(\Phi) \leq \tfrac{1}{2}(w(i - 1) - w(i) - w(j - 1) + w(j)).$$

This estimate remains true if $j = i + 1$ because in that case choosing position i (chance $w(i)$) always creates a match. Now specify

$$w(i) := \tfrac{i(n-i)}{w_n}, \quad w_n := \sum_{j=1}^{n-1} j(n - j)$$

and then $E\Phi \leq -\frac{j-i}{w_n}$. This leads to

$$E_{(\mathbf{x},\mathbf{y})}\rho(X_1, Y_1) \leq \left(1 - \tfrac{1}{w_n}\right)\rho(\mathbf{x}, \mathbf{y})$$

for adjacent $(\mathbf{x}, \mathbf{y})$. We are thus in the setting of Lemma 7, which shows

$$\bar{d}(t) \leq \Delta_n \exp(-t/w_n).$$

Since $\Delta_n = O(n^2)$ and $w_n \sim n^3/6$ we obtain

$$\tau_1^{\text{disc}} = (\tfrac{1}{3} + o(1))\, n^3 \log n.$$

2   Notes on Chapter 4-3

Coupling has become a standard tool in probability theory. The monograph of Lindvall [12] contains an extensive treatment, and history. In brief, Doeblin [5] used the idea of running two copies of the chain independently until

they meet, in order to prove the convergence theorem (Chapter 2 Theorem 2) for finite-state chains, and this is now the textbook proof ([14] Theorem 1.8.3) of the convergence theorem for countable-state chains. The first wide-ranging applications were in the context of infinite-site interacting particles in the 1970s, where (e.g. Liggett [11]) couplings were used to study uniqueness of invariant distributions and convergence thereto. Theory connecting couplings and variation distance is implicit in Griffeath [7] and Pitman [15], though the first systematic use to bound variation distance in finite-state chains was perhaps Aldous [1], where examples including those in sections 1.4, 1.7 and 1.9 were given.

Section 1.2. There may exist Markov couplings which are not of the *natural* form (4), but examples typically rely on very special symmetry properties. For the theoretically-interesting notion of (non-Markov) *maximal coupling* see Chapter 9 section 1 (yyy 4/21/95 version).

The coupling inequality is often presented using a first chain started from an arbitrary point and a second chain started with the stationary distribution, leading to a bound on $d(t)$ instead of $\bar{d}(t)$. See Chapter 13-4 yyy for an example where this is used in order to exploit distributional properties of the stationary chain.

Section 1.5. This chain was first studied by Jerrum [9], who proved rapid mixing under the weaker assumption $c \geq 2r$. His proof involved a somewhat more careful analysis of the coupling, exploiting the fact that "bad" configurations for the inequalities (5,6) are different. This problem attracted interest because the same constraint $c \geq 2r$ appears in proofs of the absence of phase transition in the zero-temperature anti-ferromagnetic Potts model in statistical physics. Proving rapid mixing under weaker hypotheses was first done by Bubley et al [4] in special settings and using computer assistance. Vigoda [17] then showed that rapid mixing still holds when $c > \frac{11}{6}r$: the proof first studies a different chain (still reversible with uniform stationary distribution) and then uses a comparison theorem.

Section 1.6. The chain here was suggested by Jerrum [8] in the context of a general question of counting the number of orbits of a permutation group acting on words. More general cases (using a subgroup of permutations instead of the whole permutation group) remain unanalyzed.

Section 1.10. See Luby and Vigoda [13] for more detailed study and references.

Section 1.11. Conceptually, the states in these examples are *unordered* families of words. In genetic algorithms for optimization one has an objective function $f : \{0,1\}^L \to R$ and accepts or rejects offspring words with

probabilities depending on their f-values.

Interesting discussion of some different approaches to genetics and computation is in Rabani et al [16].

Hint for (14). First match the L'th letters in each word, using the occasions when $U_i = L$ or $L+1$. This takes $O(LK)$ time.

Section 1.12. Another setting where path-coupling has been used is contingency tables: Dyer and Greenhill [6].

In the case where (16) holds for $\kappa = 1$, one might expect a bound of the form

$$\bar{d}(t) = O(\Delta_\rho^2/\alpha) \tag{18}$$

$$\alpha := \min_{(i,j)\in\mathcal{E}} P(\rho(X_1^{(i)}, X_1^{(j)}) \le \rho(i,j) - 1).$$

by arguing that, for arbitrary (i, k), the process $\rho(X_t^{(i)}, X_t^{(k)})$ can be compared to a mean-zero random walk with chance α of making a negative step. Formalizing this idea seems subtle. Consider three states i, j, k with $(i, j) \in \mathcal{E}$ and $(j, k) \in \mathcal{E}$. Suppose

$$P(\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) + 1) = P(\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) - 1) = \alpha$$

and otherwise $\rho(\cdot, \cdot)$ is unchanged; similarly for (j, k). The changes for the (i, j) process and for the (j, k) process will typically be dependent, and in the extreme case we might have

$$\rho(X_1^{(i)}, X_1^{(j)}) = \rho(i,j) + 1 \text{ iff } \rho(X_1^{(j)}, X_1^{(k)}) = \rho(j,k) - 1$$

and symmetrically, in which case $\rho(X_t^{(i)}, X_t^{(k)})$ might not change at all. Thus proving a result like (18) must require further assumptions.

Section 1.13. The Markov chain here (with uniform weights) was first studied by Karzanov and Khachiyan [10].

References

[1] D.J. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminaire de Probabilites XVII*, pages 243–297. Springer-Verlag, 1983. Lecture Notes in Math. 986.

[2] R. Bubley and M. Dyer. Path coupling: a technique for proving rapid mixing in Markov chains. In *Proc. 38'th IEEE Symp. Found. Comp. Sci.*, pages 223–231, 1997.

[3] R. Bubley and M. Dyer. Faster random generation of linear extensions. *Discrete Math.*, 201:81–88, 1999.

[4] R. Bubley, M. Dyer, and C. Greenhill. Beating the 2Δ bound for approximately counting colorings: A computer-assisted proof of rapid mixing. In *Proc. 9'th ACM-SIAM Symp. Discrete Algorithms*, pages 355–363, New York, 1998. ACM.

[5] W. Doeblin. Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique*, 2:77–105, 1938.

[6] M. Dyer and C. Greenhill. A genuinely polynomial-time algorithm for sampling two-rowed contingency tables. Technical report, University of Leeds, U.K., 1998. Unpublished.

[7] D. Griffeath. A maximal coupling for Markov chains. *Z. Wahrsch. Verw. Gebiete*, pages 95–106, 1974-75.

[8] M. Jerrum. Uniform sampling modulo a group of symmetries using Markov chain simulation. In J. Friedman, editor, *Expanding Graphs*, pages 37–48. A.M.S., 1993. DIMACS, volume 10.

[9] M.R. Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Struct. Alg.*, 7:157–165, 1995.

[10] A. Karzanov and L. Khachiyan. On the conductance of order Markov chains. *Order*, 8:7–15, 1991.

[11] T.M. Liggett. Coupling the simple exclusion process. *Ann. Probab.*, 4:339–356, 1976.

[12] T. Lindvall. *Lectures on the Coupling Method*. Wiley, 1992.

[13] M. Luby and E. Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. Technical report, ICSI, Berkeley CA, 1999.

[14] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[15] J.W. Pitman. On coupling of Markov chains. *Z. Wahrsch. Verw. Gebiete*, 35:313–322, 1976.

[16] Y. Rabani, Y. Rabinovich, and A. Sinclair. A computational view of population genetics. In *Proc. 36'th IEEE Symp. Found. Comp. Sci.*, pages 83–92, 1995.

[17] E. Vigoda. Improved bounds for sampling colorings. C.S. Dept., U.C. Berkeley, 1999.

[18] D.B. Wilson. Mixing times of lozenge tilings and card shuffling Markov chains. Unpublished, 1997.